U.S. Stock Market Crash Risk, 1926 - 2006

David S. Bates
University of Iowa and the National Bureau of Economic Research

October 9, 2008
Preliminary and incomplete; please do not cite.

Abstract
This paper applies the Bates (RFS, 2006) methodology to the problem of estimating and filtering time-changed Lévy processes, using daily data on stock market excess returns over 1926-2006. In contrast to density-based filtration approaches, the methodology recursively updates the associated conditional characteristic functions of the latent variables. The paper examines how well time-changed Lévy specifications capture stochastic volatility, the “leverage” effect, and the substantial outliers occasionally observed in stock market returns. The paper also finds that the autocorrelation of stock market excess returns varies substantially over time, necessitating an additional latent variable when analyzing historical data on stock market returns.

Henry B. Tippie College of Business
University of Iowa
Iowa City, IA 52242-1000

Tel.: (319) 353-2288
Fax: (319) 335-3690
email: david-bates@uiowa.edu
Web: www.biz.uiowa.edu/faculty/dbates
What is the risk of stock market crashes? Answering this question is complicated by two features of stock market returns: the fact that conditional volatility evolves over time, and the fat-tailed nature of daily stock market returns. Each issue affects the other. What we identify as outliers depends upon that day’s assessment of conditional volatility. Conversely, our estimates of current volatility from past returns can be disproportionately affected by outliers such as the 1987 crash. In standard GARCH specifications, for instance, a 10% daily change in the stock market has 100 times the impact on conditional variance revisions of a more typical 1% move.

This paper explores whether recently proposed continuous-time specifications of time-changed Lévy processes are a useful way to capture the twin properties of stochastic volatility and fat tails. The use of Lévy processes to capture outliers dates back at least to Mandelbrot’s (1963) use of the stable Paretian distribution, and there have been many others proposed; e.g., Merton’s (1976) jump-diffusion, Madan and Seneta’s (1990) variance gamma; Eberlein, Keller and Prause’s (1998) hyperbolic Lévy; and Carr, Madan, Geman and Yor’s (2002) CGMY process. As all of these distributions assume identical and independently distributed returns, however, they are unable to capture stochastic volatility.

More recently, Carr, Geman, Madan and Yor (2003) and Carr and Wu (2004) have proposed combining Lévy processes with a subordinated time process. The idea of randomizing time dates back to at least to Clark (1973). Its appeal in conjunction with Lévy processes reflects the increasing focus in finance – especially in option pricing – on representing probability distributions by their associated characteristic functions. Lévy processes have log characteristic functions that are linear in time. If the time randomization depends on underlying variables that have an analytic conditional characteristic function, the resulting conditional characteristic function of time-changed Lévy processes is also analytic. Conditional probability densities, distributions, and option prices can then be numerically computed by Fourier inversion of simple functional transforms of this characteristic function.
Thus far, empirical research on the relevance of time-changed Lévy processes for stock market returns has largely been limited to the special cases of time-changed versions of Brownian motion and Merton’s (1976) jump-diffusion. Furthermore, there has been virtually no estimation of newly proposed time-changed Lévy processes solely from time series data. Papers such as Carr et al (2003) and Carr and Wu (2004) have relied on option pricing evidence to provide empirical support for their approach, rather than providing direct time series evidence. The reliance on options data is understandable. Since the state variables driving the time randomization are not directly observable, time-changed Lévy processes are hidden Markov models – a challenging problem in time series econometrics. Using option prices potentially identifies realizations of those latent state variables (under the assumption of correct model specification), converting the estimation problem into the substantially more tractable problem of estimating state space models with observable state variables.

This paper provides direct time series estimates of some proposed time-changed Lévy processes, using the Bates (2006) approximate maximum likelihood (AML) methodology. AML is a filtration methodology that recursively updates conditional characteristic functions of latent variables over time given observed data. Filtered estimates of the latent variables are directly provided as a by-product, given the close link between moments and characteristic functions. The primary focus of the paper’s estimates is on the time-changed CGMY process, which nests various other processes as special cases. The approach will also be compared to the time-changed jump-diffusions previously estimated in Bates (2006).

The central issue in the study is the one stated at the beginning: what best describes the distribution of extreme stock market movements? Such events are perforce relatively rare. For instance, the ~20% stock market crash of October 19, 1987 was the only daily stock market movement in the post-World War II era to exceed 10% in magnitude. By contrast, there were seven such movements over 1929-32. Consequently, I use an extended data series of excess value-weighted stock market returns over 1926-2006, to increase the number of observed outliers.

---

1Li, Wells and Yu (2006) use MCMC methods to estimate some models in which Lévy shocks are added to various stochastic volatility models. However, the additional Lévy shocks are i.i.d., rather than time-changed.
A drawback of using an extended data set is the possibility that the data generating process may not be stable over time. Indeed, this paper identifies one such instability, in the autocorrelation of daily stock market returns. The instability is addressed directly, by treating the autocorrelation as another latent state variable to be estimated from observed stock market returns. Autocorrelation estimates appear to be nonstationary, and peaked at the extraordinarily high level of 35% in 1971, before trending downwards to the near-zero values observed since 2002.

Overall, the time-changed CGMY process is found to be a parsimonious alternative to the Bates (2006) approach of using finite-activity stochastic-intensity jumps drawn from a mixture of normals, although the fits of the two approaches are not dramatically different. Interestingly, one cannot reject the hypothesis that stock market crash risk is adequately captured by a time-changed version of the Carr-Wu (2003) log-stable process. That model’s implications for upside risk, however, are strongly rejected, with the model severely underpredicting the frequency of large positive outliers.

Section I of the paper progressively builds up the time series model used in estimation. Section I.1 discusses basic Lévy processes and describes the processes considered in this paper. Section I.2 discusses time changes, the equivalence with stochastic volatility, and further modifications of the data generating process to capture leverage effects and time-varying autocorrelations. Section I.3 describes how the model is estimated, using the Bates (2006) AML estimation methodology for hidden Markov models.

Section II describes the data on excess stock market returns over 1926-2006, and presents the estimates of parameters and filtered estimates of latent autocorrelations and volatility. Section III presents option pricing implications, while Section IV concludes.
I. Time-changed Lévy processes

I.1 Lévy processes

A Lévy process $L(t)$ is an infinitely divisible stochastic process; i.e., one that has independent and identically distributed increments over non-overlapping time intervals of equal length. The Lévy processes most commonly used in finance have been Brownian motion and the jump-diffusion process of Merton (1976), but there are many others. All Lévy processes other than Brownian motions can be viewed as extensions of jump processes. These processes are characterized by their Lévy density $k(x)$, which gives the intensity (or frequency) of jumps of size $x$. Alternatively and equivalently, Lévy processes can be described by their generalized Fourier transform

$$F(u) = E e^{uL(t)} = \exp[t f_{al}(u)], u \in D_u \subset \mathbb{C}$$ (1)

where $u$ is a complex-valued element of the set $D_u$ for which (1) is well-defined. If $\Phi$ is real, $F(i\Phi)$ is the characteristic function of $L(t)$, while $tf_{al}(\Phi)$ is the cumulant generating function of $L(t)$. Its linearity in time follows from the fact that Lévy processes have i.i.d. increments. Following Wu (2006), the function $f_{al}(u)$ will be called the cumulant exponent of $L(t)$.

The Lévy-Khintchine formula gives the mapping between jump intensities $k(x)$ and the cumulant exponent for arbitrary $u \in D_u$. Lévy processes in finance are typically specified for the log asset price, and then exponentiated: $S(t) = \exp[L(t)]$. For such specifications, it is convenient to write the Lévy-Khintchine formula in the form

$$f_{al}(u) = u\mu + \int_{\mathbb{R} - \{0\}} [e^{ux} - 1 - u(e^x - 1)]k(x)dx,$$ (2)

where $\mu = f_{al}(1)$ is the continuously-compounded expected return on the asset:

$$ES(t) = E e^{L(t)} = e^{f_{al}(1)t} = e^{\mu t}.$$ (3)

Intuitively, Lévy processes can be thought of as a drift term plus an infinite sum $\int L_x dx$ of independent point processes, each drift-adjusted to make $\exp[L_x(t)]$ a martingale:

$$dL_x = xdN_x - (e^x - 1)k(x)dt,$$ (4)

---

Carr et al (2003) call $f_{al}(i\Phi)$ the “unit time log characteristic function.” Bertoin (1996) uses the characteristic exponent, which takes the form $\Psi(\Phi) = -f_{al}(i\Phi)$. 

where $N_x$ is an integer-valued Poisson counter with intensity $k(x)$ that counts the occurrence of jumps of fixed size $x$. The log characteristic function of a sum of independent point processes is the sum of the log characteristic functions of the point processes, yielding equation (2).

As discussed in Carr et al (2002), Lévy processes are finite-activity if $\int k(x) \, dx < \infty$, and infinite-activity otherwise. Finite-activity jumps imply there is a non-zero probability that no jumps will be observed within a given time interval. Lévy processes are finite-variation if $\int \min(x^2, 1) \, k(x) \, dx < \infty$, and infinite-variation otherwise. An infinite-variation process has sample paths of infinite length – a property also of Brownian motion. All Lévy processes must have finite $\int \min(x^2, 1) \, k(x) \, dx$, in order to be well-behaved, but need not have finite variance $\int x^2 \, k(x) \, dx$ – the stable distribution being an counterexample. A priori, all financial prices must be finite-activity processes, since price changes reflect a finite (but large) number of market transactions. However, finite-activity processes can be well approximated by infinite-activity processes, and vice versa; e.g., the Cox, Ross and Rubinstein (1979) finite-activity binomial approximation to Brownian motion. Activity and variation will therefore be treated as empirical specification issues concerned with identifying which functional form $k(x)$ best fits daily stock market returns.

I will consider two particular underlying Lévy processes for log asset prices. The first is Merton (1976)’s combination of a Brownian motion plus finite-activity normally distributed jumps:

$$d \ln S_t = \mu \, dt + (\sigma \, dW_t - \frac{1}{2} \sigma^2 \, dt) + (\gamma \, dN_t - \lambda \, k \, dt)$$

where $W_t$ is a Wiener process,

$N_t$ is a Poisson counter with intensity $\lambda$.

$\gamma \sim N(\gamma^*, \delta^2)$ is the normally distributed jump conditional upon a jump occurring, and

$\bar{k} = e^{\gamma^* + \frac{3}{2} \delta^2} - 1$ is the expected percentage jump size conditional upon a jump.

The associated intensity of jumps of size $x$ is

$$k(x) = \frac{\lambda}{\sqrt{2\pi \delta}} \exp \left[ -\frac{(x - \gamma)^2}{2\delta^2} \right]$$

while the cumulant exponent takes the form
The approach can be generalized to allow alternate distributions for $\gamma$ – in particular, a mixture of normals:

$$f_{\text{Merton}}(u) = \mu u + \frac{\gamma \sigma^2 (u^2 - u)}{2} + \lambda \left( e^{\theta u} + \frac{\gamma^2 u^2}{2} - 1 - u \bar{\kappa} \right).$$

The corresponding intensity parameters $k(x)$ in (7) are

$$k(x) = \sum_{i=1}^{2} \frac{\lambda_i}{\sqrt{2\pi \delta_i^2}} \exp \left( -\frac{(x - \bar{\gamma}_i)^2}{2\delta_i^2} \right).$$

Second, I will consider the generalized CGMY process of Carr, Madan, Geman and Yor (2003), which has a jump intensity of the form

$$k(x) = \begin{cases} C_n e^{-|x|} |x|^{-1} & \text{for } x < 0 \\ C_p e^{-|x|} |x|^{-1} & \text{for } x > 0 \end{cases}$$

(8)

where $C_n, C_p, G, M \geq 0$ and $Y_p, Y_n < 2$. The associated cumulant exponent is

$$f_{\text{CGMY}}(u) = (\mu - \omega)u + V \left\{ w_n \left( \frac{(G-u)^{Y_n} - G^{Y_n}}{Y_n (Y_n-1) G^{Y_n-2}} + (1-w_n) \left( \frac{(M+u)^{Y_p} - M^{Y_p}}{Y_p (Y_p-1) M^{Y_p-2}} \right) \right) \right\}$$

(9)

where $\omega$ is a mean-normalizing constant determined by $f_{\text{CGMY}}(1) = \mu$;

$V$ is the variance per unit time, and

$w_n$ is the fraction of variance attributable to the downward-jump component.

