Volatility estimation for Bitcoin: Replication and robustness

Olivier Darné, LEMNA

Amélie Charles, Audencia Business School

Abstract

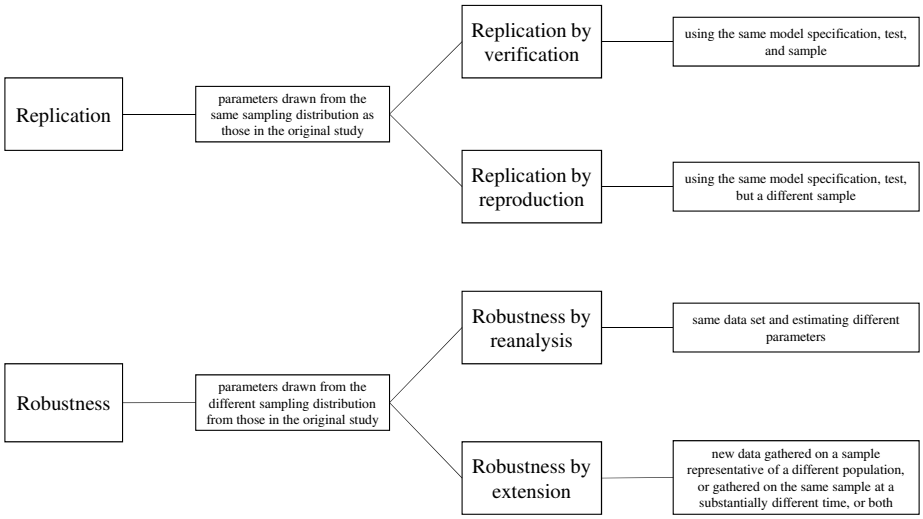
Katsiampa [Volatility estimation for Bitcoin: A comparison of GARCH mod-els. *Economics Letters*, 158, 3-6, 2017] compares several GARCH-type models to estimate volatility for Bitcoin returns. First, we propose a replication study (i) by verification, using the same sample and period (July 2010 to October 2016), and (ii) by reproduction, extending the sample until March 2018. We obtain only partially different results from those of Kasiampa (2017) on both samples. Sec-ond, we propose a robustness analysis (i) by reanalysis, using the robust QML estimator for computing the standard errors of the parameters, and (ii) by exten-sion, taking into account the presence of jumps in the Bitcoin returns. The results show that the six GARCH-type models studied, namely GARCH-type models characterized by short memory, asymmetric effects, or long-run and short-run movements, seem not to be appropriate for modelling the Bitcoin returns.

Keywords: Bitcoin, GARCH, volatility, jumps.

JEL Classification: C22, C50, G10.

1 Introduction

Reproducibility, replicability and robustness are fundamental characteristics of scientific studies. Replication, reproduction and robustness studies stimulate scientific rigor and enhance the validity and reliability of published research. Burman et al. (2010) emphasize that replication is a critical tool for scientific progress and that the absence of such studies "is particularly problematic because empirical economic research is often prone to error".[1] However, there is no consensus on definitions for the concepts of reproducibility, replicability and robustness (see, e.g., Arulampalam et al., 1997; Pesaran, 2003; Hamermesh, 2007; Clemens, 2017), which can lead to confusion. In order to avoid confusion on these concepts Clemens (2017) identifies four categories that we present briefly in Figure 1:[2]



[1] Burman et al. (2010) emphasize that the errors can arise from inadvertent and innocent mistakes by researchers or from bugs in computer programs but also from carelessness or even dishonesty. Further, McCullough and Vinod (2003) emphasize that "Research that cannot be replicated is not science, and cannot be trusted either as part of the profession's accumulated body of knowledge or as a basis for policy."

[2] We thank an anonymous referee for the importance to distinguish between reproducibility, replicability and robustness.

These issues are of increasing importance for the scientific literature as shown by the computational infrastructures (e.g., runmycode.org, Hurlin et al, 2012, codeocean.com), the websites providing information on replication studies (e.g., *Replication in Economics* or *Replication Network*), the special sections proposed by some journals (e.g., "Replication section" in *Journal of Applied Econometrics* or "Data, Tools, and Replication" in *International Economics*, the special issue on replication in *Energy Economics*), the replication policy to make data and code available by some journals on their website (e.g., *The American Economic Review*, *Econometrica*, *Review of Economic Studies*, *Journal of Applied Econometrics*, etc.)[3]

Cryptocurrencies, such as Bitcoin, are a new breed of digital currency systems built on computer cryptology and decentralized (peer-to-peer) network architecture. Bitcoin, the most popular virtual (digital) currency and cryptocurrency (Nakamoto, 2008), is an open source virtual currency, as it is managed by an open source software algorithm that uses the global internet network both to create Bitcoins as well as to record and verify its transactions.[4] While central authorities and central banks guarantee or have control over standard (fiat) currencies, Bitcoin is fully decentralized and depends on a sophisticated protocol that uses only cryptography to control transactions, manage its supply, and prevent harmful actions that may endanger the system. All transactions are stored digitally and recorded in a shared ledger data technology known as blockchain (see, e.g., Dwyer, 2015; Böhme et al., 2015).[5]
Since its inception in January 2009, Bitcoin exhibits an extreme increase of its market

[3]Note that very few journals publish replications. Duvendack et al. (2017) inventory only 28 of 333 economics journals which regularly make data and code available.

[4]The success of Bitcoin has led to the emergence of many alternative virtual currencies, called altcoins (abbreviation of "Bitcoin alternative"), such as BitShares, Ethereum, Dash, DogeCoin, LiteCoin, PeerCoin, Ripple. Most of altcoins rely on the same or similar blockchain technology as Bitcoin, and aim to either complement or improve certain BitCoin characteristics. See Ciaian et al. (2018) for a discussion on altcoins.

[5]Bitcoins are created in a mining process, in which computer network participants, called miners, i.e. users who provide their computing power, verify, validate transactions (by gathering together blocks) and record payments into a public ledger called blockchain. This work is remunerated by giving the miners (new) Bitcoins, what makes the validating costs cheaper than in a centralized system. Note that mining is the only way that new Bitcoins are introduced. The validation is made by solving some kind of algorithm. See Lee (2015) et Narayanan et al. (2016).

value, with a market capitalization exceeding $139 billion and representing more than 60% of the total cryptocurrency capitalization, and an estimated circulation to be almost 17 million in April 2018 (CoinDesk and CoinMarketCap).[6] Bitcoin prices are characterized by high volatility: at the end of 2010 Bitcoin prices rose from $1 to nearly $30 in June 2011, before falling below $2 at the end of 2011. Since 2012 Bitcoin prices rose until nearly $20,000 in December 2017, then falling below $7,000 in February 2018 (see Figure 1).

