MASTER THESIS

BITCOIN DATA ANALYSIS

Professor: Nicolas Vieille

Marta Anadón Rosinach

Higinio Raventós

Executive Summary

This paper analyses 26 time series that measure daily data for different attributes of the Bitcoin network and studies how the virtual currency behaves compared to a basket of currencies containing the Brazil Real (BRL), the Chinese Yuan (CNY), the Euro (EUR), and the Japan Yen (JPY) against the US Dollar (USD).

Basic statistics about the time series have been taken and stationarity has been studied in order to build sterilized fact data and meaningful cointegrations have been found among them. By applying a Vector Autoregressive (VAR) model, a regression has been built among the currencies and the Granger causality test has been applied in order to determine whether one time series (of a given currency) is useful in forecasting another and to observe causal relationships among the currencies studied.

Table of Contents

Section 1: Introduction

The history of virtual currencies is overlapped to that of traditional ones and its existence is just an extension caused by technology advancements trying to get rid of some of their shortcomings.

Currency is defined as something that can be used as a medium of exchange, that is, a unit of purchasing power. It can be used in exchange of something else because it has value. Once, commodity-based currencies were the standard. People could exchange goods or services for precious metals or even cacao beans. Later on, coins appeared, and were nothing more than a small piece made of a precious metal, initially gold and silver, and were followed by paper money, which could be exchanged for gold at a later time. This represented a revolution. It meant that governments could issue paper bills that were backed by something of inherent value. A single bill had – or represented – the same value of a fraction of the government's reserves of gold/silver. Citizens trusted paper money and found it convenient, so its adoption spread across the world quickly. Other advantages were that paper money reduced the transport of gold and silver, thus reducing risks; it made loaning gold or silver at interest easier, since the specie (gold or silver) never left the possession of the lender until someone else redeemed the note; and it allowed for a division of currency into credit and specie backed form. However, paper money also had many disadvantages. It has no intrinsic value and it just represents a fraction of the value of the reserves; therefore printing more paper money (increasing the money supply) would decrease its value, creating inflationary pressures. Noteworthy is the fact that in the beginning of the 20th century, most of the industrializing nations were on some form of gold standard, and by the end of the century, none of them were. As a matter of fact, the United States was one of the last countries to leave the standard and, by 1971, the Federal Reserve stopped redeeming notes in gold, silver or any other commodity.

With the invention of computers and the Internet, electronic money was born. This meant that money started being deposited and transferred electronically. Existing currencies did not need a physical form, as they could now be transferred electronically.

Central banks are the public institutions in charge of managing the different currencies and have the ability to alter the monetary base. The value of a given currency largely depends on the decision taken by the issuing central bank. The central banks need to make sure that the means of payment has a fixed purchasing power. However, many decentralized virtual currencies, such as Bitcoin, do not have a central authority and serve as an alternative currency whose monetary base is determined by an algorithm and its price solely depends on supply and demand.

The aim of this project is to analyse the behaviour of several variables related to Bitcoin, a decentralized virtual currency that is gaining popularity and presents itself as a threat to traditional currencies. Bitcoin has some other special traits such as low transaction fees, instant and irreversible transactions, high anonymity, and openness. The exchange rates depend on the supply and demand of Bitcoin, which is usually acquired in any of the available exchange services.

This virtual currency is a rather new phenomenon and there is not much research on its behaviour. However, there is already a sufficient amount of data available in order to conduct the study. Due to the lack of past research, this paper will try to start by analysing time series of several variables and building sterilized fact data.

Once the behaviour of the data is understood, the study will focus on looking for possible cointegration among the variables. Furthermore, a multivariate analysis of prices will be performed in order to see if the evolution of the Bitcoin exchange rate behaves similarly to that of a control basket that contains the Euro, the Brazilian Real, the Japanese Yen, and the Chinese Yuan against the US Dollar.

How the Bitcoin protocol works: Summary

As previously mentioned, Bitcoin is a decentralized virtual currency. Thousands of computers - called "miners" - work together to keep the system secure. In order to encourage people to join the network with their computer, the system gives incentives by giving rewards. A reward is given approximately every 10 minutes, and its value is halved after a certain amount of blocks - 210,000 - have been created. This ensures that the supply of Bitcoins increases at a given rate (not linear), until a total of 21 million Bitcoins have been be created. When the maximum is reached, miners' reward will only consist of transaction fees. A block contains information about transactions occurring in the network, and "miners" are the ones who create them. In order to do this, they have to solve a mathematical puzzle, and they can only do it by trial & error: they make a cryptographic hash of information about the current and last block, which yields a number. This number must by lower than a certain number, which is determined by the "difficulty" at the time. This "difficulty" changes as "hash rate" – the speed at which hashes are performed – changes. It usually increases when new "miners" enter or when technology enhancements occur. On the user level, individuals can acquire Bitcoins by acquiring them to other Bitcoin holders – this is usually done through an exchange. The main exchange of the Bitcoin network is called MtGox, and is based in Tokyo, Japan.

Bitcoin transactions are public. Any physical person with internet access can download the blockchain – a file that contains all the blocks, which in turn contain all the transactions in the network. Each transaction is associated with two Bitcoin addresses (one from the sender and another one from the receiver). Because if this, if any past or future transactions can be tied to an actual identity, it may be possible to guess the owner of the addresses with which there is interaction. Because of this, it cannot be said that Bitcoin is completely anonymous.

Bitcoin is secure. Or at least, the Bitcoin protocol is. The cryptographic function used for creating every block is SHA-256, designed by the U.S. National Security Agency. This paper is not going to address the technical details of the encryption methods used. However, it is not guaranteed that the exchange services are secure, as they are completely independent of the Bitcoin protocol, and only act as an intermediary between buyers and sellers.

Bitcoin is scarce. Bitcoin's scarcity is mathematically guaranteed. The Bitcoin network cannot, once the public accepts the currency, face the temptation of issuing large quantities of currency [(1)]. At the same time, it cannot have the spikes in supply that some pure commodities have, such as minerals when new mines are found. This means that there will be no unexpected inflation, which could lead to constant price deflation.

(1) "How Would the Invisible Hand Handle Money? - See Selgin, George and Larry White (1994)

Section 2: Literature Review

Because Bitcoin is a rather new phenomenon, most of the existing literature is based on explaining how it works and what their characteristics are, such as in the paper:

"Bitter to Better — How to Make Bitcoin a Better Currency" - Simon Barber, Xavier Boyen, Elaine Shi, and Ersin Uzun (http://crypto.stanford.edu/~xb/fc12/bitcoin.pdf)

The initial paper released by Satoshi Nakamoto – the pseudonymous person or group of people who designed and created the original Bitcoin software – explains how the core Bitcoin protocol works and suggests a statistical approach aiming to determine the probability of success to an attack to the Bitcoin network. In a few words, it suggests that the higher the total hash rate (computational power) of the Bitcoin network, the higher the hash rate of the attacker has to be in order to succeed.

"Bitcoin: A Peer-to-Peer Electronic Cash System" – Satoshi Nakamoto (http://bitcoin.org/bitcoin.pdf)

Some of the papers reviewed aim at analysing Bitcoin at a fundamental level, such as the following paper, which analyses Bitcoin's traits as both a currency and a commodity:

"Quasi-Commodity Money" – George Selgin
(http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2000118)

This paper also studies the time series related to the number of transactions, and looks for possible cointegrations between them and other variables. The following paper studies all the transactions until 13-May-2012 and concludes that most of the minted Bitcoins remain in dormant addresses which had never participated in any outgoing transactions, and that there is a huge number of tiny transactions which move only a small fraction of a single Bitcoin, but there are also hundreds of transactions which move more than 50,000 Bitcoins.

"Quantitative Analysis of the Full Bitcoin Transaction Graph" (http://eprint.iacr.org/2012/584.pdf)

Last but not least, in October 2012 the European Central Bank published a paper aimed at analysing virtual currencies, and concludes that they do not pose a risk to price stability, provided that money creation continues to stay at a low level. The paper also emphasizes the instability of the virtual currencies, their risks, and states that they could have a negative impact on the reputation of central banks.

"Virtual Currency Schemes" – European Central Bank
(http://www.ecb.europa.eu/pub/pdf/other/virtualcurrencyschemes201210en.pdf)

Section 3: Data

For the first two parts of the paper, the data consists of 26 time series of different measures related to the virtual currency. The data has 916 observations. The period studied comprises daily data for the dates between 17-Aug-2010 and 19-Feb-2013. All the data has been obtained from the website http://www.blockchain.info. There exist some structural breaks that only affect a few variables, which have been studied in the first part. The variables analysed are the following:

Market Price ($/BTC): USD market price from Mt.Gox (main exchange).
Total Bitcoins In Circulation: Historical total number of Bitcoins that have been mined.
Market Capitalization: Total number of Bitcoins in circulation * the market price in USD.
Total Transaction Fees: Total BTC value of transaction fees miners earn per day.

- Transaction Fees in USD
- Network Deficit - Shows the difference between transaction fees and the miners' income from block rewards

Number Of Transactions: Total number of unique Bitcoin transactions per day.

- Total Number Of transactions

Number of transactions excluding popular addresses: Total number of unique Bitcoin transactions per day excluding those that involve any of the top 100 most popular addresses popular addresses.
Number of unique Bitcoin addresses used: Number of unique Bitcoin addresses used per day.
Number of transactions per block: Average number of transactions per block.
Total Output Volume: The total value of all transaction outputs per day. This includes coins that were returned to the sender as change.
Estimated Transaction Volume: Similar to the total output volume with the addition of an algorithm which attempts to remove change from the total value. This may be a more accurate reflection of the true transaction volume.
Estimated Transaction Volume: Similar to the total output volume with the addition of an algorithm which attempts to remove change from the total value. This may be a more accurate reflection of the true transaction volume.
USD Exchange Trade Volume: USD trade volume from the top exchanges.
Trade Volume / Transaction Volume Ratio: Ratio between BTC transaction volume and USD exchange volume.
Cost % of transaction volume: Miners revenue as as percentage of the transaction volume.
Cost Per Transaction: Miners revenue divided by the number of transactions.
Hash Rate: The estimated number of Giga hashes per second the Bitcoin network is performing.
Miners Revenue: Number of Bitcoins mined per day + transaction fees * market price.
Bitcoin Days Destroyed Cumulative: Bitcoin Days Destroyed is a measure of the transaction volume of Bitcoin that consists of the multiplication of the value of the transaction by the number of days that those Bitcoins had been idle.
Bitcoin Days Destroyed: None cumulative version of Bitcoin Days Destroyed.

- Bitcoin Days Destroyed (Filtered By Min Age 1 Week) - Filtered by minimum input age of 1 week
- Bitcoin Days Destroyed (Filtered By Min Age 1 Month) - Filtered by minimum input age of 1 month
- Bitcoin Days Destroyed (Filtered By Min Age 1 Year) - Filtered by minimum input age of 1 year

Blockchain Size: The total size of all block headers and transactions. Not including database indexes.

For the last part, the data is comprised of 656 observations – after taking the weekends out – of daily data for the exchange rates of Bitcoin (BTC), Brazil Real (BRL), Chinese Yuan (CNY), Euro (EUR), and Japan Yen (JPY) against the US Dollar (USD). The data has been obtained from the software package Thomson Reuters DataStream.

Section 4: Methodology

Part 1: Time series analysis

For the first part, in which the 26 variables will be analysed, the following methodology is going to be applied for each of the variables:

1. Plot time series
2. Plot ACF/PACF
3. If stationary, show the performed Augmented Dickey-Fuller test. If not, check for the existence of structural breaks (confirm with a Chow test) and/or take differences of the data until stationary. If there is a significant trend, remove it
4. Analyse the stationary series
    - **a.** Summary Statistics
    - **b.** Plot histogram / fit the Normal distribution
    - **c.** Perform tests to compare stationary data to the normal distribution
        - i. Normal, Skewness and Excess kurtosis tests
        - ii. Jarque-Bera test

Part 2: Looking for pair-wise cointegration

Once the time series have been analysed, it would be interesting to see if there is any pair of variables that is cointegrated. The methodology to be followed is the following:

1. Decide on what pairs of variables cointegration will be studied. The variables within a pair will not represent the same measurement with different units or filter applied.
2. Perform the Engle-Granger Method –, as the purpose of this research is only determining the existence of cointegration

- Make sure that all the variables are I(1)
- Estimate the cointegrating regression using OLS ($y_t = \Phi + \gamma \cdot x_t + u_t$) and save the residuals of the cointegrating regression, $\hat{u}_t$
- Test these residuals to ensure that they are I (0)

3. Show Adjusted R-Squared of the OLS regression

Part 3: Multivariate analysis of a basket of currencies

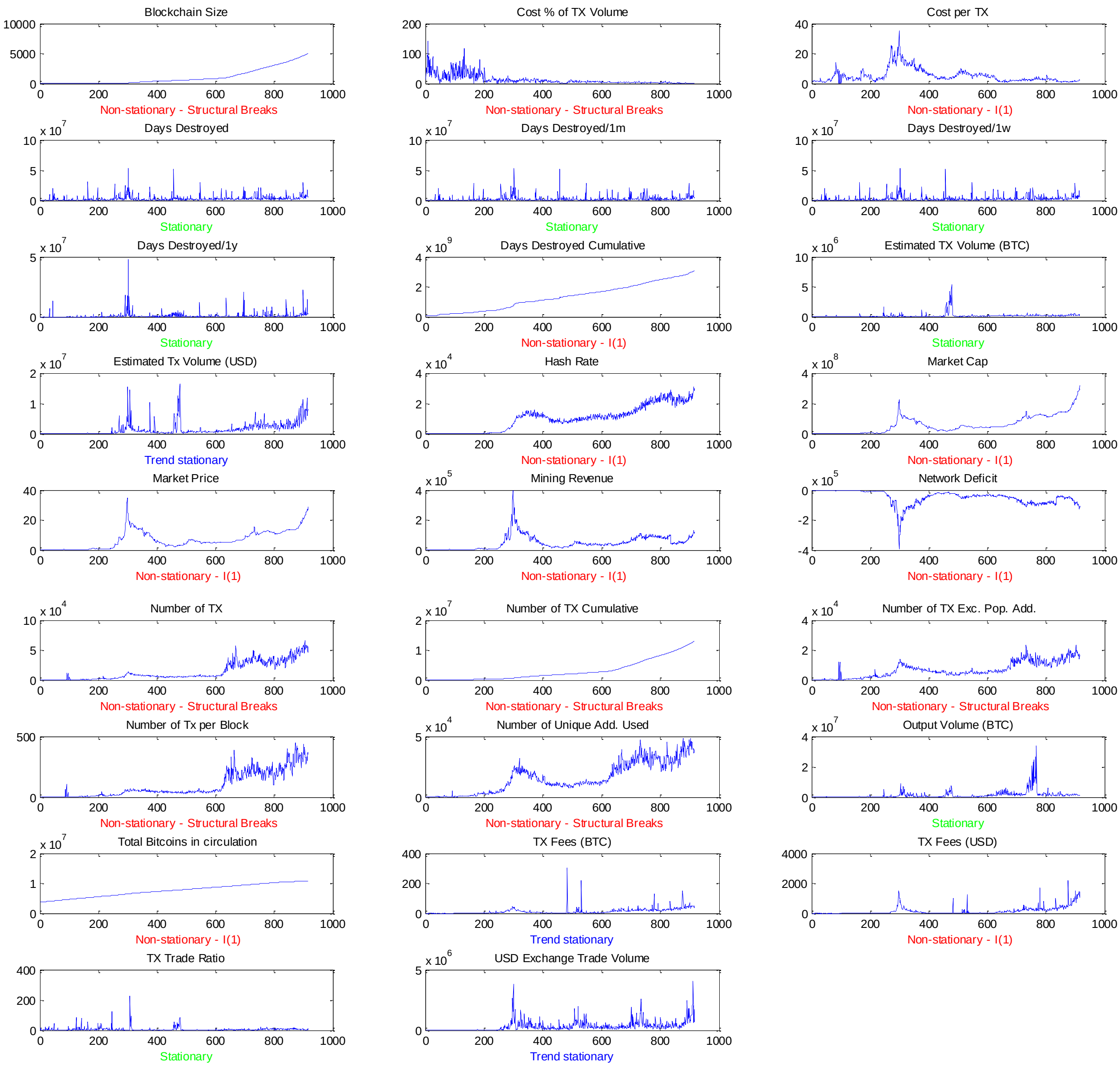
The methodology followed in this case is the following:

1. Plot raw data to see how it behaves
2. Transform the data if necessary
3. Check whether or not data is stationary by analysing ACF/PACF plots and by performing an Augmented Dickey-Fuller test
4. If data is not stationary, take the first difference (y_t- y_{t-1}) and go back to Step s3.
5. Determine the most appropriate lag length to use in a VAR model by using different information criteria, such as MAIC, MSBIC and MHQIC
6. Create the appropriate VAR model by determining, for each currency i=1:5
   $y_{i,t} = \beta_{i0} + \beta_{i1}y_{1,t-1} + \beta_{i2}y_{2,t-1} + \beta_{i3}y_{3,t-1} + \beta_{i4}y_{4,t-1} + \beta_{i5}y_{5,t-1})$
   By determining its coefficients and the corresponding standard errors
7. Perform a Granger causality test

Section 5: Results

Part 1: Time series analysis

Due to the large size of the analysis, it has been moved to the appendix and a short summary is shown in this page, showing a plot of the variables and a statement regarding whether or not the variables are stationary. If not, the order of integration is shown, unless there is a structural break, and in that case it will be indicated. For more information, refer to the appendix, where a deeper analysis is held for each of the time series.



Part 2: Looking for pair-wise cointegration

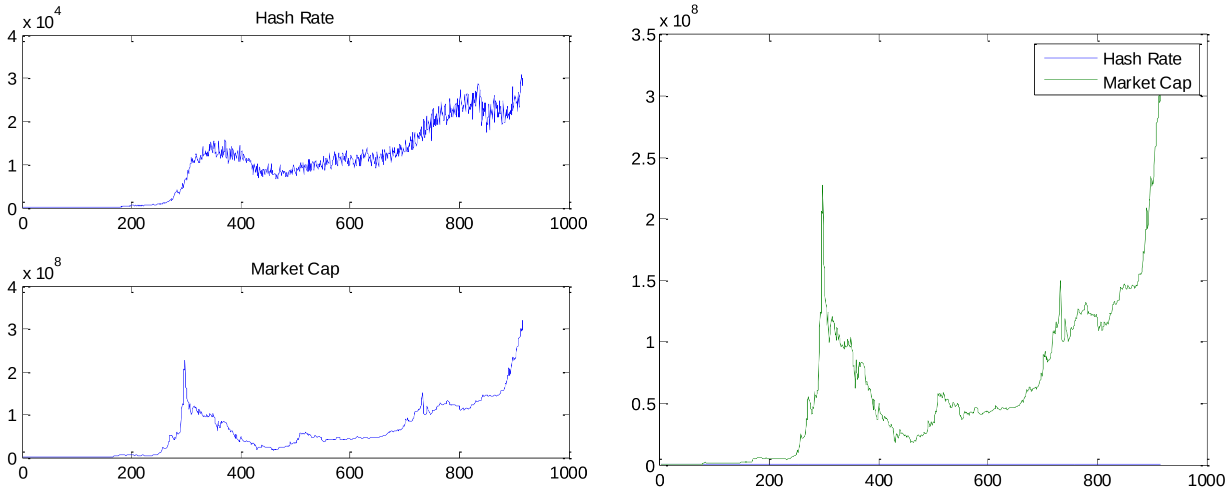
Pairs of time series to test for cointegration:

- Hash Rate – Mining Revenue
- Hash Rate – Market Cap
- Hash Rate – Market Price
- Hash Rate – Number of Unique Addresses Used
- Market Price – Number of Unique Addresses Used
- Mining Revenue – Market Price
- Mining Revenue – Network Deficit
- Number of TX – Number of TX per Block
- Number of TX – Number of Unique Addresses Used
- Number of TX Excluding Popular Addresses – Number of Unique Addresses Used

Note: Cointegrating pairs will be shown first. Non-cointegrating pairs have been moved to the end of this part.

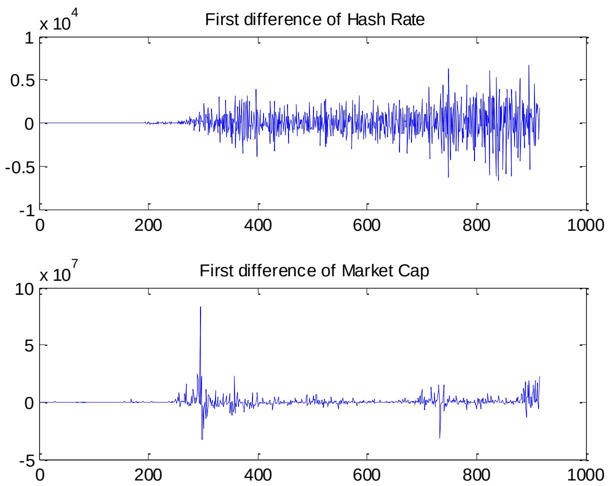
Cointegrating pairs

Pair: Hash Rate – Market Cap



	Augmented Dickey-Fuller Test	
	Hash Rate	**Market Cap**
Test stat	-1.5554	0.7913
p-value	0.8104	1.0000
Unit root?	Yes	Yes

1. Make sure both variables are I(1)

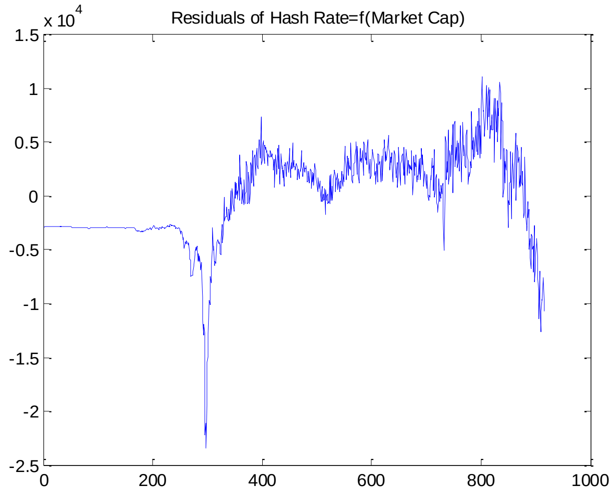


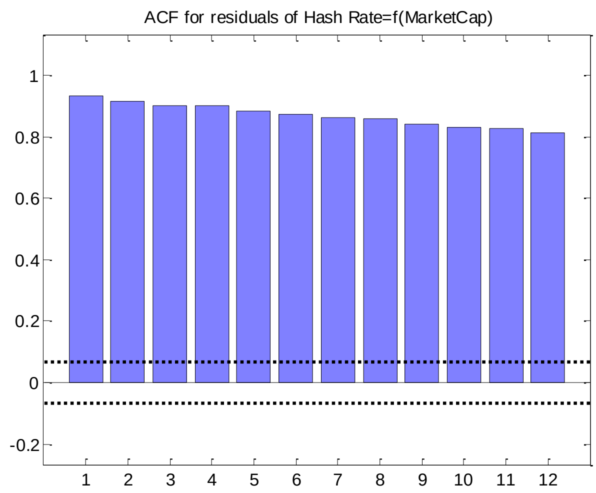
	Augmented Dickey-Fuller Test	
	Δ Hash Rate	**Δ Market Cap**
Test stat	-5.9156	-4.6097
p-value	0.0000	0.0000
Unit root?	No	No

2. Estimate cointegrating regression using OLS

$$y_t = \Phi + \gamma \cdot x_t + u_t$$

Estimated Parameters	Confidence Intervals
Φ = 2.8498e+03	[2.4561e+03 3.2435e+03]
γ = 1.1879e-04	[1.1410e-04 1.2348e-04]





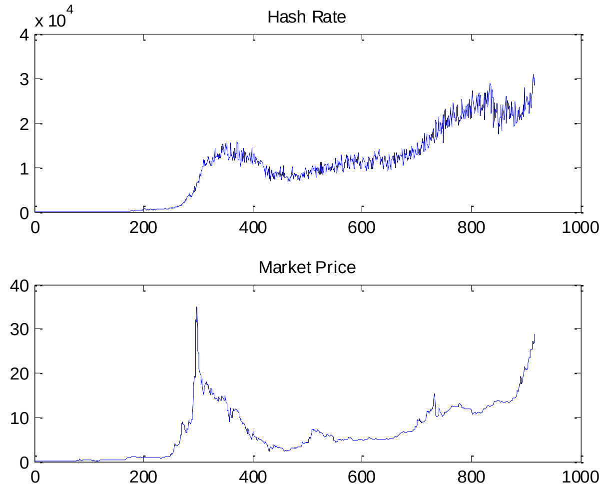
3. Check if the residuals are I(0)

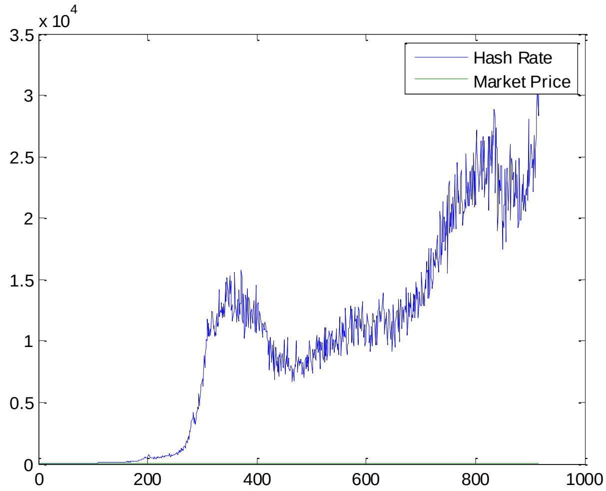
	Augmented Dickey-Fuller Test
	Residuals
Test stat	-2.2347
p-value	0.0252
Unit root?	No

The variables are cointegrated.

The R-Squared of the regression is 0.7297; therefore it can be said that the Market Cap is a good predictor for the Hash Rate. This would mean that the higher the Market Cap, the higher the Hash Rate is. The cause could be the popularity and penetration of Bitcoin, which affects both variables.

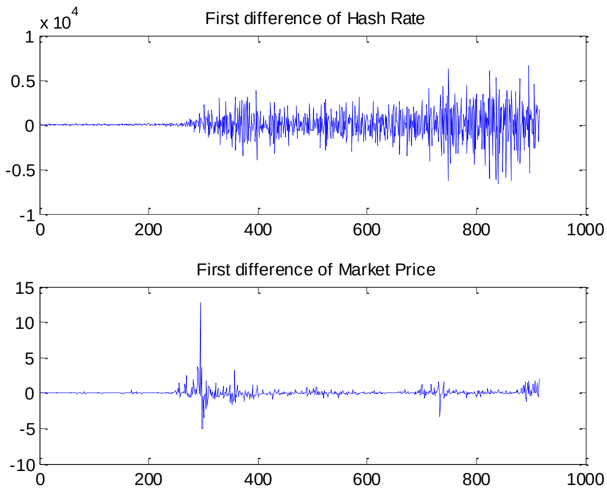
Pair: Hash Rate – Market Price





	Augmented Dickey-Fuller Test	
	Hash Rate	**Market Price**
Test stat	-1.5554	-1.1124
p-value	0.8104	0.9255
Unit root?	Yes	Yes

1. Make sure both variables are I(1)

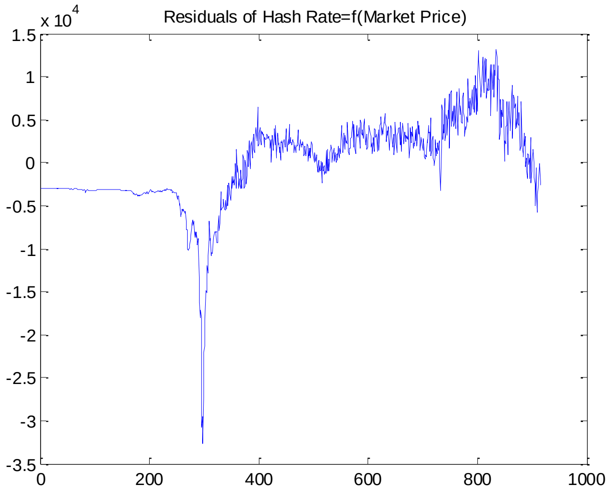


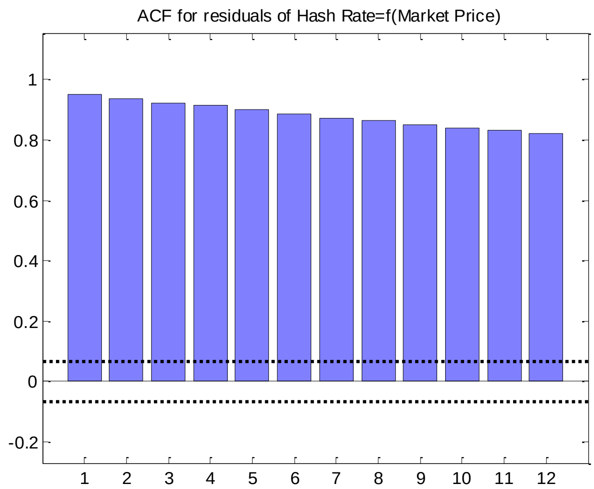
	Augmented Dickey-Fuller Test	
	Δ Hash Rate	**Δ Market Price**
Test stat	-5.9156	-10.1445
p-value	0.0000	0.0000
Unit root?	No	No

2. Estimate cointegrating regression using OLS

$$y_t = \Phi + \gamma \cdot x_t + u_t$$

Estimated Parameters	Confidence Intervals
Φ = 2.9613e+03	[2.4521e+03 3.4705e+03]
γ = 1.0323e+03	[976.4586 1.0882e+03]





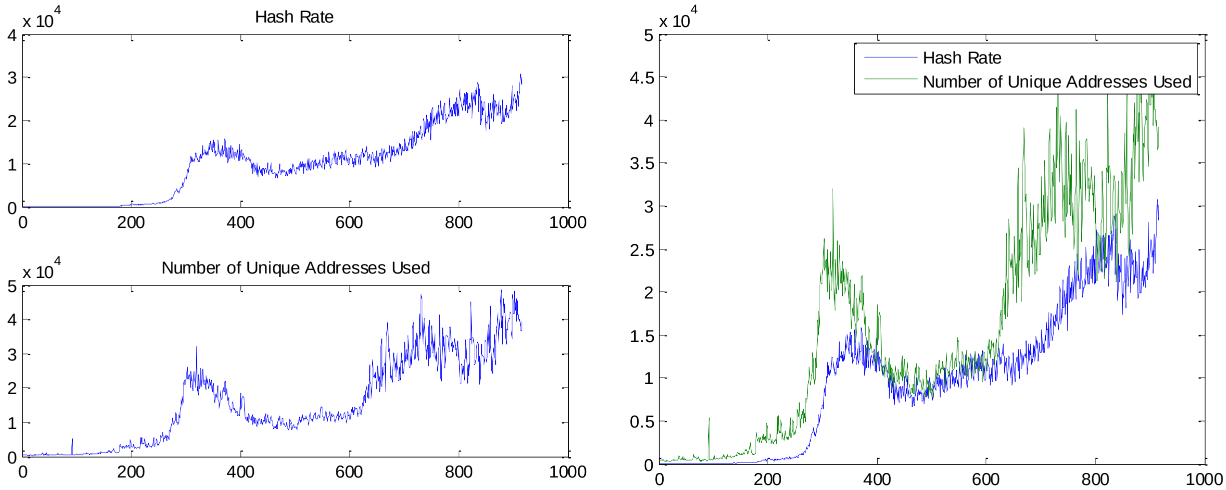
3. Check if the residuals are I(0)

	Augmented Dickey-Fuller Test
	Residuals
Test stat	-2.5004
p-value	0.0127
Unit root?	No

The variables are cointegrated.

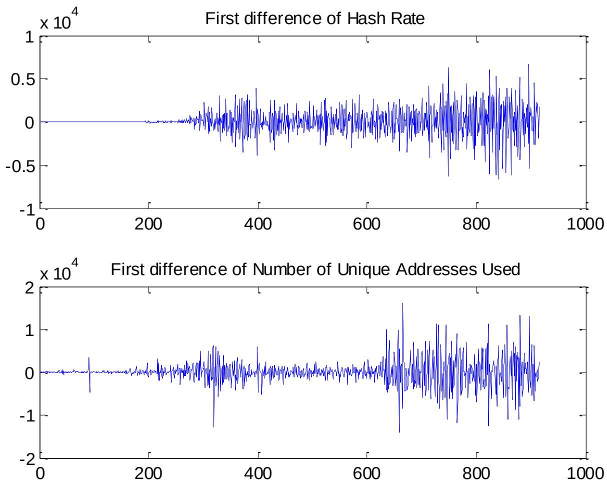
The R-Squared of the regression is 0.5899; therefore it can be said that the regression explains more than half of the total variance. However, the Market Cap seems to be a better predictor.

Pair: Hash Rate – Number of Unique Addresses Used



	Augmented Dickey-Fuller Test	
	Hash Rate	**Number of Unique Addresses Used**
Test stat	-1.5554	-2.0252
p-value	0.8104	0.5884
Unit root?	Yes	Yes

1. Make sure both variables are I(1)

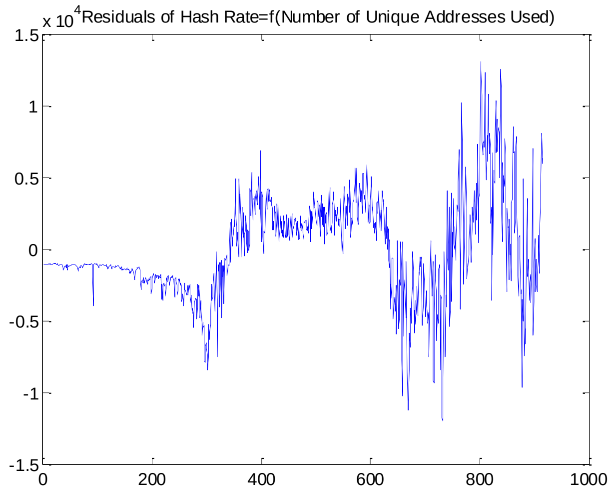


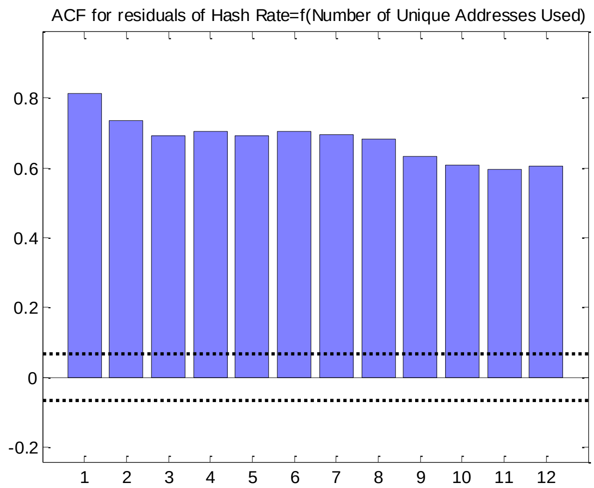
	Augmented Dickey-Fuller Test	
	Δ Hash Rate	**Δ Number of Unique Addresses Used**
Test stat	-5.9156	-8.2960
p-value	0.0000	0.0000
Unit root?	No	No

2. Estimate cointegrating regression using OLS

$$y_t = \Phi + \gamma \cdot x_t + u_t$$

Estimated Parameters	Confidence Intervals
Φ = 871.4100	[507.9673 1.2349e+03]
γ = 0.5841	[0.5660 0.6022]





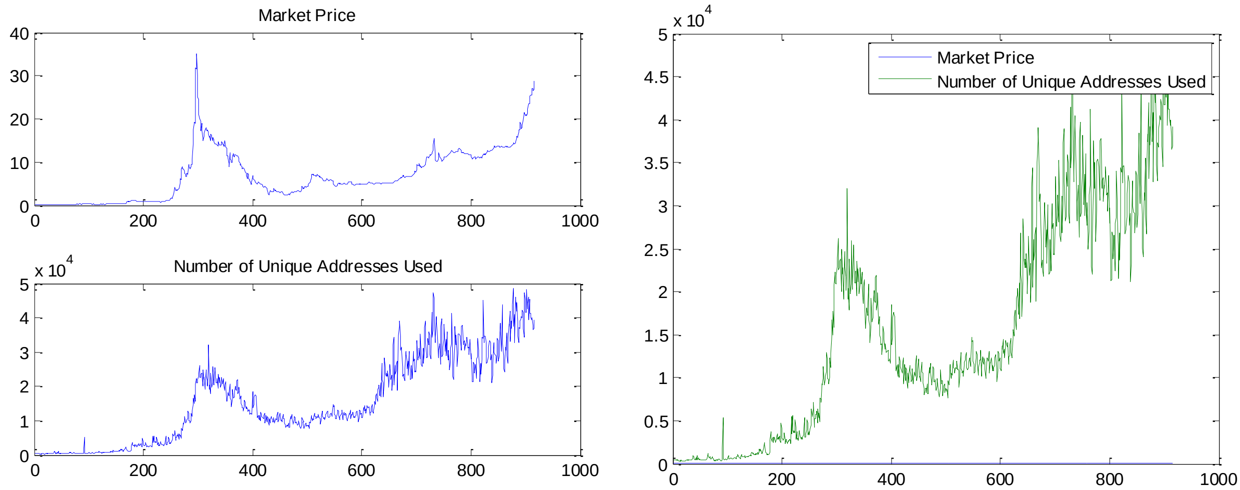
3. Check if the residuals are I(0)

	Augmented Dickey-Fuller Test
	Residuals
Test stat	-3.0218
p-value	0.0032
Unit root?	No

The variables are cointegrated.

