State Dependent Jump Models:

How do US Equity Indices Jump?

Michael Johannes and Rohit Kumar[1]
Department of Economics, University of Chicago

Nicholas G. Polson
GSB, University of Chicago

First draft: September 21, 1998

This draft: November 10, 1999

Summary

This paper introduces a class of state dependent jump (SDJ) models in which the arrival intensity and jump sizes depend on a given set of state variables, including lagged jumps. With this model, we investigate the structure of jumps to U.S. equity indices, concentrating on the predictability of jump times and sizes. Evidence for strong predictability of jumps times is found for all of the indices considered: Standard and Poor's 500 and Mid-Cap, the Russell 1000, 2000 and 3000 indices, the Wilshire 5000 and the Nasdaq 100 (NDX). Given the evidence for predictability, we show how risk management decisions are affected by state dependent jump structures. Using inferred jump times and sizes, we also detail the shortcomings of popular jump models and demonstrate how jump models fit various events such as the Crash of 1987. Using implied volatility data, we investigate the ability of implied volatility to predict both jump times and sizes and find strong evidence that implied volatility can predict both jump times and sizes.

[1]We thank Lars Hansen and Bjorn Eraker for their comments and seminar participants at Duke University, University of Chicago and participants at the 1999 North American Summer Meetings of the Econometric Society and 1999 Summer Meetings of the ASA. Address correspondence to Michael Johannes, Department of Economics, University of Chicago, 1126 East 59^{th}, Chicago, IL 60637 or via email at msjohann@midway.uchicago.edu. The paper is also available at http://student-www.uchicago.edu/users/msjohann/. All remaining errors are ours.

1 Introduction

Large, sudden movements occur periodically in nearly all financial markets. Whether referred to as crashes, devaluations, corrections, defaults or jumps, they have substantial implications, especially for risk managers and participants in derivative markets since these movements induce margin calls, option exercise and sometimes default by counter-parties. Recent examples of these movements over the last few years in equity, currency and fixed income markets include U.S. equity markets (October 1987 and 1997), the Mexican Peso (1994), the Thai Bhat (1997), the Malaysian Ringgit (1997), Russian GKO's (1998) and the Brazilian Real (1999).

Anecdotal evidence suggests that these jumps tend to persist, at least in equity markets. Two examples document this. During the crash of October 1987 the S&P 500 cash index fell 5% on Friday (Oct 16th), fell 20% on Monday (Oct 19) and rose 5% on Tuesday (Oct.20). In the mini-crash of October 1997, the NDX (Nasdaq 100) fell more than 7% on Oct 27th, only to rise almost 7% the following day, with little discernible increase in volatility before or after. As noted in Schwert (1990) in his study of daily equity index returns over the last 200 years, "there are many reversals, when large drops in stock prices have been followed by large increases in stock prices" (p. 79). This persistence in the sign and magnitude is in the opposite direction of the "usual" positive autocorrelation in daily returns reported in, for example, Campbell, Lo and MacKinlay (1997) and Schwert (1990). Furthermore, this type of movement cannot be accommodated in standard models with time varying volatility as the shocks in these models are i.i.d. and mean zero.

Although jumps have been incorporated into most theoretical fields of finance[2], surprisingly little research has been directed toward the empirical content of models with jumps. This leaves unanswered a number of important financial questions: Are jump times and sizes predictable? If so, what variables help to predict them? Do jumps follow jumps? Do positive jumps follow negative jumps? How do jumps affect the distribution of daily returns? What proportion of volatility comes from jumps?

The structure of jumps is important for a number of financial applications including risk and portfolio managment and derivatives pricing and hedging. Jumps, if present, will determine the tail behavior of returns, and thus careful modeling of jumps is required to

[2]Examples abound in various fields of finance. For example, in option pricing, see Merton (1976a,b), Bates (1996), Bakshi, Cao and Chen (1997) and Duffie, Pan and Singleton (1998), in fixed income see Duffie and Kan (1995), Bjork, Kabanov and Rungaldier (1997)), Das (1998), in portfolio selection see Merton (1972), Bardhan and Chao (1996), Aase (1984, 1986) and Jeanblanc-Pique and Pontier (1990), in risk management see Duffie and Pan (1996, 1998), Zangari (1996,1998) and Venkataraman (1997) and in general equilibrium asset pricing see Aase (1993), Naik and Lee (1990) and Bardhan and Chao (1996). Aase (1993) offers a simple model with jumps as an explanation of the equity premium puzzle.

accurate model the conditional distribution of returns. Jumps are especially important for risk management procedures such as Value-at-Risk (VaR) which identify the maximum movement in asset prices for a given coverage probability. Ironically, most model used to compute VaR explicitly rule out jumps, and thus do not contain the very features of the data they are meant to guard against. As noted by Federal Reserve Chairman Alan Greenspan ("Work that characterizes the statistical distribution of extreme events would be useful" (referenced in Embrechts et al (1998))), developing and analyzing models that explicitly account for these movements is important.

Existing state-independent jump models preclude the dependencies mentioned above typically assuming jump times and sizes are i.i.d.[3]. In order to address the issues of predictability and state-dependence, we introduce a new framework for modeling financial state variables subject to periodic jump movements. The state dependent jump (SDJ) model posits jumps as state dependent variables: the probability of a jump arriving and the size of a jump depend on given state variables, which could include previous jumps. The model produces a more flexible model capable of capturing large price movements while nesting many of the popular empirical models.

Methodologically, we provide an computationally efficient method for performing statistical inference on the SDJ model. Existing methods for inference in models with jumps focus exclusively on parameter estimation and do not attempt to estimate the latent jump times or sizes, which provide valuable model diagnostics as well as information on how the model fits various events, such as the Crash of 1987. Using Markov Chain Monte Carlo (MCMC) methods, we provide an integrated approach to estimation and inference, that simultaneously estimates parameters and the latent jump times and sizes from the observed data. Furthermore, latent variable estimation is performed accounting for parameter uncertainty. This allows us to analyze the historical structure of jumps and provides additional insight into issues regarding the sizes of neighboring jumps and the probabilities that jumps occurred.

Empirically, we analyze the structure of jumps in U.S. equity indices using the SDJ model for the following indices: S&P 500 and Mid-Cap, Russell 1000, 2000 and 3000, the Wilshire 5000 and the NDX. Using estimated jump times and sizes, we first document that standard state-indepedent models of returns such as Merton's (1976a) jump-diffusion model cannot capture the presistence in the data. Jump times are clustered, in contrast to the i.i.d. specification and jump sizes seem to exhibit temporal dependence with large negative jumps often preceding large positive jumps. To explain these findings, we turn to the SDJ model and focus on the predictability of jump times and sizes for each of the

[3]Bakshi and Madan (1999) specify a different class of jump models, although the models are still independent through time, unable to account for the well-documented persistence in return data.

indices and the structure of the jump processes across different indices. We find strong evidence for predictability and state dependence in the jump arrivals and some evidence for predictability of jump sizes. Using standard value-at-risk calculations, we demonstrate the effect of state-dependencies on VaR.

Finally, we use the model to take a new look at the information contained in implied volatility. Using a general measure of implied volatility, the VIX index, we analyze its ability to predict the arrivals and sizes of jumps. Previous research on implied volatility focuses on predictable components of implied volatility itself, the ability of implied volatility to predict future volatility and the ability of implied volatility to predict some measure of volatility, such as GARCH[4]. Evidence for the ability of implied volatility in these settings to explain realized volatility or GARCH volatility is at best mixed. We find extremely strong evidence that the VIX index can predict both jump arrivals and sizes, with important implications for option market participants and risk management procedures.

The rest of the paper is as follows. Section 2 introduces the general form of our state dependent jump model and relates it to it to the continuous time jump diffusion model. Section 3 describes our general hierarchical framework for inference using Merton's (1976a) model as an example. Section 4 introduces the specific state dependent model we implement and derives an algorithm for estimation and inference. Section 5 gives simulation results justifying our ability to extract jump times and sizes and gives the numerical results of estimation and jump times and size extraction. Section 6 has conclusions and discusses possible extensions.

2 State Dependent Jump Models

The motivation for the model we consider is very simple and is demonstrated in Figure 1 which displays the Nasdaq 100 during 1997. The main feature in Figure 1 is the large decrease (Oct 27^{th}, -7.7%) followed by a large increase, almost 7%, the following day. There are two notable features of this event. First, it is difficult to attribute this move to the standard time varying volatility models, such as ARCH, GARCH or Stochastic Volatility, because there is little evidence of increased volatility prior to larger movements. In the preceding month, the NDX 100 traded in a relatively tight band of $\pm 2\%$, giving little indication of an increase in volatility. In fact the volatility of returns over month prior was at the historical average. Using the previous months data to estimate volatility, both movements exceeded six standard deviations, extremely unlikely, especially on neighboring days.

[4]See for example, Harvey and Whaley (1992), Day and Lewis (1992), Canina and Figlewski (1993) and Lamoureoux and Lastrapes (1993).

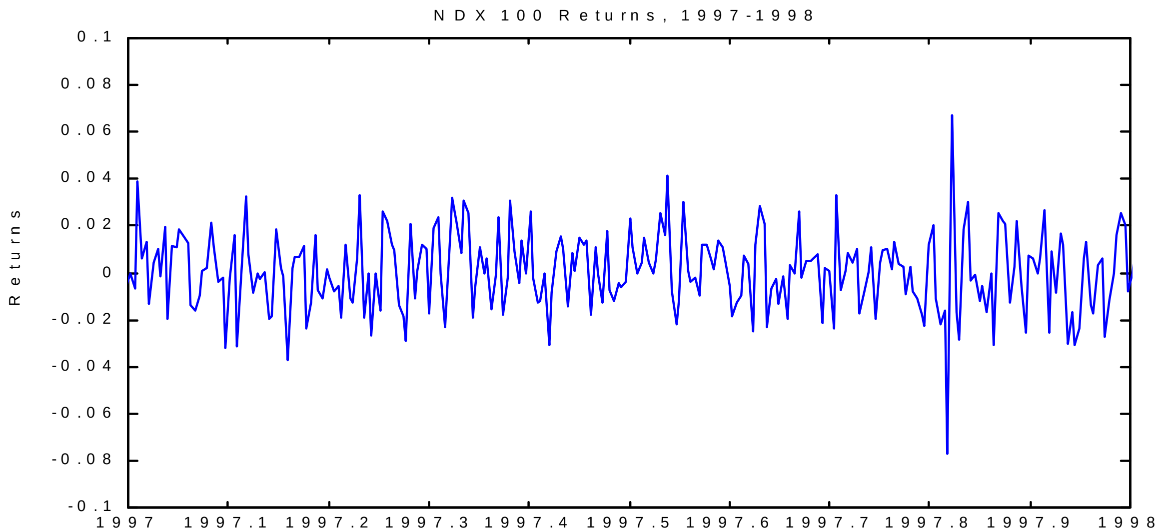

Figure 1: Nasdaq 100 continuously compounded returns during 1997.

Second, after the 7% decline on Oct 27^{th}, the loss was almost recouped the following day. This implies that a jump on one day may increase the probability of a jump the following day and the a large negative movement will be followed by a large positive movement the following day. Both of these features are not well captured by standard models of market volatility such as GARCH or Stochastic volatility, as the levels of the shocks are not related through time.

2.1 The General Model

To capture these effects, the state dependent jump (SDJ) model specifies that the state variable, X_t, evolves according to

$$X_{t+1} = X_t + \mu_t + \sigma_t \varepsilon_{t+1} + c_{t+1} J_{t+1} \tag{1}$$

where X_t could be the level or logarithm of index level.or returns. This distributional assumptions are that $\varepsilon_t \sim \mathcal{N}(0,1)$, $c_{t+1} = c(X_t, \xi_{t+1}, \Theta)$, $\xi_{t+1} \sim \Pi$ is the jump size and $J_{t+1} \in \{0,1\}$ indicates the arrival of a jump. Θ generically refers to any unknown parameters of the model. The general SDJ model has three factors: the normally distributed error term, the jump times and jump sizes (impacts) and following the continuous time literature we refer to μ, σ and c as the drift, diffusion and jump impact functions (all potentially time-varying). Through judicious choice of the time varying conditional mean and volatility, this model nests many of the popular time series models used in empirical finance, for example, ARCH, GARCH and stochastic volatility models.

The rate of arrivals is controlled by the intensity function, which we specify as

$$Prob\,[J_{t+1} = 1 | X_t, J_t, \Theta] = \Phi(\beta_0 + \beta_1 J_t + \beta_2 X_t)$$

where Φ is the standard normal cumulative distribution function. We choose this specification for two reasons. First, the cumulative normal CDF specification fits nicely into our estimation. Second, the specification retains the attractive intuition of a Probit model in which jumps times are explained through the "regressors" in the CDF. In this regard, the model provides for a vehicle to identify any variables affecting the probability of a jump.

When a jump arrives, its impact on the state variable is $c_{t+1} = c(X_t, \xi_{t+1})$. This impact is influenced in two major ways (ignoring parameterization of the function). First, if there is a systematic relationship between jump impact and state variables, the impact of the jumps can, in part, be explained by the state variables. In a degenerate case in which the random jump size, ξ_{t+1} is constant, the sizes of the jumps are perfectly explained by the state variables. Again the addition of exogenous variables is easy at this stage. The second way to influence the impact is through the distribution of the jump sizes, Π. This distribution can be either discrete or continuous depending on the application. Although for the most part the literature assumes normally distributed jumps to log-returns in equities, the model and our methodology introduced later facilitates heavy-tailed distributions such as a $t(\nu)$ distribution or other scale mixtures of normal distributions.

The simplest example of this model is a time-discretization of Merton's (1976a) model in which continuously compounded returns have state independent drift, diffusion and jump impacts:

$$r_{t+1} = \log\left(\frac{X_{t+1}}{X_t}\right) = \mu + \sigma\varepsilon_{t+1} + \xi_{t+1} J_{t+1} \tag{2}$$

where the jumps arrive with constant intensity λ. Assuming that the jumps to the continuously compounded returns are normally distributed ($\xi \sim N(\xi_0, \sigma_\xi^2)$), the distribution of log-returns is an i.i.d. mixture of normal distributions. Variants of this model have been estimated in Honore (1998), Rosenfeld and Jarrow (1983), Ball and Torous (1983,1985), Jorion (1988) and Andersen, Benzoni and Lund (1999). Merton's model is independent across time and it therefore will not account for the well-documented persistence seen in most financial time series.

The SDJ model can be seen as a time-discretization of a jump-diffusion model, where the discretization interval, for our applications, is one day. The model in equation (1), under mild regularity conditions, converges to a jump-diffusion, provided the jump intensity is independent of J_t. Since the discretization error is very small when discretization interval is small, say daily, the SDJ model and a time discretization of a jump-diffusion model can be recognized as being approximately observationally equivalent. Appendix 1 reviews this link between the SDJ model and the jump-diffusion model.

