Time Series Analysis of Bitcoin

Andrew Hencic

A Dissertation Submitted to the Faculty of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

Graduate Program in Economics
York University
Toronto, Ontario

June 2019

©Andrew Hencic, 2019

Abstract

This thesis addresses the prediction problems associated with noncausal processes in cryptocurrency markets. Chapter one provides background on Bitcoin and cryptocurrencies in general. It begins by introducing four major cryptocurrencies. Then recent developments in economic research on Bitcoin are discussed.

Chapter two introduces a noncausal autoregressive process with Cauchy errors in application to the exchange rates of the Bitcoin electronic currency against the US Dollar. The dynamics of the daily Bitcoin/USD exchange rate series display episodes of local trends, which are modelled and interpreted as speculative bubbles. The structure of the Bitcoin market is described to give context for the presence of multiple bubbles in the exchange rate. The bubbles may result from the speculative component in the on-line trading. The Bitcoin/USD exchange rates are modelled and predicted. The mixed causal-noncausal autoregressive model is shown to better fit the data than the traditional purely causal model. A forecasting exercise using the noncausal model is then presented.

Chapter three examines the performance of nonlinear forecasts of noncausal processes from closed-form functional predictive density estimators. To examine the performance, time series are simulated with different conditional means and non-Gaussian distributions. The processes considered have the mixed causal-noncausal MAR(1,1)dynamics and both finite and infinite variance. The forecasts are assessed based on the forecast error behaviour and the goodness of fit of the estimated predictive density. The persistence in the

noncausal component directly relates to the magnitude of the bubble effects in the time series and is found to have a meaningful impact on how forecastable the process is. To better predict bubbles the joint density of the forecast at horizon two is shown to be an effective graphical method to detect the outset of a bubble.

Dedication

This thesis is dedicated to my parents. Thank you for everything you've done.

Acknowledgements

This thesis would not have been completed had it not been for the contributions of many important people. I want to thank my supervisor Joann Jasiak. Despite her many commitments she always made time to talk, help, and offer words of encouragement throughout. Her dedication to her students is incredible and without it this thesis would not have been possible.

I also want to thank Christian Gourieroux for his willingness to listen to research ideas, provide feedback, and collaborate. The manuscript in Chapter 2 was published as a joint authored chapter in the book Econometrics of Risk entitled, "Noncausal Processes in Application to Bitcoin/USD Exchange Rates".

Finally, I want to thank my committee members Augustine Wong and Barry Smith for their constructive feedback, thoughtful conversations and positivity.

Contents

List of Tables

List of Figures

Introduction

This thesis focuses on two emerging branches of research. The first, cryptocurrencies, have emerged as part of the fallout of the Financial Crisis. In 2008, with distrust of the banking sector permeating society, an individual under the pseudonym Satoshi Nakamoto created an independent currency that would operate on a network beyond the purview of monetary authorities. This currency would rely on the network of users to ensure legitimacy of transactions, be issued directly to users as a reward for participation and be storable on private devices. In effect, the technology allows users to bypass the traditional financial sector. Nakamoto named this currency Bitcoin and effectively started a generation of cryptocurrencies and research on their underlying technology, the Blockchain.

The second branch of research involves noncausal processes. Recent developments have greatly simplified the estimation and forecasting of what had previously been an involved econometric exercise. Lanne and Saikonnen's (2011) work presented the simple case for how forward-looking processes can be estimated and applied to economics. A closed form estimator for the predictive density was presented by Gourieroux and Jasiak (2014) that greatly simplifies the forecasting of univariate noncausal processes.

This thesis is divided into three chapters. Chapter one will provide a brief overview of past research into cryptocurrencies (primarily Bitcoin). This area of research is in a relatively nascent stage and the chapter is to serve as background information for the

analysis in the chapters that follow.

Chapter two introduces a noncausal autoregressive process with Cauchy errors in application to the exchange rates of the Bitcoin electronic currency against the US Dollar. The dynamics of the daily Bitcoin/USD exchange rate series display episodes of local trends, which can be modelled and interpreted as speculative bubbles. The bubbles may result from the speculative component in the on-line trading. The Bitcoin/USD exchange rates are modelled and predicted.

Chapter three examines the performance of nonlinear forecasts of noncausal processes from closed-form functional predictive density estimators. The processes considered have the mixed causal-noncausal MAR(1,1) dynamics and various non-Gaussian distributions with finite and infinite variance. The forecasts are assessed based on the forecast error behaviour and the goodness of fit of the estimated predictive density.

Chapter 1

Bitcoin and Other Cryptocurrencies

1 Bitcoin and Other Cryptocurrencies

This thesis will provide some background on Bitcoin and other cryptocurrencies. Specifically, Chapters 2 and 3 involve an application of the noncausal modelling method directly to an explosive bubble period in Bitcoin/USD exchange rates. However, since the publication of the manuscript in Chapter 2 there has been much development in cryptocurrency markets. A large appreciation of prices in late 2017/early 2018 along with the accompanying news stories generated much interest and discussion in the currencies. This chapter will address some of these more recent developments.

1.1 Cryptocurrencies

Researchers at the European Central Bank (2015) define cryptocurrencies as virtual currency schemes which can be readily converted to a national currency and bought and sold according to an exchange rate. Of these, the most popular is Bitcoin. Originally unveiled in 2009 by a pseudonymous developer under the moniker Satoshi Nakamoto, it has provided the blueprint and inspiration for a new generation of cryptocurrencies.

Of the more than 2000 cryptocurrencies in existence [1] four stand apart from the rest in terms of market capitalization and volume. These are Bitcoin, Ethereum, XRP and Litecoin. As of March 27, 2019 their market capitalizations stand at $71.9 Billion, $17.8 billion, $13.0 Billion and $3.8 Billion, respectively. Each has its own unique features and purpose, which makes them an interesting case study and introduction into the world of cryptocurrencies.

[1] The website Coin Market Cap has a complete list: https://coinmarketcap.com/all/views/all/

1.2 Bitcoin

Bitcoin is the original cryptocurrency. Created in 2009 by Satoshi Nakamoto it became famous for its distributed ledger technology (called the Blockchain), fixed lifetime supply and guiding philosophy to create a financial system that ensures anonymity (Nakamoto 2009).

The basic building block of Bitcoin is the Blockchain. Bitcoin seeks to create a system where payments can be sent and received without the need for an intermediary. In order to do this, the Blockchain was developed as a "distributed ledger", so that all participants in the ecosystem can trust that payments are being processes accurately and no double spending is occurring. Chapter 2 will provide a more detailed description of the mechanism behind the Blockchain, but a brief discussion here will follow.

Owners of bitcoin download an electronic wallet where they will store their bitcoins. Each wallet is assigned a unique id. As users make payments in bitcoin, they are transferred from one wallet to another. In order to track these transactions the blockchain records them in a public ledger. Importantly, even though the ledger is public the anonymity of the users is preserved.

Members of the network compete to be able to add transactions to the ledger. In order to win the right to add to the ledger they have to solve a cryptographic problem. The difficulty of this problem adjusts such that a new 'block' is added to the blockchain every 10 minutes. Whichever participants solves the problem first is rewarded with newly created bitcoins (this is how the supply of bitcoins is increased). However, as the total supply of bitcoins is limited to 21 million units, eventually the participant that adds the new block will be able to charge a transaction fee.

1.3 Litecoin

Much like Bitcoin, Litecoin's stated mission is to facilitate peer-to-peer payments around the world[2]. Its creator, Charlie Lee, sought to improve upon Bitcoin so as to make a currency more suitable for use in many transactions. As with Bitcoin, it is built using the blockchain. However, rather than blocks being added every 10 minutes, Litecoin has blocks added every 2.5 minutes. Along with the faster transaction times, it also has a finite supply of 84 million Litecoins.

1.4 Ether

Ethereum itself is not a cryptocurrency, but a platform upon which applications looking to use blockchain technology can be built[3]. It was originally proposed by a developer by the name of Vitalik Buterin [4], and is currently the project of the Ethereum Foundation[5].

The basic premise of Ethereum is like that of Bitcoin. The blockchain still exists as a public ledger that is the official record of transactions and balances on the Ethereum network. However, unlike the Bitcoin blockchain which is essentially a list of the transactions on the network, the Ethereum blockchain contains both the list of transactions and the state of the blockchain (in the simplest case, the balances of all the accounts).

The Ethereum whitepaper provides a list of potential use cases for the platform, but the primary one is the ability to build 'smart contracts' that will execute themselves, eliminating the need for counterparties to rely on each others trustworthiness.

Ether (ETH) is the unit of account employed by the Ethereum network. For the Ethereum network to process a smart contract (or any other application) it charges a

[2]https://litecoin.org/
[3]https://www.ethereum.org/
[4]Whitepaper: https://github.com/ethereum/wiki/wiki/White-Paper
[5]www.ethereum.org

transaction fee in Ether based on how computationally complex the task is. The user whose transaction is being executed compensates the network with Ether taken from their digital wallet.

Like other cryptocurrencies, Ether relies on the blockchain to record and verify transactions. There are two major differences between Ether and Bitcoin. First, unlike Bitcoin who's total supply is capped at 21 million units, there is no fixed supply of Ether. Secondly, blocks are added to the Ether blockchain every 15 seconds (on average), unlike the approximately 10 minutes for Bitcoin[6].

The block reward for mining Ethereum (i.e. the payment received for adding transactions to the blockchain) is currently set at 5.0 ETH plus whatever the transaction cost charged was, plus and extra reward for including an orphaned chain as part of the block [7].

1.5 XRP

Unlike other cryptocurrencies, XRP was developed by the for profit firm Ripple. Ripple [8] looks to facilitate payments for institutions and XRP is a component of their payments processing system. Rather than looking to create a financial system that operates outside of the establishment, Ripple is looking to exploit blockchain technology to reduce frictions in the global payments system.

Amongst their clients Ripple lists, Santander, a major European bank. They are using the Ripple platform for a mobile application that allows users execute cross border transactions with much shorter transaction times [9].

[6] www.ethdocs.org/en/latest/mining.html

[7] The orphaned chains are called 'uncles'. For more details as to why they are included please see: www.ethdocs.org/en/latest/mining.html

[8] A description of the firm is available at: https://ripple.com/company/

[9] `https://ripple.com/files/case_study_santander.pdf`

XRP is not only limited to financial institutions but is available for purchase by the general public on cryptocurrency exchanges. In addition, it XRP not mined like other cryptocurrencies. 100 Billion XRP coins were created and this supply is to be held fixed going forward[10].

1.6 Bitcoin News and Awareness

Bitcoin was on the radar of Fed officials far before its recent price appreciations. In a working paper (Yermack 2013) Fed researchers explored its merits as a currency by discussing how well it fulfils the three functions of money: (1) a medium of exchange, (2) a unit of account, and (3) a store of value.

Yermak finds that there are serious limitations to its usefulness as a medium of exchange. At the time of writing, the author finds that registered merchants(roughly 24,000 of them) averaged less than one Bitcoin transaction per day. Additionally, the author highlights that acquiring Bitcoin is not very easy as they either have to be mined, or acquired on an exchange which exposes the user to another set of risks. Finally, Yermack notes that there are no credit products available to consumers, resulting in all transactions having to be conducted on a cash basis.

When Bitcoin is examined through the lens of a unit of account. Yermack notes the volatility in the Bitcoin/Dollar exchange rate would negatively impact price stability in the market and is ultimately, "costly to the merchant and confusing to the consumer". Citing the summary of research on consumer prices by Thomas and Morwitz (2009), Yermack notes the challenges faced by consumers in comparing prices. Extending this notion to Bitcoin, whose divisibility is often cited as a feature, Yermack remarks that these issues could very well be exacerbated when prices are listed in scientific notation.

[10] `https://bitcoinmagazine.com/guides/what-ripple`

Lastly, Yermack notes two limitations in Bitcoin's usefulness as a store of value. First, is the very high volatility relative to gold and national currencies. Secondly, there is no real form of deposit insurance. Given the relatively frequent security lapses and robberies of Bitcoin exchanges (notably that of Mt.Gox in 2014), it seems the latter point is crucial to long term viability of Bitcoin as a currency.

Schilling and Uhlig (2018b) develop a model to derive the fundamental value of Bitcoin in a world where agents can decide which currency they transact in. The authors present a model of an exchange economy with two infinitely lived agents who alternate their consumption and production between even and odd periods. The pricing equation for Bitcoin is shown to be a martingale if conditions for no speculation are satisfied. Schilling and Uhlig (2018a) then present a model for currency choice in the presence of transaction costs (in the case of Bitcoin) and value added taxes (for traditional currencies). The relative costs of transacting in the currencies determine the choice of the medium of exchange for agents.

The rising popularity of cryptocurrencies has caught the attention of monetary authorities around the world. A European Central Bank (2015) research paper provides a comprehensive review of virtual currency schemes (and the cryptocurrencies that form a subset of them). Of the roughly 500 cryptocurrencies operating at the time of writing the report notes that only Bitcoin had emerged as being used for payments. Bitcoin is also credited as being important to the emergence of other cryptocurrencies as it is an open-source project whose protocol forms the basis for many of the new "altcoins", including other major cryptocurrencies Ethereum, XRP, and Litecoin.

The report provides four main characteristics along which virtual currency schemes can be differentiated. First, the transaction validating system they employ. The two main methods are proof of work systems, like that used by Bitcoin, and proof of stake

systems [11]. The proof of stake system purports to be more energy efficient and process transactions faster than the proof of work system. Secondly, the algorithms used by the currencies can be differentiated by the type and quality of hardware required to execute the algorithm. Thirdly, the supply of coins differs across products. Some, like Bitcoin and Litecoin, have a fixed total supply, while others do not. Lastly, some virtual currency schemes don't necessarily operate as currencies, but rather offer services that employ the blockchain technology. More detail in the measure of differentiation and about the potential applications and pitfalls of virtual currency schemes can be found in the 2015 report.

The Bank of Canada has also been conducting research on the applications and awareness of cryptocurrencies. Given its leading position in the market place the BoC has run a so called "Bitcion Omnibus Survey" of the general public to ascertain their awareness of, and involvement with, Bitcoin. The first edition was conducted in 2016, while the 2017 edition cointains an addendum with a questionnaire to gauge the knowledge of the participants.

In line with the rapid price appreciation from 2016 to 2017 the authors find that the awareness of Bitcoin grew by 21 percentage points, with the highest levels of awareness amongst those with incomes over $ 70,000. The knowledge test showed that both owners and non-owners showed an increase in average test scores. The authors computed two test scores, including one that deducted points for wrong answers to discourage guessing, and in both instances the average score was higher for owners of Bitcoin than for non-owners. This indicates an increased familiarity for those who have a monetary investment

[11] The biggest difference between the two is that the proof of work system relies on participants to compete to, in essence, solve mathematical problems in order to verify transactions. These participants are then rewarded with newly created cryptocurrency. In a proof of stake system transactions are verified by predetermined members of the network who already have established stakes in the cryptocurrency. These members can then collect a transaction fee, rather than receiving newly created cryptocurrency.

in the cryptocurrency. It is important to note that the authors attribute an increase in scores from year to year to the improved readability of the questionnaire. Finally, the authors find that though ownership of Bitcoin had increased from 2016, that increase was attributable to newly aware individuals.

At the shorter time frequency, Urquhart (2018) employed Google Trends data that describes search popularity to examine what causes attention to Bitcoin. They construct three vector autoregressive models. The first includes search queries and realized volatility of Bitcoin returns, the second includes search queries and Bitcoin trading volume, and the third search queries and Bitcoin returns. Ultimately the author finds that after splitting the Bitcoin price sample (using a Bai and Perron (2003) test) the reaction of search queries to trading volume, returns, and realized volatility differ across subperiods. In the first period (pre October 28, 2013) none Granger cause search queries. In the second subperiod realized volatility of Bitcoin returns, Bitcoin trading volume and Bitcoin returns all Granger cause search queries at various lags.

For all of its attention there are still issues with Bitcoin markets. In their presentation to the Securities and Exchange Commission in support of their application for a Bitcoin ETF, Bitwise Asset Management illustrates one. Their report shows that up to possibly 95% of current trading volumes may not be real (slide 62) and, potentially, only 10 exchanges have actual volumes (slide 61).

1.7 Bitcoin Return Research

Much of the academic literature on Bitcoin has focused on its merits as a financial asset. Specifically, there has been a significant amount of research about the efficiency of Bitcoin prices and whether the behaviour of the price can be said to satisfy the Efficient Market Hypothesis (EMH).

Urquhart (2016) was amongst the first to examine the efficient market properties of Bitcoin. Treating it as an asset, Urquhart looks to examine whether Bitcoin returns satisfy the weak form of market efficiency proposed by Fama (1970). Urquhart calculates Bitcoin log returns at the daily frequency by taking the volume weighted average price across exchanges. Then, the author conducts a variety of tests on these returns to test the EMH.

Ultimately, the author finds that all of the tests reject the weak form of informational efficiency over the full sample (August 1, 2010 to July 31, 2016). However, when the author splits the sample, with the second sample period starting in August 2013, the Ljung-Box (Ljung and Box (1978)) and automatic variance test (Choi (1999)) tests fail to reject the null hypothesis indicating some efficiency.

Bariviera (2017) looks to investigate the inconclusive market efficiency results of Urquhart (2016) by examining the long memory properties of Bitcoin return volatility. In order to conduct the analysis, Bariviera employs the Hurst exponent statistic using two methods; (1) the R/S method and (2) the Detrended Fluctuation Analysis method. The Detrended Fluctuation analysis method is better at avoiding detection of spurious long rang dependence. The author examines both daily returns and intraday volatility for dependence (i.e. the violation of the weak form of the Efficient Markets Hypotehsis).

Similar to Urquhart, Bariviera (2017) finds that Bitcoin returns exhibited dependence earlier in the sample, but lost these features after 2014. Meanwhile, the volatility clustering at the intraday level is found to be present over the whole sample.

Katsiampa (2017) looks to apply a single model of conditional heteroscedasticity to the whole Bitcoin historical time series. In order to accomplish this, the author computes the log returns of a Bitcoin price index provided by Coindesk. Then, using the log returns, an autoregressive model of the conditional mean and various forms of the

conditional variance are modelled. Ultimately, amongst the whole suite of potential conditional variance models (GARCH, EGARCH, TGARCH, Asymmetric Power ARCH, Component GARCH, Asymmetric Component GARCH), the empirical exercise identifies that the AR(1)-CGARCH(1,1) model has the best in sample properties as measured by the maximized value of the log-likelihood function, the Akaike information criterion, Bayesian information criterion and Hannan-Quinn information criterion. The implication of these finding, like those of Bariviera (2017) are that there is information in the long run components of the historical time series, in this case directly related to the conditional variance.

Another branch of literature examines potential other drivers of bitcoin returns. Panagiotidis, Stengos and Vravosinos (2018) estimate four types of vector autoregressive processes (VAR), over two time periods, to examine the effects of shocks to explanatory variables on bitcoin returns [12]. The four frameworks they employ are; (1) the standard VAR; (2) a factor augmented VAR; (3) a VAR with factor analysis; (4) a VAR with principal component analysis. The reason for the variety of methods is that there are a large number of potential explanatory variables, so the authors use various methods to reduce the dimensions of the VAR. Ultimately they find find that popularity (as measured by Google trends data and Wikipedia searches) has a reduced impact on bitcoin returns, while there exists a connection between shocks in traditional financial markets and bitcoin. Additionally, the factor analysis allowed the authors to examine the effects of geographic markets on Bitcoin returns. In the initial period (which begins in 2010 and ends in 2016) the authors find that shocks in the Asian region (measured by a combined China and Japan factor) had the largest impact, followed by the US and then Europe.

[12]The authors provide a very extensive review of the literature on econometric research into the empirical determinants of bitcoin returns. This background forms the basis for the explanatory variables chosen in the analysis.

However, over the sample that begins in 2010 and ends in 2018, the US becomes the most important. The authors note that this is likely due to the regulatory clampdown by Chinese authorities in late 2017 that saw a large reduction in the share of Bitcoin trading done with Renminbi.

1.8 Investment Strategies

Investors have begun examining the merits of inclusion of Bitcoin in their personal portfolios. Two papers, Bouri, Molnar, Azzi, Roubaud Hagfors (2017) and Timborn, Li and Hardle (2017), introduce methods for incorporating cryptocurrencies into investment portfolios.

Bouri et. al. (2017) examine the usefuleness of Bitcoin as a hedge or diversifier against conventional asset (equity, bonds and commodities) returns. To examine which of these functions Bitcoin can fulfil, the authors employ the Dynamic Conditional Correlation (DCC) model proposed by Engle (2002). The DCC model involves modelling the bivariate return equation of Bitcoin and the other asset as an autoregressive process and the conditional variance of the returns processes as a GARCH(1,1). Then, the time-varying unconditional correlation matrix of the standardized residuals from the initial GARCH estimation is then modelled as a bivariate system.

These dynamic correlations are then employed to evaluate whether Bitcoin can serve as an effective diversifier, hedge, or safe haven asset against a wide variety of financial assets [13]. By regressing these dynamic conditional correlations against dummies that represent extreme movements in different quantiles of the return distribution of various asset classes, the hedging properties are identified.

[13] These include the S&P 500, a Bond Index, a commodity index and gold among others. See Bouri et. al. (2017) for the full list.

Using daily Bitcoin returns, the authors find that Bitcoin is a "strong hedge against movements in Japanese and Asia Pacific stocks". Meanwhile, the weekly returns demonstrate that it can only be a strong hedge for Chinese equities at this frequency. This difference in hedging and safe haven properties stands out. The authors note that this may be a product of the speculative nature of the daily returns. These speculative movements in the price then undermine Bitcoin's safe haven properties at both the daily and weekly frequency.

Finally, the authors conclude by noting that the hedging and diversification benefits of Bitcoin could change profoundly in the future as the time period they study (2011-2015) features very high return variance. Secondly, they note the limited liquidity of Bitcoin is an important caveat for its role in a portfolio, that could be improved with further financial innovation.

On this note, Trimborn et. al (2017) directly address the liquidity issue associated with investing with cryptocurrencies. The authors look to examine potential advantages of including cryptocurrencies in a diversified portfolio while directly addressing the issue of their relative illiquidity.

The authors begin by noting that there exist sufficient liquidity concerns when adding alternative assets to an investment portfolio comprised of the S&P 500 index, the DAX30 index and those listed on the Portuguese stock exchange. The advantages of the added returns and diversification could quickly be offset by an inability to liquidate the assets in time to rebalance the portfolio. In order to address this concern Trimborn et al. introduce the LIBRO (Liquidity Bounded Risk-Return Optimization) method. It is a modification of the traditional Markowitz diversification theory proposed by Markowitz (1952).

To motivate their work Trimborn et al. first present a simulation study to demonstrate that in portfolio construction the incorrect estimation of asset allocation weights is less

of a concern than the incorrect estimation of the maximum weight attributable to an individual asset. This problem is, not surprisingly, exacerbated in the cases where the covariance amongst the assets in the portfolio is higher.

To find a proxy for liquidity the authors use Turnover Value. Turnover value is defined as the sum of transaction values over a predefined time interval.

This measure is then scaled by a factor that determines the speed at which the investor wants to clear their position. In turn, dollar amount assigned to any asset is then restricted to be less than or equal to the scaled Turnover value. By implementing this new constraint into the portfolio optimization problem, the performance of a portfolio with liquidity restrictions can be evaluated.

The authors then examine the merits of employing this augmented portfolio optimization strategy to a combination of 39 cryptocurrencies, the S&P 100 Index, the DAX30 index and those listed on the Portugal Stock Exchange. They find that, in sample, the portfolios incorporating cryptocurrencies improve the risk-return trade-off at every level of risk, including at the point of the global minimum variance portfolio. A novel observation is that a greater share of the investment is allocated to altcoins relative to bitcoins.

The cumulative performance of the S&P 100 portfolio with cryptocurrencies and no liquidity constraints shows minor improvements over a small horizon. However, when the liquidity constraints are added, the cumulative returns of the portfolio are 1.12% higher over the 3-year period. Similar results hold for the portfolio comprising of DAX stocks and cryptocurrencies, with cumulative returns exceeding those of the stock portfolio by 1.5%. Examining the risk-adjusted returns with the Sharpe ratio, a similar story holds, with the liquidity constrained portfolio combining cryptocurrencies and stocks dominating those of the stock only portfolios. Finally, the portfolios comprising of Portuguese stocks and cryptocurrencies the outperformance of the portfolio with liquidity constraints was

3.4%. Therefore, by accounting for the liquidity issues and introducing cryptocurrencies into stock portfolios the authors show that meaningful improvements in the risk-return relationship can be achieved.

Using the portfolio with Portuguese stocks, the authors undertake their simulation study to demonstrate that error in the estimation of the risk of the portfolio remains small even when incorporating their proposed liquidity constraints.

2 Conclusion

The emergence of Bitcoin and other cryptocurrencies have led to an explosion of trading and speculation in once non-traditional markets. To specifically address the speculation in the market, this thesis will employ the application of noncausal processes to model explosive bubble patterns introduced by Gourieroux and Zakoian (2017). Their method will be applied to the Bitcoin exchange rate series in Chapters 2 and 3. This relatively unexplored aspect of the exchange rate series is different from the traditional market efficiency and portfolio diversification problems that have been addressed in the literature. This thesis will use the features of the noncausal processes introduced here to seek a better explanation for the dynamics of the Bitcoin exchange rate series, and bubble processes more broadly.

Chapter 2

Noncausal Model in Application to Bitcoin/USD Exchange Rates[1]

[1]The manuscript in this chapter was published as: Gourieroux, C. and A., Hencic (2015): "Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rate" , in Huynh, V., et al. eds. "Econometrics of Risk", Series: Studies in Computational Intelligence, 17-40, Springer.

1 Introduction

In late 2013 digital currencies (sometimes referred to as crypto-currencies) and their standard bearer, Bitcoin, have received a lot of public attention [see, Velde (2013)]. This can likely be attributed to two factors. Public adoption of the digital currency was beginning to become more commonplace [see, Li (2013)] and its more nefarious uses were slowly being exposed [see, Flitter (2013)].