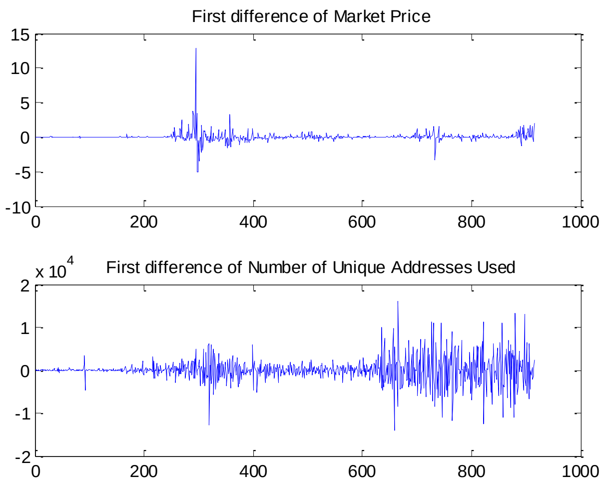
The R-Squared of the regression is 0.8139; therefore one of the variables explains a large amount of variance of the other one. This is interesting because the variable Number of Unique Addresses Used is a good proxy for the adoption of Bitcoin, so the higher the adoption, the higher the Hash Rate is. This can mean that either the number of miners increase or that advancement in technology increases the hash rate that each individual miner provides to the network.

Pair: Market Price – Number of Unique Addresses Used



	Augmented Dickey-Fuller Test	
	Market Price	**Number of Unique Addresses Used**
Test stat	-1.1124	-2.0252
p-value	0.9255	0.5884
Unit root?	Yes	Yes

1. Make sure both variables are I(1)

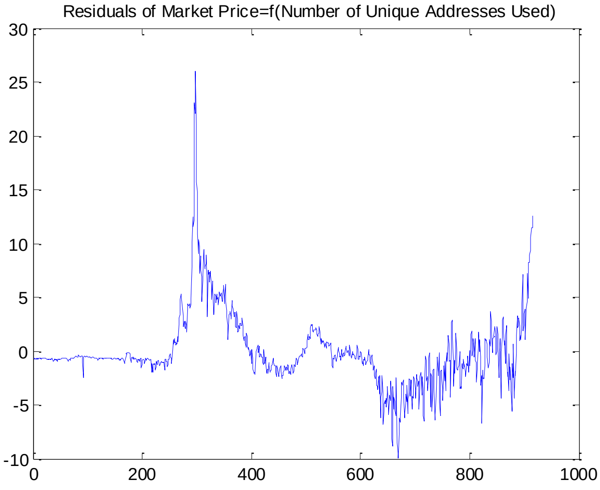


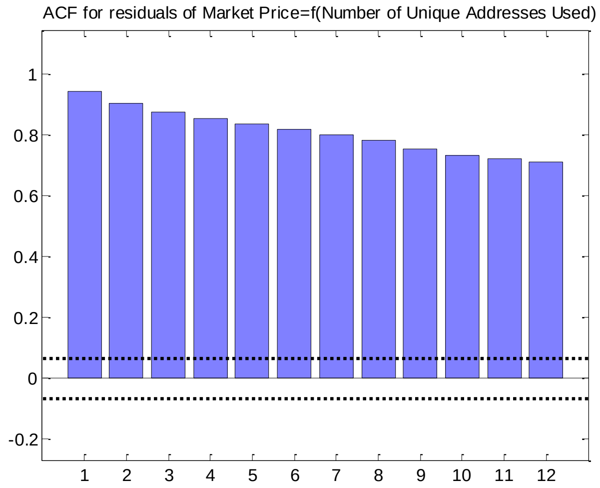
	Augmented Dickey-Fuller Test	
	Δ Market Price	**Δ Number of Unique Addresses Used**
Test stat	-10.1445	-8.2960
p-value	0.0000	0.0000
Unit root?	No	No

2. Estimate cointegrating regression using OLS

$$y_t = \Phi + \gamma \cdot x_t + u_t$$

Estimated Parameters	Confidence Intervals
Φ = 0.6173	[0.2622 0.9724]
γ = 0.0004	[0.0004 0.0004]





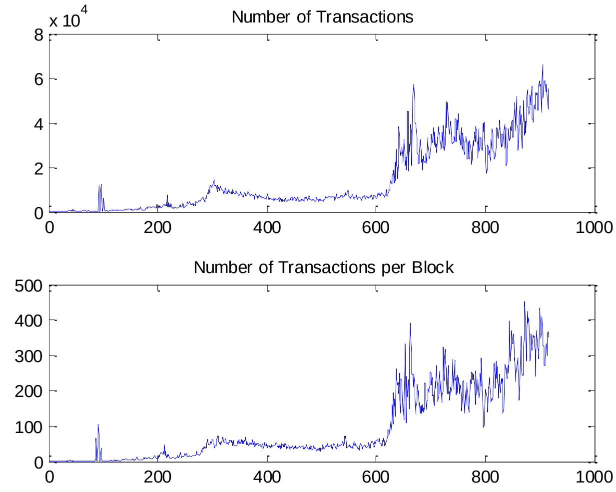
3. Check if the residuals are I(0)

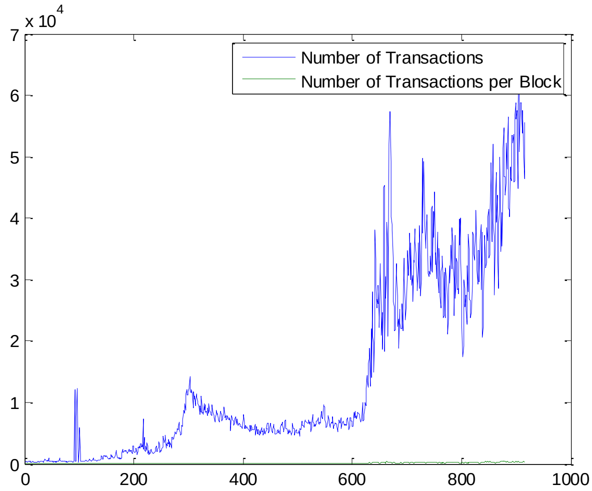
	Augmented Dickey-Fuller Test
	Residuals
Test stat	-2.6988
p-value	0.0072
Unit root?	No

The variables are cointegrated.

The R-Squared of the regression is 0.6791; therefore one of the variables explains a good amount of variance of the other one. Again, this is interesting for the same reason stated in the analysis of the previous variable: the variable Unique Addresses Used is a proxy for the adoption of Bitcoin, and this regression indicates that the higher the Number of Unique Addresses used is, the higher the Hash Rate (or the other way round).

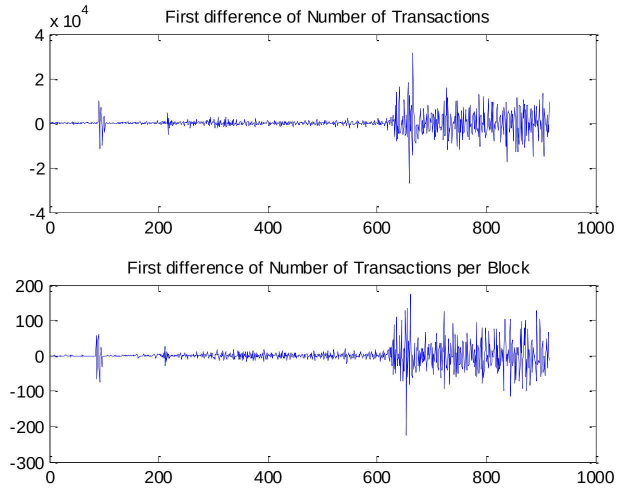
Pair: Number of TX – Number of TX per Block





	Augmented Dickey-Fuller Test	
	Number of Transactions	**Number of Transactions per Block**
Test stat	-1.3676	-1.6654
p-value	0.8706	0.7671
Unit root?	Yes	Yes

1. Make sure both variables are I(1)

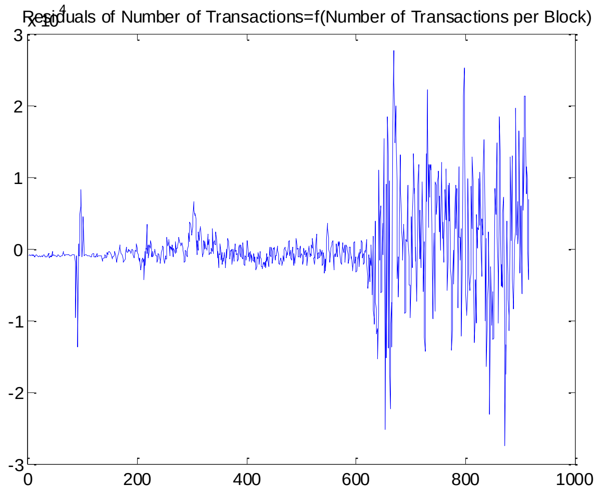


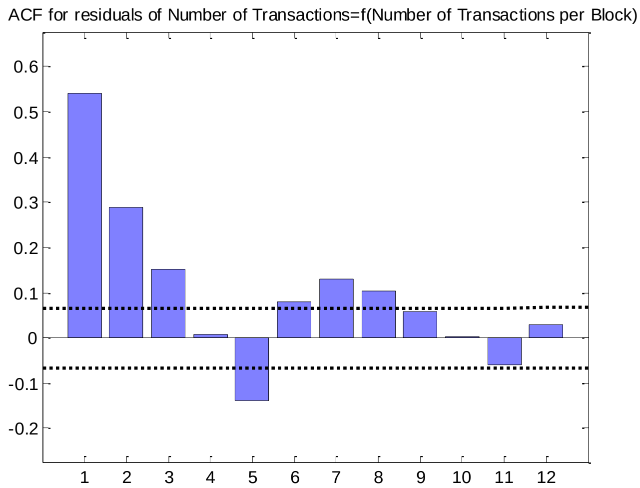
	Augmented Dickey-Fuller Test	
	Δ Number of Transactions	**Δ Number of Transactions per Block**
Test stat	-8.4387	-9.4404
p-value	0.0000	0.0000
Unit root?	No	No

2. Estimate cointegrating regression using OLS

$$y_t = \Phi + \gamma \cdot x_t + u_t$$

Estimated Parameters	Confidence Intervals
Φ = 1.1385e+03	[659.2621 1.6178e+03]
γ = 135.0692	[131.7040 138.4344]





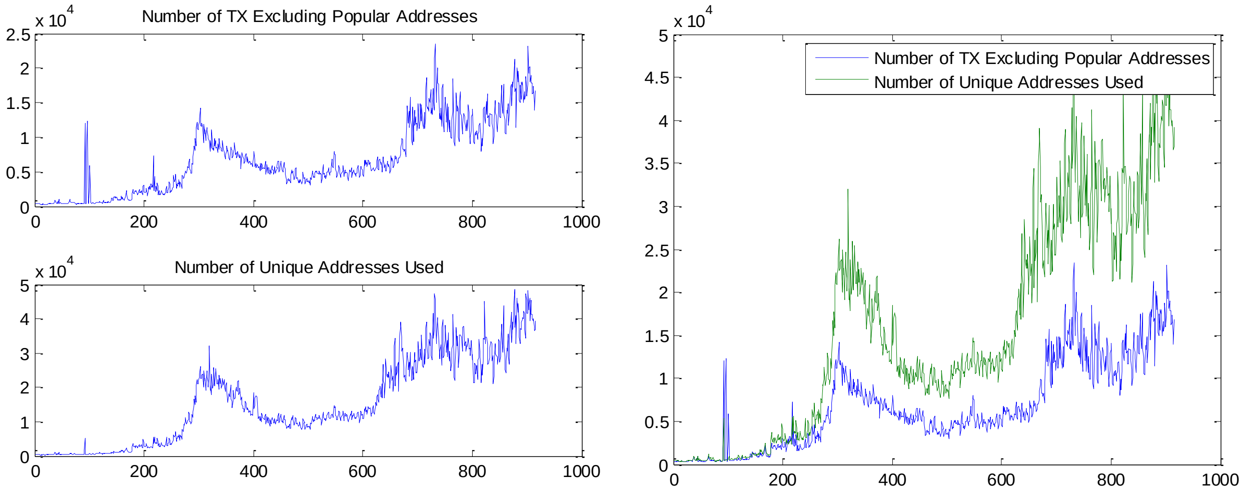
3. Check if the residuals are I(0)

	Augmented Dickey-Fuller Test
	Residuals
Test stat	-5.9264
p-value	0.0000
Unit root?	No

The variables are cointegrated.

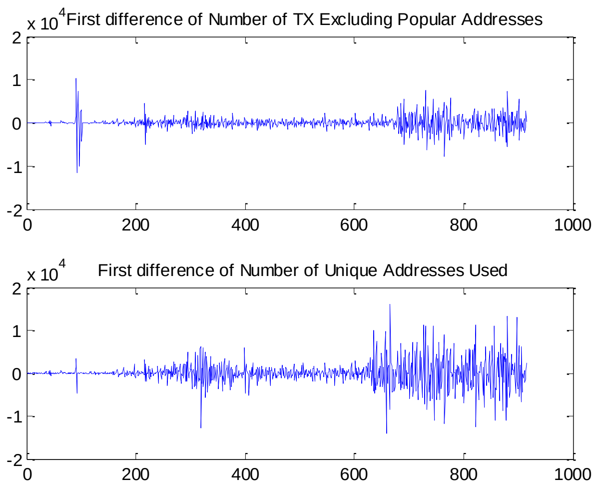
The R-Squared of the regression is 0.8716. This high value is realistic, as the larger the number of transactions, the higher the amount of transactions that should be included in each block, as blocks are generated at an approximately constant interval.

Pair: Number of TX Excluding Popular Addresses – Number of Unique Addresses Used



	Augmented Dickey-Fuller Test	
	Number of Transactions Excluding Pop. Addresses	**Number of Unique Addresses Used**
Test stat	-2.0594	-2.0252
p-value	0.5695	0.5884
Unit root?	Yes	Yes

4. Make sure both variables are I(1)

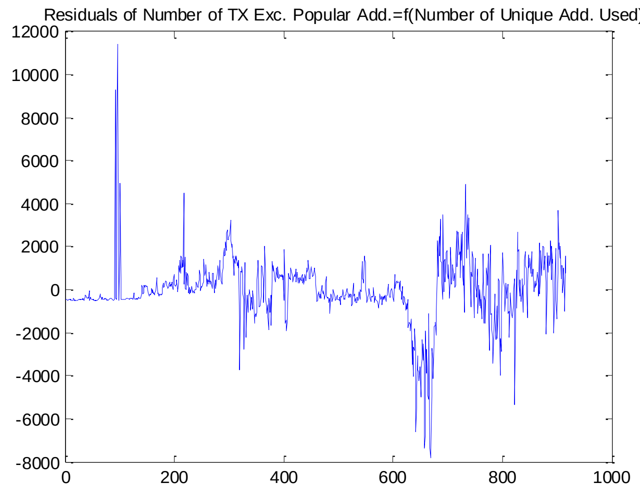


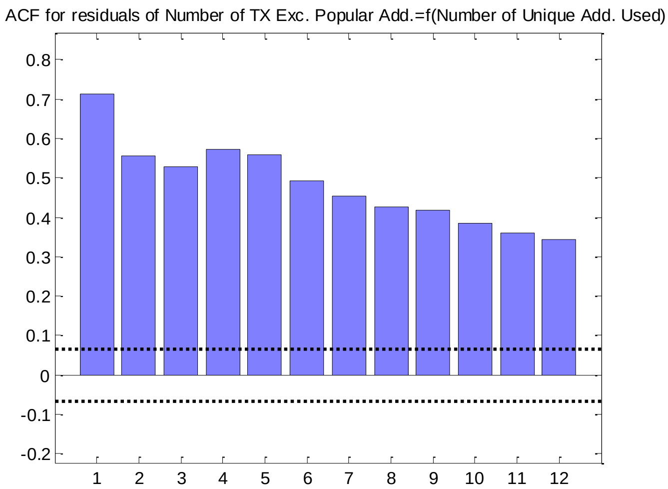
	Augmented Dickey-Fuller Test	
	Δ Number of Transactions Excluding Pop. Addresses	**Δ Number of Unique Addresses Used**
Test stat	-8.6838	-8.2960
p-value	0.0000	0.0000
Unit root?	No	No

5. Estimate cointegrating regression using OLS

$$y_t = \Phi + \gamma \cdot x_t + u_t$$

Estimated Parameters	Confidence Intervals
Φ = 671.2339	[515.7761 826.6918]
γ = 0.3913	[0.3835 0.3990]





6. Check if the residuals are I(0)

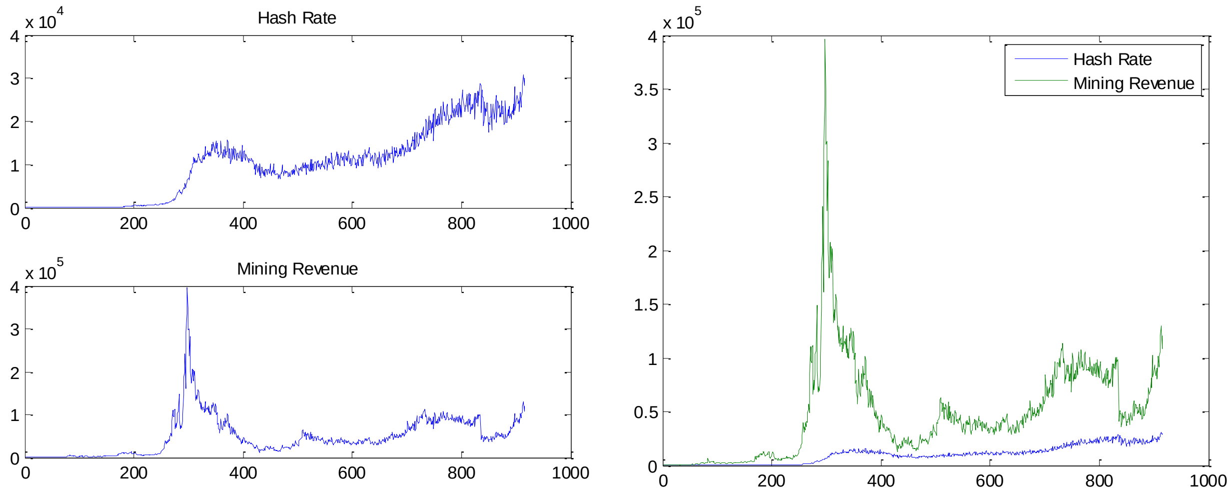
	Augmented Dickey-Fuller Test
	Residuals
Test stat	-5.3198
p-value	0.0000
Unit root?	No

The variables are cointegrated.

The R-Squared of the regression is 0.9147. In this case the variables are cointegrated (they "move" together with time) and the percentage of explained variance is higher than in the previous analysis (in which Number of Transactions was used, see Appendix). There was no cointegration between the Number of Transactions and the Number of Unique Addresses Used. However, in this case there is, and this could be due to the fact that a high amounts of transactions are held by a few Bitcoin addresses (e.g. Satoshi Dice is an online casino that currently accounts for over 50% of the daily transactions).

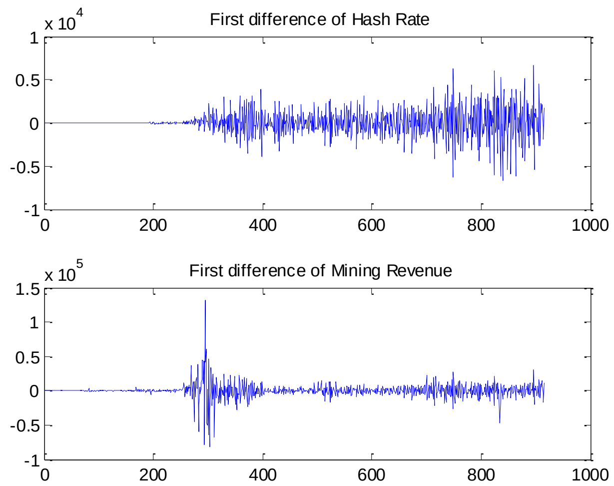
Non-cointegrating pairs

Pair: Hash Rate – Mining Revenue



	Augmented Dickey-Fuller Test	
	Hash Rate	**Mining Revenue**
Test stat	-1.5554	-2.6335
p-value	0.8104	0.5147
Unit root?	Yes	Yes

1. Make sure both variables are I(1)

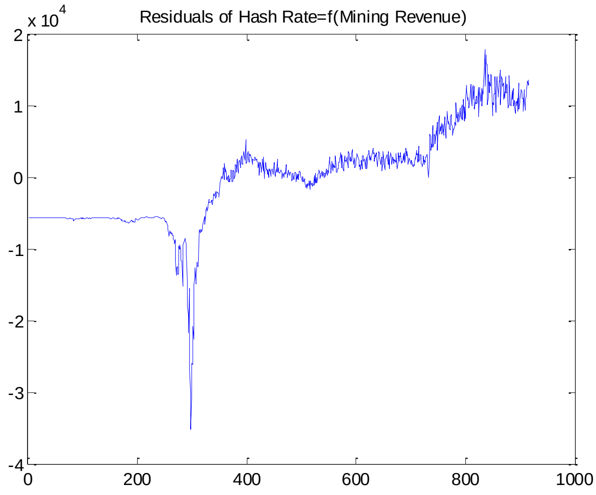


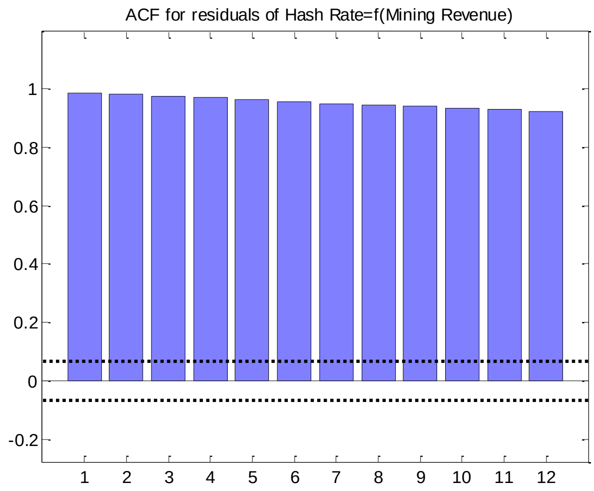
	Augmented Dickey-Fuller Test	
	Δ Hash Rate	**Δ Mining Revenue**
Test stat	-5.9156	-9.8992
p-value	0.0000	0.0000
Unit root?	No	No

2. Estimate cointegrating regression using OLS

$$y_t = \Phi + \gamma \cdot x_t + u_t$$

Estimated Parameters	Confidence Intervals
Φ = 5.6116e+03	[4.9681e+03 6.2551e+03]
γ = 0.0909	[0.0813 0.1004]





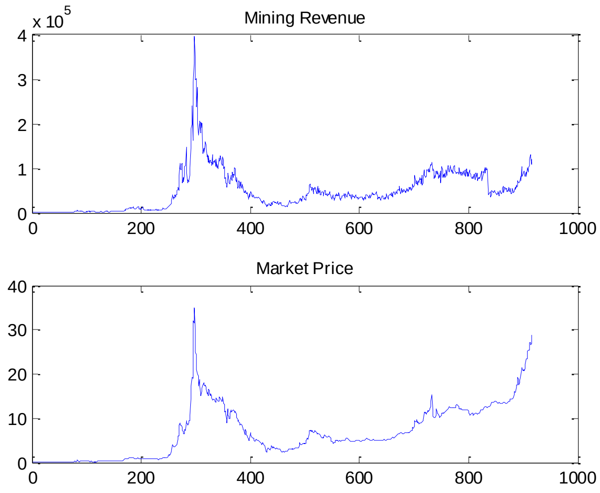
3. Check if the residuals are I(0)

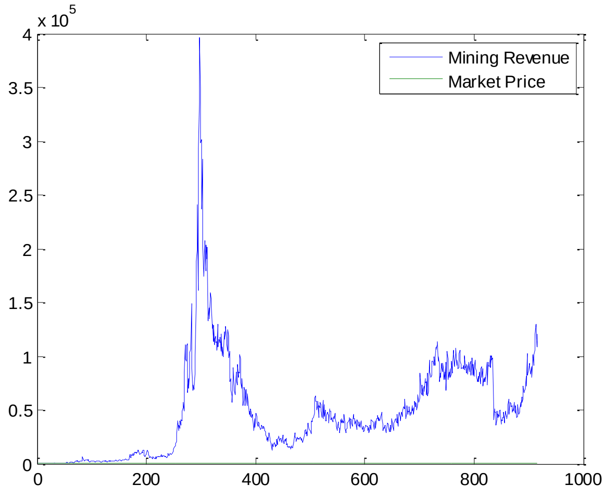
	Augmented Dickey-Fuller Test
	Residuals
Test stat	-1.5444
p-value	0.1154
Unit root?	Yes

The variables are not cointegrated.

The R-Squared of the regression is 0.2764; therefore one variable does not explain much of the variance of the other one.

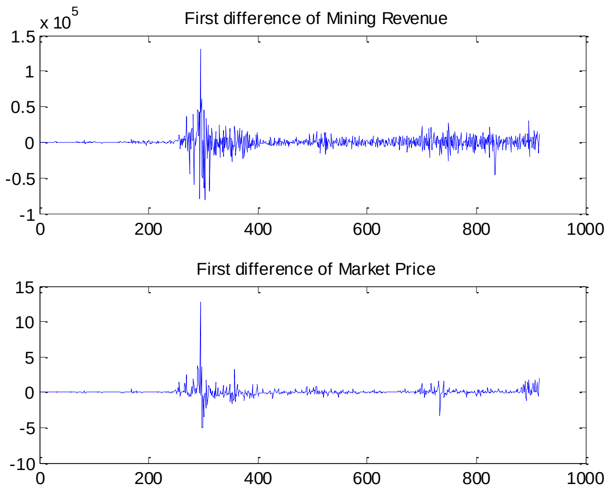
Pair: Mining Revenue – Market Price





	Augmented Dickey-Fuller Test	
	Mining Revenue	**Market Price**
Test stat	-2.6335	-1.1124
p-value	0.2673	0.9255
Unit root?	Yes	Yes

1. Make sure both variables are I(1)

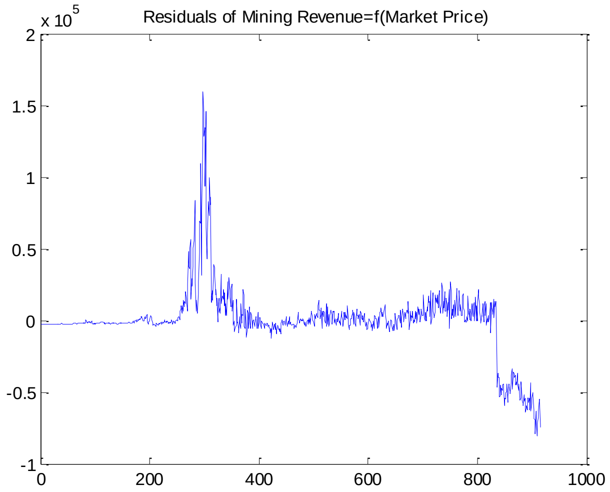


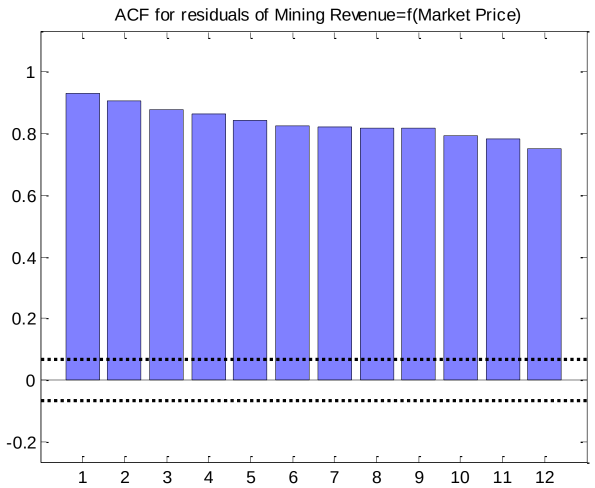
	Augmented Dickey-Fuller Test	
	Δ Mining Revenue	**Δ Market Price**
Test stat	-9.8992	-10.1445
p-value	0.0000	0.0000
Unit root?	No	No

2. Estimate cointegrating regression using OLS

$$y_t = \Phi + \gamma \cdot x_t + u_t$$

Estimated Parameters	Confidence Intervals
Φ = 2.6088e+03	[279.7147 4.9379e+03]
γ = 6.7065e+03	[6.4509e+03 6.9620e+03]





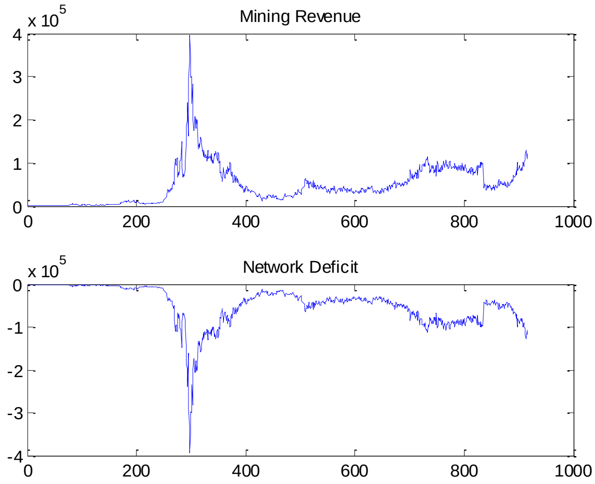
3. Check if the residuals are I(0)

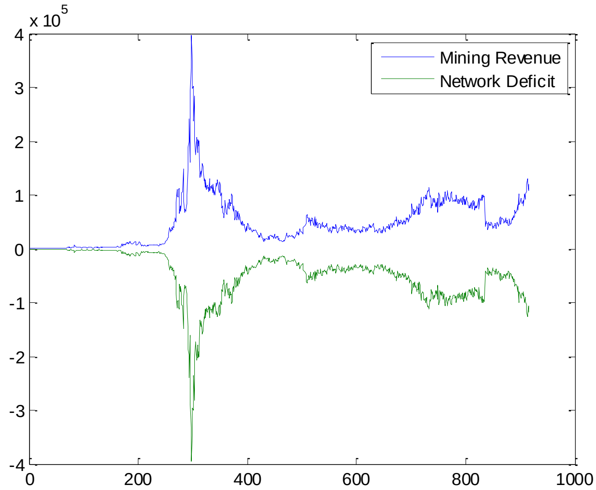
	Augmented Dickey-Fuller Test
	Residuals
Test stat	-1.3960
p-value	0.1517
Unit root?	Yes

The variables are not cointegrated.

The R-Squared of the regression is 0.7437; therefore one of the variables explains a good amount of variance of the other one. The residuals seem stationary at one point but they suddenly collapse around the observation 820. That is exactly the moment in which the reward given to miners halved (by the end of 2012).

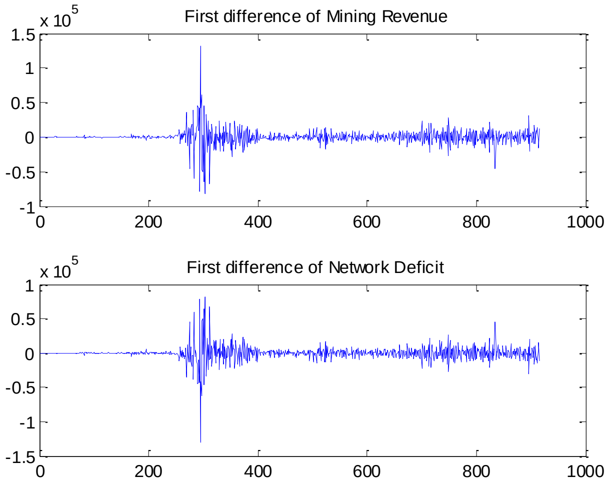
Pair: Mining Revenue – Network Deficit





	Augmented Dickey-Fuller Test	
	Mining Revenue	**Network Deficit**
Test stat	-2.6335	-2.6372
p-value	0.2673	0.2657
Unit root?	Yes	Yes

1. Make sure both variables are I(1)

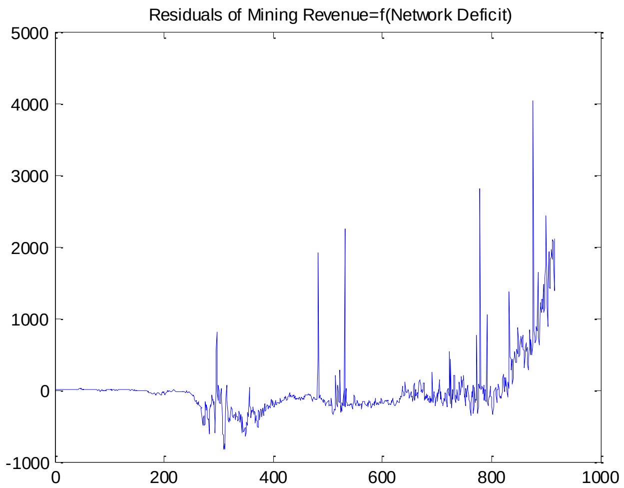


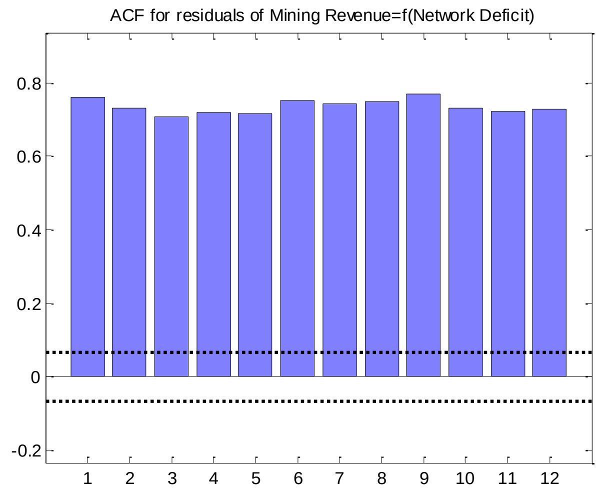
	Augmented Dickey-Fuller Test	
	Δ Mining Revenue	**Δ Network Deficit**
Test stat	-9.8992	-9.8879
p-value	0.0000	0.0000
Unit root?	No	No

2. Estimate cointegrating regression using OLS

$$y_t = \Phi + \gamma \cdot x_t + u_t$$

Estimated Parameters	Confidence Intervals
Φ = -18.1786	[-57.0242 20.6671]
γ = -1.0068	[-1.0073 -1.0062]



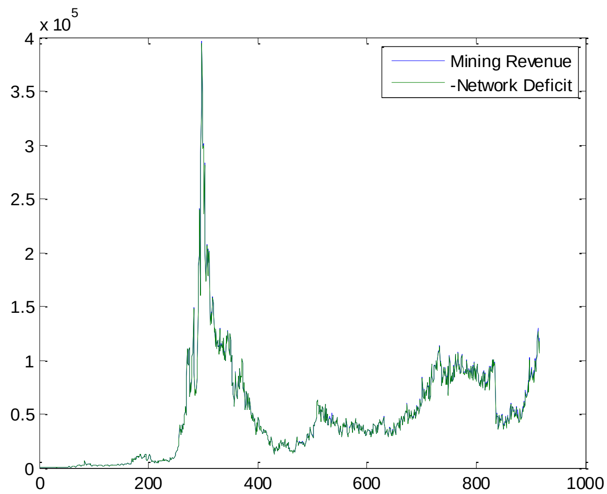


3. Check if the residuals are I(0)

	Augmented Dickey-Fuller Test
	Residuals
Test stat	2.1580
p-value	0.9925
Unit root?	Yes

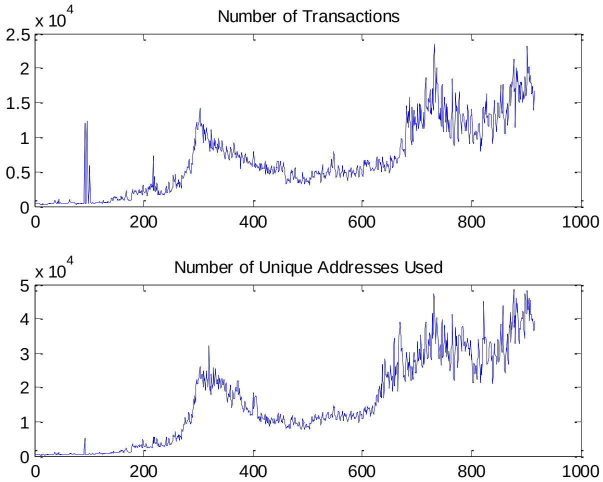
The variables are not cointegrated.

