Beyond Stochastic Volatility and Jumps in Returns and Volatility

Garland Durham[*] and Yang-Ho Park[†‡]

October 9, 2012

Abstract

While a great deal of attention has been focused on stochastic volatility in stock returns, there is strong evidence suggesting that return distributions have time-varying skewness and kurtosis as well. Under the risk-neutral measure, for example, this can be seen from variation across time in the shape of Black-Scholes implied volatility smiles. This paper investigates model characteristics that are consistent with variation in the shape of return distributions using a stochastic volatility model with a regime-switching feature to allow for random changes in the parameters governing volatility of volatility, leverage effect and jump intensity. The analysis consists of two steps. First, the models are estimated using only information from observed returns and option-implied volatility. Standard model assessment tools indicate a strong preference in favor of the proposed models. Since the information from option-implied skewness and kurtosis is not used in fitting the models, it is available for diagnostic purposes. In the second step of the analysis, regressions of option-implied skewness and kurtosis on the filtered state variables (and some controls) suggest that the models have strong explanatory power for these characteristics.

Keywords: return distributions; skewness; stock price dynamics; stochastic volatility; regime switching; option pricing; leverage effect; volatility of volatility; jump intensity

1 INTRODUCTION

Understanding volatility dynamics and improving option pricing have long been of interest to practitioners and academics. It is well-known that the volatility of many financial assets is time-varying, and an enormous amount of research has been devoted to studying this feature of financial data. But, there is strong empirical evidence suggesting that return distributions have time-varying skewness and kurtosis as well. For example, stochastic skewness in risk-neutral return distributions is implied by variation across time in the slope of the Black-Scholes implied volatility smile. Stochastic kurtosis is related to variation across time in the curvature of the Black-Scholes implied volatility smile. These are important features of observed option prices and are only weakly correlated with variation in option-implied volatility. Understanding variation in the shape of return distributions (and the shape of the implied volatility smile) is important in many applications, such as hedging and risk management.

The objective of this paper is to investigate model characteristics that are consistent with time-varying skewness and kurtosis in return distributions as is observed empirically in the options market. In particular, we look at models with additional state variables that allow for time-variation in volatility of volatility, correlation between innovations in prices and volatility (leverage effect), and jump intensity, all of which are able to generate variation in the shape of return distributions that is independent of the level of volatility. We find strong evidence in favor of these features.

An important aspect of our analysis is that the models are estimated using returns and implied volatility, but additional information about the shape of the return distribution embedded in option prices (e.g., higher order moments of the risk-neutral measure) is not used in fitting them. By withholding this information from the model estimation, we are able to use it for diagnostic purposes. Toward this end, we look at some regressions to examine whether the implied state variables have explanatory power for option-implied skewness and kurtosis and find strong evidence that they do. This is important because it suggests that variation in the shape of risk-neutral return distributions (and of the Black-Scholes implied volatility smile) is not just due, for example, to changes in risk premia, but is associated with changes in related characteristics of the physical dynamics.

This paper builds on a substantial body of previous work. Das and Sundaram (1999) show that both volatility of volatility and correlation between the innovations in an asset's price and its volatility (leverage effect) affect the shape of the volatility smirk. They show that the size and intensity of jumps in returns do so as well, though the effect is primarily at short terms to maturities (alternative specifications for the jump dynamics may be able to generate similar effects at longer horizons). There has been some work toward implementing these ideas in empirical work. For example, the two-factor stochastic volatility model of Christoffersen, Heston, and Jacobs (2009) is able to generate time-varying correlation, while Santa-Clara and Yan (2010) allow the jump intensity to be stochastic. Carr and Wu (2007) propose a stochastic skew model for foreign exchange rates with positive and negative jumps driven by independent Levy processes. Johnson (2002) looks at a stochastic volatility model with time-varying correlation between return and volatility innovations. Jones (2003) proposes a constant elasticity of variance model that incorporates a time-varying leverage effect. Harvey and Siddique (1999) look at GARCH models that incorporate time-varying skewness.

The underlying modelling framework is based on a standard single-factor stochastic volatility model. Although models of the affine (or affine-jump) class are often used in work of this kind due to their analytical tractability, these models have trouble fitting the data (e.g., Jones 2003; Ait-Sahalia and Kimmel 2007; Christoffersen, Jacobs, and Mimouni 2010). But since the techniques applied in this paper do not rely upon the analytical tractability of the affine models, we are able to choose among classes of models based on performance instead. We have found that log volatility models provide a useful starting point. We allow for contemporaneous jumps in both returns and volatility. We build on this framework by adding a regime-switching feature for the parameters corresponding to volatility of volatility, leverage effect and jump intensity. This idea is motivated by the fact that changes in any of these three variables, at least under the risk-neutral measure, are capable of generating variation in the shape of the Black-Scholes implied volatility smirk.

Our empirical work uses S&P 500 index (SPX) option data. Figure 1 shows time-series plots of option-implied volatility, skewness, and kurtosis estimated using the model-free approach of Bakshi, Kapadia, and Madan (2003). Figure 2 shows scatter plots of option-implied skewness and kurtosis

versus volatility. It is evident from these plots that there is substantial and persistent variation in skewness and kurtosis, and that this variation is only weakly correlated with the level of volatility. Models with only a single state variable are hard to reconcile with the empirical features of these data. The inclusion of additional state variables in the model (such as the regime states used in this paper) is needed to break this lock-step relationship between volatility, skewness and kurtosis.

The analysis is comprised of two main parts. First, we fit the models using SPX prices and option-implied volatility. We compare models based on log likelihoods, information criteria, and other diagnostics. The second step involves testing the explanatory power of the implied regime states for option-implied skewness and kurtosis.

Regarding the first step of the analysis, including jumps in the model provides a huge improvement relative to the base model with no jumps. The log likelihood increases by over 300 points. Other diagnostics of model fit are also greatly improved. Including the regime-switching feature provides additional large improvements. The best of these models uses regime switching in volatility of volatility. This model provides an increase of 110 points in log likelihood relative to the model without regime switching, with improvements in other diagnostics of model fit as well.

In the second step of the analysis, regressions testing whether the implied states have explanatory power for option-implied skewness and kurtosis are also decisive. The slope coefficients are strongly significant and in the expected directions. Option-implied skewness tends to be more negative and option-implied kurtosis tends to be more positive when volatility of volatility is high or the leverage effect is more pronounced (more negative correlation between price and volatility innovations). Our regression results are robust to inclusion of control variables such as VIX index, variance risk premium, and jump variation. While regression of option-implied skewness on the control variables alone has an adjusted R^2 of only 9.2%, adding the implied regime states to the regression gives an adjusted R^2 of over 32%. Slope coefficients for the regime state variables have t-statistics with absolute value greater than 10.

The remainder of the paper is organized as follows: Section 2 describes the models under consideration; Section 3 describes the methodology; Section 4 reports parameter estimates and
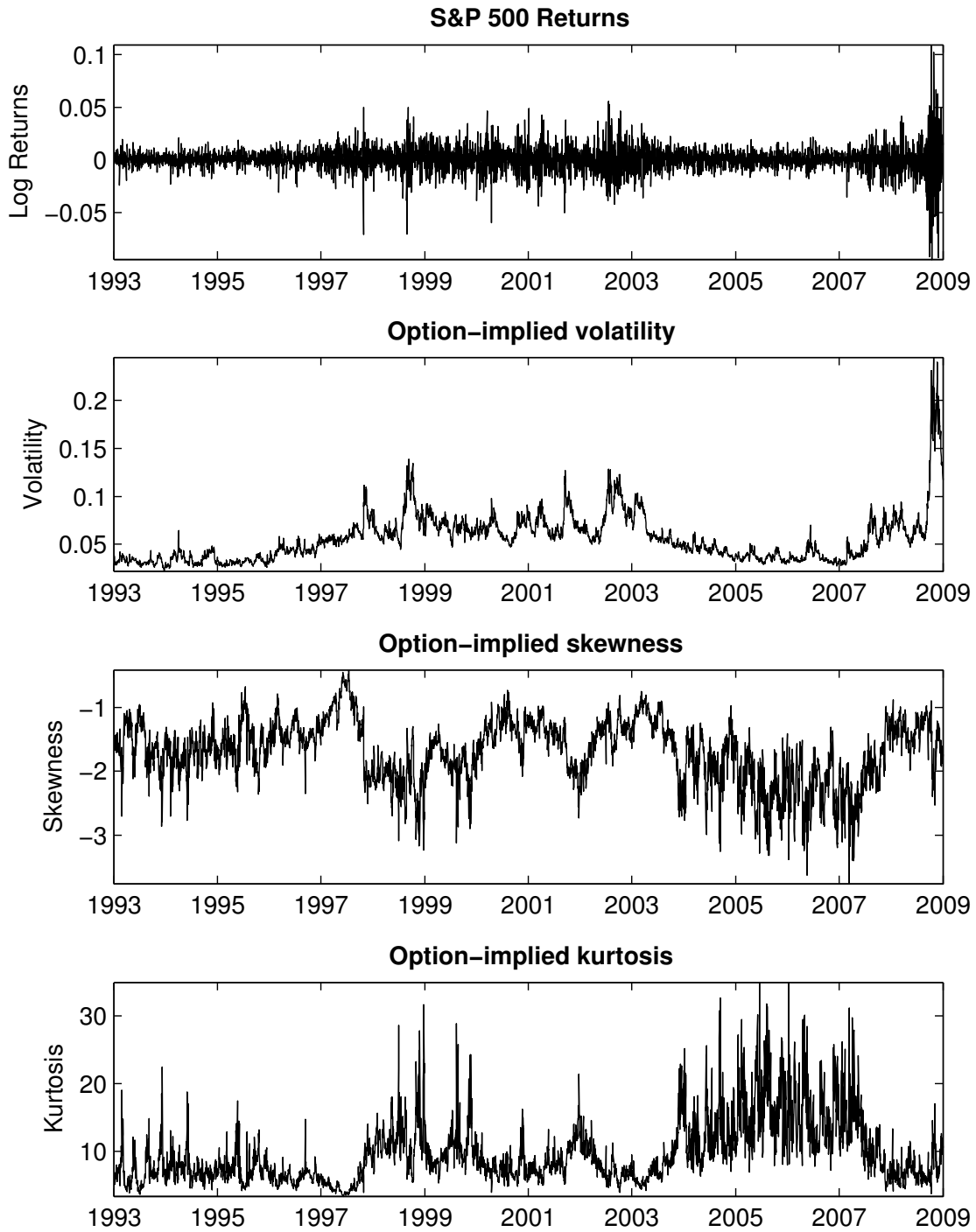
Figure 1: Time series of S&P 500 returns, option-implied volatility, option-implied skewness, and option-implied kurtosis.
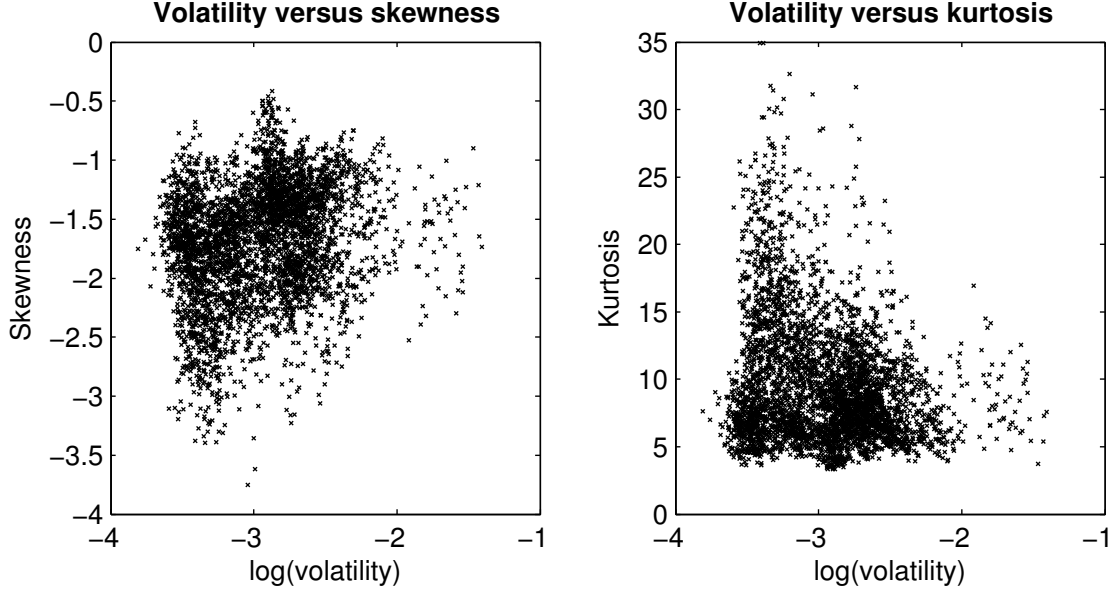
Figure 2: Scatter plots of option-implied skewness and kurtosis against option-implied volatility.

