Testing for jumps in GARCH models, a robust approach

Sébastien Laurent* Christelle Lecourt† Franz C. Palm‡

June 10, 2011

Abstract

Financial series occasionally exhibit large changes. Assuming that the observed return series consist of a standard normal ARMA-GARCH component plus an additive jump component, we propose a new test for additive jumps in an ARMA-GARCH context. The test is based on standardised returns, where the first two conditional moments are estimated in a robust way. Simulation results indicate that the test has very good finite sample properties, i.e. correct size and much higher proportion of correct jump detection than Franses and Ghijsels's (1999) test. We apply our test on the YEU-USD exchange rate and find twice as much jumps as Franses and Ghijsels's (1999) test.

*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
We are also grateful to Kris Boudt for helpful comments and suggestions. The usual disclaimer applies.

1 Introduction

It is well known that high-frequency returns of most financial assets exhibit volatility clustering but also large jumps caused by big surprises (e.g. news announcements). Andersen, Bollerslev, and Diebold (2007), Harvey and Chakravarty (2008) and Muler and Yohai (2008), among others, found that these jumps affect future volatility less than what standard volatility models would predict. In a realized volatility context, Andersen, Bollerslev, and Diebold (2007) show that in an autoregressive (AR) model conditioning also on past jumps improves the predictions of future realized volatility.

In a univariate GARCH context, 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 Gaussian quasi-maximum likelihood (QML) estimates of GARCH models tend to overestimate the volatility for the days following the jumps but also produce upward biased estimates of the long-term volatility.

The impact of jumps has been modeled assuming a Poisson or a Bernoulli jump distribution which when combined with a normal distribution for the Brownian motion part leads to Poisson or Bernoulli mixtures of distributions for financial returns (see e.g. Ball and Torous, 1983). Alternatively some studies assume fat tail distributions such as the student-t or the generalized error distribution to account for the occurrence of large changes in returns. This literature was not aimed at jump detection and testing for jump.

The effect of jumps on multivariate GARCH models has also been investigated recently by Boudt and Croux (2010) and Boudt, Daníelsson, and Laurent (2010b), respectively in the BEKK and dynamic conditional correlation (DCC) frameworks. Boudt, Daníelsson, and Laurent (2010b) show that the unconditional and conditional correlations given by the constant conditional correlation (CCC) model of Bollerslev (1990) and the DCC model of Engle (2002) are strongly affected by these jumps. They also compare the conditional covariance forecasts of obtained for various multivariate GARCH models including the DCC and their robust version with ex post covariance estimates based on high-frequency data (i.e. the realized covariance of the EUR/USD and Yen/USD exchange rates over the period 2004-2009). Using the model confidence set methodology, proposed by Hansen, Lunde, and Nason (2010), they find that their robust DCC model always belongs to the set of superior forecasting models. Moreover, for most forecast horizons, their covariance forecasts are significantly better than all other models considered.

Our goal in this paper is to propose a new statistical test procedure to detect additive jumps and to study its statistical properties. The performance of the test is investigated by means of a Monte Carlo simulation and it is compared with that of the test proposed by Franses and Ghijsels (1999). We apply our and the Franses-Ghijsels tests to daily returns for the YEN-USD exchange rate for the period January 2005 to May 2011.

The main advantages of the new test over the one of Franses and Ghijsels (1999) are that

1. all jumps are detected at once in a single test;
2. critical values do not need to be simulated as the asymptotic distribution of the test does not depend on nuisance parameters;
3. we can control for the type-I error (probability of rejecting the null of no jump in the sample, under the null);
4. the proportion of correct jump detection is much higher, that is more powerful than the procedure by Franses and Ghijsels (1999).

While being designed for data observed at lower frequencies, our test is much in the spirit of the nonparametric test put forward by Lee and Mykland (2009) for high frequency data. Similar to Lee and Mykland (2008), who use standardize their nonparametric test for high frequency data by a consistent estimate of instantaneous volatility (they use bipower variation to estimate instantaneous volatility), we standardize our test using the conditional volatility based on a GARCH-model (in absence of jumps) and using a robustified GARCH volatility estimate (in case jumps are likely to affect the GARCH process). Our test therefore incorporates the idea that when spot or instantaneous volatility is high (also in the absence of jumps), returns may also be high, even as high as that due to jumps. Franses and Ghijsels (1999) standardize their test statistics using unconditional residual variances estimates. Our test is expected to be useful especially when intraday data are not available and thus when realized volatility estimates cannot be computed.

Finally, Hotta and Tsay (1998) propose a Lagrange multiplier test for additive levels outliers and for additive volatility outliers. Doornik and Ooms (2005) propose a likelihood ratio test, to test first the occurrence and timing of an outlier and then in a second step to determine the type of additive outlier, either in the mean or in volatility. As these types of tests require the specification of a distribution of the data under the null hypothesis, they are likely to be less robust than tests based on the Quasi-ML-method. Charles and Darné (2005) extend the test for additive outliers proposed by Franses and Ghijsels (1999) 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.

In an application to the Yen-USD exchange rate, it appears that the jumps that our test procedure detects ate related to news and interventions by the Bank of Japan.

2 Model and test

2.1 Data generating process

The data generating process (DGP) assumes that the observed return series r_t^* $(t = 1, \ldots, T)$ consist of a standard normal ARMA(p, q)-GARCH(1,1) component r_t and an additive jump component, i.e.

$$r_t^* = r_t + a_t I_t \tag{2.1}$$

$$\phi(L)(r_t - \mu) = \theta(L)\varepsilon_t \text{ where } \varepsilon_t \equiv \sigma_t z_t \text{ and } z_t \overset{i.i.d.}{\sim} N(0,1) \tag{2.2}$$

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

where a_t corresponds to the size of the jump, I_t is a dummy variable that takes the value 1 if there is a jump at time t and 0 otherwise, L is the lag operator while $\phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$ and $\theta(L) = 1 - \sum_{i=1}^{q} \theta_i L^i$ with roots outside the unit circle.

Let $\lambda(L) = \phi^{-1}(L)\theta(L) = 1 + \sum_{i=1}^{\infty} \lambda_i L^i$. Equation (2.2) can be rewritten as follows

$$r_t = \mu_t + \varepsilon_t, \quad (2.4)$$

$$\mu_t = \mu + \sum_{i=1}^{\infty} \lambda_i \varepsilon_{t-i}, \quad (2.5)$$

where μ_t is the conditional mean of r_t.

2.2 Jumps detection

2.2.1 Franses and Ghijsels (1999)

One of the most popular method for additive jumps detection in a GARCH framework 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 GARCH models.

