

Bayesian Estimation and Option Mispricing

Alberto Vargas Mendoza*

Massachusetts Institute of Technology

December 5, 2011

Abstract

When the Mexican Stock Exchange introduced options over its main index (the IPC) in 2004 it chose Heston's (1993) "square root" stochastic volatility model to price them on days when there was no trading. I investigate whether Heston's model is a good specification for the IPC and whether more elaborate models produce significantly different option prices. To do so, I use an MCMC technique to estimate four different models within the stochastic volatility family. I then present both classical and Bayesian diagnostics for the different models. Finally, I use the transform analysis proposed by Duffie, Pan and Singleton (2000) to price the options and show that the prices implied by the models with jumps are significantly different from those implied by the model currently used by the exchange.

1 Introduction

The Mexican derivatives market (MexDer) introduced options on the main index of the Mexican stock exchange (the IPC) in 2004. Banks and other financial institutions buy these options in order to hedge their positions on the IPC and often hold them until maturity. Thus, there is little trading in these products and in fact most days there is no trading for many of the options. In spite of this, by law, MexDer is required to publish a price for the options each day, even if there was no trading. These prices are then used by banks and investment firms to mark-to-market their positions.

MexDer bases its pricing procedure on Heston's (1993) "square root" stochastic volatility model (MexDer (2008)). This specification offers more flexibility than the classical Black-Scholes (Black

*I am grateful to Victor Chernozhukov, Anna Mikusheva, Jerry Hausman, Isaiah Andrews and the participants of MIT's econometrics lunch seminar for many useful suggestions. I gratefully acknowledge support from CONACyT doctoral grant 205834.

and Scholes (1973)) setting. While the latter assumes a constant level of volatility, Heston's model assumes that volatility follows a mean reverting diffusion process and thus allows for episodes of volatility-clustering. Further, Heston's model retains one of the most useful characteristics of the Black-Scholes setting: a closed-form option pricing formula. Thus, Heston's model becomes a natural candidate when modeling stocks or indices that exhibit volatility-clustering.

Nonetheless, there is more than one way in which asset dynamics can depart from the geometric Brownian-motion assumption of Black-Scholes. In particular, stocks and stock indices tend to have days with extremely large returns (either positive or negative). This fact has led to a large literature including the work by Bakshi et al. (1997), Bates (2000), Pan (2002) and Eraker et al. (2003) that shows that models without jumps tend to be misspecified for stock indices. Although most of this research was conducted on data from US markets, the presence of days with crashes and large positive swings in the IPC suggests that models without jumps will also be misspecified for this index. Hence, a model that accurately describes the dynamics of the IPC will most likely have to include jumps in the return series or simultaneous jumps in the return and volatility dynamics.

In this paper I investigate whether the model used by the exchange is, in fact, misspecified for the IPC. Next, I test whether models with jumps represent a better alternative. Finally, I examine whether the more elaborate models imply significantly different option prices.

To formally show that jumps are necessary to fit the IPC data I estimate models with and without jumps for the IPC and test their relative performance. However, estimation in the context of stochastic volatility models is not trivial. The main complication arises from the fact that the volatility series is not observed. Including jumps complicates the estimation further in two directions. First, jumps increase the number of unobserved state variables in the model. Second, and perhaps more importantly, it is sometimes difficult for estimation algorithms to disentangle the variation arising from the diffusive element of the dynamics and those arising from the jump component (see e.g. Honore (1998)) .

In spite of these difficulties, several techniques have been developed to estimate and test this family of models. These have included calibration (Bates (1996)), Implied State GMM (Pan (2002)), Efficient Method of Moments (Andersen et al. (1999)) and Markov Chain Monte Carlo (Jacquier et al. (1994) and Eraker et al. (2003)).

When choosing an empirical technique for this problem, an important requirement was that I could use the output to price options. Hence, I need not only estimates of the parameters of the

models, but also of the unobserved state variables. The MCMC technique of Eraker et al. (2003) does precisely this: it produces estimates of both the parameters and unobserved variables, and thus serves as the appropriate tool for the purpose of this paper.

Beyond the actual estimation, I am interested in the relative performance of the different models under consideration. To investigate this, I perform two kinds of tests. The first is a standard normality test of residuals, while the second is based on Bayes factors and is particularly useful for comparing nested models. Just like option valuation, this procedure depends crucially on estimating the latent variables as well as the parameters, and so can only be conducted after certain estimation procedures.

From the estimation and model selection results, I conclude that the current stochastic volatility model used by MexDer is misspecified. Further, I find that the model with jumps only in returns is still misspecified, but is an improvement over Heston's model. Finally, I show that models with simultaneous jumps in volatility and returns represent a much better fit for the index dynamics.

Using the transform analysis introduced by Duffie et al. (2000), I then estimate option prices and show that those implied by the stochastic volatility and jumps models are significantly different than those implied by Heston's model. Hence, I find that the option prices implied by a better fitting model for IPC dynamics are significantly different than those currently used to mark-to-market positions.

The rest of the paper is organized as follows: section 2 introduces the family of stochastic volatility and jumps models that we will be working with, section 3 presents the estimation methodology, section 4 describes the data, section 5 presents the estimation and model selection results, section 6 presents the option pricing results, and section 7 concludes.

2 Models of Price Dynamics

The Black-Scholes model introduced in 1973 is one of the most widely-used tools in finance. Not only does Black-Scholes provide a set of reasonable assumptions for the dynamics of asset prices, it also has the great advantage of implying a closed-form formula for the pricing of European-style options. Formally, the Black-Scholes model assumes that the prices (S_t) of the underlying asset

follow a Geometric Brownian Motion, which implies:

$$dS_t = mS_t dt + \sigma S_t dW_t,$$

where W is a standard Brownian motion process and m and σ are fixed parameters. By a simple application of Ito's lemma it implies that $Y_t = \log(S_t)$ behaves according to:

$$dY_t = \mu dt + \sigma dW_t.$$

This formulation provides an immediate interpretation of the model: dY_t can be thought of as an instantaneous measure of log-returns. Thus the model tells us that log-prices grow according to a deterministic trend (μdt) that is perturbed by a Brownian motion with volatility σ . Putting together these two assumptions we have a very reasonable starting point for modeling asset dynamics. Further, under no arbitrage assumptions, the Black-Scholes model implies that the price of a European call that matures at time T with strike price K is given by:

$$S_t N(d_1) - e^{-r(T-t)} K N(d_2), \quad (1)$$

where N is the standard normal cumulative distribution function, r is the risk free rate and d_1 and d_2 are functions of $K, S_t, (T-t), \sigma$ and r .

However, the dynamics of stocks and stock indices exhibit phenomena that are not consistent with the assumptions of the Black-Scholes model. In particular, stock indices tend to experience volatility-clustering; that is, there are periods when the volatility of returns is very high and periods when it is low.

Further, all the variables in the option pricing formula are observable, except σ . Thus, if option prices are observed, σ can be calculated from the option pricing formula, and if the Black-Scholes model were true, the estimated σ should be the same for every option regardless of its strike price or time to maturity. However, a well-known result is that the implied volatility is highest when the strike price is very high or very low. This result, known as "volatility smiles" or "volatility smirks", is extensively documented in Stein (1989), Aït-Sahalia and Lo (1998) and Bakshi et al. (2000).

Thus, stochastic volatility models were introduced as generalizations of the Black-Scholes framework that allows for volatility-clustering and potentially provides a solution to the "volatility smiles"

issue. Some of the first models to include a volatility term that could vary with time were those of Scott (1987), Hull (1987), and Wiggins (1987). However, it was Heston's 1993 model that became widely used because it shared a crucial characteristic with the Black-Scholes model: it offered a closed-form solution to pricing European options. Formally, Heston's model assumes the following asset dynamics:

$$\begin{aligned} dS_t &= mS_t dt + \sqrt{V_t} S_t dW_t^1 \\ dV_t &= \kappa (\theta - V_t) dt + \sigma \sqrt{V_t} dW_t^2, \end{aligned}$$

where W_t^1 and W_t^2 are Brownian motion processes with correlation ρ . Thus, Heston's model is a generalization of Black-Scholes, where the volatility of the price diffusion is allowed to vary with time. In particular, the volatility is allowed to follow a diffusion similar to an Ornstein–Uhlenbeck process, but where the volatility of the diffusive element is proportional to $\sqrt{V_t}$. In spite of this difference, it is important to note that a feature that the volatility process shares with Ornstein–Uhlenbeck processes is the mean reversion of the deterministic drift. This gives an immediate interpretation to θ as a long-term value the process tends to return to and κ as the speed with which the process will tend to return to θ .

Precisely because volatility is allowed to vary, this specification can accommodate periods when returns will be very volatile, but the model also implies that volatility will tend to return to its long term value θ and thus periods of high volatility will be transient.

This added flexibility, together with the availability of a closed-form pricing formula for European-style options, makes the Heston model a considerable improvement over Black-Scholes. However, the added flexibility provided by the introduction of stochastic volatility does not seem to be enough.

Episodes of "crashes" cannot be captured by these dynamics. For instance, under Heston's model, a crash like that of 1987 when the S&P 500 lost more than 20% in a single day would imply an observation of more than 8 standard deviations away from the mean. In order to capture these crashes models with jumps in the return series were introduced. These jumps take the form of occasional shocks to the return series which will explain large (either positive or negative) returns in a single day which could not be explained by a continuous diffusion such as those in Black-Scholes' or Heston's specification.

The use of jumps in returns goes back to at least Merton's 1976 model which augments the Black-Scholes model by adding normally distributed jumps that occur according to a Poisson process. So,

following the notation I have been using, Merton's model implies:

$$dY_t = \mu dt + \sigma dW_t + \xi^y dN_t^y,$$

where N is a Poisson process, and ξ^y is a Gaussian random variable.

This family of models will be able to explain days with sudden large returns. However, since it assumes σ to be fixed throughout time, it will not be able to describe processes where there are periods of high volatility and periods of low volatility. Thus, a natural idea is to combine stochastic volatility and jumps in the return series. Following this direction, Bates (1996) introduces the following model:

$$\begin{aligned} dY_t &= \mu dt + \sqrt{V_t} dW_t + \xi^y dN_t^y \\ dV_t &= \kappa (\theta - V_t) dt + \sigma \sqrt{V_t} dW_t^2. \end{aligned}$$

This gives us a specification that can accommodate both volatility-clustering via the stochastic volatility element and sudden large returns through the jump component. So far, we have a model built around Black-Scholes, with added characteristics tailored to explain the phenomena that Black-Scholes cannot. Hence, we would expect this new model to provide a good fit to the data of stock indices. In spite of this, Bakshi et al. (1997), Bates (2000) and Pan (2002) find that models with jumps only in the return series are still misspecified for stock index data. The rationale is the following: having jumps in the return series can explain an isolated episode of a very large return. However, data shows that frequently, when there is a day with a very large loss, the next day tends to have a very large gain. In contrast, under Bates' specification, jumps arrive according to a Poisson process with arrivals a few times per year, thus observing consecutive days with jumps on several occasions would once more imply that we are observing very low probability events.

A natural extension of Bates' model is to add a jump process to the dynamics of the volatility series. This will allow for a sudden increase in the volatility of returns, therefore allowing for periods when there are very large returns on consecutive days. Although this is achieved partially by introducing stochastic volatility, by introducing jumps in V we allow for a sudden "burst" in the returns instead of a gradual increase in their dispersion. Thus, Duffie et al. (2000) introduce a

family of models that generalize Bates' as follows:

$$\begin{aligned} dY_t &= \left(m - \frac{1}{2}V_t\right) dt + \sqrt{V_t}dW_t + \xi_t^y dN_t^y \\ dV_t &= \kappa(\theta - V_t) dt + \sigma\sqrt{V_t}dW_t^2 + \xi_t^v dN_t^v, \end{aligned} \tag{2}$$

where N^v is a Poisson process, and ξ^v represents the magnitude of jumps in volatility at time t . Although Duffie et al. (2000) allow for a variety of distributions for ξ^v , we will be interested in the case where ξ_t^v is exponentially distributed. Clearly, the model described by system (2) is a very general specification that allows for jumps in both returns and volatility. Thus, sudden "bursts" in the volatility of returns can be described in this framework. Further, since it retains jumps in returns, it can potentially encompass all the phenomena that we have described above and that simpler models are not capable of capturing.

For the rest of this paper, we will work with a variation on model (2). We begin by presenting this general stochastic volatility with jumps model and later describe restrictions that will allow us to test whether jumps are necessary for the particular index we are modeling.

2.1 General Stochastic Volatility with Jumps Model

Let S_t denote the price of the index at time t and let $Y_t = \log(S_t)$. The most general form of stochastic volatility model with jumps which we are interested in can be expressed as:

$$\begin{aligned} dY_t &= \mu dt + \sqrt{V_{t-}}dW_t^1 + \xi_t^y dN_t^y \\ dV_t &= \kappa(\theta - V_{t-}) dt + \sqrt{V_{t-}}\sigma_v \left(\rho dW_t^1 + \sqrt{1-\rho^2}W_t^2\right) + \xi_t^v dN_t^v, \end{aligned} \tag{3}$$

where $V_{t-} = \lim_{s \uparrow t} V_s$, (W^1, W^2) is a standard Brownian Motion in \mathbb{R}^2 , N_t^y and N_t^v are Poisson processes with constant intensities λ_y and λ_v respectively and $\xi_t^v \sim \exp(\mu_v)$ and $\xi_t^y \sim N(\mu_y + \rho_J \xi_t^v, \sigma_y^2)$ are the jump sizes in volatility and returns respectively.

