

2017

Bubble Modeling by Mixed Casual-Noncausal Autoregressive Processes

Anton Lobach
University of Rhode Island, antonlobach@uri.edu

Follow this and additional works at: <https://digitalcommons.uri.edu/theses>

Terms of Use

All rights reserved under copyright.

Recommended Citation

Lobach, Anton, "Bubble Modeling by Mixed Casual-Noncausal Autoregressive Processes" (2017). *Open Access Master's Theses*. Paper 1138.
<https://digitalcommons.uri.edu/theses/1138>

This Thesis is brought to you by the University of Rhode Island. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu. For permission to reuse copyrighted content, contact the author directly.

BUBBLE MODELING BY MIXED CAUSAL-NONCAUSAL
AUTOREGRESSIVE PROCESSES

BY
ANTON LOBACH

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
STATISTICS

UNIVERSITY OF RHODE ISLAND

2017

MASTER OF SCIENCE THESIS
OF
ANTON LOBACH

APPROVED:

Thesis Committee:

Major Professor Gavino Puggioni

Steffen Ventz

Georges Tsafack

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2017

ABSTRACT

Some financial time series exhibit short periods of explosive local trends followed by an abrupt decline. Such trends can be a result of speculative bubble phenomena. A bubble is formed when investors' future profits expectations influence the present market value of securities. Mixed causal-noncausal autoregressive processes (MAR) are able to better capture such behavior in comparison to standard causal ARIMA models. In the first part of this work we propose an alternative distribution (Voigt) to model the disturbances in the MAR processes. The Voigt, a convolution of Gaussian and Cauchy distributions, is used in atomic and molecular spectroscopy, and is more flexible than other heavy-tail distributions. The second part of this work extends the MAR models to Markov switching mixed causal-noncausal autoregressive processes (MSMAR) with Cauchy distributed errors to account for changes in regime at different times. Parameter estimation of both models MAR with Voigt errors and MSMAR is performed in a Bayesian framework via MCMC algorithms. The models are tested for performance with a simulation study and then applied to Bitcoin/USD exchange rate data.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Gavino Puggioni for his patience and time that he spent with me discussing statistical methods, giving me valuable advises and ideas regarding my thesis, teaching me time series and Bayesian statistics classes that contributed a lot to my knowledge of statistical methods. I would like to thank Dr. Steffen Ventz, Dr. Leonard Kahn and Dr. Georges Tsafack for being my committee members, reviewing my work and providing useful comments and insights. I would also like to thank Dr. Liliana Gonzales and Dr. Natalia Katenka for teaching me statistical methods and theory. I want to thank Lorraine Berube for reminding me about all necessary forms and documents needed for my graduation. I would like to thank my friends: Gabriel DePace for discussing topics in computer science and programming, Linden Wyatt and Nicholas Bianchi for discussing physics and mathematics and good times spent at URI.

Lastly, I would like to thank my grandfather who, unfortunately, passed away, for opening the world of engineering and science for me, that became my passion throughout my life.

DEDICATION

I am dedicating this work to my parents and my wife. My success is largely attributed to you. Thank you for raising, loving and supporting me throughout my life!

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iii
DEDICATION	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	vii
LIST OF TABLES	viii
CHAPTER	
1 Introduction	1
List of References	3
2 Mixed Causal-Noncausal Autoregressive Processes with Voigt Errors (MARV)	4
2.1 Voigt Distribution	5
2.2 Estimation of the Parameters of the MAR Processes	8
2.3 Simulation of the MAR Processes	12
2.4 Simulation Study	13
2.5 Analysis of the Bitcoin/USD Exchange Rate Data	18
List of References	28
3 Markov Switching Mixed Causal-Noncausal Autoregressive Processes (MSMAR)	30
3.1 The MSMAR processes	30
3.2 Parameter Estimation	31

	Page
3.3 Simulation of the MSMAR processes	36
3.4 Simulation Study	38
3.5 Application to Bitcoin/USD Exchange Rate	40
List of References	46
4 Conclusion and Future Research Directions	47
List of References	48
BIBLIOGRAPHY	49

LIST OF FIGURES

Figure		Page
1	Shapes of the Voigt density.	7
2	MARV(1, 1) fit and the residuals.	15
3	Simulation study of the MARV processes with the extreme ratios of r	17
4	Bitcoin/USD exchange rate over a period of 02/20/2013 - 07/20/2013	18
5	Global trend and autocorrelation structure in the Bitcoin/USD exchange rate series.	21
6	The MARV(2, 2) fit added to the global trend along with the Bitcoin/USD exchange rate series over a period of 02/20/2013 -07/20/2013.	22
7	Bitcoin/USD exchange rate and the corresponding fit summaries.	23
8	Visual summary of the Bitcoin/USD exchange rate series over a period of 2010-07-17 - 2014-02-25	26
9	Visual summary of the best MARC fit in application to the Bitcoin/USD exchange rate series over a period of 2010-07-17 - 2014-02-25	27
10	Simulated MSMAR process.	39
11	Posterior modes for the regimes of the Bitcoin/USD exchange rate data	44
12	Visual summaries of the fit of the best MSMAR model.	45

LIST OF TABLES

Table		Page
1	Posterior means and 95% quantile based coverage intervals for the parameters of various MAR models with Voigt errors fitted to the simulated data.	14
2	Posterior means and 95% quantile based coverage intervals for the parameters of the MAR(1, 1) models for the simulated data with the distribution of errors $\mathcal{V}(0, 2, 0.1)$	16
3	Posterior means and 95% quantile based coverage intervals for the parameters of the MAR(1, 1) models for the simulated data with the distribution of errors $\mathcal{V}(0, 0.2, 2)$	17
4	Posterior means and 95% quantile based coverage intervals for the parameters of various MAR models with Cauchy and Voigt errors in application to the Bitcoin/USD exchange rate.	20
5	Posterior means and 95% quantile based coverage intervals for the parameters of various MAR models with Cauchy and Voigt errors.	28
6	Posterior means and 95% quantile based coverage intervals for the parameters of various MSMAR models applied to simulated data.	40
7	Posterior means and 95% quantile based coverage intervals for the parameters of various MSMAR models in application to the Bitcoin/USD exchange rate series.	46

CHAPTER 1

Introduction

Bubble effects have been known primarily as a financial phenomenon. In the *New Palgrave: A Dictionary of Economics* a bubble is defined as:

“A sharp rise in the price of an asset or a range of assets in a continuous process, with the initial rise generating expectations of further rises and attracting new buyers - generally speculators interested in profits from trading in the asset rather than its use or earnings capacity.” [1]

Peter M. Garber provided an overview of several bubbles in his “Famous First Bubbles” paper [2]. One of them and arguably the most famous is “tulipmania” which took place in Netherlands during 1634-1637. Rare tulip varieties were sold at high prices on the market created by flower growers and those who enjoyed tulips. By 1636 the rapid price growth of the flowers attracted people who thought that they could make money buying and selling bulbs. This resulted in a further price increase during the period between November 1636 and January 1637. At this time, the price of one rare tulip bulb exceeded the annual salary of a wealthy person. The rapid price growth got to the point when buyers refused to pay for tulips. In February 1637 the market suddenly crashed. This resulted in economic hardship in Netherlands for many years. There are many examples of bubbles. An overview of the Mississippi Bubble (1719 - 1720) and the South Sea Bubble (1720) can be found in [2].

It is important to study bubbles as they have a significant effect on human life and national economy, as it can be seen from “tulipmania” example.

According to [3] there are three parts in a bubble phenomenon: an object that lies in the center of the phenomenon, and two object’s attributes which are

called external and driving values. In the tulipmania example, the object is the tulip. The external value of the tulip is its market value, the price payed for the bulb. The tulip's driving value is the people's expectation of the profit that can be made by reselling it in the near future. The external value and driving value are not related to the "true" value of a tulip, which is called a "fundamental value". A bubble appears when a positive feedback loop is formed between the object's driving value and its external value [3]. In other words, people's expectations of future profits influence the present market value of an object and vice-versa. A feedback loop is a fundamental notion in system theory that emerges in fields such as organizational theory, electrical engineering, epidemiology, etc. [3].

Mixed causal-noncausal autoregressive processes take into account the positive feedback by allowing present values to depend explicitly on future and past values. In contrast, purely causal autoregressive processes force the variable to depend only on past values. Due to explicit dependence on future the noncausal autoregressive processes often provide a better fit to economic time series where future expectations play a central role, such as bubble phenomena, in comparison to standard purely causal ARMA models [4]. Several authors attempted to model bubble effects by mixed and noncausal autoregressive processes. In [5] a Bitcoin/USD exchange rate was modeled by noncausal autoregressive process with Cauchy distributed errors. An adequacy of the noncausal Cauchy AR model on Nasdaq composite price index was tested in [6]. As an alternative to Cauchy, this work will introduce Voigt distribution to model disturbances of the MAR processes.

An attempt to model noncausal dynamics by nonlinear method such as neural network can be found in [7].

Regular MAR models are not flexible enough to account for regime switching in temporal dependence. To extend the regular MAR models we propose

a Markov-switching mixed causal-noncausal autoregressive processes (MSMAR), that take into account the regime switching in the correlation structure. In this work the regime switching of the MSMAR is modeled in the MAR coefficients and parameters of the error distributions but not in the level, although it can be also incorporated in the model. A similar process that tries to model the dependence of the parameters of the noncausal processes on time was proposed by [8] and is called the "time-varying AR" (TVP-AR).

This work will address performance of the proposed models by comparing MAR with Voigt distributed errors (MARV) to MAR with Cauchy distributed errors (MARC). The performance of MSMAR will be compared to MARC models by analyzing the Bitcoin/USD data [5].

