Stock Market Crash Risk, 1926 - 2006

David S. Bates

University of Iowa and the
National Bureau of Economic Research

April 23, 2008

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

How should we model 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.[1] 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.

---

[1]Li, 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 are found 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 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 arbitrary time intervals. 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 arrival rate) of jumps of size x. Alternatively and equivalently, Lévy processes can be described by their generalized Fourier transform

$$F(u) \equiv Ee^{uL(t)} = \exp[t f_{dL}(u)], \; u \in D_u \subset \mathbb{C} \tag{1}$$

where u is a complex-valued element of the set D_u for which (1) is well-defined. If Φ is real, $F(i\Phi)$ is the characteristic function of $L(t)$, while $tf_{dL}(\Phi)$ 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 $f_{dL}(u)$ will be called the *cumulant exponent* of $L(t)$.[2]

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

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

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

$$ES(t) = Ee^{L(t)} = e^{f_{dL}(1)t} = e^{\mu t}. \tag{3}$$

Intuitively, Lévy processes can be thought of as a drift term plus an infinite sum of independent point processes, each an exponential martingale of the form

$$d\ln L_x = -(e^{x} - 1)k(x)dt + x dN_x, \tag{4}$$

---

[2]Carr et al (2003) call $f_{dL}(i\Phi)$ the "unit time log characteristic function." Bertoin (1996) uses the *characteristic exponent*, which takes the form $\Psi(\Phi) \equiv -f_{dL}(i\Phi)$.

where N_x is a Poisson counter with intensity $k(x)$ that counts the number 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 $\int |x|\, k(x)\,dx < \infty$, 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 quadratic variation $\int x^2\, k(x)\,dx$, in order to be well-behaved. 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:

$$d\ln S_t = \mu\, dt + (\sigma\, dW_t - \tfrac{1}{2}\sigma^2 dt) + (\gamma\, dN_t - \lambda \bar{k}\, dt) \qquad \textbf{(5)}$$

where W_t is a Wiener process,

N_t is a Poisson counter with intensity λ,

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

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

The associated intensity of jumps of size x is

$$k(x) = \frac{\lambda}{\sqrt{2\pi\delta^2}} \exp\left[-\frac{(x-\bar{\gamma})^2}{2\delta^2}\right] \qquad \textbf{(6)}$$

while the cumulant exponent takes the form

$$f_{Merton}(u) = (\mu - \lambda\bar{k})u + ½\sigma^2(u^2 - u) + \lambda(e^{\bar{\gamma}u + ½\delta^2u^2} - 1).$$

The approach can be generalized to allow alternate distributions for γ – in particular, a mixture of normals:

$$k(x) = \sum_{i=1}^{2} \frac{\lambda_i}{\sqrt{2\pi\delta_i^2}} \exp\left[-\frac{(x - \bar{\gamma}_i)^2}{2\delta_i^2}\right]. \qquad \textbf{(7)}$$

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

$$k(x) = \begin{cases} C_n e^{-G|x|} |x|^{-1-Y_n} \; \textit{for } x < 0 \\ C_p e^{-M|x|} |x|^{-1-Y_p} \; \textit{for } x > 0 \end{cases} \qquad \textbf{(8)}$$

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

$$f_{CGMY}(u) = (\mu - \omega)u + V\left\{w_n \frac{(G - u)^{Y_n} - G^{Y_n}}{Y_n(Y_n - 1)G^{Y_n - 2}} + (1 - w_n)\frac{(M + u)^{Y_p} - M^{Y_p}}{Y_p(Y_p - 1)M^{Y_p - 2}}\right\} \qquad \textbf{(9)}$$

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

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

$$C_n = \frac{w_n V}{\Gamma(2 - Y_n)G^{Y_n - 2}}, \quad C_p = \frac{(1 - w_n)V}{\Gamma(2 - Y_p)M^{Y_p - 2}} \qquad \textbf{(10)}$$

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, $Y = -1$ 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

$$f_{CGMY}(u) = (\mu - \tfrac{1}{2}V)u + \tfrac{1}{2}Vu^2. \tag{11}$$

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 ($w_n = 1$). 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 $w_1 k_1(x) + w_2 k_2(x)$ of Lévy densities for $w_1, w_2 \geq 0$ is also a valid Lévy density, and generates an associated weighted cumulant exponent of the form $w_1 f_1(u) + w_2 f_2(u)$.

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

$$\begin{aligned} \ln F(u) &= f_{dL}(u)t \\ &= \frac{f_{dL}(u)}{f_{dL}^{\prime\prime}(0)} V t \end{aligned} \tag{12}$$

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. This article will initially assume that the log asset price $s_t \equiv \ln S_t$ follows a process of the form

$$\begin{aligned} ds_t &= (\mu_0 + \mu_1 V_t)dt + \left(\rho_{sv}\sqrt{V_t}\,dW_t - \tfrac{1}{2}\rho_{sv}^2 V_t\,dt\right) + (dL_t - \omega V_t\,dt) \\ dV_t &= (\alpha - \beta V_t)dt + \sigma\sqrt{V_t}\,dW_t \end{aligned} \qquad \textbf{(13)}$$

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 $(1 - \rho_{sv}^2)V_t\,dt$. The term $\omega V_t\,dt \equiv E_t e^{dL_t} - 1$ 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 $y_{t+\tau} \equiv \int_t^T ds$ as the discrete-time return observed over horizon $\tau = T - t$, and define $f_{dL}(u) \equiv (1 - \rho_{sv}^2) V_t \, g_{dL}(u)$ as the cumulant exponent of $dL_t - \omega V_t \, dt$. By construction, $g_{dL}(u)$ is a standardized cumulant exponent, with $g_{dL}(1) = 0$ and variance $g_{dL}^{//}(0) = 1$. A key property of affine models is the ability to compute the conditional generalized Fourier transform of $(y_{t+\tau}, V_T)$. This can be done by iterated expectations, conditioning initially on the future variance path:

$$
\begin{aligned}
F(\Phi, \psi \mid V_t, \tau) &\equiv E\left(e^{\Phi y_{t+\tau} + \psi V_T} \mid V_t\right) \\
&= E\left\{E\left[e^{\Phi \mu_0 \tau + \int_t^T \Phi\left((\mu_1 - \frac{1}{2}\rho_{sv}^2) V_s ds + \rho_{sv}\sqrt{V_s} dW_s + (dL_s - \omega V_s ds)\right) + \psi V_T} \mid \{V_s\}_{s=t}^{T}\right] \mid V_t\right\} \\
&= E\left[e^{\Phi \mu_0 \tau + \int_t^T \left[\mu_1 + \frac{1}{2}\rho_{sv}^2(\Phi^2 - \Phi) + (1 - \rho_{sv}^2) g_{dL}(\Phi)\right] V_s ds + \psi V_T} \mid V_t\right] \\
&\equiv E\left[e^{\Phi \mu_0 \tau + h(\Phi) \int_t^T V_s ds + \psi V_T} \mid V_t\right]
\end{aligned}
\qquad \textbf{(14)}
$$

for $h(\Phi) \equiv \mu_1 + \frac{1}{2}\rho_{sv}^2 (\Phi^2 - \Phi) + (1 - \rho_{sv}^2) g_{dL}(\Phi)$. This is the generalized Fourier transform of the future spot variance V_T and the *average* future variance $\overline{V}_{t \to T} \equiv \frac{1}{\tau}\int_t^T V_s ds$. 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, $F(\bullet \mid V_t, \tau)$ solves the Feynman-Kac partial differential equation

$$
-F_\tau + (\alpha - \beta V_t) F_V + \tfrac{1}{2}\sigma^2 V_t F_{VV} = -h(\Phi) V_t F \qquad \textbf{(15)}
$$

