Hafner, Christian M.

Working Paper

Testing for bubbles in cryptocurrencies with time-varying volatility

IRTG 1792 Discussion Paper, No. 2018-005

Provided in Cooperation with:
Humboldt University Berlin, International Research Training Group 1792 "High Dimensional Nonstationary Time Series"

Suggested Citation: Hafner, Christian M. (2018) : Testing for bubbles in cryptocurrencies with time-varying volatility, IRTG 1792 Discussion Paper, No. 2018-005, Humboldt-Universität zu Berlin, International Research Training Group 1792 "High Dimensional Nonstationary Time Series", Berlin

This Version is available at:
https://hdl.handle.net/10419/230716

IRTG 1792 Discussion Paper 2018-005

Testing for bubbles in cryptocurrencies with time-varying volatility

Christian M. Hafner *

* Université catholique de Louvain, Belgium

This research was supported by the Deutsche Forschungsgemeinschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

Testing for bubbles in cryptocurrencies with time-varying volatility

Christian M. Hafner*

January 18, 2018

Abstract

The recent evolution of cryptocurrencies has been characterized by bubble-like behavior and extreme volatility. While it is difficult to assess an intrinsic value to a specific cryptocurrency, one can employ recently proposed bubble tests that rely on recursive applications of classical unit root tests. This paper extends this approach to the case where volatility is time varying, assuming a deterministic long-run component that may take into account a decrease of unconditional volatility when the cryptocurrency matures with a higher market dissemination. Volatility also includes a stochastic short-run component to capture volatility clustering. The wild bootstrap is shown to correctly adjust the size properties of the bubble test, which retains good power properties. In an empirical application using eleven of the largest cryptocurrencies and the CRIX index, the general evidence in favor of bubbles is confirmed, but much less pronounced than under constant volatility.

JEL: C22, C58, G20

Keywords: cryptocurrencies, speculative bubbles, wild bootstrap, volatility

JEL classification: **C14, C43, Z11**

*Institut de statistique, biostatistique et sciences actuarielles, and CORE, Université catholique de Louvain, Voie du Roman Pays, 20, 1348 Louvain-la-Neuve, Belgium, christian.hafner@uclouvain.be

1 Introduction

The question whether the evolution of some or all of the cryptocurrencies (cryptos) is driven by speculative bubbles has been conversely debated both in public media and academics. Some of the currently available cryptos such as Bitcoin have similarities with gold in that they are mined through complicated procedures and have a fixed global supply, see e.g. Gronwald (2014) for a comparison between gold and Bitcoin. However, while gold has a lower bound of its fundamental value given by its use in various physical applications, there is no such lower bound for cryptos, whose intrinsic value crucially hinges upon their acceptance as a payment device. The view that the fundamental value of Bitcoin is zero is supported empirically by the analysis of Cheah and Fry (2015) for the early Bitcoin market until 2014. Moreover, several studies cast doubt on the perception of cryptos, in particular Bitcoin, as a currency. For example, Yermack (2013) argues that, while Bitcoin is increasingly used as a medium of exchange, it can hardly be used as a unit of account and a store of value, mainly hampered by its excessive volatility.

As of today, Bitcoin and other cryptos seem to share more characteristics of alternative or speculative assets than currencies, see e.g. Glaser et al (2014). Indeed, given their low correlation with traditional currencies and financial assets, they can be considered as diversifiers in portfolios, safe haven investments and hedging instruments, a role traditionally played by gold. Bouri et al (2017) study this question and find that Bitcoin is a poor hedge but suitable for diversification purposes. Accepting the role of alternative assets, understanding the market and price dynamics is of primary importance for investors. Urquhart (2016) shows that the Bitcoin market is not weak-form market efficient, while Nadarajah and Chu (2017) use odd-power transformations of returns to obtain market efficiency. In various studies, price and return dynamics have been found to be intrinsically complex, with extreme volatility, spikes and jumps. For example, Gronwald (2014) uses an autoregressive jump-intensity GARCH model to descrive Bitcoin return dynamics until 2014, while Scaillet et al (2017) analyze the jump behavior of Bitcoin transaction data from 2011 to 2013.

Speculative bubbles may be rational when investors are aware of the fact that prices have moved away from fundamental values, but consider the probability of selling at a higher price more important than a possible crash. Irrational bubbles are mainly driven by psychological factors, herd instincts, etc. Famous examples of irrational bubbles are the Dutch tulip mania in 1636/37 and the south sea bubble in 1719-1721, see e.g. Dale et al. (2005). A more recent example is the dot come bubble 1997-2001. The current market of cryptocurrencies shows similarities with these historical events at least in its extreme price movements over short periods of time. It is remarkable that the three historical examples showed the evolution of the bubble and its burst over periods of roughly three years, while we have seen a dramatic increase in crypto prices over the period 2015-2017, but so far without a significant burst. The recent hype in public media attracts more and more people to the market driven by psychological factors, suggesting a possible irrational bubble.

As there seems to be consensus in the academic literature about the possible presence of speculative bubbles in cryptocurrencies, it is useful to formulate models that allow for bubble behavior while taking into account the complex dynamic features observed in cryptos. Hencic and Gouriéroux (2015) apply a non-causal autoregressive process with Cauchy innovations to the Bitcoin-US Dollar rate, which reproduces the bubble-like behavior of the rate over the period February to July 2013. Cheah and Fry (2015) use a continuous time model to identify bubbles via anomalous behaviors of the drift and volatility components. A possibly nonstationary behavior of volatility does not seem to have been taken into account explicitly for tests of bubbles in cryptocurrencies.

In this paper we follow the idea of the bubble test developed by Phillips et al (2011), PWY in the following, which essentially tests whether the price dynamics are characterized by a unit root versus explosive behavior. As shown by Harvey et al (2015), the original test of PWY is biased in the presence of non-stationary volatility. For cryptocurrencies we may expect volatilities to be nonstationary for various reasons, one of them being the typically high volatility shortly after the initial coin offering (ICO) and a possible decline once the crypto is becoming more mature and accepted as a payment medium. Harvey

et al (2015) suggest a wild bootstrap procedure to correct the size of the PWY test and show that is has good power properties. While they only consider deterministic volatility functions, it is important to also allow for a possible volatility clustering due to stochastic effects, such as GARCH or stochastic volatility. Moreover, we adjust the algorithm to account for serial correlation and highly significant non-zero skewness coefficients of daily log returns.

