Risk Preferences Under Changing Volatility Regimes – A Markov Switching Approach to Pricing Kernel Estimation[*]

Emily Johnston Ross [†]

October 2012

Abstract

Since the 2008 meltdown in financial markets, the role of risk appetite as a factor in market vulnerability has received increasing attention. This paper contributes to this area of research by examining risk preference behavior in financial markets. We use a modified approach to pricing kernel estimation that addresses known conceptual difficulties in the pricing kernel ratio. Evidence shows that introducing regime-switching into the conditional variance of the underlying returns better reflects this period of extreme volatility. It also appears to have a significant impact on characterizations of pricing kernels and risk preference behavior. Results indicate that conventional single-regime models may mistakenly identify changes in risk preferences by failing to account for possible regime changes in conditional variance over time.

[†]Ph.D. candidate at The Graduate Center, City University of New York, Department of Economics, 365 Fifth Avenue New York, NY 10016, phone: 212-817-8255, e-mail: ejohnston@gc.cuny.edu.

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 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 must vary as well. This would imply in some cases that assumptions of constant relative risk aversion in models of investor behavior may be inappropriate. Yet uncovering the true nature of risk preferences is tricky as preferences themselves are not directly observable. Instead, preferences for risk 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 methodologies 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 comparing state-price densities and state probability densities. These densities can be calculated from observed options prices and from the observed returns on the underlying asset. Variations of this strategy can be seen throughout the existing literature[1].

But there are questions as to whether the changes captured in these estimates can truly be interpreted as preference-driven. State-price densities are typically estimated non- or semi-parametrically from cross-sections of options data[2]. The range of strike prices existing on a single option allows estimation of a full probability density that can vary over time. But unlike options, only a single price at a time exists for the underlying asset. There-

---

[1]Bartunek and Chowdhury (1997), Jackwerth (2000), and Gai and Vause (2006), for example.

[2]See Jackwerth (2004) for discussion.

fore underlying returns densities are typically estimated from an aggregated time-series of returns data. One common strategy[3] uses GARCH methods to allow for stochastic volatility dynamics in the underlying returns. Bliss and Panigirtzoglou (2004) point out however that there is nevertheless an implicit stationarity in the parameters of the underlying GARCH process. In some cases this may mean that the time-varying state-price density is compared with a series of average state probability densities. In others it may mean forecasting from a fixed-parameter process assumed to hold constant for long periods of time.

From a conceptual viewpoint it is problematic that the state-price densities would vary with time but the underlying returns densities not. The danger here is that all of the changes in the state-price to state probability ratios may be interpreted as changes in risk preferences, when to some degree they could be picking up on changes in the underlying returns process itself. The issue of nonstationarity in underlying returns calls into question the reliability of preference-driven interpretations for these results. This has yet to be resolved in the literature.

In contribution to this literature, we estimate time-varying empirical pricing kernels for capturing preferences towards risk. We do this in the form of ratios of state-price densities to state probability densities. We modify common assumptions for these density estimations to see how it may alter pricing kernel estimates and behaviors. We pose the question: Does nonstationarity in the underlying returns process affect common preference-driven explanations for pricing kernel behavior?

For state-price density estimation we employ a shape-constrained least squares estimator along the lines of Aït-Sahalia and Duarte (2003) and Yatchew and Härdle (2006). This appropriately fits the relevant second derivative of the option-pricing function, not just the smoothed observa-

---

[3]For example, in Scheicher (2003), Tarashev, Tsatsaronis and Karampatos (2003), Tarashev and Karampatos (2006) and Gai and Vause (2006).

tions. Also, favorable small-sample properties ensure that sensitivity is not averaged out. For state probability density estimation, the critical intent here is to modify assumptions of stationarity in the underlying returns process but without needing to overly restrict risk preference itself. We preserve a GARCH-like structure in which volatility can be interpreted as an error-driven process. However, we additionally introduce elements of nonstationarity in the form of Markov switching in the underlying conditional variance process.

The data for this paper consist of options on the S&P 500 index as well as the prices on the S&P 500 index itself. The sample extends from 2005 to 2011, encompassing the lead-up, duration and aftermath to the 2008 crisis. Pricing kernels are estimated on a monthly basis in order to maintain a constant time-to-maturity for options contracts. Single-regime and Markov switching models are then compared using goodness-of-fit and posterior predictive performance. After comparing fit, we investigate how pricing kernel behavior differs under our new assumptions. Results show that the implications for risk preference behavior differ rather significantly from those of conventional estimation strategies.

The goal of this paper is to allow nonstationarity in the underlying returns process and see how it influences findings for risk preference behavior, especially surrounding periods of financial turmoil. Stationary conditional variance specifications can only capture a kind of average process, but by introducing high and low volatility states into the conditional variance this paper introduces greater sensitivity in measuring pricing kernels across periods of stability and distress. We hope that this will help in the estimation and interpretation of empirical pricing kernels, contributing to a more complete picture of how markets may react in times of distress. Understanding more about the empirical characteristics of risk preferences would be valuable for modeling and understanding market behavior. What is more, the behavior of risk preferences 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[4]. Thus achieving a better understanding of the empirical characteristics of investor risk preferences is becoming increasingly important for matters of financial market and economic stability.

2   Conceptual background

An advantage of the options-derived pricing kernel 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. The options-derived measure rather allows examination of preferences extracted from the pricing of marginal trades in the market itself. Therefore, the interpretation is emphasized throughout this paper of a marginal investor conducting trades in the market. It may be after all that the risk preference of the marginal investor in markets behaves somewhat differently than what would be assumed is reasonable for the individual investor. Risk preferences of the marginal investor might change over time because preferences of individual investors have changed, for example, or alternatively because the composition of investors in the market with particular levels of risk preference has changed.

Options-derived empirical 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 subsequent periods for consumption that is certain. Along with the expected payoff x_{t+1}, the pricing kernel m_{t+1} helps form the central pricing equation in asset pricing theory: $p_t = E_t[m_{t+1}x_{t+1}]$.