A prime example of the first point is the University of Nicosia in Cyprus. The University is the largest private university in Cyprus and it began accepting bitcoins as tuition payment in 2015. The university's reasoning was that they wished to be at the forefront of global commerce, but there may have been other reasons at play. Cyprus had gone through significant financial stress and many of the country's depositors were likely to face significant losses [see, Tagaris (2013)]. Mistrust of the established financial system might have lead institutions to begin accepting alternative means of payment.

As for the nefarious uses of bitcoins, the story about the raid on the website The Silk Road can speak to the dark side of digital and anonymous currency. In October of 2013 the FBI shut down The Silk Road for allegedly selling illegal drugs and charged its owner with a whole host of offences. Critics of digital currencies say that the anonymity provided to their users is dangerous and should be further regulated. The government of the United States has responded to these concerns by implementing rules to attempt to curb the use of digital currencies in money laundering [see, Sparshott (2013)]. With the market capitalization of bitcoin surpassing $12 Billion USD [see, 11], and its ever increasing adoption, further study of the uses, threats and mechanisms that govern digital currencies is needed.

The objective of this chapter is to examine the dynamics of the Bitcoin/USD exchange

rate and to predict its future evolution. The dynamics of the series are characterized by the presence of local trends and short-lived episodes of soaring Bitcoin/USD rates, followed by sudden almost vertical declines. These patterns are referred to as bubbles. In economics, bubbles in asset prices have been introduced in the context of the rational expectation hypothesis in the seventies and as a result of the speculative behavior of traders. The bubbles in the Bitcoin/USD rate may originate from a) the fact that the bitcoin market is still an emerging market with a lot of speculative trading, b) the asymmetric information and crowd phenomena [see, e.g. Bunnermeier (2001) for the analogous on Nasdaq], c) the lack of a centralized management and control of exchange rate volatility, d) the deterministic supply of bitcoins and the evolution of the volume over time. As the volume of bitcoins available on the market is exogenously determined, this enhances the bitcoin price and exchange rate volatility.

Due to the presence of local explosive trends, depicted as bubbles, the Bitcoin/USD exchange rate cannot be modelled by any traditional ARIMA or ARCH models [see, e.g. Andrews et al (2009)]. In this chapter, we use the mixed causal-noncausal autoregressive process with Cauchy errors [Gourieroux, Zakoian (2013)] to estimate and predict the Bitcoin/USD exchange rate.

The structure of the chapter is as follows. In Section 2, we describe the bitcoin as an electronic currency, and we explain the mechanisms of bitcoin trading and storage. Next, we describe the data and the period of interest that includes a bubble burst and crash. A speculative bubble is a nonlinear dynamic feature that can be accommodated by the aforementioned noncausal autoregressive process. In Section 3, we review the properties of noncausal processes and introduce the associated inference and prediction methods. The application to the Bitcoin/ US Dollar exchange rates recorded on the Mt. Gox [1]

[1]Formerly Magic: The Gathering Online Exchange.

exchange market is presented in Section 4. The noncausal model is used to predict the occurrence of the bubble in the Bitcoin/USD exchange rate. Section 5 concludes the chapter.

2 The Bitcoin/USD Exchange Rate

2.1 Bitcoin Currency

Bitcoin (BTC) is an electronic currency originally created by a developer under the pseudonym Satoshi Nakamoto in 2009 [see, Davis (2011)]. The electronic currency is distributed on a peer-to-peer network anonymously between any two accounts. There is no formal denomination or name for units of the currency other than 1.00 BTC being referred to as a bitcoin and the smallest possible denomination, 10^{-8} BTC, being a "satoshi".

The bitcoin can be purchased on a virtual exchange market, such as mtgox.com against the US Dollar or other currencies [2]. Users of the currency store it on a private digital "wallet". This wallet has no personal identification with an individual and is comprised of three components: an address, a private key, and a public key. There is nothing that connects a wallet to an individual. This level of anonymity has been one of the driving forces behind the currency's popularity. The bitcoin can be used to purchase a number of goods and services that are listed on the Bitcoin website.

Three types of wallets exist: the software wallet, the mobile wallet and the web wallet. Software wallets are installed directly on a computer and allow the user complete control over the wallet. Mobile wallets are installed on mobile devices and operate the same way.

[2] The transactions on this market have been suspended as of February 25, 2014. The reason is yet to be revealed, but an attack by hackers has been declared.

Web wallets host an individual's bitcoins online. All of these wallets can be accessed with just the private key assigned to the address. Again, there is nothing to associate a physical human being with a Bitcoin address other than if the person owns the hardware on which the wallet is installed.

As of December 2, 2013 the total market capitalization of bitcoin is approximately $12 billion USD [see, 11]. Bitcoin is traded 24 hours a day on various exchanges, the largest of which include Mt. Gox (based in Japan) [3] and BTC China (recently the world's largest BTC exchange [Liu (2013)]). The former is a real time exchange whereas BTC China is a fixed rate exchange [see, 9]. Bitcoins are denominated in USD on Mt. Gox and in Renminbi on BTC China. After a clarification by the People's Bank of China on Bitcoin's status at the beginning of December 2013, the exchanges on BTC China can only be done in Chinese Yuan, and the users have to now provide their identity using, for example, a passport number. Trading and use of Bitcoin is forbidden in Thailand.

The trading volume on Mt. Gox has slowly increased over time as adoption of the currency has increased. On its first day of trading on Mt. Gox, the total volume of bitcoin traded was 20 BTC units. Obviously this is a very small number in comparison to the 3,436,900 bitcoins in circulation at that time. However, trading volume has gradually increased since then as Bitcoin has become more generally accepted and garnered more attention. Trading volume reached an all-time high on April 15, 2013 with 572,185.7 bitcoins changing hands on Mt. Gox. At the time there were approximately 11,027,700 units in existence, meaning that on this day approximately 5 % of all bitcoins in circulation were traded on Mt. Gox.

The long term supply of BTC will never exceed 21,000,000 units. In comparison, the daily volume traded on the platforms can be much smaller. For example, on December 8,

[3] It represented 12% of the trades before it collapsed.

2013, the traded volume on Mt. Gox was 31 800 BTC units. The evolution of the traded volume of bitcoins between February and July 2013 is displayed in Figure 2.1. Figure 2.1 compares the daily volume of bitcoins exchanged against USD and the daily volume of bitcoins used for real transactions that is for the sale and purchase of goods and services offered in bitcoin. The daily volumes are small compared to the capitalization of bitcoin, showing that this market may encounter liquidity problems.

Bitcoins are produced in such a way that the volume of new bitcoins produced will be halved every four years until the volume of new coins produced decays to zero. At this point the final supply of bitcoins will be fixed (the exact amount of units varies depending on rounding, but it will be less than 21 million units) [see, 8]. Bitcoins are produced in a process referred to as "mining". Computers on the Bitcoin network solve complex mathematical problems and are rewarded for their work with a predetermined amount of bitcoins, referred to as a "block reward", and a transaction fee. The current block reward is 25 bitcoins [see, 7]. In order to control the supply of bitcoins being produced, the difficulty of these problems is automatically adjusted so that the time between solutions averages 10 minutes.

2.2 Bitcoin Transactions

To ensure the security of transactions, the Bitcoin system uses public key cryptography. Each individual has one or more addresses with an associated private and public key. The system is totally anonymous and balances are only associated with an address and its keys. Only the user with the private key can sign a transfer of bitcoins to another party, whereas anybody in the network can validate the signature and transaction using the user's public key [see, 7]. When a transaction occurs, one user sends another an amount of bitcoins and signs the transaction with their private key. The user who sends

the bitcoins announces a public key and it falls on the network to verify the signature. The user then broadcasts the transaction on the Bitcoin network. In order to prevent double spending the details about a transaction are sent to as many other computers on the network as possible in a block. Each computer on this network has a registry of these blocks called a "block chain". In order for the newest block to be accepted into the chain, it must be valid and must include proof of work (the solution to the aforementioned math problem). When a block is announced the miners work to verify the transaction by solving the math problem. When a solution is reached it is verified by the rest of the network. This allows for the tracking of the life of every individual bitcoin produced.

Thus for any individual to double spend their Bitcoins, their computing power would have to exceed the combined computing power of all other Bitcoin computers.

Alternatives to Bitcoin had begun to spring up in 2013. The largest competitor was Litecoin, which as of December 2, 2013 has a market capitalization of $695,376,891 USD [8]. Litecoin seeks to be an improvement over Bitcoin by attempting to overcome some of the more technical issues facing Bitcoin.

2.3 The Data

In our empirical study, we consider the Bitcoin/USD exchange rate from the first part of year 2013 that includes a bubble, which burst on April 10, 2013.

More specifically, the sample consists of 150 observations [4] on the daily closing values of the Bitcoin/USD exchange rate over the period February 20 -July 20, 2013. The dynamics of the data is displayed in Figure 2.2.

We observe a nonlinear trend as well as the bubble that peaked at the virtual time

[4] All data were retrieved from Blockchain.info. The public ledger of all bitcoin transactions is recorded and available for public use on the site.

t=50. The sample median, interquartile range and total range [5] are 103.27, 46.69 and 208.21, respectively. For comparison, the sample mean and variance are 96.98 and 1327.63, respectively. These standard summary statistics can be misleading, since their usual interpretation assumes the stationarity of the process. This assumption is clearly not satisfied for the Bitcoin/USD exchange rate. Figure 2.3 shows the histogram and a kernel-based density estimate of the sample marginal density.

Both estimates display fat tails, as suggested by the fact that the total range is five times greater than the interquartile range. Also, the histogram indicates a trough in the left tail, which appears as an almost bimodal pattern in the kernel-smoothed density estimate.

3 The Model

This section presents the mixed causal-noncausal autoregressive process and explains how it can accommodate the bubble effects observed in the Bitcoin/USD exchange rate series. The estimation and inference methods are also discussed.

3.1 The Noncausal and Mixed Autoregressive Process

A mixed (causal-noncausal) autoregressive process is a stochastic process $\{y_t;\ t = 0, \pm 1, \pm 2, ...\}$, defined by:

$$\Psi(L^{-1})\Phi(L)y_t = e_t, \tag{3.1}$$

where $\Psi(L^{-1})$ and $\Phi(L)$ are polynomials in the negative (resp. positive) powers of the lag operator L, such that $\Psi(L^{-1}) = 1 - \psi_1 L^{-1} - \cdots - \psi_s L^{-s}$ and $\Phi(L) = 1 - \phi_1 L - \cdots - \phi_r L^r$.

[5] The difference between the sample max and min

The roots of both polynomials are assumed to lie outside the unit circle, and error terms e_t are identically and independently distributed. When $\phi_1 = \cdots = \phi_r = 0$, model (3.1) defines a pure noncausal autoregressive process of order s, while for $\psi_1 = \cdots = \psi_s = 0$, the process y_t is the traditional pure causal AR(r) process. When some of the coefficients of both polynomials are non-zero, we obtain a mixed process that contains both the lags and leads of y_t. Under the above assumptions, there exists a unique stationary solution to equation (3.1). This solution admits a strong, two-sided moving average representation:

$$y_t = \sum_{j=-\infty}^{\infty} \xi_j e_{t-j},$$

where the ξ_j's are the coefficients of an infinite order polynomial in positive and negative powers of the lag operator L and such that: $\Xi(z) = \sum_{j=-\infty}^{\infty} \xi_j z^j = [\Psi(z^{-1})]^{-1}[\Phi(z)]^{-1}$.

When errors e_t are normally distributed, the causal and noncausal components of the dynamics cannot be distinguished, and model (3.1) is not identifiable. However, the causal and noncausal autoregressive coefficients are identifiable when the process (e_t) is not Gaussian [6]. For example, Lanne, Saikkonen (2010) consider the t-Student distributed errors, while Gourieroux and Zakoian (2013) introduce the Cauchy distributed errors, characterized by long-tailed marginal density function. The density of a Cauchy distributed random variable X with location μ and scale γ is:

$$g(e_t; \mu, \gamma) = \frac{1}{\pi}\left[\frac{\gamma}{(x-\mu)^2 + \gamma^2}\right] \tag{3.2}$$

Expression (3.1) with errors $e_t \sim Cauchy(0, \gamma)$ defines a mixed autoregressive model with Cauchy distributed errors. A particular feature of the Cauchy distribution is that the

[6] See e.g. Cheng (1992), Rosenblatt (2000), Theorem 1.3.1. for errors with finite variance, Breidt, Davis (1992) for errors with finite expectation and infinite variance, Gourieroux, Zakoian (2015) for errors without finite expectation, as the Cauchy errors considered in the application.

expected value as well as all population moments of any higher order do not exist, due to the occurrence of extreme values.

3.2 The Bubble Effect

The trajectory of the Bitcoin/USD exchange rate displays repetitive episodes of upward trends, followed by instantaneous drops, which are called the bubbles. In general, a bubble has two phases: 1) a phase of fast upward (or downward) departure from the stationary path that resembles an explosive pattern and displays an exponential rate of growth, followed by 2) a phase of sudden almost vertical drop (or upspring) back to the underlying fundamental path. There exist several definitions of a bubble in the economic literature. The first definition was introduced by Blanchard (1979) in the framework of rational expectation models. The formal definition by Blanchard as well as the later definitions by Blanchard and Watson (1982) and Evans (1991) all assume a nonlinear dynamic models of x_t (say) with two components, one of which depicts the fundamental path of x_t, while the second one represents the bubble effect. The economic explanation of this phenomenon is as follows: a bubble results from the departure of a price of an asset from its fundamental value. In the context of the Bitcoin/USD exchange rate, the bubbles may result from the speculative trading that makes the rate deviate quickly above its trend, although it is hard to say if the trend is representative of the fundamental value of the bitcoin. Indeed, the bitcoin is a virtual currency, which is backed neither on a real asset, nor on the performance of a firm or a national economy.

So far, the bubbles were considered in the time series literature as nonstationary phenomena and treated similarly to the explosive, stochastic trends due to unit roots. In fact, the existing tests for the presence of a bubble are essentially tests of a breakpoint in the general explosive stochastic trend of a nonstationary process [see e.g. Phillips, Wu

and Yu (2011) Phillips, Shi and Yu (2012)].

Gourieroux and Zakoian (2013a) propose a different approach and assume that the bubbles are rather short-lived explosive patterns caused by extreme valued shocks in a noncausal, stationary process. Formally, they propose to model that process as a noncausal AR(1) model with Cauchy distributed errors, because the noncausal model in reverse time allows for accommodating the asymmetric pattern of the bubble. The merit of the Cauchy distributed errors is in replicating the sudden spike in the reverse time trajectory that is observed as a bubble burst from the calendar time perspective. Such a noncausal or mixed process has to be examined conditionally on the information of the current and past rates. It is known that a noncausal, linear autoregressive process also has a nonlinear causal autoregressive dynamics, except in the Gaussian case. This is the special nonlinear feature, which makes it suitable for modelling the bubbles in Bitcoin/USD exchange rate. Moreover the noncausal autoregressive model allows for forecasting the occurrence of a future bubble and the time of bubble burst. The methodology of forecasting is discussed in Section 3.4 and illustrated in the application in Section 4.

3.3 Estimation and Inference

The traditional approach to the estimation of causal time series models relies on the Box-Jenkins methodology that consists of three steps: identification, estimation and diagnostics. In application to noncausal and mixed processes, most of the traditional Box-Jenkins tools of analysis need to be interpreted with caution. The reason is that most of the traditional estimators are based on the first- and second-order sample moments of the process and rely on the Gaussian approximation of its density, while the noncausal processes need to be non-Gaussian to solve the aforementioned identification problem and may have infinite moments of order one and/or two.

a) *Identification*

The autocorrelation function (ACF) is the basic tool for detecting temporal dependence. By construction, the ACF estimators rely on a implicit normality assumption, as they are computed from the sample moments up to order two. Due to the aforementioned nonidentifiability problem, the ACF cannot reveal whether a time series is causal or not, as it yields identical results in either case. It remains however a valid tool for detecting serial dependence in variables with infinite variances [see Andrews and Davis (2013)]. In particular Andrews and Davis show that the total autoregressive order $p = r + s$ can be inferred from the autocorrelation function, while r and s need to be inferred from the estimated models by comparing their fit criteria, computed from the sample.

For variables with infinite variance, Davis and Resnick (1985) established the asymptotic properties of the sample autocorrelation $\hat{\rho}$ at lag l defined as:

$$\hat{\rho}(l) = \frac{\widehat{cov}(l)}{\widehat{cov}(0)},\ \widehat{cov}(l) = \frac{1}{T}\sum_{t=1}^{T-l}(y_t - \bar{y})(y_{t+l} - \bar{y}),\ l > 0,$$

where T is the sample size and $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$.

In the presence of Cauchy errors, the standard confidence intervals of the ACF are no longer valid as the sample ACF is no longer asymptotically normally distributed and has a nonstandard rate of convergence. By using the results of Davis and Resnick (1985), Gourieroux and Zakoian (2013a) (Proposition 6) show that the sample autocorrelations of a noncausal AR(1) process with Cauchy errors and autoregressive coefficient ρ have a limiting stable distribution and a rate of convergence that is different from the standard $\sqrt{T}$ rate. More specifically, let us denote the vector of sample autocorrelations up to lag M: by $\hat{\rho}_T = (\hat{\rho}_T(1), ..., \hat{\rho}_T(M))'$ and consider the true values $\rho = (\rho, ..., \rho^M)'$. Then,

$$\frac{T}{\ln T}(\hat{\rho}_T - \rho) \xrightarrow{d} Z = (Z_1, ..., Z_M),'$$

where for $l = 1, ..., M$, $Z_l = \sum_{j=1}^{\infty}[\rho^{j+l} - \rho^{|j-l|}]S_j/S_0$, and $S_1, S_2...$ is an i.i.d. sequence of symmetric 1-stable random variables independent of the positive 1/2 stable random variable S_0. The limiting true values can be interpreted as pseudo-autocorrelations, as the autocorrelations themselves do not exist in a process with infinite variance.

b) *Estimation*

The standard Gaussian quasi-maximum likelihood approach can no longer be used to estimate the autoregressive parameters due to the Gaussian-specific identification problem. However, when the distribution of the errors is non-Gaussian, the estimation of the parameters in noncausal and mixed processes can be based on the maximum likelihood estimator [see, Lanne, Saikkonen (2010)]. The maximum likelihood method differs slightly from that used in causal processes. It is called the "approximate maximum likelihood" for the reason that the sample used in the approximate likelihood is reduced to $T - (r + s)$ observations[7]. Indeed, the first error to be included in the likelihood function that can be written without a value of y_t prior to the sample is e_{r+1}. To see that, assume $\psi_1 = \cdots = \psi_s = 0$ and write:

$$e_{r+1} = y_t - \phi_1 y_{t-1} - \cdots - \phi_r y_{t-r}.$$

Suppose now that $\phi_1 = \cdots = \phi_r = 0$ The last error in the sample to be included in the

[7] The approximate likelihood disregards the first r state variables that summarize the effect of shocks before time r and the last s state variables that summarize the effect of shocks after time $T-s$ [Gourieroux, Jasiak (2013)] and is therefore constructed from shocks $e_{r+1}, ...e_{T-s-1}$ only.

likelihood function that can be written without the values of y_t posterior to the sample is

$$e_{T-s-1} = y_T - \psi_1 y_{T+1} - \cdots - \psi_s y_{T+s}$$

The Approximate Maximum Likelihood (AML) estimator is defined as:

$$(\hat{\Psi},\, \hat{\Phi},\, \hat{\theta}) = Argmax_{\Psi,\Phi,\theta} \sum_{t=r+1}^{T-s} \ln\, g[\Psi(L^{-1})\Phi(L)y_t;\theta], \tag{3.3}$$

where $g[.;\theta]$ denotes the probability density function of e_t.

c) *Diagnostics*

The diagnostic checking consists of testing if the estimated shocks $\hat{e}_t = \hat{\Psi}(L^{-1})$ $\hat{\Phi}(L)y_t$ of the model are strong white noise. The asymptotic distribution of the sample autocorrelation of the residuals is different from the standard one derived for processes with finite variance. For instance, for a noncausal Cauchy autoregressive process of order 1, the limiting distribution of the residual autocorrelation estimator at lag 1 is:

$$\frac{T}{\ln T}\hat{\rho}_T(1) \xrightarrow{d} \rho^*(1+2\rho^*)S_1/S_0,$$

where ρ^* is the noncausal autoregressive coefficient of process Y. Contrary to the standard process with finite variance, the limiting distribution depends on ρ^*.

3.4 Forecasting

Due to the different dynamics of non-Gaussian processes in the calendar and the reverse times, the "backcasting" algorithm in the spirit of Newbold [Biometrika (1974)] is no longer valid. Nevertheless, it is possible to extend the concept of the Kalman filter and make it applicable to noncausal and mixed processes. The approach involves the errors and

the unobserved causal and noncausal components of the process. For example, the mixed causal-noncausal process (3.1) can be written as a function of error e_t and decomposed into a causal component v and a "noncausal" component u as follows:

$$y_t = \frac{1}{\Phi(L)} \frac{1}{\Psi(L^{-1})} e_t, \tag{3.4}$$

where the noncausal component is defined as $u_t = \Phi(L)y_t$ and satisfies $e_t = \Psi(L^{-1})u_t$, and the causal component is written as $v_t = \Psi(L^{-1})y_t$ and satisfies $e_t = \Phi(L)v_t$.

It follows that the noncausal component u_t (resp. the causal component v_t) is a combination of current and future values (resp. of the current and lagged values) of the error e_t.

The noncausal component u_t is next used for forecasting the future values $y_{T+1}, y_{T+2}, ...,$ y_{T+H}. More specifically, the forecasting procedure is based on the functional closed-form estimator of the predictive density of the unobserved component u. Next, that estimated predictive density is used to generate the future values of processes (u) and (y) over a given horizon H [see, Gourieroux, Jasiak (2014)]. In practice, the procedure consists of the following steps:

Step 1: Estimate the unknown coefficients Φ, Ψ and γ, compute the residuals $\hat{e}_t$ and filter the causal and noncausal state components of the process $\hat{v}_t$ and $\hat{u}_t$.

Step 2: Estimate the predictive density of the noncausal component u_t. For example for $s = 1$, the predictive distribution at horizon H is:

$$
\begin{aligned}
&\hat{\Pi}(u_{T+1}, \ldots, u_{T+H} | \hat{u}_T) \\
&= \left\{ \hat{g}(\hat{u}_T - \hat{\psi} u_{T+1}) \hat{g}(u_{T+1} - \hat{\psi} u_{T+2}) \ldots \hat{g}(u_{T+H-1} - \hat{\psi} u_{T+H}) \sum_{t=1}^{T-1} \hat{g}(u_{T+H} - \hat{\psi} \hat{u}_t) \right\} \\
&\quad \left\{ \sum_{t=1}^{T-1} \hat{g}(\hat{u}_T - \hat{\psi} \hat{u}_t) \right\}^{-1}
\end{aligned}
\tag{3.5}
$$

where $\hat{g}$ is the estimated error density.

Step 3: Simulate the future noncausal components $u^s_{T+1}, u^s_{T+2}, ..., u^s_{T+H}$ by using a sampling importance resampling (SIR) algorithm [see, Appendix A] and compute the future values of the process $y^s_{T+1}, y^s_{T+2}, ..., y^s_{T+H}$ and future unobserved components e and v by using the recursive formulas that define them and are given above.

This methodology is used in Section 4.4 to derive the Bitcoin/USD rate predictions and prediction intervals at short and medium horizons of up to one week.

4 Application

4.1 ACF Analysis

The traditional Box-Jenkins approach starts from the analysis of the sample autocorrelation function (ACF). The ACF provides information on the possible linear serial dependence in the series, but its interpretation can be rather misleading in the case of extreme events.

The standard confidence interval for testing the statistical significance of the auto-

correlations is based on the approximate limiting standard normal distribution of the autocorrelation estimator at a given lag, under the null hypothesis that the true value of that autocorrelation is zero. Hence, with the sample size of 150, the statistically significant autocorrelations exceed 0.16 in absolute value.

In order to establish the confidence interval for Cauchy distributed errors, we approximate the limiting distribution of the pseudo-autocorrelation estimator given in Section 3.3 by simulations. We draw independent standard normals Z_1, Z_2, Z_3 and build the ratio

$$SR = \frac{S_1}{S_0} = \frac{Z_1 Z_3^2}{Z_2}$$

where $S_1 = \frac{Z_1}{Z_2}$ is a symmetric 1-stable random variable and $S_0 = \frac{1}{Z_3^2}$ is a symmetric 0.5 stable random variable. From the 25th and 975th order statistics in a sample of 1000 values of SR multiplied by $\frac{\ln(150)}{150}$, we obtain the confidence interval $[-0.36, 0.36]$. Under the null hypothesis of zero pseudo-autocorrelation at lag l, the statistically significant autocorrelation at lag l is less than 0.36 in absolute value with the asymptotic probability of 95%.

In Figure 2.4, we plot the ACF of the data with the standard confidence interval and the interval adjusted for infinite variance.

The ACF displays slow, linear decay, which resembles the patterns observed in unit root processes. Moreover, the Dickey-Fuller and the Augmented Dickey-Fuller ADF(4) tests fails to reject the null hypothesis of a unit root in the data with p-values 0.4 and 0.6, respectively. However, it is easily checked that the standard procedure of transforming the data into first differences and estimating a stationary ARMA cannot accommodate the asymmetric pattern of the bubble and other nonlinear features of the series.

4.2 Global and Local Trends

In the Bitcoin/USD exchange rate series, it is important to disentangle the fundamental and the bubble components. The fundamental component is modelled as a nonlinear deterministic trend [8] and the bubble component as a noncausal autoregressive process with Cauchy errors. Accordingly, we define the Bitcoin/USD rate as:

$$rate_t = trend_t + y_t,$$

a) *Estimation of the trend and detrended series*

In order to remove the trend, we fit a nonlinear function of time by regressing the data on a 3rd degree polynomial in time. The detrended series, obtained as the following series of residuals:

$$y_t = rate_t + 3.045 - 3.854 trend_t + 3.499 trend_t^2 - 0.866 trend_t^3$$

is calculated and plotted in Figure 2.5. The marginal density of y_t is shown in Figure 2.6.

We observe that the detrended series no longer displays the bimodal pattern, while it preserves the peaked and long-tailed shape of the density of the Bitcoin/USD rate.