diagnostics for the various models; Section 5 investigates the explanatory power of implied regime states for option-implied skewness and kurtosis; and Section 6 concludes.

2 MODELS

The modeling framework used in this paper is based on a standard stochastic volatility model. In light of findings by e.g. Eraker, Johannes, and Polson (2003), we allow for jumps in both returns and volatility. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and information filtration $\{\mathcal{F}_t\}$, the ex-dividend log stock price, y_t, is assumed to evolve as

$$
\begin{aligned}
dy_t &= \left[\mu - \overline{\mu}_{1,Jt}\lambda_1(s_t) - \frac{1}{2}\exp(v_t)\right]dt + \exp(v_t/2)dW_{1t} + J_{1t}dN_{1t} \\
dv_t &= \left[\kappa(\overline{v} - v_t) - \overline{\mu}_{2,Jt}\lambda_1(s_t)\right]dt + \sigma(s_t)dW_{2t} + J_{2t}dN_{1t} \\
ds_t &= (1 - 2s_t)dN_{2t}
\end{aligned}
\tag{1}
$$

where v_t and s_t are the volatility state and the regime state, respectively. The regime state is either 0 or 1. W_{1t} and W_{2t} are standard Brownian motions with regime-dependent correlation $\rho(s_t)$. N_{1t} and N_{2t} are Poisson processes with intensity $\lambda_1(s_t)$ and $\lambda_2(s_t)$, respectively.

6

We consider two different forms for the jump structure, depending upon whether jumps are scaled by the volatility state or not. The unscaled models (UJ) assume that jump innovations are *i.i.d.* bivariate normal with $J_{1t} \sim N\left(\mu_{1J}, \sigma_{1J}^2\right)$, $J_{2t} \sim N\left(\mu_{2J}, \sigma_{2J}^2\right)$, and $\text{corr}(J_{1t}, J_{2t}) = \rho_J$. Similar jump specifications have been used in the existing literature (e.g., Eraker, Johannes, and Polson 2003; Eraker 2004; Broadie, Chernov, and Johannes 2007). In contrast, the scaled models (SJ) assume that jumps scale in proportion to the volatility of the diffusive component of the process. That is, J_{1t} and J_{2t} are bivariate normal with $J_{1t}/\exp(v_t/2) \sim N\left(\mu_{1J}, \sigma_{1J}^2\right)$, $J_{2t}/\sigma(s_t) \sim N\left(\mu_{2J}, \sigma_{2J}^2\right)$, and $\text{corr}(J_{1t}, J_{2t}) = \rho_J$. By generating larger jumps when volatility is higher, the SJ model is potentially capable of providing more realistic dynamics (this hypothesis is indeed confirmed in the empirical section). In either case, we denote expected jump size (conditional on the volatility state for the SJ models) by $\overline{\mu}_{1Jt} = E(e^{J_{1t}} - 1)$ and $\overline{\mu}_{2Jt} = E(J_{2t})$.

The regime-dependent parameters, σ, ρ and λ_1, allow for variation across time in the volatility of volatility, leverage effect and jump intensity, respectively. These are the mechanisms by which it is possible to generate time-varying skewness and kurtosis. The regime-dependence of λ_2 lets the regimes differ in persistence.

For simplicity, we only look at models with two possible regimes, although extending our techniques to more regimes is straightforward (at the cost of more free parameters to estimate). The regime state process is the continuous-time analog of a discrete-time Markov switching model in which the probability of switching from state s to state $1 - s$ is $\lambda_2(s)\Delta t$ ($s = 0, 1$).

While the dynamics of the underlying asset are described by the above model, options are priced according to a risk-neutral measure, $\mathbb{Q}$. We assume that under this measure the model takes the form

$$
\begin{aligned}
dy_t &= \left[r_t - q_t - \overline{\mu}_{1Jt}\lambda_1(s_t) - \frac{1}{2}\exp(v_t) \right] dt + \exp(v_t/2)dW_{1t}^{\mathbb{Q}} + J_{1t}^{\mathbb{Q}}dN_{1t}^{\mathbb{Q}} \\
dv_t &= \left[\kappa(\overline{v} - v_t) - \eta(s_t)v_t - \overline{\mu}_{2Jt}\lambda_1(s_t) \right] dt + \sigma(s_t)dW_{2t}^{\mathbb{Q}} + J_{2t}^{\mathbb{Q}}dN_{1t}^{\mathbb{Q}} \\
ds_t &= (1 - 2s_t)dN_{2t}^{\mathbb{Q}}
\end{aligned}
\tag{2}
$$

where r_t and q_t denote the risk-free rate and the dividend rate, respectively. $W_{1t}^{\mathbb{Q}}$ and $W_{2t}^{\mathbb{Q}}$ are standard Brownian motions with regime-dependent correlation $\rho(s_t)$ under the risk-neutral measure.

Table 1: Summary of models

SV	No jumps, no regime switching
SJ	Volatility-scaled jumps, no regime switching
UJ	Volatility-unscaled jumps, no regime switching
SJ-RS-Vol	Volatility-scaled jumps, regime switching for σ
UJ-RS-Vol	Volatility-unscaled jumps, regime switching for σ
SJ-RS-Lev	Volatility-scaled jumps, regime switching for ρ
UJ-RS-Lev	Volatility-unscaled jumps, regime switching for ρ
SJ-RS-Jmp	Volatility-scaled jumps, regime switching for λ_1
UJ-RS-Jmp	Volatility-unscaled jumps, regime switching for λ_1

Jump parameters are the same under physical and risk-neutral measures. In other words, we do not attempt to identify any jump risk premia. The models and data used in this paper have limited power to separately identify jump risk and diffusive volatility premia, and our results do not depend on being able to do so.

In empirical work, the variance risk premium is generally found to be negative (e.g., Coval and Shumway 2001; Bakshi and Kapadia 2003; Carr and Wu 2009). This finding can be understood in the framework of classic capital asset pricing theory (e.g., Heston 1993; Bakshi and Kapadia 2003; Bollerslev, Gibson, and Zhou 2011). But theory also suggests that the variance risk premium should be dependent on both the volatility of volatility and the leverage effect (correlation between returns and changes in volatility state). We allow for this possibility by letting the volatility risk premium parameter $\eta(s_t)$ be regime-dependent.

The application looks at several variants of this model, summarized in Table 1. We also looked at a number of alternative specifications with different regime states, jump dynamics, and risk premia, but do not report results for such experiments to keep the presentation manageable.

In the empirical work, we use an Euler scheme approximation to the model. For the physical

model (and analogously for the risk-neutral model), the approximation is given by

$$
\begin{aligned}
y_{t+1} &= y_t + \mu - \overline{\mu}_{1Jt}\lambda_1(s_t) - \frac{1}{2}\exp(v_t) + \exp(v_t/2)\varepsilon_{1,t+1} + \sum_{j>N_{1,t}}^{N_{1,t+1}} \xi_{1j} \\
v_{t+1} &= v_t + \kappa(\overline{v} - v_t) - \overline{\mu}_{2Jt}\lambda_1(s_t) + \sigma(s_t)\varepsilon_{2,t+1} + \sum_{j>N_{1,t}}^{N_{1,t+1}} \xi_{2j}
\end{aligned}
\tag{3}
$$

where $\varepsilon_{1,t+1}$ and $\varepsilon_{2,t+1}$ are standard normal with correlation $\rho(s_t)$, and ξ_{1j} and ξ_{2j} have the same distribution as the jumps in the continuous-time model. The regime state, s_t, follows the discrete-time Markov process with $p(s_{t+1} = i | s_t = j) = P_{ij}$, corresponding to the transition matrix

$$
\mathbf{P} = \begin{pmatrix} \pi_0 & 1 - \pi_1 \\ 1 - \pi_0 & \pi_1 \end{pmatrix}.
$$

For computational purposes, we impose an upper bound constraint of five jumps in a single day.

3   METHODOLOGY

The introduction of the regime-switching feature means that the models used in this paper require the development of new estimation techniques. While the estimation strategies used in similar work often rely heavily on computationally costly simulation methods, the approach we propose in this paper runs in several minutes on a typical desktop PC. Our approach consists of three steps: (1) back out volatility states from observed option prices; (2) filter regime states using a Bayesian recursive filter; and (3) optimize the likelihood function using the volatility and regime states obtained in the previous two steps. As by-products, the algorithm provides a series of generalized residuals, which we make use of for model diagnostics, and estimates of the volatility and regime states, which are used for the regressions in Section 5. A detailed description of each step of the procedure is provided below.

9

3.1 Extracting the volatility states

Building on the work of Chernov and Ghysels (2000), Pan (2002), Ait-Sahalia and Kimmel (2007), and others, we make use of observed option prices in addition to the price of the underlying asset to estimate the models. Estimates of option-implied volatility are obtained using the model-free approach of Bakshi, Kapadia, and Madan (2003).