In the state-preference model of Arrow (1964) and Debreu (1959), con-

---

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

tingent claims are 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-price probabilities weighted by security payout x in state s. Time preference can be incorporated by discounting the probability-weighted payouts by $1/(1+r)$, while multiplying state-prices $p^*(s)$ by $(1+r)$ in order to maintain the sum of probabilities to 1. Transforming then by multiplying numerator and denominator by underlying state probabilities $\phi(s)$ gives a representation of the pricing kernel within the basic pricing equation framework:

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

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 risk preference, or the more risk-averse the investor. This 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. What differentiates the state-price and state probability densities is the effect that investor risk preferences have on the valuation of these state-contingent payouts. Pricing kernels are

reflective of risk preferences in the extent to which the preference-weighted state-price densities differ from expected state probability densities.

3   Methodology

Having outlined the theory behind pricing kernel ratios of state-price to state probability densities, these ratio components must now be estimated. For the state-price density, we use a shape-constrained least squares estimator with the options price data. For the state probability density, the intent within this paper is to modify assumptions of stationarity in the density of the underlying returns. A GARCH-like structure is preserved in which volatility can be interpreted as an error-driven process. GARCH models additionally are flexible in their ability to accommodate observed asset return characteristics. However, we introduce nonstationarity in the form of Markov switching in the conditional variance of the underlying returns. This section briefly surveys some of the relevant literature to explain our methodological reasoning.

3.1   Choice of state-price density estimator

Breeden and Litzenberger (1978) showed that risk-neutral densities can be derived from options prices by taking the second derivative of the call option-pricing function with respect to strike price. This can be done in practice without prior assumptions of distributional form by estimating and smoothing an implied volatility function for observed discrete options with identical maturity dates. 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 of the underlying asset. Shimko (1993) implements this approach by fitting an implied volatility curve to options data, finding analytic expressions for

the distribution of the underlying asset, and using these expressions to numerically evaluate the moments of the distribution. Others such as Bliss and Panigirtzoglou (2004), Glatzer and Scheicher (2005), and Gai and Vause (2006) expand on this method using varying interpolation techniques. In another recent paper, Figlewski (2008) develops an approach incorporating modified smoothing techniques with a fourth-order polynomial and Generalized Extreme Value tails.

In contrast Aït-Sahalia and Lo (1998) estimate a nonparametric kernel regression fitting realized options price data. The state-price density then is found by taking the second derivative of the resulting estimator. This method requires no functional pre-suppositions about options prices and volatility, nor does it require assumptions of a prior distribution. It does require that certain assumptions be made regarding smoothness criteria, and it can also be rather 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 implied volatility alone. Even with this reduced adjustment, they aggregate over a yearly timeframe 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 example, convergence can be slow especially for higher-order derivatives, so large sample sizes tend to be required. In addition the authors point out that estimating the pricing function with options prices across time in order to achieve large samples introduces potential nonstationarity or regime changes into the data, not to

mention 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 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 these estimators display plausible excess skewness and kurtosis in the resulting densities. They also satisfy constraints for non-negativity, a requirement sometimes violated through other approaches. Simulations show that the constrained nonparametric method produces better root-integrated-mean-squared-error measures compared with other unconstrained nonparametric methods. Favorable small sample properties make this method 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 constrained nonparametric estimators. They draw further attention to fitting not just 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 use a nonparametric least squares procedure incorporating restrictions on functional shape, similar to those used in Aït-Sahalia and Duarte (2003). Simulation results indicate that the constrained nonparametric 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 in this paper 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 across time.

3.2   Choice of state probability density estimator

The estimated state probability density reflects expected returns, which is compared in a ratio with the state-price density to measure risk preferences. Options-derived risk-neutral densities are forward-looking, and the related state probabilities are 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 by using a history of observations to estimate GARCH parameters and then forecasting into the future with random disturbances. It is not necessary that the forecasted returns precisely match 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 process to estimate state probability densities. This is based upon the GJR-specification of Glosten, Jagannathan and Runkle (1993). It then employs Monte Carlo simulations of S&P 500 index returns to calculate the predicted densities. Variations on this strategy are echoed throughout the literature, as in Scheicher (2003); Tarashev, Tsatsaronis and Karampatos (2003); Gai and Vause (2006); and Tarashev and Tsatsaronis (2006). 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 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 auto-correlated clustering. Modified GARCH processes in the literature have also been able to improve the capture of other observed characteristics of equity return processes. For example, the GJR-specification captures the observed tendency of equity return volatility to be disproportionately large when shocks are negative. Elsewhere Barone-Adesi, Engle and Mancini (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 these GARCH methods is that they do 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 stationary in the parameters of the underlying stochastic GARCH process. They question the effectiveness of a model that assumes state-price densities to be time-varying but the shape of the state probability densities not. Though key variables may vary with time, the parameters of the underlying stochastic process are constant. This means additionally 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 to time-vary. But in tradeoff they must make strong functional assumptions regarding pricing kernel form. These assumptions necessarily impose a constant coefficient of risk aversion, which conceals any time-varying behavior in investor preferences towards risk.

The critical intent in this paper is to modify implicit assumptions of stationarity in the underlying returns process, but without overly restricting risk preference behavior itself. We consider the Markov switching GARCH

model introduced in Hamilton and Susmel (1994) and Cai (1994), where a discrete Markov chain controls the parameters of the conditional variance process. By comparing the suitability of a regime-switching model with a single-regime model, we can then investigate whether common pricing kernel estimates come up short in only revealing average effects in state probability estimation. This would have broader implications for measuring and interpreting pricing kernels in the literature.

We argue that the single-regime model for state probability 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 of a single-regime state probability density, especially when state-price densities in comparison 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 density on the other hand could be very informative. For example, variation in state probability skewness would suggest that the perceived probability of downward market movements 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 or less likely than before. The intent in this paper then is to obtain a picture of how loosening these stationarity restrictions may impact risk preference implications within the context of the literature.

4 Estimation procedure

4.1 State-price density estimation

4.1.1 Least squares estimator

The nonparametric locally polynomial estimator for state-price densities in Aït-Sahalia and Duarte (2003) 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 observation 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

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

subject to restrictions on slope and convexity.