The corresponding intensity parameters $C_n, C_p$ in (8) are

$$C_n = \frac{w_n V}{\Gamma(2 - Y_n) G^{Y_n-2}}, \quad C_p = \frac{(1-w_n) V}{\Gamma(2 - Y_p) M^{Y_p-2}}$$

(10)

where $\Gamma(z)$ is the gamma function.

As discussed in Carr et al (2002), the $Y$ parameters are key in controlling jump activity near 0, in addition to their influence over tail events. The process has finite activity for $Y_p, Y_n < 0$, finite variation for $Y_p, Y_n < 1$, but infinite activity or variation if $\min(Y_p, Y_n)$ is greater or equal to 0 or...
1, respectively. The model conveniently nests many models considered elsewhere. For instance, 
\( Y_n = Y_p = -1 \) includes the finite-activity double exponential jump model of Kou (2002), while 
\( Y_n = Y_p = 0 \) includes the variance gamma model of Madan and Seneta (1990). As \( Y_p \) and \( Y_n \) 
approach 2, the CGMY process converges to a diffusion, and the cumulant exponent converges to 
the corresponding quadratic form

\[
\mathcal{L}_{\text{CGMY}}(\mu) = \mu \mu + \frac{\gamma}{2} \nu (\mu^2 - \mu). \tag{11}
\]

As \( G \) and \( M \) approach 0 (for arbitrary \( Y_p, Y_n \)), the Lévy density (8) approaches the infinite-
variance log stable process advocated by Mandelbrot (1963), with a “power law” property for 
asymptotic tail probabilities. The log-stable special case proposed by Carr and Wu (2003) is the 
limiting case with only negative jumps (\( \nu = 1 \)). While infinite-variance for log returns, percentage 
returns have finite mean and variance under the log-stable specification.

One can also combine Lévy processes, to nest alternative specifications within a broader 
specification. Any linear combination \( \nu_1 \mathcal{L}_1(\mu) + \nu_2 \mathcal{L}_2(\mu) \) of Lévy densities for \( \nu_1, \nu_2 \geq 0 \) is also 
a valid Lévy density, and generates an associated weighted cumulant exponent of the form 
\( \nu_1 \mathcal{L}_1(\mu) + \nu_2 \mathcal{L}_2(\mu) \).

I.2 Time-changed Lévy processes and stochastic volatility

Time-changed Lévy processes generate stochastic volatility by randomizing time in equation (1). 
Since the log transform (1) can be written as

\[
\ln F(u) = \frac{\mathcal{L}(\mu) \nu}{\mathcal{L}(\mu)(0)} \mu \nu \tag{12}
\]

for any finite-variance Lévy process, randomizing time is fundamentally equivalent to randomizing 
variance. As the connection between time changes and stochastic volatility becomes less transparent 
once “leverage” effects are added, I will use a stochastic volatility (or stochastic intensity) 
representation of stochastic processes.
The leverage effect, or correlation between asset returns and conditional variance innovations, is captured by directly specifying shocks common to both. I will initially assume that the log asset price $S_t$ follows a process of the form

$$dS_t = (\mu_0 + \mu_1 V_t)dt + \left[p_\sigma \sqrt{V_t} dW_t - \frac{1}{2} p_\sigma^2 V_t dt\right] + (dL_t - \omega V_t, dt)$$

$$dV_t = (\alpha - \beta V_t)dt + \sigma \sqrt{V_t} dW_t$$

(13)

The log increment $dS_t$ consists of the continuously-compounded return, plus increments to two exponential martingales. $dW_t$ is a Wiener increment, while $dL_t$ is a Lévy increment independent of $dW_t$, with instantaneous variance $(1 - p_\sigma^2) V_t dt$. The term $\omega V_t dt = E_{t} e^\Delta t - 1$ is a convexity adjustment that converts $dL_t - \omega V_t dt$ into an exponential martingale. Further refinements will be added below, to match properties of stock market returns more closely.

This specification has various features or implicit assumptions. First, the approach allows considerable flexibility regarding the distribution of the instantaneous shock $dL_t$ to asset returns, which can be Wiener, compound Poisson, or any other fat-tailed distribution. Three underlying Lévy processes are considered:

1) a second diffusion process $W_{2t}$ independent of $W_t$ [Heston (1993)];
2) finite-activity jumps drawn from a normal distribution or a mixture of normals; and
3) the generalized CGMY (2003) Lévy process from (8) above.

Combinations of these processes will also be considered, to nest the alternatives.

Second, the specification assumes a single underlying variance state variable $V_t$ that follows an affine diffusion, and which directly determines the variance of diffusion and jump components. This approach generalizes the stochastic jump intensity model of Bates (2000, 2006) to arbitrary Lévy processes.

Two alternate specifications are not considered, for different reasons. First, I do not consider the approach of Li, Wells and Yu (2006), who model log-differenced asset prices as the sum of a Heston (1993) stochastic volatility process and a constant-intensity fat-tailed Lévy process that captures outliers. Bates (2006, Table 7) found the stochastic-intensity jump model fits S&P 500 returns better than the constant-intensity specification, when jumps are drawn from a finite-activity
normal distribution or mixture of normals. Second, the diffusion assumption for \( V_t \) rules out volatility-jump models, such as the exponential-jump model proposed by Duffie, Pan and Singleton (2000) and estimated by Eraker, Johannes and Polson (2003). Such models do appear empirically relevant, but the AML filtration methodology described below is not yet in a form appropriate for such processes.

Define \( y_{t,T} = \int_t^T ds \) as the discrete-time return observed over horizon \( \tau = T-t \), and define \( f_{\alpha}(\omega) = (1 - \omega^2) V_t \omega_{\alpha}(\omega) \) as the cumulant exponent of \( dL_t - \omega V_t \, d\tau \). By construction, \( \omega_{\alpha}(\omega) \) is a standardized cumulant exponent, with \( \omega_{\alpha}(1) = 0 \) and variance \( \omega'_{\alpha}(0) = 1 \). A key property of affine models is the ability to compute the conditional generalized Fourier transform of \( (y_{t,T}, V_T) \).

This can be done by iterated expectations, conditioning initially on the future variance path:

\[
F(\Phi, \psi \mid V_T, \tau) = E( e^{\psi y_{t,T} + \psi V_T} \mid V_t) = E(E(e^{\psi y_{t,T} + \psi V_T} \mid V_T) \mid V_t)
\]

\[
= E(e^{\psi \mu_T + \int_t^T \Phi((\mu_1 - \frac{1}{2} \sigma^2) V_T + \rho \sigma \bar{V}_T + (\alpha - \omega) \alpha V_T) + \psi V_T} \mid V_T) \mid V_t)
\]

\[
= E(e^{\psi \mu_T + \int_t^T \Phi((\mu_1 - \frac{1}{2} \sigma^2) \Phi^2 - \Phi) + (1 - \rho^2) \omega_{\alpha}(\Phi)) V_T + \psi V_T} \mid V_t)
\]

\[
= E(e^{\psi \mu_T + h(\Phi) \int_t^T V_T + \psi V_T} \mid V_t)
\]

for \( h(\Phi) = \mu_1 \Phi + \frac{1}{2} \rho^2 \Phi^2 - \Phi) + (1 - \rho^2) \omega_{\alpha}(\Phi) \). This is the generalized Fourier transform of the future spot variance \( V_T \) and the average future variance \( \bar{V}_{t,T} = \frac{1}{\tau} \int_t^T V_T \, d\tau \). This is a well-known problem (see, e.g., Bakshi and Madan (2000)), with an analytic solution if \( V_t \) follows an affine process. For the affine diffusion above, \( F(\Phi, \psi \mid V_T, \tau) \) solves the Feynman-Kac partial differential equation

\[
\begin{aligned}
-F_T + (\alpha - \beta V_T) F_T + \frac{1}{2} \sigma^2 F_T F_{V_T} &= -h(\Phi) F_T F \\
\end{aligned}
\]

subject to the boundary condition \( F(\Phi, \psi \mid V_T, 0) = \exp(\psi V_T) \). The solution is

\[
F(\Phi, \psi \mid V_T, \tau) = \exp[C(\tau; \Phi, \psi) + D(\tau; \Phi, \psi) V_T] \]

where
I.3 Autocorrelations, and other refinements

Conditional variance is not the only latent state variable of relevance to stock market returns. It will be shown below that daily stock market returns were substantially autocorrelated over much of the 20th century; and that the autocorrelation was persistent and nonstationary. To capture this, I will explore below two alternate time series models for daily log-differenced stock index excess returns $y_t$:

$$y_{t+1} = \rho_t y_t + \eta_{t+1} \quad \text{(Model 1)}$$

or

$$y_{t+1} = \rho_t y_t + (1 - \rho_t) \eta_{t+1} \quad \text{(Model 2)}$$

where
\[\eta_{t+1} = \int_{t}^{t+\tau_t} ds_t\]

\[V_{t+1} = V_t + \int_t^{t+\tau_t} dV_u\]

\[\rho_{t+1} = \rho_t + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim N(0, \sigma^2_p) \text{ and i.i.d.}\]

and \(\tau_t\) is the effective length of a business day,

\(\rho_t\) determines the daily autocorrelation of stock index returns,

\(ds_t\) is the instantaneous intradaily underlying shock to log asset prices, and

\(V_t dt = Var_1(ds_t)\) is the instantaneous conditional variance of \(ds_t\).

The intradaily shocks \((ds_t, dV_t)\) are given by (13) above.

Both models add an autocorrelation state variable \(\rho_t\) that captures the fact that autocorrelations of stock market returns are not constant over time. Following the literature on time-varying coefficient models, the autocorrelation is modeled as a simple random walk, to avoid constraining estimates of \(\rho_t\). Estimation of the autocorrelation volatility parameter \(\sigma_p\) endogenously determines the appropriate degree of smoothing to use when filtering the current autocorrelation value \(\rho_t\) from past data.

The two models differ in ease of use, in their implications for the interaction between volatility and autocorrelation, and in the pricing of risks. Model 1 assumes the stock excess return residual \(\eta_{t+1} = y_{t+1} - \rho_t y_t\) is stationary (i.e., with a stationary variance process), and that the current value of \(\rho_t\) affects only the conditional mean of \(y_{t+1}\). This implies that excess returns \(y_{t+1}\) are nonstationary, given nonstationary \(\rho_t\). This model is more convenient for estimation, in that it generates a “semi-affine” structure that can be directly estimated using the methodology of Bates (2006).