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

$$
F(\Phi, \psi \mid V_t, \tau) = \exp[C(\tau; \Phi, \psi) + D(\tau; \Phi, \psi) V_t] \qquad \textbf{(16)}
$$

where

$$C(\tau;\,\Phi,\,\psi,\,\xi) \;=\; \mu_0\Phi\tau \;-\; \frac{\alpha\tau}{\sigma^2}\,(\rho_{sv}\sigma\Phi \,-\, \beta \,-\, \gamma)$$
$$-\;\frac{2\alpha}{\sigma^2}\,\ln\left[1 \,+\, \tfrac{1}{2}(\rho_{sv}\sigma\Phi - \beta - \gamma)\,\frac{1-e^{\gamma\tau}}{\gamma}\right] \;-\; \frac{2\alpha}{\sigma^2}\,\ln[1 \,-\, K(\Phi)\psi] \tag{17}$$

$$D(\tau;\,\Phi,\,\psi,\,\xi) \;=\; \frac{-2(\mu_1 - \tfrac{1}{2})\Phi \,-\, \Phi^2}{\rho_{sv}\sigma\Phi \,-\, \beta \;+\; \gamma\,\dfrac{1+e^{\gamma\tau}}{1-e^{\gamma\tau}}} \;+\; \frac{\Lambda(\Phi)\,\psi}{1 \,-\, K(\Phi)\,\psi} \tag{18}$$

$$\gamma \;=\; \sqrt{(\rho_{sv}\sigma\Phi \,-\, \beta)^2 \;-\; 2\sigma^2 h(\Phi)} \tag{19}$$

$$\Lambda(\Phi) \;=\; \frac{\left(\dfrac{e^{\gamma\tau}+1}{e^{\gamma\tau}-1}\right)^2 \,-\, 1}{\left(\dfrac{e^{\gamma\tau}+1}{e^{\gamma\tau}-1} \,+\, \dfrac{\beta - \rho\sigma\Phi}{\gamma}\right)^2} \tag{20}$$

$$K(\Phi) \;=\; \frac{\sigma^2}{\gamma\dfrac{e^{\gamma\tau}+1}{e^{\gamma\tau}-1} \,+\, \beta \,-\, \rho\sigma\Phi}\,. \tag{21}$$

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. Consequently, it is assumed that daily log-differenced stock index excess returns y_t can be described by the following stochastic process:

$$\begin{aligned} y_{t+1} &= \rho_t y_t \,+\, \int_t^{t+\tau_t} ds_t \\ V_{t+1} &= V_t \,+\, \int_t^{t+\tau_t} dV_t \\ \rho_{t+1} &= \rho_t \,+\, \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim N(0,\sigma_\rho^2) \;\; \textit{and} \;\; \textit{i.i.d.} \end{aligned} \tag{22}$$

where τ_t is the effective length of a business day,

ρ_t is the daily autocorrelation of stock index returns,

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

$V_t\, dt \equiv Var_t(ds_t)$ is the instantaneous conditional variance of ds_t.
The intradaily shocks (ds_t, dV_t) are given by (13) above.

The model adds an autocorrelation state variable ρ_t that captures the fact that autocorrelations of stock market returns are not constant over time.[3] Following the literature on time-varying coefficient models, the autocorrelation is modeled as a simple random walk, to avoid constraining estimates of ρ_t. Estimation of the autocorrelation volatility parameter σ_ρ endogenously determines the appropriate degree of smoothing to use when filtering the current autocorrelation value ρ_t from past data.

Furthermore, the length τ_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 τ_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).[4] 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%.

Given the above stochastic process, the cumulant generating function of future returns and state variable realizations *conditional* upon current values is analytic, and of the semi-affine form

$$\begin{aligned} \ln F(\Phi, \xi, \psi \mid y_t, \rho_t, V_t) &\equiv \ln E\left[e^{\Phi y_{t+1} + \xi \rho_{t+1} + \psi V_{t+1}} \mid y_t, \rho_t, V_t\right] \\ &= C(\tau_t; \Phi, \xi, \psi) + (\xi + \Phi y_t)\rho_t + D(\tau_t; \Phi, \psi) V_t \end{aligned} \qquad \textbf{(23)}$$

where $C(\tau; \xi, \Phi, \psi) = C(\tau; \Phi, \psi) + \frac{1}{2}\sigma_\rho^2 \xi^2$, and

---

[3]See, e.g., Andersen, Benzoni and Lund (2002, Table I), who estimate different autocorrelations for 1953-96 and 1980-96.

[4]Gallant, Rossi and Tauchen (1992) use a similar approach, and also estimate monthly seasonals.

$C(\tau;\, \Phi,\, \psi)$ and $D(\tau;\, \Phi,\, \psi)$ are given in (17) and (18) above.

I.3 Filtration and maximum likelihood estimation

If the state variables $(\rho_t,\, V_t)$ 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 $F(i\Phi,\, i\xi,\, i\psi \mid y_t,\, \rho_t,\, V_t)$, and to use this in a maximum likelihood procedure to estimate the parameters of the stochastic process. However, since $(\rho_t,\, V_t)$ are latent rather than directly observed, this is a *hidden Markov* model that must be estimated by other means.

