Identifying Jumps in Financial Assets: a Comparison between Nonparametric Jump Tests

First version: April 20, 2010

This version: October 14, 2011

Ana -Maria DUMITRU

School of Economics, Faculty of Business, Economics and Law, University of Surrey

Guildford, Surrey, GU2 7XH (UK). E-mail a.dumitru@surrey.ac.uk &

Department of Economics and Technology Management, University of Bergamo (Italy)

Giovanni URGA

Centre for Econometric Analysis, Faculty of Finance, Cass Business School, 106 Bunhill Row,

London EC1Y 8TZ (UK). E-mail: g.urga@city.ac.uk &

Hyman P. Minsky Department of Economic Studies, University of Bergamo (Italy)

Abstract

We perform a comprehensive Monte Carlo comparison between nine alternative procedures available in the literature to detect jumps in financial assets using high-frequency data. We evaluate size and power properties of the procedures under alternative sampling frequencies, persistence in volatility, jump size and intensity, and degree of contamination with microstructure noise. The overall best performance is showed by the Andersen et al. (2007) and Lee and Mykland (2008) intraday procedures (ABD-LM), provided the price process is not very volatile. We propose two extensions to the existing battery of tests. The first regards the finite sample improvements based on simulated critical values for the ABD-LM procedure. The second regards a procedure that combines frequencies and tests able to reduce the number of spurious jumps. Finally, we report an empirical analysis using real high frequency data on five stocks listed in the New York Stock Exchange.

Keywords: jumps, nonparametric tests, high frequency data, stochastic volatility, Monte Carlo simulations

1 INTRODUCTION

There is a large consensus in the financial literature, theoretical and applied, that modeling return dynamics requires the specification of a stochastic volatility component, which accommodates the persistence in volatility, and of a jump component, which takes care of the unpredictable, large movements in the price process. The identification of the time and the size of jumps has profound implications in risk management, portfolio allocation, derivatives pricing (Aït-Sahalia, 2004).

One of the main advances in high frequency econometrics over the last decade was the development of nonparametric procedures to test for the presence of jumps in the path of a price process during a certain time interval or at certain point in time. Such methods are very simple to apply, they just require high frequency transaction prices or mid-quotes. Moreover, they are developed in a model free framework, incorporating different classes of stochastic volatility models. In addition to the seminal contribution of Barndorff-Nielsen and Shephard (2006), in this paper we consider eight other tests proposed by Andersen et al. (2007), Lee and Mykland (2008), Aït-Sahalia and Jacod (2008), Jiang and Oomen (2008), Andersen et al. (2009) [two tests based on the minimum and median realized variance], Corsi et al. (2010) and Podolskij and Ziggel (2010). All tests rely on CLT-type results that require an intraday sampling frequency that tends to infinity. The test statistics are based on robust to jumps measures of variation in the price processes which are estimated by using one of the following types of estimators: realized multi-power variations (Barndorff-Nielsen et al., 2006), threshold estimators (Mancini, 2009), the median or the minimum realized variation (Andersen et al., 2009), the corrected realized threshold multipower variation (Corsi et al., 2010).

Given such a variety of nonparametric methodologies to identify jumps, one might wonder which procedure should be preferred, or whether there are data characteristics for which it is recommended to use one test instead of the others. The main objective of this paper is to perform a thorough comparison among the various testing procedures, based on a comprehensive set of Monte Carlo simulations, which embody the main features of financial data. It is important to establish whether the performance of the tests is related to some features of the data, such as different sampling frequencies, different levels of volatility, varying persistence in volatility, varying contamination with microstructure noise, varying jump size and jump intensity. Based on the findings of the simulation exercise, we aim to provide guidelines to users of nonparametric tests for jumps.

To the best of our knowledge, there are two other papers that deal with similar objectives. Theodosiou and Žikeš (2010) perform an extensive Monte Carlo simulation exercise to evaluate the performance of different jump detection procedures, with a special interest in the effect of illiquid data on the behaviour of the various tests. Schwert (2009) instead relies only on real data to conclude that different procedures pick up different sets of jumps. Our paper is more comprehensive in terms of testing procedures included in our comparison. Moreover, we make two additional contributions to the existing literature. First, we propose using simulated critical values for the Andersen et al. (2007) and Lee and Mykland (2008) tests; second, and most importantly, we show that combining various procedures greatly improves the performance of the tests in terms of spurious jump detection. In addition, the tests are applied to five selected stocks listed in the New York Stock Exchange. We find that both the percentage of detected jumps, as well as the estimated quadratic variation of the jump process, vary considerably with the employed testing procedure, which advocates for the use of combinations of tests and frequencies.

The paper is organized as follows. In Section 2, we review the nine nonparametric tests for jumps available in the literature. Section 3 describes the Monte Carlo setup and reports the main findings of the simulations. Section 4 describes two extensions: the finite sample approximations for the intraday procedures and alternative procedures that combine tests and frequencies. Section 5 reports an empirical exercise using stock data. Section 6 concludes.

2 JUMP TESTS

In this section, we describe the available jump detection procedures. The dynamics of the logarithmic price process, p_t, is usually assumed to be a jump-diffusion process of the form:

$$\mathrm{d}p_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t + \mathrm{d}J_t \tag{1}$$

where μ_t represents the drift, σ_t the diffusion parameter, and W_t a Brownian motion at time t. J_t is the jump process at time t, defined as $J_t = \sum_{j=1}^{N_t} c_{t_j}$ where c_{t_j} represents the size of the jump at time t_j and N_t is a counting process, representing the number of jumps up to time t.

The quadratic variation of the price process up to a certain point in time t (QV_t), usually a trading day, can be defined as follow:

$$QV_t = \int_0^t \sigma_s^2 \mathrm{d}s + \sum_{j=1}^{N_t} c_{t_j}^2, \tag{2}$$

where $\int_0^t \sigma_s^2 \mathrm{d}s = IV_t$ is the integrated variance or volatility. Thus, QV_t is made up of a part coming from the diffusion component and another one caused by the jump component. The two components have a different nature and should be separately analyzed and modelled. The integrated volatility is characterized by persistence, whereas jumps, apart from a possible drift, have an unpredictable nature.

The recent literature in the field of high frequency econometrics has developed several estimators for both the quadratic variance and the integrated volatility of the price. Most of these estimators are based on equally spaced data. Thus, the interval $[0, t]$ is split into n equal subintervals of length δ. The j-th intraday return r_j on day t is defined as $r_j = p_{t-1+j\delta} - p_{t-1+(j-1)\delta}$.

QV_t can be estimated by the realized variance (RV_t), defined as (Andersen and Bollerslev, 1998):

$$RV_t = \sum_{j=1}^{n} r_j^2 \xrightarrow{p} QV_t, \qquad \text{for} \quad \delta \to 0 \tag{3}$$

where $\xrightarrow{p}$ stands for convergence in probability.

To measure IV_t one can use a wide range of estimators, such as multipower variations, threshold estimators, medium and minimum realized variance. All these quantities are robust to jumps in the limit. Most of the jump detection procedures are based on the comparison between RV_t, which captures the variation of the process generated by both the diffusion and the jump parts, and a robust to jumps estimator.

It is important to note that none of these procedures can test for the absence or presence of jumps in the model or data generating process. They merely supply us with information on whether within a certain time interval or at a certain moment, the realization of the process is continuous or not. Andersen et al. (2007) and Lee and Mykland (2008) assume the null of continuity of the sample path at time t_j; for the other procedures, the null is of continuity of the sample path during a certain period, such as a trading day. For all tests, the alternative hypothesis implies discontinuity of the sample path, that is the occurrence of at least one jump.

We turn now to a short presentation of the procedures and we refer to the original papers for details.

2.1 Barndorff-Nielsen and Shephard (2006) test (BNS)

Barndorff-Nielsen and Shephard (2004) propose the first robust to jumps estimator of the integrated variance, the realized bipower variation (BV_t), constructed to reduce the impact of jump returns on the volatility estimate by multiplying them with adjacent jump-free returns:

$$BV_t = 1.57 \sum_{j=2}^{n} |r_j||r_{j-1}| \tag{4}$$

The BNS test infers whether jumps occur during a time interval (usually a trading day) by comparing RV_t, as an estimator for the quadratic variance, with BV_t. Following simulation studies reported in BNS and in Huang and Tauchen (2005), in this paper we use the ratio test defined as:

$$\frac{1 - \frac{BV_t}{RV_t}}{\sqrt{0.61\,\delta\,\max\left(1, \frac{TQ_t}{BV_t^2}\right)}} \xrightarrow{L} \mathcal{N}(0,1) \tag{5}$$

where $\xrightarrow{L}$ stands for convergence in law. TQ_t represents the realized tripower quarticity that consistently estimates the integrated quarticity, i.e. $\int_0^t \sigma_u^4\,\mathrm{d}u$, and is defined as follows:

$$TQ_t = n\,1.74\left(\frac{n}{n-2}\right)\sum_{j=3}^{n} |r_{j-2}|^{4/3}|r_{j-1}|^{4/3}|r_j|^{4/3}. \tag{6}$$

2.2 Andersen et al. (2007) and Lee and Mykland (2008) tests (ABD-LM)

The ABD and LM procedures test for jumps by comparing standardized intraday returns to a threshold. Both tests are constructed under the null that there is no jump in the realization of the process at a certain time, t_j. This enables users to identify the exact time of a jump, as well as the number of jumps within a trading day. We call these two procedures "intraday" tests.