Therefore, financial-market participants would benefit from a better understanding of Bitcoin volatility over time. In this framework, Katsiampa (2017) in "Volatility estimation for Bitcoin: A comparison of GARCH models" [*Economics Letters*, 158, 3-6, 2017] compares six GARCH-type (GARCH, EGARCH, TGARCH, APARCH, CGARCH and ACGARCH) models to estimate volatility for Bitcoin returns, covering the period from July 18, 2010 to October 01, 2016, and find that " the best model is the AR-CGARCH model". We propose a replication study by verification on the same sample, period and GARCH-type models. Our results are partially different from those of Kasiampa (2017) in terms of values of the GARCH coefficients and t-ratios, and the differences are mainly coming from the method used to compute the return series. We also propose a replication study by reproduction, using a different sample extended until March 2018, and find similar results to those for the shorter period.
We revisit the Katsiampa's (2017) results through a robustness test by reanalysis that allows for changes in empirical specifications and/or estimation methods (Arulampalam et al., 1997; Clemens, 2017). For that we use a more appropriate estimator with the robust QML estimator (Bollerslev and Wooldridge, 1992) which is robust to conditional non-Normality to produce standard errors of the estimated parameters. The results of the reanalysis show that all the GARCH-type models are rejected for the both samples.

It is well-known that financial returns are subject to a number of drastic shocks (called large shocks, outliers or jumps), and the Bitcoin returns do not make discordant note neither (see Figures 1 and 2). For instance, on June 11, 2011, the Bitcoin recorded a negative returns of 49% because of the hacking of Mt.Gox exchange platform. These

[6]The information is available at: www.coindesk.com (CoinDesk) and https://coinmarketcap.com (CoinMarketCap).

shocks may pose difficulties for the identification and estimation of GARCH models governing the conditional volatility of returns (e.g., Carnero et al., 2001, 2007; Charles and Darné, 2006; Laurent et al., 2016). We thus extend the study of Katsiampa (2017) to a robustness level by detecting first the jumps in the Bitcoin returns via the semi-parametric test for jumps in GARCH models proposed by Laurent et al. (2016).[7], and estimating then GARCH-type models on filtered data. We find that the Bitcoin returns are strongly characterized by the presence of jumps mainly due to attacks, hacks, thefts, closings or bankruptcies of Bitcoin exchange platforms as well as technical issues. The results on the filtered returns show that the six GARCH-type models studied seem not to be appropriate for modelling the Bitcoin returns.

The remainder of this article is organized as follows. Section 2 presents the replication of the Katsiampa's (2017) study (by verification and reproduction) whereas the robustness tests (by reanalysis and extension) are presented in Section 3. Finally, Section 4 concludes.

2 Replication study

In this section we follow Katsiampa (2017) and consider the returns series r_t, described by a Normal AR(1)-GARCH(1,1) model:

$$r_t = \mu_t + \varepsilon_t, \qquad \mu_t = c + \sum_{i=1}^{\infty} \xi_i \varepsilon_{t-i}, \tag{1}$$

$$\varepsilon_t = \sigma_t z_t, \qquad z_t \sim i.i.d.N(0,1), \tag{2}$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{3}$$

where ε_t is the error term, z_t is a white noise process, σ_t is the conditional standard deviation, and ξ_i are the coefficients of $\xi(L) = \phi^{-1}(L) = 1 + \sum_{i=1}^{\infty} \xi_i L^i$, with L the lag operator and $\phi(L)$ the AR polynomial of order 1.

For the conditional variance σ_t^2 we consider six different GARCH-type models, namely GARCH (Bollerslev, 1986), EGARCH (Nelson, 1991), GJR-GARCH (Glosten

[7] In his definition of 'new data' for robustness test by extension Clemens (2017) includes "dropping influential observations, since a truncated sample cannot represent the same population" (p. 327).

et al., 1993)[8], Asymmetric Power ARCH (APARCH, Ding et al., 1993), Component GARCH and Asymmetric Component GARCH (CGARCH and ACGARCH, respectively, Engle and Lee, 1993) models. A brief presentation of the models is given in Table 2.[9]

2.1 Replication by verification

We use the same data and sample period as in Katsiampa (2017), covering July 18, 2010 to October 01, 2016 (2 268 observations).[10] The Bitcoin returns are calculated as the first differences of the logarithmic prices, defined by $r_t = \log P_t - \log P_{t-1}$, where P_t is the observed price at time t. The graphical representations of prices, returns and volatility are in Figure 1.

The replication of the summary statistics for Bitcoin returns are presented in Table 1 (Panel A). However, the values are not equal to Katsiampa (2017), whatever the descriptive statistics. We then compute the descriptive statistics on the Bitcoin returns calculated as arithmetic returns defined by $r_t^g = (P_t - P_{t-1})/P_{t-1}$ and find exactly the same values to those displayed in Katsiampa (2017). Even if the different approaches to calculate the returns do not change the main characteristics of Bitcoin

[8]Note that Katsiampa (2017) estimates a TARCH model whereas we estimate a GJR-GARCH model. However, the estimated parameters are the same because EViews proposes the estimation of a TARCH model but its TARCH model is estimated on σ_t^2 and not on σ_t as defined in Rabemananjara and Zakoian (1993). Its TARCH specification is rather similar to the GJR-GARCH model of Glosten et al. (1993).

[9]The parameters of the volatility models are estimated by the quasi-maximum likelihood (QML) method and the quasi-likelihood function is maximized using the quasi-Newton method of Broyden, Fletcher, Goldfarb and Shanno (BFGS) with the Marquardt correction, namely the same code and software as in Katsiampa (2017). In a previous version of the study we have used the Berndt-Hall-Hall-Hausman (BHHH) algorithm with the Marquardt correction, from the default option of EViews 7.0. Katsiampa (2017) did not give information about the software package used in estimation, what the default estimation methods are, such as method to optimize the likelihood function or how the standard errors are computed. We have tried different software packages and only EViews produced similar results to those obtained in Katsiampa (2017). However, in an author's reply to this previous study she mentions that the BFGS optimization method with Marquardt step method was employed, and the standard errors were computed from the outer product of the gradient (OPG) estimator, from the default option of EViews 9.0.