As a model of the conditional distribution of equity returns, the SDJ model has important implications for risk management and derivative pricing. To see the conditional structure, suppose it was known whether or not a jump occurred ($J_{t+1} = 1$) on a given day and the size of the jump (ξ_{t+1}). Then the state variables are normally distributed:

$$p(X_{t+1} - X_t | X_t, J_{t+1}, c_{t+1}, \Theta) \sim \mathcal{N}(\mu_t + c_{t+1} J_{t+1}, \sigma_t). \tag{3}$$

In this sense, the model is similar in form to a regime switching model. However, since the jump impacts depend on both state and the random outcome of the jump, the state is not necessarily a finite state Markov process, which is typically assumed in the literature.

Introducing uncertainty regarding the times and sizes of jumps produces an extremely flexible class of mixture distributions as the jump times and sizes are integrated out of equation (3). Particularly convenient assumptions at this stage that lead to closed form densities are either discrete or normally distributed jumps and state independent jump impacts ($c(X_t, \xi_{t+1}) = \xi_{t+1}$) which implies, in the case of Merton's model:

$$p(r_{t+1}|\Theta) \sim \lambda \mathcal{N}(\xi_0 + \mu, \sigma^2 \Delta + \sigma_\xi^2) + (1 - \lambda)\mathcal{N}(\mu, \sigma^2 \Delta).$$

This model is quite flexible along certain dimensions since mixtures of normal distributions generate a broad class of distributions (multi-modal, fat-tailed, asymmetric).

It is important to note that the SDJ model has quite different conditional implications than ARCH or GARCH models. Unlike these models which are conditionally normally distributed, the SDJ model has both conditional skewness and kurtosis. This has important implications for portfolio management and option pricing, but are perhaps most transparent for risk management. Duffie and Pan (1996 and 1998), Zangari (1996 and 1997), and Venkataraman (1997) argue that jumps improve on standard Value-at-Risk, or VaR, calculations used in risk management. VaR identifies the maximum movement over are given period of time with a certain probability. Given that conditional return distributions exhibit skewness and kurtosis, careful modeling of the tails of the these distributions is important (see Greenspan quote in the introduction). When present, jumps will dominate the other components in the tails and any predictability in the jump components will have important implications for VaR.

As an example, suppose there was a strong persistence in jumps ($\beta_1 > 0$) and that a jump occurred at time t. In this case, the probability of a jump tomorrow is much higher and tomorrow's observation will be dominated by the jump component. Fitting a simple state independent model would never capture this and could result in severe miscalculations in VaR using standard models.

2.2 A Specific SDJ model for Equity Indices

For empirical application, we need to specify the functional forms of the drift, diffusion and jump impact function. In order to concentrate on the predictability of jump times and sizes, we retain the state independent diffusion function. When the state variable is the return on an equity index, we consider

$$\begin{aligned}
&r_{t+1} = \mu_0 + \mu_1 r_t + \sigma\varepsilon_{t+1} + c_{t+1}J_{t+1} \\
&Prob\left[J_{t+1} = 1|r_t, J_t\right] = \Phi\left(\alpha + \beta_1 J_t + \beta_2|r_t|\right) \\
&c_{t+1} = \xi_0 + \xi_1 r_t + \xi_{t+1}, \\
&\xi_{t+1} \sim \mathcal{N}\left(0, \sigma_\xi^2\right).
\end{aligned}$$

The main advantage is the ability of the model to provide for a mechanism to explain jump times and sizes. Consider first the jump sizes. Conditional upon a jump arriving, the impact of a jump is measured by c_{t+1} which contains three components: a constant term, ξ_0, a component related to the previous periods return, $\xi_1 r_t$ and a random component, ξ_{t+1}. The coefficient ξ_1 determines whether or not the jump sizes have a predictable component. For example, if $\xi_1 < 0$, jumps sizes will be larger if returns in the previous day were negative which captures an overreaction effect and would be consistent with the episode in Figure 1. At this stage it is easy to add additional variables that can explain the sizes of the jumps. Later, we add implied volatility to explain jump sizes.

The jump intensity function, discussed in more above, also provides a means to explain jump times and identify variables affecting the arrival rates of jumps. We also add lagged returns to the drift in order to account for the autocorrelation in returns and to avoid the possibility that ξ_1 would be spuriously capturing omitted autocorrelation from the drift. This specification provides for time varying conditional means, volatility, skewness and kurtosis, unlike ARCH and GARCH models. Furthermore, the nature of the persistence captured in the jump components decays much quicker than standard time-varying volatility models.

Although we retain a constant volatility in order to focus on the structure of jumps, it is straightforward (if somewhat computationally burdensome) to add time-varying volatility to the model. It is unclear if time-varying volatility is required as the jump process can capture persistence in not only the volatility but also in the conditional kurtosis and skewness. In this regard, the SDJ model is more flexible than the standard time-varying volatility models.

The main advantage of this specification is that it provides a intuitive setting to study predictability of jump components. Although, to our knowledge, this issue has not been addressed to date, it should not be a surprise that there may be predictable components in the jump process. Virtually all financial financial series, to some extent, have been shown to have predictable components in expected returns and volatility. This model provides

3 Estimation and Inference in SDJ models

3.1 Previous Methods and Studies

A number of papers have studied the structure of jumps to equity indices, although all of them consider state-independent jump times and sizes. For example, a partial list of these papers include Press (1967), Beckers (1981), Ball and Torous (1983, 1985), Jarrow and Rosenfeld (1982), Jorion (1988), Honore (1998), Jiang (1998) and Andersen, Benzoni and Lund (1998). Another related literature, considers studying the importance of jumps in option pricing models.

The general conclusion of the papers estimating state-independent models is that jumps play a significant role in asset returns, although the exact impact varies from study to study due to different estimation methods and data sets. Although these papers a unaninimous regarding the gains to adding jumps to the standard, Black-Scholes log-normal model, most of them do not question the ability of the state-independent jump model to fit the observed data. This may be due to a shortcoming in the estimation procedures used. None of these papers, and the estimation methods contained within, consider estimation of the latent jump times and sizes.

It is important to note that in models with latent variables, there are two goals: estimating both parameters and jump times and sizes. Much like stochastic volatility models where estimates of the volatility process provide important information, estimates of the latent jump times and sizes are a crucial part of the inference problem as they provide partial answers to many of the questions of interest. The estimation techniques previously applied to jump models (MLE, GMM, EMM and various EM algorithms) deal exclusively with parameter estimation and do not consider estimating the latent jump times and sizes. Since the parameters of the system are unknown, estimation of the jump times and sizes must be done taking into account parameter uncertainty.

Thus in considering an estimation method for the SDJ model, we require that it perform both portions of the estimation problem. We argue that all of the existing methods fail in one of two ways: first, they apply only in simple cases and second, they cannot infer the jump times and sizes in the presence parameter risk. The first part of the inference problem, parameter estimation, is straightforward in simple jumps models such as Merton's, but is much more difficult in models with state dependent jump arrivals. In Merton's model, MLE, EMM, GMM , the classical method of moments or an EM approach all apply. In the state dependent model, these models are not feasible.

For example, MLE in the SDJ model is infeasible as the likelihood function involves mixture distributions whose weights contain telescoping terms due to the lagged jump times in the intensity function.[5] For even moderately sizes data sets, the likelihood function is intractable and it is unclear if the required regularity conditions on the likelihood function will hold. Other methods based on moments (GMM, classical method of moments) are intractable even in the simplest SDJ model given the nonlinearity in the jump intensity.

Another problem is that in many simple, state independent models, the likelihood function (or the moments) are not available in closed form if the distribution of the jumps changes. To see this, consider a variant of Merton's model where the jump sizes have a $t(\nu)$-distribution. The model is still an i.i.d. mixture, but the likelihood function, the distribution of the data given the unknown parameters is

$$p(r_{t+1}|\Theta) = \int (\lambda\phi(\sigma\varepsilon_{t+1} + \xi|\Theta)d\xi + (1-\lambda)\,\phi(\sigma\varepsilon_{t+1}|\Theta)$$

where $\phi(x)$ is the standard normal pdf. The integral in the first term is not available in closed form and thus MLE is infeasible. The same shortcomings apply to GMM, classical method of moments and various EM algorithms.

Regarding the second part of the inference problem, estimating the jump times and sizes, all the previously mentioned methods fail in that they cannot infer the jump times and sizes in the presence of unknown parameters. This applies not only to MLE, but also EM and EMM. This step is paramount to the analysis of jump models and since the parameters are unknown (they have to be estimated), their estimates cannot just be plugged into a filtering formula. For example, in Merton's model, if the parameters were known with certainty, the probability of jump could easily be estimated using simple odds ratio. Furthermore, if jump times and parameters were known, then the size of the jumps could be estimated. However, since the goal is to estimate the latent variables and the parameters and other latent variables are unknown, a new methodology must be developed. The next section does this.

3.2 Hierarchical Structure

To implement the SDJ model, we re-interpret the model in a hierarchical framework that facilitates estimation of parameters, inference on latent jump times and sizes and prediction. The hierarchical framework is a powerful method for statistical analyses of models with

[5] Computing the impact of each of the mixture weights on the likelihood function for the n^{th} data point requires evaluating complicated products of 2^{n+1} normal CDF's.

latent variables as it allows the researcher to act as if the latent variables are known for certain steps in the inference and extraction procedure. Furthermore, as prior restrictions on parameter values are often required for estimation in models with jumps (see Honore (1998) and Andersen, Benzoni and Lund (1999)), the hierarchy provides a natural stage for incorporating this information.

To create the hierarchical structure, consider treating the jump times and sizes as observed along with the state variable. Much like factoring a joint density into the product of conditionals, the hierarchy breaks the model into different levels, each characterized by conditional distributions: observable variables conditional on knowing parameters and latent variables, latent variables conditional on parameters and then parameters. The hierarchy is:

$$\begin{aligned}
\text{Level 1} \quad &: \quad X_{t+1}|\xi_{t+1}, J_{t+1}, X_t, \Theta \sim \mathcal{N}\left(\mu_t + c_{t+1}J_{t+1}, \sigma_t^2\right) \\
\text{Level 2} \quad &: \quad Prob\left[J_{t+1} = 1|X_t, J_{t,}\Theta\right] = \Phi(\beta_0 + \beta_1 J_t + \beta_2 X_t) \\
&: \quad \xi_{t+1}|\Theta \sim \Pi(\Theta) \\
\text{Level 3} \quad &: \quad \Theta \sim p(\Theta)
\end{aligned}$$

The third level specifies the prior distribution of the assumed random parameters. Note that in Levels 1 and 2 by using the unobserved variables as conditioning arguments, we "act" as if the complete data is $\left\{X_{t+1}, J_{t+1}, \xi_{t+1}\right\}_{t=0}^{T-1}$. Via the hierarchy, it is easy to accommodate time-varying volatility. For example, see Eraker, Johannes and Polson (1999) for models with jumps to returns, stochastic volatility and jumps to volatility.

The hierarchy exploits the conditional independence of the Gaussian error term, jump sizes and jump times to create an intuitive structure that breaks a complicated mixture model into a number of simpler problems. For example, if the jump sizes and jump times were known, Level 1 is a nonlinear normal regression model with heteroskedasticity. The conditional independence is also useful as it simplifies adjustments by allowing one level to be changed without affecting other levels. For example, changing the distribution of jump sizes from a normal distribution to a t-distribution requires modifying only the second stage in level 2.

Another advantage of this structure is in the third level, via the specification of the prior distributions. Since SDJ models involve mixtures of distributions and parameters governing latent processes, additional a priori information is often required to identify certain parameters in sample. For example, in Merton's model standard estimation techniques such as MLE or EMM require additional parameter restrictions to identify parameters in sample. For the likelihood function to be well behaved, MLE requires a lower bound on the diffusive volatility (see (Honore (1998) for a discussion). An EMM implementation by Andersen, Benzoni and Lund (1999) of Merton's models required constraining the mean jump size to be zero, which is a strong form of prior information. Furthermore, EMM procedures also require the specification of an auxiliary model. The prior distributions in the third level provide a systematic way to incorporate prior information.

In many cases the hierarchy can be simplified. To see this, consider Merton's (1976a) model from section 2. The parameters of interest are $\lambda, \mu, \xi_0, \sigma^2$ and σ_ξ^2 and define $\underline{R} = [r_1, ..., r_T]$. Estimation of this model can be significantly simplified re-parameterizing the variance terms. This induces a Metropolis step into the MCMC algorithm described below, but significantly increases the speed of the algorithm by eliminating a step in the second level of the hierarchy. Defining $\xi_t = \xi_0 + \sigma_\xi z_t,\ z_t \sim \mathcal{N}(0,1)$ and $\tau^2 = \sigma_\xi^2/\sigma^2$, we decouple the normally distributed jump size into its components and define a measure of the relative variance of the components (recall that σ^2 is the annualized variance, while σ_ξ^2 does not have any time units associated with it). It is important to impose prior information on τ^2 as a bounded τ^2 prevents the likelihood function from blowing up. τ^2 also has important implications for option pricing (Merton (1976b)). The importance of this measure lies in the fact that as the relative variance of the jump sizes increases vis-a-vis the diffusion coefficient, the deviations of the option prices of this model increase relative to those of Black-Scholes model.

After the re-parameterization, the hierarchy collapses to

$$\begin{aligned}
\text{level 1} &: r_{t+1}|\xi_0, J_{t+1}, \mu, \sigma^2, \tau^2 \sim \mathcal{N}\left(\mu + \xi_0 J_{t+1}, \sigma^2(1+\tau^2 J_{t+1})\right) \\
\text{level 2} &: J_{t+1}|\lambda \sim Ber(\lambda) \\
\text{level 3} &: \xi_0, \mu, \sigma^2, \tau^2, \lambda \sim p(\xi_0, \mu, \sigma^2, \tau^2, \lambda).
\end{aligned}$$

This simplification is not unique to normally distributed jump sizes[6]. If we were not concerned with the efficiency of our algorithm, we could also keep the jump sizes in the second level. As can be clearly seen from the hierarchy, if the jump times were known with certainty, level 1 is just the normal regression model with heteroskedastic errors and the second level is equally simple. The prior distributions are chosen, in most cases, to be conjugate. Their specification and the choice of prior parameters is discussed in Appendix 3. For all parameters except τ^2 we have proper, but uninformative priors. The choice of prior distribution and parameters for τ^2 must be more informative and this issue is discussed in detail in Appendix 3. Without additional prior information, the likelihood based approach we use degenerates. In a classical MLE setting the problem is discussed in Kiefer (1978), Lindren (1978) and Honore (1998).

Using the modular nature of the hierarchical structure, the state dependent model can be built directly on the foundation laid by Merton's model in the previous section. We

[6] Since the $t(\nu)$-distribution is a scale mixture of normal distributions, we could incorporate $t(\nu)$-distributed jumps easily (see Carlin and Polson (1991)).

first focus on the jump intensity. To incorporate the probit specification into the MCMC algorithm, we draw on the analysis given in Carlin and Polson (1992) and Albert and Chib (1993). We define a latent factor Z_t which is related to the jump times by

$$\begin{aligned} Z_{t+1} &> 0 \text{ implies } J_{t+1} = 1 \\ Z_{t+1} &< 0 \text{ implies } J_{t+1} = 0 \end{aligned}$$

where we assume that

$$Z_{t+1} = \beta_0 + \beta_1 J_t + \beta_2 X_t + v_{t+1} \tag{4}$$

and v_{t+1} is standard normal. Z_{t+1} has the interesting interpretation as the "cause" of the jump arrivals. In this form, state dependent jump intensities can be added directly to our hierarchy, without affecting the first level. The constant is only identified relative to the threshold for the latent variable. Following standard discrete choice models, we include and estimate it.