In Model 2, \(\eta_{t+1}\) is the permanent impact of daily shocks to stock index excess returns, and is again assumed stationary. The model assumes that infrequent trading in the component stocks (proxied by \(\rho_t\)) slows the incorporation of such shocks into the observed stock index, but that the

\[\text{See, e.g., Andersen, Benzoni and Lund (2002, Table I), who estimate different autocorrelations for 1953-96 and 1980-96.}\]
index ultimately responds fully once all stocks have traded.\(^4\) It will be shown below that this model is more consistent with LeBaron’s (1992) observation that stock market volatility and autocorrelations appear inversely related. Furthermore, the model is more suitable for pricing risks; i.e., identifying the equity premium, or the (affine) risk-neutral process underlying options. The current value of \(\rho_t\) affects both the conditional mean and higher moments of \(\eta_{t+1}\), resulting in a significantly different filtration procedure for estimating \(\rho_t\) from past excess returns. The time series model is not affine, but this paper develops a transformation of variables that makes filtration and estimation tractable.

Both models build upon previous time series and market microstructure research into stock market returns. For instance, the effective length \(\tau_t\) of a business day is allowed to vary based upon various periodic effects. In particular, day-of-the-week effects, weekends, and holidays are accommodated by estimated time dummies that allow day-specific variation in \(\tau_t\). In addition, time dummies were estimated for the Saturday morning trading available over 1926-52, and for the Wednesday exchange holidays in the second half of 1968 that are the focus of French and Roll (1986).\(^5\) Finally, the stock market closings during the “Bank Holiday” of March 3-15, 1933 and following the September 11, 2001 attacks were treated as \(\frac{12}{365}\) - and \(\frac{7}{365}\)-year returns, respectively. Treating the 1933 Bank Holiday as a 12-day interval is particularly important, since the stock market rose 15.5% when the market re-opened on March 15. September 17, 2001 saw a smaller movement, of -4.7%.

For Model 1, the cumulant generating function of future returns and state variable realizations conditional upon current values is analytic, and of the semi-affine form

\(^4\)Jukivuolle (1995) distinguishes between the “observed” and “true” stock index when trading is infrequent, and proposes using a standard Beveridge-Nelson decomposition to identify the latter. This paper differs in assuming that the parameters of the ARIMA process for the observed stock index are not constant.

\(^5\)Gallant, Rossi and Tauchen (1992) use a similar approach, and also estimate monthly seasonals.
where, and are given in (17) and (18) above.

For model 2, the cumulant generating function is of the non-affine form given the shocks to are scaled by .

I.3 Filtration and maximum likelihood estimation

If the state variables were observed along with returns, it would in principle be possible to evaluate the joint transition densities of the data and the state variable evolution by Fourier inversion of the joint conditional characteristic function , and to use this in a maximum likelihood procedure to estimate the parameters of the stochastic process. However, since are latent rather than directly observed, this is a hidden Markov model that must be estimated by other means.

For Model 1, the assumptions that the cumulant generating function (23) is affine in the latent state variables implies that the hidden Markov model can be filtered and estimated using the approximate maximum likelihood (AML) methodology of Bates (2006). The AML procedure is a filtration methodology that recursively updates the conditional characteristic functions of the latent variables and future data conditional upon the latest datum. Define as the data observed up through period , and define

\[ G_{1|t}(i\xi, i\psi) = \mathbb{E}[e^{i\xi \rho_t + i\psi \nu_t} | \mathbf{y}_t] \]  

(27)
as the joint conditional characteristic function that summarizes what is known at time $t$ about $(\rho_t, \Psi_t)$. The density of the observation $y_{t+1}$ conditional upon $Y_t$ can be computed by Fourier inversion of its conditional characteristic function:

$$p(y_{t+1} | Y_t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{t+1}(\xi, \phi, 0) e^{i \xi y_{t+1} + \phi T_{t+1}(\xi, \phi, 0)} d\phi.$$  (28)

Conversely, the joint conditional characteristic function $G_{t+1|t}((\xi, \phi)$ needed for the next observation can be updated given $y_{t+1}$ by the characteristic-function equivalent of Bayes’ rule:

$$G_{t+1|t}(\xi, \phi) = \frac{1}{2\pi p(y_{t+1} | Y_t)} \int_{-\infty}^{\infty} G_{t+1}(\xi + i\phi y_{t+1}, D(\tau, i\phi, 0)) e^{C(\tau, i\phi, 0) - i\phi y_{t+1}} d\phi.$$(29)

The algorithm begins with an initial joint characteristic function $G_{1|1}(\cdot)$ and proceeds recursively through the entire data set, generating the log likelihood function $\sum \ln p(y_{t+1} | Y_t)$ used in maximum likelihood estimation. Filtered estimates of the latent variables can be computed from derivatives of the joint conditional moment generating function, as can higher conditional moments:

$$E[\rho_{t+1} | Y_{t+1}] = \left. \frac{\partial^{m+n} G_{t+1|t+1}(\xi, \phi)}{\partial \xi^m \partial \phi^n} \right|_{\xi = \phi = 0}.$$  (30)

The above procedure, if implementable, would permit exact maximum likelihood function estimation of parameters. However, the procedure would require storing and updating the entire function $G_{t+1}(\cdot)$ based on point-by-point univariate numerical integrations. As such a procedure would be slow, the AML methodology instead approximates $G_{t+1}(\cdot)$ at each point in time by a moment-matching joint characteristic function, and updates the approximation based upon updated estimates of the moments of the latent variables. Given an approximate prior $\hat{G}_{t+1}(\cdot)$ and a datum $y_{t+1}$, (27) is used to compute the posterior moments of $(\rho_{t+1}, \Psi_{t+1})$, which are then used to create an approximate $\hat{G}_{t+1|t+1}(\cdot)$. The overall procedure is analogous to the Kalman filtration procedure of updating conditional means and variances of latent variables based upon observed data, under the assumption that those variables and the data have a conditional normal distribution. However, the equations (26) and (27) identify the optimal nonlinear moment updating rules for a given prior $G_{t+1}(\cdot)$, whereas Kalman filtration uses linear rules. It will be shown below that this modification in
filtration rules is important when estimating latent autocorrelations and variances under fat-tailed Lévy processes. Furthermore, Bates (2006) proves that the iterative AML filtration is numerically stable, and shows that it performs well in estimating parameters and latent variable realizations.

Autocorrelations can be negative or positive, while conditional variance must be positive. Consequently, different two-parameter distributions were used for the conditional distributions of the two latent variables: Gaussian for autocorrelations, gamma for variances. Furthermore, since volatility estimates mean-revert within months whereas autocorrelation estimates evolve over years, realizations of the two latent variables were assumed conditionally independent. These assumptions resulted in an approximate conditional characteristic function of the form

$$\ln \hat{G}_t(\xi, \psi) = [\hat{\rho}_t \xi + \frac{1}{2} W_t \xi^2] - v_t \ln(1 - \kappa_t \psi).$$

The following summarizes key features of joint conditional distributions of the latent variables.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Autocorrelation $\rho_t$</th>
<th>spot variance $V_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_t$</td>
<td>$\hat{\rho}_t</td>
<td>Y_t \sim N(\hat{\rho}_{t</td>
</tr>
<tr>
<td>ln $E[e^{\xi N_t}</td>
<td>Y_t] = \hat{\rho}_{t</td>
<td>t} \xi + \frac{1}{2} W_{\xi} \xi^2$</td>
</tr>
<tr>
<td>$\kappa_t v_t = \hat{V}_{t</td>
<td>t}, \kappa_t^2 v_t = \hat{P}_{t</td>
<td>t}$</td>
</tr>
<tr>
<td>initial CGF</td>
<td>$\rho_1</td>
<td>Y_t \sim N(0, 10^2)$</td>
</tr>
</tbody>
</table>

$\rho_t, V_t | Y_t$ assumed independent for all $t$.

Initial variance was assumed drawn from its unconditional gamma distribution, with the parameters $(\kappa_1, v_1)$ given above. Since autocorrelations were assumed nonstationary, no unconditional distribution exists. Consequently, the AML algorithm was initiated using a relatively diffuse conditional distribution for the initial autocorrelation – one much wider than the plausible (-1, +1) range.
The parameters $\theta_t = (\hat{\rho}_{t\mid t}, \hat{W}_{t\mid t}; \kappa_t, \nu_t)$ – or, equivalently the moments $(\hat{\rho}_{t\mid t}, \hat{W}_{t\mid t}; \hat{\nu}_{t\mid t}, \hat{P}_{t\mid t})$ – summarize what is known about the latent variables. These were updated daily using the latest observation $y_{t+1}$ and equations (26) - (27). For each day, 5 univariate integrations were required: 1 for the density evaluation in (26), and 4 for the mean and variance evaluations in (27). An upper $\Phi_{\text{max}}$ was computed for each integral which upper truncation error would be less than $10^{-10}$ in magnitude. The integrands were then integrated over $(-\Phi_{\text{max}}, \Phi_{\text{max}})$ to a relative accuracy of $10^{-9}$, using IMSL’s adaptive Gauss-Legendre quadrature routine DQDAG and exploiting the fact that the integrands for negative $\phi$ are the complex conjugates of the integrands evaluated at positive $\phi$. On average between 234 and 448 evaluations of the integrand were required for each integration.6

The non-affine specification $y_{t+1} = \rho_t y_t + (1 - \rho_t) \eta_{t+1}$ in Model 2 necessitates additional restrictions upon the distribution of latent $\rho_t$. In particular, it is desirable that the scaling factor $1 - \rho_t$ be nonnegative, so that the lower tail properties of $\eta_{t+1}$ originating in the underlying Lévy specifications do not influence the upper tail properties of $y_{t+1}$. Consequently, the distribution of latent $1 - \rho_t$ for Model 2 is modeled as inverse Gaussian – a 2-parameter unimodal distribution with conditional mean $1 - \hat{\rho}_{t\mid t}$ and variance $\hat{W}_{t\mid t}$. Appendix A derives the resultant filtration procedure for this model, exploiting a useful change of variables procedure. The filtration is initiated at $\rho_0 \sim (0, .5^2)$, and it is again assumed that $\rho_t$ and $\nu_t$ are conditionally independent.

II. Properties of U.S. stock market returns, 1926 - 2006

II.1 Data

The data used in this study are daily cum-dividend excess returns on the CRSP value-weighted index over January 2, 1926 through December 29, 2006; a total of 20,919 excess returns. The CRSP value-weighted returns are very similar to returns on the (value-weighted) S&P Composite Index, which began in 1928 with 90 stocks and was expanded on March 1, 1957 to its current 500-stock structure. Indeed, the correlation between the CRSP value-weighted returns and S&P 500 returns was .9987 over 1957-2006. The CRSP series was preferred to S&P data partly because it begins two years earlier, but also because the S&P Composite Index is only reported to two decimal places,

---

6By contrast, the FFT approach used in Carr et al (2002) requires 16,384 functional evaluations.
which creates significant rounding error issues for the low index values observed in the 1930’s.
CRSP daily returns for each month were converted to daily log excess returns using Ibbotson and Associates’ data on monthly Treasury bill returns, and the formula
\[
y_t = \ln(1 + R_t) - \frac{\ln(1 + i)}{N} n_t
\]
where \( R_t \) is the daily CRSP cum-dividend return;
\( i \) is that month’s return on Treasury bills of at least 1 month to maturity;
\( N \) is the number of calendar days spanned by the monthly Treasury bill return; and
\( n_t \) is the number of calendar days spanned by the “daily” return \( R_t \).
The monthly interest rate data were downloaded from Ken French’s Web site, and extended backwards through 1926 using data in Ibbotson and Associates’ *SBBI Yearbook*.