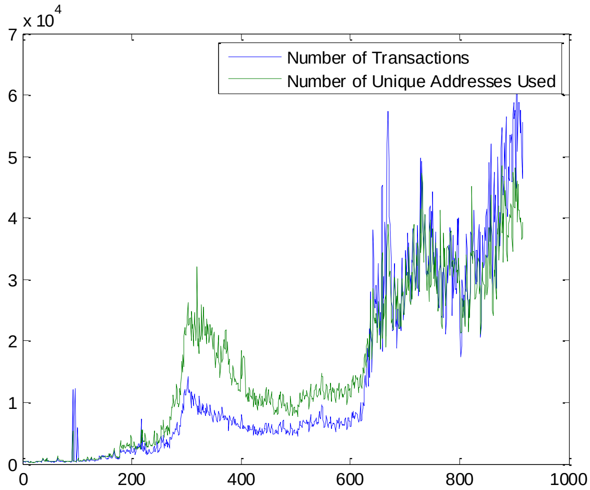
The R-Squared of the regression is 0.9999, extremely high. This is because one variable is almost the other one flipped around the x-axis:



They seem to overlap. Then, why are the residuals non-stationary? Because as time passes, the difference between one variable and (the negative) of the other increases. This is, in fact, the expected behaviour. As explained in Part 1, miners' revenue will increasingly come from transaction fees (instead of block rewards), thus the network deficit increases faster than mining revenue.

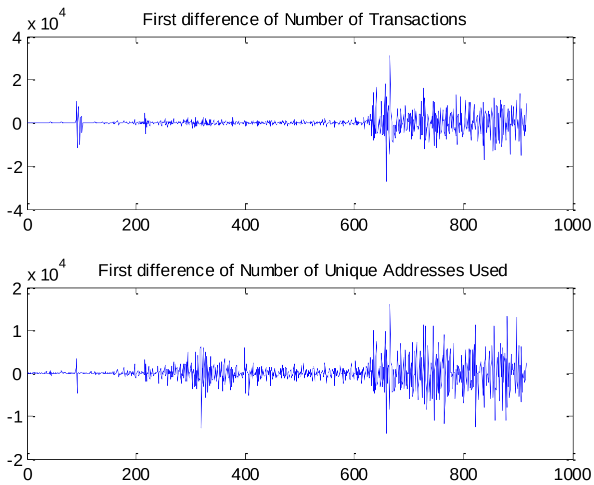
Pair: Number of TX – Number of Unique Addresses Used





	Augmented Dickey-Fuller Test	
	Number of Transactions	Number of Unique Addresses Used
Test stat	-1.3676	-2.0252
p-value	0.8706	0.5884
Unit root?	Yes	Yes

1. Make sure both variables are I(1)

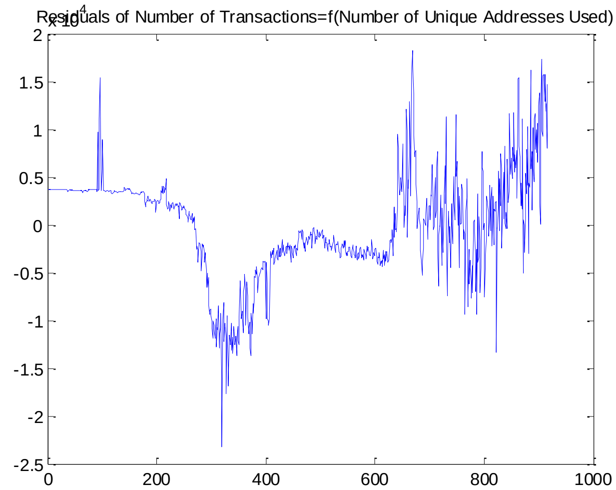


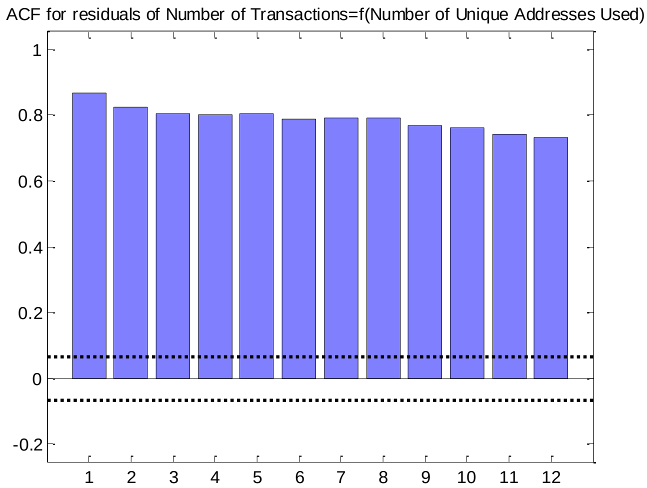
	Augmented Dickey-Fuller Test	
	Δ Number of Transactions	Δ Number of Unique Addresses Used
Test stat	-8.4387	-8.2960
p-value	0.0000	0.0000
Unit root?	No	No

2. Estimate cointegrating regression using OLS

$$y_t = \Phi + \gamma \cdot x_t + u_t$$

Estimated Parameters	Confidence Intervals
Φ = -3.7702e+03	[-4.3398e+03 -3.2005e+03]
γ = 1.1383	[1.1099 1.1667]





3. Check if the residuals are I(0)

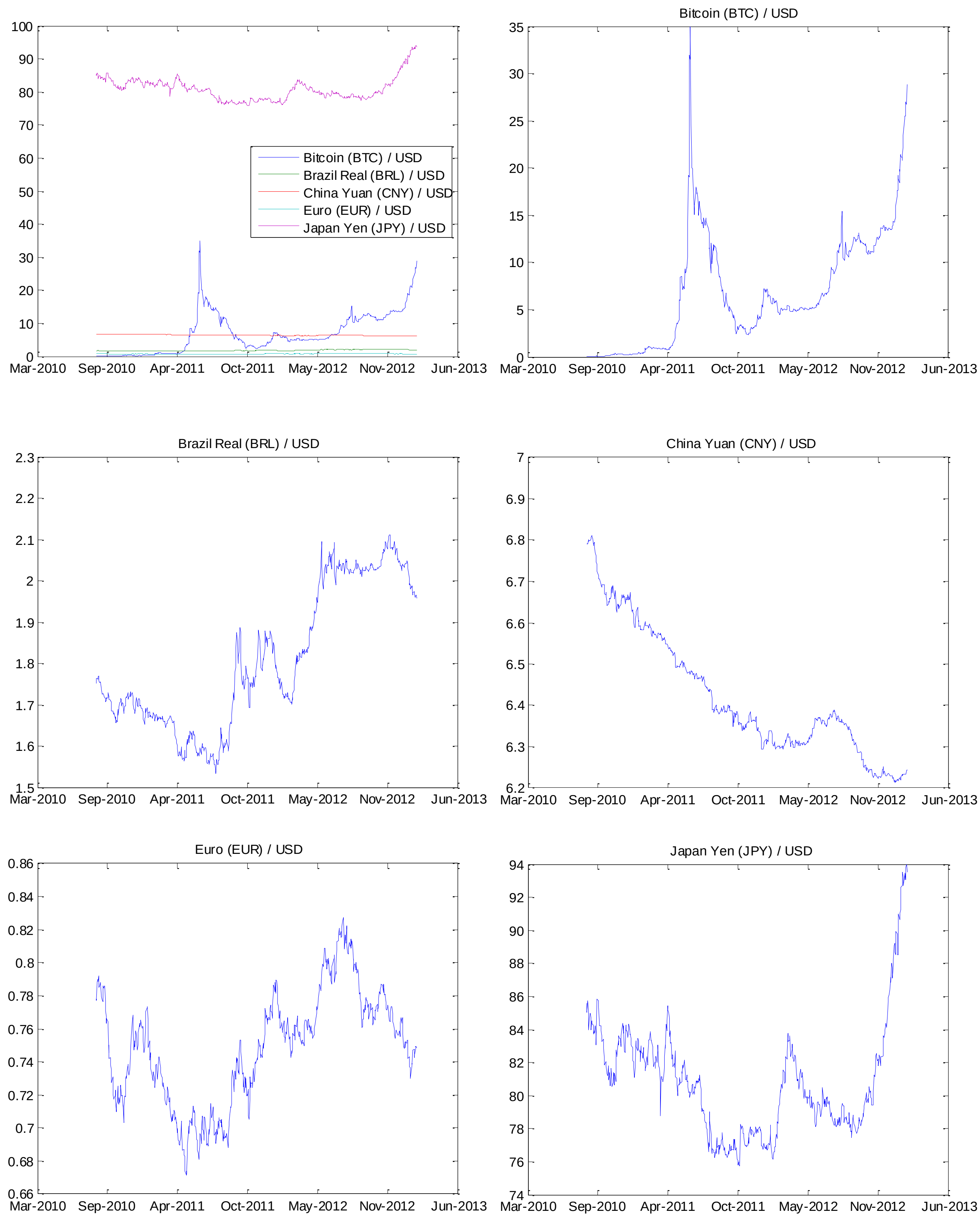
	Augmented Dickey-Fuller Test
	Residuals
Test stat	-1.4962
p-value	0.1264
Unit root?	Yes

The variables are not cointegrated.

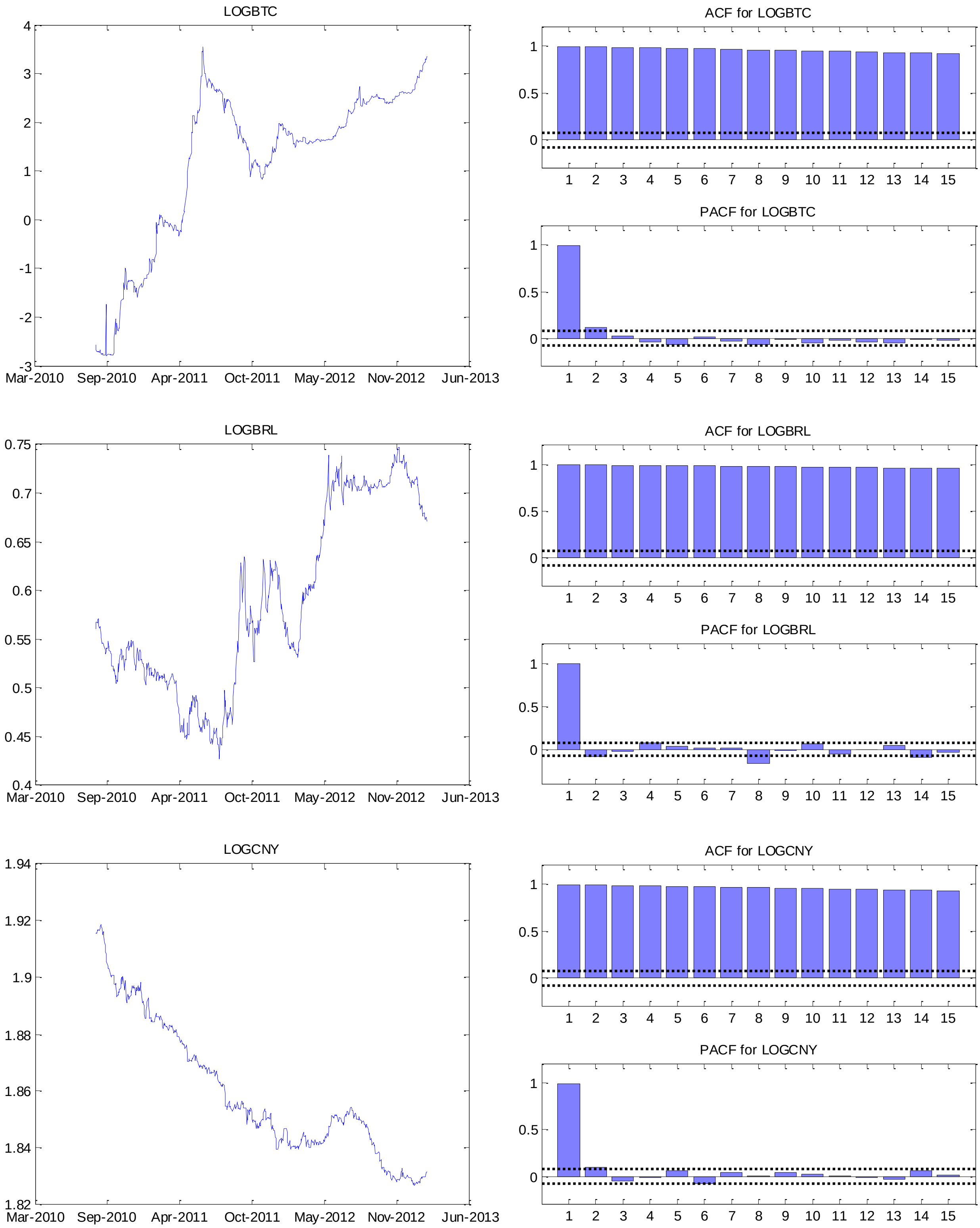
The R-Squared of the regression is 0.8712. This value should be high (as it is) due to the fact that it is sensible to imply that the higher the number of unique addresses used, the higher the number of users inside the bitcoin network, and the higher the number of transactions will be. However, it could be that a high amount of transactions are held by the same Bitcoin address (e.g. Satoshi Dice is an online casino that currently accounts for over 50% of the daily transactions).

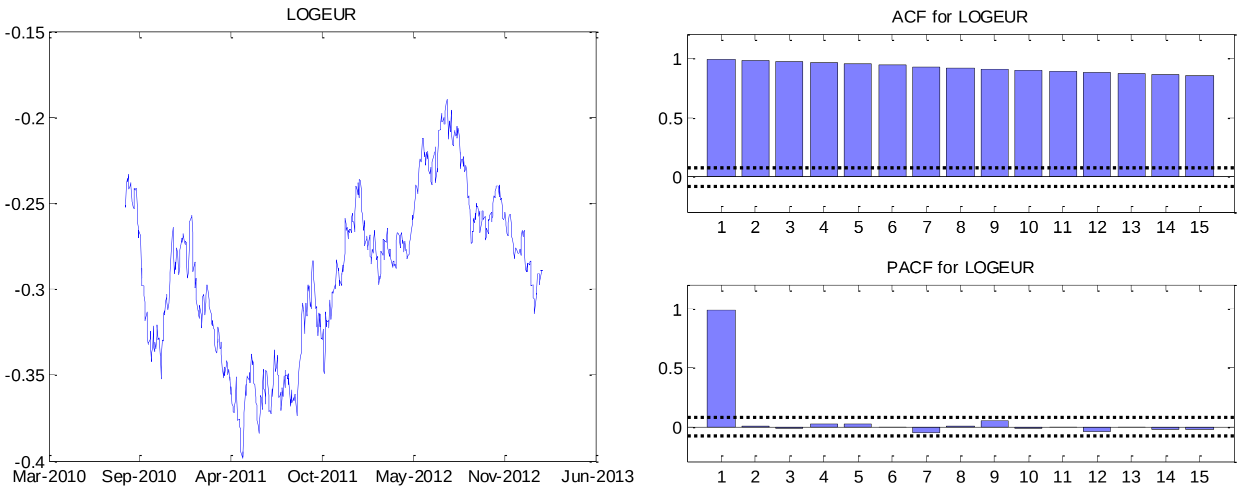
Part 3: Multivariate analysis of a basket of currencies

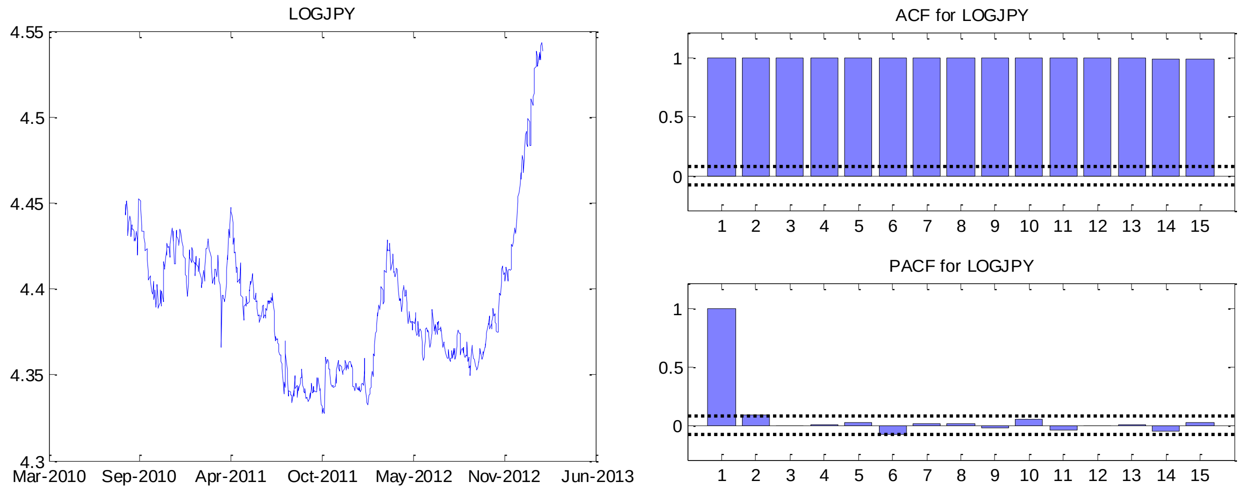
Plots of raw data



From the above figures, it seems that the data is not stationary. The log of the different currencies is going to be taken to try to make it linear – and as the focus lies on relative changes rather than absolute ones. Furthermore, the ACF and PACF graphs will be presented allied with an Augmented Dickey-Fuller test in order to conclude whether or not the series are stationary.

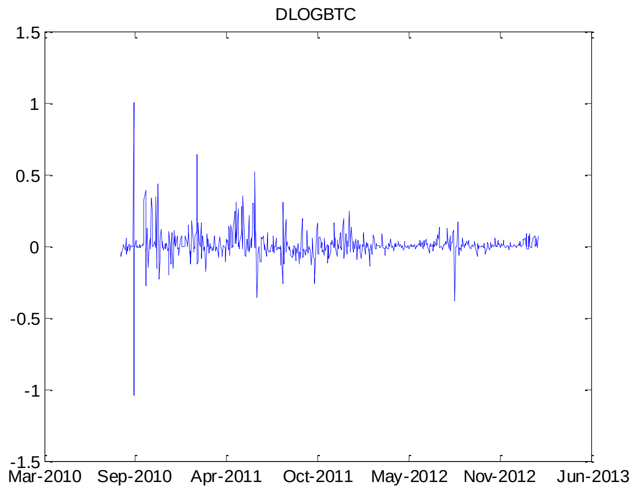


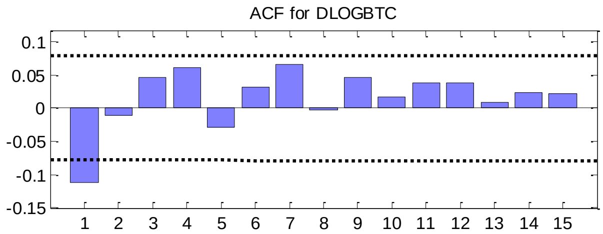


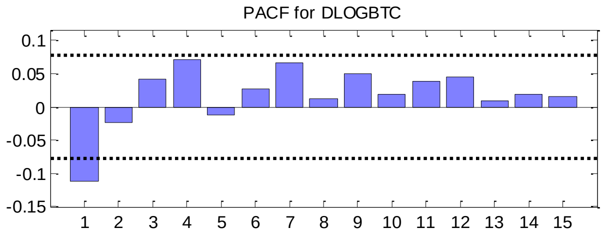


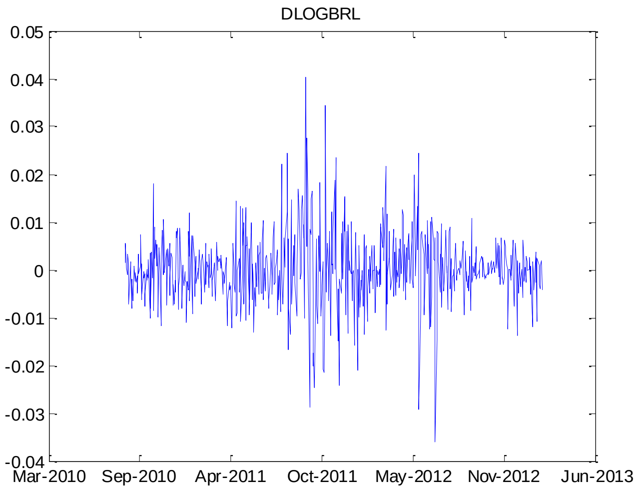
	Augmented Dickey-Fuller Test				
	LOGBTC	**LOGBRL**	**LOGCNY**	**LOGEUR**	**LOGJPY**
Test stat	-2.6677	-2.4992	-2.0874	-2.4810	0.1721
p-value	0.2531	0.3316	0.5545	0.3408	0.9971
Unit root?	Yes	Yes	Yes	Yes	Yes

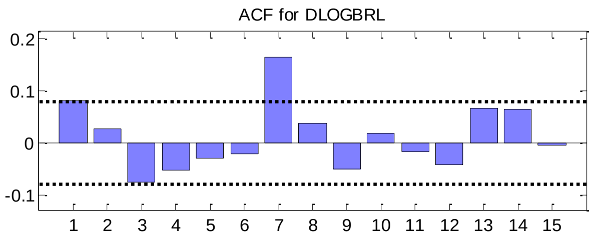
From the above graphs, it is possible to conclude that the series are not stationary. Both the autocorrelograms and the Augmented Dickey Fuller tests suggest that all the time series have a unit root (in the latter, , the null hypothesis cannot be rejected with a significance of over 5% in all cases). In order to make these series stationary, the first difference of each time series is taken.

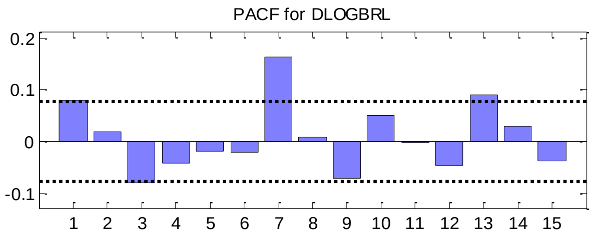

DLOGBTC
Mar-2010
Sep-2010
Apr-2011
Oct-2011
May-2012
Nov-2012
Jun-2013

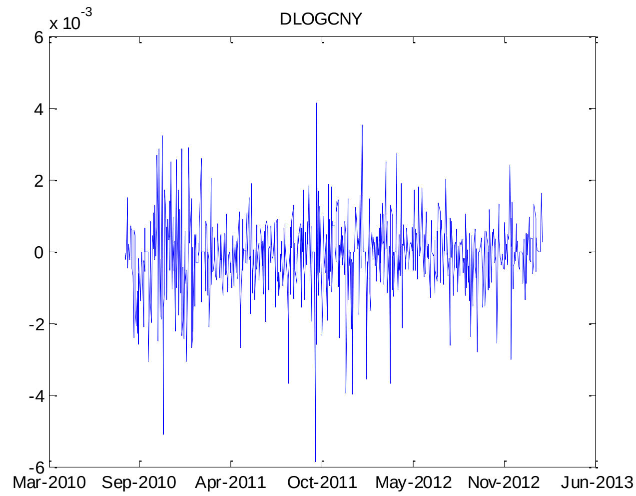

ACF for DLOGBTC

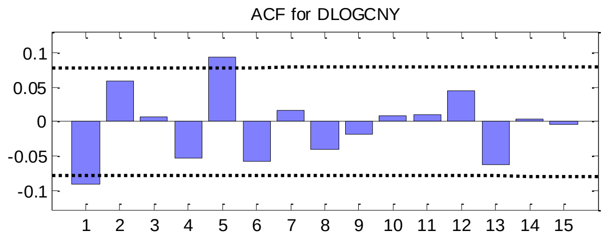

PACF for DLOGBTC

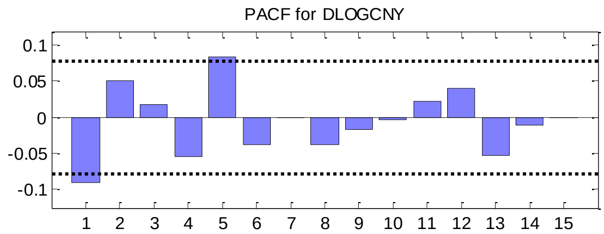

DLOGBRL
Mar-2010
Sep-2010
Apr-2011
Oct-2011
May-2012
Nov-2012
Jun-2013

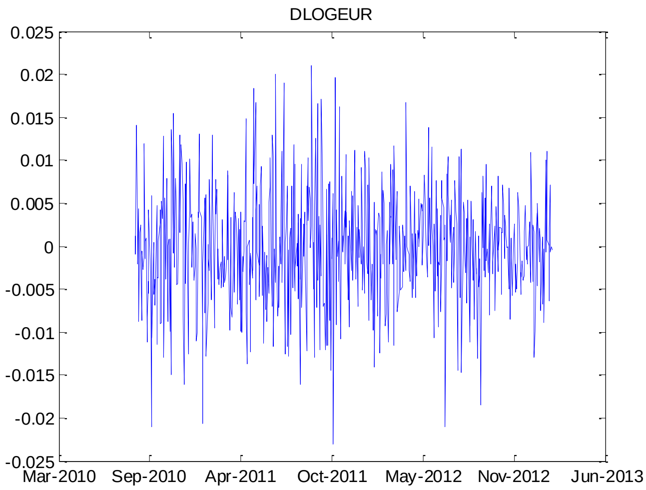

ACF for DLOGBRL

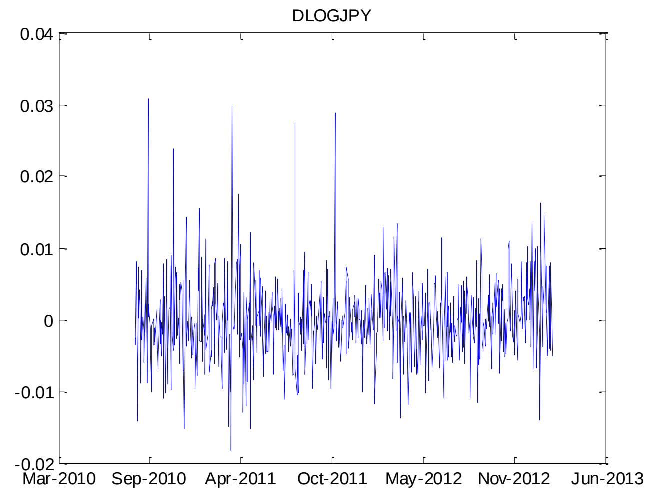

PACF for DLOGBRL


x 10-3
DLOGCNY
Mar-2010
Sep-2010
Apr-2011
Oct-2011
May-2012
Nov-2012
Jun-2013


ACF for DLOGCNY


PACF for DLOGCNY





	Augmented Dickey-Fuller Test				
	DLOGBTC	**DLOGBRL**	**DLOGCNY**	**DLOGEUR**	**DLOGJPY**
Test stat	-3.7313	-8.5584	-10.1880	-25.9080	-27.9036
p-value	0.0000	0.0000	0.0000	0.0000	0.0000
Unit root?	No	No	No	No	No

The first difference makes all the time series stationary. The ACFs decay quickly to zero and the Augmented Dickey-Fuller test rejects all the null hypotheses of a unit root. Now the study can proceed with the multivariate analysis of the first differences of the log of the exchange rates.

VAR Lag Length Selection

VAR(i)	MAIC	MSBIC	MHQIC
i=1	**-48.8986**	**-48.6932**	**-48.8190**
i=2	-48.8600	-48.4834	-48.7140
i=3	-48.8397	-48.2919	-48.6273
i=4	-48.7948	-48.0759	-48.5160
i=5	-48.7657	-47.8756	-48.4206

When minimizing the output obtained by each information criteria, all three of them suggest that the VAR(1) is the one more appropriate. The high number of coefficients required is highly penalized and all three criterions favour a model with the least amount of lags.

Resulting VAR(1) model

(**DLOGBTC**) $y_{1,t} = \beta_{10} + \beta_{11}y_{1,t-1} + \beta_{12}y_{2,t-1} + \beta_{13}y_{3,t-1} + \beta_{14}y_{4,t-1} + \beta_{15}y_{5,t-1}$

(**DLOGBRL**) $y_{1,t} = \beta_{20} + \beta_{21}y_{1,t-1} + \beta_{22}y_{2,t-1} + \beta_{23}y_{3,t-1} + \beta_{24}y_{4,t-1} + \beta_{25}y_{5,t-1}$

(**DLOGCNY**) $y_{1,t} = \beta_{30} + \beta_{31}y_{1,t-1} + \beta_{32}y_{2,t-1} + \beta_{33}y_{3,t-1} + \beta_{34}y_{4,t-1} + \beta_{35}y_{5,t-1}$

(**DLOGEUR**) $y_{1,t} = \beta_{40} + \beta_{41}y_{1,t-1} + \beta_{42}y_{2,t-1} + \beta_{43}y_{3,t-1} + \beta_{44}y_{4,t-1} + \beta_{45}y_{5,t-1}$

(**DLOGJPY**) $y_{1,t} = \beta_{50} + \beta_{51}y_{1,t-1} + \beta_{52}y_{2,t-1} + \beta_{53}y_{3,t-1} + \beta_{54}y_{4,t-1} + \beta_{55}y_{5,t-1}$

The coefficients for the estimated model are, in the previous order:

Constant		BTC	BRL	CNY	EUR	JPY
0.010393	**BTC**	-0.105376	-0.274384	-0.466084	0.092926	-2.342119
0.000116	**BRL**	-0.004176	0.083609	-0.461744	0.006982	0.106900
-0.000145	**CNY**	-0.000526	0.015383	-0.136681	0.028643	0.017028
-0.000071	**EUR**	-0.002587	-0.007135	-0.205590	-0.013679	0.067303
0.000200	**JPY**	-0.002492	-0.048406	0.151264	-0.038806	-0.081322

And their standard errors:

Constant		BTC	BRL	CNY	EUR	JPY
0.003454	**BTC**	0.139339	0.534789	5.369861	0.706282	1.385150
0.000288	**BRL**	0.002978	0.061288	0.274188	0.058271	0.061509
0.000041	**CNY**	0.000601	0.007632	0.054615	0.007853	0.008266
0.000259	**EUR**	0.002371	0.044796	0.258259	0.045446	0.047842
1.4357	**JPY**	0.001701	0.036725	0.249347	0.039300	0.044988

Test statistics:

Constant		BTC	BRL	CNY	EUR	JPY
3.008714	**BTC**	-0.756253	-0.513069	-0.086796	0.131570	-1.690878
0.402414	**BRL**	-1.402511	1.364190	-1.684041	0.119813	1.737969
-3.572076	**CNY**	-0.874945	2.015512	-2.502654	3.647648	2.059982
-0.274477	**EUR**	-1.090857	-0.159276	-0.796061	-0.300987	1.406768
0.901406	**JPY**	-1.464863	-1.318069	0.606641	-0.987431	-1.807623

And their corresponding p-values:

Constant		BTC	BRL	CNY	EUR	JPY
0.002624	**BTC**	0.449497	0.607903	0.930833	0.895324	0.090860
0.687380	**BRL**	0.160763	0.172508	0.092174	0.904632	0.082216
0.000354	**CNY**	0.381604	0.043851	0.012327	0.000265	0.039400
0.783718	**EUR**	0.275336	0.873452	0.425997	0.763425	0.159496
0.367373	**JPY**	0.142958	0.187480	0.544089	0.323432	0.070665

From the previous results, and with a significance of 5%, for the first equation none of the coefficients is significant. In fact, the only significant coefficients are $\beta_{32}, \beta_{33}, \beta_{34}, \beta_{34}$ and β_{35}. Therefore the first lag of the difference of the log of the Brazil Real (DLOGBRL), the Euro (DLOGEUR), the Japan Yen (DLOGJPY) and the Chinese Yuan (DLOGCNY) itself are significant when explaining part of the variance of the difference of the log of the Chinese Yuan (DLOGCNY). However, the explained variance by the predictors is not very high in any case, as can be seen by the R^2 obtained for each equation.

	BTC	BRL	CNY	EUR	JPY
R^2	0.0311	0.0209	0.0804	0.0063	0.0188

Granger causality test

Test statistics

	BTC	BRL	CNY	EUR	JPY
BTC	3.1582	0.324985	0.009377	0.022814	11.015159
BRL	2.6033	2.372486	3.739506	0.020215	5.121300
CNY	1.7461	4.786532	7.726186	14.688728	5.438579
EUR	1.8671	0.035165	0.937405	0.126786	3.185389
JPY	3.7251	1.916988	0.486030	1.039372	3.712703

P-values

	BTC	BRL	CNY	EUR	JPY
BTC	0.0755	0.568627	0.922858	0.879942	0.000904
BRL	0.1066	0.123490	0.053140	0.886937	0.023634
CNY	0.1864	0.028683	0.005443	0.000127	0.019697
EUR	0.1718	0.851250	0.332946	0.721788	0.074299
JPY	0.0536	0.166189	0.485704	0.307968	0.054000

In the table above, it is possible to say with a significance of 5% that:

- DLOGJPY Granger-causes DLOGBTC (p-value of 0.000904 → cannot reject H_o)
- DLOGJPY Granger-causes DLOGBRL (p-value of 0.023634 → cannot reject H_o)
- DLOGBRL Granger-causes DLOGCNY (p-value of 0.028683 → cannot reject H_o)
- DLOGCNY Granger-causes DLOGCNY (p-value of 0.005443 → cannot reject H_o)
- DLOGEUR Granger-causes DLOGCNY (p-value of 0.000127 → cannot reject H_o)
- DLOGJPY Granger-causes DLOGCNY (p-value of 0.019697 → cannot reject H_o)

There is no significant bi-directional causality among the time series.

Section 6: Conclusions

In the first part, an extensive analysis has been performed to 26 time series. The study found that 8 of them present structural breaks, possibly related to Bitcoin's mid-2011 crash. Noteworthy is the fact that, during the analysis, it was found that the behaviour of the variable *Number of Addresses Used* is very close to that of *Number of Transactions Excluding Popular Addresses*. It has also been seen how fast transactions are increasing with time as well as transaction fees, which increase while the block reward is halved every 210,000 blocks.

In the second part, many interesting cointegration relationships have been found. While the variables *Number of Transactions* and *Number of Unique Addresses Used* are not cointegrated, the variables *Number of Transactions Excluding Popular Addresses* and *Number of Unique Addresses Used* are. This suggests that a high amount of transactions are held by a few Bitcoin addresses. Other cointegration relationships occur between the variable *Hash Rate* and the variables *Market Cap*, *Market Price*, and *Number of Unique Addresses Used*. This last variable is also cointegrated with *Market Price*. Therefore, the number of users of the network "moves" together in time with the exchange rate. Another meaningful cointegration has been found between the variable *Number of Transactions* and *Number of Transactions per Block*. This was to be expected, as blocks are processed at a nearly-constant interval – one block approximately every 10 minutes – as specified in the Bitcoin protocol.

As for the third part, in which a basket of currencies including the Bitcoin has been analysed, the unit root test (Augmented Dickey-Fuller test) on logarithmic series shows that all of them have a unit root and hence are not stationary. When the first difference is taken, the results of the test show significant stationarity. The Granger causality tests do not show any significant bi-directional causality, but show that there is unidirectional causality between JPY and BTC (all currencies are measured in USD), in the sense that JPY is useful to forecast BTC, while the converse is not true. Furthermore, JPY is also useful in forecasting BRL. Last but not least, the BRL, EUR, JPY (and the CNY itself) are useful in forecasting the CNY exchange rate.

Since this paper started being written, Bitcoin has become increasingly popular. The exchange rate as of 03-Apr-2013 is 128.7304 USD/BTC. It would be interesting for further research to focus on **scalability of the Bitcoin network** (1) as well as on its liquidity.