4.1.2 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:

$$C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) = \int_0^{+\infty} \max(S_t - X, 0) p^*(S_T | S_t, \tau, r_{t,\tau}, \delta_{t,\tau}) dS_T \tag{3}$$

where S_t is the price of the underlying asset at time t; X is the strike price; τ 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 vari-

ous conditional outcomes of the asset price at maturity. As risk-free interest rates $r_{t,\tau}$ we use interpolated historical Treasury bill rates[5].

The first derivative of the function is:

$$\frac{\partial C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial X} = -e^{-r_{t,\tau}t} \int_X^{+\infty} p^*(S_T|S_t, \tau, r_{t,\tau}, \delta_{t,\tau})dS_T \qquad (4)$$

which can also be expressed in the form of an inequality:

$$0 \geq \frac{\partial C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial X} \geq -e^{-r_{t,\tau}t}. \qquad (5)$$

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

The second derivative is:

$$\frac{\partial^2 C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial^2 X} = e^{-r_{t,\tau}t} p^*(X) \geq 0. \qquad (6)$$

This shows the second constraint for a convex option pricing function, with the second derivative positive with respect to 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 strike price. Normalization is required, multiplying the second derivative by $e^{r_{t,\tau}t}$.

4.1.3 Local polynomial smoothing

Aït-Sahalia and Duarte (2003) show transformed data points z_i are then used in a smoothing step[6]. They assume regression function $z(X)$ is continuous as a second-order derivative. Then according to local Taylor approximation methodology, function $z(X)$ can be estimated locally as a polynomial of order

---

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

[6]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).

p, with d denoting the order of the derivative:

$$z(X) \approx z(X_0) + z'(X_0)(X - X_0) + \ldots + \frac{z^p(X_0)(X - X_0)^p}{p!}$$

$$= \sum_{d=0}^{p} \frac{z^{(d)}X}{d!}(X_i - X)^d = \sum_{d=0}^{p} \beta_d(X)(X_i - X)^d \qquad (7)$$

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:

$$\min_{\beta} \sum_{i=1}^{N} (z_i - \sum_{d=0}^{p} \beta_d(X)(X_i - X)^d)^2 K_h(X_i - X) \qquad (8)$$

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 for approximated strike prices X_i in the neighborhood of observed 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 observed points. For asymmetric weights we specify the squared inverse of the distance between the index level and strike prices.

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

$$\min_{\boldsymbol{\beta}} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{W} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \qquad (9)$$

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

where $\mathbf{W}=diag\{K_h(X_1 - X), \ldots, K_h(X_N - X)\}$, $\mathbf{y}=[z_1, \ldots, z_N]$, and $\mathbf{X} =$

15

$$\begin{pmatrix} 1 & X_1 - X & \cdots & (X_1 - X)^p \\ 1 & X_2 - X & \cdots & (X_2 - X)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_N - X & \cdots & (X_N - X)^p \end{pmatrix}.$$

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

$$\hat{z}^{(d)}(X) = d!\hat{\beta}_d(X). \tag{10}$$

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.

4.2 State probability density estimation

4.2.1 Specification 1: Single-regime GJR-GARCH

GARCH models are able to accommodate heteroskedastic characteristics in financial returns. They assume that error term u_t can be decomposed as $u_t = \epsilon_t h_t^{1/2}$, where ϵ_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 ϵ_t, and it is distributed conditional on the information set $\boldsymbol{\Psi}$ at $t - 1$ as $u_t|\Psi_{t-1} \sim N(0, \sigma_t^2)$.

The GJR-GARCH model 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 in this paper as a baseline for comparison. To better capture fatter asset return tails, 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 inno-

vations have an additional latent variable, λ_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 can be described as follows:

$$y_t = a_0 + ay_{t-1} + u_t, \qquad u_t = \epsilon_t(\lambda_t h_t)^{1/2} \tag{11}$$
$$u_t \sim t(\nu), \quad \epsilon_t \sim N(0,1), \quad \lambda_t \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2})$$
$$h_t = \alpha_0 + \alpha(|u_{t-1}| - \gamma u_{t-1})^2 + \beta h_{t-1}$$

for $t = 1, \ldots, T$. 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 a constant term, and a the regression coefficient. We attempt to fit various ARMA formulations to the y_t process for our sample but find the estimated coefficient high density posterior intervals all include zero-values. We therefore simplify by assuming $y_t = u_t$. For the t distribution, λ_t is a scaling factor and ν is the degrees of freedom characterizing the density of λ_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. 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 $\alpha - \alpha(\gamma^2 - 1) + \beta < 1$ so that the process is bounded and does not explode. Parameter γ introduces a leverage effect into the model, taking into account the stylized fact in financial markets that negative shocks may tend to result in higher volatility than positive ones. When γ 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. With α_0 as a kind of baseline variance, coefficients α and β 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 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.

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 stochastic process of the state probability density. We will explore this matter further by comparing pricing kernels incorporating this single-regime GJR-GARCH formulation with the regime-switching model in our second specification below.

We estimate the single-regime GJR-GARCH model in this paper 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:

$$\hat{\boldsymbol{\theta}} = E[\boldsymbol{\theta}|\boldsymbol{Y} = \boldsymbol{y}] = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} \tag{12}$$

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

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{13}$$

The likelihood function is defined as:

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{y}) \propto (\det \boldsymbol{\Sigma})^{-1} \exp\left[-\frac{1}{2}\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right] \qquad (14)$$

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:

$$p(\alpha_0) \propto N(\mu_{\alpha_0}, \sigma_{\alpha_0}^2)I\{\alpha_0 > 0\} \qquad (15)$$
$$p(\alpha) \propto N(\mu_{\alpha}, \sigma_{\alpha}^2)I\{\alpha \geq 0\}$$
$$p(\gamma) \propto N(\mu_{\gamma}, \sigma_{\gamma}^2)I\{-1 < \gamma < 1\}$$
$$p(\beta) \propto N(\mu_{\beta}, \sigma_{\beta}^2)I\{\beta \geq 0\}.$$

We initiate vague parameter values at 0.1. Parameters are then re-estimated using dispersed initial values to confirm that the results are not influenced by this choice.