List of References

- [1] J. J. Siegel, "What is an asset price bubble? an operational definition," *European financial management*, vol. 9, no. 1, pp. 11–24, 2003.
- [2] P. M. Garber, "Famous first bubbles," *The Journal of Economic Perspectives*, vol. 4, no. 2, pp. 35–54, 1990.
- [3] Y. Nov and O. Nov, "Living in a bubble? toward a unified bubble theory," *International Journal of General Systems*, vol. 37, no. 5, pp. 627–635, 2008.
- [4] C. Gouriéroux, J.-M. Zakoian, *et al.*, "Explosive bubble modelling by noncausal process," Tech. Rep., 2013.
- [5] A. Hencic and C. Gouriéroux, "Noncausal autoregressive model in application to bitcoin/usd exchange rates," in *Econometrics of Risk*. Springer, 2015, pp. 17–40.
- [6] C. Gouriéroux and J.-M. Zakoian, "Local explosion modelling by noncausal process," 2016.
- [7] Y. Ouyang and H. Yin, "Time series prediction with a non-causal neural network," in *Computational Intelligence for Financial Engineering & Economics (CIFEr), 2104 IEEE Conference on*. IEEE, 2014, pp. 25–31.
- [8] M. Lanne and J. Luoto, "A noncausal autoregressive model with time-varying parameters: An application to us inflation," 2013.

CHAPTER 2

Mixed Causal-Noncausal Autoregressive Processes with Voigt Errors (MARV)

As was mentioned in chapter 1, the mixed causal-noncausal autoregressive processes (MAR) allow present observations to depend on future as well as the past values, making it more flexible in modeling financial time series where future expectations are believed to have an effect on present observations. More specifically, an MAR of order (s, r) is a discrete-time stochastic process $\{y_t : t \in \mathbb{Z}\}$ defined as:

$$\Psi(L^{-1})\Phi(L)y_t = \epsilon_t, \quad \epsilon_t \stackrel{\text{i.i.d.}}{\sim} f_\epsilon(\cdot|\boldsymbol{\eta}) \quad (1)$$

Where L and L^{-1} are the “lag-backward” $Ly_t = y_{t-1}$ and the “lag-forward” $L^{-1}y_t = y_{t+1}$ operators, respectively. The disturbances ϵ_t are assumed to be a white noise with some distribution $f_\epsilon(\cdot|\boldsymbol{\eta})$ parameterized by $\boldsymbol{\eta}$. In order for the process (1) to be stationary the roots of the noncausal $\Psi(L^{-1}) = 1 - \sum_{j=1}^s \psi_j L^{-j}$ and causal $\Phi(L) = 1 - \sum_{j=1}^r \psi_j L^j$ polynomials must lie outside of the unit circle [1]. The noncausal and causal polynomials can be inverted and the MAR process (1) can have an infinite two-sided moving average representation in terms of the disturbances:

$$y_t = \frac{\epsilon_t}{\Psi(L^{-1})\Phi(L)} = \sum_{k=-\infty}^{\infty} \omega_k \epsilon_{t-k}$$

The weighted average at time t is over the disturbances at current, past and future times. The errors ϵ_t can have a distribution with infinite variance and expectation [2]. Noncausal and causal parameters of the MAR process (1) are non identifiable in the case when disturbances $\{\epsilon_t\}$ are Gaussian white noise [1], see [3] for discussion.

Purely causal or (backward looking) autoregressive processes, as considered in Box-Jenkins methodology [4], are widely used in time series analysis and are special

cases of the mixed AR processes. This can be easily seen by setting the noncausal coefficients ψ to 0 in equation 1, so it reduces to a purely causal autoregressive process,

$$Y_t = \sum_{j=1}^p \phi_j Y_{t-j} + \epsilon_t, \quad \epsilon_t \stackrel{\text{i.i.d.}}{\sim} f_\epsilon(\cdot)$$

On the other hand, when all of the ϕ coefficients are equal to 0, the process (1) reduces to a purely noncausal (forward looking) autoregressive process,

$$Y_t = \sum_{l=1}^d \psi_l Y_{t+l} + \epsilon_t, \quad \epsilon_t \stackrel{\text{i.i.d.}}{\sim} f_\epsilon(\cdot)$$

2.1 Voigt Distribution

As was mentioned previously, the causal and noncausal coefficients of the MAR processes are non identifiable under the Gaussian assumption for the errors. Because of the explosive behavior of bubbles and the identifiability problem a typical choice for the distribution of errors found in the literature is Cauchy and t-distribution, see for example [2], [1], [5]. In this work we want to compare the performance of MAR with Voigt errors to the MAR with Cauchy errors.

Voigt distribution is named after Woldemar Voigt, a German physicist. It is used in physics for atomic and molecular spectroscopy [6]. In particular, the Voigt line shape is used to model the distribution of photoelectron energies in x-ray photoelectron spectroscopy (XPS) where the data generating process is assumed to be a convolution of an instrumental function and the intrinsic line broadening mechanism.

Specifically, let $X_1 \sim G(\mu, \sigma)$ and $X_2 \sim C(0, \gamma)$ be independent Gaussian and Cauchy distributed random variables, then a sum of X_1 and X_2 , $\tilde{X} = X_1 + X_2$ is Voigt distributed. It's density has the following form:

$$\mathcal{V}(\tilde{x}; l, \sigma, \gamma) = f_C \star f_G = \int_{-\infty}^{\infty} f_C(x; l, \gamma) f_G(\tilde{x} - x; 0, \sigma) dx, \quad \text{for } \tilde{x} \in \mathbb{R} \quad (2)$$

Where $f_C(x; l, \gamma)$, $f_G(\tilde{x} - x; 0, \sigma)$ are Cauchy and Gaussian densities, respectively. Since the Cauchy and Gaussian densities are normalized, the corresponding convolution (Voigt) is also normalized.

The integral (2) is a real part of the Faddeeva function [7] $\omega(z)$ evaluated at $z = \frac{\tilde{x} - l + i\gamma}{\sigma\sqrt{2}}$,

$$\mathcal{V}(\tilde{x}; l, \sigma, \gamma) = \frac{\text{Re}(\omega(z))}{2\pi}$$

where

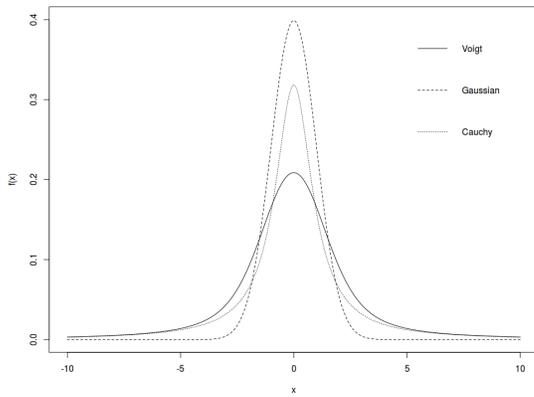
$$\omega(z) = e^{-z^2} \left(1 + \frac{2i}{\sqrt{\pi}} \int_0^z e^{t^2} dt \right) \quad (3)$$

The Faddeeva (3) and Voigt(2) functions do not have an analytical expression but can be numerically approximated, see [8], [9], [10], [11], [12] for discussion.

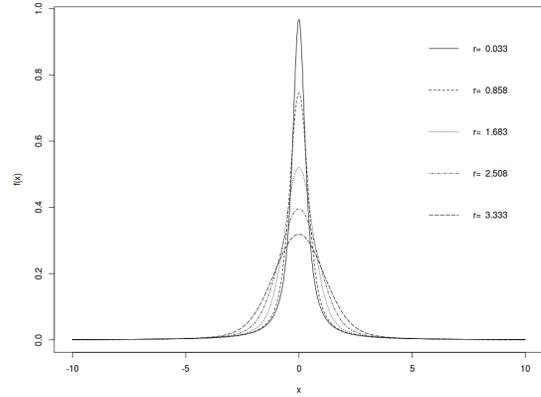
Similarly to the Cauchy, the Voigt distribution does not have finite moments. The parameter l can be interpreted as the location of the maximum of the density. In the physics literature, as a measure of the spread of the Voigt distribution is taken a half-width at half-maximum (HWHM).

The Voigt density as a convolution of $\mathcal{C}(0, 1)$ and $G(0, 1)$ is shown in figure 1a and it can be seen that it is lower than Gaussian and Cauchy near the center of the distribution and larger than Cauchy and Gaussian in its tails. Far in the tails the difference between Cauchy and Voigt is negligible. The reason for such behavior is that the density of the Gaussian distribution decreases exponentially away from the center, so the contribution of the Gaussian density becomes negligible in comparison to the tails of Cauchy.

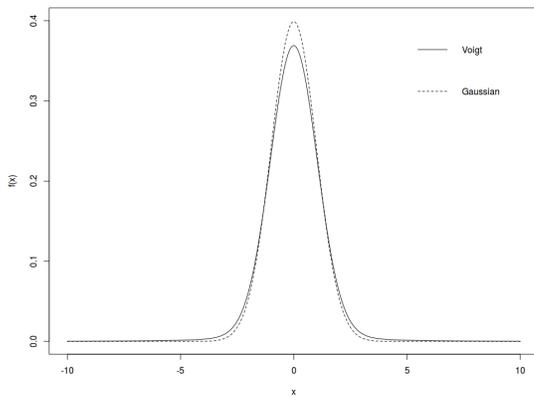
The possible advantage of using Voigt distribution can be understood from Figures 1c, 1d, 1b. What determines the shape of the Voigt distribution is the ratio between the Gaussian standard deviation σ and the Cauchy scale γ , that is $r = \frac{\sigma}{\gamma}$ [6]. It can be seen that for $r \gg 1$ the Voigt distribution is approaching Gaussian



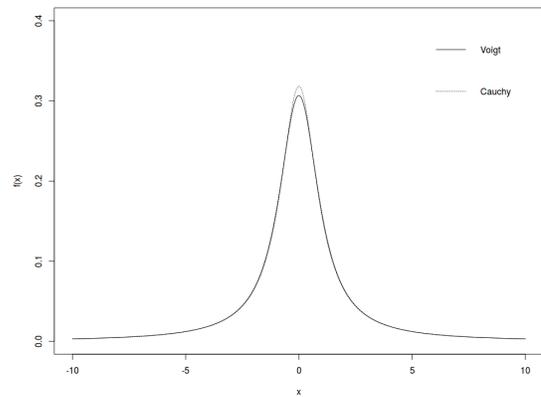
(a) $\mathcal{V}(0, 1, 1)$ and it's components $\mathcal{C}(0, 1)$ and $G(0, 1)$



(b) Shape of the Voigt distribution for various values of r .