II.2 Parameter estimates
Table 1 describes and provides estimates of the time dummies from the time-changed CGMY model,\(^7\) with Wednesday returns (Tuesday close to Wednesday close) arbitrarily selected as the benchmark day. Daily variance tended to be highest at the beginning of the week and decline thereafter, but day-of-the-week effects do not appear to be especially pronounced. The major exception is the Saturday morning (10 AM to noon) trading generally available over 1926-52.\(^8\) Saturdays were effectively 43\% as long as the typical Wednesday. Total weekend variance (Friday close to Monday close) was \((.43 + 1.05) / 1.10 - 1 = 34.5\%\) higher when Saturday trading was available (over 1926-52) than when it was not (primarily over 1945-2006).\(^9\) This is qualitatively similar to but less pronounced than the doubling of weekend variance found by Barclay,

\(^7\)Estimates from other specifications were virtually identical, with estimates typically within \(\pm 0.01\) of the CGMY model’s estimates.

\(^8\)Saturday trading was standard before 1945. Over 1945-51, it was eliminated in summer months, and was permanently eliminated on June 1, 1952.

\(^9\)As the time dummy estimates are estimated jointly with the volatility and autocorrelation filtrations, the estimates of weekend variances with versus without Saturday trading control for any divergences in volatility and autocorrelation levels in the two samples.
Litzenberger and Warner (1990) in Japanese markets when Saturday half-day trading was permitted. Barclay et al lucidly discuss market microstructure explanations for the increase in variance.

Holidays also did not have a strong impact on the effective length of a business day – with the exception of holiday weekends spanning 4 calendar days. Consistent with French and Roll (1986), 2-day returns spanning the Wednesday exchange holidays in 1968 (Tuesday close to Thursday close) had a variance not statistically different from a typical 1-day Wednesday return, but substantially less than the $1 + .94 = 1.94$ two-day variance observed for returns from Tuesday close to Thursday close in other years. Overall, the common practice of ignoring day-of-the-week effects, weekends, and holidays when analyzing the time series properties of daily stock market returns appears to be a reasonable approximation, provided the data exclude Saturday trading.

Table 2 reports estimates for various specifications, while Figure 1 presents associated normal probability plots for model 1. As noted above, all models capture the leverage effect by a correlation $\rho_\nu$, with the diffusion shock to conditional variance. The models diverge in their specifications of the Lévy shocks $dL_t$ _orthogonal_ to the variance innovation. The first two models (SVJ1, SVJ2) have a diffusion for small asset return shocks, plus finite-activity normally-distributed jumps to capture outliers. The other models examine the generalized time-changed CGMY model, along with specific parameter restrictions or relaxations.

The SVJ1 and SVJ2 results largely replicate the results in Bates (2006). The SVJ1 model has symmetric normally-distributed jumps with standard deviation 3% and time-varying jump intensities that occur on average $\lambda_1(\alpha/\beta) = 3.2$ jumps per year. As shown in Figure 1, this jump risk assessment fails to capture the substantial 1987 crash. By contrast, the SVJ2 model adds a second jump component that directly captures the 1987 outlier. The resulting increase in log likelihood from 75,044.60 to 75,049.07 is statistically significant under a likelihood ratio test, with a marginal significance level of 3.0%.

The various CGMY models primarily diverge across the specification of the $\gamma_p, \gamma_s$ parameters – whether they are set to specific levels, and whether they diverge for the intensities of positive versus negative jumps. The DEXP model with $\gamma_p = \gamma_s = -1$ is conceptually similar to the
jump-diffusion model SVJ1, but uses instead a finite-activity double exponential distribution for jumps. Despite the fatter-tailed specification, Figure 1 indicates the DEXP model has difficulties comparable to SVJ1 in capturing the ‘87 crash. The VG model replaces the finite-activity double exponential distribution with the infinite-activity variance process \((Y_p = Y_n = 0)\), and does marginally better in fit. Both models include a diffusion component, which captures 73-74% of the variance of the orthogonal Lévy shock \(dL_t\).

Models Y, YY, YY_J, and LS involve pure-jump specifications for the orthogonal Lévy process \(L_t\), without a diffusion component. Overall, higher values of \(Y\) fit the data better – especially the 1987 crash, which ceases to be an outlier under these specifications. Relaxing the restriction \(Y_p = Y_n\) leads to some improvement in fit, with the increase in log likelihood (YY versus Y) having a P-value of 1.8%. Point estimates of the jump parameters \((w_n, G, Y_n)\) governing downward jump intensities diverge sharply from the parameters \((1-w_n, M, Y_p)\) governing upward jump intensities when the \(Y_p = Y_n\) restriction is relaxed, although standard errors are large. The dampening coefficient \(G\) is not significantly different from zero, implying one cannot reject the hypothesis that the downward-jump intensity is from a stochastic-intensity version of the Carr-Wu (2003) log-stable process.\(^{10}\) By contrast, the upward intensity is estimated as a finite-activity jump process – which, however, still overestimates the frequency of big positive outliers (Figure 1, sixth panel).

Motivated by option pricing issues, Carr and Wu (2003) advocate using a log-stable distribution with purely downward jumps. An approximation to this model generated by setting \(G = .001\) and \(w_n = 1\) fits stock market returns very badly. The basic problem is that while the LS model does allow positive asset returns, it severely underestimates the frequency of large positive returns. This leads to a bad fit for the upper tail (Figure 1, last panel). Furthermore, it will be shown below that volatility and autocorrelation estimates are adversely effected following large positive

---

\(^{10}\)This was also tested by imposing \(G = .001\) in the YY model and optimizing over other parameters. The resulting log likelihood was 75,052.72, insignificantly different from the unconstrained 75,052.90. While setting \(G\) to zero was not permitted, given the assumption of finite variance, a value of \(G = .001\) implies negligible exponential dampening of the intensity function (8) over the \([-0.20, 0]\) observed range of negative log stock market excess returns, and is therefore observationally equivalent to the log stable specification.
returns by the model’s assumption that such returns are unlikely. However, the YY estimates indicate that the Carr-Wu specification can be a useful component of a model, provided the upward jump intensity function is modeled separately.

Some nested models were also estimated, to examine the sensitivity of the YY model to specific features of the data. For instance, unrestricted CGMY models generate at least one Y parameter in the infinite-activity, infinite-variation range [1, 2], and typically near the diffusion value of 2. This suggests that the models may be trying to capture considerable near-zero activity. However, adding an additional diffusion component to the time-changed YY Lévy specification to capture that activity separately (model YY_D) led to no improvement in fit. Similarly, the possibility that YY estimates might be affected substantially by the extreme 1987 crash was tested by adding an independent finite-activity normally-distributed jump component capable (as in the SVJ2 model) of capturing that outlier. The resulting fit (model YY_J) was not a statistically significant improvement over the YY model.

Apart from the LS model, all models have similar estimates for the parameters determining the conditional mean and stochastic variance evolution. The parameter $\mu_1$ is not significantly different from zero, indicating no evidence over 1926-2006 that the equity premium depended upon the level of conditional variance. Latent variance mean-reverts towards an estimated average level $(.143)^2 - (.159)^2$, with a half-life about 2 months, and a volatility of variance estimate at about .36. The half-life estimates are similar to those in Bates (2006, Table 8) for excess stock market returns over 1953-1996. However, the level and volatility of variance are higher than the 1953-96 estimates of $(.130)^2$ and .25, respectively. The divergence is almost assuredly attributable to differences in data sets – in particular, to the inclusion of the turbulent 1930’s in this study.

Overall, Figure 1 suggests the differences across the alternate fat-tailed specifications are relatively minor. The models SVJ1, DEXP, VG, and LS appear somewhat less desirable, given their failure to capture the largest outliers. However, the SVJ2, Y, and YY specifications appear to fit about the same. Furthermore, all models appear to have some specification error (deviations from linearity) in the $z \in [-2.5, -1.5]$ range and in the upper tail ($z > 2$). The sources of specification error are not immediately apparent. One possibility is that the jump intensity functions $h(x)$ are too
tightly parameterized, given the large amount of data. Another explanation is the data generating process may have changed over time, and that data from the 1930's and 1940's have little relevance for stock market risk today. Some support for this latter explanation is provided by Bates (2006, Figure 3), who finds less evidence of specification error for models estimated over 1953-96. These alternate possibilities will be explored further in future versions of this paper.

II.3 Autocorrelation estimates

That stock indexes do not follow a random walk was recognized explicitly by Lo and MacKinlay (1988), and implicitly by various earlier practices in variance and covariance estimation designed to cope with autocorrelated returns; e.g., Dimson (1979)’s lead/lag approach to beta estimation. The positive autocorrelations typically estimated for stock index returns are commonly attributed to stale prices in the stocks underlying the index. A standard practice in time series analysis is to pre-filter the data by fitting an ARMA specification; see, e.g., Jukivuolle (1995). Andersen, Benzoni and Lund (2002), for instance, use a simple MA(1) specification to remove autocorrelations in S&P 500 returns over 1953-96; a data set subsequently used by Bates (2006).

The approach of prefiltering the data was considered unappealing in this study, for several reasons. First, the 1926-2006 interval used here is long, with considerable variation over time in market trading activity and transactions costs, and structural shifts in the data generating process are probable. Indeed, Andersen et al (2002, Table 1) find autocorrelation estimates from their full 1953-96 sample diverge from estimates for a 1980-96 subsample. Second, ARMA packages use a mean squared error criterion that is not robust to the fat tails observed in stock market returns. Consequently, autocorrelations were treated as an additional latent variable, to be estimated jointly with the time series model (22).

Given that the prior distribution $\mathbf{p}_t | \mathbf{Y}_t$ is assumed $\mathcal{N}(\mathbf{\beta}_{\text{fit}}, \mathbf{W}_{\text{fit}})$, it can be shown that the autocorrelation filtration algorithm (27) for Model 1 updates conditional moments as follows:

$$\hat{\mathbf{p}}_{t+1} | s_{t+1} = \hat{\mathbf{p}}_{\text{fit}} - \mathbf{W}_{\text{fit}} \frac{\partial \ln p(\mathbf{y}_{t+1} | \mathbf{Y}_t)}{\partial y_{t-1}}$$  \hspace{1cm} (33)

$$\mathbf{W}_{t+1 | s_{t+1}} = \sigma_p^2 + (\mathbf{y}_t \mathbf{W}_{\text{fit}})^2 \frac{\partial^2 \ln p(\mathbf{y}_{t+1} | \mathbf{Y}_t)}{\partial y_{t-1}^2}$$  \hspace{1cm} (34)
If \( y_{t+1} \mid Y_t \) were conditionally normal, the log density would be quadratic in \( y_{t+1} \), and (30) would be the linear updating of Kalman filtration. More generally, the conditionally fat-tailed properties of \( y_{t+1} \) are explicitly recognized in the filtration.\(^{11}\) The partials of log densities can be computed numerically by Fourier inversion.

Figures 2 illustrates the autocorrelation filtrations estimated under various models. For model 1, the autocorrelation revision is fairly similar to a Kalman-filtration approach for observations within a ±2% range – which captures most observations, given a unconditional daily standard deviation around 1%. However, the optimal filtration for fat-tailed distributions is to downweight the information from returns larger than 2% in magnitude. The exception is the Carr-Wu log-stable specification (LS). Since that model assumes returns have a fat lower tail but not a particularly fat upper tail, its optimal filtration downweights the information in large negative returns but not in large positive returns.

The autocorrelation filtration under Model 2 is substantially different. Since \( y_{t+1} = \rho_t + (1 - \rho_t) \eta_{t+1} \) in that model, large observations of \( y_{t+1} \) are attributable either to large values of \( 1 - \rho_t \) (small values of \( \rho_t \)), or to large values of the Lévy shocks captured by \( \eta_{t+1} \). The resulting filtration illustrated in the lower panels of Figure 2 is consequently sensitive to medium-size movements in a fashion substantially different from the Model-1 specifications.

