

# **No-Arbitrage Semi-Martingale Restrictions for Continuous-Time Volatility Models subject to Leverage Effects and Jumps: Theory and Testable Distributional Implications\***

**Torben G. Andersen**  
Kellogg School, Northwestern University and NBER

**Tim Bollerslev**  
Duke University and NBER

**Dobrislav Dobrev**  
Kellogg School, Northwestern University

**PRELIMINARY AND INCOMPLETE FIRST DRAFT**

**September 2005**

---

\* This research was supported by a grant from the National Science Foundation to the NBER. We thank participants at the International Finance Conference at the University of Copenhagen, September 2005, and seminar participants at the Robert H. Smith School, University of Maryland for comments.

<sup>a</sup> Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208, and NBER,  
phone: 847-467-1285, e-mail: t-andersen@kellogg.northwestern.edu

<sup>b</sup> Department of Economics, Duke University, Durham, NC 27708, and NBER,  
phone: 919-660-1846, e-mail: boller@econ.duke.edu

<sup>c</sup> Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208,  
phone: 224-392-0782, e-mail: d-dobrev@kellogg.northwestern.edu

## 1. Introduction

Modeling financial market volatility has been a thriving research area over the last couple of decades. The topic speaks to fundamental risk and asset pricing issues with important applications in areas relating to portfolio allocation, risk management and measurement of systematic macroeconomic risk exposures. At the same time, it also provides a unique set of challenges to time series modeling through the vast amount of available high-frequency intraday data, the pronounced longer-run interday temporal persistence in higher order return moments, coupled with the existence of relatively frequent apparent extreme outliers. The importance of the field was recognized by the recent award of the 2003 Nobel Prize in Economics to Robert F. Engle for his seminal work on autoregressive conditionally heteroskedastic (ARCH), Engle (1982).

ARCH or closely related stochastic volatility model specifications applied at a daily data frequency remains the most common approach to practical volatility modeling. This is true even if we now have a decades worth of high-frequency intraday data available for a broad cross-section of actively traded financial assets. This reflects the very limited progress that has been made in utilizing the intraday data directly for volatility modeling and forecasting over longer daily, weekly and monthly horizons. Of course, a variety of market microstructure and announcement studies use high-frequency data to great effect, but it does not alter the fact that the information in these dense data sets have not been harnessed successfully in lower frequency return volatility studies. Meanwhile, some promising alternatives that rely on summary statistics extracted from the intraday data have been entertained, for example models using daily ranges, e.g., Garman and Klass (1980), Parkinson (1980), Gallant, Hsu and Tauchen (1997) and Alizadeh, Brandt and Diebold (2002) and - very recently - the so-called *realized volatility* measures, e.g., Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold and Labys (henceforth ABDL) (2001a,b, 2003), Barndorff-Nielsen and Shephard (henceforth BN-S) (2002a,b), and Meddahi (2002). Substantial advances have been documented from the adoption of such approaches, and the development of improved techniques for construction of daily volatility measures from ultra-high frequency data is currently a very active research area, e.g., Aït-Sahalia, Mykland and Zhang (2005), Bandi and Russell (2004a,b), Barndorff-Nielsen, Hansen, Lunde and Shephard (2005), Hansen and Lunde (2006), Oomen (2004), Zhang, Aït-Sahalia and Mykland (2005), among many others. Nonetheless, the relationship between these

methods and the standard daily ARCH type modeling paradigm is not yet fully understood, neither theoretically nor empirically.

This article seeks to shed further light on the characteristics of high-frequency asset return and volatility processes and their implications for daily return distributions. We shall not survey the literature on ARCH and stochastic volatility or high-frequency data based volatility modeling as a number of other sources already cover that ground; e.g., Bollerslev, Engle and Nelson (1994), Ghysels, Harvey and Renault (1996), Andersen, Bollerslev and Diebold (henceforth ABD) (2004), Andersen, Bollerslev, Christoffersen and Diebold (2005), Engle and Russell (2004), and Shephard (2005). Instead we focus on extending recent developments within the realized volatility literature in a direction that allow us to elaborate on general features of high-frequency asset return processes and link their properties more directly to the stylized facts from the burgeoning empirical volatility literature. Specifically, we demonstrate that the standard jump-diffusion models associated with arbitrage-free modeling in financial economics offer a flexible setting for exploring and rationalizing the properties of daily asset return data. In fact, we relate the real-time evolution of quantities studied in the realized volatility literature with characteristics of the daily return distribution. As such, we provide new empirical evidence on the nature of the intraday return generating process. This should set the stage for further improvements in the construction of high-frequency based volatility measures and their use in practical forecasting and real-time financial decision-making.

Our contribution is best appreciated in the context of the aforementioned, widely documented finding that the conditional distribution of the daily return innovations in standard volatility models invariably is heavy tailed and often possess extreme outliers. It is furthermore known that ARCH models - as the underlying data is sampled at ever finer frequencies and subject to standard regularity conditions - provide consistent volatility filters for extracting the correct conditional variance process from return series driven by a continuous-time diffusion; see Nelson (1990, 1992) and Drost and Werker (1996). Within this setting, the returns are locally conditionally Gaussian, so one may intuitively reason that the daily returns, appropriately standardized by the (realized) volatility over the course of the trading day, should be Gaussian as well. This is indeed correct for some interesting and popular special cases. Moreover, empirically, it has been found that this procedure produces normalized returns that are, to a close

approximation, Gaussian although formal tests still typically reject normality fairly convincingly; see ABDL (2000, 2001), and Andersen, Bollerslev Diebold and Ebens (2001). This points to the potential usefulness of the above result, but it also seems to indicate that there are features in the actual data which invalidate the above intuition.

One such critical feature is the presence of an asymmetric relation between the high-frequency return and volatility innovations, as implied by the so-called leverage or volatility feedback effects; see, e.g., the recent discussion in Bollerslev, Litvinova and Tauchen (2005). If such an asymmetric relation is at work, the result should fail in theory, even if the underlying process is a continuous semi-martingale. However, through the “time-change theorem for continuous local martingales” we may formally restore the Gaussianity of appropriately standardized trading day returns by sampling the underlying asset prices in “event time” or “financial time” as measured by equal sized increments to the volatility process rather than in calendar time as given by equidistant time intervals.<sup>1</sup>

The above scenarios largely exhaust the relevant possibilities when the underlying asset return process evolves as a *continuous* semi-martingale. Meanwhile, there is an increasing body of empirical work which concludes that continuous-time models must incorporate *jumps* or discontinuities in order to provide an satisfactory characterization of the daily return process; see, among others, Aït-Sahalia (2002), Andersen, Benzoni and Lund (2002), Bates (2000), Chan and Maheu (2002), Chernov, Gallant, Ghysels and Tauchen (2003), Drost, Nijman and Werker (1998), Eraker (2004), Eraker, Johannes and Polson (2003), Johannes (2004), Maheu and McCurdy (2004), and Pan (2002).<sup>2</sup> Although the jump-diffusion setting is fully compatible with the standard no-arbitrage framework of financial asset pricing theory, as detailed in e.g., Back (1991), the presence of jumps take us outside the domain of the statistical framework and the corresponding theorems discussed above. However, recent advances in the realized volatility literature include nonparametric data-driven procedures explicitly designed to identify jumps

---

<sup>1</sup> This procedure has recently been implemented as a part of a test for whether the underlying return process can be seen as a continuous semi-martingale by Peters and de Vilder (2004); see also the related studies by Zhou (1998) and Ané and Geman (2000).

<sup>2</sup> Earlier influential studies based on time-invariant diffusions allowing for jumps include Merton (1976) and Ball and Torous (1980).

from underlying high-frequency return series; see BN-S (2004, 2005), ABD (2005), and Huang and Tauchen (2005). This allows for the possibility that, initially, we can test for, and subsequently eliminate, the impact of discontinuities in the price path. Following such an preliminary jump detection and extraction step, we may then apply the above reasoning to explore if the appropriately normalized trading day returns, cleaned for jumps, are Gaussian. In combination, our approach constitutes a novel sequential procedure for exploring, and informally testing for, whether a jump-diffusion offers a reasonable characterization of the underlying return generating process in continuous time. It also raises the question of how the standardized returns in event time will behave if the underlying price path exhibit jumps. That is, how do jumps manifest themselves in the conditional return distribution if they are ignored? Moreover, do we have the power to detect their existence through the realized volatility based jump detection techniques?

More generally, information regarding the strength of the jump intensities and sizes, the significance and magnitude of potential leverage effects, along with direct estimates of the time series of diffusive volatility, are of immediate import for a whole array of key financial economics questions, including analysis of the causes behind extreme return realizations, the general risk-return tradeoff and the associated pricing of financial assets, the portfolio allocation problem, the construction of improved risk management techniques, and derivatives pricing. The possibility that we may gain insights into these issues through direct statistical analysis of the intraday return series under minimal auxiliary assumptions is intriguing. We address these issues both through an empirical illustration based on a 2-minute intraday S&P 500 futures return series covering a relatively long sample period from 1988 to 2004, and through an extensive simulation study.

The paper progresses as follows. Section 2 provides additional motivation and more formally outlines the relevant theoretical framework. Section 3 explores the finite sample behavior of our new sequential test procedure for satisfactorily assessing the adequacy of a no-arbitrage jump-diffusion model. We explore the behavior of the tests for empirically calibrated jump-diffusion series under a variety of different scenarios including pure diffusions and jump-diffusions both with and without a leverage effects. Section 4 presents our empirical analysis of the high-frequency S&P500 returns. We find strong suggestive evidence for the presence of both

jumps and leverage effects. Following our sequential procedure for generating appropriately standardized event time returns, excluding the identified jumps, we are unable to reject the null hypothesis that the resulting time series is i.i.d. Gaussian. As such, our findings are consistent with the premise that the underlying returns follow an arbitrage-free continuous-time jump-diffusive process.<sup>3</sup> Section 5 concludes.

## **2. Theoretical Background**

The continuous trade and quote activity taking place on financial markets with instantaneous information transmission renders a continuous time specification for the underlying price process natural. Moreover, the sudden release of news or the arrival of large buy or sell orders will often induce a distinct large change, or jump, in the asset price. Hence, a standard approach within the financial economics literature is to let the logarithmic asset price process evolve continuously according to a generic jump-diffusion process. Even if the underlying prices cannot be observed at every instant, the recorded quote and transaction prices may be seen as, possibly noisy, observations from this continuously evolving process. This formulation has a strong theoretical underpinning as the price process under standard regularity conditions will constitute a special semi-martingale and hence not allow for arbitrage opportunities, see, e.g., Back (1991). Moreover, it allows for trades and quotes to occur at any time, mimicking the continuous operation of a financial market during trading hours, and it enables us, at least in principle, to derive the distribution of discretely observed returns at any frequency through appropriate aggregation, or integration, of the increments to the underlying continuous price process. Importantly, it is also an extremely flexible setting that has the potential to accommodate all major characteristics of daily financial return series, including pronounced volatility persistence, asymmetric return distributions, intraday patterns and jumps or discontinuities.

Our foremost interest here is in gaining insight into the descriptive validity of the semi-martingale representation for asset prices as embodied within the general jump-diffusion setting outlined above. The main limitation is that we exclude Lévy jump style processes with an infinite

---

<sup>3</sup> At the same time, we scrutinize the related conclusion of Peters and de Vilder (2004), that the S&P500 returns may be characterized adequately through a pure continuous diffusion process.

jump intensity, as in e.g., Carr, Geman, Madan and Yor (2002), as our setting only allows for “rare” jumps occurring at a finite expected rate per unit time interval. Hence, a key issue is to what extent the jump-diffusion representation is consistent with empirical data and what features of the specification are necessary in order to adequately describe observed return processes. If the overall strategy is successful, the strength of the various features of the return process may in turn be assessed directly during the distinct phases of the diagnostic procedure that we develop below.

## 2.1. Quadratic Variation, Realized Volatility, and Trading Day Return Distributions

For simplicity, we focus on the univariate case. Let  $p(t)$  denote the time  $t$  logarithmic asset price. The generic jump-diffusion process may then be expressed in stochastic differential equation (sde) form,

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t), \quad 0 \leq t \leq T, \quad (1)$$

where  $\mu(t)$  is a continuous and locally bounded variation process, the stochastic volatility process  $\sigma(t)$  is strictly positive and càglàd,<sup>4</sup>  $W(t)$  denotes a standard Brownian motion,  $dq(t)$  is a counting process with  $dq(t)=1$  corresponding to a jump at time  $t$  and  $dq(t)=0$  otherwise with (possibly time-varying) jump intensity  $\lambda(t)$ , and  $\kappa(t)$  refers to the size of the corresponding jumps. The quadratic variation for the cumulative return process,  $r(t) \equiv p(t) - p(0)$ , is given by

$$[r, r]_t = \int_0^t \sigma^2(s) ds + \sum_{0 < s \leq t} \kappa^2(s). \quad (2)$$

Of course, in the absence of jumps, the second term on the right-hand-side disappears, and the quadratic variation simply equals the integrated volatility.

