2-1-2014

# Essays on Financial Market Volatility: Applications of Time-Varying Dynamics

Emily Johnston
*Graduate Center, City University of New York*

# ESSAYS ON FINANCIAL MARKET VOLATILITY: APPLICATIONS OF TIME-VARYING DYNAMICS

by

EMILY B. JOHNSTON

A dissertation submitted to the Graduate Faculty in Economics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2014

This manuscript has been read and accepted for the
Graduate Faculty in Economics in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Professor Randall K. Filer

Date — Chair of Examining Committee

Professor Merih Uctum

Date — Executive Officer

Professor Kevin R. Foster

Professor Devra L. Golbe

Professor George Vachadze

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

# Abstract

ESSAYS ON FINANCIAL MARKET VOLATILITY: APPLICATIONS OF TIME-VARYING DYNAMICS

by

Emily B. Johnston

Advisor: Professor Randall K. Filer

This dissertation examines time-variation in asset volatility surrounding periods of financial market distress. In the first chapter we give a brief introduction of the overall theme of the project, and we outline the models used. The next chapters individually focus on the application of time-varying volatility to important themes in the literature. These include: the behavior of investor risk preferences across periods of stability and distress; inconsistencies in options pricing with regard to the behavior of the underlying asset; and the characterization of time-varying volatility dynamics in equity returns.

The second chapter of this dissertation examines the impact of changing asset volatility on the estimation of investor risk preferences. We ask whether prior findings of time-varying behavior for risk preferences may be due in part to a failure to account for changes in volatility. This is an important issue, because there is evidence in the existing literature that suggests a contributing role of risk preferences during periods of crisis and contagion. We use a regime-switching GARCH model for pricing kernel estimation to show that much of the variation in estimated investor risk preferences can be explained by changing volatility instead.

In the third chapter we examine stochastic volatility as an additional uncertainty factor regarding the future state of the market. We explore whether this inclusion affects prior findings of options pricing inconsistencies in the literature. Options mispricing is an important topic in debates concerning the role of investor sentiment in market behavior and asset pricing. Results from our investigation indicate that including this additional uncertainty factor does not fully explain away the inconsistencies. Our findings thus appear to support the existing evidence of options mispricing with respect to the behavior of the underlying asset.

In the fourth and final chapter of this dissertation, we examine asset volatility dynamics over a long historical time frame from 1871-2013. We demonstrate best fit for the number of distinct volatility regimes and characterize these separate dynamics. There is growing evidence that some economic relationships themselves may change between periods of high and low volatility – understanding changing volatility dynamics is crucial for understanding these economic relationships as well. We also show in this chapter how the estimated high-volatility state matches up with well-documented historical financial market events.

# ACKNOWLEDGEMENTS

This dissertation would not be possible without the constant support and counsel of my advisor, Professor Randall K. Filer. His timely and thoughtful comments have made all the difference throughout this process, and I am forever grateful for all of his encouragement. I am also grateful to my other committee members, Professor George Vachadze, Professor Kevin R. Foster, and Professor Devra L. Golbe, who have provided me with enormously helpful feedback at numerous points along the way.

I would like to sincerely thank the Graduate Center for research and travel support provided during this writing process. Acknowledgements are also due to seminar participants at the Graduate Center for their excellent comments and suggestions. My sincere thanks go to the Economics Department itself, which has been so good to me.

I would like to thank my parents, Jerry and Leslie Johnston, for their endless support and encouragement. I would also like to thank my in-laws, Chris and Helen Ross, who have been there every step of the way as well.

Finally, I am grateful to my husband, Stephen Ross, for his levity, patience and love. This dissertation would not have been possible without his support.

# Contents

**1 Introduction** **1**

**2 Risk Preferences Under Changing Volatility Regimes: A Markov Switching Approach to Pricing Kernel Estimation** **5**

2.1 Introduction . . . . . 6

2.2 Conceptual background . . . . . 9

2.3 Methodology . . . . . 11

2.3.1 Choice of state-price density estimator . . . . . 11

2.3.2 Choice of state probability density estimator . . . . . 14

2.4 Estimation procedure . . . . . 17

2.4.1 State-price density estimation . . . . . 17

2.4.2 State probability density estimation . . . . . 21

2.5 Data . . . . . 34

2.6 Analysis . . . . . 36

2.6.1 Comparison of models . . . . . 36

2.6.2 Pricing kernel behavior . . . . . 42

2.7 Conclusion . . . . . 45

**3 Are Options Prices Empirically Consistent with the Time Series of the Underlying Asset? 48**

3.1 Introduction . . . . . 49

3.2 Conceptual background . . . . . 51

3.3 Model . . . . . 55

3.3.1 Risk-neutral underlying asset dynamics . . . . . 55

3.3.2 Risk-neutral state-price densities . . . . . 59

3.4 Estimation . . . . . 62

3.4.1 Stochastic volatility model . . . . . 62

3.4.2 State-price densities . . . . . 67

3.5 Data . . . . . 70

3.6 Analysis . . . . . 73

3.7 Conclusion . . . . . 81

**4 Characterizing Volatility Regimes in the S&P 500 Index** **83**

4.1 Introduction . . . . . 83

4.2 Conceptual background . . . . . 86

4.3 Model . . . . . 89

4.4 Estimation . . . . . 90

4.5 Analysis . . . . . 100

4.6 Conclusion . . . . . 107

**Bibliography** **109**

# List of Tables


2.1 S&P 500 log-return summary statistics . . . . . 35

2.2 Conditional variance parameter estimates . . . . . 39

2.3 Deviance information criterion, full sample . . . . . 40

2.4 Deviation information criterion, by state . . . . . 41

2.5 Posterior predictive results . . . . . 41

2.6 Pricing kernels, by state . . . . . 44

2.7 Pricing kernels, by time period . . . . . 45

3.1 Moment regression coefficients and standard errors . . . . . 79

4.1 S&P 500 log-return summary statistics . . . . . 101

4.2 Single-regime parameter estimates . . . . . 102

4.3 Two-regime parameter estimates . . . . . 102

4.4 Three-regime parameter estimates . . . . . 103

4.5 Deviance information criterion . . . . . 105

# List of Figures

2.1 S&P 500 index and log-returns . . . . . 36

2.2 Pricing kernel function estimates. . . . . 43

3.1 S&P 500 index prices, log returns, and volatility. . . . . 71

3.2 Mean scatter plot . . . . . 74

3.3 Variance scatter plot . . . . . 75

3.4 Skewness scatter plot . . . . . 76

3.5 Kurtosis scatter plot . . . . . 77

4.1 S&P 500 index log-prices and log-returns . . . . . 101

4.2 S&P 500 index and smoothed probabilities for three volatility states . . . . . 106

# Chapter 1

# Introduction

This dissertation examines time-variation in asset volatility, particularly surrounding periods of financial market distress. The implications of this research reach beyond descriptive and forecasting purposes. First, inferences made from existing models in the literature may be called into question because of a failure to address fundamental changes in volatility dynamics over time. Second, there is increasing evidence that certain economic relationships themselves may change under high and low volatility conditions. Understanding the nature of changing volatility conditions is crucial for understanding these changing economic relationships as well. The following chapters of this dissertation can be taken separately, but they are aligned in their application of changing volatility dynamics to important questions in the economic literature.

There are good reasons for allowing volatility dynamics to change over time in models of financial market behavior. As pointed out in Hamilton and Susmel (1994), large and small shocks to asset returns may have different causes, not to mention different consequences for subsequent

volatility behavior. Some shocks are transitory, while others have a longer-lasting impact. Though time-varying volatility can be commonly found in the literature, it is often applied deterministically through shocks to asset prices, which may be insufficient for capturing major changes in market conditions over time.

To see this, consider a GARCH representation of conditional variance for asset returns:

$$h_t = \alpha_0 + \alpha u_{t-1}^2 + \beta h_{t-1} \tag{1.1}$$

for $t = 1, \ldots, T$. $h_t$ is the conditional variance, which captures a degree of persistence in the asset volatility through lags to shocks and prior conditional variances. With $\alpha_0$ as a kind of baseline variance, coefficients $\alpha$ and $\beta$ describe the influence of past shocks $u_{t-1}$ to asset prices and past conditional variances $h_{t-1}$ on today's $h_t$. Thus we can interpret a kind of baseline component and a memory component contributing to the persistence of the conditional variance process.

Time variation in the GARCH process is driven through prior shocks $u_{t-1}$ and prior conditional variances $h_{t-1}$, which vary over time but work their way through a constant-parameter process. This means that the same parameters that decribe these dynamics when markets are calm are supposed to describe market dynamics in crisis as well – there is only one set of parameters estimated to describe volatility under all market conditions. This really becomes more of a weighted-average expression for volatility dynamics. Large but infrequent transitory shocks can bias estimates of volatility persistence upwards during calmer financial market conditions. At the same time, these estimates can greatly under-represent volatility under more extreme conditions.

A Markov switching GARCH formulation on the other hand specifies random hidden states $s_{[1,T]}$ that control the parameters of the underlying GARCH process. These are part of a stationary, irreducible Markov process that allows instantaneous shifts in volatility dynamics. The Markov chain $\Delta$ has $1,2,\ldots,k$ state spaces and a $k \times k$ transition matrix $\mathbf{\Pi}$:

$$\mathbf{\Pi} = [\pi_{i,j}] = [\pi(\Delta_t = j | \Delta_{t-1} = i)] \tag{1.2}$$

for $i, j = 1,2,\ldots,k$, giving us the transition probabilities $\pi_{i,j}$ of moving from state $i$ at time $t-1$ to state $j$ at time $t$. The conditional probability of the current state at $t$ only depends upon the previous state at $t-1$. Regime variances are given by:

$$\boldsymbol{h}_t = \boldsymbol{\alpha_0} + \boldsymbol{\alpha} u_{t-1}^2 + \boldsymbol{\beta} \boldsymbol{h}_{t-1} \tag{1.3}$$

where $\boldsymbol{h}_t$=$[h_{1t},\ldots,h_{kt}]$ is a $k \times 1$ vector of regime variances, $\boldsymbol{\alpha_0}$=$[\alpha_{01},\ldots,\alpha_{0k}]$, $\boldsymbol{\alpha}$=$[\alpha_1,\ldots,\alpha_k]$, and $\boldsymbol{\beta}$=$diag(\beta_1,\ldots,\beta_k)$. The formulation results in separate parameter sets that reflect volatility dynamics in the market according to each state. This clear interpretation of distinct volatility states is a major advantage of the discrete-time Markov switching GARCH model. It is an important feature in our examination of investor risk preferences in Chapter 2 and characteristic volatility behavior in Chapter 4. It also allows deeper examination of these separate volatility dynamics – we can observe structural breaks in the volatility when $\alpha_{0i} \neq \alpha_{0j}$, or whether the memory response to prior shocks may be different between regimes when $\alpha_i \neq \alpha_j$, for example.

Next consider the following continous-time diffusion process estimated in the options mispricing literature:

$$dS_t = \mu S_t dt + \sigma S_t dW_t. \tag{1.4}$$

In this formulation $\mu$ is a drift term, $\sigma$ is the volatility in asset prices, and $W_t$ represents Brownian motion. Uncertainty about the future state of the market has only a single dimension, driven by random Brownian movements in the underlying asset price $S_t$.

Yet a second form of uncertainty about the future state of the market could be achieved through variation in the volatility process itself, so that volatility $\sigma$ is stochastic in its own right. This can be done using a model of stochastic volatility along the lines of Heston (1993):

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_{1t} \tag{1.5}$$

$$dV_t = \kappa(\theta - V_t)dt + \sigma_v \sqrt{V_t} dW_{2t} \tag{1.6}$$

where $dV_t$ denotes the volatility process, $\theta$ is its long-run mean, and $\kappa$ is the mean-reversion speed. The term $\sigma_v$ is interpreted as the so-called "volatility of the volatility." Here the volatility process itself is stochastic through its own random Brownian movements, introducing a second dimension of uncertainty – $dW_{1t}$ as well as $dW_{2t}$ – to the future state of the market. We use a model of stochastic volatility in Chapter 3 of this dissertation in order to address limitations in the options mispricing literature. Our purpose is to extend volatility as an additional uncertainty factor, investigating prior findings of options mispricing which are based upon a single source of market uncertainty alone.

# Chapter 2

# Risk Preferences Under Changing Volatility Regimes: A Markov Switching Approach to Pricing Kernel Estimation

Since the 2008 meltdown in financial markets, the role of risk appetite as a factor in market vulnerability has received increasing attention. Our research contributes to the literature by using a modified approach for pricing kernel estimation to examine risk preferences in financial markets. Evidence shows that introducing regime-switching volatility into the model better reflects this period of extreme volatility, and also appears to have a significant impact on the characterization of pricing kernels and risk preference behavior. Our results indicate that conventional single-regime models may mistakenly identify changes in risk preferences by failing to account for possible regime changes in the conditional variance of returns over time.

# 2.1. Introduction

In periods of financial market distress, casual observations that risk appetite "seemingly vanishes" abound. Such perceptions of market behavior draw attention to the complicated matter of modeling investor preferences towards risk. If preferences for risk can vary with extreme market events, then the relative weights given to down- and upside-risks in markets would vary as well. Yet uncovering the true nature of risk preferences is tricky as preferences themselves are not directly observable. Instead, they are generally presented within the literature according to basic assumptions of what we think is reasonable based upon economic theory.

A number of studies have explored empirical methods for examining risk preference behavior in financial markets. Aït-Sahalia and Lo (2000) and Rosenberg and Engle (2002) were influential contributors, constructing estimates from well-studied aspects of asset pricing. Asset pricing theory suggests that investor risk preferences can be derived by constructing ratios of state-price densities to state probability densities. These densities can be estimated from observed options prices and from the prices on their underlying assets. Variations of this strategy can be seen throughout the existing literature.[1]

However, there are questions as to whether the changes captured in these ratios can truly be interpreted as preference-driven. State-price densities are typically estimated from cross sections of options data.[2] The range of existing strike prices for a single option allows the estimation of a full probability density at a given point in time. But unlike options, only a single price at any point in time exists on the underlying asset. Therefore, state probability densities are typically

[1] Bartunek and Chowdhury (1997), Jackwerth (2000), and Gai and Vause (2006), for example.
[2] See Jackwerth (2004) for further background and discussion.

estimated from an aggregated time series of underlying asset prices. In some cases this may mean a time-varying options-derived density in the numerator is compared with a series of average state probability densities in the denominator. In others, it may mean forecasting the denominator from a fixed-parameter process assumed to hold constant for long periods of time.

It is problematic that the state-price densities in the numerator would vary with time but the state probability densities in the denominator remain implicitly fixed. The danger is that all of the changes in these pricing kernel ratios may be interpreted as changes in risk preferences, when to some degree they could be picking up on changes in the underlying asset dynamics in the denominator itself. This would call into question the reliability of preference-driven interpretations for pricing kernel behavior, and has yet to be resolved in the literature.

We follow the existing literature by estimating pricing kernels as ratios of state-price densities to state probability densities. We contribute by posing the following question: Do regime-switching dynamics in the underlying asset affect common preference-driven explanations for pricing kernel behavior? By introducing high and low volatility states into the conditional variance of the underlying asset, we introduce greater sensitivity in measuring pricing kernels across periods of stability and distress. We then investigate how this affects interpretations of risk preference behavior in financial markets.

Our data consist of S&P 500 index options and the prices on the S&P 500 index itself. The sample time frame is 2005-2012, encompassing the lead-up, duration, and aftermath to the 2008 financial crisis. We estimate pricing kernels on a monthly basis in order to maintain a constant time-to-maturity for short-term options contracts. We use GARCH models to describe single-regime and

regime-switching behavior in the underlying asset. We compare the models through goodness-of-fit and posterior predictive performance, and then investigate how pricing kernel behavior differs in each case. Results show that the regime-switching model provides a better description of the sample period, and that the implications for risk preference behavior differ significantly from those of a conventional single-regime estimation strategy.

Better understanding the empirical characteristics of pricing kernels will be valuable for modeling market behavior and risk preferences. What is more, the behavior of risk preferences themselves has become an especially timely and important issue since the 2008 financial meltdown. A number of recent papers have suggested an important contributing role for investor attitudes towards risk in periods of crisis and contagion.[3] Thus, examining the empirical characteristics of investor risk preferences is becoming increasingly important for matters of economic and financial market stability.

This chapter proceeds as follows: section 2.2 presents a conceptual background for pricing kernels and their relation to risk preferences; section 2.3 outlines the basic reasoning and methodology used in our estimation of pricing kernels; section 2.4 gives further details on how the estimation is carried out; section 2.5 provides an overview of the data; section 2.6 reports parameter estimates and interprets how the models differ in their impact on pricing kernel behavior; and section 2.7 concludes.

[3] González-Hermosillo (2008) and Danielsson, Shin and Zigrand (2009), for example.

## 2.2. Conceptual background

In this paper we use pricing kernels to estimate risk preference behavior in financial markets. An advantage of a pricing kernel strategy is that it allows examination of preferences with respect to equity-return states across time. Other approaches may tend to examine cross-sectional compositions of portfolios across wealth, taking the viewpoint of a representative investor for market behavior. Pricing kernels rather allow for the examination of preferences extracted from the pricing of marginal trades in the market itself. Therefore, we emphasize throughout this paper the interpretation of a marginal investor conducting trades in the market. It may be after all that the risk preference for the marginal investor in markets behaves somewhat differently than what would be assumed as reasonable for the individual investor. Risk preference for the marginal investor might change over time because the collective preferences of individual investors have changed, for example. Alternatively, risk preference could change perhaps because the composition of investors in the market with particular levels of risk preference has changed.

Pricing kernels are suited for this investigation because of their well-established foundation in asset pricing. Pricing kernels are random variables that capture the rate at which investors are willing to substitute uncertain consumption in future periods for consumption that is certain. Along with a random future payoff $x_{t+1}$, the pricing kernel $m_{t+1}$ helps form the central pricing equation in asset pricing theory:[4]

$$p_t = E_t[m_{t+1}x_{t+1}]. \tag{2.1}$$

[4]See Cochrane (2000).

To further understand pricing kernels, consider the state-preference model of Arrow (1964) and Debreu (1959). According to this model, contingent claims may be defined as securities that make one unit of payment if a specified state $s$ is reached but in alternative states pay nothing. The subjective value investors assign to this tradeoff of uncertain, state-contingent future consumption for certain consumption is embodied in contingent claims markets through "state prices." The arbitrage-free price $p$ of the underlying security is equal to the sum of these state prices $p^*$ weighted by the security payout $x$ associated with each state $s$:

$$p = \sum_{s=1}^{S} p^*(s)x(s). \tag{2.2}$$

Time preference can be incorporated by discounting the payoffs by $1/(1+r)$, while multiplying state prices $p^*(s)$ by $(1+r)$ in order to maintain the sum to 1. Transforming then by multiplying numerator and denominator by underlying state probabilities $\phi(s)$ – or the probabilities of each state $s$ actually occurring – gives a representation of the pricing kernel within the basic pricing equation framework:

$$p = \frac{1}{1+r} \sum_{s=1}^{S} (1+r)\phi(s)m(s)x(s) \tag{2.3}$$

where $m(s) \equiv p^*(s)/\phi(s)$ is the pricing kernel. This reformulation of the central pricing equation allows us to see the pricing kernel $m(s)$ interpreted as state prices $p^*(s)$ scaled by state probabilities $\phi(s)$.

The key insight here is that state-price densities and state probability densities should be identical if preferences were not a factor in contingent claims pricing. Asset pricing theory tells us that assets which produce higher payouts in situations when wealth is lower tend to be valued

more highly. The higher the value given to payouts in unfavorable states of wealth, the lower the preference for risk, or the more risk averse the investor. This point emphasizes state prices as "preference-weighted." If investors were indifferent between payouts received in various alternative states of wealth, then these payouts should not be valued any differently than the actual likelihood of the underlying asset movements would warrant. In this case the state-price density would be identical to the state probability density, and the pricing kernel ratio would simply be equal to 1. What differentiates the state-price and state probability densities is the effect that investor risk preferences have on the valuation of payouts received in different states of wealth. Pricing kernels reflect investor risk preferences, based upon how much the preference-weighted state-price densities in the numerator differ from the actual probabilities in the denominator.

## 2.3. Methodology

Having outlined the theory behind pricing kernel ratios of state-price to state probability densities, we must now estimate these ratio components. This section briefly surveys some of the relevant literature to explain our methodological reasoning.

### 2.3.1. Choice of state-price density estimator

Estimation of state-price densities from options prices was introduced in Breeden and Litzenberger (1978). The authors show this can be done by taking the second derivative of the call option-pricing function with respect to the strike price. This requires no prior assumptions of distributional form,

and in practice can be estimated by smoothing an implied volatility function for observed options with identical maturity dates. The smoothed volatilities are used to generate smooth options prices for all possible outcomes, based on an assumed relationship between options prices and the underlying asset volatility. The resulting option-pricing function is then twice differentiated to obtain the state-price density.[5]

Extending this method nonparametrically, Aït-Sahalia and Lo (1998) estimate a kernel regression fitting realized options price data. The state-price density is then found by taking the second derivative of the resulting estimator. Their method requires no functional pre-suppositions about options prices and volatility, nor does it require assumptions of a prior distribution for the options. It does require that certain assumptions be made regarding smoothness criteria, and can also be data intensive. Small samples may be uninformative due to over-smoothing and may produce a biased estimator. In practice the authors suggest a semi-parametric adjustment, reducing the required sample size by estimating other variables in the option pricing function parametrically but reserving nonparametric estimation for the implied volatility alone. Even with this reduced adjustment, they aggregate over a yearly time frame to obtain a set of 14,431 options-price observations for use in a single density estimation. This results in what is really a succession of average rather than time-varying densities. With this in mind Aït-Sahalia and Lo (2000) acknowledge that their estimator may reflect cross-sectional observations better at some dates than at others.

Aït-Sahalia and Duarte (2003) address some of the challenges associated with nonparametric estimation strategies for state-price densities. Convergence can be slow, especially for higher-order

[5] Variations on this method can be seen in Shimko (1993), Bliss and Panigirtzoglou (2004), Glatzer and Scheicher (2005), Gai and Vause (2006), and Figlewski (2008), for example.

derivatives, so large sample sizes tend to be required. In addition, the authors point out that estimating the pricing function with options across time in order to achieve large samples introduces potential nonstationarity or regime changes into the data, and other model variables may not remain stable over this same time period. The authors propose a modification imposing shape restrictions on the first and second derivatives of the nonparametrically-estimated pricing function to ensure monotonicity and convexity with respect to the strike price. They use a method of constrained least squares regression and locally polynomial kernel smoothing, achieving estimators that satisfy the prescribed constraints not just asymptotically but also within sample.

In application, Aït-Sahalia and Duarte's (2003) estimator displays plausible excess skewness and kurtosis in the resulting state-price densities. It also satisfies constraints for non-negativity, a requirement sometimes violated through other approaches. Simulations show that their constrained method produces better root-integrated-mean-squared errors compared with other unconstrained estimators. Also, favorable small sample properties make this well-suited for the typical small size of options price cross sections. Hence the authors are able to estimate state-price densities on a more sensitive daily basis as opposed to in larger, time-aggregated average segments.

Yatchew and Härdle (2006) similarly find promising results using a constrained nonparametric estimator. They draw further attention to fitting not just the smoothness of the "true" option-pricing function, but also its first and second derivatives. The authors show that a smooth point-mean function for example may appear a good fit with the data points themselves, but that it tends to gyrate wildly in its first and second derivatives around those of a sample data-generating process. This has worrying implications since it is the second derivative in particular that yields the state-

price density. In order to take the fit of higher-order derivatives into account in their estimation, they incorporate restrictions on functional shape similar to those used in Aït-Sahalia and Duarte (2003). Simulations indicate that the shape-constrained estimator is able to greatly improve capture of the sample process and its derivatives.

With these considerations in mind, we use a shape-constrained least squares estimator for estimating the state-price density along the lines of Aït-Sahalia and Duarte (2003) and Yatchew and Härdle (2006). This way the second derivative of the option-pricing function can be appropriately fitted, not just the smoothed observations. Also, favorable small-sample properties will ensure that sensitivity is not averaged out due to large scale aggregation of options across time. In the estimation procedure presented later in section 2.4.1, we outline specific details for how this estimator is constructed.