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 $\boldsymbol{\alpha_{MH}} = \min\{\frac{f(\boldsymbol{\theta}^*|\boldsymbol{y})}{f(\boldsymbol{\theta}|\boldsymbol{y})}\frac{q(\boldsymbol{\theta})}{q(\boldsymbol{\theta}^*)}, 1\}$, where $f()$ is the likelihood of the model producing this observation, $q()$ the parameter proposal density, $\boldsymbol{\theta}$ the existing parameter set, and $\boldsymbol{\theta}^*$ the parameter set updated with the candidate parameter draw.

4.2.2 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 dynamics. Error terms for process y_t will follow $u_t = \epsilon_t(\lambda_t h_{\Delta_t,t})^{1/2}$, where Δ_t is a Markov chain with $1, \ldots, k$ state spaces and a $k \times k$ transition matrix $\boldsymbol{\Pi}$:

$$\boldsymbol{\Pi} = [\pi_{i,j}] = [\pi(\Delta_t = j | \Delta_{t-1} = i)] \tag{16}$$

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 are given by:

$$\boldsymbol{h}_t = \boldsymbol{\alpha_0} + \boldsymbol{\alpha}(|u_{t-1}| - \boldsymbol{\gamma} u_{t-1})^2 + \boldsymbol{\beta} \boldsymbol{h}_{t-1} \tag{17}$$

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 $\alpha_{01} \neq \alpha_{02}$, for example, or whether the asymmetric response may be different between regimes when $\gamma_1 \neq \gamma_2$.

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 in the following steps:

20

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

2) Sample new state based upon drawn transition probability and 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. For a two-state example Henneke, Rachev, Fabozzi and Nikolov (2011) write the posterior distribution of $\pi_{i,i}$, the probability of remaining in state i from one observation to the next, as:

$$p(\pi_{i,i}|\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{\theta}) \propto p(\pi_{i,i})p(\boldsymbol{s}, \boldsymbol{y}|\boldsymbol{\theta})$$
$$\propto p(\pi_{i,i})p(\boldsymbol{s}|\boldsymbol{\theta}). \tag{18}$$

The probability of independent state observations is driven by the cumulated transitions over the sample:

$$p(\boldsymbol{s}|\boldsymbol{\theta}) = \prod_{t=1}^{T} p(s_{t+1}|s_t, \boldsymbol{\theta}) = (\pi_{i,i})^{\eta_{i,i}}(\pi_{i,j})^{\eta_{i,j}}(\pi_{j,j})^{\eta_{j,j}}(\pi_{j,i})^{\eta_{j,i}}$$
$$= (\pi_{i,i})^{\eta_{i,i}}(1 - \pi_{i,i})^{\eta_{i,j}}(\pi_{j,j})^{\eta_{j,j}}(1 - \pi_{j,j})^{\eta_{j,i}} \tag{19}$$

where $\pi_{i,j}$ are the cumulative transitions from state i to state j. This makes the Beta density a natural choice for conjugate prior in each row of the transition probability matrix: $(\pi_{i,i})^{((\eta_{i,i}+H_{i,i})-1)}(1 - \pi_{i,i})^{((\eta_{i,j}+H_{i,j})-1)}$. It can be generalized to a Dirichlet distribution to accommodate cases of more than

two states. This motivates sampling of $\pi_{i,i}$ and $\pi_{j,j}$ as follows:

$$\pi_{i,i}|s_{[1,T]} \sim Dirichlet(H_{i,i} + \eta_{i,i}, H_{i,j} + \eta_{i,j}) \tag{20}$$

$$\pi_{j,j}|s_{[1,T]} \sim Dirichlet(H_{j,j} + \eta_{j,j}, H_{j,i} + \eta_{j,i}).$$

At each new state observation, parameters $\pi_{i,j}$ are updated to reflect the cumulated history of transitions. As in Ardia (2009) we initiate hyperparameters at $H_{i,i} = 2$ and $H_{i,j} = 1$ with the idea that observations are more likely to remain in the prevailing state than to transition into a new state. The results are not sensitive to these initial values.

Step 2: Sample current state

Chib (1996) shows a typical point in the joint density for states $p(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{\theta}) = p(s_T|\boldsymbol{y}, \boldsymbol{\theta}) \cdot \ldots \cdot p(s_t|\boldsymbol{y}, \boldsymbol{S^{t+1}}, \boldsymbol{\theta}) \cdot \ldots \cdot p(s_1|\boldsymbol{y}, \boldsymbol{S^2}, \boldsymbol{\theta})$ can be expressed[7] in terms of the probability mass function and model likelihood along with the transition probabilities:

$$p(s_t|\boldsymbol{y}, \boldsymbol{S^{t+1}}, \boldsymbol{\theta}) \propto p(s_t|\boldsymbol{Y_t}, \boldsymbol{\theta}) f(\boldsymbol{Y^{t+1}}, \boldsymbol{S^{t+1}}|\boldsymbol{Y_t}, s_t, \boldsymbol{\theta})$$

$$\propto p(s_t|\boldsymbol{Y_t}, \boldsymbol{\theta}) p(s_{t+1}|s_t, \boldsymbol{\theta}) f(\boldsymbol{Y^{t+1}}, \boldsymbol{S^{t+2}}|y_t, s_t, s_{t+1}, \boldsymbol{\theta})$$

$$\propto p(s_t|\boldsymbol{Y_t}, \boldsymbol{\theta}) p(s_{t+1}|s_t, \boldsymbol{\theta}) \tag{21}$$

since the final term $f(\boldsymbol{Y^{t+1}}, \boldsymbol{S^{t+2}}|y_t, s_t, s_{t+1}, \boldsymbol{\theta})$ is independent of s_t. $p(s_t|\boldsymbol{Y_t}, \boldsymbol{\theta})$ is the probability mass function for the current state given information up to time t, and $p(s_{t+1}|s_t, \boldsymbol{\theta})$ is the transition probability π between state observations s_t and s_{t+1}. He further breaks down the probability mass function

---

[7]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, $\boldsymbol{S_t}$ is the history of states to time t, and $\boldsymbol{S^{t+1}}$ is the future procession of states through final observation T.

