

Stochastic Volatility and Jumps: Exponentially Affine Yes or No? An Empirical Analysis of S&P500 Dynamics

Katja Ignatieva[‡] Paulo J. M. Rodrigues[‡] Norman Seeger[‡]

This version: April 30, 2009

Abstract

This paper analyzes exponentially affine and non-affine stochastic volatility models with jumps in returns and volatility. Markov Chain Monte Carlo (MCMC) technique is applied within a Bayesian inference framework to estimate model parameters and latent variables using daily returns from the S&P 500 stock index. There are two approaches to overcome the problem of misspecification of the square root stochastic volatility model. The first approach proposed by Christoffersen, Jacobs and Mimouni (2008) suggests to investigate some non-affine alternatives of the volatility process. The second approach consists in examining more heavily parameterized models by adding jumps to the return and possibly to the volatility process. The aim of this paper is to combine both model frameworks and to test whether the class of affine models is outperformed by the class of non-affine models if we include jumps into the stochastic processes. We conclude that the non-affine model structure have promising statistical properties and are worth further investigations. Further, we find affine models with jump components that perform similar to the non affine models without jump components. Since non affine models yield economically unrealistic parameter estimates, and research is rather developed for the affine model structures we have a tendency to prefer the affine jump diffusion models.

Keywords: Stochastic volatility, Markov Chain Monte Carlo (MCMC), Bayesian inference Deviance information criteria (DIC), Bayes factor.

JEL: G11, G12

[‡] House of Finance, Goethe University, D-60323 Frankfurt am Main, Germany. E-mail: ignatieva|rodrigues|seeger@finance.uni-frankfurt.de

1 Introduction

This paper analyzes exponentially affine and non-affine stochastic volatility models with jumps in returns, and jumps in returns and volatility. One of the main research topics in finance is to find models which fully capture the statistical properties of asset returns. In particular, the model framework for stock prices has evolved dramatically over the past decades. Starting from the fairly simple assumption of constant volatility in the Black-Scholes-Merton framework, model complexity and sophistication have increased in order to capture stylized facts of the data. These include for example crash events in stock markets, volatility clustering, smile patterns in options data, and the leverage effect. Including further risk factors such as stochastic volatility, jumps in returns and jumps in volatility allows the modeling of these stylized facts in a comprehensive way. A general framework that allows for jumps in returns, stochastic volatility, and jumps in returns and volatility can be found in Duffie, Pan and Singleton (2000). This model nests the specifications analyzed by Merton (1976), Heston (1993), and Bates (1996).

Empirical testing of these continuous time stock price processes has been carried out in several studies. Stochastic volatility models were analyzed in Jacquier, Polson, and Rossi (1994, 2004). Models with jumps in returns and stochastic volatility can be found in Pan (2002), Bakshi, Cao and Chen (1997) and Chernov, Gallant and Ghysels (2003). Conclusions drawn from these studies are that the Heston (1993) type stochastic volatility model is severely misspecified. Although models including additionally jumps in the return process have a better empirical performance compared to stochastic volatility models, it has been found by, e.g., Bakshi et al. (1997), Bates (2000), and Pan (2002) that these models are still potentially misspecified.

Recent work by Eraker, Johannes and Polson (2003) and Christoffersen et al. (2008) test two approaches to overcome the problem of misspecification. The first study uses the framework laid out by Duffie et al. (2000) and analyzes several models using the S&P 500 and the NASDAQ 100 stock market indices. The models under consideration are the stochastic volatility model, the stochastic volatility model containing jumps in returns, and two stochastic volatility models with jumps in returns and

volatility. The first stochastic volatility model assumes that jumps in returns and volatility are correlated, whereas the second models assumes jump components to be independent. Eraker et al. (2003) find significant better performance of models including jumps in volatility compared to models without jumps in volatility. The second approach proposed by Christoffersen et al. (2008) suggests to leave the class of the exponentially affine models by investigating some alternative non-affine specifications for the volatility process without considering jumps. They test several alternative specifications for the stochastic volatility process using the data of S&P 500 index returns. In particular, they change the speed of mean reversion and the diffusion term in the specification of the stochastic volatility process. They find significant improvements resulting in less misspecification and better empirical performance of stochastic volatility models when leaving the class of affine models.

A further distinctive feature of both studies is related to the estimation approach. The study by Christoffersen et al. (2008) uses the Maximum Likelihood Importance Sampling (MLIS) algorithm for parameter estimation. This method uses a resampling algorithm for the particle filter developed by Pitt (2002) that yields a smooth likelihood function, and therefore ensures that standard numerical procedures can be used in the maximization routine. The study by Eraker et al. (2003) uses a bayesian framework for parameter estimation. They break down the high dimensional posterior distribution into a series of complete conditionals in order to use Markov Chain Monte Carlo (MCMC) methods to estimate the posterior moments.

The aim of this paper is to combine both model frameworks and to test whether the class of non-affine models still outperforms the class of affine models if we include jumps into the stochastic processes. In particular, we use a class of non-affine stochastic volatility (SV) models of Christoffersen et al. (2008) and extend it to a model class which includes jumps in returns or jumps in returns and volatility. Incorporating jumps in the return process leads to the class of SVJ models. Incorporating jumps in returns and in volatility (which arrive contemporaneously and are correlated) constitutes the class of the SVCJ models. We would expect that a more heavily parameterized model with jumps will outperform a simple nested

model. But the central question is how a non-affine parsimonious model (i.e. without jumps in returns and volatility) performs compared to an affine more heavily parameterized model? In other words, is it necessary to leave a class of the affine models after adding jumps in returns and in volatility?

We use the MCMC technique to estimate model parameters from SV, SVJ and SVCJ model classes. We choose MCMC for two reasons. First, MCMC approach allows to account for multivariate latent processes, whereas the resampling method used in the MLIS estimation procedure assumes that the latent process is univariate. By including jumps into the model, the latent variable process becomes multivariate, and thus, the resampling algorithm can no longer be applied. Second, as shown in a Monte Carlo study by Andersen, Chung and Sørensen (1999), the MCMC approach outperforms several other estimation approaches, resulting in a smaller bias and a smaller root mean squared error.

For model comparison we choose the following methodology. The first procedure used to investigate model misspecification consists in inspection of the quantile-to-quantile (QQ) plots of the return residuals. Further, Bayes factors are used as a test statistic in order to compare nested models, in particular jump models compared to SV models. Finally, to compare non-nested models we use the deviance information criterion (DIC) proposed by Spiegelhalter, Best, Carlin and van der Linde (2002).

Based on the performance of the QQ plots and the DIC statistics we conclude that the non-affine stochastic volatility models provide a good fit to the data. The best performing non-affine models use a non-linear drift specification for the stochastic volatility. However, the non-linear drift specification is problematic, since it implies an long run mean of zero for the volatility. For the affine models we find that including jumps allows to reduce model misspecification and we observe no significant differences in the QQ plots compared to the best non-affine models. Further, Bayes factors and the DIC statistics clearly reject the SV model class in favor of model classes with jump diffusion specification. Since asset pricing literature is rather developed for the affine model structures, we tend to favor the affine jump diffusion models, even though they perform less well in terms of their DIC statistics compared

to their non-affine counterparts.

The paper is organized as follows: Section 2 introduces the jump diffusion model used for modeling asset price and volatility dynamics. Section 3 discusses the MCMC estimation approach. Model selection criteria and model testing are developed in Section 4. Our empirical results using time series of the S&P 500 index daily returns are presented in Section 5. Finally, Section 6 summarizes the findings.

2 Model Description

Models with stochastic volatility and jumps in returns have been studied in many papers, e.g. Bakshi et al. (1997), Bates (2000), Pan (2002), Duffie et al. (2000). While modeling of stochastic volatility and jumps in returns is crucial, Bakshi et al. (1997) and Pan (2002) found that an additional factor which drives conditional volatility movements is needed to avoid inconsistency with the diffusion specification. Adding jumps to the diffusive component in the volatility process allows to model a rapidly moving and persistent conditional volatility of returns. Stochastic volatility models with jumps in the return process, as well in the volatility process have been studied by Eraker et al. (2003). Recently, Christoffersen et al. (2008) proposed another approach to overcome the problem of misspecification of the Heston (1993) model which suggests to consider several alternative volatility specifications. We combine both approaches which leads to the following equations for modeling the logarithm of asset price $Y_t = \log S_t$ and the volatility process V_t :

$$\begin{aligned} \begin{pmatrix} dY_t \\ dV_t \end{pmatrix} &= \begin{pmatrix} \mu \\ \kappa V_t^a (\theta - V_t) \end{pmatrix} dt + \begin{pmatrix} \sqrt{V_t} & 0 \\ 0 & V_t^b \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \rho \sigma_v & \sqrt{(1 - \rho^2)} \sigma_v \end{pmatrix} d\mathbf{W}_t \\ &+ \begin{pmatrix} \xi^y dN_t^y \\ \xi^v dN_t^v \end{pmatrix}. \end{aligned} \quad (1)$$

In (1) \mathbf{W}_t denotes a standard bivariate Brownian motion, N_t^y and N_t^v are Poisson processes both with constant intensity λ , and ξ^y and ξ^v are jump sizes in returns and

a	b	Name	Features
0.0	0.5	SQR	square root diffusion, variance drift is linear in variance
1.0	0.5	SQRN	square root diffusion, variance drift is nonlinear in variance
0.0	1.0	VAR	linear diffusion, variance drift is linear in variance
1.0	1.0	VARN	linear diffusion, variance drift is nonlinear in variance
0.0	1.5	3/2	3/2 diffusion, variance drift is linear in variance
1.0	1.5	3/2N	3/2 diffusion, variance drift is nonlinear in variance

Table 1: Stochastic volatility model specifications. For each model class SV, SVJ and SVCJ we vary parameter values a and b as indicated in the first two columns of the table.

volatility, respectively. Further, for $a = 0$ θ controls the long-run mean of variance V_t and κ is the speed of mean reversion. Note that for $a = 1$ the parameters θ and κ are not directly comparable to the previous case. For this case we have a specification for the variance of $dV_t = \kappa\theta V_t dt + \kappa V_t^2 dt + \sigma_v V_t^b dW_t$. It can be seen that in the case that V_t hits zero both components (drift and diffusion) vanish. This implies a long run mean of zero, which is a problematic property from an economic perspective. Further, ρ denotes correlation between Brownian increments for stock and volatility and σ_v^2 determines the variance of variance. Finally, parameter value a allows to specify a variance drift in the model: setting a equal to zero (one) leads to a variance drift which is linear (nonlinear) in the level of variance V_t . Varying parameter value b allows to control diffusion in the model: setting b equal to 0.5, 1.0, or 1.5 leads to a square root, a linear, or a 3/2 diffusion, respectively. Thus, we think of six different models for all combinations of parameters $a \in \{0, 1\}$ and $b \in \{1/2, 1, 3/2\}$. These models, named in accordance with their diffusion and variance drift specifications as SQR, SQRN, VAR, VARN, 3/2 and 3/2N, are indicated in Table 1. In the following, SV, SVJ and SVCJ are referred to as *model classes*, whereas a *model* is defined by the choice of the model class and one particular combination of a and b .¹ Note that

¹For example, SVCJ-SQR denotes a model from the model class SVCJ having SQR parameter specification $a = 0$ and $b = 1/2$.

Eraker et al. (2003) consider system (1) using parameter values $a = 0$ and $b = 1/2$ and include jumps, whereas Christoffersen et al. (2008) consider all combinations of a and b but restrict their analysis to pure stochastic volatility models without jumps.

System (1) is the model setup which nests many simpler models. In particular, setting jump sizes $\lambda = 0$ it reduces to the SV model class. Setting $\lambda > 0$ leads to the SVJ model class. Thereby, jumps in returns are assumed to be normally distributed: $\xi^y \sim N(\mu_y, \sigma_y^2)$. The most general model class referred to as SVCJ allows for jumps in returns $\xi^y \sim N(\mu_y, \sigma_y^2)$ and volatility $\xi^v \sim \exp(\mu_v)$. Thereby, we assume that jumps arrive contemporaneously ($N_t^y = N_t^v = N_t$) and that jump sizes are correlated, that is, $\xi^v \sim \exp(\mu_v)$ and $\xi^y|\xi^v \sim N(\mu_y + \rho_j \xi^v, \sigma_y^2)$.

For the model estimation we consider time-discretization of system (1) with discretization interval $\Delta = 1$ corresponding to one day. Denoting $R_{t+1} = Y_{t+1} - Y_t$ the log-returns of asset price, we can write the discretized version for the SV model class as follows:

$$\begin{aligned} R_{t+1} &= \mu + \sqrt{V_t} \varepsilon_{t+1}^y \\ V_{t+1} &= V_t + \kappa V_t^a (\theta - V_t) + \sigma_v V_t^b \varepsilon_{t+1}^v, \end{aligned}$$

where shocks to returns and volatility, $\varepsilon_{t+1}^y = W_{t+1}^y - W_t^y$ and $\varepsilon_{t+1}^v = W_{t+1}^v - W_t^v$, are random variables which follow a bivariate normal distribution with zero expectation, unit variance, and correlation ρ . For this model we estimate the latent variables $\{V_t\}_{t=1}^T$ and the unknown parameters $\rho, \kappa, \theta, \sigma_v, \mu$.

Now we introduce jumps in the return processes which constitutes the SVJ model class. The resulting discretized version is given by the following two equations system:

$$\begin{aligned} R_{t+1} &= \mu + \sqrt{V_t} \varepsilon_{t+1}^y + \xi_{t+1}^y J_{t+1} \\ V_{t+1} &= V_t + \kappa V_t^a (\theta - V_t) + \sigma_v V_t^b \varepsilon_{t+1}^v, \end{aligned}$$

where $J_{t+1} = N_{t+1} - N_t$ indicates jump arrivals in the return process, this is a Bernoulli random variables with intensity λ . For the jump sizes ξ^y we assume that

these are normally distributed random variables: $\xi^y \sim N(\mu_y, \sigma_y^2)$. The set of latent variables to estimate includes $\{V_t, J_t, \xi_t^y\}_{t=1}^T$, and the unknown parameters are $\rho, \kappa, \theta, \sigma_v, \mu, \mu_y, \sigma_y, \lambda$.

