Testing for jumps in GARCH models, a robust approach

Sébastien Laurent[*]   Christelle Lecourt[†]   Franz C. Palm[‡]

March 29, 2013

Abstract

Financial series occasionally exhibit large changes. To deal with those events, we assume that the observed return series consists of a conditionally Gaussian ARMA-GARCH (or -GJR) model contaminated by an additive jump component. In this framework, we propose a new test for additive jumps. The test is based on standardised returns, where the first two conditional moments of the non-contaminated observations are estimated in a robust way. Simulation results indicate that the test has very good finite sample properties, i.e. correct size and high proportion of correct jump detection. We apply our test on daily returns and detect less than 1% of jumps for the three exchange rates and between 1 and 3% of jumps for about 50 large capitalization stocks from the NYSE. Once jumps have been filtered out, all series are found to be conditionally Gaussian. We also find that simple GARCH-type models estimated on filtered returns deliver better out-of-sample forecasts of the conditional variance than GARCH and GAS models estimated on raw data.

[*]Maastricht University, School of Business and Economics, The Netherlands and CORE, Belgium. E-mail: s.laurent@maastrichtuniversity.nl

[†]CeReFim, Université de Namur, Belgium and Maastricht University, School of Business and Economics, The Netherlands. E-mail: christelle.lecourt@fundp.ac.be

[‡]Maastricht University, School of Business and Economics, The Netherlands. E-mail: f.palm@maastrichtuniversity.nl

1   Introduction

The distributional properties of speculative prices have been extensively studied in the finance literature. High frequency returns of most financial assets exhibit volatility clustering and large jumps caused by the arrival of important surprising news (e.g. news announcements). Daily stock prices are often treated as discretely sampled observations from a jump-diffusion process with time-varying volatility (often assumed to follow a GARCH-type model). Mixtures of normal distributions have been used to account for the effect of jumps. For instance, Ball and Torous (1983) and Jorion (1988) use jump processes to study the impact of big shocks in common stock prices or foreign exchange rates while Vlaar and Palm (1993) propose Poisson or Bernoulli mixtures of normal distributions. Such a parametric approach requires one to further specify and estimate a model governing the time-varying jump intensity and also the jump size. Notice that those papers aim at accounting for jumps, not at testing for jumps.

In this paper, we propose a new semi-parametric statistical procedure to detect additive jumps in financial series. It is similar to the non-parametric tests for jumps proposed by Lee and Mykland (2008) and Andersen, Bollerslev, and Dobrev (2007b). However, in many cases, those tests are not applicable because they require intraday data which are not always available and, if they are, the asset might not be liquid enough for the test to be applicable. In those cases a test for jumps based on daily returns is needed because, as shown by Andersen, Bollerslev, and Diebold (2007), the largest shocks have a relatively smaller effect on future volatility than smaller shocks. Therefore, GARCH models neglecting jumps usually overestimate the volatility during several days, if not several weeks, after the occurrence of those jumps.[1]

Similar to Lee and Mykland (2008) and Andersen, Bollerslev, and Dobrev (2007b), who standardize their non-parametric test statistics by a robust to jumps estimate of instantaneous volatility (based on realized bipower varia-

---

[1]See for instance the example of the stock price of Apple in Boudt, Daníelsson, and Laurent (2013).

tion), we standardize our test statistic using the conditional volatility based on a robustified GARCH volatility estimate and a robust conditional mean estimate as daily data will likely have a non-zero mean. Our test therefore incorporates the idea that when spot or instantaneous volatility is high (also in absence of jumps), returns may also be high, even as high as those due to jumps.

Franses and Ghijsels (1999) proposed a sequential test for jumps in a GARCH framework. However, their critical values have to be simulated and depend on some unknown parameter values of the GARCH model. This test suffers from the so-called outlier masking problem because it is based on a Quasi-Maximum Likelihood estimate of the GARCH model, which is know to be non-robust to additive jumps (see our Monte-Carlo simulation). In our case, critical values need not be simulated as the asymptotic distribution of the test does not depend on nuisance parameters, its type-I error can be controlled for and Monte-Carlo simulation results suggest that it is more powerful.

Our test builds upon the same theoretical framework as, for instance, Franses and Ghijsels (1999), Lee and Mykland (2008) and Andersen, Bollerslev, and Dobrev (2007b), i.e. that returns are conditionally Gaussian on days without jumps. This assumption is however incompatible with, for instance, the GARCH model with Student-t innovations (with finite degree of freedom) of Bollerslev (1987), which is another way of accounting for the presence of large unexpected shocks.[2] To overcome the problem of overestimating volatility after the occurrence of extremely large shocks caused by once-off events, Creal, Koopman, and Lucas (2012) and Harvey and Chakravarty (2008) independently proposed a novel way to deal with large returns in conditional variance models. Their models rely on a potentially non-normal distribution for the innovations and a GARCH-type equation for the conditional variance derived from the conditional score of the assumed distribution with respect to the second moment. Therefore, they are called respectively Generalized Au-

---

[2]GARCH-type models have been successfully combined with several fat-tail distributions. Giot and Laurent (2003) show, for several assets, that GARCH-type models with skewed-Student innovations outperform models based on normal and Student-t distributions when forecasting the Value-at-Risk.

toregressive Score (GAS) and Dynamic Conditional Score (DCS) models for volatility.

Whether data are compatible with our theoretical framework or the one of GARCH and GAS/DCS models with non-normal distributions is an empirical question that we investigate in a comparison using simulated data and in an application on 3 exchange rates and about 49 US stocks. If jumps are detected by our test, we propose to filter out those jumps before forecasting volatility using standard GARCH models. Results suggest that simple GARCH-type models estimated on filtered returns deliver better out-of-sample forecasts of the true variance (approximated by the realized variance computed from 5-minute returns) than GARCH and GAS models estimated on raw data.

Interestingly, once jumps have been filtered out, all series are found to be conditionally Gaussian. Importantly, a rejection of the conditional normality of the jump-corrected data would be taken as an indication that our theoretical framework is inappropriate. To investigate this issue, we simulated data following a GAS model with Student-t innovations. This model is able to generate volatility clustering and large once-off events that might be indistinguishable from additive jumps. However, this model does not contain additive jumps and therefore our test should not be applied in this case. Results suggest that while a small proportion of jumps is detected by our test, a Jarque-Bera test has very good power (greater than 95% for $v \leq 7$) to reject the assumption of conditional normality of the jump-corrected data.

The rest of this article is organized as follows. In Section 2, the theoretical framework and the model setting are described and the proposed semi-parametric test and its asymptotic distribution are presented. In Section 3, simulation results comparing our test with alternative tests in the literature are presented. Section 4 contains the findings of an empirical study of jump detection in daily data. Finally, Section 5 concludes the paper.

2   Model and test

Andersen, Bollerslev, and Diebold (2007), Harvey and Chakravarty (2008)

and Muler and Yohai (2008) among others found that jumps affect future asset return volatility less than what standard return volatility models predict. Using realized volatilities Andersen, Bollerslev, and Diebold (2007) show that conditioning also on past jumps in an autoregressive (AR) model tempers the persistence in the volatility forecasts, indicating that jumps in asset prices tend to lead the short-lived increases in volatility. In a univariate GARCH setting, Sakata and White (1998), Franses and Ghijsels (1999), Carnero, Pena, and Ruiz (2007, 2008), Charles and Darné (2005) and Muler and Yohai (2008) show that, in the presence of additive jumps, the Gaussian Quasi-Maximum Likelihood (QML) estimator of GARCH models tends to overestimate the volatility for the days following a jump and to produce upward biased estimates of the long-run volatility.[3]

In a recent paper, Nyberg and Wilhelmsson (2009) decompose the risk measure Value-at-Risk (Var) of a portfolio of stocks into a jump and a continuous component. The continuous component is associated with market risk and the jump component is assumed to measure (extreme) event risk. Using the parametric Normal Inverse Gaussian (NIG) distribution combined with the GARCH-Poisson-Jump model with autoregressive jump intensity (GARJI) model proposed by Maheu and McCurdy (2004), Nyberg and Wilhelmsson (2009) find that event risk constitutes 30% of the risk for a portfolio of small cap stocks but less that 1% for a portfolio of large cap stocks.

These empirical findings about the nature of the impact of jumps in financial asset prices stress the importance of reliable statistical methods for jump detection. Relying on seminal work by Andersen and Bollerslev (1998b) on realized volaltility and by Barndorff-Nielsen and Shephard (2004, 2006)

---

[3]The effects of jumps on multivariate GARCH models have also been investigated by Boudt and Croux (2010) and Boudt, Daníelsson, and Laurent (2013), respectively in BEKK and dynamic conditional correlation (DCC) frameworks. Boudt, Daníelsson, and Laurent (2013) show that unconditional and conditional correlations from the constant conditional correlation (CCC) model of Bollerslev (1990) and the DCC model of Engle (2002) are strongly affected by the presence of jumps. They also find that conditional covariance forecasts obtained from various multivariate models, including the DCC model, are frequently outperformed by their jump-robust version based on ex post realized covariance estimates from high-frequency EUR-USD and YEN-USD exchange rates over the period 2004-2009.

on bipower variation, an important segment of the literature uses realized moments from high frequency data to estimate the jump component by substracting realized bipower variation from realized volatility and to provide non-parametric jump detection statistics for testing its significance at a lower observation frequency (see e.g. Andersen, Bollerslev, Frederiksen, and Nielsen, 2010). Multipower extensions have been considered by Barndorff-Nielsen and Shephard (2006) and Huang and Tauchen (2006). Specification tests for volatility processes in the presence of jumps, (e.g. Lee and Mykland, 2008; Aït-Sahalia and Jacod, 2004, 2009) are indirect ways of checking for jumps. Aït-Sahalia and Jacod (2009) provide a test for semi-martingales which has different limits depending on the presence or absence of jumps, as the sampling frequency decreases. Aït-Sahalia and Jacod (2012b) provide a version of the previous test that is robust to the presence of market microstructure noise in high-frequency data. Finally, Aït-Sahalia and Jacod (2012a) propose a simple yet powerful methodology to decompose high frequency returns into drift, continuous, small jumps and large jumps components and to analyse their characteristics such as their relative magnitude and the degree of activity of the jumps.

2.1 Data generating process

One of the most popular model for financial return data sampled at say the daily frequency is certainly the GARCH(1,1) of Bollerslev (1986).

A random variable r_t follows a normal-ARMA(p,q)-GARCH(1,1) model if it can be described by the system (2.1)-(2.3):

$$\phi(L)(r_t - \mu) = \theta(L)\varepsilon_t \text{ or } r_t = \mu_t + \varepsilon_t \tag{2.1}$$

$$\varepsilon_t \equiv \sigma_t z_t \text{ and } z_t \overset{i.i.d.}{\sim} N(0,1) \tag{2.2}$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{2.3}$$

where L is the lag operator, $\phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$ and $\theta(L) = 1 - \sum_{i=1}^{q} \theta_i L^i$ are polynomials of orders p and q respectively (with roots outside the unit circle) such that $\mu_t = \mu + \sum_{i=1}^{\infty} \lambda_i \varepsilon_{t-i}$ is the conditional mean of r_t (where the λ_i's are the coefficients of $\lambda(L) = \phi^{-1}(L)\theta(L) = 1 + \sum_{i=1}^{\infty} \lambda_i L^i$) and σ_t^2 is the

conditional variance of r_t.

If we add an independent jump component $a_t I_t$ to r_t with jump size a_t, we obtain

$$r_t^* = r_t + a_t I_t, \tag{2.4}$$

where r_t^* denotes observed financial returns and I_t is generated by some jump process such as a Poisson process. Notice that more complicated mixture distributions could be considered to generate the jumps and the 'continuous' components. As we focus on the testing for jumps and eliminating them, our procedure will not suffer from the difficulty of getting reliable estimates for the jump process from which mixture models suffer when only few jumps have occurred.

The model for r_t^*, resulting from (2.1)-(2.4), is a combination of an approximation of a smooth and slowly reverting continuous sample path process and a much less persistent jump component, that according to Andersen, Bollerslev, Frederiksen, and Nielsen (2010) best describes many (log-)return processes. It is also an approximation of the decomposition of asset returns sampled at high frequency into their base components (continuous, small jumps, large jumps) proposed by Aït-Sahalia and Jacod (2012a). The model for r_t^* has the properties that a jump $a_t I_t$ will not affect σ_{t+1}^2 (the conditional variance of r_{t+1}), and it allows for non-Gaussian fat-tailed conditional distributions of r_t^*.

2.2 Jump detection tests

A popular method for additive jumps detection in a Data Generating Process satisfying Equations (2.1)-(2.4) is the test proposed by Franses and Ghijsels (1999). They adapt the procedure of Chen and Liu (1993) for additive outlier detection in ARMA models to make it applicable for our DGP. To test $H_0 :$ $a_t I_t = 0$ for $t = \tau$, they compute a sequence of t-statistics for $t = 1, \ldots, T$, by running regressions of residuals $\hat{v}_t = (r_t^* - \hat{\mu}_t)^2 - \hat{\sigma}_t^2$ (where $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ are the QML estimates of μ_t and σ_t^2 obtained from r_t^*) on a quantity measuring the impact of a potential jump $a_\tau I_\tau$ at time t on $\hat{v}_{\tau+s}$, $s \geq 0$, and compare the maximum of the sequence of t-statistics with a critical value obtained by

simulation. See Appendix 1 for more details on Franses and Ghijsels' test for jumps.

Franses and Ghijsels (1999) propose to estimate $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ by Gaussian QML which is not robust to additive jumps of the type (2.4) and therefore their test is likely to suffer from the so-called outlier masking problem (as illustrated in our Monte-Carlo experiment). Another drawback of this test is that the critical values depend on the unknown parameters α_1 and β_1 of the GARCH model and therefore the size of the test cannot be controlled.

Our test radically differs from the one described above and does not suffer from those drawbacks. It is similar in spirit to the non-parametric tests for jumps proposed simultaneously by Andersen, Bollerslev, and Dobrev (2007b) and Lee and Mykland (2008) for high-frequency data.