using Bayes theorem and the law of total probability:

$$p(s_t|\boldsymbol{Y_t},\boldsymbol{\theta}) \propto p(s_t|\boldsymbol{Y_{t-1}},\boldsymbol{\theta})f(y_t|\boldsymbol{Y_{t-1}},\boldsymbol{\theta})$$

$$\propto [\sum_{i=1}^{k} p(s_t|\boldsymbol{Y_{t-1}}, s_{t-1} = i, \theta)p(s_{t-1} = i|\boldsymbol{Y_{t-1}},\boldsymbol{\theta})]f(y_t|\boldsymbol{Y_{t-1}},\boldsymbol{\theta})$$

$$\propto [\sum_{i=1}^{k} p(s_t|s_{t-1} = i, \boldsymbol{\theta})p(s_{t-1} = i|Y_{t-1},\boldsymbol{\theta})]f(y_t|Y_{t-1},\boldsymbol{\theta}). \qquad (22)$$

This is the transition probability $p(s_t|s_{t-1} = i,\boldsymbol{\theta})$ multiplied by the probability mass function at $p(s_{t-1} = i|\boldsymbol{Y_{t-1}},\boldsymbol{\theta})$, multiplied by the likelihood $f(y_t|\boldsymbol{Y_{t-1}},\boldsymbol{\theta})$ of the current observation, all given our estimates made from the prior observation. Current state s_t is then sampled with this updated probability mass $p(s_t|\boldsymbol{Y_t},\boldsymbol{\theta})$. We initiate the process by calling the first state observation "1" and setting the first probability mass observation as Chib (1996) does to the stationary distribution of the chain, or the left eigenvector for an eigenvalue of 1.

Step 3: Sample λ_t and ν

For a prior on λ_t, Jacquier, Polson and Rossi (2004) and Henneke, Rachev, Fabozzi and Nikolov (2011) use a conjugate inverse-gamma density. They show:

$$p(\lambda_t|\nu) \propto \lambda_t^{-\nu/2+1}e^{-\nu/2\lambda_t} \sim \nu/\chi^2(\nu) \qquad (23)$$

with the conditional distribution of λ_t given by

$$p(\boldsymbol{\lambda}|\boldsymbol{y},\boldsymbol{h},\boldsymbol{\theta},\nu) = \prod_{t=1}^{T} p(\lambda_t|y_t, h_t, \nu) \qquad (24)$$

and

$$p(\lambda_t|y_t, h_t, \nu) \equiv p(\lambda_t|\frac{y_t}{\sqrt{h_t}}) \propto p(\frac{y_t}{\sqrt{h_t}}|\lambda_t, \nu)p(\lambda_t|\nu). \qquad (25)$$

23

The conditional posterior becomes:

$$p(\lambda_t|h_t, y_t, \nu) \propto \lambda_t^{-(\nu+1)/2+1} e^{-(y_t^2/h_t+\nu)/2\lambda_t} \sim \mathcal{IG}(\frac{\nu}{2}, \frac{y_t^2/h_t+\nu}{2}). \qquad (26)$$

Jacquier, Polson and Rossi (2004) then sample ν from this distribution. Alternatively we follow Henneke, Rachev, Fabozzi and Nikolov (2011) who sample $x_t \sim \chi^2(\nu+1)$ and calculate $\lambda_t = (\frac{y_t^2}{h_t} + \nu)/x_t$.

Jacquier, Polson and Rossi (2004) and Henneke, Rachev, Fabozzi and Nikolov (2011) explain that since $p(y_t|h_t, \nu) \sim t(\nu)$, the posterior distrbution of ν is proportional to the product of t-distributional ordinates:

$$p(\nu|\boldsymbol{h}, \boldsymbol{\theta}, \boldsymbol{y}) \propto p(\nu)p(\boldsymbol{y}|\boldsymbol{h}, \nu) = p(\nu) \prod_{t=1}^{T} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}(1 + \frac{y_t^2}{h_t\nu})^{-\frac{(\nu+1)}{2}}. \qquad (27)$$

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

$$p(\nu|\boldsymbol{h}, \boldsymbol{\theta}, \boldsymbol{y}) = \frac{1}{37} \frac{\prod_{t=1}^{T} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}(1 + \frac{y_t^2}{h_t\nu})^{-\frac{(\nu+1)}{2}}}{\prod_{t=1}^{T} \sum_{\nu=3}^{40} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}(1 + \frac{y_t^2}{h_t\nu})^{-\frac{(\nu+1)}{2}}}. \qquad (28)$$

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 con-

24

ditional 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 $j = 1$:

$$s_j \sim p(\boldsymbol{s}|\boldsymbol{y}, \alpha_0^{[j-1]}, \alpha^{[j-1]}, \gamma^{[j-1]}, \beta^{[j-1]}) \tag{29}$$
$$\alpha_0^{[j]} \sim p(\alpha_0|\boldsymbol{y}, s_j, \alpha^{[j-1]}, \gamma^{[j-1]}, \beta^{[j-1]})$$
$$\alpha^{[j]} \sim p(\alpha|\boldsymbol{y}, s_j, \alpha_0^{[j]}, \gamma^{[j-1]}, \beta^{[j-1]})$$
$$\gamma^{[j]} \sim p(\gamma|\boldsymbol{y}, s_j, \alpha_0^{[j]}, \alpha^{[j]}, \beta^{[j-1]})$$
$$\beta^{[j]} \sim p(\beta|\boldsymbol{y}, s_j, \alpha_0^{[j]}, \alpha^{[j]}, \gamma^{[j]})$$

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 α_0^1 is estimated using previous values $\alpha^0, \gamma^0, \beta^0$. After α_0^1 is simulated and updated, it is used together with γ^0 and β^0 to estimate α^1. Current values of α_0^1, α^1 and the previous β^0 value are used to estimate γ^1. Finally all three updated parameters are used in estimating β^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.

5   Data

The data used in these estimations include historical options[8] prices on the S&P 500 index from 2005 to 2011 and prices on the underlying index itself.