Compared to the SVJ model class, the SVCJ model class additionally includes jumps in the volatility process. The discretized SVCJ version is given by:

$$\begin{aligned} R_{t+1} &= \mu + \sqrt{V_t} \varepsilon_{t+1}^y + \xi_{t+1}^y J_{t+1} \\ V_{t+1} &= V_t + \kappa V_t^a (\theta - V_t) + \sigma_v V_t^b \varepsilon_{t+1}^v + \xi_{t+1}^v J_{t+1}. \end{aligned}$$

It assumes that the jump sizes in returns ξ^y and volatility ξ^v are correlated: $\xi^v \sim \exp(\mu_v)$ and $\xi^y | \xi^v \sim N(\mu_y + \rho_j \xi^v, \sigma_y^2)$. J_{t+1} is a Bernoulli random variable with intensity λ . $J_{t+1} = 1$ indicates that a jump occurs between time t and $t + 1$. The set of latent variables to estimate extends to $\{V_t, J_t, \xi_t^y, \xi_t^v\}_{t=1}^T$, and the unknown parameters are now given by $\rho, \kappa, \theta, \sigma_v, \mu, \mu_y, \sigma_y, \lambda, \mu_v, \rho_j$.

The underlying discretization scheme may lead to some discretization bias, since we assume that at most one jump occurs per time period which is one day. However, the following example highlights that due to the fact, that jumps are rare events, the discretization bias is typically very small. Using $P(N_{t+1} - N_t = j) = \frac{\exp\{\lambda\} \lambda^j}{j!}$ and assuming jump intensity to be $\lambda = 0.1$, the probability for observing more than one jump per day would be 0.0045. Note that our estimation results indicate estimates for λ smaller than 0.1.

3 Markov-Chain Monte-Carlo estimation

The underlying problem setup involves estimation of latent variables such as volatility, jump times and jump sizes. In a Bayesian context each of these unobserved variables is treated as a parameter to estimate. This leads to a high dimensional posterior distribution which is not known. In order to compute the moments of the posterior, we would have to compute a high dimensional integral, which is not feasible. Therefore, we rely on the Markov-Chain Monte-Carlo (MCMC) method to compute the moments of the parameter values and latent variables conditional

on the observed data. These moments are used as a point estimator for the parameters and the latent variables. Bayesian MCMC method for stochastic volatility models have been developed in Jacquier, Polson and Rossi (2004). MCMC generates samples from a given target distribution, in our case² $p(\Theta, \mathbf{V}, \xi^y, \xi^v, \mathbf{J}|\mathbf{R})$ - the joint distribution of the parameter vector $\Theta = (\mu, \theta, \kappa, \sigma_v, \rho, \mu_y, \sigma_y, \rho_j, \mu_v, \lambda)^\top$ and the state variables $\{\mathbf{V}, \xi^y, \xi^v, \mathbf{J}\}$, given the observed returns. The basis for MCMC estimation is provided by the Hammersley-Clifford theorem which claims that under certain regularity conditions, the joint posterior distribution can be completely characterized by the complete conditional distributions. In order to compute the posterior distribution note that the joint distribution of (R_t, V_t) follows a bivariate normal with the following parameters:

$$\begin{aligned} \boldsymbol{\mu} &= \begin{pmatrix} \mu + \xi_t^y J_t \\ V_{t-1} + \kappa(V_{t-1})^a(\theta - V_{t-1}) + \xi_t^v J_t \end{pmatrix} \\ \boldsymbol{\Sigma} &= \begin{pmatrix} V_{t-1} & \rho\sigma_v(V_{t-1})^{b+0.5} \\ \rho\sigma_v(V_{t-1})^{b+0.5} & \sigma_v^2(V_{t-1})^{2b} \end{pmatrix} \end{aligned} \quad (2)$$

Given the normality of the joint distribution we can determine the conditional distribution of returns $p(R_t|V_t, V_{(t-1)}, \xi_t^y, \xi_t^v, \Theta)$. It follows a normal distribution with parameters

$$\begin{aligned} \mu_{R_t|V_t} &= \mu + \xi_t^y J_t + \rho\sigma_v^{-1}(V_{t-1})^{0.5-b} \times \\ &\quad \times \{V_t - V_{t-1} - \kappa(V_{t-1})^a(\theta - V_{t-1}) - \xi_t^v J_t\} \\ \sigma_{R_t|V_t}^2 &= V_{t-1}(1 - \rho^2). \end{aligned} \quad (3)$$

On the other hand, the conditional distribution $p(V_t|R_t, V_{(t-1)}, \xi_t^y, \xi_t^v, \Theta)$ is also normal with parameters:

$$\begin{aligned} \mu_{V_t|R_t} &= V_{t-1} + \kappa(V_{t-1})^a(\theta - V_{t-1}) + \xi_t^v J_t \\ &\quad + \rho\sigma_v\{R_t - \mu\Delta - \xi_t^y J_t\} \\ \sigma_{V_t|R_t}^2 &= \sigma_v^2(V_{t-1})^{2b}(1 - \rho^2). \end{aligned} \quad (4)$$

²In the following, we will focus on the SVCJ model class since it has the most complex distributional structure.

In general, the posterior is given by the following expression:

$$\begin{aligned} p(\Theta, \mathbf{V}, \xi^y, \xi^v, \mathbf{J}|\mathbf{R}) &\propto p(\mathbf{R}|\Theta, \mathbf{V}, \xi^y, \xi^v, \mathbf{J})p(\Theta, \mathbf{V}, \xi^y, \xi^v, \mathbf{J}) \\ &= p(\mathbf{R}|\Theta, \mathbf{V}, \xi^y, \xi^v, \mathbf{J})p(\mathbf{V}|\Theta, \xi^v, \mathbf{J})p(\mathbf{J}|\lambda)p(\xi^y|\xi^v, \Theta)p(\xi^v|\Theta)p(\Theta) \end{aligned} \quad (5)$$

where $p(\mathbf{R}|\Theta, \mathbf{V}, \xi^y, \xi^v, \mathbf{J})$ denotes the likelihood and $p(\Theta, \mathbf{V}, \xi^y, \xi^v, \mathbf{J})$ is the prior distribution. Assuming that the jump times and the jump sizes are independent and denoting $\theta_{VR} = \{\mu, \sigma_v, \rho, \kappa, \theta\}$ and $\theta_{\xi_t^y} = \{\mu_y, \rho_j, \sigma_y^2\}$, we can rewrite (5) as follows:

$$p(\Theta, \mathbf{V}, \xi^y, \xi^v, \mathbf{J}|\mathbf{R}) \propto \prod_{t=1}^T p(R_t, V_t|V_{t-1}, \xi_t^y, \xi_t^v, J_t, \theta_{VR})p(J_t|\lambda)p(\xi_t^y|\xi_t^v, \theta_{\xi_y})p(\xi_t^v|\mu_v)p(\Theta) \quad (6)$$

where we used the fact that

$$\begin{aligned} p(\mathbf{R}|\Theta, \mathbf{V}, \xi^y, \xi^v, \mathbf{J})p(\mathbf{V}|\Theta, \xi^v, \mathbf{J}) &\propto \prod_{t=1}^T p(R_t|V_t, V_{t-1}, \xi_t^y, \xi_t^v, \Theta)p(V_t|V_{t-1}, \xi_t^v, \Theta) \\ &= \prod_{t=1}^T p(R_t, V_t|V_{t-1}, \xi_t^y, \xi_t^v, \Theta). \end{aligned} \quad (7)$$

In (7) we can write $p(\mathbf{R}|\Theta, \mathbf{V}, \xi^y, \xi^v, \mathbf{J}) = \prod_{t=1}^T p(R_t|V_t, V_{t-1}, \xi_t^y, \xi_t^v, \Theta)$ by conditional independence and $p(\mathbf{V}|\Theta, \xi^v, \mathbf{J}) \propto \prod_{t=1}^T p(V_t|V_{t-1}, \xi_t^v, \Theta)$ by the Markov property. To update estimated parameters value in each iteration, MCMC algorithm draws from its posterior distribution conditional on the current values of all other parameters and state variables. Therefore, in order to reduce the influence of the starting point and to assure that stationarity is achieved, the general approach is to discard a burn-in period of the first h iterations. The iterations after the burn-in period provide a representative sample from the joint posterior, and averaging over the non-discarded iterations provides an estimate for posterior means of parameters and latent variables.

Sampling from the conditional posterior can be implemented by either using Gibbs sampler introduced by Geman and Geman (1984) or the Metropolis-Hasting algorithm, see Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953). Gibbs sampler is applied if the complete conditional distribution to sample from is known. The MCMC algorithm with the Gibbs step samples iteratively drawing from the

following conditional posteriors, see Gilks (1995):

$$\text{Parameters : } p(\Theta_i | \Theta_{-i}, \mathbf{J}, \xi^y, \xi^v, \mathbf{V}, \mathbf{R}), i = 1, \dots, k$$

$$\text{Jump times : } p(J_t | \Theta, \mathbf{J}_{-t}, \xi^y, \xi^v, \mathbf{V}, \mathbf{R}), t = 1, \dots, T$$

$$\text{Jump sizes : } p(\xi_t^y | \Theta, \mathbf{J}, \xi_{-t}^y, \xi^v, \mathbf{V}, \mathbf{R}), t = 1, \dots, T$$

$$p(\xi_t^v | \Theta, \mathbf{J}, \xi_{-t}^v, \xi^y, \mathbf{V}, \mathbf{R}), t = 1, \dots, T$$

$$\text{Volatility : } p(V_t | \Theta, V_{t+1}, V_{t-1}, \mathbf{J}, \xi^v, \xi^y, \mathbf{R}), t = 1, \dots, T$$

If some conditional distributions cannot be sampled directly, as in the case of the variance and the correlation ρ , we apply the Metropolis-Hasting algorithm. For details on the Metropolis-Hasting algorithm refer to Johannes and Polson (2006).

To start the procedure running, we have to specify the prior distributions. When possible we assume so-called conjugate priors which after multiplying with the likelihood lead to a posterior distribution belonging to the same family of distributions as the prior itself. Wherever possible, we choose standard conjugate priors, which allow to draw from the conditional posteriors directly. For the model parameters we choose the following conjugate priors proposed by Eraker et al. (2003): $\mu \sim \mathcal{N}(1, 25)$, $\kappa\theta \sim \mathcal{N}(0, 1)$, $\kappa \sim \mathcal{N}(0, 1)$, $\sigma_v^2 \sim \mathcal{IG}(2.5, 0.1)$, $\lambda \sim \mathcal{B}(2, 40)$, $\mu_y \sim \mathcal{N}(0, 100)$, $\sigma_y^2 \sim \mathcal{IG}(5, 20)$, $\mu_v \sim \mathcal{G}(20, 10)$, $\rho_j \sim \mathcal{N}(0, 4)$, where \mathcal{N} denotes the normal distribution, \mathcal{IG} is the inverse gamma distribution, \mathcal{B} denotes the beta distribution and \mathcal{G} is the gamma distribution. For the correlation we have not a conjugate standard uniform prior: $\rho \sim \mathcal{U}(-1, 1)$.

4 Model Testing

This section aims to present diagnostic tools which allow to quantify the model performance. Model comparison can be done using the fit of the model to the data and the complexity of the model as a penalty factor. The model fit is measure by a deviance statistic and the complexity is represented by the number of effective parameters. In a non-Bayesian setting deviance is used as a quantity which estimates

the number of degrees of freedom in the underlying model: it refers to the difference in log-likelihoods between the fitted and the saturated model (that is, the one which yields perfect fit of the data). Obviously, increasing the complexity of the model by, e.g., incorporating jumps will lead to a better fit of the model to the data. Therefore, one should incorporate a penalty term for complexity.

In analogy, Dempster (1997) and Spiegelhalter et al. (2002)³ have developed the deviance information criterion (DIC) as a Bayesian model choice criterion. DIC solves the problem of comparing complex hierarchical models when the number of parameters is not clearly defined. Further, Bayes factors for nested models are used for calculation of posterior odds, which are applied to examine whether model classes with jumps (SVJ, SVCJ) provide better fit for the return data than the more parsimonious model class (SV). An overview on the calculation of DICs and Bayes factors is presented below.

4.1 Deviance information criterion

The DIC value consists of two components: a term \bar{D} that measures goodness of fit and a penalty term p_D accounting for model complexity:

$$\text{DIC} = \bar{D} + p_D. \quad (8)$$

The first term can be calculated as follows:

$$\bar{D} = E_{\Theta|\mathbf{R}}\{D(\Theta)\} = E_{\Theta|\mathbf{R}}\{-2 \log f(\mathbf{R}|\Theta)\}$$

where \mathbf{R} denotes the returns data and Θ is a vector of parameters. The better the model fits the data, the larger is the likelihood, i.e., smaller values of \bar{D} will indicate a better model fit. In fact, since \bar{D} already includes a penalty term p_D , it could be better thought of as a measure of “model adequacy” rather than a measure of fit, although these terms can be used interchangeably.

³An application of this criterion in financial econometrics can be found in Berg, Meyer and Yu (2004).