We now describe the mechanics of how volatility and regime states are backed out conditional on an observed value of option-implied volatility (together with a candidate model and parameter vector). Following Bakshi, Kapadia, and Madan (2003), let IV_t^T denote the square root of the expected integrated variance of log returns on the interval $(t, T]$ under the risk-neutral measure. Given a risk-neutral model for stock price dynamics and initial values for the volatility and regime states, the corresponding value for IV_t^T can be obtained by integrating the quadratic variation of the log stock price. For the SJ models (including any of the regime-switching variants), for example, we get

$$\mathrm{IV}_t^T = \sqrt{\frac{1}{T-t} E_t^{\mathbb{Q}} \left\{ \int_t^T e^{v_\tau} \ 1 + \lambda_{1\tau}(\mu_{1J}^2 + \sigma_{1J}^2) \ d\tau \right\}}.$$

The integrand is equal to the instantaneous variance, which is comprised of terms reflecting the diffusion and jump components of returns. The expectation in the above expression can be computed by means of Monte Carlo simulations,

$$\mathrm{IV}_t^T \approx \sqrt{\frac{1}{(T-t)S} \sum_{i=1}^{S} \left\{ \int_t^T e^{v_\tau^{(i)}} \ 1 + \lambda_{1\tau}^{(i)}(\mu_{1J}^2 + \sigma_{1J}^2) \ d\tau \right\}}, \tag{4}$$

where $v_\tau^{(i)}$ and $\lambda_{1\tau}^{(i)}$ ($i = 1, \ldots, S$) are obtained by simulating paths from the risk-neutral analog of Equation (3) for $t < \tau \le T$, and S denotes the number of simulation paths. The calculations for the UJ models are similar, except the integrand is $e^{v_\tau} + \lambda_{1\tau}(\mu_{1J}^2 + \sigma_{1J}^2)$, reflecting the fact that the diffusive component of returns is scaled by volatility, but not the jumps.

However, we actually want to go in the reverse direction. That is, observed values for IV_t^T are available and we need to obtain the corresponding volatility and regime states, v_t and s_t, by

10

inverting Equation (4). We begin by showing how to do this conditional on the regime state (the issue of backing out the regime state is addressed in the next subsection).

Given an initial value for the regime state, $s_t = i$, the first step is to obtain an approximation to the mapping from spot to integrated volatility,

$$\Gamma_i : \mathrm{SV} \longrightarrow \mathrm{IV},$$

where SV denotes the spot volatility, i.e., $\mathrm{SV}_t = \exp(v_t/2)$, and the subscript i denotes the conditioning on initial regime state. The simplest way to do this is to evaluate (4) on some grid of initial values for SV and then use some curve fitting technique to approximate Γ_i. That is, let $\widehat{\mathrm{SV}}_1 < \widehat{\mathrm{SV}}_2 < \cdots < \widehat{\mathrm{SV}}_G$ be the grid, where G is the number of grid points and we use hats to indicate that these are grid points rather than data. For each $g = 1, \ldots, G$, evaluate $\widehat{\mathrm{IV}}_g = \Gamma_i(\widehat{\mathrm{SV}}_g)$ using Monte Carlo methods as described above (note that while the initial regime state is given, it evolves randomly thereafter). Then approximate Γ_i based on the collection of pairs $\{(\widehat{\mathrm{SV}}_g, \widehat{\mathrm{IV}}_g)\}_{g=1}^G$. As long as this mapping is monotonic, it is equally straightforward to approximate the inverse, $\Gamma_i^{-1} : \mathrm{IV} \longrightarrow SV$, which is what we are really interested in. Let $\widehat{\Gamma}_i^{-1}$ denote the approximation. While there are many curve fitting schemes one could use, we have found that simply fitting a cubic polynomial to the collection $\{(\widehat{\mathrm{IV}}_g, \widehat{\mathrm{SV}}_g)\}_{g=1}^G$ using nonlinear least squares works well. We use $G = 15$ for the empirical work reported in Sections 4 and 5. We also tried more grid points, higher order polynomials, splines, and various other interpolation schemes, but none provided noticeable improvements. Approximation errors are negligible for any reasonable scheme.

Given an observed value for option-implied integrated volatility, IV_t and assuming $s_t = i$, one evaluates $\widehat{\Gamma}_i^{-1}$ to obtain SV_t. Then, the volatility state itself is given by $v_t = 2\log(\mathrm{SV}_t)$. The important thing to notice here is that computing $\widehat{\Gamma}_i^{-1}$, which is the costly step, need only be done once (for each candidate parameter vector). Once this is accomplished, the evaluation step is fast.

3.2 Filtering the regime states

Given an observed value of IV_t, we now have two possible values for SV_t (and v_t), one for each regime state. Let v_t^j denote the volatility state corresponding to regime j. The second step of the estimation involves applying a filter to compute $p_t^j = p(s_t = j | \mathcal{F}_t) = p(v_t = v_t^j | \mathcal{F}_t)$.

The filter is constructed recursively using standard techniques. Let $p_t = (p_t^0, p_t^1)'$ for each $t = 0, \ldots, n$, and initialize the filter by setting p_0^j equal to the marginal probability of state j $(j = 0, 1)$. Now, suppose that p_t is known. The problem is to compute p_{t+1}. This is given by

$$p_{t+1}^j = \frac{\sum_{i=0}^1 p(y_{t+1}, v_{t+1}^j | y_t, v_t^i, s_t = i) \cdot p(s_{t+1} = j | s_t = i) \cdot p_t^i}{\sum_{k=0}^1 \sum_{i=0}^1 p(y_{t+1}, v_{t+1}^k | y_t, v_t^i, s_t = i) \cdot p(s_{t+1} = k | s_t = i) \cdot p_t^i}, \qquad j = 0, 1.$$

The third factor in the summand is known from the previous step of the recursion. The second factor is determined by the Markov transition matrix of the regime state process. For the first factor, since we allow for the possibility of more than a single jump per day, it is necessary to sum over the potential number of jumps,

$$p(y_{t+1}, v_{t+1}^j | y_t, v_t^i, s_t = i) = \sum_{k=0}^{\text{NJ}_{\max}} p(y_{t+1}, v_{t+1}^j | y_t, v_t^i, s_t = i, \text{NJ}_t = k) p(\text{NJ}_t = k)$$

where NJ_t is the number of potential jumps on day t, $\text{NJ}_{\max}$ is the maximum number of allowable jumps in a single day, and $p(\text{NJ}_t = k) = \lambda_1^k e^{-\lambda_1}/k!$ is given by the Poisson distribution with intensity λ_1. The distribution of $y_{t+1}, v_{t+1}^j | y_t, v_t^i, s_t = i, \text{NJ}_t = k$ is bivariate normal with mean and variance given by summing the means and variances of the diffusive part of the process and k jumps in (3).

It is sometimes useful to speak of the filtered regime state. By this we mean the expected value of s_t conditional on information available at time t,

$$\hat{s}_t = E_t(s_t) = p_t^0 \cdot 0 + p_t^1 \cdot 1 = p_t^1. \tag{5}$$

3.3 Maximum likelihood estimation

Having backed out volatility states and computed filtered regime state probabilities, computing the log likelihood is straightforward. Given a candidate parameter vector, θ,

$$\log L(\{y_t\}_{t=1}^n, \{IV_t\}_{t=1}^n; \theta) \approx \sum_{t=1}^{n-1} \sum_{i=0}^{1} \sum_{j=0}^{1} \Bigg[\log p(y_{t+1}, v_{t+1}^j | y_t, v_t^i, s_t = i) + \log p(s_{t+1} = j | s_t = i) \\ + \log p(s_t = i) + \log \Lambda_{t+1}^j \Bigg], \tag{6}$$

where $\Lambda_{t+1}^j = \left| dv_{t+1}^j / dIV_{t+1} \right|$ is the Jacobian corresponding to regime state j. Recall that the mapping from volatility state, v_t^j, to IV_t is given by $IV_t = \Gamma_j[\exp(v_t^j/2)]$. The Jacobian is obtained from the derivative of the inverse of this. As in the preceding subsection, $p(y_{t+1}, v_{t+1}^j | y_t, v_t^i, s_t = i)$ must be computed by summing across the number of potential jumps. The maximum likelihood estimator is obtained by optimizing (6) across candidate parameter vectors. Note that the inversion from option-implied volatility to volatility states must be computed at each evaluation of the likelihood function.

3.4 Diagnostics for assessing model fit

We examine diagnostics based on generalized residuals constructed using the probability integral transform, as proposed by Diebold, Gunther, and Tay (1998) and Diebold, Hahn, and Tay (1999). Let $\{z_t\}_{t=1}^n$ be a sequence of random vectors generated from some model with cumulative distribution functions $G_t(z|\mathcal{F}_{t-1})$ $(t = 1, \ldots, n)$. Let $u_t = G_t(z_t|\mathcal{F}_{t-1})$ denote the probability integral transform of z_t. Then, $\{u_t\}$ must be *i.i.d.* uniform$(0, 1)$.

Given data and a candidate model, it is typically straightforward to compute the corresponding sequence of probability integral transforms, $\{u_t\}$. Shortcomings in the model's ability to generate predictive distributions that reflect the observed data can be detected by looking at diagnostics based on this sequence.

It is often more useful to look at diagnostics based on

$$\widetilde{u}_t = \Phi^{-1}(u_t), \qquad t = 1, \ldots, n \tag{7}$$

where Φ is the standard normal distribution function. In this case, the transformed residuals $\{\widetilde{u}_t\}$ should be *i.i.d.* standard normal under the hypothesis of correct model specification. It is these that we shall refer to as generalized residuals.

For the models in this paper, the generalized residuals are computed in a manner similar to equation (6), using

$$u_{t+1} = \sum_{i=0}^{1} \sum_{j=0}^{1} P(y_{t+1}, v_{t+1}^j | y_t, v_t^i, s_t = i) \cdot p(s_{t+1} = j | s_t = i) \cdot p_t^i,$$

where $P(\cdot)$ denotes a cdf. These residuals correspond to the joint distribution of price and volatility innovations. Following Diebold, Hahn, and Tay (1999), we have found it more useful to study marginal residuals corresponding to price and volatility innovations separately,

$$u_{y,t+1} = \sum_{i=0}^{1} P(y_{t+1} | y_t, v_t^i, s_t = i) \cdot p(s_t = i)$$

$$u_{v,t+1} = \sum_{i=0}^{1} \sum_{j=0}^{1} P(v_{t+1}^j | y_t, v_t^i, s_t = i) \cdot p(s_{t+1} = j | s_t = i) \cdot p(s_t = i).$$

In the diagnostics reported in our application, we always use the generalized residuals obtained by applying the inverse normal cdf to these, $\tilde{u}_{y,t} = \Phi^{-1}(u_{y,t})$ and $\tilde{u}_{v,t} = \Phi^{-1}(u_{v,t})$.