(c) Density of $\mathcal{V}(0, 1, 0.1)$ and it's component and $G(0, 1)$



(d) Density of $\mathcal{V}(0, 0.2, 1.0)$ and it's component $\mathcal{C}(0, 1)$

Figure 1: Shapes of the Voigt density.

and for values of $r \ll 1$ the Voigt distribution is approaching Cauchy. Such property of the Voigt density makes it more flexible than Cauchy and Gaussian in modeling distribution of disturbances.

2.2 Estimation of the Parameters of the MAR Processes

In this work parameter estimation was performed within Bayesian framework. An alternative method for parameter estimation of the MAR processes using maximum likelihood can be found in [13].

Bayesian formalism requires specification of an observational equation which is called "likelihood" if viewed as a function of the parameters and prior distributions which express our uncertainty about the parameters of interest before the data is observed.

Combining the distribution of errors and the definition of the MAR process (1), the joint likelihood of the mixed causal-noncausal processes has the following form:

$$p(\mathbf{y}, |\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\eta}) = \prod_{t=1}^T f_{\epsilon_t}(\Psi(L^{-1})\Phi(L)y_t|\boldsymbol{\eta}) \quad (4)$$

where $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_s)^T$ and $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_r)^T$ are vectors of noncausal and causal autoregressive coefficients respectively. The distribution of disturbances is parameterized by $\boldsymbol{\eta}$. Due to the dependence of the present observations on the past and future observations in MAR(s, r), the error terms ϵ_t can not be computed for the first r and last s observations, unless the marginal distribution is available at these time points. For a relatively long time series this problem can be solved by approximating the likelihood (4) by conditioning on initial r and last s observations leading to the so called conditional likelihood:

$$p(\mathbf{y}|\tilde{\mathbf{y}}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\eta}) = \prod_{t=r+1}^{T-s} f_{\epsilon_t}(\Psi(L^{-1})\Phi(L)y_t|\boldsymbol{\eta}) \quad (5)$$

where, $\tilde{\mathbf{y}}$ is the vector of the first r and last s observations. In case of the Voigt distributed errors ϵ_t the conditional likelihood (5) becomes:

$$p(\mathbf{y}|\tilde{\mathbf{y}}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\eta}) = \prod_{t=r+1}^{T-s} \mathcal{V}_{\epsilon_t}(\Psi(L^{-1})\Phi(L)y_t; 0, \sigma, \gamma) \quad (6)$$

The unknown parameters of the MAR process are causal $\boldsymbol{\phi}$ and noncausal $\boldsymbol{\psi}$ coefficients and a vector of nuisance parameters $\boldsymbol{\eta} = (\sigma, \gamma)$ that parameterize the Voigt distribution. A prior specification for parameters of the MAR $p(\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\eta})$ completes the Bayesian specification and leads to the following posterior distribution:

$$p(\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\eta}|\{y_t\}_{t=r+1}^{T-s}) \propto p(\mathbf{y}|\tilde{\mathbf{y}}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\eta})p(\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\eta}) \quad (7)$$

The prior in (7) is assumed to be have an independent structure so it factorizes as $p(\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\eta}) = p(\boldsymbol{\psi})p(\boldsymbol{\phi})p(\boldsymbol{\eta})$.

As mentioned previously the Voigt distribution does not have an analytic expression, so conjugate priors are not available. A particular choice of priors is based on the range that the parameters can take as well as computational convenience. In this work the prior uncertainty in causal and noncausal parameters is reflected by the multivariate normal distributions, that is:

$$\boldsymbol{\psi} \sim \mathcal{N}(\mathbf{0}, \sigma_{\boldsymbol{\psi}}^2 \mathbf{I}) \quad (8)$$

$$\boldsymbol{\phi} \sim \mathcal{N}(\mathbf{0}, \sigma_{\boldsymbol{\phi}}^2 \mathbf{I}) \quad (9)$$

In order for the MAR process to be stationary, solutions of the causal and noncausal polynomials must lie outside of the unit circle. Thus the priors must be chosen in such a way that they put a high probability mass on a range of causal and noncausal parameters that lead to stationary MAR processes. In the case of MAR(1, 1) a reasonable prior for both causal and noncausal coefficients would be normal with mean $\mu = 0$ and variance $\sigma^2 = 0.25$ that puts a 95% of the probability mass on the range of values between -0.980 and 0.980.

Since σ and γ must be strictly positive that is $\boldsymbol{\eta} = (\sigma, \gamma)^T \in \mathbb{R}_{>0}^2$, reasonable priors are Gamma distributions with shapes ν_1, α_1 and rates ν_2, α_2 :

$$\sigma \sim \mathcal{G}(\nu_1, \nu_2) \quad (10)$$

$$\gamma \sim \mathcal{G}(\alpha_1, \alpha_2) \quad (11)$$

Particular values of the hyperprior parameters ν_1, α_1 and ν_2, α_2 are chosen in such a way that the prior puts a high probability mass on a reasonable range of values. For example, if we believe that the reasonable value for the Gaussian standard deviation is 1 with an uncertainty in this value of about 0.25, then by choosing $\nu_1 = 16$ and $\nu_2 = 16$ a 95% of the probability mass would be concentrated on the range of values between 0.5716 and 1.5463.

Combining the priors (8) - (11) and the likelihood (6) we are interested in estimating the following posterior distribution:

$$p(\boldsymbol{\psi}, \boldsymbol{\phi}, \sigma, \gamma | \{y_t\}_{r+1}^{T-s}) \propto \prod_{t=r+1}^{T-s} \mathcal{V}_{\epsilon_t}(\epsilon_t | 0, \sigma, \gamma, \tilde{\mathbf{y}}) \mathcal{N}(\boldsymbol{\psi} | \cdot) \mathcal{N}(\boldsymbol{\phi} | \cdot) \mathcal{G}(\sigma | \cdot) \mathcal{G}(\gamma | \cdot) \quad (12)$$

Since the Voigt density does not have a closed form, it is not possible to get an analytic expression for the posterior (12), although it can be approximated by Monte Carlo methods. In particular, a hybrid of Metropolis, Metropolis-Hastings and blocked-Gibbs algorithms can be used to sample from the posterior distribution (12). The dimensionality of the joint posterior distribution (12) can be broken by considering the following full conditional distributions:

$$p(\boldsymbol{\psi}|\cdot) \propto \prod_{t=r+1}^{T-s} \mathcal{V}_{\epsilon_t}(\epsilon_t|0, \sigma, \gamma, \tilde{\mathbf{y}}) \mathcal{N}(\boldsymbol{\psi}|\mathbf{0}, \boldsymbol{\Lambda}_{\psi}) \quad (13)$$

$$p(\boldsymbol{\phi}|\cdot) \propto \prod_{t=r+1}^{T-s} \mathcal{V}_{\epsilon_t}(\epsilon_t|0, \sigma, \gamma, \tilde{\mathbf{y}}) \mathcal{N}(\boldsymbol{\phi}|\mathbf{0}, \boldsymbol{\Lambda}_{\phi}) \quad (14)$$

$$p(\sigma|\cdot) \propto \prod_{t=r+1}^{T-s} \mathcal{V}_{\epsilon_t}(\epsilon_t|0, \sigma, \gamma, \tilde{\mathbf{y}}) \mathcal{G}(\sigma|\nu_1, \nu_2) \quad (15)$$

$$p(\gamma|\cdot) \propto \prod_{t=r+1}^{T-s} \mathcal{V}_{\epsilon_t}(\epsilon_t|0, \sigma, \gamma, \tilde{\mathbf{y}}) \mathcal{G}(\gamma|\alpha_1, \alpha_2) \quad (16)$$

Sampling from the full conditionals (13) - (16) follows the structure of a blocked Gibbs algorithm. Unfortunately none of the full conditional distributions (13) - (16) have a closed form, thus a Metropolis algorithm with normal proposals can be used to sample from (13), (14) and Metropolis-Hastings to sample from (15), (16). Reasonable proposal distributions for σ and γ would be Gamma.

The MAR processes have two free parameters, noncausal and causal orders s and r that require selection. In this work, the selection of s and r is based on minimizing the Deviance Information Criterion (DIC). The DIC tries to estimate the out-of-sample predictive performance of the model. It can only provide an insight about the model performance relative to another model. There are several definitions of the DIC, the one that is used in this work can be found in [14] and is defined as:

$$DIC = \mathbb{E}_{\boldsymbol{\eta}}(D(\mathbf{y}|\boldsymbol{\eta})) + 2p_d$$

$$p_d = \frac{1}{2} \widehat{Var}(D(\mathbf{y}|\boldsymbol{\eta})) \quad (17)$$

where $D(\mathbf{y}) = -2 \log(p(\mathbf{y}|\boldsymbol{\eta}))$ is deviance and p_d can be interpreted as the effective number of parameters, or a measure of complexity. The advantage of defining the measure of complexity p_d as in (17) is discussed in [15].