Thus, we are assuming that log-prices behave according to a Brownian motion with drift μ and volatility \sqrt{V} . The departure from standard Brownian motion with drift is that this process is hit by normally distributed shocks with arrival times that follow a Poisson process. Furthermore, the volatility \sqrt{V} itself behaves according to a mean-reverting diffusion with long term value θ that is hit by exponentially distributed shocks. These shocks also occur according to a Poisson process. Finally, the Brownian motions driving both diffusions are assumed to have a correlation of ρ .

This general specification includes the Black and Scholes (1973) and the Heston (1993) specification as particular cases: if $\sigma_v = 0$ and $N_t^y \equiv N_t^v \equiv 0$ for all t , Y would follow a geometric Brownian motion, as described by the Black-Scholes model. If instead we only restrict $N_t^y \equiv N_t^v \equiv 0$ for all t we have Heston's "square root" stochastic volatility model (SV model from now on). In addition, we will also be interested in the cases where:

1. $N_t^v \equiv 0$ the case where there are jumps only in the returns and not in the volatility (SVJ model).
2. $N_t^y \equiv N_t^v$ the case where there are correlated jumps in both returns and volatility, but they are restricted to occurring simultaneously (SVCJ model).
3. The case where there are jumps in both returns and volatility and these jumps occur at independent times (SVIJ model).

3 Estimation Methodology

With all the models presented above, we face a common problem: all of these models describe continuous time dynamics, but the data will invariably be discretely sampled. Thus the estimation of the parameters will always be for a discrete time analog of the continuous time model we are interested in. Nonetheless, evidence that goes back at least as far as Merton (1980) shows that estimates of parameters based on discretization of continuous time models can be biased. Further, Melino (1994) offers a broad overview of the possible problems that may arise when estimating discretized versions of continuous-time models. However, through simulation exercises, Eraker et al. (2003) show that, for the range of parameters we are interested in, these biases become negligible when the discretization is at the daily level.

The structure of the Black-Scholes model or Merton's Jump-Diffusion lead to moment conditions and likelihood functions which are straightforward to construct, thus allowing for GMM estimation as in Fisher (1996) et al.. The estimation of stochastic volatility models is more challenging, however, since the volatility series cannot be observed. Thus, several authors have estimated the unobserved parameters by using data contained in the prices of options. For instance, Ledoit et al. (2002) use the Black-Scholes implied volatilities of at-the money short-maturity options as a proxy for volatility. A more elaborate method presented in Pan (2002) calculates the implicit volatility from option prices,

but using the option pricing model that corresponds to the asset dynamics she is assuming instead of the "naive" Black-Scholes implied volatilities. This then allows her to construct moment conditions, and form a variation of GMM, known as implied state GMM.

Another problem of having only discretely-observed data for a continuous-time model, is that the likelihood function of the observations is usually not explicitly computable. However, Aït-Sahalia (2002) proposes a maximum-likelihood method that approximates the likelihood function of the observations. To do so he uses Hermite polynomials to construct a sequence of closed-form functions that converge to the unknown likelihood function. This method was generalized to stochastic volatility models in Aït-Sahalia and Kimmel (2007). However, a clear drawback of this generalization is that it requires the prices of options or some proxy for the unobserved volatility.

Since the purpose of this paper is to show that the prices published by MexDer come from a misspecified model, we can not use the time series of option prices as part of the data used for inference. Thus, the methods described so far will not be useful in our context. Instead, we must explore the options available for estimation when the volatility series is not observable and option data is not available either.

These methods fall broadly into two categories. The first category comprises variations of the simulated method of moments proposed by Pakes (1989) and McFadden (1989). In particular, Gallant and Tauchen (1996) constructed a variation known as the efficient method of moments that has the significant advantage of being as efficient as maximum likelihood under certain conditions. This method has been successfully used by Andersen et al. (1997) and Andersen et al. (1999) to estimate a wide variety of models for the dynamics of stock indices.

The second approach, which we follow, uses Bayesian MCMC estimators. Their use dates back to Jacquier et al. (1994) but also includes Kim et al. (1999), Eraker (2001) and Eraker et al. (2003). A fundamental advantage of these methods is that they can produce estimates of the unobserved state variables and not only of the parameters. This is crucial for our purposes, since we are interested in using the results of the estimation to price options, and the (unobserved) volatility is an input required by the option-pricing formulas.

The rest of this section will be devoted to describing the Bayesian technique used for the estimation of the models of interest. This technique is a Markov Chain Monte Carlo procedure based on Eraker et al. (2003).

3.1 Estimated Model and Priors

We begin by constructing the Euler discretization of model (3), and we take it as a true model in the sense that the discretized model will have the same likelihood function as the continuous time model sampled at discrete time intervals:

$$\begin{aligned} Y_t - Y_{t-1} &= \mu + \sqrt{V_{t-1}}\varepsilon_t^y + \xi_t^y J_t^y \\ V_t - V_{t-1} &= \kappa(\theta - V_{t-1}) + \sigma_v \sqrt{V_{t-1}}\varepsilon_t^v + \xi_t^v J_t^v, \end{aligned} \tag{4}$$

where $J_t^y, J_t^v = 1$ indicates a jump arrival and $\varepsilon_t^y, \varepsilon_t^v$ are standard normal random variables with correlation ρ .

Now, let Θ denote the collection of parameters we wish to estimate so:

$$\Theta \subseteq \{\mu, \kappa, \sigma_v, \theta, \rho, \mu_y, \sigma_y, \rho_J, \mu_v, \lambda_y, \lambda_v\}.$$

In the context of Bayesian estimation, in addition to the model specification described by model (4) we will also need a set of priors for the parameters. Priors are usually thought of as containing some information about the parameters that is not contained in the data, but a statistician believes should be included in the estimation. However, for most of the parameters we wish to impose no such prior information. Hence, "uninformative" priors are used.

So, let $p(\Theta)$ denote the prior distributions for the parameters. These prior distributions will be:

$$\begin{aligned} \mu &\sim N(1, 25), & \mu_y &\sim N(0, 100), \\ \kappa &\sim N(0, 1), & \sigma_y^2 &\sim IG(5, 20), \\ \kappa\theta &\sim N(0, 1), & \mu_v &\sim \Gamma(20, 10), \\ \sigma_v^2 &\sim IG(2.5, 0.1), & \rho_J &\sim N(0, 4), \\ \rho &\sim U(-1, 1), & \lambda_v = \lambda_y &\sim \beta(2, 40). \end{aligned}$$

Notice, for instance, that the prior for μ has a standard deviation of 5. Later, we will see that our data has an average daily return of around 0.1%. Therefore, by imposing a prior with a standard deviation several orders of magnitude larger than the initial guess for the parameter, we are essentially imposing no prior information. Analogous arguments can be constructed for most of the parameters except for the case of λ and σ_y^2 . These priors restrict the models with jumps so

that jumps are infrequent and of magnitudes larger than what could be explained by the stochastic element of the diffusions. We could think that large returns observed infrequently could be part of a jump process with a much higher rate of occurrence and the possibility of either large or small jump magnitudes. However, by introducing frequent jumps the estimation procedure might have problems distinguishing between variation attributable to the diffusive element of the process and the jump element. These difficulties are described in Honore (1998) and in Aït-Sahalia (2009) and the references therein. Thus, it is reasonable to impose an informative prior for these two parameters.

Finally, the choice of the parametric forms of the priors is standard in the literature. Later we will see that the reason that these distributions have become standard is that the priors and their posteriors are conjugate distributions.

3.2 Estimation Algorithm

Let Ω be the set of parameters together with the latent variables. Thus Ω includes all the values we wish to estimate. Further, if we denote V, J, ξ^y and ξ^v as the entire vectors of V_t, J_t, ξ_t^y and ξ_t^v sampled at the same discrete intervals as Y_t , then:

$$\Omega = \{\Theta, J, \xi^y, \xi^v, V\}.$$

Thus, the objective of our Bayesian estimation is to find the posterior distribution of Ω given the observed log prices Y . A simple application of Bayes' theorem tells us that:

$$p(\Omega|Y) \propto p(Y|\Omega)p(\Omega).$$

However, $p(\Omega|Y)$ is not a standard distribution so drawing from it is not a trivial task. Therefore, to find $p(\Omega|Y)$ we use a hybrid - MCMC algorithm from Eraker et al. (2003).

3.2.1 MCMC Procedure

The general idea of MCMC procedures is to construct a Markov Chain that has as its stationary distribution the distribution we are interested in i.e. $p(\Omega|Y)$. We begin by dividing Ω into subsets (which will be discussed later). Each element of Ω is then assigned a starting value $\Omega^0 = (\omega_1^0, \dots, \omega_r^0)$. Then, the c 'th iteration of the algorithm consists of obtaining a draw for each $\omega_i \in \Omega$. Specifically,

we draw the i 'th component ω_i^c from the distribution of ω_i^c conditional on the data and the most current values of all the other components of Ω :

$$\omega_i^c \sim p\left(\omega_i | \tilde{\Omega}_{-i}^c, Y\right),$$

where

$$\tilde{\Omega}_i^c = (\omega_1^c, \omega_2^c, \dots, \omega_{i-1}^c, \omega_{i+1}^{c-1}, \dots, \omega_r^{c-1}).$$

Hence, in each step we are drawing on the distribution of ω_i conditional on the data, and on the rest of parameters taking the values of the most recent draws we have made of those parameters. Note that here $p\left(\omega_i | \tilde{\Omega}_{-i}^c, Y\right)$ is the true conditional distribution.

These steps are then repeated a large number of iterations until we believe that the chain has converged. The output of the algorithm can then be taken as a sample drawn from $p(\Omega|Y)$. A detailed description of these methods as well as proofs of the convergence of the algorithm can be found in Johanes and Polson (2009).

The importance of the priors that we chose for the parameters becomes clear once we note that for each of the $\omega_i \in \Omega$ we will have¹:

$$\begin{aligned} \mu | \Omega_{-\mu} &\sim N(\mu_\mu, \sigma_\mu^2), & \rho_J | \Omega_{-\rho_J} &\sim N(\mu_{-\rho_J}, \sigma_{-\rho_J}^2), \\ (\kappa, \kappa\theta) | \Omega_{-(\kappa, \kappa\theta)} &\sim N(M, \Sigma), & \lambda_v | \Omega_{-\lambda_v} &\sim \beta(\alpha_{\lambda_v}, \beta_{\lambda_v}), \\ \sigma_v^2 | \Omega_{-\sigma_v^2} &\sim IG(\alpha_v, \beta_v), & \lambda_v | \Omega_{-\lambda_v} &\sim \beta(\alpha_{\lambda_v}, \beta_{\lambda_v}), \\ \mu_y | \Omega_{-\mu_y} &\sim N(\mu_{\mu_y}, \sigma_{\mu_y}^2), & \xi_t^v | \Omega_{-\xi_t^v} &\sim \mathbf{1}_{\xi_t^v > 0} N(\mu_{\xi_t^v}, \sigma_{\xi_t^v}^2), \\ \sigma_y^2 | \Omega_{-\sigma_y^2} &\sim IG(\alpha_y, \beta_y), & \xi_t^y | \Omega_{-\xi_t^y} &\sim N(\mu_{\xi_t^y}, \sigma_{\xi_t^y}^2), \\ \mu_v | \Omega_{-\mu_v} &\sim \exp(\lambda_{\mu_v}), & J_t | \Omega_{-J_t} &\sim \text{Bernoulli}(p_{J_t}). \end{aligned}$$

Thus we can now see clearly the advantage of the prior distributions that were chosen: they have conjugate posteriors that are easy to draw from. Notice however, that we are missing the posteriors for ρ and V_t .

These elements of Ω unfortunately do not have posteriors that follow a distribution that is known in closed form. Thus, in order to draw from them, accept-reject methods must be used. In the case of ρ we use an independence Metropolis-Hastings algorithm to draw from $\rho | \Omega_{-\rho}$, and in the case of V_t a

¹A detailed derivation of the posteriors and the procedure used to draw from ρ and V_t can be found in the appendix.

random walk Metropolis-Hastings algorithm is used. These Metropolis-Hastings algorithms belong to the family of accept-reject methods used to draw from non-standard distributions. A candidate ω_i^c is drawn from a known distribution and it is accepted or rejected with a certain probability that is a function of the ratio of its likelihood and the likelihood evaluated at ω_i^{c-1} . A disadvantage of these methods is that if each new draw takes steps that are too far from the previous iteration, the algorithm will tend to reject new draws very often and thus convergence will be very slow. On the other hand, if steps are too small, the algorithm will not reject often, but convergence will still be slow, since it will take the program a long time to explore all the potential values where the posterior distribution of ω_i has positive mass. Therefore, when using these Metropolis-Hastings algorithms one must calibrate them, and the rule of thumb suggested in Johanes and Polson (2009) is that 30%-60% of the candidates should be rejected.