Let the discretely sampled  $\Delta$ -period returns be denoted by,  $r_{t,\Delta} \equiv p(t) - p(t-\Delta)$ . For ease of

---

<sup>4</sup> Note that this assumption allows for discrete jumps in the stochastic volatility process. Recent related work on Lévy-driven stochastic volatility models include BN-S (2001), Carr, Geman, Madan and Yor (2003), and Todorov and Tauchen (2005).

notation we normalize the daily trading day time interval to unity and label the corresponding discretely sampled trading day returns by a single time subscript,  $r_{t+1} \equiv r_{t+1,1}$ . Also, we define the daily realized volatility by the summation of the corresponding  $1/\Delta$  high-frequency intraday squared returns,

$$RV_{t+1}(\Delta) \equiv \sum_{j=1}^{1/\Delta} r_{t+j\Delta,\Delta}^2, \quad (3)$$

where without loss of generality  $1/\Delta$  is assumed to be an integer. Then, as emphasized in the series of recent papers by Andersen and Bollerslev (1998), ABDL (2001, 2003), BN-S (2002a,b) and Comte and Renault (1998), among others, by the theory of quadratic variation this realized volatility converges uniformly in probability to the increment to the quadratic variation process defined above as the sampling frequency of the underlying returns increases. Specifically, under weak regularity conditions, and for  $\Delta \rightarrow 0$ ,

$$RV_{t+1}(\Delta) \rightarrow [r,r]_{t+1} - [r,r]_t = \int_t^{t+1} \sigma^2(s) ds + \sum_{t \leq s \leq t+1} \kappa^2(s). \quad (4)$$

In the absence of jumps, and even in the presence of a leverage type effect, the realized volatility is therefore consistent for the integrated variance that figures prominently in the stochastic volatility option pricing literature and, importantly, if jumps are present then the realized volatility is consistent for the sum of the integrated variance and the cumulative sum of squared jumps. Hence, the realized volatility approximates (for  $\Delta > 0$ ) the total (ex-post) return variability, whether the source is the diffusive or the jump component of the return process.

### 2.1.1. No Leverage or Jumps in the Return Generating Process

The quadratic variation represents the cumulative variability of the continuously evolving return process. As such, it is the natural basic concept of the realized return variation over the interval  $[0,t]$ , as emphasized by ABD (2004) and ABDL (2003). This is particularly transparent in the case of a pure diffusive return process with no leverage style effect, where by assumption the drift and volatility processes,  $\mu(t)$  and  $\sigma(t)$ , are independent of the return innovation process  $W(t)$ ,

whereby it follows that,

$$r(t) | \sigma \{ [\mu(\tau), \sigma(\tau)]_{0 \leq \tau \leq t} \} \sim N \left( \int_0^t \mu(s) ds, \int_0^t \sigma^2(s) ds \right), \quad (5)$$

where  $\sigma \{ \mu(\tau), \sigma(\tau) \}_{0 \leq \tau \leq t}$  denotes the  $\sigma$ -field generated by the sample paths of  $\mu(\tau)$  and  $\sigma(\tau)$  for  $0 \leq \tau \leq t$ . The integrated variance thus provides a natural measure of the true latent  $t$ -period return variability. Notice furthermore that the expected mean component in equation (5) typically is negligible over shorter time horizons such as a trading day or trading week.

It is important to keep in mind that the integrated variance term in equation (5) represents the *ex-post* or realized return variability. *Ex-ante*, letting the relevant information set at time  $s$  be denoted by  $\mathcal{F}(s)$ , the corresponding concept of return variability is given by the conditionally expected future return variability over the forecast horizon,

$$V(t) = E \left[ \int_0^t \sigma^2(s) ds \mid \mathcal{F}(0) \right]. \quad (6)$$

Since the volatility process generally is genuinely stochastic (see the discussion in Andersen, 1992), the realized or integrated variance will equal the expected variance,  $V(t)$ , *plus* an innovation term. Consequently, even when the correct model is used to predict future return variability, in accordance with equation (6), the standardized returns will be fat-tailed relative to a Gaussian benchmark. For simplicity assuming that the mean is equal to zero, or  $\mu(s) \equiv 0$ ,

$$r(t) \cdot V^{-1/2}(t) \sim N(0, 1).$$

On the other hand,

$$r(t) \cdot \left[ \int_0^t \sigma^2(s) ds \right]^{-1/2} \sim i.i.d. N(0, 1), \quad (7)$$

so that returns normalized appropriately by the realized return variability are truly Gaussian. In the presence of a non-zero expected return component, the returns should still - to a very good approximation - be Gaussian over short time intervals in this no-leverage pure diffusion case.

This result provides a possible rationalization for why financial returns normalized by volatility forecasts from standard ARCH and stochastic volatility models almost invariably exhibit fat-tails relative to the normal distribution; see, e.g., Bollerslev (1987), Nelson (1991), Chib, Nardari and Shephard (2002), and Forsberg and Bollerslev (2002). However, the result in (7) may also appear abstract and impractical as the requisite volatility scaling is obviously random and not measurable with respect to the time 0 information set. Moreover, it relies on standardization with the *true* integrated volatility, which is latent and hence by its very nature not directly observable. Nonetheless, the result does provide inspiration for the development of precise ex-post measurements of the realized return variability. This is exactly what the realized volatility measures seek to accomplish. The basic insight is that by appealing to the general consistency result in equation (4), high quality intraday price and quote data allow for a vastly improved assessment of the actual trading period return variability. However, a number of practical complications arise in actually implementing these ideas.

In principle, we should use all available price and quote observations so as to mimic the limiting operation,  $\Delta \rightarrow 0$ , as best possible. However, the assumption that the transaction or quote prices follow a semi-martingale is blatantly violated in practice at the very finest sampling frequencies where the discrete price grid and the bouncing between bid and ask prices implies that recorded price changes are either zero or “large” relative to the expected return variability over very small time intervals. The average time between ticks for liquid securities often amounts to just a few seconds. The average return volatility over such short intervals is very small and typically an order of magnitude less than the lowest feasible price change as dictated by the available or commonly used price grid. Therefore, we would only expect the semi-martingale property to provide a decent approximation over somewhat longer intraday return horizons such as one- or five minutes. Moreover, it is clear that the “optimal frequency” also will depend upon the liquidity, price grid and specific market structure. As already noted above, these issues are the subject of quite intense scrutiny within a rapidly expanding literature, see, e.g., the studies by Aït-Sahalia, Mykland and Zhang (2005), ABDL (2000, 2003), Bandi and Russell (2004a,b),

Bollen and Inder (2002), Corsi, Zumbach, Müller and Dacorogna (2001), Hansen and Lunde (2006), Oomen (2004), Zhang, Aït-Sahalia and Mykland (2005), and Zhou (1996) among others. We shall not pursue any of the more refined procedures recently proposed in this literature in the present paper. Instead, we simply rely on a “sensible” choice of intraday sampling frequency for providing a robust and acceptable compromise between obtaining additional information through more frequent sampling on the one hand and avoiding excessive noise through the accumulation of microstructure distortions in the observed price process (relative to the frictionless diffusive ideal) at the very highest frequencies on the other.

Taken together, the results discussed above inspire a practical high-frequency data based strategy for a nonparametric test of the hypothesis that the given return process may be treated as arising from a pure diffusion without leverage effects which should be valid under minimal auxiliary assumptions. The idea is to construct the realized volatility measures along the lines indicated in equation (3) and simply substitute the resulting estimate of the integrated variance into equation (7) and then test whether the resulting standardized trading period return series is statistically distinguishable from a sequence of i.i.d. draws from a  $N(0,1)$  distribution. Of course, this does involve a joint hypothesis as any rejection also could arise from the fact that the integrated variance is estimated with error due to the presence of the market microstructure “noise” in the high frequency return observations as well as the use of a discrete intraday sampling frequency.<sup>5</sup> Of course, if the underlying results are to be used for practical purposes they must be shown to provide a reasonable guide to the distributional properties of actual return series. In fact, ABDL (2000, 2001) and Andersen, Bollerslev, Diebold and Ebens (2001) do find that the realized volatility normalized returns are much closer to the ideal of an i.i.d.  $N(0,1)$  series than is the case for daily returns normalized by the corresponding daily return based volatility forecast.<sup>6</sup> Nonetheless, it is generally found that such realized volatility standardized series differ significantly from the Gaussian ideal as formal normality tests almost always reject the null hypothesis of i.i.d.  $N(0,1)$  quite overwhelmingly. We shall shed some additional light on these

---

<sup>5</sup> The asymptotic (for  $\Delta \rightarrow 0$ ) theory in BN-S (2002a) and Andersen, Bollerslev and Meddahi (2005) provides a framework for assessing the latter effect.

<sup>6</sup> This was further exploited by ABDL (2003) in designing reduced-form time series modeling and forecasting procedures for realized volatilities of daily foreign exchange series.

issues in the empirical and simulation based investigations below.

### 2.1.2. The Impact of Leverage

The preceding section explored some of the distributional implications of one of the more commonly assumed return generating processes in financial economics. However, as previously noted, there is now compelling evidence that the return process for many important asset markets, including those for equity indices, display a pronounced asymmetric relationship between return and volatility innovations. This is known under the acronym of a “leverage effect” although the origin of the asymmetry in the return dynamics arguably has very little, if anything, to do with the underlying financial leverage of the traded assets; see Black (1976) and Christie (1992), and the subsequent reasoning in Campbell and Hentschel (1992), Bekaert and Wu (2000), and Bollerslev, Litvinova and Tauchen (2005), among others. In this case, the results from Section 2.1.1 are generally not valid. Of course, one obvious question is whether this actually makes a practical difference in terms of the distribution of the standardized returns. A second question is whether there is any way to restore some general distributional results for this case. We explore these issues in this section.

We retain the pure diffusion assumption, so equation (1) remains valid and the term representing the jump component is identically zero; i.e.,  $\kappa(t) dq(t) \equiv 0$ . However, in contrast to the results discussed in the previous subsection, we do not require the stochastic volatility process,  $\sigma(t)$ , to be independent of the return innovation process,  $W(t)$ . In particular, the existence of a leverage type effect generally induces a negative correlation between the innovations to the return and volatility processes. As such, knowledge of the daily integrated variance (or the associated realized volatility measure) will be informative regarding the sign of the daily return innovation. Hence, equations (5) and (7) are no longer true. Nonetheless, the Dambis-Dubins-Schwartz theorem (see Dambis, 1965, and Dubins and Schwartz, 1965) ensures that an appropriately time-changed continuous martingale will become a Brownian Motion.<sup>7</sup> The

---

<sup>7</sup> Formally, as noted in a similar context by Peters and de Vilder (2004), any continuous local martingale (started at the origin), say  $Y$ , can be decomposed as  $Y = B \circ Q$  where  $B$  denotes a standard Brownian motion and  $Q$  represents the quadratic variation of  $Y$ , see, e.g., Karatzas and Shrieve (1991), Theorem 4.6. This same idea has also been explored empirically in a more informal setting by Zhou (1998).

implication is that appropriately sampled return series will be Gaussian even in the leverage case.

In particular, for simplicity assuming again that  $\mu(t) \equiv 0$ , it then follows from this general result that the time series of returns defined by equation (1) with no jumps will be i.i.d. Gaussian if sampled in equidistant increments as dictated by the corresponding quadratic variation process. Specifically, for a fixed positive period of “financial” or “event” time  $\tau^*$ , we seek to sample the logarithmic price process in calendar time points,  $0 = t_0, t_1, t_2, \dots, t_k, \dots$ , where the calendar time sampling points are defined by

$$t_k = \inf_{t>0} ( [r,r]_t - [r,r]_{t_{k-1}} > \tau^* ), \quad k = 0, 1, \dots ,$$

so that returns are computed over intervals of identical quadratic variation,  $\tau^*$ . Note that while all of these return horizons span the identical amount of underlying return variability, they will, of course, reflect potentially highly variable calendar time intervals.<sup>8</sup>

In order to facilitate comparisons with some of our other distributional results, a natural choice is to calibrate the event time step,  $\tau^*$ , such that the average calendar period associated with the event-time sampled returns equals one trading day. Denoting the corresponding sequence of returns sampled in financial time by

$$R_k \equiv p(t_k) - p(t_{k-1}), \quad k = 0, 1, 2, \dots ,$$

the following distributional result thus remains valid, even in the case of leverage,

$$R_k / \sqrt{\tau^*} \sim \text{i.i.d. } N(0,1), \quad k = 0, 1, 2, \dots . \quad (8)$$

---

<sup>8</sup> The notion of “financial” or “event” time is related to the so-called Mixture-of-Distributions Hypothesis (MDH) originally proposed by Clark (1973), and further developed by Epps and Epps (1986), Tauchen and Pitts (1983), Andersen (1996), and Andersen and Bollerslev (1997) among others. The main gist of the MDH, namely that the trading process (along with the return volatility process) is driven by an underlying latent activity process, is notably absent from equation (8). Our strategy of using the high-frequency data for the construction of an observable proxy for the “financial event time” also deviates from the empirical approaches in the MDH literature. Our approach is furthermore similar in spirit to the concept of theta-time advocated by Olsen and Associates (see, e.g., Dacorogna et al, 2001), which also relies on high-frequency data for the construction of a deformed time-scale.