---

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

Table 1: S&P log return summary statistics

Observations	1762
Mean	3.053e-17
Standard deviation	1.4617
Skewness	-0.2855
Kurtosis	8.8511

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 to be used 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. 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, choosing 30-day contracts set to expire the same month in 20 days' time (or closest thereto if this happens to fall on a non-trading day).

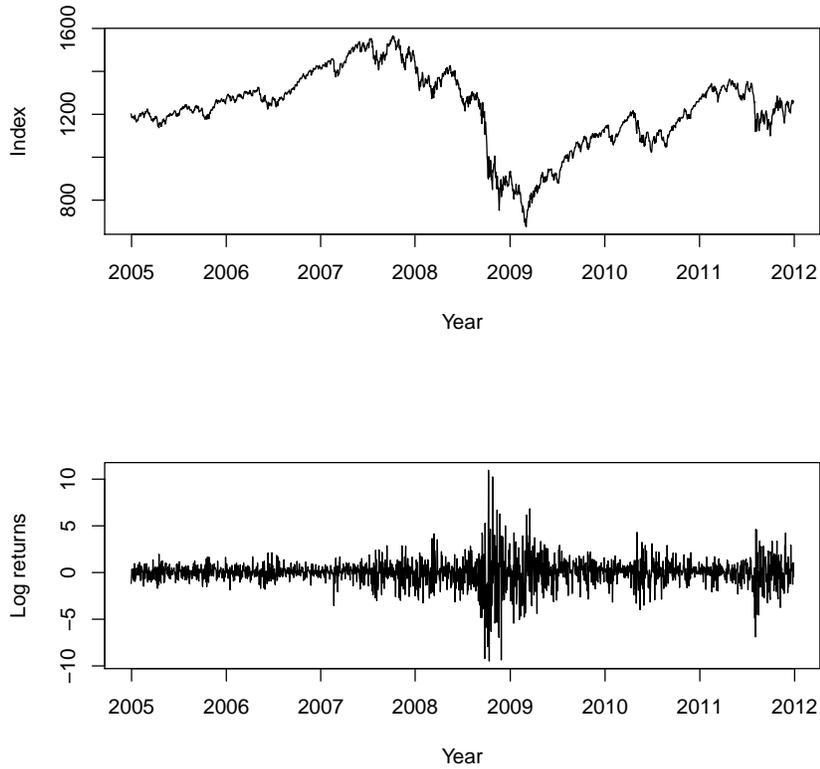For state probability density estimation, we transform S&P 500 index

Figure 1: S&P 500 index and log returns

prices into demeaned daily log-returns. Summary statistics are given in Table 1, with time-series plots of S&P 500 index levels and log-returns for years 2005 to 2011 shown in Figure 1. State probability densities are estimated on a monthly basis using daily closing values on the the S&P 500 index. Because our state-price densities are calculated each month using forward-looking contracts set to expire in 20 days' time, we maintain similar forward-looking expectations by forecasting state probability density estimates 20 days into the future from the final observation date used for estimation.

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 to 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 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. Eventually observations far enough in the past will provide a negligible weight in forming expectations made today.

6   Analysis

6.1   Comparison of models

We estimate the single-regime and Markov switching GJR-GARCH models for the entire 2005-2011 sample in order to compare fit and inferences. For identification purposes we restrict $\alpha_{01} < \alpha_{02}$ in the Metropolis-Hastings simulation step. Fruhwirth-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 α_0, and scatterplots of α_0 in the two states show a majority of the observations

clustered together but with a large amount of higher-ranging observations scattered separately. This geometry suggests the use of α_0 as a valid candidate for labeling the subspaces.

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 for us to look for evidence that convergence in the chains has been achieved[9]. 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 $\rho_{\alpha_1} = 0.5806$ in the Markov switching 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 indicate good mixing and appear to converge quickly. Acceptance rates for parameter chain draws are 83% 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 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 a large discrepancy indicative of 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 contains the results from our estimation. The degrees of freedom parameter ν is estimated to be around 21 in both models. This suggests that the use of t-distributed innovations in our sample is indeed warranted

---

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

Table 2: Conditional variance parameter estimates

Model	Parameters	Mean	Median	95% HPDI
Single regime	α_0	2.6845	2.6580	[2.5878,2.8505]
	α	0.1350	0.1350	[0.1333,0.1369]
	γ	3.634e-05	-0.0002	[-0.0115,0.0107]
	β	0.3929	0.3933	[0.3899,0.3967]
	ν	21.5806		
Two regimes	α_{01}	2.0799	2.0900	[2.0382,2.1310]
	α_1	0.1428	0.1428	[0.1407,0.1451]
	γ_1	6.972e-05	-0.0001	[-0.0120,0.0113]
	β_1	0.3731	0.3736	[0.3688,0.3784]
	α_{02}	4.3766	4.329	[4.1754,4.8136]
	α_2	0.1432	0.1431	[0.1409,0.1453]
	γ_2	0.0001	5.784e-05	[-0.0104,0.0104]
	β_2	0.3731	0.3731	[0.3689,0.3782]
	π_{11}	0.8029	0.8119	[0.7273,0.8692]
	π_{22}	0.3541	0.2832	[0.1507,0.5929]
	ν_1	21.2174		
	ν_2	21.2031		

to better capture tail behavior; values of 30 or above for ν would essentially revert to capturing a distribution that is normal. Parameters α_0, α and β are significant in both models, but results do not show the leverage parameter γ in either model to be significantly different from zero. Thus we do not find evidence within our sample of a difference in the conditional variance response to positive and negative shocks. The main difference between models lies in the baseline variance parameter for α_0. The separate regime parameters α_{01} and α_{02} in the Markov switching model appear to shrink towards their mean in the single-regime representation. The mean-shrinking behavior of α_0 also appears to cause much smaller but still significant differences in the impact parameter α and the memory parameter β. This results in a very slight understatement of the shock impact parameter α and a slight overstatement of the memory parameter β relative to the Markov switching model.