(1) On 12-Mar-2013, a large number of miners agreed to upgrade to Bitcoin version 0.8 as the previous version had scalability issues regarding a software limit on the size of blocks. During the process, it has been proved that a successful double spend took place (https://bitcointalk.org/index.php?topic=152348.0). Another important issue is the fact that a single exchange service, MtGox, accounts for around 70% of the whole exchange market (http://bitcoincharts.com/charts/volumepie as). It seems that, although the Bitcoin protocol is decentralized, the exchange market is not. A problem in the service offered by MtGox could have a big impact on the currency and its value. **UPDATE:** On April 10th 2013, MtGox's trade engine started lagging, which meant that buy and sell orders were not being executed for a period of time. This caused panic and the bitcoin price plummeted from $266 to $120. The next day, trading was suspended for a few hours. Even though there was speculation of a DDOS attack towards MtGox's servers, the company attributed the problem to their own success. This event confirmed our thoughts regarding Bitcoin's scalability issues.

References

Brooks, Chris (2008). Introductory Econometrics for Finance. 2nd Edition

Enders, Walter (2004). Applied Econometric Time Series. 2nd Edition

Mikael Apel and Staffan Viotti (1998). Why is an independent central bank a good idea?

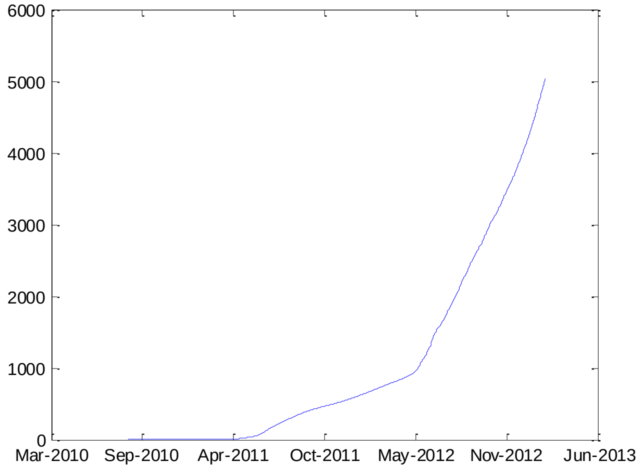
See Selgin, George and Larry White (1994) - "How Would the Invisible Hand Handle Money?

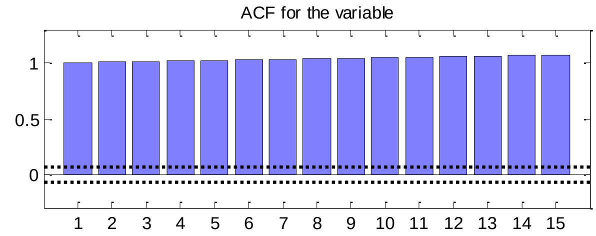
Appendices

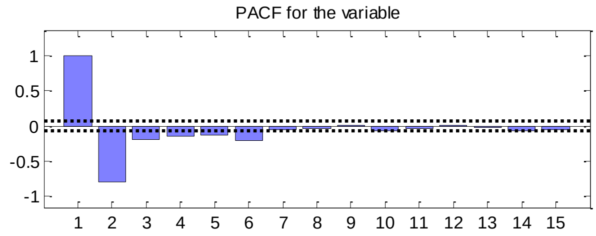
Part 1: Time series analysis

First variable analysed on the next page.

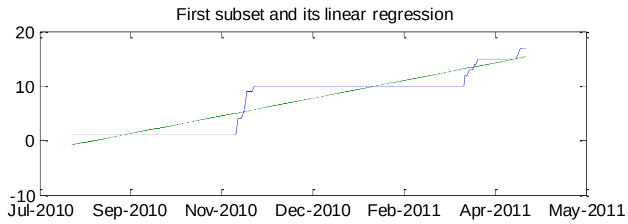
Variable Name: Blockchain Size

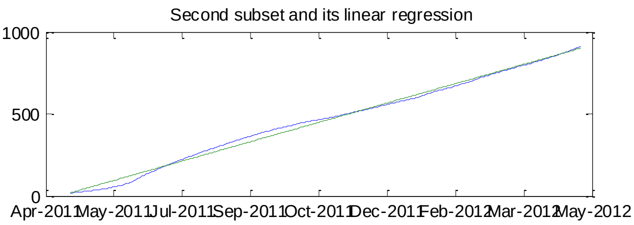


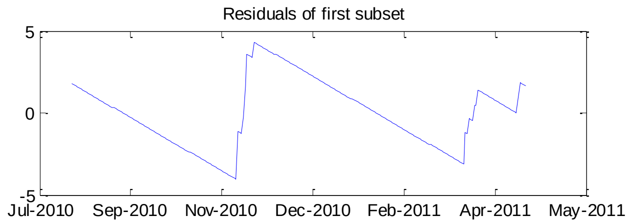


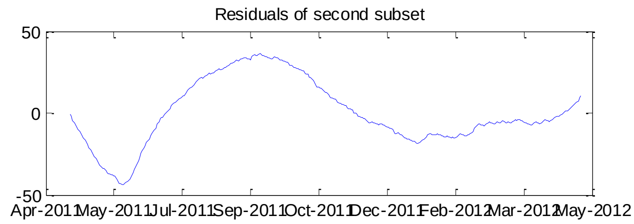


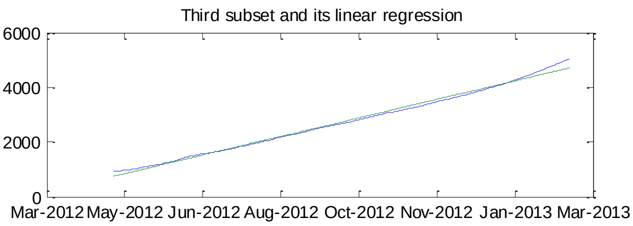
This variable is clearly not stationary. However, there seem to be unexpected shifts to its behaviour on 25-Apr-2011 (day 250 in the variable 'date') and on 02-May-2012 (day 625 in the variable 'date'). In order to test whether or not structural breaks exist, the data is split in three subsamples and two Chow tests are performed in order to determine if the coefficients among the different regressions of the subsamples can be considered significantly different.



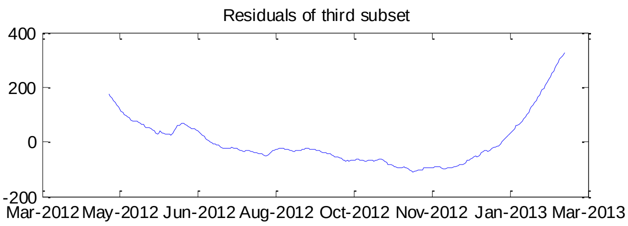




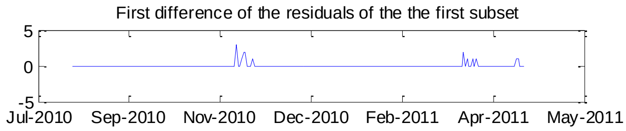


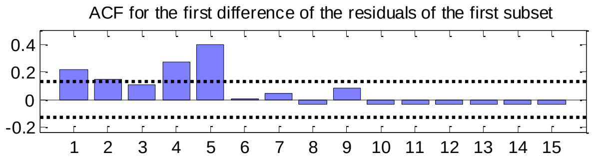


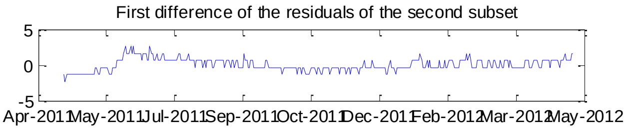
Chow Test		
Test statistic	**p-value**	**Break?**
F = 1.2063e+04	0.0000	Yes
F = 2.4134e+04	0.0000	Yes

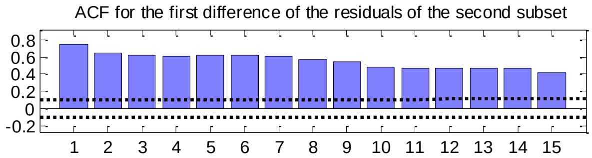


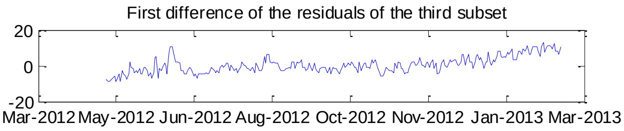
The test statistics for the Chow test take extremely high values, suggesting that there the two structural breaks indeed exist. However, the residuals obtained are still not stationary. If the first difference is taken:

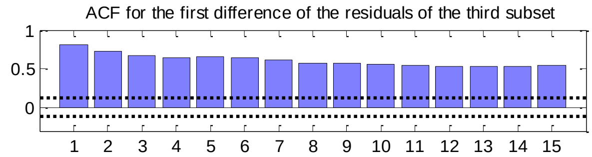




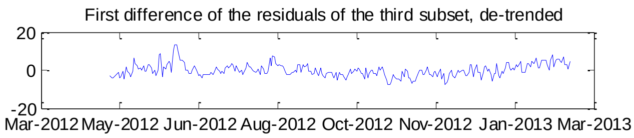


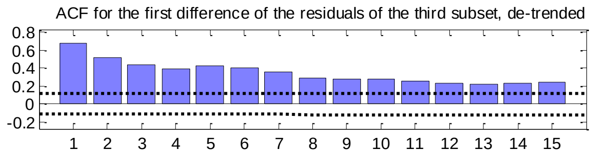






Looking at the autocorrelations, the subsets are stationary. In addition, there seems to be a flat line on the first difference of the residuals of the first subset. This means that, for most values in this range, the change in the Blockchain Size was zero. This is because the Blockchain is measured in Megabytes, and by 2010, the amount of transactions was so low that, to measure its change, a smaller unit would be needed (such as Kilobytes). In addition, there seems to be a time trend in the third subset. If a regression is made, the resulting time-dependent coefficient is 0.0359 and has a 95% confidence interval of [0.0312, 0.0405], therefore the time trend is significant. Therefore the third subset is de-trended.

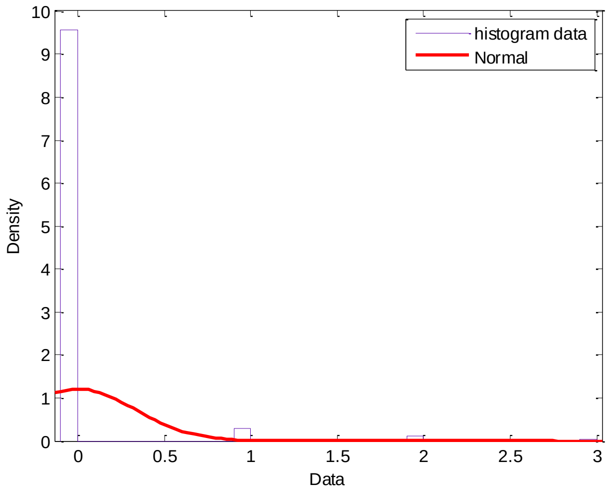


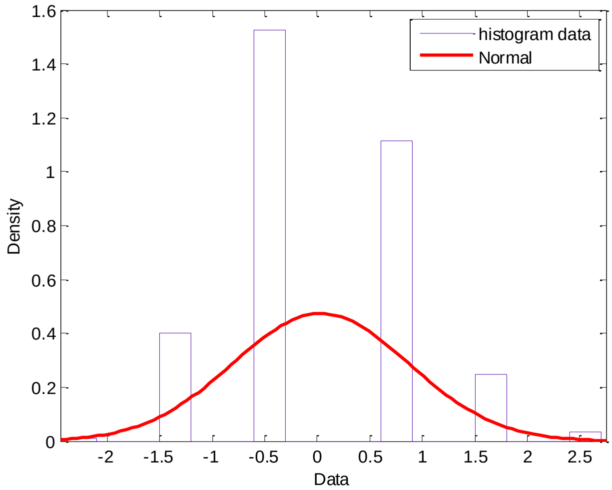


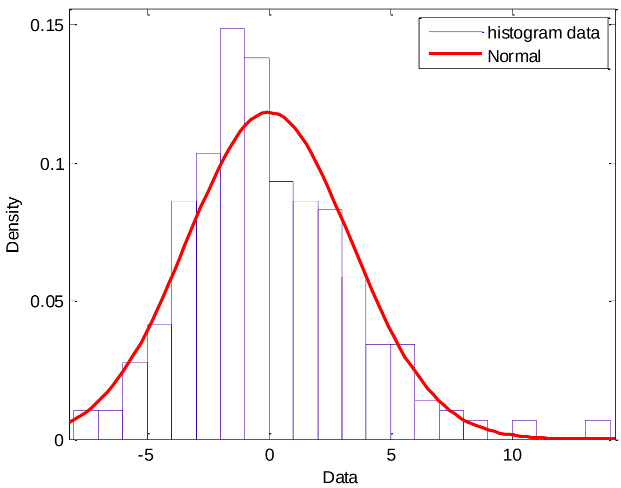
The Dickey Fuller Test confirms our hypothesis.

	Augmented Dickey-Fuller Test		
	Test statistic	**p-value**	**Unit root?**
First subset	-4.8094	0.0000	No
Second subset	-2.4277	0.0160	No
Third subset	-3.5472	0.0000	No

Summary statistics:	First difference of the first subset	First difference of the second subset	First difference of the third subset (de-trended)
Mean	-0.0004	0.0306	0.0000
Minimum	-0.0647	-2.3598	-7.5979
25th Percentile	-0.0647	-0.3598	-2.1485
Median	-0.0647	-0.3598	-0.4052
75th Percentile	-0.0647	0.6402	2.0589
Maximum	2.9353	2.6402	13.4968
Variance	0.1088	0.7105	11.3998
Standard Deviation	0.3298	0.8429	3.3764
Skewness	5.9406	0.3201	0.7482
Kurtosis	41.6801	3.1715	4.3900
Excess kurtosis	38.6801	0.1715	1.3900







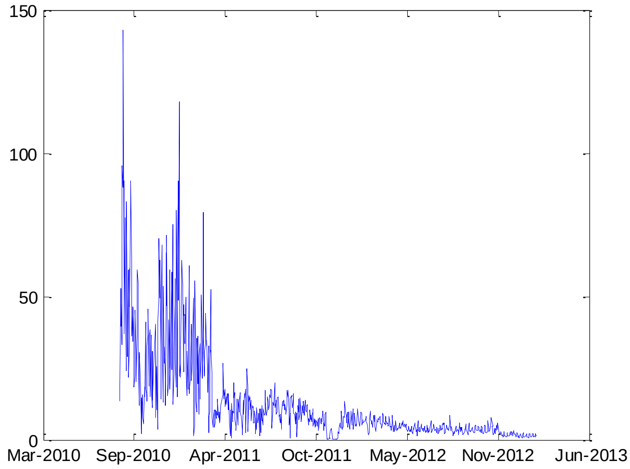
Looking at the histograms, it is clear that the time series for the first two data sets do not follow a normal distribution. This is due to the reason explained earlier – the size of the Blockchain is measured in Megabytes, and is not continuous. For the third subset, however, the amounts are larger so the behaviour is similar to a continuous variable, and the resemblance to a normal distribution is higher. The following tests are conducted only to the first difference of the de-trended third subset:

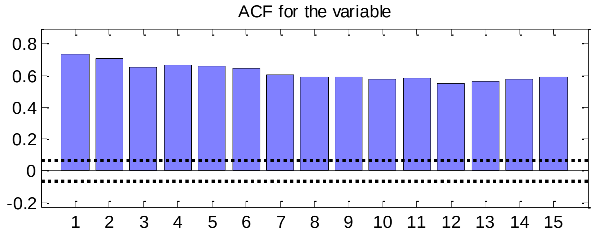
	Hypothesis	**Test statistic**	**Distribution**	**p-value**	**Significantly ≠ 0?**
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	t = 2.0298e-14	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	1.000	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	t_s = 5.2019	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	1.9724e-07	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	t_k = 4.8320	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	1.3519e-06	Yes

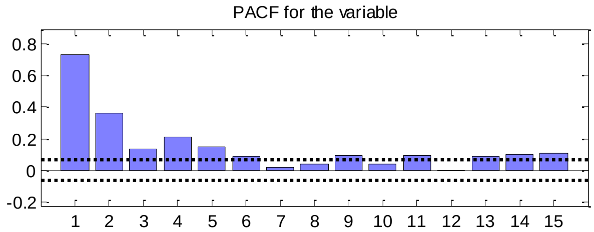
Our tests suggest that the third subset (the most recent one) – although not continuous – has a skewness and excess kurtosis not significantly different from zero.. Because of this, it can't be said that it behaves similarly to a Normal distribution, as the Jarque-Bera test suggests (although the sample size may be too small to conduct this test).

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or $k \neq 3$	JB = 50.4079	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	1.1326e-11	Yes

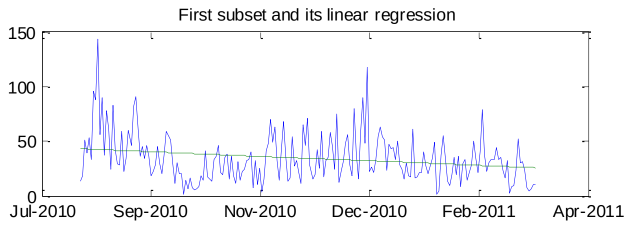
Variable Name: Cost % of Transaction Volume

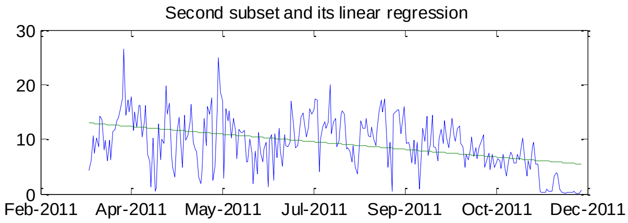


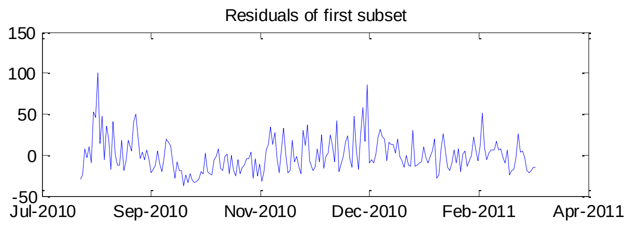


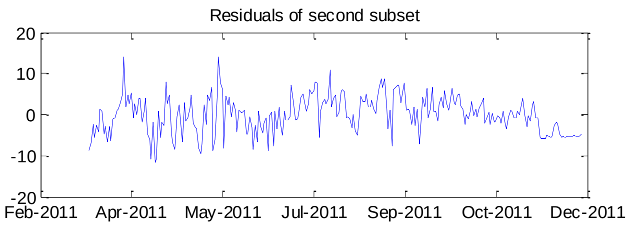


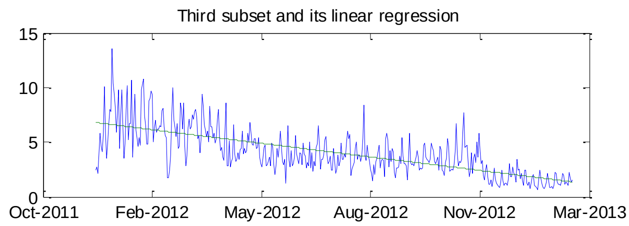
This variable does not look stationary, as shown in both the regular plot and the ACF. Again, there seem to be a couple of structural breaks. The behaviour of the variable seems to change abruptly on 13-Mar-2011 (day 209 in the variable 'date') and on 26-Jun-2012 (day 680 in the variable 'date'). Let's split the data and conduct Chow tests in order to confirm our hypotheses:



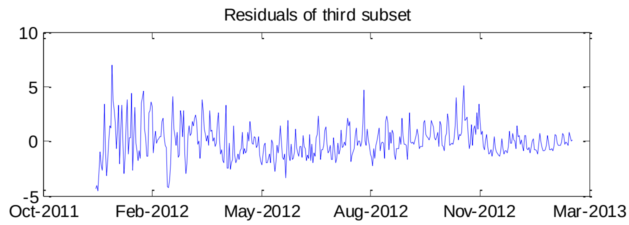




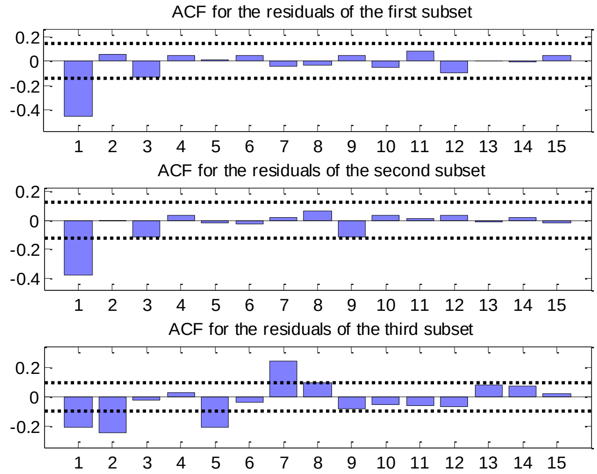




Chow Test		
Test statistic	**p-value**	**Break?**
F = 18.7383	2.9346e-08	Yes
F = 17.7502	6.0331e-08	Yes



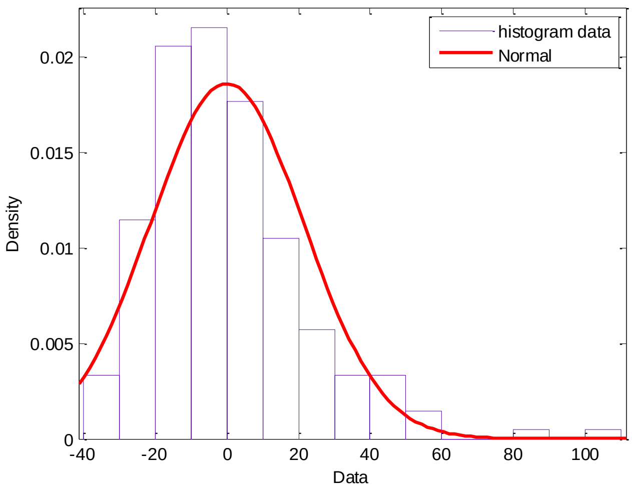
Again, the Chow test suggests that two structural breaks exist. The residuals of the de-trended subsets seem to be stationary. Let's have a look at their ACF:

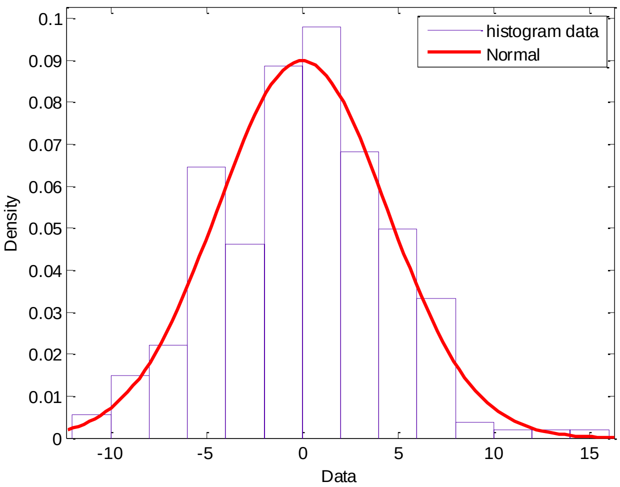


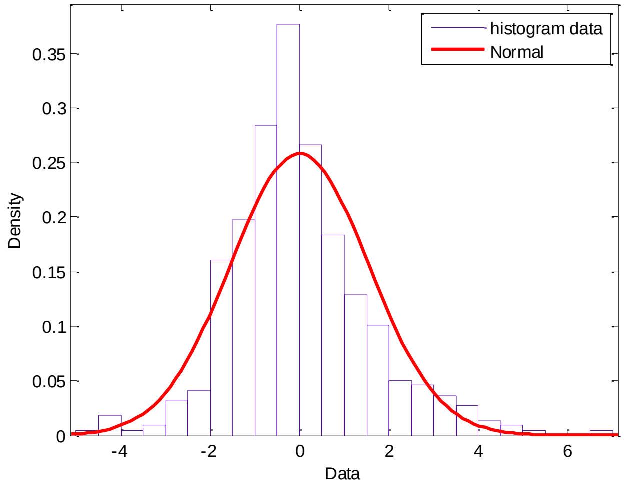
Indeed, the ACF of the three subsets confirm our hypotheses of stationarity.

	Dickey-Fuller Test		
	Test statistic	p-value	Unit root?
First subset	-6.7538	0.0000	No
Second subset	-6.9015	0.0000	No
Third subset	-11.7610	0.0000	No

Summary statistics:	First subset	Second subset	Third subset
Mean	0.0000	0.0000	0.0000
Minimum	-37.1748	-11.5627	-4.6003
25th Percentile	-15.4303	-2.9304	-0.8965
Median	-3.3018	0.4471	-0.2023
75th Percentile	10.5717	2.9292	0.7287
Maximum	100.8034	14.1211	6.9874
Variance	459.1743	19.7148	2.3913
Standard Deviation	21.4284	4.4401	1.5464
Skewness	1.2387	-0.0018	0.5847
Kurtosis	5.6807	2.9797	4.5279
Excess kurtosis	2.6807	-0.0203	1.5279







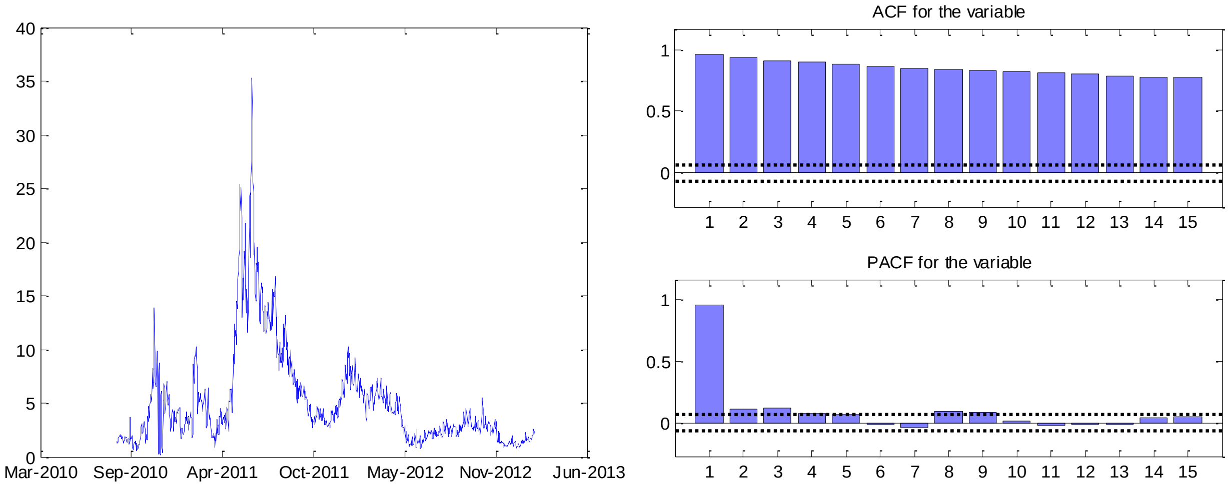
Skewness can be appreciated in the first subset and, to a lesser extent, in the third one. Both of these subsets also seem to have excess kurtosis. Let's confirm our findings through the appropriate statistics:

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t_1 =$ -9.9602e-15	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	1.0000	No
		$t_2 =$ -2.2358e-15		1.0000	No
		$t_3 =$ -3.6797e-14		1.0000	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_{s,1} =$ 7.3107	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
		$t_{s,2} =$ -0.0122		0.9903	No
		$t_{s,3} =$ 4.9844		0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_{k,1} =$ 7.9107	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes
		$t_{k,2} =$ -0.0684		0.9455	No
		$t_{k,3} =$ 6.5121		0.0000	Yes

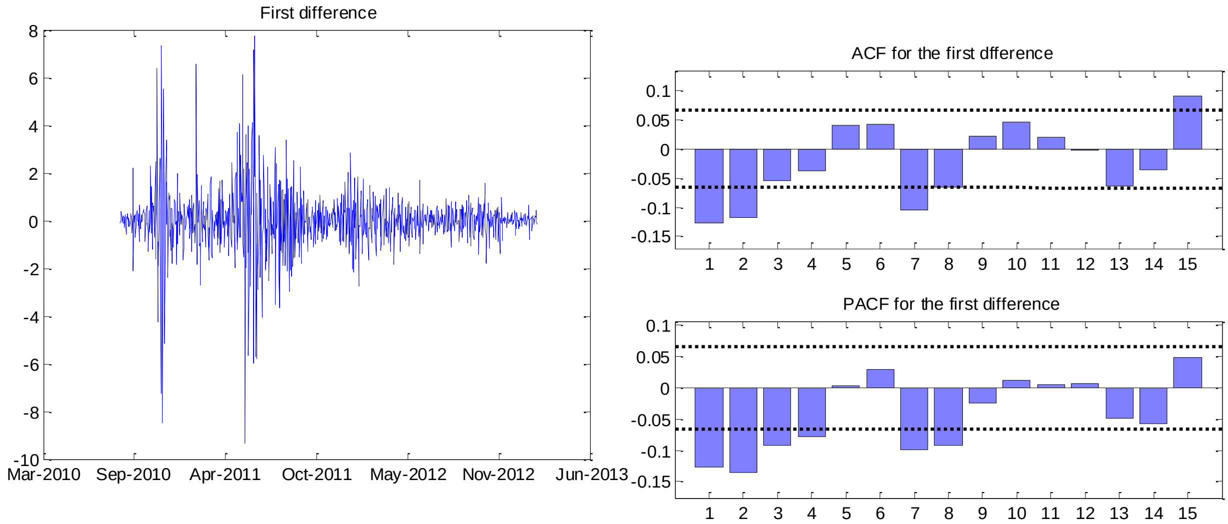
Clearly, both the first and the last subset seem to behave more differently than the normal distribution. The middle one, however, is significantly closer.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	$JB_1 =$ 116.0251	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.0000	Yes
		$JB_2 =$ 0.0048		0.9976	No
		$JB_3 =$ 67.2515		2.4425e-15	Yes

Variable Name: Cost per Transaction (USD)

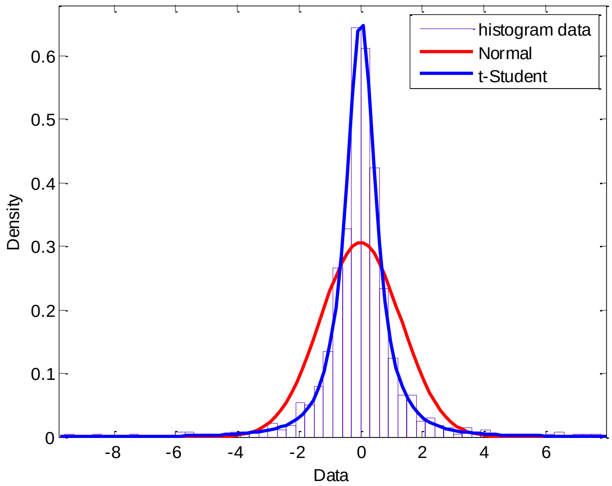


This variable is not stationary. Its behaviour keeps changing over time in an – a priori – unpredictable way. However, in the beginning the cost per transaction tended to increase until reaching a maximum. Since then, the cost has been overall reduced. Let's not forget that the cost is the same as the miners' revenue, which heavily depends on the exchange rate ($/BTC). If we take the first difference:



The result is stationary, although the variance changes over time. In fact, it seems that the variance tends to decrease with time – except in the beginning, where there is a period with small volatility. The Dickey-Fuller test confirms the hypothesis that there is no unit root.

	Augmented Dickey-Fuller Test		
	Test statistic	p-value	Unit root?
First difference	-9.1422	0.0000	No



Summary statistics:	First difference
Mean	0.0008
Minimum	-9.3501
25th Percentile	-0.4380
Median	0.0032
75th Percentile	0.4347
Maximum	7.7567
Variance	1.6941
Standard Deviation	1.3016
Skewness	-0.2234
Kurtosis	15.6810
Excess kurtosis	12.6810

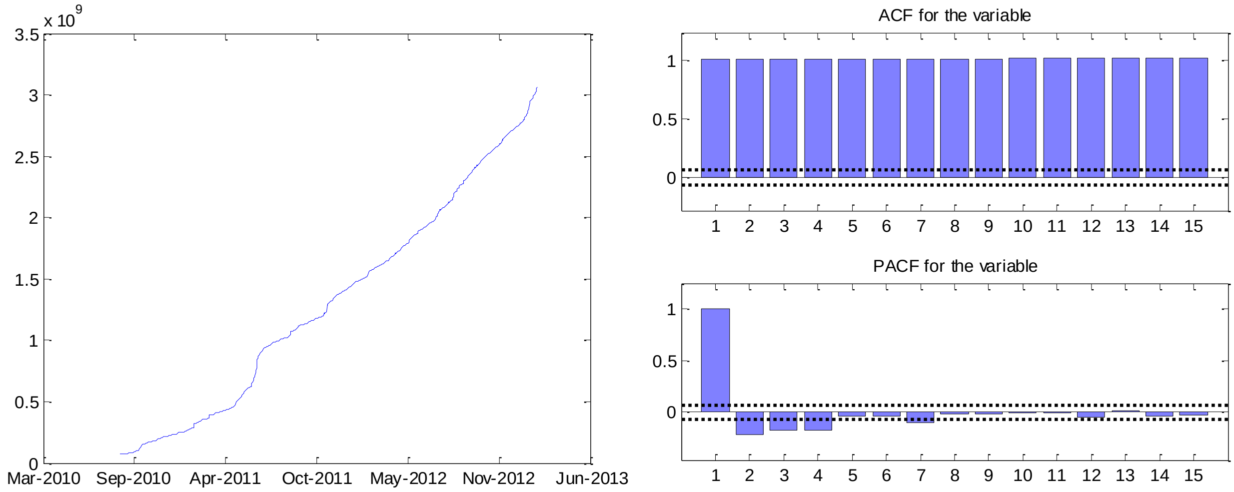
The first difference has high excess kurtosis. It behaves very similarly to a t-Student distribution. If the time series is compared with the normal distribution:

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t = 0.0189$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.9849	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = -2.7592$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0058	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 78.2994$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

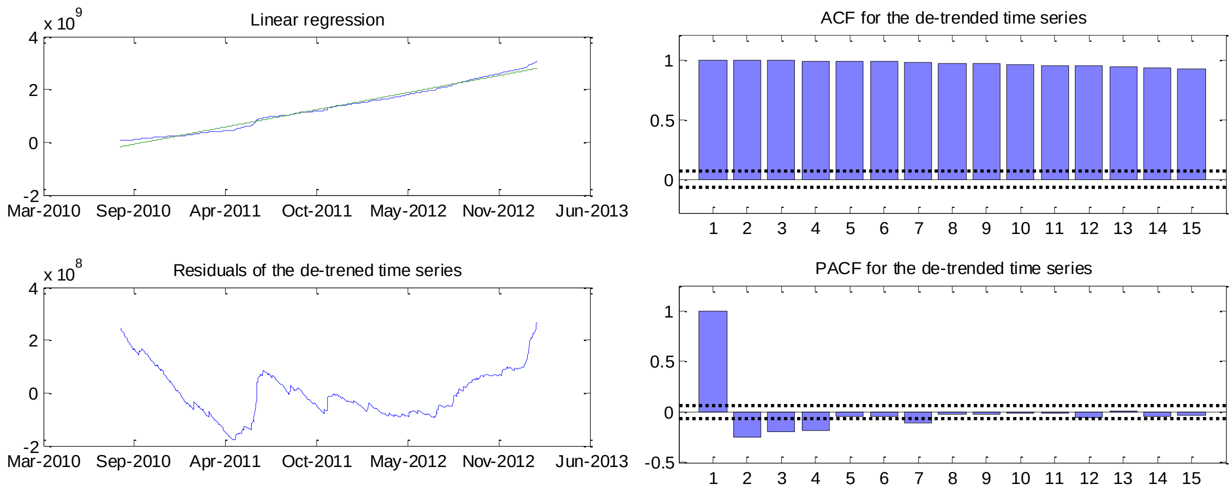
The series indeed has significant excess kurtosis. It is also significantly skewed. The Jarque-Bera test shows us that the first difference of the time series behaves significantly different than a normal distribution, due to its high skewness and excess kurtosis.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k ≠ 3	JB = 6.1384e+03	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.0000	Yes

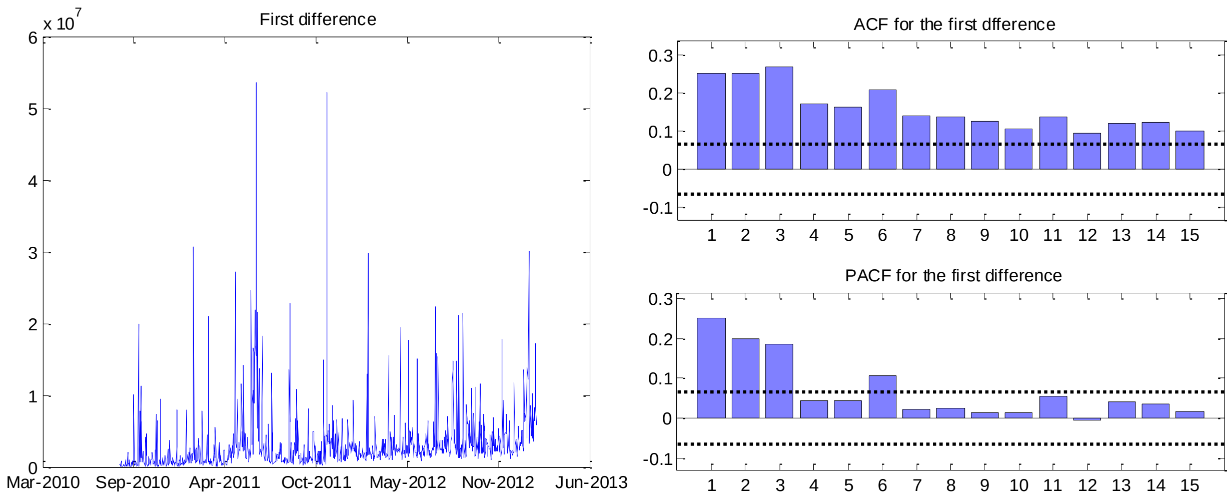
Variable Name: Bitcoin Days Destroyed – Cumulative



There seems to be a time trend in the current time series. Let's take it out and look at the result.