Since the cumulant generating function (23) is affine in the latent state variables $(\rho_t,\, V_t)$, 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 $\boldsymbol{Y}_t \equiv \{y_1, y_2, \ldots, y_t\}$ as the data observed up through period *t*, and define

$$G_{t|t}(i\xi,\; i\psi) \;\equiv\; E\left[e^{i\xi\rho_t + i\psi V_t} \mid \boldsymbol{Y}_t\right] \tag{24}$$

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

$$p(y_{t+1} \mid \boldsymbol{Y}_t) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} G_{t|t}[\Phi y_t,\, D(\tau_t;\, i\Phi,\, 0)]\, e^{C(\tau_t;\, i\Phi,\, 0,\, 0) \,-\, i\Phi y_{t+1}}\, d\Phi. \tag{25}$$

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

$$G_{t+1|t+1}(i\xi,\, i\psi) \;=\; \frac{1}{2\pi p(y_{t+1} \mid \boldsymbol{Y}_t)}\int_{-\infty}^{\infty} G_{t|t}[i\xi + i\Phi y_t,\, D(\tau_t;\, i\Phi,\, i\psi)]\, e^{C(\tau_t,\, i\Phi,\, i\xi,\, i\psi) \,-\, i\Phi y_{t+1}}\, d\Phi. \tag{26}$$

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 $\Sigma \ln p(y_{t+1} | \boldsymbol{Y}_t)$ 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:

$$E[\rho_{t+1}^m V_{t+1}^n \mid \boldsymbol{Y}_{t+1}] = \left. \frac{\partial^{m+n} G_{t+1|t+1}(\xi, \psi)}{\partial \xi^m \partial \psi^n} \right|_{\xi = \psi = 0} . \qquad \textbf{(27)}$$

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 $G_{t|t}(\cdot)$ based on point-by-point univariate numerical integrations. As such a procedure would be slow, the AML methodology instead approximates $G_{t|t}(\cdot)$ 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 $\hat{G}_{t|t}(\cdot)$ and a datum y_{t+1}, (27) is used to compute the posterior moments of (ρ_{t+1}, V_{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 $G_{t|t}(\cdot)$, 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

$$\ln \hat{G}_{t|t}(\xi, \psi) = [\hat{\rho}_{t|t}\xi + \tfrac{1}{2} W_{t|t}\xi^2] - \nu_t \ln(1 - \kappa_t \psi). \tag{28}$$

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

	Autocorrelation ρ_t	spot variance V_t
Distribution	$\rho_t \mid \boldsymbol{Y}_t \sim N(\hat{\rho}_{t\mid t}, W_{t\mid t})$	$V_t \mid \boldsymbol{Y}_t \sim (\hat{V}_{t\mid t}, P_{t\mid t})$
conditional cumulant generating function	$\ln E\left[e^{\xi\rho_t} \mid \boldsymbol{Y}_t\right] = \hat{\rho}_{t\mid t}\xi + \tfrac{1}{2} W_{t\mid t}\xi^2$	$\ln E\left[e^{\psi V_t} \mid \boldsymbol{Y}_t\right] = -\nu_t \ln(1 - \kappa_t\psi)$ $\kappa_t\nu_t = \hat{V}_{t\mid t},\ \kappa_t^2\nu_t = \hat{P}_{t\mid t}$
initial CGF	$\rho_1 \mid \boldsymbol{Y}_t \sim N(0,\ 10^2)$	$\kappa_1 = \dfrac{\sigma^2}{2\beta},\ \nu_1 = \dfrac{2\alpha}{\sigma^2}$

$\rho_t, V_t \mid \boldsymbol{Y}_t$ assumed independent for all t.

Initial variance was assumed drawn from its unconditional gamma distribution, with the parameters (κ_1, ν_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 $\theta_t \equiv (\hat{\rho}_{t|t}, W_{t|t}; \kappa_t, \nu_t)$ – or, equivalently the moments $(\hat{\rho}_{t|t}, W_{t|t}; \hat{V}_{t|t}, P_{t|t})$ – 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 Φ_{max} was computed for each integral which upper truncation error would be less than 10^{-10} in magnitude. The integrands were then integrated over $(-\Phi_{max}, \Phi_{max})$ 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 Φ are the complex conjugates of the integrands evaluated at positive Φ. On average between 234 and 448 evaluations of the integrand were required for each integration.[5]

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, 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

$$y_t = \ln(1 + R_t) - \frac{\ln(1+i)}{N} n_t \tag{29}$$

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*.

[5] By contrast, the FFT approach used in Carr et al (2002) requires 16,384 functional evaluations.

II.2 Parameter estimates

Table I describes and provides estimates of the time dummies from the time-changed CGMY model,[6] 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.[7] 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).[8] This is qualitatively similar to but less pronounced than the doubling of weekend variance found by Barclay, 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 II reports estimates for various models, while Figure 1 presents associated normal probability plots. As noted above, all models capture the leverage effect by a correlation ρ_{sv} with

---

[6]Estimates from other specifications were virtually identical, with estimates typically within ±0.01 of the CGMY model's estimates.

[7]Saturday trading was standard before 1945. Over 1945-51, it was eliminated in summer months, and was permanently eliminated on June 1, 1952.

[8]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.

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 Y_p, Y_n parameters – whether they are set to specific levels, and whether they diverge for the intensities of positive versus negative jumps. The DEXP model with $Y_p = Y_n = -1$ 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.[9] 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 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.

---

[9]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 [-.20, 0] observed range of negative log stock market excess returns, and is therefore observationally equivalent to the log stable specification.

Apart from the LS model, all models have similar estimates for the parameters determining the conditional mean and stochastic variance evolution. The parameter μ_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 $k(x)$ 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 $\rho_t \mid \boldsymbol{Y}_t$ is assumed $N(\hat{\rho}_{t|t}, W_{t|t})$, it can be shown that the autocorrelation filtration algorithm (27) updates conditional moments as follows:

$$\hat{\rho}_{t+1|t+1} = \hat{\rho}_{t|t} - y_t W_{t|t} \frac{\partial \ln p(y_{t+1} \mid \boldsymbol{Y}_t)}{\partial y_{t+1}} \tag{30}$$

$$W_{t+1|t+1} = \sigma_\rho^2 + (y_t W_{t|t})^2 \frac{\partial^2 \ln p(y_{t+1} \mid \boldsymbol{Y}_t)}{\partial y_{t+1}^2} \tag{31}$$

If $y_{t+1} \mid \boldsymbol{Y}_t$ 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.[10] The partials of log densities can be computed numerically by Fourier inversion.

Figure 2 illustrates the autocorrelation filtrations estimated under various models. 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

[10]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.

upper tail, its optimal filtration downweights the information in large negative returns but not in large positive returns.

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.[11] After 1971, autocorrelation estimates fell steadily, and became insignificantly different from zero after 2002.

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,[12] 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 $E_{t+1}\sqrt{V_{t+1}}$ is updated for the various models. The conditional volatility revisions use median parameter values $(\kappa_t, \nu_t) = (.00295, 5.85)$

---

[11]See LeBaron (1992, Figure 1) for annual estimates of the daily autocorrelation of S&P composite index returns over 1928-1990.

[12]I am indebted to Ruslan Goyenko for pointing this out.

for the prior gamma distribution of V_t, implying a conditional mean $\kappa_t \nu_t = (.131)^2$ that is close to the $(.129)^2$ median value observed for $\hat{V}_{t|t}$ estimates from the YY model.[13] 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).

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.[14] 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, the associated conditional standard deviation, and the deviations from the YY estimates for other models.[15] Volatility estimates from all models except LS are similar – as,

---

[13]As $\hat{V}_{t|t}$ estimates have substantial positive skewness, the median is substantially below the mean estimate of $(.159)^2$ reported in Table 2.