Having constructed these generalized residuals, models can be assessed using standard time series techniques. In this paper, we look at normal-quantile plots and Jarque-Bera test statistics to assess normality, and correlograms and Ljung-Box test statistics to detect the presence of auto-correlation. We look at correlograms and Ljung-Box statistics for both the residuals and squared residuals (diagnostics based on the squared residuals allow us to detect unexplained stochastic volatility in returns and stochastic volatility of volatility).

4 EMPIRICAL RESULTS

4.1 Data

The application uses daily S&P 500 index (SPX) option data from Jan 1, 1993 through Dec 31, 2008 ($N = 4025$). These data were obtained directly from the CBOE. To address the issue of nonsynchronous closing times for the SPX index and option markets, SPX close prices are computed using put-call parity based on closing prices for at-the-money options (see, e.g., Ait-Sahalia and Lo 1998). Option prices are taken from the bid-ask midpoint at each day's close. Options with zero bid/ask prices or where the bid-ask midpoint is less than 0.125 are discarded. We also eliminate options violating the usual lower bound constraints. That is, we require $C(t, \tau, K) \geq \max(0, x_t \exp(-q_t \tau) - K \exp(-r_t \tau))$ and $P(t, \tau, K) \geq \max(0, K \exp(-r_t \tau) - x_t \exp(-q_t \tau))$ where $C(t, \tau, K)$ and $P(t, \tau, K)$ are the time t prices of call and put options with time-to-maturity τ and strike price K, x is the index price, q is the dividend payout rate, and r is the risk-free rate. Finally, we require that valid prices exist for at least two out-of-the-money call and put options for each day. Options are European, so there is no issue regarding early exercise premium.

Time-series of one-month risk-neutral volatility, skewness, and kurtosis are computed using SPX option prices following the model-free approach of Bakshi, Kapadia, and Madan (2003). Following Carr and Wu (2009), we use the two closest times to maturity greater than eight days and linearly interpolate to construct 30-day constant maturity series. Jiang and Tian (2007) report the possibility of large truncation and discretization errors in the VIX index. To reduce such errors, we follow the approach of Carr and Wu (2009) in interpolating/extrapolating option prices on a fine grid across moneyness. We use a grid with \$5 increments in strike prices, interpolating between observed prices based on Black-Scholes implied volatilities and extrapolating beyond the last observed strike price using that option's Black-Scholes implied volatility out to the last strike price at which the corresponding option price is 0.125 or greater (see Carr and Wu 2009 for additional detail).

Five-minute intraday S&P 500 index returns are used to compute measures of the variance

risk premium and jump risk. Since these variables are possibly related to option-implied skewness and kurtosis, we include them as control variables in the regressions of option-implied skewness and kurtosis on the regime state (Section 5). The high-frequency data were obtained from Tick-Data.com.

Following Andersen, Bollerslev, Diebold, and Ebens (2001), Andersen, Bollerslev, Diebold, and Labys (2003), and Barndorff-Nielsen and Shephard (2002), daily realized volatility is obtained by summing the squared intraday returns over each day,

$$\text{RV}_t^{(d)} \equiv \sum_{j=1}^{1/\Delta} \Big(y_{t-1+j\Delta} - y_{t-1+(j-1)\Delta} \Big)^2$$

where Δ is the sampling interval for the intraday data (we use five minute intervals). For each date t, we then sum the daily realized volatilities over the previous month (rolling samples), $\text{RV}_t \equiv \sum_{i=0}^{21} \text{RV}_{t-i}^{(d)}$ to get a monthly measure.

Following Carr and Wu (2009), we define the variance risk premium as the log difference between monthly realized variance and option-implied variance, $\text{VRP}_t \equiv \log\big(\text{RV}_t/\text{VIX}_t^2\big)$, where VIX_t is the VIX index, divided by $\sqrt{12}$ to get a monthly volatility measure comparable to RV_t. We use the log difference because we find that it provides a better measure than the difference in levels.

A measure of jump risk is obtained using the approach of Barndorff-Nielsen and Shephard (2004). The bipower variation is given by

$$\text{BV}_t^{(d)} \equiv \frac{\pi}{2} \sum_{j=2}^{1/\Delta} \big| y_{t-1+j\Delta} - y_{t-1+(j-1)\Delta} \big| \big| y_{t-1+(j-1)\Delta} - y_{t-1+(j-2)\Delta} \big| .$$

The daily jump variation is defined by subtracting the daily bipower variation from the daily realized volatility, $\text{JV}_t^{(d)} \equiv \max(\text{RV}_t^{(d)} - \text{BV}_t^{(d)}, 0)$. And, finally, for each date t we sum the daily jump variations over the previous month, $\text{JV}_t \equiv \sum_{i=0}^{21} \text{JV}_{t-i}^{(d)}$ to get a monthly measure.

One- and three-month Treasury bill rates (obtained from the Federal Reserve website), inter-

Table 2: Summary statistics.

The sample period covers January 1993 to December 2008. $\Delta \log(\text{SPX}_t)$ refers to S&P 500 index log returns. VIX_t is the VIX index, divided by $\sqrt{12}$ to get a monthly volatility measure for comparison. IV_t, SKEW_t, and KURT_t denote the one-month option-implied volatility, skewness, and kurtosis, computed using the model-free approach of Bakshi, Kapadia, and Madan (2003). RV_t and JV_t are the realized volatility and the jump variation, calculated using five-minute high-frequency data over the past 22 trading days. $\text{VRP}_t \equiv \log(\text{RV}_t/\text{VIX}_t^2)$ denotes the variance risk premium. $\text{AR}(i)$ means the i-lagged autocorrelation.

	Mean	Min	Max	STD	Skewness	Kurtosis	AR(1)	AR(5)	AR(22)
$\Delta \log(\text{SPX}_t)$	0.00	-0.09	0.11	0.01	-0.24	13.11	-0.06	-0.04	0.05
VIX_t (%)	5.73	2.69	23.34	2.40	2.23	11.91	0.98	0.94	0.78
IV_t (%)	5.62	2.22	24.46	2.55	2.26	12.31	0.98	0.94	0.79
SKEW_t	-1.69	-3.75	-0.41	0.49	-0.54	3.25	0.90	0.80	0.63
KURT_t	9.55	3.34	34.94	4.56	1.62	6.22	0.83	0.73	0.57
RV_t (%^{2})	21.79	1.58	457.27	40.45	6.72	57.28	1.00	0.97	0.68
JV_t (%^{2})	3.33	0.12	83.09	6.89	7.19	67.06	0.99	0.95	0.65
VRP_t	-0.88	-2.16	0.82	0.42	0.16	3.07	0.96	0.83	0.50

polated to match option maturity, are used as a proxy for the risk-free rate. Dividend rates are obtained from the Standard and Poor's information bulletin. Actual dividend payouts are used as a proxy for expected payouts.

Throughout, time is measured in trading days. SPX returns and option-implied moments are plotted in Figure 1. Summary statistics are reported in Table 2.

4.2   Parameter estimates and model comparisons

Parameter estimates and log likelihoods for all of the models under consideration are shown in Table 3. As found in previous work, including jumps in returns and volatility gives a large improvement in log likelihood relative to a model with no jumps (over 300 points in log likelihood). We also looked at models with jumps in returns only, but these models work less well (including jumps in volatility is critical). The SJ model (jumps scaled by volatility state) is strongly preferred over the UJ model (unscaled jumps). The improvement is over 40 points in log likelihood with the same number of parameters. In the scaled jump model jumps are larger on average when volatility is high than when it is low, whereas in the unscaled jump model jump sizes are identically distributed across time and unaffected by the volatility state. Although models with unscaled jumps have been

17

commonly used in previous work, (e.g., Pan 2002; Eraker, Johannes, and Polson 2003; Broadie, Chernov, and Johannes 2007), this specification is not consistent with the data.

Including the regime switching feature in the model provides additional large improvements. The SJ model with regime switching in volatility of volatility (SJ-RS-Vol) offers an improvement of 110 points in log likelihood relative to the corresponding model without regime switching (SJ) at the cost of four additional parameters. Including regime switching in jump intensity (SJ-RS-Jmp) provides an improvement of 67 points in log likelihood (relative to SJ), while including regime switching in leverage effect (SJ-RS-Lev) provides a smaller but still important gain of 32 points in log likelihood (relative to SJ).

A useful way to compare models is by using some form of information criterion. Akaike and Schwarz information criteria are common choices. These are based on comparison of log likelihood minus some penalty based on the number of free parameters in the model. The Akaike information criterion uses a penalty equal to the number of free parameters, while the Schwarz criterion (also known as the Bayesian information criterion) uses a penalty equal to the number of free parameters times $\ln(n)/2$, where n is the number of observations. For either of these, the results are the same: SJ models are always preferred over their UJ counterparts. For both SJ and UJ models, the regime-switching variants are all preferred over their non-regime-switching counterparts. Among SJ models, the ranking is SJ-RS-Vol $\succ$ SJ-RS-Jmp $\succ$ SJ-RS-Lev $\succ$ SJ. The rankings are all decisive and provide strong evidence in favor of the regime-switching models. We note also that while the basic SJ model does not include regime switching, it does include jumps and is itself vastly better than the models with unscaled jumps (or no jumps at all) commonly used in the existing literature.

Time series plots of filtered values for the regime-dependent parameters in several models are shown in Figure 3. For the models with regime switching in volatility of volatility, for example, σ_t takes on two possible values (σ^0 or σ^1). The filter described in Section 3.2 provides probabilities for each state conditional on available information, $p_t^j = p(s_t = j | \mathcal{F}_t)$. Filtered estimates of σ_t are thus given by $\hat{\sigma}_t = E_t[\sigma(s_t)] = p_t^0 \sigma^0 + p_t^1 \sigma^1$ $(t = 1, \ldots, n)$, with filtered estimates $\hat{\rho}_t$ and $\hat{\lambda}_{1t}$ computed analogously.

The sample period covers January 1993 through December 2008 ($N = 4025$). Standard errors are in parentheses. Time is measured in trading days.