### 2.3.2. Choice of state probability density estimator

The denominator of the pricing kernel ratio is estimated from state probabilities, reflecting expected movements in the underlying asset. Options-derived state prices are forward-looking, and their associated state probabilities should be forward-looking as well. This poses a significant modeling problem since the inner expectations of market participants concerning future returns are not observable. Much of the related literature deals with this problem by using historical price observations to estimate model parameters, and then forecasting into the future with random disturbances. It is not necessary that the forecasted returns precisely match future realized returns, just that the forecasts reasonably capture expectations at the time. Under assumptions of rational

expectations and market efficiency, these market expectations should be unbiased predictors of future market behavior.

A common strategy in the literature uses an asymmetric GARCH model based upon the GJR specification of Glosten, Jagannathan and Runkle (1993) to estimate state probability densities.[6] The conditional variance $h_t$ can be represented by:

$$h_t = \alpha_0 + \alpha(|u_{t-1}| - \gamma u_{t-1})^2 + \beta h_{t-1} \tag{2.4}$$

with $u_{t-1}$ as prior shocks to asset prices and $h_{t-1}$ prior conditional variances. Together with an estimated mean process, Monte Carlo simulations of S&P 500 index behavior can then be used to calculate the predicted densities.

Originally developed in Bollerslev (1986) as a generalization of Engle's (1982) autoregressive conditional heteroskedasticity model, GARCH models treat conditional variance as a function of past errors. In effect they allow the underlying variance to change stochastically over time through shocks to asset prices rather than assume that the variance remains constant. This is appealing as error terms in financial time series often appear larger in some periods than in others, and with patterns of autocorrelated clustering. Modified GARCH processes in the literature have been able to improve the capture of other observed characteristics of financial time series as well. For example, the $\gamma$ parameter in the GJR specification above captures the observed tendency of volatility to be disproportionately large when shocks are negative. Elsewhere Barone-Adesi, Engle and Mancini

[6] Variations on this strategy are common throughout the literature, as in Scheicher (2003), Tarashev, Tsatsaronis and Karampatos (2003), Gai and Vause (2006), and Tarashev and Tsatsaronis (2006).

(2008) incorporate innovations drawn from an empirical density function, accommodating characteristics of non-normality in equity returns for state probability density estimation.

An advantage of GARCH modeling for state probability densities is that it does not require imposition of a pre-specified functional form on the pricing kernel itself. However as Bliss and Panigirtzoglou (2004) point out, there is nevertheless an implicitly restricting assumption of stationarity in the parameters of the underlying stochastic GARCH process. Though key variables may vary with time, the parameters of the underlying stochastic process are constant. The authors question the effectiveness of a model that assumes state-price densities to be time-varying but the state probability densities implicitly not. In this case all of the variation in the pricing kernel ratio may be attributed to changes in risk preferences, when it could in part be driven by a failure to account for changes in the underlying asset process in the denominator itself.

Additionally this means that the higher moments of the process are held constant, yet there is no reason to necessarily assume that these would not vary over time as well. As an alternative, Bliss and Panigirtzoglou (2004) impose restrictions rather on the utility function underlying the pricing kernel, thereby allowing the statistical distribution in the denominator to time-vary. But as a tradeoff they must make strong functional assumptions regarding the pricing kernel form. These assumptions necessarily impose a constant coefficient of risk aversion, which would conceal time-varying behavior in investor preferences towards risk.

We argue that conventional GARCH models for state probability density estimation may be overly restrictive, and behaviorally less realistic than a regime-switching approach. We share Bliss and Panigirtzoglou's (2004) contention with the stationary higher moments in conventional esti-

mates, especially when state-price densities are widely accepted as variable over time. Considering market behavior, this would imply that market participants keep constant their perceived probabilities of relative price movements, regardless of recent experience or circumstances. Allowing variability in the higher moments of the state probability densities on the other hand could be very informative. For example, variation in state probability skewness would suggest that the perceived probability of downward movements in the market relative to upward ones may have changed. Variation in kurtosis with longer, fatter tails would imply that extreme price movements may be perceived as more likely than before. Thus it is important that considerations be made for variation in the underlying state probability densities as well. We explain our proposed regime-switching modifications in more detail in the estimation procedure in section 2.4.2 below.

## 2.4. Estimation procedure

### 2.4.1. State-price density estimation

#### Least squares estimator

The nonparametric locally polynomial estimator for state-price densities developed in Aït-Sahalia and Duarte (2003) is obtained through a combination of constrained least squares regression and smoothing. The idea is to find $z_1$,$z_2$,...,$z_N$, the least squares values closest to an observation set of

call option prices $c_1$,$c_2$,...,$c_N$, while satisfying specified restrictions. That is, vector **z** is solved for vector **c** such that:

$$\min_{z\in\Re}\sum_{i=1}^{N}(z_i-c_i)^2=\min_{z\in\Re}\|\boldsymbol{z}-\boldsymbol{c}\|^2 \tag{2.5}$$

subject to restrictions on slope and convexity.

**Restrictions on estimator**

Restrictions on Aït-Sahalia and Duarte's (2003) estimator are motivated by theory - that call option-pricing functions are decreasing and convex with respect to the strike price. These restrictions rule out arbitrage and guarantee that the density will be positive and integrate to 1. They are enforced via inequality constraints on the first two derivatives of the option-pricing function.

The call option-pricing function itself is given by:

$$C(S_t,X,\tau,r_{t,\tau},\delta_{t,\tau})=e^{-r_{t,\tau}t}\int_0^{+\infty}\max(S_T-X,0)p^*(S_T|S_t,\tau,r_{t,\tau},\delta_{t,\tau})dS_T \tag{2.6}$$

where $S_t$ is the price of the underlying asset at time $t$, $X$ is the strike price, $\tau$ is the time-to-expiration such that $T=t+\tau$, $r_{t,\tau}$ is the risk-free interest rate for that maturity, $\delta_{t,\tau}$ is the dividend yield on that asset, and $p^*(S_T|S_t,\tau,r_{t,\tau},\delta_{t,\tau})$ is the state-price density assigning probabilities to various conditional outcomes of the asset price at maturity. For risk-free interest rates $r_{t,\tau}$ we use historical Treasury bill rates.[7]

[7]H.15 historical selected interest rates from www.federalreserve.gov.

The first derivative of the call option-pricing function is:

$$\frac{\partial C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial X} = -e^{-r_{t,\tau}t} \int_X^{+\infty} p^*(S_T | S_t, \tau, r_{t,\tau}, \delta_{t,\tau}) dS_T \tag{2.7}$$

which satisfies the inequality:

$$0 \geq \frac{\partial C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial X} \geq -e^{-r_{t,\tau}t}. \tag{2.8}$$

This shows the first constraint for a monotone decreasing option-pricing function, where the first derivative is negative with respect to the strike price.

The second derivative of the call option-pricing function is:

$$\frac{\partial^2 C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial^2 X} = e^{-r_{t,\tau}t} p^*(X) \geq 0. \tag{2.9}$$

This shows the second constraint for a convex option-pricing function, where the second derivative is positive with respect to the strike price. It is also possible to see here Breeden and Litzenberger's (1978) finding that the state-price density $p^*$ can be obtained from the second derivative of the call option-pricing function with respect to the strike price. Normalization is then required, multiplying the second derivative by $e^{r_{t,\tau}t}$.

**Local polynomial smoothing**

Aït-Sahalia and Duarte (2003) show transformed data points $z_i$ for call option prices are then used in a smoothing step.[8] The idea is to estimate a smooth options-pricing function from transformed call prices, $\boldsymbol{z}$. The authors assume $z(X)$ as a function of the strike price is continuous as a second-order derivative. Then according to local Taylor approximation methodology, $z(X)$ can be estimated locally as a polynomial of order $p$, with $d$ denoting the order of the derivative:

$$\begin{aligned} z(X) &\approx z(X_0) + z'(X_0)(X - X_0) + \ldots + \frac{z^p(X_0)(X - X_0)^p}{p!} \\ &= \sum_{d=0}^{p} \frac{z^{(d)}(X)}{d!}(X_i - X)^d = \sum_{d=0}^{p} \beta_d(X)(X_i - X)^d \end{aligned} \tag{2.10}$$

for strike prices $X_i$ in the neighborhood of a fixed point $X$, where $\beta_d(X) = \frac{z^{(d)}(X)}{d!}$.

This can be used to set up a weighted least squares minimization of the following:

$$\min_{\beta} \sum_{i=1}^{N} (z_i - \sum_{d=0}^{p} \beta_d(X)(X_i - X)^d)^2 K_h(X_i - X) \tag{2.11}$$

with respect to $\beta_d(X)$ to obtain the locally polynomial estimate $\hat{\beta}_d(X)$. Essentially this is a generalized least squares regression of $z_i$ on powers of $(X_i - X)$. The expression is asymmetrically weighted with kernel function $K_h(X_i - X)$ to give greater importance to strike prices that lie within closer proximity to $X$. For weights we specify the squared inverse of the distance between the index level and strike prices.

[8] We use assistance from the *KernSmooth* package in R. The optimal bandwidth $h$ for the smoothing kernel is chosen automatically based upon the direct plug-in method described in Ruppert, Sheather, and Wand (1995).

The procedure can also be written in basic weighted least squares notation as:

$$\min_{\boldsymbol{\beta}} (\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{W} (\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \tag{2.12}$$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{y}$$

where $\mathbf{X} = \begin{pmatrix} 1 & X_1 - X & \cdots & (X_1 - X)^p \\ 1 & X_2 - X & \cdots & (X_2 - X)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_N - X & \cdots & (X_N - X)^p \end{pmatrix}$, $\mathbf{W}$=$diag\{K_h(X_1 - X), K_h(X_2 - X), \ldots, K_h(X_N - X)\}$, and $\mathbf{y}$=$[z_1, z_2, \ldots, z_N]$.

Estimated coefficients $\hat{\beta}_d(X)$ are then used to find the option-pricing function $z(X)$ and its derivatives, using:

$$\hat{z}^{(d)}(X) = d! \hat{\beta}_d(X). \tag{2.13}$$

From Fan and Gijbels (1996), derivatives of order $d$ can be found using local polynomials of optimal order $p = d + 1$. Because it is the second derivative of the option-pricing function that gives us the state-price density, the polynomial should be estimated up to order 3.

## 2.4.2. State probability density estimation

### Specification 1: Single-regime GJR-GARCH

GARCH models are commonly used to accommodate heteroskedastic characteristics in financial time series. These models assume that error term $u_t$ can be decomposed into $u_t = \varepsilon_t h_t^{1/2}$, where

$\varepsilon_t$ are i.i.d. random variables with zero-mean and unit variance. Then $u_t$ guarantees a white noise property through its reliance on $\varepsilon_t$, and it is distributed conditional on the information set $\mathbf{\Psi}$ at $t-1$ as $u_t|\Psi_{t-1} \sim N(0,h_t)$.

The GJR-GARCH specification of Glosten, Jagannathan and Runkle (1993) introduces asymmetry into the basic GARCH formulation. It is a popular form for estimating state probability densities in the literature, so it will be taken as our baseline for comparison. To capture fatter tails in asset behavior, we incorporate error terms with a Student-*t* distribution. Assumptions of *t*-distributed innovations are widely known to have advantages over the normal distribution in financial modeling, particularly in the presence of outlier observations. GARCH models with *t*-distributed innovations have an additional latent variable, $\lambda_t$ in the model outlined below, which provides an extra source of flexibility for responding to extreme observations. This means that the conditional variance is not as volatile with *t*-distributed innovations as it is with normally distributed innovations. It also increases the stability of volatility states in our later regime-switching model.

The GJR-GARCH model with *t*-distributed innovations can be described as follows:

$$\begin{aligned} y_t &= a_0 + a y_{t-1} + u_t, \quad u_t = \varepsilon_t (\lambda_t h_t)^{1/2} \\ u_t &\sim t(\nu), \quad \varepsilon_t \sim N(0,1), \quad \lambda_t \sim \mathscr{IG}(\frac{\nu}{2}, \frac{2}{\nu}) \\ h_t &= \alpha_0 + \alpha(|u_{t-1}| - \gamma u_{t-1})^2 + \beta h_{t-1} \end{aligned} \tag{2.14}$$

for $t = 1, 2, \ldots, T$. Here $y_t$ is a scalar dependent variable representing the series of log-returns on the underlying S&P 500 index, $y_{t-1}$ is its lag, $a_0$ is a constant term, and $a$ is the regression coefficient.

We attempt to fit various ARMA formulations to the demeaned $y_t$ process for our sample but find the estimated coefficient high density posterior intervals include zero. We therefore simplify by assuming $y_t = u_t$. For the $t$ distribution, $\lambda_t$ is a scaling factor and $\nu$ is the degrees of freedom parameter characterizing the shape of the distribution. $\lambda_t$ is chosen with an inverse-gamma prior distribution because, as pointed out in Jacquier, Polson and Rossi (2004), the $t$ distribution can be interpreted as a mixture of normal and inverse-gamma distributions.

In our estimation, we follow Jacquier, Polson and Rossi (2004) and Henneke, Rachev, Fabozzi and Nikolov (2011) who conveniently rescale $y_t^* = y_t/\sqrt{\lambda_t}$ to be able to sample from normal proposal distributions later in the Metropolis-Hastings simulations. $h_t$ is the conditional variance, where it is assumed $\alpha_0 > 0$, $\alpha$,$\beta \geq 0$, and $-1 < \gamma < 1$ in order to ensure that the conditional variance is positive. We also restrict $\alpha(1-\gamma^2)+\beta < 1$ so that the process is bounded and does not explode.[9] Parameter $\gamma$ introduces a leverage effect into the model, taking into account the observation in financial markets that negative shocks may tend to result in higher volatility than positive shocks. We can see that when $\gamma$ is positive, the conditional variance $h_t$ will have a larger response when $u_{t-1}$ is negative. If the leverage effect is not significant the equation simply reverts back to a classic GARCH model. With $\alpha_0$ as a kind of baseline variance, coefficients $\alpha$ and $\beta$ describe the influence of past shocks and past variance on today's variance. So we can interpret a baseline component and a memory component to the conditional variance process. In a single-regime GARCH model, the coefficient impacts of these components are fixed, just with time-varying new shocks filtering through a constant-parameter process.

[9] See Peters (2001).

The criticism from Bliss and Panigirtzoglou (2004) stems from this implicit stationarity in the parameters of the underlying stochastic process when estimating state probability densities. The authors point out conceptual difficulties in having a pricing kernel ratio where the state-price density is free to vary over time, but the state probability density not. In this case estimated changes in the pricing kernel ratio may be attributed in full to preferences, when it may really in part be picking up time variation in the underlying process of the state probability density. We explore this issue by comparing pricing kernels incorporating this common GJR-GARCH formulation with the regime-switching approach in our second specification described below.

We estimate the GJR-GARCH model according to Bayesian methods. Though maximum likelihood would also be appropriate here, a Bayesian approach is chosen to accommodate later difficulties integrating a likelihood function over multiple states in the regime-switching model. In Bayesian terms the optimal estimator is the posterior mean:

$$\hat{\boldsymbol{\theta}} = E[\boldsymbol{\theta}|\boldsymbol{Y} = \boldsymbol{y}] = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} \tag{2.15}$$

where $\boldsymbol{\theta}$ denotes the model parameters $\{\alpha_0,\alpha,\gamma,\beta\}$ and $p(\boldsymbol{\theta}|\boldsymbol{y})$ is the posterior density. The posterior density itself is determined by the likelihood $f(\boldsymbol{y}|\boldsymbol{\theta})$ for observed data and the prior density $p(\boldsymbol{\theta})$:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{\int f(\boldsymbol{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}}. \tag{2.16}$$

The likelihood function is defined as:

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{y}) \propto (\det \boldsymbol{\Sigma})^{-1} \exp\left[-\frac{1}{2} \boldsymbol{y}' \boldsymbol{\Sigma}^{-1} \boldsymbol{y}\right] \tag{2.17}$$

with $\boldsymbol{y}$ as the zero-mean vector of log-returns and $\boldsymbol{\Sigma}$ a $T \times T$ diagonal matrix of conditional variances $\{h_t(\theta)\}_{t=1}^{T}$. Priors on the parameters are assumed to be independent with truncated normal densities:

$$
\begin{aligned}
p(\alpha_0) &\propto N(\mu_{\alpha_0}, \sigma^2_{\alpha_0})\mathbb{I}\{\alpha_0 > 0\} \\
p(\alpha) &\propto N(\mu_{\alpha}, \sigma^2_{\alpha})\mathbb{I}\{\alpha \geq 0\} \\
p(\gamma) &\propto N(\mu_{\gamma}, \sigma^2_{\gamma})\mathbb{I}\{-1 < \gamma < 1\} \\
p(\beta) &\propto N(\mu_{\beta}, \sigma^2_{\beta})\mathbb{I}\{\beta \geq 0\}
\end{aligned}
\tag{2.18}
$$

where $\mathbb{I}\{\cdot\}$ denotes indicator functions that equal 1 if the specification in brackets is satisfied and 0 otherwise.

Mechanically speaking, we simulate the joint posterior density using a Metropolis-Hastings algorithm. The Metropolis-Hastings algorithm is appropriate in cases such as this when the conditional distribution of parameter estimates is unknown. The idea is to find an appropriate "transition kernel" - a conditional distribution function that tells us the probability of moving from one observation, $x$, to another, $y$ - from a target density generating the desired sample. In practice we start with a candidate-generating proposal density $q(x,y)$, and when the process is at point $x$, point $y$ is generated from this density.[10] If $q(x,y)$ satisfies certain conditions the move is accepted as a value from the target distribution, but if not the process reverts back to $x$ for the next draw. Eventually this process should converge to the desired target density. The acceptance ratio for evaluating each draw is constructed as: $\boldsymbol{\alpha}_{MH} = \min\{\frac{p(\boldsymbol{\theta}^*|\boldsymbol{y})}{p(\boldsymbol{\theta}|\boldsymbol{y})}\frac{q(\boldsymbol{\theta})}{q(\boldsymbol{\theta}^*)}, 1\}$, where the first fraction is the posterior proba-

[10] We follow Ardia's (2009) influence in derivation of the proposal densities.

bility ratio, the second fraction is the ratio of proposal densities, $\boldsymbol{\theta}^*$ is the parameter vector updated with proposed candidate draws, and $\boldsymbol{\theta}$ is the parameter vector from the previous cycle.

**Specification 2: Markov switching GJR-GARCH**

A Markov switching GARCH model specifies random hidden states $s_{[1,T]}$ that control the parameters of the underlying GARCH process. These are assumed part of a stationary, irreducible Markov process that allows instantaneous shifts in volatility. In the formulation used here, error terms follow $u_t = \varepsilon_t(\lambda_t h_{\Delta_t,t})^{1/2}$, where $\Delta_t$ is a Markov chain with $1,2,\ldots,k$ state spaces and a $k \times k$ transition matrix:

$$\boldsymbol{\Pi} = [\pi_{i,j}] = [\pi(\Delta_t = j|\Delta_{t-1} = i)] \tag{2.19}$$

for $i, j = 1,2,\ldots,k$, representing the probabilities of transitioning from state $i$ at time $t-1$ to state $j$ at time $t$. The conditional probability of the current state at $t$ only depends upon the previous state at $t-1$. Regime variances are given by:

$$\boldsymbol{h}_t = \boldsymbol{\alpha_0} + \boldsymbol{\alpha}(|u_{t-1}| - \boldsymbol{\gamma}u_{t-1})^2 + \boldsymbol{\beta}\boldsymbol{h}_{t-1} \tag{2.20}$$

where $\boldsymbol{h}_t$=$[h_{1t},\ldots,h_{kt}]$ is a $k \times 1$ vector of regime variances, $\boldsymbol{\alpha_0}$=$[\alpha_{01},\ldots,\alpha_{0k}]$, $\boldsymbol{\alpha}$=$[\alpha_1,\ldots,\alpha_k]$, $\boldsymbol{\gamma}$=$[\gamma_1,\ldots,\gamma_k]$, and $\boldsymbol{\beta}$=$diag(\beta_1,\ldots,\beta_k)$. The expression allows examination of structural breaks in the volatility dynamics of the process when $\alpha_{0i} \neq \alpha_{0j}$, or whether the asymmetric response may be different between regimes when $\gamma_i \neq \gamma_j$, for example.

Because of the difficulties integrating and evaluating a likelihood function over multiple states using maximum likelihood methods, we employ a Bayesian approach to estimate these parameters. The necessary parameters of the model are computed then from the posterior means of the parameters' simulated stationary distributions. Following Henneke, Rachev, Fabozzi and Nikolov (2011), we proceed in our estimation with the following steps:

1) Draw transition probability $\pi_{i,j}$ of moving from the current state observation to the next.

2) Sample new state $s_t$ based upon the transition probability and the likelihood for current observation $y_t$.

3) Sample latent variable $\lambda_t$ and degrees of freedom parameter $\nu$ for $t$-distributed errors.

4) Sample GARCH parameters sequentially, conditional upon state.

Then we return to step 1 to draw the transition probabilities for the new observation and repeat throughout the remaining sample, updating parameters all along. These individual steps are explained more fully in the exposition below.

*Step 1: Draw transition probabilities*

Transition probabilities are drawn from a Dirichlet distribution, which is a generalization of the Beta distribution. To see why it is a natural choice for this context, consider the two-state example given in Henneke, Rachev, Fabozzi and Nikolov (2011). The authors write the posterior distribution of $\pi_{i,i}$, the probability of remaining in state $i$ from one observation to the next, as:

$$p(\pi_{i,i}|\boldsymbol{y},\boldsymbol{s},\boldsymbol{\theta}) \propto p(\pi_{i,i}) f(\boldsymbol{y},\boldsymbol{s}|\boldsymbol{\theta}). \tag{2.21}$$

This shows the posterior conditional probability of $\pi_{i,i}$ as proportional to its prior $p(\pi_{i,i})$ times the likelihood $f(\boldsymbol{y},\boldsymbol{s}|\boldsymbol{\theta})$ for the observed innovations $\boldsymbol{y}$ and states $\boldsymbol{s}$, given the estimated parameters $\boldsymbol{\theta}$. At this stage, the existing state $s$ is carried over from the prior observation made at time $t-1$; it can therefore be considered independent of the current observation for $y$. Then the likelihood portion $f(\boldsymbol{y},\boldsymbol{s}|\boldsymbol{\theta})$ in this expression can be further simplified to obtain:

$$p(\pi_{i,i}|\boldsymbol{y},\boldsymbol{s},\boldsymbol{\theta}) \propto p(\pi_{i,i}) f(\boldsymbol{s}|\boldsymbol{\theta}). \tag{2.22}$$

This second term $f(\boldsymbol{s}|\boldsymbol{\theta})$ is driven by the cumulated transitions over the sample from one period to the next:

$$\begin{aligned} f(\boldsymbol{s}|\boldsymbol{\theta}) = \prod_{t=1}^{T} p(s_t|s_{t-1},\boldsymbol{\theta}) &= (\pi_{i,i})^{\eta_{i,i}} (\pi_{i,j})^{\eta_{i,j}} (\pi_{j,j})^{\eta_{j,j}} (\pi_{j,i})^{\eta_{j,i}} \\ &= (\pi_{i,i})^{\eta_{i,i}} (1-\pi_{i,i})^{\eta_{i,j}} (\pi_{j,j})^{\eta_{j,j}} (1-\pi_{j,j})^{\eta_{j,i}} \end{aligned} \tag{2.23}$$

where $\eta_{i,j}$ are the cumulative transitions from state $i$ to state $j$. This makes the Beta density a natural choice for the conjugate prior in each row of the transition probability matrix: $(\pi_{i,i})^{((H_{i,i}+\eta_{i,i})-1)}(1-\pi_{i,i})^{((H_{i,j}+\eta_{i,j})-1)}$ and $(\pi_{j,j})^{((H_{j,j}+\eta_{j,j})-1)}(1-\pi_{j,j})^{((H_{j,i}+\eta_{j,i})-1)}$. It can be generalized to a Dirichlet distribution to accommodate cases of more than two states. It motivates sampling of matrix rows $\boldsymbol{\pi_i} = [\pi_{i,i} \;\; \pi_{i,j}]$ and $\boldsymbol{\pi_j} = [\pi_{j,i} \;\; \pi_{j,j}]$ as follows:

$$\begin{aligned} \boldsymbol{\pi_i}|s_{[1,T]} &\sim Dirichlet(H_{i,i}+\eta_{i,i}, H_{i,j}+\eta_{i,j}) \\ \boldsymbol{\pi_j}|s_{[1,T]} &\sim Dirichlet(H_{j,j}+\eta_{j,j}, H_{j,i}+\eta_{j,i}). \end{aligned} \tag{2.24}$$

At each new state observation, parameters $\eta_{i,j}$ are updated to reflect the cumulated history of transitions. As in Ardia (2009), we initiate hyperparameters at $H_{i,i}, H_{j,j} = 2$ and $H_{i,j}, H_{j,i=}1$ with the idea that observations are more likely to remain in the prevailing state than to switch into a new state.