[14]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.

[15]"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.

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-factor stochastic variance model is too simple, and suggests that multifactor specifications of variance evolution are worth exploring.[16]

II.5 Unconditional distributions

A final 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 Paretian 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

[16]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.

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 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.[17] Unsurprisingly, the confidence intervals on extreme tail events are quite wide. However, the underestimation of moves of 3% - 7% in magnitude is again apparent, and is statistically significant. Thus, the best-fitting YY model fails to capture important features of stock market excess returns. Given that the normal probability plots in Figure 1 indicate the models are capturing *conditional* distributions reasonably well, it seems likely that model errors are entering via the distributional randomization involved when going from conditional to unconditional distributions; parameter instability, for instance, or misspecification of the variance process. In particular, I suspect that the 1-factor diffusive variance process in equation (13) is failing to capture the actual distribution of variance realizations, leading to more 3-7% moves than predicted by the model.

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

[17]Conditional 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.

$$
\begin{aligned}
K(y) \equiv \int_{-\infty}^{y} k(x)\,dx &= C_n G^{Y_n} \Gamma(-Y_n, G|y|) \\
&\approx C_n \frac{y^{-Y_n}}{Y_n} \;\; for \;\; G \approx 0
\end{aligned} \qquad \textbf{(46)}
$$

where $C_n = w_n(1-\rho_{sv}^2)\,\tau\,\alpha/[\beta\Gamma(2-Y_n)G^{Y_n-2}]$ and $\Gamma(a,z)$ 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.

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.

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. 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

Andersen, Torben G., Luca Benzoni, and Jesper Lund (2002). "An Empirical Investigation of Continuous-Time Equity Return Models." *Journal of Finance* **57**, 1239-1284.

Bakshi, Gurdip and Dilip B. Madan (2000). "Spanning and Derivative-Security Valuation." *Journal of Financial Economics* **55**, 205-238.

Barclay, Michael J., Robert H. Litzenberger, and Jerold B. Warner (1990). "Private Information, Trading Volume, and Stock-Return Variances." *Review of Financial Studies* **3**, 233-254.

Bates, David S. (2006). "Maximum Likelihood Estimation of Latent Affine Processes." *Review of Financial Studies* **19**, 909-965.

Bertoin, Jean (1996). *Lévy Processes,* Cambridge: Cambridge University Press.

Carr, Peter, Hélyette Geman, Dilip B. Madan, and Marc Yor (2002). "The Fine Structure of Asset Returns: An Empirical Investigation." *Journal of Business* **75**, 305-332.

Carr, Peter, Hélyette Geman, Dilip B. Madan, and Marc Yor (2003). "Stochastic Volatility for Lévy Processes." *Mathematical Finance* **13**, 345-382.

Carr, Peter and Liuren Wu (2003). "The Finite Moment Log Stable Process and Option Pricing." *Journal of Finance* **58**, 753-777.

Carr, Peter and Liuren Wu (2004). "Time-changed Lévy Processes and Option Pricing." *Journal of Financial Economics* **71**, 113-141.

Clark, Peter K. (1973). "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices." *Econometrica* **41**, 135-155.

Cox, John C., Stephen A. Ross, and Mark Rubinstein (1979). "Option Pricing: A Simplified Approach." *Journal of Financial Economics* **7,** 229-263.

Dimson, Elroy (1979). "Risk Measurement When Shares are Subject to Infrequent Trading." *Journal of Financial Economics* **7**, 197-226.

Duffie, Darrell, Jun Pan, and Kenneth J. Singleton (2000). "Transform Analysis and Asset Pricing for Affine Jump-Diffusions." *Econometrica* **68**, 1343-1376.

Eberlein, Ernst, Ulrich Keller, and Karsten Prause (1998). "New Insights into Smile, Mispricing, and Value at Risk: The Hyperbolic Model." *Journal of Business* **71**, 371-405.

Eraker, Bjorn, Michael Johannes, and Nicholas G. Polson (2003). "The Impact of Jumps in Volatility and Returns." *Journal of Finance* **58**, 1269-1300.

French, Kenneth R. and Richard Roll (1986). "Stock Return Variances: The Arrival of Information and the Reaction of Traders." *Journal of Financial Economics* **17**, 5-26.

Gallant, A. Ronald, Peter E. Rossi, and George Tauchen (1992). "Stock Prices and Volume." *Review of Financial Studies* **5**, 199-242.

Hentschel, Ludger (1995). "All in the Family: Nesting Symmetric and Asymmetric GARCH Models." *Journal of Financial Economics* **39**, 71-104.

Heston, Steve L. (1993). "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options." *Review of Financial Studies* **6**, 327-344.

Jukivuolle, Esa (1995). "Measuring True Stock Index Value in the Presence of Infrequent Trading." *Journal of Financial and Quantitative Analysis* **30**, 455-464.

Kou, Steve (2002). "A Jump Diffusion Model for Option Pricing." *Management Science* **48**, 1086-1101.

LeBaron, Blake D. (1992). “Some Relations between Volatility and Serial Correlations in Stock Returns.” *Journal of Business* **65**, 199-219.

Li, Haitao, Martin T. Wells, and Cindy L. Yu (2006). "A Bayesian Analysis of Return Dynamics with Stochastic Volatility and Lévy Jumps." University of Michigan working paper, January.

Lo, Andrew W. and A. Craig MacKinlay (1988). "Stock Market Prices Do Not Follow Random Walks: Evidence from a New Specification Test." *Review of Financial Studies* **1**, 41-66.

Maheu, John M. And Thomas H. McCurdy (2004). “News Arrival, Jump Dynamics and Volatility Components for Individual Stock Returns.” *Journal of Finance* **59**, 755-793.

Madan, Dilip B. and Eugene Seneta (1990). "The Variance Gamma (V.G.) Model for Share Market Returns." *Journal of Business* **63**, 511-525.

Mandelbrot, Benoit B. (1963). "The Variation of Certain Speculative Prices." *Journal of Business* **36**, 394-419.

Mandelbrot, Benoit B. and Richard L. Hudson (2004). *The (mis)Behavior of Markets: A Fractal View of Risk, Ruin, and Reward*. New York: Basic Books.

Masreliez, C. J. (1975). “Approximate Non-Gaussian Filtering with Linear State and Observation Relations.” *IEEE Transactions on Automatic Control* **20**, 107-110.

Merton, Robert C. (1976). "Option Pricing When Underlying Stock Returns are Discontinuous." *Journal of Financial Economics* **3**, 125-144.

SBBI Yearbook, 2006. Chicago: R. G. Ibbotson Associates.

Schick, Irvin C. And Sanjoy K. Mitter (1994). “Robust Recursive Estimation in the Presence of Heavy-tailed Observation Noise.” *The Annals of Statistics* **22**, 1045-1080.

Wu, Liuren (2006). "Dampened Power Law: Reconciling the Tail Behavior of Financial Security Returns." *Journal of Business* **79**, 1445-1473.