We combine deterministic and stochastic components by applying a version of the so-called Spline-GARCH model of Engle and Rangel (2008) to show that the long-term average levels of volatility change for most of the analyzed crytos, being typically higher at the beginning of their existence than at later stages, while short term volatility is driven by volatility clustering effects similar to classical financial assets. We then show in a simulation study that the wild bootstrap PWY test has good size and power properties in sample sizes typically available for the largest cryptos. We then apply this test to the eleven largest cryptos by market capitalization on the last day of our sample, 12/31/2017, as well as the CRIX index, see Trimborn and Härdle (2016). Our findings suggest that, out of eleven cryptos, for eight do we reject the null hypothesis of a unit root in favor of the explosive alternative at the 5% significance level. However, the obtained p-values are much larger than those of the original PWY distribution, suggesting that the evidence in favor of speculative bubbles is weaker under the presence of nonstationary volatility. Indeed, at the 1% level we can reject the unit root only in the case of Bitcoin, which hence shows the strongest evidence in favor of a bubble.

The remainder of the paper is organized as follows. The next section introduces the data used in this study and gives some summary statistics. Section 3 provides evidence in favor of time-varying volatility, decomposed into deterministic and stochastic effects. Section 4 presents the test for explosive speculative bubbles in a general framework allowing for time-varying volatility. Section 5 applies the test to the data and attempts to date-stamp the identified bubbles, and Section 6 concludes.

2 Data and summary statistics

Throughout the paper we will use the eleven largest cryptos on 12/31/2017, the last day of our sample, according to the data provider Coinmarketcap.com. Additionally, we use the crypto index CRIX developed jointly by Humboldt University Berlin, Singapore Management University and the data provider CoinGecko, see http://crix.hu-berlin.de/ and Trimborn and Härdle (2016) for details. Table 1 gives an overview of the data under study. In most, but not all cases, the start date corresponds to the date of the ICO. An exception is Bitcoin, which was launched already on January 3, 2009. Bitcoin cash is a so-called hard fork of the original Bitcoin, augmenting the blocksize of the latter from 1 to 8 MB and thus enhancing transaction times. Most of the cryptos use the Blockchain technology, but for example IOTA does not.

Table 2 provides summary statistics of the returns of each crypto. In terms of performance, two measures are given, the total return over the sample period and the annualized average log return. Because the sample periods differ, the best performer in total returns (XEM with 425,519%) is not the same as the one for annualized average log returns (ADA with 1,348.16%). These very large numbers, unusual for any other market, illustrate the general steep upward trend since the inception of cryptos. However, they are accompanied by equally large numbers for average annualized volatilities, such that the Sharpe ratios are more in line with typical values in other markets. Note that the volatilities tend to be lower for more mature cryptos and higher for more recent ones. Bitcoin as the oldest crypto has the lowest volatility, while the three most recent ones, i.e. BCH, ADA and IOTA, have the highest volatilities, suggesting that volatility is a function of the elapsed time since ICO. We will investigate this point further in the following section. As may be expected by similar findings for financial indices, the CRIX index as a weighted average of leading cryptos has the lowest average volatility, close to Bitcoin. It has mainly been dominated by Bitcoin so that the other statistics of the CRIX are also similar to those of Bitcoin.

Unlike for typical equity returns, sample skewness coefficients are mostly positive and

significantly different from zero, a fact that may be explained by larger extrema on the positive side than on the negative side, comparing the min and max values, and noting that higher order moments are very sensitive to outliers. The sample kurtosis coefficients, on the other hand, are comparable to typical values in high frequency financial assets and reflect the fat tails of return distributions. In all cases, classical normality tests such as Jarque-Bera reject normality at all conventional significance levels.

Serial correlation coefficients are small but often significant. To take this into account in the following, we fit an autoregressive model of order p, where p is chosen according to the Bayesian information criterion (BIC). The orders are reported in the last column of Table 2.

3 Time-varying volatility

In this section we give evidence for time-varying volatility of cryptos. We have seen in the previous section that volatility seems to be higher for cryptos whose ICO is in the recent past, and lower for more mature cryptos. This feature can be embedded into a volatility which is an explicit deterministic function of the number of elapsed trading days, or simply of time. This function could take into account other effects, for example the emergence of forks such as bitcoin cash in July 2017, or the introduction of futures contracts on the Bitcoin at the CBOE and CME in December 2017, which may have effects on the volatility of bitcoin prices. As in most cases these events are anticipated, we may model them using a deterministic function of time that will be estimated for each crypto.

Apart from deterministic effects, we also observe stochastic volatility changes, and the volatility clustering that is typically encountered in financial markets. Periods of high volatility alternate with those of low volatility in a random way. We will incorporate both features of volatility changes into a model very similar to the spline-GARCH model of Engle and Rangel (2008). To introduce the notation, we will call y_t the log return of a crypto on day t, and $g(t/T)$ a smooth deterministic function of time that captures the

effects due to calendar and anticipated events. For the stochastic part of the model, we consider two asymmetric GARCH versions, the threshold GARCH (TGARCH) model of Glosten et al. (1993) and the exponential GARCH (EGARCH) model of Nelson (1991). It is a priori not clear whether a leverage effect, typical for equity returns, should be present also in cryptos. Common findings in traditional exchange rates is that leverage effects are, if at all present, much weaker than in equity returns. However, given the current consideration of cryptos as speculative assets rather than currencies, we may suspect some asymmetry in the news impact functions. The complete model takes the form

$$\begin{aligned} \varepsilon_t &= g(t/T)\sqrt{h_t}\xi_t, \quad \xi_t \sim i.i.d.(0,1) && (1)\\ \text{TGARCH}: \quad h_t &= \omega + \alpha\frac{\varepsilon_{t-1}^2}{g(\frac{t-1}{T})^2} + \theta\frac{\varepsilon_{t-1}^2}{g(\frac{t-1}{T})^2}D_t + \beta h_{t-1} && (2)\\ \text{EGARCH}: \quad \log h_t &= \omega + \alpha(|\xi_{t-1}| - \mathrm{E}|\xi_{t-1}|) + \theta\xi_{t-1} + \beta \log h_{t-1} && (3) \end{aligned}$$

where ε_t is the error term of an AR(p) model fitted to log returns obtained as the first difference of log prices y_t, with the order p given in Table 2. D_t is a dummy variable that takes the value 1 if ε_{t-1} is negative, and zero otherwise, α, β and θ are parameters, which in the TGARCH case are restricted by $\alpha, \beta > 0$ and $\alpha + \beta + \theta\eta < 1$, where $\eta = \mathrm{E}(\xi_t < 0)$, and in the EGARCH case by $\beta < 1$. The TGARCH restriction does not assume that the distribution of ξ_t is symmetric, hence η is not necessarily equal to 1/2. We will only assume that $\mathrm{E}|\xi_t|^p < \infty$ for some $p > 2$.

Without further constraints, model (1)-(3) is not identified. We can obtain identifiability by assuming that the unconditional variance is given by $g(t/T)^2$ and therefore restricting $\mathrm{E}[h_t] = 1$, which in the TGARCH case means that $\omega = 1 - \alpha - \beta - \theta\eta$. In the EGARCH case, similar restrictions could be found but are more complicated as they depend on the distribution of ξ_t, see Nelson (1991). Note that this restriction is not necessary in the case of sequential estimation, estimating first the unconditional variance $g(t/T)^2$, then dividing through by the estimate of $g(t/T)$ and estimate h_t as in (2) or (3) without further constraint.