The ACF function of the detrended series, given in Figure 2.7, shows considerably less persistence than the original series and indicates short linear memory.

b) *Noncausal analysis of the detrended series*

The detrended series is modelled as a mixed autoregressive process with Cauchy distributed errors. In order to find the optimal fit, we explore several specifications. We report the results of the estimation of two pure noncausal models, a pure causal model,

[8] Alternatively, it can be represented by a model with a stochastic trend, assumed independent of the shocks that create the speculative bubble.

and three mixed causal-noncausal specifications below. The parameter estimate of all models are given in Table 2.1. As the limiting distribution of the AML estimators of the autoregressive coefficients in polynomials Φ and Ψ are intractable due to Cauchy errors [see, Andrews, Calder, Davis(2009), Th. 3.2], the standard errors are computed by bootstrap [see, Davis, Wu (1997), Andrew Calder, Davis (2009) Th. 3.4] and given in parentheses. The values of the t-ratios can be interpreted as measures of the relative accuracies of the estimators . The AML estimator of the variance parameter γ is asymptotically normal and standard inference methods can be applied.

The models are compared in terms of the in-sample median squared error, mean squared error (MSE), and the value of the log-likelihood function at the maximum. The values are given in the last rows of Table 2.1.

The pure noncausal Cauchy AR(1) process is:

$$y_t = \psi y_{t+1} + e_t, \tag{4.6}$$

where e_t are independent and Cauchy distributed with location 0 and scale γ. In the AR(1) case the noncausal polynomial is $\Psi(L^{-1}) = 1 - \psi_1 L^{-1}$ [9]. The model is estimated by maximizing the approximated log-likelihood function, based on the Cauchy density function:

$$\ln L(\psi, \gamma) = (T-1)[-\ln(\pi) + \ln(\gamma)] - \sum_{t=1}^{T-1}[\ln((y_t - \psi y_{t+1})^2 + \gamma^2)]. \tag{4.7}$$

The residuals of the noncausal AR(1) display no significant serial autocorrelation. Their pattern indicates that the asymmetry due to the bubble has been successfully

[9] It is interesting to compare the trajectory of y_t with the simulated path of a noncausal AR(1) with the autoregressive coefficient 0.9, as displayed in Gourieroux, Zakoian (2013a), Figure 4. It is clear that the trajectory of the transformed Bitcoin/USD rate resembles the simulated series.

removed.

The pure noncausal AR(2) model is:

$$y_t = \psi_1 y_{t+1} + \psi_2 y_{t+2} + e_t, \tag{4.8}$$

The estimates indicate that both roots of the noncausal polynomial $\Psi(L^{-1}) = 1 - \psi_1 L^{-1} - \psi_2 L^{-2}$ are outside the unit circle. The residuals of the noncausal AR(2) are also white noise, as their autocorrelations are not statistically significant. Moreover, the noncausal AR(2) slightly outperforms the AR(1) in term of the in-sample MSE.

The next specifications considered are mixed autoregressive models. The MAR(1,1) has both causal and noncausal orders equal to 1:

$$(1 - \phi L)(1 - \psi L^{-1}) y_t = e_t. \tag{4.9}$$

Both roots of the estimated polynomials lie outside the unit circle.

Note that the noncausal AR(2) process and the MAR(1,1) process are not equivalent from the modeling point of view. Noncausal parameters ψ_j (resp. causal parameters ϕ_j) have a significant impact on the rate of increase of the bubble (resp. decrease of the bubble). The noncausal AR(2) model is able to fit bubbles with two possible rates of increase, corresponding to ψ_1 and ψ_2, but with very sharp decrease due to the absence of causal autoregressive parameter. The mixed MAR(1,1) model is flexible enough to fit any asymmetric bubbles.

The MAR(2,2) model is:

$$(1 - \phi_1 L - \phi_2 L^2)(1 - \psi_1 L^{-1} - \psi_2 L^{-2}) y_t = e_t. \tag{4.10}$$

The estimated autoregressive polynomials have real-valued roots outside the unit circle.

This model is outperformed by the MAR(1,1) in terms of MSE. Given that the value of $\hat{\psi}_2$ does not exceed twice the value of the standard error, we also estimate the MAR(2,1):

$$(1 - \phi_1 L - \phi_2 L^2)(1 - \psi_1 L^{-1})y_t = e_t. \tag{4.11}$$

The estimated autoregressive polynomial in past $y's$ of this model has real-valued roots out of the unit circle. Its MSE and median squared error are higher than that of the MAR(2,2).

A causal AR(2) model is also reported for comparison. The roots of the autoregressive polynomial both lie outside of the unit circle and the residuals are white noise. The value of the MSE is similar to that of the MAR(1,1) however the median squared error is greater than that of the mixed causal-noncausal models.

By comparing all the estimated models, we find that the best fit in terms of the in-sample MSE is provided by the MAR(1,1) and the noncausal AR(1) models. Upon examining the median of the squared errors, the MAR(1,1) model has a smaller value (6.63 vs. 7.24).

The residuals $\hat{e}_t$ and the filtered components $\hat{u}_t$ and $\hat{v}_t$ (see Section 3.4) of the MAR(1,1) are shown in Figure 2.8. We observe that extremely large $\hat{e}_t$ accommodate the bubble and eliminate its asymmetry. The noncausal component $\hat{u}_t$ has an effect on the growth of the bubble and captures the forward looking persistence. In the following section,this filtered component along with the observed value of the series will be used for out-of-sample forecasting at horizons up to one week.

Table 2.1: Parameter Estimated and In-Sample Fit

	NAR(1)	NAR(2)	MAR(1,1)	MAR(2,1)	MAR(2,2)	AR(2)
ψ_1	0.9122	1.316	0.678	0.632	0.739	
	(0.053)	(0.062)	(0.063)	(0.046)	(0.025)	
ψ_2		-0.401			0.032	
		(0.061)			(0.023)	
ϕ_1			0.717	0.664	0.501	1.350
			(0.066)	(0.037)	(0.063)	(0.073)
ϕ_2				0.033	0.144	-0.423
				(0.033)	(0.027)	(0.076)
γ	2.734	2.433	2.559	2.481	2.413	2.513
	(0.113)	(0.112)	(0.109)	(0.112)	(0.114)	(0.109)
$-\ln L$	496.135	478.709	479.402	470.658	471.507	480.92
MSE	131.754	126.234	129.423	138.391	135.5145	128.311
Median SE	7.235	10.758	6.632	6.582	6.291	7.427

4.3 Prediction

This section illustrates the forecasting from the estimated noncausal models. First we discuss the forecasting at short horizons from the MAR(1,1) model that outperformed other models in terms of the parsimony and the in-sample fit to the data . Next, we compare the predictive properties of all models estimated in the previous section at short and medium horizons of up to one week. The forecasting is based on the method outlined in Section 3.4, which involve estimating the predictive density of $u_{T+1}, ..., u_{T+H}|\hat{u}_T$ and then using the SIR method in order to simulate future noncausal components from which the future y are computed. The sample sizes for sampling and resampling are of 1000 and 300, respectively.

4.4 Short-Term Forecasting from MAR(1,1)

Let us first consider the estimated MAR(1,1) model and predict the density of detrended process y at horizon $H = 1$.

Figure 2.9 displays the predicted density of y_{T+1}. We observe some asymmetry in the predicted density of y_{T+1}, which would not have been detected if a standard Gaussian ARMA model was used. It is centered at $y_{T+1} = 13.6$ and displays a long left tail suggesting a small but non-zero probability of a sudden significant drop in the Bitcoin exchange rate.

Figure 2.10 shows the estimated joint predicted density of the bitcoin series at horizons $H = 1$ and $H = 2$, which is also displayed as the contour plot in Figure 2.11.

The joint predicted density of y_{T+1} and y_{T+2} is very different from the Gaussian and displays strong dependence in extreme future risks in some directions [10]. For the MAR(1,1) process, that predicted joint density depends on two state variables, which are y_T, u_T, or equivalently y_T, y_{T-1} and is centered at (y_T, y_T). The values of the state variables for the Bitcoin data are : $u_T = 2.87$, y_{T-1}=9.64, $y_T = 12.27$. Thus, at the end of the sampling period, the detrended series is increasing. We see that there exist several different patterns of future exchange rate dynamics which can be a continued increase, an increase followed by a slow decrease, or a sharp decrease. To clarify the various future patters, we provide a more intuitive display in Figure 2.12. By considering Figure 2.12, we see that the probability of a continuing increase of y (North-East orthant) is rather high, but so is the probability of a sharp downturn at date $T + 1$ (South-West orthant). However, the probability of a downturn at date $T + 2$ (South-East orthant) is small. Thus the joint predictive density can be used to recognize the future pattern

[10] Its associated copula is close to an extreme value copula [see e.g. Balrema, Embrechts and Nolde (2013) for examples of extreme value copulas].

of y by comparing the likelihood of the different scenarios, in particular to evaluate the probability of the downturn at dates T+1, T+2, etc. The above discussion based on the graphical representation is limited to horizon 2. The probabilities of the different types of future patterns can be evaluated numerically at larger H.

4.5 Short- and Medium-Term Forecasting

In this section, we assess the predictive power of the estimated processes up to one week ahead, or equivalently, up to horizon $H = 7$. To do that, we predict in one step the densities of future values of the series up to $H = 7$ from all processes estimated in the previous section. This allows us to compute and compare the forecasts from the MAR(1,1), MAR(2,2), MAR(2,1), the pure noncausal AR(1) and AR(2) processes, and the causal AR(2) process. At each horizon H, the out-of-sample point prediction is represented by the median and the prediction interval is represented by the 10th and 90th quantiles from the predicted density of y at each horizon. Next, the point predictions are compared to the true detrended Bitcoin/USD exchange rates at dates $T + 1, ..., T + H$, as reported on Blockchain.info. The extended sample is detrended by using the same procedure as outlined in Section 4.2.

Table 2.2 below shows the forecast errors as the differences between the point predictions and the true values of the Bitcoin series and reports their cumulated squared values.

We see that the MAR(1,1) outperforms the other models in terms of the predictive power up to $H = 7$.

For illustration, the forecasts of the MAR(1,1) process along with the corresponding seven true values of future bitcoin series, up to $H = 7$, are plotted in Figure 2.13.

The solid line represents the point predictions estimated from the median (50th quan-

Table 2.2: Cumulative Out-of-Sample Square Forecast Errors

Forecast Horizon	NAR(1)	NAR(2)	MAR(1,1)	MAR(2,1)	MAR(2,2)	AR(2)
1	0.6380	8.3233	1.4434	0.0616	0.3029	3.9943
2	1.5641	10.9769	10.9091	0.0940	2.3469	14.2747
3	70.7152	51.1850	11.0605	6.2228	62.4054	92.80178
4	179.3388	275.2328	96.1129	99.7600	582.2711	313.2267
5	462.2530	1013.6831	132.6133	372.2187	969.0964	690.0930
6	658.6162	2521.4349	299.8194	909.1155	1478.3370	1371.627
7	828.5468	2521.8963	524.4793	1042.5181	2679.3096	2089.915

tile), the dotted lines indicate the prediction intervals (the 10th and 90th quantiles) and the dashed line shows the true future values of the detrended Bitcoin/USD exchange rates.

We see that the true values of the bubble component of the Bitcoin series are close to the point forecasts, up to $H = 3$. The forecasted values deviate from the true values of the bubble component of the Bitcoin/USD exchange rate at longer horizons while remaining within the prediction interval. The prediction interval behaves as expected, gradually widening as the forecast horizon increases due to the changes in the left tail of the predictive density at longer horizons. Overall the forecasts perform well as the true values up to $H = 7$ all lie within the prediction interval.

Figure 2.14 displays the term structure of forecasts, i.e. the predicted univariate densities of future values $y_{T+1}, ..., y_{T+H}$.

The term structure shows the extent to which the model predicts sustained positive values in the bubble component. In fact, the peaks for the forecasted distributions are all located to the right of 0, forecasting sustained growth in the Bitcoin/USD exchange rate. Over time, the sample variance of the forecasts also increases. This is likely not due to the presence of extreme values as long tails in the distributions are only observed at y_{T+4} and at y_{T+7}. In particular, at y_{T+7} a small lobe in the left tail appears at about $y_{T+7} = -50$, implying a considerable chance of an occurrence of an extreme negative value

in the bubble process and a bubble burst.

5 Conclusion

The causal-noncausal autoregressive models have been proposed as nonlinear dynamic models for processes with speculative bubbles. We applied this methodology to examine the Bitcoin/USD exchange rates over the period February-July 2013, during which speculative bubbles have occurred. We have estimated pure and mixed model with the causal and noncausal orders up to two, by the Approximated Maximum Likelihood and filtered the underlying unobserved components of the process to better understand the type of observed bubbles. Next, the series of Bitcoin/USD exchange rates has also been used to compare the fit and forecasting power of various mixed and pure noncausal processes that capture bubble phenomena.

We also forecasted the Bitcoin/USD exchange rates at daily horizons of up to one week. The provided forecasts included the univariate predicted density of the future values, the point predictions and the prediction intervals. Moreover, the joint density of predictions for the next two consecutive days was used to discuss all possible future patterns of the process. This is kind of a model-based chartist approach, which can in particular be used to evaluate the likelihood of a future downturn.

The Bitcoin/USD exchange is an example of a highly speculative emerging market. Recently, several exchange platforms have closed, temporarily [11] or definitely. Other platforms were submitted to regulations. There is clearly a need for regulation and protection of the investors against the theft of bitcoins and the speculative behavior of large bitcoin holders. Such future regulations may reduce the speculative component in the Bitcoin

[11] The French platform Bitcoin-Central has been closed for five months in 2013 due to hackers attack. Nevertheless the customers had still the possibility to withdraw their bitcoins.

trades and cause the disappearance of the speculative bubbles or even of the market for this electronic currency itself.

However, there will still exist a large number of financial markets, not necessarily emerging, with frequently appearing bubbles. Examples are the markets for commodity futures, and the markets with high frequency trading. These are other potential applications for the causal-noncausal model presented in this chapter.

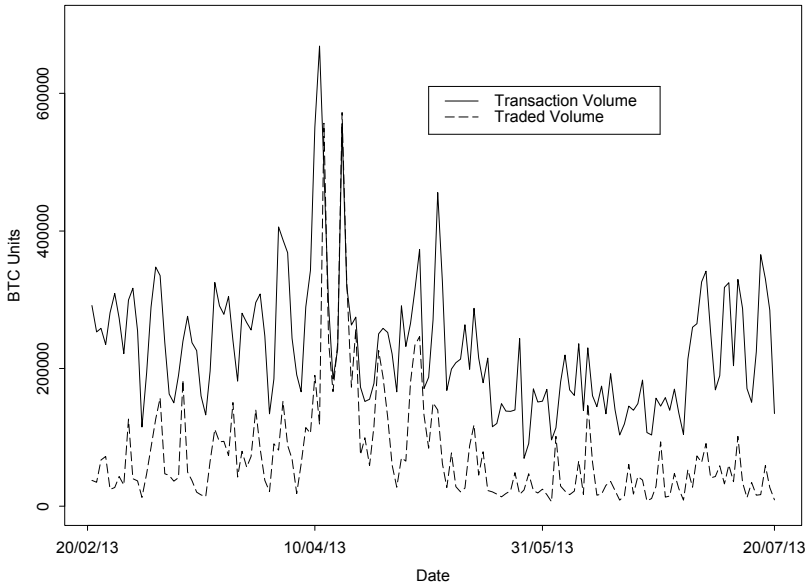


Figure 2.1: Bitcoin Daily Trading Volume and Transaction Count in BTC Units, Feb-July 2013

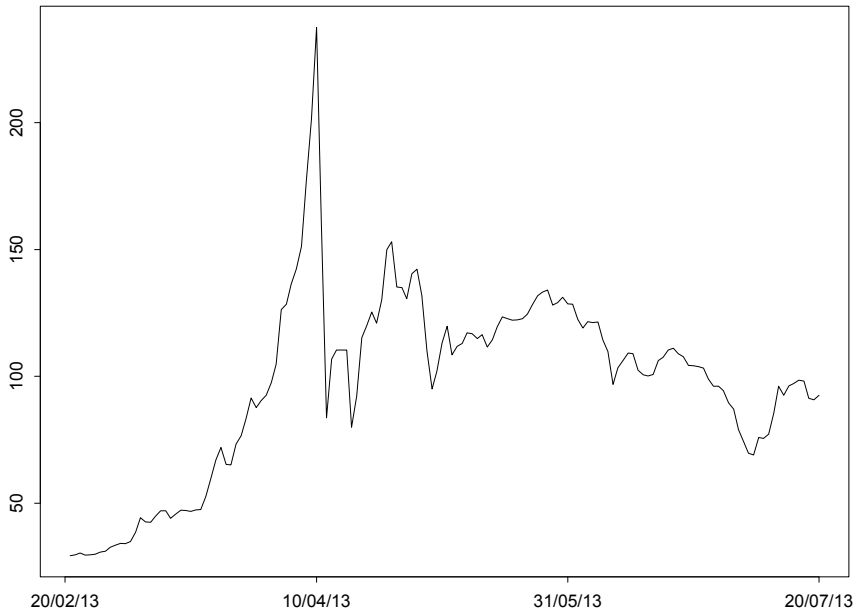


Figure 2.2: Bitcoin/USD Daily Exchange Rate, Feb-July 2013

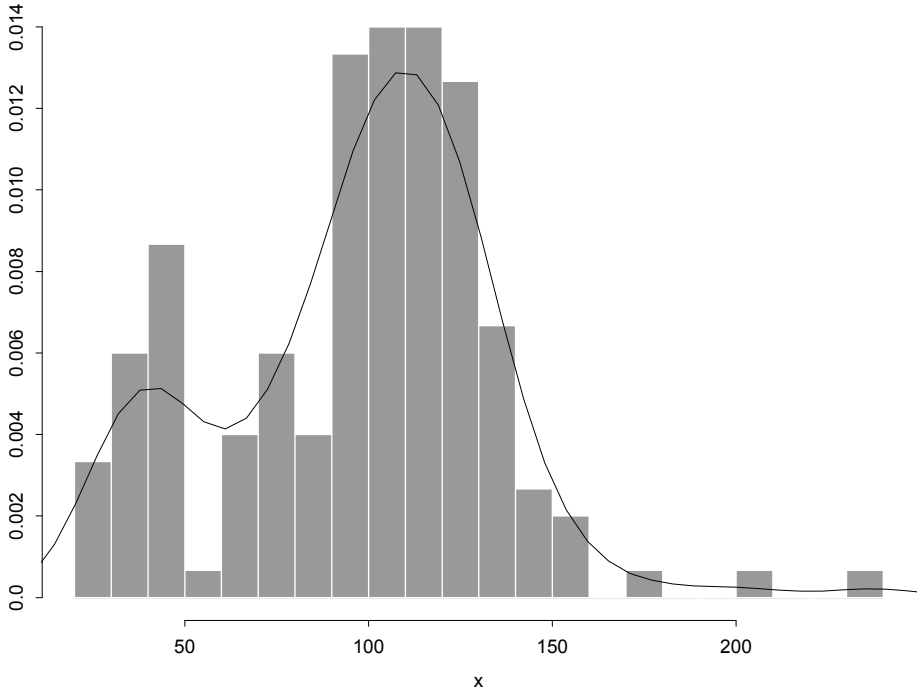


Figure 2.3: Bitcoin/USD, Histogram

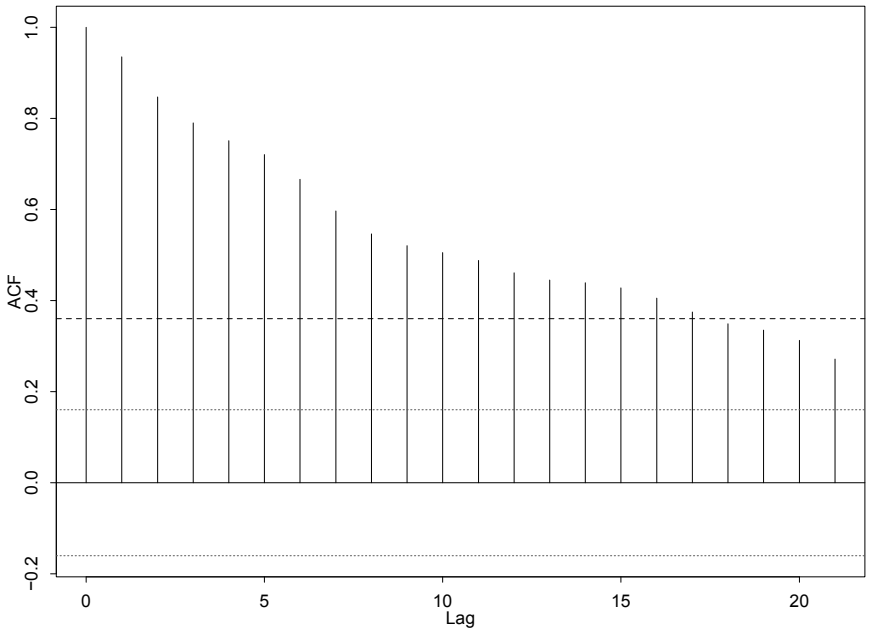


Figure 2.4: ACF, Bitcoin

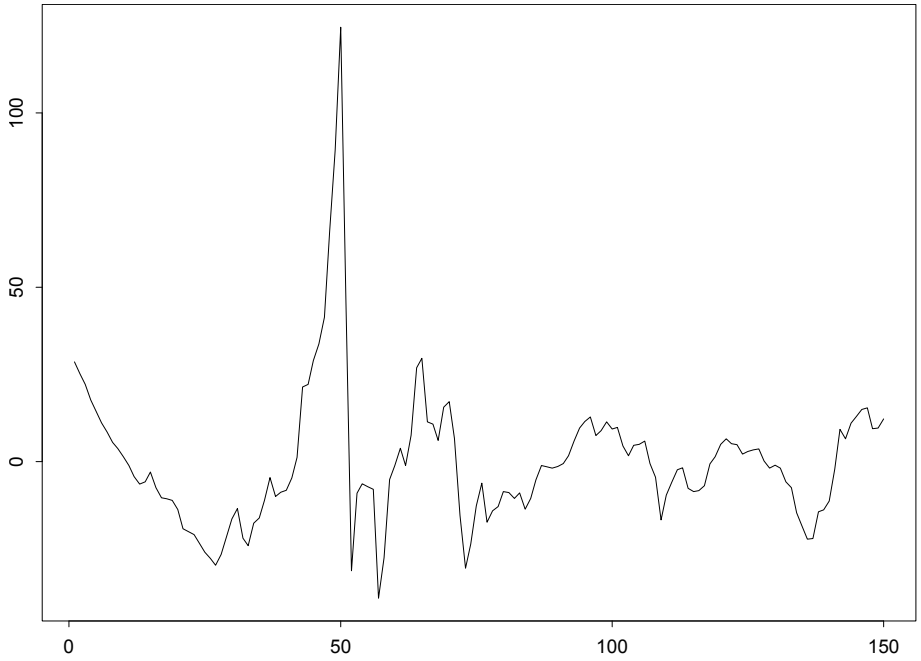


Figure 2.5: Detrended series

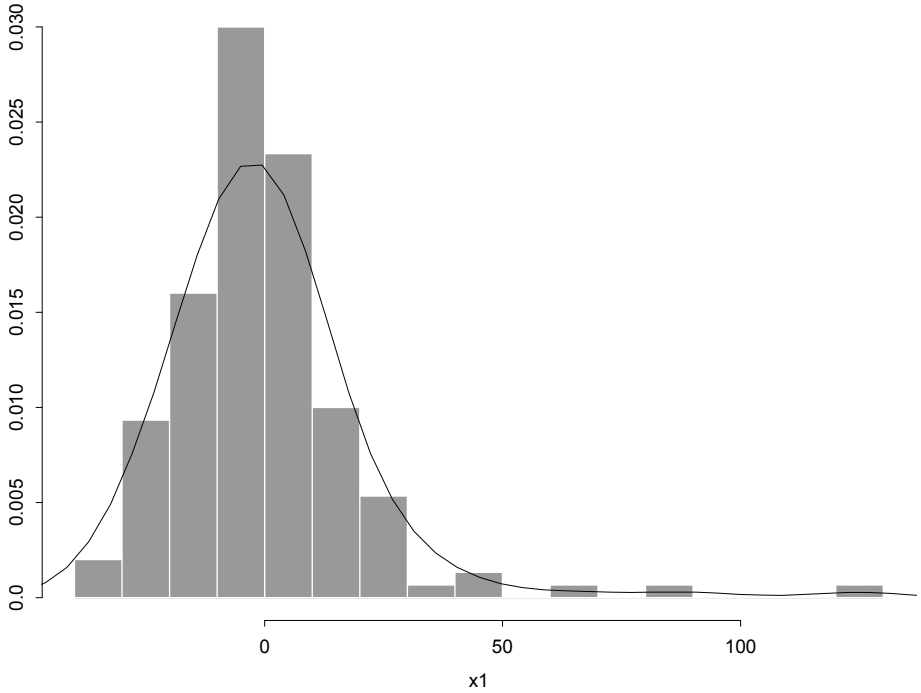


Figure 2.6: Detrended series, histogram

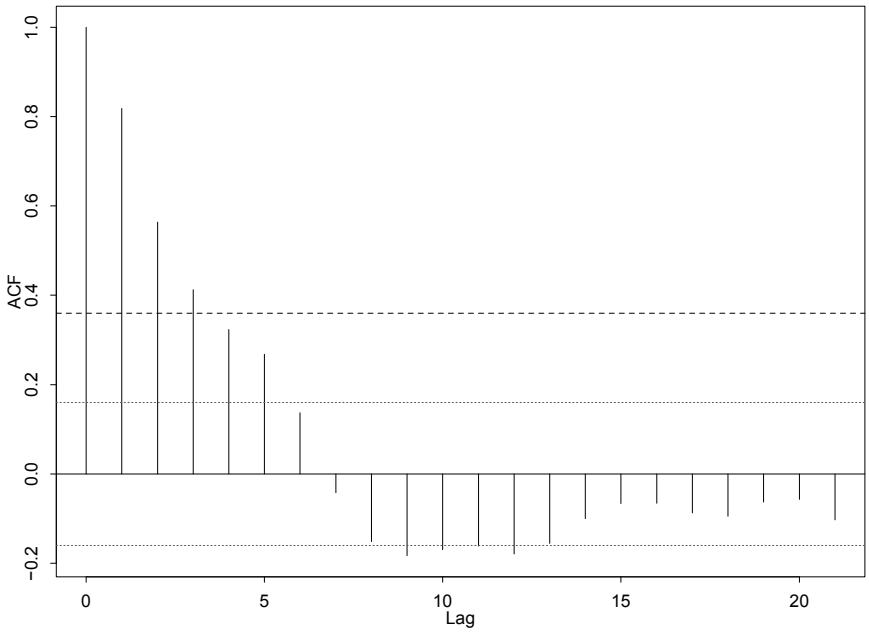


Figure 2.7: Detrended series, ACF

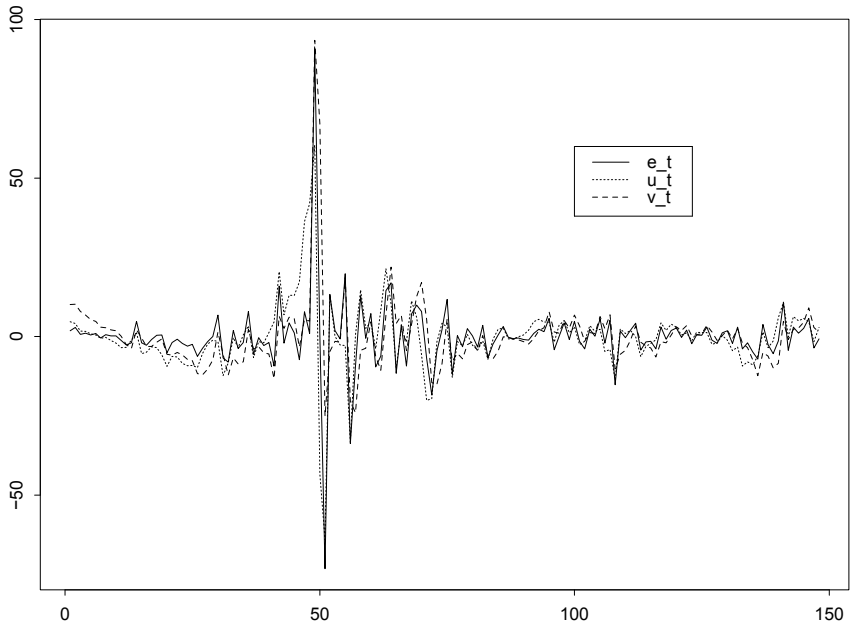


Figure 2.8: Components series of the MAR(1,1)