Let us denote by $\tilde{\mu}_t$ and $\tilde{\sigma}_t^2$ estimates of μ_t and σ_t^2 in model (2.1)-(2.4) that are robust to the potential presence of the additive jumps $a_t I_t$ (i.e. estimated on r_t^* and not r_t). Denote by

$$\tilde{J}_t = \frac{r_t^* - \tilde{\mu}_t}{\tilde{\sigma}_t}, \qquad (2.5)$$

the standardised return on day t. If $a_t I_t = 0$ on day t, $\tilde{J}_t$ follows a standard normal distribution and thus standardised returns $\tilde{J}_t$ that are too large to plausibly come from this distribution must reflect jumps.[4]

This suggests the following jumps detection rule:

$$\tilde{I}_t = I\left(|\tilde{J}_t| > k\right), \qquad (2.6)$$

where $I(\cdot)$ is the indicator function and k is a suitable critical value defined below. The rule described in (2.6) implies that $\tilde{I}_t = 1$ when a jump is detected at observation t and $\tilde{I}_t = 0$ otherwise. $\tilde{I}_t$ is thus an estimate of the unknown quantity I_t in Equation (2.4). Given $\tilde{\mu}_t$ and $\tilde{I}_t$, detected jumps can be filtered

---

[4]Note that since one single very large jump can artificially blow up the conditional variance during several weeks after its occurrence, it is crucial to use jump-robust estimators of both μ_t and σ_t^2.

out from r_t^* as follows

$$\tilde{r}_t = r_t^* - (r_t^* - \tilde{\mu}_t)\tilde{I}_t, \tag{2.7}$$

where filtered returns $\tilde{r}_t$ should be conditionally Gaussian under Model (2.1)-(2.4). Note here the crucial role of the assumption of conditional normality of r_t. However nothing guarantees to find conditional normality of the filtered returns after applying our jump removing procedure on real data. We will show in the section devoted to the Monte-Carlo simulation that a simple Jarque-Bera test of conditional normality of $\tilde{r}_t$ (i.e. test of normality on the standardized residuals of a normal-ARMA-GARCH model estimated on $\tilde{r}_t$) has very good power to reject conditional normality (e.g. when the DGP has no jumps but Student-t innovations) and thereby to reject model (2.1)-(2.4). But when we do not reject conditional normality of $\tilde{r}_t$, the loop is closed and the procedure is consistent with the data which assume the continuous component to follow a Gaussian Brownian motion.

A straightforward jump detection rule is that return r_t^* is taken as being affected by a jump if $|\tilde{J}_t|$ exceeds the $1 - \lambda$ quantile ($> .5$) of the standard Gaussian distribution. This rule has a probability of type I error of λ (i.e. of detecting that r_t^* is affected by jumps, if in reality $r_t^* = r_t, \forall t = 1 \dots, T$). But its disadvantage is that the expected number of false positives over the whole estimation sample is equal to λT under the null of no jump which can be large when the sample size is large. For instance, with $T = 1000$ and $\lambda = 0.05$, 50 spurious jumps are expected under the null of no jump. Lee and Mykland (2008) call these false positives "spurious jump detections".

Andersen, Bollerslev, and Dobrev (2007b) use a Bonferroni correction to control for the number of spurious jumps detected. This corresponds to choosing a higher quantile of the standard normal distribution, e.g. $\lambda = 0.001$ or 0.0001. Instead, we propose to follow Lee and Mykland (2008) and control for the size of the multiple jump tests using the extreme value theory result that the maximum of T i.i.d. realizations of the absolute value of a standard normal random variable is asymptotically (for $T \to \infty$) Gumbel distributed. More specifically, in the absence of jumps, the probability that the maximum

9

of any set of T independent J-statistics $|\tilde{J}_t|$ exceeds

$$g_{T,\lambda} = -\log(-\log(1-\lambda))b_T + c_T, \qquad (2.8)$$

with $b_T = 1/\sqrt{2\log T}$ and $c_T = (2\log T)^{1/2} - [\log \pi + \log(\log T)]/[2(2\log T)^{1/2}]$, equals λ. Therefore, following (2.6), all returns for which $|\tilde{J}_t| > g_{T,\lambda}$ should be declared as being affected by jumps.

As mentioned above, our statistic for jumps is similar to the non-parametric tests of Andersen, Bollerslev, and Dobrev (2007b) and Lee and Mykland (2008) but our estimates of the conditional moments differ because our test is designed for data sampled at a daily or lower frequency while their test is applied on intraday-data (e.g. 5-minute returns). The price to pay is therefore that $\tilde{\mu}_t$ and $\tilde{\sigma}_t^2$ have to be estimated parametrically.

For the variance, they rely on the (non-parametric) bipower variation estimator of Barndorff-Nielsen and Shephard (2004) computed on returns belonging to a local window around the tested time (e.g. they assume that the spot volatility is constant during a period ranging from a few minutes to about one day). Obviously one cannot make similar assumptions on daily data, reason why a parametric model is needed to estimate the first two conditional moments "in a robust way".

For the conditional mean $\tilde{\mu}_t$, Andersen, Bollerslev, and Dobrev (2007b) and Lee and Mykland (2008) use the trivial estimator $\tilde{\mu}_t = 0 \; \forall t$ because at high frequencies, the drift is negligible. In our case the problem lies in the fact that μ_t is not necessarily 0, follows an ARMA model as in Equation (2.1) and importantly that this model is specified for the non-contaminated observations r_t while one observes the (potentially) contaminated series r_t^*. A robust estimate of $\tilde{\mu}_t$ is therefore needed.

Muler, Pena, and Yohai (2009) (MPY) introduce a new class of estimates for ARMA models that is robust to additive jumps. They downweight the effect of past values of r_t^* in the infinite MA representation of the conditional mean of the ARMA model. This leads to the following specification to approximate

the conditional mean of r_t^*:[5]

$$\tilde{\mu}_t = \mu + \sum_{i=1}^{\infty} \lambda_i w_{k_\delta}^{MPY}(\tilde{J}_{t-i}), \qquad (2.9)$$

where $\tilde{J}_{t-i}$ is given in Equation (2.5), while $\tilde{\sigma}_{t-i}$ will be defined below.

The weight function $w_{k_\delta}^{\mathrm{MPY}}(\cdot)$ in Equation (2.9) plays a key role in the robustification of the ARMA model. To obtain robust and efficient estimates of the ARMA coefficients, MPY show that $w_{k_\delta}^{\mathrm{MPY}}(\cdot)$ needs to be bounded. More specifically, they propose the following weight function

$$w_{k_\delta}^{\mathrm{MPY}}(u) = \operatorname{sign}(u) \min\left(|u|, k_\delta\right). \qquad (2.10)$$

The auxiliary model underlying (2.9) with weight function (2.10) is called *Bounded Innovation Propagation* (BIP)–ARMA since the effect of jumps (i.e. $a_t I_t$) on future values of μ_t is bounded.

Using the same argument as above, $\tilde{J}_{t-i}$ follows a standard normal distribution in absence of jumps at time $t - i$. It is natural to suspect the presence of a jump in r_{t-i}^* when $|\tilde{J}_{t-i}|$ exceeds k_δ, the δ quantile of the standard normal distribution. Typical values for δ are 0.95 and 0.975. Note that we expect $T(1 - \delta)$ residuals in each sample of size T to be downweighted even if there is no jump.[6]

A similar idea is used by Muler and Yohai (2008) (MY) to limit the effect of $a_t I_t$ on the estimation of the parameters of the GARCH model. In this case the Gaussian QML is not appropriate because $a_{t-1} I_{t-1}$ has no impact on σ_t^2 in Equation (2.3) while assuming a GARCH(1,1) for r_t^* would imply (if for simplicity $\mu_t = 0$) $\sigma_t^2 = \omega + \alpha_1(r_{t-1} + a_{t-1}I_{t-1})^2 + \beta_1\sigma_{t-1}^2$, i.e., a large and slowly decaying effect of $a_{t-1}I_{t-1}$ on future volatility predictions.

---

[5]Note that in MPY, $\tilde{\sigma}_{t-i}$ is assumed to be constant and replaced by a robust M-scale estimate of ε_t.

[6]An alternative would be to compare $|\tilde{J}_{t-i}|$ with the critical value of the Gumbel distribution (2.8) like in (2.6). This would naturally lead to a smaller number of observations that are falsely downweighted. However, Monte-Carlo simulation results (not reported here to save space) suggest that, for the purpose of estimating the parameters, downweighting too many observations is less damageable for the efficiency of this method than neglecting small jumps.

MY propose the following auxiliary GARCH(1,1) model with weights on extremes:

$$\tilde{\sigma}_t^2 = \omega + \alpha_1 \tilde{\sigma}_{t-1}^2 c_\delta w_{k_\delta}^{\text{MPY}} \left(\tilde{J}_{t-1} \right)^2 + \beta_1 \tilde{\sigma}_{t-1}^2. \tag{2.11}$$

The factor c_δ has been proposed by Boudt, Daníelsson, and Laurent (2013) to ensure the conditional expectation of the weighted squared unexpected shocks to be the conditional variance of r_t in absence of jumps. Note that $c_\delta = 1.0185$, 1.0465, 1.0953, 1.2030 respectively for $\delta = 0.99, 0.975, 0.95$, and 0.90.

Equation (2.11) corresponds to Equation (**??**) with $w(z_{t-1}) = c_\delta w_{k_\delta}^{\text{MPY}} \left(\tilde{J}_{t-1} \right)^2$ and is called *Bounded Innovation Propagation* (BIP)–GARCH(1,1).

As for the BIP-ARMA, (squared) residuals that are suspected to be contaminated by additive outliers are downweighted in the BIP-GARCH equation. Again, typical values for δ are 0.95 and 0.975.

Extensions of the BIP–GARCH to higher GARCH orders or other GARCH-type specifications (e.g. FIGARCH) are possible but not discussed in detail here to save space. The only extension we consider is the BIP–GJR(1,1), i.e. the robust version of the GJR(1,1) model of Glosten, Jagannathan, and Runkle (1993) that accounts for the so-called leverage effect, i.e.

$$\tilde{\sigma}_t^2 = \omega + \alpha_1 \tilde{\sigma}_{t-1}^2 c_\delta w_{k_\delta}^{\text{MPY}} \left(\tilde{J}_{t-1} \right)^2 + \gamma_1 D_{t-1} \tilde{\sigma}_{t-1}^2 c_\delta w_{k_\delta}^{\text{MPY}} \left(\tilde{J}_{t-1} \right)^2 + \beta_1 \tilde{\sigma}_{t-1}^2, \tag{2.12}$$

where $D_{t-1} = 1$ if $\tilde{J}_{t-1} < 0$ and 0 otherwise.

For a comparison between the GARCH(1,1), GAS(1,1) with (Skewed) Student-t innovations and the BIP-GARCH(1,1) model, see Figure 1. To make the comparison with the other models, $c_\delta w_{k_\delta}^{\text{MPY}} (z_t)^2 - 1$ is plotted as a function of z_t for $z_t \in (-5; 5)$. Figure 1 suggests that the downweighting mechanism in a GAS model with ST(0,1,5) innovations is very similar to the one of a BIP-GARCH specification with $\delta = 0.975$.

MPY and MY show respectively that QML estimation of a BIP–ARMA model with constant variance and a BIP–GARCH model with zero conditional mean are not efficient in the presence of large outliers (jumps). They recommend using a M–estimator that minimizes the average value of an objective function $\rho(\cdot)$, evaluated at the log–transform of squared standardised returns,

i.e. in our case

$$\hat{\theta}^{\mathrm{M}} = \mathrm{argmin}_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} \rho \left(2 \log \left| \frac{r_t^* - \tilde{\mu}_t}{\tilde{\sigma}_t} \right| \right), \qquad (2.13)$$

where $\tilde{\mu}_t$ and $\tilde{\sigma}_t^2$ are given respectively in (2.9) and (2.11).

For robustness, this ρ–function needs to downweight the extreme observations and hence the jumps. The choice of $\rho(\cdot)$ trades off robustness vs. efficiency. MY recommend $\rho_1(z) = 0.8m(g_0(z)/0.8)$, where the m–function is a smoothed version of $m_1(x) = xI(x \leq 4) + 4I(x > 4)$ and $g_0(z) = \frac{1}{\sqrt{2\pi}} \exp[-(\exp(z) - z)/2]$. Based on a comparison of several candidate ρ–functions for (2.13), Boudt, Daníelsson, and Laurent (2010, 2013) recommend the one associated with the Student-t density function with 4 degrees of freedom (t_4):

$$\rho_{t_4}(z) = -z + 0.8260 \, (1 + \nu) \log \left(1 + \frac{\exp(z)}{\nu - 2} \right).$$

It is important to highlight that we do not assume the innovations to follow a Student-t_4 but assume that in absence of jumps $z_t \sim N(0, 1)$. The t_4 is only used in (2.13) to downweight the extreme observations and hence the jumps in the objective function (2.13).

To sum up, we perform the estimation of the BIP–ARMA-BIP–GARCH (or GJR) model in one step by minimising the objective function (2.13) with $\delta = 0.975$ in the weight function $w_{k_\delta}^{\mathrm{MPY}}(\cdot)$ and $\rho(\cdot) = \rho_{t_4}(\cdot)$. We denote by $\tilde{\mu}_t$ and $\tilde{\sigma}_t^2$ the robust estimates of μ_t and σ_t^2 obtained by this method. They are used in Equation (2.6) with critical value k given by (2.7) to test the presence of jumps in the sample. Then, jumps (if detected) are filtered out from the returns using (2.8) to finally obtain $\tilde{r}_t$.

For the M-estimators for GARCH models which minimize the objective function in (2.13), MY have shown consistency for stationary GARCH-processes. Normality of the data is not required. These M-estimates are less sensitive to outliers than the QML-estimate and they satisfy Huber (1981)'s first requirement for a robust estimate, that is the estimate should be highly efficient when the observations are not subject to outliers. MY propose a modification of the

M-estimator, called bounded M-estimator (BM). The BM-estimator includes an additional mechanism that bounds the propagation of the effect of an outlier on the subsequent predictions of the conditional variance. The BM-estimator is also consistent and asymptotically normally distributed. In addition to satisfying Huber (1981)'s first requirement for M-estimators, it also satisfies his second requirement that replacing a small fraction of observations by outliers should produce a small change in the estimator. Therefore, as shown by MY, the BM-estimator has a high efficiency. In view of their findings, the second robust method that we propose is expected to be more efficient that our first method.

MPY propose robust (M-) estimates for ARMA models. On p. 826, they write '*We conjecture that similar results, consistency and asymptotic normality, hold when the observations follow a BIP-ARMA model.*' Similar properties are expected to be found for the BIP-GARCH process. They would underpin the proposed use of an ARMA-GARCH model for filtered return.

In the next section we shall provide simulation results for our test and compare it with the test procedure of Franses and Ghijsels (1999) and with the GAS approach.

3 Simulation

3.1 Data Generating Processes

In the Monte-Carlo simulation we simulate 1000 samples of size $T = 2000$ following a normal-AR(1)-GARCH(1,1) model with additive jumps as described in Equations (2.1)-(2.4), with $p = 1$ and $q = 0$, $\mu = 0.05, \phi_1 = 0.3, \omega = 0.05$, and $\alpha_1 = 0.02, 0.03, \ldots, 0.1$ and $\beta_1 = 0.95 - \alpha_1$.[7]

The size of the jump process a_t in Equation (2.4) is specified as follows:

$$a_t = \text{sign}(r_t) m \sigma_t, \tag{3.1}$$

---

[7]Similar results have been obtained for other sample sizes, i.e. $T = 500, 1000$ and 3000 but results are not reported to save space.

i.e. m times the conditional standard deviation of r_t (i.e., σ_t), where m takes any integer value between 0 and 6 to simulate very small jumps to large jumps. Note that either $m = 0$ or $I_t = 0$ $\forall t$ correspond to the case of no jump.