This result is considerably more general than the previous distributional result in equation (7), and importantly applies for *any* continuous martingale. Moreover, the result should provide a very good approximation for shorter return horizons, even if the expected return is non-zero as formally assumed in (8). As such, this provides a novel way of gauging the importance or strength of the leverage effect by comparing the distributional properties of the return series standardized by realized volatility versus the returns obtained from sampling in financial time. This is a fully nonparametric approach, independent of any specific modeling choices for the leverage effect and/or the diffusive volatility component.<sup>9</sup> Of course, in order to render it practical one must approximate the fixed increments of the latent quadratic variation process by an observable estimator thereof. A natural candidate is the realized volatility as defined in equation (3). Once an appropriate choice of  $\tau^*$  has been made, high-frequency returns can be used to split the sample in equal-sized financial time steps of length  $\tau^*$ . In practice this will, of course, induce some measurement error into the procedure, as the realized volatility only provides a noisy measure of the true underlying quadratic variation. We explore the implications of these practical complications within our simulation setting later on.

### **2.1.3. The Impact of Jumps**

The preceding sections report results under the maintained assumption that the price process is generated by a continuous sample path diffusion. Hence, there are no discontinuities in the price path. However, as previously noted, several recent studies involving the direct estimation of continuous time stochastic volatility models along the lines of equation (1) have highlighted the importance of explicitly incorporating jumps in the price process; see Andersen, Benzoni and Lund (2002), Eraker, Johannes and Polson (2003), Eraker (2004), Johannes, Kumar and Polson (1999), Maheu and McCurdy (2004), among others. More generally, ruling out jumps a priori is also theoretically unsatisfactory as the existence of discontinuities in the price path is entirely consistent with the foundation for continuous time finance as derived from basic no-arbitrage

---

<sup>9</sup> In this regard it is noteworthy that even though the MDH may provide a satisfactory empirical description of the joint return volatility-trading volume relationship at the daily frequency, the hypothesis typically fails when it is explored at higher intraday frequencies. Since equation (8) is based on a different set of assumptions, requiring largely that the return process is arbitrage-free along with the assumption of no jumps, it is entirely distinct from the MDH, and its empirical performance will inform us about different features of the return generating process.

principles. In fact, in an efficient market setting the release of significant “news” *should* induce an immediate jump in the price.<sup>10</sup>

Once we allow for a jump component in the general return specification (1), the distributional results for the pure diffusion case discussed above break down. The question is, can we still derive testable distributional implications based on the assumption of an arbitrage-free price process? At first glance, this might appear impossible without additional (auxiliary) restrictions, as the jump process can be endowed with an arbitrary finite intensity rate for jumps and the associated jump distribution may be of almost any type. Hence, the logic from the pure diffusive case based on the local Gaussian behavior of the return process cannot be restored, but will apply only to the diffusive part of the price process. One potential solution is to directly identify the jumps in the price path, thus decomposing the return process into a jump and diffusive part, and then investigate the distributional properties of each component separately. This approach turns out to be feasible given the recent powerful asymptotic results (for  $\Delta \rightarrow 0$ ) in BN-S (2004, 2005) that allow for separate (non-parametric) identification of the two components of the quadratic variation process.

Specifically, on defining the standardized realized bi-power variation measure,

$$BV_{t+1}(\Delta) \equiv \mu_1^{-2} \sum_{j=2}^{1/\Delta} |r_{t+j\Delta, \Delta}| |r_{t+(j-1)\Delta, \Delta}|, \quad (9)$$

where  $\mu_1 \equiv \sqrt{(2/\pi)}$ , it follows that for  $\Delta \rightarrow 0$ ,

$$BV_{t+1}(\Delta) \rightarrow \int_t^{t+1} \sigma^2(s) ds. \quad (10)$$

Consequently, the bipower variation (asymptotically) annihilates the contribution of the jumps to the quadratic variation and only measures the integrated volatility attributable to the diffusive volatility component. Hence, as noted by BN-S (2004, 2005), combining the results in equations

---

<sup>10</sup> This is also consistent with a recent and rapidly expanding literature documenting almost instantaneous price reactions in response to the release of a number of perfectly timed macroeconomic news announcements; see, e.g., Andersen, Bollerslev, Diebold and Vega (2003, 2005) and the many references therein.

(4) and (10), the contribution to the quadratic variation process due to the discontinuities (jumps) in the underlying price process may be consistently estimated by

$$RV_{t+1}(\Delta) - BV_{t+1}(\Delta) \rightarrow \sum_{t \leq s \leq t+1} \kappa^2(s). \quad (11)$$

Of course, in the absence of jumps both measures provide consistent estimates of the integrated variance so, for any given finite number of intraday return observations, the expression in (11) may well turn out to be negative due to regular small-sample ( $\Delta > 0$ ) variation. Hence, it is sensible, at a minimum, to impose a non-negativity truncation on the empirical squared jump measurements,

$$J_{t+1}(\Delta) \equiv \max[RV_{t+1}(\Delta) - BV_{t+1}(\Delta), 0]. \quad (12)$$

In addition, BN-S (2004) provide an asymptotic theory (for  $\Delta \rightarrow 0$ ) for the joint asymptotic distributions of the realized volatility and bipower variation measures under the null hypothesis of a continuous sample path, in turn allowing for the construction of formal statistical tests for significant jumps based on the appropriately scaled difference between the two measures. Hence, again recognizing that small squared jump measures implied by the statistic in (12) may well be due to finite sample variation, it has alternatively been suggested only to designate those days for which the corresponding jump statistic appears highly significant under the null hypothesis of a diffusion process as actual jump days. These general insights and considerations have inspired the construction of a variety of practical jump detection techniques by ABD (2005), BN-S (2005) and Huang and Tauchen (2005).<sup>11</sup>

Still, when analyzing the distributional features of the price process, once a day has been designated as containing a jump, an additional step is required if we want to identify the exact location and size of the jump - or even multiple jumps - during the designated day. This remains a research area in its infancy and little is known about the best practical approach for actually

---

<sup>11</sup> An alternative but related nonparametric continuous record asymptotic jump detection scheme based on the Paul Lévy Law for the modulo of continuity for the sample path of a Brownian Motion have recently been developed in a series of papers by Mancini (2004, 2005a,b).

identifying the exact jump times and sizes. We provide a more detailed discussion of the specific approach we adopt in the empirical section below; see also Andersen, Bollerslev, Frederiksen and Nielsen (2005). For now, we simply take as given the ability to obtain nonparametric (albeit noisy) empirical measures of the jumps over the full trading day sample.

This ability to perform statistical inference regarding the timing and size of the jumps allow us to devise a fully non-parametric strategy for deriving useful distributional implications for appropriately adjusted and standardized return series within the general jump-diffusion setting. First, we subject the intraday return series to a jump identification scheme and remove the identified jumps from the trading day return series. These series are then seen, approximately, as generated from a pure diffusion process, so that we can apply the techniques suitable for that case as discussed in the preceding sections. Overall, this provides a non-parametric strategy for gauging the validity of the jump diffusion framework for a given financial return series. However, before we assess the empirical merit of this approach, we first discuss some pertinent implementation issues.

## **2.2. Testing for Distributional Features of Jump-Adjusted, Standardized Returns**

Under ideal circumstances, including frictionless markets and perfect jump detection and extraction techniques, the appropriately adjusted and realized volatility standardized trading day returns should asymptotically, for ever finer sampling frequencies, be identically and independently distributed as standard normal random variables, as indicated in equations (7) and (8) for the pure diffusion case without and with a leverage effect, respectively. This is indeed the property that will serve as a benchmark for our empirical investigation concerning the descriptive validity of the jump-diffusion setting, based upon the actual return distributions calculated from the limited number of intraday trading day returns at our disposal.

### **2.2.1. Some General Properties of Standardized Trading Day Returns**

An important first observation is that trading day returns standardized by realized volatility will tend to be thin tailed by construction. The basic argument is straightforward. Assume that there are  $n \equiv 1/\Delta$  continuously compounded intraday return observations available for a specific trading day and there is at least one recorded price change over the course of the trading day so

that the realized volatility is strictly positive.<sup>12</sup> Next, let the sum of these intraday returns be denoted by  $c$ . Obviously, if  $c = 0$  the standardized return is also zero. Thus, the absolute standardized trading day returns can be arbitrarily small. In contrast, they are bounded from above. To see this, note that if the trading day return equals  $c$ , the vector of intraday returns must belong to the set  $\mathbf{X} = \{(x_1, x_2, \dots, x_n) : \sum_{j=1}^n x_j = c\}$ . The maximum attainable absolute

standardized return for trading day  $t+1$ , say, is then given by solving the following simple optimization problem,

$$\text{Max} \sum_{j=1}^n r_{t+j,\Delta,\Delta} / \sqrt{RV_{t+1}(\Delta)} = c \cdot \left( \sum_{j=1}^n r_{t+j,\Delta,\Delta}^2 \right)^{-1/2},$$

subject to the  $n \times 1$  vector of intraday returns  $(r_{t+\Delta,\Delta}, r_{t+2\Delta,\Delta}, \dots, r_{t+(n-1)\Delta,\Delta}, r_{t+1,\Delta})$  belonging to the set  $\mathbf{X}$ . The solution and the associated maximum standardized (absolute) return are readily determined as,

$$r_{t+j,\Delta,\Delta} = c/n, \quad j = 1, 2, \dots, n,$$

and

$$c \cdot \left( \sum_{j=1}^n r_{t+j,\Delta,\Delta}^2 \right)^{-1/2} = \sqrt{n},$$

respectively. Intuitively, the absolute standardized return is maximized when the realized volatility is minimized, and this occurs when the intraday return process is as smooth as possible subject to the total daily return constraint. It is noteworthy that the maximal value is independent of  $c$ , so that the upper bound on the standardized return is strictly a function of the number of intraday returns employed in the construction of the realized volatility measure. Consequently, the distribution of the standardized returns will have finite support, or truncated tails, as it is impossible to observe any realizations outside the  $[-\sqrt{n}, \sqrt{n}]$  interval. This result holds for *all*

---

<sup>12</sup> Otherwise normalization with the realized volatility is not meaningful, although we may proceed by defining the standardized return to be zero in this degenerate case.

return generating processes and in particular remains valid in the presence of jumps. In fact, it follows from the above reasoning that, for a given trading day return  $c$ , the absolute standardized return will be low if the corresponding intraday return series is relatively volatile, or “choppy.” Extending this logic, if there is a jump present, the absolute standardized return will tend to be relatively low. This same line of reasoning also suggests that the removal of jumps from the intraday return series will render it less thin tailed.

The above conjecture may be studied more formally. Assume that we have a given set of intraday returns denoted  $(x_1, x_2, \dots, x_n)$ , and that, without loss of generality suppose that,

$\sum_{j=1}^n x_j = c > 0$ , so that the trading day return is positive. Now, imagine that one of these intraday

returns, say  $x_i$ , actually represent a jump and that this price jump was positive - as it typically will be in this situation if it is a big move with a marked impact on the overall return for the day. What happens to the standardized trading day return as we increase the jump size marginally?

According to the above conjecture it should shrink as the additional (marginal) “choppiness” tend to increase the realized volatility, and hence lower the standardized return. However, we are no longer holding the overall trading day return fixed at  $c$ , but instead allowing it to increase in step with the jump size,  $x_i$ , invalidating this simple reasoning. Nonetheless, taking the partial derivative of the standardized return,

$$\sum_{j=1}^n x_j \cdot \left( \sum_{j=1}^n x_j^2 \right)^{-1/2}$$

with respect to  $x_i$ , it follows readily that the overall impact is negative if and only if,

$$x_i > \sum_{j=1}^n x_j^2 / c . \tag{13}$$

This condition is trivially satisfied if  $x_i$  represents a jump as assumed above, since the largest intraday return must exceed the right-hand-side limit in (13) unless all of the intraday returns are identically equal to  $c/n$ , in which case the relationship in (13) holds as an equality. Likewise, if we increase a positive intraday return that is less than the quantity in (13), then the standardized trading day return will increase as the effect is proportionally larger for the daily return than the

realized volatility.

These simple arithmetic arguments lend direct support to the conjecture that jumps tend to render the standardized returns thin tailed. Of course, it is possible to construct counterexamples where, e.g., the jump happens to be of the opposite sign of the overall trading day return, so that the impact will go in the opposite direction. Similarly, if one jump helps offset another large jump on the same day, the effect is generally unpredictable and may well go in the opposite direction. Thus, even though we would anticipate that the identification and elimination of jumps from the intraday return series will render the standardized returns (net of jumps) less thin tailed, this is not a general theoretical result. We explore this further within our simulation and empirical sections.