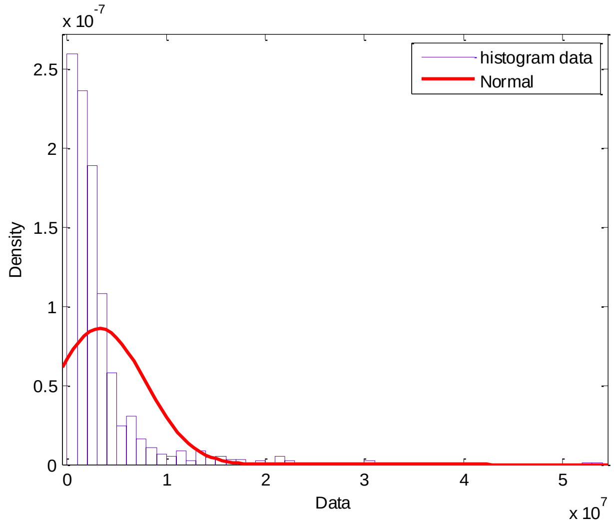
The autocorrelation function of the residuals starts at 1 at the first lag and decays very slowly - suggesting that the residuals of the de-trended time series is not stationary. Let's analyse he first difference.



There seem to be some spikes in the variance but the autocorrelations suggest that the variable is indeed stationary.

	Augmented Dickey-Fuller Test		
	Test statistic	p-value	Unit root?
First difference	-7.6826	0.0000	No

The Dickey-Fuller test indicates that the time series does not have a unit root.



Summary statistics:	First difference
Mean	3.2671e+06
Minimum	25227
25th Percentile	9.6378e+05
Median	2022038
75th Percentile	3.4392e+06
Maximum	53532635
Variance	2.1532e+13
Standard Deviation	4.6403e+06
Skewness	4.7523
Kurtosis	37.5647
Excess kurtosis	34.5647

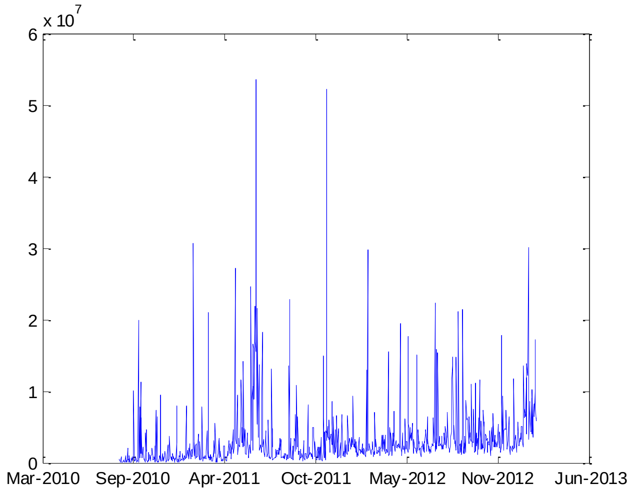
The variable under study does not follow a normal distribution. The variable always grows since the total number of Bitcoin days destroyed is a cumulative function. Therefore, the first difference only has positive values (the minimum value of the sample is 25227).

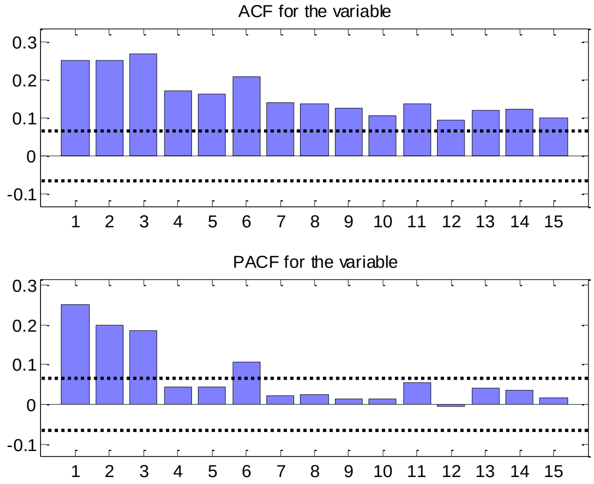
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t = 21.2978$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.0000	Yes
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = 58.6869$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 213.4214$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

The tests reject all the hypothess. The Jarque-Bera test does nothing but confirm that the variable does not follow a normal distribution.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	JB = 4.8993e+04	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.000	Yes

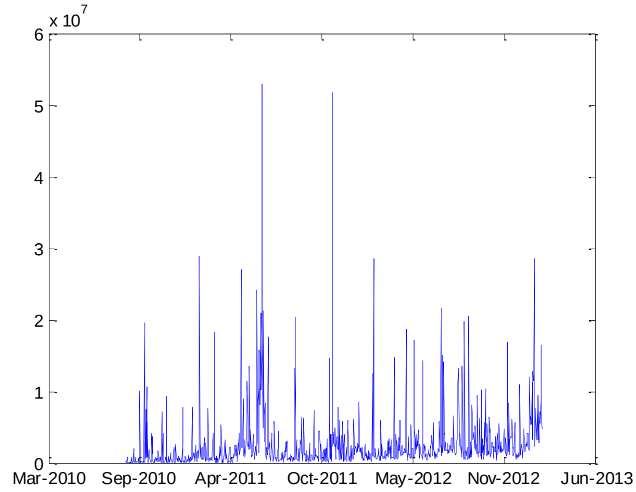
Variable Name: Bitcoin Days Destroyed

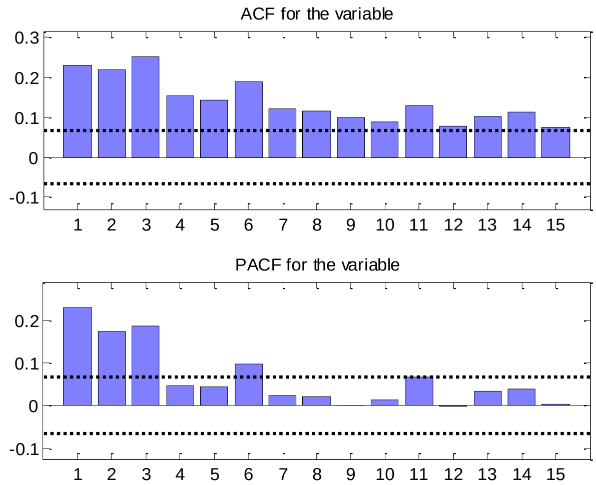




This variable is exactly the first difference of the last variable analysed 'Bitcoin Days Destroyed - Cumulative'.

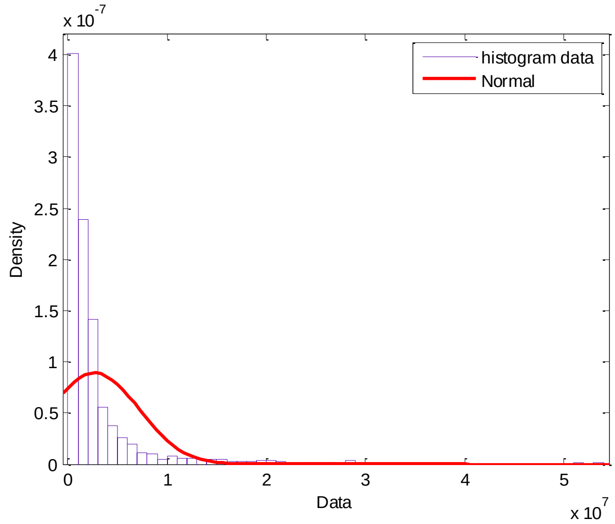
Variable Name: Bitcoin Days Destroyed – Filtered by 1 month





In this case, the bitcoin days destroyed are only accounted for those bitcoins that had not moved since at least 1 month. The autocorrelations are small and the Augmented Dickey-Fuller rejects the null hypothesis of a unit root.

	Augmented Dickey-Fuller Test		
	Test statistic	**p-value**	**Unit root?**
Variable	-3.0039	0.0033	No



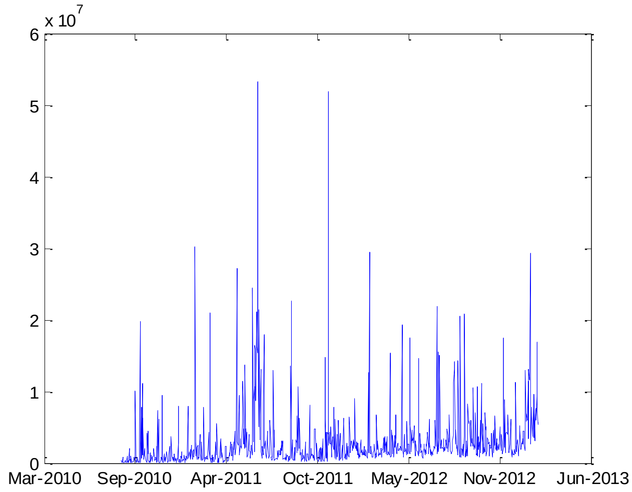
Summary statistics:	Variable
Mean	2.6716e+06
Minimum	3182
25th Percentile	5.8061e+05
Median	1379302
75th Percentile	2704865
Maximum	53013542
Variance	2.0007e+13
Standard Deviation	4.4729e+06
Skewness	5.1392
Kurtosis	42.7692
Excess kurtosis	39.7692

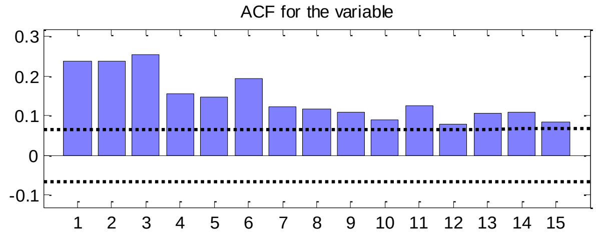
The time series clearly does not follow the Normal distribution.

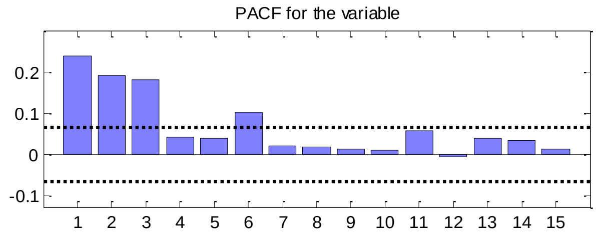
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t = 18.0770$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.0000	Yes
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = 63.4989$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 245.6907$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or $k \neq 3$	JB = 6.4396e+04	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.000	Yes

Variable Name: Bitcoin Days Destroyed – Filtered by 1 week

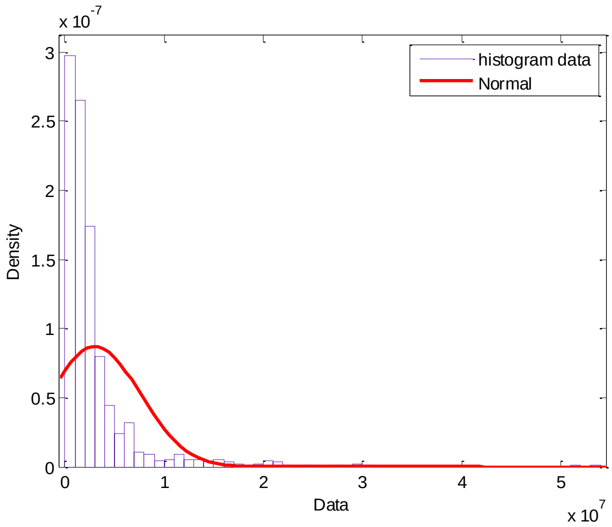






Same as the previous variable, but now filtered by 1 week. The variable is already stationary.

	Augmented Dickey-Fuller Test		
	Test statistic	p-value	Unit root?
Variable	-2.8495	0.0046	No



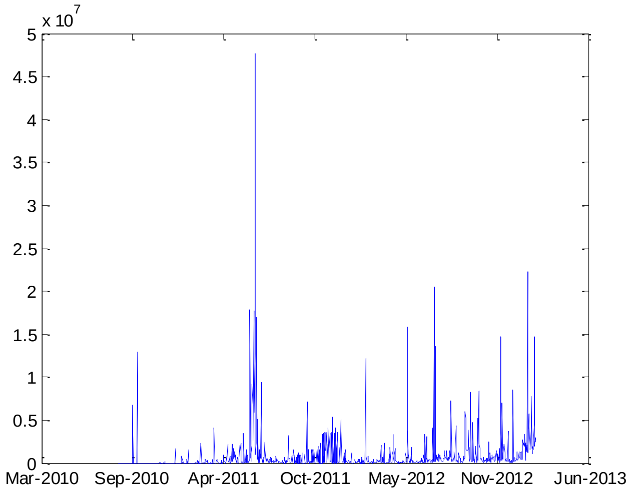
Summary statistics:	Variable
Mean	3.0235e+06
Minimum	16720
25th Percentile	8.0933e+05
Median	1.7739e+06
75th Percentile	3.0924e+06
Maximum	53207248
Variance	2.1045e+13
Standard Deviation	4.5875e+06
Skewness	4.8771
Kurtosis	39.0193
Excess kurtosis	36.0193

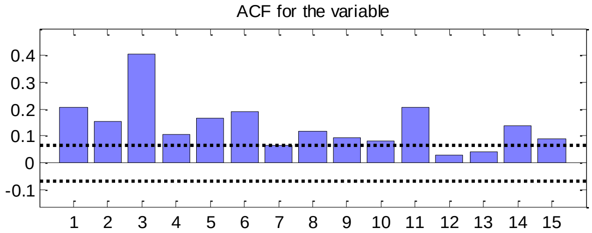
The time series clearly does not follow the Normal distribution.

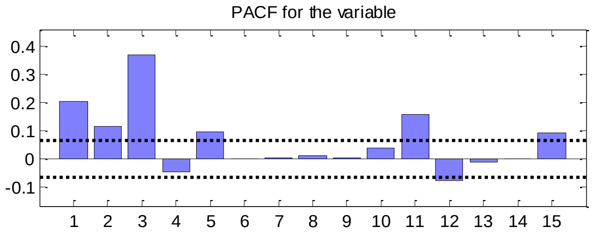
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t = 19.9474$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.0000	Yes
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = 60.2611$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 222.5242$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or $k \neq 3$	JB = 5.3148e+04	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.000	Yes

Variable Name: Bitcoin Days Destroyed – Filtered by 1 year

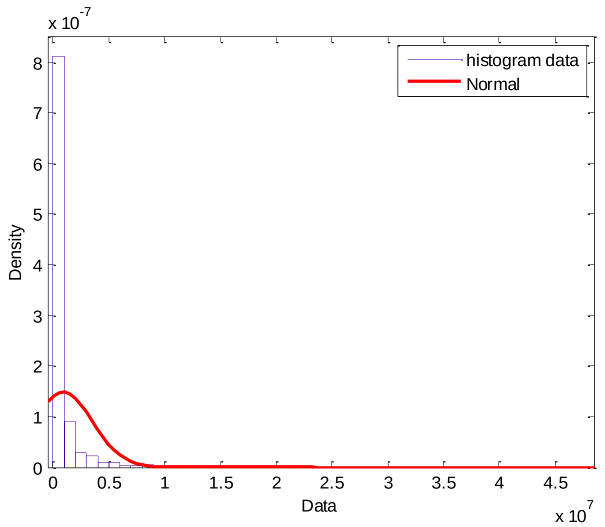






Same as the previous variable, but now filtered by 1 year. The variable is already stationary.

	Augmented Dickey-Fuller Test		
	Test statistic	p-value	Unit root?
Variable	-3.8970	0.0000	No



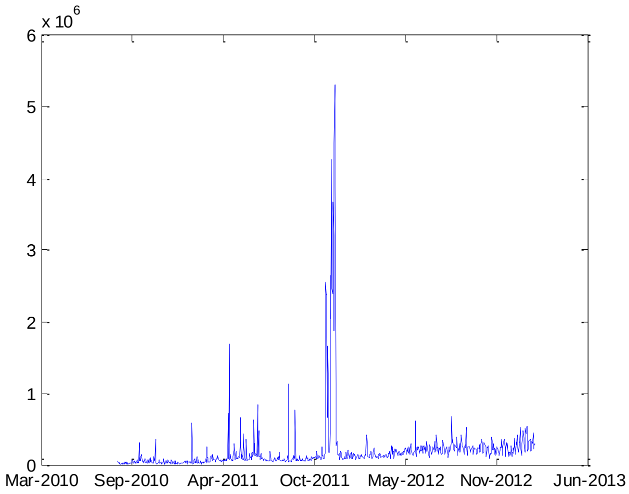
Summary statistics:	Variable
Mean	8.8091e+05
Minimum	0.0000
25th Percentile	11.5000
Median	1.3367e+05
75th Percentile	646590
Maximum	47664897
Variance	7.1807e+12
Standard Deviation	2.6797e+06
Skewness	8.7987
Kurtosis	119.3442
Excess kurtosis	116.3442

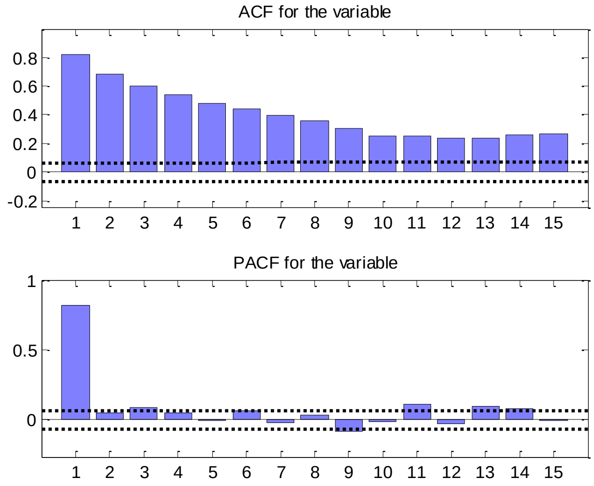
The time series clearly does not follow the Normal distribution.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t = 9.9493$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.0000	Yes
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = 108.7155$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 718.7652$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	JB = 5.2844e+05	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.000	Yes

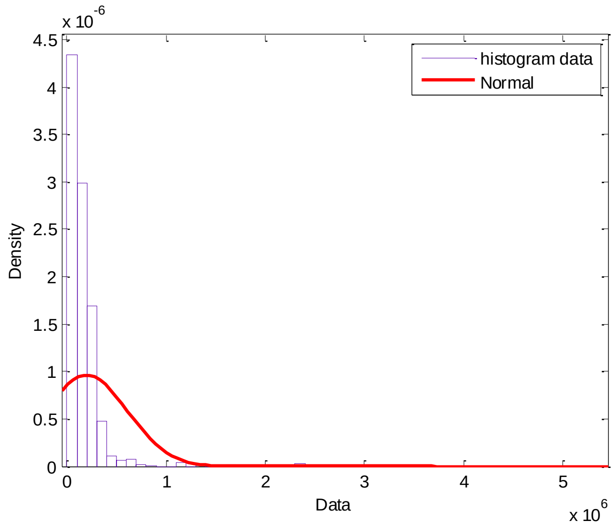
Variable Name: Estimated Transaction Volume (BTC)





The time series is stationary. It has some important spikes in late 2011. The autocorrelations are not too high and the Augmented Dickey-Fuller test rejects the null hypothesis of a unit root.

	Augmented Dickey-Fuller Test		
	Test statistic	p-value	Unit root?
Variable	-4.0810	0.0000	No



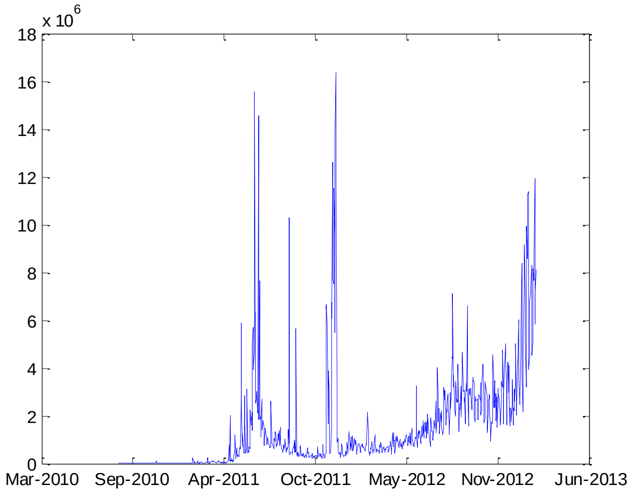
Summary statistics:	Variable
Mean	1.9661e+05
Minimum	6221
25th Percentile	58563
Median	1.1828e+05
75th Percentile	2.1129e+05
Maximum	5312433
Variance	1.7176e+11
Standard Deviation	4.1443e+05
Skewness	7.5996
Kurtosis	71.0081
Excess kurtosis	68.0081

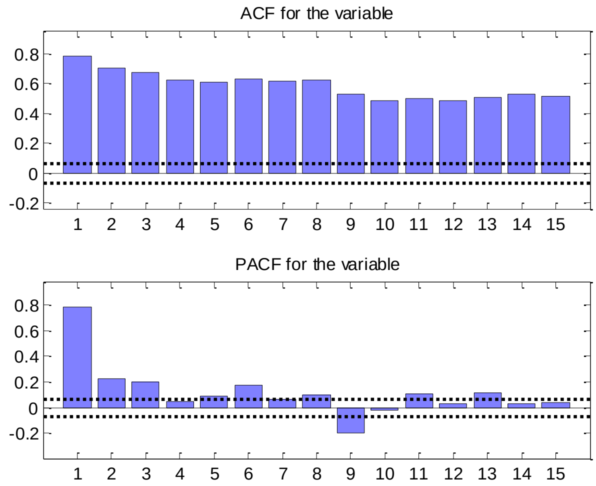
The time series clearly does not follow the Normal distribution.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t = 14.3579$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.0000	Yes
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = 93.8995$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 420.1486$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

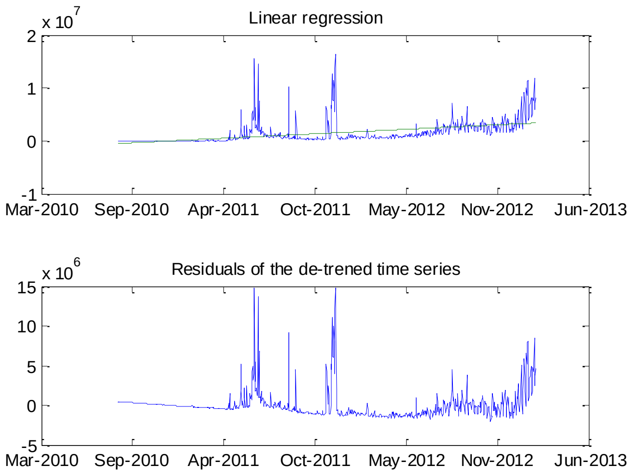
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	JB = 1.8534e+05	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.000	Yes

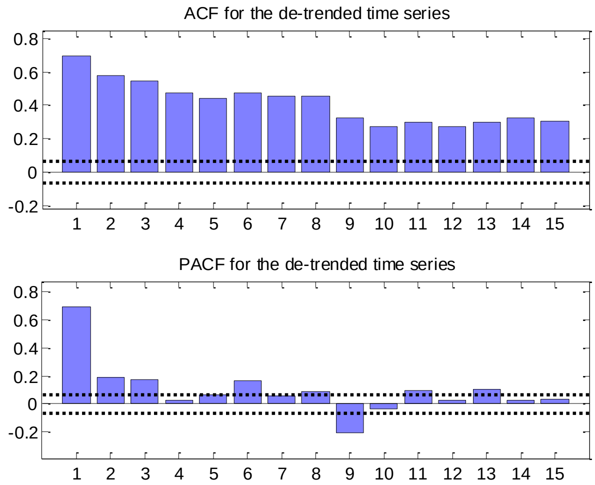
Variable Name: Estimated Transaction Volume (USD)





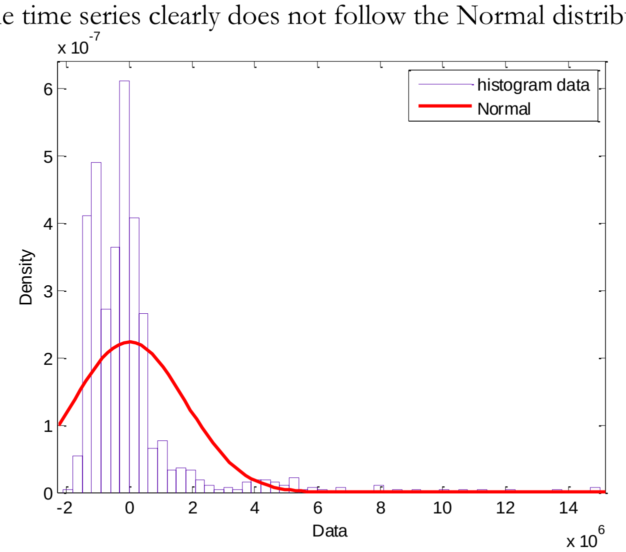
The time series is trend-stationary. If the trend is removed, the time series becomes:





	Augmented Dickey-Fuller Test		
	Test statistic	**p-value**	**Unit root?**
Variable	-3.8053	0.0000	No

The time series clearly does not follow the Normal distribution.



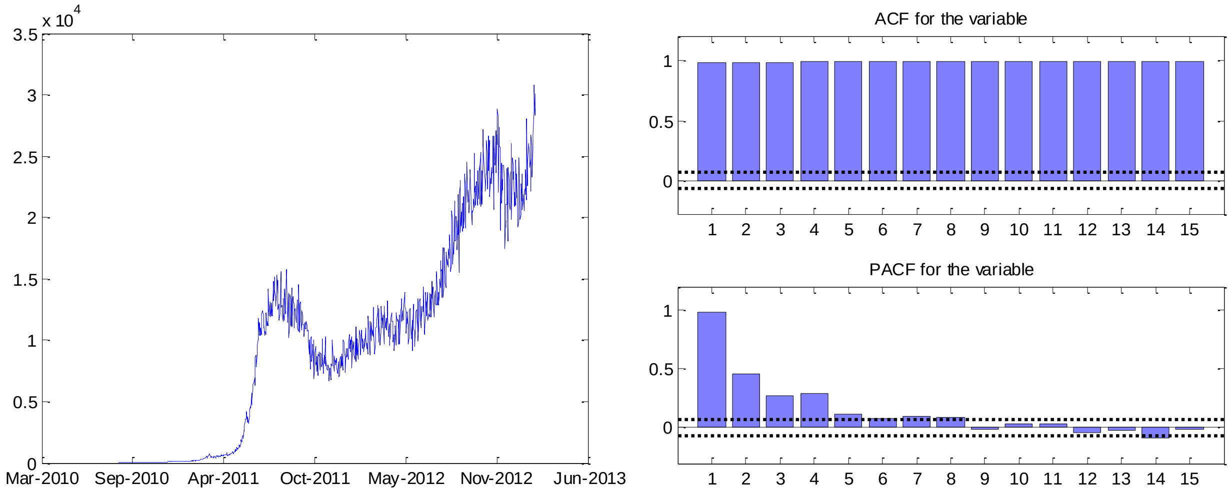
Summary statistics:	Variable
Mean	-7.1781e-10
Minimum	-2.0431e+06
25th Percentile	-9.9557e+05
Median	-2.7174e+05
75th Percentile	2.1599e+05
Maximum	1.4826e+07
Variance	3.1872e+12
Standard Deviation	1.7853e+06
Skewness	4.0748
Kurtosis	25.7220
Excess kurtosis	22.7220

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t =$ -1.2169e-14	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	1.0000	Yes
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s =$ 50.3474	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k =$ 140.3744	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

The tests confirm what has been suggested.

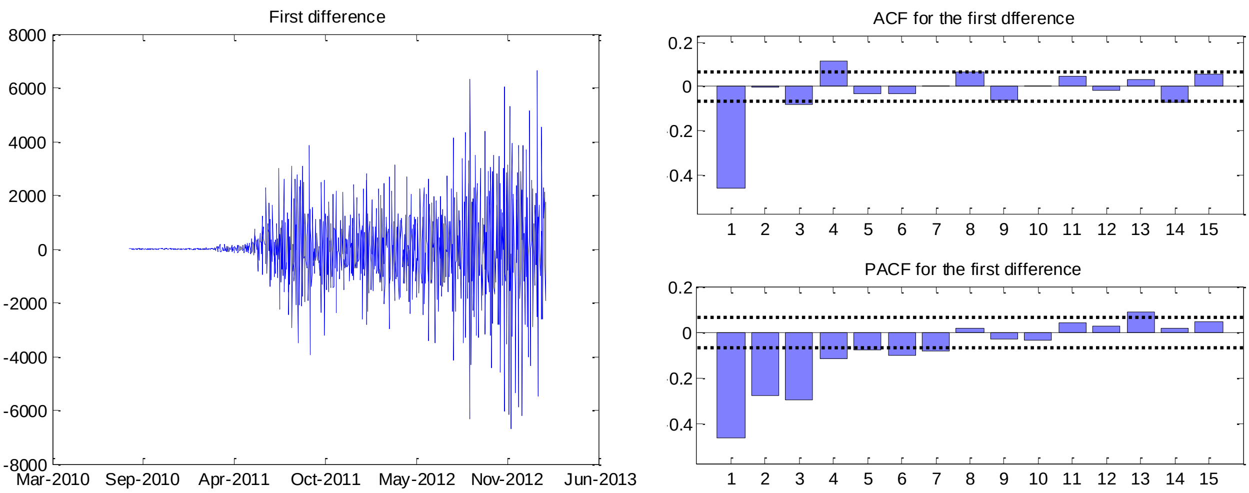
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	JB = 2.2240e+04	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.000	Yes

Variable Name: Hash Rate



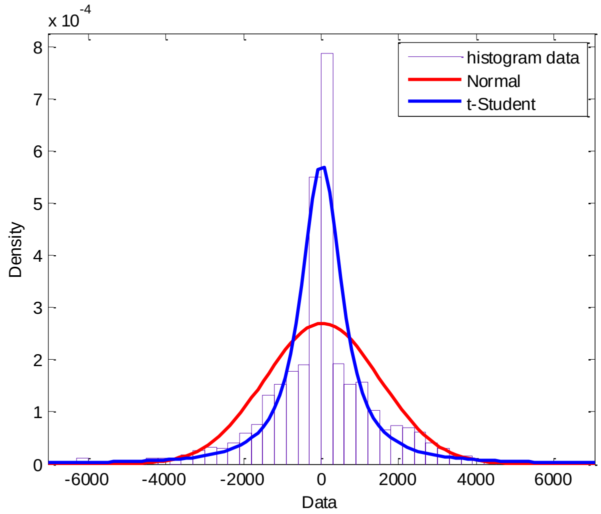
The hash rate seems to evolve in many different patterns across time. After some research, a possible explanation is that technology advancements (such as GPU mining in 2011 and the adoption of ASICs in early 2013) cause the hash rate to increase at a very high pace. The increase of popularity of the currency also explains the increase of the hash rate. After these sudden increases, mining becomes unprofitable for those using older technology, so miners leave until equilibrium is reached. Another factor that should be taken into consideration is the fact that the reward given to miners is halved every 210,000 blocks. When that happens, mining becomes less profitable so some miners may leave until the equilibrium is found.

A look at the first difference yields:



The variance seems to significantly grow with time. This fact could be explained by the fact that the amount of computational power increases with time, due to new technology and new miners. Because of this, the hash rate, overall, increases in orders of magnitude, so changes (increases due to the reasons mentioned and decreases due to readjustments) also increase. The autocorrelations suggest that the series is stationary, and the following test confirms it.

	Augmented Dickey-Fuller Test		
	Test statistic	p-value	Unit root?
First difference	-5.9156	0.0000	No



Summary statistics:	First difference
Mean	32.9229
Minimum	-6.6670e+03
25th Percentile	-524.4485
Median	0.8031
75th Percentile	582.9520
Maximum	6.6410e+03
Variance	2.1841e+06
Standard Deviation	1.4779e+03
Skewness	-0.1993
Kurtosis	6.6761
Excess kurtosis	3.6761

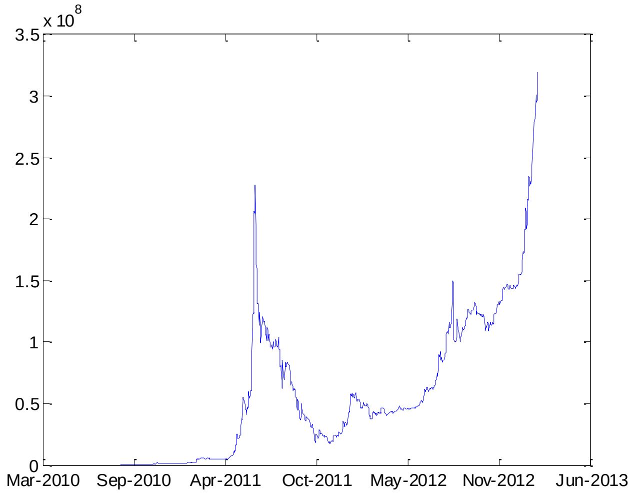
The first difference of this time series seems to have high excess kurtosis.

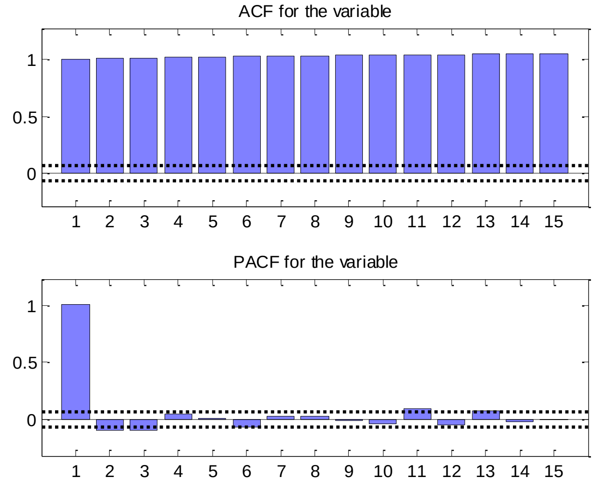
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	t = 0.6739	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.5004	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	t_s = -2.4613	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0138	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	t_k = 22.6983	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

The tests confirm what was inferred from the histogram. However, the sample also appears to have significant skewness. It does not behave like a Normal distribution.