Under the above assumptions, h_t is weakly and strictly stationary and $g(t/T)^2$ is the function towards which the conditional variance, $g(t/T)^2 h_t$ reverts. The two main

differences with respect to Engle and Rangel (2008) is that rather than using a symmetric plain vanilla GARCH we use two more flexible asymmetric versions, and instead of using splines for the estimation of $g(t/T)$, we use a local linear kernel estimator.

Estimation is done in two steps, where first the nonparametric part $g(t/T)$ is estimated using a local linear kernel estimator, see e.g. Fan and Gijsels (1996). We use the Epanechnikov kernel and the bandwidth h chosen by least squares cross-validation to obtain an estimate $\hat{g}_h(\cdot)$. In the second step, the parameters of h_t are estimated using Gaussian quasi maximum likelihood, which provides consistent, if not efficient parameter estimates. Table 4 reports estimation results for the parametric part of the model, as a well as the ratio of the maximum and minimum values of $\hat{g}_h(\cdot)$. The latter statistic gives an indication for the variability of unconditional volatilities over the sample period and takes values in the range of 1.97 for XMR and 4.52 for DASH. Thus, the average volatility level varies over the sample period by a factor of roughly two to four and a half. Figures 1 and 2 show the estimated functions for BTC and CRIX. BTC has a typical shape, representative of many other cryptos that are not shown to save space. The typical shape is a high level in 2013, reflecting uncertainty at an early stage of the crypto as to whether it will be accepted by the market, a lower level during the consolidation period 2014 to 2016, and an increase in 2017, paralleling increasing media coverage, the introduction of Bitcoin Cash in July, futures contracts in December, and increasing concerns about market regulation, the closing of online exchanges in China and uncertainty about other countries such as South Korea.

The results of the parametric part in Table 4 can be summarized as follows:

(i) Volatility clustering, measured by the size effects α, is important and significant.

(ii) Persistence of shocks in volatility, measured by $\alpha + \beta + \theta\eta$ for TGARCH and β for EGARCH, is high but not as high as in standard GARCH models applied to financial returns. This is due to the presence of $g(t/T)$ which incorporates already a large part of the variation of volatility.

(iii) Unlike equities, cryptos do not systematically show an asymmetry of the news im-

pact function. In many cases, asymmetry is not significant. The cases with significant asymmetry have negative asymmetry for BTC, and positive for LTC, XLM and XMR.

(iv) EGARCH is the preferred model in eight cases, TGARCH in only four. This may be explained by the exponential increase of the news impact function as opposed to only quadratic of TGARCH, so that the often occurring extreme events in cryptos tend to be better captured by EGARCH.

To summarize these results, we find strong evidence for time-varying volatility in two ways: a change in unconditional volatilities, viewed as deterministic functions of time, and in conditional volatilities being driven by short term effects and stochastic factors. This will have an impact on the performance of statistical tests for speculative bubbles, to which we now turn.

4 Testing for explosive speculative bubbles

We follow the approach of Phillips et al (2011) to test for speculative bubbles, extended to our framework. Under the null hypothesis, log prices y_t have a unit root and follow, in the simplest case ignoring serial correlation of the error terms, a random walk, $y_t = y_{t-1} + \varepsilon_t$, where ε_t is white noise. Under the alternative, there is at least one sub-period of the sample for which the process is explosive, i.e. $y_t = (1+\delta)y_{t-1} + \varepsilon_t$ with $\delta > 0$, which may be followed, or not, by a collapse, i.e., a period with $\delta < 0$.

Classical unit root asymptotics of Dickey and Fuller (1979) continue to hold under heteroskedasticity as long as a so-called global homoskedasticity condition holds, see Cavaliere and Taylor (2009, CT in the following for brevity). This is the case, for example, for stationary GARCH processes. If global homoskedasticity does not hold, as in our case of model (1)- (3), the results of Dickey and Fuller (1979) no longer hold. CT provide a generalized framework, allowing for so-called global heteroskedasticity, and derive the unit-root asymptotics for this general case. Our model (1)- (3) fits into their framework

as we show in the following.

First, Assumption 1 of CT holds because ε_t is a martingale difference sequence w.r.t. $\mathcal{F}_{t-1}$, and $\sup_t \mathrm{E}[|\varepsilon_t|^p|\mathcal{F}_{t-1}] \leq Q_\varepsilon < \infty,\ a.s.$ for some $p > 2$ because of our assumption $\mathrm{E}|\xi_t|^p < \infty$ for some $p > 2$, where $\mathcal{F}_{t-1}$ is the σ-algebra generated by $\{y_1, \ldots, y_{t-1}\}$. Denote by $\sigma^2_{T,t-1} := \mathrm{Var}(\varepsilon_t|\mathcal{F}_{t-1}) = g(t/T)^2 h_t$ the conditional variance of ε_t given $\mathcal{F}_{t-1}$. Then we can show that Assumption 2 of CT holds by finding a strictly positive deterministic sequence $\{a_T\}$ such that $a_T^{-1}\sigma_T(\cdot) \to_d \Omega(\cdot)$ with $\int_0^1 \Omega(s)^2 ds > 0,\ a.s.$ For example, in the GARCH case, i.e. $\theta = 0$ in (2), we can set $a_T = 1$, and assuming near-integratedness as T increases, the process $\Omega(s)$ is given by

$$d\Omega(s)^2 = (g(s)^2 - \lambda\Omega(s)^2)ds + \phi\Omega(s)^2 dB(s)$$

with $B(s)$ a standard Brownian motion, and where $\phi = \lim_{T\to\infty}(2T)^{1/2}\alpha_T$ and $\lambda = \lim_{T\to\infty} = T(1 - \alpha_T - \beta_T),\ \lambda > -\phi^2/2$, see Nelson (1990). This is an extension of example 4 of CT, who treat the stationary case with a constant unconditional variance. Analogous results can be found for the TGARCH and EGARCH process using the results of Duan (1997) for the weak diffusion limits of the augmented GARCH model, which nests both as special cases. Finally, Assumption 3 of CT also holds in our case. We are therefore in a position to proceed with the test for speculative bubbles developed by Phillips et al (2011).