Putting all the steps together we have the following:

Algorithm 1 *Estimation of stochastic volatility and jumps model.*

1. *Set starting values for Ω*
2. *For $c = 1 \dots C$*
 - *For $i = 1 \dots \#\Theta$*
 - *Draw parameter i from: $p(\Theta_i | \Theta_{-i}, J, \xi_t^y, \xi_t^v, V, Y)$*
 - *Draw jump sizes:*
 - *For $t = 1 : T$ draw from: $p(\xi_t^y | \Theta, J_t = 1, \xi_t^v, V, Y)$*
 - *For $t = 1 : T$ draw from: $p(\xi_t^v | \Theta, J_t = 1, V, Y)$*
 - *Draw jump times:*
 - *For $t = 1 : T$ draw from: $p(J_t = 1 | \Theta, \xi_t^y, \xi_t^v, V, Y)$*
 - *Draw volatility:*
 - *For $t = 1 : T$ draw from: $p(V_t | \Theta_i, J, \xi_t^y, \xi_t^v, V_t, V_{t-1}, Y)$*
3. *Discard the first \tilde{C} iterations.*

By writing the algorithm in this expanded form, we can clearly appreciate that this method is very computationally intensive. Drawing the parameters is not the issue, since the number of

parameters is less than 11 for all models. However for each iteration of the algorithm at least T latent variables have to be drawn, and for the models with jumps in both equations, up to $5T$ state variables have to be simulated for each iteration. Now, T is frequently in the thousands, and convergence criteria are usually not met unless C is in the tens of thousands. Hence, it becomes clear that computational cost will be one potential drawback of this method. Nonetheless, this procedure remains the best estimating option available, in spite of its computational cost, precisely because we need the draws of the state variables for our option pricing application and for the Bayesian model selection procedure we describe below.

Finally, Johanes and Polson (2009) show that this algorithm will in fact produce a Markov chain whose stationary distribution will be the posterior distribution $p(\Omega|Y)$ that is of interest here. Precisely because we are interested in the stationary distribution, we discard the first \tilde{C} iterations and take the draws of Ω^c from \tilde{C} to C as a sample from the stationary distribution.

Still, we need some criteria to know that the algorithm has achieved convergence. Although there are no set rules on how to guarantee convergence, there are still several things we can check. First, we can use very different starting points and check that the estimators that the algorithm produces are sufficiently close regardless of the starting point. Second, assuming we believe that C iterations are enough, we can run the procedure for a considerably longer time (up to an order of magnitude larger) and compare the distribution of the last C draws of the chain with the C draws where we thought that convergence had occurred.

4 Data

The IPC is the main index of the Mexican Stock Exchange. The exchange describes it as a "representative indicator of the Mexican stock market" (Bolsa Mexicana de Valores (2011)). It includes 35 stocks chosen for their market capitalization and liquidity. Nonetheless, a handful of stocks dominates the index; in 2010-2011, the two most heavily-weighted stocks comprised over 36% of the index. These were America Movil (Telecommunications) having a weight of 25% and Walmart-Mexico (Retail) having a weight of 11.5%. However, having single stocks with very large weights is not unusual for stock indices; a common example being the Nasdaq 100, where Apple Inc. had a 20% weight on the index during 2010. The rest of the IPC is well diversified sectorwise, including firms as diverse as Peñoles (the world's largest silver producer) and FEMSA (a beverage company

producer of beer and soft drinks).

Moreover, a crucial characteristic of the IPC is that it is used as a benchmark by banks and investment firms in Mexico, and it is one of the most closely followed economic indicators in the country. Thus, when the Mexican stock exchange introduced options in 2004, European-style calls and puts on the IPC were some of the first contracts to be issued.

The data used for this paper consists of 6189 daily prices of the index ranging from 1987 to 2011. Although the index existed before, 1987 the methodology used for the construction of the index was very different, so only data from 1987 onward was used. Still, the period considered comprises a wide range of economic conditions. Near the beginning of the series, in October 1987, one of the most volatile periods in world financial history occurred. For the Mexican case this was compounded by internal problems that led to an annual inflation of over 100% for that year. Next in 1994, Mexico experienced one of the most politically turbulent years in recent history, which was the prelude to the 1994-95 financial crisis. Later, we will pay special attention to this period and what some of the models can tell us about the behavior of the index under these extreme circumstances.

Other periods of worldwide financial instability are also reflected on the IPC. These include the Asian and Russian crises of 1997-98, the burst of the .com bubble in the early 2000s and the global financial crisis beginning in 2008. Finally, between these highly volatile times, there are periods of low volatility, and thus the returns exhibit the usual volatility-clustering that is common for stock indices.

4.1 Descriptive Statistics

Let us begin the analysis of the data by presenting simple descriptive statistics. Table 1 shows that mean returns for the index were almost 27% per year with an average annual volatility of 29%, both numbers higher than US stock indices for this period. More importantly, the series exhibits negative skewness and kurtosis that considerably exceeds 3. This is a common finding for financial returns, indicating that log returns are not normally distributed and thus not consistent with the Black-Scholes model.

Table 1: Summary Statistics for Daily Returns

Mean (Annualized)	0.106% (26.82%)
Volatility (Annualized)	1.84% (29.32%)
Skewness	−0.508
Kurtosis	21.13
Min.	−20.24%
Max.	23.58%

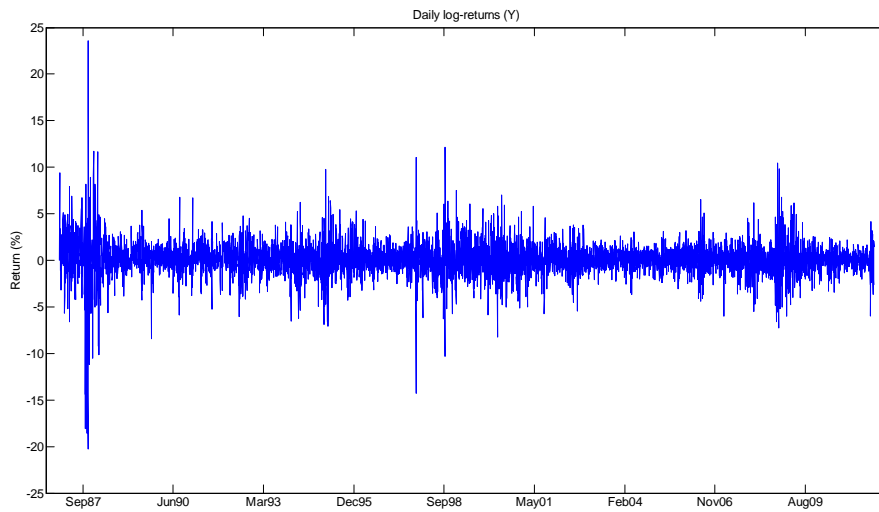


Figure 1: Daily log-returns of the IPC index 1987-2011

Figure 1 shows the time series for the returns. We can see that the data exhibits periods of very high volatility, particularly in 1987, but also in 1994-95, 1997-98 and 2008-09. This presence of volatility-clustering is yet another suggestion that the Black-Scholes model will not be able to describe the dynamics of the index and that stochastic volatility models may be appropriate. However, one can also observe days of extreme returns with a magnitude that exceeds 10 standard deviations above or below the mean return. This is an indication that a model without jumps will not be able to fully capture the dynamics of the index. Further, and especially 1987, we observe periods when there are returns with very large magnitudes during consecutive days, suggesting that jumps in returns might not be sufficient, and a correctly specified model for this data will also have to include jumps in the volatility series.

5 Estimation Results and Model Selection

5.1 Estimation Results

The parameters are estimated by setting $C = 150,000$ and discarding the first 100,000 iterations. Table 2 shows the median of the posterior distributions for each of the four estimated models.

Table 2: Medians of the Posterior Distributions.				
-Standard deviation of the posterior distributions in brackets.				
	SV	SVJ	SVCJ	SVIJ
μ	0.1170(0.017)	0.1174(0.019)	0.1178(0.018)	0.1188(0.018)
σ_v	0.0430(0.001)	0.0417(0.001)	0.0423(0.001)	0.0425(0.001)
ρ	-0.0274(0.015)	-0.02510(0.015)	-0.02894(0.016)	-0.0283(0.018)
κ	0.0444(0.005)	0.0462(0.006)	0.04488(0.005)	0.0453(0.005)
θ	3.051(0.230)	2.782(0.212)	2.958(0.222)	2.962(0.225)
σ_y		0.364 (0.12)	0.446 (0.24)	0.453 (0.27)
μ_y		-0.124 (0.83)	-2.592 (1.00)	-3.420 (3.82)
λ_y		0.017 (0.006)	0.003 (0.001)	0.002 (0.001)
μ_v			4.54 (16.61)	20.81 (9.85)
λ_v				0.007 (0.004)
ρ_J			-0.423 (1.14)	

The parameters that have an immediate interpretation are μ and θ . The estimates for μ indicate a deterministic drift that remains close to 30% annually regardless of the model. Meanwhile the estimations for θ imply a long term annual volatility of 27.72% for the SV model but only 26.48% in the SVJ model. These reductions in volatility are to be expected: under the SV specification all the variation in the return series is driven by the volatility of its diffusive element. However in the SVJ model, part of this variation is now explained by jumps in the return series. This can also be seen in the estimates of V . Figure 2 shows the SV, SVJ and SVIJ estimates for V for two of the periods of particularly high volatility. Once more, we see that the volatility series is lower when part of the variation of the returns can be attributed to jumps. However, it is also important to note that volatility is higher when jumps in the volatility series are considered. The explanation for this is

that under the SVJ specification very large returns will be attributed to jumps in the return series, without volatility having to rise much. However, once we allow for jumps in the volatility series, a large return can be the consequence of a sudden increase in volatility and not necessarily a jump in the return series.

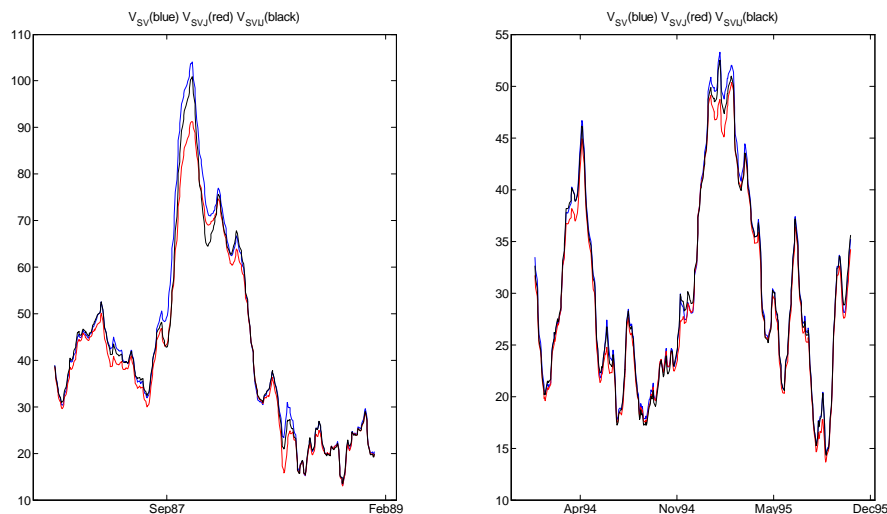


Figure 2: Estimated volatility for two sub-periods under the SV, SVJ and SVIJ specifications

Finally, the estimates of $\lambda = (\lambda_y, \lambda_v)$ also have an immediate interpretation. The frequency of the jumps, the estimations of λ_y imply 4.25 jumps in the return series are expected every year under the SVJ specification, but only 2 jumps every 3 years are expected under the SVCJ and SVIJ specifications. Once again, this result is what we would expect: by introducing jumps in volatility, large returns can be attributed to a sudden increase in V , without the need for jumps in the return series.

5.2 Implied Jump Process

5.2.1 Results for the SVJ Model

In Figure 3 we can see that the days with high probability of a jump are not necessarily the days with the largest returns, but days with large returns and relatively low levels of volatility. Thus, for high volatility periods, the model still assigns an important part of the variation in the returns to the diffusive volatility. In fact, in Figure 4 we can observe that on dates when a high probability of

a jump is assigned, practically all the magnitude of the return can be attributed to the jump.

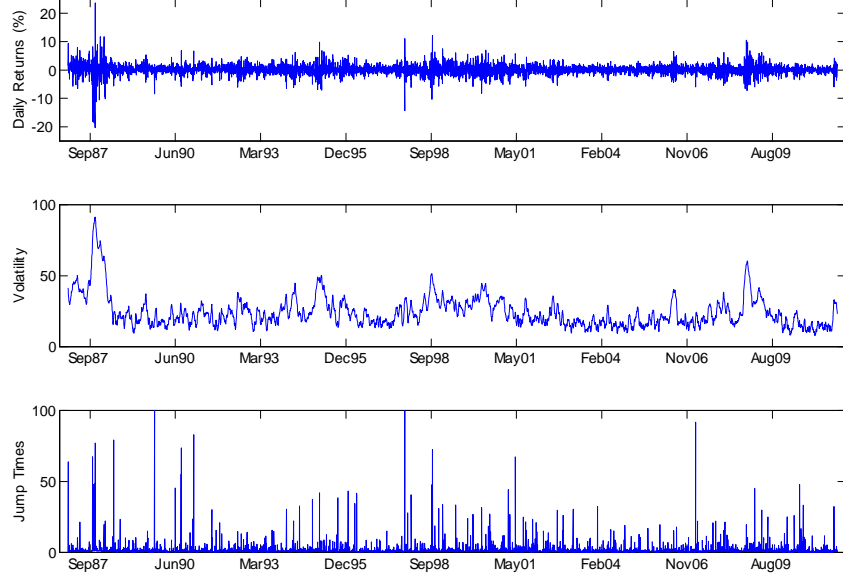


Figure 3: Volatility estimates and probability of jump times under the SVJ specification