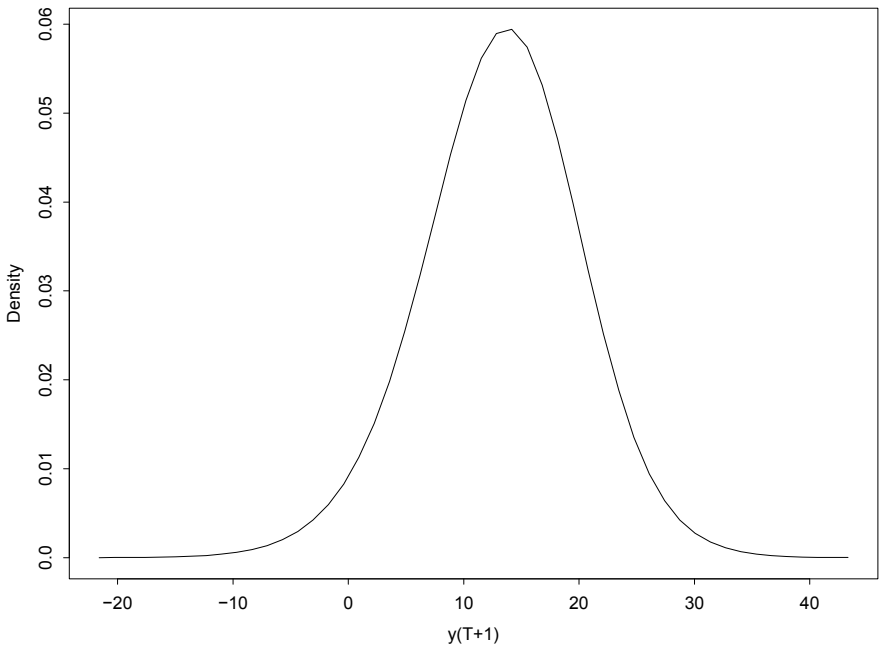


Figure 2.9: Predicted Density of y_{T+1}, MAR(1,1)

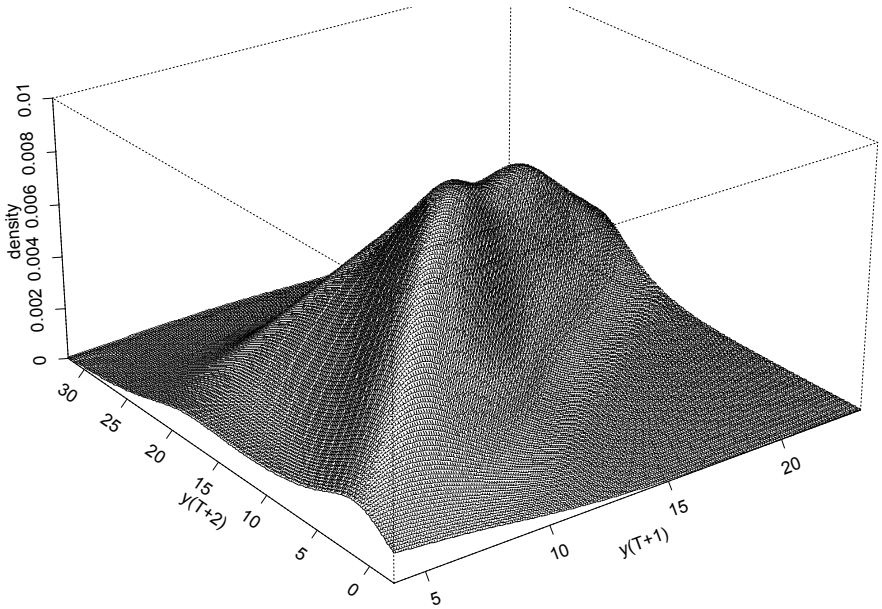


Figure 2.10: Joint Predicted Density of y_{T+1} and y_{T+2}, MAR(1,1)

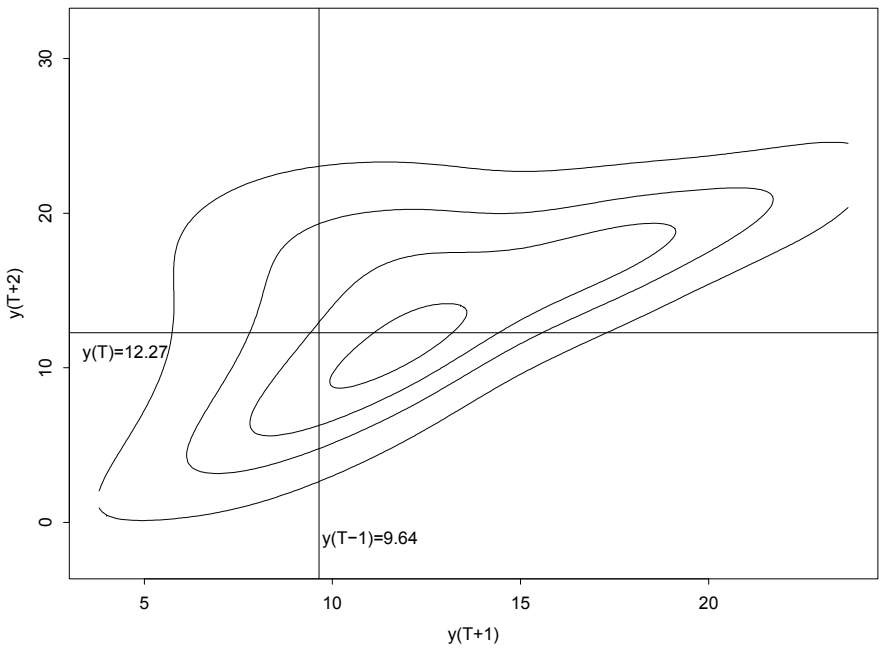


Figure 2.11: Contour Plot of Joint Predicted Density, MAR(1,1)

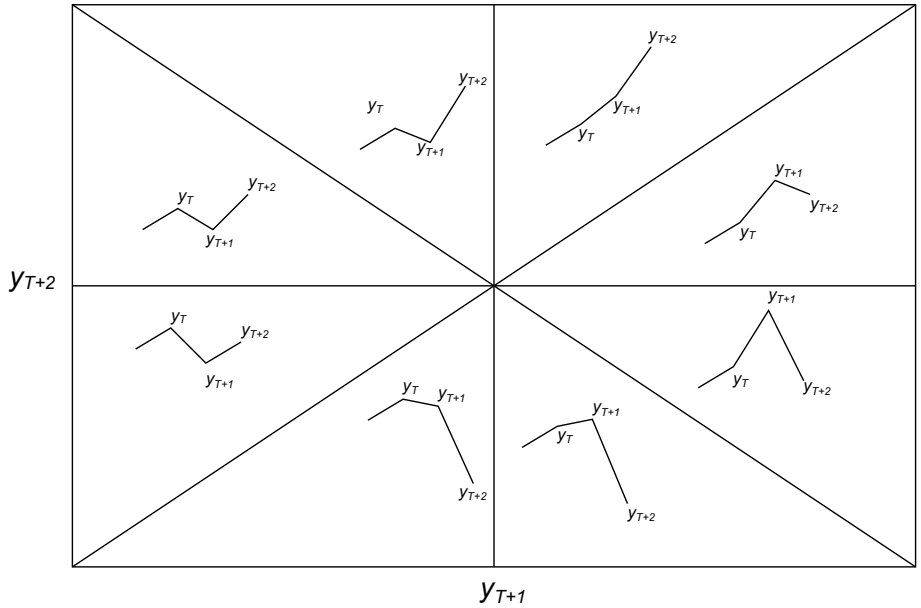


Figure 2.12: Predicted Dynamics of MAR(1,1)

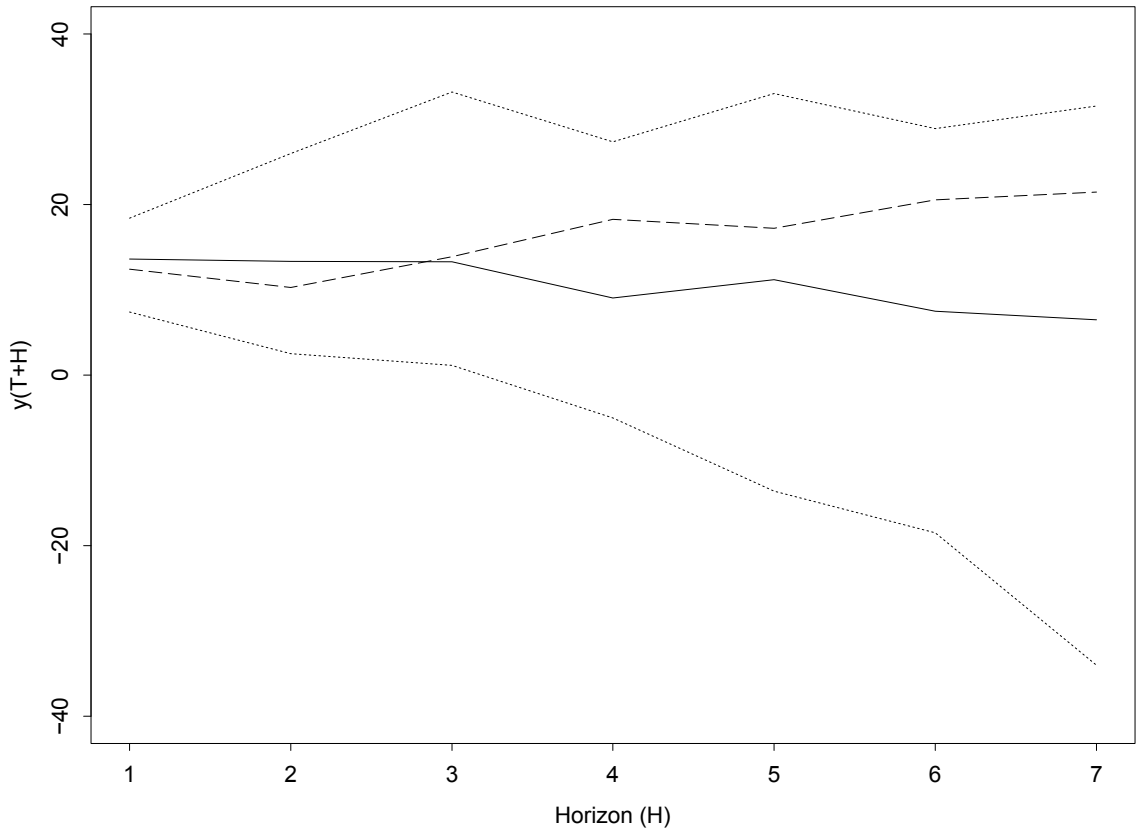


Figure 2.13: MAR(1,1) Out of Sample Forecasts at Horizon H=7 (one week)

Solid Line: Point Prediction, Dotted Line: Prediction Interval, Dashed Line: True Future Values

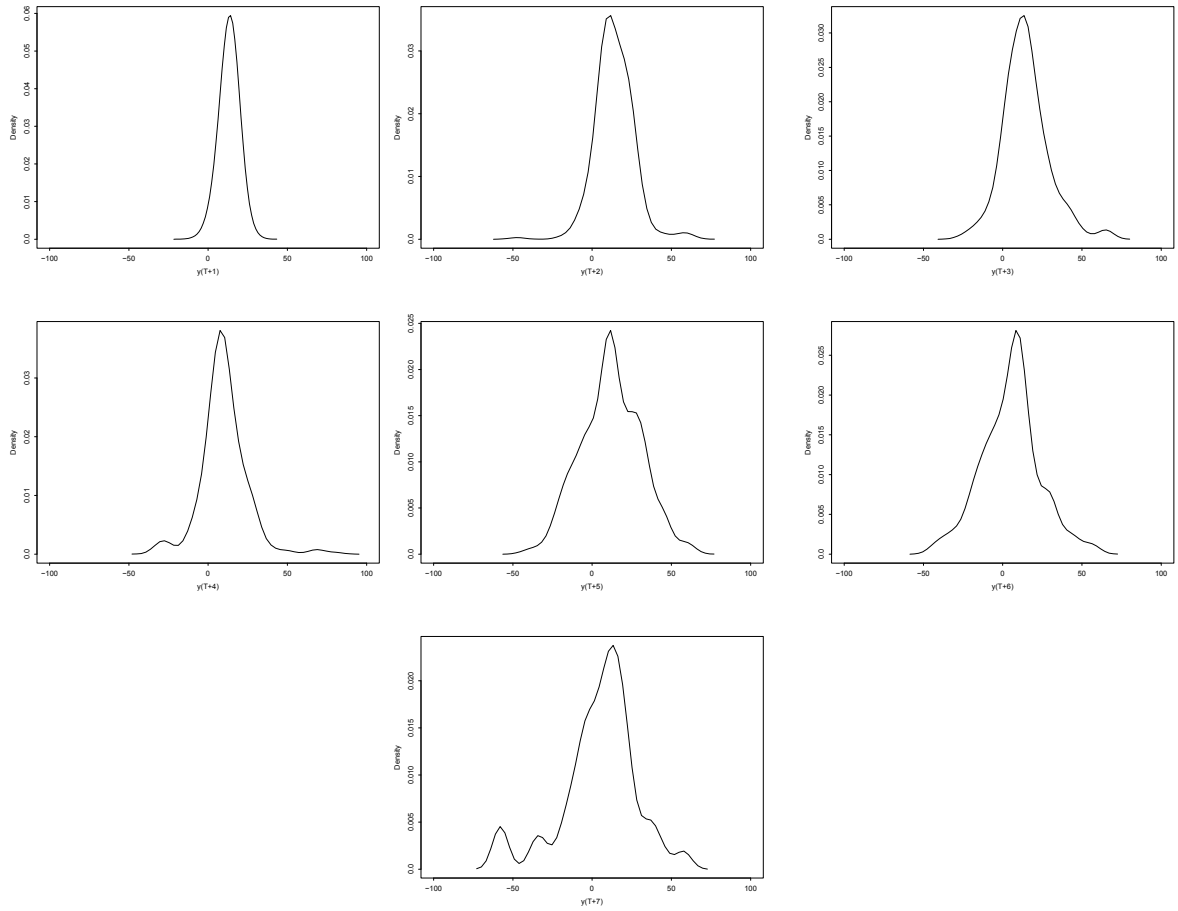


Figure 2.14: Term Structure of MAR(1,1) Forecasts

Chapter 3

Forecast Performance in Noncausal MAR(1,1) Processes

1 Introduction

This chapter examines the forecast performance of noncausal processes obtained from closed-form functional estimators of the predictive density. This forecasting method provides simulation-based prediction intervals at multiple finite horizons in one step. It has been proposed by Gourieroux and Jasiak (2016) as an alternative to the point prediction methods based on conditional expectations, which have no closed-form expressions and need to be approximated by simulations. These point forecasting methods have been introduced for univariate noncausal processes by Lanne, Luoto and Saikkonen (2012) and for constrained noncausal multivariate processes by Lanne, Saikkonen (2011).

The objective of the chapter is to distinguish the effects of noncausal persistence and the error distribution of the process on forecast accuracy. We capture the persistence effect by varying the conditional mean parameters of predicted models. In addition, we assess the forecasts from processes with infinite error mean and variance distribution and compare them to forecasts from processes with finite moments. Our study includes noncausal processes with the Cauchy error distribution and t-student error distributions with degrees of freedom less than 4.

In addition, the predictive densities at horizon 2 are explored. These predictive densities can be illustrated by contour plots which change shape at the outset of a bubble or spike in the trajectory. That finding suggests that the contour plots can be used as methods of graphical display for bubble detection.

The chapter is organized as follows.

Section 1 provides some additional motivation and background on noncausal processes. Section 2 recalls the definition of the noncausal MAR(1,1) process, the filtering and simulation methods as well as the forecasts from closed-form predictive density [Gourieroux,

Jasiak (2016)].

Section 3 investigates the effect of noncausal persistence and error distribution on the forecast error in one-step ahead predictions.

Section 4 examines the goodness of fit of the predictive density estimators in one-step ahead predictions. We compare the performance of the Cauchy specific closed-form predictive density that has been derived for the Cauchy noncausal process with the closed-form general formula of predictive density that can be applied to noncausal processes with any error distribution. In the context of the general formula of predictive density, we study the effect of the number of simulations on forecast accuracy.

Section 5 studies the accuracy of point predictions and the size of prediction interval in multiple-step ahead predictions and provides the analytical formulas of the iso-density curves of predictive densities at horizon 2. In particular, it examines the detection of bubble outset from two-steps ahead predictive densities. Section 6 concludes. Additional results are given in Appendices 1 and 2.

2 Motivation

The literature on noncausal processes has grown in recent years due to innovations in estimation and forecasting. Various applications have been demonstrated ranging from the estimation of processes with expectations [Lanne and Saikonnen 2011] to the modelling and forecasting of bubbles [Gourieroux and Zakoian (2016), Hencic and Gourieroux (2015)]. The ability to identify and accommodate roots within the unit circle makes it a unique tool for analyzing processes that had previously been empirically challenging to forecast.

The forecasting problem for noncausal processes is relatively complex given its non-

linear nature. To the authors' knowledge three approaches for univariate processes have been proposed. The first, by Lanne, Luoto and Saikkonen (2012), is a simulation based approached. They rely on an approximation of the expectation of the error term from a noncausal autoregressive model to recursively calculate future observations. Alternatively, Lanne, Luoma, and Luoto (2012) propose a Bayesian approach for forecasting noncausal processes. In their framework the unknown forward and backward error terms are additional parameters that are estimated and then used to generate forecasts. Importantly, Lanne, Luoma and Luoto (2012) rely heavily on numerical methods to solve for the parameters of interest as they note that there is no closed form solution for the posterior moments of the parameters.

The issue with these first two approaches is that they suffer from a "truncation bias". To approximate the noncausal components the infinite order of the future errors are cut off at some large enough order. The bias can be made smaller by increasing the order of the errors used, however this comes with added computational burden. Additionally, if the noncausal polynomial has a root close to unity, the number of future errors must be large to avoid a significant bias.

Conversely, Gourieroux and Jasiak (2016) proposed a closed-form estimator of the predictive density, that avoids the need to produce long simulations of future errors. Their forecasting procedure involves filtering the process and constructing a forward-looking predictive density at any horizon h.

The predictive density forms the most complete forecast for a process and the merits of its use in financial econometrics were highlighted by Diebold, Gunther and Tay (1998). This chapter explores the density, interval and point forecasts of the noncausal MAR(1,1) process produced by the closed form estimator under both Cauchy and t-distributed errors.

A key feature of the closed form estimator is its ability to generate multi-step density forecasts and joint forecasts for future values. In fact, by using the joint forecasts, a probable path for the process featuring upturns and downturns can be produced. However, the effectiveness of the approach is limited when the noncausal persistence approaches unity or in the presence of large tails of the forecast error due to infinite variance and means.

The paper will also provide formulas for the iso-density curves of the noncausal Cauchy AR(1) process. These iso-density curves can then form the basis of graphical analysis for bivariate densities to study the behaviour of processes with bubbles. Specifically, by studying their behaviour the outset of bubbles can be identified.

3 Noncausal MAR(1,1) Process

This section reviews the definition of the MAR(1,1) process and the nonlinear method of filtering and simulation. The closed-form formulas of predictive density used for forecasting at various horizons are also shown..

3.1 Definition

The mixed noncausal autoregressive MAR(1,1) process is defined as:

$$(1 - \phi L)(1 - \psi L^{-1})y_t = \varepsilon_t \tag{3.1}$$

where the errors are independent, identically distributed and such that $E(|\varepsilon_t|^\delta) < \infty$ for $\delta > 0$. Parameters ϕ and ψ are two autoregressive coefficients that are strictly less than one. The coefficient ϕ represents the standard causal persistence while the coefficient ψ depicts noncausal persistence.

When $\psi = 0$, equation (3.1) defines a purely causal autoregressive process. If $\phi = 0$, the equation above defines a purely noncausal process. If both polynomials contain non-zero coefficients, then equation (3.1) describes a mixed causal-noncausal MAR(1,1) process. The mixed process contains both leads and lags of y_t.

It follows from Lanne, Saikkonen (2011), and Lanne, Luoto, Saikkonen (2012), that process (y_t) has the following unobserved components u_t, v_t defined by:

$$u_t \equiv (1 - \phi L)y_t \leftrightarrow (1 - \psi L^{-1})u_t = \varepsilon_t, \tag{3.2}$$

and

$$v_t \equiv (1 - \psi L^{-1})y_t \leftrightarrow (1 - \phi L)v_t = \varepsilon_t, \tag{3.3}$$

which can be interpreted as the "causal" and "noncausal" components of process (y_t). Gourieroux, Jasiak (2016) show that i) u_t is ε-noncausal and y-causal and ii) v_t is ε-causal and y-noncausal. Process y_t has the following representations based on unobserved components:

$$y_t = \frac{1}{1 - \phi\psi}(\phi v_{t-1} + u_t). \tag{3.4}$$

In this representation, y_t is a linear function of the first lag of the ε-causal component v and of the current value of the ε-noncausal component u. We also have:

$$y_t = \frac{1}{1 - \phi\psi}(v_t + \psi u_{t+1}). \tag{3.5}$$

In the above representation, y_t is a linear function of the current value of the ε-causal component v and of the first lag of the ε-noncausal component u.

The above unobserved component representations (3.4) and (3.5) of process y_t can be used for filtering and forecasting. It can be shown that when the errors are Cauchy distributed, the stationary distribution of the process (y_t) is Cauchy as well (see Appendix B).

3.2 Filtering and Simulation

The filtering procedure allows us to compute the unobserved components given the observations on process (y_t), over a period of length T. Let $(y_1, \ldots, y_T)$ denote the observed sequence.

The values of unobserved components u and v and errors ε can be computed from a set of observations $(y_1, \ldots, y_T)$ as follows:

(i) From equation (3.1) for $t = 2, \ldots, T-1$, we obtain the values $\varepsilon_2, \ldots, \varepsilon_{T-1}$ as functions of $(y_1, \ldots, y_T)$.

(ii) From equation (3.2) : $u_t = (1 - \phi L)y_t, t = 2, \ldots, T$, we obtain $u_2, \ldots, u_T$.

(iii) From equation (3.3) : $v_t = (1 - \psi L^{-1})y_t, t = 1, \ldots, T-1$, we obtain $v_1, \ldots, v_{T-1}$.

When an additional observation y_{T+1} becomes available, the set of unobserved components can be updated by computing ε_T, u_{T+1} and v_T.

The above formulas can be used to simulate the trajectories of process (y_t) as follows:

Step 1: Simulate a path of i.i.d. errors $\varepsilon_t^{sim},\ t = 1, ..., T$.

Step 2: Use formulas (3.2)-(3.3) to obtain the simulated paths of the ε-causal and ε-noncausal components :

$$u_t^{sim} = \varepsilon_t^{sim} + \psi u_{t+1}^{sim},\ t = 1, ..., 2T,$$

$$v_t^{sim} = \varepsilon_t^{sim} + \phi v_{t-1}^{sim},\ t = -T, ..., T,$$

starting from a far terminal condition (resp. far initial condition) $u_{2T}^{sim} = u_0$, say (resp. $v_{-T}^{sim} = v_0$).

Step 3: The simulated trajectory (y_t^{sim}) is obtained from either one of the two partial fraction representations given below:

$$y_t^{sim} = \frac{1}{1-\phi\psi}(u_t^{sim} + \phi v_{t-1}^{sim}) = \frac{1}{1-\phi\psi}(v_t^{sim} + \psi u_{t+1}^{sim}),\ t = 1, ..., T. \tag{3.6}$$

As an illustration, four simulated trajectories of MAR(1,1) processes with Cauchy distributed errors with location coefficient 0 and scale 1 and fixed coefficient $\phi = 0.3$ are plotted in Figure 3.3 for sample size $T = 200$ and the following four values of noncausal persistence coefficients $\psi = 0.0, 0.3, .0.5, .0.9$. We observe that the simulated series display large spikes in their trajectories. These spikes can be interpreted as bubbles with a phase of growth followed by a sudden burst. The bubbles grow at a faster or slower rate depending on the value of noncausal coeffficient ψ.