The second component measures the complexity of the model due to the effective number of parameters:

$$\begin{aligned} p_D &= \bar{D} - D(\bar{\Theta}) = E_{\Theta|\mathbf{R}}\{D(\Theta)\} - D\{E_{\Theta|\mathbf{R}}(\Theta)\} \\ &= E_{\Theta|\mathbf{R}}\{-2\log f(\mathbf{R}|\Theta)\} + 2\log f(\mathbf{R}|\bar{\Theta}). \end{aligned}$$

Clearly, since p_D is considered to be the posterior mean of the deviance (average of log-likelihood ratios) minus the deviance evaluated at the posterior mean (likelihood evaluated at average), it can be used to quantify the number of free parameters in the model (the number of degrees of freedom). Further, defining $-2\log f(\mathbf{R}|\Theta)$ to be the residual information in data \mathbf{R} conditional on Θ , and interpreting it as a logarithmic penalty, or uncertainty, see Kullback and Leibler (1951), Bernardo (1979), p_D can be regarded as the expected excess value of the true over the estimated residual information in the return data \mathbf{R} conditional on Θ , and thus, can be thought of as the expected reduction in uncertainty. From (9) we obtain: $\bar{D} = D(\bar{\Theta}) + p_D$, and thus, DIC can be rewritten as the estimate of the fit plus twice the number of effective parameters:

$$\text{DIC} = D(\bar{\Theta}) + 2p_D.$$

4.2 Computing the Bayes factors

Suppose we want to examine whether the richer parameter model class with jumps (SVCJ) provides better fit for the return data than the more parsimonious model class (SV), that is, we want to compare the posterior probabilities for both model classes conditioning on the observed return data \mathbf{R} . If the model classes are nested (which is the case when comparing SV with SVJ, or SVCJ), this can be done by means of the posterior odds, see Eraker et al. (2003):

$$\frac{p(\text{SV}|\mathbf{R})}{p(\text{SVJ}|\mathbf{R})} = \frac{p(\mathbf{R}|\text{SV})}{p(\mathbf{R}|\text{SVJ})} \cdot \frac{p(\text{SV})}{p(\text{SVJ})}. \quad (9)$$

In (9) the first term on the r.h.s. is known as the Bayes factor and the second term is the prior odds. If we have no prior preferences for any of model classes, we set

prior odds equal to one. Eraker et al. (2003) showed that the odds ratio for the model class SV relative to the model class SVJ can be calculated as:

$$\frac{p(\text{SV}|\mathbf{R})}{p(\text{SVJ}|\mathbf{R})} = \frac{\mathbb{B}(\alpha_0, \beta_0)}{\mathbb{B}(\alpha_0, T + \beta_0)} \cdot \frac{1}{G} \sum_{g=1}^G \frac{\mathbb{B}(\alpha_0 + \sum_{t=1}^T J_t^g, \beta_0 + 2T - \sum_{t=1}^T J_t^g)}{\mathbb{B}(\alpha_0 + \sum_{t=1}^T J_t, \beta_0 + T - \sum_{t=1}^T J_t)}.$$

with $\mathbb{B}(\alpha_0, \beta_0)$ denoting the Beta function and G the number of non-discarded iterations in the MCMC sample. Clearly, the higher is the odds ratio, the stronger is the evidence against a model. Note however, that the odds ratio do not necessarily favor more complex models, since they contain a penalty for using more heavily parametrization. Following Kass and Raftery (1995), the evidence against a model is positive if log odds ratio is between 2 and 6, strong if it is between 6 and 10, and very strong if it is greater than 10.

5 Empirical Analysis

Using a time series of continuously compounded daily returns of the S&P 500 ranging from January 2, 1986 to July 31, 2008, we compare parameter estimates for different models. To run the estimation procedure we set the number of draws equal 500,000 and choose a burn-in period of 200,000 iterations which guarantees the convergence of the algorithm for all models.

In order to compare the performance of the models we use a combination of the test statistics presented in Section 4. First, we consider the QQ plots of the return residuals in order to detect the degree of model misspecification for each of the estimated models. Second, models are compared for all combinations of a and b using the DIC statistic (see Section 4.1). To choose between jump diffusion and stochastic volatility models within each a and b specification we rely on log odds ratios (see Section 4.2).

5.1 The SV Model Class

For all specifications within the SV model class we report means and standard deviations (numbers in parentheses), as well as the DIC values and the p_D values

Model (a,b)	SQR (0.0;0.5)	SQRN (1.0;0.5)	VAR (0.0;1.0)	VARN (1.0;1.0)	3/2 (0.0;1.5)	3/2N (1.0;1.5)
μ	0.0370 (0.0110)	0.0371 (0.0110)	0.0415 (0.0112)	0.0399 (0.0110)	0.0414 (0.0138)	0.0425 (0.0109)
θ	1.0423 (0.2309)	0.1842 (7.5817)	1.1060 (0.0857)	2.5509 (0.6256)	1.3136 (0.1062)	6.7684 (2.5292)
κ	0.0055 (0.0017)	0.0010 (0.0006)	0.0326 (0.0179)	0.0027 (0.0012)	0.2102 (0.0558)	0.0068 (0.0019)
σ_v	0.0835 (0.0072)	0.0803 (0.0063)	0.1694 (0.0385)	0.0877 (0.0093)	0.2471 (0.0347)	0.2254 (0.0595)
ρ	-0.5569 (0.0484)	-0.5211 (0.0488)	-0.5043 (0.0450)	-0.5472 (0.0510)	-0.3277 (0.0441)	-0.4316 (0.0661)
DIC	14281.59 (4)	14401.61 (5)	14263.09 (3)	14240.15 (2)	15355.37 (6)	14199.65 (1)
p_D	1712.36	1508.73	1391.82	1650.56	635.63	1099.34
\bar{D}	12569.23	12892.88	12871.28	12589.59	14719.75	13100.31

Table 2: Parameter estimates and their standard errors (in parentheses) for the SV model class. The reported parameter estimates are obtained using data on the S&P 500 daily returns from January 2, 1986 to July 31, 2008.

with a preference order (numbers in parentheses) in Table 2. Our estimate of the parameter ρ which captures correlation between shocks to return and variance ranges between -0.5569 for the SV-SQR model and -0.3277 for the SV-3/2 model which is consistent with the results obtained by Eraker et al. (2003). As discussed in Section (2) θ can be interpreted as long run mean only for the models where $a = 0$. Therefore we find for the SV-SQR, the SV-VAR and the SV-3/2 model a long-run mean of the variance which is close to one (ranges from 1.04 to 1.31). This corresponds to the annualized long-run volatility $\sqrt{252 \cdot \theta}$ of 16.3 to 18.2 percent. Similarly, the parameter κ is also not directly comparable between linear ($a = 0$) and nonlinear ($a = 1$) models. For a fixed value of a and increasing value of b we observe that the estimated values for κ and θ increase. The diffusion parameter σ_v is comparable within a given diffusion specification (i.e. for a fixed b), but not across diffusion specification. For square root diffusion ($b = 0.5$) we observe σ_v corresponding to roughly 0.08. It increases up to 0.247 for the 3/2 model.

We find that the SV-SQR model is misspecified, which is in line with the results of Eraker et al. (2003) and Christoffersen et al. (2008). This can be seen by inspecting Figure 1, which shows the QQ probability plot of the residuals in returns for each model specification within the SV model class. The QQ plot contrasts the quantiles

of the estimated residuals in returns with the quantiles from the standard normal distribution. A deviation of the residual data from the 45 degrees line indicates strong non-normality of residuals and thus, the evidence of misspecification. In the upper left graph of Figure 1, we observe for the SV-SQR model fatter tails than the normal distribution would suggest, which clearly shows that the model is misspecified. Altogether, we see that moving from the linear towards nonlinear variance drift (i.e. increasing a from 0 to 1) and from the square root diffusion towards linear or 3/2 diffusion (i.e. increasing b from 1/2 to 3/2) allows to capture the behavior of outliers at tails. In the lower right graph of Figure 1 we see that the SV-3/2N model performs best in capturing the normality assumption of the return residuals. The QQ plot shows no severe deviation from the 45 degrees line.

To analyze the differences in models in more details, we additionally examine volatility paths, annualized daily conditional leverage paths, and conditional volatility of the variance paths. Figure 2 plots annualized daily spot volatility path $\sqrt{V_t}$ during 1986-2008 for all model specifications. We observe the same pattern in the volatility path across models. However, the higher is the diffusion parameter b , the sharper is an increase in volatility. The SV-VAR model experiences more spikes compared to the SV-SQR model and the SV-3/2 model exhibits more spikes in the volatility pattern than the SV-SQR model. Figure 3 plots the annualized daily conditional leverage path $\rho\sigma_v V_t^{b+1/2}$ defined as the conditional covariance between shocks to returns and shocks to variance. Figure 4 plots the conditional volatility of the variance path σV_t^b , defined as the square root of the conditional variance of the variance of returns. Again, we observe the same pattern across models, but the spikes become more pronounced with increasing b .

In the following, we use DIC statistic to rank the models. Comparing DIC values across different models we observe that the SV-3/2N and the SV-VARN models with underlying nonlinear variance drift are ranked first, respectively second. The SV-SQR model is ranked fourth. Altogether, based on the QQ probability plots and the DIC values we conclude that the non square root diffusion provides a better fit for the returns data than the model with underlying square root diffusion. The

parsimonious affine SV-SQR model without jumps is outperformed by the non-affine models such as SV-3/2N, SV-VARN and SV-VAR. This is in line with a conclusion made by Christoffersen et al. (2008) who suggest to leave the class of the affine stochastic volatility models.

5.2 The SVJ Model Class

Model (a,b)	SQR (0.0;0.5)	SQRN (1.0;0.5)	VAR (0.0;1.0)	VARN (1.0;1.0)	3/2 (0.0;1.5)	3/2N (1.0;1.5)
μ	0.0468 (0.0108)	0.0466 (0.0109)	0.0480 (0.0110)	0.0481 (0.0108)	0.0435 (0.0143)	0.0488 (0.0110)
θ	1.0973 (0.5869)	1.5300 (4.4548)	1.0487 (0.1339)	1.9155 (0.3169)	1.3581 (0.1144)	9.6390 (4.1848)
κ	0.0034 (0.0016)	0.0005 (0.0016)	0.0123 (0.0051)	0.0039 (0.0014)	0.2254 (0.0562)	0.0058 (0.0025)
σ_v	0.0660 (0.0063)	0.0621 (0.0058)	0.1029 (0.0179)	0.0758 (0.0079)	0.2281 (0.0358)	0.2313 (0.0610)
μ_y	-3.0109 (1.1093)	-2.9915 (1.1675)	-3.9204 (1.1687)	-3.6157 (1.2273)	-4.0990 (1.7825)	-1.8774 (0.7179)
σ_y	4.4211 (0.7230)	4.5056 (0.7060)	3.1901 (0.7024)	3.5792 (0.7443)	2.1810 (0.4703)	1.8406 (0.3189)
ρ	-0.5290 (0.0457)	-0.5196 (0.0510)	-0.5353 (0.0421)	-0.5439 (0.0465)	-0.3303 (0.0431)	-0.4333 (0.0636)
λ	0.0054 (0.0016)	0.0052 (0.0015)	0.0039 (0.0014)	0.0046 (0.0016)	0.0018 (0.0011)	0.0068 (0.0032)
DIC	14177.90 (4)	14212.51 (3)	14103.33 (5)	14125.84 (6)	15474.34 (2)	14062.19 (1)
p_D	1636.77	1552.10	1666.93	1697.18	622.57	1201.24
\bar{D}	12541.13	12660.41	12436.40	12428.67	14851.77	12860.96

Table 3: Parameter estimates and their standard errors (in parentheses) for the SVJ model class. The reported parameter estimates are obtained using data on the S&P 500 daily returns from January 2, 1986 to July 31, 2008.

For the SVJ model class, besides estimating the parameters μ , θ , κ , σ_v and ρ , in addition, three jump parameters are estimated: the intensity of the jump λ , the mean jump size μ_y and the variance of the jump size σ_y^2 . First, we observe that adding jumps to the return processes leads to an increase in estimated values of μ which can be regarded as a compensation for the negative effect of jumps. For all but the SVJ-3/2N model θ ranges between 1.05 and 1.92. This corresponds to a long run mean of annualized volatility of 16.3 to 18.5 percent for the models where $a = 0$. For the 3/2N model θ rises to the value of 9.64. Similar to the SQRN-SV model, the SQRN-SVJ model produces high standard errors for the estimate

of θ . Further, the variance persistence in the model measured through κ increases with increasing diffusion parameter b across linear ($a = 0$) and nonlinear ($a = 1$) models. Parameter estimates for σ_v and ρ remain nearly unchanged compared to the SV model class. For the mean jump size in returns μ_y we obtain negative values ranging between -4.1 and -1.88 , i.e., the S&P 500 index exhibits negative expected jumps significantly different from zero. Figure 9 plots jump times and jump sizes in returns for different model specifications within the SVJ model class. The estimated standard deviation of the jump size σ_y is larger for more negative jumps, and the jump intensity λ ranges between 0.0018 and 0.0068 which corresponds to an expected number of jumps between 0.5 and 1.7 per year.

We observe that adding jumps to the square root diffusion models (SVJ-SQR and SVJ-SQRN), or to the linear diffusion models (SVJ-VAR and SVJ-VARN) allows to capture outliers in the low tail of the QQ plot (compare the upper and the middle panel of Figure 1 and Figure 5). However, increasing b to $3/2$ (low panel of Figure 1 and Figure 5) shows that the more complex model with jumps fails to capture outliers at tails resulting in a strong deviation of the residual data from the 45 degrees line. This indicates that the assumption of residual normality is violated and thus, the SVJ-3/2 and the SVJ-3/2N models are misspecified. In particular, we note that a jump component in the SVJ-3/2N model worsens the model fit considerably compared to the case of the SV-3/2N model. This indicates that jumps in returns are not only a superfluous but even a destructive factor for this model.

The volatility path $\sqrt{V_t}$ (Figure 6), the leverage path $\rho\sigma_v V_t^{b+1/2}$ (Figure 7) and conditional volatility of the variance path σV_t^b (Figure 8) follow the same pattern for the SVJ as for the SV model class. However, spikes in volatility for the SVJ model class become damped compared to the SV model class due to the presence of jumps.

Comparing the SVJ models for different a and b values in terms of their DIC performance, we observe that the SVJ-3/2N model is ranked first, similar to the SV case. However, the inspection of the QQ plots indicates that the SVJ-3/2N model is misspecified. Thus, it is difficult to assess the reliability of the DIC statistic in this

case. Regarding the affine SVJ-SQR model, it is ranked fourth based on its DIC value, yet we observe considerable improvement in its QQ plot after including jumps in the return process. Comparing the SVJ with the SV model class in terms of their DIC statistics, we observe a considerable improvement for all models within the SVJ model class. We conclude that adding jumps to the diffusion models from the SV class makes their performance better compared to their parsimonious counterparts without jumps.