Additionally, the fact that the SVJ model very clearly identifies dates with high probabilities of jumps allows us to look at one of the most interesting periods in Mexican economic history and link historic events with predictions of jumps (or lack thereof). In Figure 5 we can see the results of the SVJ model for the 1994-1995 period. 1994 was one of the most politically unstable years in recent Mexican history. On New Year's Day 1994 Zapatista guerillas appeared in the southeastern state of Chiapas. This represented the first bout of civil insurgency in two decades in Mexico. Not surprisingly, our estimates assign a large probability of a jump having occurred on January 2nd (the first day of trading that year). The next date with a large probability of a jump is immediately after March 23rd. On this date the presidential candidate of the official party was assassinated. Given the political structure at the time, the candidate of the official party was practically assured to be the next president. Therefore, his assassination represented a particularly important shock to the confidence in Mexican financial markets. The next instances of high jump probabilities correspond to June 11th when the Zapatista guerrillas rejected the terms of a proposed peace treaty and later on November 1994. The latter might be thought of as the beginning of the chain of events that triggered

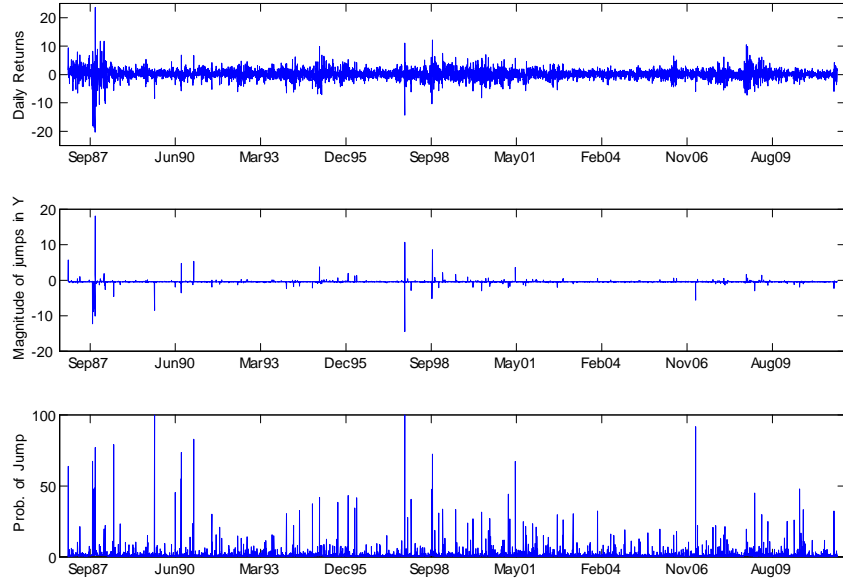


Figure 4: Magnitude and probability of jumps under the SVJ specification.

the 1995 financial crisis (see e.g. Aspe (1995)). On November 18th the central bank experienced one of the largest declines in foreign reserves on record. Confidence had been deteriorating throughout the year, but on November 15th a high ranking government official accused the official party of orchestrating a political assassination earlier that year. Given the political structure of Mexico during the 1990s, instability in the ruling party was perceived to be equivalent to instability in the government. Further, historically, the party had been characterized by ferocious loyalty of party members (see e.g. Preston and Dillon (2004)), so public accusations of serious misconduct coming from a high ranking party member were interpreted as a sign of deep instability within the party. Given the other problems discussed above, and that rumors of a potential sovereign default were beginning to circulate, it is no surprise that the model identifies a jump on November 18th.

Finally, the last date marked explicitly in Figure 5 is December 19, 1994. On this date the Mexican president announced rather abruptly that the peso would be devalued shortly. This episode has become known colloquially as the "error de diciembre" or December's mistake. It is somewhat surprising that the model does not detect a jump on this date or the dates immediately afterward. Nonetheless, if we look at the return series we see that there were in fact very large returns around that date. However, after inspecting the volatility estimates, we see that volatility had been increas-

ing from November onward. Thus, by the time the announcement was made, volatility was so high that jumps are not needed to explain the very large returns observed in late December and early 1995.

During early 1995, while the crisis was unfolding, we observe that volatility tended to be high and thus no jumps are needed to explain large swings in prices. Nonetheless, volatility decreased through the first half of the year and once it reached its lowest point of the year in August, a jump is once more detected. At this point volatility begins to increase once more, explaining the large returns during late 1995, and thus no further jumps are detected until 1996, which coincides with the end of the sub-period of interest.

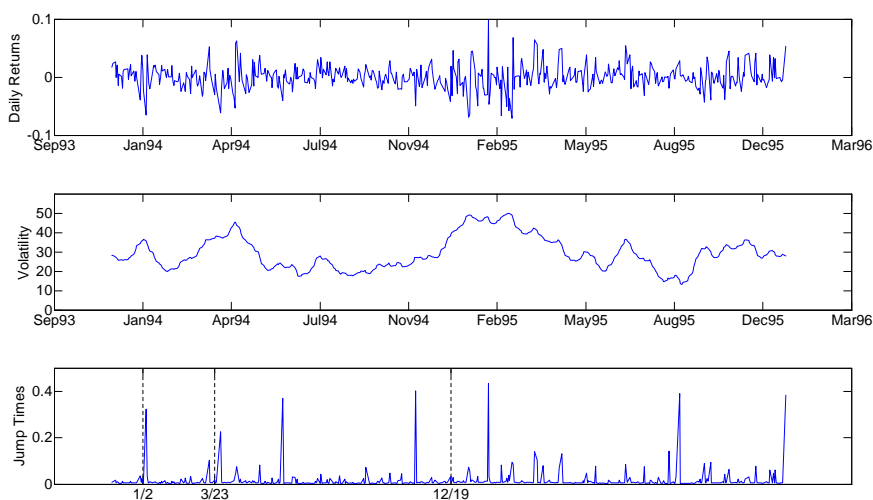


Figure 5: Volatility estimates and probability of jump times under the SVJ specification for the 1994-95 period.

5.2.2 Results for Models with Jumps in Both Returns and Volatility

In Figure 6 we can observe graphically how under the SVCJ a much lower frequency of jumps is estimated (when compared to the result under the SVJ model). However, we still find that jumps are particularly important in explaining large returns during periods of low volatility. Also, it is important to note that, as we would expect, periods of high volatility tend to be preceded by large jumps in the volatility series.

In Figure 7 we can see that under the SVIJ specification the probability of jumps in the return

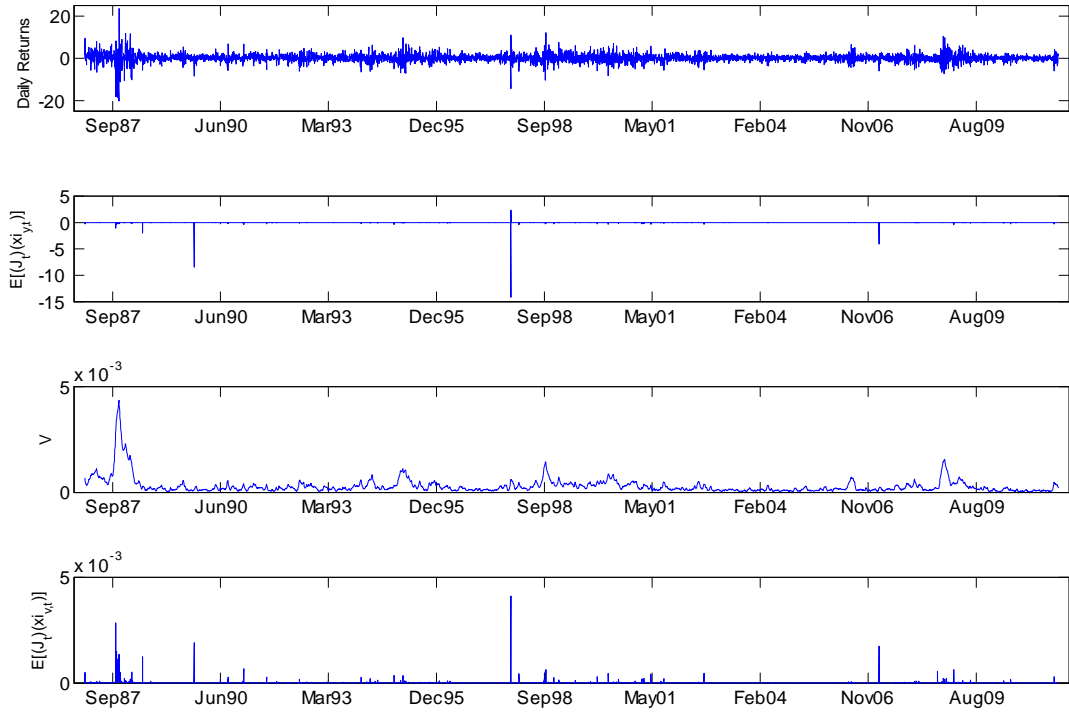


Figure 6: Returns, volatility and expected magnitude of jumps in both series under the SVCJ specification.

series remains very high for very specific days. In contrast, the probability of jumps remains almost uniformly low for the volatility series. Nonetheless, this does not imply that there is little information in this series of probabilities. Since what is plotted in the last graph of Figure 7 is the probability of a jump occurring on each particular day, having a period of time with above average probabilities of jumps for each of those days tells us that there is a high probability of a jump having occurred in that period. So, instead of identifying precise days when a jump occurred (as we can with the return series) for the volatility series we can only identify an interval of days where it seems very likely that there was a jump. Further, Figure 8 shows the product of the probability of a jump with its expected magnitude. This gives us a sense of the expected jump on that day. Therefore, by looking at a given period we can get a measurement of the expected accumulated contributions of jumps to the volatility.

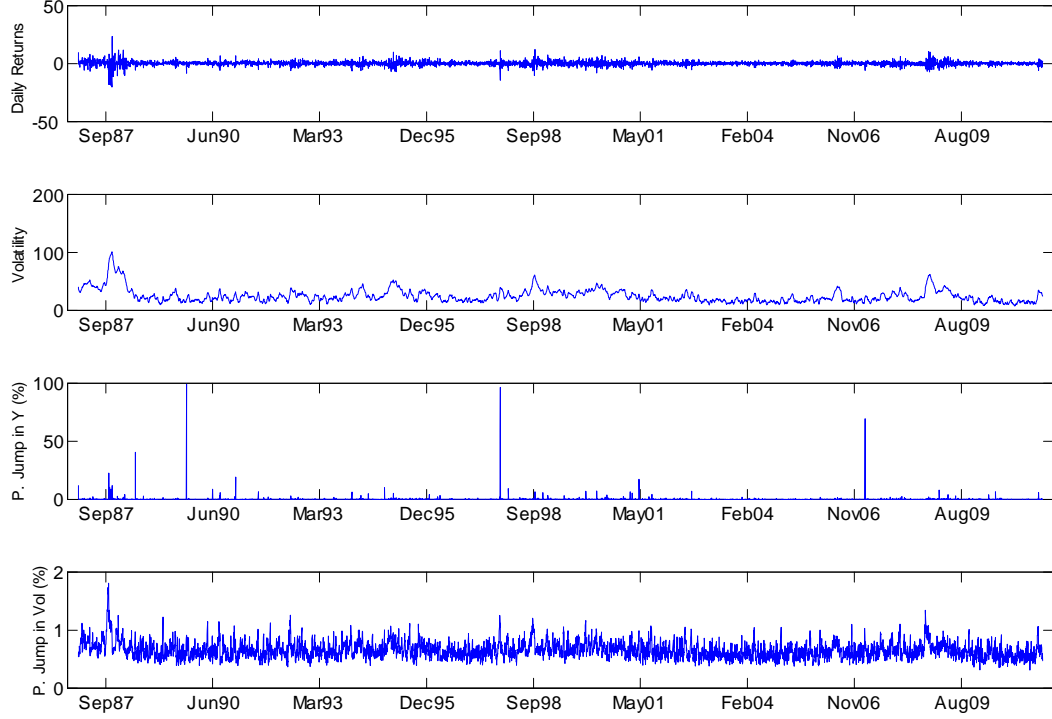


Figure 7: Returns, volatility, probability of jumps in returns and probability of jumps in volatility under the SVIJ specification.

5.3 Model Selection - Test of Residuals

Through the MCMC procedure, we completely characterize the posterior distributions of the parameters. However, knowing the probability of the parameters lying in a certain interval will not be sufficient to test the suitability of the models or their relative performance. Thus, in order to test how well specified the models are, the following residuals are calculated:

$$\hat{\varepsilon}_t = \frac{Y_t - Y_{t-1} - \hat{\mu} - \hat{J}_t^y \hat{\xi}_t^y}{\sqrt{\hat{V}_{t-1}}}.$$

Note that, regardless of the model being considered, under the hypothesis that the model describes the dynamics of the returns, this residual should follow a standard normal distribution. Thus, a first approach to testing the fit of the different models is to construct a QQ plot for the residuals.