The first step in applying both ABD and LM procedures is to compute a robust to jumps local (spot) volatility estimate at time t_j, $\hat{V}_j$, and then standardize the intraday returns as follows:

$$z_j = |r_j|/\sqrt{\hat{V}_j}, \tag{7}$$

where $\hat{V}_j = BV_{t_j}/(K-2)$, with K the window size on which BV_{t_j} is calculated.

Given that z_j is asymptotically normal, one can identify jumps by comparing z_j with a normal threshold (see ABD). This comparison must be performed for each of the n intraday returns to decide whether jumps occurred during a trading day or not. This raises the issue of multiple testing, well-documented the statistical literature. ABD propose to use the Šidák approach to deal with the false discovery rate problem which may arise in this context. Once a nominal daily size, α, is fixed, the corresponding size for each intraday test is defined as $\beta = 1 - (1-\alpha)^{\delta}$. We reject the null of continuity of the sample path if $z_j > \Phi_{1-\beta/2}$, where Φ is the standard normal cdf.

LM use a slightly different approach. The usual 95% and 99% quantiles from the normal distribution prove too permissive, leading to an over-rejection of the null. To overcome this limitation, the authors propose using critical values from the limit distribution of the maximum of the test statistics. They show that this maximum converges, for $\delta \to 0$, to a Gumbel variable:

$$(\max(z_j) - C_n)/S_n \xrightarrow{L} \xi, \qquad \mathbf{P}(\xi) = \exp(-e^{-x}) \tag{8}$$

where $C_n = \frac{(2\log n)^{1/2}}{0.8} - \frac{\log\pi + \log(\log n)}{1.6\,(2\log n)^{1/2}}$ and $S_n = \frac{1}{0.8\,(2\log n)^{1/2}}$.

The test can be conducted by comparing z_j, standardized as $\max(z_j)$ in (8), to the critical value from the Gumbel distribution. LM propose computing $\hat{\sigma}_j$ on a window size of K observations that precede time t_j. They show that K depends on the choice of the sampling frequency and suggest to take $K = \sqrt{252 * n}$, where n is the daily number of observations, whereas 252 is the number of days in the (financial) year.

The ABD test requires very low nominal sizes (10^{-5}) with respect to all other procedures, which use a 5% significance level. Because of this and because the ABD and LM tests differ only in terms of the choice of the critical values, in this paper we do not distinguish between the two procedures, and we report the results under the acronym 'ABD-LM' based on the critical values of LM.

Whenever we make comparisons with the other tests which are applied on time intervals equal to one trading day, we compute the 'ABD-LM' test statistics for every time t_j within a trading day and then pick up the maximum statistic as the final test for that day.

2.3 The Aït-Sahalia and Jacod (2008) test (AJ)

AJ base their procedure on the following argument. Let $B(m,\delta)_t = \sum_{j=1}^{n} |r_j|^m$ be the realized power variation for data sampled every δ observations. If one computes the above estimator on two different time scales, δ and $k\delta$, the ratio between the two is proportional in the limit with some power of the ratio of the time scales, provided no jumps are present:

$$S(\widehat{m,k,}\delta)_t = \frac{B(m,k\delta)_t}{B(m,\delta)_t} \xrightarrow{p} k^{m/2-1}, \qquad \text{for} \quad \delta \to 0 \tag{9}$$

where $B(m,\delta)_t$ and $B(m,k\delta)_t$ are the realized power variations of order $m > 2$, computed on time scales δ and $k\delta$, $k \in \mathbb{N}$, $k \geq 2$. The realized power variation for data sampled every δ observations is defined as: $B(m,\delta)_t = \sum_{j=1}^{n} |r_j|^m$.

The following test statistic is proposed to test for the null of no jumps:

$$\frac{\widehat{S(m,k,\delta)_t} - k^{m/2-1}}{\sqrt{M(m,k)\frac{A(2m)_t}{A(m)_t}}} \xrightarrow{L} \mathcal{N}(0,1), \tag{10}$$

where $A(p) = \int_o^t |\sigma_s|^p \mathrm{d}s$, $p = m,\ 2m$ and $M(m,k) = \frac{1}{\mu_m^2}\left(k^{m-2}(1+k)\mu_{2m} + k^{m-2}(k-1)\mu_m^2 - 2k^{m/2-1}\mu_{k,m}\right)$, with $\mu_m = E|U|^m$, $\mu_{2m} = E|U|^{2m}$, and $\mu_{k,m} = E(|U|^m|U + \sqrt{k-1}V)|$, for U and V independent standard normals. To estimate $A(p)$, one can use either multipower variations, defined as $\frac{n^{\frac{qr}{2}-1}}{\mu_r^q}\sum_{j=q+1}^n \prod_{i=1}^q |r_{j-i+1}|^r \xrightarrow{p} A(qr)$, for $\delta \to 0$ and $p = qr$, or threshold estimators, i.e. $\frac{n^{\frac{p}{2}-1}}{\mu_p}\sum_{j=1}^n |r_j|^p \mathbb{I}_{\{|r_j| \le c\delta^w\}} \xrightarrow{p} A(p)$, where $1_{\{|r_j| \le c*\delta^w\}}$ is an indicator function for absolute returns lower than a threshold fixed to $c \cdot \delta^w$, while μ_p and μ_r are absolute moments of orders p and r of a standard normal. In this paper, we employ both approaches.

2.4 Jiang and Oomen (2008) test (JO)

The JO test exploits the difference between arithmetic and logarithmic returns:

$$SwV_t(\delta) = 2\sum_{j=1}^{[t/\delta]} (R_j - r_j) \tag{11}$$

where R_j denotes the j-th arithmetic intraday return, while r_j is the log return. The absence of jumps makes the difference between SwV_t and the realized variance equal to 0:

$$\operatorname*{plim}_{\delta \to 0}(SwV_t - RV_t) = \begin{cases} 0 & \text{no jumps in}[0,t] \\ 2\int_0^t \underline{J}_u \, \mathrm{d}q_u - \int_0^t J_u^2 \, \mathrm{d}q_u & \text{jumps in}[0,t] \end{cases} \tag{12}$$

where $\underline{J}_u = exp(J_u) - J_u - 1$, with J the jump process. Under the alternative, in the limit, the difference $SwV_t - RV_t$ captures jumps in exponential form. Thus, the test statistic becomes very large in the presence of large returns, enabling jump identification. The test statistic is defined as:

$$\frac{nBV_t}{\sqrt{\Omega_{SwV}}}\left(1 - \frac{RV_t}{SwV_t}\right) \xrightarrow{L} \mathcal{N}(0,1). \tag{13}$$

Ω_{SwV} can be estimated as $\hat{\Omega}_{SwV} = 3.05\,\frac{n^3}{n-3}\sum_{i=0}^{n-4}\prod_{k=1}^4 |r_{i+k}|^{6/4}$.

2.5 Andersen et al. (2009) tests based on MinRV and MedRV tests(Min and Med)

Andersen et al. (2009) propose to estimate integrated volatility in the presence of jumps based on the nearest neighbour truncation. The minimum realized variance ($MinRV_t$) and median realized variance

($MedRV_t$) eliminate jumps by taking respectively the minimum and the median over adjacent returns:

$$MinRV_t = 2.75 \, \tfrac{n}{n-1} \textstyle\sum_{j=2}^{n} \min(|r_j|, |r_{j-1}|)^2$$
$$MedRV_t = 1.42 \, \tfrac{n}{n-2} \textstyle\sum_{j=3}^{n} \operatorname{med}(|r_j|, |r_{j-1}|, |r_{j-2}|)^2. \tag{14}$$

The Min and Med tests are based on the same argument as the BNS procedure, i.e. the comparison between a robust to jumps estimator and RV_t:

$$\frac{1 - \frac{MinRV_t}{RV_t}}{\sqrt{1.81 \, \delta \, \max\left(1, \frac{MinRQ_t}{MinRV_t^2}\right)}} \xrightarrow{L} \mathcal{N}(0,1) \quad \text{and} \quad \frac{1 - \frac{MedRV_t}{RV_t}}{\sqrt{0.96 \, \delta \, \max\left(1, \frac{MedRQ_t}{MedRV_t^2}\right)}} \xrightarrow{L} \mathcal{N}(0,1), \tag{15}$$

where $MinRQ_t = 2.21 \frac{n^2}{n-1} \sum_{j=2}^{n} \min(|r_j|, |r_{j-1}|)^4$ is the minimum realized quarticity and $MedRQ_t = 0.92 \frac{n^2}{n-2} \sum_{j=3}^{n} \operatorname{med}(|r_j|, |r_{j-1}|, |r_{j-2}|)^4$ the median realized quarticity which estimate the integrated quarticity.

2.6 Corsi et al. (2010) test (CPR)

The CPR test is also based on a comparison between RV_t and a robust to jumps estimator. However, the procedure employs the corrected realized threshold bipower variation, as an alternative to the BV_t. This new estimator discards jumps given that it is built as a bipower variation and truncates returns over a certain threshold. The following test statistic is employed:

$$\frac{1 - \frac{C-TBV_t}{RV_t}}{\sqrt{0.61 \, \delta \, \max\left(1, \frac{C-TTriPV_t}{C-TBV_t^2}\right)}} \xrightarrow{L} \mathcal{N}(0,1), \tag{16}$$

where $C-TBV_t$ and $C-TTriPV_t$ represent the corrected realized threshold bipower and tripower variation, respectively, defined as:

$$C - TBV_t = 1.57 \, \textstyle\sum_{j=2}^{n} Z1(r_j, \vartheta_j) Z1(r_{j-1}, \vartheta_{j-1}),$$
$$C - TTriPV_t = 1.74 \, \textstyle\sum_{j=3}^{n} Z1(r_j, \vartheta_j) Z1(r_{j-1}, \vartheta_{j-1}) Z1(r_{j-2}, \vartheta_{j-2}) \tag{17}$$

where $Z1(r_j, \vartheta_j) = \begin{cases} |r_j|, & r_j^2 < \vartheta_j \\ 1.094 \, \vartheta_j^{\frac{1}{2}}, & r_j^2 > \vartheta_j \end{cases}$ is a function of the return at time t_j and a threshold $\vartheta_j = c_\vartheta^2 \cdot \hat{V}_j$. c_ϑ^2 is a scale free constant and $\hat{V}_j$ a local volatility estimator.