5.3 The SVCJ Model Class

Model	SQR	SQRN	VAR	VARN	3/2	3/2N
(a,b)	(0.0;0.5)	(1.0;0.5)	(0.0;1.0)	(1.0;1.0)	(0.0;1.5)	(1.0;1.5)
μ	0.0532 (0.0110)	0.0490 (0.0118)	0.0488 (0.0112)	0.0458 (0.0108)	0.0441 (0.0136)	0.0483 (0.0120)
θ	0.5492 (0.0621)	0.6274 (0.4895)	0.9150 (0.0677)	1.6646 (0.1752)	1.2471 (0.0901)	6.2169 (1.9563)
κ	0.0243 (0.0056)	0.0493 (0.0341)	0.0357 (0.0101)	0.0162 (0.0059)	0.1859 (0.0516)	0.0082 (0.0022)
σ_v	0.0692 (0.0077)	0.0943 (0.0287)	0.1409 (0.0231)	0.1169 (0.0172)	0.2357 (0.0305)	0.2371 (0.0470)
μ_y	-2.8198 (0.5730)	-0.4239 (1.5785)	-1.5729 (0.7991)	-2.0876 (0.8179)	-2.8561 (1.3542)	-2.0182 (1.1673)
ρ_j	0.1176 (0.2060)	-1.0813 (0.8548)	-2.8339 (0.8341)	-3.1298 (0.8357)	-0.8319 (1.5433)	-1.2554 (1.4142)
σ_y	2.1237 (0.3875)	2.4124 (0.4644)	1.9026 (0.3133)	1.8568 (0.3250)	2.0868 (0.3934)	1.8444 (0.3286)
μ_v	1.3347 (0.2431)	1.3064 (0.3101)	0.7514 (0.1724)	0.6904 (0.1570)	0.5338 (0.1145)	0.4884 (0.1150)
ρ	-0.4866 (0.0551)	-0.4422 (0.0645)	-0.5030 (0.0413)	-0.5385 (0.0433)	-0.3433 (0.0448)	-0.4191 (0.0688)
λ	0.0085 (0.0021)	0.0127 (0.0045)	0.0044 (0.0014)	0.0039 (0.0012)	0.0022 (0.0011)	0.0057 (0.0054)
DIC	14144.14 (3)	14222.77 (5)	14177.38 (4)	14023.51 (1)	15373.22 (6)	14091.67 (2)
p_D	1496.30	1271.41	1492.55	1630.40	712.90	1037.55
\bar{D}	12647.84	12951.35	12684.83	12393.11	14660.32	13054.12

Table 4: Parameter estimates and their standard errors (in parentheses) for the SVCJ model. class The reported parameter estimates are obtained using data on the S&P 500 daily returns from January 2, 1986 to July 31, 2008.

For the SVCJ model class, in addition to the set of parameters already included in the SVJ model class, we estimate two additional parameters which include the mean jump size in volatility μ_v and the coefficient ρ_j . We observe that the mean jump sizes in volatility range between 0.49 and 1.33 for different model specifications. Conditional on jumps in volatility, we observe that the mean jump size in return ξ^y

become less negative compared to the SVJ model class. Note that when interpreting the expected jump size for the return process, we have to take into account that conditional on the current jump size in volatility, mean jump size in return is given by $\mu_y + \rho_j \xi^v$. The unconditional expected value for the jump size in returns can be calculated as $\mu_y + \rho_j \mu_v$, which leads to significantly negative values ranging from -3.7 to -1.8 . Regarding other parameter estimates, we observe quantitatively similar results for the mean return μ and the correlation coefficient ρ . For all models except the SVCJ-3/2, the estimated values for κ increase and the parameter θ decreases compared to the SVJ case. We observe that the volatility of the variance σ_v ranges between 0.07 and 0.24. Volatility of the return σ_y decreases compared to the SVJ model class and the jump intensity λ increases for all models except the SVCJ-VARN model.

From the QQ plots in Figure 10 we observe that adding jumps to the volatility process allows to better capture outliers at tails, compared to the SV model class and to some model specifications of the SVJ model class. The SVCJ-SQRN and SVCJ-VAR models provide a clearly better fit than the corresponding models from the SVJ model class. Further, similar to the SVJ case, the SVCJ-3/2 and SVCJ-3/2N models perform worse compared to the corresponding models from the SV class.

We observe similar pattern for the volatility path, the leverage path, and the conditional volatility of the variance path (Figures 11-13) across and between the models. However, spikes in the path patterns become more pronounced for the SVCJ class compared to the SVJ class.

Based on the DIC statistics we observe that the SVCJ-VARN model is ranked first, followed by the SVCJ-3/2N model. SVCJ-SQR model is ranked third. Comparing the SVCJ model class with the SVJ model class, we observe an improvement in DIC statistics for the SQR-SVCJ, VARN-SVCJ and 3/2-SVCJ models. For the remaining models SQRN-SVCJ, VAR-SVCJ and 3/2N-SVCJ the performance of the SVCJ class worsens compared to their counterparts in the SVJ model class.

5.4 Model Comparison

To perform an overall comparison, we use the following criteria: the QQ plots, the DIC statistics, and the Bayes factors. Considering each criteria on its own leads to the following conclusions.

Comparing the QQ plots, we find that the SV-3/2N, the SVJ-SQR, and SVCJ-SQR models perform best compared to all other models. They slightly differ in capturing the normality assumption for the return residuals in the tail of the distribution: the SV-3/2N model performs better in the lower tail whereas the SVJ-SQR and SVCJ-SQR models performs better in the upper tail.

In general, DIC statistics allow to compare complex hierarchical models. Overall, based on the DIC values we observe that the SV model class is outperformed by the SVJ and the SVCJ model classes. Table 5 shows that the SVCJ-VARN model with nonlinear variance drift is the best performing model, followed by the SVJ-3/2N and the SVCJ-3/2N models.

Model class choice for a given a and b specification can be performed based on the Bayes factors. For every a and b combination, the SV model class is outperformed by both jump diffusion model classes SVJ and SVCJ. Further, the SVJ model class is preferred to the SVCJ model class for the following combinations of parameters: $a = 0.0$, $b = 0.5$, and $a = 0.0$, $b = 1.0$, and $a = 0.0$, $b = 1.5$. For the parameter combinations $a = 1.0$, $b = 0.5$, and $a = 1.0$, $b = 1.5$, the SVCJ model class is preferred to the SVJ model class.

Based on the results obtained using each test criteria we can observe that the SV model class is outperformed by the SVJ and the SVCJ model classes. In particular, the Bayes factors clearly reject the SV model class, and the DIC statistic favors the SVCJ-VARN and the SVJ-3/2N models. Further, in order to judge on the model performance, we will not merely rely on the formal tests listed above, but we will additionally take into account other factors, e.g., economic plausibility⁴.

⁴This approach follows Spiegelhalter et al. (2002), who point out that an overformal approach to model selection is not appropriate, and one should take into account all features of the model.

Model	DIC	p_D	\bar{D}
SVCJ-VARN	14023.51	1630.40	12393.11
SVJ-3/2N	14062.19	1201.24	12860.96
SVCJ-3/2N	14091.67	1037.55	13054.12
SVJ-VAR	14103.33	1666.93	12436.40
SVJ-VARN	14125.84	1697.18	12428.67
SVCJ-SQR	14144.14	1496.30	12647.84
SVCJ-VAR	14177.38	1492.55	12684.83
SVJ-SQR	14177.90	1636.77	12541.13
SV-3/2N	14199.65	1099.34	13100.31
SVJ-SQRN	14212.51	1552.10	12660.41
SVCJ-SQRN	14222.77	1271.41	12951.35
SV-VARN	14240.15	1650.56	12589.59
SV-VAR	14263.10	1391.82	12871.28
SV-SQR	14281.59	1712.36	12569.23
SV-SQRN	14401.61	1508.73	12892.88
SV-3/2	15355.37	635.63	14719.75
SVCJ-3/2	15373.22	712.90	14660.32
SVJ-3/2	15474.34	622.57	14851.77

Table 5: This table shows models across all model classes ranked by their DIC statistics. Smaller values of DIC statistics indicate better model fit.

Therefore, our main findings can be summarized as follows: In accordance with Christoffersen et al. (2008) we observe a good fit to the data of the non-affine SV-3/2N model. After adding jumps in returns or jumps in returns and volatility, the affine models SVJ-SQR and SVCJ-SQR show a good fit to the data also. The best models in terms of the Deviance Information Criterion are given by non-affine model specifications with nonlinear drift term (see Table (5)). Since these models imply an long run mean of zero for the volatility, the usage might be problematic in a finance context. This problem can be overcome by conditioning the process

(a; b)	SVJ vs. SV	SVCJ vs. SV	SVCJ vs. SVJ
(0.0;0.5)	19.74	4.57	-15.17
(1.0;0.5)	19.58	41.36	21.78
(0.0;1.0)	27.82	25.66	-2.15
(1.0;1.0)	25.05	26.50	1.45
(0.0;1.5)	40.46	38.70	-1.76
(1.0;1.5)	34.03	43.30	9.26

Table 6: This table shows the resulting Bayes Factors for comparing model classes for each a and b specification. Large values indicate evidence in favor of the first model. For example, 19.74 shows very strong preference for the SVJ-SQR model compared to the SV-SQR model.

on not hitting zero, however, how this can be applied in the finance context gives rise to further research. In terms of QQ plots, the affine jump diffusion models have a similar performance as the non-affine models. Furthermore, the affine jump diffusion models have the second best DIC statistic when considering models with a linear drift specification. With respect the question to jump or not to jump we find that jump models are clearly preferred by Bayes factors as well as in terms of DIC statistics.

Provided the good statistical fit of the non-affine models based on the DIC, we would suggest to investigate further topics related to, e.g., option pricing implications, and the statistical and mathematical properties of non-affine models. Further, since asset pricing literature is rather developed for the affine model structures, we tend to favor the affine jump diffusion models, even though they perform less well compared to their non-affine counterparts in terms of their DIC statistics.

6 Conclusion

In this paper we combine two approaches to overcome the problem of misspecification of the Heston (1993) stochastic volatility model. The first approach, suggested

by Christoffersen et al. (2008), recommends to leave the class of exponentially affine stochastic volatility models. The second approach, investigated by Eraker et al. (2003), improves on the Heston (1993) model by adding jumps to the return and to the volatility process. Both approaches lead to a satisfying statistical model fit. Our primary research question is how the non-affine stochastic volatility specifications perform after adding jumps to the return and volatility processes. This is of importance, since jumps constitute an important component in order to explain certain stylized facts in the data. In particular, we are interested in whether these models have a better performance compared to well studied affine models.

We consider stochastic volatility models with jump components in returns and in volatility (SVCJ model class), or just in returns (SVJ model class). The MCMC technique is applied within a Bayesian inference framework to estimate model parameters and latent variables using daily returns from the S&P 500 stock index. When modeling stochastic volatility, we consider different specifications for the diffusion and the variance drift parameters. We leave the class of the exponentially affine models by choosing linear or $3/2$ rather than square root diffusion, and specifying variance drift which is nonlinear in the variance level.

Overall our main findings can be summarized as follows: Using the QQ probability plots we identify the SV- $3/2$ N as the less misspecified non-affine model and the SVJ-SQR and SVCJ-SQR as the less misspecified affine models, whereas, the performances are in the same range. Further, we use statistical (DIC, Bayes factors) as well as economical criteria in order to rank the models according to their performance. Overall, we find that the SV model class is clearly rejected based on the DIC statistics and the Bayes factors in favor of jump diffusion models. In particular, we observe that the SVCJ-VARN and the SVJ- $3/2$ N models are ranked first, respectively second in terms of DIC statics. But since these models have a nonlinear drift specification, which implies a long run mean of zero for the volatility, it is not clear how applicable these models can be for theoretical finance applications. Further, using the QQ probability plots we identify the SV- $3/2$ N as the best performing non-affine model and the SVJ-SQR and SVCJ-SQR models as the best performing

affine models.

We conclude that the non-affine models possess promising statistical properties and are worth of further investigation. Moreover, we find affine models with jump components (SVJ-SQR, SVCJ-SQR) that perform well compared to the best non-affine model (SVCJ-VARN, SVJ-3/2N). Since asset pricing literature is rather developed for the affine model structures, we tend to favor the affine jump diffusion models.

References

- Andersen, T. G., Chung, H.-J. and Sørensen, B. E.: 1999, Efficient method of moments estimation of a stochastic volatility model: A monte carlo study, *Journal of Econometrics* **91**, 61–87.
- Bakshi, G., Cao, C. and Chen, Z.: 1997, Empirical performance of alternative option pricing models, *Journal of Finance* **52**, 2003–2049.
- Bates, D.: 1996, Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options, *Review of Financial Studies* **9**, 69–107.
- Bates, D.: 2000, Post-'87 crash fears in the s&p 500 futures option market, *Journal of Econometrics* **94**, 181–238.
- Berg, A., Meyer, R. and Yu, J.: 2004, Deviance information criterion for comparing stochastic volatility models, *Journal of Business and Economic Statistics* **22**, 107–119.
- Bernardo, J.: 1979, Expected information as expected utility., *Annals of Statistics* pp. 686–690.
- Chernov, M., Gallant, A. and Ghysels, E.: 2003, Alternative models for stock price dynamics, *Journal of Econometrics* **116**, 225–257.
- Christoffersen, P., Jacobs, K. and Mimouni, K.: 2008, Models for s&p 500 dynamics: Evidence from realized volatility, daily returns, and option prices, Working Paper.
- Dempster, A.: 1997, The direct use of likelihood for significance testing, *Statistics and Computing* **7(4)**, 247–252.
- Duffie, D., Pan, J. and Singleton, K.: 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* **68**, 1343–1376.
- Eraker, B., Johannes, M. and Polson, N.: 2003, The impact of jumps in volatility and returns, *Journal of Finance* **58(3)**, 1269–1300.