### 2.2.2. Finite Sample Results for the Diffusion Case

For simplicity, we now consider the pure diffusive setting - or equivalently assume that all jumps have been perfectly identified and removed from the return series - and that as before  $n$  equidistant intraday return observations are available from the underlying logarithmic price process.

Moreover, assume that the diffusive volatility component is constant over the day. This is obviously not a valid characterization of the true intraday volatility process, but it may serve as a useful benchmark and for better understanding the finite sample behavior. In particular, it follows from Peters and de Vilder (2004) that in this situation, the density function for the standardized returns,

$$\tilde{R}_{t+1} \equiv \sum_{j=1}^n r_{t+j,\Delta,\Delta} / \sqrt{RV_{t+1}(\Delta)}$$

takes the explicit form,

$$f_{\tilde{R}}(\tilde{r}) = \frac{\Gamma(n/2)}{\sqrt{\pi n} \Gamma((n-1)/2)} \left(1 - \frac{\tilde{r}^2}{n}\right)^{\frac{(n-3)}{2}} 1_{(-\sqrt{n}, \sqrt{n})}(\tilde{r}), \quad \tilde{r} \in \mathbb{R}. \quad (14)$$

A number of observations are in order. First, the support of the distribution is obviously  $[-\sqrt{n}, \sqrt{n}]$  in accordance with our results for the general case in the preceding section. Hence,

the finite sample distribution is truncated and will have thin tails. Second, if instead of sampling equidistant high frequency returns in calendar time, the return period is defined through the increment to realized volatility, as described in Section 2.1.2, the number of underlying “intraday” observations, or  $n$ , may vary widely across “days,” ranging from only a few to as high as a thousand. We provide direct evidence on this in our empirical work later on. Third, from equations (7) and (8) it is clear that the density function in (14) will converge to the standard normal distribution for  $n \rightarrow \infty$ . This is indeed the case, as illustrate in Figure 1 for different values of  $n$ . It is evident that the normal approximation is exceptionally poor when based on a low number of high frequency observations. Only for  $n \geq 48$  does the approximation work reasonably well in the center of the distribution, and the tail behavior is only close to that of the normal for even higher values of  $n$ . Fourth, it is worth keeping in mind that the finite sample distribution in (14) does not constitute an exact representation of the true sampling distribution, but instead relies on the problematic assumption of constant volatility within each trading “day,” or period. When there are only a handful of intraday return observations over the (financial) trading period in question, some of these high frequency returns must, by construction, be rather extreme. In such instances, it is unlikely that the underlying assumption of i.i.d. normal returns within the trading period affords a satisfactory description, and we may expect the analytic distribution for the standardized returns in (14) to provide an especially poor approximation in these situations. Again, these are issues that we will explore in the simulation setting.

### **2.2.3. Finite Sample Biases and Statistical Tests**

One popular set of normality tests is based on comparison of higher order sample moments with the corresponding theoretical values under the null hypothesis of Gaussianity. In the current context, we have the sharp asymptotic i.i.d.  $N(0,1)$  null hypothesis for the appropriately standardized daily returns rather than a more generic  $N(\mu, \sigma^2)$  null. Hence, we may compare the third and fourth sample moments directly to the corresponding theoretical standard normal values without first demeaning and scaling the observed series as is done, for example, when applying the usual Jarque and Bera (1980) (henceforth JB) test. The focus on the exact null hypothesis is likely to bring about important improvements in both the size and power of our test procedures

compared to regular normality tests. However, it may potentially bring about an excessive amount of power as the finite sample moments are likely to have a downward bias. Specifically, as noted by Peters and de Vilder (2004), the benchmark finite sample distribution for the pure diffusive case in (14) implies that the second and fourth moment of the standardized trading day returns should equal unity (as for the standard normal) and  $3n/(n+2)$ , respectively. Hence, the thin tailed finite sample distribution invariably manifests itself in a kurtosis below the Gaussian value of 3. When the number of intraday observations is relatively low, this finite sample “bias” can be substantial. In practice, the intraday returns are unlikely to be Gaussian, so this issue may be even more pertinent than suggested by the above computations. In recognition of these issues, we will report test statistics that seek to alleviate such biases, along with the more standard JB and related Empirical Distribution Function (EDF) tests for i.i.d.  $N(0,1)$ .

### **2.3. Summary of Theoretical Implications and Testing Strategy**

According to Section 2.1 the (perfectly) jump adjusted and appropriately standardized trading day returns should to a good approximation constitute an i.i.d.  $N(0,1)$  series if the underlying return generating process is a semi-martingale given as a jump-diffusion. The obvious approach is to subject this null hypothesis to a battery of tests to check whether the result holds for actual financial data and this is indeed what we will do. However, as noted above a variety of issues can render direct tests problematic. These issues are tied to the general problem that the procedure is based on multiple nonparametric estimation steps that all must be performed before the final return series can be constructed. Taking the steps one at a time, we first require a jump detection and extraction scheme for the intraday return series. By construction, our approach only identifies rather extreme jumps so smaller jumps may remain in the jump-adjusted series. Perhaps less importantly, we will also erroneously eliminate, with a small probability, some extreme returns as jumps even if they arise from a pure diffusive process. So the initial jump adjustment step is invariably plagued by measurement errors that are hard to assess without imposing additional structure on the problem - something we do not want to do as we seek to retain the model-free nonparametric spirit of the procedure. The hope is, of course, that truly significant jumps will be correctly identified, enabling us to eliminate the major distortions induced by the presence of jumps. Second, we construct realized volatility measures from a finite set of noisy intraday

returns. These will not be perfect due to the standard finite sample variation stemming from the use of a finite number of observations, as well as the presence of market microstructure frictions that induce an additional layer of noise into the measurements. While we can attempt to control for the finite sample bias, it is harder to assess the impact of microstructure frictions. We tailor our empirical application accordingly by choosing a return series we expect to be only minimally impacted by the latter issues. Third, the computation of event time sampled returns provides a separate source of error for our empirical results related to the standardized “financial time” return series, as the cut-off point given by the predetermined  $\tau^*$  will never be hit precisely given the discrete nature of the price observations and sampling frequency. Hence, the financial time will not be constant, but instead vary somewhat from one financial trading “day” to the next. Also, given the pronounced time variation in volatility, there will invariably be some standardized financial time returns that are constructed from very few underlying intraday observations, and these are likely to be poorly described by our approximating distributions. Finally, for real-time data there are inevitable data errors, partial market closures, failures of the data transmission systems, and so on. In short, there are many reasons why we may reject the sharp null hypothesis of i.i.d. standard normal returns. Therefore, it is particularly important to assess the proposed procedures within an ideal, albeit somewhat realistic, simulation setting as well as for an empirical setting that is likely to minimize the impact of the type of distortions listed above. These are the issues to which we now turn.

### **3. Simulation Evidence**

This section provides evidence on the finite sample distribution of the jump-adjusted and realized volatility standardized return series introduced in Section 2. We assume that the underlying high-frequency returns are generated by a jump diffusion that is calibrated to correspond roughly to the empirical features documented in recent empirical work based on the S&P 500 equity index.

Hence, the base scenario features a strong leverage effect, pronounced volatility persistence and fairly frequent jumps. We also explore the separate impact of the various features by studying the case of a pure diffusion without leverage, a pure diffusion with leverage, and a jump diffusion without any leverage effect. Finally, we study the impact of different jump intensities and average jump sizes. All experiments are conducted for a range of financial time trading periods,  $\tau^*$ , and

for different return horizons along with the associated sample sizes.

### 3.1. A Standard One-Factor Stochastic Volatility Jump-Diffusion Model

Our simulation evidence is based on the one-factor stochastic volatility jump-diffusion model estimated by Andersen, Benzoni and Lund (2002), henceforth ABL, from daily S&P 500 data, but the calibration of the jump component is also influenced by subsequent empirical work in ABD (2005). Although this is not necessarily the best model available it captures the dominant features of the equity index returns and the structure is sufficiently simple to allow for direct interpretation of the impact of the various components of the return generating process. The model takes the following form,

$$\begin{aligned} dp(t) &= \sigma(t) dW_1(t) + \kappa(t) dq_t \\ d\sigma^2(t) &= \eta(\theta - \sigma^2(t))dt + \nu \sigma(t) dW_2(t) . \end{aligned} \tag{15}$$

This is a standard affine (latent) stochastic volatility model augmented by a jump component. It is, of course, a special case of the general model class given in equation (1) where, for simplicity, the drift coefficient is set to zero. If the Wiener processes are independent and there are no jumps, we have the simple case of an affine stochastic volatility diffusion with the three volatility parameters  $\theta$ ,  $\eta$ , and  $\nu$  controlling, respectively, the unconditional (daily) return variance, the strength of mean reversion in the volatility process, and the volatility of volatility. Obviously, in this setting equation (7) applies. The additional real-world complications of primary interest are introduced if we allow for jumps and correlation between the return and volatility innovations. We capture these features in a parsimonious manner through the following representation,

$$dq_t \sim Po(\lambda); \quad \kappa(t) \sim N(0, \vartheta^2); \quad corr(dW_1(t), dW_2(t)) = \rho . \tag{16}$$

Hence, we consider Poisson jumps with a constant intensity rate of  $\lambda$  and with a lognormal jump size so that the jumps in the return process are normally distributed with mean zero and a variance given by  $\vartheta^2$ . Finally, the strength of the leverage effect is governed by  $\rho$ .

To keep the simulation manageable the diffusive volatility parameters are fixed throughout

at  $(\theta, \eta, \nu) = (1, 0.01, 0.1)$ . They imply an unconditional daily return variance of 1% and a strength of mean reversion for daily stock returns within the usual range and roughly consistent with ABL (2002). The leverage effect is also, whenever present, fixed at a value that reflect recent empirical studies for the U.S. equity index including ABL (2002), namely  $\rho = -0.5$ . Finally, the jump parameters are inspired by the contribution of jumps to overall daily return volatility in ABD (2005) although we experiment with different combinations of jump intensities and jump sizes to gauge the impact of this critical component for the properties of the standardized return series.

We provide results for a simulated sample size of 5,000 trading days with 195 intraday return observations corresponding to the use of two-minute returns over a 6½ hour trading day, reflecting our actual implementation with the S&P 500 futures data in Section 4. Details of the simulation design are provided in the appendix. We produce a total of 1,000 simulated samples. We also vary the length of the trading period over which we construct the standardized trading day returns from ½ day up to 2½ days in increments of ½ day. These are obtained from the original 5,000 trading days, so there are 10,000 half trading days, 3,333, trading periods of length 1½ trading day, 2,500 periods with a duration of 2 trading days, and 2,000 biweekly (2½ trading days) ones. We keep the number of total trading days fixed at 5,000 to account for the fact that intraday data typically are not available for financial assets before 1985, leaving a maximum of about twenty years of data. This configuration implies that there is a type of finite sample trade-off present in the choice of a longer versus shorter trading period return. The shorter trading period utilizes fewer intraday returns so the associated standardized return measures are relatively more noise but there are also more of these trading periods available over the 5,000 trading days, so the sampling variation across trading period returns can be assessed better than for the longer trading periods. For every such return series we also construct the corresponding jump-adjusted intraday return series and aggregate to the various horizons to obtain jump-adjusted trading period returns. The jump detection procedure is calibrated to obtain an (approximate) chance of 0.001% of falsely identifying a jump on a given trading day. Hence, we employ a very conservative jump extraction technique that seeks to control only for price moves that we are quite certain represent actual discontinuities. Finally, both the original (physical) intraday return series and the jump-adjusted series is converted into financial period return series. These financial time return observations cover a varying (calendar) time interval, but we calibrate the financial time clock via the choice of

$\tau^*$  to obtain five separate series which cover an average trading period matching those for the physical return series. The procedures employed for jump detection and adjustment and for financial return conversion are detailed in the appendix.

### 3.2. Simulation Results

For each simulation scenario we provide a variety of distributional tests and some illustrative plots conveying information about the quality of fit in different dimensions. The most basic information is conveyed by the descriptive statistics for the standardized returns and sizes associated with the i.i.d.  $N(0,1)$  null hypothesis tests or tests inspired by the approximating finite sample distribution in (14). We include three well-known empirical distribution function (EDF) tests for i.i.d.  $N(0,1)$ , namely the Kolmogorov-Smirnov (KS) test, the Anderson-Darling (AD) test and the Cramer-von Mises (CVM) test. The former is sensitive to deviation between the empirical distribution and the standard normal over the entire support while the latter two pay more attention to the behavior in the tails. We also incorporate the standard JB test which is a generic test for normality based on the third and fourth sample moments. As pointed out by Peters and de Vilder (2004), this test is likely to be incorrectly sized because of the thin tails of the finite sample distribution of the standardized returns. Under the simplifying assumptions of Section 2.2.2 we have an explicit expression for the expected fourth sample moment. Following Peters and de Vilder we therefore also incorporate a modified JB test based on an adjusted fourth moment that coincides with the expected value determined from the finite sample approximating distribution.