Within the Markov switching model, the parameter results show that the majority of variation for the data can be attributed to regime shifts between baseline variance parameters rather than to changes in the impact of prior shocks or prior variances. This may be due to the extreme changes in volatility over the sample which show up as major structural shifts that outweigh the other more subtle conditional variance parameters. The probability of remaining in the low volatility state is estimated to be $\pi_{11} = 81\%$, which suggests a fair amount of stability in the low volatility state in our sample. The probability of remaining in the high volatility state is less stable at $\pi_{22} = 35\%$, suggesting a higher degree of mixing between the states. This means that the high volatility state is typically more fleeting – once we are in a state of high volatility there is a higher likelihood of returning back to the more stable low volatility state.

To compare goodness-of-fit we estimate deviation information criterion statistics for each of the models. The deviation information criterion 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 esimation, taking both fit and complexity into account:

$$DIC = \overline{D} + 2p_D = E[-2lnf(y|\boldsymbol{\theta})] + 2E[-2lnf(y|\boldsymbol{\theta})] - 2lnf(y|\overline{\boldsymbol{\theta}}). \quad (30)$$

$\overline{D}$ 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: $lnf(y|\boldsymbol{\theta})$ is evaluated with parameters estimated at each observation, and $lnf(y|\overline{\boldsymbol{\theta}})$ is estimated 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. A lower relative DIC value is indicative of model preference.

Table 3: Deviance information criterion

Model	DIC	95% CI
Single regime	6592.695	[6589.601,6595.790]
Two regimes	5980.182	[5974.259,5986.105]

To add further meaning to our comparisons we also use a resampling procedure described in Ardia (2009) to estimate confidence intervals. This involves allowing random perturbations in the estimated parameters and re-calculating samples of new DIC values to estimate an entire distribution for these statistics. The Markov switching model appears to be statistically favored by DIC comparisons, lying outside of a 95% confidence interval from the single-regime statistic. Table 3 reports these estimates.

For insight into model fit according to the separate volatility states, we next filter these estimates according to whether the observation occurs in the low volatility state *(1)* or the high volatility state *(2)*. The resulting statistics along with their estimated confidence intervals are presented in Table 4. These show that the Markov switching model significantly improves fit over the single-regime model in both the low and the high volatility states.

Table 4: Deviation information criterion, by state

Model	State	DIC	95% CI
Single regime	*(1)*	5129.429	[5127.387,5131.470]
	(2)	1463.266	[1460.940,1465.592]
Two regimes	*(1)*	4641.372	[4637.763,4644.981]
	(2)	1338.810	[1334.114,1343.506]

Posterior predictive performance is also examined. We replicate the data using each of the two models, and then measure these deviations from the actual observed returns. These are reported in the form of mean squared errors (MSE) and mean absolute errors (MAE). The results shown in Table 5 indicate that the Markov switching model has slightly smaller predictive

Table 5: Posterior predictive results

Model	Measurement	Statistic	95% CI
Single regime	MSE	7.8224	[7.2415,8.4033]
	MAE	2.1829	[2.1002,2.2656]
Two regimes	MSE	7.46484	[6.8968,8.0400]
	MAE	2.0885	[2.0113,2.1657]

errors than the single-regime model, which can be interpreted as the Markov switching model having a slightly better ability to capture the actual range of the observed data. Through resampling we estimate 95% confidence intervals for these measures. These intervals indicate that the differences are likely not statistically significant. Comparable predictive performance between the two models is not entirely unexpected however, due to the stochastic nature of the error-driven data generating process.

6.2 Pricing kernel behavior

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

To investigate, we estimate pricing kernel functions under single-regime and Markov switching models for comparison. We take the ratios of state-price and state probability densities, measuring the perceived probabilities of negative events as the tail areas of these densities below particular changes in log-returns. When the pricing kernel ratio in a particular state of wealth is greater than 1, it indicates that 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. With asset pricing theory telling us that payouts
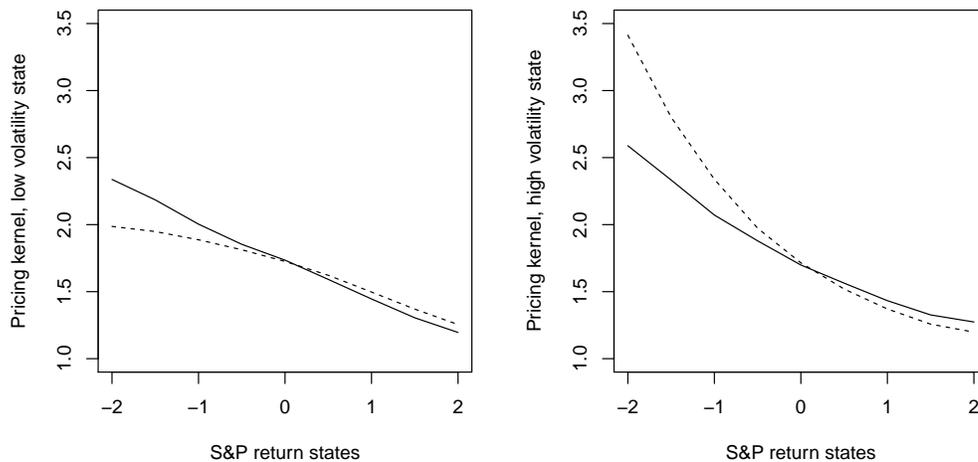
Figure 2: Pricing kernel function estimates. Dashed lines represent estimations from the single-regime model; solid lines are estimated using the Markov switching model.

in lower states of wealth are valued more highly for the risk averse investor, these higher preference-weighted valuations relative to expected price movements reflect lower aggregate market preferences for risk. When the pricing kernel ratio in a particular state of wealth is less than 1 on the other hand, it indicates that investors may be underestimating downward risks in the market. They are pricing options as if those downward movements are actually less likely than the historical likelihood would suggest. In these cases risk preferences in markets may be running high.

Sample pricing kernel estimates are shown in Figure 2. The pricing kernels on the left are estimated during the low volatility state[10], and the pricing kernels on the right are estimated during the high volatility state[11]. Dashed lines represent the single-regime model estimate, and solid lines represent the

---

[10]Estimated for November 2010.
[11]Estimated for December 2011.