The test of Phillips et al (2011) is essentially the supremum of a forward recursive sequence of right-tailed Dickey-Fuller tests, and the procedure is the following. First, the minimum proportion of the sample to be used in the first DF test is chosen as $r_0 \in (0, 1)$, so that the first DF statistic is computed using the observations $1, \ldots, \lfloor r_0 T \rfloor$. This sample is sequentially augmented until the end of the full sample, T, each time recalculating the DF statistic. Then, the test statistic proposed by Phillips et al (2011) is given by

$$PWY := \sup_{r\in[r_0,1]} DF_r, \tag{4}$$

where DF_r is the augmented Dickey-Fuller test statistic, i.e. the t-ratio for the OLS

estimator of b_r in the regression

$$\Delta y_t = a_r + b_r y_{t-1} + \sum_{j=1}^{p} \alpha_j \Delta y_{t-j} + \varepsilon_{t,r}, \quad t = 1, \ldots, \lfloor rT \rfloor.$$

The alternative hypothesis is not stationarity but explosiveness, so that rejections occur not in the left tail but in the right tail of the distribution under the null. Table 1 of Phillips et al (2011) give critical values of this distribution, but Harvey et al (2015) show that these are incorrect if the error process ε_t is not globally homoskedastic in the sense of Cavaliere and Taylor (2009). They propose a wild bootstrap procedure to correct for the size of the PWY test for the case $p = 0$ using perturbations of residuals by standard normal random variables, see e.g. Mammen (1993). Their procedure has a good performance for deterministic volatility functions and symmetric innovation distributions.

In our case, as we have seen in Section 2 and Table 2, return distributions are often highly skewed to the right, due *inter alia* to positive extremes being more frequent than negative ones. Moreover, as we have seen in Table 2, serial correlation of returns is often not negligible so that we would like to reproduce this autocorrelation also in the bootstrap samples. To better approximate the distribution of the PWY distribution under skewed returns and using augmented DF tests, we therefore propose the following generalization of the algorithm of Harvey et al (2015):

Step 1. Generate independent $u_t \sim N(0,1)$ and $v_t \sim N(0,1)$, $t = 1, \ldots, T$, and compute $w_t = u_t/\sqrt{2} + (v_t^2 - 1)/2$. By construction, $\mathrm{E}[w_t] = 0$, $\mathrm{E}[w_t^2] = 1$, and $\mathrm{E}[w_t^3] = 1$.

Step 2. Generate T bootstrap innovations ε_t^* as: $\varepsilon_1^* = 0$, $\varepsilon_t^* = w_t \hat{\varepsilon}_t$, $t = 2, \ldots, T$, where $\hat{\varepsilon}_t$ are the OLS residuals of the fitted AR(p) models. By construction, $\mathrm{E}[\varepsilon_t^*] = 0$, $\mathrm{E}[\varepsilon_t^{*2}] = \hat{\varepsilon}_t^2$, and $\mathrm{E}[\varepsilon_t^{*3}] = \hat{\varepsilon}_t^3$.

Step 3. Set $y_t^* = 0, t \leq 0$, and generate recursively the bootstrap sample as $y_t^* = \hat{a} + y_{t-1}^* + \sum_{j=1}^{p} \hat{\alpha}_j \Delta y_{t-j}^* + \varepsilon_t^*$, $t = 1, \ldots, T$.

Step 4. Compute the bootstrap test statistic

$$PWY^* := \sup_{r\in[r_0,1]} DF_r^*,$$

where DF_r^* is the t-ratio for the OLS estimator of b_r^* in the regression

$$\Delta y_t^* = a_r^* + b_r^* y_{t-1}^* + \sum_{j=1}^{p} \alpha_j \Delta y_{t-j}^* + \varepsilon_{t,r}^*, \quad t = 1, \ldots, \lfloor rT \rfloor.$$

Step 5. Repeat steps 1 to 4 B times to obtain bootstrap test statistics PWY_b^*, $b = 1, \ldots, B$. Then, calculate the simulated bootstrap p-value as $B^{-1}\sum_{b=1}^{B} I(PWY_b^* > PWY)$.

If the test rejects the null hypothesis, i.e., the bootstrap p-value is smaller than a nominal significance level, then it is possible to estimate the timings of the bubble and its collapse, if it occurs, following the procedure of Phillips et al.(2011). In particular, they consider the time series of sequential Dickey-Fuller statistics DF_r, and define the estimate of the beginning of the bubble period as $r_1 := \inf_{r\geq r_0}\{r : DF_r > cv_{\alpha_T}(r)\}$, and the burst, should it occur, as $r_2 := \inf_{r\geq r_1}\{r : DF_r > cv_{\alpha_T}(r)\}$, where $cv_{\alpha_T}(r)$ is a critical value at r corresponding to a significance level α_T. In order to eliminate the Type I error probability, $cv_{\alpha_T}(r)$ should increase to infinity at a slow rate, or equivalently, α_T should decrease to zero. Similar to PWY, the distribution of DF_r is also affected by global heteroskedasticity, but can also be simulated using the wild bootstrap for given observations.

In the following we provide simulation evidence for the consistency of the above algorithm in our model framework, i.e., allowing for both deterministic and stochastically varying volatility, and skewed innovation distributions. The setup is as follows. We generate the process $y_t = (1 + \delta J_t)y_{t-1} + \varepsilon_t$, where $J_t = I(t \geq T/2)$ such that the process is explosive in the second part of the sample if $\delta > 0$, and the process does not collapse. Under the null, $\delta = 0$, and for the alternative we set $\delta \in \{0.01, 0.02, 0.03\}$. The error process ε_t is generated as in (1) with ξ_t drawn from a centered and normalized negative

log-χ^2 distribution[1]. The smooth, deterministic part of volatility is set as

$$g(t/T) = 0.05(1 + c\cos(\pi t/T)^2)$$

where $c \in \{1, 2\}$ is a constant to determine the degree of global heteroskedasticity. For the parametric part, we use a standard GARCH model, i.e. in (2) we set $\theta = 0$, while α is set to 0.1 and β to 0.85. The sample size is set to $T = 100$ and $T = 200$. The number of bootstrap replications B is set to 5,000, while each process is generated $N = 500$ times. The simulation results are reported in Table 4 for a nominal level of five %.

Note that the test is undersized in small samples, with the bias increasing with the degree of global heteroskedasticity. The empirical size converges however to the nominal size as the sample size increases. The power of the test is slightly larger for the case $c = 2$, i.e. large variation of $g(t/T)$, and increases as expected both in T and in δ. For the case $\delta = 0.03$ and $T = 500$ the estimated power is virtually equal to one. This confirms the good performance of the wild bootstrap test also in the more general framework of time-varying volatility with deterministic and stochastic effects, as well as skewed innovations, i.e. a realistic framework in the light of our results of Sections 2 and 3. We will now apply this test to our set of cryptos and the CRIX.