Given the number of posterior draws B of all the unknown parameters $\{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \dots, \boldsymbol{\theta}^B\}$, it is straightforward to estimate the DIC,

$$\bar{D}(\mathbf{y}) = \frac{1}{B} \sum_{l=1}^B D(\mathbf{y}, \boldsymbol{\theta}^l) \quad (18)$$

$$DIC = \bar{D}(\mathbf{y}) + \frac{1}{(B-1)} \sum_{b=1}^B (\bar{D}(\mathbf{y}) - D(\mathbf{y}, \boldsymbol{\theta}^b))^2 \quad (19)$$

In case of the conditional Voigt likelihood (6) the deviance becomes:

$$D(\mathbf{y}_{r+1}^{T-s}, \boldsymbol{\theta}^b) = -2 \sum_{t=r+1}^{T-s} \log(\mathcal{V}_{\epsilon_t}(\Psi^b(L^{-1})\Phi^b(L)y_t; 0, \sigma^b, \gamma^b)) \quad (20)$$

Finally, the DIC can be calculated by plugging eq. (20) into eq. (18) and (19).

2.3 Simulation of the MAR Processes

The simulation of the mixed autoregressive processes is based on partial fraction decomposition [2]. The MAR can be decomposed into two unobserved processes. Let $u_t = \Phi(L)y_t$ and $v_t = \Psi(L^{-1})y_t$, the equation (1) can be rewritten as:

$$\begin{aligned} \Psi(L^{-1})u_t &= \epsilon_t \\ \Phi(L^1)v_t &= \epsilon_t \end{aligned} \quad (21)$$

The process u_t is y_t -causal and ϵ_t -noncausal, the process v_t is y_t -noncausal and ϵ_t -causal. Mixed AR process y_t can be simulated by recovering from the unobserved components u_t and v_t . Using partial fraction decomposition and the unobserved processes $\{u_t\}$ and $\{v_t\}$, the mixed AR processes has two equivalent representations which are useful for simulations:

1. Causal representation

$$y_t = L^s g_1(L)v_t + g_2(L)u_t$$

2. Noncausal representation

$$y_t = f_1(L^{-1})v_t + L^{-r} f_2(L^{-1})u_t$$

For example, using the causal representation of the mixed AR processes, the simulation of the MARV(1, 1) can be performed in the following steps [2]:

1. For $t = t_0, \dots, T$, where T is sufficiently large, sample:

$$\epsilon_t \sim \mathcal{V}(0, \sigma, \gamma)$$

2. For $t = T - 1, T - 2, \dots, t_0$ compute:

$$u_t = \psi u_{t+1} + \epsilon_t$$

3. For $t = t_0 + 1, \dots, T$ compute:

$$v_t = \phi v_{t-1} + \epsilon_t$$

4. For $t = t_0 + 1, \dots, T - 1$, obtain y_t by computing:

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

2.4 Simulation Study

The simulation study is first performed on the data that was simulated from the MAR(1, 1) process with $\mathcal{V}(0, 1, 1)$ errors, to assess the behavior of the MARV models under the model misspecification and to check if the DIC will be able to select the true model. The shape of the $\mathcal{V}(0, 1, 1)$ can be seen from the figure 1a and is a convolution of the standard Gaussian and standard Cauchy densities. More specifically, for this simulation study the true data generating process is the following:

$$(1 - 0.85L^{-1})(1 - 0.25L)y_t = \epsilon, \quad \epsilon \sim \mathcal{V}(0, 1, 1) \quad (22)$$

Assuming that the true parameters of the process (22) are unknown the models that we choose to fit are:

1. MARV(0, 1) which is a regular causal AR with Voigt distributed errors,

$$(1 - \phi L)y_t = \epsilon \quad (23)$$

2. A purely noncausal AR process, MARV(1, 0)

$$(1 - \psi L^{-1})y_t = \epsilon \quad (24)$$

3. Mixed causal-noncausal AR processes, MARV(1, 1)

$$(1 - \phi L)(1 - \psi L^{-1})y_t = \epsilon \quad (25)$$

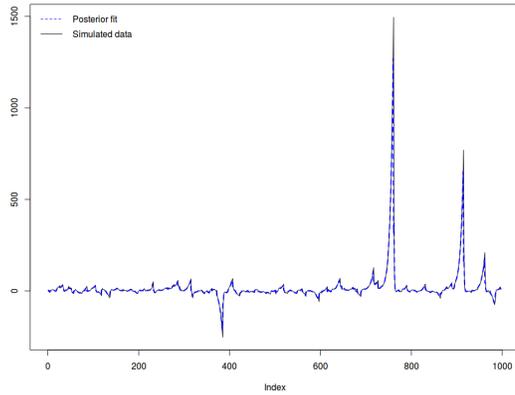
4. A higher, or second order mixed causal-noncausal AR process, MARV(1, 2)

$$(1 - \psi L^{-1})(1 - \phi_1 L - \phi_2 L^2)y_t = \epsilon \quad (26)$$

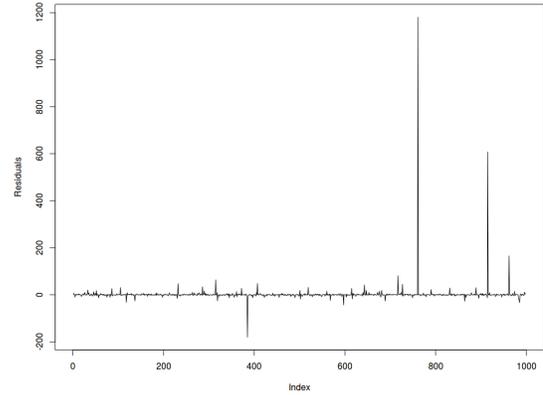
After running an MCMC algorithm, the estimated parameters of the models (23) - (26) along with 95% quantile based coverage intervals are shown in table 1. The smallest DIC corresponds to an over parameterized model MARV(1, 2), but the difference with the DIC of the MARV(1, 1) is very small. Also, one of the estimated causal parameter ϕ_2 of the MARV(1, 2) model is not significantly different from 0. Based on these observations the best model would be MARV(1, 1) which was the true data generating process. Also, notice that all estimated coverage intervals include the true values of the parameters. The simulated data along with the best fit MARV(1, 1) and the estimated residuals are shown in figure 2.

Parameter	ψ_1	ψ_2	ϕ_1	ϕ_2	γ	σ	$\ln \mathcal{L}$	DIC
MARV(0,1)	-	-	1.177 (1.175 1.179)	-	1.433 (1.232 1.639)	1.352 (1.069 1.625)	-3035.629 (-3038.855 -3034.258)	6074.325
MARV(1,0)	0.849 (0.848 0.851)	-	-	-	1.221 (1.054 1.396)	1.171 (0.933 1.412)	-2872.608 (-2875.437 -2871.327)	5747.732
MARV(1,1)	0.849 (0.848 0.851)	-	0.249 (0.246 0.252)	-	0.958 (0.812 1.118)	1.131 (0.929 1.326)	-2673.323 (-2677.059 -2671.475)	5350.978
MARV(1, 2)	0.849 (0.847 0.851)	-	0.001 (-0.002 0.005)	0.249 (0.246 0.252)	0.959 (0.811 1.107)	1.133 (0.925 1.329)	-2671.654 (-2675.884 -2669.389)	5349.427

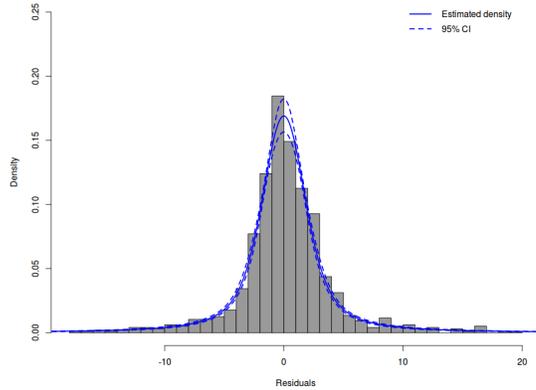
Table 1: Posterior means and 95% quantile based coverage intervals for the parameters of various MAR models with Voigt errors fitted to the simulated data.



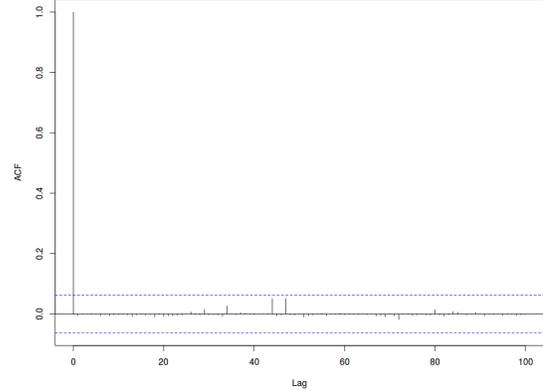
(a) Simulated MARV(1,1) with parameters $\psi = 0.85, \phi = 0.25, \sigma = 1, \gamma = 1$ and the best model fit.



(b) Time series of the residuals of the MARV(1, 1) fit.



(c) A histogram of the residuals and an estimated Voigt density with 95% quantile based coverage interval.



(d) An autocorrelation function of the residuals.

Figure 2: MARV(1, 1) fit and the residuals.