- Geman, S. and Geman, D.: 1984, Stochastic relaxation, gibbs distributions, and the bayesian restoration of images, *IEEE Tran. on Pattern Analysis and Machine Intelligence* **6**, 721–741.
- Gilks, W.: 1995, Full conditional distributions, in W. Gilks, S. Richardson and D. Spiegelhalter (eds), *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London, chapter 5, pp. 75–88.
- Heston, S.: 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* **6**(2), 327–343.
- Jacquier, E., Polson, N. and Rossi, P.: 1994, Bayesian analysis of stochastic volatility models, *Journal of Business & Economic Statistics* **12**(4), 371–417.
- Jacquier, E., Polson, N. and Rossi, P.: 2004, Bayesian analysis of stochastic volatility models with fat-tails and correlated errors, *Journal of Econometrics* **122**, 185–212.
- Johannes, M. and Polson, N.: 2006, Mcmc methods for financial econometrics, in Y. Aït-Sahalia and L. Hansen (eds), *Handbook of Financial Econometrics*, Elsevier.
- Kass, R. and Raftery, A.: 1995, Bayes factors, *Journal of the American Statistical Association* **90**, 773–795.
- Kullback, S. and Leibler, R.: 1951, On information and sufficiency, *Annals of Mathematical Statistics* **22**, 79–86.
- Merton, R.: 1976, Option pricing when underlying stock returns are discontinuous, *Journal of Financial Economics* **3**, 125–144.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E.: 1953, Equation of state calculations by fast computing machines, *Journal of Chemical Physics* **21**(6), 1087–1092.

- Pan, J.: 2002, The jump-risk premia implicit in options: Evidence from an integrated time-series study, *Journal of Financial Econometrics* **63**, 3–50.
- Pitt, M.: 2002, Smooth particle filters for likelihood evaluation and maximization, Manuscript, University of Warwick.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A.: 2002, Bayesian measures of model complexity and fitting, *Journal of the Royal Statistical Society B* **64** (4), 583–639.

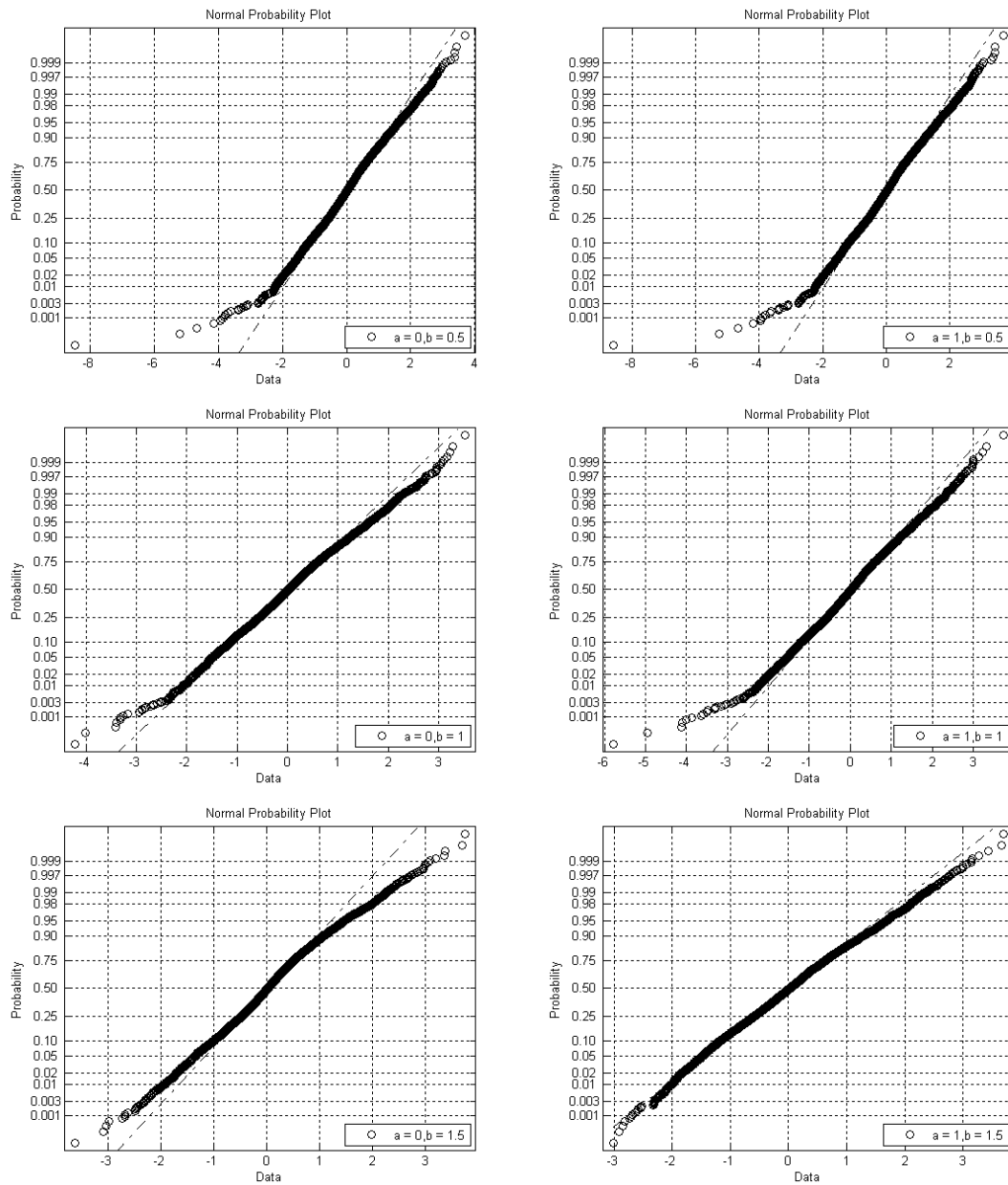


Figure 1: The QQ probability plot: plot quantiles of the daily changes in returns against quantiles from the normal distribution, using residuals in returns from 1986 through 2008 for each SV model specification.

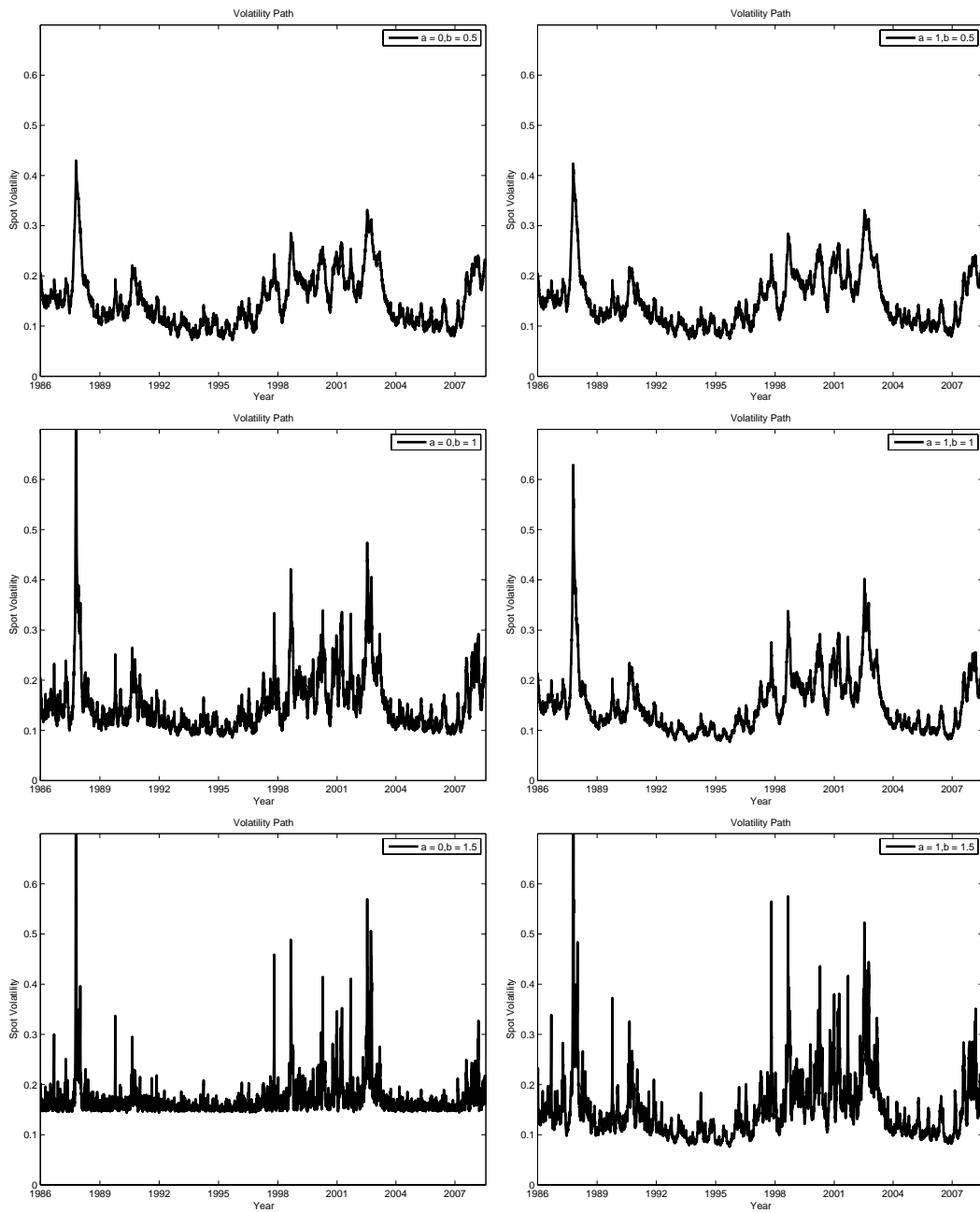


Figure 2: Annualized daily spot volatility path $\sqrt{V_t}$ during 1986-2008 for each SV model specification.

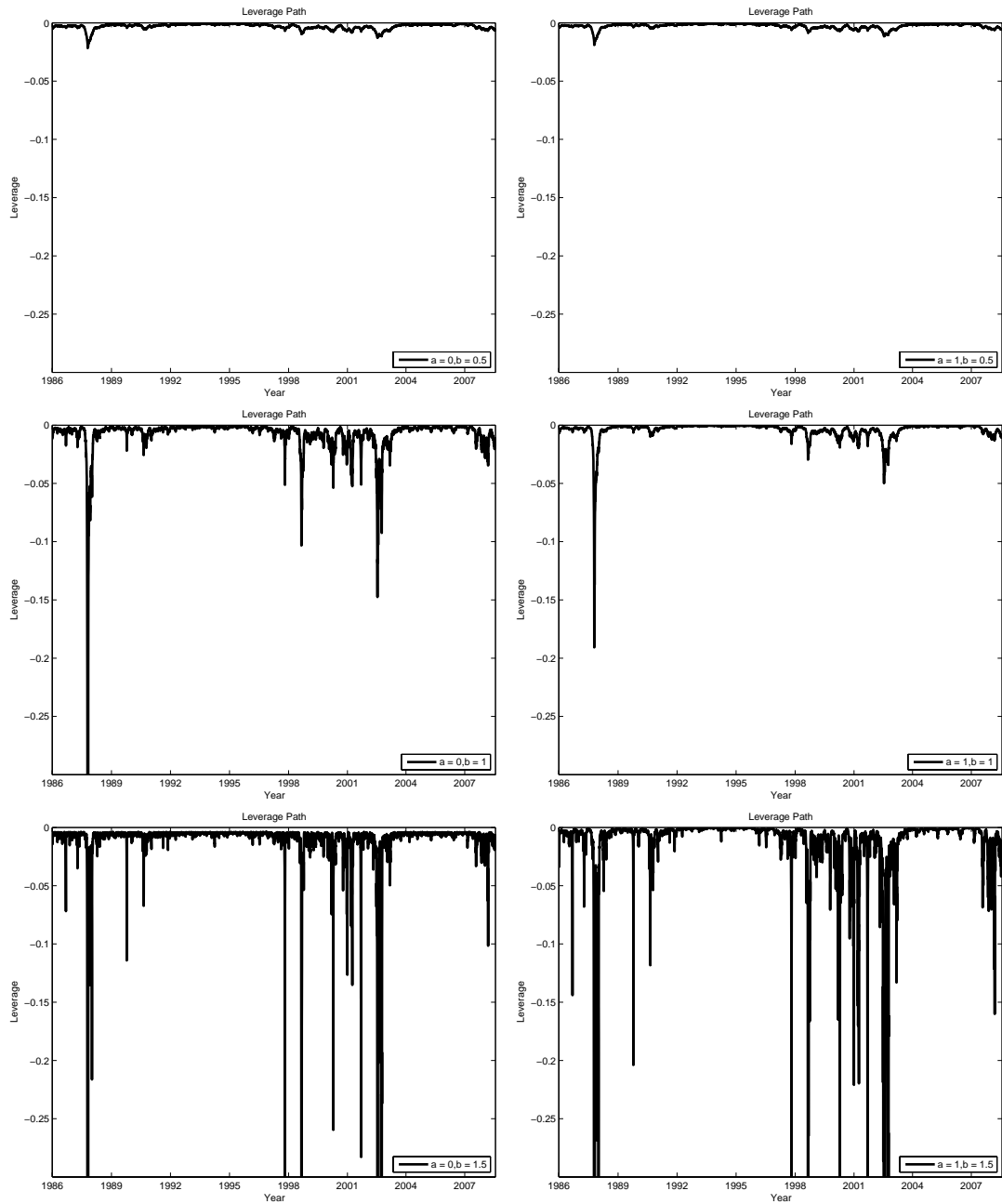


Figure 3: Annualized daily conditional leverage path defined as the conditional covariance between shocks to returns and shocks to variance, $\rho\sigma_v V_t^{b+1/2}$ during 1986-2008 for each SV model specification.