Following authors' recommendation, to compute the threshold, ϑ_j, we take $c_\vartheta = 3$. For the auxiliary local volatility estimate, $\hat{V}_j$, we employ the non-parametric filter proposed by CPR that removes jumps from data in several iterations.

2.7 Podolskij and Ziggel (2010) test (PZ)

The PZ procedure is based on a modified version of Mancini (2009)'s threshold estimator as an alternative to BV_t in testing for jumps. However, in order to derive a limiting theory, authors define the test statistics as a difference between a realized power variation estimator and a threshold estimator perturbed by some external positive i.i.d. random variables, η_j, with $E[\eta_j] = 1$ and finite variance, $Var[\eta_j]$:

$$T(m,\delta)_t = n^{\frac{m-1}{2}} \sum_{j=1}^{n} |r_j|^m (1 - \eta_j \mathbb{I}_{\{|r_j| \leq c\delta^w\}}), \qquad m \geq 2, \tag{18}$$

where $1_{\{|r_j| \leq c*\delta^w\}}$ is an indicator function for absolute returns lower than a threshold fixed to $c \cdot \delta^w$, with $c = 2.3\sqrt{BV_t}$ and $w = .4$. The test statistic is defined as follows:

$$\frac{T(m,\delta)_t}{\sqrt{Var[\eta_j] n^{\frac{2m}{2}-1} \sum_{j=1}^{[t/\delta]} |r_j|^{2m} \mathbb{I}_{\{|r_j| \leq c\delta^w\}}}} \xrightarrow{L} \mathcal{N}(0,1), \tag{19}$$

PZ recommend to sample η_j from the distribution $P^\eta = \frac{1}{2}(\varsigma_{1-\tau} + \varsigma_{1+\tau})$, where ς is the Dirac measure, and τ is a constant chosen relatively small, e.g. $\tau = 0.1$ or 0.05.

3 MONTE CARLO ANALYSIS

In this section, we report the results of an extensive comparison among the testing procedures presented in the previous section. The exercise is based on a comprehensive set of Monte Carlo simulations, which embody several features of financial data. To quantify the size for all tests, our simulations are based on stochastic volatility models with varying persistence. To evaluate their power property, we consider stochastic volatility models with jumps of different sizes arriving with varying intensity.

3.1 Simulation design

Following Huang and Tauchen (2005), we simulated several stochastic volatility processes with leverage effect, different levels of persistence in volatility, to which we add jumps arriving with different intensities and variances.

The benchmark model for our simulations is a stochastic volatility model with one volatility factor (SV1F). The volatility factor enters the price equation in an exponential form, as suggested in Chernov et al. (2003):

$$\begin{aligned} \mathrm{d}p_t &= 0.03\mathrm{d}t + \exp[0.125\,\upsilon_t]\mathrm{d}W_{p_t}, \\ \mathrm{d}\upsilon_t &= \alpha_\upsilon \upsilon_t \mathrm{d}t + \mathrm{d}W_{\upsilon_t}, \qquad \mathrm{corr}(\mathrm{d}W_p, \mathrm{d}W_\upsilon) = -0.62 \end{aligned} \tag{20}$$

where p_t is the log-price process, the W's are standard Brownian motions, υ_t the volatility factor, α_υ the mean reversion parameter of the volatility process with values in the following set: {-0.137e^{-2}, -0.100, -1.386}. This is the process that we simulate under the null hypothesis of no jumps.

Chernov et al. (2003) show that it is possible to generate dynamics similar to jumps by using a two factor stochastic volatility model. A first volatility factor controls for the persistence in the volatility process, while the second factor generates higher tails in a similar manner to a jump process. Thus, a second stochastic volatility model (SV2F) is defined as:

$$\begin{aligned} \mathrm{d}p_t &= 0.03\mathrm{d}t + \exp[-1.20 + 0.04\upsilon_{1_t} + 1.50\upsilon_{2_t}]\mathrm{d}W_{p_t} \\ \mathrm{d}\upsilon_{1_t} &= (-0.137e^{-2})\,\upsilon_{1_t}\mathrm{d}t + \mathrm{d}W_{\upsilon_{1_t}} \\ \mathrm{d}\upsilon_{2_t} &= -1.386\,\upsilon_{2_t}\mathrm{d}t + [1 + 0.25\,\upsilon_{2_t}]\mathrm{d}W_{\upsilon_{2_t}} \end{aligned} \tag{21}$$

with $\mathrm{corr}(\mathrm{d}W_p, \mathrm{d}W_{\upsilon_1}) = -0.30$ and $\mathrm{corr}(\mathrm{d}W_p, \mathrm{d}W_{\upsilon_2}) = -0.30$. The values for the coefficients for both SV1F and SV2F models are as in Huang and Tauchen (2005).

SV2F can generate extreme returns, without having a jump component. We simulate this model only under the null hypothesis. Our objective is to understand whether the various tests for jumps maintain a reasonable size in extremely volatile periods.

To assess the power of the tests, we augment SV1F with rare compound Poisson jumps, arriving with intensity λ between 0 and 2 and having normally distributed sizes with mean 0 and standard deviation σ_{jump} that ranges between 0 and 2.5.

In general, empirical works apply these tests at a daily level, in order to be able to conclude whether jumps occurred during the trading day. Therefore, we evaluate the statistical properties of all jump tests based on data simulated for 10,000 trading days, for all models and under both hypotheses of continuity and discontinuity. For the simulation of each path, we use an Euler discretization scheme based on increments of 1 second. We then perform a sampling at 1, 5, 15 and 30 minutes.

We report results using a 5% significance level. The results for alternative significance levels, such as 1%, 0.1% and 0.01%, are in line with the ones at 5%. We report size and size adjusted power. The latter is computed by centering the test statistic under the alternative with the average of the statistic under the null.

3.2 Monte Carlo results

3.2.1 Size and power of the tests for stochastic volatility models

SIZE For SV1F, we consider three alternative values of the mean reversion parameter of the volatility factor. In all cases, the empirical size tends to slightly decrease with the increase in the mean reversion parameter, without affecting the ranking of the tests. In Table 1, we only report the empirical size for the medium mean reversion case ($\alpha_v = -0.100$). The results for values of $\alpha_v = -0.137e^{-2}, \ -1.386$ are available upon request.

[Insert Table 1 here]

The biggest size distortion is encountered in the case of the JO test, where, for a 1 second sampling frequency, we have a size of 6.5%, which increases when the sampling frequency diminishes. As already mentioned, the JO test statistic captures jumps in exponential form in the limit and thus, gets very large in the presence of high returns, leading to an over-rejection of the null. A similar pattern can be seen for the PZ procedure, which displays a size close to the nominal one when sampling is performed every second, but then gets rapidly and highly oversized. This may be due to that the magnitude of the threshold estimator in the PZ test is reduced under the influence of two simultaneous effects: the truncation of returns and the dependence of the threshold on the realized bipower variation, which smooths volatility out and lowers the threshold. Consequently, the test statistic becomes large, leading to over-rejection.

The BNS, Med, Min and CPR behave very similarly. The best performance is shown by the Med and BNS tests. Both tests display a size very close to the nominal one at a sampling frequency of 1 second, i.e. 5.1% for the Med and 4.8% for BNS. The size tends to slowly increase with the decrease in the sampling frequencies. The Min test behaves well at 1 second with a size of 4.7%, but has a tendency to become undersized at lower frequencies, getting to 3.5% at 30 minutes. The ranking between the BNS, Med and Min tests can be explained by the difference in the efficiency of the integrated variance estimators that these tests employ. Both BPV_t and $MedRV_t$ display smaller asymptotic variances than $MinRV_t$. The CPR has a size close to the nominal one at 1 second, but then becomes oversized with the decrease in the sampling frequency and displays a size equal to 7.5% for 30 minutes data. For this procedure, the magnitude of the robust to jumps estimator ($C-TBV_t$) is reduced by that the estimator is built as a multipower variation and because of the truncation of returns. The intraday ABD-LM procedure tends to be oversized at all sampling frequencies. Its size distortion is not very high though, varying around 1-1.5% from the nominal size. This may be due to that the realized bipower variation, used to standardize returns, tends to smooth out the local volatility, resulting in a higher test statistic. The AJ test statistic was standardized with standard deviations based on both power variations and threshold estimators. In both cases, at a sampling frequency of 1 second, the test seems slightly undersized. However, when diminishing the sampling frequency, the behavior of the test statistics differs. The test becomes rapidly oversized when its variance is based on realized power variations and severely undersized when threshold estimators are used to estimate its variance. This test too seems to work well at higher frequencies.

Table 2 reports the empirical size for the SV2F model. If we look at all sampling frequencies, the best performance is displayed by the Min test, followed by BNS. For 1 second sampling frequency, size is equal to 5.2% and 5.4%, which increases at lower sampling frequencies though less dramatically than the other tests. The Med, CPR and JO tests behave similarly to BNS and Min, but become more rapidly oversized. The AJ(power var) has a size close to the nominal one when sampling is done every second, but then becomes rapidly oversized. When the AJ(threshold) is considered, the test gets severely undersized at lower sampling frequencies. The PZ and the intraday procedures display by far the poorest performance, being severely oversized even when we sample every second (99.3% for the intraday tests and 70.1% for PZ).