Recall from section 2.1 that the shape of the Voigt density depends on the ratio between the Gaussian standard deviation and Cauchy scale. In the next simulation study we are trying to check how well would the Cauchy density approximate the Voigt in two extreme cases when the Gaussian standard deviation is much greater than the Cauchy scale $r \gg 1$ and vice-versa $r \ll 1$. For the first scenario, the data was generated from the MARV(1, 1) process with the errors distributed as $\mathcal{V}(0, 2, 0.1)$. For the second scenario the data was generated with the errors distributed

according to $\mathcal{V}(0, 0.2, 2)$. The values of the causal and noncausal coefficients in both cases are $\psi = 0.856$, $\phi = 0.732$. The datasets were fitted with the MARV(1, 1) and MARC(1, 1) models. The summary of the posterior distributions of the parameters and the DIC, are shown in tables 2 - 3. The visual summaries of the MARV and MARC fits are presented in figure 3.

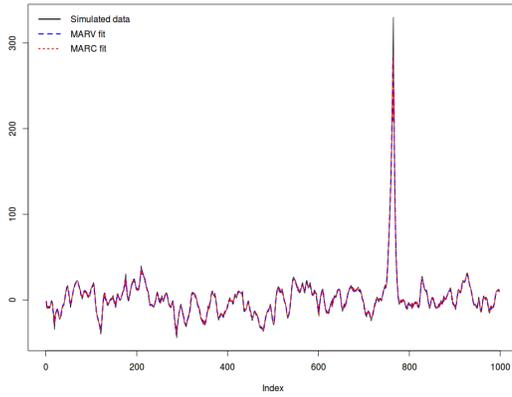
From the estimates of the DIC criterion, the MARV fits the data better when the simulated process has the Gaussian standard deviation much greater than the Cauchy scale. On the other hand, the MARC model fits the data equally well when the Cauchy scale is much greater than the Gaussian standard deviation in the simulated process. The reason for such performance can be understood by visually inspecting the “true” distribution of errors and the corresponding estimated densities presented in figures 3c - 3d. Regardless of the distribution of errors, the causal and noncausal parameters estimates under the MARC and the MARV models are very similar. In the first case the DIC selects the true MARV(1, 1) model, as expected. In the second case the DIC selects the MARC(1, 1) model, since the Cauchy density has fewer parameters and acts as a very good approximation to the Voigt density for $r \ll 1$.

Parameter	ψ_1	ϕ_1	γ	σ	$\ln \mathcal{L}$	DIC
MARV(1,1)	0.854 (0.839 0.869)	0.739 (0.720 0.758)	0.133 (0.072 0.211)	1.985 (1.869 2.105)	-2233.475 (-2237.111 -2231.641)	4471.418
MARC(1,1)	0.861 (0.844 0.880)	0.745 (0.729 0.763)	1.287 (1.186 1.392)	- -	-2366.979 (-2370.284 -2365.592)	4737.151

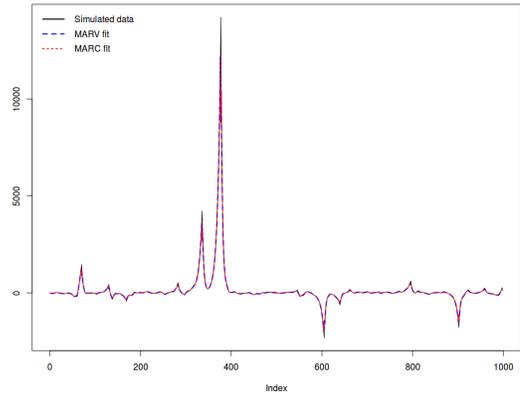
Table 2: Posterior means and 95% quantile based coverage intervals for the parameters of the MAR(1, 1) models for the simulated data with the distribution of errors $\mathcal{V}(0, 2, 0.1)$.

Parameter	ψ_1	ϕ_1	γ	σ	$\ln \mathcal{L}$	DIC
MARV(1,1)	0.856	0.732	2.062	0.223	-3246.622	6496.88
	(0.856 0.857)	(0.731 0.732)	(1.879 2.254)	(0.073 0.438)	(-3250.283 -3245.078)	
MARC(1,1)	0.856	0.732	2.078	-	-3246.571	6496.57
	(0.856 0.857)	(0.731 0.732)	(1.899 2.275)	-	(-3249.834 -3245.038)	

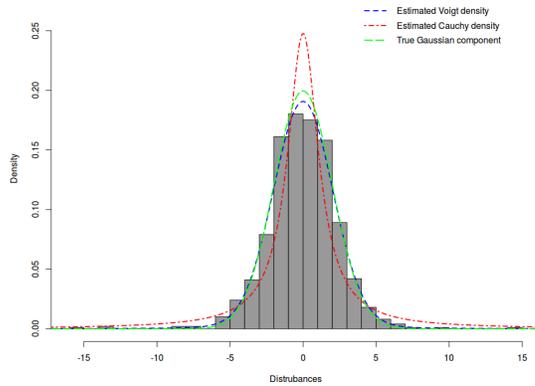
Table 3: Posterior means and 95% quantile based coverage intervals for the parameters of the MAR(1, 1) models for the simulated data with the distribution of errors $\mathcal{V}(0, 0.2, 2)$



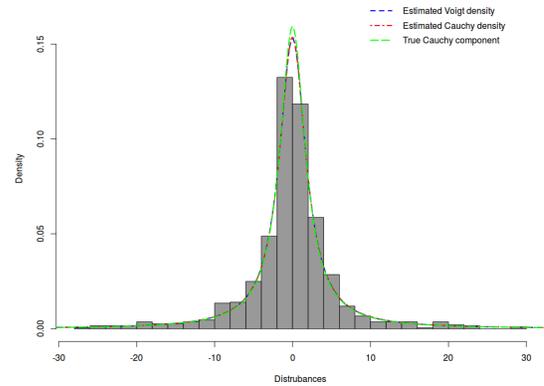
(a) Simulated MARV(1,1) with parameters $\psi = 0.856$, $\phi = 0.732$, $\sigma = 2$, $\gamma = 0.1$ and the MAR fits.



(b) Simulated MARV(1,1) with parameters $\psi = 0.856$, $\phi = 0.732$, $\sigma = 0.2$, $\gamma = 2$ and the MAR fits.



(c) A histogram of the true disturbances with estimated densities of the MARV and MARC models for the case when $r \gg 1$.



(d) A histogram of the true disturbances with estimated densities of the MARV and MARC models for the case when $r \ll 1$.

Figure 3: Simulation study of the MARV processes with the extreme ratios of r .

2.5 Analysis of the Bitcoin/USD Exchange Rate Data

In this section, the MARV and MARC models are applied to the Bitcoin/USD exchange rate, in order to compare their performance on a real data. The Bitcoin(BTC) is an electronic currency created by Satoshi Nakamoto and is one of the so called “cryptocurrencies”. These type of electronic/alternative currencies are named after the fact that the cryptography is used to secure the on line transactions and make them anonymous. The list of cryptocurrencies is quite exhaustive and includes: Auroracoin, Bitcoin, Burstcoin, Litcoin, etc. The bitcoins can be purchased with the US dollars or other currencies on virtual exchange markets. A very detailed overview of the bitcoins can be found in [1].

We analyzed the Bitcoin/USD exchange rate for a period over 02/20/2013 -07/20/2013 as in [1]. The data consists of 150 observations of the daily closing values and is shown in figure 4. ¹

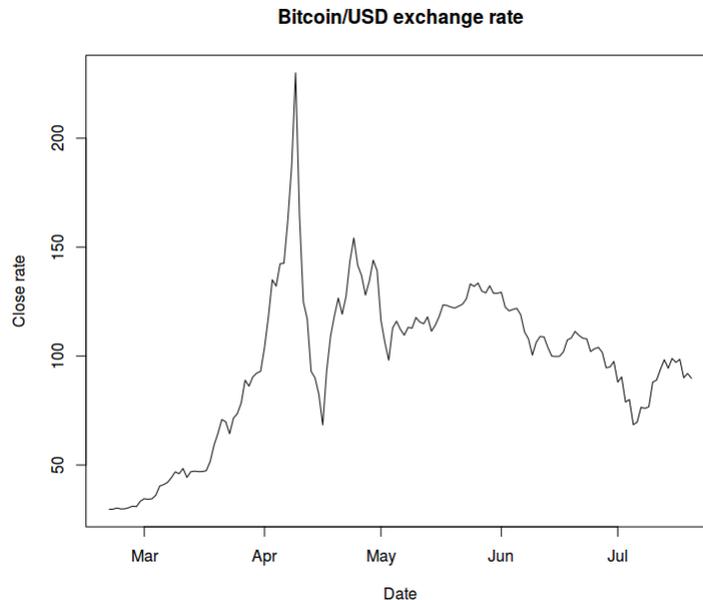


Figure 4: Bitcoin/USD exchange rate over a period of 02/20/2013 -07/20/2013

¹The data was downloaded from <https://www.quandl.com/>

It can be seen from figure 4 that the exchange rate series have a global trend with locally explosive behavior and a bubble that burst at the beginning of April 2013. Since the MAR is a stationary process, the global trend in the data needs to be subtracted before the MAR model can be used to fit the data. The global trend is modeled with a polynomial function of the time as in [1].