[10]Bitcoin price data is downloaded from coindesk.com.

returns, namely highly non-Normal, evidence of significant excess skewness and kurtosis, and strong conditional heteroscedasticity, except for excess skewness which is with the opposed sign, these differences could have implications in the estimation of GARCH-type models.

Table 3 shows the estimation results of the GARCH-type models, computed using arithmetic (Panel A) and logarithmic returns (Panel B). The comparison between the volatility models is effected via various in-sample criteria: Log-Likehood (LL), Akaike (AIC) and Hannan-Quinn (HQ). Each time, the best model appears in bold showing the highest value of LL and the lowest values of AIC and HQ. The residual tests are also shown to see whether the chosen volatility model is the most appropriate.

When estimating GARCH-type models on returns computed as arithmetic returns we find exactly the same values for GARCH parameters, t-ratios, in-sample criteria and residual tests as in Katsiampa (2017), whatever the GARCH-type models, except for the ACGARCH models where the estimations are slightly different. Nevertheless, we also find that the Bitcoin volatility is best modelled by the AR(1)-CGARCH model. When estimating GARCH-type models on returns computed as logarithmic returns we find similar values of GARCH coefficients and t-ratios for all the GARCH-type models. The AR(1)-CGARCH model stays the best specification for the Bitcoin returns. Overall, our results are only partially different from those of Kasiampa (2017) due to the difference in computing the returns.

2.2 Replication by reproduction

We now propose a replication by reproduction of the Katsiampa's (2017) study by estimating the GARCH-type models on a longer period, namely until March 22, 2018, producing $2,803$ observations (see Figure 2). For this analysis we use Bitcoin logarithmic returns, r_t.

Table 1 (Panel B) presents the summary statistics and shows the same main characteristics as those for the shorter period, namely highly non-Normal, evidence of significant excess skewness and kurtosis, and strong conditional heteroscedasticity.

After the estimation of the models, the AR(1)-CGARCH model seems to be again the best specification (Table 5).

3 Robustness analysis

In this section we first reanalyze the study of Katsiampa (2017) by computing robust standard errors of the parameters in the GARCH-type models, and second we extend her study by taking into account the presence of jumps, which can be classified as robustness tests.

3.1 Robustness by reanalysis

Since the Bitcoin returns are characterized by non-Normality we apply a more appropriate estimator for computing standard errors rather than the outer product of the gradient (OPG) estimator. For that, the so-called QML "sandwich" estimator (also known as robust covariance matrix estimator or heteroscedasticity-consistent covariance matrix estimator) proposed by Bollerslev and Wooldridge (1992) is used to compute robust standard errors of the parameters in the GARCH-type models.
The covariance matrix is estimated using both matrices of second order derivatives (the Hessian matrix, $\mathbf{H}(.)$) and outer products of first derivatives (or OPG, $\mathbf{B}(.)$) of the log-likelihood function:

$$\mathbf{Q}(\widehat{\theta}) = \mathbf{H}(\widehat{\theta}_{QML})^{-1}\mathbf{B}(\widehat{\theta}_{QML})\mathbf{H}(\widehat{\theta}_{QML})^{-1},$$

where $\widehat{\theta}_{QML}$ is the robust QML estimator of θ, and the coefficient standard errors as the (robust) Bollerslev-Wooldridge standard errors.[11]

The results are displayed in brackets in Table 3 and show that the asymmetric parameters are become not statistically significant for all the asymmetric models. This finding is not consistent with the presence of asymmetry detected from asymmetric volatility models by Katsiampa (2017), Bouri et al. (2017) and Baur et al. (2018).[12] The GARCH model is the only volatility model for which all the parameters are

[11] Katsiampa (2017) uses the OPG estimator for computing standard errors, given by $\mathbf{I}(\widehat{\theta}) \approx E\left[\mathbf{B}(\widehat{\theta})\right] = E\left[\sum_{t=1}^{T}\mathbf{G}_t(\widehat{\theta})\mathbf{G}_t'(\widehat{\theta})\right] = E\left[\frac{\partial lnL}{\partial\theta}\frac{\partial lnL}{\partial\theta'}\right]$, with $\mathbf{G}(.)$ the gradient matrix.

[12] Bouri et al. (2017) fit GRJ-GARCH(1,1) and EGARCH(1,1) models from August 18, 2011 to April 29, 2016, and Baur et al. (2018) estimate a GJR-GARCH(1,1) model from July 19, 2010 to July 14, 2017. Note that Baur et al. (2018) use Bitcoin prices from CoinDesk whereas Bouri et al. (2017) use data from Bitstamp.

statistically significant, however its stationarity and fourth-moment conditions are not satisfied.

In Table 5 we estimate the GARCH-type models on the longer period (2010-2018) and the robust QML t-statistics are given in brackets. The results confirm those obtained for the shorter period.

3.2 Robustness by extension

We apply the semi-parametric procedure to detect jumps proposed by Laurent et al. (LLP) (2016). Their test is similar to the non-parametric tests for jumps in Lee and Mykland (2008) and Andersen, Bollerslev, and Dobrev (2007) for daily data.

The returns r_t are described by the AR(1)-GARCH(1,1) model of equations (1)-(3). Consider the return series with an independent jump component $a_t I_t$, defined as:

$$r_t^* = r_t + a_t I_t, \tag{4}$$

where r_t^* denotes the observed returns, I_t is a dummy variable for a jump on day t, and a_t is the jump size. In equation (4) a jump $a_t I_t$ will not affect σ_{t+1}^2 (the conditional variance of r_{t+1}), so that we can have non-Gaussian fat-tailed conditional distributions of r_t^*.

The bounded innovation propagation (BIP)-AR(1) proposed by Muler, Peña and Yohai (2009) and the BIP-GARCH(1,1) of Muler and Yohai (2008) is used to obtain robust estimates of μ_t and σ_t^2 respectively in equations (1) and (3). These are denoted by $\widetilde{\mu}_t$ and $\widetilde{\sigma}_t$ and are robust to potential jumps $a_t I_t$ (i.e. they are estimated on r_t^* and not on r_t). The BIP-AR(1) and BIP-GARCH(1,1) are defined as:

$$\widetilde{\mu}_t = \mu + \sum_{i=1}^{\infty} \xi_i \widetilde{\sigma}_{t-i} \omega_{k_\delta}^{MPY}(\widetilde{J}_{t-i}), \tag{5}$$

$$\widetilde{\sigma}_t^2 = \omega + \alpha_1 \widetilde{\sigma}_{t-1}^2 c_\delta \omega_{k_\delta}^{MPY} \left(\widetilde{J}_{t-1}\right)^2 + \beta_1 \widetilde{\sigma}_{t-1}^2, \tag{6}$$

where ξ_i are the coefficients from the AR(1) polynomial defined in equation (1), $\omega_{k_\delta}^{MPY}(.)$ is the weight function, and c_δ a factor ensuring that the conditional expectation of the weighted square of unexpected shocks equals the conditional variance of r_t in the absence of jumps (Boudt et al., 2013).