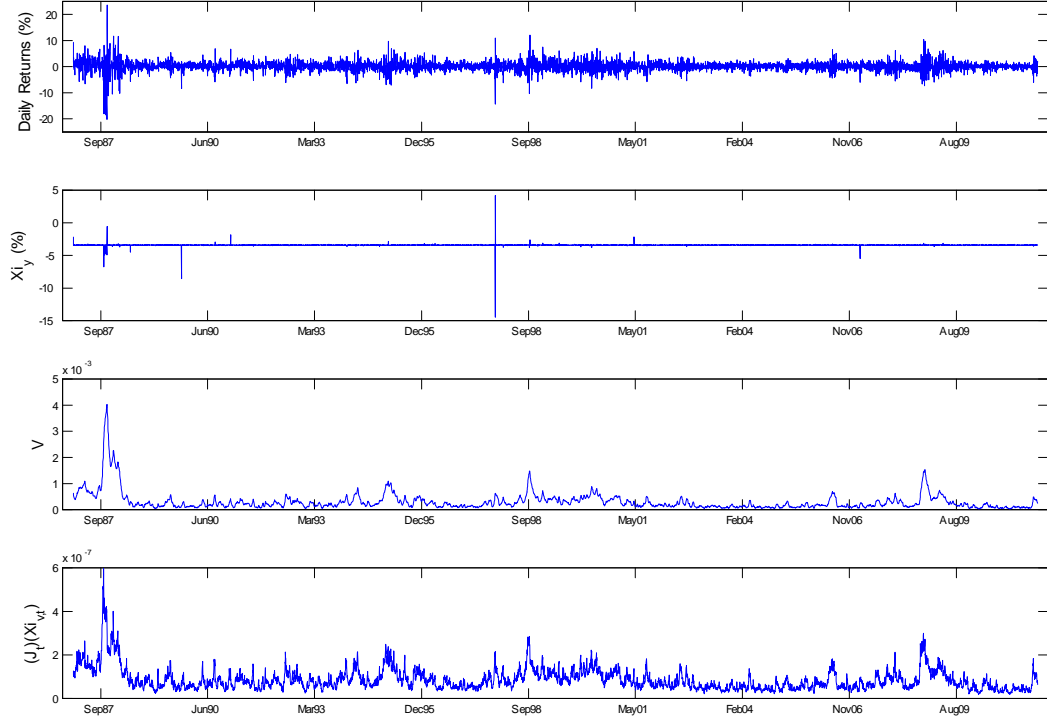


Figure 8: Returns, magnitude of jumps in returns, volatility and expected size of jumps in volatility under the SVIJ specification.

The QQ plots in Figure 9 show us how the distribution of the residuals is closer to normality for the SVJ than for the SV model and even closer for the models with jumps in both series. Moreover, beyond the graphical data represented by the QQ plots, we can perform a formal statistical test of the normality of $\hat{\varepsilon}_t$ for each of the models. Table 3 shows the p -values for a Jarque-Bera normality test of the residuals.

Table 3: Jarque-Bera Test

Model	SV	SVJ	SVCJ	SVIJ
p -value	<0.001	0.0199	0.0763	0.0680

For the SV case the normality of the residuals can be rejected at the 99.9% level, confirming that in fact this model provides a very poor fit for the data. At the 95% level we can also reject the

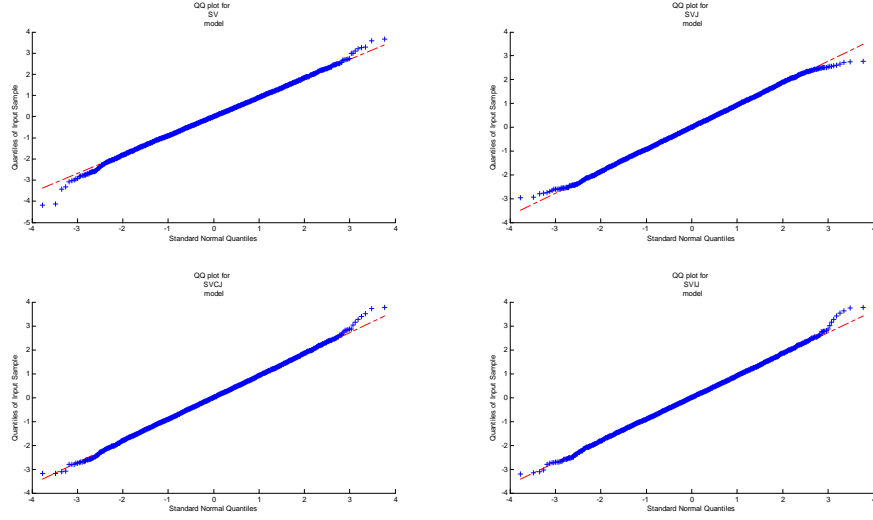


Figure 9: Figure 9: QQ plots for the residuals under the four models.

SVJ model, but not of the models with jumps in volatility. Thus models with jumps in both series represent a considerable improvement relative to the SV model. This confirms the initial conjecture: that models with jumps are better specified for the IPC than Heston's model.

5.4 Model Selection - Bayes Factors

So far we have seen that the models with jumps offer a better fit than the standard Heston (SV) model. However, when assessing the merits of the different models, we still must take into account the fact that the SV model is the most tightly parametrized within the set. This should lead us to ask whether the improvements in fit outweigh the loss parsimony.

Fortunately, our estimation procedure produces estimates for all the latent variables. This, together with the fact that the SV model can be expressed as a restricted version of any of the other models, allows us to compare the performance of the models using Bayes' factors.

In general, for two models M_1 and M_2 and data Y , the posterior odds:

$$B(M_1, M_2) = \frac{p(M_1|Y)}{p(M_2|Y)}$$

represent the relative likelihood of the two models conditional on the data. Further, if we assume

prior ignorance by setting $p(M_1) = p(M_2)$, a simple application of Bayes' theorem yields:

$$B(M_1, M_2) = \frac{p(Y|M_1) p(M_1)}{p(Y|M_2) p(M_2)} = \frac{p(Y|M_1)}{p(Y|M_2)},$$

which is simply the ratio of the likelihoods of the data under the two model specifications. So, assuming we are comparing the SV model to any of the alternatives (SVX), $B(SV, SVX)$ can be expressed as:

$$\frac{p(Y|SV)}{p(Y|SVX)} = \frac{\int p(Y|\Omega, SV) p(\Omega|SV) d\Omega}{\int p(Y|\Omega, SVX) p(\Omega|SVX) d\Omega}.$$

Further, SV is equivalent to restricting the jump vector (or vectors) to zero in any of the other models, so that $p(Y|\Omega, SV) = p(Y|\Omega, SVX, J = 0)$. Thus, a further application of Bayes' formula tells us that:

$$\frac{p(J = 0|Y, SVX)}{p(J = 0|SVX)} = \frac{p(Y|J = 0, SVX)}{p(Y|SVX)} = \frac{p(Y|SV)}{p(Y|SVX)},$$

where the denominator can be computed directly from the prior distribution of λ (the rate of the jump process) and the numerator will be:

$$P(J = 0|Y, SVJ) = \int_0^1 p(J_v = 0|\lambda, Y, SVX) p(\lambda|Y, SVX) d\lambda. \quad (5)$$

Now, since the Markov chain constructed for estimation completely characterizes $p(\lambda|Y, SVX)$, we can use the draws from this chain to perform a Monte Carlo integration of equation (5).

Following Smith and Spiegelhalter (1980), in table 2 we report $2 \ln(B)$ and the interpretation suggested in that same reference.

Table 4: Bayesian Factors		
Models	$2 \ln(B)$	Interpretation
SV vs. SVJ	-7.5	Positive against SVJ
SV vs. SVCJ	9.1794	Positive for SVCJ
SV vs. SVIJ	15.164	Positive for SVIJ
SVCJ vs. SVIJ	5.9846	Positive for SVIJ

The negative value in SV vs. SVJ indicates that under this criterion the improvements in fit from SV to SVJ do not compensate for the less parsimonious model. However, the positive values

for the rest of the factors indicate that there is otherwise positive evidence in favor of the more complex models.

6 Option Pricing Application

The starting point of this paper was the fact that MexDer uses Heston's model to price options on the days when there has been no trading. Now, under the naming convention used in the previous sections, Heston's model corresponds to the SV model, and so far I have shown that it is a poor choice for modeling the dynamics of the IPC and that models with jumps offer a much better description of the asset dynamics. However, we still have to check whether the option prices implicit in the more elaborate models will significantly differ from those estimated using Heston's model.

Let us remember that one of the characteristics which make Heston's model is so attractive is that it has a closed form formula for the valuation of European-style options. In contrast, with more elaborate models we cannot always be sure that such valuation formulas exist. Fortunately, the four models estimated and tested above belong to the larger class of affine-jump diffusions. Duffie et al. (2000) show that the pricing formula for European options on assets that follow an affine-jump diffusion has a closed form (up to the solution of a system of complex-valued differential equations). Concretely, for an option expiration date T , at time t the conditional characteristic function is defined as:

$$CCF_t(z) = E \left[e^{izY_T} | Y_t \right].$$

Using this definition, Duffie et al. (2000) show that if Y follows an affine process then $CCF_t(z)$ will be of the form:

$$CCF_t(z) = e^{\phi_{0t} + \phi_{Yt} Y_t}$$

where ϕ_{0t} and ϕ_{Yt} satisfy a pair of complex-valued ordinary differential equations. Since the CCF_t is fundamental for option pricing, in principle every time we wanted to price an option, one would have to solve (ϕ_{0t}, ϕ_{Yt}) by some numerical method (e.g. as Runge-Kutta). However, for the models considered in this paper ϕ_{0t} and ϕ_{Yt} are known explicitly as a function of the parameters of the jump diffusions and the parameters of the options (see e.g. Sepp (2003)).

Further, Sepp (2003) also shows that the particular models estimated above yield a pricing formula that can be expressed as a weighted sum of the current price of the underlying asset and

the present value of the strike of the option. Specifically, the price (F) of a European option over an asset with current price S_t , and strike price K is given by:

$$F = \varphi \left[S_t P_1(\varphi) - e^{-r(T-t)} K P_2(\varphi) \right],$$

where $\varphi = 1$ if the option is a call and $\varphi = -1$ if the option is a put and:

$$P_j(\varphi) = \frac{1-\varphi}{2} + \varphi \left(\frac{1}{2} + \frac{1}{\pi} \int_0^\infty \mathcal{R} \left[\frac{\phi_j(e^{iky})}{ik} \right] dk \right), \quad (6)$$

where \mathcal{R} denotes the real part of a complex number and (ϕ_1, ϕ_2) are functions of the CCF_t and parameters of the option being evaluated. When compared to equation (1) we see why Sepp (2003) calls this a Black-Scholes-style formula.

Using this pricing procedure and the parameters estimated above, prices are calculated for every call option issued between 2004 and 2009. Then, the prices implied by Heston's model are compared to those implicit in the SVJ model.

Figures 10 and 11 show the percentage difference in price of call options under the SVJ specification relative to the prices under Heston's model. The graphs show the difference in price as it varies with the ratio of the strike price of the option to the current price (a measure of the "moneyness" of the options). In each case the set of options is separated into those that will expire in under 45 days, those that will expire in between 150 and 250 days, and those that will expire in more than 300 days. Figure 10 shows that for options that are out of the money, the difference in prices is moderate, although it increases as the expiration of the options becomes longer, and it also grows as the options come close to being in the money (as K/S_t approaches 1).

These differences in prices would be significant if one could actually trade in these options, but since the purpose of the prices is to mark to market them, the difference in prices is not that important.

However, in Figure 11 we see that for options that are in the money, the price differences are much larger and can substantially change the apparent value of the options that banks hold in their balance sheets. This clearly makes the choice of the model for the dynamics of the underlying assets a very important decision when imputing prices to options with no market price.

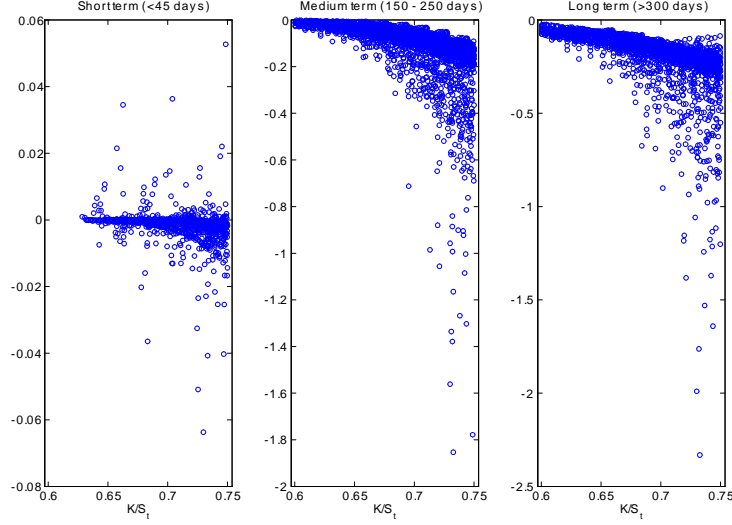


Figure 10: Percentage difference in price implied by the SVJ model relative to those implied by Heston's model for out of the money options.

7 Conclusions and Direction of Future Research

I have shown that Heston's model is misspecified for the IPC and that models which incorporate jumps offer a better description of the dynamics of the index. Further, I showed that models with jumps in both returns and volatility improve the fit sufficiently to compensate for the loss in sparseness. Finally, transform pricing techniques allowed me to show that option prices will be significantly different under the assumption of jump-diffusions than under Heston's model.

The remaining issue is the suitability of the method in a real-life situation. The estimation procedure is very intensive computationally and may not be practical if the model needs to be estimated every day. Thus, implementing these results would require an estimation method that could incorporate new information, such as new prices at the end of each day, in a computationally-efficient manner. I conjecture that by combining the MCMC procedure presented here with Bayesian filtration methods such as those proposed in Johannes et al. (2009), a methodology for pricing the options can be constructed that is both based on a well-specified model for the asset dynamics and implementable in a real setting. This would then enable MexDer to provide more reliable prices on a daily basis.