$$u_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + y_t \quad (27)$$

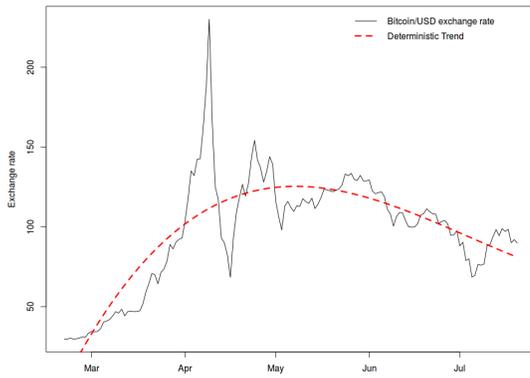
The residuals y_t of the model (27) are assumed to follow an MAR process of order (s, r) i. e. $\Psi(L^{-1})\Phi(L)y_t = \epsilon$ where $\epsilon \sim \mathcal{C}(0, \gamma)$ or $\epsilon \sim \mathcal{V}(0, \sigma, \gamma)$. Notice that the errors ϵ do not depend on time. The estimated posterior means and 95% quantile based coverage intervals are provided in table 4. Based on the DIC criterion, MAR models with Voigt distributions slightly outperform MAR models with the Cauchy distribution. The best MARV model, based on the DIC, is MARV(2, 2), with an addition of the estimated global trend, the full model becomes:

$$\begin{aligned} u_t &= -1.496 + 3.784t - 0.034t^2 + 0.0000852t^3 + y_t \\ (1 - 0.608L^{-1} - 0.049L^{-2})(1 - 0.109L - 0.469L^2)y_t &= \epsilon_t \\ \hat{f}_\epsilon(x) &= \mathcal{V}(x; 0, 0.918, 3.782) \end{aligned}$$

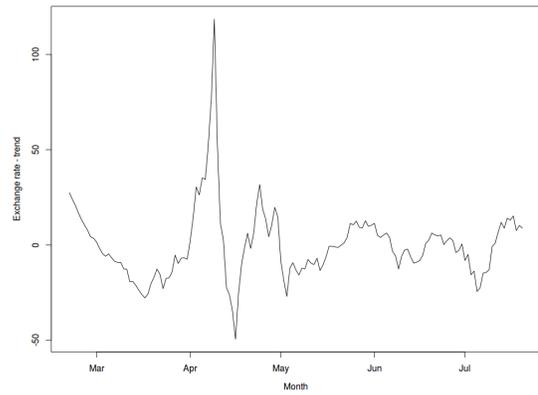
The MARV(2, 2) model was able to remove the autocorrelation in the data that can be seen by comparing the ACF of the detrended series, figure 5d with the ACF of the residuals from the best model fit shown in figure 7c.

Parameter	ψ_1	ψ_2	ϕ_1	ϕ_2	γ	σ	$\ln \mathcal{L}$	DIC
MARV(0,1)	-	-	0.904 (0.838 0.973)	-	1.789 (0.904 2.808)	2.783 (0.597 4.497)	-510.303 (-513.742 -508.696)	1024.149
MARC(0,1)	-	-	0.885 (0.831 0.941)	-	2.707 (2.189 3.296)	-	-511.37 (-514.889 -510.02)	1026.187
MARV(0,2)	-	-	-0.241 (-0.386 -0.108)	1.136 (0.989 1.289)	1.493 (0.732 2.547)	3.277 (1.219 4.732)	-500.883 (-504.709 -498.927)	1006.352
MARC(0,2)	-	-	-0.186 (-0.287 -0.087)	1.068 (0.952 1.18)	2.668 (2.153 3.274)	-	-503.266 (-507.187 -501.436)	1011.291
MARV(1,0)	0.818 (0.748 0.895)	-	-	-	0.933 (0.421 1.635)	4.55 (3.563 5.521)	-504.047 (-507.157 -502.674)	1011.022
MARC(1,0)	0.901 (0.803 0.989)	-	-	-	2.993 (2.432 3.645)	-	-515.116 (-518.594 -513.626)	1034.038
MARV(1,1)	0.656 (0.581 0.732)	-	0.488 (0.375 0.581)	-	0.835 (0.307 1.589)	4.076 (2.937 5.113)	-480.555 (-484.231 -478.825)	965.2042
MARC(1,1)	0.645 (0.586 0.708)	-	0.531 (0.457 0.595)	-	2.593 (2.096 3.175)	-	-488.399 (-492.44 -486.44)	982.1
MARV(1,2)	0.625 (0.555 0.7)	-	0.165 (0.045 0.277)	0.462 (0.33 0.549)	0.941 (0.419 1.676)	3.657 (2.598 4.613)	-474.783 (-478.877 -472.552)	955.1045
MARC(1,2)	0.612 (0.547 0.664)	-	0.215 (0.089 0.3)	0.501 (0.415 0.561)	2.43 (1.938 3.006)	-	-480.929 (-485.328 -478.624)	968.0657
MARV(2,0)	1.029 (0.869 1.197)	-0.188 (-0.318 -0.054)	-	-	0.961 (0.43 1.703)	4.315 (3.214 5.28)	-497.218 (-500.922 -495.411)	998.7647
MARC(2,0)	1.218 (1.08 1.346)	-0.319 (-0.42 -0.2)	-	-	2.722 (2.196 3.321)	-	-503.417 (-507.526 -501.553)	1011.783
MARV(2,1)	0.597 (0.494 0.686)	0.111 (0.014 0.223)	0.482 (0.388 0.567)	-	0.807 (0.301 1.498)	4.076 (3.001 5.082)	-475.866 (-479.822 -473.769)	956.9749
MARC(2,1)	0.622 (0.532 0.686)	0.061 (-0.034 0.188)	0.527 (0.443 0.59)	-	2.606 (2.1 3.166)	-	-484.642 (-489.348 -482.038)	976.7998
MARV(2,2)	0.608 (0.513 0.694)	0.049 (-0.07 0.19)	0.109 (-0.077 0.26)	0.469 (0.343 0.557)	0.918 (0.383 1.647)	3.782 (2.673 4.822)	-472.626 (-476.779 -470.175)	951.4516
MARC(2,2)	0.606 (0.532 0.668)	-0.035 (-0.145 0.082)	0.203 (0.047 0.3)	0.499 (0.423 0.56)	2.441 (1.947 3.001)	-	-478.758 (-483.447 -476.134)	965.1865

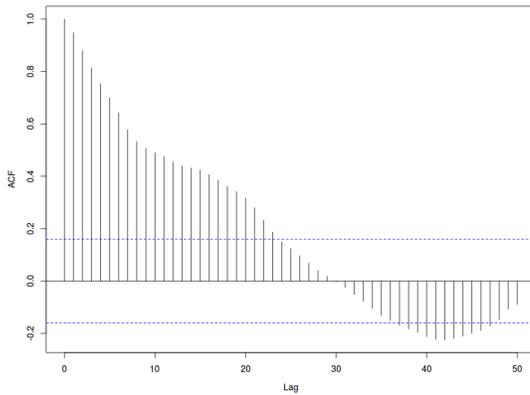
Table 4: Posterior means and 95% quantile based coverage intervals for the parameters of various MAR models with Cauchy and Voigt errors in application to the Bitcoin/USD exchange rate.



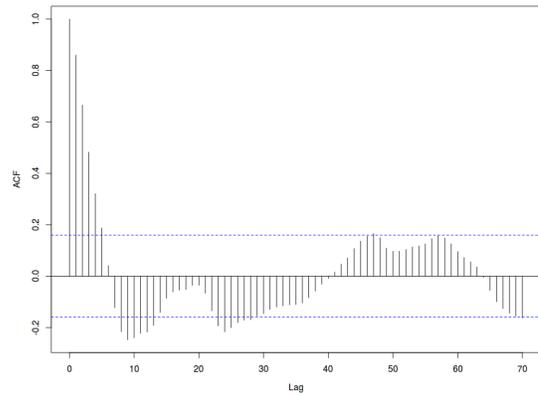
(a) Bitcoin/USD exchange rate over a period of 02/20/2013 - 07/20/2013 with an estimated global trend.



(b) Detrended Bitcoin/USD exchange rate series over a period of 02/20/2013 - 07/20/2013.



(c) The autocorrelation function of the “raw” data.



(d) The autocorrelation function of the detrended data.

Figure 5: Global trend and autocorrelation structure in the Bitcoin/USD exchange rate series.