[Insert Table 2 here]

POWER We now evaluate the power of the tests by adding to the continuous stochastic volatility process SV1F jump processes with alternative intensities and jump sizes.

Varying jump intensity In order to examine how jump detection changes as the number of jumps grows, we consider Poisson jump arrival times depending on the following varying jump intensities (λ): .014, .058, .089, .118, .5, 1, 1.5, and 2. These intensities can be interpreted as the average number of jumps per day and generate the following total number of jumps: 148, 560, 754, 1208, 5081, 10052, 15058 and 20200. For all these scenarios, we consider a jump size that is normally distributed with mean 0 and standard deviation equal to 1.5%. We did not impose any restrictions on the maximum number of jumps per day. Thus, more than one jump may occur during a trading day.

In Table 3, we report the size corrected power of the tests for $\lambda = 0.5$ and $\lambda = 2$. Results for alternative values of λ, reported in Dumitru and Urga (2011), confirm the findings discussed here. The frequency of correctly identified jumps increases as the jump intensity raises.

[Insert Table 3 here]

In general, tests which display a higher size are also better ranked in terms of power. The best tests are the intraday ABD-LM procedures and the PZ test. For the intraday procedures, when $\lambda = 0.5$, the size adjusted power is 99% for a sampling frequency of 1 second and then gradually diminishes as the sampling frequency decreases. At lower frequencies, the power for ABD-LM is 91%, for a sampling frequency of 1 minute, 80% for 5 minutes data, 66% at 15 minutes and finally 54% at 30 minutes. For the PZ procedure we observe a very high power (99% at 1 sec) which decreases with the sampling frequency, and remains higher than the other procedures (except the intraday tests) for data sampled at 1, 5 and 15 minutes. It is worth mentioning that at 30 minutes the power of PZ is (very close to) 0 for all values of λ. Note that size adjusted power implies centering with the average test statistic under the null, which in this case is extremely high, i.e. $3.29 \cdot 10^{12}$. The JO test displays a very high power (97% for $\lambda = .5$) at 1 second and can be ranked after the PZ, ABD-LM and AJ tests. However, at lower frequencies, its power becomes slightly lower than the other tests, except AJ. Power becomes 85% at 1 minute, 73% at 5 minutes, 57% at 15 minutes and finally 45% for data sampled every 30 minutes. The intuition for this results can be explained by considering the Taylor series expansion of the test statistic, which contains sums of returns at powers higher than 3. At lower frequencies, jumps are offset by summation with returns of opposite sign. Both versions of the AJ test display a high power at 1 second, which plummets at lower frequencies. For instance, for $\lambda = .5$, the power decreases at around 80% when sampling is done every minute, for both versions of the test, followed by a fall at a level of 23% for the version based on threshold estimators and 32% for the test based on power variations, for a sampling frequency of 5 minutes. If we look at lower frequencies, the test based on power variation-type estimators displays a gradual decrease in power, which gets to a value of 24% for a 30 minutes sampling frequency, while the version based on threshold estimators displays a very low power of 0.6% at 30 minutes. This test relies on the asymptotic scale proportionality of two realized power variations computed on two different time scales. At lower frequencies, the above relationship does no longer hold, leading to the poor performance of the test.

The BNS, CPR, Med and Min tests display a very similar behaviour. They all exhibit very good power properties, with a power ranging between 95% and 96% when sampling at every second for $\lambda = .5$, which then decreases with the decrease in the sampling frequency, with values below the ones observed for the intraday and PZ tests. Over all frequencies, the highest power is displayed by CPR, followed by Med, BNS and Min.

Varying jump size A further insight on the ability of all these procedures to identify jumps can be attained by varying the jump size. We now fix the number of jumps for the entire sample and vary the jump size. But, we maintain its nondeterministic character, by drawing it from a normal distribution with mean 0 and a standard deviation (σ_{jump}) that ranges between 0 and 2 with a growth rate of .5. Table 4 reports the power of the jump detection procedures for $\sigma_{jump} = 0.5$ and $\sigma_{jump} = 2$. Results for alternative values of σ_{jump} reported in Dumitru and Urga (2011) confirm the findings reported here.

[Insert Table 4 here]

Overall, the performance of all tests increases with the size of the jumps. The ranking of the tests is in line with what was found for the case of varying jump intensity. There is a confirmation about the very good ability of the ABD-LM and PZ tests to detect jumps, with power ranging between 95% and 99% at 1 second, which gradually decreases with the sampling frequency. Just as in the case of varying jump intensity, the JO procedure exhibits a very high power at 1 second sampling frequency, ranging between 89% and 98%. However, at lower frequencies, the procedure loses power in front of all other tests with the exception of AJ. We observe the same ranking as in the previous section for the CPR, Med, BNS and Min procedures. At the highest frequency, they exhibit a power ranging between 84% and 88% for the lowest levels of jump sizes ($\sigma_{jump} = .5$). When σ_{jump} takes its highest value, 2, power is around 97% for all 4 procedures at 1 second. For lower frequencies, the performance of these tests decays. The AJ does again very well for the highest frequency and ranks itself immediately after the PZ and ABD-LM procedures. However, at lower frequencies, we observe a dramatic decrease in power.

3.2.2 Size and power of the tests in the presence of microstructure noise

So far, our analysis was based on the assumption that prices are generated as a continuous time jump diffusion process. However, when we deal with financial assets, the observed price process is a discrete one. It is either constant, generating zero returns, or changes a lot from one transaction to another. We say that prices are contaminated with microstructure noise, which may obstruct our viewing of the real price process.

In what follows, we simulate i.i.d. microstructure noise normally distributed with mean 0 and a varying variance. The noise is then added to the SV1F model with medium mean reversion to study its effects on the statistical properties of the tests for jumps.

SIZE The values for the standard deviation of the noise (σ_{noise}) are: .027, .052, and 0.080, which are equivalent to 0.025, 0.047 and 0.073 in terms of signal-to-noise ratios, defined as $\frac{\sigma_{noise}}{IV_t^{avg}}$, where IV_t^{avg} stands for the daily average integrated variance. Table 5 reports the frequencies of spuriously detected jumps for all tests, under alternative sampling frequencies and a medium level of contamination with microstructure noise, i.e. $\sigma_{noise} = .052$. The full set of results is reported in Dumitru and Urga (2011).

[Insert Table 5 here]

Apart from the AJ and JO tests, all tests become severely undersized in the presence of microstructure noise with an increasing size distortion as the variance of the noise grows. AJ(threshold) does better than

AJ(power var) when lower sampling frequencies are considered. If sampling is made every 15 minutes, the size of AJ(threshold) gets close to the nominal one. When $\sigma_{noise} = 0.052$ (Table 5), size is 3.7% for the version based on threshold estimators, whereas for the other version of the test, it reaches a very high level of 10.9%. The JO procedure generally displays a very high size in the presence of noise at 1 second, which increases with the variance of the noise. However, when sampling is done at lower frequencies (from 1 minute onward), size decreases abruptly in the beginning and then, moderately increases again. The large size at 1 second is due to the fact that the distribution of the test statistic shifts to the right in the presence of microstructure noise. JO notice this problem in the original paper and propose corrections for the test statistics in the presence of i.i.d. noise. The least affected by noise is the PZ procedure, which, at the highest sampling frequency, displays a size close to the nominal one even for the highest values of σ_{noise}. This is a consequence of its higher and rapidly increasing size, which turns out to be an advantage in this case, as it compensates the downward bias caused by the presence of noise. The intraday tests, ABD-LM, also behave very well in the presence of i.i.d. noise, being less underbiased than other procedures at high frequencies. The BNS, CPR, Med and Min tests are severely undersized at very high frequencies. Then their size increases with the decrease in the sampling frequency. For the levels of noise reported here, BNS, CPR and Med tend to reach a size level close to the nominal only at 15 minutes. The Min procedure, which tends to be undersized in the absence of noise, displays size levels lower than the nominal one for all frequencies. Except the PZ test which has a size close to the nominal one at 1 second and 1 minute sampling frequency, as if the noise was not present, all other tests tend to get close to the nominal size as the sampling frequency diminishes: JO somewhere between the 5 and 15 minutes sampling frequencies, AJ, BNS, CPR and Med generally at 15 minutes, and ABD-LM somewhere between 15 and 30 minutes.

POWER In this section, we examine how the ability of the tests to detect jumps changes in the presence of microstructure noise. We consider a jump process with intensity $\lambda = .5$ and jump sizes randomly drawn from a $\mathscr{N}(0, 1.5)$. The size adjusted power for all tests and for a medium level of contamination with noise, i.e. $\sigma_{noise} = .052$, is reported in Table 6. Results for $\sigma_{noise} = .027$ and .080 are reported in Dumitru and Urga (2011).

[Insert Table 6 here]

The hierarchy of the tests in terms of power remains close to the one for the case of no noise. As the size of the noise standard deviation increases, we observe a decrease in power. The intraday and PZ procedures display again the best power. ABD-LM displays the same tendency of decreasing power with the decrease in the sampling frequency, as if the noise were not present. PZ seems to be affected by noise at 1 second, but then regains power at 1 minute (84%). Power at 30 minutes is again extremely low, just as in the case without noise. BNS, CPR, Med and Min tend to behave similarly again. They suffer a significant loss in power at 1 second, but then tend to regain it. All these tests exhibit a very fast power recovery, occurring at 1 minute. The highest power at 1 minute (79%) is showed by CPR. It is followed by Med, BNS and Min, with closed values for power. Though we notice improvements, however power is lower than in the absence of noise. JO displays a similar pattern to the above tests, but an overall slightly weaker performance. It tends to be better ranked for lower levels of noise. There is a decrease in the corrected power at 1 second, followed by a slight recovery of power up to 1 minute or 5 minutes. Power at 1 minute is 76% and decreases again at lower frequencies. By far the worst performance is observed for the AJ tests, which lose their power at 1 second. For lower frequencies, we notice a slight increase in power. AJ(power var) performs somewhat better than the test based on threshold estimators.