The extreme values in the trajectory of the MAR(1,1) processes with Cauchy distributed errors complicate the forecasting from these processes.

3.3 Forecasting

The information set $(y_1, \ldots, y_T)$ is equivalent to the information set $(v_1, \ldots, v_2, \varepsilon_2, \ldots, \varepsilon_{T-1}, u_T, \ldots, u_T)$, as shown in Gourieroux, Jasiak (2016).

Therefore, the information contained in $(y_1, \ldots, y_{T+H})$ is equivalent to the information in $(v_1, \varepsilon_2, \ldots, \varepsilon_{T+H-1}, u_{T+1}, \ldots, u_{T+H})$, and it is also equivalent to that in $(v_1, \varepsilon_2, ..., \varepsilon_{T-1}, u_T, ..., u_{T+H})$, because $(1 - \psi(L^{-1}))u_t = \varepsilon_t, t = T, \ldots, T + H - s$ by formula (3.2).

Thus, instead of predicting the future value of y, at horizon H, we can equivalently predict the future value of the ε-noncausal component u, by finding the predictive density $\hat{\Pi}$ at horizon H for a noncausal process of order 1:

For a given error density g and for known values of coefficients ϕ, ψ we get :

$$\begin{aligned} & \hat{\Pi}(u_{T+1}, \ldots, u_{T+H}|\hat{u}_T) \\ \equiv \quad & \frac{g(\hat{u}_T - \psi u_{T+1}) g(u_{T+1} - \psi u_{T+2}) \ldots g(u_{T+H-1} - \psi u_{T+H}) \sum_{t=1}^{T} g(u_{T+H} - \psi \hat{u}_t)}{\sum_{t=1}^{T} g(\hat{u}_T - \psi \hat{u}_t)}. \end{aligned} \tag{3.7}$$

where $\hat{u}_t,\ t = 1+1, ..., T$ are the filtered values of the ϵ-noncausal component, that are functions of $y_1, ..., y_T$ and of coefficients ϕ, ψ.

The predictive density given above has a closed-form representation when the error density g is known. In particular, when ϵ follows a Cauchy distribution, the one-step ahead predictive density is:

$$\pi^*(u_{T+1}|\hat{u}_T) = \frac{1}{\pi} \frac{1}{1 + (\hat{u}_T - \psi u_{T+1})^2} \frac{1 + (1 - |\psi|)^2 \hat{u}_T^2}{1 + (1 - |\psi|)^2 u_{T+1}^2}, \tag{3.8}$$

Also, the predictive joint distribution of two future values is:

$$\begin{aligned} \pi^*(u_{T+1}, u_{T+2}|\hat{u}_T) \quad = \quad & \frac{1}{\pi^2} \frac{1}{1 + (\hat{u}_T - \psi u_{T+1})^2} \frac{1}{1 + (u_{T+1} - \psi u_{T+2})^2} \\ & \times \frac{1 + (1 - |\psi|)^2 \hat{u}_T^2}{1 + (1 - |\psi|)^2 u_{T+2}^2}. \end{aligned} \tag{3.9}$$

The estimator of predictive density $\hat{\Pi}$ of the ε-noncausal component u given above

can be used to generate the future values or future paths of the observable process y and its unobservable causal and noncausal components over a given horizon H.

The future values of y are computed from the future values of u that are drawn in $\hat{\Pi}$ by applying a sampling importance resampling (SIR) method [see Rubin (1988), Geldfand, Smith (1992)], or alternatively a Metropolis-Hasting algorithm. More specifically, the procedure is as follows:

Step 1 : Use data $(y_1, \ldots, y_T)$ to compute the filtered values of in-sample unobserved components u: $\hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_{T-1},\ \hat{v}_1, \ldots, \hat{v}_{T-1},\ \hat{u}_2, \ldots, \hat{u}_T$.

Step 2 : Compute the predictive density $\hat{\Pi}$.

Step 3: Use the SIR method to simulate future u's: $u^{sim}_{T+1}, \ldots, u^{sim}_{T+H}$.

Step 4: Use formulas (3.2-3.4) to compute the future values $y^{sim}_{T+1}, ..., y^{sim}_{T+H}$, $\epsilon_{T+1-s}, ..., \epsilon_{T+H-s}$, $v_{T+1-s}, ..., v_{T+H-s}$.

The estimator of the predictive density (3.5) can be used to forecast y_{T+1} as follows. Given that

$$y_{T+1} = u^{sim}_{T+1} + \phi y_T$$

one can compute the predictive density $\hat{\pi}(y_{T+1}|y_T, \hat{u}_T)$ by shifting the predictive density $\hat{\Pi}(u_{T+1}|\hat{u}_T)$ by ϕy_T. The predictive density provides the most complete forecast. It can be used for computing simple one-step ahead forecasts or for predicting entire future trajectories.

The predictive distribution in (3.7) can be alternatively rewritten as a joint distribution of future values of the process y_{T+1}, y_{T+2}, conditional on $\hat{u}_T$ and y_T, as follows:

$$\pi^*(y_{T+1}, y_{T+2}|y_T, \hat{u}_T) = \frac{1}{\pi^2}\frac{1}{1+(\hat{u}_T - \psi(y_{T+1} - \phi y_T))^2}\frac{1}{1+(y_{T+1} - \phi y_T - \psi(y_{T+2} - \phi y_{T+1}))^2} \times \frac{1+(1-\psi)^2\hat{u}_T^2}{1+(1-\psi)^2(y_{T+2} - \phi y_{T+1})^2}. \tag{3.10}$$

See Appendix C.

4 One-step-ahead Forecasts

This section examines the effect of noncausal persistence and error distribution on one-step ahead out-of-sample predictions. First, we examine the predictive density and point predictions. Next, we study the size and coverage of the prediction interval based on selected quantiles of the predictive density.

The following one-step ahead forecasting techniques are based on the predictive density $\hat{\pi}(y_{T+1}|y_T)$, where

$$\hat{\pi}(y_{T+1}|y_T, \hat{u}_T) = \hat{\Pi}(u_T|\hat{u}_T) + \phi y_T.$$

A point forecast $\hat{y}_{T+1}$ is the mode of the estimated predictive density and interpreted as the most likely future value of the process.

$$\hat{y}_{T+1} = mode[\hat{\pi}(y_{T+1}|y_T, \hat{u}_T)] \tag{4.11}$$

The mode is a nonparametric and robust estimator that exists for any distribution of the forecasted time series, including the Cauchy distribution.

The forecast error associated with the point forecast is:

$$fe_{T+1} = y_{T+1} - \hat{y}_{T+1} \tag{4.12}$$

i.e. the difference between the true and predicted y.

A robust interval forecast is defined from the quantiles of the estimated predictive density as follows:

$$P[y_{T+1} \in (Q_{T+1,\alpha/2}, Q_{T+1,1-\alpha/2})] = 1 - \alpha \tag{4.13}$$

where $Q_{.}$ is the quantile of $\hat{\pi}(y_{T+1}|y_T, \hat{u}_T)$. The probability α can be set depending on the desired accuracy.

4.1 Predictive Density

To study the most complete forecast provided by the entire predictive density, let us consider the four trajectories of MAR(1,1) processes with Cauchy errors displayed in Figure 3.3 and denote them by $y_t^1, y_t^2, y_t^3, y_t^4$. The predictive one-step ahead densities of $y_{T+1}^i, i = 1, ..., 4$ are computed from each of these processes given their respective values y_T^i. The values of the processes at $T = 200$ and their out-of-sample point predictions are as follows:

Process 1: $y_T^1 = 1.770$, $\hat{y}_{T+1}^1 = 0.702$

Process 2: $y_T^2 = 25.037$, $\hat{y}_{T+1}^2 = 3.055$

Process 3: $y_T^3 = 70.185$, $\hat{y}_{T+1}^3 = 11.430$

Process 4: $y_T^4 = 253.947$, $\hat{y}_{T+1}^4 = 74.468$

Figure 3.4 plots the predictive density under the assumption of four starting points and four levels of persistence. As the value of the starting point of the forecast and the amount of persistence increase, so too does the error in the forecast. Note how the upper tail of the predictive density changes from a smoothly decaying tail (the upper left panel) to a jagged one with no discernable decay (the lower right panel). In addition, the scale of the four plots changes, showing how the predictive density becomes more dispersed as the last observed y_T and persistence increase. These plots serve to show the difficulty of forecasting noncausal processes with a high degree of noncausal persistence and large values of y_T.

4.2 Point Predictions

Let us now examine the performance of point predictions in finite samples and study how it is affected by noncausal persistence. The MAR(1,1) processes with Cauchy errors have infinite marginal moments and finite conditional moments. To account for the effect of error distribution, we compare the MAR(1,1) with Cauchy errors with MAR(1,1) processes with the following error distributions:

- t-student with 3 degrees of freedom with finite variance and infinite moments of higher orders

- t-student with 4 degrees of freedom with finite skewness and infinite kurtosis

Samples of length 200 of each of the three types of processes are generated and replicated 1000 times. The coefficients are fixed and equal to $\phi = 0.3$ and $\psi = 0.0, 0.3, 0.5, 0.9$, respectively. At each replication, we predict $\hat{y}_T$ and compute the one step ahead forecast errors for all processes as a difference between $\hat{y}_T$ and y_T.

The densities of the forecast errors are plotted in Figures 2-4.

Figure 3.5 plots the densities corresponding to the forecast errors from the Cauchy MAR(1,1) process. All of the error densities are centered at zero. However, due to the extreme values of the forecast errors the tails of the distribution become asymmetric across different values of the persistence parameters making them difficult to compare.

Figures 3.6 and 3.7 show the forecast errors densities for the MAR(1,1) processes with Student-t distributed processes with 3 and 4 degrees of freedom, respectively. The tails of the error densities for the Student-t with 3 degrees of freedom appear to grow slightly as the persistence parameter is increased from 0.0 to 0.9. Note also the changing asymmetry of the forecast error density as the noncausal persistence parameter is increased. A similar pattern is observed in the case with the Student-t with 4 degrees of freedom. As the noncausal persistence parameter is increased, the distribution of the error terms becomes more symmetrical, even though the dispersion increases.

For a more detailed comparison, Table 3.1 shows the values of the median, interquartile range, and range of the forecast error densities of the MAR(1,1) processes with Cauchy, t-student-3 and t-student-4 error distributions displayed in three panels and for noncausal coefficients $\psi = 0.0, 0.3, 0.5, 0.9$ shown in columns 2 to 4. The values of the sample mean are reported for MAR(1,1) processes with t-distributed errors.

Under all specifications, the medians (means) of the forecast errors are close to 0 for all processes. The interquartile range for the processes with Cauchy errors are larger than their t-distributed counterparts. This is to be expected as the Cauchy process has larger tails corresponding to more extreme values. The total range reported in the fourth row shows the magnitude of this effect. The Cauchy distributed process exhibits a range up to several thousand times the size of that of the t-distributed processes. Another interesting feature is that the range of the Cauchy process depends on the size of the

Table 3.1: Forecast Errors of MAR(1,1)

Cauchy Distribution				
ψ	0.0	0.3	0.5	0.9
Median	0.039	-0.017	-0.026	0.055
Interquartile Range	2.155	2.820	3.598	2.745
Range	3,366.22	3,501.232	15,901.116	430,769.815
t-Distribution 3 d.f.				
ψ	0.0	0.3	0.5	0.9
Median	0.045	-0.024	-0.000	-0.021
Mean	-0.044	-0.030	-0.016	0.126
Interquartile Range	1.479	1.501	1.727	1.611
Range	10.449	11.035	11.648	12.855
t-Distribution 4 d.f.				
ψ	0.0	0.3	0.5	0.9
Median	0.050	0.006	-0.037	-0.016
Mean	-0.030	-0.036	-0.040	-0.013
Interquartile Range	1.427	1.464	1.637	1.477
Range	15.417	15.113	15.343	13.649

noncausal parameter, increasing along with the degree of noncausal persistence. In fact, as the noncausal parameter increases from 0.3 to 0.9, the range increases by a factor of roughly 120.

The behaviour of the range of the forecast errors is notably different between the Cauchy and t-distributed processes. In contrast to the Cauchy MAR(1,1) process, the range of the t-distributed processes appear to be relatively unaffected by the degree of noncausal persistence. Though the ranges are much different depending on the distribution of the process, the interquartile ranges are not. It is obvious that the Cauchy distributed MAR(1,1) forecast error has a larger interquartile range, however it is not orders of magnitude larger than that of the t-distributed processes like the range is.

Table 3.2: Prediction intervals of MAR(1,1)

Cauchy Distribution			
mean LB	mean UB	mean length	freq in
-3.523	4.532	8.061	0.64
t-Distribution 3 d.f.			
-1.468	1.393	2.882	0.73
t-Distribution 4 d.f.			
-1.347	1.447	2.794	0.71

4.3 Interval Predictions

Let us now examine interval predictions obtained for fixed values of persistence coefficients $\phi = 0.3, \psi = 0.5$. Again, we perform 1000 replications of the MAR(1,1) process with errors that follow the Cauchy and t-student distributions with 3 and 4 degrees of freedom.

Table 3.2 below shows the prediction intervals at 70 % provided by the 15th and 85th quantiles of the predictive distribution (3.4). Note that 5% of computed prediction intervals for the Cauchy process are incomplete. This can be explained by computational difficulty in fitting the tails of the predictive distribution for some extreme values of y_T, illustrated in Figure 3.2. In those cases, the support of the predictive density predetermined in the numerical procedure turns out to be insufficient and one of the quantiles is reported as a missing value. In contrast 0.2% and 0.7% of Student-3 and Student-4 simulations gave incomplete prediction intervals.

The accuracy of interval predictions can be assessed from the number of times the true value y_{T+1} is inside the prediction interval, which is given in column 4 of Table 3.2. The empirical size of the prediction interval is very close to the theoretical size of 70% in the processes with t-distributed errors. It is slightly less than 70% in the Cauchy MAR(1,1) process. Moreover, the average length of the prediction interval decreases across the rows

of Table 3.2.

The properties of the forecast depend on the conditioning variable, i.e. the value of the last observation in the sample. Figures 3.6, 3.7 and 3.8 show how the properties of the forecast vary as functions of the value of the conditioning variable y_T in the processes with the Cauchy errors and student-t errors with 3 and 4 degrees of freedom and $\phi = 0.3, \psi = 0.5$.

For each process, the 1000 simulated values y_T are ranked and divided into deciles. From each decile, the computed mean squared forecast error, mean interval length and coverage count are displayed in the 3 panels of each Figure.

In the case of the Cauchy process (Figure 3.8) , the forecast errors are noticeably worse when the conditioning value is in the first and tenth decile. The mean interval length rises rapidly from the ninth to tenth decile, while the mean coverage count drops precipitously. Taken together, this indicates a rapid deterioration of forecast quality. Figure 3.9 shows the results for the process with Student-t distributed errors with 3 degrees of freedom. In line with the results from Table 3.2, the mean interval length is shorter at all deciles than that of the Cauchy process, while the mean coverage count that is relatively similar. Importantly, the behaviour of the mean interval length clearly illustrates how the forecast errors increase with extreme values of y_T. The plot rises as the deciles move from the fifth decile out to the extremes. As expected, the mean coverage count is also highest in the middle deciles and declines in the extremes. Lastly, Figure 3.10 shows the plots for the Student-t process with 4 degrees of freedom. As is the case for the other processes, the mean squared errors are largest for the first and tenth deciles. The range of the interval lengths are shorter than those of either the Cauchy or Student-t 3d.f., but follow the same patter with larger values in the first and tenth deciles. Finally, the mean coverage count is lower at the tenth decile. These results are in line with the results plotted in Figure

3.4 where we know the upper tail of the predictive distribution does not behave well with large, positive, starting values. Overall, we observe that the quality of the forecast diminishes as the interval widens up when y_T takes extreme values and is in the extreme deciles 1 and 10.

4.4 Rolling Forecast

To further examine the forecast in the most challenging MAR(1,1) process (Cauchy errors) we simulate one trajectory with $\phi = 0.3$ and $\psi = 0.5$. 100 values of that trajectory are forecasted out-of sample from one-step ahead point and interval forecasts based on samples of length 200. The result is displayed in Figure 3.11. The solid, dashed and dotted lines indicate the true value of the series, its one-step ahead prediction and the prediction-interval of size 70%. We observe that the quality of the forecast is rather high when the process takes standard values and deteriorates when the series takes extreme values.

5 Multi-step Ahead Forecast

The multi-step forecasting techniques are based on the predictive density $\hat{\pi}(y_{T+1}...y_{T+H}|y_T)$ obtained by shifting $\hat{\Pi}(u_{T+1}, \ldots, u_{T+H}|\hat{u}_T)$ by $\phi y_T, ..., \phi\hat{y}_{T+H-1}$. We obtain in one step the following summaries of the future trajectory of the process up to horizon H.

The trajectory forecast up to horizon H is:

$$\hat{y}_{T+1}, ..., \hat{y}_{T+H} = mode[\hat{\pi}(y_{T+1}|y_T, \hat{u}_T)], ..., mode[\hat{\pi}(y_{T+H}|y_T, \hat{u}_T)] \tag{5.14}$$

The trajectory forecast error up to horizon H is:

$$fe_{T+1}, ..., fe_{T+H} = Y_{T+1} - \hat{y}_{T+1}, ..., Y_{T+H} - \hat{y}_{T+H} \tag{5.15}$$

The bivariate predictive densities $\hat{\pi}(y_{T+1}, y_{T+2})$ shown in Figures 3.12-3.15 are computed for Processes 1-4 introduced in Section 3.1 and plotted in Figure 3.2. Recall that in these processes, the noncausal persistence takes values 0.0, 0.3. 0.5 and 0.9, respectively and the causal persistence coefficient $\phi = 0.3$ is fixed. The contour plots are conditioned at the last observed value y_T of each process and the unobserved component $u_T = y_T - \phi y_{T-1}$. The pairs (y_{T-1}, y_T) used in the computation of u_T for each process are (2.344, 1.770), (9.852, 25.037), (37.433, 70.185) and (230.893, 253.947), respectively. In each Figure, the x- and y-axis measure the values of Y_{T+1} and Y_{T+2}. In Figure 3.12, the vertical and horizontal lines indicate the coordinates of (y_{T-1}, y_T) for comparison.

We observe that the shape of the contour plots become more elliptical and elongated when the noncausal persistence coefficient increases. The range of admissible future values increases as well. In Process 4, we observe the appearance of multiple local maxima.

5.1 Joint Forecast at Horizon 2

This section reports the study of two-step ahead point forecasts from Processes 1 to 4. The bivariate point forecasts at steps 1 and 2 are computed jointly by rolling. Forecasting starts from t=50 in order to ensure a sufficiently large set of conditioning variables that are $u_t = y_t - \phi y_{t-1}$ and y_t. The forecasted values are $\hat{y}_t,\ t = 50, ..., 200$ for the four Cauchy distributed processes. The forecast errors at horizon 1 and 2 are computed and their summary statistics are reported in Table 3.3 below. They are calculated conditional on the values of variables u_t and y_t being outside or inside the interquartile range of the entire sample. Values that are outside the interquartile range are referred to as extreme

and values inside the interquartile range are referred to as standard. Note that in many instances the extreme values in u_t and y_t coincide and thus the values across columns may be the same.

Beginning with the process with $\psi = 0.0$ we observe that for standard values of u_t and y_t the median forecast errors are smaller for both the one-step ahead and two-step ahead forecasts. This is in line with results presented in prior sections. Also, note how extreme values of u_t appear to have a larger effect on the interquartile range of the errors than extreme values of y_t. This is a feature of both the one-step and two-step ahead forecasts. Also, note that even with standard u_t and y_t the forecast errors grow along with the magnitude of the noncausal persistence parameter. Additionally, the difference between the median forecast errors with standard u_t and y_t and those with extreme u_t, y_t or both appear to increase as the value of ψ increases.

The trajectories of forecasted values of the processes and their forecasts are displayed in Figure 3.16 for one-step ahead and in Figure 3.17 for two step ahead forecasts. We observe that the quality of forecast depends on the noncausal persistence coefficient, as higher noncausal persistence results in larger spikes that are difficult to predict. The forecasts of standard values 1 and 2 steps ahead are mostly satisfactory. As expected, the forecasts 1 step ahead are better than those 2 steps ahead, along the trajectory. Both forecasts are unable to adjust to the sudden spike when the process reaches a value 150 to 200 units away from its location. The forecasts of values on the largest bubble at horizon 1 are better than the forecast at horizon 2.

Table 3.3: Joint One and Two-Step Ahead Forecast Errors of MAR(1,1)

one-step ahead forecast				
$\psi = 0.0$	u_t extreme	y_t extreme	u_t and y_t extreme	u_t and y_t standard
Median	1.225	0.183	0.255	0.040
Interquartile Range	4.078	2.025	3.366	1.228
Range	247.383	60.013	247.383	234.077
two-step ahead forecast				
$\psi = 0.0$	u_t extreme	y_t extreme	u_t and y_t extreme	u_t and y_t standard
Median	0.592	0.477	0.584	0.173
Interquartile Range	2.215	2.035	2.240	2.128
Range	96.896	25.381	96.896	250.383
one-step ahead forecast				
$\psi = 0.3$	u_t extreme	y_t extreme	u_t and y_t extreme	u_t and y_t standard
Median	1.621	0.504	0.542	0.300
Interquartile Range	6.407	4.063	5.818	1.678
Range	243.181	243.181	243.181	16.330
two-step ahead forecast				
$\psi = 0.3$	u_t extreme	y_t extreme	u_t and y_t extreme	u_t and y_t standard
Median	0.876	0.627	0.796	0.465
Interquartile Range	6.738	5.829	5.831	2.625
Range	316.339	316.339	316.339	45.965
one-step ahead forecast				
$\psi = 0.5$	u_t extreme	y_t extreme	u_t and y_t extreme	u_t and y_t standard
Median	-0.241	0.606	0.370	0.672
Interquartile Range	11.281	6.166	9.151	2.133
Range	242.517	242.517	242.517	12.263
two-step ahead forecast				
$\psi = 0.5$	u_t extreme	y_t extreme	u_t and y_t extreme	u_t and y_t standard
Median	0.917	0.629	0.916	1.0668
Interquartile Range	12.492	7.103	9.667	3.176
Range	280.668	280.668	280.668	28.740
one-step ahead forecast				
$\psi = 0.9$	u_t extreme	y_t extreme	u_t and y_t extreme	u_t and y_t standard
Median	2.900	-0.372	1.094	-6.0755
Interquartile Range	89.926	89.894	11.858	30.619
Range	260.115	240.578	260.115	75.324
two-step ahead forecast				
$\psi = 0.9$	u_t extreme	y_t extreme	u_t and y_t extreme	u_t and y_t standard
Median	16.977	-2.891	-0.537	-26.442
Interquartile Range	118.050	116.688	118.891	58.885
Range	367.661	320.292	367.661	136.896

5.2 Iso-Density Curves

The shapes of bivariate predictive densities $\hat{\pi}(y_{T+1}, y_{T+2})$ illustrated above for processes with Cauchy distributed errors can be examined analytically. To do that we need to determine the isodensity curves that define the contour plots in Figures 3.10-3.14. For ease of exposition, let us set the causal autoregressive coefficient $\phi = 0$ and consider the noncausal AR(1) process with Cauchy distributed errors:

$$y_t = \psi y_{t+1} + \epsilon_t, t = 1, ..., T, T+1, ..., T+H$$

The joint density $l(y_t, y_{t+1}) = l(y_t|y_{t+1})l(y_{t+1})$ is :

$$l(y_t, y_{t+1}) = l(y_t|y_{t+1})l(y_{t+1}) = \frac{(1-\psi)}{\pi^2}[1 + (y_t - \psi y_{t+1})^2]^{-1}[1 + (1-\psi)^2 y_{t+1}^2]^{-1}$$

It follows that the isodensity curve is defined by:

$$l(y_t, y_{t+1}) = \text{const} \iff [1 + (y_t - \psi y_{t+1})^2][1 + (1-\psi)^2 y_{t+1}^2] = c,$$

where $c > 1$ is a constant. We observe that

$$(y_t - \psi y_{t+1})^2 + (1-\psi)^2 y_{t+1}^2 = c - 1 - (y_t - \psi y_{t+1})^2 (1-\psi)^2 y_{t+1}^2 < c - 1$$

Proposition 1: The isodensity curve is located inside Ellipse 1 defined by:

$$(y_t - \psi y_{t+1})^2 + (1-\psi)^2 y_{t+1}^2 = c - 1, \tag{5.16}$$

and outside Ellipse 2, defined by:

$$(y_t - \psi y_{t+1})^2 + (1-\psi)^2 y_{t+1}^2 = 2(\sqrt{c} - 1), \tag{5.17}$$

where $c - 1 > 0$. The isodensity is tangent to Ellipse 1 at points such that

$$y_{t+1} = 0 \text{ or } y_t - \psi y_{t+1} = 0$$

Since $c - 1 - 2(\sqrt{c} - 1) = (\sqrt{c} - 1)^2 \geq 0$, Ellipse 1 is outside Ellipse 2 and the isodensity is tangent to Ellipse 2 at points

$$y_t = y_{t+1} \text{ or } y_t = (2\psi - 1) y_{t+1}$$

.

Proof: see Appendix C.

The above result allows us to determine the shape of isodensity curves as follows. Let us consider first the noncausal autoregressive coefficient $\psi = 0$. Then, Ellipses 1 and 2 are circles centered at the origin and defined by:

$$y_t^2 + y_{t+1}^2 = c - 1, \quad \text{and } y_t^2 + y_{t+1}^2 = 2(\sqrt{c} - 1),$$

respectively. These patterns are displayed in Figure 3.1.

When $\psi \neq 0$, the isodensities are located inside ellipses instead of circles.