Franses and Ghijsels (1999) 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 like Franses and Ghijsels (1999) 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^*:

$$r_t^* = r_t(1 - I_t) + (\mu_t + \varepsilon_t^*)I_t \quad (2.6)$$

$$= (\mu_t + \varepsilon_t) + (\mu_t + \varepsilon_t^* - \mu_t - \varepsilon_t)I_t \quad (2.7)$$

$$= r_t + (\varepsilon_t^* - \varepsilon_t)I_t, \quad (2.8)$$

where r_t is defined as in (2.2)-(2.3). Note that Equation (2.8) is a particular case of Equation (2.1), 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 (2.2)-(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

$$t_{\hat{\xi}(\tau)} = (1/\hat{s}) \left(\sum_{t=\tau}^{T} x_t^2 \right)^{-1/2} \left(\sum_{t=\tau}^{T} x_t \hat{v}_t \right) \; \forall \tau = 1, \ldots, T,$$

   where $x_t = 0$ for $t < \tau$, $x_\tau = 1$ and $x_{\tau+k} = -\pi_k$ for $k = 1, \ldots$ and finally $\pi(L) = (1 - \beta_1 L)^{-1}(1 - (\alpha_1 + \beta_1)L)$ for a GARCH(1,1). $t_{\hat{\xi}(\tau)}$ corresponds to the t-statistic for the estimated slope coefficient $\hat{\xi}(\tau)$ of the regression of $\hat{v}_t$ on x_t while $\hat{s}$ is 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. Franses and Ghijsels (1999) also propose to clean the original series for the detected additive outliers by replacing $r_{\hat{\tau}}^*$ by $\hat{\mu}_\tau + \text{sign}(\hat{\varepsilon}_\tau)\sqrt{(r_\tau^* - \hat{\mu}_\tau)^2 - \hat{\xi}(\tau)}$, where $\text{sign}(x) = 1$ 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.

For the choice of the critical value, Franses and Ghijsels (1999) recommend using $C = 4$ 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. For $T = 250$ these quantiles equal 9.67 and 13.96, respectively.

2.2.2 Our jumps detection rule

The intuition behind our jump test is similar to the one proposed simultaneously by Andersen, Bollerslev, and Dobrev (2007b) and Lee and Mykland (2008). Let us denote by $\tilde{\mu}_t$ and $\tilde{\sigma}_t^2$ estimates of μ_t and σ_t^2 in model (2.1)-(2.3) that are robust to the potential presence of the additive jumps $a_t I_t$ (see Sections 2.3, 2.4 and 2.5).

Denote by $\tilde{J}_t = \frac{r_t^* - \tilde{\mu}_t}{\tilde{\sigma}_t}$ the standardised return on day t. If $I_t = 0$ on day t, $\tilde{J}_t$ should be standard normally distributed and thus standardised returns $\tilde{J}_t$ that are too large to plausibly come from a standard normal distribution must reflect jumps.

This suggests the following jumps detection rule:

$$\tilde{I}_t = I\left(|\tilde{J}_t| > k\right), \tag{2.9}$$

where $I(\cdot)$ is the indicator function and k is suitable critical value defined below. The rule described in (2.9) 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 (2.1).

A straightforward jump detection rule is that return r_t^* is taken as being affected by a jump if $|\tilde{J}_t|$ exceeds the quantile $\delta(> .5)$ of the standard Gaussian distribution. This rule has a probability of type I error (detect that r_t^* is affected by jumps, if in reality r_t^* is not) equal to $(1 - \delta)$. But its disadvantage is that the expected number of false positives over the whole estimation sample is equal to $(1-\delta)T$ under the null of no jump which can be large for large T. For instance, with $T = 1000$ and $\delta = 0.95$, 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. $\delta = 0.999$ or 0.9999. 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 of any set of T independent J-statistics $|\tilde{J}_t|$ exceeds

$$g_{T,\delta} = -\log(-\log(\delta))b_T + c_T, \tag{2.10}$$

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 $1-\delta$. All returns for which the $|\tilde{J}_t|$ exceeds $g_{T,\delta}$ should be declared as being affected by jumps.

As mentioned above, $|\tilde{J}_t|$ requires estimates of μ_t and σ_t^2 that are robust to jumps. Sections 2.3, 2.4 and 2.5 deal precisely with this.

2.3 BIP-ARMA

Muler, Pena, and Yohai (2009) (MPY) introduces a new class of estimates for ARMA models (i.e. for μ_t) that is robust to additive jumps. To robustify the estimation of the ARMA model, MPY propose to replace Equation (2.4)-(2.5) by a family of auxiliary models for the (potentially) contaminated returns r_t^*:[1]

$$r_t^* = \mu_t + \varepsilon_t + a_t I_t, \quad (2.11)$$

$$\mu_t = \mu + \sum_{i=1}^{\infty} \lambda_i \sigma_{t-i} w_{k_\delta}^{\mathrm{MPY}}\left(J_{t-i}\right), \quad (2.12)$$

where $J_{t-i} = \frac{r_{t-i}^* - \mu_{t-i}}{\sigma_{t-i}}$.

The weight function $w_{k_\delta}^{\mathrm{MPY}}(\cdot)$ in Equation (2.12) 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

$$w_{k_\delta}^{\mathrm{MPY}}(u) = \mathrm{sign}(u) \min\left(|u|, k_\delta\right). \quad (2.13)$$

Model (2.11) with the estimator of μ_t in (2.12) and weight function (2.13) is called *Bounded Innovation Propagation* (BIP)–ARMA since the effect of I_t on future values of μ_t is bounded.

Since J_{t-i} is standard normally distributed in absence of jumps at time $t-i$, it is natural to suspect the presence of a jump in r_{t-i}^* if $|J_{t-i}|$ 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. An alternative would be to compare $|J_{t-i}|$ with the critical value of the Gumbel distribution like in the previous section. We did not pursue this direction because Monte-Carlo simulation results (not reported here to save place) suggest that downweigthing too many observations is less damageable for the efficiency of this method than neglecting some small jumps.

2.4 BIP-GARCH

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.

[1]Note that in MPY, σ_{t-i} is assumed to be constant and replaced by a robust M-scale estimate of ε_t.

In this case the Gaussian QML is not appropriate because $a_{t-1}I_{t-1}$ has no impact on σ_t^2 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.

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

$$\sigma_t^2 = \omega + \alpha_1\sigma_{t-1}^2 w_{k_\delta}^{\text{MPY}}\left(J_{t-1}\right)^2 + \beta_1\sigma_{t-1}^2. \tag{2.14}$$

Model (2.14) is called *Bounded Innovation Propagation* (BIP)–GARCH(1,1). Note that extensions of the BIP-GARCH to higher GARCH orders or other more general GARCH-type specifications are trivial and not discussed here to save space.

