

Feedback comments on Draft Dissertation Paper – Hamoor Oberai

IMPORTANT: As part of the uploaded file package of your final report, please include:

a) within the text of your paper, in blue colour: all the changes you have made compared to this submitted draft;

b) in a copy of this Feedback comments file, also in blue: your response to the “Specific Comments” listed below, immediately below each comment.

[NB: when using LaTeX, colour can be introduced using the command `\textcolor{blue}{TEXT}`.]

General comments:

Report received on time.

Spelling of “its”, “stationarity” as noun

ADF test is classical tool. It seems a missed opportunity to not investigate and apply alternatives and study literature on competitive tests, more recent results.

Terms and concepts explained in general terms/definitions, insufficient application/interpretation for the Bitcoin/market problem at hand.

The paper seems to be written at a fairly basic level, explaining basic terms extensively, lack of precision/quantifying or concepts, more recent research results and trends from literature.

Comprehensive data analysis, though still some way to go to arrive at “Bitcoin prediction”.

Some highly controversial/wrong statements made, especially regarding nonstationarity.

Specific comments:

Abstract

“non stationarity, which needs to be constant”: how can this be?

No abbreviations to be used in index terms.

I mean that our data is not stationary. it needs to be converted to stationary. Then only we can use the data in SARIMA model.

Apologies for the misleading term. Rectified.

Sec. I:

A:

“democratized system”: how can it be democratic if it requires access to computer calculation power for mining?

A democratized system is a system that is accessible to everyone. Blockchain network has no central authority. All the information is public. Each Bitcoin address is unique, and it can't be forged. It has the ability to uniquely identify every person and guarantee that the person is how they say they are. Personal information can be masked, yet there is a digital footprint.

Bitcoin mining is the act of processing transactions on the bitcoin network. The miners have the ability to accept/ reject a transaction. They can choose to refuse to process the transaction.

Example: An individual miner can choose to not process a transaction, but someone else will.

But what if more than half the miners decided to stop processing transactions from an entity they disagree with? Now the choice of those miners has an impact. If more than half of the network decides something does not belong on the network, the transactions will never make it into the

blockchain, effectively being ignored by the bitcoin universe. This makes the network a democratic system.

The miner requires computer power/calculation to process the transaction in return of a small percentage of bitcoin as a reward.

B:

“The cleaned data set will have stationarity removed” please explain the rationale

Typo Error : The cleaned data after transformation will be stationary. So the non-stationarity will be removed.

Sec. II.A: contains repetition of ideas in I.A.

This section is too colloquially written. In a dissertation, I expect a more expert approach, which coding and error correction schemes are used, encryption methods, etc.

Modified

“The amount charged by the artist will be so small...” How are IPR and piracy handled/protected by Bitcoin?

The immutable ledger will find its’ use in IPR and piracy. The network will create a timestamp on ledger providing proof of ownership. The information will be public and will be readily available. Thus, immune from tampering and modification.

“transparency”: How can Bitcoin prevent money laundering? What is the role and importance of cyber-attacks?

Transparency in this context means:

A person’s identity is hidden via complex cryptography and represented only by their public address. So, if you were to look up a person’s transaction history, you will not see “Bob sent 1 BTC” instead you will see “1MF1bhsFLkBzzz9vpFYEmvwT2TbyCt7NZJ sent 1 BTC”.

So, while the person’s real identity is secure, you will still see all the transactions that were done by their public address. It adds that extra, and much needed, level of accountability which is required by some of these biggest institutions.

Bitcoin can’t prevent money laundering. Criminals use crypto money laundering to hide the illicit origin of funds, using a variety of methods. The personal identify of the person hidden which makes it difficult to trace the originality of the funds.

C: “They act in a regular uncertain manner.”: unclear, potentially self-contradictory.

REMOVED

“The trended time series observes the ACF plot to have positive values which will gradually decrease as the lags increase.” – not true in general

This statement was related to the ACF plot I created for the trended historical data. I have added the figure name to the data.

“Average is the best forecast model for White noise.” – not correct. Most of the statements in Sec. II.E and F are imprecise, misleading or wrong.