*Step 2: Sample current state*

Chib (1996)[11] shows that a typical point in the joint density for states $p(\boldsymbol{s}|\boldsymbol{y},\boldsymbol{\theta}) = p(s_T|\boldsymbol{y},\boldsymbol{\theta}) \cdot \ldots \cdot p(s_t|\boldsymbol{y},S^{t+1},\boldsymbol{\theta}) \cdot \ldots \cdot p(s_1|\boldsymbol{y},S^2,\boldsymbol{\theta})$ can be expressed in terms of the probability mass function, the likelihood, and the state transition probabilities:

$$\begin{aligned} p(s_t|\boldsymbol{y},S^{t+1},\boldsymbol{\theta}) &\propto p(s_t|Y_t,\boldsymbol{\theta}) f(Y^{t+1},S^{t+1}|Y_t,s_t,\boldsymbol{\theta}) \\ &\propto p(s_t|Y_t,\boldsymbol{\theta}) p(s_{t+1}|s_t,\boldsymbol{\theta}) f(Y^{t+1},S^{t+2}|Y_t,s_t,s_{t+1},\boldsymbol{\theta}). \end{aligned} \tag{2.25}$$

Here $p(s_t|Y_t,\boldsymbol{\theta})$ is the probability mass function in the current state given information up to time $t$, and $p(s_{t+1}|s_t,\boldsymbol{\theta})$ is the probability of the new state at $t+1$ given the prior state at $t$. Since future observations of $y$ and states at $t+2$ and beyond are independent of the state at time $t$, the likelihood term $f(Y^{t+1},S^{t+2}|Y_t,s_t,s_{t+1},\boldsymbol{\theta})$ can be dropped above to obtain:

$$p(s_t|\boldsymbol{y},S^{t+1},\boldsymbol{\theta}) \propto p(s_t|Y_t,\boldsymbol{\theta}) p(s_{t+1}|s_t,\boldsymbol{\theta}). \tag{2.26}$$

[11] Remaining consistent with Chib's notation, capital letters with subscripts denote history up to indicated period, and capital letters with superscripts denote future through the remaining sample. For example, $S_t$ is the history of states to time $t$, and $S^{t+1}$ is the future procession of states through final observation $T$.

Chib (1996) further breaks down the probability mass function into its prior and likelihood, using Bayes theorem and the law of total probability:[12,13]

$$\begin{aligned} p(s_t|Y_t,\boldsymbol{\theta}) &\propto p(s_t|Y_{t-1},\boldsymbol{\theta})f(y_t|Y_{t-1},\boldsymbol{\theta}) \\ &\propto [\sum_{i=1}^{k} p(s_t|Y_{t-1},s_{t-1}=i,\boldsymbol{\theta})p(s_{t-1}=i|Y_{t-1},\boldsymbol{\theta})]f(y_t|Y_{t-1},\boldsymbol{\theta}) \\ &\propto [\sum_{i=1}^{k} p(s_t|s_{t-1}=i,\boldsymbol{\theta})p(s_{t-1}=i|Y_{t-1},\boldsymbol{\theta})]f(y_t|Y_{t-1},\boldsymbol{\theta}). \end{aligned} \tag{2.27}$$

This is the transition probability $p(s_t|s_{t-1}=i,\boldsymbol{\theta})$ multiplied by the probability mass function from the prior observation $p(s_{t-1}=i|Y_{t-1},\boldsymbol{\theta})$, multiplied by the likelihood $f(y_t|Y_{t-1},\boldsymbol{\theta})$ of the current observation $y_t$, all given our parameter estimates made from the last observation at $t-1$. Current state $s_t$ is then sampled using this updated probability mass function $p(s_t|Y_t,\boldsymbol{\theta})$.

*Step 3: Sample* $\lambda_t$ *and* $\nu$

In this step, we follow the work of Jacquier, Polson and Rossi (2004) and Henneke, Rachev, Fabozzi and Nikolov (2011). For a prior on $\lambda_t$, the authors use a conjugate inverse-gamma density. This takes the form:

$$p(\lambda_t|\nu) \propto \lambda_t^{-\nu/2-1} e^{-\nu/2\lambda_t} \sim \mathscr{IG}(\frac{\nu}{2},\frac{2}{\nu}) \sim \nu/\chi^2(\nu). \tag{2.28}$$

[12]Bayes' theorem states that the conditional probability of $y$ given $x$ is equal to the conditional probability of $x$ given $y$ times the marginal probability of $y$ and divided by the marginal probability of $x$, or $P(y|x) = \frac{P(x|y)P(y)}{P(x)}$.

[13]The law of total probability states that the marginal probability of $y$ is equal to the weighted average of the conditional probabilities of $y$ given $x_i$ times the marginal probability of $x_i$, or $P(y) = \sum_i^n P(y|x_i)P(x_i)$.

The conditional distribution of $\lambda_t$ is given by:

$$p(\boldsymbol{\lambda}|\boldsymbol{y},\boldsymbol{h},\boldsymbol{\theta},\nu) = \prod_{t=1}^{T} p(\lambda_t|y_t,h_t,\nu) \tag{2.29}$$

and can be re-written as:

$$p(\lambda_t|y_t,h_t,\nu) \equiv p(\lambda_t|\frac{y_t}{\sqrt{h_t}},\nu) \propto p(\lambda_t|\nu)p(\frac{y_t}{\sqrt{h_t}}|\lambda_t,\nu). \tag{2.30}$$

The conditional posterior becomes:

$$p(\lambda_t|h_t,y_t,\nu) \propto \lambda_t^{-(\nu+1)/2-1} e^{-(y_t^2/h_t+\nu)/2\lambda_t} \sim \mathscr{IG}(\frac{\nu+1}{2}, \frac{2}{\nu+y_t^2/h_t}). \tag{2.31}$$

Jacquier, Polson and Rossi (2004) then sample $\nu$ from this distribution. Alternatively we follow Henneke, Rachev, Fabozzi and Nikolov (2011) who make use of the relationship between inverse-gamma and chi-squared distributions. They show $(\frac{y_t^2}{h_t}+\nu)(\frac{1}{\lambda_t}) \sim \chi^2(\nu+1)$, sampling $x_t \sim \chi^2(\nu+1)$ and calculating $\lambda_t = (\frac{y_t^2}{h_t}+\nu)/x_t$.

The posterior distrbution of $\nu$ is proportional to the product of *t*-distributional ordinates:

$$p(\nu|\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{y}) \propto p(\nu)p(\boldsymbol{y}|\boldsymbol{h},\nu) = p(\nu)\prod_{t=1}^{T} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}(1+\frac{y_t^2}{h_t\nu})^{-\frac{(\nu+1)}{2}}. \tag{2.32}$$

Using a flat prior and restricting $\nu$ within the discrete range $\{3,...,40\}$, the posterior distribution of $\nu$ can then be sampled from:

$$p(\nu|\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{y}) = \frac{1}{37}\frac{\prod_{t=1}^{T}\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}(1+\frac{y_t^2}{h_t\nu})^{-\frac{(\nu+1)}{2}}}{\prod_{t=1}^{T}\sum_{\nu=3}^{40}\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}(1+\frac{y_t^2}{h_t\nu})^{-\frac{(\nu+1)}{2}}}. \tag{2.33}$$

*Step 4: Sample conditional variance parameters given current state*

We employ a technique known as "data augmentation" seen elsewhere in the literature for latent class models. In this method, state observations $s_{[1,T]}$ are included as a block to be estimated within the parameter set $\boldsymbol{\theta} = \{\alpha_0, \alpha, \gamma, \beta, \boldsymbol{s}\}$. Conventional sampling methods draw from each parameter's distribution, conditional on the current values of the data and all remaining parameters. However the full conditional distribution of the state $p(s_t|\boldsymbol{y},\boldsymbol{s}_{j\neq t},\boldsymbol{\theta})$ may not be tractable across a multiple-state space because it requires knowledge of the entire sequence of past as well as future states. The data augmentation method instead allows each state to be drawn from the joint distribution of states $p(s_1, s_2, \ldots, s_T|\boldsymbol{y},\boldsymbol{\theta})$ rather than the full conditional distribution. Crucially, this means that states can be simulated with knowledge only of the prior state and transition probabilities into the next state, without necessarily knowing the entire sequence of past and future states. Furthermore,

the states can be treated collectively as a single block so that only one additional block is required in the sampler. The sampler then looks something like this, starting with $j = 1$:

$$
\begin{aligned}
s_j &\sim p(\boldsymbol{s}|\boldsymbol{y}, \alpha_0^{[j-1]}, \alpha^{[j-1]}, \gamma^{[j-1]}, \beta^{[j-1]}) \\
\alpha_0^{[j]} &\sim p(\alpha_0|\boldsymbol{y}, s_j, \alpha^{[j-1]}, \gamma^{[j-1]}, \beta^{[j-1]}) \\
\alpha^{[j]} &\sim p(\alpha|\boldsymbol{y}, s_j, \alpha_0^{[j]}, \gamma^{[j-1]}, \beta^{[j-1]}) \\
\gamma^{[j]} &\sim p(\gamma|\boldsymbol{y}, s_j, \alpha_0^{[j]}, \alpha^{[j]}, \beta^{[j-1]}) \\
\beta^{[j]} &\sim p(\beta|\boldsymbol{y}, s_j, \alpha_0^{[j]}, \alpha^{[j]}, \gamma^{[j]})
\end{aligned}
\tag{2.34}
$$

looping back to set $j = j + 1$, and so on through all $T$ observations. The current state is sampled based upon the last observation's parameter values. Given the current state $s_1$, parameter $\alpha_0^1$ is estimated using previous values $\alpha^0$,$\gamma^0$,$\beta^0$. After $\alpha_0^1$ is simulated and updated, it is used together with $\gamma^0$ and $\beta^0$ to estimate $\alpha^1$. Current values of $\alpha_0^1$,$\alpha^1$ and the previous $\beta^0$ value are used to estimate $\gamma^1$. Finally all three updated parameters are used in estimating $\beta^1$ to complete the simulation cycle for the current observation. It then starts over again with transition probabilities for the next state observation, and parameters are estimated in like fashion conditional upon the new sampled state. At the end of the full sample we have estimates for each parameter conditional upon which state we are in.

## 2.5. Data

Our data include options prices[14] on the S&P 500 index from 2005-2012 and prices on the underlying index itself. S&P 500 options are appealing for this study on several accounts. First, they comprise a large and active index option market - in 2004 the Chicago Board Options Exchange estimated the underlying notional value of trading in these options as more than $20 billion per day. In November 2010, S&P 500 options had a reported 15.3 million open interest, indicating a high degree of liquidity. Additionally the underlying index is capitalization-weighted according to total market value of outstanding shares, with 500 component stocks from a wide range of industries. It is a leading benchmark for investors, portraying a reasonably diversified picture of market developments. Finally, S&P 500 options are European-style, meaning they can only be exercised on the last business day before expiration. This greatly simplifies computational issues with possible early exercise.

Since state-price densities are derived from the call option-pricing function, the options sample must be comprised specifically of call prices. However we follow others from the literature in using the put-call parity condition to convert put prices into call prices, including these within our sample as well.[15] We believe this inclusion strategy to be more representative of market sentiment as a whole. We also weight call prices by open interest numbers to give greater importance to more liquid and active contracts in deriving the full density. We maintain a constant time-to-maturity for the data, using short-term 30-day contracts in estimation.

[14] Obtained through the Chicago Board Options Exchange's MDX service.

[15] Jackwerth and Rubinstein (1996), for example.

**Table 2.1**
**S&P 500 log-return summary statistics**

| | |
|---|---|
| Observations | 1762 |
| Mean | 3.053e-17 |
| Standard deviation | 1.461 |
| Skewness | -0.2855 |
| Kurtosis | 8.851 |

For state probability density estimation, we transform S&P 500 index prices into demeaned daily log-returns. Summary statistics are given in Table 2.1, with time series of S&P 500 index levels and log-returns for years 2005-2012 shown in Figure 2.1. The state probability densities are estimated on a monthly basis using daily closing values on the S&P 500 index. Because our state-price densities are calculated each month using forward-looking contracts, we maintain similar forward-looking expectations by forecasting state probability density estimates into the future as well to match the options expiries.

Historical prices are typically used to model the GARCH parameters, though prior papers differ in their lengths of historical perspective included. Rosenberg and Engle (2002) for example use a 25-year period from 1970-1995, and Gai and Vause (2006) go all the way back to the year 1920 in their estimations. Our concern however is there is so much history that entirely new market developments receive very little weight in the parameter estimation. It seems odd that today's observation would exert the exact same influence on present expectations for market movements as an observation that occurred decades prior. Certainly history matters in forming expectations of possible market movements, but how to balance this with the force of recent developments on market expectations? We choose to address this problem through an inverse-weighting scheme of

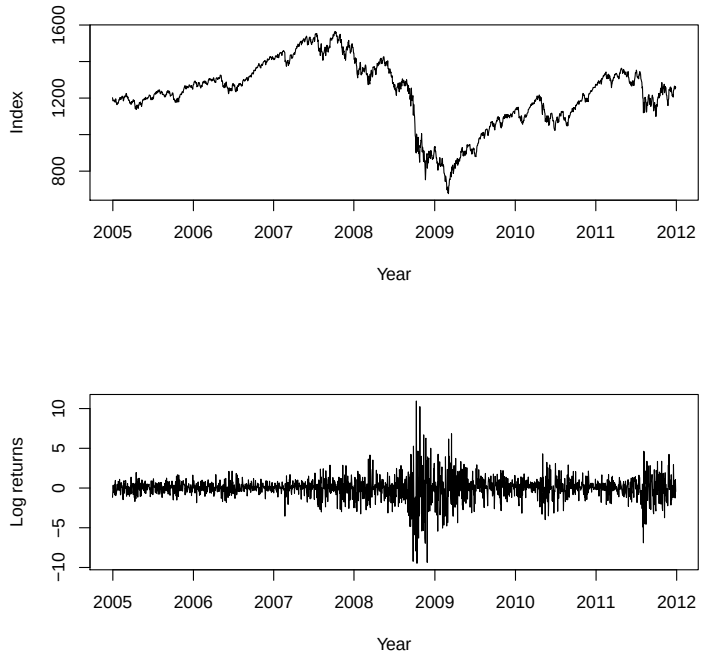


Figure 2.1: S&P 500 index and log-returns

the parameter estimates over time. Observations far into the past receive less and less weight in forming expectations, and those that are more recent receive a greater share of the weight.

## 2.6. Analysis

### 2.6.1. Comparison of models

We estimate single-regime and Markov switching GJR-GARCH models for the entire 2005-2012 sample in order to compare fit and inferences. For identification purposes we restrict $\alpha_{01} < \alpha_{02}$ in the Metropolis-Hastings simulation step. Frühwirth-Schnatter (2001) examines unidentifiability

in switching models where there is no unique unconstrained way to label the states. She finds in such cases that parameter estimation can switch between states, leading the parameters to shrink towards a multimodal mean. She shows how a unique labeling of the state spaces can be achieved however by taking into account the geometry of the unconstrained posterior. Examination of our unconstrained posteriors shows evidence of bimodality in $\alpha_0$, and scatterplots of $\alpha_0$ in the two states show a majority of observations clustered together but with a large amount of higher-ranging observations scattered separately. This suggests the use of $\alpha_0$ as a valid candidate for labeling the subspaces.

In Monte Carlo simulations of 10,000 draws, we discard the first 1,000 for burn-in purposes. Parameters are estimated as the posterior means of simulated chains, so it is important to look for evidence that convergence in the chains has been achieved.[16] We check for autocorrelation in the parameter draws, which can lead to convergence delays due to slow movement around the parameter space. These all decay quickly, with the highest one-lag autocorrelation being $\rho_\alpha = 0.677$ from the single-regime model. Trace plots of iteration number versus parameter value at each draw show that the simulated chains explore the parameter space well. Running-mean plots of iterations against the means of draws summed up to that iteration also appear to converge quickly. Acceptance rates for parameter chain draws are around 70% and above. We inspect Geweke diagnostic results for the parameter chains, which take two non-overlapping parts of each chain and compare means in each part to test the null hypothesis that they are both taken from the same distribution. The resulting statistics are all below critical values and fail to reject this null at the 5% significance level, suggesting that convergence in the chains is likely. Gelman-Rubin diagnostic tests exam-

[16] We use R package *coda* to assist in posterior analysis.

ining parallel chains with dispersed initial values are also run to see if they converge to the same target distribution. These tests compare the variance within chains to the variance between chains, with large discrepancies indicative of a failure to converge properly. The resulting potential scale reduction factors are all close to 1, suggesting that convergence is achieved and that our results are not sensitive to dispersed priors.

Table 2.2 contains results from our single-regime and Markov switching model estimations, reporting parameter mean and median values and high posterior density intervals.[17] Parameters $\alpha_0$, $\alpha$ and $\beta$ are significant in both models, but our results do not show the leverage parameter $\gamma$ to be statistically significant. Therefore we do not find evidence here of volatility being systematically higher in response to negative shocks relative to positive ones. Our estimates do show Markov switching parameters $\alpha_{01}$ and $\alpha_{02}$ to be significantly different between states. Thus the main difference between regimes appears to lie in the shifting baseline volatility. This may be largely due to the extreme changes over our sample period which show up as major structural shifts that outweigh the other more subtle conditional variance parameters.

The degrees of freedom parameter $\nu$ for the Student-*t* distribution is estimated to be 10.86 in the single-regime model and 12.35 in the regime-switching model. This suggests that the use of *t*-distributed innovations in our sample is indeed warranted to better capture tail behavior; values of 30 or above for $\nu$ would essentially revert to capturing a distribution that is normal. We estimate the transition probabilities from one state to the next to be $[\pi_{11} \;\; \pi_{12}] = [0.7135 \;\; 0.2865]$ in the low

[17] High posterior density intervals (HPDI) can be seen as a Bayesian counterpart to confidence intervals, capturing a specified range of the estimated parameter's posterior probability distribution.

**Table 2.2**
**Conditional variance parameter estimates**

| Model | Parameters | Mean | Median | 95% HPDI |
|---|---|---|---|---|
| Single-regime | $\alpha_0$ | 0.0593 | 0.0485 | [0.0352,0.1105] |
| | $\alpha$ | 0.0997 | 0.0860 | [0.0381,0.1798] |
| | $\gamma$ | 0.0006 | -1.889e-05 | [-0.0044,0.0046] |
| | $\beta$ | 0.8741 | 0.8843 | [0.7914,0.9399] |
| Regime-switching | $\alpha_{01}$ | 0.0525 | 0.0441 | [0.0270,0.0912] |
| | $\alpha_1$ | 0.0544 | 0.0403 | [0.0293,0.0945] |
| | $\gamma_1$ | 0.0012 | 0.0005 | [-0.0036,0.0040] |
| | $\beta_1$ | 0.9202 | 0.9371 | [0.8598,0.9533] |
| | $\alpha_{02}$ | 0.1616 | 0.1603 | [0.1021,0.2878] |
| | $\alpha_2$ | 0.0781 | 0.0827 | [0.0483,0.1637] |
| | $\gamma_2$ | 0.0011 | -0.0002 | [-0.0096,0.0055] |
| | $\beta_2$ | 0.9068 | 0.9141 | [0.8128,0.9396] |

volatility state, and $[\pi_{21} \ \ \pi_{22}] = [0.3049 \ \ 0.6951]$ in the high volatility state, implying a fair amount of mixing between states during our sample period.

To compare goodness-of-fit, we compute deviation information criteria for each of the models. The deviation information criterion (DIC) from Spiegelhalter, Best, Carline and van der Linde (2002) can be viewed as the Bayesian counterpart to maximum likelihood measures of fit. It compares models based upon the output from Bayesian estimation, taking both fit and complexity into account:

$$DIC = \overline{D} + p_D = E[-2lnf(y|\theta)] + \{E[-2lnf(y|\theta)] - -2lnf(y|\overline{\theta})\}. \tag{2.35}$$

$\overline{D}$ is defined as the posterior expectation of the deviance, and $p_D$ is the posterior expectation of the deviance minus the deviance evaluated at the posterior parameter means. These are calculated as by-products from our models by using estimated likelihood functions: $lnf(y|\theta)$ is evaluated with

**Table 2.3**
**Deviance information criterion, full sample**

| Model | DIC | 95% confidence interval |
|---|---|---|
| Single-regime | 5364.692 | [5348.765,5380.619] |
| Regime-switching | 4959.293 | [4946.853,4971.733] |

parameters estimated at each observation, and $lnf(y|\overline{\theta})$ is calculated with the estimated posterior parameter means. $\overline{D}$ can be interpreted as how much the model fitted for each observation deviates from the data itself, and $p_D$ serves as a kind of penalty for model complexity. Including additional parameters will automatically decrease the deviance evaluated at the posterior parameter means, resulting in a higher penalty that is weighed against the increased explanatory power. A lower relative DIC value is indicative of model preference.

To add further meaning to our comparisons we also use a resampling procedure described in Ardia (2009) to estimate confidence intervals for our DIC calculations. This involves allowing random perturbations in the estimated parameters and recalculating samples of new DIC values to estimate an entire distribution of values. Table 2.3 reports these estimates, indicating preference for the regime-switching model over the single-regime model.

For insight into model fit according to regimes, we next filter these estimates according to whether the observation occurs in the low volatility state or in the high volatility state.[18] The resulting criterion values along with their estimated confidence intervals are presented in Table 2.4. These indicate that the regime-switching model appears to significantly improve fit over the single-regime model in both low and high states of volatility.

[18] As designated by the Markov switching model.

**Table 2.4**
**Deviation information criterion, by state**

| Model | Volatility state | DIC | 95% confidence interval |
|---|---|---|---|
| Single-regime | low | 2146.811 | [2139.196,2154.426] |
| | high | 3217.882 | [3204.789,3230.975] |
| Regime-switching | low | 2049.636 | [2045.700,2053.572] |
| | high | 2909.656 | [2889.227,2920.085] |

Posterior predictive performance of the models is also examined. We replicate the data using each of the two models, and then measure their deviations from the actual observed returns. These are reported in the form of mean squared errors (MSE) and mean absolute errors (MAE). As fundamentally error-driven models, there is an essential randomness to their predictive results. Therefore we use a resampling procedure here as well to obtain mean values and 95% confidence intervals from an entire distribution of these measures. The results shown in Table 2.5 indicate that the regime-switching model tends to have smaller predictive errors than the single-regime model. We interpret that the regime-switching model has a somewhat better ability to capture the actual range of the observed data compared with the single-regime approach.

**Table 2.5**
**Posterior predictive results**

| Model | Measurement | Mean value | 95% confidence interval |
|---|---|---|---|
| Single-regime | MSE | 4.354 | [3.461,5.246] |
| | MAE | 1.551 | [1.398,1.703] |
| Regime-switching | MSE | 2.136 | [2.028,2.244] |
| | MAE | 0.9324 | [0.9063,0.9585] |

### 2.6.2. Pricing kernel behavior

Our results presented thus far add to the existing financial literature on Markov switching models of volatility. But what effect does regime-switching have on pricing kernel estimation? Though the evidence regarding fit appears to favor the regime-switching approach, we still need to ask: Does regime-switching volatility impact pricing kernel estimates compared with a conventional single-regime model?

To investigate, we estimate pricing kernel functions under both single- and regime-switching formulations as described above. We measure the probabilities of negative events as the tail areas of the densities below a given percent change in log-returns. When the pricing kernel is greater than 1, this means that the state-price density in the numerator is assigning a higher probability of that state occurring than the actual physical probability in the denominator. In that case, options on the S&P 500 index are being priced as if a downward movement in prices is more likely than the historical likelihood would warrant. These higher preference-weighted valuations relative to actual price movements reflect a higher aversion to risk. Alternatively, when the pricing kernel ratio is less than 1, the state-price density is assigning a lower probability of that state occurring than the actual probability suggests. Investors may be underestimating downward risks in the market, pricing options as if those downward movements are actually less likely than the historical likelihood would suggest. In such cases, aversion to risk in financial markets may be running low.