Using the same reasoning 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.

Boudt, Daníelsson, and Laurent (2010b) propose a slightly different weight function than $w_{k_\delta}^{\text{MPY}}(\cdot)$ in a GARCH context that ensures the conditional expectation of the weighted squared unexpected shocks to be the conditional variance of r_t^* in absence of jumps, i.e.

$$w_{k_\delta}^{\text{BDL}}(u) = c_\delta^{1/2} w_{k_\delta}^{\text{MPY}}(u). \tag{2.15}$$

Boudt, Daníelsson, and Laurent (2010b) report the following values for c_δ: 1.0185, 1.0465, 1.0953, 1.2030 (for $\delta = 0.99, 0.975, 0.95,$ and 0.90 respectively).

2.5 Estimation

MPY and MY show that QML estimation of BIP-ARMA and BIP-GARCH models are not efficient in 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

$$\hat{\theta}^{\mathrm{M}} = \mathrm{argmin}_{\theta\in\Theta}\frac{1}{T}\sum_{t=1}^{T}\rho\left(2\log\left|\frac{r_t^* - \mu_t}{\sigma_t}\right|\right), \tag{2.16}$$

where μ_t and σ_t^2 are given respectively in (2.12) and (2.14).

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 Boudt, Daníelsson, and Laurent (2010a,b) recommend the one associated with the Student t_4 density function:

$$\rho_{t_4}(z) = -z + 0.8260\rho_{t_4}(\exp(z)),$$

where

$$\rho_{t_\nu}(u) = (1+\nu)\log\left(1 + \frac{u}{\nu - 2}\right). \tag{2.17}$$

To sum up, we perform the estimation of the BIP-ARMA–BIP-GARCH model in one step by minimising the objective function (2.16) 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.

Given $\tilde{\mu}_t$ and $\tilde{\sigma}_t^2$, one can apply the test for additive jumps described in (2.9) for $k = g_{T,\delta}$ and then obtain $\tilde{I}_t$, an estimate of I_t.

We propose a second robust estimation method that uses the extra information contained in $\tilde{I}_t$ about the additive jumps.

Let us denote by r_t^{**} the filtered return series obtained by replacing the returns r_t^* for which we detected a jump by a robust estimate of the conditional expectation of $r_t^* - a_t I_t$, i.e. $r_t^{**} = r_t^*(1 - \tilde{I}_t) + \tilde{\mu}_t\tilde{I}_t$.

For the M-estimators for GARCH models which minimize the average value of the objective function in (2.16), 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 proces. They would underpin the proposed use an an ARMA-GARCH model for filtered return.

3 Simulation

3.1 Data Generating Processes (DGP)

In the Monte-Carlo simulation study we simulate 5000 samples of size $T = 500, 1000, 2000$ or 3000 following an AR(1)-GARCH(1,1) model with additive jumps as described in Equations (2.1)-(2.3), with $p = 1$ and $q = 0$, $\mu = 0.05, \phi_1 = 0.3, \omega = 0.3, \alpha_1 = 0.2$ and $\beta_1 = 0.7$.

The size of the jump process a_t in Equation (2.1) is specified as follows:[2]

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

[2] We also considered the case where $a_t = (\varepsilon_t^* - \varepsilon_t)$, with $\varepsilon_t^* = \text{sign}(\varepsilon_t)\sqrt{\varepsilon_t^2 + w_t I_t}$ and $w_t = m^2 \sigma_t^2$ to simulate jumps in the spirit of Franses and Ghijsels's (1999) DGP (see Equation (2.8)). Results were found to be qualitatively the same and thus 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 8 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 qualitatively the same we only report those for the equidistant jumps in order to save place. The number of jumps per sample of T observations is set to $1, 2, 5, 10$ or 20.

3.2 Global spurious detection of jumps

Monte-Carlo simulation results reported in Franses and van Dijk (2000) state that the 95% quantile of $t_{max}(\hat{\xi})$ under the null assumption of no jump in a sample of $T = 500$ observations equals 16.93 when $\alpha_1 = 0.2$ and $\beta_1 = 0.7$. Our own Monte-Carlo simulations support this finding. For $T = 1000, 2000$ and 3000 we obtained the following values critical values (C): 21.60, 27.13 and 31.41.

These critical values are chosen such that one expects to reject

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

in 5% of the cases (type I error) when the null is true. The percentage of global spurious detection under the null of no jump (type I error) is in this case 5%.

The main drawback of this 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).

To get the same expected type I error for our proposed test in (2.9), we set k to $g_{T,0.95}$, i.e. $3.95, 4.10, 4.25$ and 4.34 for $T = 500, 1000, 2000$ and 3000 respectively. The rejection frequencies of H_0 over the 5000 replications for our test are 5.80, 5.42, 5.46 and 5.58% for these four considered sample sizes.[3] This

[3]Results reported in this paper are based on programs written by the authors using Ox

suggests that there is no evidence of 'size' distortion for our proposed test.

3.3 Ability to detect actual jumps

Another question of interest is whether the two tests have sufficient 'power' to detect actual jumps. We define the proportion of correct (resp. false) jump detections as the average number (over the 5000 replications) of correctly (falsly) detected jump days.

Figures 1 plots the proportion of correct jump detections as a function of m (jump size) for $T = 500$ and 2000 (the figures corresponding to $T = 1000$ and 3000 are available upon request but are not reported here to save space). Recall that jumps are equally spaced and the number of jumps per sample equals $1, 2, 5, 10$ or 20.

This figure clearly suggests that our test (right side) has a much higher power to detect the actual jumps than Franses and Ghijsels's (1999) test (left side).

For instance, when $T = 500$ (upper panel) the proportions of correct jump detection in presence of one additive jump of size of 3 and 4 standard deviation ($m = 3$ and 4) equal respectively 32.98% and 88.90% for our test. These proportions equal 8.53% and 23.56% for Franses and Ghijsels's (1999) test. Note that choosing a smaller quantile δ to determine the critical value $g_{T,\delta}$ would naturally lead to a higher proportion of correct jump detection, e.g. 62.10% and 77.10% instead of 32.98% for $g_{500,0.75}$ and $g_{500,0.50}$ respectively when $m = 3$.

Furthermore, it emerges from these figures that unlike Franses and Ghijsels's (1999) test, our test is not sensitive to the actual number of jumps. 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 statistical literature as outlier masking as in the presence of jumps the estimated standard-errors are large compared to the estimate of

version 6.0 (Doornik, 2009) and G@RCH version 6.0 (Laurent, 2009).

a_t rendering the test insignificant).