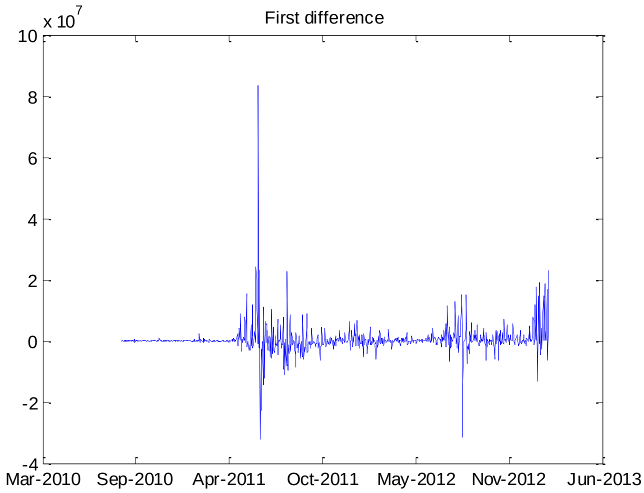
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	JB = 521.2712	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.000	Yes

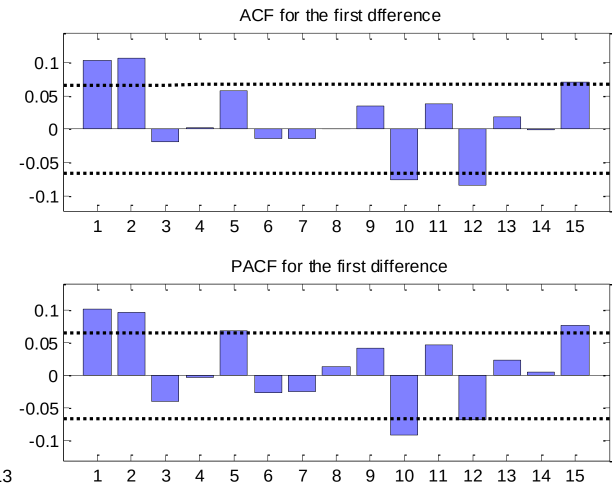
Variable Name: Market Capitalization





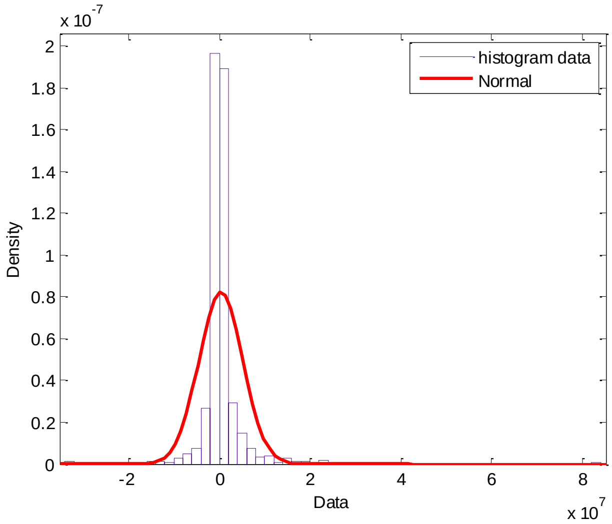
This time series is extremely close to that of the Exchange Rate ($/BTC). Therefore it behaves in the same way. It is also possible to see how the price reached its maximum in 2011 until it crashed due to panic in the market after the main exchange service (MtGox) had been hacked. Some months later, however, the market started recovering. The last peak is higher in this case, however, because the supply of bitcoins is higher – let's not forget that market capitalization is the bitcoin price multiplied by the total supply of bitcoins. Knowing this, there is evidence for structural breaks in the data. However, the data would have to be cut in 4 (3 structural breaks), leaving the first three subsets with a very low data size. For this reason, it has been decided not to proceed with this method but instead study the first difference of the whole sample.





The variance is not constant and there are some important spikes. The first ones clearly show how fast the market capitalization grew in mid-2011 and how quickly it crashed. There is no presence of a unit root in the data.

	Augmented Dickey-Fuller Test		
	Test statistic	p-value	Unit root?
First difference	-4.6097	0.0000	No



Summary statistics:	First difference
Mean	3.4929e+05
Minimum	-3.2291e+07
25th Percentile	-5.5897e+05
Median	1.0997e+03
75th Percentile	5.4833e+05
Maximum	8.3328e+07
Variance	2.3461e+13
Standard Deviation	4.8437e+06
Skewness	5.2935
Kurtosis	106.0691
Excess kurtosis	103.0691

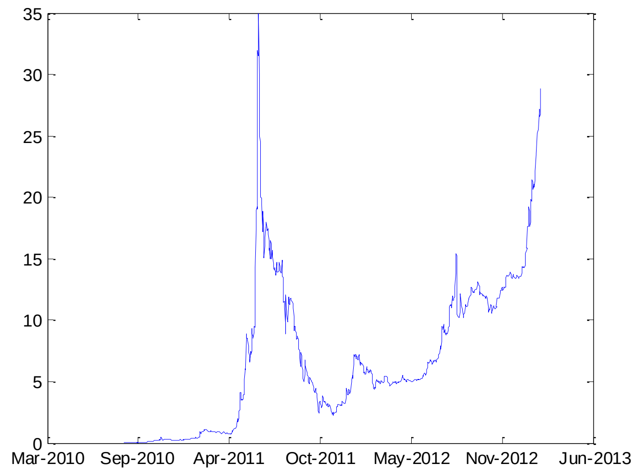
The studied sample has excess kurtosis and seems to also be skewed. Some of the kurtosis could be due to the first period (until Apr-2011) in which there was a very stable market capitalization with minor changes.

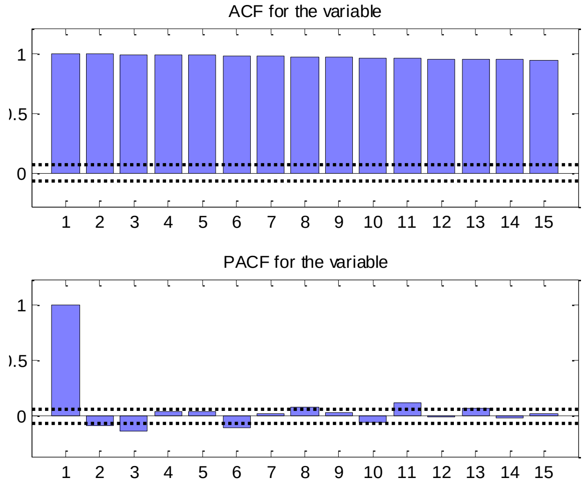
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t = 2.1813$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.0292	Yes
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = 65.3701$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 636.4048$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

The conducted tests confirm that all of the statistics are significant. The data does not behave like the Normal distribution.

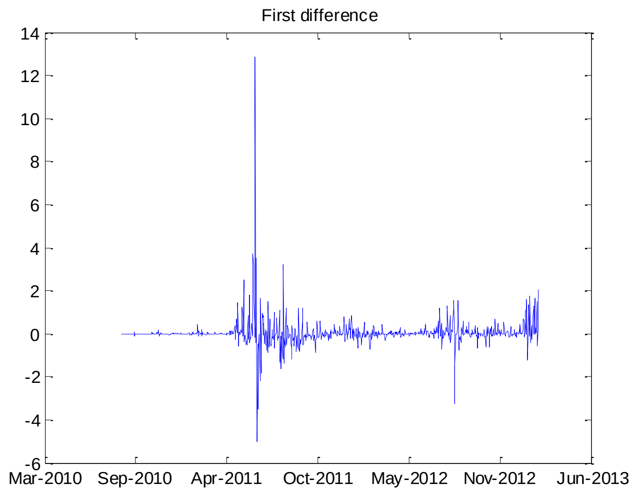
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k ≠ 3	JB = 4.0928e+05	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.000	Yes

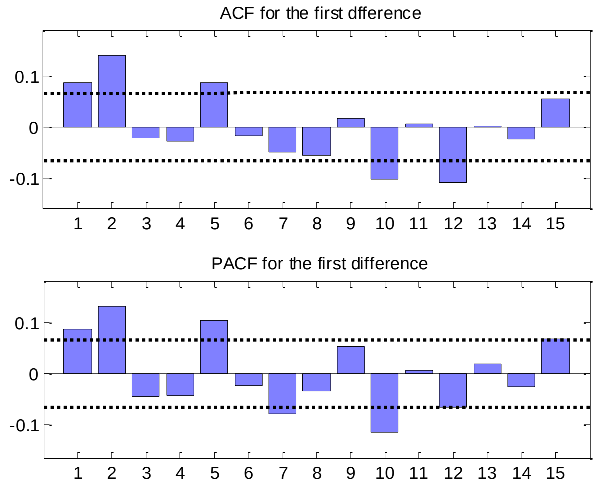
Variable Name: Market Price (Exchange Rate $/BTC)



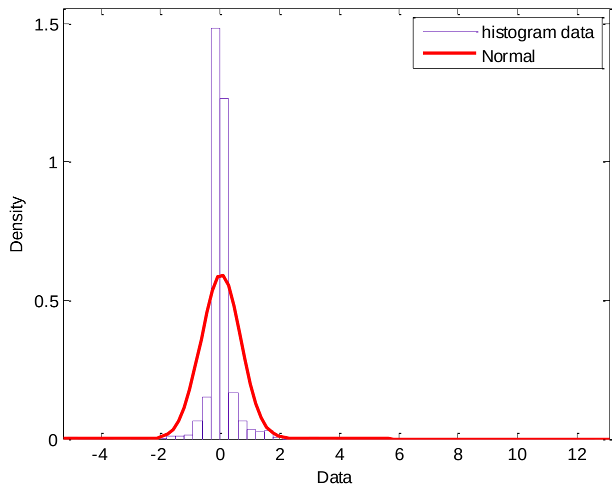


As had been suggested, this time series follows a similar pattern to that of Market Capitalization, so the analysis is going to be conducted in the same way.





In this case the spike of the sudden growth in mid-2011 is more noticeable. Such an excessive growth was followed by a crash.



<u>Summary statistics:</u>

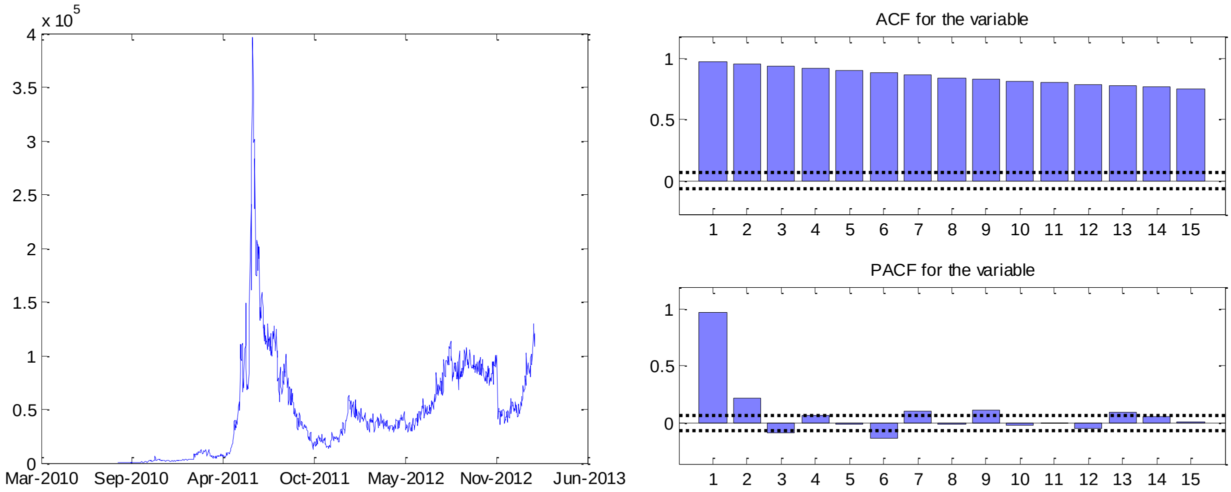
	First difference
Mean	3.4929e+05
Minimum	-3.2291e+07
25th Percentile	-5.5897e+05
Median	1.0997e+03
75th Percentile	5.4833e+05
Maximum	8.3328e+07
Variance	2.3461e+13
Standard Deviation	4.8437e+06
Skewness	5.2935
Kurtosis	106.0691
Excess kurtosis	103.0691

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t = 1.4148$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.1571	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = 87.8528$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 956.9008$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

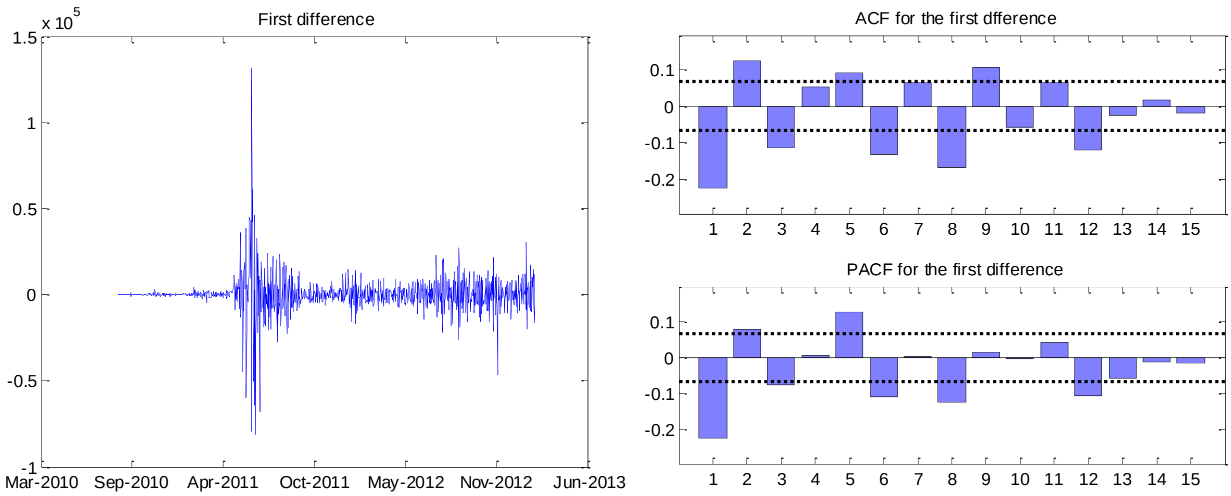
Although the mean is not significantly different from zero, the skewness and the excess kurtosis are extremely high.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	JB = 9.2338e+05	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.000	Yes

Variable Name: Miners' Revenue (USD)

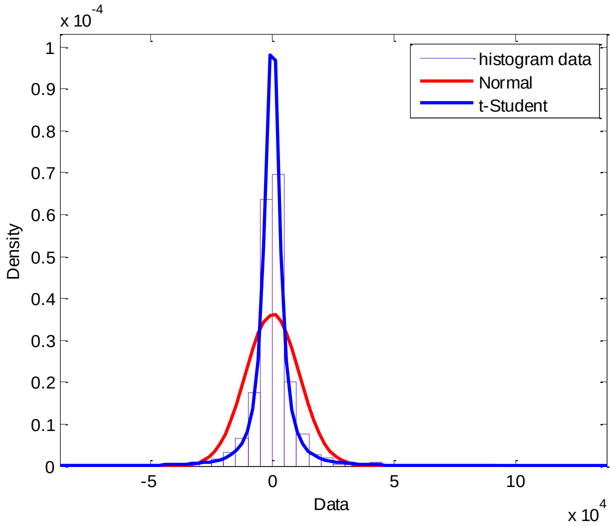


The mining revenue is the sum of the rewards given to miners each day in BTC multiplied by the market price that day. The total rewards given each day are approximately constant, and this value is halved every 210,000 blocks – currently, it takes many months for this to happen. Therefore, this time series should be very similar to that of the market price, except that it is halved at a certain point. Let's look at the first difference.



The first similar is also similar to the exchange rate, and doesn't have a unit root.

	Augmented Dickey-Fuller Test		
	Test statistic	**p-value**	**Unit root?**
First difference	-9.8992	0.0000	No



Summary statistics:	First difference
Mean	131.7083
Minimum	-8.1623e+04
25th Percentile	-2.8207e+03
Median	40.9850
75th Percentile	2.9754e+03
Maximum	1.3163e+05
Variance	1.2139e+08
Standard Deviation	1.1018e+04
Skewness	0.6501
Kurtosis	37.5614
Excess kurtosis	34.5614

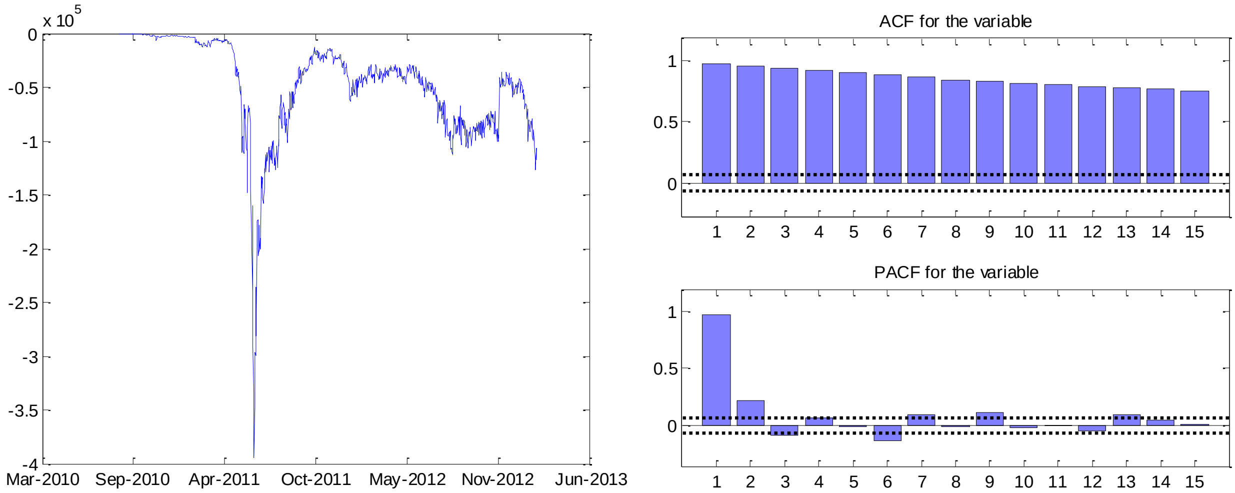
The first difference has a high kurtosis, but not as much as the t-Student distribution would require. It clearly does not follow the Normal distribution.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t = 0.3616$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.7176	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = 8.0287$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	8.8818e-16	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 213.4010$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

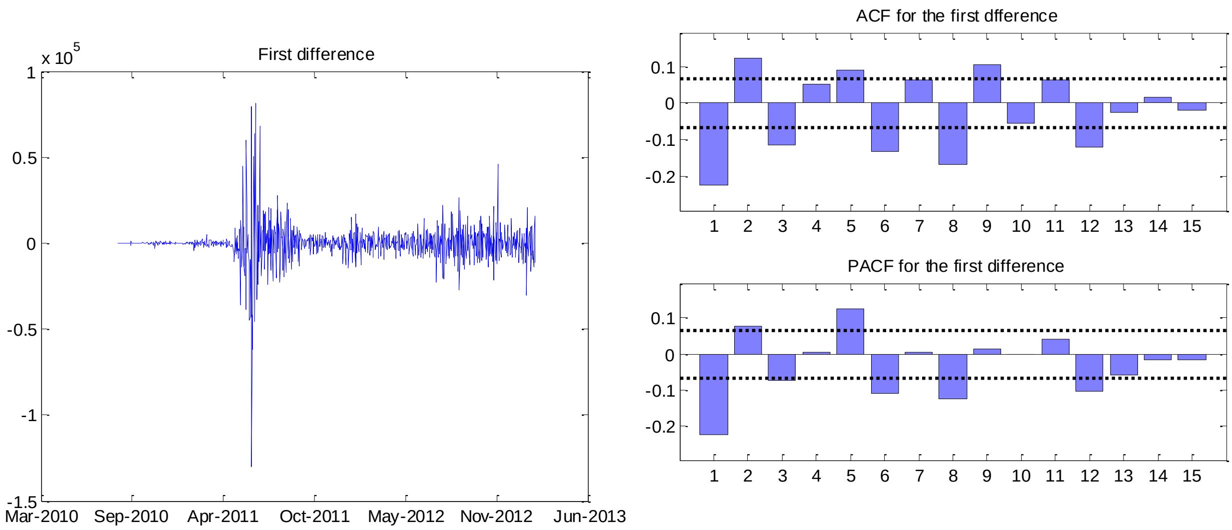
The Jarque-Bera test is used to confirm this.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or $k \neq 3$	JB = 4.5604e+04	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.000	Yes

Variable Name: Network Deficit

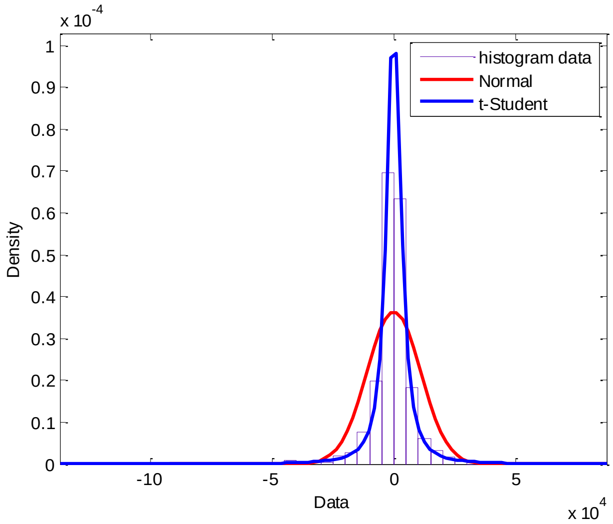


The network deficit shows the difference between transaction fees and the miners' income from block rewards. Remember that miners' revenue comes from both transaction fees and block rewards, the latter being the most significant part. It is expected, for now, that transaction fees do not cover the cost of mining (as they are relatively small compared to the block rewards). However, as time passes and the reward gets smaller (remember that it is halved every 210,000 blocks), transaction fees will become of increasing importance, and the network deficit will become positive at a certain point. Let's have a look at the first difference.



Once again, although the series seems to converge around zero, it is possible to see how the variance changes with time.

	Augmented Dickey-Fuller Test		
	Test statistic	p-value	Unit root?
First difference	-9.8879	0.0000	No



Summary statistics:	First difference
Mean	-128.5431
Minimum	-1.3021e+05
25th Percentile	-3.0148e+03
Median	-40.9070
75th Percentile	2.8141e+03
Maximum	8.1186e+04
Variance	1.2027e+08
Standard Deviation	1.0967e+04
Skewness	-0.6108
Kurtosis	37.1622
Excess kurtosis	34.1622

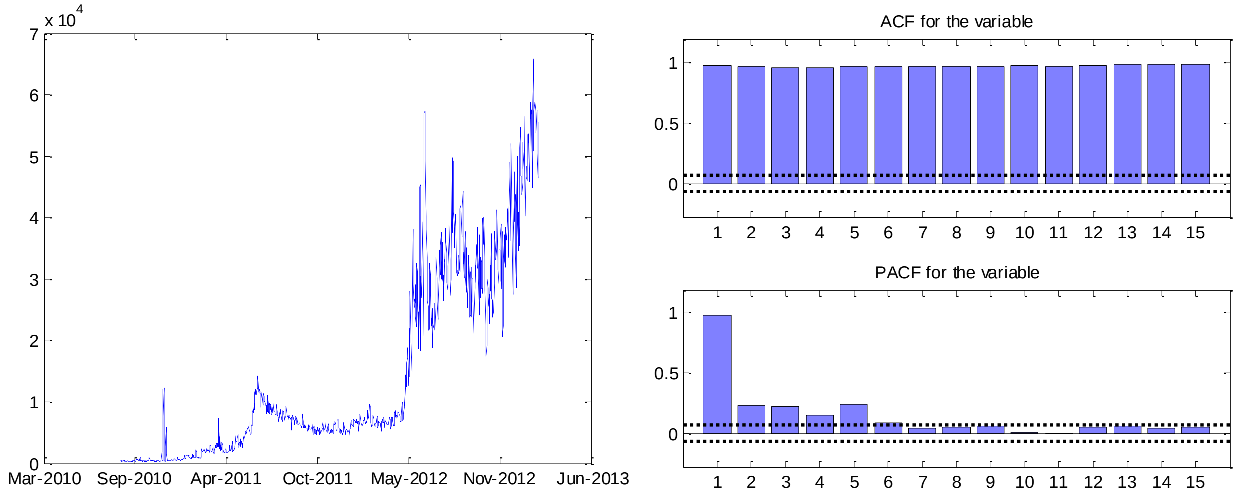
The excess kurtosis seems sufficient in order to determine that the variable does not behave like a normal distribution. There seems to be some skewness as well.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	t = -0.3546	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.7229	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	t_s = -7.5429	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	4.5963e-14	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	t_k = 210.9361	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

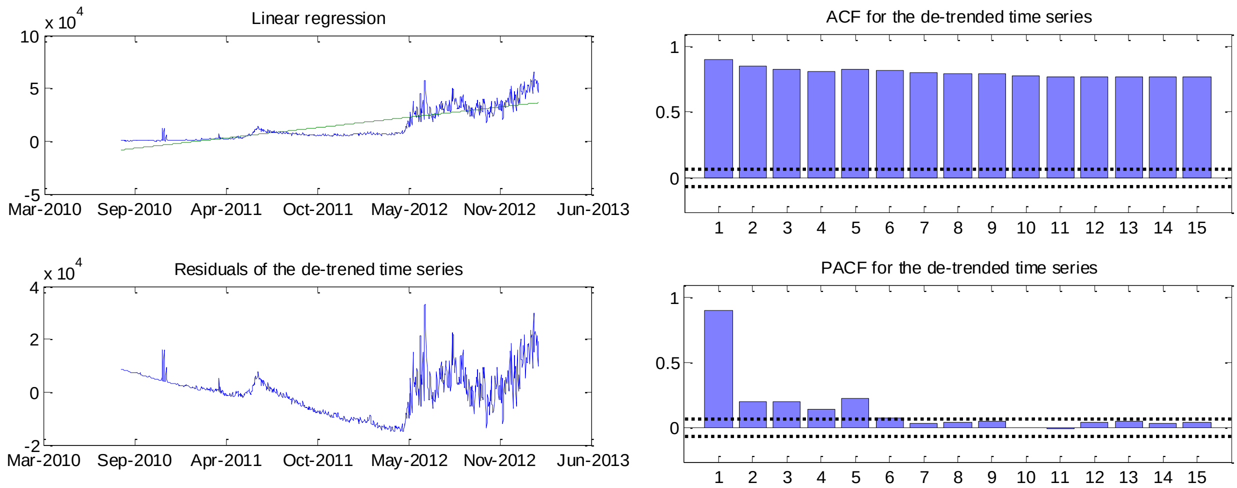
The tests confirm our hypotheses.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	JB = 4.4551e+04	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.000	Yes

Variable Name: Number of Transactions

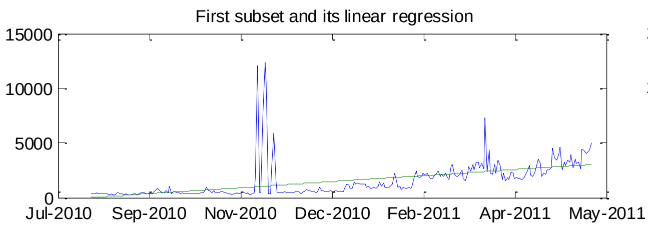


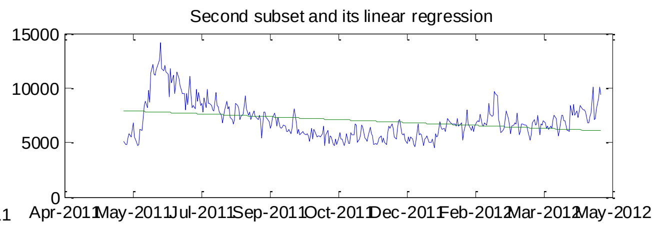
The number of daily transactions has increased significantly over the course of time. Let's see if the de-trended series is stationary.

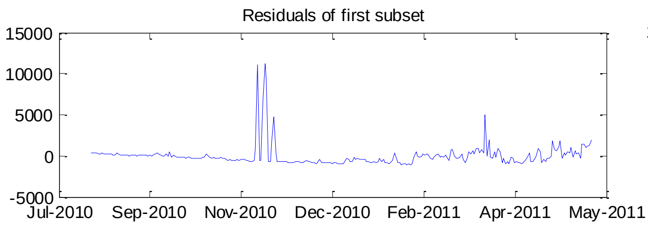


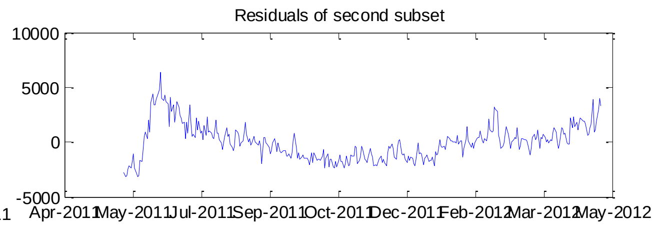
	Augmented Dickey-Fuller Test		
	Test statistic	**p-value**	**Unit root?**
First difference	-1.4265	0.1436	Yes

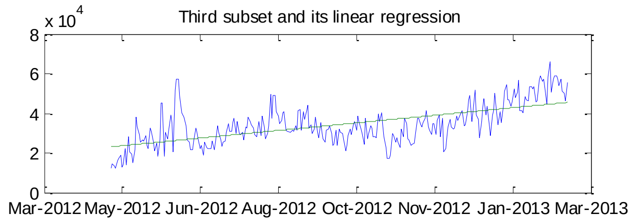
The hypothesis that there is a unit root cannot be rejected. Looking at the graphs again, it seems that there could be two structural breaks, as there seem to be unexpected shifts to its on 18-May-2011 (day 275 in the variable 'date') and on 02-May-2012 (day 625 in the variable 'date'). Let's split the data and conduct Chow tests.



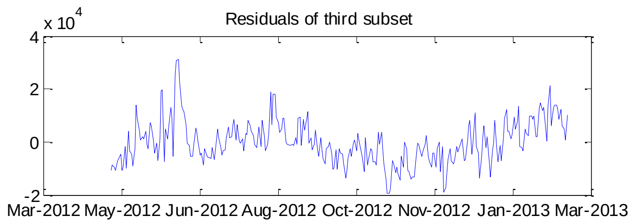




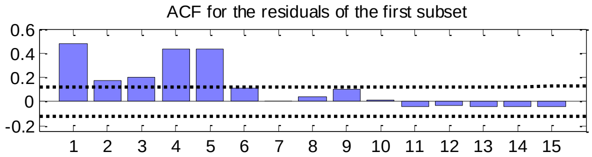


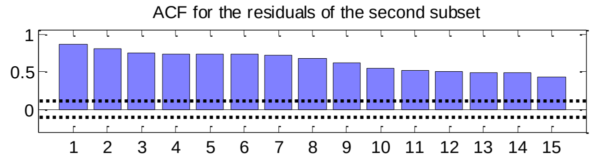


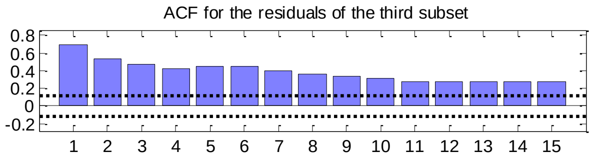
Chow Test		
Test statistic	**p-value**	**Break?**
F = 330.4316	0.0000	Yes
F = 372.0319	0.0000	Yes



The Chow test confirms the existence of the structural breaks. Furthermore, the autocorrelations of the first and third subset suggest stationarity (see below). A further test would be needed to decide about the second subset, and the Augmented Dickey-Fuller test confirms that there is a significant unit root in it.

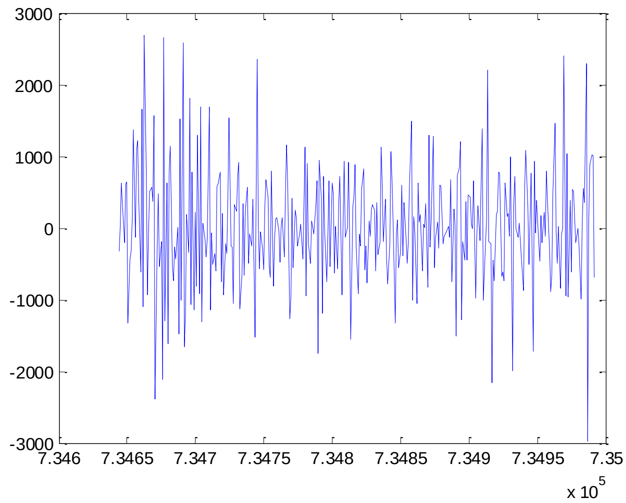


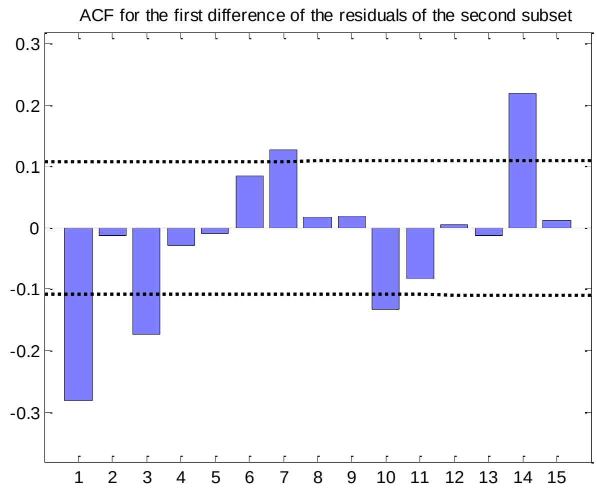




	Augmented Dickey-Fuller Test		
	Test statistic	**p-value**	**Unit root?**
First subset	-4.4808	0.0000	No
Second subset	-1.2007	0.2124	Yes
Third subset	-3.4613	0.0000	No

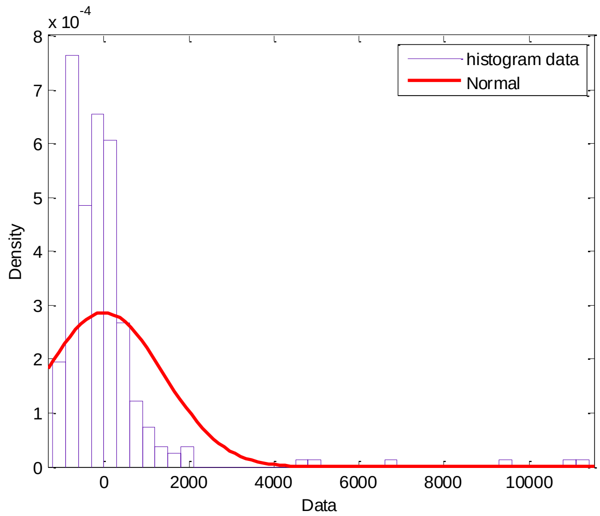
The first difference of the second subset is stationary.

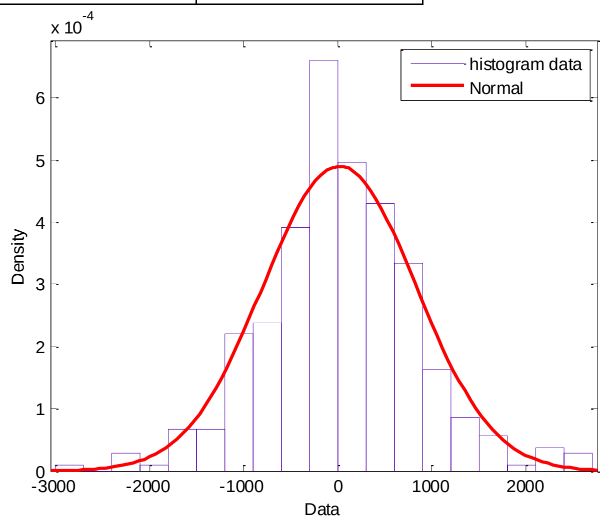


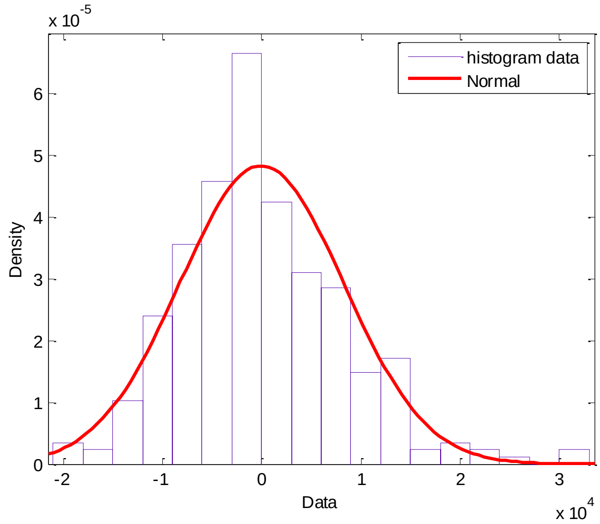


Even though analysing the subsets by taking the difference of only one of them may seem strange due to lack of comparability, this paper analyses the stationarized time series, therefore this is performed next.