Sample pricing kernel estimates are shown in Figure 2.2. The pricing kernel function on the left is estimated during the low volatility state, and the pricing kernel function on the right is estimated during the high volatility state. S&P 500 return states here are measured as -2%,-1%,0%,1% and

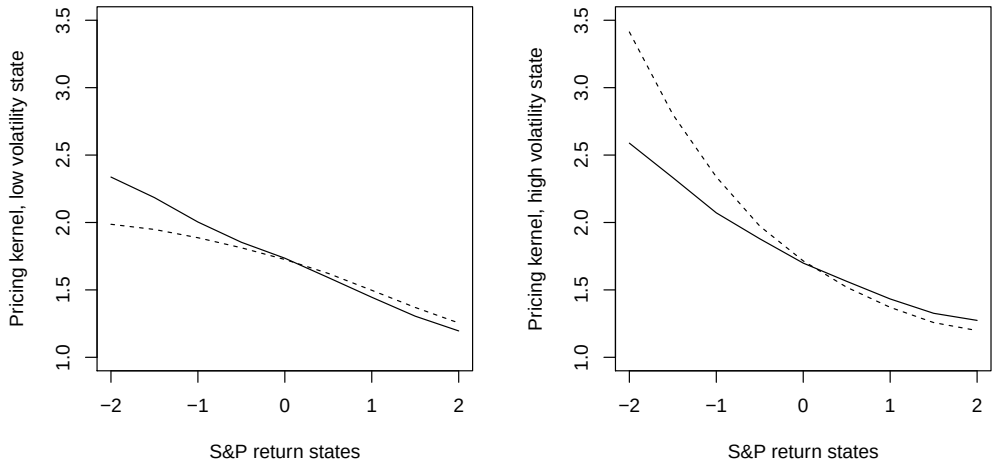


Figure 2.2: Pricing kernel function estimates.

Dashed lines from single-regime model, solid lines from regime-switching model.

2% changes in the S&P 500 index level. Dashed lines represent single-regime model estimates, and solid lines represent regime-switching model estimates. Qualitatively, in the low volatility state to the left, the single-regime model appears to flatten the pricing kernel function relative to the regime-switching model. Pricing kernel slopes tell us about the relationship between risk preferences and changes in the state of wealth. The steeper pricing kernel slope from the regime-switching model indicates that preferences are more highly correlated with changes in the state of wealth in low states of volatility than the single-regime model would suggest. In contrast, on the right the single-regime model overestimates the slope of the pricing kernel function relative to the regime-switching model. This suggests that the single-regime model over-represents the correlation between risk preferences and changes in the state of wealth in the high volatility state.

Examining across time, we separate pricing kernel estimates according to volatility regime across the entire sample. Tables 2.6 shows mean pricing kernel values for a 1% fall in the state

**Table 2.6**
**Pricing kernels, by state**

| Model | Volatility state | Mean | 95% confidence interval |
|---|---|---|---|
| Single-regime | low | 1.048 | [0.8973,1.199] |
| | high | 1.654 | [1.410,1.899] |
| Regime-switching | low | 1.179 | [1.006,1.351] |
| | high | 1.247 | [1.039,1.454] |

of wealth and their 95% confidence intervals. The single-regime model identifies mean pricing kernel estimates across states as significantly different between regimes, but the regime-switching model does not. This suggests that the single-regime model may be mistakenly identifying risk preference behavior in failing to distinguish between changing regimes of volatility.

Next we divide the sample by time with a breakpoint at the end of September 2008. This is when the most extreme spikes in volatility occur in our sample, representing an acute period of distress in financial markets. Mean pricing kernel values for the time intervals before and after this breakpoint and their 95% confidence intervals are reported in Table 2.7. While the regime-switching model did not identify a significant difference in pricing kernel values between states, it does identify a significant difference in pricing kernel values before and after the extreme events of September 2008. The pricing kernel mean prior to the end of September 2008 is below 1, suggesting that aggregate preferences for risk were rather high. The pricing kernel mean from October 2008 onwards on the other hand is significanty greater than prior to this breakpoint. The mean here is greater than 1, suggesting that aggregate preferences for risk have dropped. The fact that both volatility states are interspersed throughout the second period indicates that preferences for risk are lower on average despite reverting back to a low state of volatility. So rather than being

**Table 2.7**
**Pricing kernels, by time period**

| Model | Time period | Mean | 95% confidence interval |
|---|---|---|---|
| Single-regime | pre-Sept 2008 | 0.8366 | [0.6710,1.002] |
| | post-Sept 2008 | 1.572 | [1.413,1.899] |
| Regime-switching | pre-Sept 2008 | 0.8813 | [0.7061,1.057] |
| | post-Sept 2008 | 1.553 | [1.375,1.730] |

closely tied to movements in volatility, aggregate risk preferences appear here to outlast the actual structural shifts in volatility itself.

## 2.7. Conclusion

In this paper we find evidence to support the existing literature on Markov switching models of volatility in financial returns. We examine the time frame surrounding the recent 2008 financial meltdown to gain insight into this period of extreme volatility. Results show an improvement in fit for the regime-switching model relative to a single-regime model of conditional variance for S&P 500 log-returns. The large changes in volatility throughout this period appear to be mainly explained by major structural shifts in the baseline variance parameter of the conditional variance process.

We extend these findings to examine the impact of regime-switching volatility on pricing kernel functions and risk preferences. This addresses known conceptual difficulties in the literature regarding pricing kernel estimation. Our results show that existing concerns are warranted, not just conceptually, but for empirical reasons as well. For one, the introduction of Markov switching parameters to capture high and low states of volatility leads to qualitative differences for pricing

kernels and risk preference behavior. Relative to the regime-switching model, the single-regime model underestimates the pricing kernel in negative states of wealth during times of low volatility. Conversely, it overestimates the pricing kernel in negative states of wealth during times of high volatility. This makes preferences for risk appear higher than they should be in the low volatility state and lower than they should be in the high volatility state. Thus risk preferences appear to change much more between volatility states according to the single-regime model than may really be the case.

Pricing kernel slopes also tell us about the correlation between risk preferences and changes in the state of wealth. By flattening the pricing kernel slope at times in the low volatility state, the single-regime model under-represents the correlation between risk preferences and changes in the state of wealth. When overestimating the slope in the high volatility state, it over-represents this relationship. This suggests that risk preferences may be more sensitive to changes in states of wealth than previously indicated in the low volatility state, but less sensitive than previously indicated in the high volatility state. This may be perhaps because downward movements in the state of wealth are unexpected within the low volatility state, but come as less of a surprise when already in the high volatility state. An alternative explanation might be that the implied composition of the market is different between models – the Markov switching model could be more suggestive that some investors with low levels of risk preference are exiting the market in the high volatility state.

Regarding the relationship between volatility and risk preferences over time, the two models have separate implications here as well. The single-regime model shows a significant change in risk preferences across time between states of volatility, yet the regime-switching model does not.

This suggests that the single-regime model may indeed be mistakenly identifying such changes in risk preferences by failing to account for possible regime changes in volatility over time. This lies at the heart of existing concerns regarding the implicitly fixed state probability estimation for pricing kernels. The regime-switching model does however pick up a significant change in risk preferences after the financial meltdown in September 2008 through the end of the sample in 2012. This means that aggregate risk preferences are remaining low even after the market transitions back to the low volatility state. This behavior suggests that risk preference may have a long memory after the impact of extreme events, outlasting the actual estimated structural shift in volatility itself.

# Chapter 3

# Are Options Prices Empirically Consistent with the Time Series of the Underlying Asset?

This paper investigates pricing inconsistencies in options markets by comparing the prices of S&P 500 index options with the time series of the underlying index. We extend the model of Aït-Sahalia, Wang and Yared (2001), in which uncertainty about the future state of the market is introduced through stochastic movements in the underlying asset price. We examine whether or not their finding of systematic inconsistencies in options pricing may disappear when considering a less restrictive formulation. We adjust their approach to introduce a second form of uncertainty by including stochastic volatility in the model. Our results indicate that incorporating this additional uncertainty factor does not fully explain away the pricing inconsistencies. Therefore our findings

support the existing evidence for mispricing in options markets with respect to the behavior of the underlying asset.

## 3.1. Introduction

Are options prices empirically consistent with the time series of their underlying assets? As argued in Aït-Sahalia, Wang and Yared (2001), under assumptions of no arbitrage and dynamically complete markets, options should be tied to the prices of the assets upon which they are written. Options prices then should reflect the risks related to the stochastic movements of those underlying assets, but not anything else. If options are being priced consistently in financial markets, their implied forecasts on average should reflect the underlying asset behavior. One would expect to find incidental discrepancies of course due to a lack of perfect foresight. However, rational investors would be expected to take away the advantages of any such differences that are systematic in nature. Any systematic differences found remaining between the two would reflect in the options prices irrational beliefs concerning the behavior of the underlying asset.

This question of consistency between options prices and the underlying asset behavior is part of a larger existing literature that considers mispricing in options markets. For example, Jackwerth (2000) identifies mispricing in options markets as the most likely explanation for the pricing kernel puzzle, in which portions of estimated risk aversion functions are negative or increasing in wealth. Bondarenko (2003) investigates observed high prices of put options on the S&P 500 index. He concludes that no existing class of models can explain this anomaly, and that investors may not be acting fully rationally in their pricing of options. Han (2008) regresses various measures of investor

sentiment on risk-neutral skewness in the S&P 500 index, finding evidence that sentiment may have a significant impact on options pricing. Elsewhere, Constantinides, Jackwerth and Perrakis (2009) identify mispricing in options markets through widespread violations of stochastic dominance, where traders would be able to increase their expected utility through zero-net-cost trading. Such evidence of aggregate errors in options pricing is important in a broad sense, because it suggests potential limitations in assumptions of rational asset pricing.

We contribute to this literature by extending the approach developed in Aït-Sahalia, Wang and Yared (2001), in which the authors examine whether options-derived risk-neutral densities match the risk-neutral densities derived from the underlying assets. A key advantage of their approach is that it is directly testable; others may tend to offer mispricing as an untested suggestion for otherwise unexplained behavior in options markets. The authors show that the options prices do not appear to be empirically consistent with the behavior of the underlying assets, citing investor fear of rare but catastrophic events as a likely factor in the observed options mispricing.

In their model, Aït-Sahalia, Wang and Yared (2001) describe the underlying asset dynamics through an Itō stochastic differential process[1] composed of a drift and a volatility term. Uncertainty concerning the future state of the market is driven by stochastic movements in the underlying asset price. This means that the volatility may change over time through its dependence on the asset price, but not as a separate process with its own random dynamics. Our concern is whether this limitation on uncertainty in the asset pricing dynamics may result in the appearance of pricing

[1]Commonly used in the financial literature, this is a class of stochastic processes that satisfies an integral equation of the following general form: $X(t) = X(0) + \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s$. It is driven by Brownian motion $W$, a real-valued continuous stochastic process with independent and stationary increments. For further reference see Fouque, Papanicolaou and Sircar (2006), for example.

inconsistencies when compared with an options-derived density that is known to be highly variable over time. We address this issue by introducing a second uncertainty factor into the model through stochastic volatility in the asset prices.

The chapter proceeds as follows: section 3.2 provides a conceptual foundation for the model used in Aït-Sahalia, Wang and Yared (2001) and gives a context for the present paper's contribution; section 3.3 provides a framework for the models used in this paper; section 3.4 sets out the estimation strategy; section 3.5 describes the data; section 3.6 interprets the estimation results; and section 3.7 concludes.

## 3.2. Conceptual background

We begin by providing some background on the approach used in Aït-Sahalia, Wang and Yared (2001) for investigating this question of options mispricing. They compare probability densities forecasted from S&P 500 index options with those estimated from the underlying index realizations. The idea is that systematic differences between the two would reflect irrational beliefs priced into the options regarding the expected behavior of the underlying asset.

The essential problem the authors face in comparing the two densities is that, while options-derived probability densities are known to be risk-neutral, those derived from the time series of the underlying assets are not. Risk-neutral probabilities are the probability-weights that equate the

expected future payout structure, discounted at the risk-free rate of interest, with the price of the claim for payout:

$$p = \frac{1}{1+r^f} \int \pi^*(S_T) x(S_T) dS_T. \tag{3.1}$$

Here $p$ is the claim price, $r^f$ is the risk-free rate of interest, $S_T$ denotes the underlying state $S$ occurring at the time of claim maturity $T$, $\pi^*(\cdot)$ are the risk-neutral probabilities of each state occurring, and $x(\cdot)$ are the state-contingent payouts. The probabilities derived from observed underlying asset prices will only be risk-neutral if market participants are actually discounting the future at the risk-free rate of interest. If the participants themselves do not have neutral preferences for risk, then the options-derived and underlying asset-derived densities will automatically differ by the extent of the participants' aversion to risk. Therefore, without some kind of prior assumptions regarding the risk preferences of market participants, it is not possible to meaningfully reconcile these two densities and derive implications about their pricing consistency.

To address this problem, the authors propose estimating transformed densities for the underlying assets that are risk-neutral as well. This is important, because risk-neutrality does not require any assumptions regarding investor risk preferences. Rather, under a common framework of risk-neutrality, the prices of different assets can all be determined using the same risk-free rate of interest. Risk-neutrality in this sense does not imply that market participants in fact have neutral preferences for risk. It simply imposes a kind of pricing consistency between assets. If markets are truly pricing options based only upon the risks tied to the stochastic movements of their underlying assets, then the risk-neutral densities for these two instruments should match. By comparing two

risk-neutral densities, the authors are in effect neutralizing preferences to be able to meaningfully focus on whether the options-derived expectations match the realized price movements.

Yet this is all further complicated by the fact that we do not actually observe risk-neutral prices on the underlying asset. In order to estimate underlying asset dynamics that are risk-neutral as well, Aït-Sahalia, Wang and Yared (2001) use an approach based upon Girsanov (1960). Consider an Itō stochastic differential process describing the underlying asset price dynamics:

$$dS_t = \mu S_t dt + \sigma S_t dW_t. \tag{3.2}$$

$S_t$ is the observed asset price at time $t$, $\mu$ is the drift term telling us about the change in mean value of the stochastic process over time, $\sigma$ is the volatility term telling us about the dispersion of the stochastic process, and $W_t$ are independent Brownian motions describing random successive movements of the asset's dynamics. Girsanov (1960) shows that the density of this continuous stochastic process should be identical under different measures of probability. Therefore the volatility term will be the same, whether describing the asset's dynamics under assumptions of risk-neutrality or not.

Crucially this means that an estimator for the volatility term can be constructed from observed asset prices. This same estimator can then be used to describe the volatility process under unseen risk-neutral conditions as well; we do not actually have to observe risk-neutral prices in this case in order to estimate the risk-neutral density. Under an alternative measure of risk-neutrality, only the drift term $\mu$ will differ from that of the observed dynamic process. The risk-neutral drift term coincides with the rate that would occur if all investors themselves were risk-neutral; under no-

arbitrage it would be the cost of carrying the underlying asset, or the instantaneous rate of return minus dividends. This can be estimated using the spot-forward parity condition:

$$F_{t,\tau} = S_t e^{(r_{t,\tau}-\delta_{t,\tau})\tau}$$
$$r_{t,\tau} - \delta_{t,\tau} = \frac{ln(F_{t,\tau}/S_t)}{\tau} \tag{3.3}$$

where $r_{t,\tau}$ is the rate of return, $\delta_{t,\tau}$ is the rate of continuous dividends, $F_{t,\tau}$ is the observed futures price at time $t$ on a contract expiring at $T = t + \tau$, and $S_t$ is the observed spot price of the asset. Thus we can estimate the alternative risk-neutral formulation for the observed process in equation (3.2) as:

$$dS_t^* = (r_{t,\tau} - \delta_{t,\tau})S_t^* dt + \sigma S_t^* dW_t^* \tag{3.4}$$

where $S_t^*$ and $W_t^*$ denote risk-neutral characterizations of the asset prices and Brownian motions. Volatility term $\sigma$ in the risk-neutral process is identical in the observed physical process.

Then, under a risk-neutral framework, the options- and underlying asset-derived densities can be meaningfully compared to try to identify systematic inconsistencies. Aït-Sahalia, Wang and Yared's (2001) results show evidence of systematic differences, particularly in the form of excess skewness and kurtosis of the options-derived densities relative to the underlying asset-derived densities. They propose an explanation regarding investor irrationality in the form of a peso problem, with assets being priced as if negative events are much more likely to occur than in reality.[2]

[2]A *peso problem* describes a phenomenon in asset pricing where expectations for rare events are high despite relatively infrequent actual occurrences. The phenomenon derives its name from being noted in Mexican currency markets in the 1970s, reportedly first described in Rogoff (1977) and in Krasker (1980).

They also extend their model to examine price jumps, but find similar results of options pricing inconsistencies.

However, the authors acknowledge that their model is limited in that it only represents a univariate diffusion process. This means that uncertainty is driven solely through stochastic movements in the underlying asset price $S_t$. A second form of uncertainty in the specification could be achieved through variation in the volatility process itself, so that volatility $\sigma$ is stochastic in its own right. We contribute to their existing analysis by considering this second form of stochastic uncertainty in the diffusion process to see how it impacts prior findings of options pricing inconsistencies below.

## 3.3. Model

### 3.3.1. Risk-neutral underlying asset dynamics

In Heston's (1993) model of stochastic volatility, the asset price dynamics are expressed under physical probability measure $\mathbb{P}$ as:

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_{1t} \tag{3.5}$$

$$dV_t = \kappa(\theta - V_t)dt + \sigma_v \sqrt{V_t} dW_{2t} \tag{3.6}$$

where $dV_t$ denotes the volatility process,[3] $\theta$ is its long-run mean, and $\kappa$ is the mean-reversion speed. The term $\sigma_v$ is interpreted as the so-called "volatility of the volatility."[4]

Heston's (1993) model allows much more flexibility in the distributional form of the underlying process. For example, a positive correlation between error terms from the volatility and the S&P 500 index processes will increase the variance as well as create positive skewness in the right tail. Conversely, a negative correlation in these error terms will increase the variance but create negative skewness in the left tail. The volatility of the volatility term, $\sigma_v$, can allow further flexibility. When $\sigma_v$ is zero, the volatility itself is deterministic and the distribution is normal. However, when $\sigma_v$ is non-zero it increases the kurtosis of the distribution, resulting in fatter tails.

To directly compare the probability density derived from this model with the risk-neutral density derived from options, it must be estimated under a risk-neutral measure as well. It must be noted however that a straight-forward Girsanov (1960) application as used in Aït-Sahalia, Wang and Yared (2001) is unsuitable when incorporating stochastic volatility into the model. This is because a unique risk-neutral measure requires that markets are complete; that is, we must be able to use a portfolio of cash and assets to perfectly replicate any payoff. With a second dimension of uncertainty coming from stochastic volatility, an additional independent Brownian motion is introduced into the process - $W_{1t}$ in equation $(3.5)$ and $W_{2t}$ now in equation $(3.6)$. Therefore, trading in the stock alone will not be able to replicate all possible options payoffs. We would need to be

[3] More precisely, $dV_t$ is the instant conditional variance in the asset price dynamics. We will refer in very general terms to the overall volatility process, so as to reflect a focus on modeling the volatility $\sqrt{V_t}$ of the asset price. We hope our readers are not offended by this generalization of the terminology.

[4] Heston (1993), p.338.

able to account for the valuation of volatility itself as a separate asset with its own risk premium. Otherwise, the risk-neutral measure is not necessarily a unique one.

Heston (1993) approaches this problem by expressing the risk-neutral dynamics of the stochastic volatility model as follows:[5]

$$dS_t^* = (r_{t,\tau} - \delta_{t,\tau})(S_t^*)dt + \sqrt{V_t^*}(S_t^*)dW_{1t}^* \tag{3.7}$$

$$dV_t^* = \kappa^*(\theta^* - V_t^*)dt + \sigma_v\sqrt{V_t^*}dW_{2t}^*. \tag{3.8}$$

Here $\theta^*$ and $\kappa^*$ are the risk-neutral parameters of the volatility process. They are defined as $\theta^* = \frac{\kappa\theta}{\kappa+\lambda}$ and $\kappa^* = \kappa+\lambda$, where $\lambda = \gamma cov(dV_t, dC_t/C_t)$, $\gamma$ is the parameter of relative risk aversion, and $C_t$ is the level of consumption. Parameter $\lambda$ helps capture investor attitudes towards the volatility risk. When $\lambda$ is positive, it is capturing a positive covariance between volatility and consumption; markets are up at the same time that they are displaying a higher volatility. This is associated with a lower risk-neutral mean volatility relative to the statistical volatility: $\theta^* = \frac{\kappa\theta}{\kappa+\lambda^{(+)}} < \theta$. The general implication is that investors may under-weight the volatility of the asset price when times are good. However when $\lambda$ is negative, it is capturing a negative covariance between volatility and consumption; markets are down at the same time that they are displaying a higher volatility. This is associated with a higher risk-neutral mean volatility relative to the statistical volatility: $\theta^* = \frac{\kappa\theta}{\kappa+\lambda^{(-)}} > \theta$. The implication is that investors may over-weight the volatility of the asset price when times are bad.

[5] We carry over the risk-neutral drift rate here in the asset price as specified in equation (3.3) above.

Bates (1996; 2000) shows another way to relate the risk-neutral parameters to their physical counterparts with the following physical measure of the volatility process:

$$dV_t = (\alpha - \beta V_t)dt + \sigma_v \sqrt{V_t} dW_t. \tag{3.9}$$

Corresponding to equation (3.6) above, $\alpha = \kappa\theta$ and $\beta = \kappa$. His risk-neutral formulation is:

$$dV_t^* = (\alpha - \beta V_t^* + \Phi_v)dt + \sigma_v \sqrt{V_t^*} dW_t^* \tag{3.10}$$

where $\Phi_v = cov(dV_t^*, \frac{dM_t}{M_t})$ is the volatility risk premium,[6] and $M_t$ denotes the marginal utility of wealth.

From asset pricing theory, we know that the marginal utility of a dollar changes with the state of wealth - when markets are down we have a higher marginal utility for payouts than we do when markets are up. What equation (3.10) indicates then, is that investors may give greater weight to the volatility of asset prices when in lower states of wealth. For example, when there is an increase in the marginal utility of wealth, $dM_t/M_t$ is positive. If the volatility has also increased, then $\Phi_v$ is positive and increases the risk-neutral volatility drift rate relative to its statistical value: $(\alpha - \beta V_t)dt < (\alpha - \beta V_t^* + \Phi_v^{(+)})dt$. This higher risk-neutral volatility is applied in the asset price

[6]This corresponds to Heston's (1993) formulation as follows:

$$\begin{aligned} dV_t^* &= \kappa^*(\theta^* - V_t^*)dt + \sqrt{V_t^*} dW_t^* \\ &= (\kappa+\lambda)(\frac{\kappa\theta}{\kappa+\lambda} - V_t^*)dt + \sqrt{V_t^*} dW_t^* \\ &= (\kappa\theta - (\kappa+\lambda)V_t^*)dt + \sqrt{V_t^*} dW_t^* \\ &= (\alpha - \kappa V_t^* - \lambda V_t^*)dt + \sqrt{V_t^*} dW_t^* \\ &= (\alpha - \beta V_t^* + \Phi_v)dt + \sqrt{V_t^*} dW_t^*, \end{aligned}$$

so $\Phi_v$ is a function of $\lambda$, specifically $\Phi_v = -\lambda V_t^*$.

dynamics in equation (3.7). Investors then would be behaving in negative states of wealth as if the volatility in asset prices is greater than it is in actuality. Similarly, when observing higher volatility in higher states of wealth with a lower marginal utilty, investors may behave as if the volatility in asset prices is less than it is in actuality.

### 3.3.2. Risk-neutral state-price densities

Once we have estimated probability densities based upon the underlying asset dynamics described above, we must estimate the options-derived densities for comparison. Options-derived probability densities provide information about expected movements in the underlying asset, because an option's payout is contingent upon the value of that asset at maturity. Therefore, the payout probabilities are directly linked to the expected probabilities of various movements in the underlying asset itself.