In the previous section we studied the size property of our test by computing the percentage of global spurious detection (i.e. no jump in the sample) under the null of no jump. Figure 2 plots the proportion of false jump detections. This figure suggests that the proportion of false jump detections for our test (right panel) is close to 5% irrespectively of the number of jumps. Franses and Ghijsels's (1999) test is found to be too conservative when the number and/or magnitude of jumps increases, explaining the low proportion of correct jump detections in these cases. For the test statistic $\tilde{I}_t$ however, the proportion of correct jump detection is close to 100% when $m \geq 4$.

3.4 Bias, MSE and 95% coverage probability

In this subsection we investigate the finite sample properties of four estimation methods both in presence and absence of jumps, i.e.

- Gaussian quasi-maximum likelihood;
- Gaussian maximum likelihood on filtered returns using the jump test of Franses and Ghijsels (1999);
- M-Estimation of the BIP-ARMA–BIP-GARCH as previously discussed in Sections 2.3, 2.4 and 2.5;
- Gaussian maximum likelihood estimation on filtered returns r_t^{**} using our proposed jump test $\tilde{I}_t$.

In order to save space we only report the results for $T = 500$ in presence of 1 jump or 5 jumps (per sample) of magnitude $m\sigma_t$ with $m = 0, 1, \ldots, 8$.

Figures 3 and 4 plot the empirical bias of $\mu, \phi_1, \omega, \alpha_1$ and β_1 over the 5000 replications as a function of the jump size m (see Section 3.1). The empirical bias of parameter θ is defined as $\frac{1}{5000}\sum_{i=1}^{5000}(\theta_0 - \hat{\theta}_i)$, where θ_0 denotes the true parameter value and $\hat{\theta}_i$ its estimate obtained at the ith iteration. We observe that the M-Estimator of the BIP-ARMA–BIP-GARCH (denoted BIP) and the Gaussian maximum likelihood on filtered returns using our proposed jump test (denoted ML on filtered returns) are more robust than the others. Interestingly, for these two methods, the bias is found to be limited for each parameter and independent of the magnitude of the jumps. In the presence of 1 jump, the bias associated with the MLE of Franses and Ghijsels's (1999) filtered returns is also limited but this method is found to be as non-robust as the QML in presence of 5 (or more) jumps.

Figures 5 and 6 plot the mean square errors (MSE) of each parameter over the 5000 replications as a function of the jump size m for DGP_1 as well. The MSE of parameter θ is defined as $\frac{1}{5000}\sum_{i=1}^{5000}(\theta_0 - \hat{\theta}_i)^2$. These two figures also suggest that the M-Estimator of the BIP-ARMA–BIP-GARCH and the Gaussian maximum likelihood on filtered returns using our proposed jump test perform better. The loss of efficiency compared to the (Q)ML is very limited in absence of jumps and they appear to be much more efficient than the other two methods in presence of jumps when $m \geq 4$ as expected given the theoretical properties of the estimators obtained by MY.

Finally, Figures 7 and 8 plot the 95% coverage probabilities for the five parameters as a function of m. 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 (5000 in our case). 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 no jump. Our simulation set-up being more general (because $\mu_t \neq 0$ and $a_t \neq 0$), we therefore do not report the 95% coverage probabilities for this estimation method.

These two figures suggest that the Gaussian maximum likelihood estimation on filtered returns r_t^{**} using our proposed jump test $\tilde{I}_t$ has a 95% coverage probability close to the theoretical value of 95% for each parameter, irrespective of the size of the jumps and the number of jumps in the sample.[4] As expected the 95% coverage probabilities of the QML deviate from their theoretical value when m increases, even in presence of 1 jump in the sample. The

[4] We obtained similar figures for 2, 10 and 20 jumps per sample and different sample size.

Gaussian maximum likelihood estimation on filtered returns using Franses and Ghijsel's (1999) test has a 95% coverage probability close to the theoretical value of 95% for each parameter in presence of 1 jump but these confidence intervals are found to be too conservative in presence of more jumps and thus any statistical inference based on this method would be misleading.

4 Application

In this section we apply the two tests for additive jumps in ARMA-GARCH models described in Section 2.2. Our objective is to examine whether $t_{max}(\hat{\xi})$ and $\tilde{I}_t$ behave differently when applied on real data and whether the detected jumps have an economic explanation.

The analysis has been carried out on the Japanese yen US dollar (Yen-USD) exchange rate over the period January 2005 - May 2011 (i.e. $T = 1598$ observations). The data have been downloaded from the FRED (Federal Reserve Economic Data) website. We choose the Yen-USD exchange rate for our empirical analysis 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 as described by the size of the different jumps selected by our method in Table 1. 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. One type of news that causes discontinuities in exchange rate prices is the occurrence of central bank interventions in the FX market 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 behavior, conducting to some discontinuities in prices. Unlike other central banks, The Bank of Japan has continued to intervene actively during these last ten years and very recently.

Figure 9 plots the daily returns in % of the Yen-USD exchange rate (solid line) and the detected jumps. Returns being identified as contaminated by an additive jump by the $t_{max}(\hat{\xi})$ (resp. $\tilde{I}_t$) statistic are highlighted by a square (resp. triangle). Returns being identified as contaminated by an additive jump by the two methods are highlighted by a circle. For the $t_{max}(\hat{\xi})$ statistic, we chose a critical value of $C = 10$ which corresponds almost to the 50% quantile of $t_{max}(\hat{\xi})$ for $T = 1598$, $\alpha = 0.043181$ and $\beta = 0.949687$ (the M-estimates of the BIP-ARMA–BIP-GARCH(1,1) model).[5] The critical value of $\tilde{I}_t$ giving the same type-I error is $g_{1598,0.5} = 3.52724$. Note that the probability of finding at least one spurious additive jump while there is no jump in the data is thus 50% for both tests.

Table 1 reports the dates of all the detected jumps, the jump statistics $t_{max}(\hat{\xi})$ and $\tilde{I}_t$ as well as an indication about the significance of these two statistics. The last column, labelled 'Event', reports real-time financial news and information released around jump arrival days using the Factiva database in order to examine their association with jump arrivals.[6] Sources used in the Factiva search include Dow Jones and Reuters newswires.

The main findings are that our test identifies twice as many jumps as the $t_{max}(\hat{\xi})$ statistic for the same expected level of type-I error and that all jumps detected by the latter are also detected by the former. Importantly all the detected jumps have been largely documented by the newswires services and all news reports extracted the same day than 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, on the 18th of March 2011. The Biggest jumps detected in 2008 are related to the credit crisis period. This suggests that jumps detected by our method are likely not to be spurious.