Adding the additional latent variable and state dependencies, the hierarchy is given by:

$$\begin{aligned} \text{level 1} &: \quad r_{t+1}|\theta, r_t, J_{t+1}, \sigma^2, \tau^2 \sim \mathcal{N}\left(\mu(\theta, r_t, J_{t+1}), \sigma^2(1+\tau^2 J_{t+1})\right) \\ \text{level 2} &: \quad J_{t+1}|Z_{t+1} = \mathbf{1}_{[Z_{t+1}>0]}, \\ & \quad\; Z_t|r_t, J_t \sim \mathcal{N}\left(\beta_0 + \beta_1 J_t + \beta_2 \left|r_t\right|, 1\right) \\ \text{level 3} &: \quad \xi_0, \xi_1, \mu_0, \mu_1, \sigma^2, \tau^2, \alpha, \beta \sim p(\xi_0, \xi_1, \mu_0, \mu_1, \sigma^2, \tau^2, \alpha, \beta). \end{aligned}$$

where $\theta = (\xi_0, \mu_0, \mu_1)$ and $\mu(\xi_0, r_t, J_{t+1}, \mu_0, \mu_1) = \mu_0 + \mu_1 r_t + (\xi_0 + \xi_1 r_t) J_{t+1}$. Note that level 2 implies the Probit intensity. Note that except for altering the drift, the first level does not change. This is the advantage of the hierarchical structure. The parameters that are in common with Merton's retain the same prior structure. For the additional parameters, $\xi_1, \mu_1 \alpha$ and β we use proper, but uninformative priors discussed in Appendix 3.

3.3 Parameter inference

Returning to the general model, complete inference requires not only statements regarding the values of the parameters, Θ, but also statements regarding the unobserved jump times and jump sizes. The solution to both of these problems is found via the full posterior distribution of the parameters and unobserved data, which is given by Bayes Rule as

$$p\left(\Theta, \underline{J}, \underline{\xi}|\underline{X}\right) \propto p\left(\underline{X}|\Theta, \underline{J}, \underline{\xi}\right) p(\Theta, \underline{J}, \xi) \tag{5}$$

where $\underline{X} = [X_1, ..., X_T], \underline{J} = [J_1, ..., J_T]$ and $\underline{\xi} = [\xi_1, , , .\xi_T]$ are vectors containing the state variables, jump times and jump sizes. The posterior summarizes all of the information regarding the parameters, the jump times and sizes contained in the data.

To see how the posterior is used to solve our problems, consider inference on the j^{th} element of $\Theta, \Theta^{(j)}$. The information in the data regarding $\Theta^{(j)}$ is given by marginal posterior distribution of $\Theta^{(j)}$ given the data, $p(\Theta^{(j)}|\underline{X})$. To find this object, integrate out the other "nuisance" parameters from the posterior conditional distribution:

$$p(\Theta^{(j)}|\underline{X}) = \sum_{t=1}^{T} \int p\left(\Theta, J_{t+1}, \xi_{t+1}|\underline{X}\right) dJ_{(t+1)} d\xi_{t+1} d\Theta_{-1}^{(j)} \tag{6}$$

where $\Theta_{-1}^{(j)}$ contains all of the elements of Θ except the j^{th} element of Θ.

The joint posterior is a $2T + k$ dimensional density (k is the number of parameters) and computing integrals against it are very difficult. To sample from $p\left(\Theta, \underline{J}, \underline{\xi}|\underline{X}\right)$ we use now standard MCMC methods (see Appendix 1 for a brief discussion of MCMC methods). MCMC simulation is a conditional simulation strategy (as opposed to an unconditional simulation strategy) algorithm creates G samples from $p\left(\Theta, \underline{J}, \underline{\xi}|\underline{X}\right)$, $\left\{\Theta^{(g)}, \underline{J}^{(g)}, \underline{\xi}^{(g)}\right\}_{g=1}^{G}$, which are used to evaluate the integrals given above. MCMC methods are extensively used in statistics as a computationally efficient way to solve high dimensional integration problems and been previously used in finance applications by, for example, Jacquier, Polson and Rossi (1994), Stambaugh (1999) and Stambaugh and Pastor (1999).

As discussed in Appendix 2, the inputs into the MCMC algorithm are the conditional posteriors of the parameters and latent variables. For Merton's model they are stated in Appendix 3 where the choice of prior distribution and the parameters of the prior distribution are also stated.

3.4 Inferring Jump Times and Sizes

Returning to the general state dependent model, the Brownian increment, the jump times and the jump sizes are all unobservable and separating their impact and reconstructing the jump times and sizes is very difficult, especially given that the parameters are unknown. The goal is to be able to answer questions such as the following:

When did the jumps occur?

What are the inferred sizes of the jumps?

Do jumps follow jumps?

We answer these questions by computing the following marginal posterior probabilities

$$p(J_{t+1} = 1|\underline{X}),\ p(\xi_{t+1}|\underline{X}) \text{ and } p(J_{t+1}|J_t, \underline{X}),$$

To compute the posterior for the jump at time $t+1$, we use the jump times generated by the MCMC algorithm. Recall that at each iteration of our algorithm, we draw the entire vector of jump times along with the other parameters. This is the key to inferring the latent jump times and sizes. The algorithm iterates between drawing parameters, jump times and jump sizes from their respective conditional posteriors. For example, suppose it has generated parameters and jump sizes. The algorithm then draws updated jump times taking into account the previous draws for the parameter and jump sizes, consistent with the model structure. Then, given jump times, sizes and data the algorithm updates the parameters. After iterating back and forth, the algorithm converges: it draws from the joint distribution of parameter values, jump times and jump sizes.

The following theorem (a direct application of the Proposition in the Appendix 2) shows how we identify the jump times using the sample averages of draws $J_t^{(g)}$ from our algorithm.

Theorem: *Let* $\overline{J_{t+1}} = \frac{1}{G}\sum_{g=1}^{G} J_{t+1}^{(g)}$ *be the sample average of draws for* J_{t+1}*. Then*

$$Prob\left[\overline{J_{t+1}} \longrightarrow p(J_{t+1}=1|\underline{X})|\Theta^{(0)}=\theta^{(0)}\right]=1$$

for π*-almost all starting points of the chain,* $\theta^{(0)}$*.*

For every time period we compute the posterior probability that a jump occurred, for example, say, $p(J_{t+1}=1|\underline{X})=57\%$. This implies that we estimated a jump in 57% of our draws from our algorithm or that with posterior probability 0.57, $J_{t+1}=1$. To reconstruct the jump times we need a simple decision rule to decide if a jump occurred. The simplest rule is to infer a jump provided the posterior probability is larger than some specified value:

$$J_{t+1}^{*}=\begin{cases} 1 \text{ if } p(J_{t+1}=1|\underline{X})>\omega \\ 0 \text{ if } p(J_{t+1}=1|\underline{X})\leq\omega \end{cases}. \tag{7}$$

To infer the jump sizes, i.e. the computation of $p\left(\xi_{t+1}|\underline{X}\right)$, we can use the output of the MCMC algorithm to compute the following integral

$$p(\xi_{t+1}|\underline{X})=\sum_{J_{t+1},J_t}\int p\left(\xi_{t+1}|X_{t+1},X_t,J_{t+1},J_t,\Theta\right)p(\Theta,J_{t+1},J_t|\underline{X})d\Theta.$$

The details are given in the appendix. We report the posterior mean of this distribution, $E\left[\xi_{t+1}|\underline{X}\right]$, as our estimate of the latent jump size.

Although we do not focus on it, the prediction problem can also be solved using the output of our algorithm. There are three predictive distributions of interest, the predictive distribution of returns $(p(r_{t+1}|\underline{X}))$, the predictive distribution of next periods jump

$(p(J_{t+1}|\underline{X}))$ and the predictive distribution of the jump size $\left(p(\xi_{t+1}|\underline{X}\right)$. The distributions and their MCMC estimates are easily obtained. For example, in Merton's model, predictive distribution of next periods returns is, taking into account parameter uncertainty is given by:

$$p\left(r_{t+1}|\underline{X}\right) \approx \frac{1}{G}\sum_{g=1}^{G}\mathcal{N}\left(\mu^{(g)}+\xi_0^{(g)}J_{t+1}^{(g)},\sigma^{2(g)}(1+\tau^{2(g)}J_{t+1}^{(g)})\right).$$

These conditionals, as shown later, have important implications for Value-at-Risk calculations. Clearly, the predictive distribution for next periods returns can be heavy-tailed and skewed as opposed to conditional normality in the standard ARCH/GARCH framework.

4 Empirical results

This section evaluates the evidence for the presence of jumps in a number of prominent equity indices: S&P 500 and S&P Mid-Cap indices, the Russell 1000, 2000 and 3000 indices, the Wilshire 5000 index and the Nasdaq (NDX 100) index. Most of the indices have options and/or futures contracts written on them and are benchmarks for the mutual fund industry.

We use daily data for estimation. If a goal of inference is to estimate jump times and sizes, daily data (or even higher frequency data) is required. To why, consider a situation where returns jumps down on one a given day (Monday, Oct 19, 1997) and then jump up in a following day (Oct 20^{th} and 21^{st}). These jumps would not be seen if the observation interval is weekly or longer. It is very difficult to identify jump times and sizes with daily data, and we doubt that reliable extraction of jump times and sizes could be obtained using less frequently sampled data.

Table 1. Summary statistics for U.S. Equity Index returns

Estimates of the mean and volatility are annualized on the basis of a 252 day year and N is the number of continuously compounded return observations.

Index	Sample Period	N	Mean	Std. Dev.	Skewness	Kurtosis
S&P 500	12.30.83 - 10.1.98	3849	0.117	0.155	-3.74	88.12
NDX 100	12.30.83 - 10.1.98	3849	0.147	0.209	-0.87	15.23
S&P Mid-Cap	06.12.91 - 10.1.98	1906	0.115	0.125	-0.89	9.98
Russell 1000	01.01.88 - 10.1.98	2804	0.120	0.130	-0.80	11.20
Russell 2000	01.01.88 - 10.1.98	2803	0.094	0.113	-1.19	10.79
Russell 3000	01.01.88 - 10.1.98	2812	0.118	0.125	-0.86	11.46
Wilshire 5000	01.02.84 - 10.1.98	3848	0.109	0.137	-3.02	54.29

The data source was Datastream and the starting points of the samples vary. Since our results are exact finite sample results (conditional on the prior), the fact that some of our series are longer than others is not a problem. Summary statistics for the log-return series are given in Table 1. As indicated by the measures of skewness and kurtosis, the log-returns are extremely non-normal, although the deviation from normality appear to vary from index to index.

Before we report estimates, we assess the ability of our algorithm to accurately estimate parameters, jump times and jump sizes via a small Monte Carlo simulation study. Since, to our knowledge, there have been no efforts to estimate jump times and sizes under parameter uncertainty, it is important to document the efficacy of our method in this regard.

4.1 Simulation study

The goal of the simulation study is twofold. First, to assess the ability of our algorithm to accurately estimate the parameters $(\xi_0, \mu, \lambda, \sigma^2, \sigma_\xi^2)$. Since this model is i.i.d. and the sample size is large, we expect the method to be extremely accurate. Second, and more importantly, we want to assess the ability of the MCMC algorithm to accurately infer the jump times and jump sizes from the data, as this has not been done before. We also examine the sensitivity of our results to the prior parameters and initial values in the algorithm.

The simulation study uses 100 simulated series with 2000 observations from the state independent model with $\Delta = 1/252$ (an example of which is given in the top panel of Figure 2). The parameters used in the simulations are given in left-hand column of Table 2. The study was also performed drawing at a daily frequency from a sample path simulated at a higher frequency ($\Delta = 1/1260$). The results were virtually unchanged due to the state independence of the drift and diffusion. This implies that even though a time-discretization is used, we can reliably estimate the parameters of the continuous time process.

Table 2. Simulation results

The prior parameters are given in Appendix 2 and the model estimated is:

$$y_{t+1} = \mu + \sigma\varepsilon_{t+1} + \xi_{t+1}J_{t+1},\ \xi_{t+1} \sim N\left(\xi_0, \sigma_\xi^2\right), P\left[J_{t+1} = 1\right] = \lambda$$

Parameter	True Values	Average Posterior Mean	Average Posterior Standard Deviation
μ	0.1777	0.1770	0.043254
ξ_0	-0.0084	-0.0086	0.0044768
σ	0.114	0.114	0.00221
τ	0.332	0.354	0.1450
λ	0.05	0.04908	0.0089409
σ_ξ	0.0387	0.0401	-

We report the means and standard deviations of the posterior marginal distributions for each parameter in Table 2[7]. The parameters are estimated with a high degree of accuracy: μ, ξ_0, σ^2 and λ are within 1% of the true value while τ^2 (and thus σ_ξ^2) is off by 13%. The true values are well within a standard deviation of our posterior means. That we can estimate ξ_0 and σ_ξ^2 so accurately is surprising since the algorithm only learns about the parameters of the jump sizes when there is a jump and the simulated process only jumps on average 100 times per sample. The results are insensitive to choice of prior parameters for μ, ξ_0, λ and σ^2. As discussed in the Appendix, some care was required in the choice of the prior parameters for τ^2. This is not peculiar to our estimation methodology. Honore (1998) discusses the importance of prior constraints on this parameter in the setting of maximum likelihood.

We turn to the performance of the algorithm in estimating jump times and sizes. The top panel of Figure 2 displays a representative sample from the simulation study, while the lower two panels display the estimated jump sizes and times. To determine how many jumps we actually infer (since we only estimate a probability), we set $\omega = 50\%$. Note that there are two types of errors: not inferring a jump when one occurred $(J_t = 1, J_t^* = 0)$ or inferring a jump when there was not one $(J_t = 0, J_t^* = 1)$. Over the 100 samples, there were on average 100.97 jumps.

The results indicate that we did not infer a jump that occurred 43.39 times on average. Thus we missed about 40% of the jumps that occurred. On average we inferred a jump that did not occur 3.64 times per sample. These results are not surprising. Since the jump sizes are normally distributed with a small negative mean, many of the jumps will be close to zero. As expected, we can not pick up very small jumps since they are too small to "be noticed" by our algorithm. To see this, note that the mean of the actual absolute jump sizes inferred was 0.0325 compared to 0.0130 for the jumps that we did not correctly infer. Thus when we do not infer a jump, the actual jumps were too small. It is important to note that despite this error, we still accurately estimate λ. The reason is that in this setting, we do not have the usual estimate of λ which is $\sum_{t=1}^T J_t/T$, if the jumps were known.