Breeden and Litzenberger (1978) show that it is possible to derive a forward-looking probability density for the underlying asset using options, by taking the second derivative of the call option-pricing function. To see this, suppose $P(S_T)$ is the price of a claim paying \$1 if specified state $S$ is reached at the time of maturity $T$, and pays nothing otherwise.[7] $P(S_T)$ is commonly referred to as a "state price." The authors demonstrate that the price of this claim may be expressed in terms of the cost of a portfolio that replicates the claim's payout:

$$P(S_T) = \frac{[c(S_T - \Delta S) - c(S_T)] - [c(S_T) - c(S_T + \Delta S)]}{\Delta S}. \tag{3.11}$$

[7] In the context of this paper, state $S$ corresponds to the state of wealth as measured by the S&P 500 index level.

Here $\Delta S$ is an arbitrary step size in the aggregate state, and $c(X)$ denotes the price of a European call option with the strike price $X$ in parentheses. If the underlying asset reaches the strike price of $X = S_T$ at maturity, options with strike prices greater than or equal to $S_T$ will not be profitable and will be allowed to expire. The option with a strike price of $X = S_T - \Delta S$ on the other hand will be profitable: it can be used to purchase the underlying asset for $S_T - \Delta S$, and the investor could immediately turn around and sell it for $S_T$. This would result in a net gain of $\Delta S$. Dividing by $\Delta S$ as numeraire gives us an identical payout of $\frac{\Delta S}{\Delta S} = \$1$ from the constructed portfolio. In the limit as $\Delta S \to 0$, equation $(3.11)$ can be expressed as:

$$\lim_{\Delta S \to 0} \frac{P(S_T)}{\Delta S} = \frac{\partial^2 c(X)}{\partial X^2}|_{X=S_T} = c_{XX}(X = S_T) \tag{3.12}$$

where $c_{XX}(X = S_T)$ denotes the second derivative of the call option-pricing function with respect to the strike price.

Of course the future state realization of the underlying asset is unknown at time $t$, and $S_T$ could take on a number of different possible values. For each expected future value of $S_T$ and its related payout, there will be a separate price $P(\cdot)$. The overall price $p$ of this derivative security, taking all possible future states and their payouts into account, is:

$$p = \int P(S_T)x(S_T)dS_T = \int c_{XX}(X = S_T)x(S_T)dS_T. \tag{3.13}$$

The payout values $x(S_T)$ at maturity for any future state $S_T$ greater than the strike price $X$ will be $x(S_T) = S_T - X$, and will expire as $x(S_T) = 0$ otherwise. Equation $(3.13)$ tells us that the derivative

security price $p$ is equal to the sum of all possible state-contingent payouts $x(S_T)$ weighted by their state prices $P(S_T)$, which are equal to the second derivative of the call option pricing function with respect to the strike price, or $c_{XX}(X = S_T)$. These state price weights can be interpreted as probabilities for each contingent payout.

Another way of understanding state prices as probabilities is as follows. If an investor holds contingent securities or combinations thereof at time $t$ that reflect every possible state outcome at $T$, then payout will be certain and the value of these summed contingent securities should be priced equal to that certain payout. Payout can be set to \$1 by definition, or state prices divided by payout value as numeraire, so that the state prices can be made to sum to one. As such we can interpret state prices as a sort of mapping of outcomes to probability measures over all possible states. The arbitrage-free price $p$ of the derivative security is equal then to its state-price probabilities $P(S_T)$ weighted by security payout structure $x(S_T)$. Time preference can be incorporated with a positive risk-free interest rate, discounting the probability-weighted payout structure by $\frac{1}{1+r^f}$. To maintain the sum of probabilities to one in the state-price density, state prices $P(S_T)$ are multiplied by $(1+r^f)$, where now we have:

$$p = \frac{1}{1+r^f}\int x(S_T)(1+r^f)P(S_T)dS_T = \frac{1}{1+r^f}\int x(S_T)\pi^*(S_T)dS_T. \tag{3.14}$$

Here $\pi^*(S_T) = (1+r^f)P(S_T)$ is known as the risk-neutral probability of state $S_T$. We can view risk-neutral probabilities $\pi^*(S_T)$ as the probabilities that equate the security price $p$ with the state-contingent payout structure $x(S_T)$, discounted with the risk-free interest rate $r^f$.

## 3.4. Estimation

We estimate the underlying asset-derived probability densities from S&P 500 index price observations, using Bayesian techniques to estimate the parameters of the stochastic volatility model. We estimate the options-derived state-price densities from options written on the S&P 500 index, using a constrained least squares estimator derived in Aït-Sahalia and Duarte (2003). As explained above in section (3.2), these state-price probabilities are by nature risk-neutral. However, the observed underlying asset-derived probabilities are not. This requires a modification of the estimated parameters before we can compare the asset price-derived probability densities with the state-price densities estimated from options. The following sections provide further details about the mechanics of our estimation.

### 3.4.1. Stochastic volatility model

We base our estimation approach upon Eraker, Johannes and Polson (2003) and Li, Wells and Yu (2008). The model is specified to solve the following system of stochastic differential equations:

$$\begin{pmatrix} dS_t \\ dV_t \end{pmatrix} = \begin{pmatrix} \mu \\ \kappa(\theta - V_t) \end{pmatrix} dt + \sqrt{V_t} \begin{pmatrix} 1 & 0 \\ \rho\sigma_v & \sqrt{(1-\rho^2)}\sigma_v \end{pmatrix} dW_t. \tag{3.15}$$

Here $S_t$ is the log asset price and $\rho$ is the correlation in the error terms of each process, $\varepsilon_t^S$ from $dS_t$ and $\varepsilon_t^V$ from $dV_t$. The error terms are assumed to have a standard normal distribution.

To estimate the model parameters, Eraker, Johannes and Polson (2003) and Li, Wells and Yu (2008) use a Markov Chain Monte Carlo approach that addresses some of the complications that can arise. One advantage is that a Markov Chain Monte Carlo simulation strategy is able to explore posterior distributions for parameters whose joint densities are unknown and whose likelihood functions may be computationally intractable. In addition, Jacquier, Polson and Rossi (1994) show that this outperforms other related approaches such as method of moments and quasi-maximum likelihood estimators in the estimation of stochastic volatility models. It can also be used to obtain estimates for unobserved state variables, and can easily be extended to accommodate additional latent factors that may otherwise complicate estimation such as jumps in price or in volatility.

A Markov Chain Monte Carlo simulation strategy generates random samples from a target distribution given the observed asset prices. In Bayesian estimation terms, the optimal estimator is the posterior mean:

$$\{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{V}}\} = E[\boldsymbol{\theta}, \boldsymbol{V} | \boldsymbol{S}] = \int (\boldsymbol{\theta}, \boldsymbol{V}) p(\boldsymbol{\theta}, \boldsymbol{V} | \boldsymbol{S}) d\boldsymbol{V} d\boldsymbol{\theta} \tag{3.16}$$

where $\boldsymbol{\theta}$ denotes the vector of model parameters $(\mu, \kappa, \theta, \sigma_v, \rho)$, $\boldsymbol{V}$ are the latent volatilities, and $\boldsymbol{S}$ are the asset price observations. From Bayes' Theorem,[8] the joint posterior density itself is determined by the likelihood of the observed data $f(\boldsymbol{S}|\boldsymbol{V}, \boldsymbol{\theta})$, together with the latent variable density $p(\boldsymbol{V}|\boldsymbol{S})$ and the prior density of the parameters $p(\boldsymbol{\theta})$:

$$p(\boldsymbol{\theta}, \boldsymbol{V} | \boldsymbol{S}) = \frac{f(\boldsymbol{S}|\boldsymbol{V}, \boldsymbol{\theta}) p(\boldsymbol{V}|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{\int f(\boldsymbol{S}|\boldsymbol{V}, \boldsymbol{\theta}) p(\boldsymbol{V}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{V} d\boldsymbol{\theta}}. \tag{3.17}$$

[8] Bayes' theorem states that the conditional probability of $y$ given $x$ is equal to the conditional probability of $x$ given $y$ times the marginal probability of $y$ and divided by the marginal probability of $x$, or $P(y|x) = \frac{P(x|y)P(y)}{P(x)}$.

In order to remain consistent with the discrete steps of a Markov chain, Eraker, Johannes and Polson (2003) and Li, Wells and Yu (2008) consider a discretized version of equations (3.5) and (3.6) above:[9]

$$S_{t+1} - S_t = \mu dt + \sqrt{V_t}\varepsilon^S_{t+1} \tag{3.18}$$

$$V_{t+1} - V_t = \kappa(\theta - V_t) + \sigma_v\sqrt{V_t}\varepsilon^V_{t+1}. \tag{3.19}$$

Conditioning on $V_t$, changes in $S_t$ and $V_t$ are assumed to follow a joint bivariate normal distribution:

$$\begin{pmatrix} S_{t+1} - S_t \\ V_{t+1} - V_t \end{pmatrix} | V_t \sim N\left(\begin{pmatrix} \mu \\ \kappa(\theta - V_t) \end{pmatrix}, V_t \begin{pmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{pmatrix}\right). \tag{3.20}$$

Then if we define $\boldsymbol{Y} = (S_{t+1} - S_t, V_{t+1} - V_t)$, and $\boldsymbol{\Sigma}$ as the $2 \times 2$ variance-covariance matrix containing $V_t(1, \rho\sigma_v, \rho\sigma_v, \sigma_v^2)$, the full-information likelihood function[10] can be defined as:

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{V}|\boldsymbol{S}) \propto (\det\boldsymbol{\Sigma})^{-1} \exp[-\frac{1}{2}(\boldsymbol{Y} - \overline{\boldsymbol{Y}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \overline{\boldsymbol{Y}})]. \tag{3.21}$$

[9] Eraker, Johannes and Polson (2003) show that the potential parameter bias introduced through discretization is insignificant when using daily observation frequencies.

[10] Part of the advantage of the Markov Chain Monte Carlo approach is that the full-information likelihood $f(\boldsymbol{S}|\boldsymbol{V}, \boldsymbol{\theta})$ is used here and can be estimated, rather than the marginal likelihood $f(\boldsymbol{S}|\boldsymbol{\theta})$ which would involve integrating over all states and is usually unknown.

Priors on the parameters are assumed to be independent with normal and truncated-normal distributions:

$$p(\mu) \sim N(\mu_S, \sigma_S^2) \tag{3.22}$$

$$p(\kappa) \sim N(\mu_\kappa, \sigma_\kappa^2)\mathbb{I}\{\kappa \geq 0\}$$

$$p(\theta) \sim N(\mu_\theta, \sigma_\theta^2)\mathbb{I}\{\theta \geq 0\}$$

$$p(\sigma_v) \sim N(\mu_{\sigma_v}, \sigma_{\sigma_v}^2)\mathbb{I}\{\sigma_v > 0\}$$

where $\mathbb{I}\{\cdot\}$ denotes indicator functions that equal 1 if the specification in brackets is satisfied and 0 otherwise. Parameters $\kappa$, $\theta$, and $\sigma_v$ are all constrained to be positive to ensure that the conditional variance itself is positive. We initiate parameter values with randomized initial draws as in Li, Wells and Yu (2008). Parameters are later re-estimated using dispersed initial values to confirm that our results are not influenced by these initial choices.

We simulate the joint posterior density of returns with a Metropolis-Hastings algorithm, a sampling technique commonly used when the conditional distribution of parameters is unknown. The idea is to find an appropriate "transition kernel" - a conditional distribution function that tells us the probability of moving from one observation, $x$, to another, $y$ - from a target density generating the desired sample. In practice we start with a candidate-generating density $q(x,y)$, and when the process is at point $x$, point $y$ is generated from this density. If $q(x,y)$ satisfies certain conditions the move is accepted as a value from the target distribution, but if not the process reverts back to $x$ for the next draw. Eventually this process should converge to the desired target density. The acceptance ratio for evaluating each draw is constructed as $\boldsymbol{\alpha}_{MH} = \min\{\frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}|\mathbf{y})}\frac{q(\boldsymbol{\theta})}{q(\boldsymbol{\theta}^*)}, 1\}$, where the

first fraction is the posterior probability ratio, the second fraction is the ratio of proposal densities, $\boldsymbol{\theta}^*$ is the parameter vector updated with proposed candidate draws, and $\boldsymbol{\theta}$ is the parameter vector from the previous cycle.

The sampling algorithm involves updating the model parameters from observation $t-1$ at observation $t$ as follows:

$$\begin{aligned}
\mu^{[t]} &\sim p(\mu|\boldsymbol{S},\boldsymbol{V},\kappa^{[t-1]},\theta^{[t-1]},\sigma_v^{[t-1]},\rho^{[t-1]}) \\
\kappa^{[t]} &\sim p(\kappa|\boldsymbol{S},\boldsymbol{V},\mu^{[t]},\theta^{[t-1]},\sigma_v^{[t-1]},\rho^{[t-1]}) \\
\theta^{[t]} &\sim p(\theta|\boldsymbol{S},\boldsymbol{V},\mu^{[t]},\kappa^{[t]},\sigma_v^{[t-1]},\rho^{[t-1]}) \\
\sigma_v^{[t]} &\sim p(\sigma_v|\boldsymbol{S},\boldsymbol{V},\mu^{[t]},\kappa^{[t]},\theta^{[t]},\rho^{[t-1]}) \\
\rho^{[t]} &\sim p(\rho|\boldsymbol{S},\boldsymbol{V},\mu^{[t]},\kappa^{[t]},\theta^{[t]},\sigma_v^{[t]})
\end{aligned} \tag{3.23}$$

and then updating the volatility:

$$V_t \sim p(V_t|V_{t+1},V_{t-1},\boldsymbol{S},\mu^{[t]},\kappa^{[t]},\theta^{[t]},\sigma_v^{[t]},\rho^{[t]}).$$

The estimation process then loops back to set $t = t+1$, moving through all $T$ observations.

Once the stochastic volatility model parameters are estimated under physical measure $\mathbb{P}$, we must transform them under a risk-neutral measure. We follow Aït-Sahalia, Wang and Yared (2001) in using the spot-future parity condition to transform the asset price drift term. The volatility premium $\Phi_v$ in the risk-neutral volatility drift is modeled according to Bates's (1996; 2000) specification in equation (3.10) above.

### 3.4.2. State-price densities

For the options-derived densities we use an estimator described in Aït-Sahalia and Duarte (2003). Their nonparametric locally polynomial estimator is obtained through a combination of constrained least squares regression and smoothing.

The idea is to find $z_i$, the least squares values closest to an observed set of call option prices $c_1$, $c_2$ ... $c_N$ while satisfying specified restrictions. Vector $\mathbf{z}$ is solved for vector $\mathbf{c}$ such that

$$\min_{z\in\Re}\sum_{i=1}^{N}(z_i-c_i)^2=\min_{z\in\Re}\|\boldsymbol{z}-\boldsymbol{c}\|^2 \tag{3.24}$$

subject to restrictions on slope and convexity. Restrictions on the estimator are motivated by theory - that call option-pricing functions are decreasing and convex with respect to the strike price. These restrictions rule out arbitrage and guarantee that the density will be positive and integrate to 1. They are enforced via inequality constraints on the first two derivatives of the option-pricing function.

The call option-pricing function itself is given by:

$$C(S_t,X,\tau,r_{t,\tau},\delta_{t,\tau})=e^{-t_{t,\tau}t}\int_0^{+\infty}\max(S_T-X,0)\pi^*(S_T|S_t,\tau,r_{t,\tau},\delta_{t,\tau})dS_T \tag{3.25}$$

where $S_t$ is the price of the underlying asset at time $t$, $X$ is the strike price, $\tau$ is the time-to-expiration such that $T=t+\tau$, $r_{t,\tau}$ is the risk-free rate of interest, $\delta_{t,\tau}$ is the dividend yield on that asset, and $\pi^*(S_T|S_t,\tau,r_{t,\tau},\delta_{t,\tau})$ is the state-price density assigning probabilities to various conditional outcomes of the asset price at maturity. As risk-free interest rates $r_{t,\tau}$ we use Treasury bill rates.[11]

[11] H.15 historical selected interest rates from www.federalreserve.gov.

The first derivative of the call option-pricing function is:

$$\frac{\partial C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial X} = -e^{-r_{t,\tau}t} \int_X^{+\infty} \pi^*(S_T | S_t, \tau, r_{t,\tau}, \delta_{t,\tau}) dS_T \tag{3.26}$$

which satisfies the inequality:

$$0 \geq \frac{\partial C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial X} \geq -e^{-r_{t,\tau}t}. \tag{3.27}$$

This shows the first constraint for a monotone decreasing option-pricing function, where the first derivative is negative with respect to the strike price.

The second derivative of the call option-pricing function is:

$$\frac{\partial^2 C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial^2 X} = e^{-r_{t,\tau}t} \pi^*(X) \geq 0. \tag{3.28}$$

This shows the second constraint for a convex option-pricing function, with the second derivative positive with respect to the strike price. It is also possible to see here Breeden and Litzenberger's (1978) finding that the state-price density $\pi^*$ can be obtained from the second derivative of the call option-pricing function with respect to the strike price. Normalization is required, multiplying the second derivative by $e^{r_{t,\tau}t}$.

Aït-Sahalia and Duarte (2003) show transformed data points $z_i$ are then used in a smoothing step.[12] They assume regression function $z(X)$ is continuous as a second-order derivative. Then

[12]We use assistance from the *KernSmooth* package in R. The optimal bandwidth $h$ for the smoothing kernel is chosen automatically based upon the direct plug-in method described in Ruppert, Sheather, and Wand (1995).

according to local Taylor approximation methodology, function $z(X)$ can be estimated locally as a polynomial of order $p$, with $d$ denoting the order of the derivative:

$$\begin{aligned} z(X) &\approx z(X_0) + z'(X_0)(X - X_0) + \ldots + \frac{z^p(X_0)(X - X_0)^p}{p!} \\ &= \sum_{d=0}^{p} \frac{z^d(X)}{d!}(X_i - X)^d = \sum_{d=0}^{p} \beta_d(X)(X_i - X)^d \end{aligned} \tag{3.29}$$

for $X_i$ in a neighborhood of $X$, where $\hat{\beta}_d(X) \approx \frac{z^d(X)}{d!}$ .

This becomes a weighted least squares formulation minimizing the following:

$$\min_{\beta} \sum_{i=1}^{N} (z_i - \sum_{d=0}^{p} \beta_d(X)(X_i - X)^d)^2 K_h(X_i - X) \tag{3.30}$$

with respect to $\beta_d(X)$ to find the locally polynomial estimate $\hat{\beta}_d(X)$. At each fixed strike price point $X$, this is essentially a generalized least squares regression of $z_i$ on powers of $(X_i - X)$, or on the difference between approximated and fixed strike prices. In other words, regression function $z(X)$ is being approximated locally in the neighborhood of fixed strike prices $X$. This is asymmetrically weighted with kernel function $K_h(X_i - X)$ to give greater importance to approximated points that lie within closer proximity to fixed points. For asymmetric weights we specify the inverse of the squared distance between the index level and strike prices.

The procedure can be written in basic weighted least squares notation as:

$$\begin{aligned} &\min_{\boldsymbol{\beta}} (\boldsymbol{z} - \boldsymbol{X\beta})^T \boldsymbol{W} (\boldsymbol{z} - \boldsymbol{X\beta}) \\ &\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{z} \end{aligned} \tag{3.31}$$

where $\mathbf{X} = \begin{pmatrix} 1 & X_1 - X & \cdots & (X_1 - X)^p \\ 1 & X_2 - X & \cdots & (X_2 - X)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_N - X & \cdots & (X_N - X)^p \end{pmatrix}$, $\mathbf{W}$=$diag\{K_h(X_1 - X), K_h(X_2 - X), \ldots, K_h(X_N - X)\}$, and $\mathbf{z}$=$[z_1, z_2, \ldots, z_N]$.

Estimated coefficients $\hat{\beta}_d(X)$ are then used to find the option-pricing function $z(X)$ and its derivatives:

$$\hat{z}^{(d)}(X) = d!\hat{\beta}_d(X). \tag{3.32}$$

From Fan and Gijbels (1996), derivatives of order $d$ can be found using local polynomials of optimal order $p = d + 1$. Because it is the second derivative of the option pricing function that gives the state-price density, the polynomial should be estimated up to order 3.

## 3.5. Data

The time period for our study is 2000-2012. As the primary data component in our analysis, we use daily observations of the S&P 500 index. Time series plots of the S&P 500 index level and log-returns for the sample period are shown in the first and second plots of Figure 3.1 below. The third plot displays the latent volatility variable $V_t$. This is estimated using the stochastic volatility model from equations (3.5) and (3.6) for the full sample under physical probability measure $\mathbb{P}$.

For risk-neutral transformations of the asset price drift term $\mu$, we use S&P 500 index futures as a second major data component in our analysis. S&P 500 index futures are written on a quarterly

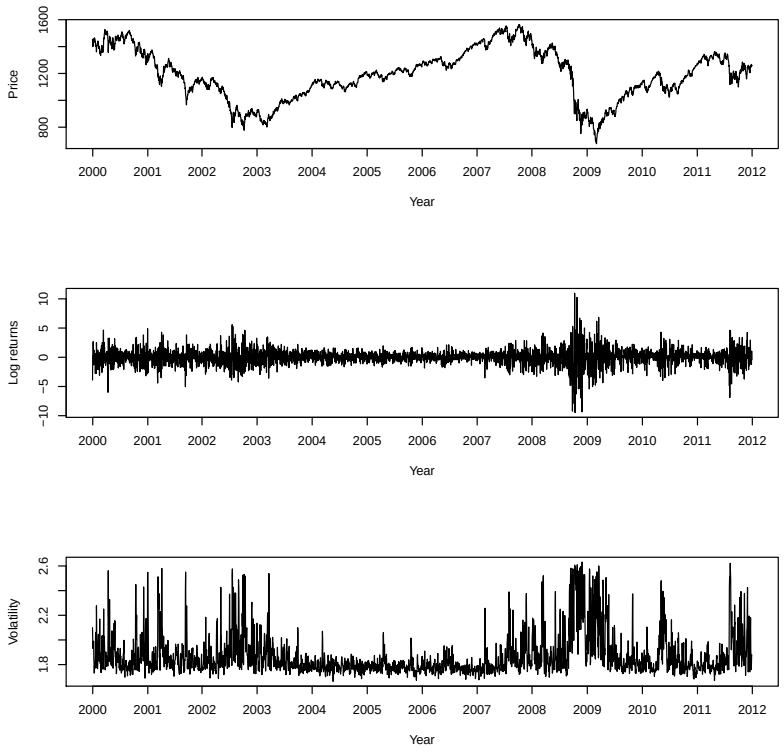


Figure 3.1: S&P 500 index prices, log returns, and volatility.

basis, expiring the third Friday of the settlement months of March, June, September, and December. We use these futures price observations $F_{t,\tau}$ together with S&P 500 index prices $S_t$ in the spot-forward parity condition from equation (3.3). This allows us to compute the risk-neutral drift rate $(r_{t,\tau} - \delta_{t,\tau})$ from equation (3.7).

In addition, we must also estimate the volatility risk premium $\Phi_v$ from equation (3.10). We follow Bates' (1996; 2000) formulation, using movements in the S&P 500 index level to approximate changes in the marginal utility of wealth. The volatility risk premium variable $\Phi_v = cov(dV_t^*, \frac{dM_t}{M_t})$ becomes $\Phi_v = -cov(dV_t^*, \frac{dS_t}{S_t})$. The negative sign arises from asset pricing theory, which tells us that marginal utility and state of wealth move in opposite directions.

The third major data component used in this analysis is S&P 500 index options. These are used to estimate state-price densities for comparison with the index-derived densities. It is ultimately this question of options-pricing consistency with the underlying asset's behavior that we are investigating in this paper. To maintain a constant horizon for time preference, we use 3-month options that follow the same quarterly cycle as our S&P 500 futures data. These options expire the Saturday following the third Friday of the same settlement months of March, June, September, and December.

Cross sections of options contracts with different strike prices are taken from a single day at the beginning of the 3-month contract period. For end-of-day price observations, we use the bid-ask midpoint. Since state-price densities are derived specifically from the call option-pricing function, only call options may be used in the state-price density calculations. We follow others such as Jackwerth and Rubinstein (1996) in using the put-call parity condition to convert put prices into call prices, including this information within our sample as well. We believe this inclusion to be more representative of market sentiment as a whole. We also weight call prices by open interest numbers to give greater importance to more liquid and active contracts in deriving the state-price density. Because S&P 500 index options are European-style, there are no complications arising from early exercise. The resulting state-price densities reveal expected future movements in the underlying S&P 500 index at the date of the options' maturity.