For the dummy variable I_t determining the arrival time of the jumps, we consider either a Poisson distribution with constant intensity or fixed the arrival times ex-ante such that jumps are equidistant and do not happen at the very beginning or the end of the sample. Results being similar we only report those for the equidistant jumps in order to save space. The number of jumps per sample of T observations is set to $1, 2, 5, 10, 20$ or 40 respectively.

3.2   Global spurious detection of jumps

Table 1 reports the rejection frequencies of the null that there is no jump in a sample of T observations, i.e.

$$H_0 : a_t I_t = 0 \ \forall t \text{ for } t = 1, \dots, T,$$

for the nine combinations of α_1 and β_1 parameter values. The percentage of global spurious detection under the null of no jump (type I error) is expected to be close to the nominal size λ.

Column labelled FG corresponds to Franses and Ghijsels' test while those labelled LLP correspond to our test. The value in parenthesis is the nominal size of the test. We consider $\lambda = 5\%$ for both tests but also $\lambda = 25\%$ and 50% for our test. Note that values of $\lambda = 25\%$ and 50% imply that one spurious jump is expected to be found respectively every fourth and every second sample of size T.

The critical values (C) of Franses and Ghijsels' test have been obtained by Monte-Carlo under the unrealistic assumption that α_1 and β_1 are known.[8]

---

[8]The main drawback of their approach is that the critical values depend on T (which is known) but also on unknown parameters (α_1 and β_1 in the GARCH(1,1) case) with the undesirable consequence that on real data one cannot control the type I error (false detections). Using the true values of α_1 and β_1, the simulated critical values for C we obtained for a sample of $T = 2000$ observations and a nominal size of 5% are 12.2373, 12.8625, 13.2743, 14.064, 14.7861, 15.7598, 16.5336, 18.2659, 18.7718, respectively for the nine combinations of α_1 and β_1 parameters considered in the Monte-Carlo experiment.

Recall that critical values for our test are obtained using Equation 2.8.

Results suggest that both tests do not suffer from large size distortions irrespectively of the values of α_1 and β_1.[9]

3.3   Ability to detect actual jumps

Another issue of interest is the power of the test. We define the proportion of correct (resp. false) jump detections as the average number (over the 1000 replications) of correctly (falsely) detected jump days.

Figure 2 plots the proportion of correct jump detections as a function of m (proportionality factor of the magnitude of jumps) for $T = 2000$. The critical level of the tests is set to 5%. Recall that jumps are equally spaced and the number of jumps per sample equals $1, 2, 5, 10, 20$ or 40.

This figure clearly suggests that our test (bottom) has a much higher power to detect the actual jumps than Franses and Ghijsels's (1999) test. For instance, the proportions of correct jump detection in presence of 20 additive jumps of magnitude 4 and 5 conditional standard deviations equal respectively 18.29% and 55.17% for Franses and Ghijsels's (1999) test while they are equal to 71.12% and 99.67% for our test (at the 5% level).

Furthermore, it emerges from these figures that unlike Franses and Ghijsels's (1999) test, our test is not sensitive to percentage of jumps in the sample. Indeed, the proportion of correct jump detections of Franses and Ghijsels's (1999) test declines sharply with the number of jumps in the sample and eventually tends to zero when the number of jumps is sufficiently large (problem known in the robust statistics literature as outlier masking as in the presence of jumps the estimated standard-errors are large compared to the estimate of a_t rendering the test insignificant).

The power of the test is of course a function of its critical level. However, as explained in the previous sub-section, the expected number of spurious jumps in a sample of T observations equals λ and is therefore necessarily smaller than 1 (irrespectively of the sample size).

---

[9]Results reported in this paper were obtained using the programming language Ox version 6.0 (Doornik, 2009) and G@RCH version 6.0 (Laurent, 2009).

Figure 3 plots the power of our proposed test for $1, 2, 10$ or 40 jumps for the following critical levels: $5, 25$ and 50%. This figure suggests that power indeed increases substantially with λ, reason why in the application we choose $\lambda = 50\%$ (or even more) and not the more conventional value of 5%.

3.4   Normality of innovations

Our test for additive jumps is based on the assumption that on days without jumps, returns are conditionally Gaussian. It is thus important to use a test for conditional normality on the filtered returns in addition to our test for jumps. Indeed, a rejection of this assumption might invalidate our postulated DGP and therefore our test.

Recall that in absence of jumps (i.e. $r_t^* = r_t \; \forall t$), the expected number of spurious jumps for a sample of T observations equals $\lambda < 1$ and therefore $\tilde{r}_t = r_t$ almost surely for $t = 1, \ldots, T$. Consequently, the standardized returns, obtained by re-estimating on $\tilde{r}_t$ the postulated model for r_t (e.g. ARMA-GARCH(1,1)) by Gaussian maximum likelihood, should be $i.i.d. N(0,1)$ under the assumption of correct specification for r_t. Similarly, in the presence of jumps, and under the assumption that jumps have been correctly detected, filtered returns $\tilde{r}_t$ (see Equation (2.7)) should approximately follow the postulated model for r_t.

Table 2 reports the rejection frequencies (over 1,000 replications) of the normality test of Jarque and Bera (1987), denoted JB hereafter. The DGP corresponds to the one presented in Section 3.2 but, to save space, we only report results for $\alpha_1 = 0.02$ and $\beta_1 = 0.93$. Jumps are tested using Franses and Ghijsels' test with $\lambda = 5\%$ (column $FG(5\%)$) and our test with $\lambda = 5, 25$ and 50% (columns $LLP(5\%)$, $LLP(25\%)$ and $LLP(50\%)$ respectively). The number of jumps (in r_t^*) per sample of $T = 2,000$ observations is set to $1, 2, 5, 10, 20$ or 40. Frequencies reported in this table correspond to the rejection frequencies of the JB test on the standardized residuals of an AR(1)-GARCH(1,1) model estimated by Gaussian maximum likelihood on the filtered returns $\tilde{r}_t$ (using the corresponding jumps statistic). The critical level of the

17

JB test is set to 5%.

Results suggest that the lack of power of Franses and Ghijsels's test, at the 5% level, translates into an over-rejection of the JB test in presence of jumps in r_t^*, compared to our test. Results also suggest that for $1, 2, 5$ and 10 jumps per sample, the rejection frequencies of the JB test are very close to the nominal size for $LLP(50\%)$. For 20 and 40 jumps in the sample, the null of normality is rejected too often for rather small jumps (i.e. for $m = 2$ or 3). This is due to the fact that all tests have low power to detect jumps when $m \leq 3$ (see Figures 2 and 3). Indeed, unlike very small jumps (i.e. $m = 1$), small jumps (i.e. $m = 2$ or 3) that have not been detected and therefore not filtered out in $\tilde{r}_t$, seem to have a strong impact on the assumption of normality of the standardized residuals of the normal-AR(1)-GARCH(1,1). This result might suggest considering even higher values for λ but this is beyond the scope of this paper.

3.5   Non-normality of innovations

In this section we investigate the behaviour of our test in absence of jumps but when innovations are non-Gaussian.

Non-normal distributions have also been used to account for the presence of large shocks. For instance, Bollerslev (1987) proposed to extend the above specification by assuming z_t in (2.2) to follow a standardized Student-t distribution, i.e. $z_t \overset{i.i.d.}{\sim} ST(0, 1, \upsilon)$, where υ is the number of degrees of freedom of the Student-t distribution. The main drawback of GARCH models (even with non-normal innovations) is that they assume that each shock (i.e. ε_{t-1}^2) has the same relative impact on future volatility, regardless of its magnitude. However, this assumption is at odds with an increasing body of evidence indicating that the largest shocks have a relatively small effect on future volatility than smaller shocks. Harvey and Chakravarty (2008) and Creal, Koopman, and Lucas (2012) recently and independently proposed a modification to the GARCH model derived from the conditional score of the assumed distribution with respect to the second moment. Those models are called either Generalized

Autoregressive Score (GAS) or Dynamic Conditional Score (DCS) models and are described in Appendix 2. They assume that observed returns are described by Equations (2.1)-(2.2), where z_t is usually assumed to follow a non-normal distribution and $\sigma_t^2 = \omega + \alpha_1 w(z_{t-1}^2)\sigma_{t-1}^2 + \beta_1\sigma_{t-1}^2$, where $w(.)$ is a function that possibly downweights and bounds the effect of past shocks. In the Gaussian case, $w(z_{t-1}^2) = z_{t-1}^2$, their model reduces to the standard GARCH model and therefore the impact of a shock is unbounded. However, for non-normal distributions, large shocks are downweighted and have a smaller effect on future volatilities than in GARCH models. For instance if $z_t \sim ST(0,1,\upsilon)$, the function $w(z_{t-1}^2) = \frac{(\nu+1)z_{t-1}^2}{\nu-2+z_{t-1}^2}$ downweights and bounds the effect of large shocks (i.e. $w(z_{t-1}^2) < z_{t-1}^2$ and $\max[w(z_{t-1}^2)] = \nu + 1 < \infty$ if $\nu < \infty$). Several extensions of the GAS model are discussed in Appendix 2, including the case with Skewed-Student innovations and an EGARCH-type version with and without leverage effect.

GAS-type models with non-normal distributions do not contain additive jumps and therefore our test is not appropriate in this case. To investigate the role of the JB test discussed in then previous section, we consider the AR(1)-GAS(1,1) model of Harvey and Chakravarty (2008) with $z_t \sim ST(0,1,\upsilon)$, as described in Appendix 2, with the same parameter values as in Section 3.4 (with $\psi_1 = \alpha_1 + \beta_1$).

We simulated $1,000$ series of $T = 2,000$ observations and applied our test with $\lambda = 50\%$, relying on the BIP–AR(1)-BIP–GARCH$(1,1)$ model to estimate $\tilde{\mu}_t$ and $\tilde{\sigma}_t^2$. Recall that there is no additive jump in the DGP.

Results, reported in Table 3, suggest that, as expected, the proportion of jumps detected by our test increases with $1/\upsilon$ and is maximised for $\upsilon = 3$ (i.e. 1.372% of jumps). The JB test presented in Section 3.4 has also been applied to the standardized residuals of the AR(1)-GARCH$(1,1)$ model estimated by Gaussian ML on the filtered returns $\tilde{r}_t$. Interestingly, the rejection frequency of the test of normality is very high and is close to 100% for values of $\upsilon < 7$, i.e. values that are commonly found for real data.

To conclude, our test is based on the assumption that the non-contaminated returns r_t are conditionally Gaussian and requires the use of a robust para-

19

metric model to estimate the first two conditional moments of r_t in the likely presence of jumps. A misspecification of those conditional moments and/or a violation of the assumption of conditional normality might distort the properties of our test. We investigated a procedure involving a test of normality on the standardized returns of a model estimated on the filtered returns $\tilde{r}_t$ and found it to have nice properties: relatively good size and very good power against the GAS model of Harvey and Chakravarty (2008) with Student-t innovations.

3.6   Bias, RMSE and 95% coverage probability

We now investigate the finite sample properties of three estimation methods, i.e.

- Gaussian (Quasi-)ML (denoted QML in the graphs);

- M-estimator of the BIP-AR(1)–BIP-GARCH(1,1) as discussed in Section 2 of the paper (denoted BIP in the graphs);

- and Gaussian ML estimation on filtered returns $\tilde{r}_t$ using our proposed detection rule, i.e. $\tilde{r}_t = r_t^* - (r_t^* - \tilde{\mu}_t)\tilde{I}_t$, where $\tilde{I}_t = I\left(|\tilde{J}_t| > g_{T,\lambda}\right)$ with $\lambda = 50\%$ (denoted LLP(50%) in the graphs).

Figure 4 plots the empirical bias of $\mu, \phi_1, \omega, \alpha_1$ and β_1 over the 1,000 replications as a function of the jump size m. In order to save space, we only report the results for 10 jumps. The empirical bias of parameter θ is defined as $\frac{1}{1000}\sum_{i=1}^{1000}(\theta_0 - \hat{\theta}_i)$, where θ_0 denotes the true parameter value and $\hat{\theta}_i$ its estimate obtained from the ith sample. As expected, the Gaussian-ML estimator of the AR(1)-GARCH(1,1) (denoted Gaussian-ML on r_t^* in the graph) is not robust to additive jumps and the bias increases with m.

We observe that the M-estimator of the BIP-ARMA–BIP-GARCH with $\delta = 0.975$ and $\rho(\cdot) = \rho_{t_4}(\cdot)$ (denoted BIP) and the Gaussian-ML on filtered returns using our proposed jump test (denoted Gaussian-ML on $\tilde{r}_t$ with $\lambda = 50\%$) are robust to jumps. Interestingly, for these two methods, the bias is

found to be small for each parameter and independent of the magnitude of the jumps.

Figure 5 plots the root mean square errors (RMSE) of the five parameters. This figure also suggests that the M-estimator of the BIP-ARMA–BIP-GARCH and the Gaussian maximum likelihood on filtered returns using our proposed jump test perform very well. Furthermore, the loss of efficiency compared to the Gaussian-ML is very small in absence of jumps.

Finally, Figure 6 plots the 95% coverage probabilities for the five parameters as a function of m for the case of 10 jumps in the sample. The 95% coverage probability of parameter θ corresponds to the number of times the true value θ_0 falls within the confidence interval $\hat{\theta}_i \pm 1.96\sqrt{var(\hat{\theta}_i)}$ divided by the number of replications. Muler and Yohai (2008) have proved the asymptotic normality of the M-estimator of the BIP-GARCH(1,1) model and derived the asymptotic variance in the particular case of zero conditional mean and in absence of jumps. Our simulation set-up includes an AR(1) in the conditional mean as well as jumps and, as far as we know, the asymptotic distribution of the M-estimator is unknown. Therefore, we do not report the 95% coverage probabilities for this estimation method but concentrate on the Gaussian maximum likelihood on filtered returns (with $\lambda = 50\%$).

Figure 6 suggests that this latter estimator has a 95% coverage probability close to the theoretical value of 95% for each parameter, irrespective of the size of the jumps.[10] In absence of jumps, as expected the 95% coverage probabilities of the Gaussian-ML on the raw returns deviate from their theoretical value when m is large.

4   Application

In this section we perform two different applications. In Section 4.1, we apply our test for additive jumps to exchange rate data and conduct an event study on the days where jumps have been detected. In Section 4.2, we apply the test to about 50 US stocks and filter out the detected jumps from the raw return

---

[10]We obtained similar figures for numbers of jumps different from those considered above.

series. Then, we compare the forecasting performance of ARMA-GARCH(1,1) and ARMA-GJR(1,1) models estimated on filtered returns with other models (several GARCH but also GAS-type models) estimated on raw series.