Consider the standardized return on day t given by:

$$\widetilde{J}_t = \frac{r_t^* - \widetilde{\mu}_t}{\widetilde{\sigma}_t}. \tag{7}$$

LLP detect the presence of jumps by testing the null hypothesis $H_0 : a_t I_t = 0$ against the alternative $H_1 : a_t I_t \neq 0$. The null is rejected if

$$\max_T |\widetilde{J}_t| > g_{T,\lambda}, \qquad t = 1, \ldots, T, \tag{8}$$

where $g_{T,\lambda}$ is the suitable critical value.[13] If H_0 is rejected a dummy variable is defined as follows:

$$\widetilde{I}_t = I\left(|\widetilde{J}_t| > k\right), \tag{9}$$

where $I(.)$ is the indicator function, with $\widetilde{I}_t = 1$ when a jump is detected at time t. The filtered returns $\widetilde{r}_t$ are obtained as follows:

$$\widetilde{r}_t = r_t^* - (r_t^* - \widetilde{\mu}_t)\widetilde{I}_t. \tag{10}$$

Laurent et al. (2016) extend this test for additive jumps in AR-GARCH models, called BIP-AR-BIP-GARCH models, to AR-GJR-GARCH models, called BIP-AR-BIP-GJR models, to account for asymmetric effects.

In Table 4, detected jumps are given with their timing and t-statistics as well as the percent changes. A number of jumps is found in the daily Bitcoin returns during the whole period, giving strong proof of infrequent large shocks. This finding shows the importance to take into account these large shocks in modelling the volatility of Bitcoin returns. We also try to associate the date of each jumps with a specific event that occurred near that date, presenting the jump dates in chronological order. Overall, most of jumps seem to be associated with attacks, hacks, thefts, closings or bankruptcies of Bitcoin exchange platforms and Bitcoin services, such as the hacking of the Japanese-based Mt.Gox exchange (-49%) on June 11, 2011.[14]
We just give few detailed explanations on jumps associated with "technical" issues of Bitcoin due to its specific functioning. Bitcoin Core is the reference implementation of Bitcoin. Initially, the software was published by Satoshi Nakamoto under the name "Bitcoin", then "Bitcoin-Qt" and later renamed to "Bitcoin Core" to distinguish

[13] The critical values are defined by $g_{T,\lambda} = -\log(-\log(1-\lambda))\, b_T + c_T$, with $b_T = 1/\sqrt{2\log T}$, and $c_T = (2\log T)^{1/2} - [\log\pi + \log(\log T)]/[2(2\log T)^{1/2}]$. Laurent et al. (2016) suggest setting $\lambda = 0.5$.

[14] Note that the year 2010 is characterized by a lot of zero returns.

it from the network. On 19 March, 2012, a bug on Bitcoin-Qt implies a fall of Bitcoin (-11.9%). On 18 August, 2015, Bitcoin Core developers Mike Hearn and Gavin Andresen release a separate version of the Bitcoin client software, called Bitcoin XT. Bitcoin XT implements BIP 101[15], which proposes replacing the fixed one megabyte maximum block size with a maximum size that grows over time at a predictable rate. The release of Bitcoin XT culminated fears that the Bitcoin community may not be able to reach a consensus on the issue, and the blockchain may hard fork, resulting in two separate versions of Bitcoin's global ledger (-16%). On July 20, 2017, the Bitcoin averted a split into two currencies after its network (miners) supported an upgrade to its software, the so-called BIP 91 through a software called SegWit2x (Segregated Witness), that would enhance its ability to process an increasing number of transactions (22.6%). The announcement of the first version of the Lightning Network protocol on January 16, 2018 can explain the fall of Bitcoin (-18%). This payment protocol is touted as a solution to the Bitcoin scalability problem. It could enable (almost) instant transactions between participating, including micro-transactions, for a very low price.

Table 1 (Panel B) presents the summary statistics for the filtered returns and show that the jump-filtered returns ($\widetilde{r}_t$) also exhibit excess skewness, excess kurtosis and conditional heteroscedasticity, although the excess kurtosis and skewness decrease dramatically. Note that the original returns exhibit negative excess skewness while the filtered returns display positive excess skewness, signifying that the series have a longer right tail (extreme gains) than left tail (extreme losses).

The estimation results of the GARCH-type models on the filtered returns (Panel B, Table 5) show that the better specification for the Bitcoin returns is the GARCH model, as this is the only volatility model for which all the parameters are statistically significant. The value of α decreases and the value of β increases when the data are cleaned of jumps. This finding is in line with those of Carnero et al. (2001, 2007), showing that this behavior on volatility parameters can be due to jumps. The stationarity condition is satisfied but the existence of the fourth moment is not satisfied. Finally, our results suggest that asymmetric and component GARCH models are not appropriate for modelling the Bitcoin returns as well as the standard GARCH model.

[15] A Bitcoin Improvement Proposal (BIP) is a design document for introducing features or information to Bitcoin. This is the standard way of communicating ideas since Bitcoin has no formal structure.

4 Conclusion

This paper proposed a replication in two ways of the Katsiampa's (2017) study on volatility estimation for Bitcoin returns based on six GARCH-type models, where the best specification is the AR-CGARCH model. We first replicated by verification on the same sample and period (2010-2016), and second by reproduction on a longer period (2010-2018). Our results are only partially different from those of Kasiampa (2017) in terms of values of GARCH coefficients and t-ratios due to the difference in computing the returns. We then proposed robustness tests in two ways. We first reanalyzed her study by using robust QML estimator which is robust to conditional non-Normality to produce standard errors of the estimated parameters. We found that all the GARCH-type models are rejected because either the parameters are not statistically significant or the stationarity condition is not satisfied. Second, we extended the study by detecting jumps in volatility and analyzing their effect on volatility modelling on the Bitcoin returns. We found that the Bitcoin returns are strongly characterized by the presence of jumps. The results on the filtered returns show that the six GARCH-type models studied, namely GARCH-type models characterized by short memory, asymmetric effects, or long-run and short-run movements, are not appropriate for modelling the Bitcoin returns. Therefore, it would be interesting to extend this work to long-memory and Markow-switching multifractal models.