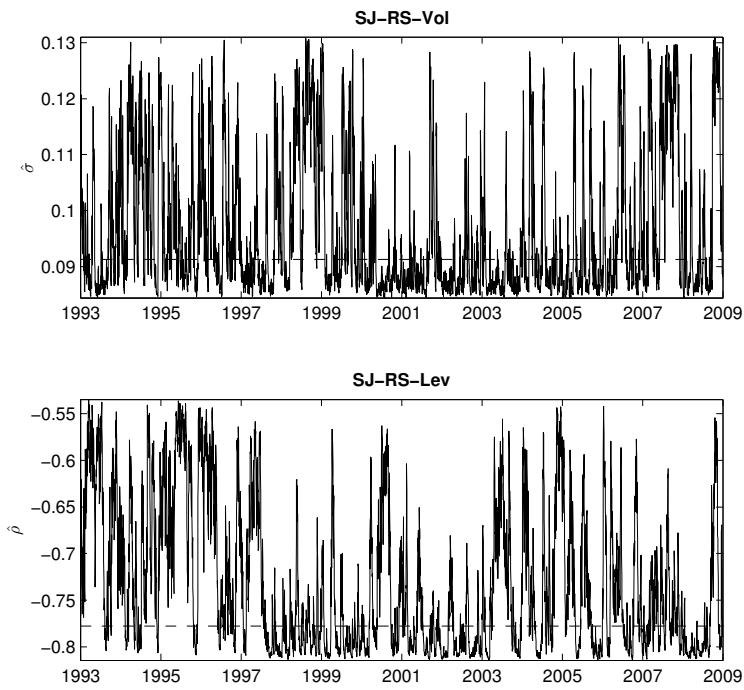|	No regime switching			RS-Jmp		RS-Vol		RS-Lev	
	SV	SJ	UJ	SJ	UJ	SJ	UJ	SJ	UJ
$\mu \times 10^4$	4.69	3.70	3.08	3.74	3.41	3.03	2.07	3.42	3.08
	(1.27)	(1.28)	(1.25)	(1.24)	(1.20)	(1.25)	(1.25)	(1.26)	(1.23)
$\kappa \times 10^3$	4.29	6.00	6.69	7.65	6.09	6.34	11.07	7.48	7.36
	(1.29)	(1.35)	(1.33)	(1.37)	(1.39)	(1.32)	(1.28)	(1.35)	(1.34)
$\overline{v}$	-9.71	-10.03	-9.70	-10.05	-10.04	-10.06	-9.63	-10.00	-9.80
	(0.55)	(0.38)	(0.36)	(0.29)	(0.41)	(0.32)	(0.20)	(0.32)	(0.33)
$\eta^0 \times 10^3$	3.01	2.84	3.84	4.22	4.56	2.86	4.18	2.49	2.72
	(0.25)	(0.36)	(0.35)	(0.40)	(0.35)	(0.42)	(0.36)	(0.47)	(0.44)
$\eta^1 \times 10^3$				-1.55	2.50	2.97	2.46	3.39	4.27
				(0.91)	(0.54)	(0.59)	(0.41)	(0.40)	(0.38)
$\sigma^0 \times 10^1$	1.32	0.91	1.23	0.94	1.19	0.84	1.18	0.99	1.25
	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
$\sigma^1 \times 10^1$						1.33	2.12		
						(0.06)	(0.06)		
ρ^0	-0.74	-0.78	-0.82	-0.78	-0.82	-0.81	-0.91	-0.53	-0.61
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.03)
ρ^1								-0.82	-0.85
								(0.01)	(0.01)
ρ_J		-0.71	-0.53	-0.72	-0.58	-0.70	-0.05	-0.77	-0.72
		(0.03)	(0.04)	(0.03)	(0.05)	(0.03)	(0.10)	(0.03)	(0.04)
λ_1^0		0.47	0.33	0.32	0.11	0.55	1.08	0.30	0.15
		(0.06)	(0.04)	(0.06)	(0.03)	(0.09)	(0.14)	(0.04)	(0.02)
λ_1^1				1.53	0.79				
				(0.25)	(0.12)				
$\mu_{1J} \times 10^2$		-1.13	0.05	-2.19	0.05	-4.52	0.08	-11.90	-0.05
		(6.59)	(0.04)	(7.25)	(0.04)	(6.38)	(0.02)	(8.78)	(0.07)
$\sigma_{1J} \times 10^1$		12.94	0.06	11.51	0.05	12.44	0.03	13.42	0.08
		(0.75)	(0.00)	(0.75)	(0.00)	(0.77)	(0.00)	(0.85)	(0.01)
$\mu_{2J} \times 10^1$		2.79	0.24	3.18	0.29	3.39	0.09	3.59	0.44
		(0.69)	(0.07)	(0.79)	(0.08)	(0.66)	(0.04)	(0.95)	(0.13)
σ_{2J}		1.54	0.15	1.37	0.15	1.15	0.05	1.72	0.21
		(0.07)	(0.01)	(0.08)	(0.01)	(0.07)	(0.01)	(0.09)	(0.01)
π_0				0.98	0.96	0.98	0.98	0.97	0.96
				(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
π_1				0.89	0.96	0.94	0.94	0.99	0.98
				(0.02)	(0.02)	(0.01)	(0.01)	(0.00)	(0.01)
log(L)	38,491	38,851	38,810	38,918	38,861	38,961	38,960	38,883	38,858

19

Figure 3: Time series of filtered values of state-dependent parameters. The dashed horizontal lines indicate the parameter estimates corresponding to the SJ model (which does not include regime switching in these parameters).

In the regime-switching models, the states are quite persistent. With SJ-RS-Vol, for example, the estimated persistence parameters are $\pi_0 = .98$ and $\pi_1 = .94$ (the probability of staying in regime 0 or regime 1, respectively, from one day to the next). The expected duration of stays is 50 days for regime 0 and 17 days for regime 1.

Expectations of future volatility of volatility and leverage effect are dramatically different depending on the current regime. The estimated volatility of volatility parameters are 0.084 for state 0 versus 0.133 for state 1 in the SJ-RS-Vol model. The estimated leverage parameters are -0.53 for state 0 versus -0.82 for state 1 in the SJ-RS-Lev model. Because of the high degree of persistence in regime states, these differences remain even over relatively long time horizons. This is not of purely theoretical interest. Any investors interested in the dynamics of volatility will find this information useful. For example, volatility options and swaps are highly dependent on the volatility of volatility. As discussed below, these persistent differences also affect the shape of the volatility smirk.

4.3 Diagnostics

QQ-plots and correlograms for return and volatility residuals are shown for several models in Figures 4 through 6. Jarque-Bera and Ljung-Box statistics are shown in Table 4.

The qq-plots and Jarque-Bera statistics provide information regarding the extent to which the models are able to capture distributional characteristics of the data. Including jumps in the model provides an enormous improvement over the model with no jumps, consistent with the previous literature. This is true regardless of the form of the jumps (SJ or UJ). However, the scaled jump models (SJ) do better than those with unscaled jumps (UJ), suggesting that they are better able to capture the non-normality in return and volatility innovations observed in the data. For the SJ models, including regime switching provides additional small improvements (for the UJ models, only regime switching in leverage effect helps much).

Turning now to the correlograms, all of the models do relatively well at eliminating autocorrelation in return residuals. On the other hand, all of the models fail with respect to the volatility
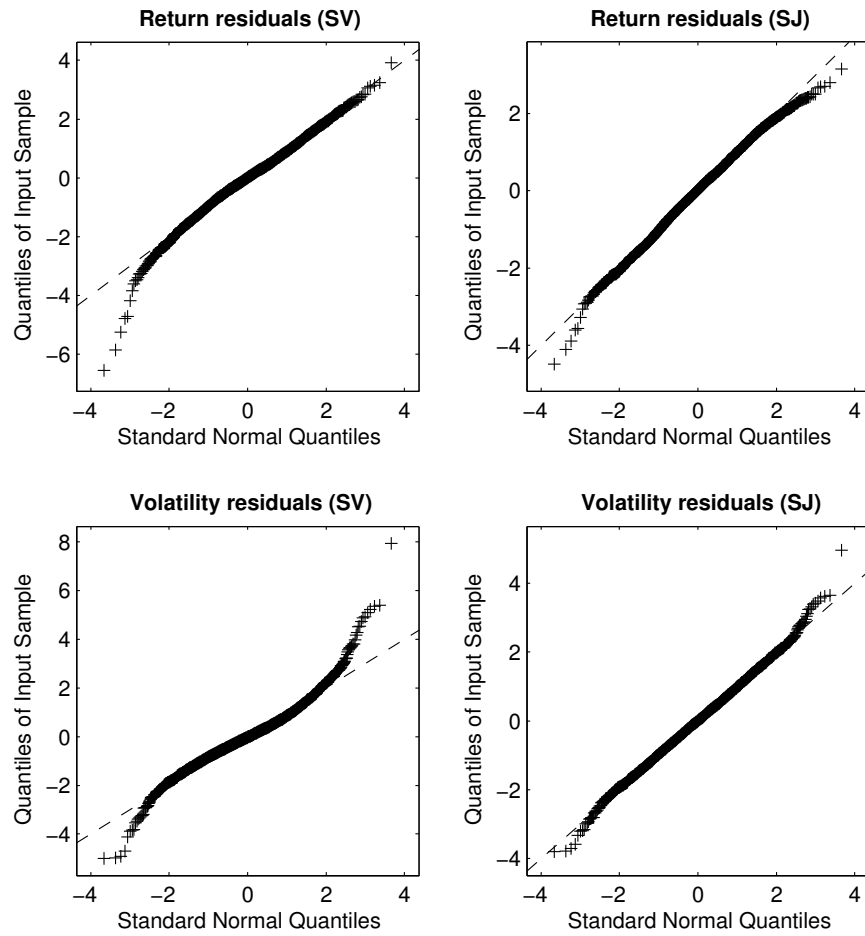
21

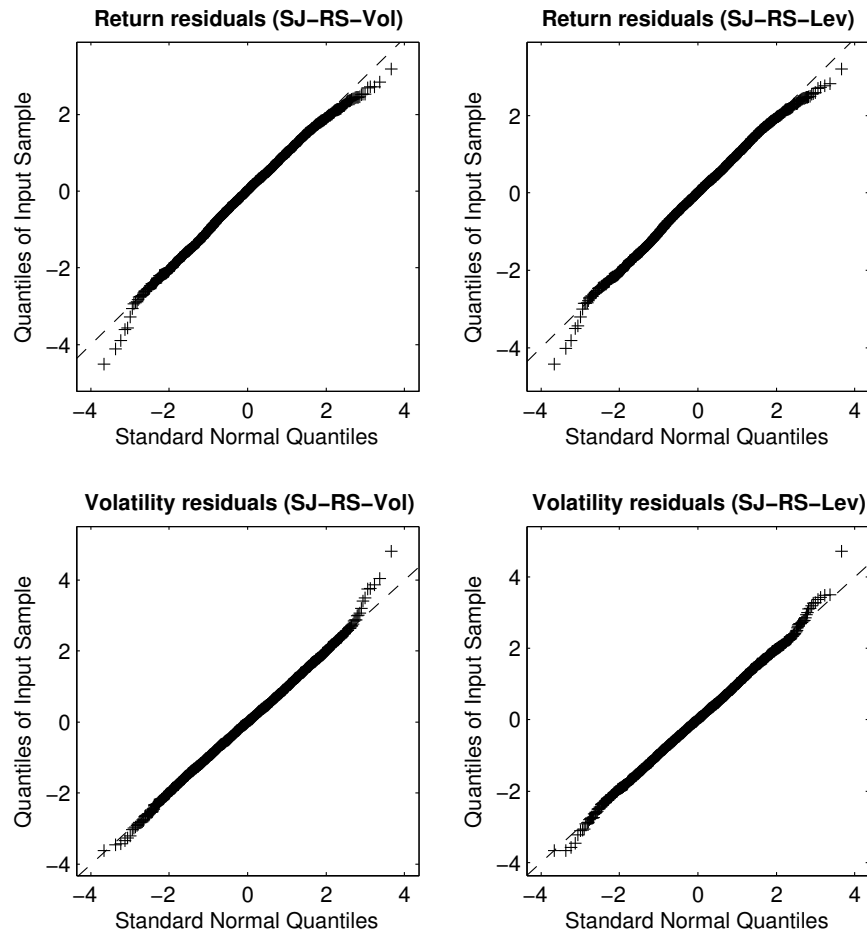Figure 4: QQ-plots for generalized residuals, SV and SJ models.