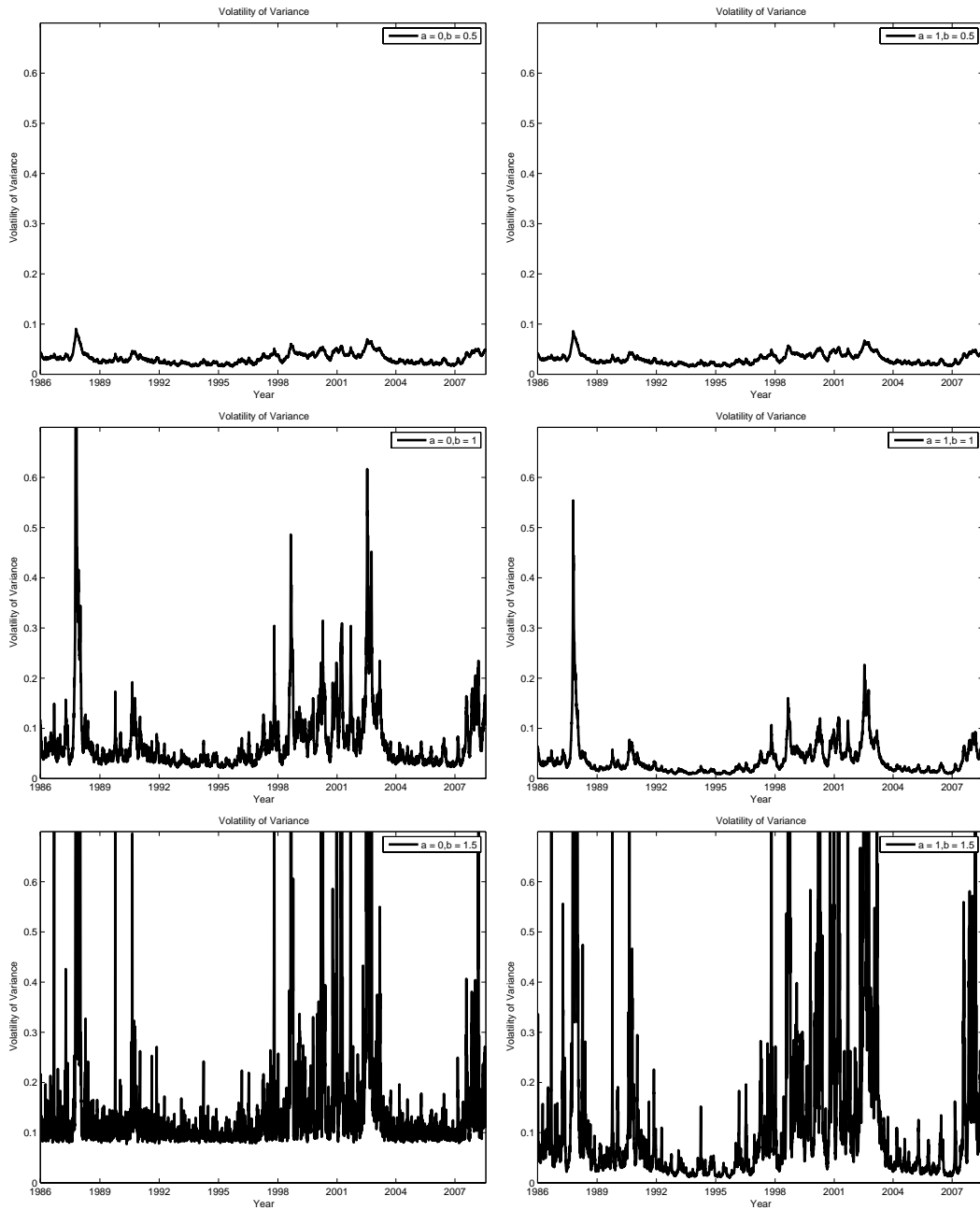


Figure 4: Conditional volatility of the variance path, defined as the square root of the conditional variance of the variance of returns, σV_t^b during 1986-2008 for each SV model specification.

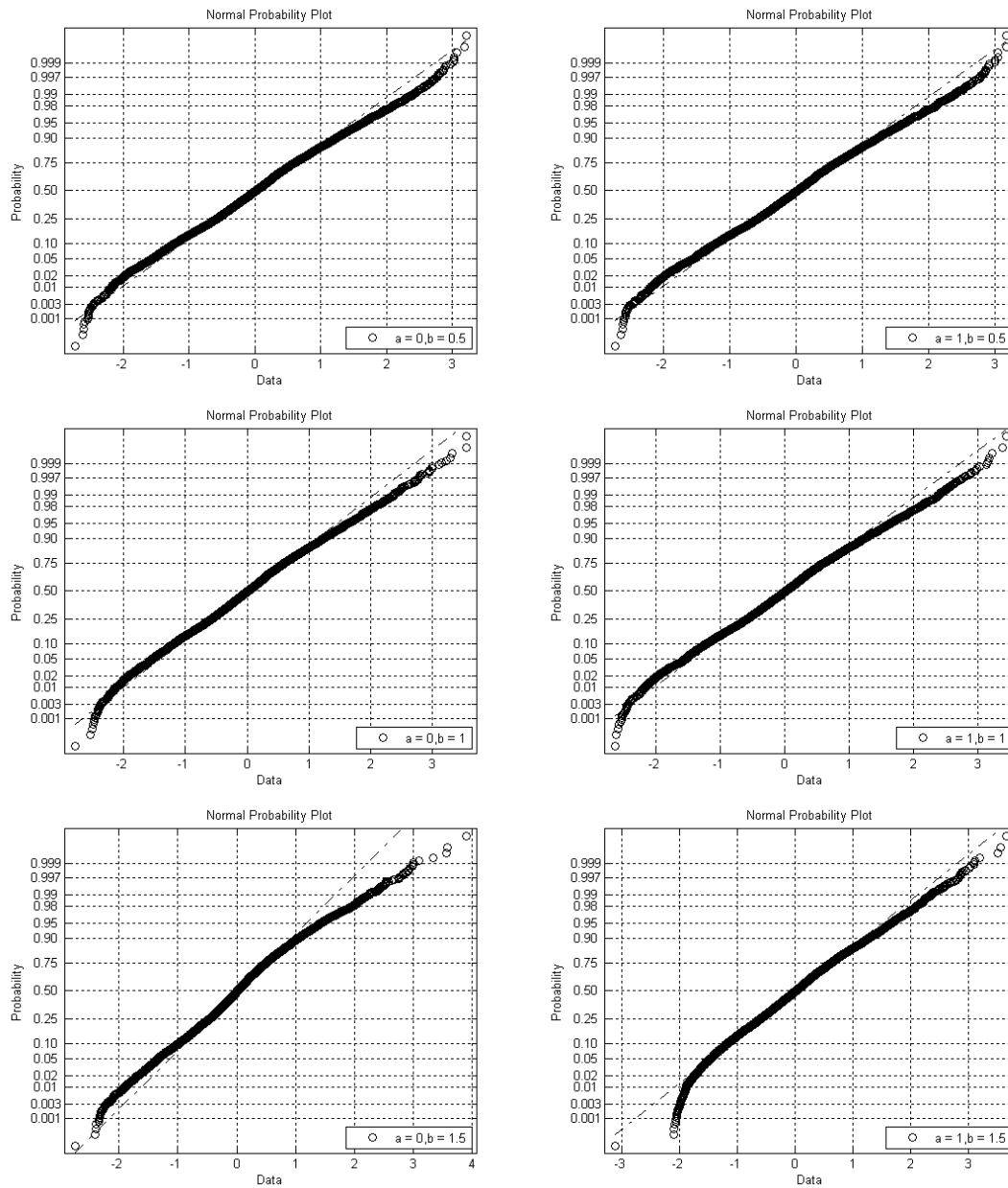


Figure 5: The QQ probability plot: plot quantiles of the daily changes in returns against quantiles from the normal distribution, using residuals in returns from 1986 through 2008 for each SVJ model specification.

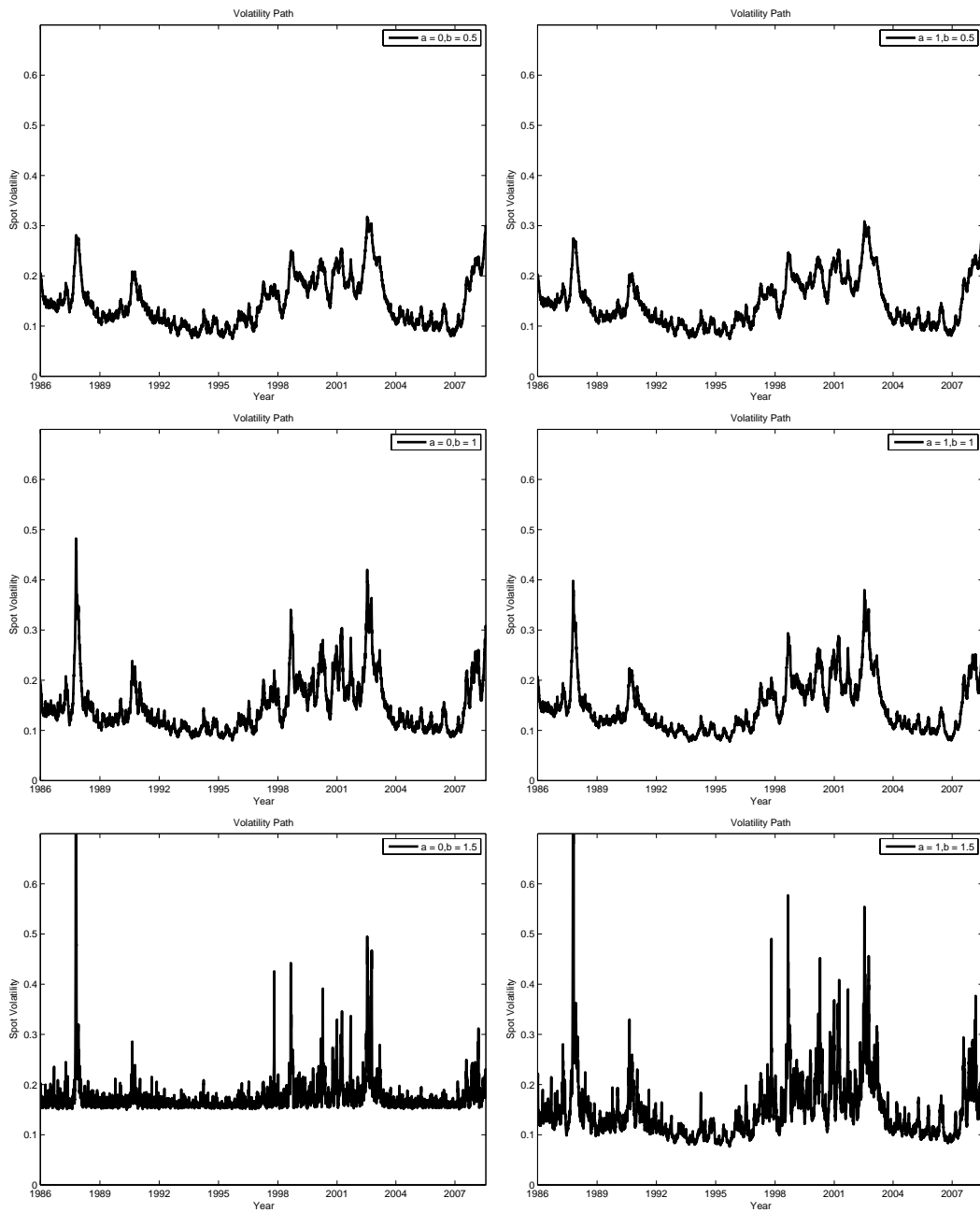


Figure 6: Annualized daily spot volatility path $\sqrt{V_t}$ during 1986-2008 for each SVJ model specification.

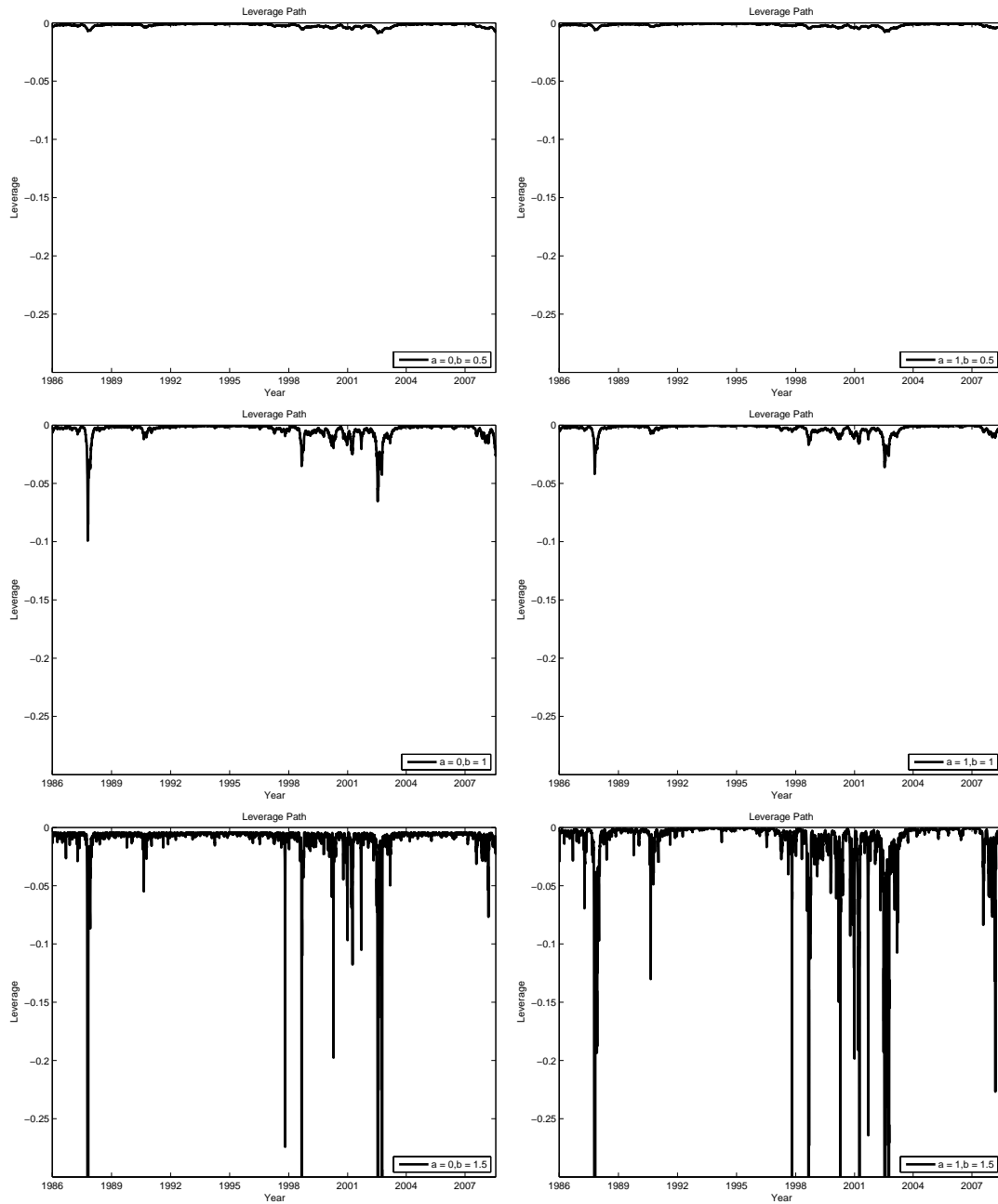


Figure 7: Annualized daily conditional leverage path defined as the conditional covariance between shocks to returns and shocks to variance, $\rho\sigma_v V_t^{b+1/2}$ during 1986-2008 for each SVJ model specification.