Figure 3 shows in detail a representative portion of the simulations with true jump sizes (+), the simulated returns when a jump occurred (*) and the estimated jump sizes when we inferred a jump (o). We comment on three episodes marked in Figure 3 as (A), (B) and (C). Episode (A) indicates that we did not infer a jump when one actually occurred. However, the process value in that increment was only -0.016 and was to small to be "seen"

[7] We ran our MCMC algorithm for 5000 iterations, discarded the first 2000 draws (burn in period), and use the last 3000 as our Monte Carlo sample. The algorithm is extremely efficient as the output mixes well and the draws converge for any reasonable starting value. The algorithm takes less than 2.5 minutes to run 5000 times on a desktop PC using MATLAB in the Windows environment.

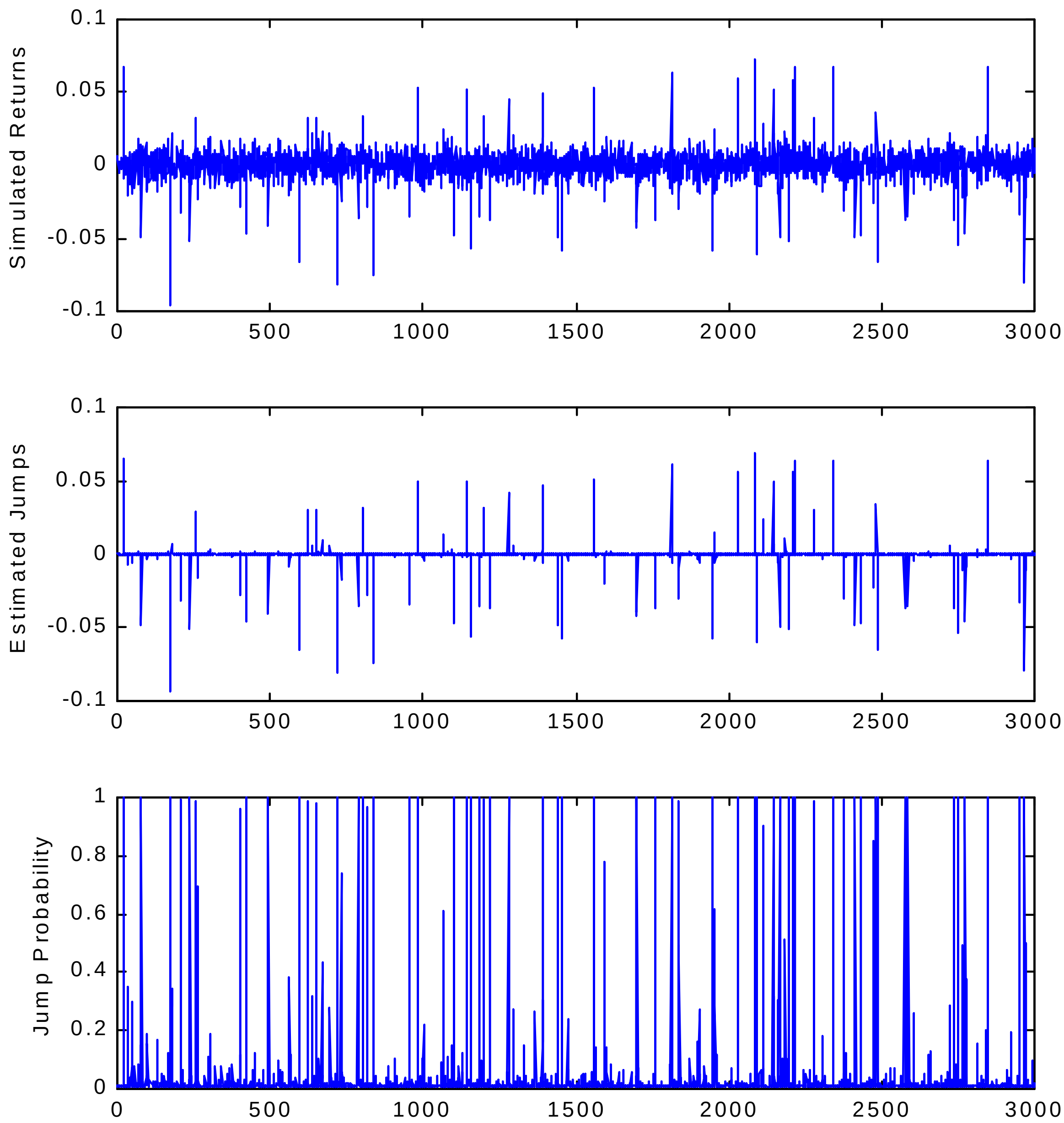

Figure 2: Summary of jump time and size estimation on simulated data. The top panel contains a representative simulated returns series, the middle panel the estimated jump sizes and the bottom panel the estimated probability that a jump occurred on a given day.

by our algorithm. Episode (B) indicates that we inferred a small jump when there was not one. Episode (C) is more indicative of our results. A 3% positive jump was nearly perfectly inferred. Notice we correctly inferred (to some degree) all of the large jumps.

Although we do not report it, we can use the same methods to extract the Brownian increments. A histogram of the inferred Brownian increments is very close to the histogram a normal random variable with mean zero and variance Δ. The extracted Brownian increments had slightly thinner tails, but still had mean zero and variance approximately Δ. The conclusion of this simulation study is that both parameter and jump time and size estimation is accurate.

4.2 The Structure of Jumps to Equity Indices

4.2.1 State Independent Model

For the indices under study, parametric inference is summarized in Table 3. Again, we report the posterior mean and standard deviation in each case. The signs and magnitudes of all the variables are as expected. In all cases, as indicated by the small posterior standard deviations, the estimates are quite accurate.

The parameter μ estimates the annualized return generated by the continuous portion of the process and is accurately estimated for all indices. In all cases, a comparison with Table 2 indicates that μ is larger than the sample mean, an implication of the negative mean jump sizes. This difference can be quite large. For the Russell 2000, the posterior mean of μ is 31.5% while the sample mean was 9.4%.

For all of the indices, the mean jump sizes are negative and much larger than the positive drifts (about 5 times as large on a daily basis). For most of the indices, the estimate of the mean jump size is negative but close to zero, although it is different from zero as in most cases the mass of the posterior is away from zero. Combined with the results for μ, the conclusion is that the indices drift upward and then, on average, fall due to jumps that are on average negative.

The jump intensity parameter, λ, is also estimated accurately and indicates indices jump with different intensities: the Nasdaq 100 jumps once every twenty days ($1/\lambda$), while the Russell 2000 jumps every six days. This is not surprising as the indices often have little overlap. For example, the Russell 2000 is a broad measure of "small-cap" stocks while the Nasdaq 100 and S&P 500 consists of much larger companies. These indices have dramatically different behavior, with the Russell 2000 jumping nearly three times as often as the S&P 500.

This asymmetry across indices is further displayed in the last column of Table 3, given a variance decomposition of the returns which shows how, in this model, the variance of

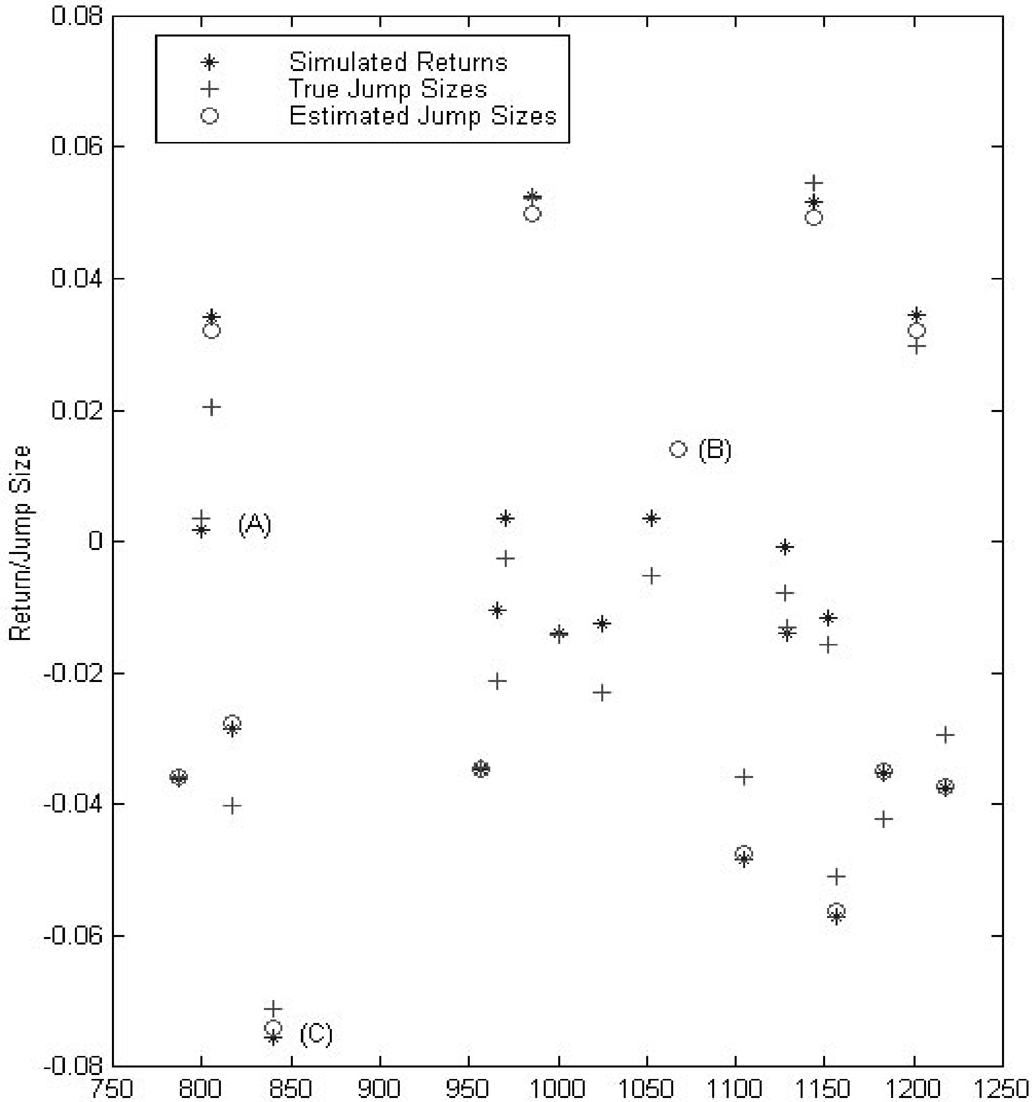

Figure 3: A detailed portion of a simulation summarizing the ability to estimate jumps.

the jumps dominates the diffusion variance for most of the indices[8]. The Russell 2000 has the highest proportion of variance from the jump component at 62% while Nasdaq has only 39%. This is an enormous difference and has important implications for option pricing. As noted in Merton (1976b), $\gamma = \lambda\sigma_\xi^2 / \left(\sigma^2 + \lambda\sigma_\xi^2\right)$, is one of the main parameters controlling the degree of mis-pricing if the Black-Scholes formula is incorrectly used when the true data generating process was a jump-diffusion. Using the results in Merton (1976b), the estimates indicate, for example, for the Russell 2000, and depending on the maturity of the option, the mis-pricing could be as much as 15-20%, a nontrivial amount. For the Nasdaq index, on the other hand, the mis-pricing would be smaller, typically less than 5%.

The structure of the jump sizes also gives insight into how jump models capture large movements. For example, for the Nasdaq 100, a 10% decline move is a 2.44 standard deviation move given the estimated jump distribution. Combining this with the probability of a jump implies that we will see 10% decline in the Nasdaq 100 occurs once every 10 years. The improvement over the geometric Brownian motion model is clear. Taking the sample standard deviation as the estimate of σ, a 10% fall is a 7.6 standard deviation move, implying it would occur every 300 billion years. This improvement should not be a surprise, as jump models are constructed to capture this effect.

Figures 4, 5 and 6 summarize the estimated jump times and sizes for the NDX 100 index. This choice was arbitrary and the other series give similar results. The top panel of Figure 4 displays the time series of NDX100 returns, the middle panel the estimated jump probabilities ($p(J_t = 1|\underline{R})$) and the bottom panel, the estimated jump sizes. Note that for many of the jumps, we have near perfect certainty that jumps occurred as $p(J_t = 1|\underline{R}) \approx 1$. Our decision rule

$$J_t^* = \begin{cases} 1 \text{ if } p(J_t = 1|\underline{R}) > \omega \\ 0 \text{ if } p(J_t = 1|\underline{R}) \leq \omega \end{cases}$$

represents a horizontal line drawn across the middle panel of Figure 4 at probability ω. A comparison of the top and bottom panels in Figure 4 indicates that periods of the high volatility (Fall 1987, 1995 and 1997) were also periods where, with high posterior probability, jumps occurred quite frequently. Also note, for the "crashes" in Oct 1987 and Oct 1997, we infer that $p(J_t = 1|\underline{R}) = 1$, which is the expected result.

The middle panel of Figure 4 gives the estimated jump sizes, while Figures 5 and 6 analyze the crashes of 1987 and 1997 in detail. From these figures, the advantage of the JD model and our estimation procedure is clear: the jump terms captures large movements while the Brownian motion captures that which it is designed to capture, small daily movements. This can be most clearly seen in Figures 5 and 6. For periods prior to and

[8] This is the exact variance decomposition is the means of the jump sizes are zero. We choose this to follow Merton (1976 B) for comparison purposes.

after the crash, we infer nearly zero jump sizes. The jump term captures both the positive and negative movements around the crash dates, as expected.

Table 3. Parameter estimates for Merton's Model

The prior parameters are given in Appendix 2 and the model estimated is the same as in Table 2. τ, σ, and μ are annualized. The jump frequency, λ, is reported in units of the probability of a jump in a given day.

	S&P 500	NDX 100	S&P MidCap	Russell 1000	Russell 2000	Russell 3000	Wilshire 5000
μ	17.4	22.1	22.6	19.0	31.5	20.1	19.0
$(\times 10^{-2})$	(3.20)	(4.79)	(4.16)	(3.43)	(3.03)	(3.31)	(2.79)
ξ_0	-4.00	-5.98	-4.93	-2.81	-5.39	-3.12	-3.46
	(2.39)	(3.44)	(1.96)	(1.54)	(0.918)	(1.40)	(1.44)
σ	11.10	17.00	9.59	9.48	7.11	9.07	9.13
$(\times 10^{-2})$	(0.292)	(0.353)	(0.397)	(0.382)	(0.266)	(0.387)	(0.259)
τ	0.273	0.226	0.199	0.206	0.199	0.204	0.243
	(0.0188)	(0.0161)	(0.0137)	(0.0117)	(0.00742)	(0.0108)	(0.0115)
λ	5.88	5.08	9.28	10.3	16.5	11.1	9.53
$(\times 10^{-2})$	(1.30)	(1.12)	(2.58)	(2.45)	(2.57)	(2.71)	(1.54)
σ_ξ	3.02	3.83	1.91	1.95	1.42	1.85	2.22
$(\times 10^{-2})$	(0.260)	(0.317)	(0.175)	(0.162)	(0.0772)	(0.150)	(0.137)
$\frac{\lambda\sigma_\xi^2}{\sigma^2\Delta+\lambda\sigma_\xi^2}$	0.523	0.394	0.476	0.527	0.621	0.537	0.586

As mentioned in earlier, the extracted jump times and sizes provide powerful diagnostics with which to analyze the model. In this case, although the jump time and size results are interesting, they indicate a potential problem with our model. The state independent model specifies that the jump times and sizes are i.i.d.. The bottom panel of Figure 4 indicates significant amounts of clustering in the jump times, while the middle panel of Figure 4 and Figure 5 and 6 indicate that large returns appear to be followed by large returns of the opposite sign.