In order to find the axes of symmetry of the ellipses, we compute the eigenvectors of the quadratic form:

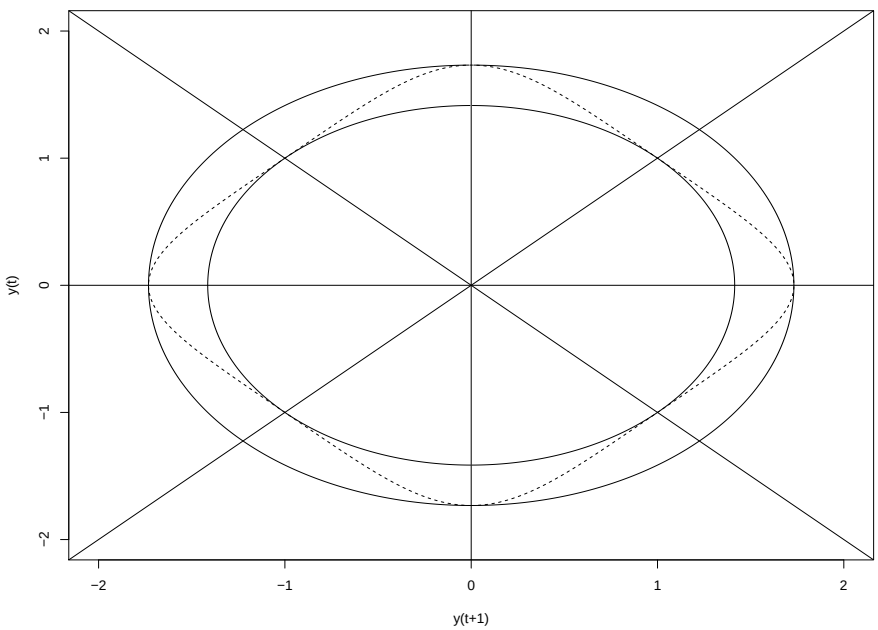


Figure 3.1: Isodensity, $\psi = 0$

$$
\begin{aligned}
&(y_t - \psi y_{t+1})^2 + (1-\psi)^2 y_{t+1}^2 \\
&= y_t^2 + [\psi + (1-\psi)^2] y_{t+1}^2 - 2\psi y_t y_{t+1} \\
&= [y_t, y_{t+1}] \begin{bmatrix} 1 & -\psi \\ -\psi & \psi^2 + (1-\psi)^2 \end{bmatrix} \begin{bmatrix} y_t \\ y_{t+1} \end{bmatrix}
\end{aligned}
$$

The eigenvalues of the quadratic form are:

$$
\begin{aligned}
&det \begin{bmatrix} 1-\lambda & -\psi \\ -\psi & \psi^2 + (1-\psi)^2 - \lambda \end{bmatrix} \\
&= (1-\lambda)[\psi^2 + (1-\psi)^2 - \lambda] - \psi^2 \\
&= \lambda^2 - \lambda[\psi^2 + (10\psi)^2 + 1] + (1-\psi)^2 = 0
\end{aligned}
$$

and are $\lambda_i = (\psi^2 - \psi + 0.5) \mp \psi\sqrt{\psi^2 - 2\psi + 2}, i = 1, 2.$

To illustrate the shape of the isodensities , let us consider the noncausal AR(1) process with coefficient $\psi = 0.3$. Figure 3.2 below shows the isodensity along with Elipse 1 and Ellipse 2. The points of tangency of the isodensity are at the intersection of Ellipse 1 and lines $y_{t+1} = 0$ and $y_t = 0.3y_{t+1}$ and Ellipse 2 and lines $y_t = y_{t+1}$ and $y_t = -0.4y_{t+1}$.

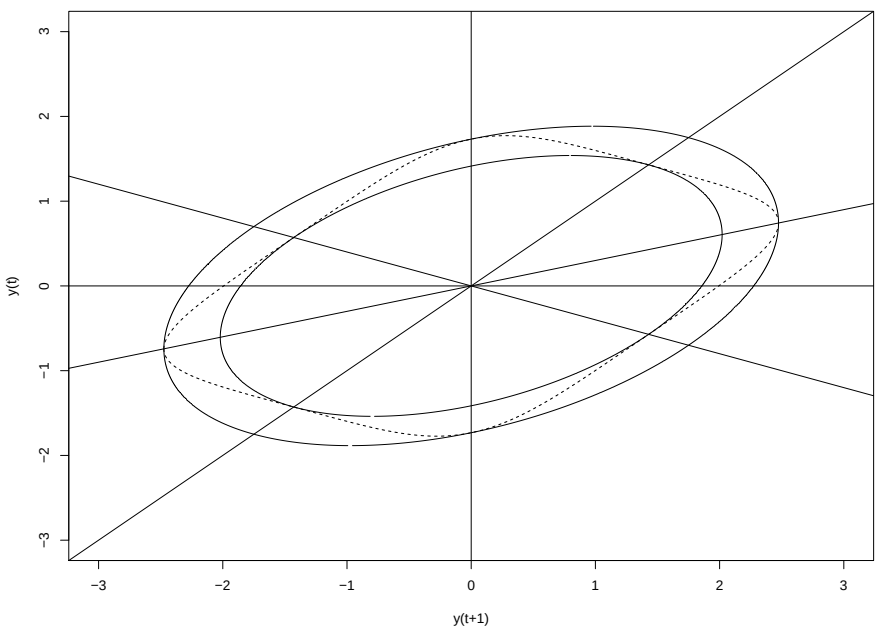


Figure 3.2: Isodensity, $\psi \neq 0$

We observe that the shapes of isodensities are similar to those displayed by copulas that represent dependence between variables with extreme values, i.e. involving high risks and illustrated in Balkema, et al. (2013).

5.3 Outset of Bubble

A challenging problem in the modelling of economic processes is in forecasting the bubbles. As shown earlier in the text, forecasts conditioned on extreme values are not reliable. Therefore forecasting a bubble burst is not promising. It is not the case with the forecasting of the outset of a bubble. The outset of a bubble can be identified from the patterns of two-step ahead predictive densities patterns.

To show that, we consider process 3 and its values between $t = 147$ and $t = 155$ which is a time interval of bubble inception. The observations and their rate of increase computed as a ratio of two consecutive values are given in the Table 3.4 below:

Table 3.4: Bubble Outset in Process 3

time	147	148	149	150	151	152	153	154	155
value	3.431	3.149	3.837	10. 862	15.394	21.236	38.583	66.563	129.605
rate	1.4	0.9	1.2	2.8	1.417	1.38	1 .72	1.81	1.94

Figure 3.18 illustrates the two-step ahead predictive densities evaluated: (1) before the bubble at time t=145 for (y(144), y(145)) = (3.260, 2.733) in panel (a); (2) when the rate of increase doubles at t=150 (interpreted as the starting point of the bubble) in panel (b); and (3) the predictive density during the bubble (at t=155) in panel (c). The plots provide the values of the conditioning variables, i.e. the value of process 3 and unobserved component evaluated at each point ($u_t = y_t$–$0.3y_{t-1}$). We observe that the outset of bubble is characterised by a degenerate predictive density at t=150 that is distinguishable from the predictive densities computed from standard values and from the values on the bubble.

5.4 Bitcoin Bubble

To see if the method of predictive densitity analysis is applicable to detect a bubble in practice, we consider the trajectory of daily closing values of the Bitcoin-US Dollar exchange rates recorded between February and July 2013. The sample consists of 150 observations. Figure 3.19 shows the trajectory of the sample.

In order to remove the trend, we fit a nonlinear function of time by regressing the data on a 3rd degree polynomial in time [Gourieroux, Hencic (2015)]. The detrended series, obtained as the following series of residuals:

$$y_t = rate_t + 3.045 - 3.854t + 3.499t^2 - 0.866t^3$$

Figure 3.20 displays the detrended sample as well as a subsample of 10 observations selected for further analysis that consists of observations at t=60 to t=70.

A MAR(1,1) model with Cauchy errors is estimated from the detrended sample, using the Approximate Maximum Likelihood method [Lanne, Saikkonen (2011)]. The model parameter estimates and their standard errors are $\hat{\psi} = 0.678(0.063), \hat{\phi} = 0.717$ (0.066) with the sample scale parameter $\hat{\gamma} = 2.559(0.109)$. The MAR(1,1) model of the detrended Bitcoin data is:

$$(1 - 0.717L)(1 - 0.678L^{-1})y_t = \varepsilon_t,$$

where ϵ_t is Cauchy distributed with scale coefficient 2.559. The spike chosen for analysis is located in the middle of the trajectory to ensure that the predictive densities are computed from sufficiently many prior fitted values of the unobserved component $\hat{u} = y_t - \hat{\phi}y_{t-1}$. Also, the starting point of this bubble is easier to locate than the outset of any bubble earlier in the sample as those grow very slowly, while the illustrated bubble starts suddenly.

The predictive densities are computed and evaluated at four consecutive points, starting from t=62. They are displayed in Figure 3.21 and the values of conditioning variables used to evaluate the predictive densities are provided under each panel.

The predictive density function changes drastically from t=62, to t=63 becoming more elliptical and changing from a density with a unique maximum to one with multiple maxima. At t=64 the predictive density breaks apart, forming three distinct components. Finally, at t=65 the density again converges into an elliptical shape similar to that at t=63. Note, however, there are now multiple maximum in the predictive density.

6 Conclusion

The chapter examined the forecast accuracy of nonlinear prediction methods for noncausal processes that display spikes and bubbles. It showed that forecasts are more accurate for processes that admit less extreme values and have strong persistence. The predictive densities at horizon 2 are proposed as a new method of graphical analysis for detection of bubble outsets, and the Bitcoin bubble episode from 2013 was presented as an empirical example.