4.1 Exchange rates

In the first application, we apply our test for additive jumps in ARMA-GARCH models to the major exchange rates.[11] Our objective is twofold. First, we examine whether, once jumps have been filtered out from r_t^*, the assumption of conditional normality of $\tilde{r}_t$ holds for the exchange rate returns. Second, we conduct an event study and check whether detected jumps have an economic explanation.

This application is carried out on the US Dollar - Euro (USD-EUR), the US dollar - British pound (USD-GBP) and the Japanese yen - US dollar (YEN-USD) exchange rates over the period January 2005 - May 2011 (i.e. $T = 1,598$ daily observations).[12] We choose to apply our test to exchange rate returns for two main reasons. First, exchange rates have known frequent and large discontinuities during the considered period and especially during the sub-prime crisis in 2008-2009. Second, the literature on jumps and announcements (see the survey of Neely, 2011 for this) concludes that many jumps appear to correspond to macroeconomic announcement news. News on the occurrence of central bank interventions in the FX market cause discontinuities in exchange rate prices, as shown by Fair (2002) and Gnabo, Laurent, and Lecourt (2009). Because this type of event is unexpected by the market, it leads market participants to adjust their trading behaviour, conducting to some discontinuities

---

[11]For foreign exchange returns, symmetry (i.e. no leverage effect) is usually assumed (see Hull, 2006 for instance). Bollerslev, Chow, and Kroner (1992) argue that "whereas stock returns have been found to exhibit some degree of asymmetry in their conditional variances, the two-sided nature of foreign exchange markets makes such asymmetries less likely". In most applications on exchange rate returns, the symmetric GARCH(1,1) is not rejected by the data (both in-sample and out-of-sample) when compared to more sophisticated models (of the same order) allowing for this asymmetric effect (e.g. GJR, EGARCH, TARCH). See among others Hsieh (1989), Diebold and Nerlove (1989), Taylor (1986), Andersen, Bollerslev, Diebold, and Labys (2001) and Hansen and Lunde (2005).

[12]Source: FRED (Federal Reserve Economic Data) website.

in prices. Interestingly, unlike other central banks, the Bank of Japan has continued to intervene actively during the last ten years.

Figure 7 plots the daily returns in % of the three exchange rates (solid line) and the detected jumps. Three critical levels are considered, i.e. $\lambda = 50, 75$ and 95%. The critical values $g_{1598,\lambda}$ are respectively 3.52724, 3.34678 and 3.14617. Returns identified as being contaminated by an additive jump at least at the $50, 75$ and 95% levels are highlighted respectively by a square, circle and triangle.

Table 4 reports the Jarque-Bera statistic (and its p-value in parentheses) computed on the standardized residuals of an ARMA-GARCH(1,1) with Gaussian innovations estimated on the raw returns and the filtered returns $\tilde{r}_t$ (for the three critical levels). Not surprisingly, the assumption of conditional normality is rejected in all cases for the models estimated on raw returns. Interestingly, this assumption is not rejected at the 5% level when this model is re-estimated on the filtered returns $\tilde{r}_t$ with $\lambda \geq 50\%$ for the USD-EUR and for $\lambda = 95\%$ for the other two series. The number of detected jumps (reported between squared brackets) is very small, i.e. between 5 (i.e. 0.31%) and 15 (i.e. 0.94%). This result suggests that the rejection of the assumption of conditional normality in the three daily exchange rate returns is due to less than 1% of large jumps.

Tables 5-7 report the dates of all the detected jumps, our jump statistic $|\tilde{J}_t|$ and in the last column, labelled 'News', real-time financial news and information released around jump arrival days using the Factiva database in order to examine their association with jump arrivals.[13] Sources used in the Factiva search include Dow Jones and Reuters newswires.

Importantly all the detected jumps have been largely documented by the newswires services and all news reports extracted the same day as jump arrivals correspond with economic events. One important event is for example the intervention of the Japanese monetary authorities in the FX market, unilaterally the 15th of September 2010 and jointly with the G7 very recently,

---

[13]The purpose of this analysis is not to identify the direction of the causality between jumps and these news. For this, we would need the timing of the discontinuities that create jumps and compare it with the timing of the arrival of these news.

on the 18th of March 2011. The largest jumps detected in 2008 and 2009 are related to the global financial crisis. The jumps detected on the 19th of March 2009 are common to the three exchange rates. It is due to a fall of the US dollar, consequence of a massive purchase of \$300 billion of long-term Treasuries securities by the US Federal Reserve.

4.2   US stocks

In this section, we apply our test on US stock return data and filter out the detected jumps using Equation (2.7). Then, we compare the forecasting performance of ARMA-GARCH and ARMA-GJR models estimated on filtered returns $\tilde{r}_t$ with that of other models (GARCH-type models but also GAS-type models) estimated on raw returns.

<u>Data:</u> The data (provided by TickData) consists of transaction prices at the 5-minute sampling frequency for $N = 49$ large capitalization stocks from the NYSE, AMEX NASDAQ, covering the period from January 4, 1999 to December 31, 2008 (2,489 trading days). The trading session runs from 9:30 EST until 16:00 EST. Several models are estimated on daily log-returns in % (obtained by summing 5-minute log-returns) on rolling windows of 980 observations.

<u>Set of competing models:</u> We consider $M = 14$ models in the set of competing models, all estimated by ML.

The models are classified in three groups.

- Group 1: GARCH. The first group is made up of the following six GARCH-type models: ARMA-GARCH(1,1) and ARMA-GJR(1,1) models with a Normal, Student-t and Skewed-Student distribution respectively (see Appendix 1 for the Skewed-Student distribution).

- Group 2: GAS. The second group of six models consists of the ARMA-GAS(1,1), ARMA-EGAS(1,1) and ARMA-AEGAS(1,1) models (see Appendix 1 for the models) with Student-t and Skewed-Student distributions.

- Group 3: BIP. Finally, the third group consists of the following two models: ARMA-GARCH(1,1) and ARMA-GJR(1,1) models with a Normal distribution estimated on the filtered returns $\tilde{r}_t$, as given in Equation (2.7), where jumps are obtained using (2.6). The conditional moments $\tilde{\mu}_t$ and $\tilde{\sigma}_t^2$ entering in the equation defining $\tilde{J}_t$ are estimated by minimizing the objective function (2.13) of the BIP version of the model, i.e. respectively a BIP–ARMA-BIP–GARCH(1,1) or BIP–ARMA-BIP–GJR(1,1) model, on the raw returns r_t^* (as described in Section 2.2). Like in the simulation study, we set δ to 0.975 in the weight function $w_{k_\delta}^{\mathrm{MPY}}(\cdot)$ and choose $\rho(\cdot) = \rho_{t_4}(\cdot)$. The critical value (k) of our jump detection rule is the one given in Equation (2.8), with $\lambda = 0.5$ meaning that the probability of finding at least one spurious additive jump under the null of no jump in the data equals 50%.

Models in Groups 1 and 2 (resp. 3) are estimated on the raw (resp. filtered) return series.

Forecasting: The models are estimated on the first 980 observations. h-step-ahead forecasts of the conditional variance are then computed for $h = 1, \ldots, 10$. Let us denote $\hat{\sigma}_{n,m,t+h|t}^2$, the h-step-ahead forecast of the conditional variance of day $t + h$ for series $n = 1, \ldots, N$ given all the information up to day t for model $m = 1, \ldots, M$. The models are re-estimated every 50 days, on a rolling window of 980 observations. The total number of h-step-ahead forecasts is therefore about $1,500$ for every model. For the models belonging to Group 3 (i.e. estimated on filtered returns), each time a new observation comes in, our test is applied to this new observation and it is filtered using Equation (2.7) if $|\tilde{J}_t| > g_{980,0.5}$ before producing the forecasts.

Proxy: The evaluation of the forecasting performance of volatility models is challenging since the variable of interest (i.e. the quadratic variation of day $t + h$ for series n, denoted $\sigma_{n,0,t+h}^2$) is unobservable and therefore a proxy is needed to rank the competing models. Given the widespread availability of intraday prices of various financial assets, Andersen and Bollerslev (1998a) have proposed to estimate the quadratic variation with the so-called realized

volatility, computed as the sum of intraday squared returns (5-minute returns in our case). The most fundamental feature of realized volatility is that it provides, under some conditions, a consistent non-parametric estimate of the price variability that has transpired over a given discrete interval.

Loss function: The h-step-ahead forecasts of the competing models are compared with realized volatility (computed over the same time-horizon). In the presence of outliers (or jumps), Preminger and Franck (2007) recommend using forecast performance evaluation criteria that are less sensitive to extreme observations. For this reason we rely on the Mean Absolute Deviation (MAD), defined as follows:

$$L_{n,m,t+h|t} = \left| \sigma^2_{n,0,t+h} - \hat{\sigma}^2_{n,m,t+h|t} \right|, \qquad (4.1)$$

where $L_{n,m,t+h|t}$ is the MAD corresponding to the h-step-ahead volatility forecasts produced by model m for series n.

Hansen and Lunde (2006), Laurent, Rombouts, and Violante (2013) and Patton (2011) show that the substitution of the underlying volatility by a proxy may induce a distortion in the ranking i.e., the evaluation based on the proxy might differ from the ranking that would be obtained if the true target was used. However, such distortion can be avoided if the loss function has a particular functional form or when the proxy is accurate enough. Monte-Carlo simulation results reported in Laurent, Rombouts, and Violante (2013) suggest that when the proxy is computed from sufficiently high frequency returns (e.g., 5-minute returns like in our case), all loss functions deliver the expected ranking (i.e., the one based on the true variance), which justifies our choice.

Statistical test on the loss differentials: The model confidence set (MCS) approach of Hansen, Lunde, and Nason (2011) is used to compare the forecasts. Given a universe of model-based forecasts, the MCS allows us to identify the subset of models that are equivalent in terms of forecasting ability (using a certain loss function), but outperform all other competing models. Implementation of this test has been done using the Ox software package MULCOM of Hansen and Lunde (2007). We set the confidence level of the MCS test

to 10% and used 10,000 bootstrap resamples (with block length of 5 observations). On equity data, Laurent, Rombouts, and Violante (2012) find that the relative performance of volatility models depends strongly on the state of the market. We therefore distinguish between the calm period of 2002–2007 and the full period 2002–2008 (that includes the more turbulent period 2007-2008).

<u>Results:</u> The proportion of jumps detected by our test, where $\tilde{\mu}_t$ and $\tilde{\sigma}_t$ are obtained using a BIP–ARMA-BIP–GJR(1,1) model[14] and λ is set to 50% is reported in Table 8. Recall that the parameters of the model are kept constant during 50 days. The model is therefore re-estimated every 50 days on a rolling window of 980 observations. The percentages reported in the table correspond to the total number of detected jumps in the whole sample divided by the total number of trading days (2,489). The percentage of detected jumps varies between about 1 and 3%.

The assumption of conditional normality in absence of jumps is crucial for our test to be applicable. Results not reported here to save space suggest that the assumption of conditional normality is rejected for all stocks for the models estimated on raw returns. Table 8 also reports the mean and variance of the p-value of the Jarque-Bera test computed on the innovations of the ARMA-GJR estimated on the filtered returns $\tilde{r}_t$. Those descriptive statistics are computed over the 1,500 p-values obtained from the innovations of the model used for producing the h-step-ahead forecasts. Interestingly, the assumption of conditional normality holds for all stocks because the means of the p-values are much greater than the conventional significance levels (for instance 60% for AAPL) and the variances of the p-values are rather small. Like for the exchange rate series, this result suggests that the rejection of conditional normality in the raw returns is due to a very small proportion of jumps that our test successfully detected.

The GARCH models, provided they are correctly specified, should deliver asymptotically unbiased forecasts of $\sigma^2_{n,0,t+h}$, that is of the quadratic variation.

---

[14]We choose to report results based on the BIP–ARMA-BIP–GJR(1,1) and not the BIP–ARMA-BIP–GARCH(1,1) because for most stocks, negative shocks have a deeper impact on volatility than positive shocks of the same magnitude.

To the extent that jumps are present and have been filtered out correctly, the GARCH models estimated on filtered returns should provide asymptotically unbiased forecasts of the conditional variance of the continuous component of the process. From the results of the MCS test in Table 9 for the full period and $h = 1$, it becomes apparent that GAS models estimated on the raw data (Group 2) and GARCH-type models estimated on filtered returns (Group 3) appear in the MCS with a much higher frequency than the GARCH-type models estimated on the raw data (Group 1). Interestingly, the performance of the GARCH-type and GAS-type models deteriorates when the forecasting horizon h increases. Indeed, for $h > 4$, models estimated on filtered returns belong to the MCS in about 94 to 98% of the time against 60 to 47% and about 63% respectively for the GARCH and GAS models estimated on raw data. The difference is even more striking for the period 2002-2007 for which those models belong to the MCS in only 12 to maximum 30% of the time. The GAS models might then suffer from some misspecification too but to a much smaller extent than the GARCH model.

An issue for future research concerns the performance of volatility forecasts at various horizons based on the sum of the forecast of the continuous and jump components.

5 Conclusion

It is well known that high-frequency returns of most financial assets exhibit volatility clustering but also large jumps caused by big surprises. However, these jumps affect future volatility less than what standard volatility models would predict (see Andersen, Bollerslev, and Diebold, 2007; Harvey and Chakravarty, 2008; Muler and Yohai, 2008 among others).

Building upon the BIP-ARMA and BIP-GARCH models of respectively Muler, Pena, and Yohai (2009) and Muler and Yohai (2008), and in line with the proposal of Aït-Sahalia and Jacod (2012a) to distinguish between a 'continuous' and a jump component when modelling financial time series, we proposed a new test for additive jumps in ARMA-GARCH models. The distribution un-

der the null hypothesis of the proposed test follows from the consistency and asymptotic normality of the parameter estimators as proved by Muler and Yohai (2007). Our Monte-Carlo simulation study suggests that the test does not suffer from any size distortion and has a very good power to detect the actual jumps in finite samples. Besides that, unlike Franses and Ghijsels's (1999) test, the critical values of our test do not depend on the unknown parameters of the GARCH model and the power of the test does not seem to depend on the number of jumps in the sample.

We highlighted the crucial role of the assumption of conditional normality of the non-contaminated returns r_t. However nothing guarantees to find conditional normality of the filtered returns $\tilde{r}_t$ after applying our jump removing procedure to real data. Our Monte-Carlo simulation results suggest that when the DGP is a GAS or GARCH model with Student-t innovations, spurious jumps are detected by our test and a Jarque-Bera test has very good power to detect conditional *non*-normality of the filtered returns.