5 Empirical results

We apply the PWY test to the set of leading cryptos and the CRIX index, taking into account time-varying volatility. We implement the test using a starting proportion of 10% of the data, i.e. $r_0 = 0.1$, which corresponds to the choice of Phillips et al (2011). For the wild bootstrap we use 5,000 replications. Table 5 reports the main results. The first column contains the obtained values of the PWY statistic (4). The next column gives an estimate of the p-value of the PWY statistic obtained by simulating 10,000 random walk processes with normally distributed error terms, as in Phillips et al (2011). The 1%

[1]Generate i.i.d. $Z_t \sim N(0, 1)$, set $\tilde{\xi}_t = -\log Z_t^2$, $t = 1, \ldots, T$, and then standardize $\tilde{\xi}_t$ to have mean zero and variance one. This r.v. is right skewed.

and 5% upper quantiles of this distribution closely correspond to the values reported in Table 1 of Phillips et al (2011). Given our discussion above and the results of CT, these p-values are not reliable in our context of global heteroskedasticity, but it will allow us to assess the impact of time-varying volatility on the results of the PWY test. The next two columns of Table 5 report estimates of the 1% and 5% upper quantiles of the distribution of the PWY statistic using the wild bootstrap with 5,000 replications. Finally, the last column gives the p-value corresponding to this distribution, which appropriately takes into account global heteroskedasticity.

Let us first compare the two columns containing p-values, with and without bootstrap correction. Taking into account global heteroskedasticity has the effect of increasing the p-values that were found to be very close to zero, which is the case for BTC, XEM, IOTA, ADA, XLM, LTC, and the CRIX. In other words, the evidence in favor of speculative bubbles is much less pronounced than when this effect is ignored. However, not all p-values increase, in particular the larger ones tend to decrease under the correction, which is the case for BCH, ETH, XMR, and DASH.

We now analyze further the p-values obtained with the wild bootstrap correction, i.e. the last column of table 5. Note that all p-values are quite small, the largest one being 0.1816 for BCH, the smallest one 0.0032 for BTC. Only one p-value is smaller than 1% (BTC with 0.0032), meaning that at the 1% level one would reject the null hypothesis in favor of a speculative bubble only for the Bitcoin. At the 5% level, this number increases to 9 (all cryptos except IOTA, BCH and XLM). Thus, if one accepts a 5% significance level, there is generally strong evidence in favor of speculative bubbles in the majority of analyzed cryptos, and the CRIX index, even after having corrected for the size of the test due to global heteroskedasticity. The effect seems strongest in the oldest and most liquid crypto, the Bitcoin.

Finally, we turn to the analysis of the time series of recursive DF_r statistics to date-stamp the bubbles, as in Phillips et al. (2011). They choose critical values that are close to the 4% level. We have to size-correct this distribution as it is affected by global heteroskedasticity, which we again do by the wild bootstrap. Using this distribution, we

choose a critical value function that at the end of sample is close to the 1% level, while slowly increasing to infinity to ensure consistency of the dating procedure. This can be achieved by the function $cv_{\alpha_T}(r) = \log(\log(Tr))/c$, where c is chosen such that $cv_{\alpha_T}(1)$ corresponds roughly to the 1% upper quantile of the simulated distribution of DF_1 under heteroskedasticity. For Bitcoin and CRIX this gives the constants $c = 2$ and $c = 1.5$, respectively, such that the range of critical values is (0.8186 1.003) for the Bitcoin, and (1.0507, 1.3096) for the CRIX.

Figure 3 shows the series of DF_r statistics together with critical values for the Bitcoin. Apart from a very short crossing beginning of 2014, there are essentially two bubble periods: the first starting on November 7, 2013 and lasting until December 18, 2013, while the second starts on November 27, 2017, and has not collapsed until the end of the sample. The sequential DF_r statistics are shown in Figure 4 for the CRIX index, together with corresponding critical values. Here we have clearly a single bubble, and the starting date is estimated as May 5, 2017. The bubble has not collapsed before the end of the sample, although it came close to the critical value twice.

To summarize our findings, at the five percent significance level there is strong evidence in favor of a presence of explosive speculative bubbles in cryptocurrencies, including the Bitcoin and the CRIX index, even taking into account non-stationarity and global heteroskedasticity of the involved time series. This evidence is much weaker, however, at the 1% percent level, where essentially only the Bitcoin displays a significant bubble behavior. Finally, the dating procedure allowed us to identify bubbles in the Bitcoin and the CRIX at the end of the sample in 2017, which have not collapsed, and which might be explained by the general media hype during this period, the introduction of Bitcoin futures contracts in December 2017, and increasing uncertainty about market regulation.

6 Conclusions

We have seen that it is possible to test for explosive speculative bubbles in a general context of nonstationarity and, in particular, global heteroskedasticity. This is important

in applications to cryptocurrencies, which due to their non-maturity still suffer from erratic behavior, extreme events, and very high and changing volatility. Thus, correcting the size of sequential Dickey-Fuller-type tests is highly recommended in applications. Our results suggest that, despite the size correction and robustness of our approach, the majority of test results remain significant at the 5% level, although not at the 1% level. The crypto that shows the strongest bubble behavior is the Bitcoin, taking size-corrected p-values as a criterion. This may be explained by the fact that it is the Bitcoin that attracts the general public as it is the best known crypto that also enjoys the widest media coverage. However, its market share has fallen steadily over the last years, and it is likely that the configuration of leading cryptos changes rapidly in the near future.

In the Bitcoin case, the date-stamping algorithm of Phillips et al (2011) suggested the presence of two bubbles. In case of multiple bubbles, the generalized approach of Phillips et al (2015a,b) refines the date-stamping strategy in several ways to obtain more precise estimates. In principle, it is possible to use these strategies also in our context of global heteroskedasticity, although computationally challenging, which we leave for future research.