Figure 3 presents filtered estimates of the daily autocorrelation from the YY model, and the divergences from those estimates for other models. The most striking result is the extraordinarily pronounced increase in autocorrelation estimates from 1941 - 1971, with a peak of 35% reached in June 1971. Estimates from other models give comparable results, as do crude sample autocorrelation estimates using a 1- or 2-year moving window.\(^{12}\) After 1971, autocorrelation estimates fell steadily.

\(^{11}\)Similar equations were derived by Masreliez (1975), while the overall moment-matching filtration methodology has been termed “robust Kalman filtration.” See Schick and Mitter (1994) for a literature review.

\(^{12}\)See LeBaron (1992, Figure 1) for annual estimates of the daily autocorrelation of S&P composite index returns over 1928-1990.
and became insignificantly different from zero after 2002. This broad pattern is observed both for Models 1 and 2, although the precise estimates diverge given the different filtration methodologies.

The reasons for the evolution in autocorrelations are unclear. Changes in trading volume would seem the most plausible explanation, given the standard stale-price explanation. However, Gallant, Rossi and Tauchen (1992, Figure 2) find that volume trended downward over 1928-43, but generally increased throughout 1943-87. LeBaron (1992) finds that autocorrelations and stock market volatility are inversely related; as is also apparent from comparing Figure 3 with Figure 5 below. Goyenko, Subrahmanyam, and Ukhov (2008, Figures 1-2) find shifts in their measures of bond market illiquidity over 1962-2006 that parallel the stock market autocorrelation estimates, suggesting the evolution involves a broader issue of overall liquidity in financial markets.

Figure 3 also illustrates that the estimates of the daily autocorrelation are virtually nonstationary, indicating that fitting ARMA processes with time-invariant parameters to stock market excess returns is fundamentally pointless. The conditional standard deviation asymptotes at about 4½%, implying a 95% confidence interval of ±9% for the autocorrelation estimates.

II.4 Volatility filtration

Figure 4 illustrates how the estimated conditional volatility $\hat{\sigma}_t = \sqrt{V_t}$ is updated for the various models. The conditional volatility revisions use median parameter values $(\kappa, \nu) = (0.00295, 5.85)$ for the prior gamma distribution of $V_t$, implying a conditional mean $\kappa \nu = 1.31$ that is close to the $(1.19)^2$ median value observed for $\hat{\sigma}_t$ estimates from the YY model. For comparability with GARCH analyses such as Hentschel (1995), Figure 4 shows the “news impact curve,” or revision in conditional volatility estimates upon observing a given excess return, using the methodology of Bates (2006, pp.931-2).

---

13I am indebted to Ruslan Goyenko for pointing this out.

14As $\hat{\sigma}_t$ estimates have substantial positive skewness, the median is substantially below the mean estimate of $(.159)^2$ reported in Table 2.
All news impact curves are tilted, with negative returns having a larger impact on volatility assessments than positive returns. This reflects the leverage effect, or estimated negative correlation between asset returns and volatility shocks. All models process the information in small-magnitude asset returns similarly. Furthermore, almost all models truncate the information from returns larger than 3 standard deviations. This was also found in Bates (2006, Figure 1) for the SVJ1 model, indicating such truncation appears to be generally optimal for arbitrary fat-tailed Lévy processes. The LS exception supports this rule. The LS model has a fat lower tail but not a fat upper tail, and truncates the volatility impact of large negative returns but not of large positive returns. The fact that volatility revisions are not monotonic in the magnitude of asset returns is perhaps the greatest divergence of these models from GARCH models, which almost invariably specify a monotonic relationship. However, since moves in excess of ±3 standard deviations are rare, both approaches will generate similar volatility estimates most of the time.

Figure 5 presents the filtered estimates of conditional annualized volatility over 1926-2006 from the YY model 2, as well as the associated conditional standard deviation. Volatility estimates from other models (except SV & LS) are similar – as, indeed, is to be expected from the similar volatility updating rules in Figure 4. The conditional standard deviation is about 2.8%, indicating a 95% confidence interval of roughly ±4.6% in the annualized volatility estimates. Because of the 81-year time scale, the graph actually shows the longer-term volatility dynamics not captured by the model, as opposed to the intra-year volatility mean reversion with 2-month half-life that is captured by the model. Most striking is, of course, the turbulent market conditions of the 1930's, unmatched by any comparable volatility in the post-1945 era. The graph indicates the 1-

---

15 An exception is Maheu and McCurdy (2004), who put a jump indicator sensitive to outliers into a GARCH model. They find that the sensitivity of variance updating to the latest squared return should be reduced for outliers, for both stock and stock index returns.

16 Annualized” volatility refers to the choice of units. Since time is measured in years, $V_t$ is variance per year, and the daily volatility estimate of a return over a typical business day of length 1/252 years is approximately $E_t\sqrt{V_t/252}$. Since variance mean-reverts with an estimated half-life of roughly 2 months, it is not appropriate to interpret Figure 5 as showing the volatility estimate for a 1-year investment horizon.
factor stochastic variance model is too simple, and suggests that multifactor specifications of variance evolution are worth exploring.\textsuperscript{17}

II.5 Unconditional distributions
A further diagnostic of model specification is the models’ ability or inability to match the unconditional distribution of returns – in particular, the tail properties of unconditional distributions. Mandelbrot (1963, 2004), for instance, argues that empirical tails satisfy a “power law”: tail probabilities plotted against absolute returns approach a straight line when plotted on a log-log graph. This empirical regularity underlies Mandelbrot’s advocacy of the stable Pareto distribution, which possesses this property and is nested within the CGMY model for $G = M = 0$.

Mandelbrot’s argument is premised upon i.i.d. returns, but the argument can in principle be extended to time-changed Lévy processes. Conditional Lévy densities time-average; if the conditional intensity of moves of size $x$ is $V_t k(x)$, the unconditional frequency of moves of size $x$ is $E(V_t) k(x)$. Since unconditional probability density functions asymptotically approach the unconditional Lévy densities for large $|x|$, while unconditional tail probabilities approach the corresponding integrals of the unconditional Lévy densities, examining unconditional distributions may still be useful.

Figures 6a provides model-specific estimates of unconditional probability density functions of stock market excess return residuals, as well as data-based estimates from a histogram. Given the day-of-the-week effects reported in Table 1, the unconditional density functions are a horizon-dependent mixture of densities, with mixing weights set equal to the empirical frequencies. The substantial impact of the 1987 crash outlier upon parameter estimates is apparent. The SVJ2 estimates treat that observation as a unique outlier, while the CGMY class of models progressively fatten the lower tail as greater flexibility is permitted for the lower tail parameter $Y_n$. As noted above, the lower tail approaches the Carr-Wu (2003) log-stable (LS) estimate. However, the LS model is unable to capture the frequency of large positive outliers. All models closely match the empirical

---

\textsuperscript{17}The inadequacies of AR(1) representations of conditional variance are already reasonably well-known in volatility research, and have also motivated research into long-memory processes.
unconditional density function in the ±3% range where most observations occur; and all models underestimate the frequency of moves of 3%-7% in magnitude.

Figure 6b provides similar estimates for unconditional lower and upper tail probabilities. In addition, 1000 sample paths of stock market excess return residuals over 1926-2006 were simulated via a Monte Carlo procedure using YY parameter estimates, in order to provide confidence intervals on tail probability estimates. Unsurprisingly, the confidence intervals on extreme tail events are quite wide. The underestimation of moves of 3%-7% in magnitude is again apparent, and is statistically significant. This rejection of the YY model does not appear attributable to misspecification of the Lévy density $k(x)$, which in Figure 1 captures conditional densities reasonably well. Rather, the poor unconditional fit in Figures 6a and 6b appears due to misspecification of volatility dynamics. Half of the 3-7% moves occurred over 1929-1935 – a sustained period with volatility exceeding 30% (see Figure 5) that simulated volatility realizations from the 1-factor variance process of equation (13) cannot match.

Figure 7 plots model-specific tail probability estimates for the YY model on the log-log scales advocated by Mandelbrot, along with data-specific quantiles for 20,004 stock market residuals that have roughly a 1-day estimated time horizon (±25%). The lower tail probability does indeed converge to the unconditional tail intensity

$$K(y) = \int_{-\infty}^{y} k(x) dx = C_n G^{y_n} \Gamma(-y_n, G | y)$$

$$= C_n y^{-y_n} \text{for } G = 0$$

(35)

where $C_n = w_n (1 - \rho^2) \alpha / [\beta \Gamma(2 - Y_n) G^{Y_n - 2}]$ and $\Gamma(\alpha, z)$ is the incomplete gamma function. Furthermore, given $G$ estimates near 0, $K(y)$ is roughly a power function, implying near linearity when plotted on a log-log scale.

---

18Conditional variance sample paths were simulated using the approach of Bates (2006, Appendix A.6), while Lévy shocks conditional upon intradaily average variance and data-based daily time horizons were generated via an inverse CDF methodology. The two observations corresponding to the market closings in 1933 and 2001 were omitted.
However, the graph indicates that the convergence of tail probabilities to the tail intensity $K(y)$ occurs only for observations in excess of 5% in magnitude – roughly 5 standard deviations. As this is outside the range of almost all data, it does not appear that log-log scales provide a useful diagnostic of model specification and tail properties. This is partly due to stochastic volatility, which significantly slows the asymptotic convergence of unconditional tail probabilities to $K(y)$ for large $|y|$. Absent stochastic volatility ($\sigma = 0$), the tail probabilities of an i.i.d. YY Lévy process converge to $K(y)$ for observations roughly in excess of 3% in magnitude (3 standard deviations).

No power law properties are observed for upper tail probabilities, given substantial estimated exponential dampening. The failure of both lower and upper unconditional tail probabilities to capture the frequency of moves of 3-7% in magnitude is again apparent, and statistically significant.

### III. Option pricing implications

Do these alternative models imply different option prices? Exploring this issue requires identifying the appropriate pricing of equity, jump, and stochastic volatility risks. Furthermore, the presence of substantial and stochastic autocorrelation raises issues not previously considered when pricing options. In particular, the observed stock index level underlying option prices can be stale, while the relevant volatility measure over the option’s lifetime is also affected. The variance of the sum of future stock market returns is not the sum of the variances when returns are autocorrelated.