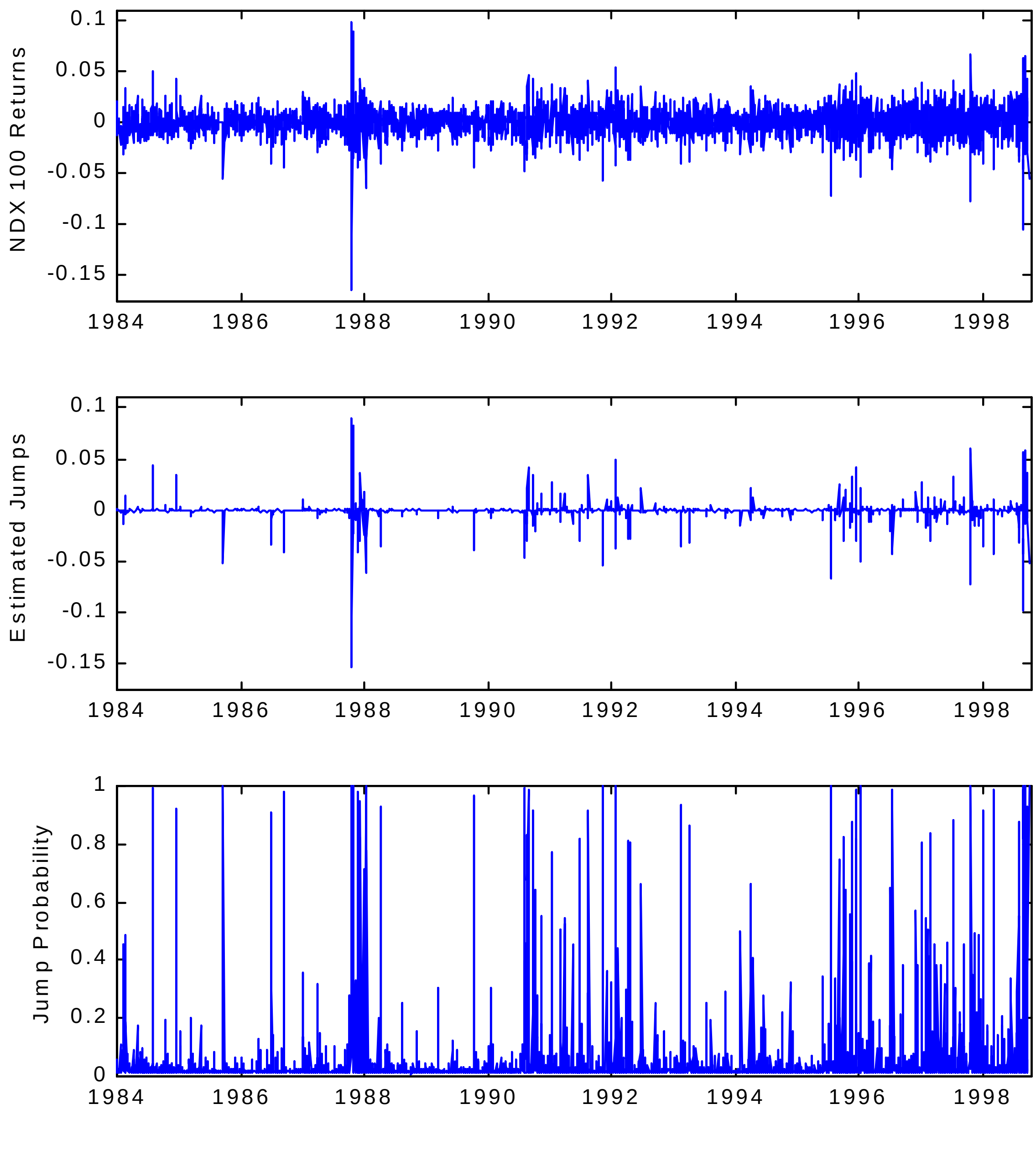

Figure 4: Summary of jump estimation for Nasdaq 100. The top, middle and bottom panels give the observed returns, estimated jump sizes and estimated jump times, respectively.

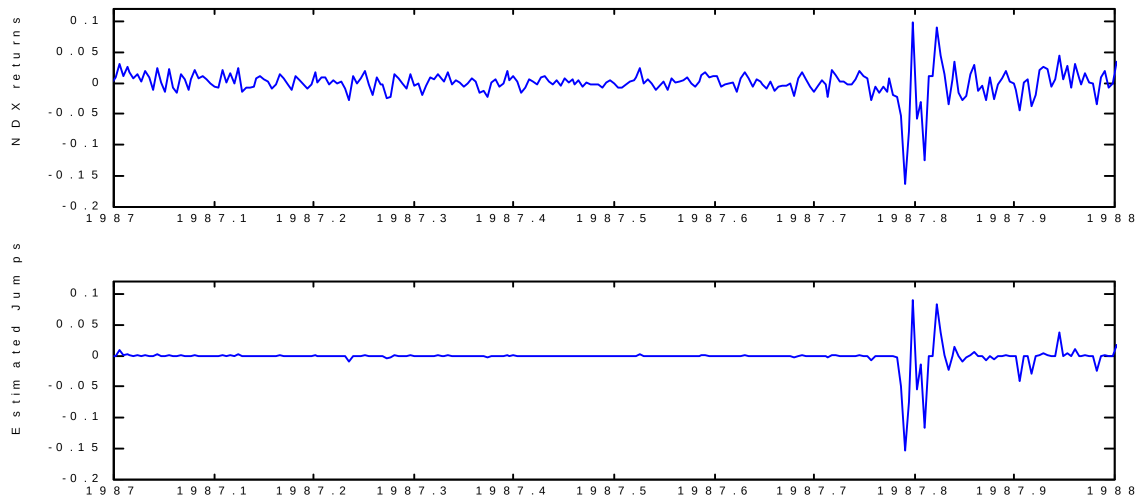

Figure 5: Returns and estimated jump sizes to the Nasdaq 100 during 1987.

This type of autocorrelation is opposite the usual positive autocorrelation documented for daily returns. It appears that there is state dependence in the jump times and maybe also in the jump sizes. To capture these state dependencies, we use the SDJ model, as described in Section 2.5.

4.3 State Dependent Model

Table 4 reports the posterior means and standard deviations for the SDJ model. The conclusions from the state independent model regarding the significance of μ_0, σ and σ_ξ still hold in the SDJ model. For all these parameters, the posterior means are well away from zero. The posterior for the volatility, σ, is very similar to that estimated in Merton's model indicating that the addition of state dependencies does not significantly alter the variance coming from the normal error term. The posterior mean of the jump variance, σ_ξ, is slightly larger in the SDJ model than Merton's model, although the posterior standard deviation is also larger.

The coefficient μ_1 is annualized and measures serial correlation of the diffusive component. In all but one case (S&P 500), the location of the posterior is well away from zero. Thus, even in the presence of jumps, there is some positive autocorrelation in daily returns

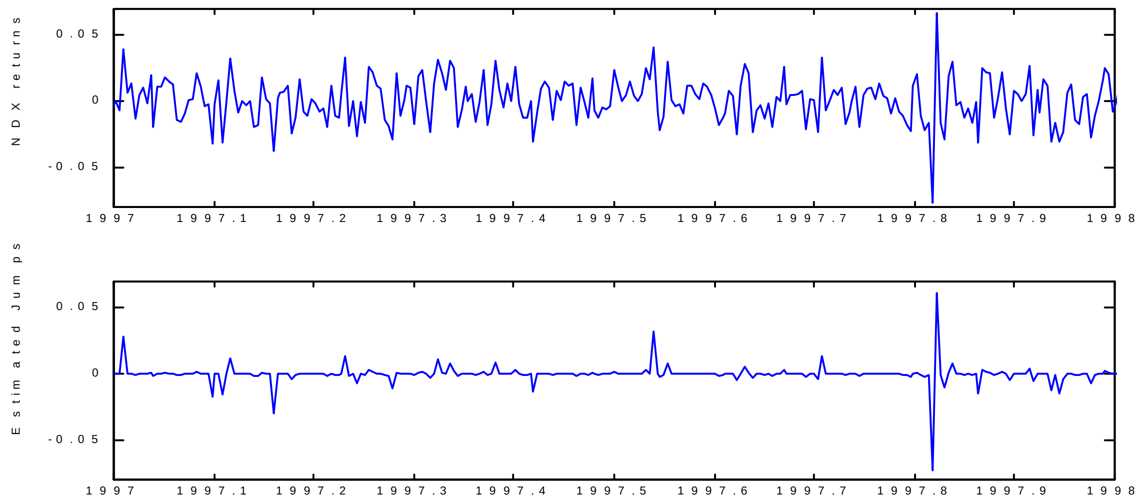

Figure 6: Returns and estimated jump sizes to the Nasdaq 100 during 1997.

and the magnitudes are roughly consistent with previous studies[9].

The mean jump size parameter, now labeled ξ_0, measures the average jump sizes conditional upon a zero return in the previous day. For all cases except the Russell 1000, significant amounts of posterior mass are below zero. The magnitudes seem quite small, usually around -0.5%. However, given the normally distributed jumps, we expect this as the jump has to accommodate both positive and negative jumps. Since from the inferred jump sizes, we know that the algorithm only picks up large jumps in absolute value, it may be advantageous to specify a more flexible jump distribution such as a mixture of two normal to capture this bi-modality.

As measured by the coefficient ξ_1 there is some evidence of predictable jump sizes for some of the indices, most notably the S&P Mid-Cap. In all cases the coefficient is negative, indicating a potential over-reaction effect: the market jumps too far down and then rebounds on the next day. The conditional volatility of jumps is roughly consistent with the results in the state independent model.

[9]See, for example, the discussion in Campbell, Lo and MacKinlay (1997), chapter 2.

Table 4. Posterior parameter summaries for the general state dependent model

The prior parameters are given in Appendix 2. σ^2, τ^2, μ_0 and μ_1 are annualized. The model estimated is given by:

$$r_{t+1} = \mu_0 + \mu_1 r_t + \sigma\varepsilon_{t+1} + c_{t+1}J_{t+1}$$
$$Prob\,[J_{t+1} = 1|r_t, J_t] = \Phi\left(\beta_0 + \beta_1 J_t + \beta_2|r_t|\right)$$
$$c_{t+1} = \xi_0 + \xi_1 r_t + \xi_{t+1},\ \ \xi_{t+1} \sim \mathcal{N}\left(0, \sigma_\xi^2\right)$$

Parameter	S&P 500	NDX 100	S&P MidCap	Russell 1000	Russell 2000	Russell 3000	Wilshire 5000
μ_0	16.42	18.59	14.83	16.52	20.97	17.38	16.77
$(\times 10^{-2})$	(3.153)	(4.860)	(4.063)	(3.433)	(2.881)	(3.375)	(2.760)
μ_1	5.949	26.462	49.740	9.627	66.445	12.580	19.045
	(4.272)	(4.366)	(6.184)	(5.102)	(5.545)	(5.139)	(4.1173)
ξ_0	-4.029	-4.975	-3.940	-2.469	-4.250	-2.812	-3.477
$(\times 10^{-3})$	(2.545)	(3.619)	(2.324)	(1.707)	(9.567)	(1.553)	(1.561)
ξ_1	-1.774	-5.713	-25.97	-1.398	-13.50	-0.7609	-4.075
$(\times 10^{-2})$	(9.361)	(10.93)	(14.49)	(1.122)	(7.705)	(10.70)	(8.384)
β_0	-1.964	-2.066	-1.833	-1.671	-1.547	-1.599	-1.642
	(0.1103)	(0.1316)	(0.1615)	(0.1178)	(0.1057)	(0.1226)	(0.09316)
β_1	0.3811	0.5975	0.5896	0.2452	0.5540	0.2665	0.3897
	(0.2388)	(0.2660)	(0.2787)	(0.2314)	(0.1604)	(0.2084)	(0.1832)
β_2	36.887	26.613	44.199	35.914	62.950	34.819	30.870
	(7.852)	(6.143)	11.930	(9.1963)	(10.521)	(9.898)	(7.0862)
σ	11.20	17.04	9.637	9.793	7.083	9.352	9.323
$(\times 10^{-2})$	(0.2318)	(0.3620)	(0.3387)	(0.2738)	(0.2184)	(0.2754)	(0.2173)
τ	28.32	23.30	21.37	21.46	20.93	21.27	25.08
$(\times 10^{-2})$	(1.852)	(1.729)	(1.665)	(1.294)	(0.8744)	(1.205)	(1.264)
σ_ξ	3.171	3.972	2.060	2.101	1.482	1.989	2.337
$(\times 10^{-2})$	(0.2509)	(0.3568)	(0.2095)	(0.1666)	(0.08906)	(0.1513)	(0.1569)

The state dependency in the jump intensity seems to have the greatest impact. Since

$$Prob\left[J_{t+1}=1|r_t, J_{t,}\Theta\right]=\Phi(\beta_0+\beta_1 J_t+\beta_2|r_t|)$$

β_1 and β_2 are very important for explaining the time series dependence of log returns. In every case, the posteriors for β_0 and β_2 are tight around the posterior mean, while the posterior for β_1 indicates that lagged jump times have an important effect for five of the indices. To interpret the impact of these results, Table 5 tabulates the daily intensity, in percentages, under different conditions on what occurred on the previous day. The first row of Table 5 gives $\Phi(\beta_0)$, which is the probability of a jump if the return on the previous day was zero and J_t was estimated to be zero. Clearly, relative to the state independent model, the probability of a jump falls drastically, with intensities being around 40% of their values in the case of Merton's model. Again, the Russell 2000 jumps most often and the NDX100 jumps the least.

TABLE 5. IMPLICATIONS OF THE STATE DEPENDENT JUMP INTENSITIES.

This table reports $\Phi(\beta_0+\beta_1 J_t+\beta_2|r_t|)$ for the various indices under different conditions

Jump Probability	S&P 500	NDX 100	S&P Midcap	Russell 1000	Russell 2000	Russell 3000	Wilshire 5000
$(r_t=J_t=0)$	2.43	2.11	3.35	4.73	6.16	5.57	5.04
$\lvert r_t\rvert=0.05, J_t=0$	45.6	22.6	64.4	54.3	94.4	55.5	46.0
$\lvert r_t\rvert=0.05, J_t=1$	60.3	44.5	82.9	64.0	99.5	65.9	61.4

4.4 Implications for Value-at-Risk

The implications of the previous estimates are quite severe for standard risk management procedures. To see this, consider the following situation: the market fell 5% in the previous day. Thus, this scenario could be the Friday before the Crash in 1987. Since jump times are predictable, the probability of a jump occurring is much higher. Table 5 summarizes the probability of a jump on the following day, using the posterior means as estimates of the parameters.