Appendix: Figures and Tables

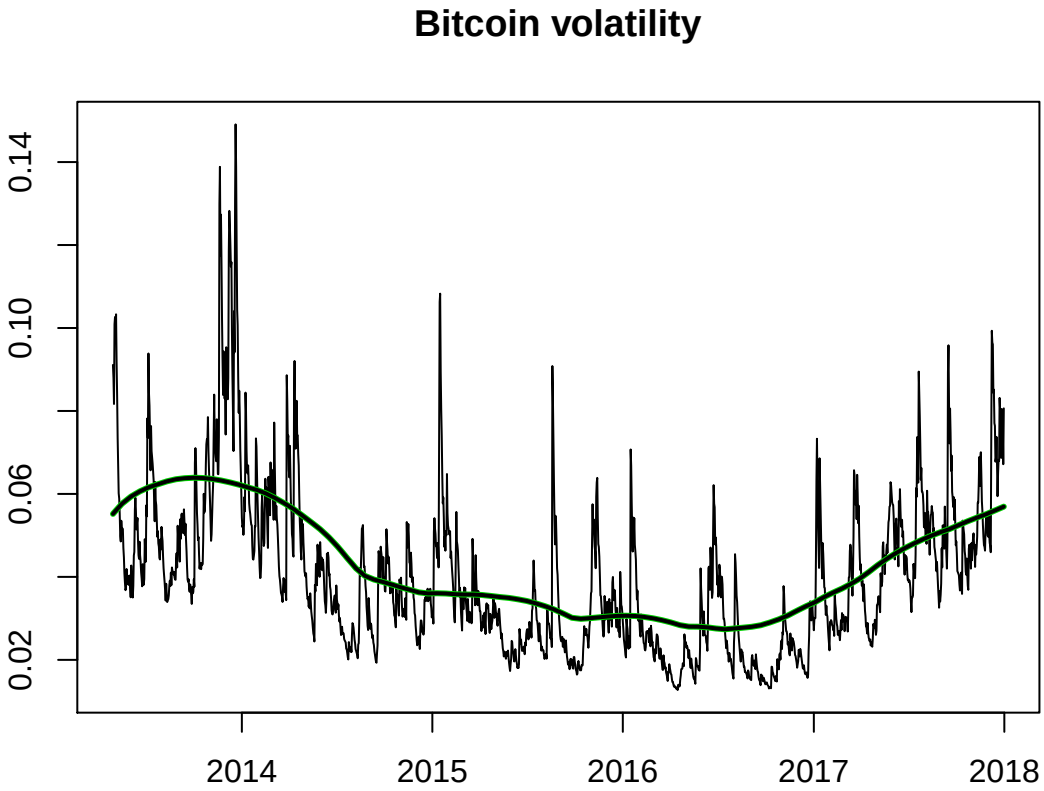

Figure 1: Estimated conditional (rough line) and unconditional (smooth line) standard deviations of daily log returns of Bitcoin.

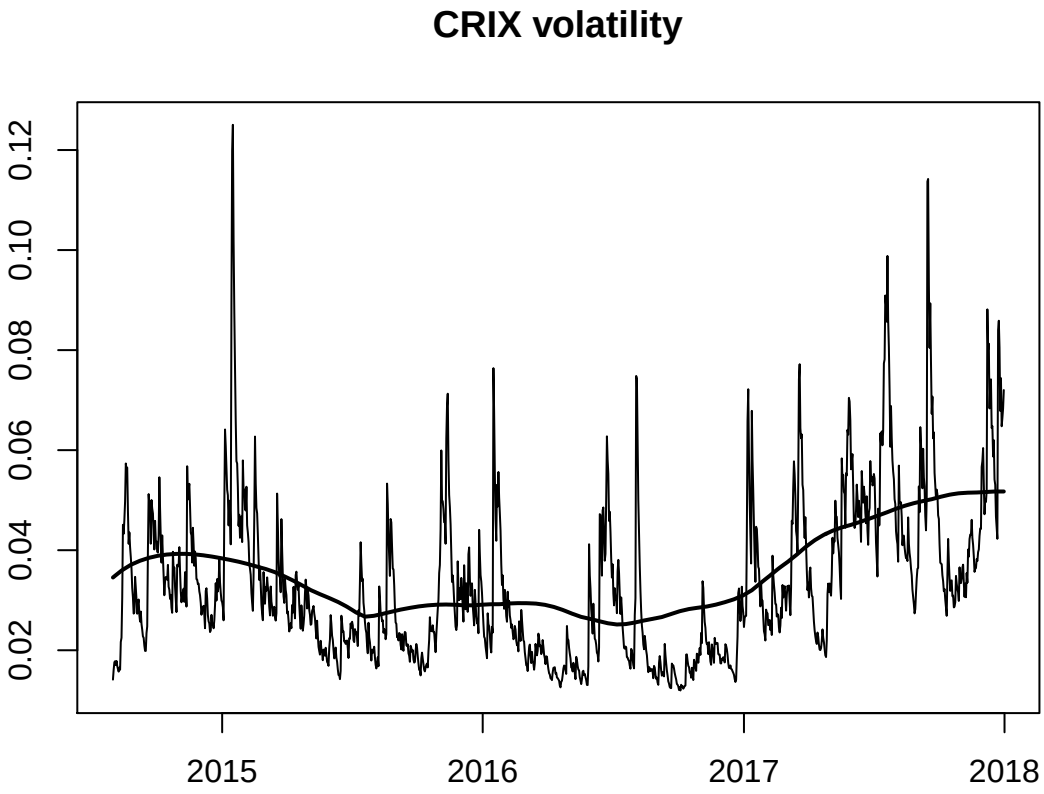

Figure 2: Estimated conditional (rough line) and unconditional (smooth line) standard deviations of daily log returns of the CRIX index.

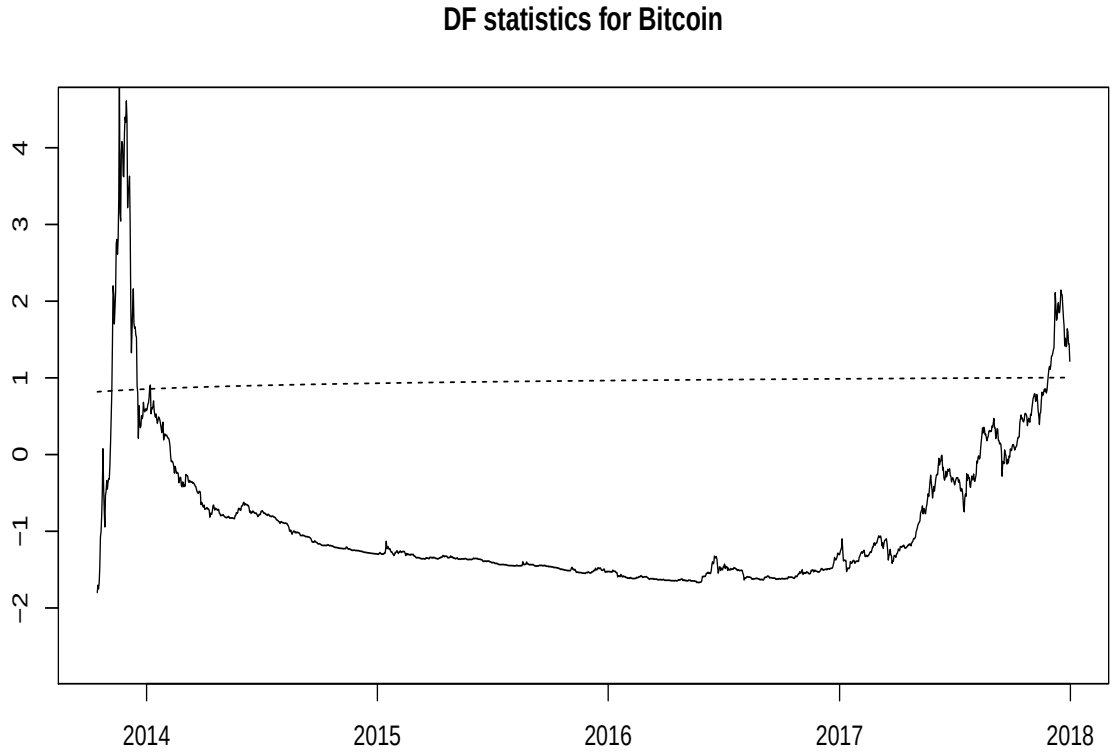

Figure 3: Time series of recursive DF_r statistics for Bitcoin. The dashed line is the function $\log(\log(Tr))/2, r \in [0.1, 1]$ giving the critical value as a function of the sample size and corresponding roughly to the 1% upper quantile of the DF_1 distribution.