[5]We found that over 1000 replications, the 50% quantile of $t_{max}(\hat{\xi})$ under the null of no jump is 9.975 for this sample size and these parameter values.

[6]The purpose of this exercise is not to identify the direction of the causality between jumps and these news, i.e. whether the created the jumps or vice-cersa. For this, we would need the timing of the discontinuities that create jumps and compare this with the timing of the arrival of these news.

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), we proposed a new test for additive jumps in ARMA-GARCH models. The distribution under the null hypothesis of the proposed test follows from the consistency and asymptotic normality of the parameters 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 much higher power to detect the actual jumps than Franses and Ghijsels's (1999) test 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.

It is interesting and of importance that in our application the detected jumps for the Yen-US Dollar exchange rate appear to be related to economic events (news and interventions by the Bank of Japan) that are reported by Dow Jones and Reuters newswires. Issues for future research are a better theoretical underpinning of the robustness of our findings for the power of the test. It would also be interesting to investigate the properties of our test for other types of models, jumps (e.g. innovative outliers) and observation frequencies.

References

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.

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

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 (2010a): "Appendix to Robust Forecasting of Dynamic Conditional Correlation GARCH Models," *Available at www.econ.kuleuven.be/kris.boudt/public.*

——— (2010b): "Robust Forecasting of Dynamic Conditional Correlation GARCH Models," *Mimeo.*

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.

Figure 1: Proportion of correct jump detections in function of m

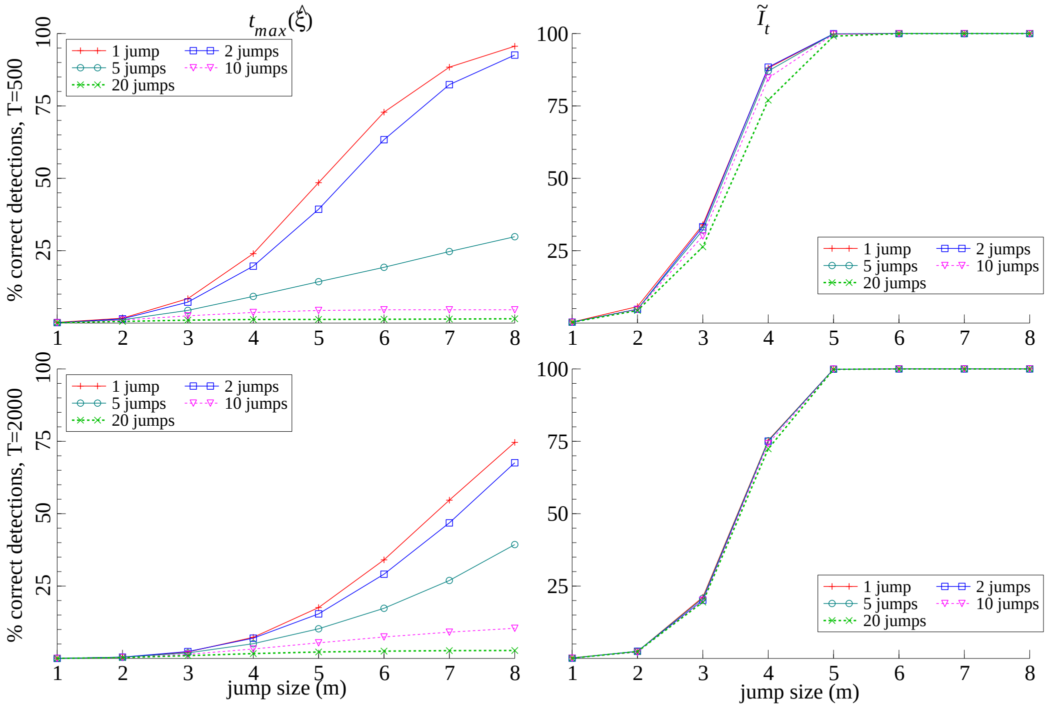

Figure 2: Proportion of false jump detections in presence of jumps

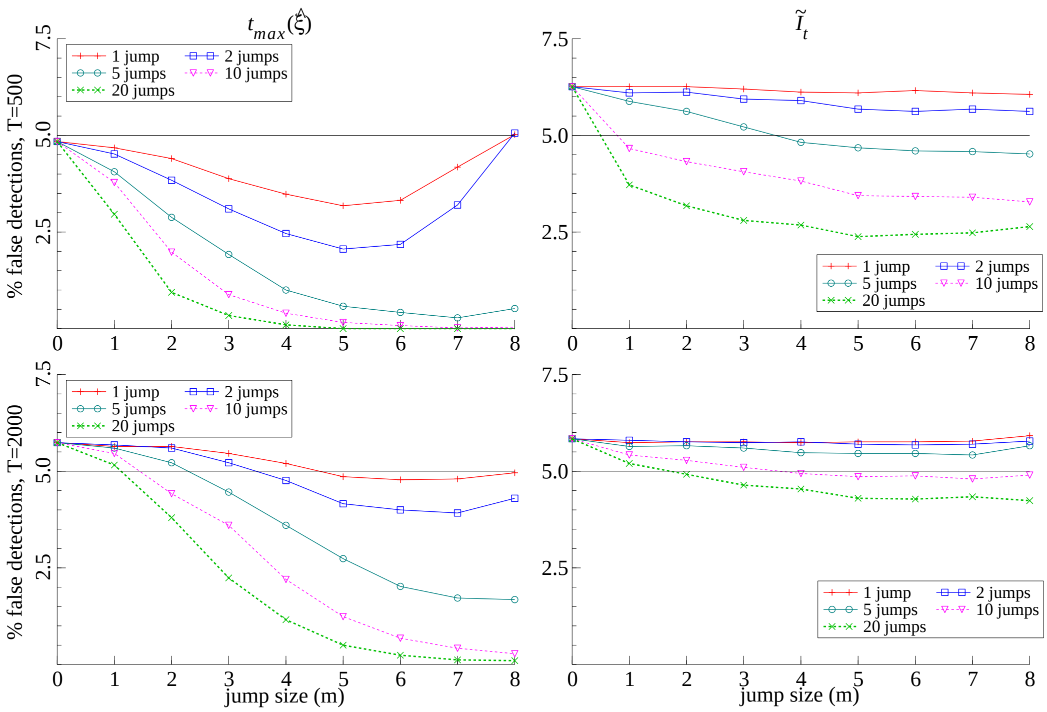

Figure 3: Bias as a function of the jump size m for the AR(1)-GARCH(1,1) with parameter values $\mu = 0.05, \phi_1 = 0.3, \omega = 0.3, \alpha_1 = 0.2$ and $\beta_1 = 0.7$ and 1 jump per sample of $T = 500$ observations