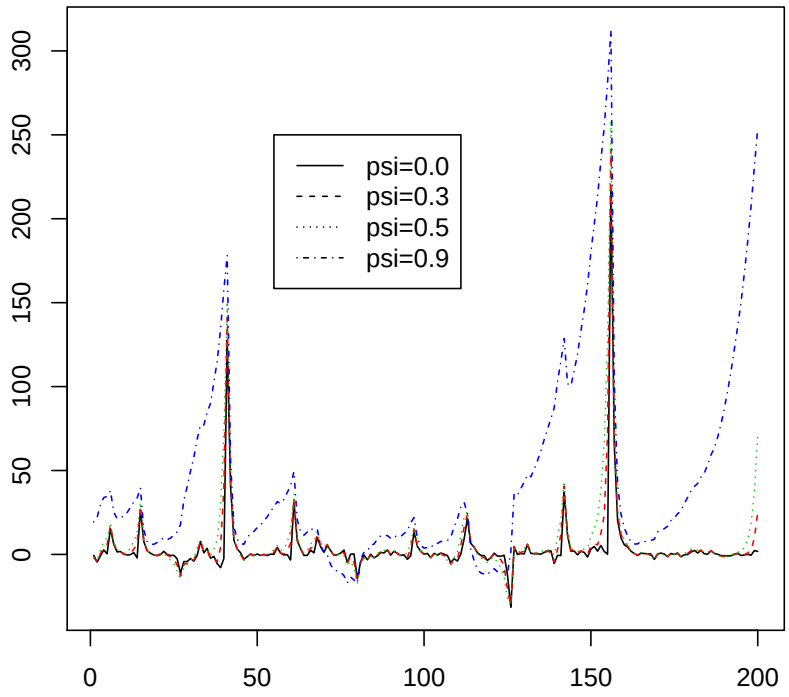


Figure 3.3: Simulated MAR(1,1) Cauchy processes

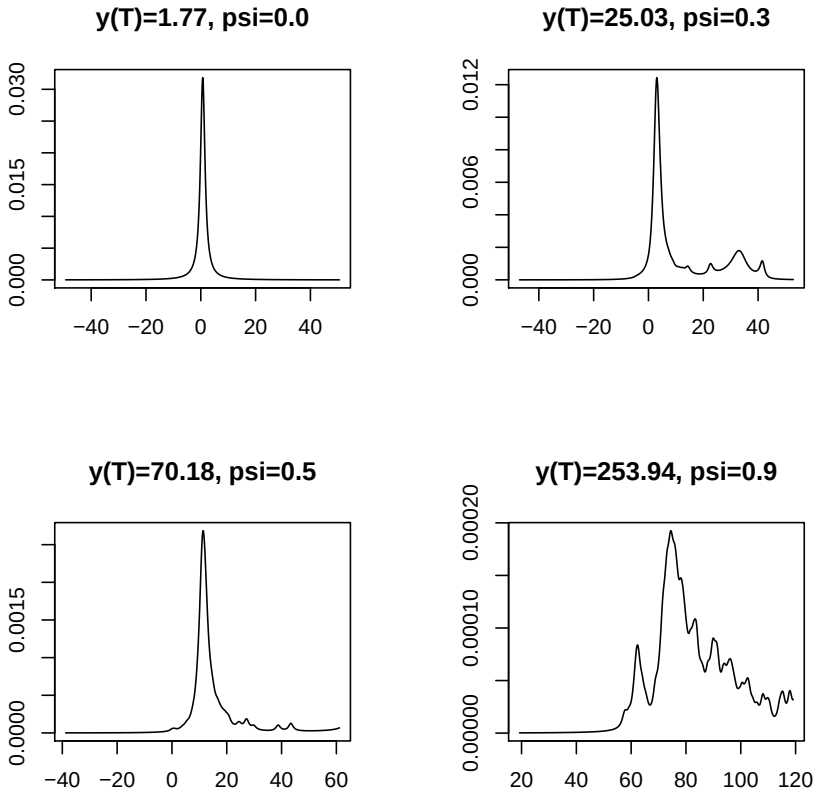


Figure 3.4: Predictive Density

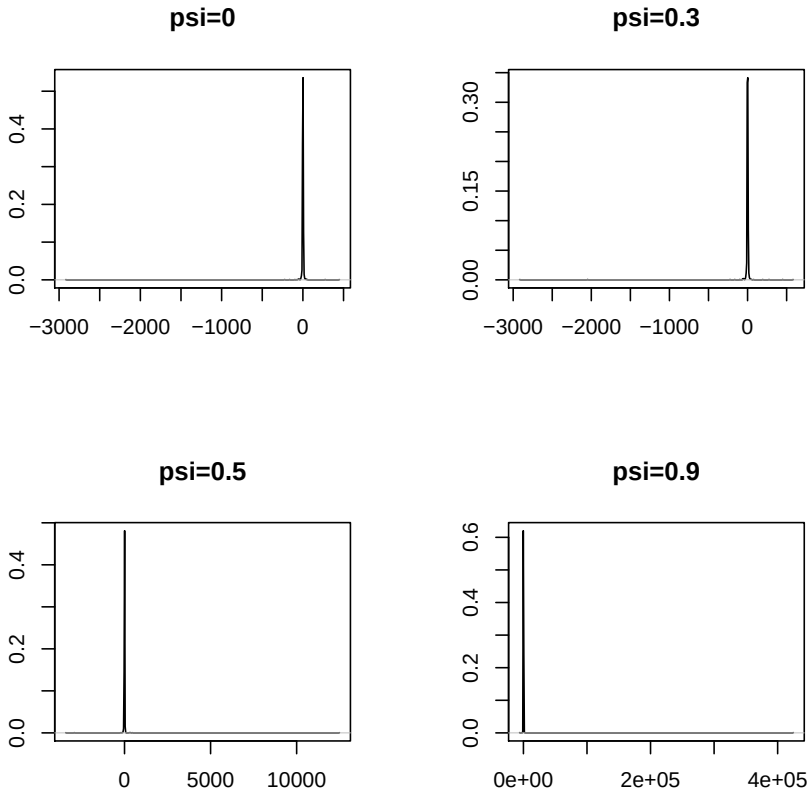


Figure 3.5: Forecast Error Density, Cauchy

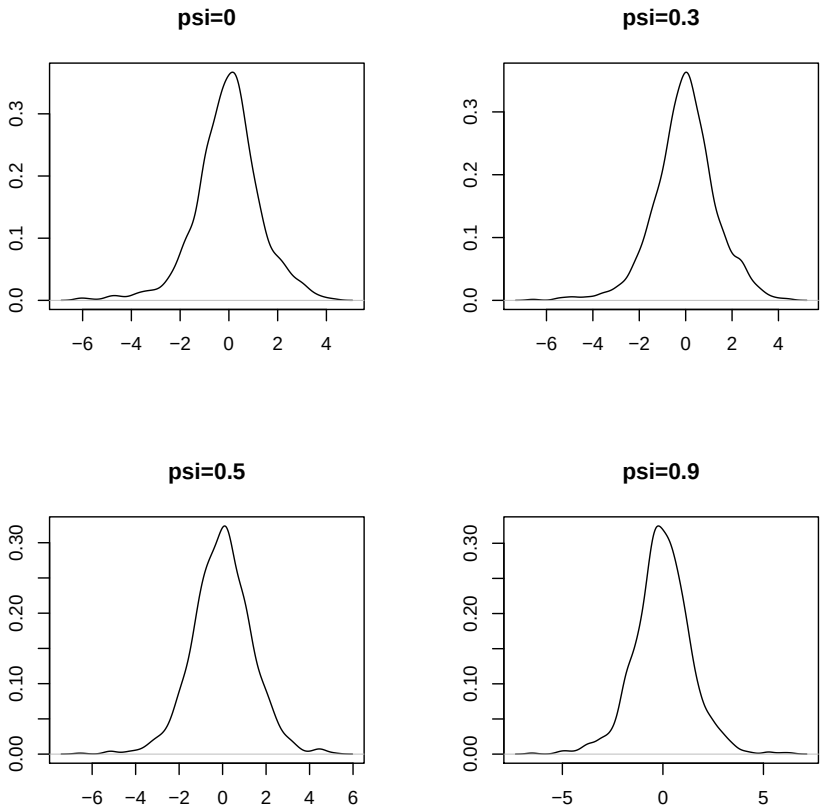


Figure 3.6: Forecast Error Density, Student-3

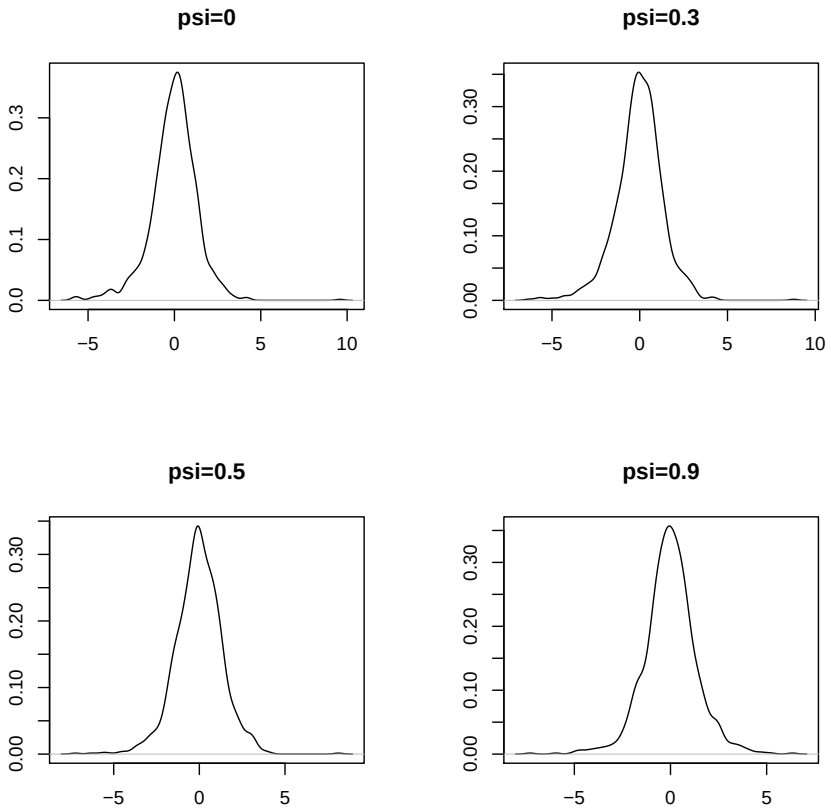


Figure 3.7: Forecast Error Density, Student-4

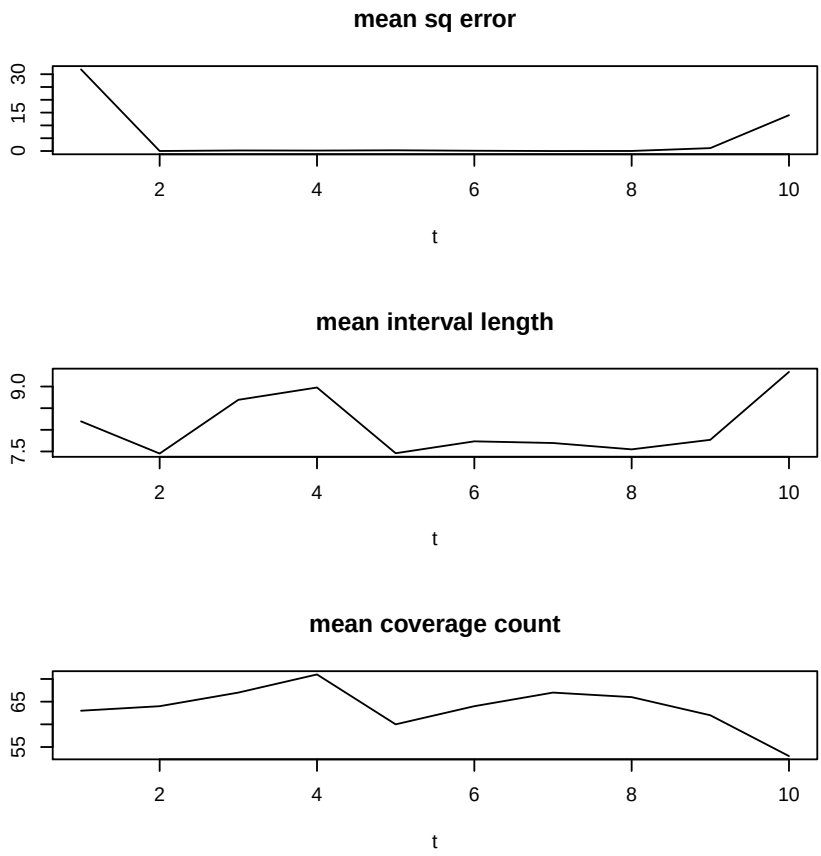


Figure 3.8: Mean Forecasts Properties, Cauchy

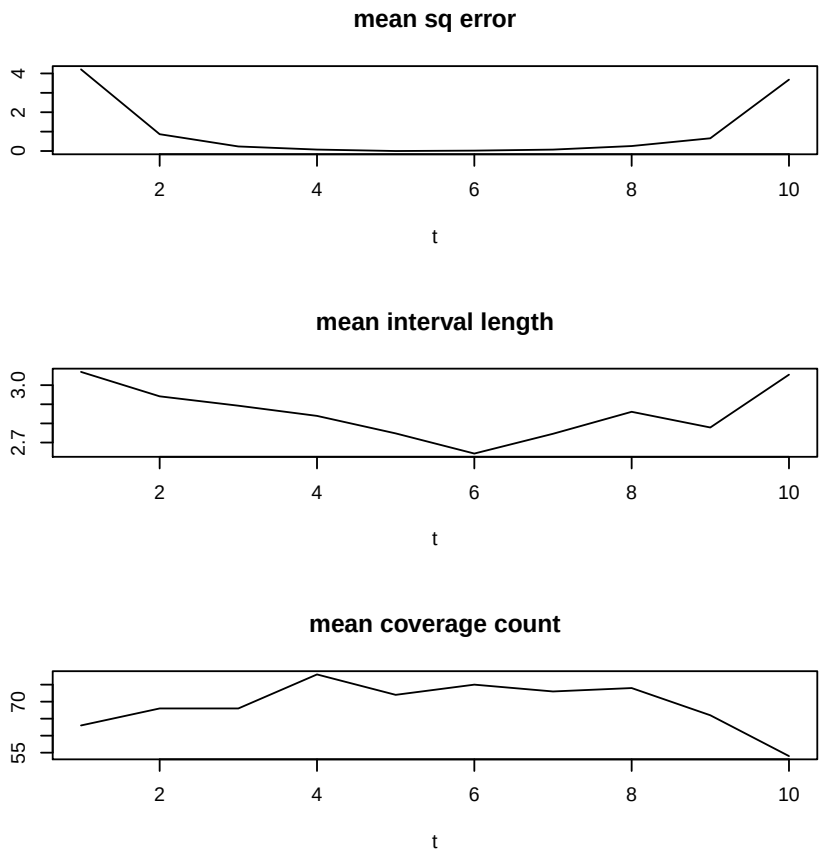


Figure 3.9: Mean Forecast Properties, Student-3

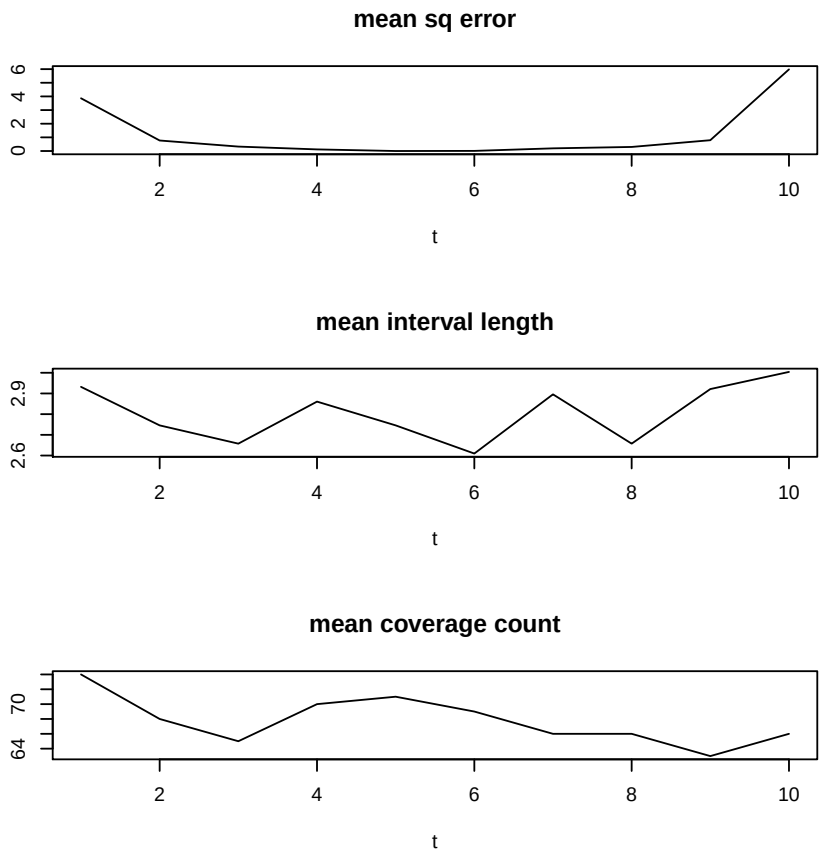


Figure 3.10: Mean Forecast Properties, Student-4

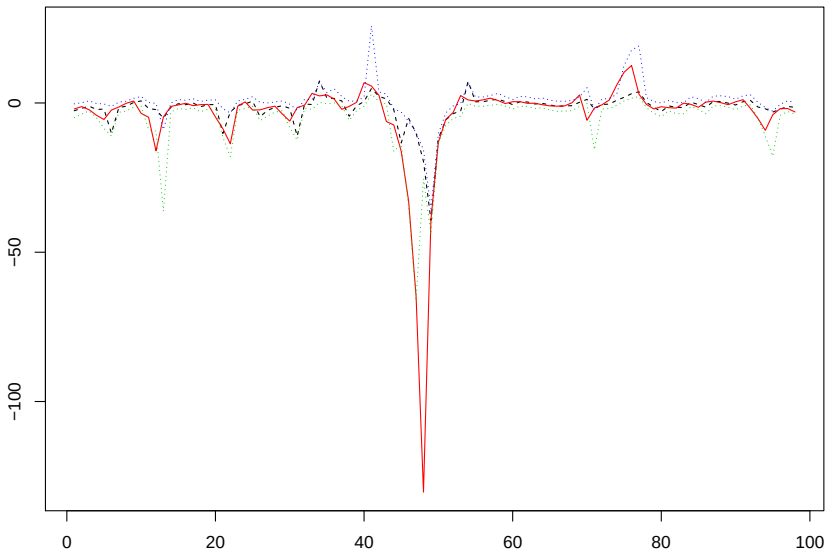


Figure 3.11: MAR(1,1) Cauchy, Forecast and Prediction Interval

lines: solid: MAR(1,1) process, dashed: forecast, dotted: prediction interval

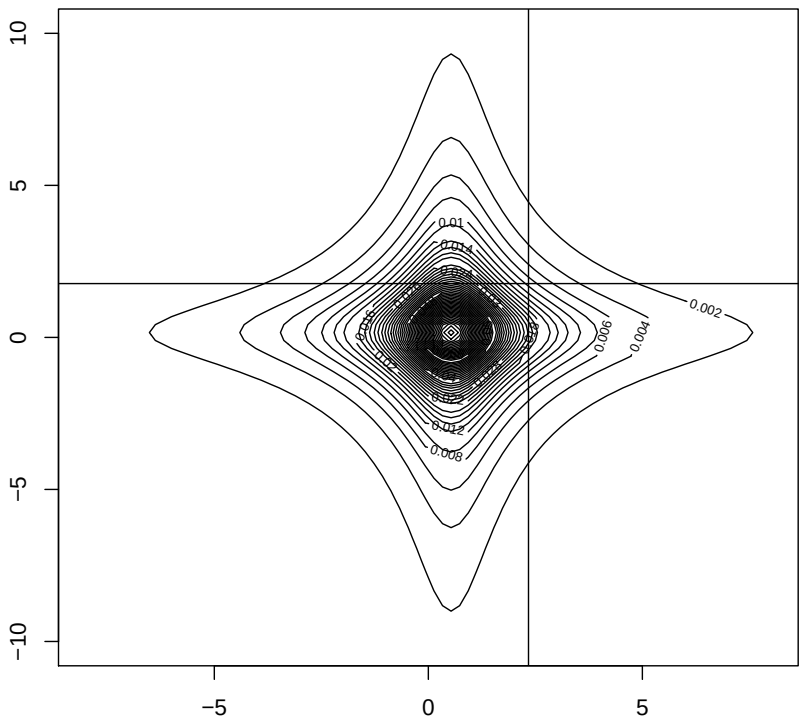


Figure 3.12: MAR(1,1) Cauchy, y(T) = 1.77, u(T) = 1.06

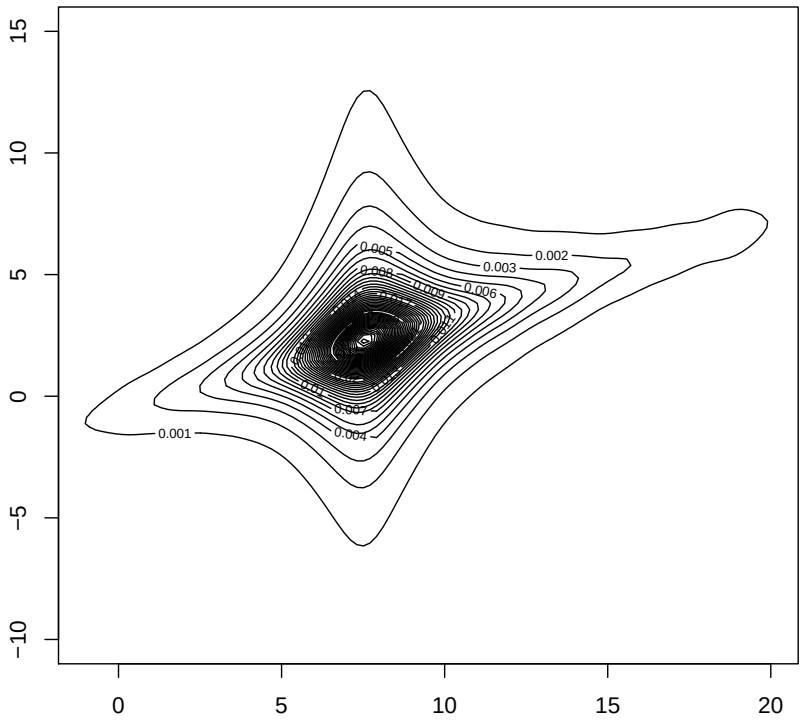


Figure 3.13: MAR(1,1) Cauchy, y(T) = 25.03, u(T) = 22.08

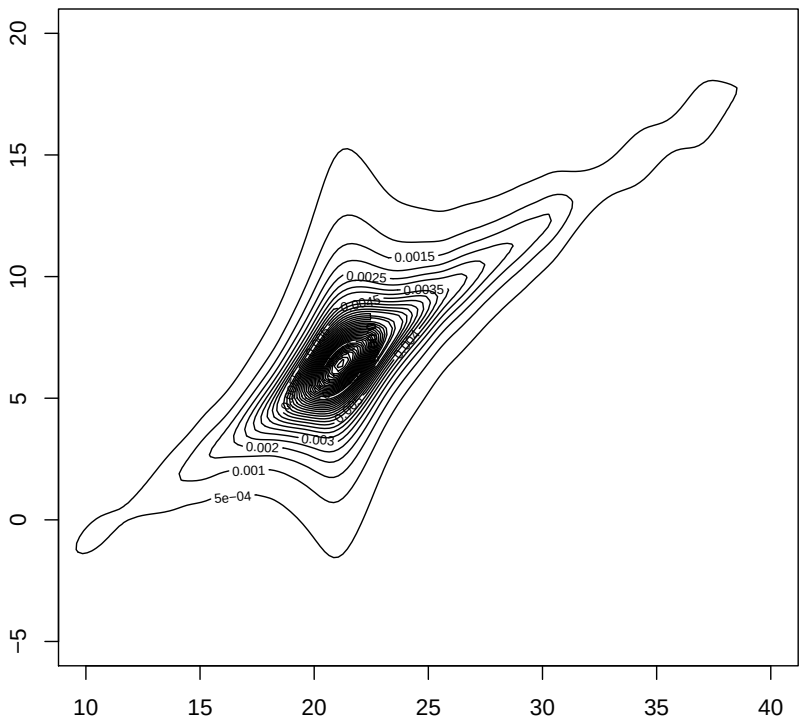


Figure 3.14: MAR(1,1) Cauchy, y(T) = 70.18, u(T) = 58.95

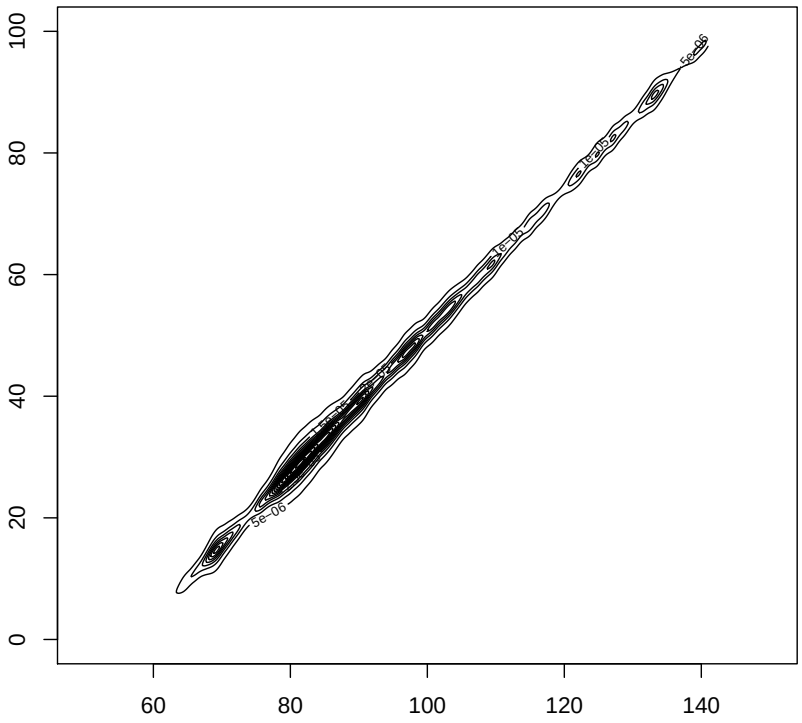


Figure 3.15: MAR(1,1) Cauchy, y(T) = 253.94, u(T) = 184.67

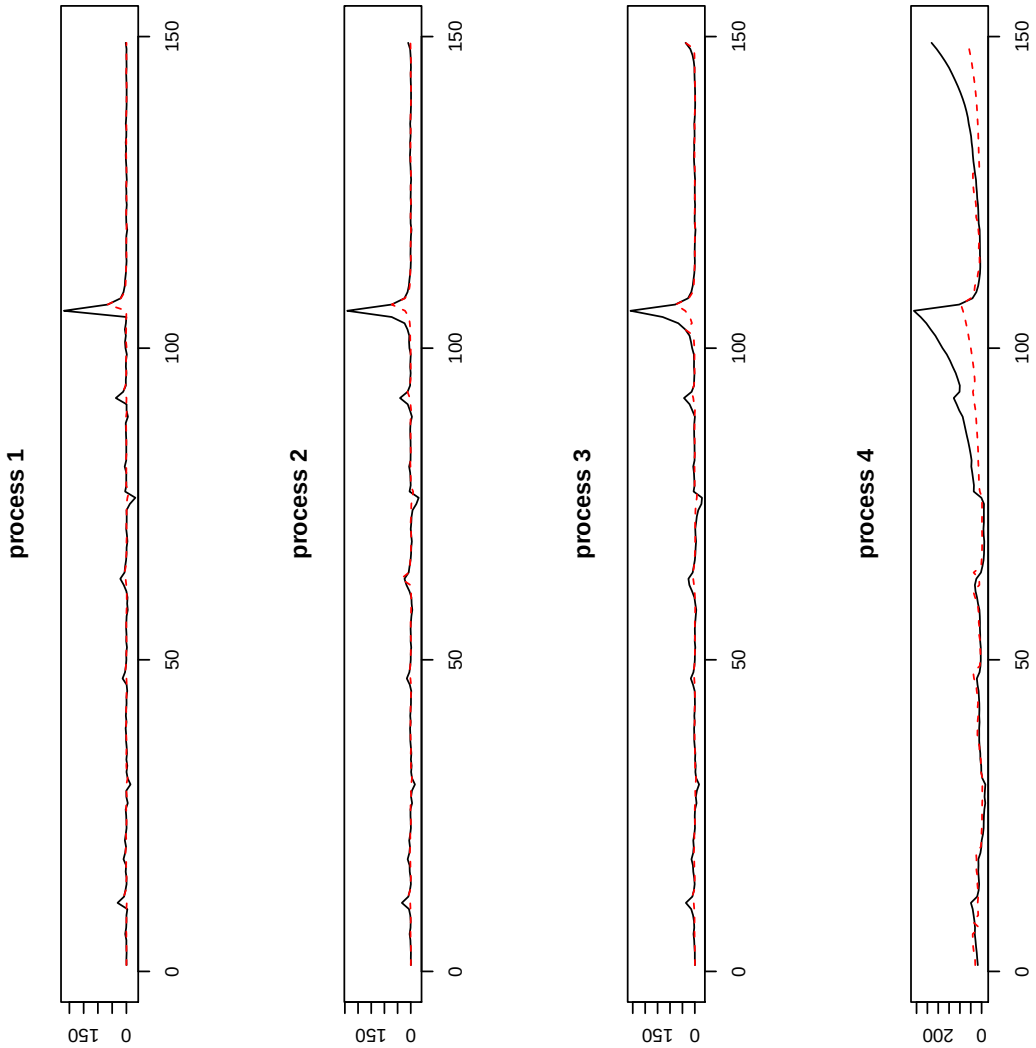


Figure 3.16: MAR(1,1) Cauchy, Joint 1-Step Ahead Forecast

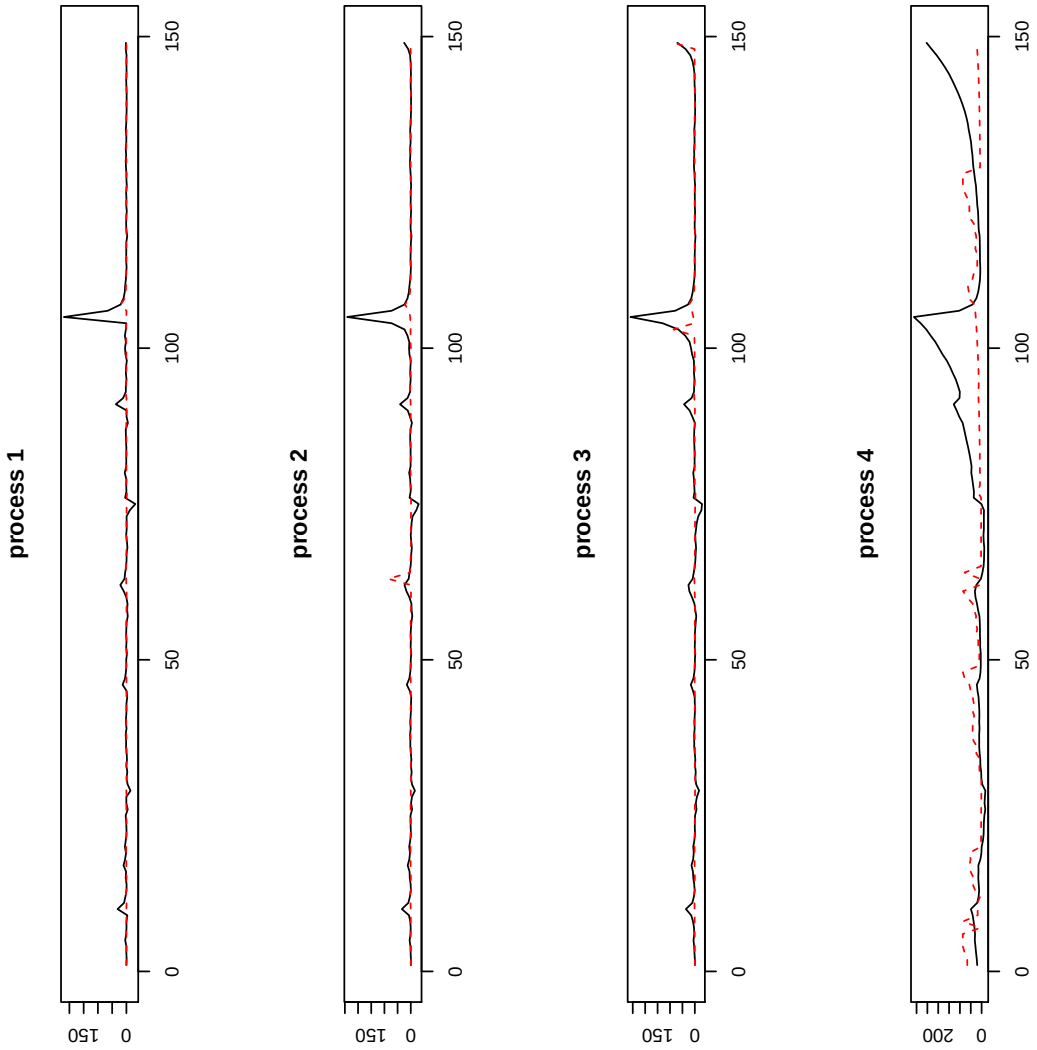
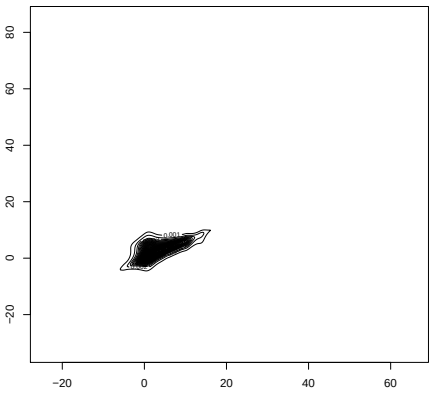
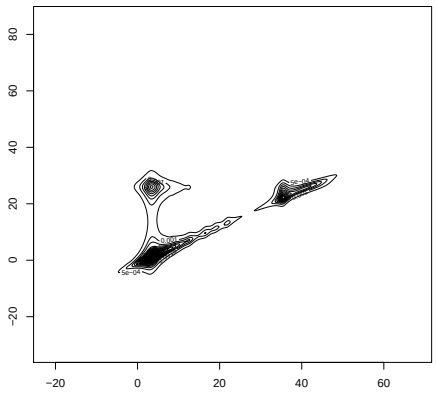


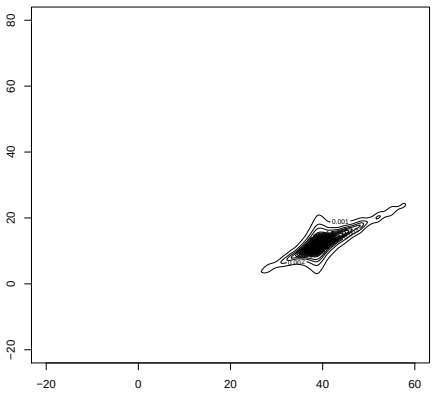
Figure 3.17: MAR(1,1) Cauchy,Joint 2-Step Ahead Forecast



(a) Y(145)=2.73, u(145) = 1.75



(b) Y(150) = 10.86, u(150) = 9.71



(c) Y(155) = 129.61, u(155) = 109.63

Figure 3.18: Outset of Bubble - process 3

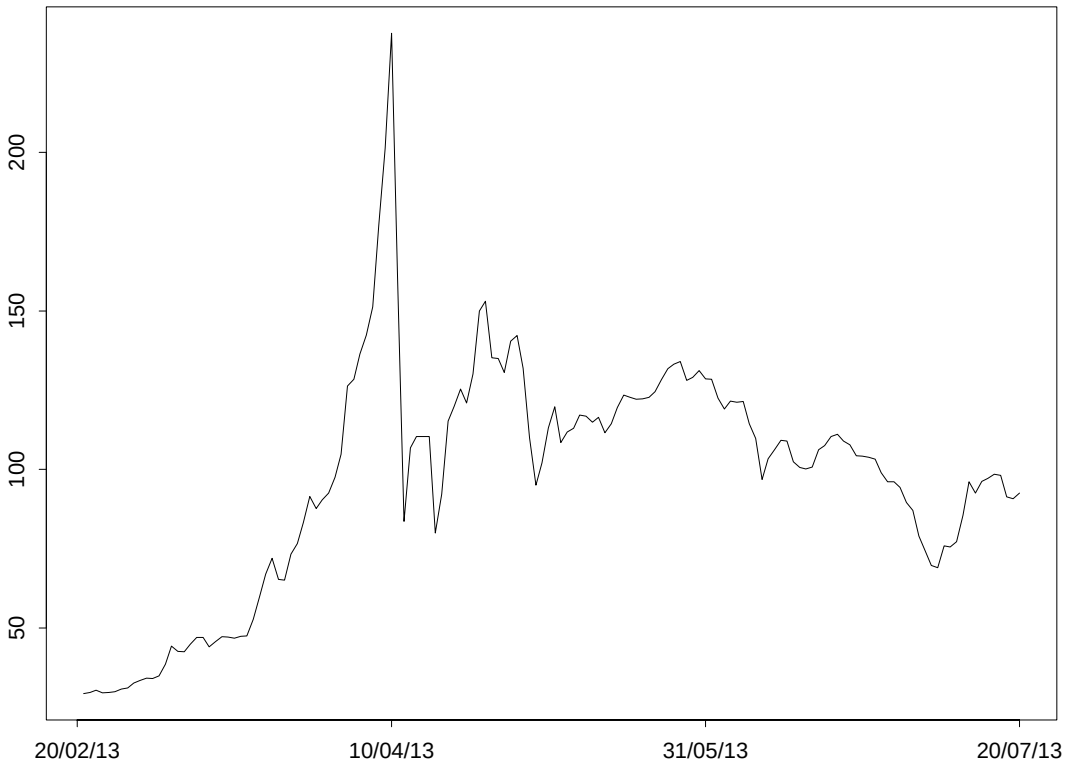
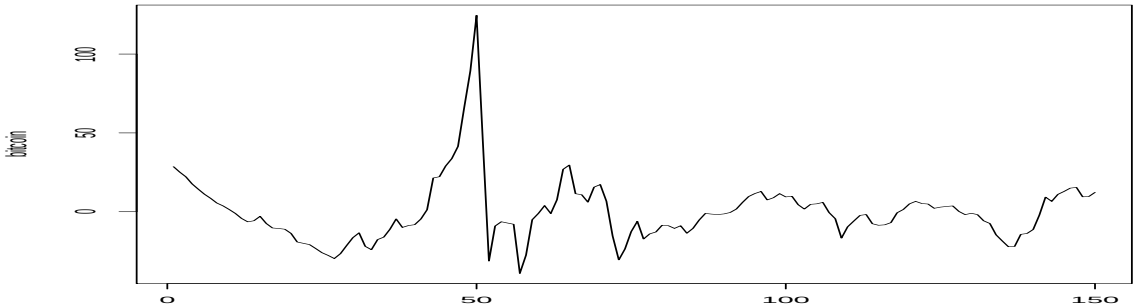
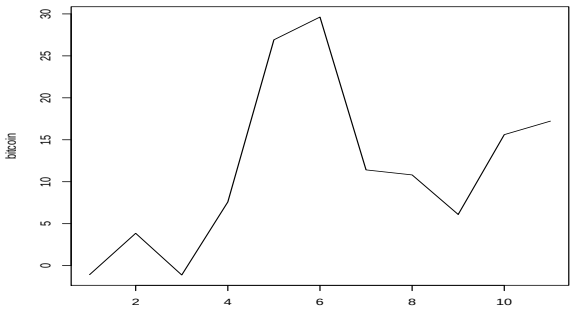


Figure 3.19: Bitcoin/US Dollar exchange rate

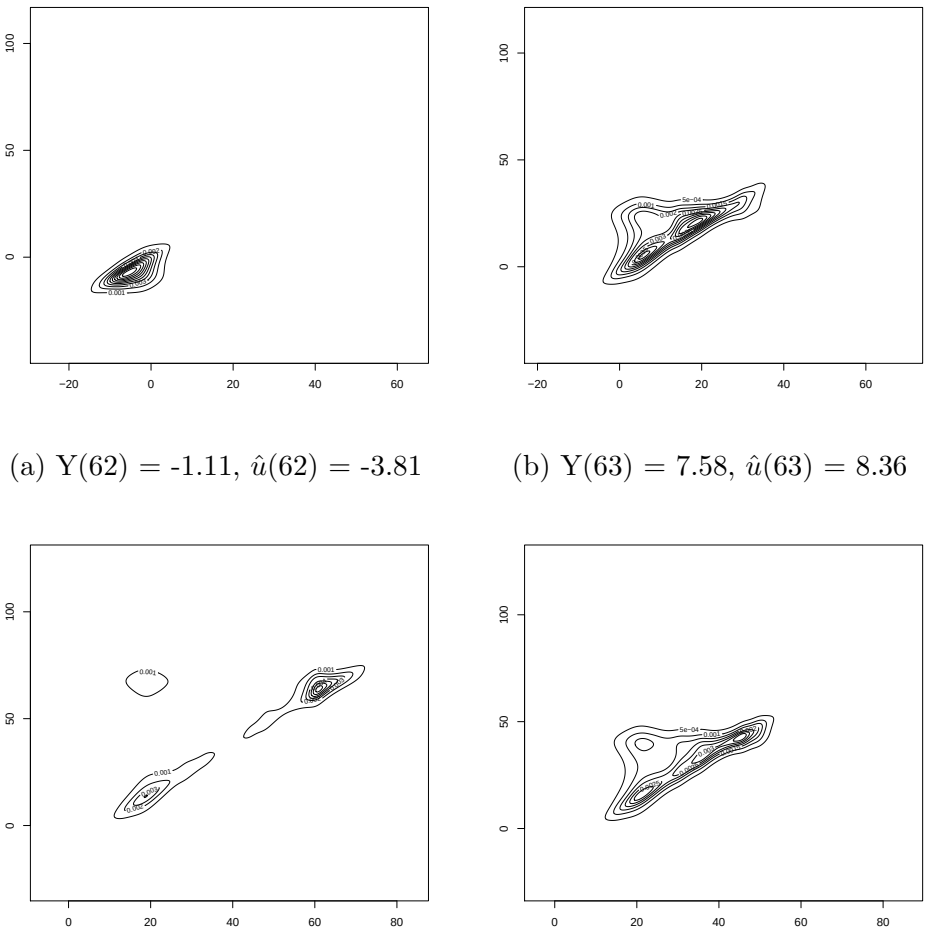


(a) trajectory of length 150



(b) observations 60-70

Figure 3.20: Bitcoin/US Dollar exchange rate - detrended series



(a) Y(62) = -1.11, $\hat{u}(62)$ = -3.81

(b) Y(63) = 7.58, $\hat{u}(63)$ = 8.36

(c) Y(64)=26.911, $\hat{u}(64)$ = 21.61

(d) Y(65) = 29.626, $\hat{u}(65)$ = 10.78

Figure 3.21: Outset of Bubble - Bitcoin

Conclusion

This thesis addressed emerging themes in the intersection of cryptocurrencies and noncausal processes. To provide context for the analysis that followed, Chapter One presented recent literature on Bitcoin and cryptocurrencies. Importantly, much of the literature on cryptocurrencies focused on their informational efficiency of the type Fama (1970) described, along with some applications to improve the portfolio optimization along the lines of Markowitz (1952). However, to the authors knowledge, none of this literature dealt with the nature of the Bitcoin/USD exchange rate to exhibit speculative bubbles. Thus, the bubble nature of Bitcoin presented a natural case of modelling and predicting explosive bubbles in cryptocurrency markets.