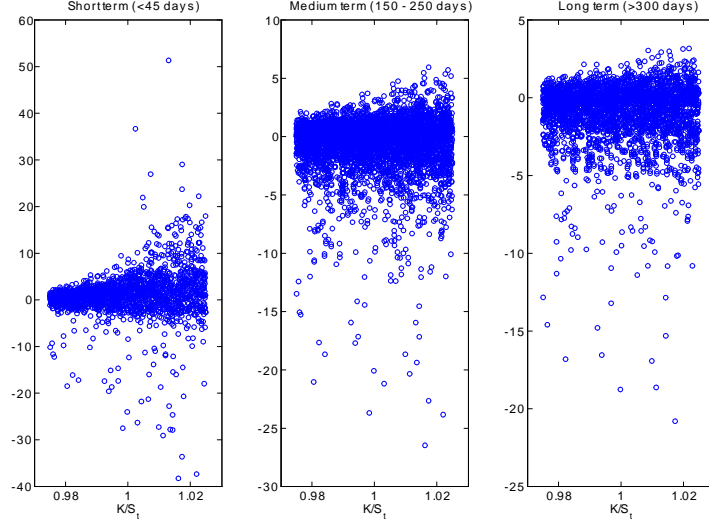


Figure 11: Percentage difference in price implied by the SVJ model relative to those implied by Heston's model for in the money options.

Appendix. Drawing from the Conditional Posteriors

Many of the results presented in this appendix can be found in standard textbooks, in Eraker et al. (2003) or in Johannes and Polson (2009). However, to ease replicability of my results, I present the derivation of all the posteriors.

We begin by re-stating the discretized model (4) as:

$$\begin{aligned}\Delta Y_t &= \mu + \sqrt{V_{t-1}}\varepsilon_t^y + \xi_t^y J_t^y \\ \Delta V_t &= \alpha + \beta V_{t-1} + \sigma_v \sqrt{V_{t-1}}\varepsilon_t^v + \xi_t^y J_t^y\end{aligned}$$

where we have, $\beta = -\kappa$ and $\alpha = \kappa\theta$, $\Delta Y_t = Y_t - Y_{t-1}$ and $\Delta V_t = V_t - V_{t-1}$. Then a distribution that will be useful throughout is:

$$p(\Delta Y_t, \Delta V_t | V_{t-1}, \Theta) = \frac{1}{2\pi\sigma_v V_{t-1} \sqrt{1-\rho^2}} \exp \left(-\frac{1}{2(1-\rho^2)} \left[\begin{aligned} &\frac{(\Delta Y_t - \mu - \xi_t^y J_t^y)^2}{V_{t-1}} \\ &+ \frac{(\Delta V_t - \alpha - \beta V_{t-1} - \xi_t^y J_t^y)^2}{\sigma_v^2 V_{t-1}} \\ &- \frac{2\rho((\Delta Y_t - \mu - \xi_t^y J_t^y)(\Delta V_t - \alpha - \beta V_{t-1} - \xi_t^y J_t^y))}{\sigma_v V_{t-1}} \end{aligned} \right] \right) \quad (7)$$

Drawing from μ

First, we have the prior for μ is:

$$p(\mu) = \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left(-\frac{(\mu - \mu_\mu)^2}{2\sigma_\mu^2}\right)$$

and applying Bayes' rule:

$$\begin{aligned} p(\mu|Y, \Omega_{-\mu}) &\propto p(Y|\mu, \Omega_{-\mu}) p(\mu) \\ &\propto \exp\left(-\frac{(\mu - \mu_\mu)^2}{\sigma_\mu^2}\right) \prod_{t=1}^T \exp\left(-\frac{(\Delta Y_t - \mu - \xi_t^y J_t^y)^2}{2V_{t-1}}\right) \\ &\propto \exp\left(-\frac{\mu_\mu^2 - 2\mu\mu_\mu}{2\sigma_\mu^2}\right) \exp\left(-\frac{1}{2}\left(\mu^2 \sum_{t=1}^T \frac{1}{V_{t-1}} - 2\mu \sum_{t=1}^T \frac{(\Delta Y_t - \xi_t^y J_t^y)}{V_{t-1}}\right)\right) \\ &\propto \exp\left(-\frac{\mu^2 - 2\mu\mu_\mu}{2\sigma_\mu^2}\right) \exp\left(-\frac{1}{2\left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1}}\left(\mu^2 - 2\mu \frac{\sum_{t=1}^T \frac{(\Delta Y_t - \xi_t^y J_t^y)}{V_{t-1}}}{\sum_{t=1}^T \frac{1}{V_{t-1}}}\right)\right) \\ &\propto \exp\left(-\frac{1}{2}\left(\mu^2 \left(\frac{1}{\sigma_\mu^2} + \frac{1}{\left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1}}\right) - 2\mu \left(\frac{\mu_\mu}{\sigma_\mu^2} + \sum_{t=1}^T \frac{(\Delta Y_t - \xi_t^y J_t^y)}{V_{t-1}}\right)\right)\right) \\ &\propto \exp\left(-\frac{1}{2\left(\frac{\sigma_\mu^2 \left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1}}{\left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1} + \sigma_\mu^2}\right)}\left(\mu^2 - 2\mu \frac{\left(\frac{\mu_\mu}{\sigma_\mu^2} + \sum_{t=1}^T \frac{(\Delta Y_t - \xi_t^y J_t^y)}{V_{t-1}}\right)}{\left(\frac{\left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1} + \sigma_\mu^2}{\sigma_\mu^2 \left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1}}\right)}\right)\right) \end{aligned}$$

and thus

$$\mu|Y, \Omega_{-\mu} \sim N\left(\frac{\left(\frac{\mu_\mu}{\sigma_\mu^2} + \sum_{t=1}^T \frac{(\Delta Y_t - \xi_t^y J_t^y)}{V_{t-1}}\right)}{\left(\frac{\left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1} + \sigma_\mu^2}{\sigma_\mu^2 \left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1}}\right)}, \frac{\sigma_\mu^2 \left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1}}{\left(\sum_{t=1}^T \frac{1}{V_{t-1}}\right)^{-1} + \sigma_\mu^2}\right)$$

Drawing from α, β and σ_v

since

$$\begin{aligned} \Delta V_t &= \alpha + \beta V_{t-1} + \sigma_v \sqrt{V_{t-1}} \varepsilon_t^v + \xi_t^y J_t^y \\ \Delta V_t|V_{t-1}, \Omega &\sim N(\alpha + \beta V_{t-1} + \xi_t^y J_t^y, \sigma_v^2 V_{t-1}) \end{aligned}$$

First, for the case of σ_v^2 we have:

$$\sigma_v^2 \sim IG(\check{\alpha}, \check{\beta}).$$

Which implies:

$$p(\sigma_v^2) \propto (\sigma_v^2)^{-\check{\alpha}-1} \exp\left(-\frac{\check{\beta}}{\sigma_v^2}\right)$$

so

$$\begin{aligned} p(\sigma_v^2 | Y, \Omega_{-\sigma_v^2}) &\propto p(V | \sigma_v^2, \Omega_{-\sigma_v^2, V}) p(\sigma_v^2) \\ &\propto (\sigma_v^2)^{-\check{\alpha}-1} \exp\left(-\frac{\check{\beta}}{\sigma_v^2}\right) \prod_{t=1}^T \frac{1}{\sigma_v} \exp\left(-\frac{(\alpha + \beta V_{t-1} + \xi_t^y J_t^y - \Delta V_t)^2}{2\sigma_v^2 V_{t-1}}\right) \\ &\propto (\sigma_v^2)^{-\check{\alpha}-1-T/2} \exp\left(-\frac{\check{\beta} + \sum_{t=1}^T \frac{(\alpha + \beta V_{t-1} + \xi_t^y J_t^y - \Delta V_t)^2}{2V_{t-1}}}{\sigma_v^2}\right) \end{aligned}$$

thus:

$$\sigma_v^2 | Y, \Omega_{-\sigma_v^2} \sim IG\left(\check{\alpha} + T/2, \check{\beta} + \sum \frac{(\alpha + \beta V_{t-1} + \xi_t^y J_t^y - \Delta V_t)^2}{2V_{t-1}}\right).$$

For the case of α and β the priors are:

$$\begin{aligned} \alpha &\sim N(\mu_a, \sigma_a^2), \\ \beta &\sim N(\mu_\beta, \sigma_\beta^2). \end{aligned}$$

So we have:

$$p(\alpha, \beta | \Omega_{-\alpha, \beta}) \propto p(V | \alpha, \beta, \Omega_{-\alpha, \beta}) p(\alpha, \beta),$$

implies:

$$p(\alpha, \beta | \Omega_{-\alpha, \beta}) \propto \exp\left(-\frac{1}{2} \left(\frac{(\alpha - \mu_a)^2}{\sigma_a^2} + \frac{(\beta - \mu_\beta)^2}{\sigma_\beta^2} \right)\right) \prod_{t=1}^T \exp\left(-\frac{1}{2} \frac{(\alpha + \beta V_{t-1} + \xi_t^y J_t^y - \Delta V_t)^2}{\sigma_v^2 V_{t-1}}\right)$$

so:

$$p(\alpha, \beta | \Omega_{-\alpha, \beta}) \propto \exp\left(-\frac{1}{2} \left(\begin{aligned} &\alpha^2 \left(\frac{1}{\sigma_a^2} + \sum_{t=1}^T \frac{1}{\sigma_v^2 V_{t-1}} \right) + \beta^2 \left(\frac{1}{\sigma_\beta^2} + \frac{1}{\sigma_v^2} \sum_{t=1}^T V_{t-1} \right) + 2 \frac{\alpha \beta T}{\sigma_v^2} \\ &- 2\alpha \left(\frac{\mu_a}{\sigma_a^2} - \frac{1}{\sigma_v^2} \sum_{t=1}^T \frac{(\xi_t^y J_t^y - \Delta V_t)}{V_{t-1}} \right) \\ &- 2\beta \left(\frac{\mu_\beta}{\sigma_\beta^2} - \frac{1}{\sigma_v^2} \sum_{t=1}^T \frac{2\beta(\xi_t^y J_t^y - \Delta V_t)}{V_{t-1}} \right) \end{aligned} \right)\right). \quad (8)$$

Now, $(\alpha, \beta|Y, \Omega_{-\alpha, \beta})$ are jointly normal if and only if:

$$p(\alpha, \beta|Y, \Omega_{-\alpha, \beta}) \propto \exp(\mathcal{E}),$$

where:

$$\mathcal{E} = -\frac{1}{2(1-\rho_{a, \beta}^2)} \left[\frac{(\alpha - \mu_\alpha)^2}{\sigma_\alpha^2} + \frac{(\beta - \mu_\beta)^2}{\sigma_\beta^2} - \frac{2\rho_{a, \beta}(\alpha - \mu_\alpha)(\beta - \mu_\beta)}{\sigma_\alpha \sigma_\beta} \right]$$

and if:

$$\tilde{\mathcal{E}} = -\frac{1}{2(1-\rho_{a, \beta}^2)} \left[\begin{aligned} & \frac{\alpha^2}{\sigma_\alpha^2} + \frac{\beta^2}{\sigma_\beta^2} - 2(\alpha\beta) \frac{\rho_{a, \beta}}{\sigma_\alpha \sigma_\beta} \\ & + 2\beta \left(\frac{\rho_{a, \beta} \mu_\alpha}{\sigma_\alpha \sigma_\beta} - \frac{\mu_\beta}{\sigma_\beta^2} \right) + 2\alpha \left(\frac{\rho_{a, \beta} \mu_\beta}{\sigma_\alpha \sigma_\beta} - \frac{\mu_\alpha}{\sigma_\alpha^2} \right) \end{aligned} \right],$$

we will still have that $p(\alpha, \beta|Y, \Omega_{-\alpha, \beta}) \propto \exp(\tilde{\mathcal{E}})$ since $\exp(\tilde{\mathcal{E}}) = K \exp(\mathcal{E})$ for some K that does not depend on α or β . So if we find:

$$p(\alpha, \beta|Y, \Omega_{-\alpha, \beta}) \propto \exp(\tilde{\mathcal{E}})$$

with:

$$\tilde{\mathcal{E}} = \alpha^2 A + \beta^2 B - 2(\alpha\beta) C + 2\beta D + 2\alpha E \quad (9)$$

all we need to do is solve the following system of equations to find the parameters of $p(\alpha, \beta|Y, \Omega_{-\alpha, \beta})$:

$$\begin{aligned} A &= \frac{1}{\sigma_\alpha^2(1-\rho_{a, \beta}^2)}, \\ B &= \frac{1}{\sigma_\beta^2(1-\rho_{a, \beta}^2)}, \\ C &= \frac{\rho_{a, \beta}}{\sigma_\alpha \sigma_\beta(1-\rho_{a, \beta}^2)}, \\ D &= \left(\frac{\rho_{a, \beta} \mu_\alpha}{\sigma_\alpha \sigma_\beta(1-\rho_{a, \beta}^2)} - \frac{\mu_\beta}{\sigma_\beta^2(1-\rho_{a, \beta}^2)} \right), \\ E &= \left(\frac{\rho_{a, \beta} \mu_\alpha}{\sigma_\alpha \sigma_\beta(1-\rho_{a, \beta}^2)} - \frac{\mu_\beta}{\sigma_\beta^2(1-\rho_{a, \beta}^2)} \right). \end{aligned} \quad (10)$$

Solving system (10) yields:

$$\begin{aligned}
\mu_\alpha &= \frac{BE+CD}{C^2-AB}, \\
\sigma_\alpha^2 &= \frac{B}{AB-C^2}, \\
\mu_\beta &= \frac{AD+CE}{C^2-AB}, \\
\sigma_\beta^2 &= \frac{A}{AB-C^2}, \\
\rho_{\alpha,\beta} &= \frac{C}{\sqrt{AB}}.
\end{aligned} \tag{11}$$

Since expression (8) is of the form described in equation (9), we can conclude that $(\alpha, \beta|Y, \Omega_{-\alpha,\beta})$ are distributed as a joint normal, and the parameters are given by matching the values of $A - E$ in equation (8) and then using result (11).