To examine these issues, I will focus upon Model 2, with its distinction between the observed CRSP value-weighted stock index and an underlying true index level that would be observed were all component stock prices continuously updated. Furthermore, I will use the myopic power utility pricing kernel specification of Bates (2006) for pricing the various risks:

$$d\ln M = \mu_m dt - R \delta_t$$

where $\delta_t$ is the permanent shock to the log stock market level given above in equations (24) - (25). This specification constrains both the equity premium estimated under the objective time series
Carr and Wu (2003) specify a log-stable process for the risk-neutral process underlying option prices. This can be generated by a time-changed CGMY process for the actual process with only downward jumps, and with ... 

\[ (\mu_0 + \mu_1 V_t) dt = -E_t \left[ e^{d_\mu t} - 1 \right] \left( e^{-R d_\mu t} - 1 \right) \]  

(37)  

which implies

\[ \mu_0 = 0 \]

\[ \mu_1 V_t = [R \rho_{sv}^2 + R(1 - \rho_{sv}^2)(1 - f_{\text{jump}})] V_t - \int_{-\infty}^{\infty} (e^x - 1)(e^{-Rx} - 1) k(x) dx \]

(38)  

where \( 1 - f_{\text{jump}} \) is the fraction of variance attributable to an orthogonal diffusion term. The approximation follows from first-order Taylor expansions, and from the fact that jumps account for a fraction \( f_{\text{jump}}(1 - \rho_{sv}^2) \) of overall variance \( V_t \). The equity premium (38) is well-defined for the SVJ1 and SVJ2 models. For the CGMY models, the restriction \( G > R \) is required for a finite equity premium; the intensity of downward jumps must fall off faster than investors’ risk aversion to such jumps. The log-stable process is inconsistent with a finite equity premium.\(^{19}\)

The change of measure from objective to risk-neutral jump intensities takes the form

\[ k^*(x) = k(x)e^{-Rx} \]

(39)  

under a myopic power utility pricing kernel. This has assorted implications for parameter transformations that depend upon the precise specification of the Lévy density \( k(x) \). For the SVJ models, as discussed in Bates (2006), this modified jump intensity shifts the mean jump size \( \bar{\gamma}_t \) by an amount \( -R \delta_t^2 \), while leaving the jump standard deviation \( \delta_t \) unchanged. For the CGMY model, the risk adjustment replaces the downward and upward exponential dampening parameters \( G \) and \( M \) by \( G - R \) and \( M + R \), respectively, while leaving the \( C \) and \( Y \) parameters unchanged.\(^{20}\) These

\(^{19}\)Carr and Wu (2003) specify a log-stable process for the risk-neutral process underlying option prices. This can be generated by a time-changed CGMY process for the actual process with only downward jumps, and with \( G = R \).

\(^{20}\)Wu (****) discusses this transformation.
risk adjustments alter the \( g_{\Phi} \) and \( h_{\Phi} \) functions in equation (14). Table (*) summarizes the various parameter transformations.

The potential impact of autocorrelations upon option prices will be addressed by examining prices of options on S&P 500 futures. I assume that stock index futures prices respond instantaneously and fully to the arrival of news, whereas lack of trading in the underlying stocks delays the incorporation of that information into the reported CRSP and S&P 500 stock index levels. Furthermore, I assume that index arbitrageurs effectively eliminate any stale prices in the component stocks on days when futures contracts expire, so that stale prices do not affect the cash settlement feature of stock index futures. MacKinlay and Ramaswamy (1988) provide evidence supportive of both assumptions.

These assumptions have the following implications under Model 2:

1. the observed futures price \( F_t \) underlying options on S&P 500 futures is not stale;
2. given the AR(1) specification for stock index excess returns in equation (22), log futures price innovations are approximately the intradaily innovations \( \Delta_\tau \) of equation (13):

\[
d\ln F_t = \Delta_\tau_t
\]  

Consequently, European options on stock index futures can be priced directly using a risk-neutral version of (40) – which is affine, simplifying option evaluation considerably. Furthermore, option prices do not depend upon \( \rho_t \), except indirectly through the impact of autocorrelation filtration upon the filtration of latent variance \( \nu_t \). Following Bates (2006), European call prices on an S&P 500 futures contract can be priced as

\[
c(F_t, T; X | Y_t) = E[c(F_t, Y_t, T; X) | Y_t, F_t] \\
= e^{-rT}F_t - \frac{1}{2} e^{-rT}X \left[ 1 + \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{2\gamma(y)0, \tau) + \gamma(D2(\Phi, 0, T) - \Phi \ln(X/F_t)) \over i(1 - i\Phi)} d\Phi \right], \tag{41}
\]
where $C^*(\cdot)$ and $D^*(\cdot)$ are risk-neutral variants\(^{21}\) of those in equations (17)-(21), using the risk-neutral parameters of Table 4; and

$$g_{Y_t}[\psi] = -v_t \ln(1 - \kappa_t \psi)$$

is the filtered cumulant generating function of $V_t$ that summarizes what is known about $V_t$ given past data $Y_t$.

The key risk aversion parameter $R$ used for change of probability measure was estimated by imposing the equity premium restrictions (36), and re-estimating all times series models. The additional parameter restriction $G \geq R$ was also imposed upon all CGMY models, and was binding for the YY model.\(^{22}\) Parameter estimates reported in Table 3 changed little relative to those in Table 2, while risk aversion was estimated at roughly 2.5 for all models. Furthermore, the restriction of a purely variance-sensitive equity premium ($\mu_0 = 0$) was not rejected for all models.

The resulting ISD’s from the various models are graphed in Figure 8, and are compared with observed ISD’s from American options on S&P 500 futures on December 29, 2006. All models generate virtually identical option prices for near-the-money options. For 21-day options maturity January 2007, for instance, option prices are about the same across models for strike prices within a range of $\pm 5\%$ around the futures price — roughly 2 standard deviations at a 21-day maturity, spanning the most active area of options trading. The divergences in tail distribution estimates across models show up only in divergences in deep OTM put options’ ISD’s.

Figure 9 chronicles the ATM ISD’s estimates over 1983-2006. The estimates are similar to those in Bates (2006, Figure 7). However, the volatility estimates taking nonstationary autocorrelations into account via model 2 are higher, lowering the overall divergence between estimated and observed ISD’s.

\(^{21}\)The parameters $(\mu_1, \alpha, \beta)$ are replaced by $(0, \alpha/(1 - \hat{\theta}_t)^2, \beta^*)$, while the function $h(\Phi)$ is replaced by a risk-adjusted function $h^*(\Phi)$ that captures the risk-adjusted jump intensities from equation (33).

\(^{22}\)Wu (2006) proposes an alternate pricing kernel with negative risk aversion for downside risk, thereby automatically imposing $G \geq R$. 
IV. Summary and Conclusions

This paper provides estimates of the time-changed CGMY (2003) Lévy process, and compares them to the time-changed finite-activity jump-diffusions previously considered by Bates 2006). Overall, both models fit stock market excess returns over 1926-2006 similarly. However, the CGMY approach is slightly more parsimonious, and is able to capture the 1987 crash without resorting to the “unique outlier” approach of the SVJ2 model. The CGMY model achieves this with a (slightly) dampened power law specification for negative jump intensities that is observationally equivalent to a time-changed Carr-Wu (2003) infinite-variance log-stable specification. However, the time-changed log-stable model is found to be incapable of capturing the substantial positive jumps also observed in stock market returns, which the more general time-changed CGMY model handles better. All models still exhibit some conditional and unconditional specification error, the sources of which have not yet been fully established.

The paper also documents some structural shifts over time in the data generating process. Most striking is the apparently nonstationary evolution of the first-order autocorrelation of daily stock market returns, which rose from near-zero in the 1930's to 35% in 1971, before drifting down again to near-zero values after 2002. Longer-term trends in volatility are also apparent in the filtered estimates, suggesting a need for multifactor models of conditional variance. Whether there appear to be structural shifts in the parameters governing the distribution of extreme stock market returns will be examined in future versions of this paper.

Finally, it is important when estimating latent state variables to use filtration methodologies that are robust to the fat-tailed properties of stock market returns. Standard GARCH models lack this robustness, and generate excessively large estimates of conditional variance after large stock market movements.
References


Appendix A. Filtration under Model 2

From equation (26), the cumulant generating function (CGF) for future \( \{y_{t+1}, \rho_{t+1}, V_{t+1}\} \) conditional upon knowing \( \{y_{t}, \rho_{t}, V_{t}\} \) is

\[
\ln F(\Phi, \xi, \psi \mid y_{t}, \rho_{t}, V_{t}) = C(\tau_{t}; (1 - \rho_{t}) \Phi, \xi, \psi) + (\xi + \Phi y_{t}) \rho_{t} + D(\tau_{t}; (1 - \rho_{t}) \Phi, \psi) V_{t}. \tag{A.1}
\]

The filtered CGF conditional upon only observing past data \( Y_{t} \) can be computed by integrating over the conditional distributions of the latent variables \( \{\rho_{t}, V_{t}\} \):

\[
P(\Phi, \xi, \psi \mid Y_{t}) = \int \int e^{C(1 - \rho; \Phi, \xi) + (\xi + \Phi y_{t}) \rho_{t} + D(1 - \rho; \Phi, \psi)V_{t} \rho_{t}} P(\rho_{t} \mid Y_{t}) P(\rho_{t}) d\rho_{t} d\Phi,
\]

where \( g_{1}(\rho) = -v_{t} \ln(1 - \rho) \psi \) is the gamma conditional CGF for latent \( V_{t} \). Under the change of variables \( \{z, x\} = \{(1 - \rho), \Phi, 1 - \rho\} \), and under the assumption that the scaling term \( x = 1 - \rho > 0 \), the Fourier inversion used in evaluating \( P(y_{t+1} \mid Y_{t}) \) from (A.2) becomes

\[
P(y_{t+1} \mid Y_{t}) = \frac{1}{\pi} \mathcal{R} e \left\{ \int_{0}^{1} \int_{e^{-\infty}}^{e^{\infty}} e^{C(1 - \rho, 0, 0) + \Phi y_{t}, \rho_{t} + \zeta_{0}[D(1 - \rho, 0, 0) - \Phi y_{t}]} P(\rho_{t} \mid Y_{t}) d\rho_{t} d\Phi \right\}
\]

where \( \mathcal{R} e \{c\} \) denotes the real component of complex-valued \( c \), and the \( 1/x \) term in the integrand reflects the Jacobean from the change of variables. It is convenient to use a unimodal inverse Gaussian distribution for \( P(x \mid Y_{t}) \):

\[
P(x \mid Y_{t}) = \sqrt{\frac{\lambda}{2\pi x^{3}}} \exp \left[ -\frac{\lambda(x - \mu)^{2}}{2\mu^{2}x} \right], x > 0 \tag{A.4}
\]

where \( \mu = \mathbb{E}[x \mid Y_{t}] \) and \( \lambda = \mu^{2} / \text{Var}[x \mid Y_{t}] \) are \( t \)-dependent parameters that summarize what is known about \( x \) (and about \( \rho_{t} \)) at time \( t \). Under this specification, the inner integration inside (A.3) can be replaced by the analytic function
A more "natural" choice would be to represent $\theta$ by a beta distribution over the range $[0, 2]$. That would constrain $\mu$, and results in an $E$ term that involves the confluent hypergeometric U-function. However, I could not find a method for evaluating that function that was fast, accurate, and robust to all parameter values.

for $a = -iz(y_{t+1} - y_t) \equiv -iz\Delta y$. Consequently, evaluating (A.3) involves only univariate numerical integration.1

Similar univariate integrations are used for filtering $V_{t+1}$ and $\rho_{t+1}$ conditional upon observing $y_{t+1}$. The noncentral posterior moments of $V_{t+1}$ are given by

$$E(V_{t+1}^m \mid Y_{t+1}) = \frac{1}{\pi p(y_{t+1} \mid Y_t)} \text{Re} \left\{ \int_{\psi = 0}^{\infty} e^{C(z, 0, 0) + \varepsilon_{\psi}D(z, 0, 0) - iz\Delta y} M_{-1}(-iz\Delta y) \, dz \right\}$$

(A.6)

where the derivatives with respect to $\psi$ inside the integrand can be easily evaluated from the specifications for $C(\psi)$ and $D(\psi)$ in equations (17) - (18). The posterior moments of $\rho_{t+1}$ can be computed by taking partials of (A.2) with respect to $\xi$, and then again using change of variables to reduce the Fourier inversion to a univariate integration. The resulting posterior mean and variance of $\rho_{t+1}$ are

$$\hat{\rho}_{t+1 \mid T+1} = 1 - \frac{1}{\pi p(y_{t+1} \mid Y_t)} \text{Re} \left\{ \int_{z = 0}^{\infty} e^{C(z, 0, 0) + \varepsilon_{\psi}D(z, 0, 0) - iz\Delta y} M_0(-iz\Delta y) \, dz \right\}$$

(A.7)

$$\omega_{t+1 \mid T+1} = \sigma_{\epsilon}^2 + \frac{1}{\pi p(y_{t+1} \mid Y_t)} \text{Re} \left\{ \int_{z = 0}^{\infty} e^{C(z, 0, 0) + \varepsilon_{\psi}D(z, 0, 0) - iz\Delta y} M_1(-iz\Delta y) \, dz \right\}$$

(A.8)

where

1A more “natural” choice would be to represent $x$ by a beta distribution over the range $[0, 2]$. That would constrain $\mu < 1$, and results in an $M_{-1}(a)$ term that involves the confluent hypergeometric U-function. However, I could not find a method for evaluating that function that was fast, accurate, and robust to all parameter values.
Finally, the conditional distribution function $P(y) = \text{Prob}[y_{i+1} < y \mid Y_i]$ that is used in QQ plots takes the form

$$P(y) = \frac{1}{2} - \frac{1}{\pi} \text{Re} \left\{ \frac{\exp[C(z, 0, 0) + g_i[T(z, 0, 0)] - iy z]}{iz} M_0(-iz \Delta y) \, dz \right\}. \quad (A.11)$$
Table 1: Effective length of a business day, relative to 1-day Wednesday returns: 1926-2006.