Table 1 Effective length of time intervals over 1926-2006, *relative* to 1-day Wednesday returns

#days	Description	NOBS	estimate	std. error
1	Monday close → Tuesday close	3831	1.02	(.04)
1	**Tuesday close → Wednesday close**	**4037**	**1**	
1	Wednesday → Thursday	3998	.94	(.03)
1	Thursday → Friday	3924	.93	(.03)
1	Friday → Saturday (1926-52)	1141	.43	(.02)
2	Saturday close → Monday close (1926-52)	1120	1.05	(.05)
2	Weekday holiday	341	1.25	(.11)
2	Wednesday exchange holiday in 1968	22	.73	(.33)
3	Weekend and/or holiday[a]	2755	1.10	(.04)
4	Holiday weekend	343	1.58	(.14)
5	Holiday weekend	6	1.31	(1.00)
		21518		

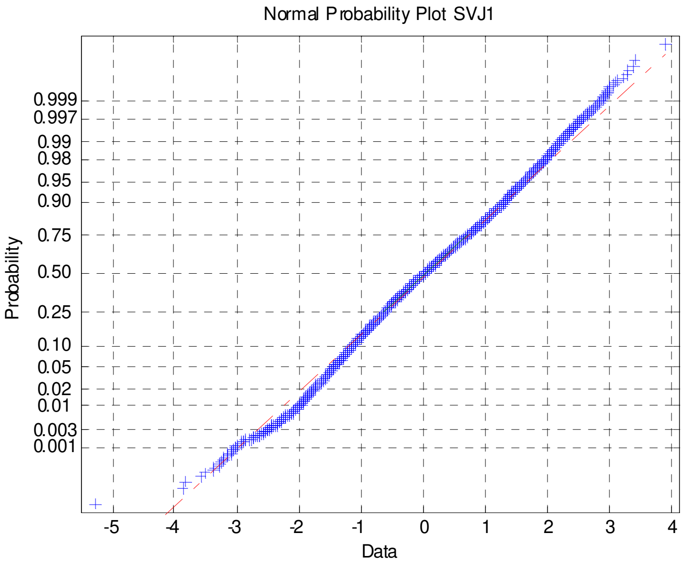
[a]Includes one weekday holiday (August 14 - 17, 1945)

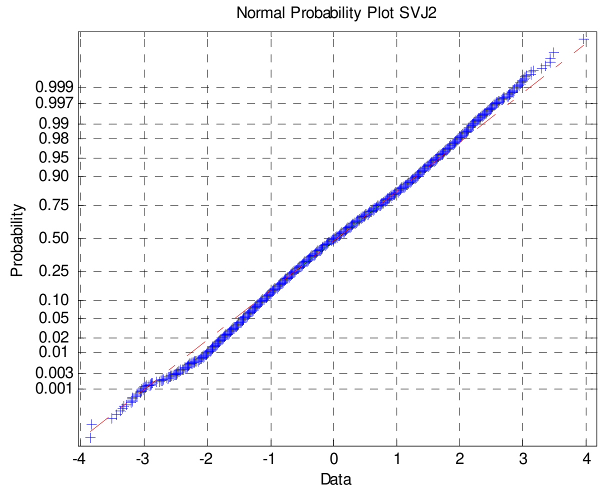
Table 2: Parameter estimates for various models. Standard errors in parentheses

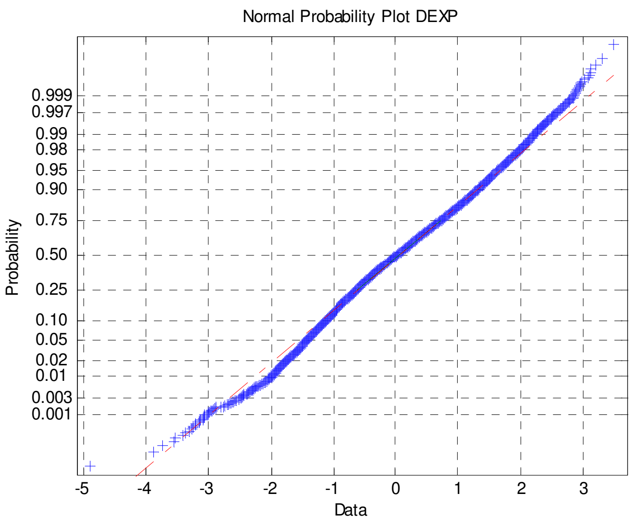
Model	Conditional mean			Stochastic volatility					
	μ_0	μ_1	$\sigma_\rho\sqrt{252}$	α	β	σ	ρ_{sv}	$\sqrt{V_{UC}}$	HL (mths)
SVJ1	.040 (.015)	.94 (.91)	.030 (.006)	.104 (.007)	4.33 (.40)	.370 (.011)	-.642 (.020)	.155 (.005)	1.9 (.2)
SVJ2	.042 (.015)	.87 (.92)	.030 (.007)	.105 (.008)	4.34 (.37)	.371 (.011)	-.642 (.020)	.155 (.005)	1.9 (.2)
DEXP	.044 (.015)	.79 (.91)	.030 (.006)	.104 (.008)	4.25 (.40)	.370 (.012)	-.588 (.020)	.156 (.005)	2.0 (.2)
VG	.042 (.015)	.91 (.91)	.030 (.006)	.103 (.008)	4.25 (.39)	.368 (.012)	-.587 (.020)	.156 (.005)	2.0 (.2)
Y	.042 (.015)	.91 (.92)	.030 (.006)	.096 (.011)	3.90 (.38)	.350 (.019)	-.577 (.032)	.157 (.008)	2.1 (.2)
YY	**.042 (.015)**	**.87 (.91)**	**.030 (.006)**	**.101 (.011)**	**4.00 (.38)**	**.362 (.019)**	**-.572 (.031)**	**.159 (.008)**	**2.1 (.2)**
YY_D	.042 (.015)	.87 (.91)	.030 (.006)	.101 (.013)	4.01 (.38)	.363 (.021)	-.572 (.036)	.159 (.010)	2.1 (.2)
YY_J	.041 (.015)	.97 (.92)	.030 (.007)	.095 (.008)	3.99 (.38)	.350 (.012)	-.586 (.020)	.154 (.005)	2.1 (.2)
LS	.019 (.015)	1.71 (.78)	.031 (.007)	.123 (.009)	4.52 (.39)	.405 (.011)	-.554 (.020)	.165 (.005)	1.8 (.2)