We applied our test on daily returns and detected less than 1% of jumps for the three exchange rates and between 1 and 3% of jumps for all the stocks considered in the application (about 50 large capitalization stocks from the NYSE). Interestingly, we failed to reject the assumption of conditional normality of the filtered returns in all cases. For US stocks, we compared the forecasting power of a group of GARCH-type models estimated on filtered returns and two groups of models estimated on raw data. Using the MCS test of Hansen, Lunde, and Nason (2011) and realized volatility to approximate the true variance, the supremacy of the former group has been established. This result suggests that standard-GARCH models estimated on filtered returns outperform other models, including GAS-type models.

Appendix 1: Franses and Ghijsels' Test for Jumps

Franses and Ghijsels (1999) adapt the procedure of Chen and Liu (1993) for additive outlier detection in ARMA models to make it applicable for GARCH models with additive jumps. They consider that if a jump occurs at time

t, instead of observing r_t, one observes the contaminated return r_t^*, where the contamination is defined through the squared error process, i.e. $(\varepsilon_t^*)^2 = (\varepsilon_t^2 + w_t I_t)$, where w_t, with $-\varepsilon_t^2 < w_t < +\infty$, is the size of the additive jump in the squared residuals. From $(\varepsilon_t^*)^2$ one can recover the contaminated returns by taking its square root and by further imposing that ε_t^* and ε_t have the same sign, i.e. $\varepsilon_t^* = \text{sign}(\varepsilon_t)\sqrt{\varepsilon_t^2 + w_t I_t}$, where $\text{sign}(x) = 1$ if $x \geq 0$ and -1 otherwise. This yields the following DGP for the observed return series r_t^*:

$$
\begin{aligned}
r_t^* &= r_t(1 - I_t) + (\mu_t + \varepsilon_t^*)I_t \\
&= (\mu_t + \varepsilon_t) + (\mu_t + \varepsilon_t^* - \mu_t - \varepsilon_t)I_t \\
&= r_t + (\varepsilon_t^* - \varepsilon_t)I_t,
\end{aligned}
\tag{5.1}
$$

where r_t is defined as in (2.1)-(2.3). Note that Equation (5.1) is a particular case of Equation (2.4), where $a_t = \varepsilon_t^* - \varepsilon_t$.

The procedure of Franses and Ghijsels (1999) to test for additive outliers in GARCH models is summarised here below:

1. Estimate an ARMA-GARCH(1,1) model by (Quasi-)Maximum Likelihood on the observed returns r_t^* by neglecting the potential presence of jumps in the data (i.e. by replacing r_t in (2.1)-(2.3) by r_t^*) and compute $\hat{\sigma}_t^2$ and $\hat{v}_t = (r_t^* - \hat{\mu}_t)^2 - \hat{\sigma}_t^2$.

2. Compute $t_{\hat{\xi}(\tau)}$, the t-statistic for the estimated slope coefficient $\hat{\xi}(\tau)$ of the regression of $\hat{v}_t$ on x_t, where $x_t = 0$ for $t < \tau$, $x_\tau = 1$ and $x_{\tau+k} = -\pi_k$ for $k = 1, \dots$ and $\pi(L) = (1 - \beta_1 L)^{-1}(1 - (\alpha_1 + \beta_1)L)$ for a GARCH(1,1). They use an estimate of the variance of the error term of this regression that is robust to the potential jump occurring at time $t = \tau$. See Franses and Ghijsels (1999) for more details.

3. Obtain $t_{max}(\hat{\xi}) \equiv \max_{1 \leq \tau \leq T} |t_{\hat{\xi}(\tau)}|$ and compare it with a critical value denoted by C. If $t_{max}(\hat{\xi}) > C$, the observation for which the t-statistic corresponds to $t_{max}(\hat{\xi})$ (say $t = \hat{\tau}$) is defined as contaminated by an additive outlier and is cleaned in the next step.

4. Clean the original series for the detected additive outliers by replacing r_τ^* by $\hat{\mu}_\tau + \mathrm{sgn}(\hat{\varepsilon}_\tau)\sqrt{(r_\tau^* - \hat{\mu}_\tau)^2 - \hat{\xi}(\tau)}$, where $\mathrm{sgn}(x) = 1$ if $x \geq 0$ and -1 otherwise.

5. Return to step 1 and re-estimate model (2.1)-(2.3) on the cleaned returns.

6. Repeat steps 1-5 until $t_{max}(\hat{\xi})$ no longer exceeds C.

Franses and Ghijsels (1999) recommend using $C = 4$ as critical value while simulation results reported in Franses and van Dijk (2000) suggest that the choice of C is not so trivial. Indeed, they show that the distribution of $t_{max}(\hat{\xi})$ under the null of no additive outliers varies not only with the number of observations T but also with the true but unknown values α_1 and β_1. For instance, for $T = 500$, $\alpha_1 = 0.1$ and $\beta_1 = 0.5$ the 95% quantile of $t_{max}(\hat{\xi})$ (based on 1,000 replications) equals 10.94 while for $\alpha_1 = 0.2$ and $\beta_1 = 0.7$ it is 16.93. Note that Charles and Darné (2005) extend the above test for additive outliers to take into account innovative outliers in a GARCH model, that is outliers that reflect an endogenous change in a series and affect all future realizations of the variable through the memory of its process.

Appendix 2: Generalized Autoregressive Score (GAS) and Dynamic Conditional Score (DCS) models

To present the new class of conditional variance models proposed by Harvey and Chakravarty (2008) and Creal, Koopman, and Lucas (2012) (let us call is GAS for simplicity), let us start with the model described in Equations (2.1)-(2.2) and define $Y_t = \{\varepsilon_1, \ldots, \varepsilon_t\}$ a vector with the demeaned returns up to time t. In GAS models, it is assumed that ε_t is generated by the very general observation density $f(\varepsilon_t|\sigma_t^2, Y_{t-1}; \theta), t = 1, \ldots, T$, where θ is a vector of unknown parameters describing the joint-distribution function of the data. Note that in Model (2.1)-(2.2), and $f(\varepsilon_t|\sigma_t^2, Y_{t-1}; \theta) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right)$.

For a GAS(1,1) model, the updating equation for the conditional variance σ_t^2 is not necessarily Equation (2.3) but the more general autoregressive updating function:

$$\sigma_t^2 = \omega + B_1 \sigma_{t-1}^2 + A_1 \kappa_{t-1}. \tag{5.2}$$

Harvey and Chakravarty (2008) and Creal, Koopman, and Lucas (2012) propose to update σ_t^2 with $\kappa_t = S_t \triangledown_t$, where $\triangledown_t$ is the score with respect to the parameter σ_t^2, i.e. $\triangledown_t = \partial \log f(\varepsilon_t | \sigma_t^2, Y_{t-1}; \theta) / \partial \sigma_t^2$ and S_t is a time dependent scaling matrix.

Note that for Model (2.1)-(2.2), $\triangledown_t = \partial - 0.5 \left(\log \sigma_t^2 + \epsilon_t^2 \sigma_t^{-2} \right) / \partial \sigma_t^2 = 0.5(z_t^2 - 1)\sigma_t^2$ which leads to Equation (2.3) if, in (5.2), $\kappa_{t-1} = S_{t-1} \triangledown_{t-1}$ $S_{t-1} = 2$, $A_1 = \alpha_1$ and $B_1 = \alpha_1 + \beta_1$. The GARCH(1,1) is therefore a GAS(1,1) when $z_t \overset{i.i.d.}{\sim} N(0,1)$.

Indeed, let us rewrite the GARCH(1,1) model in (2.3) as follows:

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1} \sigma_{t-1}^2 + (\alpha_1 + \beta_1)\sigma_{t-1}^2,$$

where $u_{t-1} = z_{t-1}^2 - 1$ is proportional to the score of the conditional distribution of ε_t with respect to σ_{t-1}^2 and therefore is a natural choice of updating scheme in a 'Newton-Raphson' sense.

The only difference between Harvey and Chakravarty (2008) and Creal, Koopman, and Lucas (2012) is the choice of S_t. Creal, Koopman, and Lucas (2012) recommend using $S_t = 1$ or $S_t = (E_{t-1} \triangledown_t \triangledown_t')^{-1}$ while Harvey and Chakravarty (2008) set S_t to a constant. We follow Harvey and Chakravarty (2008) and set S_t to 2.

This principle can be applied to any distribution. The specification of the GAS(1,1) model of Harvey and Chakravarty (2008) combined with a normal, Student-t ($ST(0,1,\upsilon)$) or Skewed-Student ($SKST(0,1,\xi,\upsilon)$) distribution[15] is given below:

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1} \sigma_{t-1}^2 + \psi_1 \sigma_{t-1}^2, \tag{5.3}$$

---

[15]The log-likelihood of the standardized (zero mean and unit variance) skewed-Student of

where

$$u_t = z_t^2 - 1 \text{ if } z_t \sim N(0,1); \tag{5.4}$$

$$u_t = \frac{(\nu+1)z_t^2}{\nu-2+z_t^2} - 1 \text{ if } z_t \sim ST(0,1,\upsilon); \tag{5.5}$$

$$u_t = \frac{(\nu+1)z_t z_t^*}{(\nu-2)g_t \xi^{L_t}} - 1 \text{ if } z_t \sim SKST(0,1,\xi,\upsilon), \tag{5.6}$$

where for the SKST,

$$z_t^* = s z_t + m \tag{5.7}$$

$$g_t = 1 + \frac{z_t^{*2}}{(\nu-2)\xi^{2L_t}} \tag{5.8}$$

$$m = \frac{\Gamma\left(\frac{\upsilon-1}{2}\right)\sqrt{\upsilon-2}}{\sqrt{\pi}\,\Gamma\left(\frac{\upsilon}{2}\right)}\left(\xi - \frac{1}{\xi}\right) \tag{5.9}$$

$$s = \sqrt{\left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2}. \tag{5.10}$$

Bauwens and Laurent (2005), denoted $SKST(0,1,\xi,\upsilon)$, is:

$$
\begin{aligned}
L_{SkSt} = \ & T\left\{\log\Gamma\left(\frac{\upsilon+1}{2}\right) - \log\Gamma\left(\frac{\upsilon}{2}\right) - 0.5\log\left[\pi(\upsilon-2)\right] + \log\left(\frac{2}{\xi+\frac{1}{\xi}}\right) + \log(s)\right\} \\
& - 0.5\sum_{t=1}^{T}\left\{\log\sigma_t^2 + (1+\upsilon)\log\left[1 + \frac{(sz_t+m)^2}{\upsilon-2}\xi^{-2L_t}\right]\right\},
\end{aligned}
$$

where

$$L_t = \begin{cases} 1 \text{ if } z_t \geq -\frac{m}{s} \\ -1 \text{ if } z_t < -\frac{m}{s} \end{cases},$$

ξ is the asymmetry parameter, υ is the degree of freedom of the distribution,

$$m = \frac{\Gamma\left(\frac{\upsilon-1}{2}\right)\sqrt{\upsilon-2}}{\sqrt{\pi}\,\Gamma\left(\frac{\upsilon}{2}\right)}\left(\xi - \frac{1}{\xi}\right),$$

and

$$s = \sqrt{\left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2}.$$

Note that when $\xi = 1$, $SKST(0,1,1,\upsilon) = ST(0,1,\upsilon)$.

Harvey and Chakravarty (2008) call the above GAS model with a Student-t distribution 'Beta-t-GARCH' because, for this distribution, $(u_t + 1)/(\nu + 1)$ has a Beta distribution.

Note that u_t being a rescaled conditional score, it holds that $E_{t-1}(u_t) = 0$ for the four distributions because it is a martingale difference sequence. Consequently, $E(\sigma_t^2) = \omega/(1 - \psi_1)$.

Figure 1 plots the conditional score function, u_t, against the shocks z_t for a GARCH model and a GAS with a Student-t distribution with $\upsilon = 5$ and a SKST$(0, 1, \exp(-0.3), 5)$ distribution. For the GARCH model or equivalently for a GAS with a normal distribution (upper solid line), an extreme observation has a huge impact on u_t and therefore, since α_1 is expected to be positive and ψ_1 is usually close to 1, this extreme event will have a large and slowly decaying effect on future volatility predictions. For the T with 5 degrees of freedom, the effect of an extreme observation is bounded and therefore has only a moderate impact. Large shocks are downweighted since z_t^2 appears both in the numerator and denominator of (5.5). For the GARCH, and GAS with Student-t distribution, the effect of a shock z_t on the conditional score function is symmetric. Interestingly, for the SKST distribution with $\xi = \exp(-0.3)$ (dashed line), negative shocks have a deeper impact on u_t and therefore on future volatility predictions than positive shocks, which is a kind of leverage effect.

Harvey and Chakravarty (2008) also considered an EGARCH-type version of this model (called Exponential GAS or EGAS). This model can be further extended to account for leverage effect as follows,

$$\log \sigma_t^2 = \omega + \alpha_1 u_{t-1} + \gamma_1 l_{t-1} + \psi_1 \log \sigma_{t-1}^2, \tag{5.11}$$

where $l_t = sgn(-z_t)(u_t + 1)$ for the three symmetric distributions and $l_t = sgn(-z_t^*)(u_t+1)$ for the SKST. Therefore $E(l_t) = 0$ for symmetric distributions and $E(l_t) = \frac{1-\xi^2}{1+\xi^2}$ for the Skewed-Student. This model is called AEGAS (for Asymmetric Exponential GAS).

References

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

———— (2009): "Testing for Jumps in a Discretely Observed Process," *Annals of Statistics*, pp. 184–222.

———— (2012a): "Analyzing the Spectrum of Asset Returns: Jump and Volatility Components in High Frequency Data," *Journal of Economic Literature*, 50, 1007–1050.

———— (2012b): "Testing for Jumps in Noisy High Frequency Data," *Journal of Econometrics*, 168, 207–222.

Andersen, T., and T. Bollerslev (1998a): "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2001): "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

Andersen, T., T. Bollerslev, P. Frederiksen, and M. Nielsen (2010): "Continuous-Time Models, Realized Volatilities, and Testable Distributional Implications for Daily Stock Returns," *Journal of Applied Econometrics*, 25(2), 233–261.

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

Andersen, T. G., T. Bollerslev, and F. Diebold (2007): "Roughing it up: Including Jump Components in the Measurement, Modelling and Forecasting of Return Volatility," *Review of Economics and Statistics*, 89, 701–720.

ANDERSEN, T. G., T. BOLLERSLEV, AND D. DOBREV (2007b): "No-arbitrage semi-martingale restrictions for continous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications," *Journal of Econometrics*, 138, 125–180.

BALL, C., AND W. TOROUS (1983): "A Simplified Jump Process for Common Stock Returns," *Journal of Financial and Quantitative Analysis*, 18, 53–65.

BARNDORFF-NIELSEN, O., 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 (2006): "Econometrics of testing for jumps in financial economics using bipower variation," *Journal of Financial Econometrics*, 4, 1–30.