It turns out that the JB tests are very conservative and under-sized because they undertake an initial centering and rescaling of the standardized return series in order to accommodate a general non-centered non-unit variance normally distributed variate. Hence, we contrast the behavior of the JB test to those of direct moment tests for standard normality as well as for the approximating finite sample distribution in equation (14), where we denote the latter (finite sample) adjusted moment tests. These are also described in further detail in the appendix. Finally, we present standard Ljung-Box (LB) tests for serial correlation in the standardized return series and the corresponding squared standardized returns. We supplement these formal tests with some visual displays in the form of QQ plots - again with the  $N(0,1)$  rather than the more commonly used  $N(\mu, \sigma^2)$  distribution as the benchmark - and various data and density plots.

### 3.2.1. The No-Leverage Pure Diffusion Case

The most basic scenario involves the pure diffusive process with no correlation between the return and volatility innovations. This is a setting where the stochastic volatility process effectively is a pure time deformation device. If we control for the integrated volatility through the corresponding realized volatility measures, we should recover standard normality (approximately) as stated in equation (7). Of course, in practice we do not know the properties of the data generation process so we also explore what happens if we transform the series into financial time and/or implement initial jump detection and extraction procedures. From Table I, which provides information about the success and failure rates of our jump detection scheme, we note that the jump adjustment is likely to have a minimal impact in this scenario since, in the absence of true jumps and given our conservative jump identification scheme, very few jumps are erroneously identified. In fact, we identify a (non-existent) jump on 0.0003% of the days or much less than once per 5,000 trading day sample. Overall, we exploit this setting both as a reference point for the more complex scenarios and for investigating the properties of the test statistics we employ throughout the study.

The first set of simulations are obtained from model (15) with the extraneous parameters in equation (16) zeroed out. The descriptive statistics for the different standardized return series are given in Table II.1A. As expected, the averages of the mean, standard deviation and skewness of the standardized return series equal the expected values for the standard normal distribution and the standard deviation across the simulated samples is fairly small in each case. In contrast, we find a noticeable downward bias in the kurtosis, even compared to the downward revised value expected when accounting for the finite sample (given in the last column). It is most pronounced for the shorter trading periods, while it almost vanishes when we reach the biweekly frequency. In addition, the bias is slightly larger for the financial trading period returns relative to the “physical” (calendar) trading period returns. This is consistent with the finite sample result (14) which implies an associated fourth moment, given  $n_t$  intraday returns, of  $3 n_t / (n_t + 2)$ . Hence, the finite sample approximation (14) works exceedingly well for this simulation design. The anticipated bias in the kurtosis induce rejections of the null hypothesis of a standard normal distribution. However, especially for lower values of  $n_t$ , this is an excessively strict null hypothesis as the finite sample distribution simply cannot be Gaussian. We further note that the varying number of intraday returns used in the computation of the kurtosis for the financial time return series will

induce an additional negative bias due to Jensen's inequality (applied to  $3n/(n+2)$ ). This observation inspires our reliance on some additional moment adjusted test statistics below.

We report evidence on the size for a set of common normality tests in Table II.1B. For parsimony, we report results for test size 5% only, but qualitatively identical results were obtained for other significance levels.. The EDF tests KS, CVM and AD all have good size properties although there is a tendency for overrejection. This is likely associated with the mild deviations from standard normality in finite samples discussed previously. The JB tests behave more erratically. As may be expected, they especially tend to over-reject for the shorter trading periods where the bias in the kurtosis is worst. However, the JB tests turn out to be undersized at daily and longer frequencies due to the initial step of centering and scaling. Since the relevant null hypothesis is  $N(0,1)$  this entails a loss of power, and the test under-rejects even if the (finite sample) distributions are not exactly  $N(0,1)$ . If one follows Peters and de Vilder (2004) in adjusting the theoretical value of the kurtosis in the JB test to equal the finite sample adjusted value the over-rejections for the half-day period is reversed to under-rejections. Moreover, these under-rejections amplify for the longer trading period returns. Here, the confounding effect of the centering and scaling is evident, causing the JB test to be systematically undersized for this (finite sample adjusted) variant of an  $N(0,1)$  test. As expected, the jump detection procedure has no visible impact on the results, whereas the conversion into financial time is a bit more problematic as the various tests now appear to be systematically, albeit mildly, oversized.

A more direct testing strategy for the standardized returns is to exploit the moments of the two approximating distributions, namely the standard normal and the finite sample approximation in equation (14), thus sidestepping the loss of power associated with the JB test. Such moment based results are presented in Tables II.1C and II.1D for each individual moment and selected combination of the moments, including a joint test based on all four moments. Table II.1C refers to the test size computed against the standard normal moments. The size of the tests based on individual moments appear sensible, while the joint moment tests are oversized for the smallest trading period. However, this problem vanishes as the trading period is increased. This is, of course, not surprising since the normal approximation is found to be quite accurate at the daily trading frequency and above in Figure 1. The results in Table II.1D based on the approximating finite sample distribution are uniformly impressive, as they also account for the downward bias in

the fourth moment over the lowest trading periods. These results also suggest that the convergence of the sampling distributions to the standard normal becomes reliable around the one to one-and-a-half trading day period. Consequently, these tests provide a convenient supplement to the empirical distribution tests for standard normality, whereas the JB tests are much less reliable. Although we report JB tests in the tables, we largely ignore them in the subsequent exposition.

We further present the size properties of standard Ljung-Box (LB) tests for uncorrelated return series in Tables II.1E and II.1F. These tests generally behave exemplary. After performing the standardization with realized volatility there is no evidence of serial correlation at lags 1, 2, 5, 10, 20 or 50. Hence, the pronounced volatility persistence of the underlying return series has been effectively captured by the empirical realized volatility measures. Again, a slight deterioration in the results for the financial time return series may be visible, but the effect is minor. The LB tests turn out to be uniformly well behaved for all our simulation settings and we do not report further simulated LB test results. These are available upon request.

Finally, we present graphical evidence for a random sample (the last) from our simulation experiment. This particular series is to a very good approximation distributed as an i.i.d. standard normal sample, as documented by the QQ plot and (log) pdf plots in Figures II.1A and II.1B.

### **3.2.2. The No-Leverage Jump-Diffusion Case**

We now introduce jumps into the model. Of course, both the impact of jumps and the difficulty in identifying and adjusting for them are likely to hinge on the specific features of the jump process. We explore two opposite benchmark cases and one positioned in-between. The parameter values are inspired by the empirical findings for futures contracts in ABD (2005). The specific parameter values chosen are  $(\vartheta^2, \lambda) = (2.5, 0.1)$ . Under this scenario, we have a jump about once every two weeks and these jumps account for 20% of the overall return volatility which is reasonably consistent with the evidence in ABD (2005), even if the jump sizes may appear slightly extreme. However, they occur with low intensity given the recent high-frequency based evidence for the equity market, and this is reinforced by the fact that many of the monthly regularly scheduled macroeconomic announcements induce an instantaneous move or jump in the market, see, e.g., ABDV (2005). Finally, this scenario should provide a good sense of how the presence of relatively large jumps impacts the distribution of the standardized returns.

Turning to the results, we first note from Table II.2A that the kurtosis for the standardized return series unadjusted for jumps now is even more downward biased. This is in line with the reasoning in Section 2.2.1 that jumps tend to induce thin tails as they impact the realized volatility more than the corresponding daily return. In contrast, there is essentially no bias for the jump-adjusted series relative to the finite sample corrected values, reflecting the fact that the jump test succeeds in capturing the majority of the large jumps. These encouraging findings are confirmed in Table II.2B where the empirical distribution function (EDF) based tests (KS, CVM, AD) for the jump-adjusted series all are correctly sized for the one day trading period and beyond, although the slight deterioration noted above again occurs for the financial time transformed series. In the absence of leverage, this transformation does have a moderately harmful effect. Not surprisingly, we now find quite severe size distortions for the series unadjusted for jumps. These conclusions are corroborated by the moment tests. The joint test based on all four moments is oversized for all scenarios except the jump-adjusted series with trading periods of one day or above in Table II.2C, while the joint adjusted moment test in Table II.2D provide the same results except that even the jump-adjusted series for the  $\frac{1}{2}$  trading day now also in correctly sized.

In summary, our sequence of jump adjustment, financial time transformation - if the presence of leverage cannot be ruled out - and standardization provides a feasible method for constructing correctly sized tests for the distributional properties implied by a diffusion extended with rare large jumps. Moreover, it is clear that the jump adjustment is critical for this procedure to be successful. However, these findings may dependent on the specific scenario studied. In fact, our jump detection scheme performs very well in this setting as seen in Table I. We extract jumps accounting for 19.86% of the overall return variability compared to the true contribution in this model of 20%. There are only minor mistakes in terms of missing jumps (0.615% of overall variability) and identifying jumps erroneously (0.54% of overall variability). Hence, it is critical to explore the performance of our procedure with alternative jump specifications.

We next consider frequent but small jumps which has the same contribution to overall volatility as above. The specific jump parameters are now  $(\vartheta^2, \lambda) = (0.25, 1)$  so the jumps are about ten times more frequent, but also ten times smaller. If anything the results are now stronger, as may be seen from Tables II.3A-II.3F. The tests for the jump-adjusted series have about the right size both for the EDF and the moment based tests, while there are noteworthy distortions for the

unadjusted returns, especially over the shorter trading periods. The joint tests based on all four moments is particularly powerful in suggesting the inadequacy of the unadjusted return series. From Table I it is evident that we fail to detect jumps more frequently within this setting, as we correctly extract jump volatility corresponding to 16.8% (= 17.9 - 1.1) of overall variability while we miss the remaining 3.2% arising from jumps and we erroneously label about 1.1% of the variability as jump volatility when it actually stems from the diffusive variation. More strikingly, we only detect jumps on 37% of the days when there on average is a jump per day. This implies that we do capture the vast majority of the large jumps which are the most critical to control for.

Finally, we explore a “moderate” jump scenario. The specification is  $(\vartheta^2, \lambda) = (0.5, 0.2)$  with a much smaller overall contribution to the return variability of 9.1%. Table I shows that we again miss a substantial fraction of the jumps but that we tend to identify the vast majority of the large jumps. Thus, not surprisingly, the tests for the jump-adjusted series are appropriately sized. However, the unadjusted series now produce somewhat similar rejections frequencies, so it is harder to distinguish the findings produced by the jump-adjustment series compared to the unadjusted series. In fact, the highest rejection rates for the unadjusted series at trading periods of one day or more in Tables II.4B-II.4D are not much above 10% for the 5% test level.

### **3.2.3. The Pure Diffusion with Leverage Case**

We now exclude jumps but introduce a strong, albeit realistic, leverage effect by letting  $\rho = -0.5$ . First, we note that the jump detection scheme is identical in the no leverage and leverage scenario so the minor discrepancies for the two cases in Table I stem solely from the different return standardization schemes used in physical and financial time. Clearly, the overall impact of the jump adjustment procedures will be very similar to what has already been discussed above for the no-leverage models. Hence, we shall not discuss this aspect of the findings much below.

Table II.5A reveals that leverage has a very different impact than jumps. Now, the mean is biased upward whereas the kurtosis accords with the theoretical value after controlling for the finite number of observations used in the computation of the trading period returns. The upward shift in the mean is due to the stronger standardization of negative returns stemming from their correlation with the volatility innovations and hence realized volatility. Table II.5B shows that the EDF tests are substantially oversized for all (physical) calendar period return series, whether jump

adjusted or not, while they are well sized for the returns computed in financial time. Hence, the financial time transformation is successful, as stipulated by equation (8). These findings are corroborated by Tables II.5C and II.5D. Another distinct difference to the jump scenario is the lack of problems with test size even for the shorter trading period in financial time so finite sample considerations are simply less pertinent here. It is evident that the realized volatility measures approximate the underlying integrated variances sufficiently well that the transformed series are remarkably close to standard normal.

#### **3.2.4. The Jump-Diffusion with Leverage Case**

Finally, we combine the leverage effect with the jump scenarios explored in Section 3.2.2. This produces a setting which arguably is most reminiscent of actual equity index return series. The results are documented in Tables II.6A-II.6.D, II.7A-II.7D and II.8A-II.8D. The descriptive statistics show that we now encounter both an upward bias in the mean and a downward bias in the kurtosis, as may be expected from a combination of the above scenarios. When the jump contribution to overall return variability is high (Tables II.6 and II.7) it is critical to perform both a jump adjustment and a financial time transformation in order to obtain correctly sized tests. The two features appear to reinforce each other to produce quite badly sized tests if either of the steps are omitted. For the moderate jump scenario (Tables II.8), we obtain the same qualitative findings but in attenuated form. This suggests that within moderate jump scenarios the financial time transformation is by far the most critical step in terms of achieving reasonably sized tests.