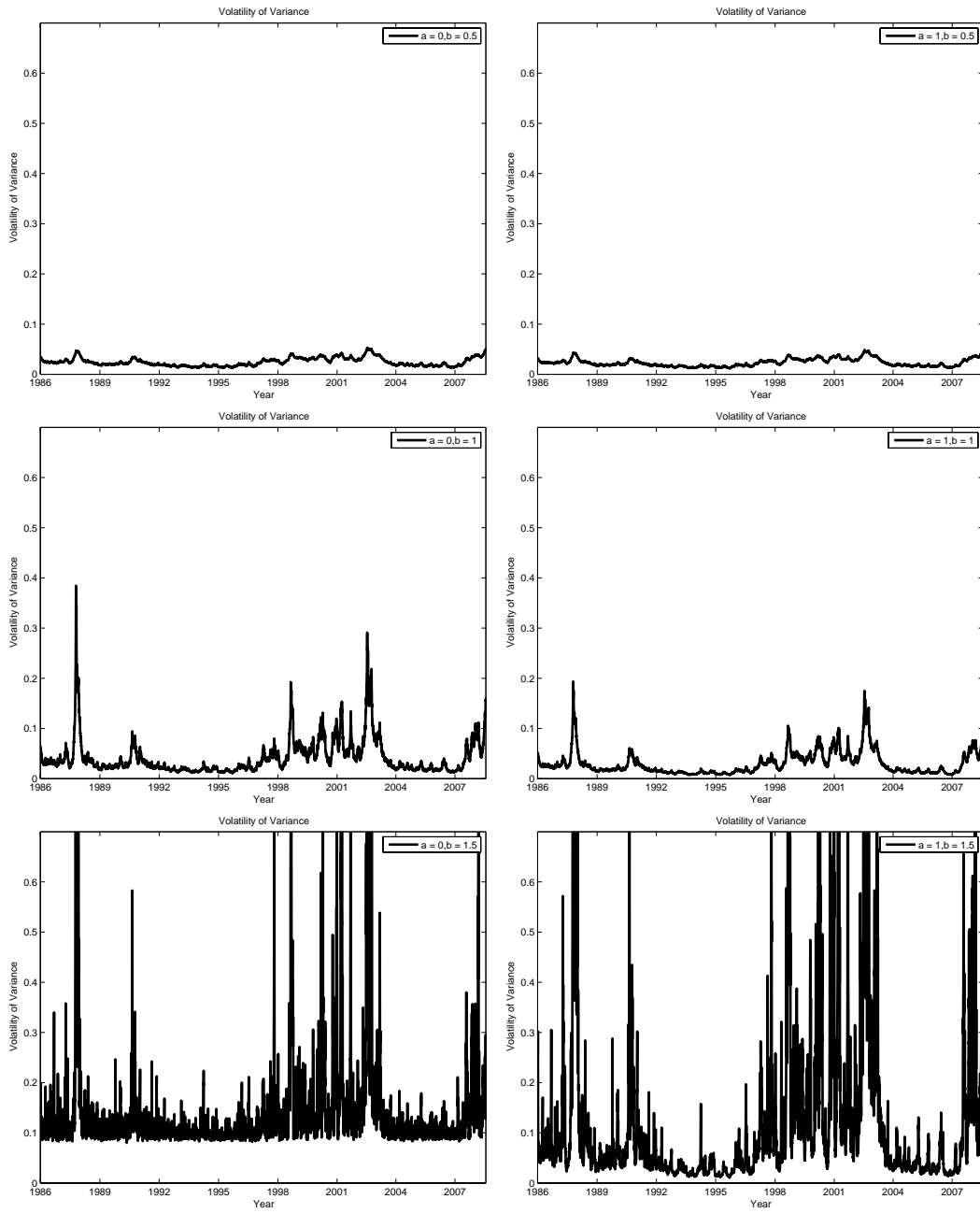


Figure 8: Conditional volatility of the variance path, defined as the square root of the conditional variance of the variance of returns, σV_t^b during 1986-2008 for each SVJ model specification.

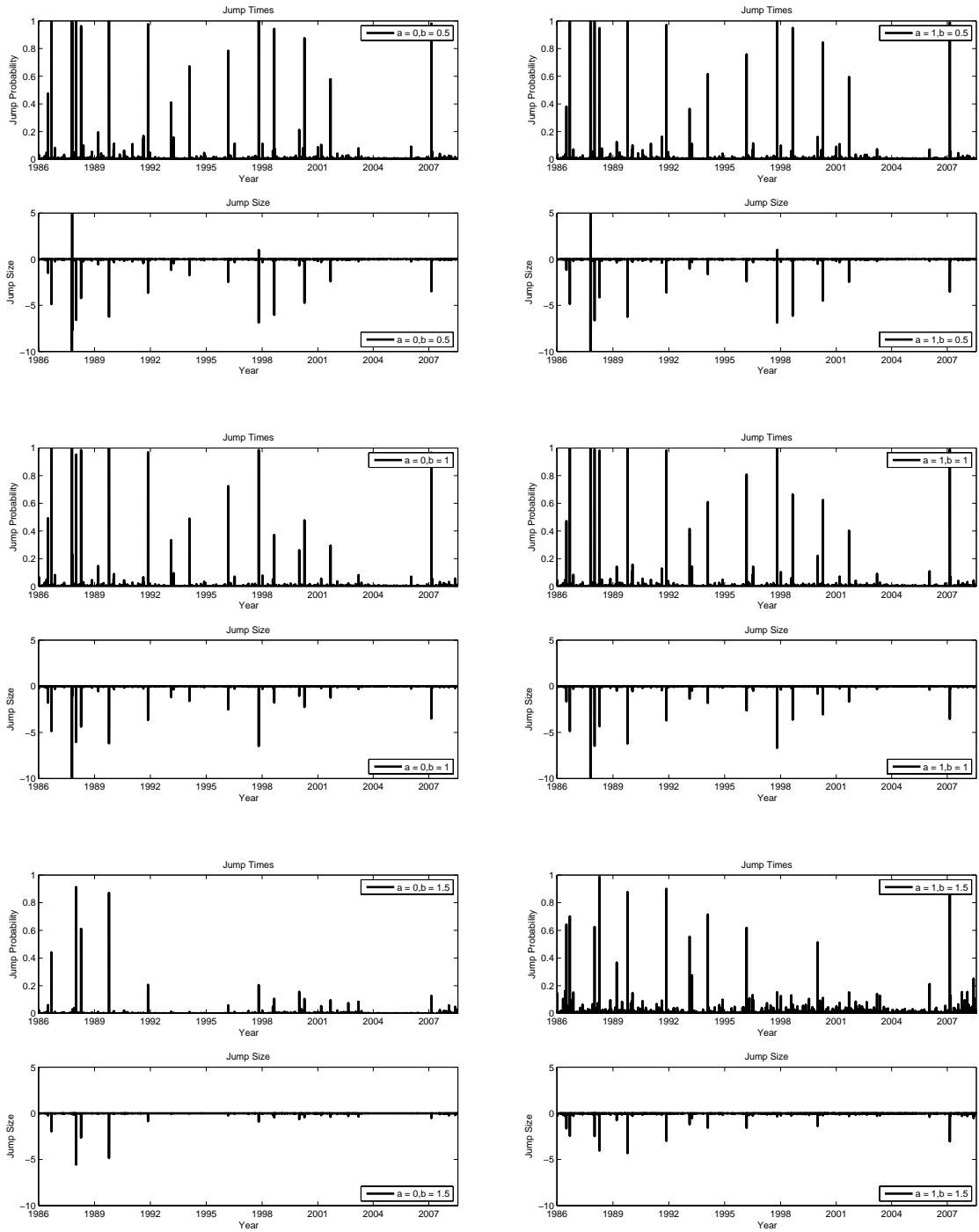


Figure 9: Jump times and jump sizes in returns during 1986–2008 for each SVJ model specification.

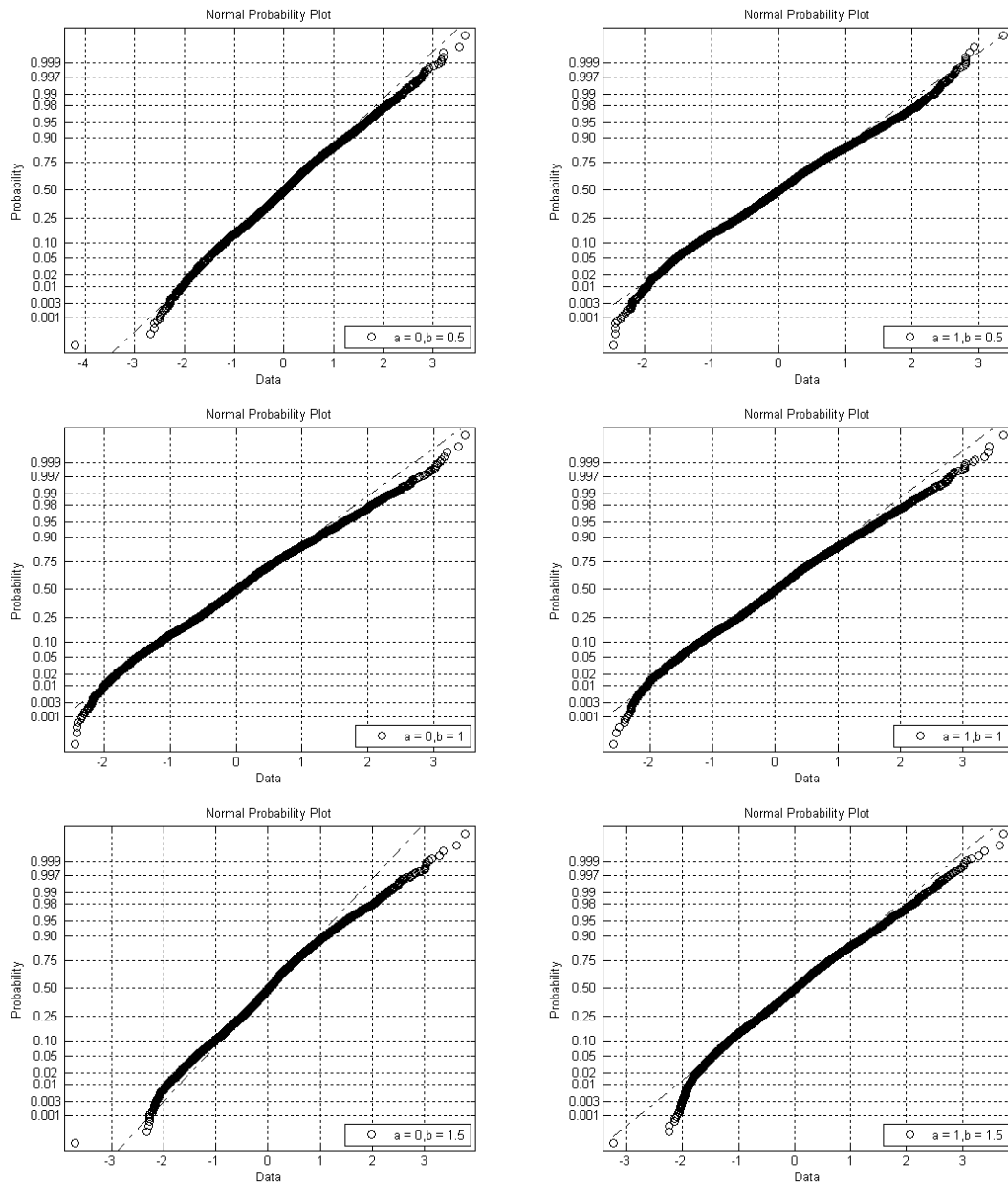


Figure 10: The QQ probability plot: plot quantiles of the daily changes in returns against quantiles from the normal distribution, using residuals in returns from 1986 through 2008 for each SVCJ model specification.