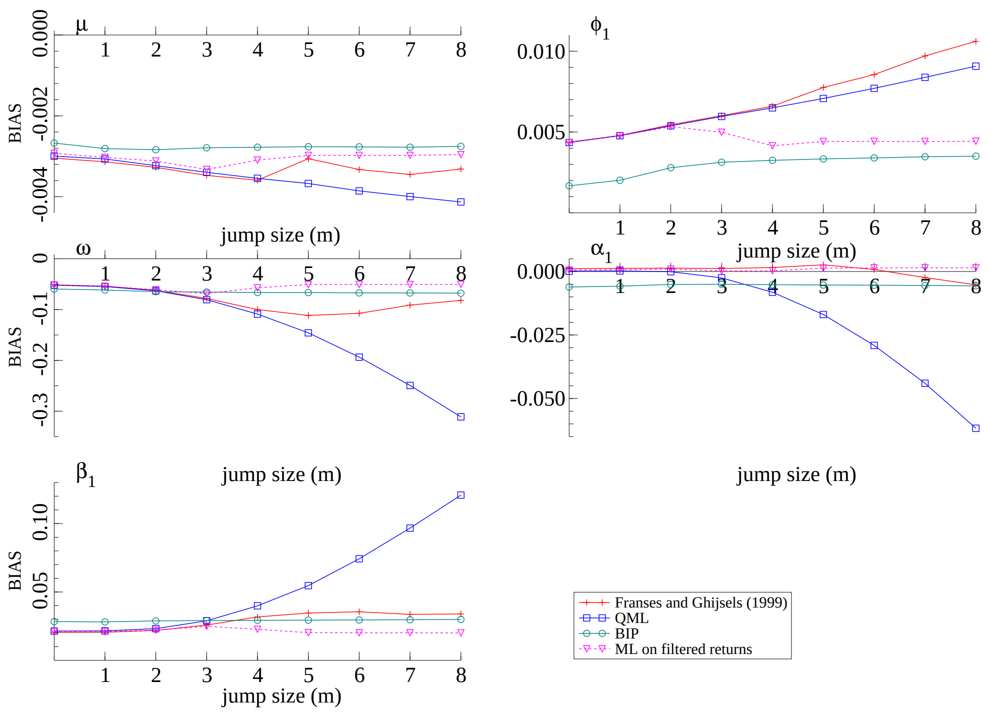

Figure 4: Bias as a function of the jump size m for the AR(1)-GARCH(1,1) with parameter values $\mu = 0.05, \phi_1 = 0.3, \omega = 0.3, \alpha_1 = 0.2$ and $\beta_1 = 0.7$ and 5 jumps per sample of $T = 500$ observations

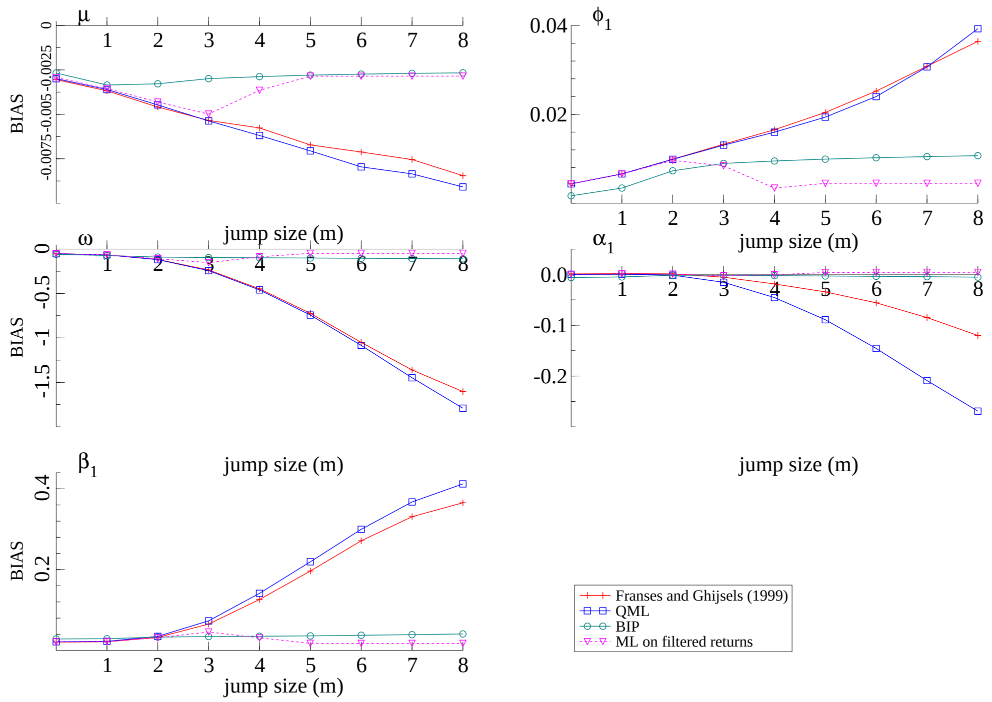

Figure 5: MSE as a function of the jump size m for the AR(1)-GARCH(1,1) with parameter values $\mu = 0.05, \phi_1 = 0.3, \omega = 0.3, \alpha_1 = 0.2$ and $\beta_1 = 0.7$ and 1 jump per sample of $T = 500$ observations