To summarize, in the presence of noise, the size of the various jump detection procedures comes close the nominal one when sampling at lower frequencies. In the case of the power, this effect is much more moderate. Power increases at 1 minute for almost all tests. At lower frequencies, power tends to decrease, just as when noise is not present. Of course, the results on the frequencies at which size and power are regained depend on the simulated data generating process and in particular on the type and amount of noise. The issue of selecting a frequency for applying tests for jumps is further discussed in Section 5.

4 EXTENSIONS

In this section, we propose two extensions to the existing battery of tests. The first regards the finite sample improvements based on simulated critical values for the ABD-LM procedure. The second regards the combinations of procedures and frequencies to reduce the number of detected spurious jumps.

4.1 Finite sample distributions for the ABD and LM tests

So far in the paper, we have used the ABD-LM procedure based on LM asymptotic critical values. In this section, we propose to use simulated critical values for the maximum of the ABD-LM tests statistics, to account for the sample size in the inference process. [We are grateful to Dobrislav Dobrev for suggesting us to explore the use of approximate finite sample distributions.] The approach we propose is based on the so-called "Monte Carlo Reality Check" proposed by White (2000). Critical values are obtained in the following way. Let n be the number of daily observations and $\widehat{V}_j$ the local volatility estimate at time t_j, obtained as in ABD-LM. At each time, t_j, we simulate 10,000 times n observations from $\mathscr{N}(0, \widehat{V}_j)$. For every run, we take the maximum over the n observations and thus, the 10,000 maxima represent the approximate finite sample distribution from which we select the critical values. Finally, the statistic in (7) is compared to the corresponding critical value.

To assess the performance of the simulation-based critical values, we simulate the SV1F model with medium mean reversion, augmented by jumps and microstructure noise. The latter is sampled from a $\mathcal{N}(0, \sigma_{noise})$, where σ_{noise} takes the values as reported in Section 3.2.2. We compare the results in terms of size and power with the ones based on the asymptotic distribution of LM test. For the 10,000 simulated trading days, we sampled data at 1 second, 1, 5, 15 and 30 minutes.

SIZE We quantify size as follows. For each of the 10,000 simulated trading days, we check whether the procedures rejected the null at least once during that day, and then, we calculate the percentage of days, out of the total number of days, with at least one jump. Figure 1 reports the nominal, asymptotic and simulated size for different sampling frequencies for the SV1F model with $\sigma_{noise} = 0.052$. The test based on simulated critical values is less undersized at very high frequencies (1 second) than the LM asymptotic counterpart. This conclusion is valid for all significance levels considered. Size for the finite sample adjustment procedure increases, but remains close to the nominal one very at 1 minute. However, at lower frequencies the procedure tends to become more oversized than its asymptotic counterpart. It is evident that the simulated-based procedure works better at high frequencies in the presence of i.i.d microstructure noise.

[Insert Figure 1 here]

POWER In order to assess the power of our finite sample adjustment, we add jumps with intensity $\lambda = 0.5$ and size with $\sigma_{jump} = 1.5\%$. We compute the power as the percentage of days the procedures were able to correctly signal that at least one jump occurred during the day.

Figure 2 illustrates the size adjusted power as a function of the sampling frequency for three levels of σ_{noise}, 0.027 (low), 0.052 (medium) and 0.08 (high) and 5%, 1% and .1% significance levels.

[Insert Figure 2 here]

We observe the size adjusted power is systematically higher when we use simulated critical values for all sampling frequencies, all significance levels and all levels of noise. The gap between the power of simulation-based and asymptotic procedures widens as the variance of the noise increases, making the use of simulated critical values very suitable in presence of higher levels of contamination with noise. Moreover, the size adjusted power for the simulation-based procedure shows a better performance as the significance levels become lower.

To summarize, the simulation-based procedure displays lower size distortions in the presence of microstructure noise only at higher frequencies, while we observe an overall better performance in terms of power. Just as in ABD, we think the use of lower significance levels (.1%) can help to correctly disentangle jumps from the price process, without generating a high number of spurious jumps.

4.2 Cross-performances of the tests

In a world without microstructure noise, users of jump tests should simply opt for the best procedure in terms of both size and power and apply it at the highest frequency available. However, as reported in Section 3.2.2, the presence of microstructure noise leads to serious distortions in both size and power at high frequencies. The sensible solution to sample less often can lead to a loss in power. In this section, we propose a procedure that combines tests and frequencies through both reunions and intersections. This new procedure retains a good level of power and, at the same time, reduces the number of detected spurious jumps.

The idea of using more sampling frequencies has been exploited before in high frequency econometrics for volatility estimation in the presence of noise. Zhang et al. (2005) and Zhang (2006) propose using at least two time scales in estimating the quadratic variation of the price process. Their methodology relies on the argument that one can capture the volatility of the prices at lower frequencies and the variation of the noise process at very high frequencies. In our case, combining tests and frequencies allows to extract as much information as possible on jumps in the price process from different procedures and time scales. We combine the advantages of more than one procedure, when combining procedures, and make use of more data ("throwing away less data"), when combining frequencies. Moreover, finding a testing procedure that reduces the percentage of spurious jumps is essential for the case in which a large number of zero returns are observed, which is very common in practice. In this case, the integrated volatility tends to be underestimated by various jump robust estimators, especially those based on the bipower variation (Andersen et al., 2009). This leads to an over-rejection of the null for most of the jump tests. The procedure proposed in this paper averages results from separate tests or time scales, and, as we show below, decreases the percentage of spurious jumps.

To analyze the performance of our procedure, we rely on data simulated from the SV1F model, augmented by jumps and microstructure noise. We consider rare compound Poisson jumps with intensity $\lambda = 0.5$ and a size distributed as a $\mathscr{N}(0, 1.5)$, while the microstructure noise is sampled from a $\mathscr{N}(0, .052)$.

First, we apply the same procedure at different sampling frequencies, i.e. 1, 5 and 15 minutes. Once the test statistics are computed, we take intersections of the results at 1 and 5 minutes and at 5 and 15 minutes.

We then take the reunion over the two sets of results. For instance, if we consider the BNS test, our decision rule can be written as $(BNS1 \cap BNS5) \cup (BNS5 \cap BNS15)$. This means that on a certain trading day, the path of the price process is considered discontinuous if one or more jumps is/ are detected by the BNS test performed at 5 minutes and at least by one of the other two BNS tests at 1 and 15 minutes.

Table 7 reports the results from combining frequencies for the BNS, CPR, ABD-LM, Med, Min, PZ and JO procedures. In each case, we computed three measures: the percentage of correctly classified jumps ('Jump'), the percentage of days that are correctly classified as having continuous paths ('No jump'), and the percentage of spurious jumps ('Spurious'). The results in Table 7 should be interpreted by contrasting them with the size and power values of the tests reported in Tables 5 and 6. The significance level for all tests is 5%.

[Insert Table 7 here]

The results suggest that our procedure manages to average the power over frequencies and/or tests, combined with a substantial decrease in the percentage of spurious jumps. For instance, in the second column of Table 7, we observe that the percentage of spuriously detected jumps becomes very low (.25%) and is combined with a very high proportion (95.74%) of days that were rightly classified as without jumps and a high proportion of correctly identified jumps (62.29%). Note that the latter percentage averages out the power of the BNS test at different sampling frequencies, i.e. 76% at 1 minute, 69% at 5 minutes, and 54% at 15 minutes (see Table 6).

Further, by comparing the results in Table 7 with those in Tables 5 and 6, we notice that one can make the most of this procedure when we combine frequencies of tests with a high power, such as PZ and ABD-LM. For instance, PZ has a very high power, but also a high size. Thus, combining frequencies allows this test to maintain a good power (77%) and at the same time, the percentage of spurious jumps decreases significantly.

In addition to combining sampling frequencies, we also combine different tests on data sampled at the same frequency. In Table 8, we report the results for a 5 minutes sampling frequency, while the results for alternative frequencies, reported in Dumitru and Urga (2011), confirm the findings below.

[Insert Table 8 here]

When we combine tests, the percentage of correctly classified jumps averages out the performance of individual tests. Moreover, there is a significant decrease in the percentage of spurious jumps. We observe that the best performance is attained when we use combinations of the most powerful tests, such as PZ and ABD-LM and, in addition, when one of the two tests is in both intersections with the other tests. For instance, the combination $(BNS5 \cap PZ5) \cup (PZ5 \cap Med5)$ intersects PZ twice with two other procedures (BNS and Med). This decision rule generates a high percentage of correctly classified jumps (72%) and a low percentage of spurious jumps (1.6%).

To summarize, the procedure combining sampling frequencies and tests performs better than the single tests. It preserves a high percentage of rightly classified jumps, with a significant decrease in the percentage of spurious jumps. A useful guideline for users is to include powerful tests in various combinations.

5 EMPIRICAL APPLICATION

In this section, we report an empirical application based on high frequency data for five stocks listed in the New York Stock Exchange, namely Procter&Gamble, IBM, JP Morgan, General Electric and Disney. Our dataset covers 5 years, running from the 3rd of January 2005 to the end of December 2009, with an average of 1250 days. In order to carry out the jump tests, we rely on transaction data, which we sample at 1, 5, 10, 15 and 30 minutes. This sampling schemes left us with an average of approximately 414 data points at 1 minute, 82 observations at 5 minutes, 40 at 10 minutes, 26 at 15 and 12 at 30 minutes.