Table 6: Pricing kernels, single-regime model

	Low volatility state	High volatility state
Mean	1.0483	1.6543
95% CI	[0.8973,1.1993]	[1.4096,1.8989]

Table 7: Pricing kernels, Markov switching model

	Low volatility state	High volatility state
Mean	1.1785	1.2465
95% CI	[1.0064,1.3506]	[1.0391,1.4539]

Markov switching model estimate. Qualitatively, in the low volatility state to the left the single-regime model appears to flatten the pricing kernel function relative to the Markov 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 Markov 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 Markov 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 their respective volatility regimes for the entire sample. Tables 6 and 7 show mean pricing kernel values over the entire sample for a 1% fall in state of wealth and their 95% confidence intervals. The single-regime model identifies mean pricing kernel estimates across time as significantly different between regimes, but the Markov switching model does not. This suggests that the single regime model may be mistakenly identifying risk preference behavior in failing to distinguish between regimes of conditional variance.

Table 8: Pricing kernels, single-regime model

	Pre-October 2008	Post-October 2008
Mean	0.8366	1.5722
95% CI	[0.6710,1.0022]	[1.4130,1.8989]

Table 9: Pricing kernels, Markov switching model

	Pre-October 2008	Post-October 2008
Mean	0.8813	1.5526
95% CI	[0.7061,1.0567]	[1.3753,1.7299]

Next we divide the sample by time with a breakpoint at the month of October 2008. This is when the most extreme spikes in volatility occur in our sample, representing an acute period of financial panic. Mean pricing kernel values for the time intervals before and after as well as their 95% confidence intervals are reported Tables 8 and 9. The Markov switching model does pick up a significant difference between mean pricing kernel estimates before and after the extreme events of September 2008. The pricing kernel average prior to October 2008 is below 1, suggesting that aggregate preferences for risk in the first period were rather high. The pricing kernel mean from October 2008 onwards on the other hand is significantly higher than in the first period. The fact that both volatility states are interspersed throughout the second period indicates that preferences for risk are lower on average, in spite of reversion to the lower volatility state. So rather than being closely tied to movements in volatility, aggregate risk preferences appear to outlast the actual structural shifts in volatility.

7   Conclusion

In this paper we find evidence to support the existing literature on Markov switching models of conditional variance in financial returns. We examine the time-frame surrounding the recent 2008 financial meltdown to gain some perspective on the nature of this period of extreme volatility. Results show an improvement in fit for the Markov switching model over the 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 Markov switching conditional variance on pricing kernel functions and risk preferences. This addresses known conceptual difficulties in pricing kernel estimation with stationary parameters in the state probability density. Our results show that the concern regarding stationary state probability densities is warranted, not just conceptually, but for empirical reasons as well. The introduction of Markov switching parameters to capture high and low states of volatility leads to some qualitative differences for pricing kernels and risk preference behavior. Relative to the Markov switching model, the single-regime model underestimates the pricing kernel in negative states of wealth during the low volatility state. Conversely it overestimates the pricing kernel in negative states of wealth during the high volatility state. 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 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 from 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 exit 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 mean risk preferences across time between states of volatility, yet the Markov switching model does not. This suggests that the single-regime model may indeed be falsely identifying these changes in risk preferences by failing to account for possible regime changes in conditional variance over time. This lies at the heart of the concern regarding stationary state probability densities in pricing kernel estimation. The Markov switching model does however find a significant change in average risk preferences after the financial meltdown in September 2008 through the end of the sample in 2011. This means that risk preferences overall are remaining low even after transitioning back to the low volatility state. The behavior suggests that aggregate risk preferences may have a long memory after the impact of extreme events, outlasting the actual estimated structural shift in volatility itself.

References

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

Aït-Sahalia, Y. and A.W. Lo (2000), "Nonparametric Risk Management and Implied Risk Aversion", *The Journal of Econometrics* 94, 9-51.

Aït-Sahalia, Y. and J. Duarte (2003), "Nonparametric Option Pricing under Shape Restrictions", *The Journal of Econometrics* 116, 9-47.

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

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

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

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

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

Bollerslev T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics* 31, 307-327.

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

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

Chib, S. (1996), "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models", *Journal of Econometrics* 75, 79-97.

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

Cochrane, J. (2005), *Asset Pricing*, Revised Edition, Princeton, NJ: Princeton University Press.

Danielsson, J., H.S. Shin, and J. Zigrand (2009), "Risk Appetite and Endogenous Risk", working paper, London School of Economics and Princeton University.

Debreu, G. (1959), *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, New Haven and London: Yale University Press, Monograph 17.

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

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

Figlewski, S. (2008), "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, UK: Oxford University Press.

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

Gai, P. and N. Vause (2006), "Measuring Investors' Risk Appetite", *International Journal of Central Banking* March, 167-188.

González-Hermosillo, B. (2008), "Investors' Risk Appetite and Global Financial Market Conditions", IMF Working Paper WP/08/85.

Glosten, L.R.; Jagannathan, R. and D.E. Runkle (1993), "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.

Hamilton J.D. and R. Susmel (1994), "Autoregressive Conditional Heteroskedasticity and Changes in Regime", *Journal of Econometrics* 64, 307-333.

Hamilton J.D. and R. Susmel (2004), "A New Approach to Markov-Switching GARCH Models", *Journal of Financial Econometrics* 2(4), 493-530.

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

Jackwerth, J.C. (2000), "Recovering Risk Aversion from Options Prices and Realized Returns", *The Review of Financial Studies* 13(12), 433-451.

Jackwerth, J.C. (2004), *Option-Implied Risk Neutral Distributions and Options Prices*, Charlotteville, USA: Research Foundation of AIMR, Inc. 13(12), 433-451.

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

Macrina, A. and P.A. Parbhoo (2011), "Randomised Mixture Models for Pricing Kernels", Quantitative Finance Papers, arXiv.org.

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

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

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

Shimko, D. (1993), "Bounds of Probability", *Risk* 6(4), 33-36.

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

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

Tarashev, N. and K. Tsatsaronis (2006), "Risk Premia Across Markets: Information from Option Prices", *BIS Quarterly Review* March, 93-103.

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