Finally, as emphasized by Koenker and Zeileis (2009) it is important that software users should mention the name (and version) of the package used in estimation, what the default estimation methods are, such as method to optimize the likelihood function, how the standard errors are computed, and so on, and (if possible) the availability of data in order to facilitate the reproduction, replication and robustness studies and thus giving confidence in the research findings.

References

[1] Andersen, T.G., Bollerslev, T., Dobrev, D. (2007). 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.

[2] Arulampalam, W., Hartog, J., MaCurdy, T., Theeuwes, J. (1997). Replication and re-analysis. Labour Economics 4, 99-105.

[3] Baur, D.G., Dimpfl, T., Kuck, K. (2018). Bitcoin, gold and the US dollar - A replication and extension. Finance Research Letters 25, 103-110.

[4] Böhme, R., Christin, N., Edelman, B., Moore, T. (2015). Bitcoin: Economics, Technology, and Governance. Journal of Economic Perspectives 29, 213-38.

[5] Bollerslev T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 31, 307-327.

[6] Bollerslev, T., Wooldridge, J. (1992). Quasi-maximum likelihood estimation inference in dynamic models with time-varying covariance. Econometric Theory 11, 143-72.

[7] Boudt, K., Danielsson, J., Laurent, S. (2013). Robust forecasting of dynamic conditional correlation GARCH Models. International Journal of Forecasting 29, 244-257.

[8] Bouri, E., Azzi, G., Dyhrberg, A.H. (2017). On the return-volatility relationship in the Bitcoin market around the price crash of 2013. Economics 11, 1-16.

[9] Burman, L.E., Reed, W.R., Alm, J. (2010). A call for replication studies. Public Finance Review 38, 787-793.

[10] Carnero, M.A., Peña, D., Ruiz, E. (2001). Outliers and conditional autoregressive heteroskedasticity in time series. Revista Estadistica 53, 143-213.

[11] Carnero, M.A., Peña, D., Ruiz, E. (2007). Effects of outliers on the identification and estimation of the GARCH models. Journal of Time Series Analysis 28, 471-497.

[12] Charles, A., Darné, O. (2006). Relevance of detecting outliers in GARCH models for modelling and forecasting financial data. Finance 26, 33-71.

[13] Ciaian, P., Rajcaniova, M., Kancs, A. (2018). Virtual relationships: Short- and long-run evidence from BitCoin and altcoin markets. Journal of International Finance Markets Institutions and Money 52, 173-195.

[14] Clemens, M.A. (2017). The meaning of failed replications: A review and proposal. Journal of Economic Surveys 31, 326-342.

[15] Ding, Z., Granger C.W.J., Engle R.F. (1993). A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83-106.

[16] Duvendack, M., Palmer-Jones, R.W. and Reed, W.R. (2015). What is meant by "replication" and why does it encounter resistance in economics? American Economic Review: Papers & Proceedings 107, 46-51.

[17] Dwyer, G.P. (2015). The economics of Bitcoin and similar private digital currencies. Journal of Financial Stability 17, 81-91.

[18] Dyhrberg, A.H. (2016). Bitcoin, gold and the dollar - A GARCH volatility analysis. Finance Research Letters 16, 85-92.

[19] Glosten, L.R., Jagannathan, R., Runkle, D.E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779-1801.

[20] Hamermesh, D.S.(2007). Viewpoint: Replication in Economics. Canadian Journal of Economics 40, 715-33.

[21] Hurlin, C., Pérignon, C., Stodden, V. (2012). RunMyCode.org: A novel dissemination and collaboration platform for executing published computational results. Available at https://halshs.archives-ouvertes.fr/halshs-00739233

[22] Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. Economics Letters 158, 3-6.

[23] Koenker, R., Zeileis, A. (2009). On reproducible econometric research. Journal of Applied Econometrics 24, 833-847.

[24] Laurent, S., Lecourt, C., Palm, F.C. (2016). Testing for jumps in GARCH models, a robust approach. Computational Statistics and Data Analysis 100, 383-400.

[25] Lee, D. (2015). The Handbook of Digital Currency, Elsevier.

[26] Lee, S.S., Mykland, P.A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. The Review of Financial Studies 21, 2535-2563.

[27] McCullough, B.D., Vinod, H.D. (2003). The numerical reliability of econometric software. The Journal of Economic Literature 37, 633-665.

[28] Muler, N., Peña, D., Yohai, V. (2009). Robust estimation for ARMA models. The Annals of Statistics 37, 816-840.

[29] Muler, N., Yohai, V. (2008). Robust estimates for GARCH models. Journal of Statistical Planning and Inference 138, 2918-2940.

[30] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

[31] Narayanan, A., Bonneau, J., Felten, E., Miller, A., Goldfeder, S. (2016). Bitcoin and Cryptocurrency Technologies, Princeton University Press.

[32] Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347-370.

[33] Pesaran, H. (2003). Introducing a Replication Section. Journal of Applied Econometrics 18, 111.

[34] Rabemananjara, R., Zakoian, J.M. (1993). Threshold arch models and asymmetries in volatility. Journal of Applied Econometrics 8, 31-49.