Drawing from μ_y , σ_y , μ_v and ρ_J

The priors of μ_y and σ_y imply that:

$$\begin{aligned}
p(\mu_y) &\propto \exp\left(-\frac{\mu_y^2}{2S}\right) \\
p(\sigma_y^2) &\propto \left(\frac{1}{\sigma_y^2}\right)^{\alpha_p+1} \exp\left(-\frac{\beta_p}{\sigma_y^2}\right).
\end{aligned}$$

Now, let $\mathcal{J} = \{t : J_t = 1\}$. Thus the set \mathcal{J} will contain the values of t where a jump has occurred and we denote $\#\mathcal{J}$ the number of elements in this set, i.e. the number of days when a jump occurred. Then, we restrict $\xi^y = \{\xi_t^y : t \in \mathcal{J}\}$ i.e. the magnitude of jumps for days when in fact there was a jump. For ease of notation, let us consider first the case where there are no jumps in the volatility series. Then:

$$p(\xi^y | \mu_y, \sigma_y^2) \propto \exp\left(-\frac{\sum_{t \in \mathcal{J}} (\xi_t^y - \mu_y)^2}{2\sigma_y^2}\right).$$

So we have:

$$p(\mu_y | \xi^y, \sigma_y^2) \propto p(\xi^y | \mu_y, \sigma_y^2) p(\mu_y),$$

which implies:

$$\begin{aligned}
p(\mu_y | \xi^y, \sigma_y^2) &\propto \exp\left(-\frac{\sum_{t \in \mathcal{J}} (\xi_t^y - \mu_y)^2}{2\sigma_y^2} - \frac{\mu_y^2}{2S}\right) \\
&\propto \exp\left(-\frac{\sum_{t \in \mathcal{J}} \mu_y^2 - \mu_y 2 \sum_{t \in \mathcal{J}} \xi_t^y}{2\sigma_y^2} - \frac{\mu_y^2}{2S}\right) \\
&\propto \exp\left(-\frac{1}{2} \left[\frac{S(\#\mathcal{J})\mu_y^2 - 2\mu_y S \sum_{t \in \mathcal{J}} \xi_t^y + \mu_y^2 \sigma_y^2}{\sigma_y^2 S} \right]\right) \\
&\propto \exp\left(-\frac{1}{\frac{2\sigma_y^2 S}{(\#\mathcal{J} \cdot S + \sigma_y^2)}} \left[\mu_y^2 - 2\mu_y \left(\frac{S \sum_{t \in \mathcal{J}} \xi_t^y}{(\#\mathcal{J} \cdot S + \sigma_y^2)} \right) \right]\right).
\end{aligned}$$

For σ_y^2 we have:

$$\begin{aligned}
p(\sigma_y^2 | \xi^y, \mu_y) &\propto p(\xi^y | \mu_y, \sigma_y^2) p(\sigma_y^2) \\
&\propto \left(\frac{1}{\sigma_y^2}\right)^{\alpha_p + 1} \exp\left(-\frac{\beta_p}{\sigma_y^2}\right) \left(\prod_{t \in \mathcal{J}} \frac{1}{\sqrt{\sigma_y^2}} \exp\left(-\frac{1}{2\sigma_y^2} (\xi_t^y - \mu_y)^2\right)\right) \\
&\propto \left(\frac{1}{\sigma_y^2}\right)^{(\alpha_p + \frac{\#\mathcal{J}}{2}) + 1} \exp\left(-\frac{1}{\sigma_y^2} \left(\frac{1}{2} \sum_{t \in \mathcal{J}} (\xi_t^y - \mu_y)^2 + \beta_p\right)\right).
\end{aligned}$$

Now, for the models with jumps in volatility, we simply substitute ξ_t^y by $\tilde{\xi}_t^y = \xi_t^y - \rho_J \xi_t^v$.

For μ_v , the prior implies:

$$p(\mu_v) \propto \left(\frac{1}{\mu_v}\right)^{\alpha_{mv} - 1} \exp\left(-\frac{\left(\frac{1}{\mu_v}\right)}{\beta_{\mu v}}\right),$$

and we also have:

$$\begin{aligned}
p(\xi_t^v | \mu_v) &\propto \frac{1}{\mu_v} \exp\left(-\frac{1}{\mu_v} \xi_t^v\right) \\
p(\xi^v | \mu_v) &\propto \left(\frac{1}{\mu_v}\right)^{\#\mathcal{J}^v} \exp\left(-\frac{1}{\mu_v} \sum_{t \in \mathcal{J}^v} \xi_t^v\right),
\end{aligned}$$

so:

$$\begin{aligned}
p(\mu_v | \xi_t^v) &\propto \left(\frac{1}{\mu_v}\right)^{\#\mathcal{J}^v + \alpha_{mv} - 1} \exp\left(-\frac{1}{\mu_v} \left(\frac{1}{\beta_{\mu v}}\right) - \frac{1}{\mu_v} \sum_{t \in \mathcal{J}^v} \xi_t^v\right) \\
&\propto \left(\frac{1}{\mu_v}\right)^{\#\mathcal{J}^v + \alpha_{mv} - 1} \exp\left(-\frac{1}{\mu_v} \left(\frac{1}{\beta_{\mu v}} + \sum_{t \in \mathcal{J}^v} \xi_t^v\right)\right) \\
&\propto \left(\frac{1}{\mu_v}\right)^{\#\mathcal{J}^v + \alpha_{mv} - 1} \exp\left(-\frac{\left(\frac{1}{\mu_v}\right)}{\frac{\beta_{\mu v}}{1 + \beta_{\mu v} \sum_{t \in \mathcal{J}^v} \xi_t^v}}\right).
\end{aligned}$$

Therefore:

$$\mu_v | \xi_t^v \sim \Gamma\left(\alpha_{mv} + \#\mathcal{J}^v, \frac{\beta_{\mu v}}{1 + \beta_{\mu v} \sum_{t \in \mathcal{J}^v} \xi_t^v}\right).$$

In the case of ρ_J we have:

$$p(\xi^y | \Theta) \propto \exp\left(-\frac{\sum_{t=2}^T ((\mu_y + \rho_J \xi_t^v) - \xi_t^y)^2}{2\sigma_y^2}\right),$$

and the prior for ρ_J implies:

$$p(\rho_J) \propto \exp\left(-\frac{1}{2\sigma_{\rho J}^2} (\rho_J)^2\right).$$

So we have:

$$\begin{aligned}
p(\rho_J | \xi_t^y) &\propto p(\xi_t^y | \Theta) p(\rho_J) \\
&\propto \exp\left(-\frac{\sum_{t=2}^T (\rho_J \xi_t^v + (\mu_y - \xi_t^y))^2}{2\sigma_y^2} - \frac{1}{2\sigma_{\rho J}^2} (\rho_J)^2\right) \\
&\propto \exp\left(-\frac{(\rho_J)^2 \sum_{t=2}^T (\xi_t^v)^2}{2\sigma_y^2} - \frac{\sum_{t=2}^T 2\rho_J \xi_t^v (\mu_y - \xi_t^y)}{2\sigma_y^2} - \frac{(\rho_J)^2}{2\sigma_{\rho J}^2}\right) \\
&\propto \exp\left(-\frac{1}{\left(\frac{\sigma_{\rho J}^2 \sum_{t=2}^T (\xi_t^v)^2 + \sigma_y^2}{\sigma_y^2 \sigma_{\rho J}^2}\right)^{-1}} \left[(\rho_J)^2 - 2\rho \frac{\sum_{t=2}^T \xi_t^v (\xi_t^y - \mu_y)}{(\sigma_{\rho J}^2 \sum_{t=2}^T (\xi_t^v)^2 + \sigma_y^2) / \sigma_{\rho J}^2} \right]\right)
\end{aligned}$$

Therefore, $\rho_J | \Omega_{-\rho_J}, Y \sim N(m_{\rho J}, s_{\rho J}^2)$ where:

$$\begin{aligned}
m_{\rho J} &= \frac{\sum_{t=2}^T \xi_t^v (\xi_t^y - \mu_y)}{(\sigma_{\rho J}^2 \sum_{t=2}^T (\xi_t^v)^2 + \sigma_y^2) / \sigma_{\rho J}^2} \\
s_{\rho J}^2 &= \left(\frac{\sigma_{\rho J}^2 \sum_{t=2}^T (\xi_t^v)^2 + \sigma_y^2}{\sigma_y^2 \sigma_{\rho J}^2}\right)^{-1}.
\end{aligned}$$

Drawing Jump Times and Jump Frequency

Drawing from λ_y and λ_v is an identical procedure except that in one case the posterior depends on J^y and in the other of J^v but both work as follows:

$$\begin{aligned}\lambda &\sim B(\alpha_\lambda, \beta_\lambda) \\ p(\lambda) &\propto \lambda^{\alpha_\lambda-1} (1-\lambda)^{\beta_\lambda-1},\end{aligned}$$

and if

$$p(J|\lambda) \propto \lambda^{\#\mathcal{J}} (1-\lambda)^{T-\#\mathcal{J}},$$

then:

$$\begin{aligned}p(\lambda|J) &\propto p(J|\lambda)p(\lambda) \\ &\propto \lambda^{\#\mathcal{J}+\alpha_\lambda-1} (1-\lambda)^{T-\#\mathcal{J}+\beta_\lambda-1}.\end{aligned}$$

Thus:

$$\lambda|J \sim B(a_\lambda + \#\mathcal{J}, \beta_\lambda + T - \#\mathcal{J}).$$

For the jump sizes, the prior distributions for ξ_t^v and ξ_t^y are given by:

$$\xi_t^v \sim \exp(\mu_v),$$

and

$$\xi_t^y|\xi_t^v \sim N(\mu_y + \rho_J \xi_t^v, \sigma_y^2).$$

Therefore,

$$p(\xi_t^y|\xi_t^v) \propto \exp\left(-\frac{((\mu_y + \rho_J \xi_t^v) - \xi_t^y)^2}{2\sigma_y^2}\right).$$

Then:

$$\begin{aligned}
p(\xi_t^y | J_t = 1, \Theta_{-\xi_t^y}, V_{t-1}, Y_t) &\propto p(Y_t | \xi_t^y, J_t = 1, V_{t-1}, \Theta) p(\xi_t^y | \Theta) \\
&\propto \exp \left(-\frac{((\Delta Y_t - \mu) - \xi_t^y)^2}{2V_{t-1}} - \frac{((\mu_y + \rho_J \xi_t^v) - \xi_t^y)^2}{2\sigma_y^2} \right) \\
&\propto \exp \left(-\left[-2\xi_t^y \left(\frac{(\Delta Y_t - \mu) \sigma_y^2 + (\mu_y + \rho_J \xi_t^v) V_{t-1}}{2\sigma_y^2 V_{t-1}} \right) + (\xi_t^y)^2 \frac{(\sigma_y^2 + V_{t-1})}{2\sigma_y^2 V_{t-1}} \right] \right) \\
&\propto \exp \left(-\frac{1}{2 \frac{\sigma_y^2 V_{t-1}}{(\sigma_y^2 + V_{t-1})}} \left[-2\xi_t^y \left(\frac{(\Delta Y_t - \mu) \sigma_y^2 + (\mu_y + \rho_J \xi_t^v) V_{t-1}}{(\sigma_y^2 + V_{t-1})} \right) + (\xi_t^y)^2 \right] \right).
\end{aligned}$$

Thus:

$$\xi_t^y | \Omega_{-\xi_t^y} \sim N \left(\frac{(\Delta Y_t - \mu) \sigma_y^2 + (\mu_y + \rho_J \xi_t^v) V_{t-1}}{(\sigma_y^2 + V_{t-1})}, \frac{\sigma_y^2 V_{t-1}}{(\sigma_y^2 + V_{t-1})} \right),$$

and for the case of ξ_t^v we have that by two successive applications of Bayes' formula:

$$\begin{aligned}
p(\xi_t^v | J_t = 1, \Theta, V_{t-1}, Y_t, \xi_t^y) &\propto p(Y_t, V_t | \xi_t^y, \xi_t^v, J_t = 1, V_{t-1}, \Theta) p(\xi_t^v | J_t = 1, \Theta, \xi_t^y) \\
&\propto p(Y_t, V_t | \xi_t^y, \xi_t^v, J_t = 1, V_{t-1}, \Theta) p(\xi_t^y | J_t = 1, \Theta, \xi_t^v) p(\xi_t^v | J_t = 1, \Theta).
\end{aligned}$$

So:

$$\begin{aligned}
p(\xi_t^v | J_t = 1, \Theta, V_{t-1}, Y_t, \xi_t^y) &\propto 1_{\xi_t^v > 0} \exp \left(-\frac{1}{2(1-\rho^2)} \left[\frac{((\Delta Y_t - \mu) - \xi_t^y)^2}{V_{t-1}} + \frac{(\alpha_v + \beta_v V_{t-1} + \xi_t^v - V_t)^2}{\sigma_v^2 V_{t-1}} \right. \right. \\
&\quad \left. \left. - \frac{2\rho((\Delta Y_t - \mu) - \xi_t^y)(\alpha_v + \beta_v V_{t-1} + \xi_t^v - V_t)}{\sigma_v V_{t-1}} \right] \right) \\
&\propto 1_{\xi_t^v > 0} \exp \left(-\frac{((\Delta Y_t - \mu) - \xi_t^y)^2}{2V_{t-1}(1-\rho^2)} - \frac{(\alpha_v + \beta_v V_{t-1} + \xi_t^v - V_t)^2}{2\sigma_v^2 V_{t-1}(1-\rho^2)} \right. \\
&\quad \left. + \frac{2\rho((\Delta Y_t - \mu) - \xi_t^y)(\alpha_v + \beta_v V_{t-1} + \xi_t^v - V_t)}{(1-\rho^2)\sigma_v V_{t-1}} \right. \\
&\quad \left. - \frac{((\mu_y + \rho_J \xi_t^v) - \xi_t^y)^2}{2\sigma_y^2} - \mu_v \xi_t^v \right) \\
&\propto 1_{\xi_t^v > 0} \exp \left(-(\xi_t^v)^2 \left[\frac{\sigma_y^2 + \rho_J^2 \sigma_v^2 V_{t-1}(1-\rho^2)}{2\sigma_v^2 \sigma_y^2 V_{t-1}(1-\rho^2)} \right] \right. \\
&\quad \left. - 2\xi_t^v \left[\frac{\rho_J(\mu_y - \xi_t^y)\sigma_v^2 V_{t-1}(1-\rho^2)}{2\sigma_y^2 \sigma_v^2 V_{t-1}(1-\rho^2)} + \frac{\mu_v \sigma_v^2 \sigma_y^2 V_{t-1}(1-\rho^2)}{2\sigma_v^2 \sigma_y^2 2V_{t-1}(1-\rho^2)} \right] \right. \\
&\quad \left. + \frac{(\alpha_v + \beta_v V_{t-1} - V_t)\sigma_y^2}{2\sigma_v^2 \sigma_y^2 V_{t-1}(1-\rho^2)} - \frac{\rho\sigma_v \sigma_y^2 ((\Delta Y_t - \mu) - \xi_t^y)}{2(1-\rho^2)\sigma_v^2 \sigma_y^2 V_{t-1}} \right] \right),
\end{aligned}$$

and therefore $\xi_t^v | J_t = 1, \Theta, V_{t-1}, Y_t, \xi_t^y$ is distributed as a truncated normal, with left truncation 0.