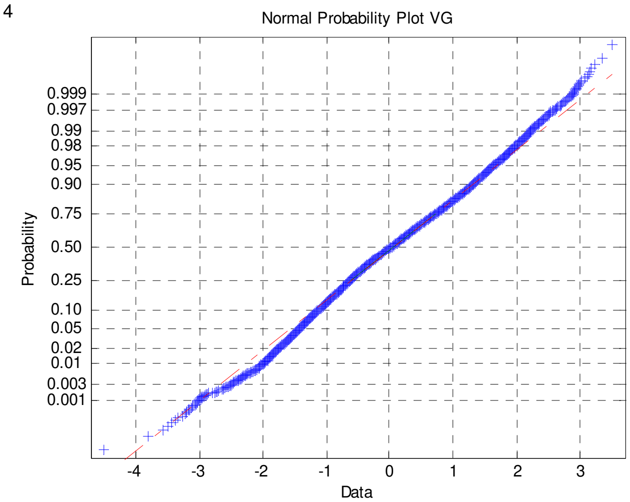
	f_{jump}	jump parameters								ln L
		w_n	G	M	Y_n	Y_p	λ_i	$\overline{\gamma}$	δ	
SVJ1	.140 (.015)						152.4 (24.8)	.000 (.002)	.030 (.002)	75,044.60
SVJ2	.156 (.022)						162.8 (29.3)	.000 (.000)	.029 (.003)	75,049.07
							0.5 (0.7)	-.189 (.083)	.005 (.028)	
DEXP	.256 (.030)	.49 (.07)	66.1 (6.0)	45.4 (10.1)	-1					75,047.62
VG	.274 (.030)	.52 (.07)	41.1 (5.4)	31.6 (9.1)	0					75,049.48
Y	1	.59 (.06)	7.0 (4.5)	2.3 (7.2)	1.87 (.03)					75,050.12
YY	**1**	**.88 (.03)**	**1.6 (4.2)**	**40.1 (31.0)**	**1.93 (.01)**	**-.24 (1.34)**				**75,052.90**
YY_D	.894 (1.82)	.86 (.29)	1.6 (6.5)	41.2 (31.6)	1.92 (.24)	-.30 (1.39)				75,052.90
YY_J	1	.87 (.02)	6.3 (5.3)	39.5 (15.4)	1.94 (.01)	-.21 (.74)	5.2 (2.6)	-.061 (.005)	.000 (1.48)	75,054.94
LS	1	1	.001		1.97 (.00)					75,005.53

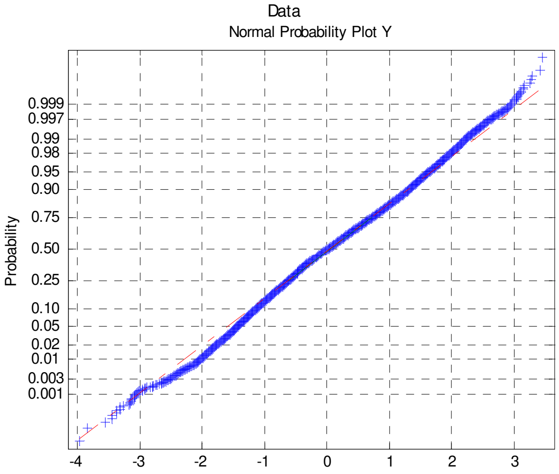
Data: daily CRSP value-weighted excess returns, 1926-2006. See equations (6) - (10), (13), and (22) for definitions of parameters.
Models with $f_{jump} < 1$ combine Lévy jump processes with an additional independent diffusion, with variance proportions $(f_{jump}, 1 - f_{jump})$, respectively.

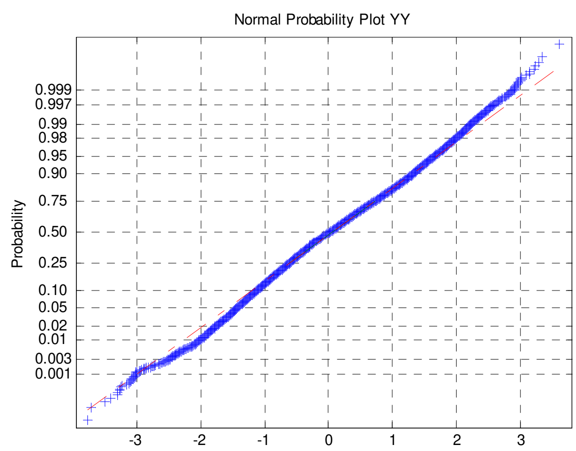












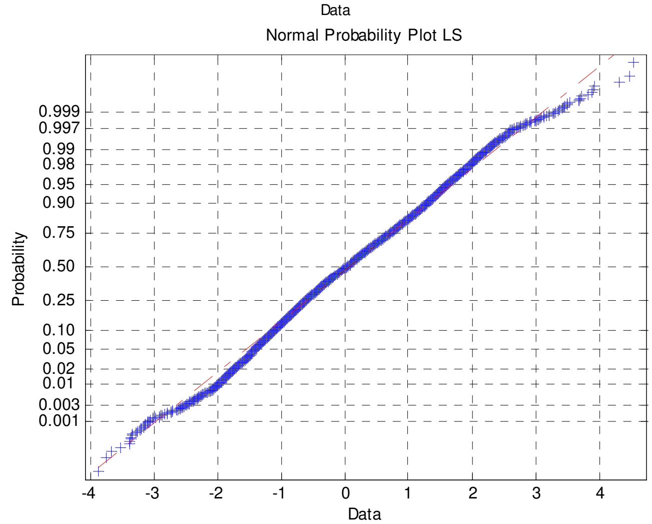


Figure 1. Normal probability plots for the normalized returns $z_{t+1} \equiv N^{-1}[CDF(y_{t+1} \mid \mathbf{Y}_t, \hat{\boldsymbol{\Theta}})]$, for different models.