### **4. Analysis of the S&P 500 Futures Returns**

This section explores two-minute transaction returns from the S&P 500 futures contract traded at the Chicago Mercantile Exchange (CME) over the period January 1, 1988 - July 26, 2004. This is the same basic data used by Peters and de Vilder (2004) in their original study of the continuous semi-martingale hypothesis, except that they rely on the January 1, 1988 - August, 2001 period to avoid potential confounding effects from the September 11, 2001 events. As such, our study may be seen as providing additional perspectives on their results by studying a number of additional tests as well as allowing for discontinuities in the underlying return generating process. We also extend their sample period up to the present so that it includes the volatile period around

September 2001. This should speak to the robustness of the findings. We will also provide estimates based on the 1988-2001 sample to ensure compatibility with earlier reported results.

An initial inspection of the high-frequency S&P 500 futures data revealed some problems with trading days that were abbreviated by missing activity or data error in the beginning or at the end of the active trading period of 9:30am - 4:00pm. Missing observations manifest themselves as zeros in the corresponding return series. While this is not a major problem for the computation of realized volatility, it will influence the length of the trading days in financial time and it will have potentially severe effects on the important finite sample approximations that we exploit. Hence, we eliminate trading days which start or end with a string of ten consecutive zero returns. This eliminates a mere 44 days, leaving 4,126 trading days over the full 1988-2004 sample. Likewise, the 1988-2001 sample period loses a total of 34 days, leaving 3,420 trading days for analysis. Additional details about the construction of the 195 two-minute returns for each trading day is given in the appendix.

#### **4.1. Analysis of the Full Sample**

We first review the findings for the full sample. The descriptive statistics in Table III.1A reveal that the mean returns in physical time are positive and highly significant, whereas the returns in financial time approximately zero. This strong discrepancy across the sampling schemes is likely indicative of a pronounced leverage effect as discussed previously. A positive bias is also evident in the skewness, while the standard deviation and kurtosis are strongly downward biased. Because the latter measures may be unduly influenced by the uncentered first and second moments, we have included a tabulation of the raw standardized return moments as well in Table III.1A'. The biases are still readily apparent across the full range of series, although they are markedly lower for the financial time return series.

Turning to the formal hypothesis tests, Table III.1B provides the EDF tests. The evidence is striking. All EDF tests reject the standard normal null hypothesis with p-values typically close to 0%. In sharp contrast, none of the EDF tests for the series sampled in financial time reject at the 5% level. The erratic behavior of the JB tests, albeit consistent with our prediction of them being severely undersized for the longer trading period returns, is consistent with our view that they are of little use in this context. Moving on to the formal moment based tests, Table III.1C underscores

the strong rejections of the standard normal hypothesis in physical time with p-values uniformly close to zero. Again, the financial time return series appear much closer to the strict Gaussian benchmark, although there are a number of small p-values for the lower trading periods and especially for the series unadjusted for jumps. This rejection pattern may reflect the inadequacy of the standard normal approximation for the lower number of intraday observations, suggesting that the basic theory may still be sensible, or it may reflect the lack of power to reject the null as there are a lower number of return observations underlying the tests as we increase the length of the trading period. Table III.1D sheds additional light on the issue by exploiting the finite sample correction associated with the approximating density in equation (14). The calendar time based return series continue to produce extraordinarily strong rejections across the board while the financial time return series now appear even closer to the standard normal benchmark. There is also some indication that the jump-adjusted series provides the closest approximation to the finite sample corrected moments as none of the corresponding tests reject for one (financial) trading day returns with all the p-values exceeding 10%, while the corresponding unadjusted return series encounter rejections at the 5% level for a few of the tests and produces a p-value of 7% for the joint test based on all four moments. As in the previous table, once we move above the one day trading period none of the tests reject the null hypothesis.

A couple of factors may explain our general findings. First, we likely have low power to reject the null hypothesis for the longer trading period returns so that the discrepancy between the jump-adjusted and unadjusted series may be relatively uninformative. Second, the actual data may resemble the moderate jump scenario where the discrepancy between the two series necessarily is very small as documented in the simulation section. Obviously, these two points would tend to reinforce each other to render a clear verdict difficult. Nonetheless, it is striking that the standard normal approximation is acceptable across many dimensions once we transform the return series to financial time. As usual, the LB tests for the trading period returns and squared returns are largely uninformative although we do see some strong rejections for the calendar period returns.

Finally, we provide some graphical illustration of the quality of fit in Figures III.1A and Figure III.1B. The QQ plots are explicitly matched to the standard normal benchmark and not a generic normal distribution. The tendency for the empirical distribution line to veer off the benchmark diagonal line is indicative of the severe distortions in the underlying distribution

relative to the theoretical benchmark. In contrast, the financial period return plots in Figure III.1B hug the line throughout with only a few deviations visible in the tail areas. It may even be possible to discern some evidence of a moderately better fit for the financial time jump-adjusted series relative to the unadjusted counterpart although the discrepancy is not compelling from these displays. Similar conclusions arise from the corresponding log density displays in the figures.

A more direct assessment of the evidence for jumps in the intraday return series may be obtained from inspection and analysis of the extracted jump series. We facilitate this step by reporting some descriptive statistics for the jump series in Table IV. We identify a total of 382 jumps, whereas we would expect to detect zero or one jump at the 0.001% level over the sample. In total, the jumps contribute about 4.4% to the overall return variability. It is evident that the intraday return series contain some extreme outliers that appear incompatible with a diffusion framework. The extracted jump series is displayed in Figure IV.1A along with the original intraday return series and the jump-adjusted series. Notice that the only remaining visible outliers in the jump-adjusted series are located during volatile periods where the jumps are found not to be statistically significant at the conservative significance level we have adopted. The same effect is visible in the corresponding daily trading period displays in Figure IV.1B. Nonetheless, the latter figure also highlights the relatively minor modification of the daily return series that has been induced through the jump extraction procedure. This partially explains the comparatively minor discrepancies we have identified between the distributional properties of the jump-adjusted and unadjusted standardized return series in financial time.

A couple of issues are worth further discussion. First, we identify relatively few jumps compared with recent studies that implement less conservative jump tests such as ABD (2005). Nonetheless, we have confirmed that even if we increase the p-values we are still somewhat short of the jump intensity that they report. The primary reason is that we exclude the generally less active morning period covering 8:20-9:30am. The main macroeconomic announcement releases take place at 8:30am, so we miss the corresponding, often large, jumps. On the other hand, this period is usually very quiet in the absence of macroeconomic news, so it is problematic to include it for every trading day, as the distributional properties of financial time and jump-adjusted return series are very sensitive to the underlying number of (active) intraday returns and the requisite finite sample adjustments are dependent on the market being “open” for the intervals included in

the trading period return computations. In order to avoid problems stemming from such issues, we exclude this heterogeneous segment of the trading day from the analysis. This is appropriate for testing the distributional predictions of interest in this paper and it facilitates a direct comparison to the Peters and de Vilder (2004) study. In contrast, if the properties of the jump series per se is of primary concern, it may be advisable to use a lower significance level in the jump tests and to include the periods surrounding the macroeconomics news releases in the analysis.

Second, it is intuitively clear that the actual return data are quite a bit more complex than the ones generated in the simulations. There are the ubiquitous microstructure effects generating an extra layer of noise and there are likely multiple volatility factors and jump processes at work in the actual data. Consequently, there may be excessive variability in the extracted return process which inflates the realized volatility measures and makes it still harder to identify smaller jumps, especially when they occur during volatile trading periods.

The positive aspect of our findings is that we are unable to reject the basic distributional implications of the assumed generic jump-diffusion setting in spite of such possible shortcomings. In fact, in light of the simulation evidence we have assembled rather compelling evidence for the presence of a pronounced leverage effect. Moreover, the case for the existence of discontinuities in the high-frequency return series is overwhelming, even if the associated adjustments have a more limited impact on the evidence regarding the suitability of the pure diffusive characterization of the equity index returns. The theoretical observations regarding the effect of jump removal as well as the extensive simulation evidence help us understand the potential lack of power that our procedure may possess versus a no-jump alternative. Moreover, as previously noted, this particular S&P 500 series is artificially short on jumps. Most relevant asset return series will be subject to larger and/or more frequent jumps than the one analyzed here. In such settings, we expect the jump detection step to be even more critical for our ability to undertake sensible inference along the lines of this paper.

#### **4.2. Analysis of the 1988-2001 Subsample**

We conclude our investigation with a brief comparison to Peters and de Vilder (2004). For that purpose we report selected results covering their sample period. The descriptive statistics in Table III.2A report kurtosis measures that seemingly are much more compatible with the theoretical

values. However, the biases in the raw return moments are also present here, as documented in Table III.2A'. The bottom line is that the EDF tests in Table III.2B and the adjusted moment tests in Table III.2D provide qualitatively identical findings to those for the full sample period. The physical return series still yield exceptionally strong rejections and even the suggestive evidence that the jump-adjusted financial time series are slightly better approximated by the standard normal is present, as the fourth moment test for the unadjusted "daily" financial time series is rejected at the 5% level whereas all jump-adjusted series for one financial trading "day" or more produce high p-values in Table III.2D. The overall contribution to return variability of the detected jumps is slightly higher here at 5.3% and there are a total of 291 identified jumps. Hence, the general conclusions are identical to those above.

How does the above evidence compare to the conclusions of Peters and de Vilder (2004)? First, they only provide evidence for the standardized financial time returns based on an average of 2½ trading days, although they mention having documented robustness for trading periods lasting half that and considerably longer than that as well. Hence, they do not present results for standardized calendar returns and they do not allow for jumps. Second, they apply only two direct normality tests, namely the KS test and the (adjusted) JB test. These are supplemented by a tail test and two tests focusing on independence of the standardized financial time return series. Our simulation evidence documents a considerable loss of power for trading period exceeding two days. Of the jump scenarios investigated, only the rare large jumps provide any hint that we can discriminate between a jump and no jump scenario for such series. For the other settings the KS statistics has essentially no power versus the no jump alternative. Of course, we found that the second test they use, namely the JB test, is close to useless. It is also easy to confirm that the tail test has low power for the sample size that they entertain. Finally, we found that tests for independence are unlikely to provide much insight as the standardization by realized volatility in-and-of-itself suffices to produce uncorrelated normalized return series. Hence, these tests also have little power to discriminate among the various features of the intraday return series.

In summary, we vastly expand upon the analysis and insights provided by Peters and de Vilder (2004). We introduce a three step analysis that allows us to consider the simplest possible stochastic volatility diffusion process separately from the diffusion with leverage and/or jumps. Moreover, we systematically explore the size of a battery of tests for i.i.d. standard normality and,

as a byproduct, observe the power of the tests to reject the null hypothesis when not all relevant features of the return series are accommodated. Equally importantly, we identify a serious loss in power when the financial time trading period is long. Finally, we are able to center our empirical study on a number of well-behaved test statistics. As a result, our findings differ in substantial ways. We conclude that jumps are important in the return generating process so that the null hypothesis of a continuous semi-martingale is incorrect. Furthermore, we present suggestive evidence that accounting for the presence of jumps allows us to draw a more firm conclusion regarding the more general null hypothesis that the underlying return process is generated by a process in the semi-martingale class. Finally, the simulation evidence suggests that there are many other return processes for which it is crucial to control for the jump features in order not to reach misleading conclusions regarding the nature of the underlying return generating process.

#### **4. Conclusion**

We have introduced a novel three step procedure to explore the general properties of the return generating process underlying a given intraday return series. Each step speaks to the importance of an important empirical feature of the corresponding daily returns, namely stochastic volatility, the presence of an asymmetric relationship between return and volatility innovations (the leverage effect) and the existence of jumps or extreme outliers in the return distribution. In combination, the procedure may be taken as an informal test for whether the underlying return process belongs to the arbitrage-free class of continuous-time semi-martingales. The properties of the relevant test statistics associated with each step of the procedure are explored through an extensive simulation study. Finally, we show that the empirical behavior of the financial time standardized returns from the S&P 500 futures equity index market is compatible with a jump-diffusion endowed with a pronounced leverage effect. The associated decomposition into a jump process, an indication of the strength of the leverage effect, and a measure of the path of integrated variance associated with the diffusive component of the return process should be useful for building daily or longer horizon conditional return distribution models based on high-frequency return observations. For now, we only claim to have shed some new light upon the relationship between the recent high-frequency realized volatility based literature and the more traditional daily conditional return distribution literature based on the discrete-time ARCH or stochastic volatility paradigm.