BAUWENS, L., AND S. LAURENT (2005): "A New Class of Multivariate Skew Densities, with Application to GARCH Models," *Journal of Business and Economic Statistics*, 23, 346–354.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

——— (1987): "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics*, 69, 542–547.

——— (1990): "Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *Review of Economics and Statistics*, 72, 498–505.

BOLLERSLEV, T., R. Y. CHOW, AND K. F. KRONER (1992): "ARCH Modeling in Finance: A Review of Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5–59.

BOUDT, K., AND C. CROUX (2010): "Robust M-Estimation of Multivariate GARCH Models," *Computational Statistics and Data Analysis*, 54, 2459–2469.

BOUDT, K., J. DANÍELSSON, AND S. LAURENT (2010): "Appendix to Robust Forecasting of Dynamic Conditional Correlation GARCH Models," *Available at www.econ.kuleuven.be/kris.boudt/public.*

———— (2013): "Robust Forecasting of Dynamic Conditional Correlation GARCH Models," *International Journal of Forecasting*, 29, 244–257.

CARNERO, M., D. PENA, AND E. RUIZ (2007): "Effects of outliers on the identification and estimation of GARCH models," *Journal of time series analysis*, 28, 471–497.

———— (2008): "Estimating and forecasting GARCH volatility in the presence of outliers," IVIE working paper.

CHARLES, A., AND O. DARNÉ (2005): "Outliers and GARCH models in financial data," *Economics Letters*, 86, 347–352.

CHEN, C., AND L. LIU (1993): "Joint Estimation of Model Parameters and Outlier Effects in Time Series," *Journal of the American Statistical Association*, 88, 284296.

CREAL, D., S. KOOPMAN, AND A. LUCAS (2012): "A General Framework for Observation Driven Time-Varying Parameter Models," Forthcoming in Journal of Applied Econometrics.

DIEBOLD, F., AND M. NERLOVE (1989): "The Dynamics of Exchange Rate Volatility: a Multivariate Latent Factor ARCH Model," *Journal of Applied Econometrics*, 4, 1–21.

DOORNIK, J. (2009): *Object-Oriented Matrix Programming Using Ox.* Timberlake Consultants Press.

ENGLE, R. (2002): "Dynamic Conditional Correlation - a Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, 339–350.

FAIR, R. (2002): "Events that shook the market," *Journal of Business*, 75, 713–731.

FRANSES, P., AND H. GHIJSELS (1999): "Additive Outliers, GARCH and Forecasting Volatility," *International Journal of Forecasting*, 15, 1–9.

FRANSES, P., AND D. VAN DIJK (2000): *Non-Linear Series Models in Empirical Finance*. Cambridge University Press, Cambridge, United Kingdom.

GIOT, P., AND S. LAURENT (2003): "Value-at-Risk for Long and Short Trading Positions," *Journal of Applied Econometrics*, 18(6), 641–664.

GLOSTEN, L., R. JAGANNATHAN, AND D. RUNKLE (1993): "On the Relation Between Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

GNABO, J.-Y., S. LAURENT, AND C. LECOURT (2009): "Does Transparency in Central Bank Intervention Policy Bring Noise to the FX Market? The Case of the Bank of Japan," *Journal of International Financial Markets, Institutions & Money*, 19, 94–111.

HANSEN, P., AND A. LUNDE (2005): "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)," *Journal of Applied Econometrics*, 20, 873–889.

——— (2006): "Consistent Ranking of Volatility Models," *Journal of Econometrics*, 131, 97–121.

——— (2007): "MULCOM 2.00, Econometric Toolkit for Multiple Comparisons," http://mit.econ.au.dk/vip_htm/alunde/mulcom/mulcom.htm.

HANSEN, P., A. LUNDE, AND J. NASON (2011): "Model Confidence Sets," *Econometrica*, 79, 453–497.

HARVEY, A., AND T. CHAKRAVARTY (2008): "Beta-t-(E)GARCH," Working Paper series, University of Cambridge.

HSIEH, D. (1989): "Testing for Nonlinear Dependence in Daily Foreign Exchange Rates," *The Journal of Business,*, 62, 339–368.

HUANG, X., AND G. TAUCHEN (2006): "The Relative Contribution of Jumps to Total Price Variance," *Journal of Financial Econometrics*, 3, 456–499.

HUBER, P. (1981): *Robust Statistics.* Wiley, New York.

HULL, J. (2006): *Options, Futures and Other Derivatives.* Pearson Prentice Hall, New Jersey.

JARQUE, C., AND A. BERA (1987): "A Test for Normality of Observations and Regression Residuals," *International Statistical Review*, 55, 163–172.

JORION, P. (1988): "On jump processes in the foreign exchange and stock markets," *Review of Financial Studies*, 1, 427–445.

LAURENT, S. (2009): *G@RCH 6. Estimating and Forecasting GARCH Models.* Timberlake Consultants Ltd.

LAURENT, S., J. ROMBOUTS, AND F. VIOLANTE (2012): "On the Forecasting Accuracy of Multivariate GARCH Models," *Journal of Applied Econometrics*, 12, 934–955.

———— (2013): "On Loss Functions and Ranking Forecasting Performances of Multivariate Volatility Models," *Journal of Econometrics*, 173, 1–10.

LEE, S. S., AND P. A. MYKLAND (2008): "Jumps in financial markets: a new nonparametric test and jump dynamics," *Review of Financial studies*, 21, 2535–2563.

MAHEU, J., AND T. MCCURDY (2004): "News Arrival, Jump Dynamics, and Volatility Components for Individual Stock Returns," *Journal of Finance*, 59, 755–793.

MULER, N., D. PENA, AND V. YOHAI (2009): "Robust Estimation for ARMA Models," *The Annals of Statistics*, 37, 816–840.

MULER, N., AND V. YOHAI (2008): "Robust Estimates for GARCH Models," *Journal of Statistical Planning and Inference*, 138, 2918–2940.

NEELY, C. (2011): "A Survey of Announcement Effects on Foreign Exchange Volatility and Jumps," *Federal Reserve Bank of St. Louis Review*, 93, 361–385.

NYBERG, P., AND A. WILHELMSSON (2009): "Measuring Event Risk," *Journal of Financial Econometrics*, 7, 265–287.

PATTON, A. (2011): "Volatility Forecast Comparison Using Imperfect Volatility Proxies," *Journal of Econometrics*, 160, 246–256.

PREMINGER, A., AND R. FRANCK (2007): "Forecasting Exchange Rates: A Robust Regression Approach," *International Journal of Forecasting*, 23, 7184.

SAKATA, S., AND H. WHITE (1998): "High breakdown point conditional dispersion estimation with application to S&P 500 daily returns volatility," *Econometrica*, 66, 529–567.

TAYLOR, S. (1986): *Modelling Financial Time Series*. Wiley, New York.

VLAAR, P., AND F. PALM (1993): "The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroskedasticity and Jumps," *Journal of Business and Economic Statistics*, 11, 351–360.

Figure 1: u_t vs. z_t for a GARCH, BIP-GARCH (with $\delta = 0.975$) and GAS models with a $ST(0, 1, 5)$, and $SKST(0, 1, exp(-0.3), 5)$

Figure 2: Proportion of correct jump detections in function of m for a nominal size $\lambda = 5\%$ and $T = 2,000$

Note: The DGP is $r_t^* = r_t + m\sigma_t I_t$, where $(1-0.3L)(r_t - 0.05) = \varepsilon_t$, $\varepsilon_t \equiv \sigma_t z_t$, $z_t \overset{i.i.d.}{\sim} N(0,1)$ and $\sigma_t^2 = 0.05 + 0.02\varepsilon_{t-1}^2 + 0.93\sigma_{t-1}^2$ and finally I_t is a dummy variable accounting for the $1, 2, 5, 10, 20$ or 40 equidistant jumps.
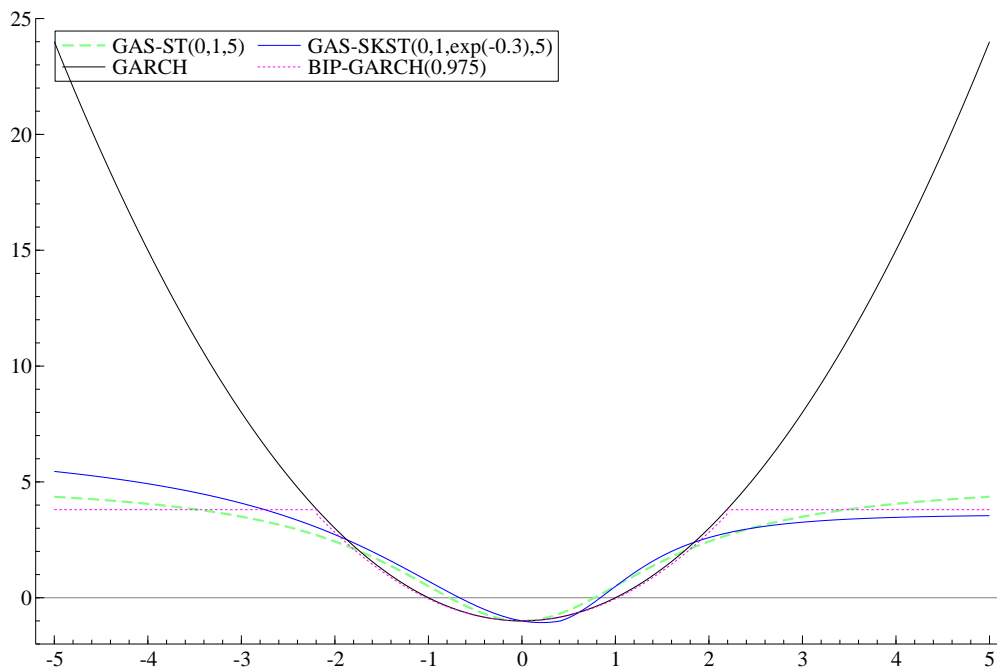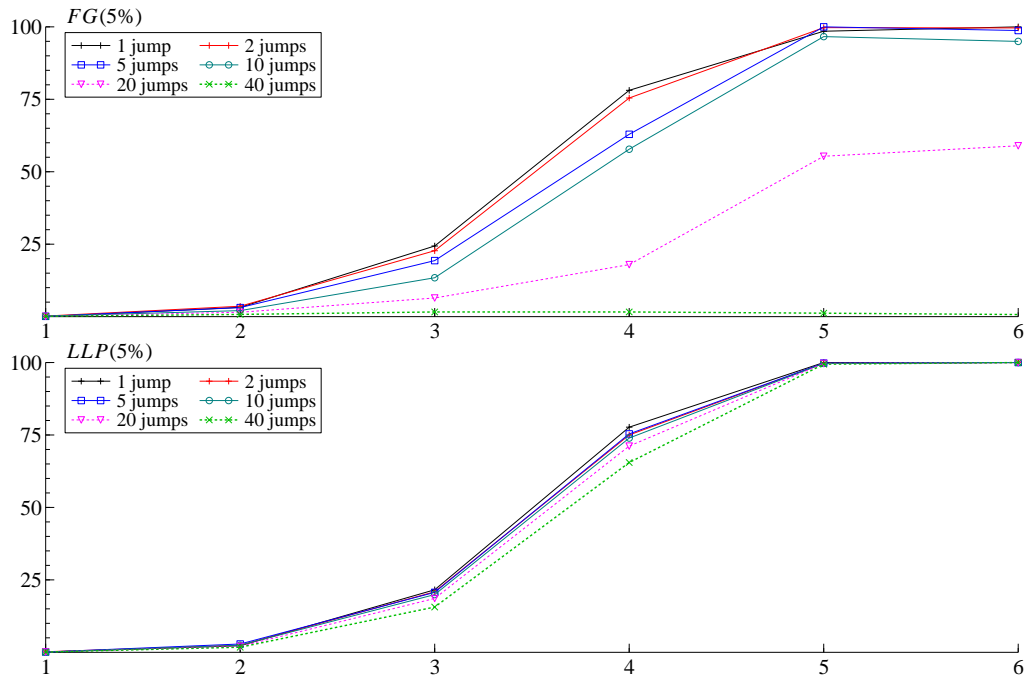
Figure 3: Proportion of correct jump detections in function of m for a nominal size $\lambda = 5$, 25 and 50% and $T = 2,000$
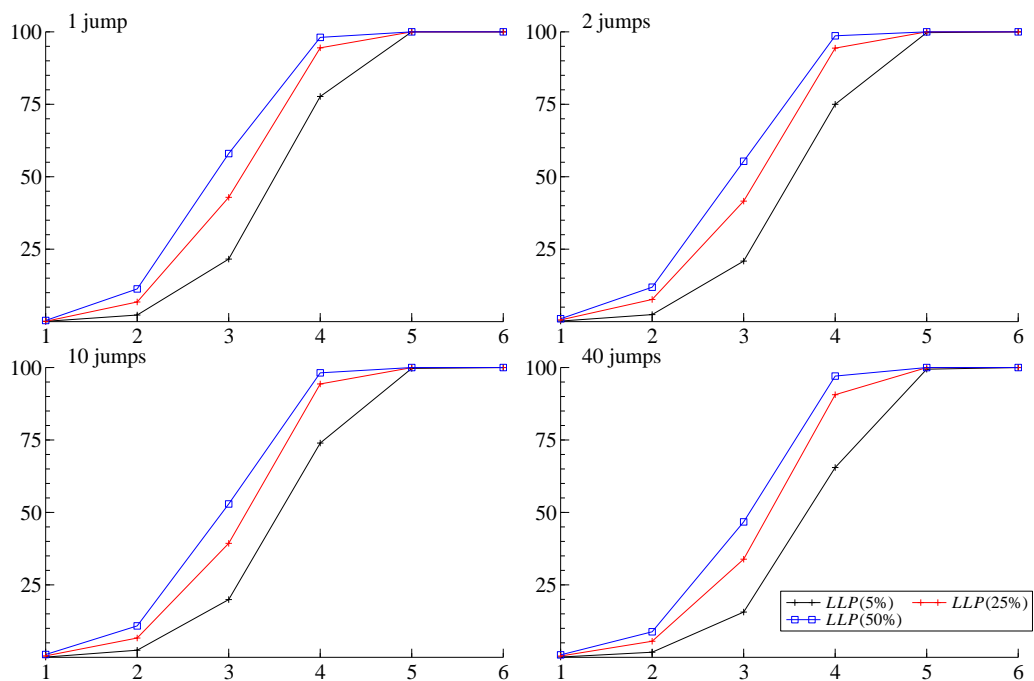
Note: The DGP is $r_t^* = r_t + m\sigma_t I_t$, where $(1-0.3L)(r_t-0.05) = \varepsilon_t$, $\varepsilon_t \equiv \sigma_t z_t$, $z_t \overset{i.i.d.}{\sim} N(0,1)$ and $\sigma_t^2 = 0.05 + 0.02\varepsilon_{t-1}^2 + 0.93\sigma_{t-1}^2$ and finally I_t is a dummy variable accounting for the $1, 2, 10$ or 40 equidistant jumps.