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 $y_t = \pm 1\%$

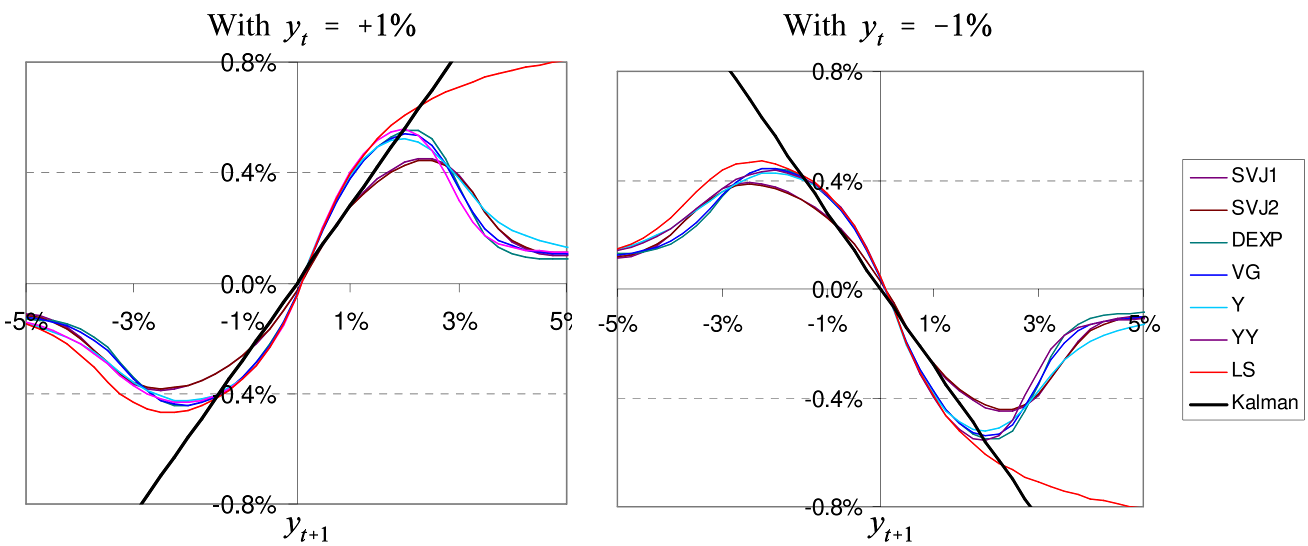


Figure 3: Autocorrelation estimates $\hat{\rho}_{t|t}$ from YY model, conditional standard deviations, and autocorrelation estimates' divergences from YY estimates for other models

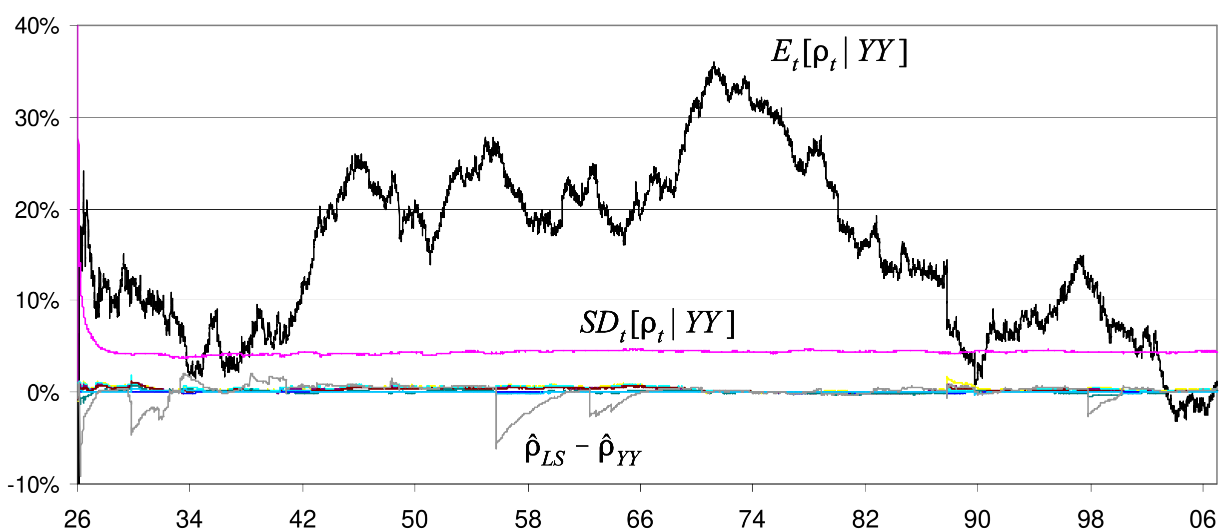
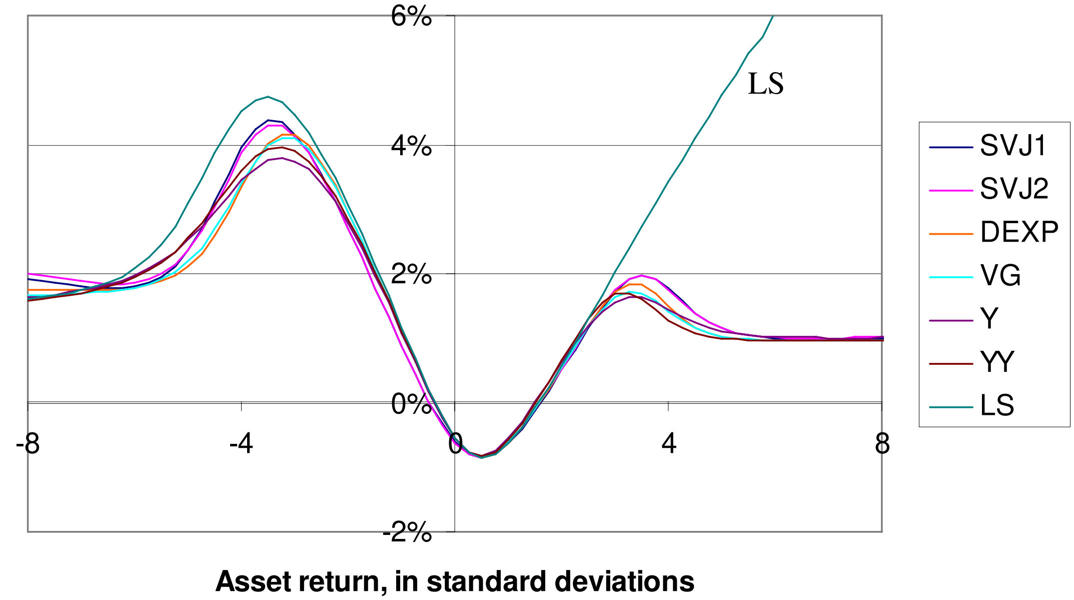


Figure 4: News impact curves for various models



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 $y_{t+1} / \sqrt{\hat{V}_{t|t} / 252}$.

Figure 5: Volatility estimates (YY model), associated conditional standard deviations, and deviations from YY estimates for other models