Summary statistics:	First subset	Second subset (first difference)	Third subset
Mean	1.4668e-12	17.4574	6.5509e-12
Minimum	-1.1024e+03	-2.9717e+03	-1.9420e+04
25th Percentile	-684.5893	-473.6974	-5.4118e+03
Median	-222.1955	-10.6974	-879.7773
75th Percentile	177.8509	511.8026	4.7991e+03
Maximum	1.1267e+04	2.6953e+03	3.0965e+04
Variance	1.9372e+06	6.6308e+05	6.8240e+07
Standard Deviation	1.3918e+03	814.2986	8.2607e+03
Skewness	5.5989	0.1545	0.5973
Kurtosis	40.9127	4.2466	3.9551
Excess kurtosis	37.9127	1.2466	0.9551





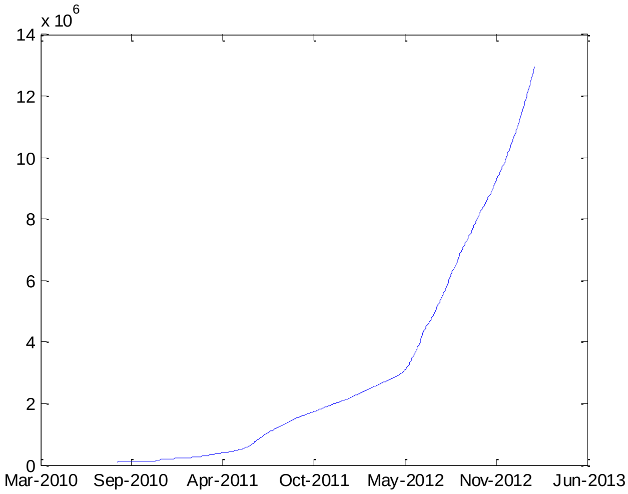


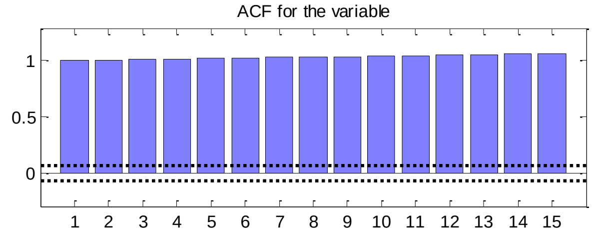
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t_1 = 0.0000$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	1.0000	No
		$t_2 = 0.4005$		0.6888	No
		$t_3 = 0.0000$		1.0000	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_{s,1} = 37.9047$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
		$t_{s,2} = 1.1783$		0.2387	No
		$t_{s,3} = 4.1599$		0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_{k,1} = 128.3350$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes
		$t_{k,2} = 4.7539$		1.9953e-06	Yes
		$t_{k,3} = 3.3257$		8.8184e-04	Yes

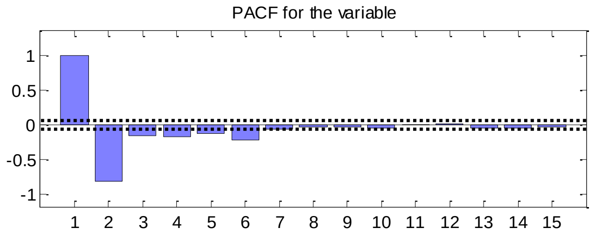
None of the variables seem to significantly behave like the Normal distribution.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	$JB_1 = 1.7907e{+}04$	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.0000	Yes
		$JB_2 = 23.9881$		6.1809e-06	Yes
		$JB_3 = 28.3652$		6.9274e-07	Yes

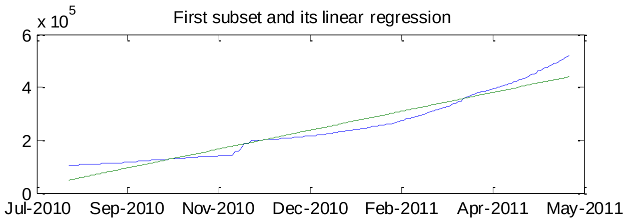
Variable Name: Number of Transactions - Cumulative

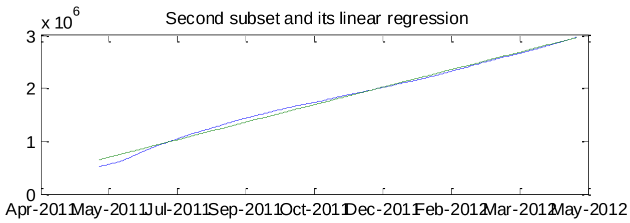


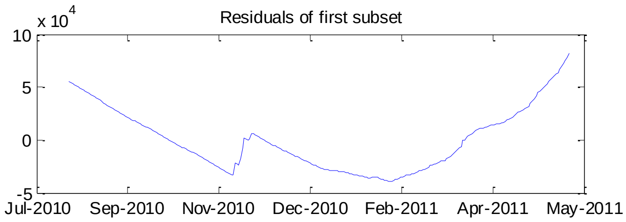


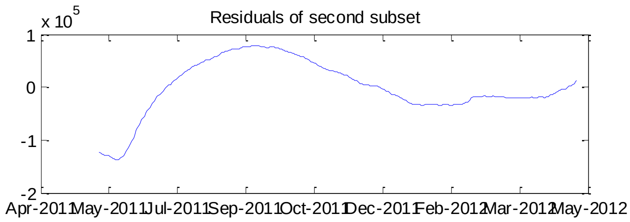


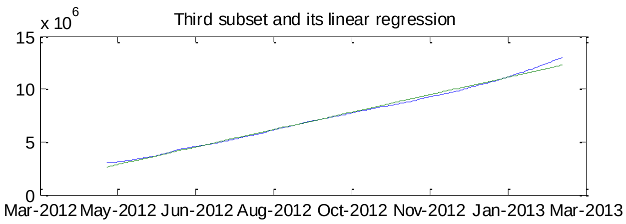
This variable is clearly not stationary. However, there seem to be unexpected shifts to its on 18-May-2011 (day 275 in the variable 'date') and on 02-May-2012 (day 625 in the variable 'date'). In order to test whether or not structural breaks exist, the data is split in three subsamples and two Chow tests are performed in order to determine if the coefficients among the different regressions of the subsamples can be considered significantly different.



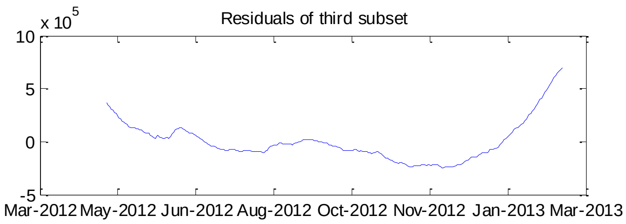




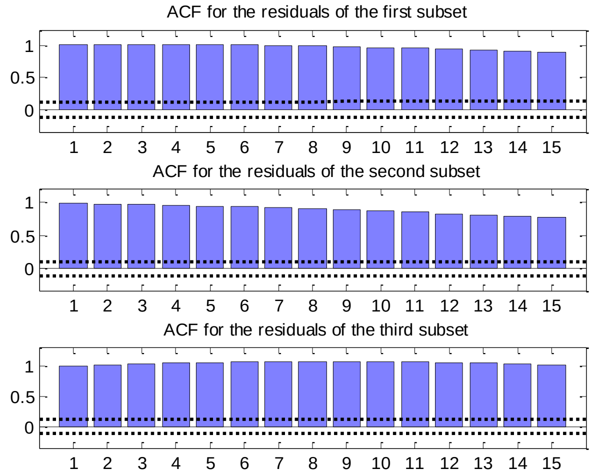




Chow Test			
Test statistic		p-value	Break?
F =	8.1570e+03	0.0000	Yes
F =	2.6276e+04	0.0000	Yes

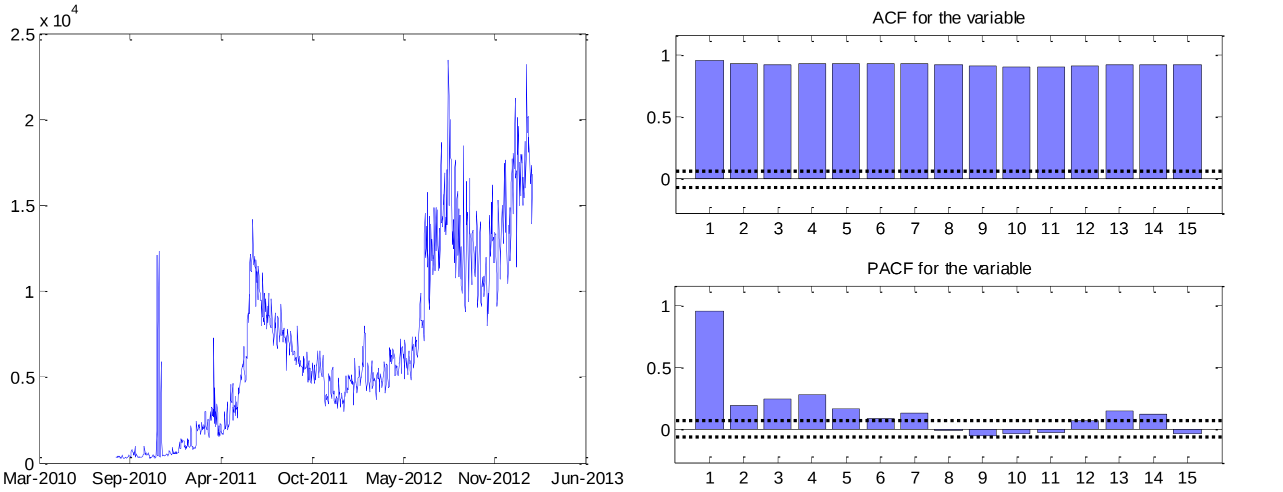


The Chow tests confirm the existence of the structural breaks. However, the subsets are not stationary:

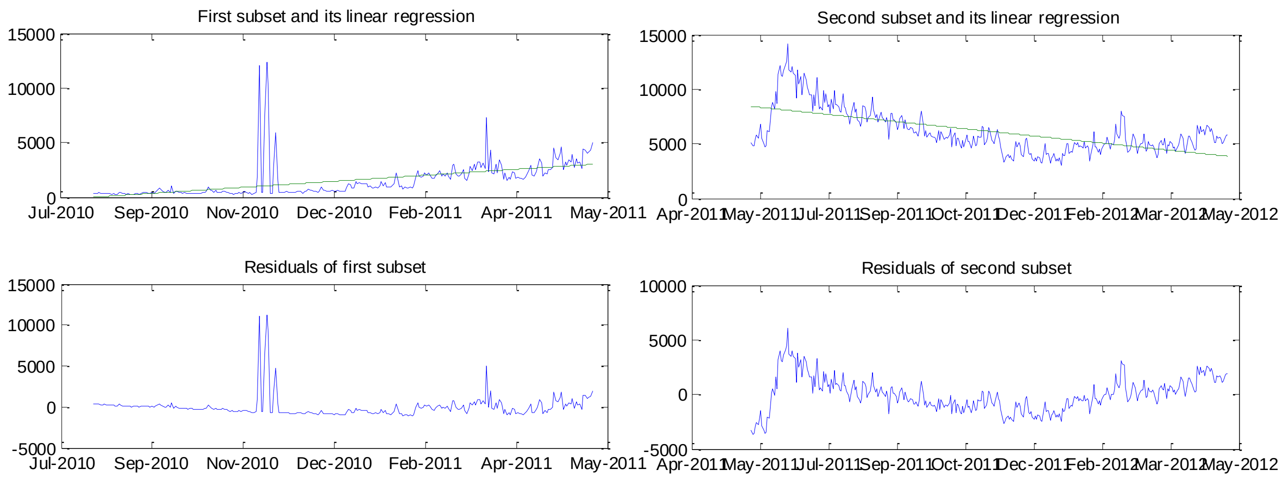


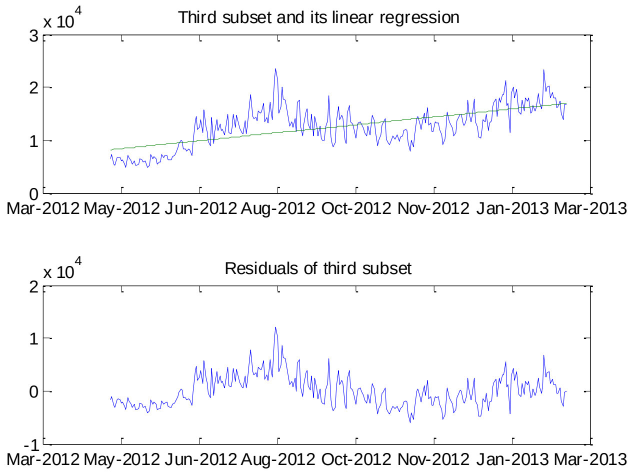
Taking the first difference for each subset, the result is exactly as the time series studied before this one, Number of Transactions.

Variable Name: Number of Transactions – Excluding popular addresses

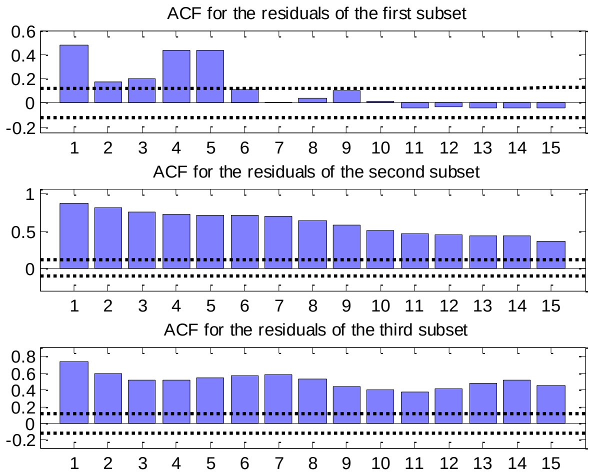


The behaviour of this time series is close to that of the series Number of Transactions- it seems to have the same structural breaks on 18-May-2011 (day 275 in the variable 'date') and on 02-May-2012 (day 625 in the variable 'date').



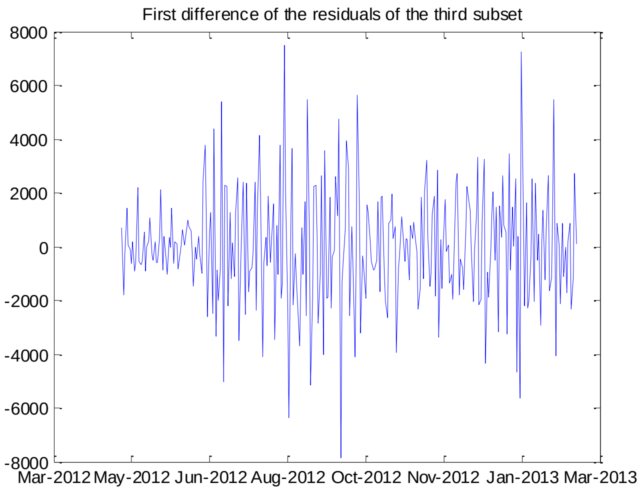


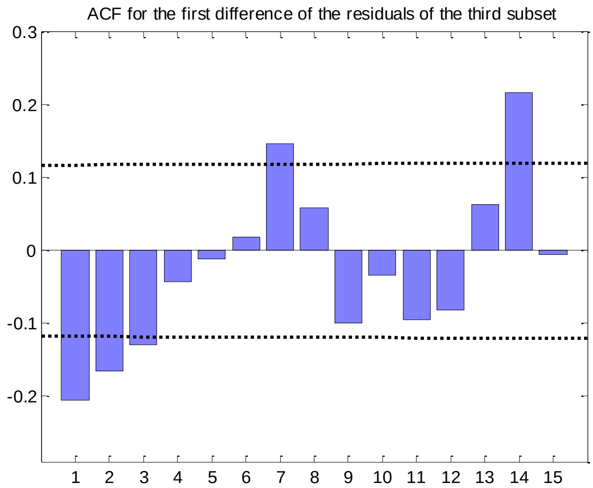
Chow Test		
Test statistic	p-value	Break?
F = 509.5650	0.0000	Yes
F = 347.7108	0.0000	Yes



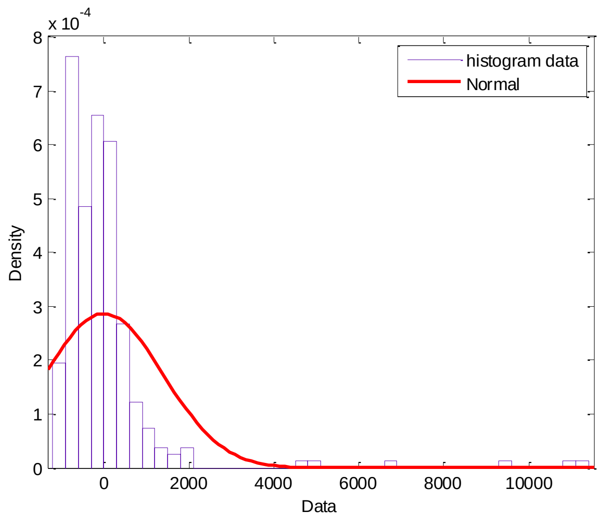
	Augmented Dickey-Fuller Test		
	Test statistic	**p-value**	**Unit root?**
First subset	-4.4808	0.0000	No
Second subset	-2.7224	0.0070	No
Third subset	-1.7128	0.0835	Yes

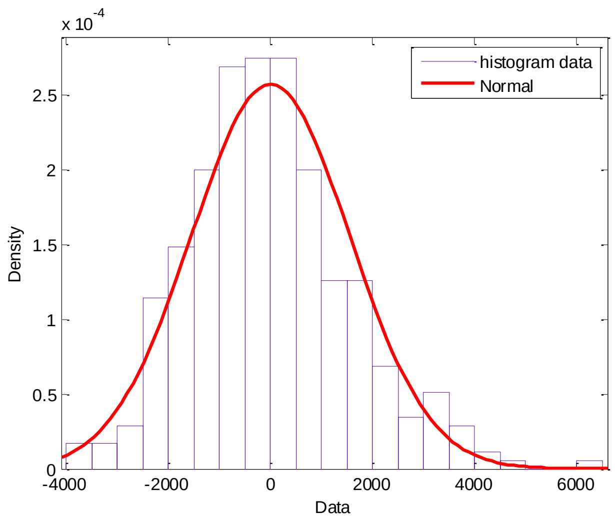
The first two subsets appear to be stationary. For the third one, however, the Augmented Dickey Fuller test cannot reject the hypothesis of a unit root. For the third subset, the first difference will be taken in order to make it stationary.

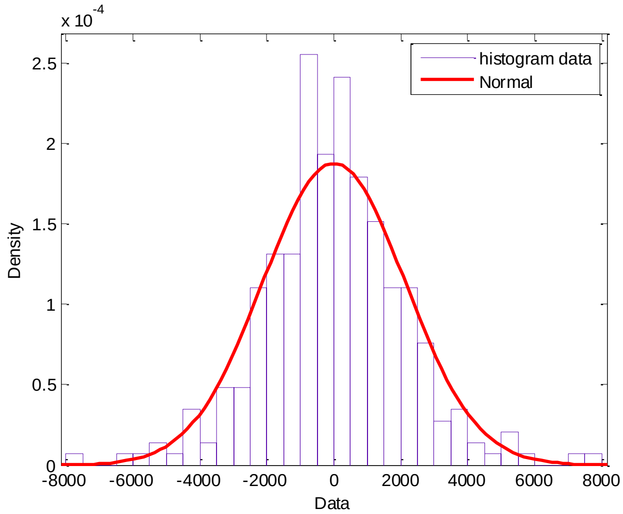




Summary statistics:	First subset	Second subset	Third subset (first difference)
Mean	1.4668e-12	-1.6345e-12	5.4700
Minimum	-1.1024e+03	-3.6148e+03	-7.8652e+03
25th Percentile	-684.5893	-1.0190e+03	-1.1472e+03
Median	-222.1955	-89.5127	-11.7473
75th Percentile	177.8509	801.4026	1.2908e+03
Maximum	1.1267e+04	6.0946e+03	7.5148e+03
Variance	1.9372e+06	2.4111e+06	4.5225e+06
Standard Deviation	1.3918e+03	1.5528e+03	2.1266e+03
Skewness	5.5989	0.4877	0.0675
Kurtosis	40.9127	3.5663	4.3326
Excess kurtosis	37.9127	0.5663	1.3326





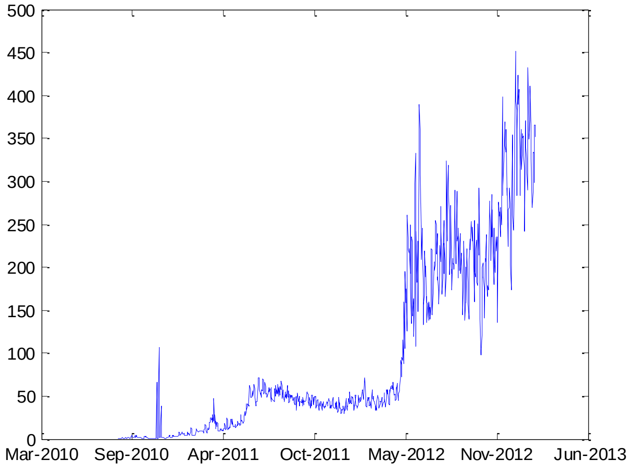


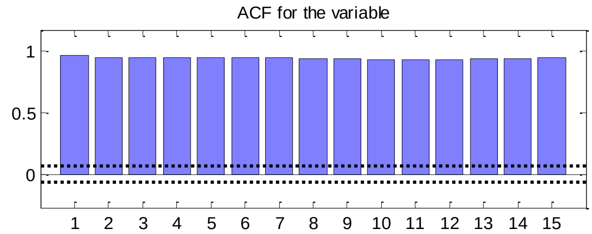
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t_1 = 0.0000$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	1.0000	No
		$t_2 = 0.0000$		1.0000	No
		$t_3 = 0.0438$		0.9651	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_{s,1} = 37.9047$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
		$t_{s,2} = 3.7249$		0.0002	Yes
		$t_{s,3} = 0.4693$		0.6388	No
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_{k,1} = 128.3350$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes
		$t_{k,2} = 2.1625$		1.9953e-06	Yes
		$t_{k,3} = 4.6323$		8.8184e-04	Yes

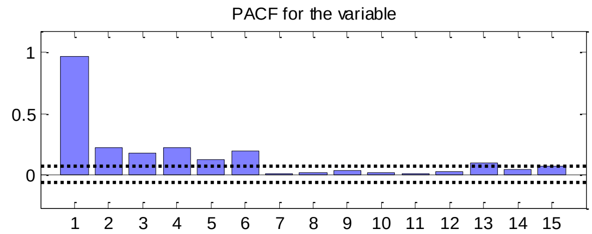
None of the variables can be considered to behave like the Normal distribution. Noteworthy is the fact that all the test statistics from the first subsets are identical to the time series Number of Transactions. In fact, the plot is exactly the same. Therefore, the popular addresses did not operate before the first cut.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or $k \neq 3$	$JB_1 = 1.7907e+04$	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.0000	Yes
		$JB_2 = 18.5513$		9.3677e-05	Yes
		$JB_3 = 21.6789$		1.9611e-05	Yes

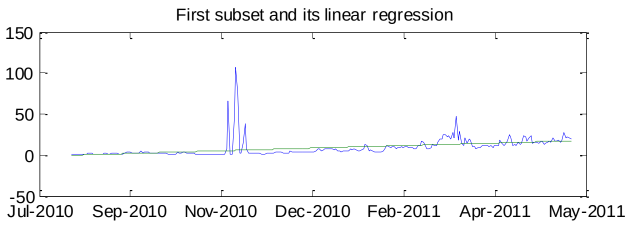
Variable Name: Number of Transactions per Block

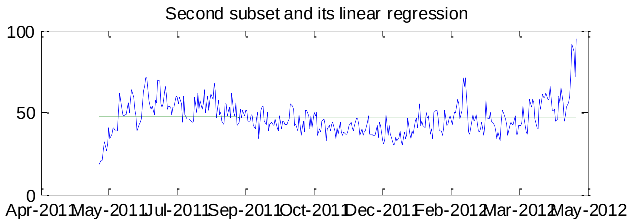


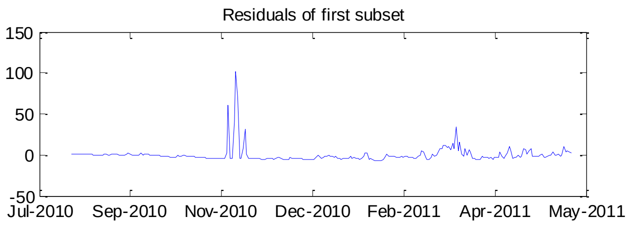


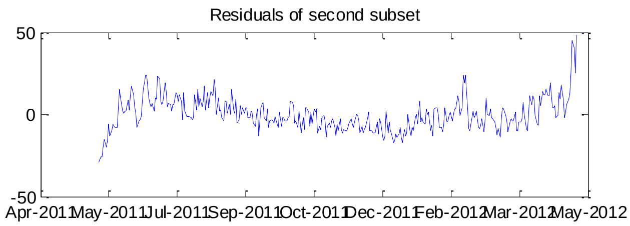


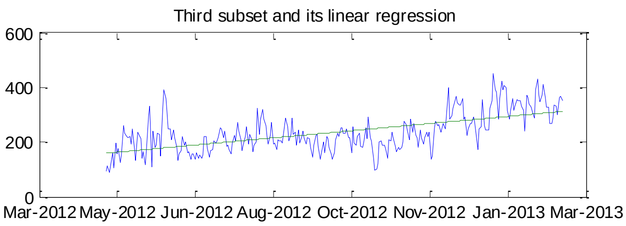
Once again, there seem to be a couple of structural breaks, on 18-May-2011 (day 275 in the variable 'date') and on 02-May-2012 (day 625 in the variable 'date').



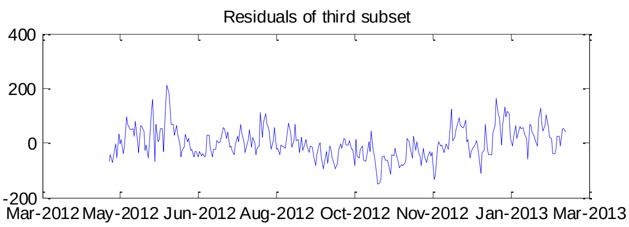




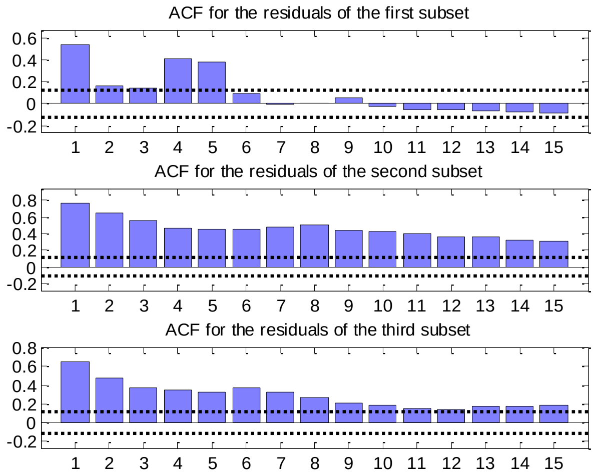




Chow Test		
Test statistic	**p-value**	**Break?**
F = 227.7205	0.0000	Yes
F = 331.9541	0.0000	Yes

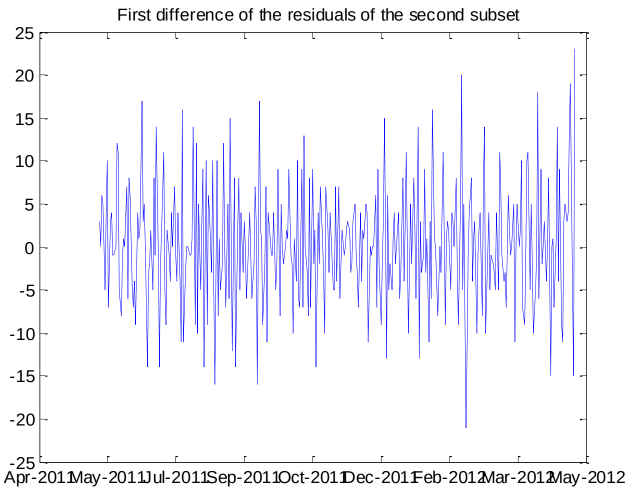


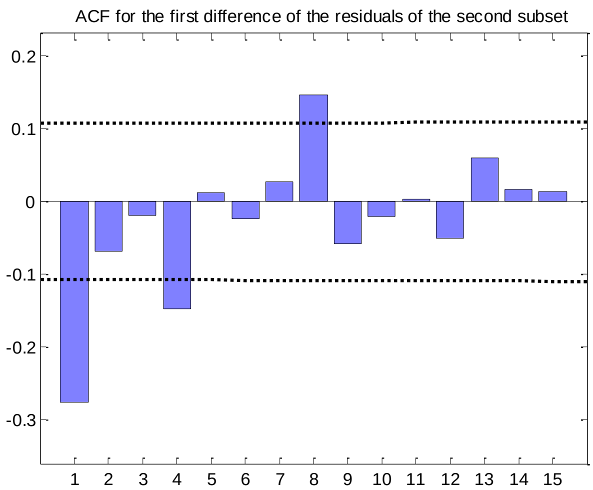
The Chow test confirms our hypothesis about the existence of structural breaks.



	Augmented Dickey-Fuller Test		
	Test statistic	**p-value**	**Unit root?**
First subset	-4.8662	0.0000	No
Second subset	-1.4678	0.1343	Yes
Third subset	-3.4809	0.0000	No

Although the autocorrelations could suggest that the three time series are stationary, the Augmented Dickey-Fuller test cannot reject the hypothesis of a unit root in the second subset.

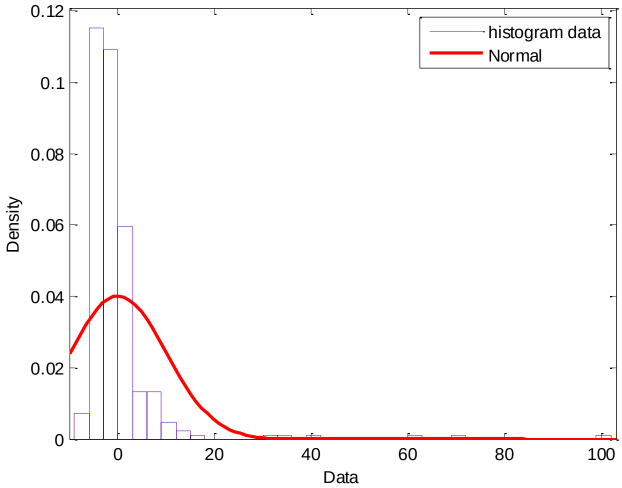


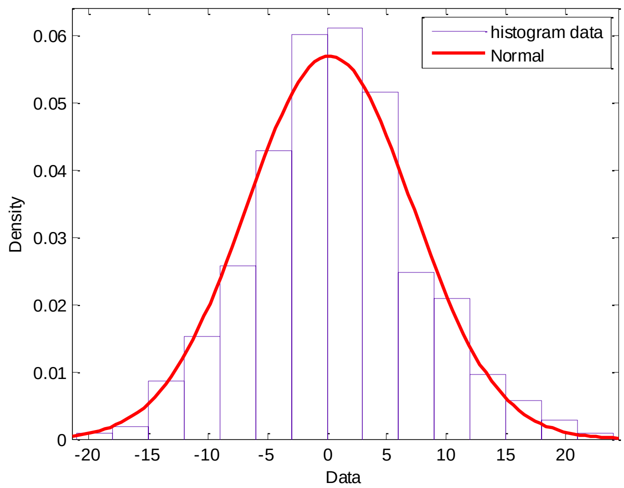


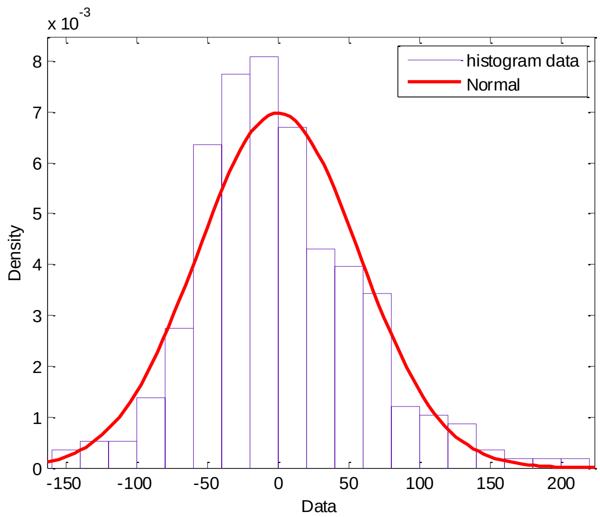
Let's study the first and third subsets along with the difference of the second one (so all are stationary).

Summary statistics:

	First subset	Second subset (first difference)	Third subset
Mean	0.0000	0.2213	3.1254e-14
Minimum	-6.7875	-20.9994	-151.6454
25th Percentile	-4.0231	-4.2494	-38.2230
Median	-1.9155	0.0006	-8.0829
75th Percentile	0.5191	4.0006	34.4080
Maximum	101.3304	23.0006	211.1239
Variance	98.6251	48.9253	3.2679e+03
Standard Deviation	9.9310	6.9947	57.1653
Skewness	6.5493	0.1874	0.4818
Kurtosis	55.9368	3.2573	3.7133
Excess kurtosis	52.9368	0.2573	0.7133





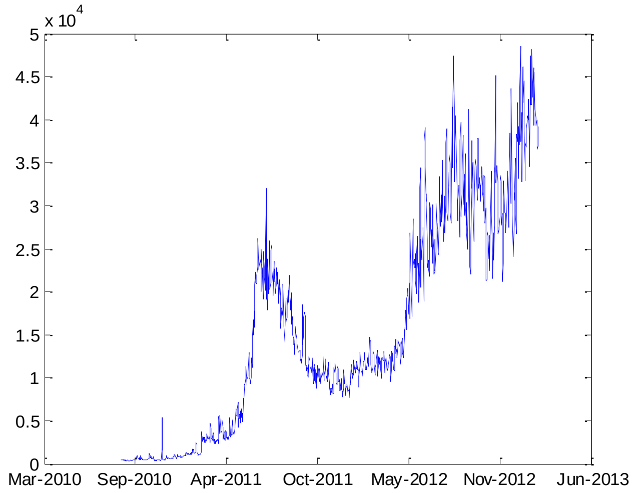


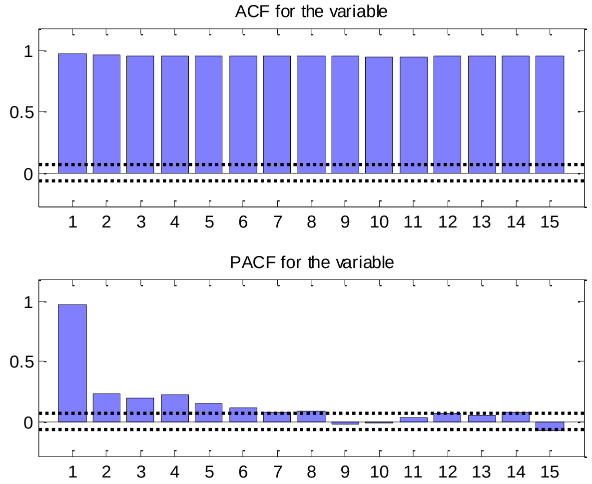
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t_1 = 0.0000$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	1.0000	No
		$t_2 = 0.5909$		0.5546	No
		$t_3 = 0.0000$		1.0000	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_{s,1} = 44.3389$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
		$t_{s,2} = 1.4291$		0.1530	No
		$t_{s,3} = 3.3552$		0.0008	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_{k,1} = 179.1920$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes
		$t_{k,2} = 0.9813$		0.3264	No
		$t_{k,3} = 2.4837$		0.0130	Yes

The first and third subsets do not seem to behave like a Normal distribution. The second one however – to which the first difference was taken in order to make it stationary – is close. A Jarque-Bera test cannot reject it.