Figure 5: QQ-plots for generalized residuals, SJ-RS-Vol and SJ-RS-Lev models.
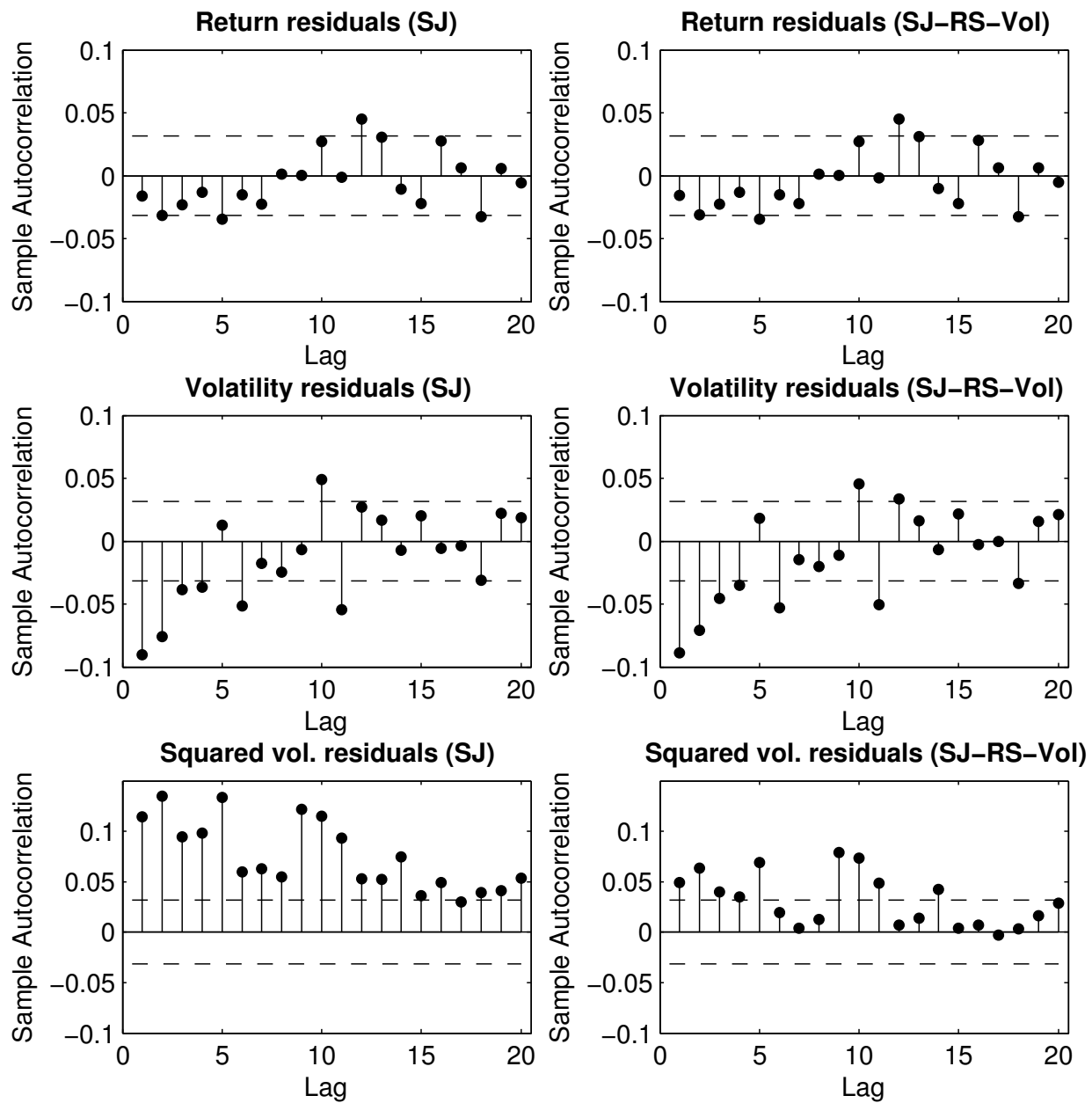
Figure 6: Correlograms for SJ and SJ-RS-Vol models. Dotted lines show the 95% confidence band.

Table 4: Diagnostic tests for generalized residuals.

Test statistics are shown with p-values in parentheses. The Jarque-Bera statistic is asymptotically $\chi^2(2)$, implying a 5% critical value of 5.99. The Ljung-Box statistic is asymptotically $\chi^2(20)$, implying a 5% critical value of 31.41.

	No regime switching			RS-Jmp		RS-Vol		RS-Lev	
	SV	SJ	UJ	SJ	UJ	SJ	UJ	SJ	UJ
Jarque-Bera statistics									
Return	445	20	113	19	117	17	245	10	50
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Volatility	2,367	22	121	7	71	20	318	8	30
	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)
Ljung-Box statistics (20 lags)									
Return	43	41	41	40	40	41	40	41	40
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Volatility	116	118	116	113	111	115	100	115	114
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Squared Vol.	419	552	424	145	220	128	74	559	446
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

residuals, which show significant negative autocorrelation for the first several lags. This is suggestive of the existence of a second volatility factor omitted by the model (e.g., Gallant, Hsu, and Tauchen 1999; Chacko and Viceira 2003; Christoffersen, Heston, and Jacobs 2009). All of the models also show a large amount of unexplained autocorrelation in squared volatility residuals. This is suggestive of stochastic volatility of volatility. Although the regime-switching model implements a very simple form of stochastic volatility of volatility, it is a substantial improvement over the other models (though still defective). Regime-switching in jump intensity also helps some in this direction.

Note that since the generalized residuals are constructed using estimated model parameters, the standard asymptotics for the test statistics reported in Table 4 are not exact (see e.g., Bai 2003 for a discussion of this issue). However, the sample size available to us is large, so parameter uncertainty would not be expected to have much impact. Monte Carlo studies that we have performed confirm that this is indeed the case. As discussed in Section 3.4, our primary interest here is using these diagnostics as an aid in the process of model exploration. If one were more interested in formal specification testing, one might alternatively try applying tests such as those

proposed by Bai (2003), Duan (2003), or Hong and Li (2005).

And finally, we note that while these diagnostics are useful in understanding directions in which the models are more or less successful in capturing various features of the data, and are informative in the process of model exploration, they do not provide a generically reliable basis for preferential ranking of the models. For this, a formal loss function is required (see, e.g., Diebold, Gunther, and Tay 1998).

5 EXPLANATORY POWER FOR OPTION-IMPLIED SKEWNESS AND KURTOSIS

In this section, we examine whether the regime states implied by our regime-switching models have any explanatory power for the shape (e.g., third and fourth moments) of the risk-neutral distribution. Recall that option-implied skewness and kurtosis are not used in the estimation, which makes this diagnostic meaningful.

Given a model for the price of an asset and values for any relevant state variables, one can easily obtain the model-implied distribution of returns at any horizon using Monte Carlo methods. We will refer to the third and fourth moments of these distributions as model-implied skewness and kurtosis. Given time series for option-implied volatility and skewness (or kurtosis) and a model with two state variables (such as our regime-switching models), it is not difficult to find model parameters and values for the state variables such that model-implied and option-implied characteristics match exactly. To the extent that implied physical and risk-neutral model parameters differ, these differences would typically be attributed to risk premia. However, since this approach confounds the information from physical dynamics and option prices, it is difficult to tell if the implied variation in states (e.g., leverage effect or volatility of volatility) is truly evident in the physical model or just an artifact of trying to match the shape of the volatility smile. The problem here is a pervasive one in the option-pricing literature. It is possible to fit even a badly misspecified model to the data. The fitted model is likely to generate spurious risk premia, but the exact source

and extent of the misspecification are difficult to diagnose.

The approach that we take in this paper is different. While we match option-implied volatility exactly, we do not use the information on option-implied skewness and kurtosis in the estimation. In particular, we make no effort to use the information in option prices to come up with model parameters and regime states such that option-implied and model-implied skewness or kurtosis match exactly. We also do not look for risk premia which cause the models to fit the shape of the implied volatility smirk on average. Rather, we fit the models using only information from returns and implied volatility. We are then able to make a clear and unambiguous examination of the extent to which the implied states are informative about variation across time in observed option-implied skewness and kurtosis.

The idea is to regress option-implied skewness and kurtosis on the regime state, controlling for other possible determinants. As described in Section 3, it is not possible based on observed data to determine the regime state conclusively. The available information consists of filtered probabilities, $p_t^j = p(s_t = j | \mathcal{F}_t)$ $(j = 0, 1)$. The operational regression equations are thus

$$
\begin{aligned}
\text{SKEW}_t &= \beta_0 + \beta_1 \hat{s}_t^{\text{RS-Vol}} + \beta_2 \hat{s}_t^{\text{RS-Lev}} + \beta_3 \hat{s}_t^{\text{RS-Jmp}} + \beta_4 \log \text{VIX}_t + \beta_5 \log \text{JV}_t + \beta_6 \text{VRP}_t + \varepsilon_t \\
\text{KURT}_t &= \beta_0 + \beta_1 \hat{s}_t^{\text{RS-Vol}} + \beta_2 \hat{s}_t^{\text{RS-Lev}} + \beta_3 \hat{s}_t^{\text{RS-Jmp}} + \beta_4 \log \text{VIX}_t + \beta_5 \log \text{JV}_t + \beta_6 \text{VRP}_t + \varepsilon_t
\end{aligned}
\tag{8}
$$

where $\hat{s}_t^{\text{RS-Vol}}$, $\hat{s}_t^{\text{RS-Lev}}$ and $\hat{s}_t^{\text{RS-Jmp}}$ denote the filtered states under the models with regime switching in volatility of volatility, leverage effect and jump intensity, respectively (as defined in Equation (5)). In interpreting the regression coefficients note that, for example, $\hat{s}_t = 0.30$ implies that the probability that $s_t = 1$ conditional on data through time t is 30%. We also include several control variables. VIX_t is the VIX index, which serves as a proxy for the volatility state. VRP_t is the variance risk premium, which Bollerslev, Gibson, and Zhou (2011) argue reflects information on variation across time in investors' risk aversion. Option-implied skewness and kurtosis may also be affected by the market expectation of jump risk. We proxy for this by using a measure of realized jump variation, JV_t. A more detailed description of these variables may be found in Section 4.1. We first run the regressions on the control variables alone, and then add the regime states to the regressions to see if there is a significant increase in explanatory power. We report Newey-West

robust t-statistics over eight lags (Newey and West 1987) and adjusted R^2. For completeness, results are reported for the regressions based on filtered regime states from both SJ and UJ models.