I have removed white Noise from the paper as I am not using in my results

G:

How is a “near-stationary series” defined and quantified?

Near stationary series is one where p value is less than 0.05; rejecting null hypothesis. Lower the value, the model is more stationary. Near-stationary was general term I added. I have removed near the paper. The purpose to add near was to signify that the mean will be close to constant.

II.C to II.H: this is a general theoretical description, not interpreted/applied to the Bitcoin study at hand.

Reason:

A general introduction to the Time series data is given showing its’ component and attributes. The concept of stationarity is explained. I choose a Seasonal ARIMA model to forecast stock prices which I wanted to explain in the literature review. The historical data is added in the SARIMA model and appropriate parameters are explained.

The methodology section have statistics about the model, steps leading to stationarity and parameters chosen for SARIMA model.

Sec. III.

III.A:

“is replaced with previous /next/median value depending upon the criteria” – rather vague. What precise rule and criterion did you use?

Any other missing attribute from Open, High, Low, Close is replaced with preceding value in the data. Figure in appendix contains function “treat_missing(df, choice)” handles all the missing values.

III.B:

“Since the force of gravity is greater than the force of elasticity,” – explain

Receoderde

When I was trading with bitcoin in my previous company, this statement was generally used to explain when you have sudden increase in stock price, it relatively falls down much steeply. I have removed this term.

III.C.1:

“Visually we can check” -- not sufficiently precise and quantitative. What criteria to be used?

Rolling Statistics: By plotting moving average or moving variance and see if it varies with time. This is more of visual technique which is demonstrated in Figure 4. Figure 4 shows the Rolling Mean in red colour and rolling standard deviation in black colour plotted against 12-month time window. The mean and variance displayed are not constant.

III.C.2:

Why “unit root”? Meaning?

A unit root is a stochastic trend in a time series. A unit root is called random walk with a drift. If a time series has unit root, it shows systematic pattern that is unpredictable. Unit root tests are test for stationary in a time series. A time series has stationarity if a shift in time doesn't cause a change in the shape of distribution. Unit root are one of the cause for non-stationarity.

“Hence, the data is not stationary.” – now what is the next step & strategy regarding data analysis, modelling and forecast?

Next step is shown in implementation where I convert my data using Box-cox transformation.

More discussion and analysis for Fig. 5 seems warranted.

I have used `seasonal_decompose` function in python that is useful in extracting trend, series and residuals from the data. Figure 5 shows some interesting facts. Bitcoin gradually increased until mid 2017. After 2017 there is an exponential positive trend observed but by April 2018, there is negative trend observed. Bitcoin gained positive confidence in 2019.

The residuals are also interesting, showing periods of low variability in early years and high variability from 2017 of the series.

IV.B:

Rationale why DF test when you have already done ADF before? Repeat values of III.C.2?

Typo Error : The results were of ADF test.

The serial correlation could be an issue of Dicky Fuller test. As a result, to handle bigger and complex model ADF test is better choice.

IV.C:

“Note that the ACF shows exponential decay. This is indicative of a stationary series.” – not correct

The ACF plot is useful in identifying a non-stationary time series. Figure 8 shows the ACF plot of data, decreasing slowly to zero, which tells us that the data is related to a previous time interval.

The ACF for stationary data will drop to zero relatively quickly, displayed in Figure 9.

“A total of 53 various parameters of a different combination of were estimated based on AIC” – please provide calculation details of your estimations

6 parameters of different combination were used to calculate the lowest AIC value. Following Parameter were used with their ranges:

1. ps: 0-3
2. qs: 0-3
3. d: 1
4. Ps: 0-3
5. D: 1
6. Qs: 0-3

The total permutation of (ps,d,qs)(PS,D,Qs) is updated and are 81.

Calculation Details added in the appendix.

IV.E:

I am not seeing any validation of the outputs generated by the tools that you have chosen. How do you know the results it generates are correct?

Table II shows the results of ADF test after box-cox transformation. The test value is less than the critical value; rejecting null hypothesis. Thus, the data has now become stationary.

LRA 2020-08-14