Table 9 reports the proportions of identified jumps. While there is no high variability of the findings from one stock to another, results vary considerably between procedures and frequencies. For each procedure and for each stock, there is a decrease in the percentage of identified jumps as we sample less frequently. At 1 minute, most of the tests detect a high percentage of jumps, which then substantially decreases at 5 minutes. From 5 minutes onward, the decrease is less pronounced, with a stabilization occurring around 10-15 minutes. There is not a clear guideline in the literature on an optimal frequency to use in applying jump tests. However, the empirical evidence in this paper suggests to apply the tests to a variety of frequencies and then choose the frequency at which the percentage of jumps stabilizes. This rule of thumb corresponds to the 10 minutes frequency.

[Insert Table 9 here]

In what follows, we comment only the results for IBM. The same comments apply for all other stocks. For IBM, PZ and ABD-LM identify 97%, followed by CPR with 88% and BNS with 77%. At lower frequencies, this percentage drastically drops. A possible explanation is that at higher frequencies, the procedures detect a spurious jumps because of the presence of high number of zero returns. Note that the above tests are based on BV_t (BNS and ABD-LM in the test statistic, and CPR and PZ in the threshold volatility estimate), which becomes downward biased in the presence of many zero returns, causing the tests statistics to increase and thus to over-reject the null. On the contrary, for tests not based on BV_t, such as Min, Med, this effect is no longer that relevant. The JO test also seems only slightly affected by zero returns at 1 minute (37% days with jumps), though the percentage of detected jumps does not change very much with the frequency. The two AJ tests display percentages of identified jumps around 55% at 1 second, but at lower frequencies AJ(threshold) detects a very small amount of jumps, whereas AJ(power var) between 24% and 33% jumps for IBM.

The high variability in the percentage of detected jumps reported in Table 9 calls for the application of the combinations of tests as we proposed in Section 4.2. Table 10 reports the proportion of jumps detected by different combinations of frequencies and procedures for IBM. There is a confirmation that combining procedures leads to a decrease in the proportion of identified jumps. Moreover, there is evidence of higher proportion of jumps when procedures with higher power, like ABD-LM and PZ, are combined. When combining frequencies, in all cases except the ABD-LM and PZ procedures, the proportion of detected jumps is lower than the proportion identified by the individual procedures on each of the combined frequencies, as reported in Table 9. In the case of the ABD-LM and PZ procedures instead, the combination of frequencies leads to a percentage of jumps in the range of what was obtained on individual procedures, due to the high individual power of the two tests. When combing different tests for jumps for a 10 minutes sampling frequency, we observe that the proportion of identified jumps is in the range of the proportions obtained in the case of individual procedures, but it tends to be closer to the lower values of the individual procedures.

[Insert Table 10 here]

A final issue we address is the relative contribution of jumps to the quadratic variation of the price process. [We wish to thank an Associate Editor for suggesting us to explore this issue.] For each test for jumps, we detect all days with discontinuities. Then, we eliminate jumps by removing the highest return in absolute value that occurs on days with jumps. We compute the realized variation on the initial price series sampled every 10 minutes, as well as on the new series without jumps. The former is a proxy for the quadratic variation of the price process, the latter for the integrated variance (RV_C), whereas the difference between the two estimates the quadratic variation of the jump process (RV_J). Table 11, Panel A, reports for each test for jumps and for all years considered in our sample the percentages of RV_C and RV_J in the yearly realized variation for IBM. The yearly RV increases from a level of 0.023 in 2005 to a peak of 0.155 in 2008, when the sub-prime crisis affected mostly the financial markets. In 2009, RV decreases to 0.058, which is still very high in comparison to tranquil years, such as 2005 and 2006. The levels of RV_C and RV_J vary a lot depending on the used jump detection procedure. Thus, during the first two calm years, 2005 and 2006, the percentage of RV_J is between 8% and 33% by the alternative procedures. However, this percentage is systematically higher in 2006 than 2005 for all tests. During the years of the financial crises, 2007-2009, this percentage drops. A minimum for almost all testing procedures is reached in 2008, the year of maximum volatility, when the percentage of RV_J varies between 4% and 22%, depending on the adopted procedure. In periods of high volatility, the ability of the tests to pick up jumps is lower, whereas in calmer periods, jumps are much more visible. Panel B shows the values for RV_C and RV_J for some combinations of frequencies and procedures. As expected, the the percentage of RV_J is generally lower when combinations are used than when individual tests are applied. When frequencies are combined, RV_J is always lower, whereas when tests are combined, RV_J is in the range of the values for individual tests. Jumps show the highest contribution when combining frequencies of the higher power tests, such as ABD-LM and PZ.

[Insert Table 11 here]

6 CONCLUSION

This paper brings three major contributions to the existing literature on non parametric jump testing procedures in high frequency data. First, we offered a robust and comprehensive comparison between nine alternative jump detection procedures. We conducted an extensive numerical analysis using alternative levels of volatility, different levels of persistence in the volatility factor(s), various jump intensities and jump sizes, and different levels of microstructure noise contamination. The main conclusion is that the ABD-LM tests show overall the best performance, though in the case of extremely volatile processes, they become highly oversized. Second, we proposed a finite sample adjustment for the ABD-LM procedure. We suggested the use of simulated critical values, as an alternative to the asymptotic critical values, leading to an improvement in size at higher sampling frequencies and an overall improvement in power. Third, given the high variability of the performance of the tests, we proposed to combine available jump tests through both intersections and reunions over alternative sampling frequencies and procedures. We showed that combining procedures with high power with other tests preserved power with a considerable reduction in the percentage of detected spurious jumps. An empirical exercise, conducted on five stocks listed in the New York Stock Exchange, confirmed the results from the simulations.

The analysis in the present paper can be extended in at least four different ways. As jumps are caused by the arrival of new information to the market, it would be interesting to examine the so-called economic significance of jumps. For the simulation set-up, the case of i.i.d. microstructure noise can be extended to account for correlation and zero returns. Another natural extension is to consider arrivals with an infinite number of jumps. Finally, to minimize the detection of spurious jumps, the combination of tests could be enriched by considering test averaging procedures using Fisher(1925)'s method of combining p-values of different tests. We leave these extensions to future research.

Guidelines for empirical work

Based on the findings of the simulation exercise and the empirical application, we offer some guidelines for empirical work when testing for jumps based on high frequency data.

If users opt for a single test, the ABD-LM procedure is a good choice, as it retains high power with a manageable size and it is also informative on the time of the jump. If the sample period is characterized by high volatility, tests with a more conservative size (BNS, Med, Min) are a better option.

However, since high frequency data is contaminated with microstructure noise, we recommend to implement combinations of tests or frequencies, as described in Section 4.2. The best way to proceed is to make use of powerful tests, such as ABD-LM and PZ, either by combining different frequencies for the same procedure or by combining powerful tests with tests with a moderate size.

To select the sampling frequency at which to apply tests for jumps, the empirical evidence suggests to implement all tests on data sampled at a variety of frequencies and then choose the frequency at which the percentage of jumps stabilizes. This strategy can also be extended to combinations of tests.

ACKNOWLEDGMENTS

We are grateful to participants in the Centre of Econometric Analysis Seminar at Cass Business School (15 June 2008), in the 6th Oxmetrics Conference (Cass Business School, 17-18 September 2008), in particular Sir David Hendry, Siem Jan Koopman and Sébastien Laurent, the PhD Workshop in Turin (Polytechnic University of Turin, 20-21 November 2008), the Workshop on Financial Econometrics (Fields Institute in Toronto, 23-24 April), in particular Yacine Aït-Sahalia, Nour Meddahi, Roberto Renó and Eric Renault for useful comments. George Tauchen provided us with useful suggestions for our simulation design. We are grateful to Mardi Dungey and Abdullah Yalama for their comments and to Brian McGlennon from ICAP, for his help in building and refining our dataset. A special thank to Dobrislav Dobrev for his extremely useful comments and suggestions on a previous version of the paper. We wish to thank the Editor, Jonathan H. Wright, an Associate Editor and two Referees for very useful comments and suggestions which greatly helped to improve the paper. The usual disclaim applies.

REFERENCES

Aït-Sahalia, Y. (2004), "Disentangling Diffusion from Jumps," *Journal of Financial Economics*, 74, 487–528.

Aït-Sahalia, Y. and Jacod, J. (2008), "Testing for Jumps in a Discretely Observed Process," *Annals of Statistics*, 37, 184–222.

Andersen, T. G. and Bollerslev, T. (1998), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International economic review*, 39, 885–905.

Andersen, T. G., Bollerslev, T., and Dobrev, D. (2007), "No-Arbitrage Semi-Martingale Restrictions for Continuous-Time Volatility Models Subject to Leverage Effects, Jumps and I.I.D. Noise: Theory and Testable Distributional Implications," *Journal of Econometrics*, 138, 125–180.

Andersen, T. G., Dobrev, D., and Schaumburg, E. (2009), “Jump-Robust Volatility Estimation using Nearest Neighbor Truncation,” NBER Working Papers 15533, National Bureau of Economic Research, Inc.

Barndorff-Nielsen, O. and Shephard, N. (2004), “Power and Bipower Variation with Stochastic Volatility and Jumps,” *Journal of Financial Econometrics*, 2, 1–48.

— (2006), “Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation,” *Journal of Financial Econometrics*, 4, 1–30.