Table 5 shows the regressions for option-implied skewness. We first discuss the contributions of the control variables toward explaining option-implied skewness. The volatility state (VIX) is highly significant for all models and regardless of whether the regime state is included in the regression. The coefficient is always positive, meaning that low volatility states are associated with more left-skewed return distributions. This is consistent with Dennis and Mayhew (2002), who report a similar finding for individual stocks. Jump risk (JV) is also highly significant for all models. The coefficient is always negative, suggesting that greater jump risk is associated with more left-skewed return distributions under the risk-neutral measure. This is consistent with intuition. In contrast, we do not find the variance risk premium (VRP) to be significant when the other control variables are included.

Although the volatility state is highly significant, it has relatively low explanatory power for option-implied skewness, with an adjusted R^2 of 7.6%. All of the control variables together are able to explain only 9.2% of the time-variation in option-implied skewness. This low explanatory power suggests that models with only a single state variable (volatility) will not be able to match time-varying patterns in the Black-Scholes volatility smirk realistically.

Including the filtered regime state in the regressions provides a dramatic improvement in explanatory power. SJ models always outperform the corresponding UJ models. Including the regime state from the models with regime-switching in volatility of volatility (RS-Vol) is slightly better than including the regime state from the models with regime-switching in leverage effect (RS-Lev). But the clear winner is the model which includes regime states from both of these. The states corresponding to volatility of volatility and regime switching each have substantial explanatory power when included in the regression alone. But, these two variables serve as complementary sources of information. For the regressions using states generated from the SJ models (together with all control variables), the adjusted R^2 is 18-19% when either the RS-Vol or RS-Lev states are included individually, but R^2 jumps to over 32% when both are included. The rankings here are

Table 5: Regressions for option-implied skewness.

The table reports the results of the following regression,

$$\mathrm{SKEW}_t = \beta_0 + \beta_1 \hat{s}_t^{\mathrm{RS\text{-}Vol}} + \beta_2 \hat{s}_t^{\mathrm{RS\text{-}Lev}} + \beta_4 \log \mathrm{VIX}_t + + \beta_5 \log \mathrm{JV}_t + \beta_6 \mathrm{VRP}_t + \varepsilon_t.$$

Newey-West robust t-statistics over eight lags are shown in parentheses. The sample period covers January 1993 to December 2008. SKEW_t denotes the one-month option-implied skewness. Filtered regime states are $\hat{s}_t^{\mathrm{RS\text{-}Vol}}$ and $\hat{s}_t^{\mathrm{RS\text{-}Lev}}$ for volatility of volatility and leverage effect respectively. We also performed regression including the regime states from the RS-Jmp models, however these were never significant. We do not report these results in the table to save space, but they are available upon request. The control variables are the VIX index, jump variation (JV), and variance risk premium (VRP). Results are shown for filtered states from both SJ (scaled jumps) and UJ (unscaled jumps) models.

Constant		$\widehat{s}_t^{\mathrm{RS\text{-}Vol}}$		$\widehat{s}_t^{\mathrm{RS\text{-}Lev}}$		$\log \mathrm{VIX}_t$		$\log \mathrm{JV}_t$		VRP_t		Adj. R^2
coeff	t-stat	coeff	t-stat	coeff	t-stat	coeff	t-stat	coeff	t-stat	coeff	t-stat	
Control variables only												
-2.78	(-16.43)					0.37	(6.53)					7.6%
-3.38	(-12.02)					0.60	(5.94)	-0.10	(-3.21)			9.2%
-2.75	(-13.84)					0.36	(4.97)			0.000	(-0.32)	7.6%
-3.39	(-11.76)					0.60	(5.37)	-0.10	(-2.19)	0.008	(0.11)	9.2%
Control variables plus filtered states from SJ models												
-2.82	(-18.07)	-0.56	(-9.02)			0.44	(8.48)					16.2%
-3.63	(-14.57)	-0.61	(-9.98)			0.74	(8.42)	-0.13	(-5.01)			19.0%
-3.17	(-16.19)	-0.59	(-9.62)			0.51	(8.87)			-0.167	(-3.73)	17.9%
-3.62	(-14.45)	-0.61	(-9.95)			0.74	(7.72)	-0.12	(-3.18)	-0.007	(-0.11)	19.0%
-2.82	(-15.39)			-0.62	(-7.37)	0.54	(7.15)					17.5%
-3.25	(-12.37)			-0.60	(-7.17)	0.69	(7.10)	-0.07	(-2.23)			18.3%
-3.01	(-15.20)			-0.61	(-7.31)	0.57	(7.87)			-0.092	(-1.91)	18.0%
-3.25	(-11.86)			-0.60	(-7.17)	0.69	(6.18)	-0.07	(-1.45)	-0.007	(-0.10)	18.3%
-2.87	(-18.05)	-0.71	(-11.53)	-0.76	(-10.66)	0.65	(10.36)					30.6%
-3.52	(-16.28)	-0.74	(-12.04)	-0.74	(-10.77)	0.89	(11.23)	-0.10	(-4.02)			32.3%
-3.17	(-18.84)	-0.73	(-11.88)	-0.75	(-10.90)	0.71	(12.00)			-0.143	(-3.35)	31.8%
-3.50	(-15.73)	-0.74	(-12.03)	-0.74	(-10.77)	0.87	(9.82)	-0.09	(-2.40)	-0.028	(-0.45)	32.3%
Control variables plus filtered states from UJ models												
-2.62	(-16.06)	-0.44	(-7.48)			0.36	(6.81)					14.2%
-3.41	(-13.30)	-0.48	(-8.59)			0.66	(7.33)	-0.13	(-4.97)			17.0%
-2.96	(-14.77)	-0.47	(-8.14)			0.43	(7.36)			-0.168	(-3.68)	15.9%
-3.41	(-13.12)	-0.48	(-8.56)			0.66	(6.65)	-0.13	(-3.14)	-0.007	(-0.10)	17.0%
-2.93	(-15.91)			-0.50	(-6.57)	0.53	(7.43)					14.0%
-3.45	(-12.68)			-0.48	(-6.56)	0.73	(7.19)	-0.08	(-2.77)			15.2%
-3.16	(-15.27)			-0.49	(-6.61)	0.58	(8.05)			-0.110	(-2.32)	14.8%
-3.45	(-12.27)			-0.48	(-6.55)	0.72	(6.39)	-0.08	(-1.82)	-0.005	(-0.07)	15.2%
-2.76	(-16.64)	-0.57	(-10.09)	-0.65	(-9.82)	0.56	(9.35)					24.7%
-3.52	(-15.05)	-0.61	(-11.07)	-0.64	(-10.25)	0.85	(10.04)	-0.12	(-4.79)			27.2%
-3.11	(-17.14)	-0.60	(-10.71)	-0.65	(-10.21)	0.64	(10.76)			-0.169	(-3.88)	26.4%
-3.50	(-14.63)	-0.62	(-11.03)	-0.64	(-10.22)	0.83	(8.92)	-0.11	(-2.91)	-0.028	(-0.44)	27.2%

consistent with the findings reported in Sections 4.2 and 4.3.

For the SJ-RS-Vol model (regime-switching in volatility of volatility), the coefficient on the regime state is highly significant and in the expected direction. For the full regression (including regime state and all control variables), the estimated slope coefficient for the regime state is -0.61, indicating that a change from state 0 to state 1 is associated with a 0.61 decrease in skewness (i.e., the distibution is substantially more left skewed in the high volatility of volatility state). The t-statistic is -9.95, corresponding to a p-value of around 10^{-22}. This model has good explanatory power, with an adjusted R^2 of 19.0% (versus 9.2% for the control variables alone). These results are both statistically and economically significant.

Results for the SJ-RS-Lev model (regime switching in leverage effect) are similar. In particular, the coefficient on the regime state is highly significant and in the expected direction. The t-statistic associated with this parameter is -7.17, corresponding to a p-value of around 10^{-15}. The model has an adjusted R^2 of 18.3%.

Table 6 shows analagous regressions for option-implied kurtosis. The results are qualitatively similar to those for option-implied skewness. The volatility state (VIX) is highly significant for all models and regardless of whether the regime state is included in the regression. The coefficient is negative, implying that low volatility states are associated with fatter-tailed return distributions. Jump risk (JV) is positively related to kurtosis. That is, the risk-neutral distribution tends to be more fat-tailed when jump risk is high, consistent with intuition. Although the variance risk premium (VRP) is highly significant when jump risk is omitted from the regression, it has little explanatory power when jump risk is included.

Including the regime state in the regression provides additional improvement in explanatory power. As with option-implied skewness, SJ models always outperform the corresponding UJ models. Slope coefficients for the regime state have the expected signs. That is, higher volatility of volatility and stronger leverage effect are both associated with more kurtotic return distributions.

Including the regime state corresponding to regime switching in leverage effect does better here than including the regime state corresponding to regime switching in volatility of volatility

Table 6: Regressions for option-implied kurtosis.