Chapter Two showed the noncausal process serves as an effective method to model processes that exhibit explosive bubble patterns. The noncausal process was found to be the most effective method to model the Bitcoin/USD exchange rate when incorporating a period of an explosive bubble pattern.

In Chapter Three we examined the forecasting properties of the nonlinear predictive density estimator. We show the nonlinear predictive density estimator has trouble predicting the timing of the bubble burst. However, by exploiting the ability of the predictive density estimator to produce forecasts at horizon two, a graphical approach can be used to predict the outset of the bubble. Specifically, as the distribution becomes disjointed

we know the bubble is beginning.

Bibliography

[1] Andrews, B., Calder, M. and R. Davis (2009): "Maximum Likelihood Estimation for α-Stable Autoregressive Processes", Annals of Statistics, 37, 1946-1982.

[2] Andrews, B. and R. Davis (2013): "Model Identification for Infinite Variance Autoregressive Processes", Journal of Econometrics, 172, 222-234.

[3] Bai, J. and P. Perron(2003): "Computation and Analysis of Multiple Structural Change Models", Journal of Applied Econometrics, 2003, vol. 18, issue 1, 1-22.

[4] Balkema, G., Embrechts, P., and N. Nolde (2013): "The Shape of Asymptotic Dependence", in Springer Proceedings in Mathematics and Statistics, Vol: Prokhorov and Contemporary Probability Theory", eds: A Shirayev, S. Varadhan and E. Presman, 33, 43-67.

[5] Bariviera, A.F., (2017): "The Inefficiency of Bitcoin Revisited: A Dynamic Approach." Economics Letters 161, 1-4.

[6] Gordon, S.: "What is Ripple?" Bitcoin Magazine. Retrieved from `https://bitcoinmagazine.com/guides/what-ripple/`

[7] Bitcoin Wiki. Introduction (2013, December 4). Retrieved December 4, 2013, from `https://en.bitcoin.it/wiki/Introduction`

[8] Bitcoin Wiki. Controlled Supply (2013, December 4). Retrieved December 4, 2013, from https://en.bitcoin.it/wiki/Controlled_Currency_Supply

[9] Bitcoin Wiki. Trade (2013, December 4). Retrieved December 4, 2013, from https://en.bitcoin.it/wiki/Trade

[10] Bitwise Asset Management: "Presentation to the U.S. Securities and Exchange Commission". March 19, 2019. Retrieved from: https://www.sec.gov/comments/sr-nysearca-2019-01/srnysearca201901-5164833-183434.pdf

[11] Blockchain. Bitcoin Market Capitalization. Retrieved from Bitcoin Block Explorer http://blockchain.info/charts/market_cap

[12] Blanchard, O. (1979): "Speculative Bubbles: Crashes and Rational Expectations", Economic Letters, 3, 387-389.

[13] Blanchard, O. and M. Watson, (1982): "Bubbles, Rational Expectations and Financial Markets", in Crisis in the Economic and Financial Structure, Ed. P. Wachtel, 295-315, Lexington, MA.

[14] Bouri, E., Molnar, P., Azzi, G., Roubaud, D., and L.I. Hagfors, (2017): "On the Hedge and Safe Haven Properties of Bitcoin: Is it Really More than a Diversifier." Finance Research Letters 20, 192-198.

[15] Breidt, F., Davis, R., Lh, K. and M. Rosenblatt (1991): "Maximum Likelihood Estimation for Noncausal Autoregressive Processes", Journal of Multivariate Analysis, 36, 175-198

[16] Breidt, F. and R. Davis (1992): "Time Reversibility, Identifiability and Independence of Innovations for Stationary Time Series", Journal of Time Series Analysis, 13, 273-390

[17] Brockwell, P.J. and R.A. Davis (1987): "Time Series: Theory and Methods", Springer-Verlag. New York.

[18] Brunnermeier, M. (2001): "Asset Pricing under Asymmetric Information: Bubbles, Crashes, Technical Analysis and Herding", Oxford University Press, Oxford.

[19] Cheng, Q. (1992): "On the Unique Representation of Non-Gaussian Linear Processes", Annals of Statistics, 20, 1143-1145.

[20] Choi, I., (1999): "Test the Random Walk Hypothesis for Real Exchange Rates", Journal of Applied Econometrics 14, 293–309.

[21] Darolles, C., Gourieroux, C., and E. Jay (2012): "Robust Portfolio Allocation with Systematic Risk Contribution Restrictions." SSRN Scholarly Paper ID 2192399. Rochester, N: Social Science Research Network

[22] Davis. J (2011): "The Crypto-Currency", The New Yorker. Retrieved from `http://www.newyorker.com/reporting/2011/10/10/111010fa_fact_davis`

[23] Davis, R. and S. Resnick (1985): "Limit Theory for Moving Averages of Random Variables with Regularly Varying Tail Probabilities", The Annals of Probability, 13, 179-195.

[24] Davis, R. and S. Resnick (1986): "Limit Theory for the Sample Covariance and Correlation Functions of Moving Averages", The Annals of Statistics, 14, 533-558.

[25] Davis, R. and L. Song (2012): "Noncausal Vector AR Processes with Application to Economic Time Series", DP Columbia University

[26] Davis, R. and A. Wu (1997): "Bootstrapping M-estimates in Regression and Autoregression with Infinite Variances", Statistica Sinica, 7, 1135-1154.

[27] Diebold F.X., Gunther T.A., and A. Tay (1998): "Evaluating Density Forecasts", International Economic Review, 39, 863-883.

[28] Efron, B. (1982) : "The Bootstrap, Jacknife and Other Resampling Plans", Philadelphia Society of Industrial & Applied Mathematics.

[29] Engle, R.F., (2002): "Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models." Journal of Business and Economics Statistics 20, 339-350.

[30] Ethereum: Blockchain App Platform. Retrieved from `https://www.ethereum.org/`.

[31] Ethereum Homestead Documentation: "Mining". Retrieved from: `www.ethdocs.org/en/latest/mining.html`

[32] Ethereum Wiki: "A Next-Generation Smart Contract and Decentralized Application Platform". Retrieved from: `https://github.com/ethereum/wiki/wiki/White-Paper`

[33] Evans, G. (1991) : "Pitfalls in Testing for Explosive Bubbles in Asset Prices", American Economic Review, 81, 922-930.

[34] European Central Bank, (2015): "Virtual Currency Schemes - A Further Analysis" European Central Bank, Retrieved from: `https://www.ecb.europa.eu/pub/pdf/other/virtualcurrencyschemesen.pdf`

[35] Fama, E.F., (1970): "Efficient Capital Markets: A Review of Theory and Empirical Work." Journal of Finance 25, 383–417.

[36] Flitter, E. (2013): FBI Shuts Alleged Online Drug Marketplace, Silk Road, Reuters. Retrieved from `http://www.reuters.com/article/2013/10/02/us-crime-silkroad-raid-idUSBRE9910TR20131002`

[37] Gelfand, A., and A., Smith (1990)a : "Sampling Based Approaches to Calculating Marginal Densities", J. Amer. Statist. Assoc., 85, 398-409.

[38] Gelfand, A., and A., Smith (1990)b : "Bayesian Statistics Without Tears : A Sampling Resampling Perspective", 46, 84-88.

[39] Gourieroux, C. and A., Hencic (2015): "Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rate" , in Huynh, V., et al. eds. "Econometrics of Risk", Series: Studies in Computational Intelligence, 17-40, Springer.

[40] Gourieroux, C. and J. Jasiak (2016): "Filtering, Prediction and Simulation Methods for Noncausal Processes", Journal of Time Series Analysis, Vol 37, 405-430

[41] Gourieroux, C. and J.M. Zakoian (2017): "Local Explosion Modelling by Non-Causal Process", Royal Statistical Society Series B, Vol 79, Issue 3, 737-756.

[42] Gourieroux, C. and J.M. Zakoian (2015): "On Uniqueness of Moving Average Representation of Heavy Tailed Stationary Processes", Journal of Time Series Analysis, Vol 36, 876–887.

[43] Henry, C., K. Huynh, G. Nicholls (2018): "Bitcoin Awareness and Usage in Canada: An Update.", Bank of Canada Staff Analytical Note

[44] Katsiampa, P., (2017): "Volatility Estimation for Bitcoin: A Comparison of GARCH Models." Economics Letters, 3-6.

[45] Lanne, M., Luoma, A. and J. Luoto (2012): "Bayesian Model Selection and Forecasting in Noncausal Autoregressive Models", Journal of Applied Econometrics, 27, 812-830.

[46] Lanne, M., Luoto, J. and P. Saikkonen (2012) : "Optimal Forecasting of Noncausal Autoregressive Time Series", International Journal of Forecasting, 28, 623-631.

[47] Lanne, M. and P. Saikkonen (2011) : "Noncausal Autoregressions for Economic Time Series", Journal of Time Series Econometrics, De Gruyter, 3, 1-32.

[48] Li, S., (2013): "Bitcoin Now Accepted as Tuition Payment at a Cyprus University", Los Angeles Times. Retrieved from `http://www.latimes.com/business/money/la-fi-mo-cyprus-university-bitcoin-20131120,0,3194094.story#axzz2mXKIff7E`

[49] Litecoin Project. Retrieved March 1, 2019, from `https://litecoin.org/`

[50] Litecoin Block Explorer. Litecoin Block Explorer Charts (2013, December 4). Retrieved December 4, 2013, from `http://ltc.block-explorer.com/charts`

[51] Liu, J. (2013): BTC China the World's Largest Bitcoin Trading Platform, ZD Net. Retrieved from $http://www.zdnet.com/btc-china-the-worlds-largest-bitcoin-trading-platform-7000023316/$

[52] Ljung, G.M., Box, G.E.P., (1978): "On a Measure of the Lack of Fit in Time Series Models." Biometrika 65 (2), 297–303.

[53] Markowitz, H. (1952): "Portfolio Selection", Journal of Finance 7.1, pp 77-91

[54] Muth, J. (1961): "Rational Expectations and the Theory of Price Movements", Econometrica, 29, 315-335.

[55] Nakamoto, S. (2009): "Bitcoin: A Peer-to-Peer Electronic Cash System", `https://bitcoin.org/en/bitcoin-paper`

[56] Newbold, P. (1974): "The Exact Likelihood Function for a Mixed Autoregressive-Moving Average Process", Biometrika, 61, 423-426.

[57] Panagiotidis, T., Stengos, T., and Vravosinos, O. (2018): "The Effects of Markets, Uncertainty and Search Intensity on Bitcoin Returns", The Rimini Centre for Economic Analysis WP.

[58] Phillips, P., Shi, S. and J. Yu (2012): "Testing for Multiple Bubbles", DP Cowles Foundation, 1843.

[59] Phillips, P., Wu, Y. and J.. Yu (2011): "Explosive Behavior in the 1990s Nasdaq: When Did Exuberance Escalate Asset Values?", International Economic Review, 52, 201-226.

[60] Ripple. "Our Company". Retrieved from: `https://ripple.com/company/`

[61] Ripple. "Santander Case Study". Retrieved from: `https://ripple.com/files/case_study_santander.pdf`

[62] Rosenblatt, M. (2000): "Gaussian and Non-Gaussian Linear Time Series and Random Fields", Springer Verlag, New York.

[63] Rubin, D. (1988) : "Using the SIR Algorithm to Simulate Posterior Distribution", in Bayesian Statistics, 3, eds. Bernando, J., De Groot, M., Lindley, D., and A., Smith, Oxford University Press.

[64] Schilling, L.M., Uhlig, H. (2018a): "Currency Substitution under Transaction Costs" forthcoming AEA Papers & Proceedings 109

[65] Schilling, L.M., Uhlig, H. (2018b): "Some Simple Bitcoin Economics", WP

[66] Sparshott, J. (2013): "Web Money Gets Laundering Rule, The Wall Street Journal" . Retrieved from http://online.wsj.com/news/articles/SB10001424127887324373204578374611351125202

[67] Tagaris, K (2013): "Cyprus Details Heavy Losses for Major Bank Customers", Reuters. Retrieved from http://www.reuters.com/article/2013/03/30/us-cyprus-parliament-idUSBRE92G03I20130330

[68] Thomas, M. and V. Morwitz, (2009): "Heuristics in Numerical Cognition: Implications for Pricing", in Vithala R. Rao ed., Handbook of Pricing Research in Marketing (Northampton, MA: Edward Elgar Publishing).

[69] Trimborn S., Li M., and K.M. Hardle, (2017): "Investing with Cryptocurrencies - A Liquidity Constrained Investment Approach"

[70] Urquhart, A., (2016): "The Inefficiency of Bitcoin." Economics Letters 148, 80-82.

[71] Urquhart, A., (2018): "What Causes the Attention of Bitcoin." Economics Letters 166, 40-44.

[72] Velde, F.R., (2013): "Bitcoin: A primer", Chicago Fed Letter. Retrieved from http://www.chicagofed.org/digital_assets/publications/chicago_fed_letter/2013/cfldecember2013_317.pdf

[73] Yermack, D., (2013): "Is Bitcoin a Real Currency? An Economic Appraisal", National Bureau of Economic Research, Inc., NBER Working Papers 19747,

[74] Yonghong, J., N. He, and R. Weihua (2018): "Time-Varying Long-Term Memory in Bitcoin Markets." Finance Research Letters 25, 280-284.

Appendices

Appendix A

The Sampling Importance Resampling (SIR) Method

This approach has been introduced in Rubin (1988) and its properties were studied in Geldfand, Smith (1992). It is a weighted variant of the bootstrap resampling procedure [Efron (1982)]. The aim of the procedure is to draw independent values in a distribution whose density f is known, but the quantile function is difficult to compute. The method requires an instrumental (or importance) distribution with known density g, in which it is easy to draw.

The steps are the following:

Step 1 : Sampling

First draw independent values $X^s, s = 1, \ldots, S$, in distribution G.

Step 2 : Importance Resampling

Then draw independent values $\tilde{Y}^1, \ldots, \tilde{Y}^S$ in the simulated set $\{X^1, \ldots, X^S\}$, with weights $f(X^s)/g(X^s), s = 1, \ldots, S$.

We have :

$$E[a(\tilde{Y}^s)] = EE[a(\tilde{Y}^s)|X^1, \ldots, X^S]$$

$$= \quad E\left\{\left[\sum_{s=1}^{S} \frac{f(X^s)}{g(X^s)} a(X^s)\right] / \left[\sum_{s=1}^{S} \frac{f(X^s)}{g(X^s)}\right]\right\}$$

$$= \quad E\left\{\frac{1}{S}\sum_{s=1}^{S}\left[\frac{f(X^s)}{g(X^s)} a(X^s)\right] / \frac{1}{S}\sum_{s=1}^{S}\left[\frac{f(X^s)}{g(X^s)}\right]\right\},$$

for any integrable function a.

If the number S of replications is large, we get :

$$E[a(\tilde{Y}^s)]$$

$$\simeq \quad E\left\{E\left[\frac{f(X)}{g(X)} a(X)\right] / E\left[\frac{f(X)}{g(X)}\right]\right\}$$

$$= \quad E\left[\frac{f(X)}{g(X)} a(X)\right] / E\left[\frac{f(X)}{g(X)}\right]$$

$$= \quad E[a(Y)].$$

We deduce that the simulated values $\tilde{Y}^1, \ldots, \tilde{Y}^S$ are drawn in distribution F, when $S \to \infty$.

Moreover, since $E[a(\tilde{Y}_s)|X^1, ..., X^S] \approx E[a(\tilde{Y}_s)]$ for S large, these simulated values are asymptotically independent.

The sample size can be as large as desired and the distribution can be multivariate. As usual in Monte-Carlo integration, the more g resembles f, the smaller the number of replications S needed to get accurate results. We can also use different simulation lengths S and S^* in the sampling and resampling steps.

Appendix B

Marginal density of y_t when errors are Cauchy distributed.

It follows from formulas (3.2) and (3.3)

$$u_t = (1 - \phi L)y_t; \quad (1 - \psi L^{-1})u_t = \varepsilon_t,$$

and

$$v_t = (1 - \psi L^{-1})y_t; \quad (1 - \phi L)v_t = \varepsilon_t,$$

that

$$\begin{aligned} u_t &= \varepsilon_t + \psi\varepsilon_{t+1} + \psi^2\varepsilon_{t+2} + \cdots \\ v_t &= \varepsilon_t + \phi\varepsilon_{t+1} + \phi^2\varepsilon_{t+2} + \cdots . \end{aligned}$$

Hence

$$\begin{aligned} y_t &= u_t + \psi u_{t-1} + \psi^2 u_{t-2} + \cdots \\ &= \varepsilon_t + \phi\varepsilon_{t+1} + \phi^2\varepsilon_{t+2} + \cdots \\ &+ \phi(\varepsilon_{t-1} + \psi\varepsilon_t + \psi^2\varepsilon_{t+1} + \cdots) \\ &+ \phi^2(\varepsilon_{t-2} + \psi\varepsilon_{t-1} + \psi^2\varepsilon_t + \cdots) \\ &+ \cdots \end{aligned}$$

$$
\begin{aligned}
y_t &= \varepsilon_t[1+\phi\psi+\phi^2\psi^2+\cdots+\phi^3\psi^3+\cdots] \\
&+ \varepsilon_{t+1}[\psi+\phi\psi^2+\phi^2\psi^3+\cdots] \\
&+ \varepsilon_{t+2}[\psi^2+\phi\psi^3+\phi^2\psi+\cdots] \\
&+ \cdots \\
&+ \varepsilon_{t-1}[\phi+\phi^2\psi+\phi^3\psi^2+\cdots] \\
&+ \cdots
\end{aligned}
$$

We get

$$
\begin{aligned}
y_t &= \sum_{j=0}^{\infty} u_{t-j}\phi^j \\
&= \sum_{j=0}^{\infty}[\phi^j \sum_{h=0}^{\infty} \psi^h \varepsilon_{t+-j+h}] \\
&= \sum_{j=0}^{\infty}\sum_{h=0}^{\infty}[\phi^j\psi^h \varepsilon_{t+-j+h}]
\end{aligned}
$$

We introduce a change of index and set $k = j - h$.

$$
\begin{aligned}
y_t &= \sum_{h=0}^{\infty}\sum_{j=0}^{\infty}[\phi^j\psi^h\varepsilon_{t+-j+h}] \\
&= \sum_{k=-\infty}^{\infty}[\varepsilon_{t-k}\sum_{j\geq 0, j-h\geq 0}\phi^j\psi^{j-k}] \\
&= \sum_{k=-\infty}^{\infty}[\varepsilon_{t-k}\psi^{-k}\sum_{j\geq 0, j-h\geq 0}(\phi^j\psi^j)] \\
&= \sum_{k=0}^{\infty}[\varepsilon_{t-k}\psi^{-k}\sum_{j\geq k}(\phi^j\psi^j)] + \sum_{k=-1}^{-\infty}[\varepsilon_{t-k}\psi^{-k}\sum_{j=0}^{\infty}(\phi^j\psi^j)] \\
&= \sum_{k=0}^{\infty}[\varepsilon_{t-k}\psi^{-k}\phi^k\psi^k/(1-\phi\psi)] + \sum_{k=1}^{\infty}\varepsilon_{t+k}\psi^k/(1-\phi\psi) \\
&= \sum_{k=0}^{\infty}\varepsilon_{t-k}\phi^{-k}/(1-\phi\psi) + \sum_{k=1}^{\infty}\varepsilon_{t+k}\psi^k 1/(1-\phi\psi)
\end{aligned}
$$

When ε_t are iid Cauchy distributed, then: $(1-\phi)u_t \sim$ Cauchy, $(1-\psi)v_t \sim$ Cauchy, and $cy_t \sim$ Cauchy, where:

$$
\begin{aligned}
1/c &= \sum_{k=0}^{\infty}\varepsilon_{t-k}\phi^{-k}/(1-\phi\psi) + \sum_{k=1}^{\infty}\varepsilon_{t+k}\psi^k 1/(1-\phi\psi) \\
&= \frac{1}{1-\phi\psi}\frac{1}{1-\phi} + \frac{1}{1-\phi\psi}\frac{\psi}{1-\psi} \\
&= \frac{1-\psi+\psi(1-\phi)}{(1-\phi\psi)(1-\phi)(1-\psi)} = \frac{1}{(1-\phi)(1-\psi)}
\end{aligned}
$$

Hence $(1-\phi)(1-\psi)y_t \sim$ Cauchy.

Appendix C

Equivalence of (3.7) and (3.8)

To derive the predictive density of y_{T+1}, y_{T+2}, conditional on $y_t, \hat{u}_T$, we observe that:

$$\begin{aligned} y_{T+1} &= u_{T+1} + \phi y_T \\ y_{T+2} &= u_{T+2} + \phi y_{T+1} \end{aligned}$$

which implies

$$\begin{aligned} u_{T+1} &= y_{T+1} - \phi y_T \\ u_{T+2} &= y_{T+2} - \phi y_{T+1} \end{aligned}$$

The Jacobian of the transformation from u to y is:

$$\frac{\partial[u_{T+1}, u_{T+2}]'}{\partial[y_{T+1}, y_{T+2}]} = |\det| \begin{matrix} 1 & 0 \\ -\phi & 1 \end{matrix} || = 1$$

Formula (3.8) follows by substitution.

Appendix D

Proof of Proposition 1

Proof:

Let us show that Ellipse 1 defined by:

$$(y_t - \psi y_{t+1})^2 + (1-\psi)^2 y_{t+1}^2 = c - 1,$$

is outside or tangent to the isodensity curve and that Ellipse 2 defined by

$$(y_t - \psi y_{t+1})^2 + (1-\psi)^2 y_{t+1}^2 = 2[\sqrt{c} - 1]$$

is inside or tangent to the isodensity curve. To do that , we consider the constrained optimization of the following objective function:

$$opt_{y_t, y_{t+1}} (y_t - \psi y_{t+1})^2 + (1-\psi)^2 y_{t+1}^2$$

under the constraint corresponding to the isodensity curve:

$$[1 + (y_t - \psi y_{t+1})^2][1 + (1-\psi)^2 y_{t+1}^2] = c.$$

Upon the following change of variables: $X = y_t - \psi y_{t+1}$, $Y = (1-\psi) y_{t+1}$, the optimization becomes:

$$opt_{X,Y} X^2 + Y^2, \text{ subject to } (1+X^2)(1+Y^2) = c$$

This optimization is a maximization or a minimization, depending on the ellipse of interest, that are Ellipse 1 and Ellipse 2, respectively.

Let us introduce the Lagrange multiplier λ and rewrite the above expression:

$$opt_{X,Y} X^2 + Y^2 - \lambda[(1+X^2)(1+Y^2) - c].$$

The FOC are:

$$\begin{aligned} \frac{\partial}{\partial X} &= 2X - 2X\lambda(1+Y^2) = 0 \\ \frac{\partial}{\partial Y} &= 2Y - 2Y\lambda(1+X^2) = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial X} &= \text{if } X = 0 \text{ or } \ 1 - \lambda(1+Y^2) = 0 \\ \frac{\partial}{\partial Y} &= \text{if } Y = 0 \text{ or } \ 1 - \lambda(1+X^2) = 0 \end{aligned}$$

The solutions are the following:

1) If $X = 0, Y = 0$, this case is impossible as it does not satisfy the constraint.

2) For $X \neq 0$ and $Y \neq 0$, we get $1 - \lambda(1+Y^2) = 0$ and $1 - \lambda(1+X^2) = 0$, which implies that $X^2 = Y^2$, i.e.

$$(y_t - \psi y_{t+1})^2 = (1-\psi)^2 y_{t+1}^2$$

In this case, the constraint is satisfied:

$$[1+(y_t-\psi y_{t+1})^2][1+(1-\psi)^2 y_{t+1}^2]=c.$$

$$\iff \quad [1+(1-\psi)^2 y_{t+1}^2]^2=c$$

$$\iff \quad (1-\psi)^2 y_{t+1}^2=\sqrt{c}$$

Hence :

$$y_t-\psi y_{t+1}=\pm(1-\psi)y_{t+1}.$$

This equality holds when either $y_t-\psi y_{t+1}=y_{t+1}-\psi y_{t+1}$ or $y_t-\psi y_{t+1}=-y_{t+1}+\psi y_{t+1}$ Therefore, the tangency points for Ellipse 2 are $y_t=y_{t+1}$ and $y_t=(2\psi-1)y_{t+1}$.

3) If $X\neq 0, Y=0$, then $X^2=c-1$ and $y_{t+1}=0$

For Ellipse 1 , the tangency points are $c-1$, $y_t=\psi y_{t+1}$ or $y_{t+1}=0$.

4) If $X=0, Y\neq 0$, then $1+Y^2=c$ and $X^2+Y^2=c-1$.

The tangency points of Ellipse 1 are $y_t=\Psi y_{t+1}$ and $y_{t+1}^2=(c-1)/(1-\psi)^2$, which implies $y_{t+1}=\pm\sqrt{c-1}/(1-\psi)$

QED