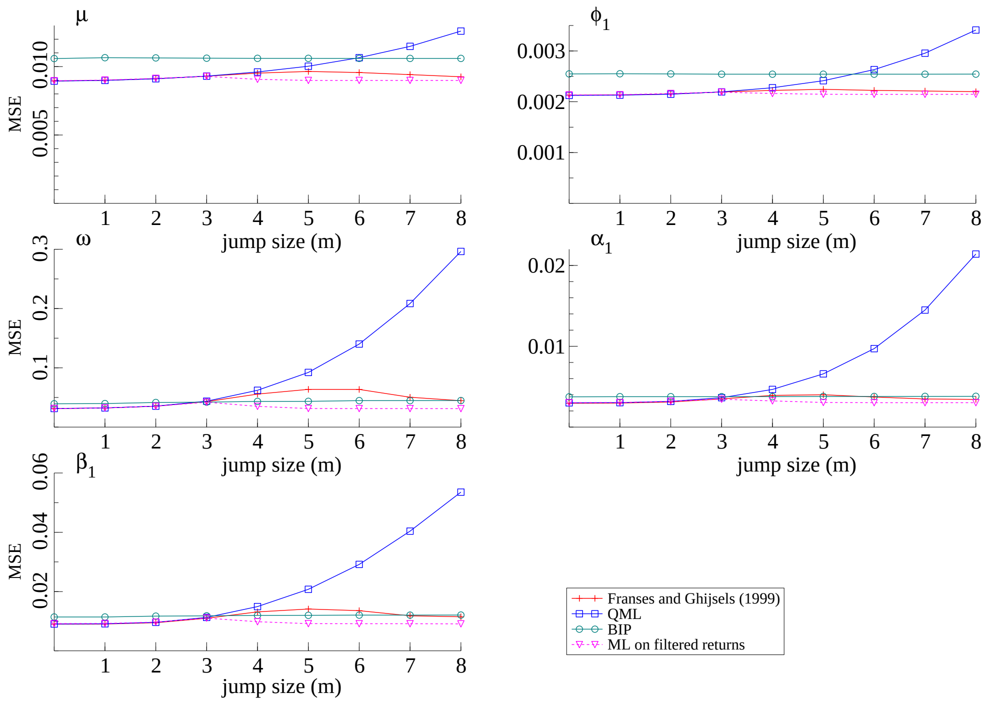

Figure 6: MSE as a function of the jump size m for the AR(1)-GARCH(1,1) with parameter values $\mu = 0.05, \phi_1 = 0.3, \omega = 0.3, \alpha_1 = 0.2$ and $\beta_1 = 0.7$ and 5 jumps per sample of $T = 500$ observations

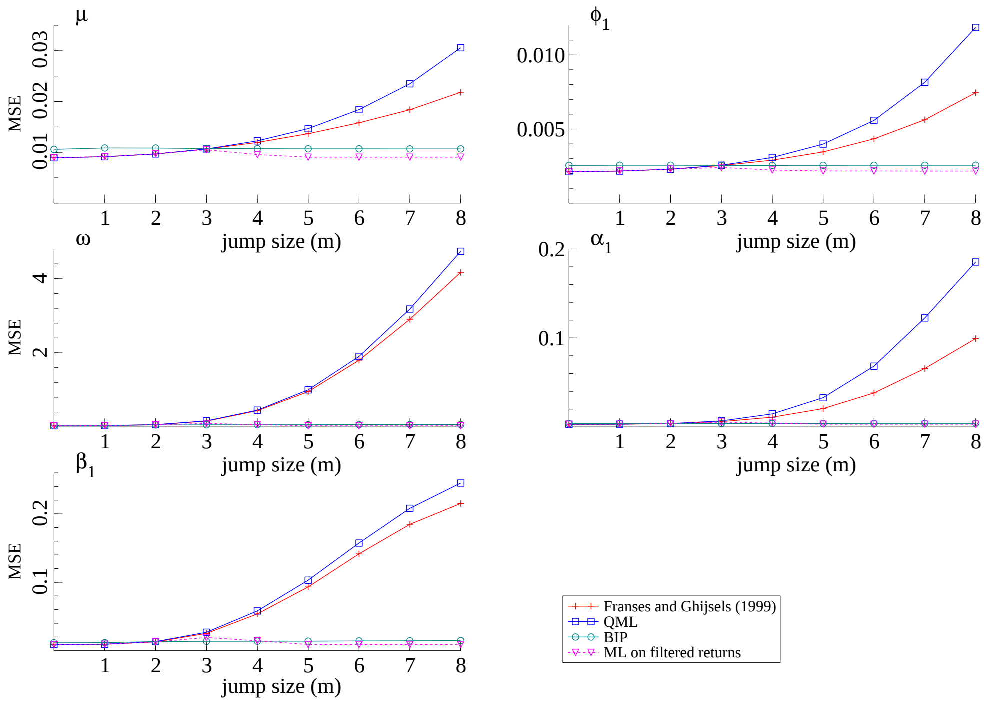

Figure 7: 95% coverage probabilities as a function of the jump size m for the AR(1)-GARCH(1,1) with parameter values $\mu = 0.05, \phi_1 = 0.3, \omega = 0.3, \alpha_1 = 0.2$ and $\beta_1 = 0.7$ and 1 jump per sample of $T = 500$ observations

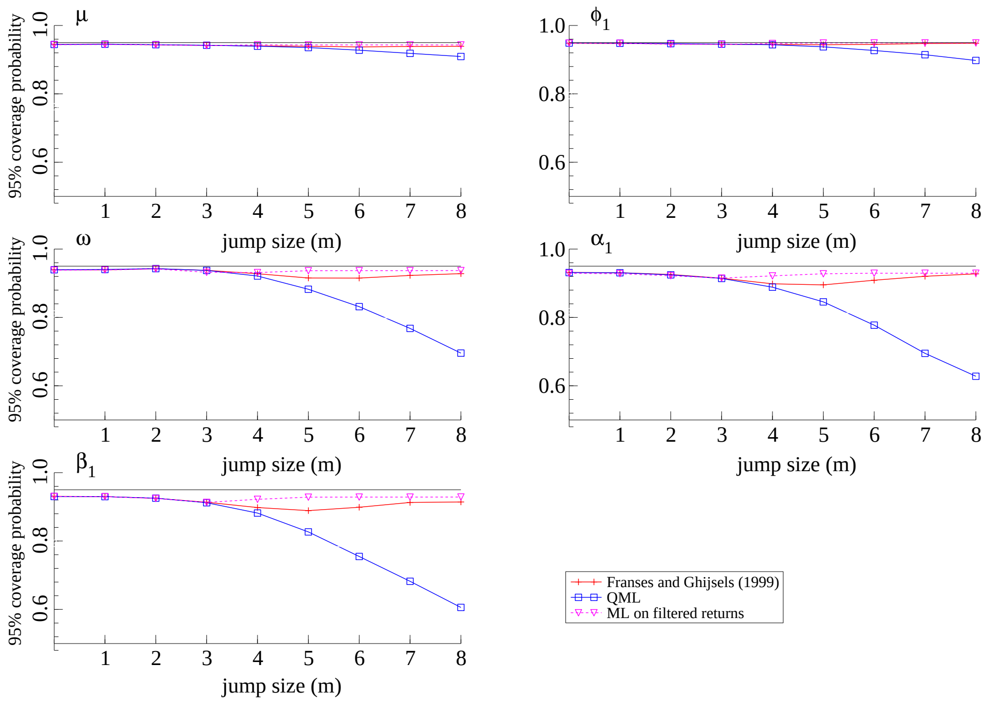

Figure 8: 95% coverage probabilities as a function of the jump size m for the AR(1)-GARCH(1,1) with parameter values $\mu = 0.05, \phi_1 = 0.3, \omega = 0.3, \alpha_1 = 0.2$ and $\beta_1 = 0.7$ and 5 jumps per sample of $T = 500$ observations