To draw from J , we use the fact that:

$$p(J_t = 1|V_t, V_{t-1}, Y_t, \xi_t^y, \Theta) \propto \lambda \cdot p(Y_t, V_t|J_t = 1, \Omega_{-J_t}) \quad (12)$$

$$p(J_t = 0|V_t, V_{t-1}, Y_t, \xi_t^y, \Theta) \propto (1 - \lambda) \cdot p(Y_t, V_t|J_t = 0, \Omega_{-J_t}) \quad (13)$$

we define $q = \Pr(J_t = 1|\Omega_{-J_t})$. Then, we can define the odds ratio:

$$O = \frac{q}{1 - q} = \frac{p(J_t = 1|V_t, V_{t-1}, Y_t, \xi_t^y, \Theta)}{p(J_t = 0|V_t, V_{t-1}, Y_t, \xi_t^y, \Theta)}$$

Now, notice that the constants that would make (12) and (13) equalities instead of proportions are indeed the same. Thus we can calculate O as:

$$O = \frac{\lambda \cdot p(Y_t, V_t|J_t = 1, \Omega_{-J_t})}{(1 - \lambda) \cdot p(Y_t, V_t|J_t = 0, \Omega_{-J_t})}$$

$$q = \frac{O}{(1 + O)}.$$

Therefore:

$$J_t|V_t, V_{t-1}, Y_t, \xi_t^y, \Theta \sim \text{Ber}(q)$$

Thus covering all the state variables and parameters except V_t and ρ .

Drawing from Non-Standard Distributions

First, for the case of V_t notice that by successive applications of Bayes' rule:

$$\begin{aligned} p(V_t|V_{t+1}, V_{t-1}, \Delta Y_{t+1}, \Theta, J, \xi^y, \xi^v) &\propto p(V_t, V_{t+1}, V_{t-1}, \Delta Y_{t+1}|\Theta) \\ &\propto p(V_{t+1}, \Delta Y_{t+1}|V_{t+1}, V_{t-1}, \Theta) p(V_t, V_{t-1}|\Theta) \\ &\propto \left[\begin{aligned} &p(V_{t+1}|V_t, V_{t-1}, \Delta Y_{t+1}, \Theta) p(\Delta Y_t|V_{t+1}, V_{t-1}, \Theta) \cdot \\ &p(V_t|\Theta, V_{t-1}) p(V_{t-1}|\Theta) \end{aligned} \right] \\ &\propto p(\Delta Y_{t+1}|V_t, \Theta) p(V_{t+1}|V_t, \Theta) p(V_t|V_{t-1}, \Theta), \end{aligned}$$

and:

$$\begin{aligned}
p(\Delta Y_{t+1}|V_t, \Theta) p(V_t|V_{t-1}, \Theta) p(V_{t+1}|V_t, \Theta) &\propto \left(\begin{array}{c} \frac{1}{V_t^{1/2}} \exp \left\{ -\frac{(Y_{t+1} - (Y_t + \mu))^2}{2V_t} \right\} \\ \frac{1}{\sigma_v V_{t-1}^{1/2}} \exp \left\{ -\frac{(V_t - (\alpha_v + \beta_v V_{t-1}))^2}{2\sigma_v^2 V_{t-1}} \right\} \\ \frac{1}{\sigma_v V_t^{1/2}} \exp \left\{ -\frac{(V_{t+1} - (\alpha_v + \beta_v V_t))^2}{2\sigma_v^2 V_t} \right\} \end{array} \right) \\
&\propto \left(\frac{1}{V_t} \exp \left\{ \begin{array}{c} -\frac{(Y_{t+1} - (Y_t + \mu))^2}{2V_t} \\ -\frac{(V_t - (\alpha_v + \beta_v V_{t-1}))^2}{2\sigma_v^2 V_{t-1}} \\ -\frac{(V_{t+1} - (\alpha_v + \beta_v V_t))^2}{2\sigma_v^2 V_t} \end{array} \right\} \right) \quad (14)
\end{aligned}$$

Thus V_t does not seem to follow any standard distribution, so to draw from it we use a random-walk, Metropolis-Hastings algorithm, where a candidate V_t^c is drawn as $V_t^{c-1} + \zeta$ where ζ is a mean 0 normal random variable. Expression (14) is then evaluated at the candidate V_t^c and at V_t^{c-1} . If the likelihood of V_t^c is greater than that of V_t^{c-1} , V_t^c is immediately accepted, otherwise it is accepted with probability equal to the ratio of the two likelihoods. Finally, the variance of the step ζ is calibrated so that candidate draws are rejected between 30% and 60% of the time.

Drawing from ρ implies a similar procedure called an independence Metropolis-Hastings. In this case, instead of drawing from a random walk, the algorithm draws candidate ρ^c s from a uniform distribution. The decision to accept or reject a given draw is then made following the same likelihood comparison step as in the case for V_t . The details of this drawing procedures can be found in Johanes and Polson (2009).

With this, we now have procedures for drawing from all the parameters and state variables.

References

- Aït-Sahalia Y. (2001): "Closed form likelihood expansions for multivariate diffusions," Working paper, Princeton university.
- Aït-Sahalia Y. (2002): "Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach," *Econometrica* Vol. 70, pp. 223-262.
- Aït-Sahalia Y. and R. Kimmel (2007): "Maximum likelihood estimation of stochastic volatility models," *Journal of Financial Economics* Vol. 83, pp. 413-452.

- Aït-Sahalia Y. and J. Jacod (2009): "Estimating the degree of activity of jumps in high frequency data," *Annals of Statistics* Vol. 37, pp. 2202-2244.
- Aït-Sahalia Y. and A. Lo (1998): "Nonparametric estimation of state-price densities implicit in financial asset prices," *Journal of Finance* Vol. 53, pp. 499-547.
- Andersen, T. G., L. Benzoni, and J. Lund (2002): "An empirical investigation of continuous-time models for equity returns," *Journal of Finance* Vol. 57, pp. 1239-1284.
- Andersen, T. G. and J. Lund (1997): "Estimating continuous time stochastic volatility models of the short term interest rate," *Journal of Econometrics* Vol. 77, pp. 343-378.
- Aspe, P. (1995): "Mexico en 1994: las razones de la política cambiaria," *Reforma*, Mexico July 15 1995.
- Bakshi G., C. Cao, and Z. Chen (1997): "Empirical performance of alternative option pricing models," *Journal of Finance* Vol. 52, pp. 2003-2049.
- Bakshi G., C. Cao, and Z. Chen (2000): "Do call prices and the underlying stock always move in the same direction?," *Review of Financial Studies* Vol. 13, pp. 549-584.
- Bates, D. (1996): "Jump and stochastic volatility: Exchange rate processes implicit in Deutsche mark options," *Review of Financial Studies* Vol. 9, pp. 69-107.
- Bates, D. (2000): "Post-'87 Crash Fears in S&P 500 Futures Options," *Journal of Econometrics* Vol. 94, pp. 181-238.
- Black, F. and M. Scholes (1973): "The pricing of options and corporate liabilities," *Journal of Political Economy* Vol. 81, pp. 637-654.
- Bakshi G., C. Cao, and Z. Chen (1997): "Empirical performance of alternative option pricing models," *Journal of Finance* Vol. 52, pp. 2003-2049.
- Bolsa Mexicana de Valores (2011): "Nota metodológica del índice de precios y cotizaciones," accessed October 30 2011, <http://www.bmv.com.mx/wb3/work/sites/BMV/resources/LocalContent/1476/7/NotaMetdIPCnov2011.pdf>.
- Duffie, D., J. Pan, and K. Singleton (2000): "Transform analysis and asset pricing for affine jump-diffusions," *Econometrica* Vol. 68, pp. 1343-1376.

- Eraker, B. (2001): "MCMC analysis of diffusion models with application to finance," *Journal of Business and Economic Statistics* Vol. 19, pp. 177-191.
- Eraker, B., M. Johannes and N. Polson (2003): "The impact of jumps in volatility and returns," *Journal of Finance* Vol. 58, pp. 1269-1300.
- Fisher M. and C. Gilles (1996): "Term premia in exponential-affine models of the term structure," Working paper, Board of Governors of the Federal Reserve System Washington DC.
- Gallant R. A. and G. Tauchen (1996): "Which moments to match?," *Econometric theory* Vol. 12 pp. 657-681.
- Heston, S. (1993): "A closed-form solution of options with stochastic volatility with applications to bond and currency options," *The Review of Financial Studies*, Vol. 6, pp.327-343.
- Honore, P. (1998): "Pitfalls in estimating jump-diffusion models," Working paper, University of Aarhus.
- Hull, J. and A. White, "The pricing of options on assets with stochastic volatilities," *Journal of Finance* Vol. 42 (1987), pp. 281-300.
- Jacquier E., N. G. Polson, and P. E. Rossi (1994): "Bayesian analysis on stochastic volatility models," *Journal of Business and Economic Statistics* Vol. 12, pp. 371-396.
- Johannes, M. and N. Polson (2009): "MCMC methods for continuous-time financial econometrics" in Aït-Sahalia, Y. and L. Hansen, Ed. *Handbook of Financial Econometrics Vol. 2*, North Holland, Oxford.
- Johannes, M., N. Polson and J. Stroud (2009): "Optimal Filtering of Jump-Diffusions: Extracting Latent States from Asset Prices," *Review of Financial Studies* Vol. 22, pp. 2759-2799.
- Kim, S., Shephard, N., and Chib, S. (1999): "Stochastic volatility: likelihood inference and comparison with ARCH models," *Review of Economic Studies* Vol. 65, pp. 361-393.
- Ledoit, O., P. Santa-Clara, and S. Yan (2002): "Relative pricing of options with stochastic volatility," Working Paper , University of California at Los Angeles.
- McFaden D. A. (1989): "Method of simulated moments for estimation of a discrete response models without numerical integration," *Econometrica* Vol. 57, pp. 995-1026.

- Melino A. (1994): "Estimation of continuous-time models in Finance" in C. A. Sims, Ed.: *Advances in Econometrics: Sixth World Congress, Vol. II* Cambridge university press, Cambridge.
- Merton, R.C. (1980): "On estimating the expected return on the market: An exploratory investigation", *Journal of Financial Economics* Vol. 8, pp. 323-361.
- MexDer (2008): "Metodología para el cálculo de precios de liquidación por medio del modelo de Heston," accessed October 30 2011, http://www.mexder.com.mx/inter/info/mexder/avisos/Metodologia_de_Precios_de_Liquidacion_por_medio_de_Heston.pdf.
- Pakes, A. and D. Pollard (1989): "Simulation and the asymptotics of optimization estimators," *Econometrica* Vol. 57, pp. 1027-1058.
- Pan, J. (2002): "The jump-risk premia implicit in options: Evidence from an integrated time-series study," *Journal of Financial Economics* Vol 63, pp. 3-50.
- Preston, J. and S. Dillon (2004): "Opening Mexico: The making of a democracy" Farrar, Straus and Giroux, New York.
- Scott, L.O. (1987): "Option pricing when the variance changes randomly: Theory, estimation, and an application", *Journal of Financial and Quantitative Analysis* Vol. 22, pp. 419-438.
- Sepp A. (2003): "Fourier transform for option pricing under affine jump-diffusions: An overview," Working paper, University of Tartu.
- Smith, A. F. M. and D. J. Spiegelhalter (1980): "Bayes factors and choice criteria for linear models," *Journal of the Royal Statistical Society. Series B (Methodological)* Vol. 42, pp. 213-220.
- Stein, J.C. (1989): "Overreactions in the options market," *Journal of Finance* Vol. 44, pp. 1011-1023.
- Wiggins, J. B. (1987): "Option values under stochastic volatilities," *Journal of Financial Economics* Vol. 19, pp. 351-372.