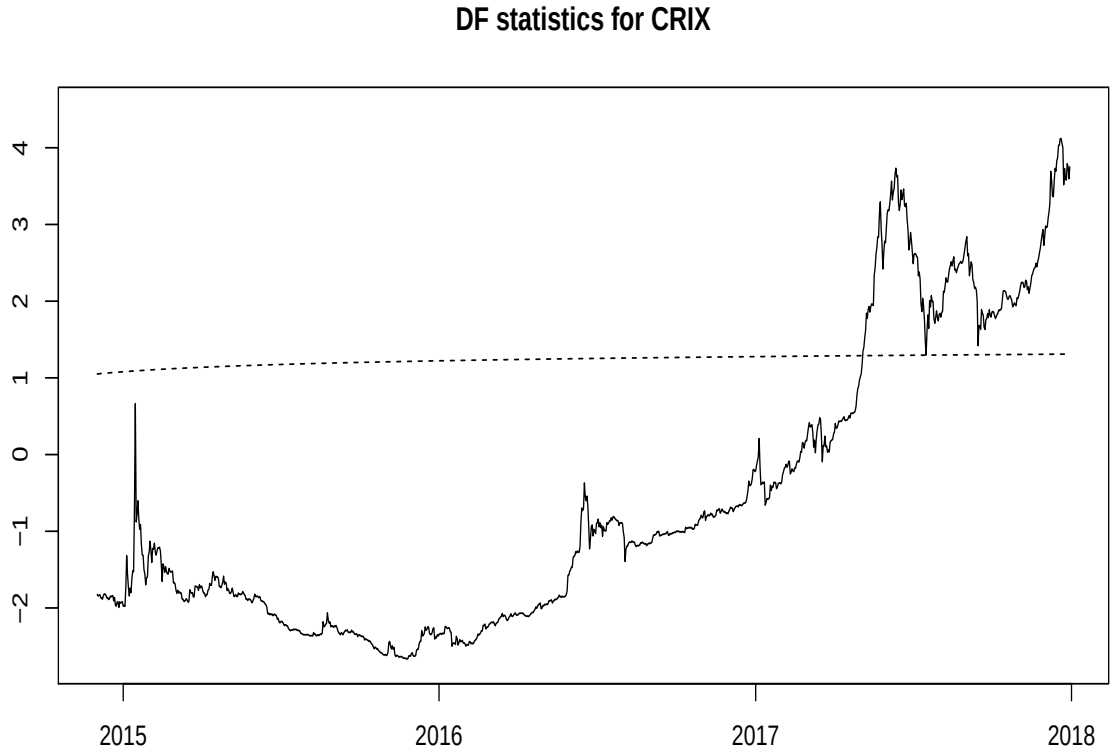

Figure 4: Time series of recursive DF_r statistics for CRIX. The dashed line is the function $\log(\log(Tr))/1.5, r \in [0.1, 1]$ giving the critical value as a function of the sample size and corresponding roughly to the 1% upper quantile of the DF_1 distribution.

	Name	Symbol	Market cap	% share	Start date	number obs
1	Bitcoin	BTC	220,903,949,498	38.59	04/28/2013	1709
2	Ripple	XRP	82,199,880,481	14.36	08/04/2013	1611
3	Ethereum	ETH	69,767,510,695	12.19	08/07/2015	878
4	Bitcoin Cash	BCH	41,526,715,510	7.25	07/23/2017	162
5	Cardano	ADA	18,030,140,406	3.15	10/01/2017	92
6	Litecoin	LTC	12,000,947,760	2.10	04/28/2013	1709
7	IOTA	MIOTA	9,564,670,064	1.67	06/13/2017	202
8	NEM	XEM	8,389,826,956	1.47	04/01/2015	1006
9	Dash	DASH	7,850,364,658	1.37	02/14/2014	1417
10	Stellar	XLM	5,756,694,225	1.01	08/05/2014	1245
11	Monero	XMR	5,255,620,533	0.92	05/21/2014	1320
		CRIX			07/31/2014	1250

Table 1: *Market cap is the market capitalization in USD as of December 31, 2017, % share is the corresponding market share, where the total market capitalization is 572,480,327,078 USD. Start date is the first day of the sample available at the data provider Coinmarketcap.com, each sample ending on 12/31/2017, and number obs is the number of observations in the sample.*

Symbol	total-ret	mean	vola	skew	kurt	min	max	p
BTC	10,447	99.53	84.10	-0.138	11.89	-0.266	0.357	6
XRP	39,002	135.31	151.95	2.256	32.12	-0.616	1.027	0
ETH	27,793	234.07	162.77	-3.72	67.63	-1.302	0.412	6
BCH	513	411.14	239.02	0.125	5.63	-0.446	0.432	2
ADA	2,782	1,348.16	316.37	2.200	11.04	-0.288	0.862	0
LTC	5,235	84.97	132.02	1.883	29.81	-0.514	0.829	0
MIOTA	501	325.89	221.44	0.150	4.882	-0.377	0.383	2
XEM	425,519	303.50	177.92	2.022	19.79	-0.361	0.995	2
DASH	281,097	204.69	164.14	3.153	44.54	-0.467	1.270	4
XLM	14,685	146.58	159.81	2.135	18.76	-0.366	0.723	1
XMR	21,714	149.03	148.88	0.750	9.09	-0.378	0.584	8
CRIX	4,736	113.37	68.58	-0.510	10.35	-0.238	0.198	0

Table 2: *Descriptive statistics of returns. total-ret is the total gross return over the sample period in %. mean is the annualized average log return in %, vola is the annualized standard deviation in %, skew, kurt, min and max are the skewness, kurtosis, minimum and maximum of daily log returns, respectively. p is the order of an AR(p) model fitted to returns, including a constant, chosen by the Bayesian information criterion (BIC).*

Symbol		TGARCH				EGARCH			
	g-ratio	α	θ	β	Lik	α	θ	β	Lik
BTC	2.37	0.159	-0.001	0.791	-2215	0.310	-0.025**	0.918	**-2206**
XRP	3.43	0.741	-0.444	0.403	**-1978**	0.615	0.173	0.680	-1989
ETH	4.22	0.299	-0.025	0.635	-1099	0.439	0.024	0.864	**-1096**
BCH	3.19	0.264	0.009	0.383	**-216**	0.343	0.058	0.250	-218
ADA	3.03	0.869	-0.894*	0.216	-120	0.171	0.512**	0.530	**-118**
LTC	3.27	0.110	-0.054**	0.857	**-2150**	0.173	0.051**	0.928	-2158
MIOTA	2.34	0.116	-0.110	0.548	-278	0.131	0.062	0.684	**-277**
XEM	2.61	0.451	0.044	0.408	-1321	0.539	-0.013	0.796	**-1317**
DASH	4.52	0.338	-0.100*	0.541	-1842	0.435	0.046	0.807	**-1838**
XLM	2.90	0.602	-0582**	0.172	**-1614**	0.348	0.231**	0.606	-1616
XMR	1.97	0.162	-0.119**	0.824	-1779	0.176	0.099**	0.928	**-1778**
CRIX	2.03	0.209	0.005	0.761	-1611	0.376	-0.018	0.911	**-1599**