Figure 4: Bias as a function of the jump size m for the AR(1)-GARCH(1,1) with parameter values $\mu = 0.05, \phi_1 = 0.3, \omega = 0.05, \alpha_1 = 0.02$ and $\beta_1 = 0.93$ and 10 jumps per sample of $T = 2,000$ observations

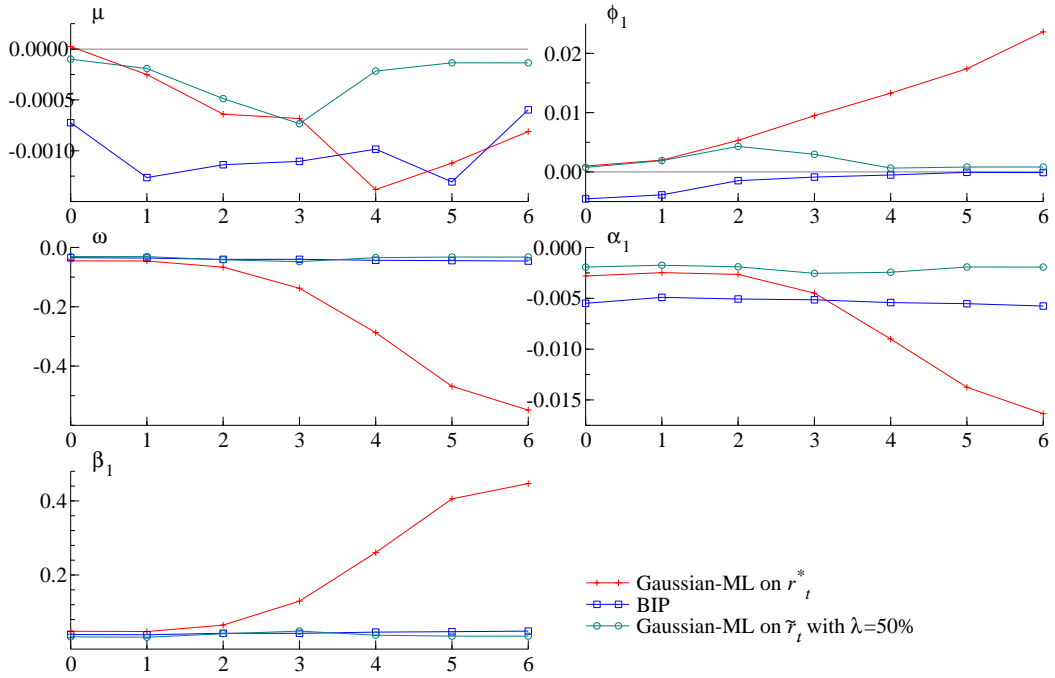
Note: The DGP is $r_t^* = r_t + m\sigma_t I_t$, where $(1-0.3L)(r_t-0.05) = \varepsilon_t$, $\varepsilon_t \equiv \sigma_t z_t$, $z_t \overset{i.i.d.}{\sim} N(0,1)$ and $\sigma_t^2 = 0.05 + 0.02\varepsilon_{t-1}^2 + 0.93\sigma_{t-1}^2$ and finally I_t is a dummy variable accounting for the 10 equidistant jumps. The estimation methods are: Gaussian-ML on r_t^*, M-estimator of the BIP-ARMA–BIP-GARCH with $\delta = 0.975$ and $\rho(\cdot) = \rho_{t_4}(\cdot)$ (denoted BIP) and Gaussian-ML on filtered returns $\tilde{r}_t$ (with $\lambda = 50\%$).

Figure 5: RMSE as a function of the jump size m for the AR(1)-GARCH(1,1) with parameter values $\mu = 0.05, \phi_1 = 0.3, \omega = 0.05, \alpha_1 = 0.02$ and $\beta_1 = 0.93$ and 10 jumps per sample of $T = 2,000$ observations
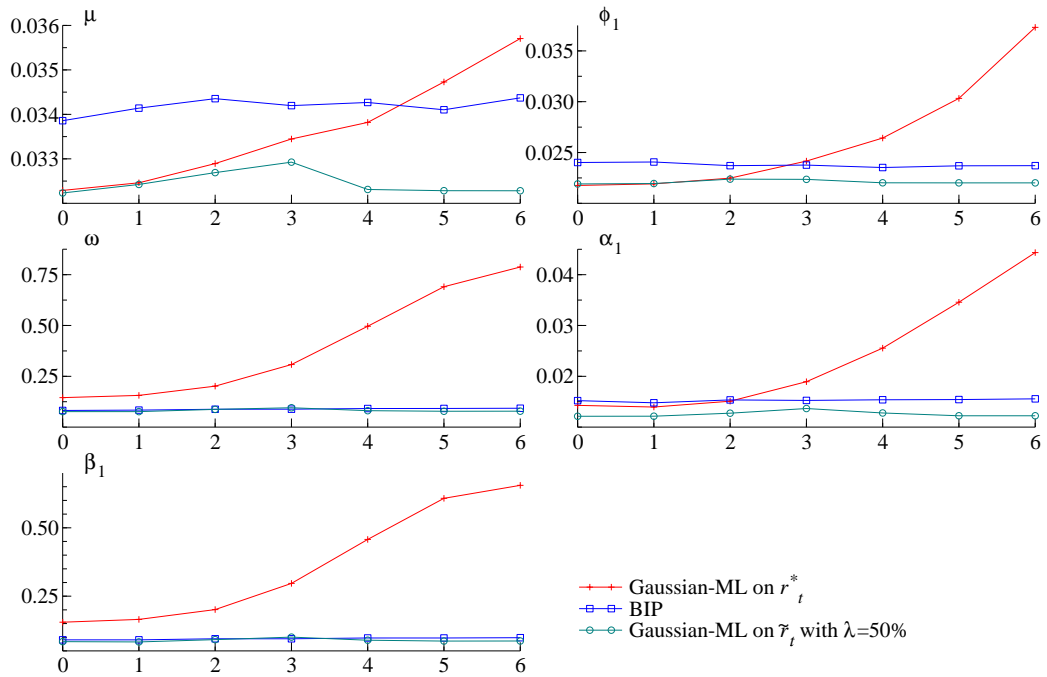
Note: See Figure 4.

Figure 6: 95% coverage probabilities as a function of the jump size m for the AR(1)-GARCH(1,1) with parameter values $\mu = 0.05, \phi_1 = 0.3, \omega = 0.05, \alpha_1 = 0.02$ and $\beta_1 = 0.93$ and 10 jumps per sample of $T = 2,000$ observations
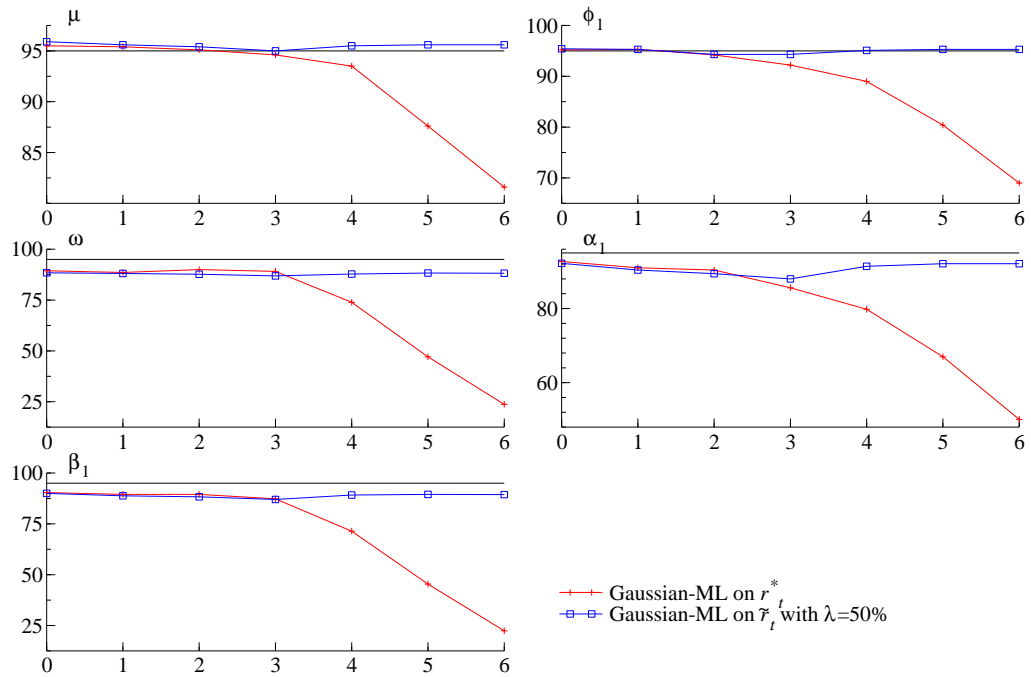

Note: See Figure 4.

Figure 7: Daily returns in % of the three exchange rates over the period January 2005 - May 2011 and detected jumps
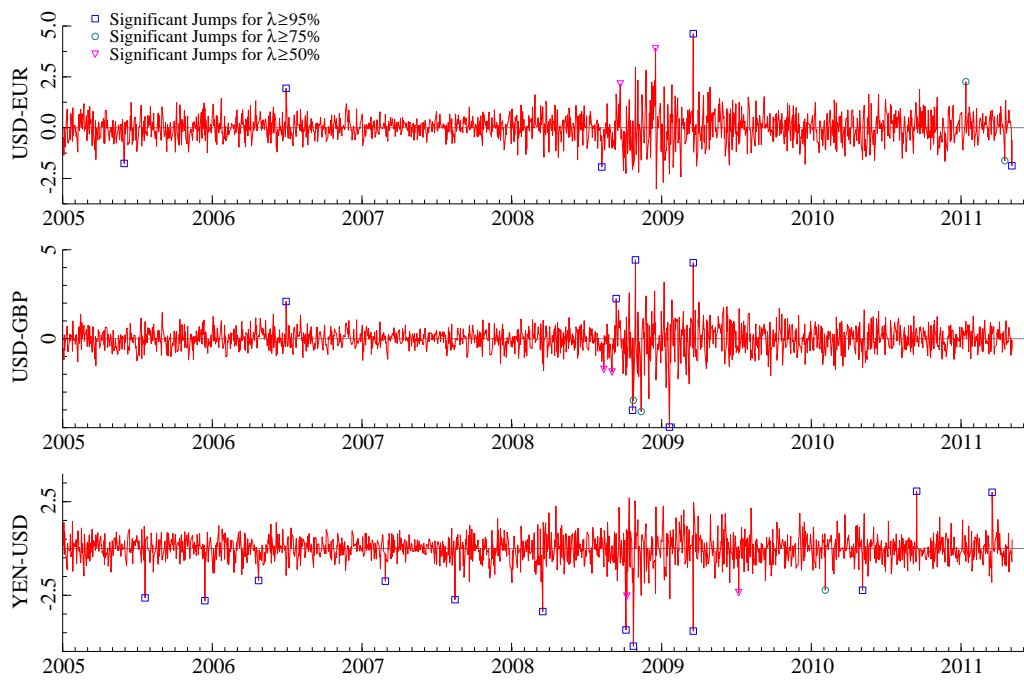
Table 1: Rejection frequencies under H_0 of no jumps for $T = 2000$

α_1	β_1	FG(5%)	LLP(5%)	LLP(25%)	LLP(50%)
0.02	0.93	7.4	5.3	26.3	48.8
0.03	0.92	7.7	4.9	25.4	49.3
0.04	0.91	5.6	4.0	26.0	49.0
0.05	0.90	7.8	3.7	25.6	49.0
0.06	0.89	6.5	4.4	22.6	47.0
0.07	0.88	7.5	5.3	24.8	50.0
0.08	0.87	6.2	5.0	24.6	48.4
0.09	0.86	8.3	4.4	25.6	48.9
0.10	0.85	6.0	3.9	25.2	49.1

Column labelled FG corresponds to Franses and Ghijsels' test while those labelled LLP correspond to our proposed test. The value in parenthesis is the nominal size (λ) of the test. The DGP is $(1 - 0.3L)(r_t - 0.05) = \varepsilon_t$, where $\varepsilon_t \equiv \sigma_t z_t$, $z_t \overset{i.i.d.}{\sim} N(0,1)$ and $\sigma_t^2 = 0.05 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.

Table 2: Rejection Frequencies of Jarque-Bera's test on the standardized residuals of a normal-AR(1)-GARCH(1,1) on $\tilde{r}_t$. Nominal size = 5%

Jumps	m	$FG(5\%)$	$LLP(5\%)$	$LLP(25\%)$	$LLP(50\%)$
1	1	4.8	4.5	3.7	3.7
	2	5.1	5.1	3.5	3.3
	3	7.0	6.4	4.2	3.8
	4	7.1	6.4	3.8	3.7
	5	5.3	4.9	3.8	3.8
	6	5.2	4.9	3.8	3.8
2	1	5.5	5.2	4.1	4.3
	2	6.1	6.0	4.2	3.7
	3	10.5	8.8	4.2	3.3
	4	10.1	8.6	3.9	3.9
	5	5.6	5.1	3.9	3.9
	6	6.0	5.1	3.9	3.9
5	1	5.0	4.4	3.7	3.6
	2	8.0	7.1	4.1	3.5
	3	34.8	30.0	8.1	4.1
	4	23.8	18.3	3.8	3.5
	5	9.1	5.4	3.7	3.6
	6	8.3	5.3	3.8	3.6
10	1	5.1	5.1	4.0	3.7
	2	20.6	18.8	8.3	3.7
	3	83.5	78.4	27.7	8.9
	4	67.2	44.9	4.7	3.0
	5	9.4	5.2	3.2	3.2
	6	13.5	4.8	3.2	3.2
20	1	5.8	5.5	3.0	3.2
	2	51.0	47.9	24.2	10.9
	3	99.9	99.8	87.8	44.7
	4	100.0	91.3	11.8	4.4
	5	53.5	6.9	3.9	3.6
	6	52.3	6.5	4.0	3.7
40	1	6.5	6.2	3.8	3.7
	2	91.1	89.8	69.6	45.8
	3	100.0	100.0	100.0	99.3
	4	100.0	100.0	57.9	9.5
	5	100.0	13.0	6.2	5.0
	6	100.0	9.1	6.1	4.8

Table 3: Simulation results for an AR(1)-GAS(1,1) model with $z_t \sim ST(0, 1, \upsilon)$.