## 3.6. Analysis

We use Monte Carlo simulations of 10,000 draws to estimate the parameters in the stochastic volatility model, discarding the first 1,000 draws for burn-in purposes. Bayesian parameters are estimated as the posterior means of the simulated chains, so it is important for us to look for evidence that convergence in the chains has been achieved.[13] Simulation histograms and running-mean plots appear to converge quickly, and traceplots indicate good mixing. Acceptance rates for parameter chain draws are around 70% and above. Geweke diagnostics, which examine two overlapping parts of each chain to test whether they are taken from the same distribution, also indicate convergence. Gelman-Rubin diagnostics are checked, which compare the variance within parallel chains with dispersed initial values to the variance between these chains. These suggest that our results are not sensitive to priors.

Based upon realized S&P 500 index observations, we estimate the stochastic volatility model for every 3-month period to coincide with the futures and options contract periods. We then transform the estimated stochastic volatility parameters under the risk-neutral measure as described in sections 3.3 and 3.4 above. The risk-neutral form is used to simulate the probability density of the asset price movements during that particular 3-month observation period. We compute the first four conditional moments of the simulated densities in order to obtain the density characteristics. Next we use the options price cross sections to compute the state-price densities for each same 3-month period. We compute the first four moments of these options-derived densities as well. The question is whether the options-derived densities - characterized by their statistical moments

[13] We use R package *coda* to assist in posterior analysis.

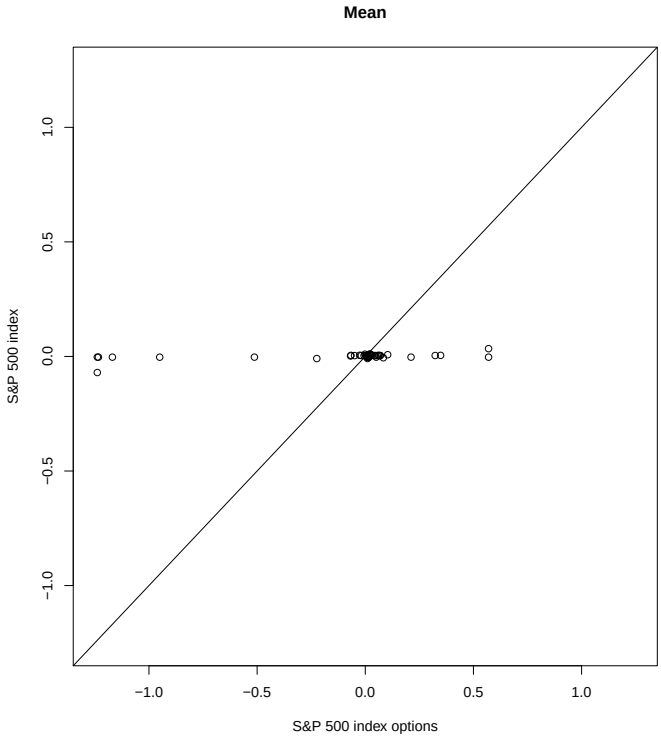


Figure 3.2: Mean scatter plot

- reflect the same general information as the realized index price densities, or if they somehow incorporate information that is inconsistent with the actual behavior of the underlying asset itself.

We plot the statistical moments calculated from the index-derived densities against those of the options-derived densities in Figures 3.2-3.5. These four scatter plots - displaying the respective mean, variance, skewness, and kurtosis calculations - are shown with 45-degree lines at the origins for comparison with the observed positions of the data points. The values along the axes represent expected changes in the index as percent changes in the index value between the start and finish of the 3-month period. These are based upon the options data for the horizontal scale and the index data for the vertical scale.

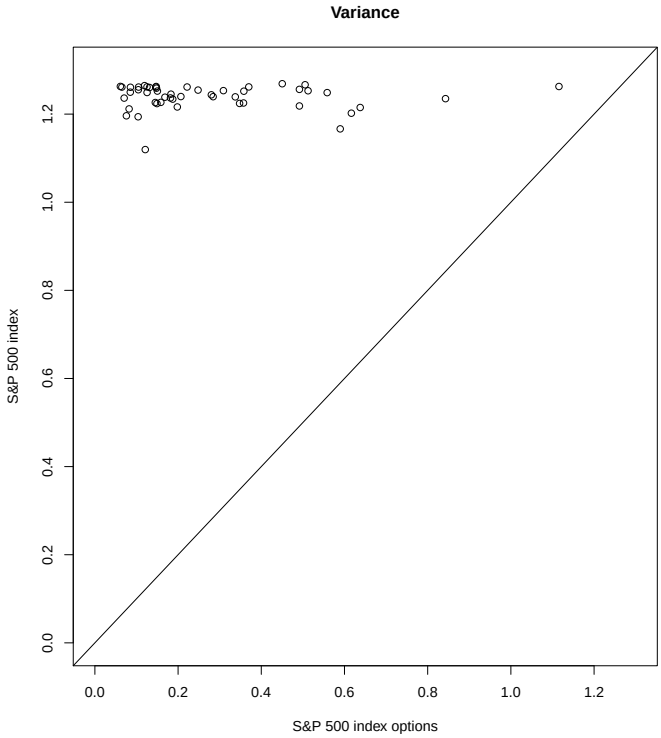


Figure 3.3: Variance scatter plot

We can see in Figure 3.2 that the densities derived from the actual behavior of the index have mean values very close to zero. The options-derived densities also have a majority of mean observations close to zero, but with several observations scattered into more extreme regions. This horizontal variation in options density mean values is found especially in the region of negative returns. It indicates that market expectations display a much higher degree of variation - especially in a negative direction - than is reflected in the behavior of the actual index itself. Market participants appear to be expecting the index to drop more often and more severely than the actual data suggests, and are pricing those options accordingly.

Figure 3.3 shows that the variance changes much more for the options-derived densities as well. This indicates a higher degree of variability in market expectations for the index return behavior

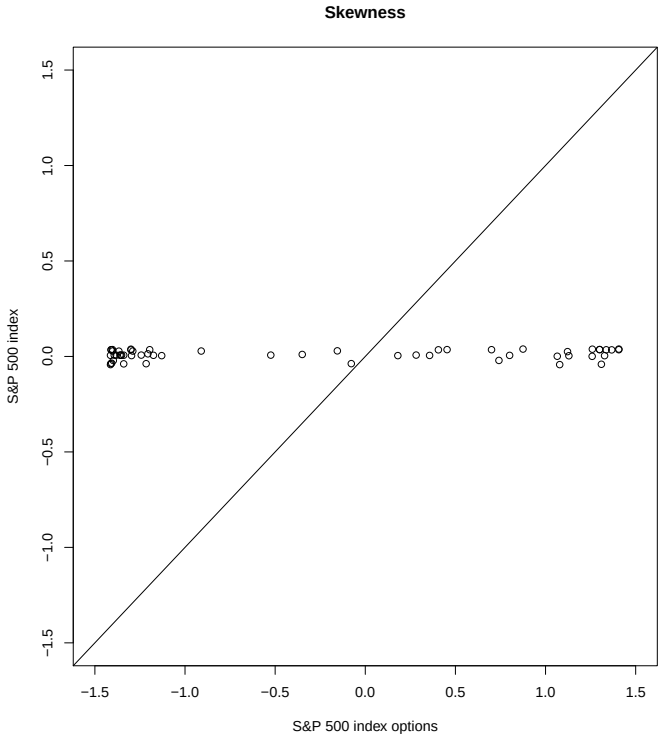


Figure 3.4: Skewness scatter plot

than is seen in the true range of behavior. The options actually appear to chronically underestimate the actual variance. This is a byproduct of weighting the options observations by their open-interest numbers to focus on more active trades when calculating the densities – the observed range of non-weighted options strike prices in fact far exceeds the range displayed by the actual returns. These open-interest weighted variances tell us that the bulk of trading activity happens fairly close to the mean, with extreme strike price observations located in the distant tails comprising relatively little of the overall trading activity.

Figure 3.4 indicates that the skewness of the options-derived densities also varies horizontally much more so than the index-derived densities. Notably, these skewness observations tend to be clustered out in distant positive and negative ends of the returns distribution rather than mostly

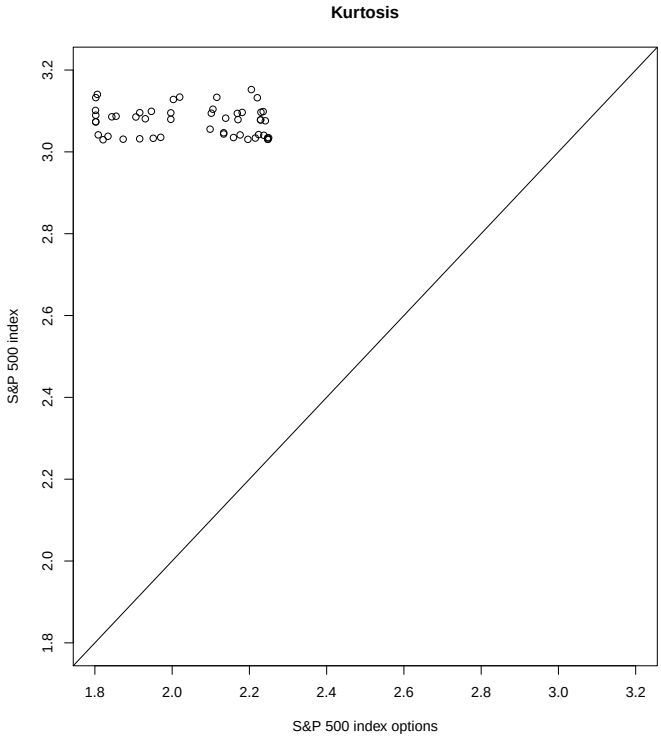


Figure 3.5: Kurtosis scatter plot

clustering near zero. Thus skewness appears to be a much larger feature of the options-derived densities than the index-derived densities. It suggests that market expectations may be systematically overweighting the actual probabilities of large movements in the index, particularly downward ones. This was noted in Aït-Sahalia, Wang and Yared (2001) as well.

In Figure 3.5 the options-derived densities display more variation in their kurtosis as well. In addition, all of these observations are clustered above the 45-degree line in the top left corner of the plot. We can see from this that the options appear to systematically underestimate the kurtosis of the index-derived densities. The index-derived densities tend to have kurtosis estimates slightly greater than 3, indicating that they are somewhat peaked with a bit of thickness to their tails. Yet these values are still rather close to a normal distribution. The options-derived densities on the

other hand show a more pronounced departure from a normal distribution. They all display a lower kurtosis, with values estimated around 2. This reflects the non-normal, fairly low and rounded peaks of the options-derived densities, with thinner tails. This under-estimation of the kurtosis is also likely influenced by open-interest weighting, because the less-frequent trading activity in extreme regions of the distribution will thin out the tails.

The greater apparent variability in moment characteristics for options-derived densities is important. As explained in section 3.3.1 above, a major benefit of introducing the stochastic volatility process is to achieve greater flexibility in the density calculations of the S&P 500 index. This is particularly so for the skewness and kurtosis estimates. However, analysis of Figures 3.2-3.5 suggests that the index behavior still exhibits far less variability than is displayed by the options-derived behavior. These figures suggest then that the stochastic volatility model still falls short of accounting for the inconsistencies between options and index density characteristics.

Following Aït-Sahalia, Wang and Yared (2001), we next take each of the four moments from the index-derived densities and regress them separately against the corresponding moments from the options-derived densities. We test the following restrictions on the regressions: *(1)* that the intercept $\beta_0$ from each moment regression should not be statistically different from a value of 0, and *(2)* that the slope $\beta_1$ from each moment regression should not be statistically different from a value of 1. These restrictions reflect the idea that, if the options are being priced consistently with their underlying asset behavior, then the coeffcient slope for each moment regression should essentially be a straight 45-degree line at the origin. We perform these regressions and test the restrictions.

**Table 3.1**
**Moment regression coefficients and standard errors**

| Moment | $\beta_0$ | S.E. | $\beta_1$ | S.E. |
|---|---|---|---|---|
| mean | 0.00238*** | (0.00067) | 0.00431*** | (0.00189) |
| variance | 1.23886*** | (0.00541) | 0.00096 | (0.01070) |
| skewness | 0.01134*** | (0.00376) | 0.00252*** | (0.00323) |
| kurtosis | 3.13498*** | (0.06332) | -0.02954 | (0.03077) |

Estimates of the intercepts and slopes along with their standard errors from each moment regression are reported in Table 3.2. We can see that the intercept estimate $\beta_0$ from each regression is statistically greater than zero. This is especially the case with the variance and kurtosis estimates. For slope $\beta_1$, the estimates for the variance and the kurtosis regressions are not statistically different from zero, so we reject the restriction that they are equal to 1. For the mean and skewness regressions, these slope estimates are statistically significant, so we perform hypothesis tests to see if they are significantly different from 1. We set up t-tests with $H_0 : \beta_1 = 1$ and $H_A : \beta_1 \neq 1$. For both tests we obtain p-values of 0.0000, so we reject the null hypothesis that these coefficients are equal to 1. Thus, each of the moment comparisons for the options- and index-derived densities appear to fail both restrictions on their intercepts and slopes. The options-derived densities do appear to have systematic differences from the index-derived densities, reflecting different information captured in the options pricing. This means that the options are not being priced consistently with the behavior of their underlying assets.

To understand the practical impact of the stochastic volatility model on density flexibility within this particular context, we estimate the original model in equation (3.2) from Aït-Sahalia, Wang and Yared (2001) as well. When we compare results using each of these models, we note some differences between them. Relative to our estimates from the original formulation, the

stochastic volatility model estimates have a higher average variance (by roughly 24%), a wider range of skewness values (by roughly 17%), and a slightly higher average kurtosis (by a little more than 3%). Most importantly, the ability of the stochastic volatility model to capture more of the skewness reflected in the options-derived densities makes it a better match for the asymmetric expectations of up- and downward movements in the index, due to its inherent flexibility. Yet in further comparison with the options-derived densities, this improvement does not appear very impressive. We find on average that the stochastic volatility model only makes up for around 2% of the differences in skewness between the options-derived densities and those in the original model. So while the stochastic volatility model indeed appears to be an improvement over the original model, it still leaves us very far from fully explaining the observed options pricing inconsistencies. In other words, there still appears to be a lot of possible room for investor sentiment in explaining options prices.

One interesting thing we would like to point out is that the skewness within the stochastic volatility model is largely dependent on the correlation coefficient $\rho$ between error terms in the volatility and index processes. Any correlation between these changes tends not to be sustained in one direction for extended periods of time, which means that large but short-lived postive or negative correlations are generally averaged out fairly close to zero. So while in theory the stochastic volatility model is able to capture skewness in the distribution due to its flexibility via the correlation coeffcient, in practice much of this advantage appears to be averaged away. The high degree of skewness in the options-derived distributions suggests the possibility that market participants could actually be overweighting the correlation between shocks in volatility and the index when

forming their expectations. Alternatively, it could be perhaps that the correlation coefficient itself is better thought of as a latent variable that changes more significantly over time.

## 3.7. Conclusion

In this paper we have taken up the question of options mispricing. We extend the approach of Aït-Sahalia, Wang and Yared (2001) to address a limitation in their univariate diffusion process. We add a second uncertainty factor by incorporating stochastic volatility into the model as well. We find that while it offers an improvement over the original model, addressing this limitation still falls short of explaining the systematic differences between options-derived and the underlying asset-derived densities. Our results therefore support the existing evidence of mispricing in options markets in the existing literature, leaving open the question of a possible role for investor sentiment in options pricing.

A further consideration that we would like to put forth here is that the estimated risk-neutral process may not be closely reflecting actual investor perceptions of the underlying asset behavior. The physical density of the underlying asset may be estimated with a certain degree of confidence since the process is observed in reality, but the risk-neutral process on the other hand is not observed. We wonder if the mismatch between options- and index-derived densities may stem in part from assumptions about the risk-neutral process.

Our estimation of the volatility risk premium in this paper is closely related to Bates' (1996; 2000) formulation: $\Phi = cov(dV_t^*, \frac{dM_t}{M_t}) = -cov(dV_t^*, \frac{dS_t}{S_t})$. Yet this really only takes into account

the way the variance changes relative to the marginal utility of wealth, not the particular direction of the change in the variance. Consider the case where the variance increases at a time that markets are down, so that we have a positive covariance between changes in the variance and the marginal utility of wealth. Then $\Phi$ is positive and leads to a higher risk-neutral valuation of the variance. Yet in the reverse case where the variance decreases and markets are up, we again have a positive covariance between changes in the variance and the marginal utility of wealth. Here again $\Phi$ is positive and leads to the same higher risk-neutral valuation of the variance. It is as if to say that investors are equally pessimistic when the market is down and the variance increases as they are when the market is up and the variance decreases.

Our findings in this paper may suggest that this does not appropriately portray how investors actually perceive asset behavior. This question of accurately estimating the volatility risk premium for the risk-neutral process in models of stochastic volatility seems a worthy issue for future research. The question of correlation between shocks in the volatility and the index could be an interesting direction of future research as well.

# Chapter 4

# Characterizing Volatility Regimes in the S&P 500 Index

We examine the behavior of returns in the S&P 500 index by estimating a regime-switching model of volatility. Our paper differs from other papers in the literature by considering volatility switching for a historical time frame, rather than focusing on recent events alone. This allows us to observe and characterize numerous transitions between relatively calm and more extreme periods in the market throughout the last century and beyond. We characterize this behavior accordingly.

## 4.1. Introduction

Volatility in financial markets can spike dramatically during periods of crisis. Such events tend to be rare and short-lived, but their dynamics can differ substantially from otherwise typical market

behavior. Especially in longer time series, models that estimate a single volatility process may fail to adequately capture extreme market conditions. This has implications not only for descriptive and forecasting purposes. Some papers have begun to ask whether certain economic relationships themselves may change when volatility rises. Edwards and Susmel (2001), for example, examine whether the correlation between stock markets increases across countries during periods of high volatility. This is an important consideration in the debate on international financial crises and market contagion. Elsewhere, Hou (2013) investigates the impact of interest rate changes on equity prices during times of high versus low volatility. This is important for monetary policymakers anticipating the market response to short-term interest rate changes, as well as for market participants choosing effective investment strategies. Such examples in the literature suggest there is much to be gained from improving our understanding of volatility dynamics in financial markets.

In this paper we estimate a regime-switching model of asset returns to examine the behavior of volatility in financial markets. Our paper differs from prior papers in the literature[1] by taking a long historical perspective in our model estimation. Others have used regime-switching to demonstrate a superior fit for a recent time series of data or to examine the impact in financial markets of a particular event of interest. However, though important questions can be addressed within a narrower range of focus, we believe that there is also much we can learn from a broad picture of regime-switching market activity. With a sample extending more than 140 years, we are able to observe and characterize numerous transitions between relatively calm and more extreme periods in the market throughout the last century and beyond.

[1]For example, Haas, Mittnik and Paolella (2004) and Ardia (2009).

For estimation we use a Markov switching GARCH model, where a discrete Markov chain controls the parameters of the conditional variance process. Markov switching GARCH models can help explain the high degree of volatility persistence that is often observed in asset returns. They also allow clear interpretation of discrete changes in volatility behavior. By disaggregating the volatility into distinct regimes, these models are able to identify separate volatility dynamics that essentially get averaged out when estimating a single-regime process. The growing popularity of Markov switching GARCH applications in recent years can be seen in papers such as Haas, Mittnik and Paolella (2004), Alexander and Lazar (2008), Ardia (2009), and Henneke, Rachev, Fabozzi and Nikolov (2011).

The data in our study has been made publically available by Robert J. Shiller.[2] Using Standard and Poor's Statistical Service *Security Price Index Record*, he has provided a monthly series for real S&P composite index prices dating all the way back to 1871. These are monthly averages of daily closing prices. Thus we are able to estimate our regime-switching volatility model over the time period of 1871-2013, providing a much longer context for more recent studies.

Our paper proceeds as follows: section 4.2 provides a conceptual foundation; section 4.3 describes our model in further detail; section 4.4 sets out our estimation strategy; section 4.5 describes our analysis; and section 4.6 concludes.

[2]See his website at: http:// www.econ.yale.edu/∼shiller/data.htm.

## 4.2. Conceptual background

Originally developed in Bollerslev (1986) as a generalization of Engle's (1982) autoregressive conditional heteroskedasticity model, GARCH models treat conditional variance as a function of past errors and past variance. In effect they allow the underlying variance to vary stochastically over time rather than assume it remains constant. This is appealing as error terms in equity return series often appear larger in some periods than in others, and with patterns of autocorrelated clustering. GARCH models accommodate heteroskedasticity by assuming that error term $u_t$ can be decomposed as $u_t = \varepsilon_t h_t^{1/2}$, where $\varepsilon_t$ are i.i.d. random variables with zero-mean and unit variance. Then $u_t$ guarantees a white noise property through its reliance on $\varepsilon_t$, and it is distributed conditional on the information set $\mathbf{\Psi}$ at $t-1$ as $u_t|\Psi_{t-1} \sim N(0,h_t)$.

GARCH models can be described as follows:

$$\begin{aligned} y_t &= a_0 + a y_{t-1} + u_t \\ u_t &= \varepsilon_t (h_t)^{1/2} \\ \varepsilon_t &\sim N(0,1) \\ h_t &= \alpha_0 + \alpha u_{t-1}^2 + \beta h_{t-1} \end{aligned} \tag{4.1}$$

for $t = 1,\ldots,T$. Here $y_t$ is a scalar dependent variable representing the series of log-returns, $y_{t-1}$ is its lag, $a_0$ a constant term, and $a$ the regression coefficient. $h_t$ is the conditional variance, which captures a degree of persistence in the asset volatility. With $\alpha_0$ as a kind of baseline variance, coefficients $\alpha$ and $\beta$ describe the influence of past shocks $u_{t-1}$ and past variance $h_{t-1}$ on today's

variance. Thus we can interpret a baseline component and a memory component contributing to the persistence of the conditional variance process.

Yet as pointed out in Hamilton and Susmel (1994), large and small shocks to asset returns may have different causes, not to mention different consequences for subsequent volatility behavior. Some shocks are transitory, and others have a longer-lasting impact on equity behavior. A single-regime estimate for the GARCH parameters cannot distinguish between these different effects. In a single-regime model the conditional variance parameter impacts are fixed, just with time-varying new shocks filtering through a constant-parameter process. In addition, models with singular volatility dynamics may lead to false inferences for asset behavior. Large but infrequent transitory shocks can bias estimates of volatility persistence upwards during more usual financial market conditions, for example. At the same time, single-regime estimates can greatly underrepresent the volatility under more extreme conditions.

Lamoureux and Lastrapes (1990) found that allowing structural shifts in the unconditional variance of stock returns could greatly improve the estimation of volatility persistence relative to a fixed-parameter GARCH process. Hamilton and Susmel (1994) and Cai (1994) extended this notion even further, introducing a Markov switching mechanism governing discrete changes in the parameters of an ARCH process. With a Markov switching mechanism the actual switching between states does not have to be specified in advance. Rather, it follows a random process with transition probabilities that evolve over time. This eliminates the need for an arbitrary pre-designation of the timing of structural shifts, unlike in the Lamoureux-Lastrapes formulation. Haas, Mittnik and Paolella (2004) made important further advances applying Markov switching to GARCH pro-

cesses. They were able to solve existing estimation difficulties by assuming independence between regimes. This means that the conditional variance only depends upon its prior history within that particular regime, as opposed to the entire history of variances integrated across all regimes.