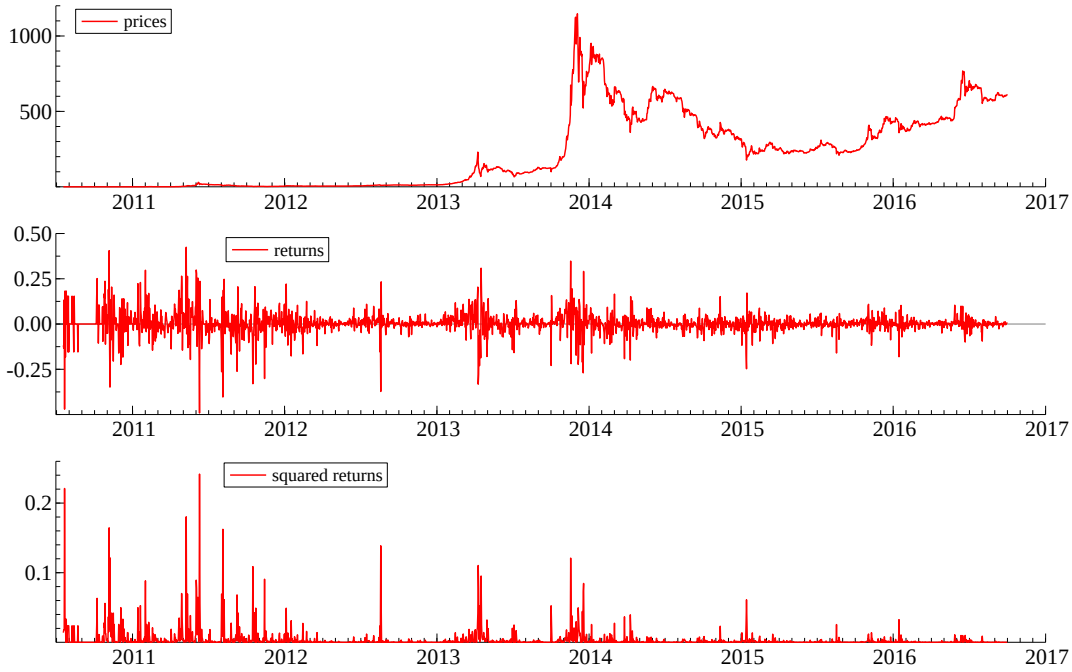


Figure 1: Bitcoin prices, returns and squared returns - July 2010 to October 2016.

Table 1: Descriptive statistics.

	Obs.	Mean	Median	Max	Min	Std.	Skew	Kurt	JB	ARCH(5)
Panel A: 2010-2016										
Growth rate	2267	0.005778	0.000290	0.528947	-0.388309	0.061737	0.857187*	15.11666*	14145.36*	54.36935*
Log 1st diff.	2267	0.003894	0.000290	0.424580	-0.491528	0.061248	-0.371206*	14.96872*	13583.25*	59.47672*
Panel B: 2010-2018										
Original	2803	0.004103	0.001427	0.424580	-0.491528	0.058896	-0.355962*	14.63687*	15874.73*	72.52037*
Filtered	2803	0.005080	0.001970	0.308564	-0.332134	0.049564	0.202438*	9.984921*	5717.31*	124.5632*

Notes: ** denotes significance at the 5% level. Growth rate and Log 1st diff. indicate that the returns are calculated as a growth rate, defined by $r_t^g = (P_t - P_{t-1})/P_{t-1}$, and as the first differences in the logs of the prices, defined by $r_t = \log P_t - \log P_{t-1}$, respectively, where P_t is the observed price at time t. Original and Filtered indicate that the descriptive statistics are calculated on the original and jump-filtered returns, respectively, and the returns are calculated as the first differences in the logs of the prices.

Table 2: GARCH(1,1)-type models.

		Constraints	
Models	Equations	positivity	stationarity
GARCH	$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$	$\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$	$\alpha + \beta < 1$
GJR-GARCH	$\sigma_t^2 = \omega + (\alpha + \gamma\, I_{t-1})\ \varepsilon_{t-1}^2 + \beta\ \sigma_{t-1}^2$	$\omega > 0$, $\alpha, \gamma, \beta \geq 0$ $\alpha + \gamma \geq 0$	$\alpha + \beta + (\gamma/2) < 1$
EGARCH	$\ln(\sigma_t^2) = \omega + g(z_{t-1}) + \beta \ln(\sigma_{t-1}^2)$ $g(z_t) = \theta_1 (\lvert z_t \rvert - E(\lvert z_t \rvert)) + \theta_2 z_t$		$\lvert\beta\rvert < 1$
APARCH	$\sigma_t^\delta = \omega + \alpha (\lvert\varepsilon_{t-1}\rvert - \gamma\varepsilon_{t-1})^\delta + \beta\sigma_{t-1}^\delta$	$\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, $-1 < \gamma < 1$	
CGARCH	$\sigma_t^2 = q_t + \alpha(\varepsilon_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1})$ $q_t = \omega + \rho(q_{t-1} - \omega) + \theta(\varepsilon_{t-1}^2 - \sigma_{t-1}^2)$	$\alpha, \beta, \theta \geq 0$ $\omega > 0$	$0 < \alpha + \beta < \rho \leq 1$ $\theta < \beta < 1$
ACGARCH	$\sigma_t^2 = q_t + \alpha(\varepsilon_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1})$ $+\gamma(\varepsilon_{t-1}^2 - q_{t-1})\ I_{t-1}$ $q_t = \omega + \rho q_{t-1} + \theta(\varepsilon_{t-1}^2 - \sigma_{t-1}^2)$	$0 \leq \alpha < \rho$, $\omega > 0$, $0 \leq \theta < \beta$ $0 \leq \beta < (\rho - \alpha)$ $0 \leq \gamma < (\rho - \alpha - \beta)$	$0 \leq \rho < 1$

Notes: where $I_{t-1} = 1$ if $\varepsilon_{t-1} < 0$, and 0 otherwise. The existence of the fourth moment implies that $E[\varepsilon_t^4] < \infty$, which is satisfied if $k\alpha^2 + 2\alpha\beta + \beta^2 < 1$ and $k\alpha^2 + 2\alpha\beta + \beta^2 + \beta\gamma + k\alpha\gamma + k\delta\gamma^2 < 1$ for the GARCH(1,1) and GJR-GARCH(1,1) models, respectively. Under a Normal distribution $k = 3$ and $\delta = \frac{1}{2}$.

Table 3: Estimates of GARCH-type models for Bitcoin returns on original data - July 2010 to October 2016.