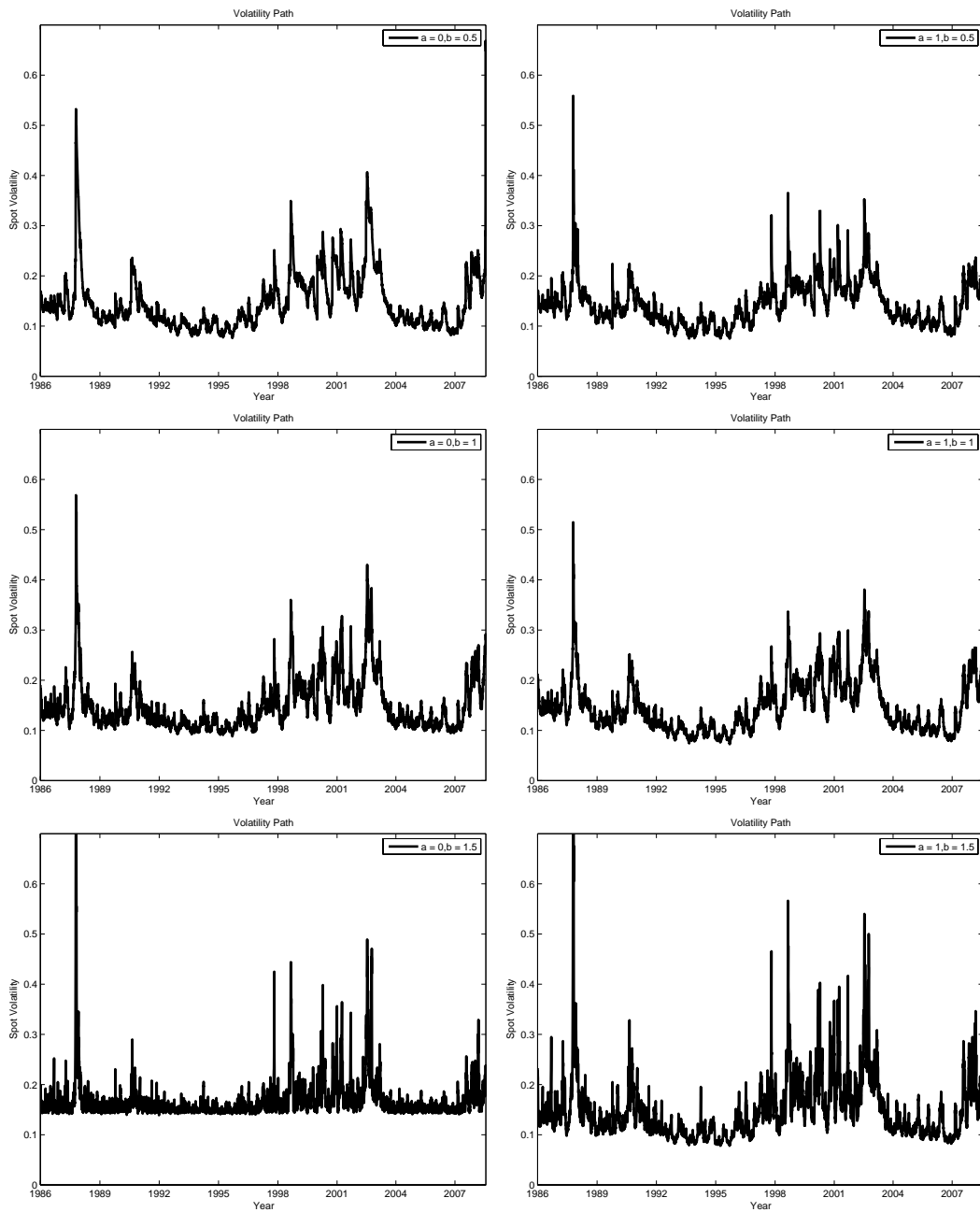


Figure 11: Annualized daily spot volatility path $\sqrt{V_t}$ during 1986-2008 for each SVCJ model specification.

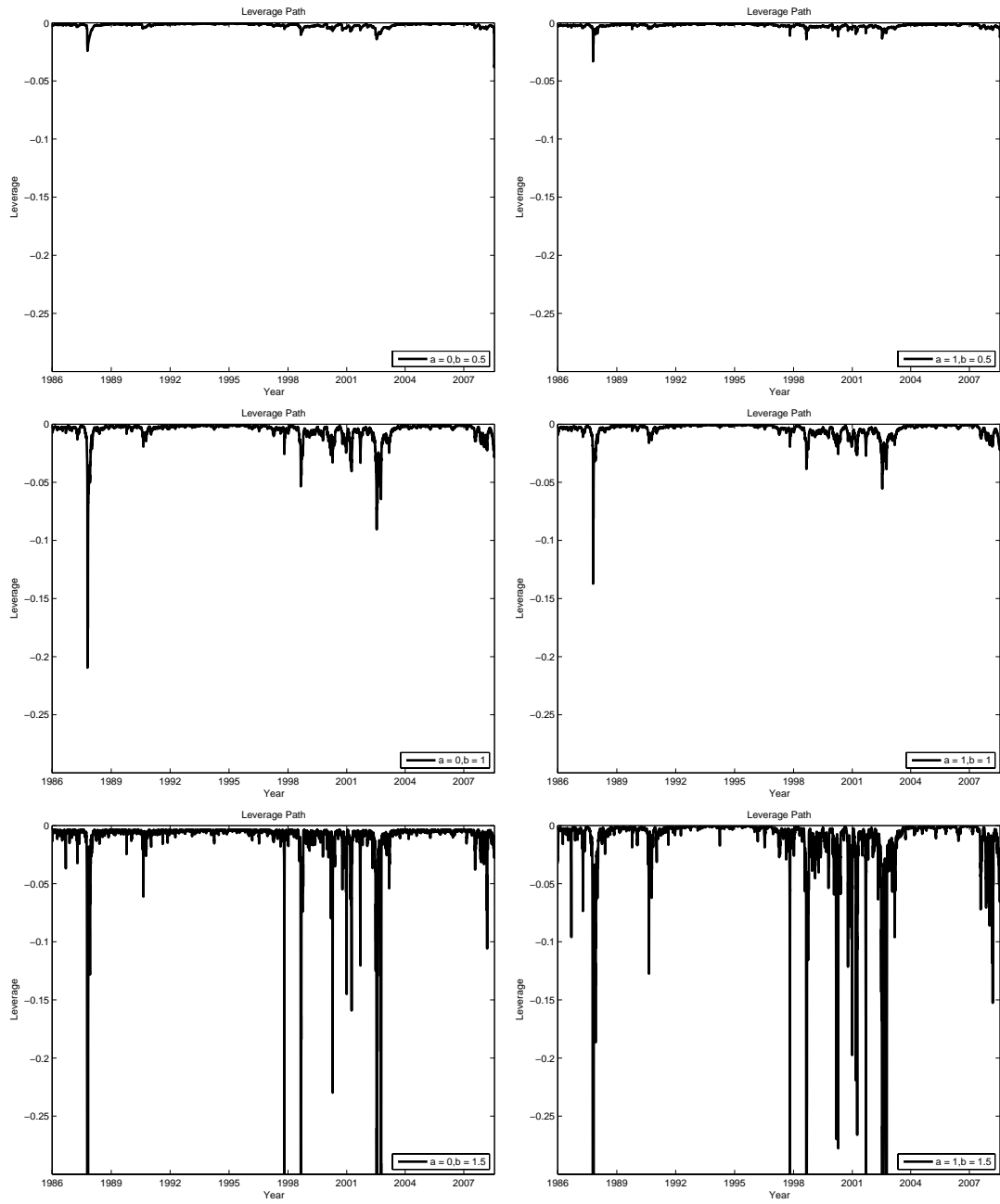


Figure 12: Annualized daily conditional leverage path defined as the conditional covariance between shocks to returns and shocks to variance, $\rho\sigma_v V_t^{b+1/2}$ during 1986-2008 for each SVCJ model specification.

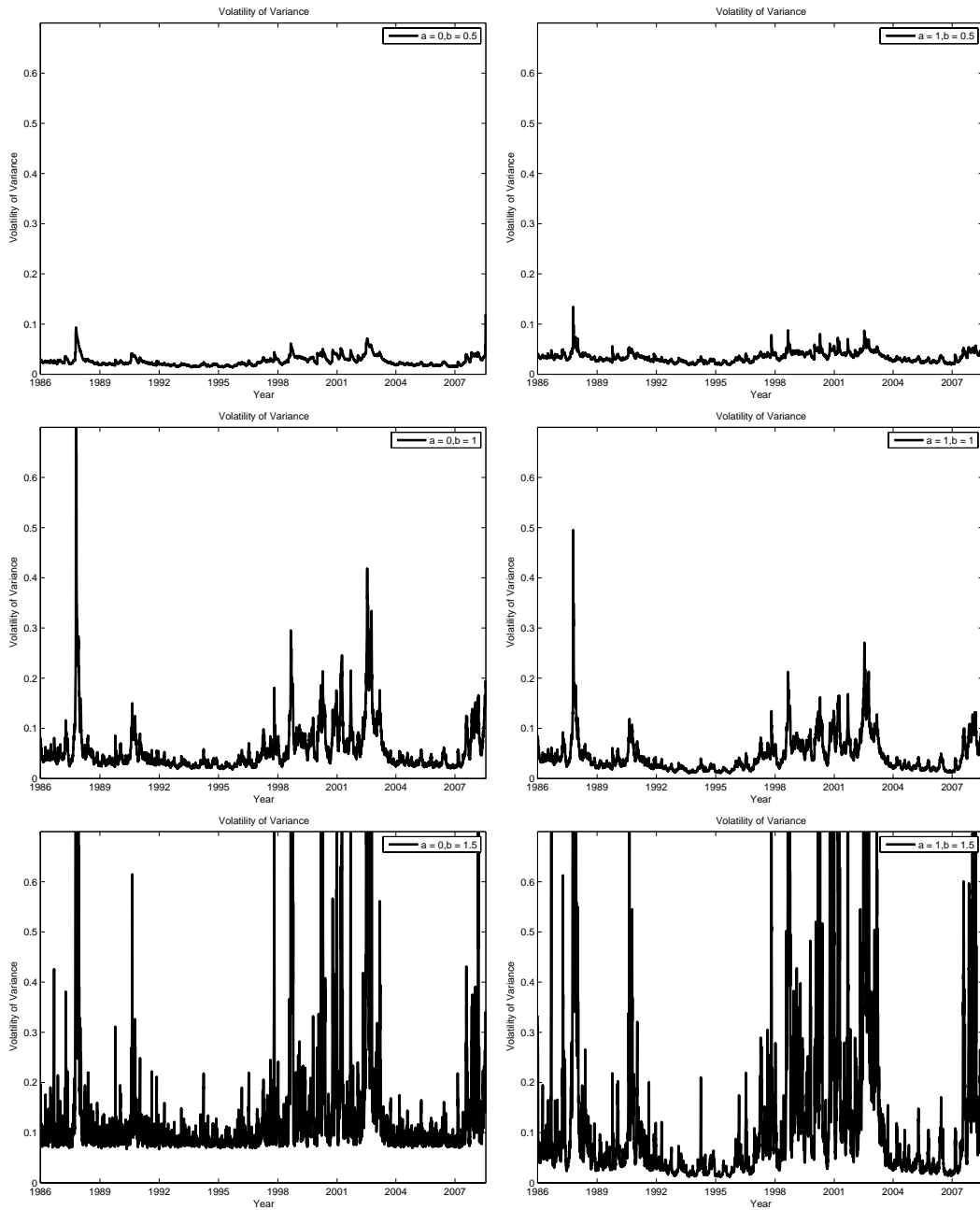


Figure 13: Conditional volatility of the variance path, defined as the square root of the conditional variance of the variance of returns, σV_t^b during 1986-2008 for each SVCJ model specification.

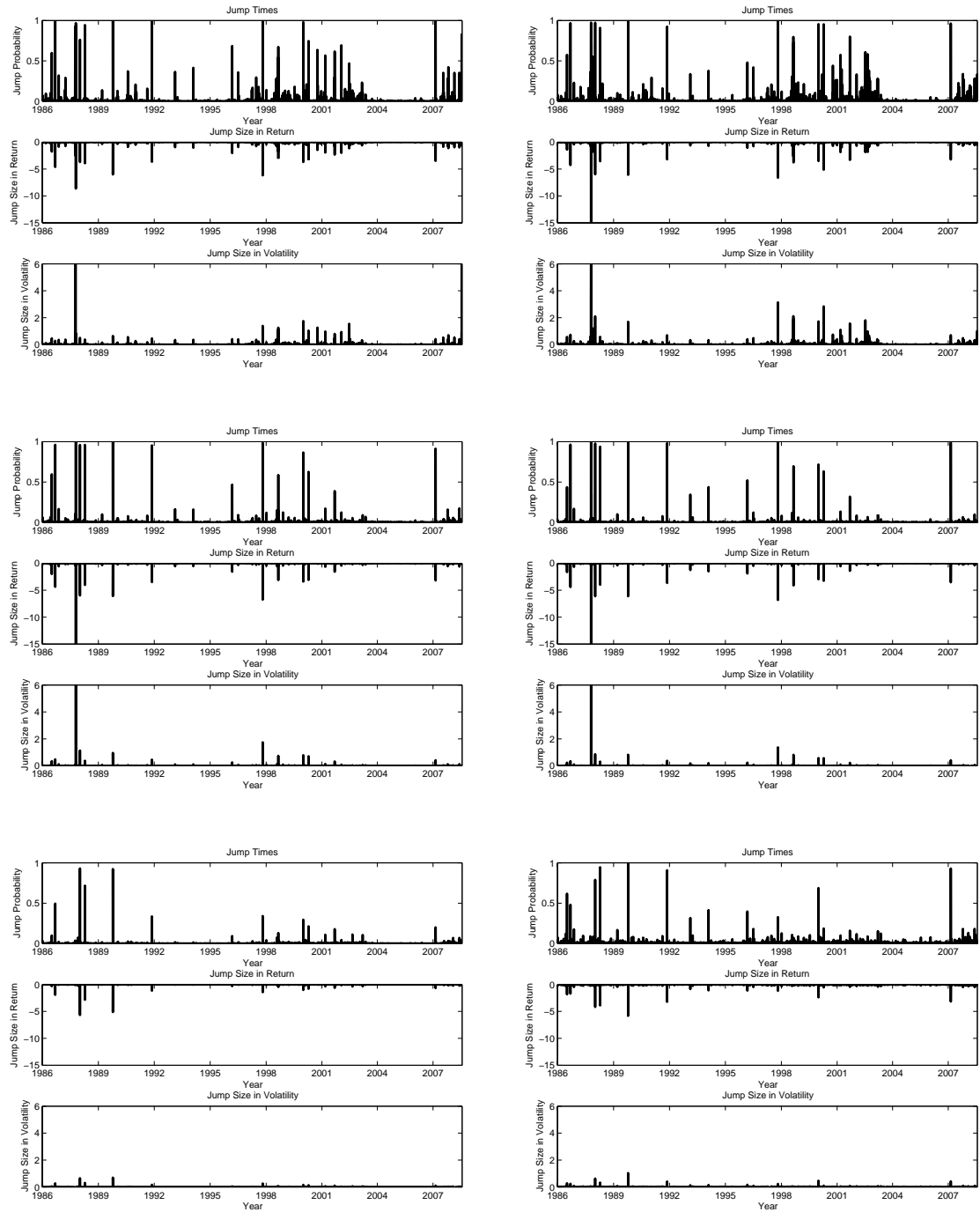


Figure 14: Jump times and jump sizes in returns and in volatility during 1986-2008 for each SVCJ model specification.