A Markov switching GARCH model specifies random hidden states $s_{[1,T]}$ that control the parameters of the underlying GARCH process. These are assumed part of a stationary, irreducible Markov process that allows instantaneous shifts in volatility dynamics. Error terms for process $y_t$ follow $u_t = \varepsilon_t(h_{\Delta_t,t})^{1/2}$, where $\Delta_t$ is a Markov chain with $1,\ldots,k$ state spaces and a $k \times k$ transition matrix $\mathbf{\Pi}$:

$$\mathbf{\Pi} = [\pi_{i,j}] = [\pi(\Delta_t = j|\Delta_{t-1} = i)] \tag{4.2}$$

for $i, j = 1,\ldots,k$. The conditional probability of the current state at time $t$ only depends upon the previous state at time $t-1$. Regime variances then are given by:

$$\boldsymbol{h_t} = \boldsymbol{\alpha_0} + \boldsymbol{\alpha} u_{t-1}^2 + \boldsymbol{\beta h_{t-1}} \tag{4.3}$$

where $\boldsymbol{h_t}$=$[h_{1t},\ldots,h_{kt}]$ is a $k \times 1$ vector of regime variances, $\boldsymbol{\alpha_0}$=$[\alpha_{01},\ldots,\alpha_{0k}]$, $\boldsymbol{\alpha}$=$[\alpha_1,\ldots,\alpha_k]$, and $\boldsymbol{\beta}$=$diag(\beta_1,\ldots,\beta_k)$. These give us separate parameter sets that reflect the changing regimes of volatility in the market. The model also allows us to examine the characteristics of these separate dynamics – we can observe structural breaks in the dynamics of the volatility process when $\alpha_{01} \neq \alpha_{02}$, or whether the memory response to shocks may be different between regimes when $\alpha_1 \neq \alpha_2$, for example.

## 4.3. Model

We estimate a slightly modified version of the GARCH model described above. First, we incorporate error terms with a Student-$t$ distribution to better capture fatter asset return tails. Assumptions of $t$-distributed innovations are widely known to have advantages over the normal distribution in financial modeling, particularly in the presence of outlier observations. Second, we consider the GJR-GARCH specification of Glosten, Jagannathan and Runkle (1993) as it introduces asymmetry into the basic GARCH formulation. This addresses what is commonly known as a "leverage effect", where volatility is observed to be higher after negative shocks than after positive ones.

The $t$-distributed innovations introduce an additional latent variable, $\lambda_t$, into the formulation introduced above. We now estimate the following:

$$
\begin{aligned}
y_t &= a_0 + a y_{t-1} + u_t \\
u_t \sim t(\nu) \ , \ u_t &= \varepsilon_t (\lambda_t h_t)^{1/2} \\
\varepsilon_t \sim N(0,1) \ , \ \lambda_t &\sim IG(\frac{\nu}{2}, \frac{2}{\nu}) \\
h_t = \alpha_0 + \alpha(|u_{t-1}| - \gamma u_{t-1})^2 &+ \beta h_{t-1}. \qquad (4.4)
\end{aligned}
$$

$\lambda_t$ acts a scaling factor, and $\nu$ is the degrees of freedom characterizing the tail behavior. This provides an extra source of flexibility for responding to extreme observations, resulting in a conditional variance that is less volatile than with normally-distributed innovations. It also increases the stability of volatility states in the regime-switching model. $\gamma$ can be included in the conditional variance as a leverage parameter. When $\gamma$ is positive, the conditional variance $h_t$ will have a larger

response when $u_{t-1}$ is negative. If the leverage effect is insignificant the equation simply reverts back to a classic GARCH model.

We attempt to fit various ARMA formulations to the $y_t$ process for our sample but find the estimated coefficient high density posterior intervals include zero. We therefore simplify by assuming $y_t = u_t$. In the estimation approach described below, we follow Jacquier, Polson and Rossi (2004) and Henneke, Rachev, Fabozzi and Nikolov (2011) who conveniently rescale $y_t^* = y_t/\sqrt{\lambda_t}$ to be able to sample from normal proposal distributions in the Metropolis-Hastings steps. We assume $\alpha_0 > 0$, $\alpha$,$\beta \geq 0$, and $-1 < \gamma < 1$ in order to ensure that the conditional variance is positive, and $\alpha(1-\gamma^2)+\beta < 1$ so that the process is bounded and does not explode.[3]

## 4.4. Estimation

Because of the difficulties integrating and evaluating a likelihood function over multiple states using maximum likelihood methods, we follow Ardia (2009) and Henneke, Fabozzi, Rachev and Nikolov (2011) in using a Bayesian Markov chain Monte Carlo approach to estimate these parameters. In Bayesian terms the optimal estimator is the posterior mean:

$$\hat{\boldsymbol{\theta}} = E[\boldsymbol{\theta}|\boldsymbol{Y} = \boldsymbol{y}] = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} \tag{4.5}$$

[3] See Peters (2001).

where $\boldsymbol{\theta}$ denotes the model parameters $\{\alpha_0,\alpha,\gamma,\beta\}$, and $p(\boldsymbol{\theta}|\boldsymbol{y})$ is the posterior density. The posterior density itself is determined by the likelihood $f(\boldsymbol{y}|\boldsymbol{\theta})$ for observed data given the model, and the prior density $p(\boldsymbol{\theta})$:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{4.6}$$

The likelihood function is defined as:

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{y}) \propto (\det\boldsymbol{\Sigma})^{-1}\exp\left[-\frac{1}{2}\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right] \tag{4.7}$$

with $\boldsymbol{y}$ as the zero-mean vector of log-returns, and $\boldsymbol{\Sigma}$ a $T \times T$ diagonal matrix of conditional variances $\{h_t(\theta)\}_{t=1}^{T}$. Priors on the parameters are assumed to be independent with truncated normal densities:

$$\begin{aligned} p(\alpha_0) &\propto N(\mu_{\alpha_0},\sigma^2_{\alpha_0})\mathbb{I}\{\alpha_0 > 0\} \\ p(\alpha) &\propto N(\mu_{\alpha},\sigma^2_{\alpha})\mathbb{I}\{\alpha \geq 0\} \\ p(\gamma) &\propto N(\mu_{\gamma},\sigma^2_{\gamma})\mathbb{I}\{-1 < \gamma < 1\} \\ p(\beta) &\propto N(\mu_{\beta},\sigma^2_{\beta})\mathbb{I}\{\beta \geq 0\}. \end{aligned} \tag{4.8}$$

We re-estimate using dispersed initial parameter values to confirm that the results are not influenced by initial value choices.

Mechanically speaking, we simulate the joint posterior density using a Metropolis-Hastings algorithm. The Metropolis-Hastings algorithm is appropriate in cases such as this when the conditional distribution of estimates is unknown. The idea is to find an appropriate "transition kernel"

- a conditional distribution function that tells us the probability of moving from one observation, $x$, to another, $y$ - from a target density generating the desired sample. In practice we start with a candidate-generating density $q(x,y)$, and when the process is at point $x$, point $y$ is generated from this density. If $q(x,y)$ satisfies certain conditions the move is accepted as a value from the target distribution, but if not the process reverts back to $x$ for the next draw. Eventually this process should converge to the desired target density. It is not necessary to have perfect knowledge in choosing the candidate density, but care must be taken as poor choices can significantly delay chances of convergence to the target density. We follow Henneke, Rachev, Fabozzi and Nikolov's (2011) influence in derivation of proposal densities. The acceptance ratio for evaluating each draw is constructed as $\boldsymbol{\alpha}_{MH} = \min\{\frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}|\mathbf{y})}\frac{q(\boldsymbol{\theta})}{q(\boldsymbol{\theta}^*)}, 1\}$, where the first fraction is the posterior probability ratio, the second fraction is the ratio of proposal densities, $\boldsymbol{\theta}^*$ is the parameter vector updated with proposed candidate draws, and $\boldsymbol{\theta}$ is the parameter vector from the previous cycle.

Following Henneke, Rachev, Fabozzi and Nikolov (2011), we proceed in our estimation in the following steps:

1) Draw transition probability of moving from the current state observation to the next.

2) Sample new state based upon the drawn transition probability and the likelihood of current observation $y_t$.

3) Sample latent variable and degrees of freedom parameter for $t$-distributed errors.

4) Sample GARCH parameters sequentially, conditional upon state.

Then we return to step 1 to draw the state for the next observation and repeat throughout the remaining sample, updating parameters all along. These individual steps are explained more fully in the exposition below.

*Step 1: Draw transition probabilities*

Transition probabilities are drawn from a Dirichlet distribution, which is a generalization of the Beta distribution. To see why it is a natural choice for this context, consider the two-state example given in Henneke, Rachev, Fabozzi and Nikolov (2011). The authors write the posterior distribution of $\pi_{i,i}$, the probability of remaining in state $i$ from one observation to the next, as:

$$p(\pi_{i,i}|\boldsymbol{y},\boldsymbol{s},\boldsymbol{\theta}) \propto p(\pi_{i,i}) f(\boldsymbol{y},\boldsymbol{s}|\boldsymbol{\theta}). \tag{4.9}$$

This shows the posterior conditional probability of $\pi_{i,i}$ as proportional to its prior $p(\pi_{i,i})$ times the likelihood $f(\boldsymbol{y},\boldsymbol{s}|\boldsymbol{\theta})$ for the observed innovations $\boldsymbol{y}$ and states $\boldsymbol{s}$, given the estimated parameters $\boldsymbol{\theta}$. At this stage, the existing state $s$ is carried over from the prior observation made at time $t-1$; it can therefore be considered independent of the current observation for $y$. Then the likelihood portion $f(\boldsymbol{y},\boldsymbol{s}|\boldsymbol{\theta})$ in this expression can be further simplified to obtain:

$$p(\pi_{i,i}|\boldsymbol{y},\boldsymbol{s},\boldsymbol{\theta}) \propto p(\pi_{i,i}) f(\boldsymbol{s}|\boldsymbol{\theta}). \tag{4.10}$$

This second term $f(\boldsymbol{s}|\boldsymbol{\theta})$ is driven by the cumulated transitions over the sample from one period to the next:

$$f(\boldsymbol{s}|\boldsymbol{\theta}) = \prod_{t=1}^{T} p(s_t|s_{t-1},\boldsymbol{\theta}) = (\pi_{i,i})^{\eta_{i,i}}(\pi_{i,j})^{\eta_{i,j}}(\pi_{j,j})^{\eta_{j,j}}(\pi_{j,i})^{\eta_{j,i}}$$
$$= (\pi_{i,i})^{\eta_{i,i}}(1-\pi_{i,i})^{\eta_{i,j}}(\pi_{j,j})^{\eta_{j,j}}(1-\pi_{j,j})^{\eta_{j,i}} \tag{4.11}$$

where $\eta_{i,j}$ are the cumulative transitions from state $i$ to state $j$. This makes the Beta density a natural choice for the conjugate prior in each row of the transition probability matrix: $(\pi_{i,i})^{((H_{i,i}+\eta_{i,i})-1)}(1-\pi_{i,i})^{((H_{i,j}+\eta_{i,j})-1)}$ and $(\pi_{j,j})^{((H_{j,j}+\eta_{j,j})-1)}(1-\pi_{j,j})^{((H_{j,i}+\eta_{j,i})-1)}$. It can be generalized to a Dirichlet distribution to accommodate cases of more than two states. It motivates sampling of matrix rows $\boldsymbol{\pi_i} = [\pi_{i,i} \;\; \pi_{i,j}]$ and $\boldsymbol{\pi_j} = [\pi_{j,i} \;\; \pi_{j,j}]$ as follows:

$$\boldsymbol{\pi_i}|s_{[1,T]} \sim Dirichlet(H_{i,i}+\eta_{i,i}, H_{i,j}+\eta_{i,j}) \tag{4.12}$$
$$\boldsymbol{\pi_j}|s_{[1,T]} \sim Dirichlet(H_{j,j}+\eta_{j,j}, H_{j,i}+\eta_{j,i}).$$

At each new state observation, parameters $\eta_{i,j}$ are updated to reflect the cumulated history of transitions. As in Ardia (2009), we initiate hyperparameters at $H_{i,i}, H_{j,j} = 2$ and $H_{i,j}, H_{j,i=}1$ with the idea that observations are more likely to remain in the prevailing state than to switch into a new state.

*Step 2: Sample current state*

Chib (1996)[4] shows that a typical point in the joint density for states $p(\boldsymbol{s}|\boldsymbol{y},\boldsymbol{\theta}) = p(s_T|\boldsymbol{y},\boldsymbol{\theta}) \cdot \ldots \cdot p(s_t|\boldsymbol{y},S^{t+1},\boldsymbol{\theta}) \cdot \ldots \cdot p(s_1|\boldsymbol{y},S^2,\boldsymbol{\theta})$ can be expressed in terms of the probability mass function, the likelihood, and the state transition probabilities:

$$\begin{aligned} p(s_t|\boldsymbol{y},S^{t+1},\boldsymbol{\theta}) &\propto p(s_t|Y_t,\boldsymbol{\theta}) f(Y^{t+1},S^{t+1}|Y_t,s_t,\boldsymbol{\theta}) \\ &\propto p(s_t|Y_t,\boldsymbol{\theta}) p(s_{t+1}|s_t,\boldsymbol{\theta}) f(Y^{t+1},S^{t+2}|Y_t,s_t,s_{t+1},\boldsymbol{\theta}). \end{aligned} \tag{4.13}$$

Here $p(s_t|Y_t,\boldsymbol{\theta})$ is the probability mass function in the current state given information up to time $t$, and $p(s_{t+1}|s_t,\boldsymbol{\theta})$ is the probability of the new state at $t+1$ given the prior state at $t$. Since future observations of $y$ and states at $t+2$ and beyond are independent of the state at time $t$, the likelihood term $f(Y^{t+1},S^{t+2}|Y_t,s_t,s_{t+1},\boldsymbol{\theta})$ can be dropped above to obtain:

$$p(s_t|\boldsymbol{y},S^{t+1},\boldsymbol{\theta}) \propto p(s_t|Y_t,\boldsymbol{\theta}) p(s_{t+1}|s_t,\boldsymbol{\theta}). \tag{4.14}$$

[4]Remaining consistent with Chib's notation, capital letters with subscripts denote history up to indicated period, and capital letters with superscripts denote future through the remaining sample. For example, $S_t$ is the history of states to time $t$, and $S^{t+1}$ is the future procession of states through final observation $T$.

Chib (1996) further breaks down the probability mass function into its prior and likelihood, using Bayes theorem and the law of total probability:[5,6]

$$
\begin{aligned}
p(s_t|Y_t,\boldsymbol{\theta}) &\propto p(s_t|Y_{t-1},\boldsymbol{\theta})f(y_t|Y_{t-1},\boldsymbol{\theta}) \\
&\propto [\sum_{i=1}^{k} p(s_t|Y_{t-1},s_{t-1}=i,\boldsymbol{\theta})p(s_{t-1}=i|Y_{t-1},\boldsymbol{\theta})]f(y_t|Y_{t-1},\boldsymbol{\theta}) \\
&\propto [\sum_{i=1}^{k} p(s_t|s_{t-1}=i,\boldsymbol{\theta})p(s_{t-1}=i|Y_{t-1},\boldsymbol{\theta})]f(y_t|Y_{t-1},\boldsymbol{\theta}). \qquad (4.15)
\end{aligned}
$$

This is the transition probability $p(s_t|s_{t-1}=i,\boldsymbol{\theta})$ multiplied by the probability mass function from the prior observation $p(s_{t-1}=i|Y_{t-1},\boldsymbol{\theta})$, multiplied by the likelihood $f(y_t|Y_{t-1},\boldsymbol{\theta})$ of the current observation $y_t$, all given our parameter estimates made from the last observation at $t-1$. Current state $s_t$ is then sampled using this updated probability mass function $p(s_t|Y_t,\boldsymbol{\theta})$.

*Step 3: Sample* $\lambda_t$ *and* $\nu$

In this step, we follow the work of Jacquier, Polson and Rossi (2004) and Henneke, Rachev, Fabozzi and Nikolov (2011). For a prior on $\lambda_t$, the authors use a conjugate inverse-gamma density. This takes the form:

$$
p(\lambda_t|\nu) \propto \lambda_t^{-\nu/2-1} e^{-\nu/2\lambda_t} \sim \mathscr{IG}(\frac{\nu}{2},\frac{2}{\nu}) \sim \nu/\chi^2(\nu). \qquad (4.16)
$$

[5]Bayes' theorem states that the conditional probability of $y$ given $x$ is equal to the conditional probability of $x$ given $y$ times the marginal probability of $y$ and divided by the marginal probability of $x$, or $P(y|x) = \frac{P(x|y)P(y)}{P(x)}$.

[6]The law of total probability states that the marginal probability of $y$ is equal to the weighted average of the conditional probabilities of $y$ given $x_i$ times the marginal probability of $x_i$, or $P(y) = \sum_i^n P(y|x_i)P(x_i)$.

The conditional distribution of $\lambda_t$ is given by:

$$p(\boldsymbol{\lambda}|\boldsymbol{y},\boldsymbol{h},\boldsymbol{\theta},\nu) = \prod_{t=1}^{T} p(\lambda_t|y_t,h_t,\nu) \tag{4.17}$$

and can be re-written as:

$$p(\lambda_t|y_t,h_t,\nu) \equiv p(\lambda_t|\frac{y_t}{\sqrt{h_t}},\nu) \propto p(\lambda_t|\nu)p(\frac{y_t}{\sqrt{h_t}}|\lambda_t,\nu). \tag{4.18}$$

The conditional posterior becomes:

$$p(\lambda_t|h_t,y_t,\nu) \propto \lambda_t^{-(\nu+1)/2-1} e^{-(y_t^2/h_t+\nu)/2\lambda_t} \sim \mathscr{IG}(\frac{\nu+1}{2}, \frac{2}{\nu+y_t^2/h_t}). \tag{4.19}$$

Jacquier, Polson and Rossi (2004) then sample $\nu$ from this distribution. Alternatively we follow Henneke, Rachev, Fabozzi and Nikolov (2011) who make use of the relationship between inverse-gamma and chi-squared distributions. They show $(\frac{y_t^2}{h_t}+\nu)(\frac{1}{\lambda_t}) \sim \chi^2(\nu+1)$, sampling $x_t \sim \chi^2(\nu+1)$ and calculating $\lambda_t = (\frac{y_t^2}{h_t}+\nu)/x_t$.

The posterior distrbution of $\nu$ is proportional to the product of *t*-distributional ordinates:

$$p(\nu|\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{y}) \propto p(\nu)p(\boldsymbol{y}|\boldsymbol{h},\nu) = p(\nu)\prod_{t=1}^{T} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}(1+\frac{y_t^2}{h_t\nu})^{-\frac{(\nu+1)}{2}}. \tag{4.20}$$

Using a flat prior and restricting $\nu$ within the discrete range $\{3,...,40\}$, the posterior distribution of $\nu$ can then be sampled from:

$$p(\nu|\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{y}) = \frac{1}{37}\frac{\prod_{t=1}^{T}\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}(1+\frac{y_t^2}{h_t\nu})^{-\frac{(\nu+1)}{2}}}{\prod_{t=1}^{T}\sum_{\nu=3}^{40}\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}(1+\frac{y_t^2}{h_t\nu})^{-\frac{(\nu+1)}{2}}}. \tag{4.21}$$

*Step 4: Sample conditional variance parameters given current state*

We employ a technique known as "data augmentation" seen elsewhere in the literature for latent class models. In this method, state observations $s_{[1,T]}$ are included as a block to be estimated within the parameter set $\boldsymbol{\theta} = \{\alpha_0, \alpha, \gamma, \beta, \boldsymbol{s}\}$. Conventional sampling methods draw from each parameter's distribution, conditional on the current values of the data and all remaining parameters. However the full conditional distribution of the state $p(s_t|\boldsymbol{y},\boldsymbol{s}_{j\neq t},\boldsymbol{\theta})$ may not be tractable across a multiple-state space because it requires knowledge of the entire sequence of past as well as future states. The data augmentation method instead allows each state to be drawn from the joint distribution of states $p(s_1, s_2, \ldots, s_T|\boldsymbol{y},\boldsymbol{\theta})$ rather than the full conditional distribution. Crucially, this means that states can be simulated with knowledge only of the prior state and transition probabilities into the next state, without necessarily knowing the entire sequence of past and future states. Furthermore,

the states can be treated collectively as a single block so that only one additional block is required in the sampler. The sampler then looks something like this, starting with $j = 1$:

$$
\begin{aligned}
s_j &\sim p(\boldsymbol{s}|\boldsymbol{y}, \alpha_0^{[j-1]}, \alpha^{[j-1]}, \gamma^{[j-1]}, \beta^{[j-1]}) \\
\alpha_0^{[j]} &\sim p(\alpha_0|\boldsymbol{y}, s_j, \alpha^{[j-1]}, \gamma^{[j-1]}, \beta^{[j-1]}) \\
\alpha^{[j]} &\sim p(\alpha|\boldsymbol{y}, s_j, \alpha_0^{[j]}, \gamma^{[j-1]}, \beta^{[j-1]}) \\
\gamma^{[j]} &\sim p(\gamma|\boldsymbol{y}, s_j, \alpha_0^{[j]}, \alpha^{[j]}, \beta^{[j-1]}) \\
\beta^{[j]} &\sim p(\beta|\boldsymbol{y}, s_j, \alpha_0^{[j]}, \alpha^{[j]}, \gamma^{[j]})
\end{aligned}
\tag{4.22}
$$

looping back to set $j = j + 1$, and so on through all $T$ observations. The current state is sampled based upon the last observation's parameter values. Given the current state $s_1$, parameter $\alpha_0^1$ is estimated using previous values $\alpha^0$,$\gamma^0$,$\beta^0$. After $\alpha_0^1$ is simulated and updated, it is used together with $\gamma^0$ and $\beta^0$ to estimate $\alpha^1$. Current values of $\alpha_0^1$,$\alpha^1$ and the previous $\beta^0$ value are used to estimate $\gamma^1$. Finally all three updated parameters are used in estimating $\beta^1$ to complete the simulation cycle for the current observation. It then starts over again with transition probabilities for the next state observation, and parameters are estimated in like fashion conditional upon the new sampled state. At the end of the full sample we have estimates for each parameter conditional upon which state we are in.

## 4.5. Analysis

In Monte Carlo simulations of 10,000 draws, we discard the first 1,000 draws for burn-in purposes. Parameters are estimated as the posterior means of simulated chains, so it is important to look for evidence that convergence in the chains has been achieved.[7] We check for autocorrelation in the parameter draws, which can lead to convergence delays due to slow movement around the parameter space. These all decay quickly. Trace plots of iteration number versus parameter value at each draw show that the simulated chains explore the parameter space well. Running-mean plots of iterations against the means of draws summed up to that iteration also indicate good mixing and appear to converge quickly. We inspect Geweke diagnostic results for the parameter chains, which take two non-overlapping parts of each chain and compare means in each part to test the null hypothesis that they are both taken from the same distribution. The resulting statistics are below critical values and fail to reject this null at the 5% significance level, suggesting that convergence in the chains is likely. Gelman-Rubin diagnostic tests examining parallel chains with dispersed initial values are also run to see if they converge to the same target distribution. These tests compare the variance within chains to the variance between chains, with large discrepancies indicative of a failure to converge properly. The resulting potential scale reduction factors are all close to 1, suggesting that convergence is achieved and that our results are not sensitive to initial priors.

Table 4.1 gives summary statistics for the sample period. Time series of the S&P 500 log-prices and log-returns for years 1871-2013 are shown in Figure 4.1. We estimate models up to four

[7] We use R package *coda* to assist in posterior analysis.

**Table 4.1**
**S&P 500 log-return summary statistics**

| | |
|---|---|
| Observations | 1704 |
| Mean | -4.473e-17 |
| Standard deviation | 4.101 |
| Skewness | -0.4236 |
| Kurtosis | 11.168 |

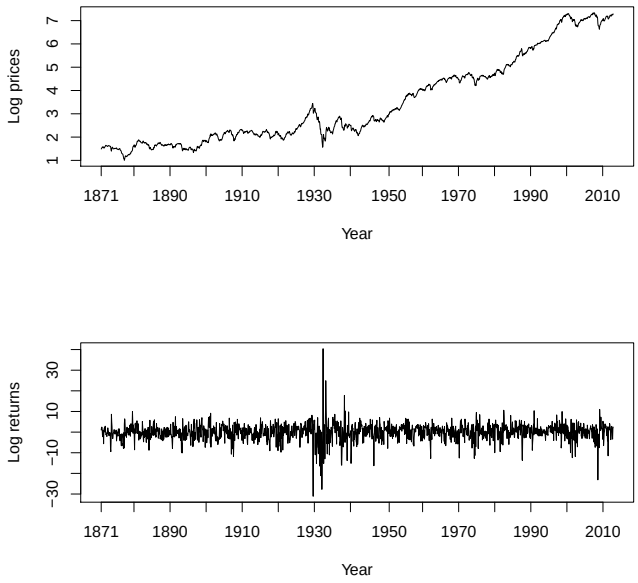


Figure 4.1: S&P 500 index log-prices and log-returns

volatility regimes, but do not find support for a distinct fourth regime. Parameter estimates from the first three models are given in Tables 4.2-4.4 below.