A number of issues call out for additional inquiry. First, how widely can these strict tests for normality of the appropriately standardized return series be expected to apply? The S&P 500 futures series is quite unique in terms of having minimal market microstructure distortions. Second, the various steps involved in our test procedure may be refined or extended in a variety of ways. Third, one should explore the impact of microstructure noise and a range of other complications within the simulation setting. Fourth, the decomposition of the return process may be related to other market activity variables that should feature similar behavior according to the MDH theory. Finally, there is work to do in terms of documenting the direct usefulness of the current results for practical financial decision making.

## References

- Aït-Sahalia, Y. (2002), "Telling from Discrete Data Whether the Underlying Continuous-Time Model is a Diffusion," *Journal of Finance*, 57, 2075-2121.
- Aït-Sahalia, Y., P.A. Mykland and L. Zhang (2005), "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18, 351-416.
- Alizadeh, S., M. Brandt and F.X. Diebold (2002), "Range-Based Estimation of Stochastic Volatility Models," *Journal of Finance*, 57, 1047-1091.
- Andersen, T.G. (1996), "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility," *Journal of Finance*, 51, 169-204.
- Andersen, T.G. (1992), "Volatility," Manuscript, Northwestern University.
- Andersen, T.G., L. Benzoni and J. Lund (2002), "Estimating Jump-Diffusions for Equity Returns," *Journal of Finance*, 57, 1239-1284.
- Andersen, T.G. and T. Bollerslev (1997), "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High-Frequency Returns," *Journal of Finance*, 52, 975-1005.
- Andersen, T.G. and T. Bollerslev (1998), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885-905.
- Andersen, T.G., T. Bollerslev, P. Christoffersen and F.X. Diebold (2005), "Volatility and Correlation Modeling," in *Handbook of Economic Forecasting* (G. Elliott, C.W.J. Granger and A. Timmermann, eds.). Elsevier Science, New York, forthcoming.

- Andersen, T.G., T. Bollerslev and F.X. Diebold (2004), "Parametric and Non-Parametric Volatility Measurement," in *Handbook of Financial Econometrics* (L.P Hansen and Y. Aït-Sahalia, eds.). Elsevier Science, New York, forthcoming.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and H. Ebens (2001), "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43-76.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2000), "Great Realizations," *Risk*, 13, 105-108.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001), "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42-55.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579-625.
- Andersen, T.G., T. Bollerslev and F.X. Diebold (2005), "Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility," Manuscript, Northwestern University, Duke University, and University of Pennsylvania.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and C. Vega (2003), "Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange," *American Economic Review*, 93, 38-62.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and C. Vega (2005), "Real-Time Price Discovery in Stock, Bond and Foreign Exchange Markets," Manuscript, Northwestern University, Duke University, University of Pennsylvania, and University of Rochester.
- Andersen, T.G., T. Bollerslev, P.H. Frederiksen and M.Ø. Nielsen (2005), "Jumps in Common Stock Returns," Manuscript, Northwestern University, Duke University, and Cornell University.
- Andersen, T.G., T. Bollerslev and N. Meddahi (2004), "Analytic Evaluation of Volatility Forecasts," *International Economic Review*, 45, 1079-1110.
- Andersen, T.G., T. Bollerslev and N. Meddahi (2005), "Correcting the Errors: Volatility Forecast Evaluation using High-frequency Data and Realized Volatilities," *Econometrica*, 72, 279-296.
- Ané, T. And H. Geman (2000), "Order Flow, Transaction Clock, and Normality of Aset Returns," *Journal of Finance*, 55, 2259-2284.
- Back, K. (1991), "Asset Prices for General Processes," *Journal of Mathematical Economics*, 20, 317-395.
- Ball, C.A. and W.N. Torous (1983), "A Simplified Jump Process for Common Stock Returns," *Journal of Financial and Quantitative Analysis*, 18, 53-65.
- Bandi, F.M. and J.R. Russell (2004a), "Microstructure Noise, Realized Volatility, and Optimal Sampling," Manuscript, University of Chicago.

- Bandi, F.M. and J.R. Russell (2004b), "Separating Microstructure Noise from Volatility" Manuscript, University of Chicago.
- Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2005), "Regular and Modified Kernel-Based Estimators of Integrated Variance: The Case with Independent Noise," Manuscript, Stanford University.
- Barndorff-Nielsen, O.E. and N. Shephard (2001), "Non-Gaussian Ornstein-Uhlenbeck-based Models and Some of Their Uses in Financial Economics," *Journal of the Royal Statistical Society, Series B*, 63, 167-241.
- Barndorff-Nielsen, O.E. and N. Shephard (2002a), "Econometric Analysis of Realised Volatility and its use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society*, 64, 253-280.
- Barndorff-Nielsen, O.E. and N. Shephard (2002b), "Estimating Quadratic Variation Using Realized Variance," *Journal of Applied Econometrics*, 17, 457-478.
- Barndorff-Nielsen, O.E. and N. Shephard (2004), "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1-37.
- Barndorff-Nielsen, O.E. and N. Shephard (2005), "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, forthcoming.
- Bates, D.S. (2000), "Post-'87 Crash fears in the S&P500 Futures Option Market," *Journal of Econometrics*, 94, 181-238.
- Bekaert, G. and G. Wu (2000), "Asymmetric Volatility and Risk in Equity Markets," *Review of Financial Studies*, 13, 1-42.
- Black, F. (1976), "Studies of Stock Price Volatility Changes," *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics*, 177-181.
- Bollen, B. and B. Inder (2002), "Estimating Daily Volatility in Financial Markets Utilizing Intraday Data," *Journal of Empirical Finance*, 9, 551-562.
- Bollerslev, T. (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics*, 69, 542-547.
- Bollerslev, T., R.F. Engle and D.B. Nelson (1994), "ARCH Models," in *Handbook of Econometrics, Volume IV* (R.F. Engle and D. McFadden, eds.). North Holland, Amsterdam.
- Bollerslev, T., J. Litvinova and G. Tauchen (2005), "Leverage and Volatility Feedback Effects in High-Frequency Data," Manuscript, Duke University.
- Campbell, J.Y. and L. Hentschel (1992), "No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns," *Journal of Financial Economics*, 31, 282-331.

- Carr, P., H. Geman, D. Madan and M. Yor (2002), "The Fine Structure of Asset Returns: An Empirical Investigation," *Journal of Business*, 75, 305-332.
- Carr, P., H. Geman, D. Madan and M. Yor (2003), "Stochastic Volatility for Lévy Processes," *Mathematical Finance*, 13, 345-382.
- Chan, W.H. and J.M. Maheu (2002), "Conditional Jump Dynamics in Stock Market Returns," *Journal of Business and Economic Statistics*, 20, 377-389.
- Chernov, M., A.R. Gallant, E. Ghysels, and G. Tauchen (2003), "Alternative Models for Stock Price Dynamics," *Journal of Econometrics*, 116, 225-257.
- Chib, S., F. Nardari and N. Shephard (2002), "Markov Chain Monte Carlo Methods for Stochastic Volatility Models," *Journal of Econometrics*, 108, 281-316.
- Christie, A.A. (1982), "The Stochastic Behavior of Common Stock Variances - Value, Leverage and Interest Rate Effects," *Journal of Financial Economics*, 3, 145-166.
- Clark, P.K. (1973), "A Subordinate Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41, 135-155.
- Comte, F. and E. Renault (1998), "Long Memory in Continuous Time Stochastic Volatility Models," *Mathematical Finance*, 8, 291-323.
- Corsi, F., Zumbach, U.A. Müller and M. Dacorogna (2001), "Consistent High-Precision Volatility from High-frequency Data," *Economic Notes*, 30, 183-204.
- Dacorogna, M.M., R. Gencay, U.A. Müller, O.V. Pictet and R.B. Olsen (2001). *An Introduction to High-Frequency Finance*. San Diego: Academic Press.
- Dambis, K.E. (1965), "On the Decomposition of Continuous Submartingales," *Theory of Probability Applications*, 10, 401-410.
- Drost, F.C. and B.J.M. Werker (1996), "Closing the GARCH Gap: Continuous Time GARCH Modeling," *Journal of Econometrics*, 74, 31-57.
- Drost, F.C., T.E. Nijman and B.J.M. Werker (1998), "Estimation and Testing in Models Containing Both Jumps and Conditional Heteroskedasticity," *Journal of Business and Economic Statistics*, 16, 237-243.
- Dubins, L.E. and G. Schwartz (1965), "On Continuous Martingales," *Proceedings of the National Academy of Sciences*, 53, 913-916.
- Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987-1008.
- Engle, R.F. and J. Russell (2004), "High-Frequency and Transaction Data," in *Handbook of Financial Econometrics* (L.P Hansen and Y. Aït-Sahalia, eds.). Elsevier Science, New York, forthcoming.

- Epps, T.W. and M.L. Epps (1976), "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis," *Econometrica*, 44, 305-321.
- Eraker, B. (2004), "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *Journal of Finance*, 59, 1367-1403.
- Eraker, B., M.S. Johannes and N.G. Polson (2003), "The Impact of Jumps in Volatility," *Journal of Finance*, 58, 1269-1300.
- Forsberg, L. and T. Bollerslev (2002), "Bridging the Gap Between the Distribution of Realized (ECU) Volatility and ARCH Modeling (of the Euro): The GARCH-NIG Model," *Journal of Applied Econometrics*, 17, 535-548.
- Gallant, A.R., C.T. Hsu and G. Tauchen (1999), "Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance," *Review of Economics and Statistics*, 81, 617-631.
- Garman, M.B. and M.J. Klass (1980), "On the Estimation of Price Volatility from Historical Data," *Journal of Business*, 53, 67-78.
- Ghysels, E., A.C. Harvey and E. Renault (1996), "Stochastic Volatility," in *Handbook of Statistics, Volume 14* (G.S. Maddala and C.R. Rao, eds.). North Holland, Amsterdam.
- Hansen, P.R. and A. Lunde (2006), "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, forthcoming.
- Huang, X. and G. Tauchen (2005), "The Relative Contribution of Jumps to Total Price Variance," *Journal of Financial Econometrics*, forthcoming.
- Jarque, C. and A. Bera (1980), "Efficient Tests for Normality, Heteroskedasticity, and Serial Independence or Regression residuals," *Economic Letters*, 6, 255-259.
- Johannes, M. (2004), "The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models," *Journal of Finance*, 59, 227-260.
- Johannes, M., R. Kumar and N.G. Polson (1999), "State Dependent Jump Models: How do US Equity Markets Jump?" Manuscript, University of Chicago.
- Karatzas, I. And S.E. Shreve (1991). *Brownian Motion and Stochastic Calculus, 2<sup>nd</sup> ed.* Springer Verlag, New York.
- Maheu, J.M. and T.H. McCurdy (2004), "News Arrival, Jump Dynamics and Volatility Components for Individual Stock Returns," *Journal of Finance*, 59, 755-793.
- Mancini, C. (2004), "Estimation of the Parameters of Jump of a General Poisson-Diffusion Model," *Scandinavian Actuarial Journal*, 104, 42-52.

- Mancini, C. (2005), "Disentangling the Jumps of the Diffusion in a geometric Jumping Brownian Motion," Manuscript, Universita di Firenze.
- Meddahi, N. (2002), "A Theoretical Comparison Between Integrated and Realized Volatility," *Journal of Applied Econometrics*, 17, 479-508.
- Merton, R.C. (1976), "Option Pricing when Underlying Stock Returns are Discontinuous," *Journal of Financial Economics*, 3, 125-144.
- Nelson, D.B. (1990), "ARCH Models as Diffusion Approximations," *Journal of Econometrics*, 45, 7-38.
- Nelson, D.B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347-370.
- Nelson, D.B. (1992), "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model," *Journal of Econometrics*, 52, 61-90.
- Oomen, R.C.A. (2004), "Properties of Realized Variance for a Pure Jump Process: Calendar Time Sampling versus Business Time Sampling," Manuscript, University of Warwick.
- Pan, J. (2002), "the Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time Series Study," *Journal of Financial Economics*, 63, 3-50.
- Parkinson, M. (1980), "The Extreme Value Method for Estimating the Variance of the Rate of Return," *Journal of Business*, 53, 61-65.
- Peters, R.T. and R.G. de Vilder (2004), "Testing the Diffusion Model for the S&P500," Manuscript, University of Amsterdam.
- Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press, Oxford UK.
- Tauchen, G. and M. Pitts (1983), "The Price Variability-Volume Relationship on Speculative Markets," *Econometrica*, 51, 485-505.
- Todorov, V. and G. Tauchen (2005), "Simulation Methods for Lévy-Driven CARMA Stochastic Volatility Models," Manuscript, Duke University.
- Zhang, L., Y. Aït-Sahalia and P.A. Mykland (2005), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, forthcoming.
- Zhou, B. (1996), "High-Frequency Data and Volatility in Foreign Exchange Rates," *Journal of Business and Economic Statistics*, 14, 45-52.
- Zhou, B. (1998), "F-Consistency, De-Volatilization and Normalization of High-Frequency Financial Data," in *Nonlinear Modelling of High Frequency Financial Time Series* (C. Dunis and B. Zhou, eds.). John Wiley Sons Ltd., London.