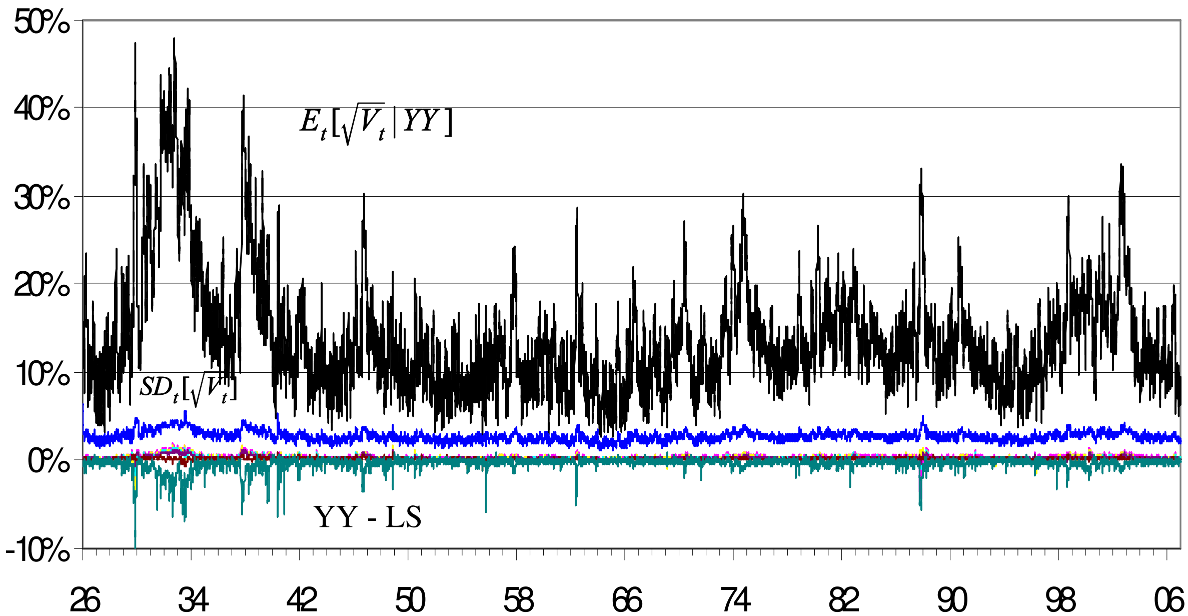


Figure 6a. Unconditional probability density function estimates from various models, and direct data-based estimates from a histogram (.25% cell width).

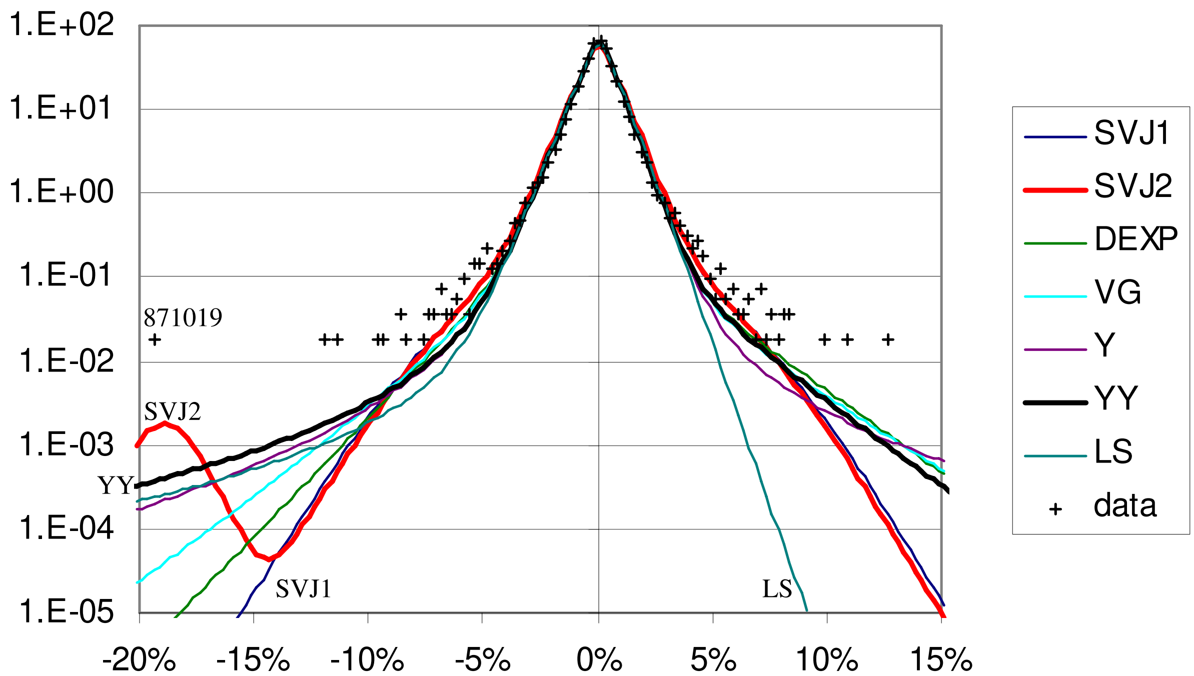


Figure 6b. Unconditional tail probability estimates. The dotted lines give 95% confidence intervals, based upon 1000 simulations of the 1926-2006 data set under YY parameter estimates.

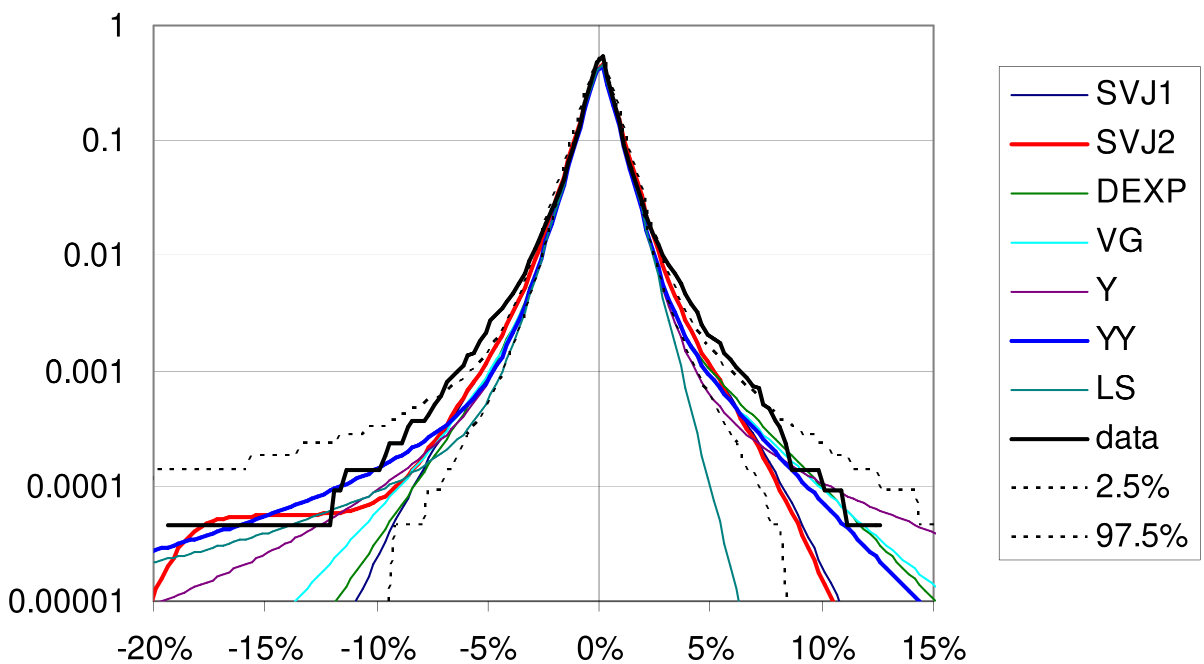


Figure7. Unconditional tail probabilities and tail intensity functions versus $|y|$; log scales on both axes. Data-based estimates from excess returns' residuals for 20,004 business days with estimated time horizons of approximately 1 day (±25%). Dotted lines give 95% confidence intervals, based upon 1000 simulated sample paths under YY parameter estimates.