<table>
<thead>
<tr>
<th>#days</th>
<th>Description</th>
<th>NOBS</th>
<th>Model 1 Estimate</th>
<th>Model 1 std. error</th>
<th>Model 2 Estimate</th>
<th>Model 2 std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Monday close → Tuesday close</td>
<td>3831</td>
<td>1.02</td>
<td>(.04)</td>
<td>1.03</td>
<td>(.03)</td>
</tr>
<tr>
<td>1</td>
<td><strong>Tuesday close → Wednesday close</strong></td>
<td>4037</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Wednesday → Thursday</td>
<td>3998</td>
<td>.94</td>
<td>(.03)</td>
<td>.94</td>
<td>(.03)</td>
</tr>
<tr>
<td>1</td>
<td>Thursday → Friday</td>
<td>3924</td>
<td>.93</td>
<td>(.03)</td>
<td>.92</td>
<td>(.03)</td>
</tr>
<tr>
<td>1</td>
<td>Friday → Saturday (1926-52)</td>
<td>1141</td>
<td>.43</td>
<td>(.02)</td>
<td>.44</td>
<td>(.02)</td>
</tr>
<tr>
<td>2</td>
<td>Saturday close → Monday close (1926-52)</td>
<td>1120</td>
<td>1.05</td>
<td>(.05)</td>
<td>1.07</td>
<td>(.05)</td>
</tr>
<tr>
<td>2</td>
<td>Weekday holiday</td>
<td>341</td>
<td>1.25</td>
<td>(.11)</td>
<td>1.26</td>
<td>(.10)</td>
</tr>
<tr>
<td>2</td>
<td>Wednesday exchange holiday in 1968</td>
<td>22</td>
<td>.73</td>
<td>(.33)</td>
<td>.81</td>
<td>(.35)</td>
</tr>
<tr>
<td>3</td>
<td>Weekend and/or holiday&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2755</td>
<td>1.10</td>
<td>(.04)</td>
<td>1.10</td>
<td>(.04)</td>
</tr>
<tr>
<td>4</td>
<td>Holiday weekend</td>
<td>343</td>
<td>1.58</td>
<td>(.14)</td>
<td>1.56</td>
<td>(.13)</td>
</tr>
<tr>
<td>5</td>
<td>Holiday weekend</td>
<td>6</td>
<td>1.31</td>
<td>(1.00)</td>
<td>1.25</td>
<td>(.93)</td>
</tr>
</tbody>
</table>

21518

<sup>a</sup>Includes one weekday holiday (August 14 - 17, 1945)
Table 2A: Estimates of parameters affecting the conditional means and volatilities. Data: daily CRSP value-weighted excess returns, 1926-2006. See equations (6) - (10), (13), and (25) for definitions of parameters. Models with $f_{\text{jump}} < 1$ combine Lévy jump processes with an additional independent diffusion, with variance proportions $(f_{\text{jump}}, 1 - f_{\text{jump}})$, respectively. Standard errors are in parentheses.

### Model 1: $y_{t+1} = \rho_t y_t + \eta_{t+1}$

<table>
<thead>
<tr>
<th>Model</th>
<th>ln $L$</th>
<th>$\mu_0$</th>
<th>$\mu_1$</th>
<th>$\sigma_\rho \sqrt{252}$</th>
<th>$\sqrt{\frac{\alpha}{\beta}}$</th>
<th>$\beta$</th>
<th>$\sigma$</th>
<th>$\rho_{sv}$</th>
<th>HL (mths)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV</td>
<td>75,044.60</td>
<td>.040 (.015)</td>
<td>.94 (.91)</td>
<td>.030 (.006)</td>
<td>.155 (.005)</td>
<td>4.33 (.40)</td>
<td>.370 (.011)</td>
<td>-.642 (.020)</td>
<td>1.9 (.2)</td>
</tr>
<tr>
<td>SVJ1</td>
<td>75,049.07</td>
<td>.042 (.015)</td>
<td>.87 (.92)</td>
<td>.030 (.007)</td>
<td>.155 (.005)</td>
<td>4.34 (.37)</td>
<td>.371 (.011)</td>
<td>-.642 (.020)</td>
<td>1.9 (.2)</td>
</tr>
<tr>
<td>DEXP</td>
<td>75,047.62</td>
<td>.044 (.015)</td>
<td>.79 (.91)</td>
<td>.030 (.006)</td>
<td>.156 (.005)</td>
<td>4.25 (.40)</td>
<td>.370 (.012)</td>
<td>-.588 (.020)</td>
<td>2.0 (.2)</td>
</tr>
<tr>
<td>VG</td>
<td>75,049.48</td>
<td>.042 (.015)</td>
<td>.91 (.91)</td>
<td>.030 (.006)</td>
<td>.156 (.005)</td>
<td>4.25 (.39)</td>
<td>.368 (.012)</td>
<td>-.587 (.020)</td>
<td>2.0 (.2)</td>
</tr>
<tr>
<td>Y</td>
<td>75,050.12</td>
<td>.042 (.015)</td>
<td>.91 (.92)</td>
<td>.030 (.006)</td>
<td>.157 (.008)</td>
<td>3.90 (.38)</td>
<td>.350 (.019)</td>
<td>-.577 (.032)</td>
<td>2.1 (.2)</td>
</tr>
<tr>
<td>YY</td>
<td><strong>75,052.90</strong></td>
<td><strong>.042 (.015)</strong></td>
<td><strong>.87 (.91)</strong></td>
<td><strong>.030 (.006)</strong></td>
<td><strong>.159 (.008)</strong></td>
<td><strong>4.00 (.38)</strong></td>
<td><strong>.362 (.019)</strong></td>
<td><strong>-.572 (.031)</strong></td>
<td><strong>2.1 (.2)</strong></td>
</tr>
<tr>
<td>YY_D</td>
<td>75,052.90</td>
<td>.042 (.015)</td>
<td>.87 (.91)</td>
<td>.030 (.006)</td>
<td>.159 (.010)</td>
<td>4.01 (.38)</td>
<td>.363 (.021)</td>
<td>-.572 (.036)</td>
<td>2.1 (.2)</td>
</tr>
<tr>
<td>YY_J</td>
<td>75,054.94</td>
<td>.041 (.015)</td>
<td>.97 (.92)</td>
<td>.030 (.007)</td>
<td>.154 (.005)</td>
<td>3.99 (.38)</td>
<td>.350 (.012)</td>
<td>-.586 (.020)</td>
<td>2.1 (.2)</td>
</tr>
<tr>
<td>LS</td>
<td>75,005.53</td>
<td>.019 (.015)</td>
<td>1.71 (.78)</td>
<td>.031 (.007)</td>
<td>.165 (.005)</td>
<td>4.52 (.39)</td>
<td>.405 (.011)</td>
<td>-.554 (.020)</td>
<td>1.8 (.2)</td>
</tr>
</tbody>
</table>

### Model 2: $y_{t+1} = \rho_t y_t + (1 - \rho_t) \eta_{t+1}$

<table>
<thead>
<tr>
<th>Model</th>
<th>ln $L$</th>
<th>$\mu_0$</th>
<th>$\mu_1$</th>
<th>$\sigma_\rho \sqrt{252}$</th>
<th>$\sqrt{\frac{\alpha}{\beta}}$</th>
<th>$\beta$</th>
<th>$\sigma$</th>
<th>$\rho_{sv}$</th>
<th>HL (mths)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV</td>
<td>74,999.87</td>
<td>-.014 (.020)</td>
<td>3.04 (.90)</td>
<td>.043 (.005)</td>
<td>.170 (.004)</td>
<td>8.01 (.57)</td>
<td>.562 (.015)</td>
<td>-.658 (.017)</td>
<td>1.0 (.1)</td>
</tr>
<tr>
<td>SVJ1</td>
<td>75,092.10</td>
<td>.033 (.020)</td>
<td>1.69 (.104)</td>
<td>.036 (.005)</td>
<td>.171 (.004)</td>
<td>5.80 (.49)</td>
<td>.457 (.015)</td>
<td>-.674 (.018)</td>
<td>1.4 (.1)</td>
</tr>
<tr>
<td>SVJ2</td>
<td>75,096.68</td>
<td>.037 (.020)</td>
<td>1.25 (.89)</td>
<td>.036 (.005)</td>
<td>.172 (.004)</td>
<td>5.71 (.49)</td>
<td>.456 (.015)</td>
<td>-.673 (.018)</td>
<td>1.4 (.1)</td>
</tr>
<tr>
<td>DEXP</td>
<td>75,094.20</td>
<td>.034 (.020)</td>
<td>1.44 (.90)</td>
<td>.036 (.005)</td>
<td>.171 (.004)</td>
<td>5.67 (.49)</td>
<td>.452 (.015)</td>
<td>-.625 (.018)</td>
<td>1.5 (.1)</td>
</tr>
<tr>
<td>VG</td>
<td>75,094.70</td>
<td>.034 (.020)</td>
<td>1.42 (.90)</td>
<td>.037 (.005)</td>
<td>.171 (.004)</td>
<td>5.56 (.48)</td>
<td>.447 (.016)</td>
<td>-.623 (.018)</td>
<td>1.6 (.1)</td>
</tr>
<tr>
<td>Y</td>
<td>75,093.68</td>
<td>.036 (.021)</td>
<td>1.35 (.90)</td>
<td>.036 (.005)</td>
<td>.172 (.007)</td>
<td>5.18 (.46)</td>
<td>.432 (.021)</td>
<td>-.613 (.027)</td>
<td>1.6 (.1)</td>
</tr>
<tr>
<td>YY</td>
<td><strong>75,097.20</strong></td>
<td><strong>.033 (.020)</strong></td>
<td><strong>1.44 (.90)</strong></td>
<td><strong>.036 (.005)</strong></td>
<td><strong>.172 (.006)</strong></td>
<td><strong>5.23 (.47)</strong></td>
<td><strong>.437 (.018)</strong></td>
<td><strong>-.613 (.022)</strong></td>
<td><strong>1.6 (.1)</strong></td>
</tr>
<tr>
<td>LS</td>
<td>75,045.48</td>
<td>.053 (.019)</td>
<td>1.50 (.76)</td>
<td>.031 (.003)</td>
<td>.174 (.005)</td>
<td>4.68 (.41)</td>
<td>.436 (.015)</td>
<td>-.576 (.019)</td>
<td>1.8 (.2)</td>
</tr>
</tbody>
</table>
### Table 2B: Estimates of jump parameters. Standard errors in parentheses.