	Parameters						*In-sample criteria*			*Residual tests*		
	ω	α	β	γ	δ/ρ	θ	LL	AIC	HQ	$Q(10)$	$Q^2(10)$	ARCH(5)
Panel A: growth rate												
GARCH	0.0001 (28.1)	0.2363 (20.8)	0.7753 (117.8)				3844.4	-3.3887	-3.3841	40.9*	6.01	0.73
EGARCH	−0.6759 (−31.8)	0.4008 27.9	0.9328 (348.8)	0.0251 (2.68)			3834.3	-3.3789	-3.3734	43.4*	3.56	0.33
GJR-GARCH	0.0001 (26.9)	0.2577 (15.6)	0.7752 (112.3)	−0.0476 (−2.67)			3845.7	-3.3890	-3.3835	39.8*	6.22	0.76
APARCH	0.0003 (3.62)	0.2391 (22.6)	0.7839 (118.5)	−0.0517 (−2.56)	1.6560 (21.5)		3847.8	-3.3900	-3.3835	39.8*	7.70	1.05
CGARCH	0.1327** (1.76)	0.1825 (19.2)	0.7855 (86.1)		0.9999 (17074.6)	0.0549 (12.4)	**3885.7**	**-3.4234**	**-3.4169**	44.3*	4.73	0.33
ACGARCH	0.0035 (3.26)	0.14434 (10.8)	0.6552 (21.8)	0.0512 (3.79)	0.9923 (448.0)	0.0915 (10.5)	3865.5	-3.4047	-3.3973	39.3*	2.58	0.05
Panel B: log first diff.												
GARCH	0.0001 (28.8)[2.63]	0.2510 (22.4)[5.61]	0.7630 (115.5)[22.2]				3835.0	-3.3804	-3.3758	36.0*	4.52	0.45
EGARCH	−0.7475 (−31.9)[−3.99]	0.4231 (28.3)[7.58]	0.9232 (321.7)[34.3]	0.0208 (2.07)[0.52*]			3830.0	-3.3751	-3.3696	35.4*	2.92	0.20
GJR-GARCH	0.0001 (27.6)[2.64]	0.2756 (14.9)[5.18]	0.7640 (109.4)[21.8]	−0.0539 (−2.76)[−0.63*]			3836.4	-3.3808	-3.3753	35.1*	4.64	0.47
APARCH	0.0004 (4.28)[0.77*]	0.2527 (24.0)[5.53]	0.7754 (118.4)[17.5]	−0.0392 (−1.89)[−0.39*]	1.5627 (23.8)[3.62]		3839.9	-3.3829	-3.3765	34.8*	5.55	0.65
CGARCH	0.1704 (2.22)[0.01*]	0.2080 (19.5)[0.06*]	0.7583 (72.0)[0.19*]		0.9999 (218173.2)[1.93]	0.0601 (13.3)[0.01*]	**3870.3**	**-3.4098**	**-3.4034**	38.6*	2.93	0.13
ACGARCH	0.0042 (3.42)[0.27*]	0.1292 (7.27)[0.40*]	0.5662 (14.2)[1.23*]	0.0852 (4.93)[0.70*]	0.9891 (333.3)[51.7]	0.1154 (11.2)[0.36*]	3854.4	-3.3949	-3.3875	37.9*	2.74	0.11

Notes: LL is the log-likelihood value, AIC and HQ correspond to the Akaike and Hannan-Quinn criteria, respectively. $Q(10)$ and $Q^2(10)$ are respectively the Box-Pierce statistics with lag 10 of the standardized and squared standardized residuals. These are asymptotically distributed as $\chi^2(k)$, where k is the lag length. ARCH(5) is the ARCH LM test with lag 5. This is distributed as $\chi^2(q)$, where q is the lag length. * and ** indicate that the null hypothesis is rejected at the 5% and 10% level, respectively. The standard and robust t-ratios appear in parentheses and brackets, respectively. The best in-sample criteria appear in bold.

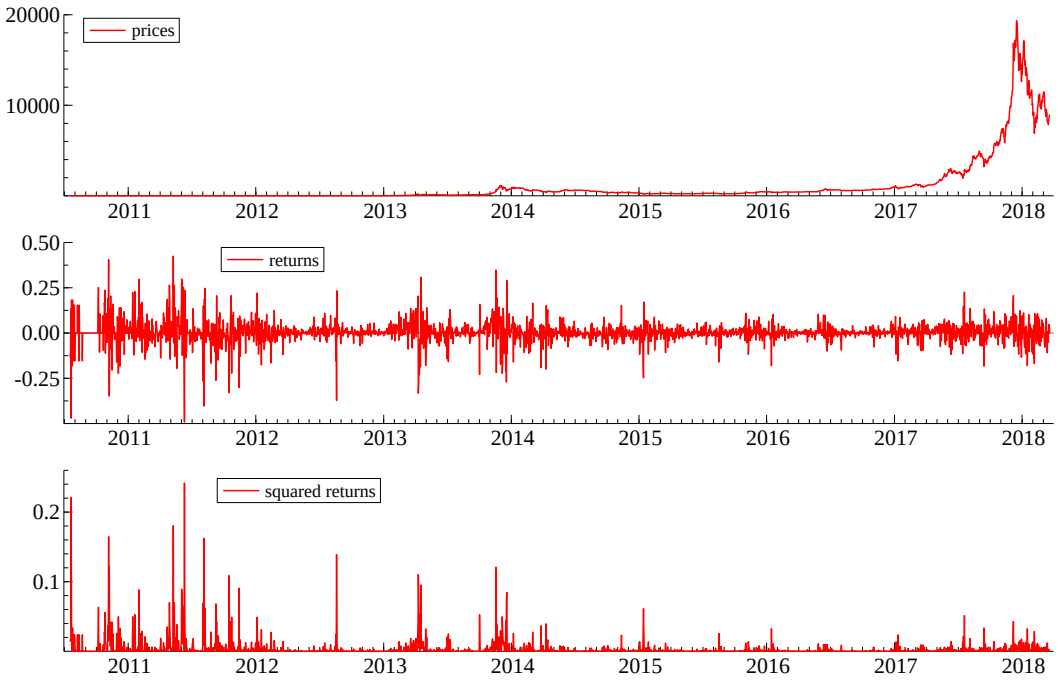


Figure 2: Bitcoin prices, returns and squared returns - July 2010 to March 2018.

Table 4: Jumps detected in the Bitcoin returns.