In the case that a jump was not inferred on that day ($J_t=0$), the probabilities increase drastically and range from a low of 22.6% for the NDX100 to a high of 94.4% for the Russell 2000. This is an indication of the extreme amounts of persistence in periods of market stress. If $J_t = 1$, this persistence is even greater and varies from a low of 44.5% for the NDX100 to a high of 99.5% for the Russell 2000. This implies that if there is a large movement on the previous day, the Russell 2000 is almost certain to jumps the following day.

This result, combined with the negative posterior means for ξ_1, indicates that the jump distribution has a positive mean for the following day. Thus a positive jump is expected following large negative movements. The estimated jump times and sizes are very similar to those reported in the state independent model in Figure 4 and are omitted. To see the impact of state dependence, consider using a conditional variance decomposition. Ignoring the mean jump sizes and drift, the conditional decomposition, conditional upon the previous days returns and jump time is given by:

$$\frac{\Phi(\beta_0 + \beta_1 J_t + \beta_2 |r_t|)\sigma_\xi^2}{\sigma^2 \Delta + \Phi(\beta_0 + \beta_1 J_t + \beta_2 |r_t|)\sigma_\xi^2}.$$

We compute this for the S&P500 and Nasdaq 100 for the three cases in Table 5 and the proportion of variance is given as $(34.0, 90.2, 92.4)$ and $(22.4, 75.6, 85.9)$ respectively. Thus, as we expect, the 5% decline yesterday implies that the probability of a jump today is high and therefore the variance of today's return is close to the variance of the jump, as the jump variance is an order of magnitude large than the daily diffusive variance.

What are the implications of this for VaR? We consider four different calculations: (1) normal model (lognormal returns) with sample mean and standard deviation, (2), Merton's model, with the estimates in Table 3, (3) the state dependent model with $r_t = -5\%, J_t = 0$ and (4) the state dependent model with $r_t = -5\%, J_t = 1$. Since the log-normal model and Merton's model are i.i.d., the VaR calculation is independent of the previous days return. The VaR for the four models are given by: $(1) - 0.0223, (2) - 0.0329, (3) - 0.0691, (4)$ -0.0729. Note that the VaR is increased by more than 45% by incorporating i.i.d. jumps ((1) to (2)). The VaR increases from 2 to 3 using Merton and the VaR in (4) is more than double the VaR in (2) and more than triple the VaR in (1).

This result can most easily be seen looking at the distribution of tomorrow's return given knowledge of today's return and J_t. Figure 7 displays the conditional density of returns in the Log-normal model (Black-Scholes), Merton's model and the SDJ model with $r_t = -5\%, J_t = 1$ for Nasdaq 100 returns. Since Black-Scholes and Merton are i.i.d., they ignore any information contained in the previous days return and jump. The conditional density for the SDJ model has extremely fat tails relative to the other models.

Although it appears that the conditional distribution for the SDJ model is centered left of zero, the conditional mean is quite small (-0.00488). This results from the estimate of ξ_1 for the NDX being close to zero implying that the jump sizes are not predictable using previous returns.

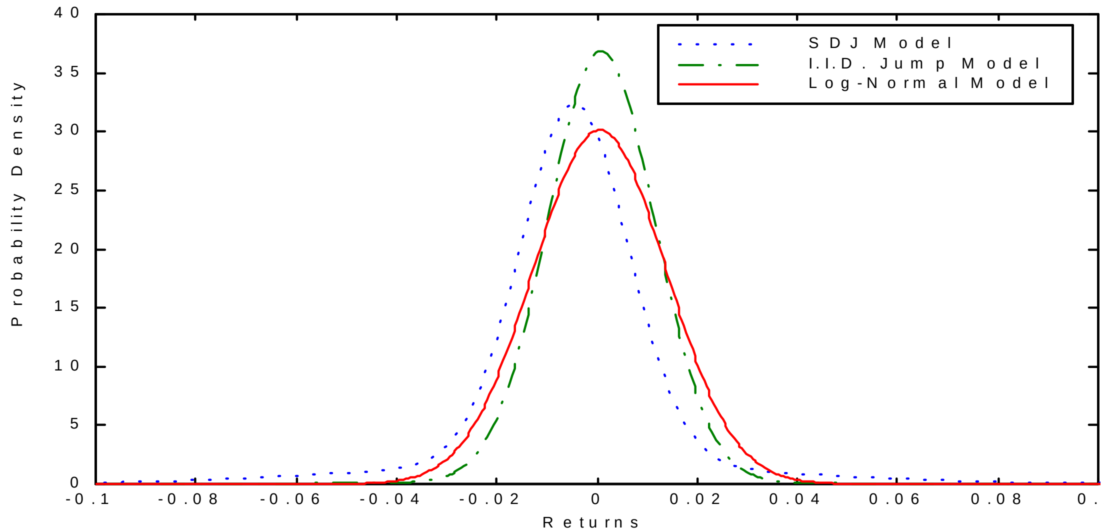

Figure 7: Conditional densities for Merton's i.i.d. jump model, the log-normal model of Black-Scholes and the SDJ model.

For risk managers who have long noted the need to fatten tails of return distributions, especially in periods of market stress, these are important results. Simple i.i.d. models, even those incorporating jumps cannot capture these extremely fat tails. Furthermore, standard GARCH models update volatility very slowly and thus volatility cannot increase quickly enough to generate these fat tails. In periods of market stress, the data is quite informative about what is going to occur next: given that there was a jump today, there will most likely be one tomorrow. Risk managers derivative market participants must be aware of this dependence and adjust their pricing models and capital reserves accordingly.

4.5 Stochastic Volatility and Jumps

The previous results relied heavily on the fact that the volatility of returns, as measured by σ, is constant. This section documents that the predictability results do not depend on this assumption. In a slightly different settings, Eraker, Johannes and Polson (1999) find that both jumps to returns and stochastic volatility are important components, although they assume a simple, state-independent jump process.

In order to address this issue in the state-dependent setting, consider a generalization with a stochastic volatility component, as given in Jacquier, Polson and Rossi (1994):

$$r_{t+1} = \mu_0 + \mu_1 r_t + \sqrt{V_{t+1}}\varepsilon_{t+1} + c_{t+1}J_{t+1}$$
$$\log(V_{t+1}) = \kappa_1 + \kappa_2 \log(V_t) + \sigma\nu_{t+1}$$

where the jump components are the same as in the previous sections. This popular specification has been used in many applications and the stochastic volatility component adds an additional source of non-normalities to the process. An advantage of our modular hierarchical structure is that we can simply add additional steps to our algorithm to accomodate the stochastic volatility. Specifically, we need to estimate three additional parameters (κ_1, κ_2 and σ) as well as the latent volatility. The only difficult step is updating the latent volatility process, as the posterior for the volatility is not standard. We use the algorithm of Shephard and Pitt (1997) to update volatility.

The estimation results are given in Appendix 3 for the S&P 500 (the other series are similar). The results indicate that although the probability of a jump given that there was no jump yesterday and no return (as measured by β_0) falls, the predictability results are robust to the addition of stochastic volatility. Lagged returns and lagged jump times are still strong predictors of jump times. As expected the variance of the jump sizes increases (many of the small jumps are now explained by time-varying volatility) and the mean jump size is smaller: around -2% compared to -0.004. Both of these results are as expected. These results are important as they document that there is a persistence in the data, not well-captured by standard stochastic volatility models.

4.6 The Predictive Power of Implied Volatility

The SDJ model has another important attribute: through the addition of exogenous state variables, it can be used to predict or explain jump times and sizes. To examine this effect, we examine the ability of implied volatility to predict jump times and sizes.

Although it has long been recognized that Black-Scholes implied volatility is an important measure, it is difficult to quantify exactly what it measures in the dynamics of the returns because it fails both as an option pricing and time series model. In spite of this objection, Black-Scholes implied volatility is an exogenous variable that contains important market based information not reflected in just the time series of returns.

Much like asset return volatility, implied volatility has predictable components as argued by Harvey and Whaley (1992). However, the ability of implied volatility to forecast future volatility is at best limited, see for example, Canina and Figlewski (1992), although in a more recent study, Christensen and Prabhala (1998) argue that it has greater ability over longer time horizons. There is also some limited evidence that implied volatility offers predictive power in standard models of time varying volatility such as GARCH or EGARCH. Day and Lewis (1992) find that implied volatilities can explain GARCH and EGARCH volatility using weekly returns from Wednesday to Wednesday, but cannot explain volatility using Friday to Friday returns. Thus there is limited, at best, evidence for implied volatility to explain volatility in standard time varying volatility models such as GARCH and EGARCH.

Our approach is different. We argue that the relevence of implied volatility is as a predictor of jumps. To investigate this we add implied volatility as a regressor in the jump intensity function and in the jump impact function:

$$\begin{aligned} Prob\,[J_{t+1}=1|r_t,J_t] &= \Phi(\beta_0+\beta_1 J_t+\beta_2\,|r_t|+\beta_3\sigma_t^{imp}) \\ c_{t+1} &= \xi_0+\xi_1 r_t+\xi_2\sigma_t^{imp}+\xi_{t+1} \end{aligned} \tag{8}$$

where σ^{imp} is the implied volatility of S&P100 index options, as measured by the Market Volatility Index (VIX) index. The VIX index is a wide measure of market volatility obtained by using a basket of options[10]. We use it to determine if it is able to predict jumps in S&P500 index returns. If jumps are present clearly this fact will be reflected by market participants in the way the price options. Abnormally high implied volatility might imply that the market anticipates that a jump (or a number of them) will occur in the near future.

This specification further allows an analysis of jump risk premia through the jump impact function. Specifically via ξ_2 we can measure if the market, is on average, compensated for taking the jump risk.This mean-variance type intuition implies that participants recieve a premia, through jumps, when the market is extremely volatility, as measured by the VIX index. In order to avoid spuriously concluding that implied volatility predicts jumps when it in actually predicts returns (ARCH in mean effect), we include the volatility as a regressor in the drift. In this case the model considered is given by (8) and the following equation

$$r_{t+1}=\mu_0+\mu_1 r_t+\mu_2\sigma_t^{imp}+\sigma\varepsilon_{t+1}+c_{t+1}J_{t+1}$$

Daily data on the VIX, beginning on January 2, 1986 through Dec. 30, 1998 was obtained from the CBOE website and the time series is given in Figure 8. To analyze the predictability of jump times and sizes with implied volatility, we ran three different versions of the model to discern what impact adding implied volatility had on the results. The results are given in Table 6. Model 1 does not include implied volatility and is used as a reference point, since the time span is different from that used previously. Model 2

[10] The exact construction of the index is detailed on the Chicago Board of Exchange website, www.cboe.com.

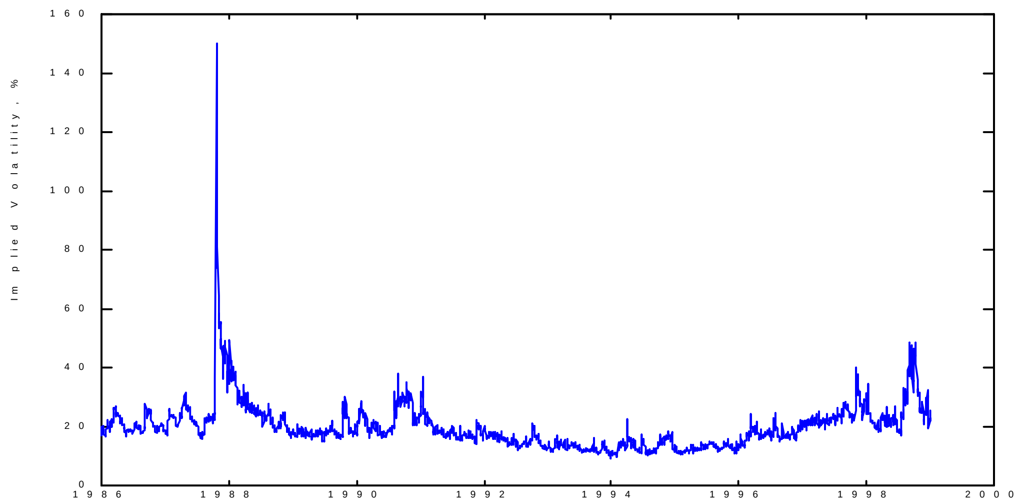

Figure 8: The VIX index, 1986-1999.

excludes implied volatility from the drift or jump impact, while Model 3 is the unrestricted model.

The results indicate a number of interesting facts. First, as indicated by Model's 2 and 3, the VIX is a strong predictor of jump times. The strength of the relationship drives lagged returns to be insignificant (compare the coefficient β_2 across the three models). Also, as measured by the coefficient ξ_2 the lagged VIX index is capable of predicting the size of the jumps. This, at first, may seem counterintuitive, however there is an intuitive explanation. Conisder the Crash in 1987. On Friday before the Crash, the VIX was at 36% which is above its historical average. On Monday, Tuesday and Wednesday, the VIX was 150%, 140% and 73%. Since our model uses lagged implied volatility to predict future jumps, the large returns in the S&P 500 on Tuesday (5%) and Wednesday (8%) coincide with the highest volatility periods. Also note, that as measured by the coefficient μ_2 there is no premium associated with implied volatility coming from the non-jump components.

Table 6. Posterior parameter summaries for the SDJ model incorporating the VIX.

The prior parameters are given the same as those in Appendix 2.

Parameter	Model 1	Model 2	Model 3
μ_0	0.1905	0.1879	0.07589
	(0.03580)	(0.03446)	(0.1443)
μ_1	6.733	6.901	8.265
	(4.586)	(4.614)	(4.761)
μ_2			0.5242
			(0.8141)
ξ_0	-0.005637	-0.002837	-0.01256
	(0.002900)	(0.001663)	(0.004667)
ξ_1	-0.03613	-0.02701	0.01530
	(0.1039)	(0.07429)	(0.08882)
ξ_2			0.02842
			(0.01184)
β_0	-1.999	-4.086	-4.296
	(0.1103)	(0.2715)	(0.2999)
β_1	0.2742	-0.1722	-0.1029
	(0.2579)	(0.2021)	(0.2369)
β_2	41.836	12.631	10.483
	(7.914)	(8.593)	(10.076)
β_3		12.397	12.384
		(1.315)	(1.328)
σ	0.1162	0.1067	0.1142
	(0.002409)	(0.002335)	(0.003154)
τ	0.2872	0.2489	0.2513
	(0.01922)	(0.01158)	(0.01476)
σ_ξ	0.03337	0.02655	0.0287
	(0.002671)	(0.001456)	(0.001879)

$$\begin{aligned} r_{t+1} &= \mu_0 + \mu_1 r_t + \mu_2 \sigma_t^{imp} + \sigma\varepsilon_{t+1} + c_{t+1}J_{t+1} \\ Prob\left[J_{t+1} = 1 | r_t, J_t\right] &= \Phi(\beta_0 + \beta_1 J_t + \beta_2 \left|r_t\right| + \beta_3 \sigma_t^{imp}) \\ c_{t+1} &= \xi_0 + \xi_1 r_t + \xi_2 \sigma_t^{imp} + \xi_{t+1} \end{aligned} \tag{9}$$

5 Conclusions and Extensions

The results given here are important for three reasons. First, we have introduced a model that is capable of capturing state dependencies in the jump process. Second, we have developed a method for estimating SDJ models (and jump-diffusion models) and extracting jump times and sizes even in the case of quite general state dependence. Furthermore, our methodology is computationally efficient. The third, and most important contribution of the paper is empirical: we find and document significant evidence supporting the existence of predictable components in jumps in the time series of equity returns. The jump times are quite predictable on the basis of past returns and jump times. The impact of this type of state-dependency for VaR is also analyzed.