<table>
<thead>
<tr>
<th>Model</th>
<th>$t_{jump}$</th>
<th>$\omega_n$</th>
<th>$G$</th>
<th>$M$</th>
<th>$Y_n$</th>
<th>$Y_p$</th>
<th>$\lambda_i$</th>
<th>$\bar{\gamma}$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVJ1</td>
<td>.140 (.015)</td>
<td>.49 (.07)</td>
<td>66.1 (6.0)</td>
<td>45.4 (10.1)</td>
<td>-1</td>
<td></td>
<td>152.4 (24.8)</td>
<td>.000 (.002)</td>
<td>.030 (.002)</td>
</tr>
<tr>
<td>SVJ2</td>
<td>.156 (.022)</td>
<td>.52 (.07)</td>
<td>41.1 (5.4)</td>
<td>31.6 (9.1)</td>
<td>0</td>
<td></td>
<td>162.8 (29.3)</td>
<td>.000 (.000)</td>
<td>.029 (.003)</td>
</tr>
<tr>
<td>DEXP</td>
<td>.256 (.030)</td>
<td>.59 (.06)</td>
<td>7.0 (4.5)</td>
<td>2.3 (7.2)</td>
<td>1.87 (.03)</td>
<td></td>
<td>0.5 (0.7)</td>
<td>-.189 (.083)</td>
<td>.005 (.028)</td>
</tr>
<tr>
<td>VG</td>
<td>.274 (.030)</td>
<td>.88 (.03)</td>
<td>1.6 (4.2)</td>
<td>40.1 (31.0)</td>
<td>1.93 (.01)</td>
<td>-.24 (1.34)</td>
<td>5.2 (2.6)</td>
<td>-.061 (.005)</td>
<td>.000 (1.48)</td>
</tr>
<tr>
<td>Y</td>
<td>1</td>
<td>1</td>
<td>.001</td>
<td>1.97 (.00)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YY_D</td>
<td>.894 (1.82)</td>
<td>1.965 (.006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YY_J</td>
<td>1</td>
<td>.59 (.05)</td>
<td>6.8 (4.2)</td>
<td>3.2 (8.2)</td>
<td>1.87 (.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>1</td>
<td>1</td>
<td>.001</td>
<td>1.965 (.006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model 1: $y_{t+1} = \rho y_t + \eta_{t+1}$

Model 2: $y_{t+1} = \rho y_t + (1 - \rho) \eta_{t+1}$
Table 3: Parameter estimates with constrained equity premium $\mu_0 = 0$, $\mu_1 = R$. Model 2: $y_{\tau+1} = \rho_y y_{\tau} + (1 - \rho_y) \eta_{\tau+1}$

<table>
<thead>
<tr>
<th>Model</th>
<th>$\ln L$</th>
<th>$R$ (p-value)</th>
<th>$\mu_1$ (p-value)</th>
<th>$\sigma_p \sqrt{252}$ (p-value)</th>
<th>$\sqrt{\frac{\alpha}{\beta}}$ (p-value)</th>
<th>$\beta$ (p-value)</th>
<th>$\sigma$ (p-value)</th>
<th>$\rho_{sv}$ (p-value)</th>
<th>HL (mths)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV</td>
<td>74,999.67</td>
<td>2.57 (.63)</td>
<td>2.57 (.63)</td>
<td>.043 (.005)</td>
<td>.171 (.004)</td>
<td>7.78 (.50)</td>
<td>.556 (.014)</td>
<td>-.653 (.015)</td>
<td>1.1 (.1)</td>
</tr>
<tr>
<td>SVJ1</td>
<td>75,091.08</td>
<td>2.52 (.62)</td>
<td>2.52 (.62)</td>
<td>.037 (.005)</td>
<td>.172 (.004)</td>
<td>6.27 (.45)</td>
<td>.438 (.014)</td>
<td>-.684 (.016)</td>
<td>1.3 (.1)</td>
</tr>
<tr>
<td>SVJ2</td>
<td>75,095.66</td>
<td>2.47 (.65)</td>
<td>2.48 (.65)</td>
<td>.037 (.005)</td>
<td>.172 (.004)</td>
<td>6.22 (.49)</td>
<td>.456 (.015)</td>
<td>-.673 (.018)</td>
<td>1.4 (.1)</td>
</tr>
<tr>
<td>DEXP</td>
<td>75,093.12</td>
<td>2.53 (.62)</td>
<td>2.51 (.62)</td>
<td>.037 (.005)</td>
<td>.172 (.004)</td>
<td>6.13 (.45)</td>
<td>.463 (.015)</td>
<td>-.636 (.016)</td>
<td>1.4 (.1)</td>
</tr>
<tr>
<td>VG</td>
<td>75,093.56</td>
<td>2.61 (.59)</td>
<td>2.61 (.59)</td>
<td>.037 (.005)</td>
<td>.172 (.004)</td>
<td>6.06 (.44)</td>
<td>.458 (.015)</td>
<td>-.634 (.017)</td>
<td>1.4 (.1)</td>
</tr>
<tr>
<td>Y</td>
<td>75,092.45</td>
<td>2.49 (.62)</td>
<td>2.49 (.62)</td>
<td>.037 (.005)</td>
<td>.173 (.007)</td>
<td>5.66 (.43)</td>
<td>.444 (.021)</td>
<td>-.625 (.027)</td>
<td>1.6 (.1)</td>
</tr>
<tr>
<td>YY</td>
<td>75,096.11</td>
<td>2.44 (.63)</td>
<td>2.44 (.63)</td>
<td>.037 (.005)</td>
<td>.173 (.005)</td>
<td>5.67 (.47)</td>
<td>.449 (.015)</td>
<td>-.624 (.017)</td>
<td>1.5 (.1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>$f_{jump}$</th>
<th>$w_n$ (p-value)</th>
<th>$G$ (p-value)</th>
<th>$M$ (p-value)</th>
<th>$Y_n$ (p-value)</th>
<th>$Y_p$ (p-value)</th>
<th>$\lambda_i$ (p-value)</th>
<th>$\bar{\gamma}$ (p-value)</th>
<th>$\delta$ (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVJ1</td>
<td>.130 (.015)</td>
<td>.54 (.07)</td>
<td>55.5 (5.4)</td>
<td>50.7 (12.3)</td>
<td>-1</td>
<td>110.3 (18.8)</td>
<td>-0.001 (.003)</td>
<td>.034 (.002)</td>
<td></td>
</tr>
<tr>
<td>SVJ2</td>
<td>.137 (.026)</td>
<td>.54 (.03)</td>
<td>41.1 (5.0)</td>
<td>31.6 (9.7)</td>
<td>0</td>
<td>122.6 (20.4)</td>
<td>.000 (.002)</td>
<td>.031 (.002)</td>
<td></td>
</tr>
<tr>
<td>DEXP</td>
<td>.172 (.004)</td>
<td>.54 (.07)</td>
<td>55.5 (5.4)</td>
<td>50.7 (12.3)</td>
<td>-1</td>
<td>0.44 (.80)</td>
<td>-.197 (.046)</td>
<td>.006 (.114)</td>
<td></td>
</tr>
<tr>
<td>VG</td>
<td>.172 (.004)</td>
<td>.54 (.03)</td>
<td>41.1 (5.0)</td>
<td>31.6 (9.7)</td>
<td>0</td>
<td>122.6 (20.4)</td>
<td>.000 (.002)</td>
<td>.031 (.002)</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>1</td>
<td>.61 (.05)</td>
<td>6.8 (4.2)</td>
<td>3.0 (8.5)</td>
<td>1.87 (.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YY</td>
<td>1</td>
<td>.89 (.03)</td>
<td>2.4</td>
<td>71.0 (60.7)</td>
<td>1.934 (.009)</td>
<td>-1.97 (2.77)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$Parameter constraint $G \geq R$ was binding.
Table 4: Change of measure under a myopic power utility pricing kernel $d \ln M = \mu_0^* dt - R ds$.

<table>
<thead>
<tr>
<th></th>
<th>Objective</th>
<th>Risk-neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equity premium</td>
<td>$\mu_0 + \mu_1 V_t$</td>
<td>$\mu_0 = 0$, $\mu_1 = R$</td>
</tr>
<tr>
<td>General jump intensity</td>
<td>$\hat{k}(x) V_t$</td>
<td>$\hat{k}^*(x) V_t = \left[ \hat{k}(x) e^{-Rx} \right] V_t$</td>
</tr>
<tr>
<td>Merton parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean jump size</td>
<td>$\bar{\gamma}$</td>
<td>$\bar{\gamma}^* = \bar{\gamma} - R \hat{\delta}^2$ unchanged</td>
</tr>
<tr>
<td>jump SD</td>
<td>$\delta$</td>
<td></td>
</tr>
<tr>
<td>CGMY parameters</td>
<td>$G$</td>
<td>$G - R$ (must be $\geq 0$)</td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>$M + R$</td>
</tr>
<tr>
<td></td>
<td>$C_n, C_p, Y_n, Y_p$</td>
<td>unchanged</td>
</tr>
<tr>
<td>Variance process</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean reversion</td>
<td>$\beta$</td>
<td>$\beta + R \sigma_p$</td>
</tr>
<tr>
<td>UC mean</td>
<td>$\alpha/\beta$</td>
<td>$\alpha/\beta^*$</td>
</tr>
</tbody>
</table>
Figure 1. Normal probability plots for the normalized returns $z_{r+1} = N^{-1}[CDF(y_{r+1} | Y_r, \hat{\Theta})]$.

for different models.

Diagonal line: theoretical quantiles conditional upon correct specification

+: Empirical quantiles
Figure 2: Autocorrelation revision \( \hat{\rho}_{t+1|t+1} - \hat{\rho}_{t|t} \) conditional on observing \( y_{t+1} \), and conditional on \( y_t = \pm 1\% \)

Model 1

With \( y_t = +1\% \)

With \( y_t = -1\% \)

Model 2

With \( y_t = +1\% \)

With \( y_t = -1\% \)
Figure 3: Autocorrelation estimates $\rho_{q\ell}$ from YY model: models 1 and 2.
Figure 4: News impact curves for various models

The graph show the revision in estimated annualized standard deviation \( (E_{t+1} - E_t) \sqrt{V_{t+1}} \) conditional upon observing a standardized return of magnitude \( y_{t+1}/\sqrt{\hat{\sigma}_t^2}/252 \).

Figure 5: Estimates of permanent volatility shocks (model 2) – YY model.
Figure 6a. Unconditional probability density function estimates from various models, and direct data-based estimates from a histogram (.25% cell width).

Figure 6b. Unconditional tail probability estimates. The dotted lines give 95% confidence intervals, based upon 1000 simulations of the 1926-2006 data set under YY parameter estimates.
Figure 7. Unconditional tail probabilities and tail intensity functions versus $|y|$; log scales on both axes. Data-based estimates from excess returns’ residuals for 20,004 business days with estimated time horizons of approximately 1 day ($\pm 25\%$). Dotted lines give 95% confidence intervals, based upon 1000 simulated sample paths under YY parameter estimates.
Figure 8: Estimated and observed ISD’s for options on S&P 500 futures: Dec. 29, 2006. 95% confidence intervals for SVJ2 model for parameter uncertainty (dark grey), and parameter & state uncertainty (light grey).