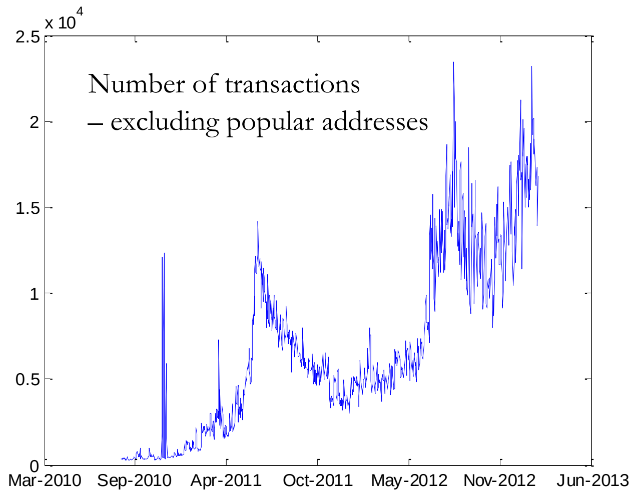
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or $k \neq 3$	$JB_1 =$ 3.4076e+04	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.0000	Yes
		$JB_2 =$ 3.0055		0.2225	No
		$JB_3 =$ 17.4259		1.6444e-04	Yes

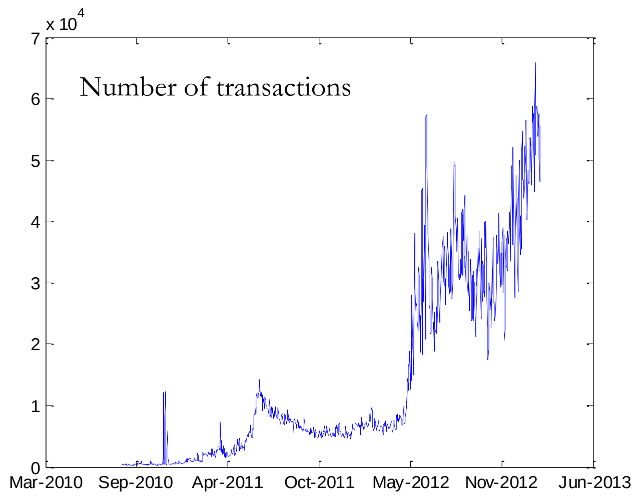
Variable Name: Number of Unique Addresses Used



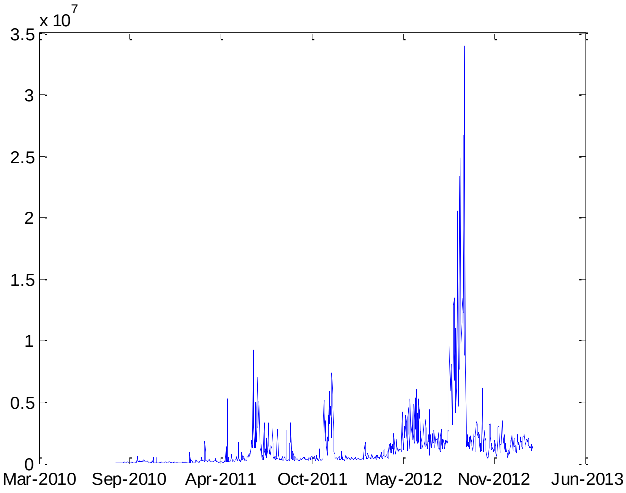


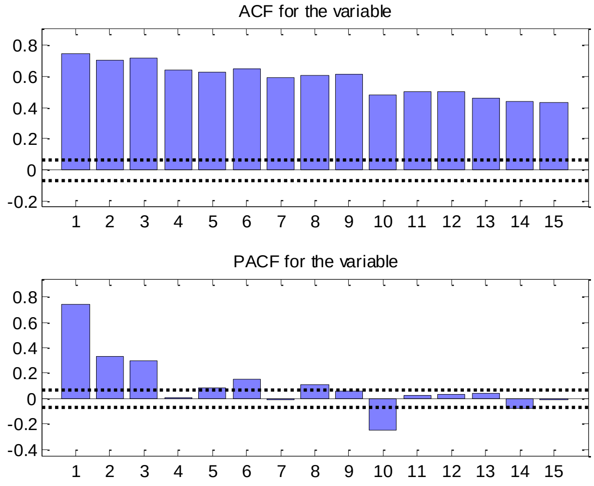
This time series is extremely close to the previous one (Number of transactions – excluding popular addresses). In order to ease comparison, it is included below, along with the time series Number of transactions. The behaviour is noticeably more similar to the former, due to the fact that that popular addresses' first transaction is on 22-Jul-2011 (at that date the operation 'Number of transactions – exclusing popular addresses' – 'Number of transactions' yields the first non-zero result).





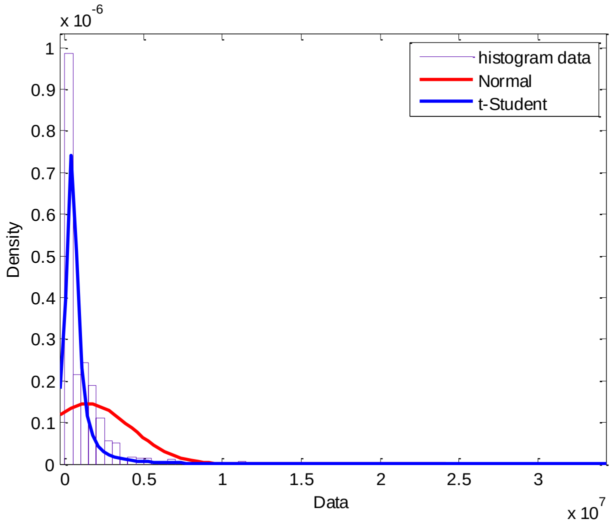
Variable Name: Output Volume (BTC)





After looking at the plot and the autocorrelations, this time series looks stationary. An Augmented Dickey-Fuller test rejects the null hypothesis of a unit root.

	Augmented Dickey-Fuller Test		
	Test statistic	p-value	Unit root?
Variable	-2.9162	0.0041	No



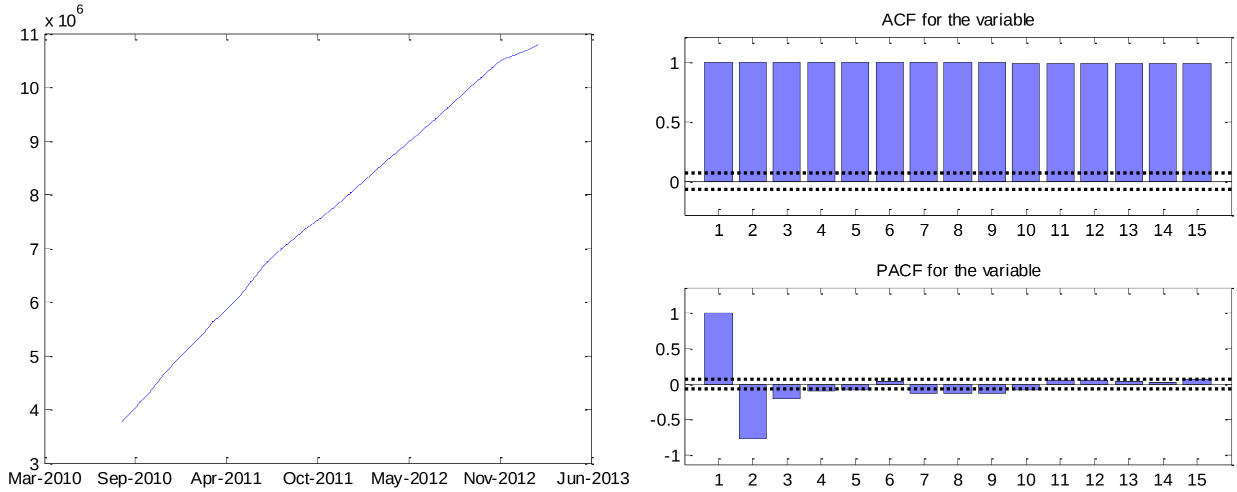
Summary statistics:	First difference
Mean	1.4232e+06
Minimum	1.7623e+04
25th Percentile	2.1235e+05
Median	5.2573e+05
75th Percentile	1.6522e+06
Maximum	3.3956e+07
Variance	7.5669e+12
Standard Deviation	2.7508e+06
Skewness	5.8562
Kurtosis	49.4104
Excess kurtosis	46.4104

The variable has high kurtosis and is highly skewed, as the output can only be a positive number. It does not behave like a normal distribution. The following tests confirm this.

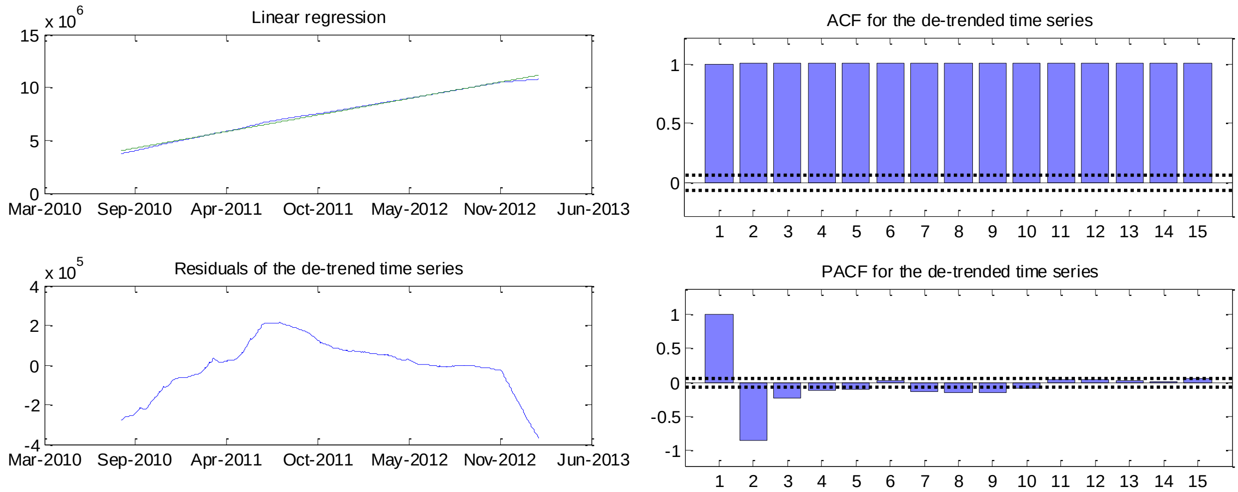
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t = 15.6587$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.0000	Yes
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = 72.3584$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 286.7195$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	JB = 8.7444e+04	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.000	Yes

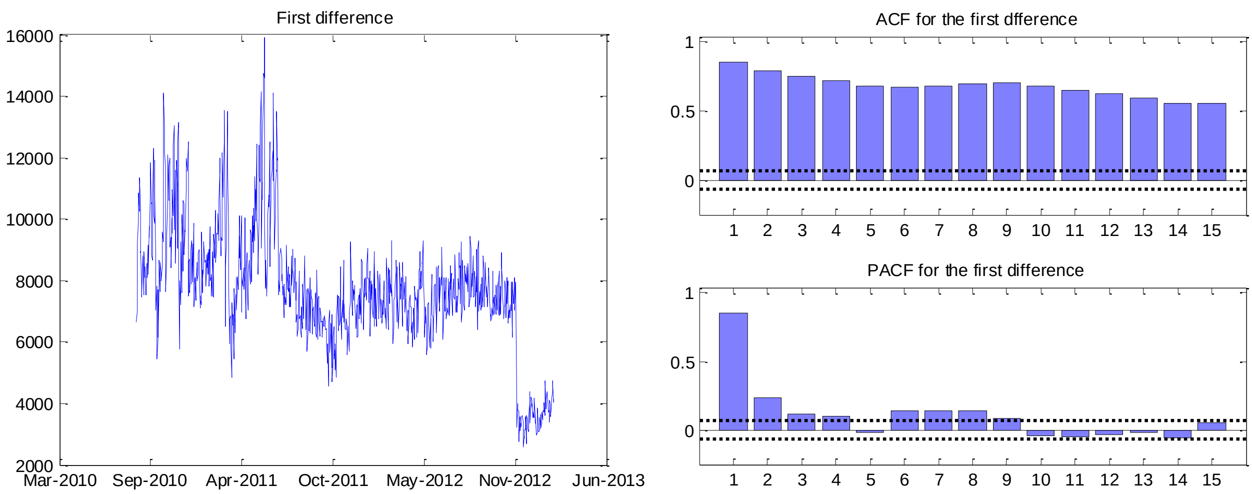
Variable Name: Total Bitcoins in Circulation



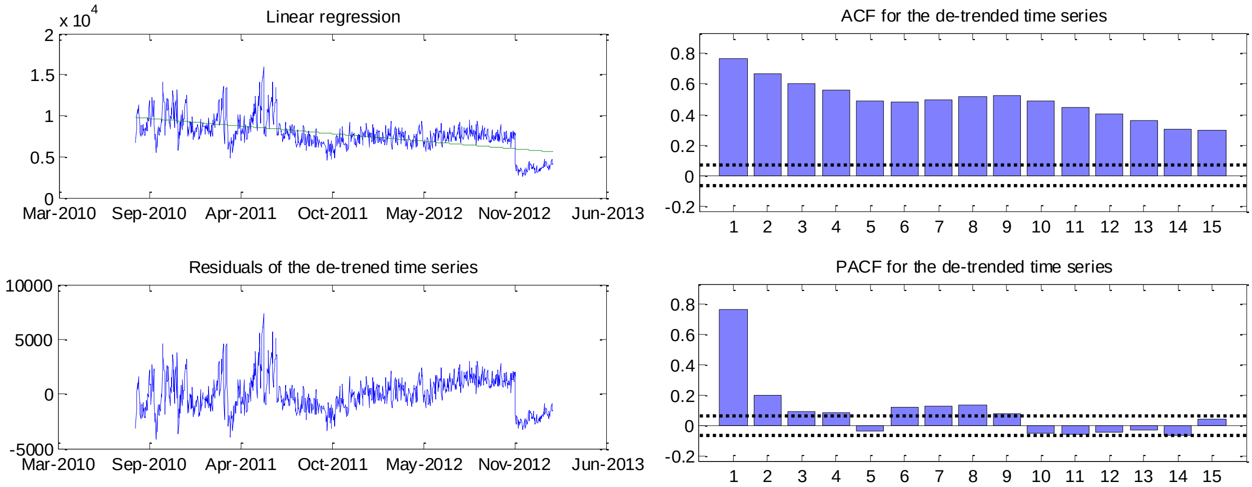
Let's de-trend the time series, as there seems to be one.



The result is clearly not stationary either. Taking the first difference:

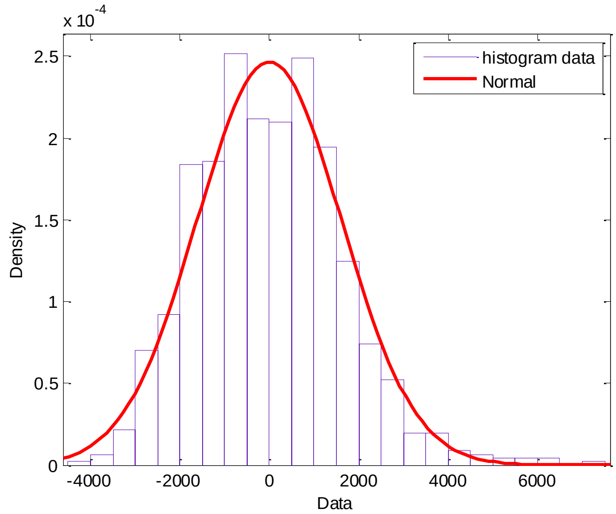


And proceed to de-trend it:



	Augmented Dickey-Fuller Test		
	Test statistic	**p-value**	**Unit root?**
De-trended difference	-3.2238	0.0016	No

The resulting time series is stationary. Let's see how it behaves.



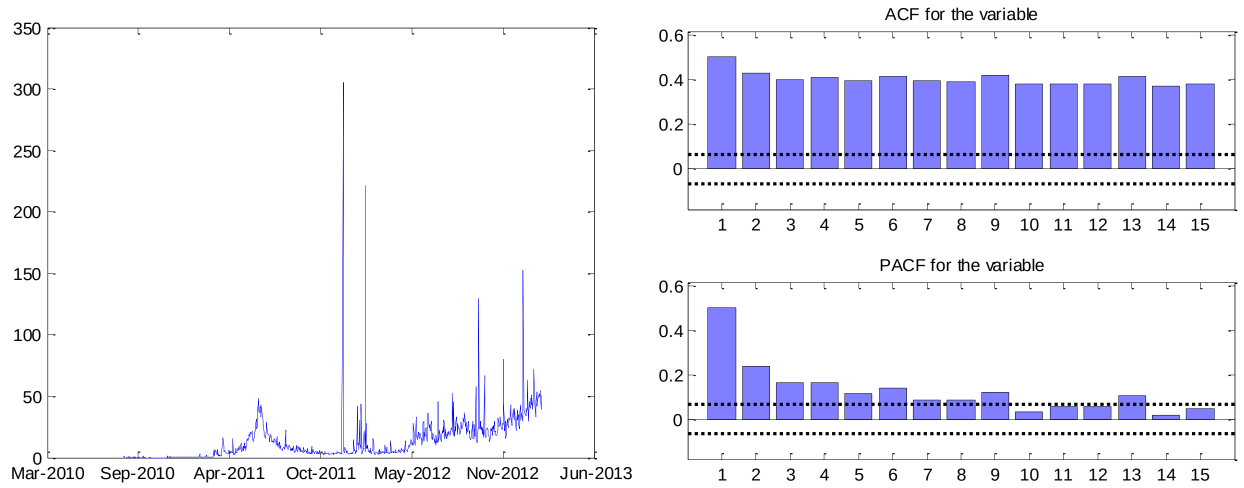
Summary statistics:	De-trended difference
Mean	1.9134e-12
Minimum	-4.1577e+03
25th Percentile	-1.1341e+03
Median	-48.0816
75th Percentile	1.0312e+03
Maximum	7.3841e+03
Variance	2.6221e+06
Standard Deviation	1.6193e+03
Skewness	0.4857
Kurtosis	3.7179
Excess kurtosis	0.7179

It seems to have some skewness and kurtosis. The following tests confirm this.

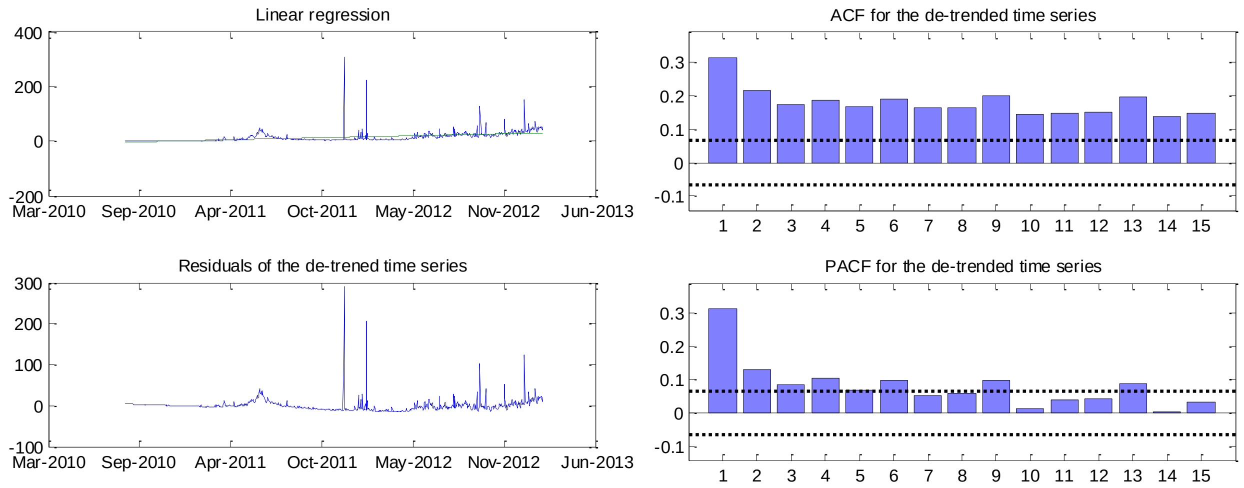
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t = $ 3.5743e-14	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	1.0000	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = 5.9979$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	1.9985e-09	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 4.4326$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	9.3088e-06	Yes

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	JB = 55.6234	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	8.3467e-13	Yes

Variable Name: Transaction Fees (BTC)

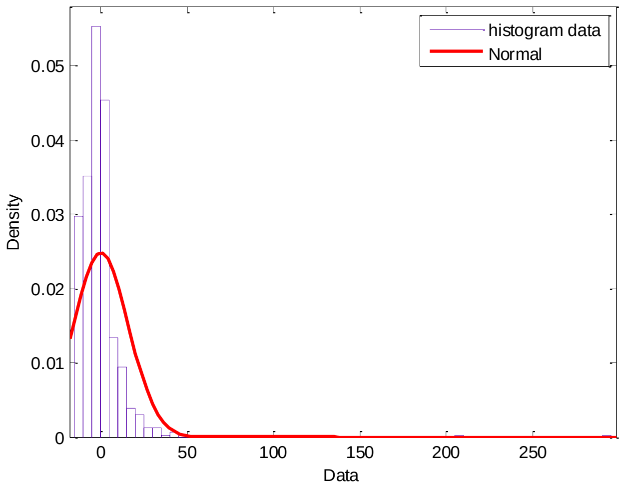


There is a noticeable trend. Let's take it out.



	Augmented Dickey-Fuller Test		
	Test statistic	**p-value**	**Unit root?**
Residuals	-4.3481	0.0000	No

The Augmented Dickey-Fuller test rejects the null hypothesis about the existence of a unit root.



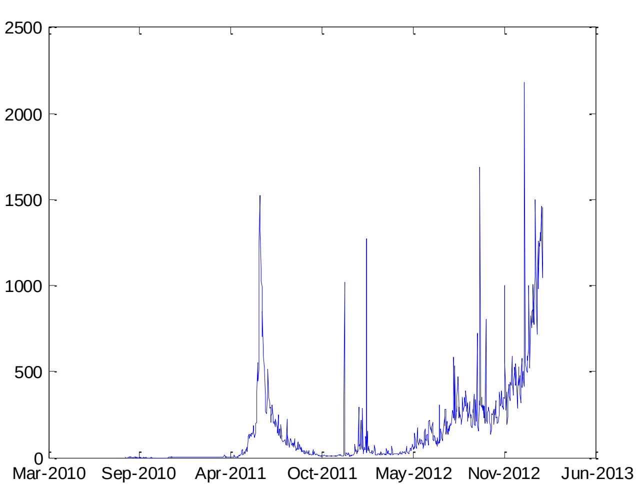
Summary statistics:	De-trended difference
Mean	0.0000
Minimum	-14.7732
25th Percentile	-7.7188
Median	-1.4134
75th Percentile	2.8442
Maximum	291.0996
Variance	258.2722
Standard Deviation	16.0709
Skewness	9.8955
Kurtosis	153.8853
Excess kurtosis	150.8853

A priori, it can already be said that the time series does not behave like the Normal distribution. The following tests confirm it.

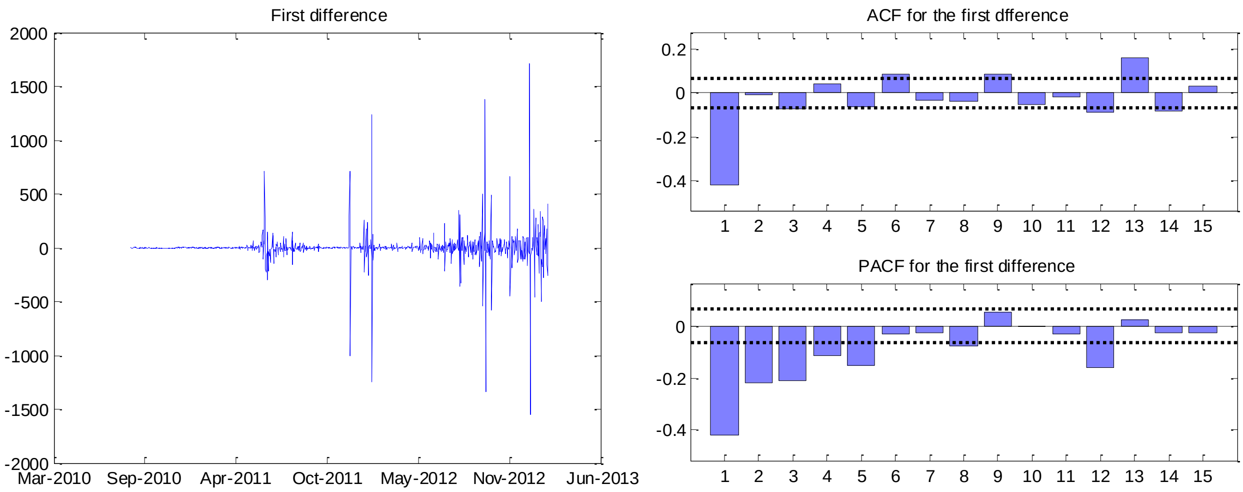
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	t = -2.8662e-14	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	1.0000	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	t_s = 122.2678	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	t_k = 932.1572	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k ≠ 3	JB = 8.8387e+05	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.0000	Yes

Variable Name: Transaction Fees (USD)

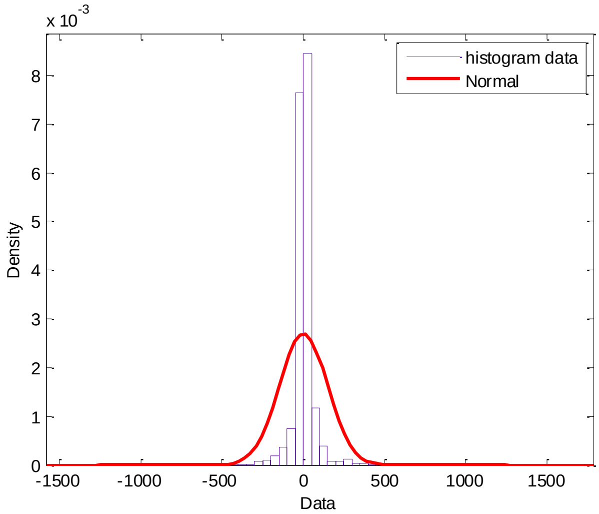


The value of this variable is calculated as the value of the previous one multiplied by the Exchange Rate. It is not stationary and de-trending it does not help, either. Let's take the first difference.



The spikes seem to increase in intensity with time. The autocorrelations are small.

	Augmented Dickey-Fuller Test		
	Test statistic	**p-value**	**Unit root?**
First difference	-12.0707	0.0000	No



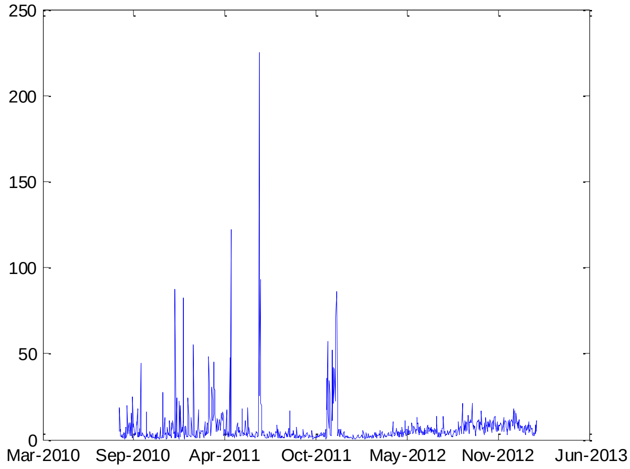
Summary statistics:	First difference
Mean	0.0000
Minimum	-14.7732
25th Percentile	-7.7188
Median	-1.4134
75th Percentile	2.8442
Maximum	291.0996
Variance	258.2722
Standard Deviation	16.0709
Skewness	9.8955
Kurtosis	153.8853
Excess kurtosis	150.8853

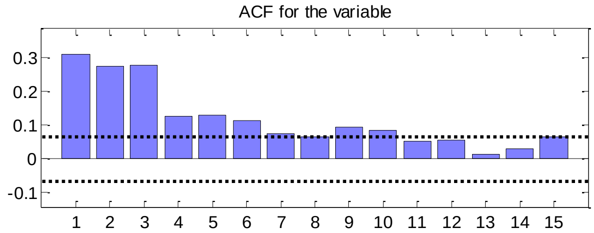
The first difference of the time series is too skewed and has too much kurtosis compared to a Normal distribution. The following tests confirm this.

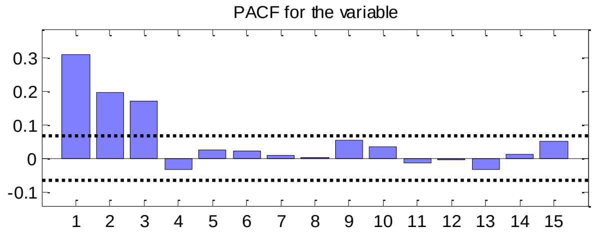
	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t = 0.3257$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.7447	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = 4.7987$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	1.5974e-06	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 382.2705$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	JB = 1.4615e+05	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.0000	Yes

Variable Name: Transaction Volume / Trade Volume Ratio

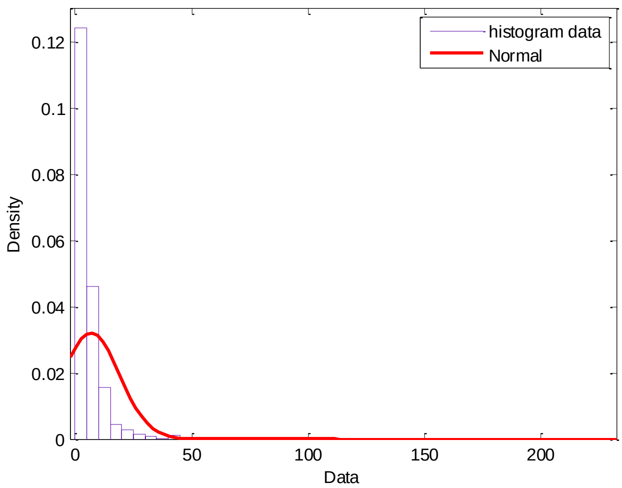






This is an interesting ratio because it tries to account for the speculation of the Bitcoin network. It compares the transaction volume with the trade volume from the top exchanges. The time series is indeed stationary. However, the data presents some of spikes which were also present in the transaction volume plot.

	Augmented Dickey-Fuller Test		
	Test statistic	p-value	Unit root?
Variable	-4.8112	0.0000	No



Summary statistics:	De-trended difference
Mean	6.6932
Minimum	0.3590
25th Percentile	1.9914
Median	3.9155
75th Percentile	7.0474
Maximum	225.1602
Variance	154.0880
Standard Deviation	12.4132
Skewness	9.0101
Kurtosis	125.1633
Excess kurtosis	122.1633

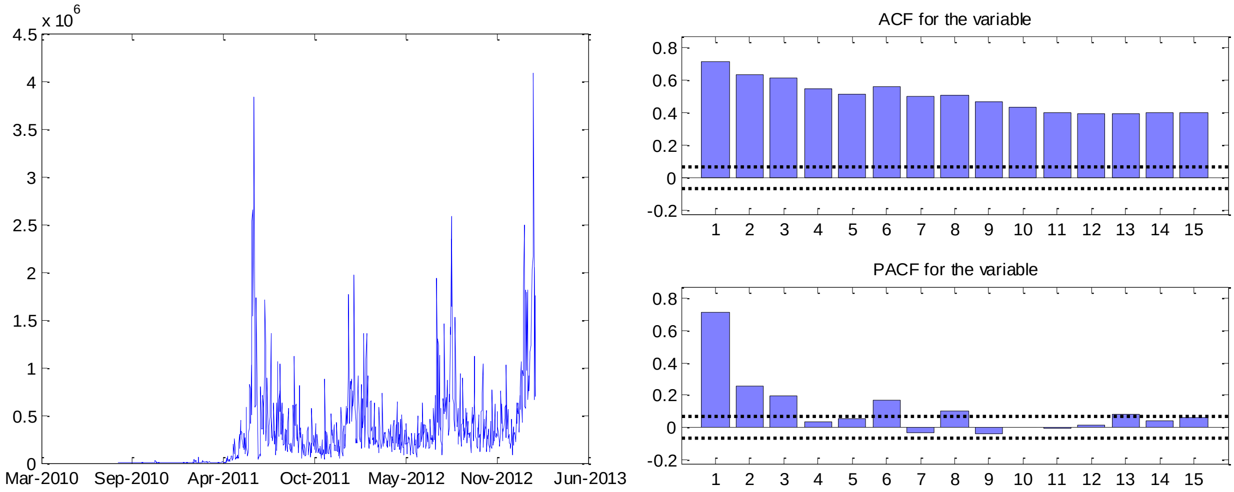
The variable is skewed and has a high kurtosis. Again, the following tests confirm that it does not behave like a Normal distribution.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	$t = 16.3191$	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	0.0000	Yes
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = 111.3277$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 754.7147$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

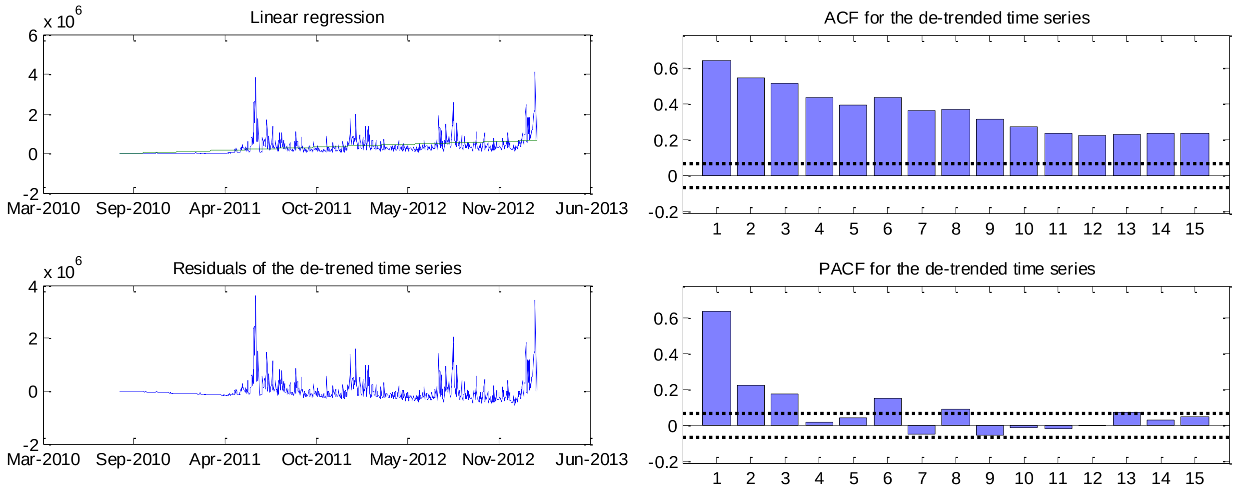
Indeed, the tests yield extreme test statistics, rejecting the null hypotheses in all the cases.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	JB = 5.8199e+05	χ^2 (2) where $JB = {t_s}^2 + {t_k}^2$	0.0000	Yes

Variable Name: USD Exchange Trade Volume

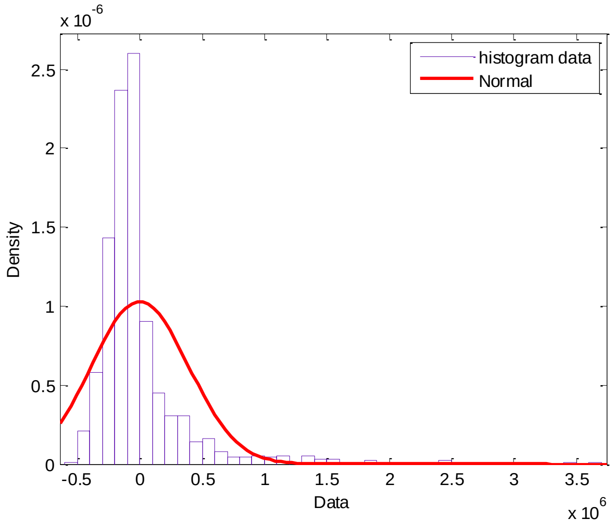


This variable represents the total amount of USD volume in the top exchanges. The autocorrelation plot suggests that the variable may be stationary. However, an ADF test suggests that there is a trend. Let's de-trend it before studying it, as in this paper the tests are conducted on the stationarized variable:



	Augmented Dickey-Fuller Test		
	Test statistic	**p-value**	**Unit root?**
Variable	-5.0161	0.0000	No

An Augmented Dickey-Fuller test with a time trend included rejects the hypothesis of a unit root (variable is trend stationary).



Summary statistics:

	De-trended variable
Mean	-2.3461e-10
Minimum	-5.3315e+05
25th Percentile	-1.7447e+05
Median	-8.6769e+04
75th Percentile	1.1458e+04
Maximum	3.6233e+06
Variance	1.5030e+11
Standard Deviation	3.8769e+05
Skewness	3.8807
Kurtosis	25.6518
Excess kurtosis	22.6518

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Mean	$H_0: \mu = 0$ $H_1: \mu \neq 0$	t = -1.8315e-14	t(T-1)~N(0,1) where $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$	1.0000	No
Skewness	$H_0: s = 0$ $H_1: s \neq 0$	$t_s = 47.9491$	N(0,1) where $t_s = \frac{\hat{s}}{\sqrt{6/T}}$	0.0000	Yes
Excess kurtosis	$H_0: k = 3$ $H_1: k \neq 3$	$t_k = 139.9409$	N(0,1) where $t_k = \frac{\hat{k}-3}{\sqrt{24/T}}$	0.0000	Yes

Clearly, the variable has significant excess kurtosis and skewness.

	Hypothesis	Test statistic	Distribution	p-value	Significantly ≠ 0?
Jarque-Bera Test	$H_0: s = 0$ and $k = 3$ $H_1: s \neq 0$ or k $\neq 3$	JB = 2.1883e+04	χ^2 (2) where $JB = t_s{}^2 + t_k{}^2$	0.0000	Yes