υ	% Jumps	Rej. Freq. $JB(5\%)$
3	1.372	100.0
4	0.812	100.0
5	0.511	99.9
6	0.363	99.2
7	0.275	96.5
8	0.206	91.6
9	0.165	86.8
10	0.134	80.4
11	0.115	73.1
12	0.095	64.7
13	0.080	60.2
14	0.073	54.2
15	0.065	49.8

Column '% Jumps' reports the %age of spurious jumps detected by our test $LLP(50\%)$ when the DGP is an AR(1)-GAS(1,1) with $ST(0, 1, \upsilon)$ innovations. Column 'Rej. Freq. JB(5%)' corresponds to the rejection frequency of the JB test (at the 5% critical level) on the standardized residuals of an AR(1)-GARCH(1,1) model estimated by Gaussian ML on the filtered returns $\tilde{r}_t$ (with $\lambda = 50\%$).

Table 4: Jarque-Bera test and number of detected jumps for the exchange rate series

	USD-EUR	USD-GBP	YEN-USD
Normal-ARMA-GARCH(1,1) on raw data	26.925 (0.0000)	18.505 (0.0000)	625.672 (0.0000)
Normal-ARMA-GARCH(1,1) on $\tilde{r}_t$ with $\lambda = 50\%$	1.615 (0.4460) [5]	11.001 (0.0041) [6]	8.862 (0.0119) [12]
Normal-ARMA-GARCH(1,1) on $\tilde{r}_t$ with $\lambda = 75\%$	0.326 (0.8493) [7]	6.523 (0.0383) [8]	7.103 (0.0287) [13]
Normal-ARMA-GARCH(1,1) on $\tilde{r}_t$ with $\lambda = 95\%$	0.324 (0.8502) [9]	4.705 (0.0951) [10]	5.438 (0.0659) [15]

Note: Jarque-Bera statistic computed on the standardized residuals of an ARMA-GARCH(1,1) estimated on the raw data and the filtered returns $\tilde{r}_t$ (using our test statistic with $\lambda = 50, 75$ and 95%). Numbers in parentheses are the p-values based on the asymptotic distribution of the JB test while numbers between squared brackets are the number of detected jumps.

Table 5: Detected jumps and financial news on the USD-EUR exchange rate over the period January 2005 - May 2011

Date	$|\tilde{J}_t|$	News
2005-05-31	3.77	The euro fell sharply on Wednesday as a report that the failure of European monetary union (EMU) had been discussed at a meeting that included senior German policy makers jangled the market's already frayed nerves (Reuters news, 31/05/2005).
2006-06-30	3.92	The dollar tumbled Thursday after the Federal Reserve surprised investors with a policy statement the market took as sign of a pause in the central bank's tightening cycle. While the Fed's decision to raise its benchmark interest rate by 25 basis points to 5.25% was largely expected, the accompanying policy statement caused the dollar to drop immediately (Dow Jones, 30/06/2006).
2008-08-08	3.77	The euro extended losses versus a broadly stronger dollar on Friday to fall more than 1 percent as concerns about the region's growth outlook weighed on the European currency. The euro has been pressured after investors interpreted comments from European Central Bank President Jean-Claude Trichet on Thursday that growth had slowed more than expected as a sign that euro zone rates are unlikely to rise (Reuters news, 8/10/2008).
2008-09-22	3.22	The U.S. dollar tumbled, hitting multiweek lows against the euro and against sterling Monday as the U.S. government's bailout plan to ease a financial credit crisis reignited worries about the country's massive budget deficit (Reuters news, 22/09/2008).
2008-12-17	3.16	The euro extended its decline against the dollar Thursday afternoon, dropping below $1.4200.The common currency fell under pressure Thursday after the European Central Bank announced that it will cut the rate of interest on deposits held at the central bank in an attempt to discourage inflows (Dow Jones, 18/12/2008).
2009-03-19	5.01	The U.S. dollar hit a two-month low on Thursday after its biggest one-day fall in at least 25 years when the U.S. Federal Reserve announced it would buy long-dated debt, a move that also lifted stock markets sharply (Reuters news, 19/03/2009).
2011-01-13	3.43	The euro notched its biggest rise against the dollar in more than six months on Thursday following solid European debt auctions and after the head of the European Central Bank cited risks of short-term inflation pressures (Reuters news, 13/01/2011).
2011-04-18	3.48	The euro fell one percent against the dollar on Monday, pressured by talk of possible debt restructuring in Greece and uncertainty about Portugal's bailout after a strong showing by a euro-sceptic party in Finnish elections (Reuters news, 18/04/2011).
2011-05-05	3.76	The euro tumbled below $1.46 Thursday and was on track for its worst day against the dollar since November after the European Central Bank signaled that interest rates were unlikely to rise next month (Reuters news, 05/05/2011).

Table 6: Detected jumps and financial news on the USD-GBP exchange rate over the period January 2005 - May 2011

Date	$|\tilde{J}_t|$	News
2006-06-30	4.14	Sterling rallied one percent on the day against the U.S. dollar on Friday, extending gains after the U.S. Federal Reserve reinforced market expectations it may be nearing the end of its two-year tightening cycle (Reuters news, 30/06/2006).
2008-08-13	3.23	Sterling was set for its biggest daily percentage loss this year on Wednesday, tumbling to a 22-month low versus the dollar after new Bank of England inflation forecasts raised expectations of an interest rate cut (Reuters news, 13/08/2008).
2008-09-02	3.20	A hefty fall in oil prices boosted the dollar to a seven-month high against a basket of currencies, putting the pound under more selling pressure after weekend comments from the UK finance minister that economic conditions are at their worst in 60 years (Reuters news, 02/09/2008).
2008-09-12	3.59	Sterling rose on Friday, bouncing off a 2 1/2-year low against the dollar as weak U.S. economic data stalled a broad rally in the U.S. currency, while investors trimmed bets on an imminent UK interest rate cut (Reuters news, 12/09/2008).
2008-10-22	4.15	Earlier in the day, sterling fell to a low of \$1.6510, its lowest since September 2003, after Bank of England Governor Mervyn King said on Tuesday that Britain's economy is probably entering its first recession in 16 years (Reuters news, 22/10/2008).
2008-10-24	3.38	The pound's sharp fall against the dollar on Friday after British output shrank for the first time in 16 years is not a condemnation of the country's economic policy, a junior finance minister said (Reuters news, 24/10/2008).
2008-10-29	4.04	Sterling extended early gains against a broadly softer dollar on Wednesday, taking a lead from rising stock markets as investors eased up on extreme risk aversion that had dominated sentiment recently (Reuters news, 29/10/2008).
2008-11-12	3.42	Sterling was pummelled broadly on Wednesday, sliding below \$1.50 versus the dollar to a near 6-1/2 year low, as the Bank of England's inflation report showed a fast-shrinking UK economy, clearing a path to more rate cuts (Reuters news, 12/11/2008).
2009-01-20	3.74	Sterling fell sharply on Tuesday as Britain's latest bank rescue plan did little to assure investors about the ailing banking sector and raised concerns about the government's ability to service its ballooning debt (Reuters news, 20/01/2009).
2009-03-19	3.68	Sterling rose to a three-week high against the dollar on Thursday, reversing early losses, as the U.S. currency weakened broadly after the U.S. Federal Reserve's surprise move to buy long-term Treasuries (Reuters news, 19/03/2009).

Table 7: Detected jumps and financial news on the YEN-USD exchange rate over the period January 2005 - May 2011

Date	$|\tilde{J}_t|$	News
2005-07-21	4.73	The dollar fell sharply against the yen in Europe Thursday on news that China has revalued its currency, the yuan (Dow Jones, 21/07/2005).
2005-12-14	6.12	The dollar tumbled after a shift in rhetoric by the Federal Reserve following its interest rate rise on Tuesday signaled that the central bank was one step closer to ending its 18-month credit tightening streak. A slightly weaker-than-expected Bank of Japan tankan survey of business confidence gave the dollar a slight boost at first, but then an array of investors stepped in to sell, particularly against the yen (Reuters, 14/12/2005).
2006-04-24	3.60	The dollar fell to a fresh three-month low against the yen on Monday, extending losses after the Group of Seven powers stepped up pressure on China to let its yuan currency appreciate (Reuters, 24/04/2006).
2007-02-27	3.75	Dollar/yen rebounds to Y118.20 after a massive yen short-covering sends the pair to Y117.50 in the previous session. Traders say expectations for Japanese corporate month-end dlr buying make speculators to trim short positions in early Tokyo trading (Reuters news, 27/02/2007).
2007-08-16	4.89	Yen vols soar as investors scramble for protection.edge funds and portfolio managers are flocking to currency options for protection against bigger yen gains as market players abandon carry trades on the deepening problems in the credit market (Reuters, 16/08/2007).
2008-03-17	4.32	Asia Forex:Dlr Falls Again As Fed Fails To Calm Markets. The dollar tumbled to its lowest point in more than 12 1/2 years, hitting Y95.77 in Asia on Monday as the Fed's discount rate cut failed to calm markets amid growing fears of more U.S. bank write-downs to come (Dow Jones, 17/03/2008).
2008-10-06	6.09	Yen holds hits huge gains against major currencies – posting biggest 1-day rise vs USD since the 1998 carry trade unwind – as the credit crisis reaches a panic stage across global markets, spurring a massive unwind of carry trades and rush to the safe-haven currency (Reuters news, 7/10/2008).
2008-10-08	3.21	The safe-haven yen edged up against the euro and other major currencies on Wednesday in calmer trade as investors awaited a coordinated response by major powers to help stem the deepening credit crisis (Reuters news, 08/10/2008).
2008-10-24	5.26	The yen jumped to a 13-year high against the U.S. dollar and a nearly six-year high versus the euro in Tokyo on Friday, as Asian stocks tumbled on worries of a prolonged global recession, leading investors to buy back the yen in a hurry to offload high-risk investments (Dow Jones, 24/10/2008).
2009-03-19	5.18	US dollar slides to 2-month low after Fed move. The U.S. dollar hit a two-month low on Thursday after its biggest one-day fall in at least 25 years when the U.S. Federal Reserve announced it would buy long-dated debt, a move that also lifted stock markets sharply (Reuters, 19/03/2009).
2009-07-08	3.21	Japan could be one step closer to intervening in the foreign exchange market for the first time in five years as a soaring yen further jeopardizes the country's chances of pulling out of recession (Reuters news, 08/07/2009).
2010-02-04	3.37	The dollar and yen soared Thursday as fears about the euro zone's worsening fiscal problems and a less-than-stellar U.S. jobless claims report drove investors to abandon riskier assets for traditional safe havens (Reuters news, 04/02/2010).
2010-05-06	3.67	The U.S. dollar extended losses against the Japanese yen Thursday to trade at a session low, amid persisting fears of financial contagion in Europe (6/05/2010, Reuters news).
2010-09-15	5.24	The yen fell sharply against the dollar Wednesday after Japan intervened in currency markets for the first time in more than six years (Dow Jones, 15/09/2010).
2011-03-18	4.37	The dollar spiked about 2 yen to above 81 yen on Friday, after the G7 agreed on joint intervention in the wake of the yen's surge to a record high the previous day (Reuters news, 18/03/2011).

Table 8: Mean and variance of the p-value of the Jarque-Bera test for the ARMA-GJR estimated on the filtered returns $\tilde{r}_t$ and %age of detected jumps for the US stocks (for $\lambda = 50\%$)

Ticker	Mean	Variance	% Jumps	Ticker	Mean	Variance	% Jumps
AAPL	0.600	0.038	1.768	JNJ	0.685	0.020	2.170
ABT	0.646	0.033	2.009	KO	0.792	0.014	1.647
AXP	0.814	0.007	1.125	LLY	0.278	0.024	1.728
BA	0.464	0.014	1.607	MCD	0.477	0.024	1.527
BAC	0.116	0.004	1.808	MMM	0.667	0.033	1.848
BMY	0.667	0.012	2.089	MOT	0.634	0.022	1.487
BP	0.819	0.010	1.647	MRK	0.227	0.017	2.089
C	0.476	0.028	2.250	MS	0.690	0.015	2.049
CAT	0.822	0.027	2.049	MSFT	0.352	0.027	1.969
CL	0.079	0.002	1.687	ORCL	0.575	0.014	1.446
CSCO	0.378	0.046	1.888	PEP	0.319	0.021	2.933
CVX	0.697	0.016	1.687	PFE	0.722	0.016	1.125
DELL	0.580	0.021	1.366	PG	0.437	0.018	2.652
DIS	0.649	0.115	1.326	QCOM	0.429	0.024	1.969
EK	0.846	0.015	2.853	SLB	0.692	0.050	1.326
EXC	0.641	0.035	1.969	T	0.819	0.017	1.728
F	0.818	0.009	1.848	TWX	0.101	0.004	2.611
FDX	0.159	0.016	1.808	UN	0.377	0.076	1.406
GE	0.204	0.010	1.487	VZ	0.249	0.013	1.045
GM	0.454	0.029	2.933	WFC	0.793	0.030	2.290
HD	0.528	0.010	1.446	WMT	0.803	0.025	1.647
HNZ	0.828	0.015	1.326	WYE	0.585	0.005	1.728
HON	0.719	0.069	1.848	XOM	0.336	0.028	1.125
IBM	0.768	0.021	1.527	XRX	0.158	0.003	1.928
INTC	0.405	0.020	1.286				

Note: The mean and variance of the p-value of the Jarque-Bera test are computed from the 1,500 p-values obtained from the innovations of the sequence of models, each based on 980 observations, and used to produce the h-step-ahead forecasts. Recall that the parameters of the model are kept constant during 50 days and that the model is therefore re-estimated every 50 days on a rolling window of 980 observations.

Table 9: Results of the MCS test

Period / h	Groups		
	GARCH	GAS	BIP
Full period (2002-2008)			
1	75.510	87.755	85.714
2	65.306	73.469	89.796
3	57.143	67.347	89.796
4	57.143	65.306	93.878
5	59.184	63.265	93.878
6	57.143	63.265	89.796
7	53.061	63.265	93.878
8	48.980	61.224	97.959
9	48.980	63.265	97.959
10	46.939	63.265	95.918
Calm period (2002-2007)			
1	24.490	30.612	97.959
2	16.327	18.367	100.000
3	16.327	16.327	100.000
4	14.286	16.327	100.000
5	14.286	14.286	100.000
6	14.286	14.286	100.000
7	14.286	14.286	100.000
8	16.327	14.286	100.000
9	14.286	14.286	100.000
10	12.245	16.327	100.000

Note: % of times that a model belonging to the corresponding group is part of the MCS (with significance level of 10% and 10,000 bootstrap resamples with block length of 5 observations). The groups are defined as follows. Group 1: GARCH = {ARMA-GARCH(1,1) and ARMA-GJR(1,1) models with a Normal, Student-t and Skewed-Student distribution.} Group 2: GAS = {ARMA-GAS(1,1), ARMA-EGAS(1,1) and ARMA-AEGAS(1,1) models with a Student-t and Skewed-Student distribution.} Group 3: BIP = {ARMA-GARCH(1,1) and ARMA-GJR(1,1) models with a Normal distribution estimated on filtered returns ($\tilde{r}_t$), as given in Equation (2.7), where jumps are obtained using (2.6).}