Barndorff-Nielsen, O. E., Shephard, N., and Winkel, M. (2006), “Limit Theorems for Multipower Variation in the Presence of Jumps,” *Stochastic Processes and their Applications*, 116, 796–806.

Chernov, M., Gallant, A. R., Ghysels, E., and Tauchen, G. (2003), “Alternative Models for Stock Price Dynamics,” *Journal of Econometrics*, 116, 225–257.

Corsi, F., Pirino, D., and Renò, R. (2010), “Threshold Bipower Variation and the Impact of Jumps on Volatility Forecasting,” *Journal of Econometrics*, 159, 276–288.

Dumitru, A.-M. and Urga, G. (2011), “Identifying Jumps in Financial Assets: a Comparison Between Nonparametric Jump Tests. [Extended Version],” Working Paper WP-CEA-1-2011, Centre of Econometric Analysis,Cass Business School.

Huang, X. and Tauchen, G. (2005), “The Relative Contribution of Jumps to Total Price Variance,” *Journal of Financial Econometrics*, 3, 456–499.

Jiang, G. J. and Oomen, R. (2008), “Testing for Jumps when Asset Prices Are Observed with Noise- a “Swap Variance” Approach,” *Journal of Econometrics*, 144, 352–370.

Lee, S. S. and Mykland, P. A. (2008), “Jumps in Financial Markets: a New Nonparametric Test and Jump Dynamics,” *Review of Financial studies*, 21, 2535–2563.

Mancini, C. (2009), “Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps,” *Scandinavian Journal of Statistics*, 36, 270–296.

Podolskij, M. and Ziggel, D. (2010), “New Tests for Jumps in Semimartingale Models,” *Statistical Inference for Stochastic Processes*, 13, 15–41.

Schwert, M. W. (2009), “Hop, Skip and Jump What Are Modern Jump Tests Finding in Stock Returns?” Working paper, Duke University.

Theodosiou, M. and Žikeš, F. (2010), “A Comprehensive Comparison of Alternative Tests for Jumps in Asset Prices,” Working paper, Imperial College London.

White, H. (2000), “A Reality Check For Data Snooping,” *Econometrica*, 68, 1097–1126.

Zhang, L. (2006), “Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach,” *Bernoulli*, 12, 1019–1043.

Zhang, L., Mykland, P. A., and Ait-Shalia, Y. (2005), “A tale of two time scales: determining integrated volatility with noisy high-frequency data,” *Journal of the American Statistical Association*, 100, 1394–1411.

Table 1. Size of the tests for jumps for the SV1F model with medium mean reversion

Procedure	1sec	1 min	5 min	15 min	30 min
AJ(threshold)	0.047	0.038	0.031	0.027	0.014
AJ(power var)	0.048	0.046	0.051	0.088	0.150
BNS	0.048	0.054	0.053	0.057	0.063
CPR	0.052	0.055	0.056	0.064	0.075
JO	0.065	0.069	0.086	0.122	0.189
ABD-LM	0.055	0.066	0.074	0.063	0.059
Med	0.051	0.050	0.052	0.053	0.064
Min	0.047	0.046	0.044	0.040	0.035
PZ	0.049	0.065	0.083	0.100	0.121

Table 2. Size of the tests for jumps for the SV2F model

Procedure	1sec	1 min	5 min	15 min	30 min
AJ(threshold)	0.127	0.094	0.039	0.020	0.008
AJ(power var)	0.052	0.077	0.121	0.205	0.255
BNS	0.054	0.073	0.097	0.113	0.119
CPR	0.062	0.165	0.150	0.168	0.247
JO	0.070	0.106	0.163	0.247	0.327
ABD-LM	0.993	0.699	0.482	0.339	0.254
Med	0.054	0.074	0.102	0.122	0.142
Min	0.052	0.063	0.084	0.082	0.080
PZ	0.701	0.648	0.448	0.305	0.239

Table 3. Size adjusted power for varying jump intensities

λ	Procedure	1sec	1 min	5 min	15 min	30 min
	AJ(threshold)	0.972	0.807	0.232	0.044	0.015
	AJ(power var)	0.972	0.811	0.322	0.216	0.319
	BNS	0.959	0.854	0.728	0.562	0.399
	CPR	0.961	0.870	0.766	0.630	0.504
0.5	JO	0.966	0.853	0.730	0.574	0.445
	ABD-LM	0.985	0.909	0.799	0.663	0.537
	Med	0.955	0.860	0.753	0.603	0.461
	Min	0.949	0.840	0.709	0.544	0.347
	PZ	0.982	0.909	0.804	0.679	0.000
	AJ(threshold)	0.992	0.858	0.192	0.030	0.009
	AJ(power var)	0.992	0.900	0.409	0.256	0.353
	BNS	0.984	0.933	0.854	0.688	0.485
	CPR	0.986	0.942	0.882	0.778	0.618
2	JO	0.988	0.919	0.823	0.655	0.489
	ABD-LM	0.995	0.957	0.883	0.728	0.533
	Med	0.983	0.930	0.845	0.688	0.515
	Min	0.981	0.924	0.829	0.645	0.420
	PZ	0.994	0.960	0.907	0.800	0.000

Table 4. Size adjusted power for a varying jump variance

σ_{jump}	Procedure	1sec	1 min	5 min	15 min	30 min
	AJ(threshold)	0.921	0.496	0.108	0.026	0.026
	AJ(power var)	0.921	0.509	0.159	0.120	0.232
	BNS	0.872	0.565	0.340	0.178	0.118
0.5	CPR	0.880	0.615	0.394	0.222	0.146
	JO	0.892	0.566	0.322	0.171	0.123
	ABD-LM	0.964	0.698	0.448	0.245	0.128
	Med	0.865	0.590	0.368	0.208	0.132
	Min	0.843	0.532	0.307	0.147	0.076
	PZ	0.950	0.720	0.482	0.262	0.001
	AJ(threshold)	0.983	0.850	0.244	0.040	0.011
	AJ(power var)	0.983	0.857	0.376	0.245	0.353
	BNS	0.970	0.891	0.799	0.660	0.501
	CPR	0.973	0.902	0.824	0.708	0.588
2	JO	0.977	0.891	0.794	0.665	0.519
	ABD-LM	0.990	0.901	0.816	0.693	0.569
	Med	0.971	0.892	0.806	0.681	0.544
	Min	0.964	0.880	0.778	0.631	0.447
	PZ	0.988	0.932	0.856	0.744	0.001

Table 5. Size in the presence of i.i.d. microstructure noise

σ_{noise}	Procedure	1sec	1 min	5 min	15 min	30 min
	AJ(threshold)	1.000	0.956	0.160	0.037	0.014
	AJ(power var)	1.000	0.948	0.187	0.109	0.165
	BNS	0.000	0.002	0.043	0.054	0.061
0.052	CPR	0.000	0.003	0.047	0.061	0.075
	JO	0.366	0.017	0.064	0.113	0.185
	ABD-LM	0.009	0.040	0.061	0.059	0.059
	Med	0.000	0.005	0.041	0.055	0.062
	Min	0.000	0.003	0.033	0.034	0.037
	PZ	0.051	0.059	0.087	0.099	0.118

Table 6. Size adjusted power of the tests in the presence of i.i.d. microstructure noise

σ_{noise}	Procedure	1sec	1 min	5 min	15 min	30 min
	AJ(threshold)	0.006	0.015	0.036	0.020	0.010
	AJ(power var)	0.019	0.032	0.119	0.161	0.252
	BNS	0.553	0.760	0.686	0.540	0.384
0.052	CPR	0.593	0.786	0.725	0.611	0.484
	JO	0.5570	0.7562	0.6846	0.5547	0.4157
	ABD-LM	0.851	0.820	0.738	0.605	0.466
	Med	0.547	0.773	0.713	0.586	0.444
	Min	0.507	0.740	0.668	0.514	0.344
	PZ	0.809	0.844	0.778	0.656	0.000

Table 7. Results from combining tests using different frequencies

Procedure	$(BNS1 \cap BNS5) \cup$ $(BNS5 \cap BNS15)$	$(CPR1 \cap CPR5) \cup$ $(CPR5 \cap CPR15)$	$(ABDLM1 \cap ABDLM5) \cup$ $(ABDLM5 \cap ABDLM15)$	$(Med1 \cap Med5) \cup$ $(Med5 \cap Med15)$	$(Min1 \cap Min5) \cup$ $(Min5 \cap Min15)$
'Jump'	0.6229	0.6772	0.7465	0.6581	0.5953
'No Jump'	0.9574	0.9554	0.9334	0.9583	0.9674
'Spurious'	0.0025	0.0022	0.0247	0.0027	0.0010
	Procedure	$(PZ1 \cap PZ5) \cup$ $(PZ5 \cap PZ15)$	$(JO1 \cap JO5) \cup$ $(JO5 \cap JO15)$		
	'Jump'	0.7744	0.7202		
	'No Jump'	0.9094	0.9324		
	'Spurious'	0.0140	0.0206		

Table 8. Results from combining different tests for jumps for data sampled every 5 minutes