Date	t-stat	Percent change	Events
07-22-2010	4.58	-47.0	Mt.Gox exchange opens
11-06-2010	3.53	40.1	Bitcoin economy passed $1 million
01-14-2011	7.93	22.3	The Silk Road website (online black market) opens
01-20-2011	4.37	23.0	The Electronic Frontier Foundation accepts Bitcoins
02-01-2011	4.63	29.7	Parity between the US dollar and the Bitcoin
04-04-2011	6.70	-13.7	Three Bitcoin exchanges open
06-11-2011	3.95	-49.2	Hacking of Mt.Gox exchange
08-03-2011	10.23	-26.3	Bitomat exchange lost wallet data
08-06-2011	6.61	-40.3	MyBitCoin (e-wallet service) shuts down
09-09-2011	3.76	-26.1	Massive sellof
10-17-2011	6.91	-33.0	ExchB exchange shuts down
11-14-2011	7.34	-30.1	High trading
02-14-2012	4.25	-16.5	Paxum (online payment service) stops accepting Bitcoins
03-19-2012	5.87	-11.9	Bug in Bitcoin-Qt (ex Bitcoin Core)
08-17-2012	5.71	-15.3	Bitcoins Savings & Trust halts payments (Ponzi scheme)
07-01-2013	5.05	-14.2	Germany plans tax on Bitcoins
10-02-2013	15.59	-22.9	The Silk Road website shuts down
11-18-2013	4.56	34.8	US Senate qualifies Bitcoin as a “legitimate” currency
02-07-2014	3.93	-10.8	Attacks against major Bitcoin exchanges
03-27-2014	8.36	-19.1	IRS declares Bitcoins to be taxed as property
04-10-2014	6.53	-19.9	Chinese Bitcoin exchanges’ bank accounts closed
05-20-2014	6.31	8.9	Thai Bitcoin Exchange is coming back online
09-18-2014	4.57	-7.6	Upcoming IPO of Alibaba
11-12-2014	5.77	15.2	Microsoft accepts Bitcoins
01-03-2015	5.40	-11.7	Hacking of Bitstamp exchange
01-13-2015	4.86	-16.3	Announcement of Bitcoin auction by the US Marshals Service
01-14-2015	5.50	-24.7	Announcement of closing Vault of Satoshi exchange
08-08-2015	5.08	-7.0	Closing exchange services to residents of New Yor state (BitLicense)
08-18-2015	9.72	-16.0	Launch of Bitcoin XT by Mike Hearn et Gavin Andresen (Bitcoin Core developers)
01-15-2016	9.90	-18.1	Mike Hearn quits Bitcoin
04-27-2016	4.32	-5.0	Payment institution status for Bitstamp
05-28-2016	5.51	10.4	Bankruptcy of KnCMiner
06-12-2016	3.98	9.9	Attack on Etherrum network
08-02-2016	4.63	-9.4	Theft on Bitfinex exchange
01-05-2017	4.01	-11.6	Intervention of the People’s Bank of China
01-11-2017	4.64	-15.4	End of margin trading on Chinese platforms
07-20-2017	4.07	22.6	Avoid a split into two cryptocurrencies of the Bitcoin
09-14-2017	4.49	-18.3	Closure of the Chinese exchanges BTCC and ViaBTC
10-12-2017	4.69	11.9	Japanese licenses to cryptocurrecncy stock exchanges
12-07-2017	4.09	20.7	High trading volume
01-16-2018	3.93	-18.0	Announcement of the Lightning Network protocol

Notes: The Gumbel critical value with $T = 2,803$ and $\alpha = 0.5$ is 3.84.

Table 5: Estimates of GARCH-type models for Bitcoin returns on original data - July 2010 to March 2018.

	Parameters						*In-sample criteria*			*Residual tests*		
	ω	α	β	γ	δ/ρ	θ	LL	AIC	HQ	$Q(10)$	$Q^2(10)$	ARCH(5)
Panel A: original												
GARCH	0.0001 (29.7)[2.75]	0.2384 (23.6)[6.54]	0.7732 (125.2)[26.1]				4777.6	-3.4066	-3.4028	36.0*	4.52	0.45
EGARCH	−0.6959 (−32.9)[−4.43]	0.4042 (28.3)[8.33]	0.9308 (362.0)[41.7]	0.0132 (1.54*)[0.38*]			–	–	–	–	–	–
GJR-GARCH	0.0001 (28.8)[2.74]	0.2555 (15.8)[5.93]	0.7742 (118.9)[25.9]	−0.0369 (−2.18)[−0.53*]			4778.6	-3.4066	-3.4020	35.1*	4.64	0.47
APARCH	0.0004 (4.45)[0.89*]	0.2404 (25.6)[6.38]	0.7859 (131.4)[20.8]	−0.0266 (−1.42*)[−0.31*]	1.5577 (24.8)[4.20]		4782.6	-3.4087	-3.4034	43.4*	7.17	0.70
CGARCH	0.1723 (2.39)[0.01*]	0.1765 (19.1)[0.06*]	0.7892 (87.9)[0.23*]		0.9999 (22106.5)[1.92**]	0.0600 (13.9)[0.01*]	**4816.4**	**-3.4329**	**-3.4275**	38.6*	2.93	0.13
ACGARCH	0.0066 (1.29*)[0.32*]	0.1643 (15.1)[5.47]	0.7817 (70.0)[15.3]	0.0218 (3.82)[0.48*]	0.9987 (985.7)[244.2]	0.0454 (4.80)[1.17*]	4814.6	-3.4308	-3.4247	37.9*	2.74	0.11
Panel B: filtered												
GARCH	0.0001 [2.10]	0.1805 [8.01]	0.8116 [37.1]				**5307.5**	**-3.7848**	**-3.7810**	29.1*	13.7	1.32
EGARCH	−0.4942 [−4.22]	0.3161 [7.19]	0.9584 [61.8]	0.0309 [1.14*]			–	–	–			
GJR-GARCH	0.0001 [2.12]	0.2024 [7.04]	0.8150 [37.9]	−0.0527 [−1.21*]			–	–	–			
APARCH	0.0001 [0.45*]	0.1723 [5.82]	0.8108 [21.3]	−0.0721 [−1.16*]	2.1410 [3.02]		–	–	–			
CGARCH	0.0045 [0.63*]	0.0325 [0.98*]	0.6108 [1.17*]	0.9909 [55.8]	0.1710 [7.00]		–	–	–			
ACGARCH	0.0044 [0.62*]	0.0364 [0.74*]	0.6070 [1.03*]	−0.0111 [−0.17*]	0.9909 [54.4]	0.1716 [6.06]	–	–	–			

Notes: LL is the log-likelihood value, AIC and HQ correspond to the Akaike and Hannan-Quinn criteria, respectively. $Q(10)$ and $Q^2(10)$ are respectively the Box-Pierce statistics with lag 10 of the standardized and squared standardized residuals. These are asymptotically distributed as $\chi^2(k)$, where k is the lag length. ARCH(5) is the ARCH LM test with lag 5. This is distributed as $\chi^2(q)$, where q is the lag length. * and ** indicate that the null hypothesis is rejected at the 5% and 10% level, respectively. The standard and robust t-ratios appear in parentheses and brackets, respectively. The best in-sample criteria appear in bold.