The methods developed for estimating jump diffusion models could be applied to other areas of research, for example interest rate and foreign exchange rate modeling. In both cases, there are a priori justifications for including jumps. In these cases, the state dependency would be via the level of the rate. There is significant evidence from the ARCH/GARCH literature on state dependencies and there is also some evidence of jumps in these markets. An application to emerging markets could also be important given the anecdotal evidence for extreme jump behavior.

Given that we have identified the importance of jumps and some indication of why they occur, a next logical step is to identify other variables that are important for explaining jump times and sizes. To this end, any evidence that explains equity crashes, currency devaluation or interest rate changes could be used in this context to explain jumps. For example, the S&P Future/Spot spread or the time series of implied volatilities for S&P options might be useful in explaining jump times and sizes in our framework. We leave these extensions for further research.

References

Aase, K., (1993). "A Jump/Diffusion Consumption-Based Capital Asset Pricing Model and the Equity Premium Puzzle." *Mathematical Finance*, Vol. 3, No. 2, 65-84.

Aase, K., (1986). "Ruin problems and myopic portfolio optimization in continuous trading." *Stochastic Process. Appl.* 21, no. 2, 213–227.

Aase, K, (1984). Optimum portfolio diversification in a general continuous-time model. Stochastic Process. Appl. 18, no. 1,81–98.

Andersen, T., L. Benzoni and J. Lund. (1998). "Estimating Jump Diffusions for Equity Returns." Working Paper.

Bakshi, G., and D. Madan, (1999). "Crash Discover in Stock and Option Markets,"working paper, University of Maryland.

Bakshi, G., C. Cao and Z. Chen, (1997). "Empirical Performance of Alternative Option Pricing Models" *Journal of Finance*, vol. 52, no 5.

Ball, C. and W. Torous, (1983)."A Simplified Jump Process for Common Stock Returns." Journal of Financial and Quantitative Analysis, vol. 18, no.1, 53-61.

Ball, C. and W.Torous, (1985). "On Jumps in Common Stock Prices and Their Impact on Call Option Pricing." *Journal of Finance*, vol XL, no.1, 155-173.

Bardhan, I. and X. Chao, (1993). "Pricing Options on Securities with Discontinuous Returns." *Stochastic Processes and their Applications*, 48, 123-137.

Bardhan, I. and X. Chao, (1996). "Stochastic multi-agent equilibria in economies with jump-diffusion uncertainty". *J. Econom. Dynam. Control*, 20, no. 1-3, 361–384.

Bates, D. (1996). "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options." *Review of Financial Studies.* v9 n1 Spring 1996, pp. 69-107.

Beckers, S., (1981). "A Note on Estimating the Parameters of the Diffusion-Jump Model." *Journal of Financial and Quantitative Analysis*, vol XVI, no. 1, 127-139

Björk, T., Y. Kabanov and W. Runggaldier, (1997)."Bond Market Structure in the Presence of Marked Point Processes." *Mathematical Finance*, Vol. 7, No. 2, 211-239.

Black, F. and M. Scholes, (1973). "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy*, 81, no. 3, 637-654.

Das, S., (1998). "A Direct Discrete-Time Approach to Poisson-Gaussian Bond Option Pricing in the Heath-Jarrow-Morton Model," *Journal of Economic Dynamics and Control*, v. 23 (3), 333-369.

Duffie, D. and J. Pan, (1997). "An Overview of Value at Risk," *Journal of Derivatives* , Spring 1997, vol. 4, 7-49.

Duffie, D. and J. Pan, (1998)."Analytical Value at Risk with Jump Diffusions," Graduate School of Business, Stanford University, December, 1998.

Duffie, D. and R. Kan, (1996). "A Yield-Factor Model of Interest Rates." *Mathematical Finance.* vol. 6, no. 4, 379-406.

Duffie, D., J. Pan and K. Singleton, (1998). "Transform Analysis and Option Pricing for Affine Jump-Diffusions." Working Paper, Graduate School of Business, Stanford University.

Gelfand, A. and A. Smith, (1990). "Sampling Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association*, 85,398-409.

Hamilton, J., (1990) "Analysis of Time Series Subject to Changes in Regime." *Journal of Econometrics*; v45 n1-2,. pp. 39-70.

Honore, P., (1998). "Pitfalls in Estimating Jump Diffusion Models." Working Paper, University of Aarhus.

Jacquier, E., N. Polson and P. Rossi, (1994)."Bayesian Analysis of Stochastic Volatility Models" *Review of Economics and Business Statistics*," with discussion, 12, 4, 371-389.

Jarrow, R. and E. R. Rosenfeld, (1984). "Jump Risks and the Intertemporal Capital Asset Pricing Model." *Journal of Business*, vol. 57, no. 3, 337-351.

Jiang, G., (1998). "Jump-Diffusion Models of Exchange Rate Dynamics: Estimation via Indirect Inference." working paper, Tilberg University.

Jeanblanc-Picqué, M. and P. Pontier. (1990). "Optimal portfolio for a small investor in a market model with discontinuous prices." *Appl. Math. Optim.* 22, no. 3, 287–310.

Jorion, P., (1988). "On Jump Processes in the Foreign Exchange and Stock Markets." *Review of Financial Studies*, vol. 1, no. 4, 427-445.

Kiefer, N., (1978). "Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model." *Econometrica*, vol. 46, no.2, 427-434.

Lindgren, G., (1978). "Markov Regime Models for Mixed Distributions and Switching Regressions." *Scandinavian Journal of Statistics*, no.5, 81-91.

Merton, R.. (1971). "Optimum consumption and portfolio rules in a continuous-time model." *J. Econom. Theory* 3, no. 4, 373–413.

Merton, R., (1976a). "Option Pricing when the Underlying Stock Returns are Discontinuous." *Journal of Financial Economics*, 3, 1235-144.

Merton, R., (1976b). "The Impact on Option Pricing of Specification in the Underlying Stock Price Returns." *Journal of Finance*, vol. 31, no.2, 333-350.

Naik, V. and L. Moon, (1990). "General Equilibrium Pricing of Options on the Market Portfolio with Discontinuous Returns." *Review of Financial Studies*; v.3 n.4 , pp. 493-521.

Platen, E. and R. Rebolledo, (1985). "Weak Convergence of Semimartinglaes and Discretization Methods." *Stochastic Processes and their Applications*, 20,41-58.

Press, S. J., (1967)."A Compound Model for Security Prices." *Journal of Business*, 40, 317-335.

Schwert, G. W., (1990). "Stock Market Volatility." *Review of Financial Studies*, v3, n1, 77-102.

Venkataraman, S. (1997). "Value at risk for a mixture of normal distributions: The use of quasi-Bayesian techniques." *Federal Reserve Bank of Chicago Economic Perspectives;* v.21 n.2 March-April 1997, 2-13.

Zangari, P. (1996). "An improved methodology of measuring VaR that allows for more realistic model of financial return tail distributions." *Risk Metrics^{TM} Monitor;* Second quarter, 1996, 7-25.

Zangari, P. (1997). "What Risk Managers should know about mean reversion and jumps in prices." *Risk Metrics^{TM} Monitor;* Fourth quarter, 1997, 12-36.

Appendix 1: Relation to Jump-diffusion models

For derivative pricing purposes, jump-diffusion models offer increased flexibility and an ability to account for many of the systematic mispricing (cross-section and term structure of implied volatility) that occurs in models with continuous state variables[11]. This section outlines the relationship of the SDJ model and the jump-diffusion model.

For simplicity, consider the case of state independent jump arrivals (suppressing parameter dependence):

$$X_{(t+1)\Delta} = X_{t\Delta} + \mu(X_{t\Delta})\Delta + \sigma(X_{t\Delta})\varepsilon_{t+1} + c(X_{t\Delta}, \xi_{t+1})J_{(t+1)\Delta}. \tag{10}$$

where $\xi_{(t+1)\Delta} \sim \Pi$ and $Prob\left[J_{(t+1)\Delta} = 1\right] = \lambda\Delta$. Weak convergence arguments (see Skorohod (1965) or Platen and Rebolledo (1985)), imply that under standard regularity conditions, the limit of (10) as $\Delta \downarrow 0$, is given as the solution of the following stochastic differential equation:

$$dX_t = \mu(X_t)dt + \sigma\left(X_t\right)dW_t + dJ_t = \int_{\Xi} c(X_{t-}, \xi)p(d\xi, dt) \tag{11}$$

where $p(d\xi, dt)$ is a Poisson random measure, W_t is a standard Brownian motion, $\xi \in \Xi$ and $\xi \sim \Pi$. The Poisson random measure sums jumps of various sizes:

$$\int_0^t \int_{\Psi} c(X_{t-}, \xi)p(d\xi, dt) = \sum_{j=1}^{N_t} c(X_{\tau_j -}, \xi_{t_{j}-})$$

The time-discretization is similar in form to the Euler approximation for diffusion process.

The accuracy of this approximation depends on two factors: the length of the interval and the frequency of the jumps. In our applications the interval between observations is one day and it is assumed that jumps are infrequent events in equity index returns (the Brownian motion captures day-to-day fluctuations), it is not unreasonable to assume that at most one jump can occur in an interval since the probability of two or more is negligible. From a practical standpoint, if a goal of inference is inferring jump time and sizes, it is clearly more difficult to identify them if more than one could occur between observations.

The probit specification with lagged jump time included has no continuous time analog. However, in the case where $\beta_1 = 0$, even in the presence of state dependencies, a continuous time analog exists. To see this (ignoring parameter dependence), the following limit must exist:

$$\lim_{\Delta \downarrow 0} \frac{1}{\Delta} Prob\left[J_{(t+1)\Delta} = 1 | X_{t\Delta} = x\right] = \lambda(x)$$

[11] See Bates (1996) and Bakshi, Cao and Chen (1997) or Das and Sundaram (1999) for more detailed analysis of these problems.

where λ is the intensity function of the jump process. For the probit specification, this limit cannot be computed in closed form as the intensity function is not absolutely continuous with respect to Lebesgue measure, however it does exist[12].

[12] However, a re-parameterization does provide the proper limit. Define α^* and β^* for each x from the following equation: $\Phi_\Delta(\beta_0 + \beta_1 x) = \frac{1}{\Delta}\left[\Phi(\Delta(\alpha^* + \beta^* x)) - \Phi(0)\right]$. As the normal CDF is monotonic, a solution exists. L'Hospital's rule implies that for each x, $\frac{1}{\Delta}\left[\Phi(\Delta(\alpha^* + \beta^* x)) - \Phi(0)\right] \underset{\Delta \downarrow 0}{\longrightarrow} \alpha^* + \beta^* x = \lambda(x)$

Appendix 2: MCMC Methods

MCMC methods are a computationally efficient method for creating samples from high dimensional distributions. The idea of MCMC is simple. In most cases, the full posterior given the data, generically $p\left(\Theta|\underline{X}\right)$, is fully characterized by its conditionals, $p\left(\Theta_j|\Theta_k : k \neq j, \underline{X}\right)$. In our case, we have $\pi(\theta) \triangleq p(\Theta|\underline{X}) = p\left(\alpha, \underline{J}, \underline{\xi}|\underline{X}\right)$. This characterization allows us to create random samples from $p\left(\Theta|\underline{X}\right)$ by sampling, in an iterative fashion, from $p\left(\Theta_j|\Theta_k : k \neq j, \underline{X}\right)$. The MCMC algorithm formalizes how to do the sampling and produces a sequence of draws whose distribution converges to that of $p\left(\Theta|\underline{X}\right)$.

Let Θ denote the state space for the Markov chain, $\Theta^{(i)}$ the position of the whole chain at iteration (i) and $\Theta_j^{(i)}$ the position of the j^{th} component of the Markov chain at iteration (i). The MCMC algorithm draws from each of the conditional posteriors in turn, updating the state space of the chain after each draw. If all of the parameter/latent variables can be drawn directly from their posteriors, the chain is referred to as the Gibbs sampler. It uses the conditional posteriors $p(\Theta_k^{(i+1)}|\Theta_j^{(i)} : j \neq k, \underline{X})$ to make the transition $\Theta^{(i)} \rightarrow \Theta^{(i+1)}$ as explained in Gelfand and Smith (1990).

Sampling from some of the conditional posteriors is often difficult or computationally intractable. In these cases a Metropolis algorithm[13] is used to generate draws. A Metropolis step generates a candidate draw from a kernel, $Q(\Theta^{(i)}, \Theta^{(i+1)})$, known "blanket." and the candidate draw is accepted with probability

$$\min\left\{\frac{\pi(\Theta^{(i+1)})/Q(\Theta^{(i)}, \Theta^{(i+1)})}{\pi(\Theta^{(i)})/Q(\Theta^{(i+1)}, \Theta^{(i)})}, 1\right\} \tag{12}$$

otherwise $\Theta^{(i+1)} = \Theta^{(i)}$.

The algorithm provides samples from $p(\Theta|\underline{X})$ by generating a sequence of random variables $\Theta^{(0)}, \Theta^{(1)}, \Theta^{(2)}$,...that forms a Markov chain. The convergence of the chain to its target is guaranteed via the following proposition:

Proposition (Tierney (1994)): *Let $(\Theta^{(0)}, \Theta^{(1)}, ...)$ be an irreducible Markov chain on the state space Θ with invariant distribution π. Let $\overline{F}_G = \frac{1}{G}\sum_{g=1}^{G} F(\Theta^{(g)})$. Then for $\pi-$almost all $\theta^{(0)}$*

$$\Pr\left[\overline{F}_G \longrightarrow \int F(\theta)\pi(\theta)d\theta|\Theta^{(0)} = \theta^{(0)}\right] = 1$$

if $\int F(\theta)\pi(\theta)d\theta < \infty$ where $\Theta^{(0)} = \theta^{(0)}$ is the starting position of the chain.

[13] See Hastings (1970) for a description of the Metropolis algorithm and Roberts and Smith (1994) for a discussion and conditions guaranteeing the convergence of the Metropolis algorithm.

Appendix 3: Conditional Posteriors and priors

This appendix derives the conditional posteriors given in the text. The posteriors combine the prior distribution with the likelihood via Bayes formula. The relevant likelihood is given by

$$\mathcal{L}(\xi_0, \mu_0, \tau^2, \sigma^2, \underline{J}|\underline{R}) = p(\underline{R}|\mu_0, \xi_0, \sigma^2, \tau^2, \underline{J})$$

and up to proportionality we have

$$\mathcal{L}(\xi_0, \mu, \tau^2, \underline{J}, \sigma^2|\underline{R}) \propto \frac{1}{\sigma^T} \prod_{t=1}^{T} \left(\frac{1}{\sqrt{\Delta + \tau^2 J_{t+1}}} \exp\left(-\frac{1}{2} \frac{(r_{t+1} - \mu\Delta - \xi_0 J_{t+1})^2}{\sigma^2(\Delta + \tau^2 J_{t+1})} \right) \right).$$

Note the advantage of decoupling the normally distributed jump sizes and treating the jump times as latent variables, our likelihood is not a discrete mixture of normal distributions.