Table 3: *Estimation results for the semiparametric volatility models. The column g-ratio reports the maximum of the estimated unconditional standard deviation $g(t/T)$ to its minimum. In the parametric part, θ characterizes asymmetry in both TGARCH and EGARCH. Asterisks indicate significance of this parameter at the 5 (*) and 1 (**) percent level. All α and β estimates are significant at the 1% level. The columns Lik give the log likelihood value for the parametric part of each model, the higher one being marked in bold.*

δ	$T = 100$	$T = 200$	$T = 500$
		$c = 1$	
0.00	0.024	0.048	0.050
0.01	0.038	0.284	0.664
0.02	0.192	0.552	0.907
0.03	0.569	0.891	0.995
		$c = 2$	
0.00	0.022	0.040	0.045
0.01	0.032	0.325	0.703
0.02	0.239	0.565	0.924
0.03	0.607	0.933	1.000

Table 4: *Monte Carlo simulation results for testing the null hypothesis* $\delta = 0$ *against* $\delta > 0$. *The entries are rejection frequencies of the null hypothesis. The sample size is* T, *each process is generated 500 times, and the number of bootstrap replications is 5,000. The degree of global heteroskedasticity is controlled by the parameter* c.

Symbol	PWY	p-value	WB 5%	WB 1 %	WB p-value
BTC	5.0347	0.0001	2.8349	4.2929	0.0032
XRP	1.9482	0.0132	1.3278	2.4430	0.0188
ETH	1.4184	0.0615	1.2800	2.1170	0.0388
BCH	0.5976	0.3205	1.3724	2.4410	0.1816
ADA	4.0663	0.0001	3.6213	6.1345	0.0354
LTC	5.8502	0.0001	4.1710	6.1244	0.0126
IOTA	1.6923	0.0291	2.4275	3.7207	0.1078
XEM	2.1801	0.0065	2.1519	3.0399	0.0468
DASH	0.5030	0.3657	0.3158	0.9317	0.0316
XLM	2.3411	0.0036	2.3807	3.6005	0.0512
XMR	1.6452	0.0330	1.0290	1.7062	0.0112
CRIX	4.1194	0.0001	2.4880	4.3962	0.0122

Table 5: *Empirical results of the PWY test applied to the cryptos and the CRIX. PWY is the value of the test statistic, p-value is the simulated p-value under the assumptions of Phillips et al (2011), WB 1 % and WB 5 % are the 1% and 5% critical values of the PWY distribution obtained by the wild bootstrap. The last column is the estimated p-value of the test using the distribution obtained by wild bootstrap.*

References

Bouri, E., Molnar, P., Azzi, G., Roubaud, D. and Hagfors, L.I. (2017). On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier? *Finance Research Letters*, 20, 192-198.

Chea, E.-T. and Fry, J. (2015). Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters*, 130, 32-36.

Dale, R.S., Johnson, J.E.V. and Tang, L.L. (2005). Irrational behavior during the South Sea bubble. *The Economic history review*, 83, 233-271.

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* 74, 427-431.

Duan (1997) Augmented GARCH(p, q) process and its diffusion limit, *Journal of Econometrics*, 79, 97-127.

Engle and Rangel (2008). The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes, *The Review of Financial Studies*, 21, 11871222.

Fa, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*, Chapman & Hall.

Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M.C. and Siering, M. (2014). Bitcoin - Asset or Currency? Revealing Users' Hidden Intentions. In: Proceedings of the twenty-second European conference on information systems, Recanati Business School, Tel Aviv University, Tel Aviv.

Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks, *The Journal of Finance*, 48, 1779-1801.

Gronwald, M. (2014). The economics of bitcoin - Market characteristics and price jumps. CESifo working paper 5121, Munich.

Harvey, D.I., Leybourne, S.J., Sollis, R. and Taylor, A.M.R. (2015). Tests for explosive financial bubbles in the presence of non-stationary volatility, *Journal of Empirical Finance*, 38, 548-574.

Hencic A. and Gouriéroux C. (2015). Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rates. In: Huynh VN., Kreinovich V., Sriboonchitta S., Suriya K. (eds) Econometrics of Risk. Studies in Computational Intelligence, vol 583. Springer.

Mammen (1993). Bootstrap and wild bootstrap for high dimensional models, *Annals of Statistics*, 21, 255-285.

Nadarajah, S. and Chu, J. (2017). On the inefficiency of Bitcoin, *Economics Letters*, 150, 6-9.

Nelson, D.B. (1990) ARCH models as diffusion approximations, *Journal of Econometrics*, 45, 7-38.

Nelson, D.B. (1991). Conditional heteoskedasticity in asset returns: A new approach. *Econometrica* 59, 347-370.

Phillips, P.C.B., Shi, S. and Yu, J. (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, 56, 10431078.

Phillips, P.C.B., Shi, S. and Yu, J. (2015b). Testing for multiple bubbles: Limit theory of real-time detectors *International Economic Review*, 56, 10791134.

Phillips, P.C.B., Wu, Y. and Yu, J. (2011). Explosive behavior in the 1990s NASDAQ: When did exhuberance escalate asset values?, *International Economic Review*, 52, 201-226.

Scaillet, O., Treccani, A. and Trevisan, C. (2018). High-frequency jump analysis of the bitcoin market. discussion paper.

Trimborn, S. and Härdle, W.K. (2016). CRIX: An index for blockchain based currencies, SFB 649 discussion paper, Humboldt University Berlin.

Urquhart, A. (2016). The inefficiency of Bitcoin, *Economics Letters*, 148, 80-82.

Yermack, D. (2013). Is bitcoin a real currency? discussion paper, NYU Stern School of Business.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

001 "Data Driven Value-at-Risk Forecasting using a SVR-GARCH-KDE Hybrid" by Marius Lux, Wolfgang Karl Härdle and Stefan Lessmann, January 2018.

002 "Nonparametric Variable Selection and Its Application to Additive Models" by Zheng-Hui Feng, Lu Lin, Ruo-Qing Zhu asnd Li-Xing Zhu, January 2018.

003 "Systemic Risk in Global Volatility Spillover Networks: Evidence from Option-implied Volatility Indices " by Zihui Yang and Yinggang Zhou, January 2018.

004 "Pricing Cryptocurrency options: the case of CRIX and Bitcoin" by Cathy YH Chen, Wolfgang Karl Härdle, Ai Jun Hou and Weining Wang, January 2018.

005 "Testing for bubbles in cryptocurrencies with time-varying volatility" by Christian M. Hafner, January 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.