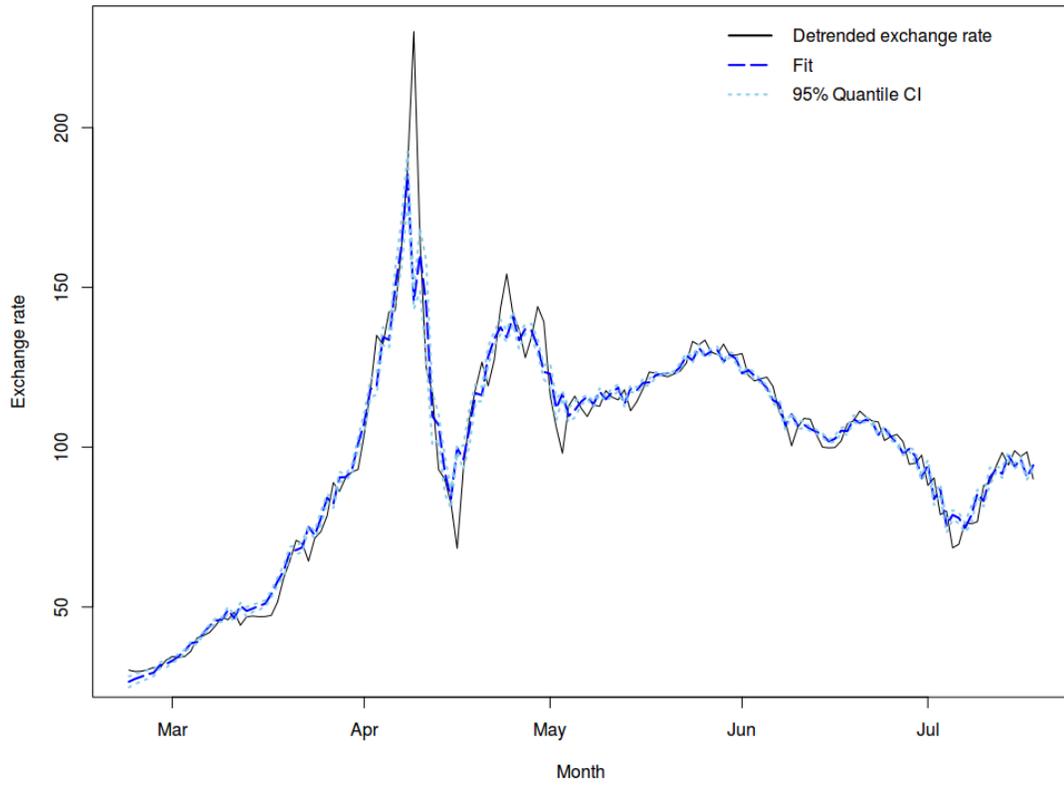
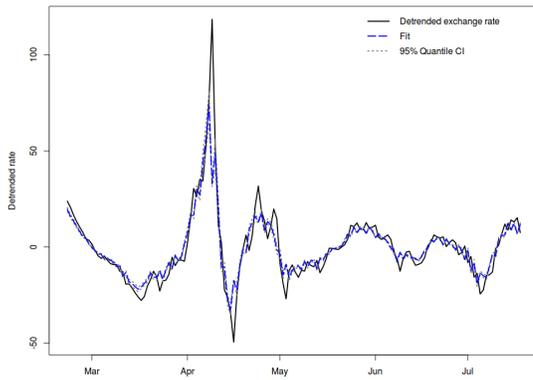
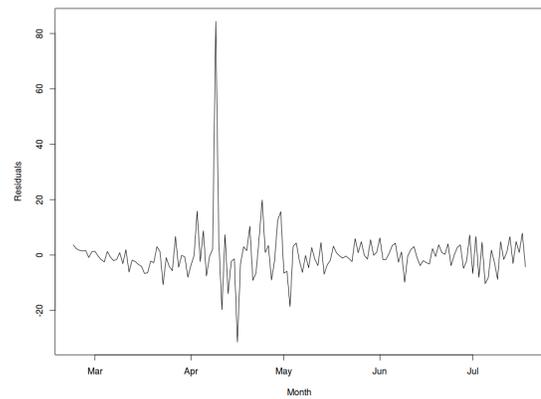


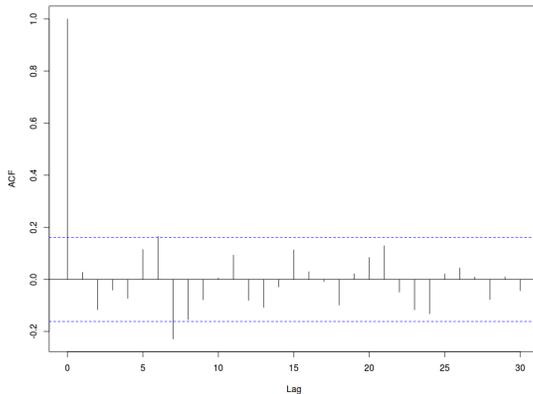
Figure 6: The MARV(2, 2) fit added to the global trend along with the Bit-coin/USD exchange rate series over a period of 02/20/2013 -07/20/2013.



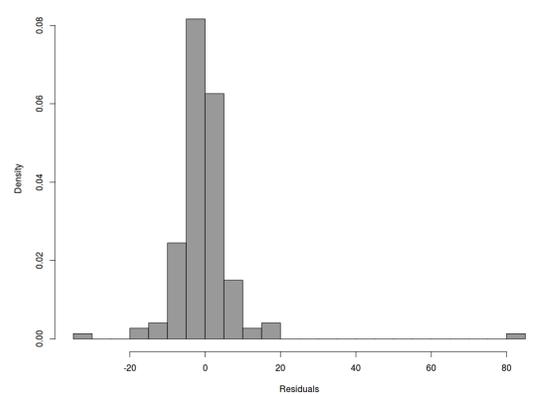
(a) Detrended Bitcoin/USD exchange rate over a period of 02/20/2013 -07/20/2013 and the best model fit.



(b) Time series of the residuals of the best model.



(c) Autocorrelation function of the residuals of the best model.



(d) A histogram of the residuals of the best model.

Figure 7: Bitcoin/USD exchange rate and the corresponding fit summaries.

For further comparison of MARC and MARV performance a much longer time series of the Bitcoin/USD exchange rate were considered, spanning a time interval from 2010-07-17 until 2014-02-25 that is shown in figure 8a.

The time series of the Bitcoin/USD exchange rate (figure 8a) exhibit a very complex dynamics. One of the features of this dataset is that there is quite long interval of a relatively linear constant trend going from 2010-07-17 until around 2013 with a small bubble around the middle of 2011. The data exhibit a series

of explosive trends starting in the middle of 2013. As in the previous analysis, in order to use the MAR models, the time series must be detrended. Since the dynamics is quite complicated, it is hard to assume any functional dependence on time as was done previously. Rather than assuming any particular form of the underlying deterministic trend, we used a nonparametric local constant least squares estimator (LCLS), also known as Nadaraya-Watson smoother. The only parameter that controls the smoothness of the LCLS estimator is a bandwidth. For detailed overview of the properties of the LCLS refer to [16], [17]. By incorporating the nonparametric trend $m(t)$ with an MAR stochastic component y_t a full specification of the model for the exchange rate u_t becomes:

$$u_t = m(t) + y_t$$

$$m(t) = \frac{\sum_{i=1}^n K_h(t_i, t) u_{t_i}}{\sum_{i=1}^n K_h(t_i, t)}$$

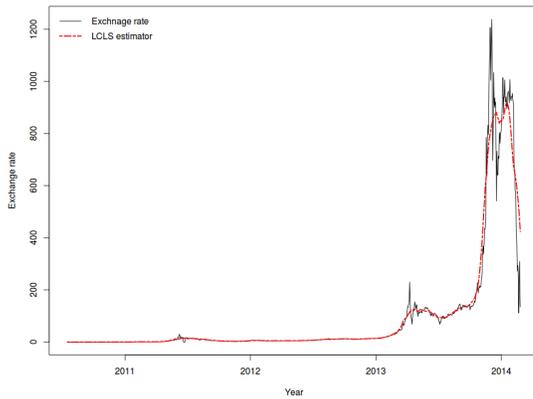
where h is a bandwidth and $K_h(\cdot)$ is a kernel (Gaussian in our case). The non-parametric trend is not a part of the MAR model, so the bandwidth and the coefficients of the MAR were not estimated simultaneously. In this settings, the LCLS smoother acts rather as a data preprocessing tool. The bandwidth h was chosen 40.5 to smooth the data just enough to account for the global trend without smoothing the bubble dynamics. The residuals after subtracting the trend, are assumed to follow an MAR process.

There were several models considered for fitting the exchange rate series. Table 5 contains posterior means and 95% quantile based coverage intervals for the parameters of the models under consideration. By comparing the DIC of the estimated models, overall MARC models slightly outperformed MARV models. Based

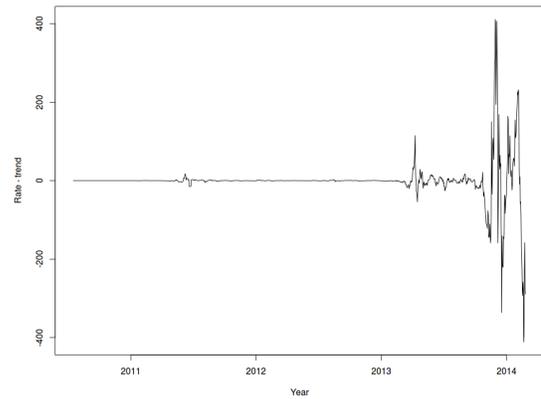
on the DIC the best performing model is purely noncausal MARC(1, 0), that is:

$$(1 - 0.881L^{-1})y_t = \epsilon_t, \quad \epsilon \sim \mathcal{C}(0, 0.164)$$

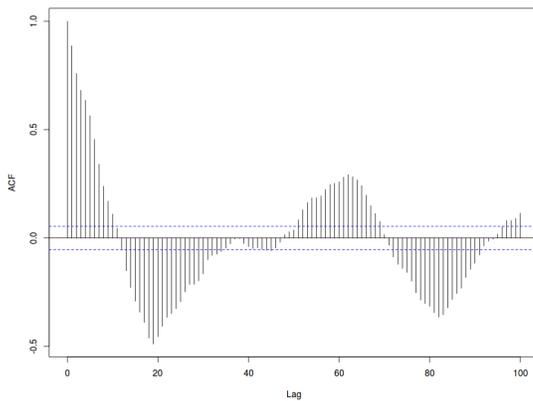
The best MARC(1, 0) fit and the residuals are shown in figure 9. Although, visually, the fit looks good, the model was not able to completely remove the autocorrelation structure in the series, which can be seen from the plot of the autocorrelation function of the residuals (figure 9e). Based on the pattern that the autocorrelation function exhibits, it is possible that there is some seasonality in the data, which is not modeled by the MARC(1, 0) model.