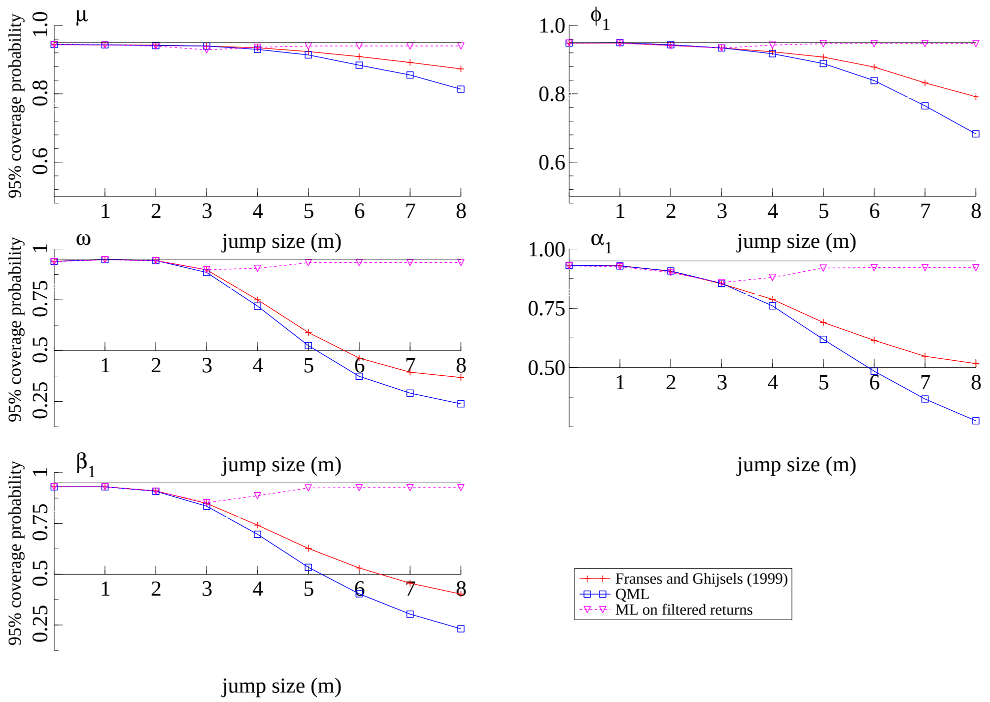

Figure 9: Daily returns in % of the YEN-USD exchange rate over the period January 2005 - May 2011 and detected jumps

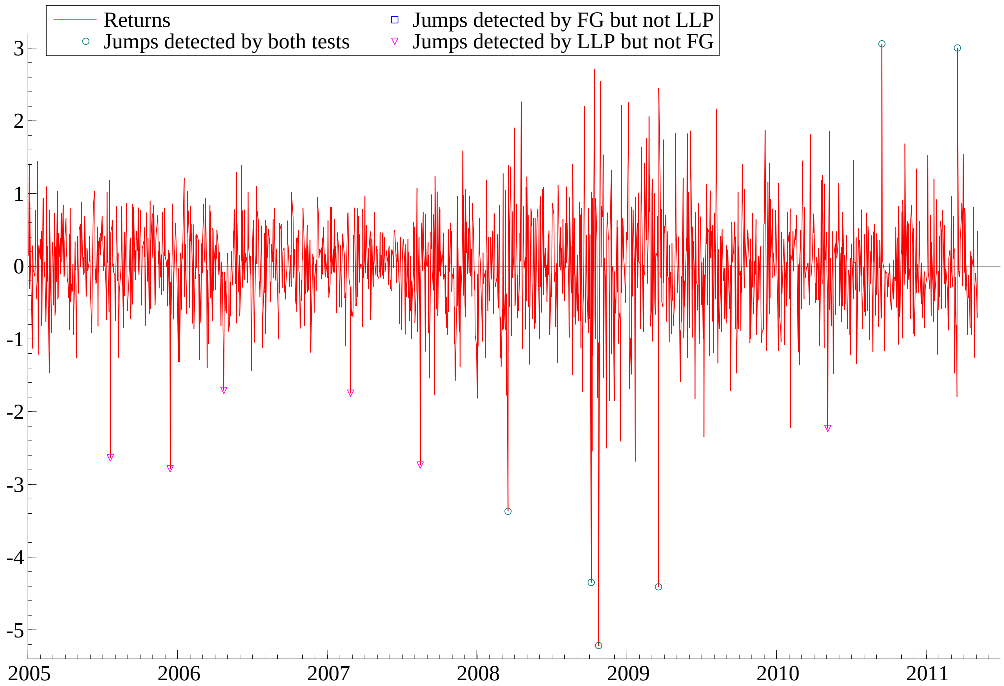

Table 1: Detected jumps

Date	Returns	$t_{max}(\hat{\xi})$	$\tilde{I}_t$	$t_{max}(\hat{\xi}) > 10$	$\tilde{I}_t > 3.527$	Event
2005-07-21	-2.635	-	4.729	no	yes	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	-2.786	-	6.117	no	yes	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	-1.708	-	3.595	no	yes	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	-1.746	-	3.751	no	yes	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	-2.733	-	4.893	no	yes	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	-3.369	12.457	4.324	yes	yes	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	-4.348	19.703	6.086	yes	yes	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).
2008-10-24	-5.216	23.374	5.264	yes	yes	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	-4.409	18.615	5.176	yes	yes	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).
2010-05-06	-2.230	-	3.672	no	yes	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	3.059	10.733	5.244	yes	yes	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	3.002	10.393	4.367	yes	yes	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).

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

Doornik, J. A., and M. Ooms (2005): "Outlier Detection in GARCH Models," 2005-W24 in Nuffield Economics Working Papers.

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

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.

Hansen, P., A. Lunde, and J. Nason (2010): "Model Confidence Sets," *Econometrica*, 79, 453–497.

Harvey, A., and T. Chakravarty (2008): "Beta-t-(E)GARCH," Working Paper series, University of Cambridge.

Hotta, L., and R. Tsay (1998): "Outliers in GARCH Process," Unpublished Manuscript.

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

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

Lee, S. S., and P. A. Mykland (2008): "Jumps in financial markets: a new nonparametric test and jump dynamics," *Review of Financial studies*, p. forthcoming.

Muler, N., D. Pena, and V. Yohai (2009): "Robust Estimateion 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.

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.