We now specify the prior distributions for the parameters $(\mu, \xi_0, \sigma^2, \tau^2, \lambda)$ in Merton's model. We assume that the variance parameters, the drift parameters and the jump frequency are independent, thus: $p(\mu, \xi_0, \sigma^2, \tau^2, \lambda) = p(\mu, \xi_0)p(\sigma^2)p(\tau^2)p(\lambda)$. For the individual components, we use standard conjugate priors:

$$\begin{aligned} \mu_0, \xi_0 &\sim \mathcal{N}(a_0, b_0), \\ \sigma^2 &\sim \mathcal{IG}(c_0, d_0), \\ \tau^2 &\sim \mathcal{IG}(e_0, f_0), \\ \lambda &\sim \mathcal{B}(\alpha_0, \alpha_1) \end{aligned}$$

where $\mathcal{B}$ and $\mathcal{IG}$ denote the Beta and inverse Gamma distributions, respectively. Here $(a_0, b_0, c_0, d_0, e_0, f_0, \alpha_0, \alpha_1)$ are all pre-specified. By varying the prior parameters we can easily perform a sensitivity analysis of our estimation procedure. Given our conjugate priors, the posterior conditionals for μ, ξ_0, λ and σ^2 are standard. μ and ξ_0 are Normally distributed, λ has a Beta posterior and σ^2 has an Inverse Gamma posterior. Generating samples from these distributions is straightforward.

The posterior distributions of the τ^2 and the jump times are given as:

$$\begin{aligned} \tau^2|\mu, \xi_0, \sigma^2, \underline{J}, \underline{R} &\propto \left(\Delta + \tau^2\right)^{-\Sigma J_{t+1}/2} \left(\tau^2\right)^{-e_0-1} \exp\left(-G_{t+1} - \frac{1}{\tau^2 f_0} \right) \\ J_{t+1}|\mu, \xi_0, \sigma^2, \tau^2, \underline{R} &\sim Ber(p_t) \\ \xi_{t+1}|\mu, \xi_0, \sigma_\xi^2, J_{t+1}, r_{t+1} &\sim \mathcal{N}(B_{t+1}b_{t+1}, B_{t+1}) \end{aligned}$$

where

$$\begin{aligned}
G_{t+1} &= \sum_{t:J_{t+1}=1} \frac{(r_{t+1} - \mu\Delta - \xi_0 J_{t+1})^2}{2\sigma^2(\Delta + \tau^2)} \\
p_{t+1} &= p(J_{t+1} = 1|\mu, \xi_0, \sigma^2, \tau^2, \underline{R}) = (1 + O_{t+1})^{-1} \\
\ln(O_{t+1}) &= \ln\left(\frac{\lambda^{\frac{3}{2}}}{(1-\lambda)\sqrt{\Delta + \tau^2}}\right) - \frac{(r_{t+1} - \mu\Delta - \xi_0)^2}{\sigma^2(\Delta + \tau^2)} + \frac{(r_{t+1} - \mu\Delta)^2}{\sigma^2\Delta} \\
B_{t+1}^{-1} &= \frac{J_{t+1}}{\sigma^2\Delta} + \frac{1}{\sigma_\xi^2},\ b_{t+1} = \frac{J_{t+1}(r_{t+1} - \mu\Delta)}{\sigma^2\Delta} + \frac{1}{\sigma_\xi^2}
\end{aligned}$$

The derivation of the posterior for τ^2 is

$$\begin{aligned}
p(\tau^2|\sigma^2, \mu, \xi_0, \underline{J}, \lambda, \underline{R}) &\propto p(\tau^2)p(\underline{R}|\tau^2, \sigma^2, \mu, \xi_0, \underline{J}) \\
&\propto \left(\frac{1}{\tau^2}\right)^{e_0+1} \exp\left(-\frac{1}{\tau^2 f_0}\right) * \\
&\quad \prod_{t=1}^{T} \frac{1}{\sqrt{\Delta + \tau^2 J_{t+1}}} \exp\left(-\frac{(r_{t+1} - \mu\Delta - \xi_0 J_{t+1})^2}{2\sigma^2(\Delta + \tau^2 J_{t+1})}\right) \\
&\propto \left(\frac{1}{\tau^2}\right)^{e_0+1} (\Delta + \tau^2)^{\sum_t J_{t+1}} \\
&\quad \exp(-\sum_{t:J_{t+1}=1} \left(\frac{(r_{t+1} - \mu\Delta - \xi_0)^2}{2\sigma^2(\Delta + \tau^2)} - \frac{1}{\tau^2 f_0}\right))
\end{aligned}$$

Note that the summations in the last expressions involve only terms with $J_{t+1} = 1$. When $\mathrm{J}_t = 0$, the relevant portion of the posterior is independent of τ^2 and is thus absorbed into the proportionality constant. Intuitively, we only learn about the parameter τ^2 when the process jumps.

The posterior for the jump times is given by the odds ratio of $\{J_t = 1\}$ versus $\{J_t = 0\}$ which is calculated by dividing the posteriors for the two cases, using Bayes rule for discrete distributions and taking the log to simplify the expression.

$$\begin{aligned}
\frac{\Pr(J_{t+1} = 1|\sigma^2, \mu, \xi_0, \tau, \underline{R})}{\Pr(J_{t+1} = 0|\sigma^2, \mu, \xi_0, \tau, \underline{R})} &= \frac{\Pr(J_{t+1} = 1|\sigma^2, \mu, \xi_0, \tau, r_{t+1})}{\Pr(J_{t+1} = 0|\sigma^2, \mu, \xi_0, \tau, r_{t+1})} \\
&= \frac{\Pr(J_{t+1} = 1|\lambda) \Pr(r_{t+1}|J_{t+1} = 1, \mu, \sigma^2, \tau^2, \xi_0)}{\Pr(J_{t+1} = 0|\lambda) \Pr(r_{t+1}|J_{t+1} = 0, \mu, \sigma^2, \tau^2, \xi_0)} \\
&= \frac{\lambda}{1-\lambda} \frac{\left(\frac{1}{\sigma^2(\Delta+\tau^2)}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\frac{(r_{t+1}-\mu\Delta-\xi_0)^2}{\sigma^2(\Delta+\tau^2)}\right)\right)}{\left(\frac{1}{\sigma^2\Delta}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\frac{(r_{t+1}-\mu\Delta)^2}{\sigma^2\Delta}\right)\right)}.
\end{aligned}$$

The conditional posterior for the jump size ξ_t given the parameters is found by using the conditional independence between current jumps and all other observations, given to days observation, $p\left(\xi_t|\mu,\xi_0,\sigma_\xi^2,\underline{J},,\underline{R}\right)=p\left(\xi_t|J_t,\mu,\xi_0,\sigma_\xi^2,r_{t+1}\right)$, reversing the conditioning and completing the square in the following expression:

$$p\left(\xi_{t+1}|r_{t+1},J_{t+1},\mu,\xi_0,\sigma_\xi^2\right)\propto p(r_{t+1}|\xi_{t+1},J_{t+1},\mu,\xi_0,\sigma_\xi^2)p(\xi_{t+1}|\sigma_\xi^2,\xi_0)$$

where

$$\begin{aligned} p(r_{t+1}|\xi_{t+1},J_{t+1},\mu,\xi_0,\sigma_\xi^2) &= \exp\left[\frac{1}{2\sigma_\xi^2\Delta}\left(r_{t+1}-\mu\Delta-\xi_{t+1}J_{t+1}\right)^2\right] \\ p(\xi_{t+1}|\sigma_\xi^2,\xi_0) &= \exp\left[\frac{1}{2\sigma_\xi^2}\left(\xi_{t+1}-\xi_0\right)^2\right]. \end{aligned}$$

We now briefly describe how the researcher sets these parameters. In our empirical work, we use proper but highly diffuse priors on $(\gamma,\xi_0,\sigma^2,\lambda)$. For τ^2 and the choice of (e_0,f_0) we argue as follows. As $\tau^2=\sigma_\xi^2/\sigma^2$, the prior parameters, (e_0,f_0), are important because they permit us to separate the variance of the jump sizes from the diffusion term. Non-informative priors on τ^2 (for example, the diffuse prior, $\tau^{-2},\tau>0$), lead to improper posteriors and thus we must be slightly informative. This requirement is similar to the truncation of the parameter space used to operationalize MLE.

The actual prior parameter distribution we use are given by:

$$\begin{aligned} \mu_0,\xi_0 &\sim \mathcal{N}(0,1000), \\ \sigma^2 &\sim \mathcal{IG}(3,25), \\ \tau^2 &\sim \mathcal{IG}(3,2.5), \\ \lambda &\sim \mathcal{B}(10,100) \end{aligned}$$

The density function for τ^2 is $p(\tau^2|e_0,f_0)\propto\left(\frac{1}{\tau^2}\right)^{e_0+1}\exp\left(-\frac{1}{\tau^2 f_0}\right)$. As an example $(e_0,f_0)=(3.0,2.5)$ implies that $Prob\left[\tau^2\in(0,0.6)\right]=.97$. For values outside this range $(\tau^2>0.6)$, we simulated series holding jump frequency constant. The sizes of the resulting jumps were too large (many of the jumps were larger in absolute value than 10%). Thus our priors are quite reasonable, and, in fact, not very informative. Sensitivity analysis confirmed this.

Given these posteriors, we require our algorithm to create the following Monte Carlo sample $\left\{\mu^{(g)},\xi_0^{(g)},\sigma^{2^{(g)}},\tau^{2^{(g)}},\lambda^{(g)},\underline{J}^{(g)}\right\}_{g=1}^{N}$. Note that at each iteration of our algorithm we draw the entire vector of jump times, $\underline{J}$. Had we not integrated the jump sizes out of the hierarchy, we would also have to draw the entire vector of jump sizes at each iteration, increasing the size of our state space from T+5 to 2T+5 dimensions. As mentioned in section (2), the inputs required to create these samples are the posterior conditional distributions of each parameter and the jump times given the observed data and the other parameters.

Given these conditional distributions, our MCMC algorithm iteratively updates the parameters. We draw $\left(\mu^{(g)}, \xi_0\right)$ and $(\underline{J})$ in blocks and σ^2, τ^2 and λ individually. To see how the algorithm works, suppose that we have the output from the first (g) iterations of our algorithm. To update, for example, σ^2, we compute, $\sigma^{2(g+1)}$ by drawing it from its posterior distribution conditional on the values of the other parameters from the previous draw and the observed data:

$$\sigma^{2(g+1)} \sim \mathcal{IG}\left(c_0 + \frac{T}{2}, \left(\frac{1}{d_0} + \frac{1}{2}\sum_{t=0}^{T} \frac{\left(r_{t+1} - \mu^{(g)}\Delta - \xi_0^{(g)} J_{t+1}^{(g)}\right)^2}{\Delta + \tau^{2(g)} J_{t+1}^{(g)}}\right)^{-1}\right)$$

Given the posterior for the jump times, J_{t+1}, it is easy to update the jump times in the algorithm. For τ^2, however, the posterior distribution is nonstandard and we therefore use a Metropolis algorithm to update τ^2. We choose the candidate density to closely approximate $p(\tau^2|\mu, \xi_0, \sigma^2, \underline{J}, \underline{R})$.

The Proposition from Tierney (1994) implies that the output of this algorithm converges in distribution to the true posterior. In order to negate any affects of the starting values, we discard the first 2000 draws from our algorithm, the "burn in" period. Thus, our algorithm creates samples that are arbitrarily close to the joint posterior distribution, $p(\mu, \xi_0, \lambda, \tau^2, \sigma^2, \underline{J}|\underline{R})$. Given the output of the algorithm, the next section describes how to use the output to solve the parametric inference and jump time and size extraction problems.

The state dependent model can easily be incorporated into the MCMC algorithm using the structure built for Merton's model. To do so, we again need to identify the conditional posteriors. For the prior structure, we assume that $\xi_0, \mu_0, \mu_1, \alpha$ and β are independent and choose conjugate priors which implies that the posteriors for ξ_0, μ_0, μ_1 are normal and the conditional posterior for τ^2 is unchanged. For the additional regression parameters (parameters in the drift, jump impact and probit), we specify the priors are mean zero and have variance equal to 1000. These priors are extremely uninformative.

Simulation of Z_{t+1} is simple as it is a normal draw. Assuming that $p(\beta_0, \beta_1, \beta_2) \sim \mathcal{N}\left(\beta_0, \sigma_\beta^2\right)$ implies that (see Albert and Chib (1993) $\beta_0, \beta_1, \beta_2|\underline{Z}, \underline{R} \sim \mathcal{N}\left(\mu_\beta, \sigma_\beta^2\right)$ where $\mu_\beta = \left(\left(\sigma_\beta^2\right)^{-1} + \underline{R}'\underline{R}\right)^{-1}\left(\left(\sigma_\beta^2\right)^{-1}\beta_0, \underline{R}'\underline{Z}\right)$ and $\sigma_\beta^2 = \left(\left(\sigma_\beta^2\right)^{-1} + \underline{R}'\underline{R}\right)$. The only existing step in our algorithm that is affected is how we draw J_{t+1} where we need only substitute the probit probabilities for λ. Our algorithm now generates samples of

$\{\mu_0, \mu_1, \xi_0, \xi_1, \sigma^2, \tau^2, \underline{J}, \beta_0, \beta_1, \beta_2, \underline{Z}\}$.

Appendix 3: Parameter estimates in the state-dependent stochastic volatility model

For comparison purposes, we estimate the model for the full sample using S&P 500 returns. This implies that the results should be compared to the first column of Table 4.

TABLE 8. PARAMETER ESTIMATES FOR THE STATE-DEPENDENT JUMP MODEL WITH STOCHASTIC VOLATILITY

	S&P 500		
μ_0	0.0006872	ξ_1	-0.2825
	(0.0001126)		(0.2162)
μ_1	0.04088	σ_ξ	0.04582
	(0.01663)		(0.009259)
κ_1	-0.1631	β_1	-2.721
	(0.03942)		(0.1250)
κ_1	-0.01658	β_2	30.01
	(0.004006)		(2.996)
σ	0.1181	β_3	0.8246
	(0.0103)		(0.4668)
ξ_0	-0.02747		
	(0.01458)		

$$r_{t+1} = \mu_0 + \mu_1 r_t + \sqrt{V_{t+1}}\varepsilon_{t+1} + c_{t+1} J_{t+1}$$
$$Prob\left[J_{t+1} = 1 | r_t, J_t\right] = \Phi\left(\beta_0 + \beta_1 J_t + \beta_2 |r_t|\right)$$
$$c_{t+1} = \xi_0 + \xi_1 r_t + \xi_{t+1},\ \ \xi_{t+1} \sim \mathcal{N}\left(0, \sigma_\xi^2\right)$$
$$\log(V_{t+1}) = \kappa_1 + \kappa_2 \log(V_t) + \sigma \nu_{t+1}$$