'Procedures'	$(Med5 \cap ABDLM5) \cup$ $(ABDLM5 \cap BNS5)$	$(CPR5 \cap BNS5) \cup$ $(BNS5 \cap Med5)$	$(CPR5 \cap BNS5) \cup$ $(BNS5 \cap PZ5)$	$(CPR5 \cap BNS5) \cup$ $(BNS5 \cap Min5)$	$(Med5 \cap BNS5) \cup$ $(BNS5 \cap ABDLM5)$
'Jump'	0.6848	0.6496	0.6658	0.6431	0.6623
'No Jump'	0.9297	0.9434	0.9525	0.9384	0.9543
'Spurious'	0.0119	0.0102	0.0160	0.0084	0.0133
'Procedures'	$(CPR5 \cap ABDLM5) \cup$ $(ABDLM5 \cap PZ5)$	$(JO5 \cap BNS5) \cup$ $(BNS5 \cap PZ5)$	$(BNS5 \cap PZ5) \cup$ $(PZ5 \cap Med5)$	$(CPR5 \cap PZ5) \cup$ $(PZ5 \cap Med5)$	
'Jump'	0.7405	0.6661	0.7165	0.7150	
'No Jump'	0.9298	0.9481	0.9122	0.9028	
'Spurious'	0.0240	0.0158	0.0158	0.0104	

Table 9. Proportion of days with jumps, at different sampling frequencies, as identified by the following procedures: AJ (both versions), BNS, CPR, JO, ABD-LM, Med, Min and PZ

Company	Procedure	1 min	5 min	10 min	15 min	30 min
	AJ(threshold)	0.552	0.109	0.050	0.024	0.014
	AJ(power var)	0.606	0.357	0.293	0.264	0.266
	BNS	0.814	0.273	0.154	0.157	0.132
PG	CPR	0.915	0.391	0.221	0.190	0.149
	JO	0.407	0.212	0.188	0.211	0.277
	ABD-LM	0.972	0.506	0.270	0.182	0.086
	Med	0.484	0.174	0.144	0.152	0.140
	Min	0.453	0.157	0.106	0.102	0.074
	PZ	0.969	0.598	0.344	0.278	0.226
	AJ(threshold)	0.534	0.094	0.043	0.020	0.014
	AJ(power var)	0.592	0.330	0.274	0.236	0.237
	BNS	0.765	0.253	0.196	0.191	0.142
IBM	CPR	0.884	0.374	0.257	0.228	0.162
	JO	0.374	0.222	0.230	0.244	0.283
	ABD-LM	0.974	0.512	0.292	0.207	0.097
	Med	0.446	0.174	0.193	0.193	0.156
	Min	0.430	0.148	0.123	0.134	0.090
	PZ	0.967	0.574	0.389	0.325	0.223
	AJ(threshold)	0.548	0.090	0.037	0.031	0.015
	AJ(power var)	0.596	0.317	0.261	0.252	0.263
	BNS	0.708	0.237	0.175	0.155	0.119
JPM	CPR	0.842	0.352	0.237	0.191	0.146
	JO	0.317	0.218	0.212	0.221	0.293
	ABD-LM	0.950	0.500	0.282	0.191	0.121
	Med	0.318	0.167	0.169	0.152	0.132
	Min	0.311	0.134	0.122	0.110	0.065
	PZ	0.953	0.566	0.346	0.269	0.202
	AJ(threshold)	0.563	0.107	0.049	0.034	0.014
	AJ(power var)	0.680	0.368	0.298	0.269	0.310
	BNS	0.754	0.213	0.137	0.153	0.118
GE	CPR	0.908	0.331	0.193	0.196	0.141
	JO	0.317	0.184	0.194	0.199	0.270
	ABD-LM	0.955	0.461	0.259	0.186	0.098
	Med	0.275	0.140	0.128	0.146	0.115
	Min	0.274	0.109	0.092	0.093	0.078
	PZ	0.951	0.510	0.319	0.259	0.194
	AJ(threshold)	0.553	0.086	0.033	0.032	0.016
	AJ(power var)	0.595	0.370	0.323	0.290	0.300
	BNS	0.840	0.327	0.196	0.179	0.135
DIS	CPR	0.923	0.423	0.263	0.217	0.151
	JO	0.385	0.241	0.227	0.246	0.302
	ABD-LM	0.978	0.541	0.271	0.179	0.101
	Med	0.466	0.184	0.188	0.188	0.150
	Min	0.431	0.163	0.118	0.115	0.073
	PZ	0.974	0.595	0.370	0.305	0.209

Table 10. Proportion of jumps identified by different combinations of sampling frequencies and procedures for IBM

Procedure	$BNS5 \cap BNS10 \cup$ $BNS10 \cap BNS15$	$(CPR5 \cap CPR10) \cup$ $(CPR10 \cap CPR15)$	$ABDLM5 \cap ABDLM10) \cup$ $(ABDLM10 \cap ABDLM15)$	$(Med5 \cap Med10) \cup$ $(Med10 \cap Med15)$
Proportion	0.105	0.173	0.258	0.098
Procedure	$(Min5 \cap Min10) \cup$ $(Min10 \cap Min15)$	$(PZ5 \cap PZ10) \cup$ $(PZ10 \cap PZ15)$	$(JO5 \cap JO10) \cup$ $(JO10 \cap JO15)$	
Proportion	0.055	0.327	0.148	
Procedure	$(Med10 \cap ABDLM10) \cup$ $(ABDLM10 \cap BNS10)$	$(CPR10 \cap BNS10) \cup$ $(BNS10 \cap Med10)$	$(CPR10 \cap ABDLM10) \cup$ $(ABDLM10 \cap PZ10)$	$(BNS10 \cap PZ10) \cup$ $(PZ10 \cap Med10)$
Proportion	0.132	0.193	0.222	0.213

Table 11. Relative contribution of the continuous part of the price process (RV_C) and of jumps (RV_J) to the total quadratic variation for IBM

Panel A

Year	Procedure	AJ(threshold)	AJ(power var)	BNS	CPR	JO	ABD-LM	Med	Min	PZ
2005	RV_C	72.8	74.9	90.3	82.7	83.5	79.9	90.7	92.4	79.8
	RV_J	27.2	25.1	9.7	17.3	16.5	20.1	9.3	7.6	20.2
2006	RV_C	67.1	68.2	81.3	77.4	81.0	74.9	81.9	83.7	74.2
	RV_J	32.9	31.8	18.7	22.6	19.0	25.1	18.1	16.3	25.8
2007	RV_C	77.3	78.5	96.1	88.4	90.1	82.3	94.6	97.0	82.0
	RV_J	22.7	21.5	3.9	11.6	9.9	17.7	5.4	3.0	18.0
2008	RV_C	78.3	78.4	94.6	90.1	91.2	87.7	93.4	95.9	88.0
	RV_J	21.7	21.6	5.4	9.9	8.8	12.3	6.6	4.1	12.0
2009	RV_C	72.6	74.4	89.6	83.9	86.6	81.7	88.3	92.2	79.4
	RV_J	27.4	25.6	10.4	16.1	13.4	18.3	11.7	7.8	20.6

Panel B

Year	Proc	$(BNS5 \cap BNS10(\cup$ $(BNS10 \cap BNS15)$	$(ABDLM5 \cap ABDLM10) \cup$ $(ABDLM10 \cap ABDLM15)$	$(Med5 \cap Med10) \cup$ $(Med10 \cap Med15)$	$(PZ5 \cap PZ10) \cup$ $(PZ10 \cap PZ15)$	$(Med10 \cap ABDLM10) \cup$ $(ABDLM10 \cap BNS10)$	$(BNS10 \cap PZ10) \cup$ $(PZ10 \cap Med10)$
2005	RV_C	95.71	82.19	93.55	81.43	90.51	89.92
	RV_J	4.29	17.81	6.45	18.57	9.49	10.08
2006	RV_C	89.17	76.15	88.08	75.04	81.89	80.66
	RV_J	10.84	23.85	11.92	24.97	18.11	19.34
2007	RV_C	97.48	83.03	97.12	82.80	95.67	94.46
	RV_J	2.52	16.97	2.88	17.20	4.33	5.54
2008	RV_C	97.31	88.38	96.03	88.82	94.27	92.58
	RV_J	2.69	11.62	3.97	11.18	5.73	7.42
2009	RV_C	91.48	82.51	91.91	81.11	88.77	86.54
	RV_J	8.52	17.49	8.09	18.89	11.23	13.47

List of Figures

Figure 1. Size based on simulated and asymptotic critical values for the SV1F model with noise variance $\sigma_{noise} = .052$ and for different significance levels: from left to right: 5%, 1%, .1%

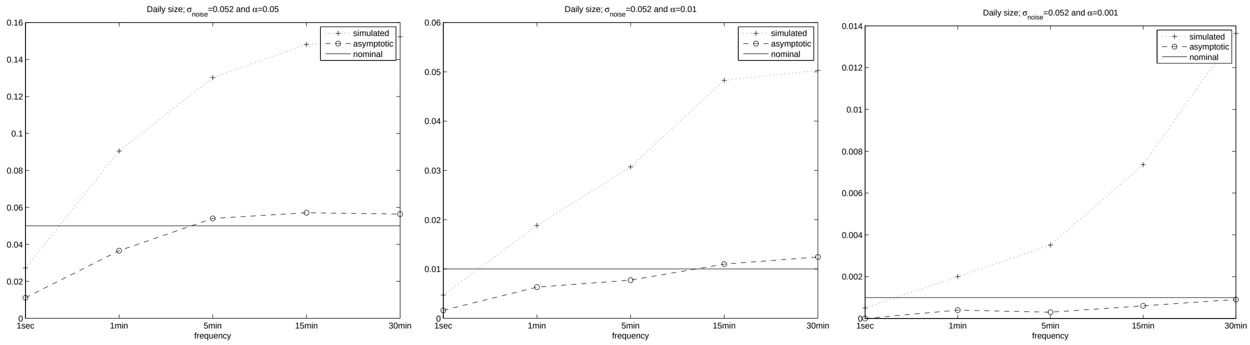

Figure 2. Size adjusted power based on simulated and asymptotic critical values for the SV1F model with jumps in the presence of noise. Significance levels: 5%, 1%, .1%