The table reports the results of the following regression,

$$\text{KURT}_t = \beta_0 + \beta_1 \hat{s}_t^{\text{RS-Vol}} + \beta_2 \hat{s}_t^{\text{RS-Lev}} + \beta_4 \log \text{VIX}_t + \beta_5 \log \text{JV}_t + \beta_6 \text{VRP}_t + \varepsilon_t.$$

Newey-West robust t-statistics over eight lags are shown in parentheses. The sample period covers January 1993 to December 2008. KURT_t denotes the one-month option-implied kurtosis. Filtered regime states are $\hat{s}_t^{\text{RS-Vol}}$ and $\hat{s}_t^{\text{RS-Lev}}$ for volatility of volatility and leverage effect respectively. We also performed regression including the regime states from the RS-Jmp models, however these were never significant. We do not report these results in the table to save space, but they are available upon request. The control variables are the VIX index, jump variation (JV), and variance risk premium (VRP). Results are shown for filtered states from both SJ (scaled jumps) and UJ (unscaled jumps) models.

| Constant | | $\hat{s}_t^{\text{RS-Vol}}$ | | $\hat{s}_t^{\text{RS-Lev}}$ | | $\log \text{VIX}_t$ | | $\log \text{JV}_t$ | | VRP_t | | Adj. R^2 |
coeff	t-stat	coeff	t-stat	coeff	t-stat	coeff	t-stat	coeff	t-stat	coeff	t-stat	
Control variables only												
20.22	(12.46)					-3.66	(-6.96)					8.4%
32.99	(13.34)					-8.40	(-9.68)	2.02	(8.02)			16.6%
19.23	(11.34)					-3.21	(-5.64)			0.02	(2.22)	8.7%
32.49	(13.23)					-7.98	(-8.74)	1.72	(4.94)	0.65	(1.08)	16.7%
Control variables plus filtered states from SJ models												
20.31	(12.67)	1.44	(2.44)			-3.81	(-7.45)					9.0%
33.86	(14.27)	2.18	(3.90)			-8.91	(-10.83)	2.14	(8.93)			18.0%
26.82	(14.40)	1.99	(3.49)			-5.18	(-9.70)			3.03	(7.22)	15.5%
33.33	(14.19)	2.19	(3.93)			-8.46	(-9.80)	1.82	(5.38)	0.70	(1.18)	18.1%
20.51	(12.11)			5.14	(7.03)	-5.00	(-7.57)					16.2%
31.98	(13.75)			4.60	(7.13)	-9.14	(-10.79)	1.82	(7.12)			22.8%
26.15	(14.73)			4.84	(7.30)	-6.06	(-10.03)			2.65	(6.33)	21.2%
31.39	(13.39)			4.62	(7.16)	-8.65	(-9.55)	1.47	(4.20)	0.76	(1.33)	22.9%
20.69	(12.63)	2.50	(4.31)	5.63	(8.11)	-5.40	(-8.68)					18.0%
33.09	(15.13)	3.09	(5.47)	5.17	(8.60)	-9.95	(-12.69)	1.96	(8.16)			25.6%
26.81	(15.80)	2.97	(5.22)	5.40	(8.74)	-6.62	(-11.87)			2.86	(7.03)	23.8%
32.44	(14.75)	3.11	(5.52)	5.19	(8.66)	-9.41	(-11.24)	1.57	(4.71)	0.85	(1.51)	25.8%
Control variables plus filtered states from UJ models												
20.00	(12.07)	0.59	(1.07)			-3.64	(-6.94)					8.5%
33.07	(13.62)	1.34	(2.64)			-8.58	(-10.20)	2.12	(8.74)			17.3%
26.18	(13.80)	1.14	(2.19)			-4.91	(-9.05)			2.99	(7.06)	14.8%
32.55	(13.52)	1.36	(2.68)			-8.13	(-9.20)	1.80	(5.28)	0.69	(1.15)	17.4%
21.47	(12.52)			4.21	(6.24)	-5.00	(-7.85)					13.7%
33.56	(13.99)			3.91	(6.57)	-9.43	(-10.79)	1.93	(7.63)			21.1%
27.40	(14.89)			4.08	(6.64)	-6.16	(-10.24)			2.80	(6.73)	19.3%
32.99	(13.74)			3.93	(6.59)	-8.95	(-9.73)	1.58	(4.56)	0.75	(1.32)	21.3%
21.01	(12.26)	1.54	(2.88)	4.63	(7.10)	-5.09	(-8.34)					14.6%
33.79	(14.66)	2.25	(4.48)	4.49	(7.87)	-9.89	(-12.00)	2.07	(8.58)			23.0%
27.21	(15.19)	2.10	(4.10)	4.64	(7.87)	-6.36	(-11.31)			3.00	(7.33)	20.9%
33.16	(14.40)	2.27	(4.54)	4.52	(7.90)	-9.35	(-10.81)	1.69	(5.06)	0.84	(1.48)	23.2%

(adjusted R^2 of 22.9% for SJ-RS-Lev versus 18.1% for SJ-RS-Vol). As is the case with implied skewness, including both regime states in the regression is better yet, with an R^2 of 25.8%. In the full model, slope coefficients are 3.11 and 5.19 (with t-statistics of 5.52 and 8.66) for the SJ-RS-Vol and SJ-RS-Lev states respectively. These are both economically and statistically significant.

6 CONCLUSION

This paper proposes a new class of models that layer regime switching on top of a standard stochastic volatility model with jumps in both returns and volatility. Motivated by the time-varying nature of option-implied skewness and kurtosis that is a prominent feature of observed data, we allow for regime switching in three parameters of the basic model: volatility of volatility, leverage effect, and jump intensity. All three parameters play important roles in determining the skewness and kurtosis of returns.

The application looks at SPX index and option data. In the first step of the analysis, we estimate the models relying only upon observations of the index price and option-implied volatility. This allows us to use observations on option-implied skewness and kurtosis for diagnostic purposes.

The models with regime switching fit the data dramatically better than those without regime switching. Accounting for time-variation in the volatility of volatility, leverage effect or jump intensity not only increases log likelihood significantly, but also provides improvements in other diagnostics of model fit.

While this first step of the analysis provides strong evidence of time-variation in model characteristics that are related to the shape of return distributions under the physical measure, the second step investigates the relationship between these characteristics and the skewness and kurtosis of returns under the risk-neutral measure (i.e, implied by observed option prices). This part of the study is carried out by running regressions of option-implied skewness and kurtosis on the regime states extracted in the estimation step (and several control variables).

We find that there is a strong relationship between model characteristics related to the shape

of return distributions under the physical measure and the skewness and kurtosis of return distributions implied by option prices. Although commonly-used models in the existing literature include volatility as the only state variable, this variable has an adjusted R^2 of only 7.6% for option-implied skewness. In contrast, regressions that include either the regime state corresponding to volatility of volatility or leverage effect (together with the control variables) have an adjusted R^2 of 18-19%. When both states are included in the regression, the adjusted R^2 is over 32%. Slope coefficients on the regime states are highly significant and in the expected direction. Option-implied return distributions tend to be more left-skewed and leptokurtic when volatility of volatility is high or leverage effect is strong. Results for option-implied kurtosis are qualitatively similar though weaker.

We also find that the specification of jump distributions is important both in fitting the time-series data of returns and volatility as well as in explaining variation in option-implied skewness and kurtosis. Volatility-scaled jumps (SJ) outperform unscaled jumps (UJ) in both respects.

References

Ait-Sahalia, Y., and Kimmel, R. (2007), "Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Financial Economics*, 83, 413–452.

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

Andersen, T., Bollerslev, T., Diebold, F., Ebens, H. (2001), "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Bai, J. (2003), "Testing Parametric Conditional Distributions of Dynamic Models," *Review of Economics and Statistics*, 85, 531–549.

Bakshi, G., and Kapadia, N. (2003), "Delta-Hedged Gains and the Negative Market Volatility Risk Premium," *Review of Financial Studies*, 16, 527–566.

Bakshi, G., Kapadia, N., and Madan, D. (2003), "Stock Return Characteristics, Skew Laws, and Differential Pricing of Individual Equity Options," *Review of Financial Studies*, 16, 101–143.

Barndorff-Nielsen, O. E., and Shephard, N. (2002), "Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models," *Journal of Royal Statistical Society B*, 64, 253–280.

——— (2004), "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–37.

Bollerslev, T., Gibson, M., and Zhou, H. (2011), "Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities," *Journal of Econometrics*, 160, 235–245.

Broadie, M., Chernov, M., and Johannes, M. (2007), "Model Specification and Risk Premia: Evidence from Futures Options," *Journal of Finance*, 62, 1453–1490.

Carr, P., and Wu, L. (2007), "Stochastic Skew in Currency Options," *Journal of Financial Economics*, 86, 213–247.

——— (2009), "Variance Risk Premiums," *Review of Financial Studies*, 22, 1311–1341.

Chacko, G., and Viceira, L. (2003), "Spectral GMM Estimation of Continuous-Time Processes," *Journal of Econometrics*, 116, 259–292.

Chernov, M., and Ghysels, E. (2000), "A Study towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation," *Journal of Financial Economics*, 56, 407–458.

Christoffersen, P., Heston, S., and Jacobs, K. (2009), "The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well," *Management Science*, 55, 1914–1932.

Christoffersen, P., Jacobs, K., and Mimouni, K. (2010), "Models for S&P500 Dynamics: Evidence from Realized Volatility, Daily Returns, and Option Prices," *Review of Financial Studies*, 23, 3141–3189.

Coval, J., and Shumway, T. (2001), "Expected Option Returns," *Journal of Finance*, 56, 983–1009.

Das, S. R., and Sundaram, R. K. (1999), "Of Smiles and Smirks: A Term Structure Perspective," *Journal of Financial and Quantitative Analysis*, 34, 211–239.

Dennis, P., and Mayhew, S. (2002), "Risk-Neutral Skewness: Evidence from Stock Options," *Journal of Financial and Quantitative Analysis*, 37, 471–493.

Diebold, F., Gunther, T., and Tay, A. (1998), "Evaluating density forecasts with applications to financial risk management," *International Economic Review*, 39, 863–883.

Diebold, F., Hahn, J., and Tay, A. (1999), "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High Frequency Returns on Foreign Exchange," *Review of Economics and Statistics*, 81, 661–673.

Duan, J.-C. (2003), "A Specification Test for Time-Series Models by a Normality Transformation," Working paper, University of Toronto.

Eraker, B. (2004), "Do Equity Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *Journal of Finance*, 59, 1367–1403.

Eraker, B., Johannes, M., and Polson, N. (2003), "The Impact of Jumps in Volatility and Returns," *Journal of Finance*, 58, 1269–1300.

Gallant, A. R., Hsu, C., and Tauchen, G. (1999), "Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance," *Review of Economics and Statistics*, 81, 617–631.

Harvey, C. R., and Siddique, A. (1999), "Autoregressive Conditional skewness," *Journal of Financial and Quantitative Analysis*, 34, 465–487.

Heston, S. L. (1993), "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327–343.

Hong, Y., and Li, H. (2005), "Nonparametric specification testing for continuous-time models with applications to term structure of interest rates," *Review of Financial Studies*, 18(1), 37–84.

Jiang, G., and Tian, Y. (2007), "Extracting Model-Free Volatility from Option Prices: An Examination of the VIX Index," *Journal of Derivatives*, 14, 35–60.

Johnson, T. C. (2002), "Volatility, Momentum, and Time-Varying Skewness in Foreign Exchange Returns," *Journal of Business and Economic Statistics*, 20, 390–411.

Jones, C. (2003), "The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets," *Journal of Econometrics*, 116, 181–224.

Newey, W. K., and West, K. D. (1987), "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Pan, J. (2002), "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study," *Journal of Financial Economics*, 63, 3–50.

Santa-Clara, P., and Yan, S. (2010), "Crashes, Volatility, and the Equity Premium: Lessons from S&P 500 Options," *Journal of Business and Economic Statistics*, 92, 435–451.