The main difference between volatility regimes in our sample appears to be driven by major structural shifts in the baseline conditional variance parameter, $\alpha_0$. We find evidence of statistically significant shifts in this baseline parameter in both two- and three-regime models. We also

**Table 4.2**
**Single-regime parameter estimates**

| Parameter | Mean | Median | 95% HPDI |
|---|---|---|---|
| $\alpha_0$ | 1.196 | 1.185 | [1.152,1.265] |
| $\alpha$ | 0.1349 | 0.1118 | [0.0607,0.2723] |
| $\gamma$ | 0.0004 | -3.036e-05 | [-0.0051,0.0054] |
| $\beta$ | 0.8408 | 0.8618 | [0.7145,0.9106] |

**Table 4.3**
**Two-regime parameter estimates**

| Parameter | Mean | Median | 95% HPDI |
|---|---|---|---|
| $\alpha_{01}$ | 0.5404 | 0.5427 | [0.5174,0.5507] |
| $\alpha_1$ | 0.0658 | 0.0596 | [0.0375,0.1239] |
| $\gamma_1$ | 0.0012 | 0.0005 | [-0.0050,0.0056] |
| $\beta_1$ | 0.9117 | 0.9171 | [0.8420,0.9429] |
| $\alpha_{02}$ | 1.682 | 1.682 | [1.633,1.771] |
| $\alpha_2$ | 0.0718 | 0.0539 | [0.0382,0.1246] |
| $\gamma_2$ | 0.0010 | 0.0000 | [-0.0043,0.0044] |
| $\beta_2$ | 0.9049 | 0.9214 | [0.8322,0.9403] |

note mean-reverting behavior for $\alpha_0$ when estimating fewer regimes. For example,[8] the single-regime baseline parameter $\alpha_0^{(1)} = 1.196$ falls between the low and high volatility state parameters $\alpha_{01}^{(2)} = 0.5404$ and $\alpha_{02}^{(2)} = 1.682$ from the two-regime model. Similarly, the lower volatility state estimate $\alpha_{01}^{(2)} = 0.5404$ from the two-regime model falls between the low and medium volatility states $\alpha_{01}^{(3)} = 0.3965$ and $\alpha_{02}^{(3)} = 0.8577$ identified in the three-regime model. However, the high volatility state estimated in the two- and three-regime models is effectively identical. Thus the two-regime model here appears to be identifying a high state of volatility that is otherwise averaged out by the single-regime model. The three-regime model identifies this same high volatility state, but also partitions the low volatility state more finely into distinct low and medium volatility states.

[8] Parameter subscripts are used here to distinguish regimes within a model, and parameter superscripts are used to denote the total number of regimes in that particular model.

**Table 4.4**
**Three-regime parameter estimates**

| Parameter | Mean | Median | 95% HPDI |
|---|---|---|---|
| $\alpha_{01}$ | 0.3965 | 0.3948 | [0.3748,0.4140] |
| $\alpha_1$ | 0.0735 | 0.0650 | [0.0452,0.1528] |
| $\gamma_1$ | 0.0013 | 0.0002 | [-0.0063,0.0068] |
| $\beta_1$ | 0.9079 | 0.9164 | [0.8090,0.9376] |
| $\alpha_{02}$ | 0.8577 | 0.8343 | [0.7497,0.9752] |
| $\alpha_2$ | 0.0935 | 0.0668 | [0.0495,0.1661] |
| $\gamma_2$ | 0.0028 | 0.0004 | [-0.0094,0.0095] |
| $\beta_2$ | 0.8936 | 0.9166 | [0.8173,0.9366] |
| $\alpha_{03}$ | 1.576 | 1.557 | [1.532,1.678] |
| $\alpha_3$ | 0.0667 | 0.0518 | [0.0349,0.1064] |
| $\gamma_3$ | 0.0020 | 0.0001 | [-0.0042,0.0040] |
| $\beta_3$ | 0.9134 | 0.9259 | [0.8473,0.9459] |

In all three models, the memory parameter $\beta$ for prior conditional variances is greater than the memory parameter $\alpha$ for prior shocks. This suggests that the overall persistence in the conditional variance process may be more largely driven by a persistence in prior variances than by lingering specific shocks. Parameters $\alpha_0$, $\alpha$, and $\beta$ are statistically significant in all three models, but the high posterior density interval for the leverage parameter $\gamma$ includes zero. This indicates that none of the models are picking up on difference in the volatility response to positive versus negative shocks.

To compare goodness-of-fit between the three models, we calculate the deviance information criterion (DIC) from Spiegelhalter, Best, Carline and van der Linde (2002). The DIC can be viewed as a Bayesian counterpart to maximum likelihood measures of fit. It compares models based upon the output from Bayesian esimation, taking both fit and complexity into account:

$$DIC = D(\overline{\boldsymbol{\theta}}) + 2p_D = \overline{D} + p_D. \tag{4.23}$$

$\overline{D}$ can be thought of as a measure of fit, capturing the residual information in the data that is left over from the model estimation. $p_D$ can be thought of as a penalty for model complexity. As with maximum likelihood measures of fit, a lower DIC value is indicative of model preference.

More specifically, $\overline{D} = E[-2lnf(y|\boldsymbol{\theta})]$ is defined as the posterior expectation of the deviance. It can be interpreted as how much the model estimated at each observation deviates from the actual data with no model at all. The lower this value, the less residual information is left over from the model estimation and the more closely the model fits the actual data. $p_D = \overline{D} - D(\overline{\boldsymbol{\theta}}) = E_{\boldsymbol{\theta}|y}[D] - D(E_{\boldsymbol{\theta}|y}[\boldsymbol{\theta}])$ is the posterior expectation of the deviance minus the deviance evaluated at the posterior parameter means. It serves as a kind of penalty for model complexity. Including additional parameters will automatically decrease the deviance evaluated at the posterior parameter means, resulting in a higher penalty. This will be weighed against the increased explanatory power. The values are calculated as by-products from our models by using the estimated likelihood functions: $lnf(y|\boldsymbol{\theta})$ is evaluated with parameters estimated at each observation, and $lnf(y|\overline{\boldsymbol{\theta}})$ is estimated with the posterior parameter means.

To add further meaning to these comparisons, we also use a resampling procedure described in Ardia (2009) to estimate confidence intervals for the DIC estimates. This involves allowing random perturbations in the model parameters and recalculating samples of new DIC values in order to estimate an entire distribution. Results in Table 4.5 show that the three-regime model

**Table 4.5**
**Deviance information criterion**

| Model | DIC | 95% confidence interval |
|---|---|---|
| One state | 9407.910 | [9392.697,9423.123] |
| Two states | 8727.575 | [8711.442,8744.708] |
| Three states | 8339.275 | [8324.188,8354.362] |

is statistically favored, lying outside of 95% confidence intervals for the single- and two-regime models.

Figure 4.2 shows smoothed probabilities of each of the three states for our sample period. These smoothed probabilities are taken from the estimated probability mass function and updated using a forward-backward algorithm. The probability mass function only incorporates past information up to time $t$, but the smoothing step allows inferences to be made for each observation based upon the full information in the entire sample. Once the probability mass function is estimated, state probabilities are smoothed by recursively moving backwards through observations $T, T-1, ..., 1$ according to:

$$p(s_t|y_T) = p(s_t|y_t)[\sum_{s_{t+1}=i}^{k} p(s_{t+1}|s_t)\frac{p(s_{t+1}|y_T)}{p(s_{t+1}|y_t)}]. \tag{4.24}$$

Intuitively, if information from the full sample does not add any new information to the existing probability mass estimates, then $\frac{p(s_{t+1}|y_T)}{p(s_{t+1}|y_t)} = 1$. In this case, smoothed probabilities will be identical to the original estimates: $p(s_t|y_T) = p(s_t|y_t)$. However, if it does provide new information, the original probability estimates are then updated.

It is interesting to see how the model matches up with well-known historical events. The time series of State 3 probabilities in Figure 4.2 represents the probability of being in the highest volatility state during our sample period. We first observe a spike in State 3 probabilities around

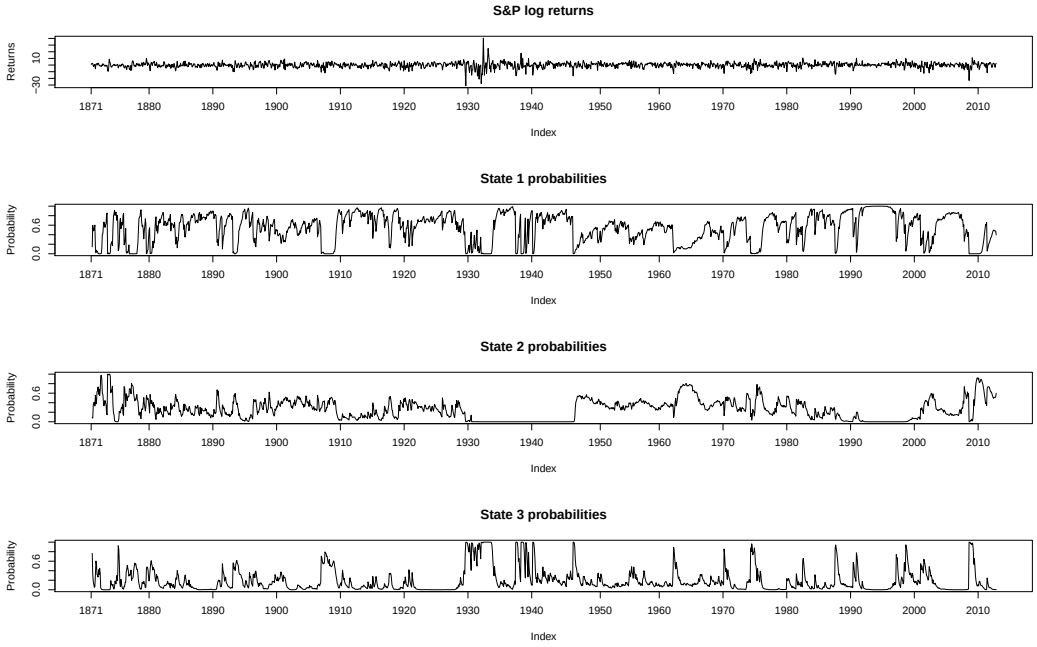


Figure 4.2: S&P 500 index and smoothed probabilities for three volatility states

the time of the Panic of 1873. More subtle increases in State 3 probabilities can be seen around the time of the Great Railroad Strike of 1877 and coinciding with the assassination of President James Garfield in 1880. There are other mild increases in State 3 probabilities seen around 1884 and 1890; Wicker (2000) identifies two banking disturbances at these times that were contained much better than the panic from 1873. We see another increase in State 3 probabilities around the time of the Banking Panic of 1893 and some mild increases through the depression that followed. In the 20th century, we see further evidence of a sustained increase in State 3 probabilities around the Panic of 1907. The most prominent spike appears around the stock market crash of 1929, and this period of instability continues through the mid-1930s. The series also displays high State 3 probabilities around the contraction in bank lending that occurred in the late 1930s as the Federal Reserve tightened its policy. The next spike appears around 1946, which Mishkin and White

(2002) identify as a market crash that occurred as post-war demand fell. Mishkin and White (2002) identify another crash in 1962 around the time of the Bay of Pigs and the Cuban Missle Crisis, seen here with a corresponding spike in State 3 probabilities. There are further upward spikes in probability around 1970 at the time of a mild U.S. recession and in the mid-1970s around the time of the oil embargo. We also see a mild increase in State 3 probabilities after the brief recession that occurred around 1982, and then another major spike around the 1987 stock market crash. We observe increased State 3 probabilities in the early 1990s, a period encompassing a recession as well as the first U.S. war in Iraq. High State 3 probabilities appear again in the late 1990s at a time of market uncertainty stemmimg from currency crises in Latin America and Asia, Russia's sovereign default, and the collapse of Long Term Capital Management. State 3 probabilities increase again in the early 2000s around the bursting of the dot-com bubble, the September 11th terrorist attacks, Argentina's economic collapse, and yet another U.S. recession. We see a final major spike in State 3 probabilities within our sample at the time of the recent financial crisis.

## 4.6. Conclusion

In this paper we find evidence to support the existing literature on Markov switching models of conditional variance in financial returns. We examine S&P 500 log-returns over a long duration from 1871-2013 to gain some perspective on the nature of volatility switching behavior through U.S. history. We estimate single-, two-, and three-regime GARCH models and compare their fit of the data. The three-regime model is clearly favored for our sample period. The models identify

regime-switching behavior here as being driven by major structural shifts in the baseline conditional variance. We also find evidence of a kind of mean-reversion for these separate parameters when estimating models of fewer regimes. This supports the idea that conventional single-regime models may be too blunt in their description of market behavior and are failing to recognize distinct dynamics for asset volatility. We compare estimated probabilities of being in the highest volatility state with historical financial disturbances and find that they appear to match up well.

# Bibliography

Y. Aït-Sahalia and J. Duarte. “Nonparametric Option Pricing under Shape Restrictions”. *The Journal of Econometrics*, 116:9-47, Sept.-Oct. 2003.

Y. Aït-Sahalia and A.W. Lo. “Nonparametric Estimation of State Price Densities Implicit in Financial Asset Prices”. *The Journal of Finance*, 53(2):499-547, Apr. 1998.

Y. Aït-Sahalia and A.W. Lo. “Nonparametric Risk Management and Implied Risk Aversion”. *The Journal of Econometrics*, 94(1-2):9-51, Jan. 2000.

Y. Aït-Sahalia, Y. Wang, and F. Yared. “Do Options Markets Correctly Price the Probabilities of the Underlying Asset?”. *The Journal of Econometrics*, 102(1):67-110, May 2001.

C. Alexander and E. Lazar. “Markov Switching GARCH Diffusion”. *ICMA Centre Discussion Papers in Finance*, 1: Mar. 2008. URL `http://www.icmacentre.ac.uk/files/discussion-papers/DP2008_01.pdf`.

D. Ardia. “Bayesian Estimation of a Markov-Switching Threshold Asymmetric GARCH Model with Student-t Innovations”. *The Econometrics Journal*, 12(1):105-126, Mar. 2009.

K. Arrow. “The Role of Securities in the Optimal Allocation of Risk-Bearing”. *The Review of Economic Studies*, 31(2):91-96, Apr. 1964.

G. Barone-Adesi, R. F. Engle, and L. Mancini. “A GARCH Option Pricing Model with Filtered Historical Simulation”. *Review of Financial Studies*, 21(3):1223-1258, May 2008.

K. S. Bartunek and M. Chowdhury. “Implied Risk Aversion Parameter from Options Prices”. *Financial Review*, 32(1):107-124, Feb. 1997.

D. S. Bates. “Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options”. *The Review of Financial Studies*, 9(1):69-107, Spring 1996.

D. S. Bates. “Post-’87 Crash Fears in the S&P 500 Futures Option Market”. *Journal of Econometrics*, 94(1-2):181-238, Jan. 2000.

F. Black and M. Scholes. “The Pricing of Options and Corporate Liabilities”. *The Journal of Political Economy*, 81(3):637-654, May-Jun. 1973.

R. R. Bliss and N. Panigirtzoglou. “Option-Implied Risk Aversion Estimates”. *The Journal of Finance*, 59(1):407-446, Feb. 2004.

T. Bollerslev. “Generalized Autoregressive Conditional Heteroskedasticity”. *Journal of Econometrics*, 31:307-327, Apr. 1986.

O. Bondarenko. “Why are Put Options so Expensive?”. *AFA 2004 San Diego Meetings*, Apr. 2003.

D. T. Breeden and R. H. Litzenberger. “Prices of State-Contingent Claims Implicit in Options Prices”. *The Journal of Business*, 51(4):621-651, Oct. 1978.

J. Cai. “A Markov Model of Switching-Regime ARCH”. *The Journal of Business and Economics Statistics*, 12(3):309-316, July 1994.

S. Chib. “Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models”. *Journal of Econometrics*, 75:79-97, Nov. 1996.

S. Chib and E. Greenberg. “Understanding the Metropolis-Hastings Algorithm”. *The American Statistician*, 49(4):327-335, Nov. 1995.

J. Cochrane. *Asset Pricing*. Revised Edition, Princeton University Press, 2005.

G. M. Constantinides, J. C. Jackwerth, and S. Perrakis. “Mispricing of S&P 500 Index Options”. *Review of Financial Studies*, 22(3):1247-1277, Mar. 2009.

J. Danielsson, H. S. Shin, and J. Zigrand. “Risk Appetite and Endogenous Risk”. AXA working paper series no. 2, discussion paper no. 647, Feb. 2010. URL `http://www.imf.org/external/pubs/ft/wp/2008/wp0885.pdf`.

G. Debreu. *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. Yale University Press, Monograph 17, 1959.

S. Edwards and R. Susmel. “Volatility Dependence and Contagion in Emerging Equity Markets”. *Journal of Development Economics*, 66(2):505-532, Dec. 2001.

R. Engle. “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation”. *Econometrica*, 50(4):987-1007, July 1982.

J. Fan and I. Gijbels. *Local Polynomial Smoothing and Its Applications*. Chapman and Hall, 1996.

S. Figlewski. “Estimating the Implied Risk Neutral Density for the U.S. Market Portfolio”. *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle* (eds. Tim Bollerslev, Jeffrey R. Russell and Mark Watson), Oxford University Press, 2008.

J. Fouque, G. Papanicolaou, and K. R. Sircar. *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge University Press, 2000.

S. Früwirth-Schnatter. “Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models”. *Journal of the American Statistical Association*, 96(453):194-209, 2001.

P. Gai and N. Vause. (2006), “Measuring Investors’ Risk Appetite”. *International Journal of Central Banking*, 167-188, Mar. 2006.

I. V. Girsanov. “On Transforming a Certain Class of Stochastic Processes by Absolutely Continuous Substitution Measures”. *Theory of Probability and Its Applications*, 5(3):285-301, 1960.

N. Gisiger. “Risk-Neutral Densities Explained”. 2010. URL `http://ssrn.com/abstract=1395390`.

E. Glatzer and M. Scheicher. “Modelling the Implied Probability of Stock Market Movements”. European Central Bank working paper no. 212, Jan. 2003. URL `http://www.imf.org/external/pubs/ft/wp/2008/wp0885.pdf`.

L. R. Glosten, R. Jagannathan, and D. E. Runkle. “On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks”. *The Journal of Finance*, 48(5):1779-1801, Dec. 1993.

B. González-Hermosillo. “Investors’ Risk Appetite and Global Financial Market Conditions”. IMF Working Paper WP/08/85, Apr. 2008. URL `http://www.imf.org/external/pubs/ft/wp/2008/wp0885.pdf`.

M. Haas, S. Mittnik, and M. S. Paolella. “A New Approach to Markov-Switching GARCH Models”. *Journal of Financial Econometrics*, 2(4):493-530, Fall 2004.

J. D. Hamilton and R. Susmel. “Autoregressive Conditional Heteroskedasticity and Changes in Regime”. *Journal of Econometrics*, 64:307-333, Sept.-Oct. 1994.

B. Han. “Investor Sentiment and Options Prices”. *The Review of Financial Studies*, 21(1):387-414, Jan. 2008.

W. K. Hastings. “Monte Carlo Sampling Methods Using Markov Chains and their Applications ”. *Biometrika*, 57(1):97-109, Apr. 1970.

J. S. Henneke, S. T. Rachev, F. J. Fabozzi, and M. Nikolov. “MCMC-Based Estimation of Markov Switching ARMA-GARCH Models”. *Applied Economics*, 43(3):259-271, 2011.

S. Heston. “A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options”. *The Review of Financial Studies*, 6(2):327-343, Summer 1993.

A. J. Hou. “EMU Equity Markets’ Return Variance and Spill Over Effects from Short-Term Interest Rates”. *Quantitative Finance*, 13(3):451-470, 2013.

J. C. Jackwerth. “Recovering Risk Aversion from Options Prices and Realized Returns”. *Review of Financial Studies*, 13(2):433-451, Summer 2000.

J. C. Jackwerth. *Option-Implied Risk Neutral Distributions and Options Prices*. Research Foundation of AIMR, 2004.

J. C. Jackwerth and M. Rubinstein. “Recovering Probability Distributions from Options Prices ”. *The Journal of Finance*, 51(5):1611-1631, Dec. 1996.

E. Jacquier, N. G. Polson, and P. E. Rossi. “Bayesian Analysis of Stochastic Volatility Models with Fat-Tails and Correlated Errors”. *Journal of Econometrics*, 122:185-212, Sept. 2004.

C. J. Kim. “Dynamic Linear Models with Markov-Switching”. *Journal of Econometrics*, 60:1-22, Jan.-Feb. 1994.

W. S. Krasker. “The ‘Peso Problem’ in Testing the Efficiency of Foreign Exchange Markets”. *Journal of Monetary Economics*, 6(2):269-276, Apr. 1980.

C. G. Lamoureux and W. D. Lastrapes. “Persistence in Variance, Structural Change, and the GARCH Model”. *Journal of Business and Economic Statistics*, 8(2):225-234, Apr. 1990.

A. Macrina and P. A. Parbhoo. “Randomised Mixture Models for Pricing Kernels”. arXiv preprint, Dec. 2011. URL `http://arxiv.org/pdf/1112.2059.pdf`.

F. Mainardi and S. Rogosin. “The Origin of Infinitely Divisible Distributions: from de Finetti’s Problem to Lévy-Khintchine Formula”. arXiv preprint, Jan. 2008. URL http://arxiv.org/pdf/0801.1910v1.pdf.

M. C. Merton. “Option Pricing When Underlying Stock Returns are Discontinuous”. *Journal of Financial Economics*, 3(1-2):125-144, Jan.-Mar. 1967.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, and A.H. Teller. “Equation of State Calculations by Fast Computing Machines ”. *The Journal of Chemical Physics*, 21(6):1087-1092, June 1953.

F. S. Mishkin and E. N. White. “U.S. Stock Market Crashes and their Aftermath: Implications for Monetary Policy”. NBER working paper series, no. w8992, June 2002. URL http://www.nber.org/papers/w8992.

J. P. Peters. “Estimating and Forecasting Volatility of Stock Market Indices Using Asymmetric GARCH Models and (Skewed) Student-t Densities”. preprint, University of Liege, Belgium, Mar. 2001.

K. Rogoff. “Rational Expectations in the Foreign Exchange Market Revisited”. unpublished manuscript, MIT, Feb. 1977.

J. V. Rosenberg and C. F. Engle. “Empirical Pricing Kernels”. *Journal of Financial Economics*, 64(3):341-372, June 2002.

D. Ruppert, S. J. Sheather, and M. P. Wand. “An Effective Bandwidth Selector for Local Least Squares Regression”. *Journal of the American Statistical Association*, 90(432):1257-1270, Dec. 1995.

M. Scheicher. “What Drives Investor Risk Aversion? Daily Evidence from the German Equity Market”. *BIS Quarterly Review*, 67-74, June 2003.

R. J. Shiller. *Market Volatility*. The MIT Press, 1989.

R. J. Shiller. *Irrational Exuberance*. Princeton University Press, 2000.

D. Shimko. “Bounds of Probability”. *Risk*, 6(4):33-36, Apr. 1993.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. “Bayesian Measures of Model Complexity and Fit”. *Journal of the Royal Statistical Society B*, 64(4):583-639, Oct. 2002.

N. Tarashev, K. Tsatsaronis and D. Karampatos. “Investors’ Attitude Towards Risk: What Can We Learn from Options?”. *BIS Quarterly Review*, 57-65, June 2002.

N. Tarashev and K. Tsatsaronis. “Risk Premia Across Markets: Information from Option Prices”. *BIS Quarterly Review*, 93-103, Mar. 2006.

E. Wicker. *Banking Panics of the Gilded Age*. Cambridge University Press, 2000.

A. Yatchew and W. Härdle. “Nonparametric State Price Density Estimation Using Constrained Least Squares and the Bootstrap”. *Journal of Econometrics*, 133(2):579-599, Aug. 2006.

C. L. Yu, H. Li, and M. T. Wells. “A Bayesian Analysis of Return Dynamics with Lévy Jumps”. *Review of Financial Studies*, 21(5):2345-2378, Sept. 2008.

C. L. Yu, H. Li, and M. T. Wells. “MCMC Estimation of Lévy Jump Models Using Stock and Option Prices”. *Mathematical Finance*, 21(3):383-422, 2011.