## APPENDIX I. Moment-Based Tests

Let  $\tilde{R}_t$ ,  $t = 1, 2, \dots, T$  be independent and distributed with pdf

$$f_{\tilde{R}_t}(\tilde{r}) = \frac{\Gamma(n_t/2)}{\sqrt{\pi n_t} \Gamma((n_t-1)/2)} \left(1 - \frac{\tilde{r}^2}{n_t}\right)^{\frac{(n_t-3)}{2}} 1_{(-\sqrt{n_t}, \sqrt{n_t})}(\tilde{r}), \quad \tilde{r} \in \mathbb{R}, n_t \in \mathbb{N}.$$

Let the raw moments of  $\tilde{R}_t$  be defined as  $m_{k,t} = \mathbf{E}[\tilde{R}_t^k]$

**Lemma 1.**  $m_{2k-1,t} = 0$ ,  $m_{2k,t} = \frac{n_t^k \cdot (2k-1) \cdot \dots \cdot 1}{(n_t+2k-2) \cdot \dots \cdot n_t}$ ,  $k=1, 2, \dots$

**Proof.** See Peters & deVilder (2004)

**Lemma 2.** For a sample  $\tilde{R}_t$ ,  $t = 1, 2, \dots, T$  of independent observations define the s-th raw sample

moment  $\bar{M}_s = \frac{1}{T} \sum_{t=1}^T \tilde{R}_t^s$  and its population value  $M_s = \mathbf{E}[\bar{M}_s] = \frac{1}{T} \sum_{t=1}^T m_{s,t}$ . Then

$$M_1 = 0, M_2 = 1, M_3 = 0, M_4 = \frac{1}{T} \sum_{t=1}^T \frac{3n_t}{n_t+2}, M_6 = \frac{1}{T} \sum_{t=1}^T \frac{15n_t^2}{(n_t+4)(n_t+2)}, M_8 = \frac{1}{T} \sum_{t=1}^T \frac{105n_t^3}{(n_t+6)(n_t+4)(n_t+2)}$$

and the following multivariate CLT holds:

$$\sqrt{T} \begin{pmatrix} \bar{M}_1 \\ \bar{M}_2 - M_2 \\ \bar{M}_3 \\ \bar{M}_4 - M_4 \end{pmatrix} \xrightarrow{d} N \left( 0, \begin{bmatrix} M_2 & 0 & M_4 & 0 \\ 0 & M_4 - M_2^2 & 0 & M_6 - M_2 M_4 \\ M_4 & 0 & M_6 & 0 \\ 0 & M_6 - M_2 M_4 & 0 & M_8 - M_4^2 \end{bmatrix} \right)$$

**Proof.** Available upon request.

**Corollary 1.** The following chi-square test statistics provide different moment-based tests of the null hypothesis that the sample  $\bar{R}_t$ ,  $t = 1, 2, \dots, T$  of independent observations comes from a distribution with the moments specified above:

$$T\bar{M}_1^2 \sim \chi^2(1), T\frac{(\bar{M}_2-1)^2}{M_4-1} \sim \chi^2(1), T\frac{\bar{M}_3^2}{M_6} \sim \chi^2(1), T\frac{(\bar{M}_4-M_4)^2}{M_8-M_4^2} \sim \chi^2(1)$$

$$T[\bar{M}_1^2 + \frac{(\bar{M}_2-1)^2}{M_4-1}] \sim \chi^2(2), T[\bar{M}_1^2 + \frac{(\bar{M}_4-M_4)^2}{M_8-M_4^2}] \sim \chi^2(2), T[\frac{(\bar{M}_2-1)^2}{M_4-1} + \frac{\bar{M}_3^2}{M_6}] \sim \chi^2(2), T[\frac{\bar{M}_3^2}{M_6} + \frac{(\bar{M}_4-M_4)^2}{M_8-M_4^2}] \sim \chi^2(2)$$

$$T\left(\begin{matrix} \bar{M}_1 & \bar{M}_2 - 1 & \bar{M}_3 & \bar{M}_4 - M_4 \end{matrix}\right) \left[ \begin{matrix} 1 & 0 & M_4 & 0 \\ 0 & M_4 - 1 & 0 & M_6 - M_4 \\ M_4 & 0 & M_6 & 0 \\ 0 & M_6 - M_4 & 0 & M_8 - M_4^2 \end{matrix} \right]^{-1} \left( \begin{matrix} \bar{M}_1 \\ \bar{M}_2 - 1 \\ \bar{M}_3 \\ \bar{M}_4 - M_4 \end{matrix} \right) \sim \chi^2(4)$$

**Corollary 2.** As  $n_t$  gets large, uniformly for all  $t = 1, 2, \dots, T$ , the moment-based tests become asymptotically equivalent to those for the special limiting case of i.i.d. normal  $\bar{R}_t$ :

$$T\bar{M}_1^2 \sim \chi^2(1), T\frac{(\bar{M}_2-1)^2}{2} \sim \chi^2(1), T\frac{\bar{M}_3^2}{15} \sim \chi^2(1), T\frac{(\bar{M}_4-M_4)^2}{96} \sim \chi^2(1)$$

$$T[\bar{M}_1^2 + \frac{(\bar{M}_2-1)^2}{2}] \sim \chi^2(2), T[\bar{M}_1^2 + \frac{(\bar{M}_4-M_4)^2}{96}] \sim \chi^2(2), T[\frac{(\bar{M}_2-1)^2}{2} + \frac{\bar{M}_3^2}{15}] \sim \chi^2(2), T[\frac{\bar{M}_3^2}{15} + \frac{(\bar{M}_4-M_4)^2}{96}] \sim \chi^2(2)$$

$$T\left(\begin{matrix} \bar{M}_1 & \bar{M}_2 - 1 & \bar{M}_3 & \bar{M}_4 - 3 \end{matrix}\right) \left[ \begin{matrix} 1 & 0 & 3 & 0 \\ 0 & 2 & 0 & 12 \\ 3 & 0 & 15 & 0 \\ 0 & 12 & 0 & 96 \end{matrix} \right]^{-1} \left( \begin{matrix} \bar{M}_1 \\ \bar{M}_2 - 1 \\ \bar{M}_3 \\ \bar{M}_4 - 3 \end{matrix} \right) \sim \chi^2(4)$$

## APPENDIX II. Jump Detection Procedure

A common implicit feature of a variety of jump detection techniques recently considered in ABD(2005), BN-S (2005), and Huang and Tauchen (2005) is that jumps are closely associated with the largest few outliers among the intraday returns on days that reveal statistically significant discrepancy between  $RV_{t+1}(\Delta)$  and  $BV_{t+1}(\Delta)$ . Identifying multiple jumps with these procedures involves some kind of recursion on the largest remaining intraday outliers until the discrepancy between  $RV_{t+1}(\Delta)$  and  $BV_{t+1}(\Delta)$  drops below a chosen daily significance level  $\alpha$ .

Here we implement a jump detection procedure that also looks into the largest intraday outliers on a given day but avoids the need for recursion by introducing a uniform decision rule that identifies simultaneously all statistically significant jump outliers on that day. The rule is driven exclusively by  $BV_{t+1}(\Delta)$ , which annihilates the contribution of jumps and only measures the diffusive volatility component. Hence, under the null hypothesis of no jumps and ignoring for a moment the potentially large intraday variations of volatility, the distribution of each intraday return

$r_{t+j\Delta, \Delta}$ ,  $j = 1, 2, \dots, \frac{1}{\Delta}$  may be approximated by  $N(0, \Delta \cdot BV_{t+1}(\Delta))$ . This approximation implies

that multiple jumps can be identified simultaneously as intraday returns that fall outside of a sufficiently large confidence interval of  $N(0, \Delta \cdot BV_{t+1}(\Delta))$ . Thus, the resulting jump detection procedure can be described as follows:

1. Choose the size  $\alpha$  of the jump test at the daily level and set  $\beta = (1-\alpha)^\Delta - 1$  to be the level of the corresponding  $(1-\beta)$  confidence interval for the intraday diffusive returns distributed approximately  $N(0, \Delta \cdot BV_{t+1}(\Delta))$

2. Detect possibly multiple intraday jumps  $\kappa_s(\Delta)$  based on the rule:

$$\kappa_s(\Delta) = r_{t+s\Delta, \Delta} \cdot \mathbf{I}[ |r_{t+s\Delta, \Delta}| > \Phi_{1-\beta/2} \cdot \sqrt{\Delta \cdot BV_{t+1}(\Delta)} ], s = 1, 2, \dots, \frac{1}{\Delta}, \text{ where } \Phi \text{ is the } N(0,1) \text{ cdf}$$

What makes the procedure sensible even in the case of a somewhat variable intraday volatility is a very conservative choice of  $\alpha$  and hence  $\beta$ . Table I below documents quite satisfactory performance for  $\alpha=10^{-5}$  for the various jump specifications that we simulated, both with and without leverage, hence we use the same daily test size  $\alpha=10^{-5}$  on the S&P 500 returns as well.

## APPENDIX III. Data Description

<i>Data Type</i>	<i>Set #</i>	<i>Size (days)</i>	<i>Description</i>
Simulated	1	5,000,000	No leverage, no jumps
Simulated	2	5,000,000	No leverage, large rare jumps
Simulated	3	5,000,000	No leverage, frequent small jumps
Simulated	4	5,000,000	No leverage, moderate jumps
Simulated	5	5,000,000	Leverage, no jumps
Simulated	6	5,000,000	Leverage, large rare jumps
Simulated	7	5,000,000	Leverage, frequent small jumps
Simulated	8	5,000,000	Leverage, moderate jumps
Real Data	1	4,126	S&P 500 most liquid futures from 1/1/1988 to 7/26/2004
Real Data	2	3,420	S&P 500 most liquid futures from 1/1/1988 to 8/31/2001

### 1. Simulated data.

We generate samples from eight specifications of a one-factor jump-diffusion model with Poisson jumps of constant intensity  $\lambda$  and lognormal jump size  $\kappa$  with variance  $\vartheta^2$ :

$$dp(t) = \sigma(t)dW_1(t) + \kappa dq_t,$$

$$d\sigma^2(t) = \eta(\theta - \sigma^2(t))dt + \nu\sigma(t)dW_2(t)$$

The diffusive volatility parameters are set to  $\theta = 1$ ,  $\eta = 0.01$ ,  $\nu = 0.1$ . We simulate four jump scenarios first without leverage and then with leverage by setting  $corr(dW_1(t), dW_2(t)) = -0.5$ :

<i>Jump scenario</i>	<i>Jump parameters</i>	<i>Jump contribution to total volatility</i>
No jumps	$\vartheta^2 = 0, \lambda = 0$	0 %
Large rare jumps	$\vartheta^2 = 2.5, \lambda = 0.1$	20%
Small frequent jumps	$\vartheta^2 = 0.25, \lambda = 1.0$	20%
Moderate jumps	$\vartheta^2 = 0.5, \lambda = 0.2$	9.1%

The simulations are based on an Euler scheme, the simulation frequency is 1 second, while the sampling frequency is 120 seconds with a total of 195 sample returns per day. We simulate 5,000,000 days providing 1,000 high-frequency data samples for 5,000 days of data each, which is about the largest sample size encountered with real data. Thanks to this special design we are able to provide a fair Monte Carlo assesment of the machinery used for real data analysis.

## **2. Real data.**

We use the R & C Research tick transaction data for the S&P 500 stock index futures traded on the Chicago Merchantile Exchange. The two periods we consider are from January 1, 1988 to July 26, 2004 and from January 1, 1988 to August 31, 2001. The latter avoids the highly volatile post 9/11 period and is also considered by Peters and de Vilder (2004).

We construct two minute returns by the previous tick method from the recorded transaction prices between 9:30 and 16:00 CET for the most liquid futures contract on a given day (with the shortest maturity above five business days). We exclude days beginning or ending with ten or more zero two-minute returns to eliminate shorter trading days around major holidays. After excluding 44 such days from the 1988-2004 sample we are left with 4,126 trading days, while the remaining trading days in the 1988-2001 sample are 3,420 (after filtering out 34 days).

## **3. Sampling procedure in feasible financial time**

Given a sample of intraday returns we measure feasible financial time in units of the average daily realized volatility for the sample. Then we sample in periods of length 0.5, 1, 1.5, 2, and 2.5 units of financial time by the following procedure:

1. Set the financial time clock to zero;
2. Increment the financial time clock by the square of the next intraday return in the sample;
3. Repeat step 2 until the clock reaches/exceeds the chosen period length
4. Sample, increment the period count, and return to step 1

The sample points obtained following this procedure define "days" in financial time. However, unlike in physical time, the days in financial time have different numbers of intraday returns and almost the same volatility level unless diffusive return outliers or undetected large jumps induce substantial spikes in the measurement of financial time.