(a) Bitcoin/USD exchange rate over a period of 2010-07-17 - 2014-02-25

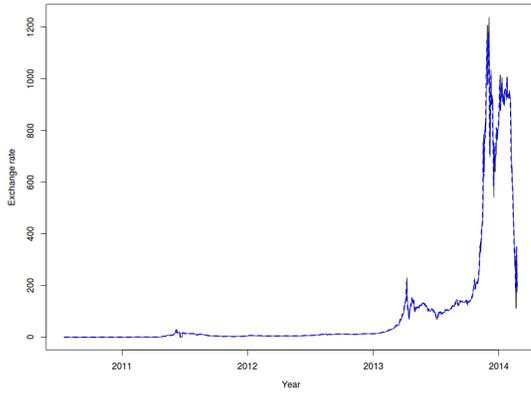


(b) Detrended bitcoin/USD exchange rate over a period of 2010-07-17 - 2014-02-25

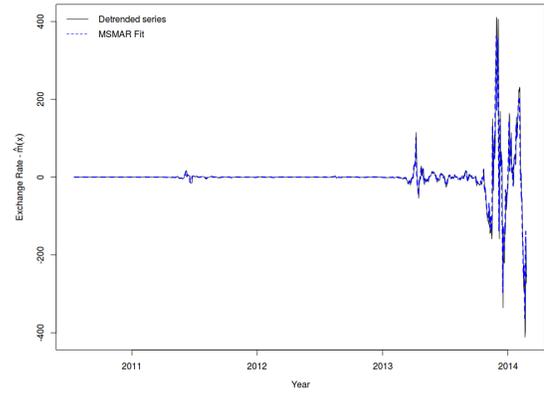


(c) An autocorrelation function of the Bitcoin/USD exchange rate series.

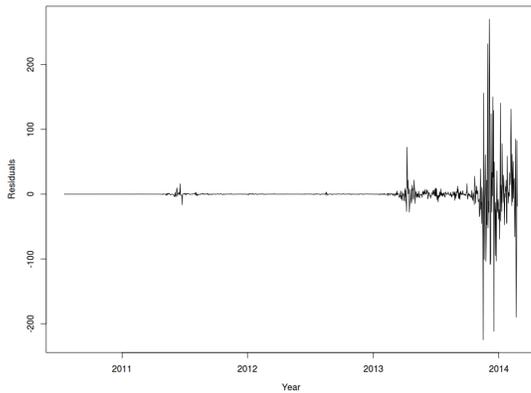
Figure 8: Visual summary of the Bitcoin/USD exchange rate series over a period of 2010-07-17 - 2014-02-25



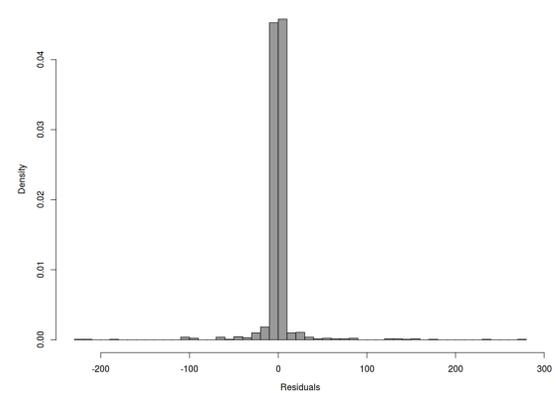
(a) The Bitcoin/USD exchange rate with the MARC(1, 0) fit added to the nonparametric trend.



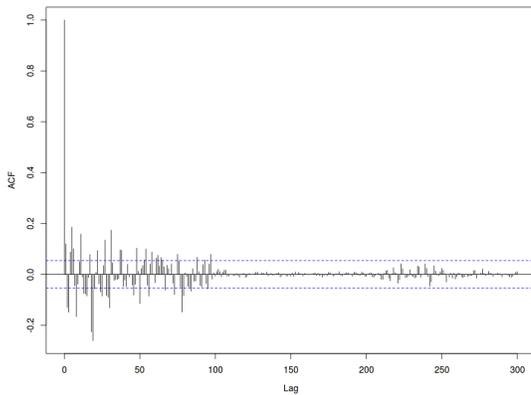
(b) The MARC(1, 0) fit and the detrended Bitcoin/USD exchange rate.



(c) Residuals of the MARC(1, 0) fit.



(d) A histogram of the residuals of the MARC(1, 0) fit.



(e) An autocorrelation function of the residuals of the MARC(1, 0) fit.

Figure 9: Visual summary of the best MARC fit in application to the Bitcoin/USD exchange rate series over a period of 2010-07-17 - 2014-02-25

Parameter	ψ_1	ψ_2	ϕ_1	ϕ_2	γ	σ	$\ln \mathcal{L}$	DIC
MARV(0,1)	-	-	0.952 (0.95 0.957)	-	0.161 (0.146 0.178)	0.013 (0.005 0.024)	-2412.704 (-2417.493 -2410.114)	4832.982
MARC(0,1)	-	-	0.952 (0.95 0.957)	-	0.16 (0.145 0.177)	-	-2410.681 (-2414.013 -2409.599)	4824.332
MARV(1,0)	0.881 (0.879 0.883)	-	-	-	0.165 (0.149 0.182)	0.013 (0.005 0.024)	-2404.954 (-2409.678 -2402.319)	4817.252
MARC(1,0)	0.881 (0.877 0.883)	-	-	-	0.164 (0.148 0.181)	-	-2402.963 (-2407.103 -2401.66)	4809.988
MARV(1,1)	0.754 (0.751 0.757)	-	0.329 (0.324 0.334)	-	0.184 (0.167 0.203)	0.015 (0.006 0.028)	-2449.582 (-2454.675 -2446.603)	4907.961
MARC(1,1)	0.754 (0.751 0.757)	-	0.329 (0.324 0.334)	-	0.184 (0.167 0.202)	-	-2447.53 (-2451.362 -2445.731)	4899.446

Table 5: Posterior means and 95% quantile based coverage intervals for the parameters of various MAR models with Cauchy and Voigt errors.

List of References

- [1] A. Hencic and C. Gouriéroux, “Noncausal autoregressive model in application to bitcoin/usd exchange rates,” in *Econometrics of Risk*. Springer, 2015, pp. 17–40.
- [2] C. Gouriéroux and J. Jasiak, “Filtering, prediction and simulation methods for noncausal processes,” *Journal of Time Series Analysis*, 2015.
- [3] F. J. Breidt and R. A. Davis, “Time-reversibility, identifiability and independence of innovations for stationary time series,” *Journal of Time Series Analysis*, vol. 13, no. 5, pp. 377–390, 1992.
- [4] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [5] A. Hecq, L. Lieb, and S. Telg, “Identification of mixed causal-noncausal models in finite samples,” *Annals of Economics and Statistics/Annales d’Économie et de Statistique*, no. 123/124, pp. 307–331, 2016.
- [6] F. Schreier, “Comments on a common misunderstanding about the voigt line profile,” *Journal of the Atmospheric Sciences*, vol. 66, no. 6, pp. 1860–1864, 2009.
- [7] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Corporation, 1964, vol. 55.
- [8] G. Poppe and C. Wijers, “More efficient computation of the complex error function,” *ACM Transactions on Mathematical Software (TOMS)*, vol. 16, no. 1, pp. 38–46, 1990.
- [9] K. L. Letchworth and D. C. Benner, “Rapid and accurate calculation of the voigt function,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 107, no. 1, pp. 173–192, 2007.

- [10] F. Lether and P. Wenston, “The numerical computation of the voigt function by a corrected midpoint quadrature rule for $(-,)$,” *Journal of computational and applied mathematics*, vol. 34, no. 1, pp. 75–92, 1991.
- [11] M. Kuntz, “A new implementation of the humlicek algorithm for the calculation of the voigt profile function,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 57, no. 6, pp. 819–824, 1997.
- [12] F. Schreier, “Optimized implementations of rational approximations for the voigt and complex error function,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 112, no. 6, pp. 1010–1025, 2011.
- [13] R. A. Davis, K.-S. Lii, and D. N. Politis, “Maximum likelihood estimation for noncausal autoregressive processes,” in *Selected Works of Murray Rosenblatt*. Springer, 2011, pp. 396–419.
- [14] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, “Bayesian data analysis. texts in statistical science series,” 2004.
- [15] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Linde, “The deviance information criterion: 12 years on,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 76, no. 3, pp. 485–493, 2014.
- [16] D. J. Henderson and C. F. Parmeter, *Applied nonparametric econometrics*. Cambridge University Press, 2015.
- [17] W. Härdle, M. Müller, S. Sperlich, and A. Werwatz, *Nonparametric and semi-parametric models*. Springer Science & Business Media, 2012.

CHAPTER 3

Markov Switching Mixed Causal-Noncausal Autoregressive Processes (MSMAR)

3.1 The MSMAR processes

Introduced in this chapter, the MSMAR process is a more general case of the MAR. The mixed causal-noncausal processes allow for an explicit dependence of the present values on the past as well as the future. A MAR of order (s, r) is defined as:

$$\Psi(L^{-1})\Phi(L)y_t = \epsilon_t, \quad \epsilon_t \stackrel{\text{i.i.d.}}{\sim} f_\epsilon(\cdot|\boldsymbol{\eta}) \quad (28)$$

To account for different regimes, in the correlation structure, at different times, the MAR process can be extended to MSMAR process by introducing a hidden regime (state) indicator. More specifically, let $S_t \in \{1, 2, \dots, K\}$ be a latent state indicator with a transition matrix $\boldsymbol{\xi}$, then the MSMAR of order $(\mathbf{s}, \mathbf{r}, K)$ is a stochastic process $\{y_t : t \in \mathbb{Z}\}$ defined as:

$$\Psi_{S_t}(L^{-1})\Phi_{S_t}(L)y_t = \epsilon_t, \quad \epsilon_t \sim f_{\epsilon_t}(\cdot|\boldsymbol{\eta}_{S_t}) \quad (29)$$

The transition matrix of the latent state variable S_t has the following form:

$$\boldsymbol{\xi} = \begin{pmatrix} \xi_{1,1} & \xi_{1,2} & \cdots & \xi_{1,K} \\ \xi_{2,1} & \xi_{2,2} & \cdots & \xi_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{K,1} & \xi_{K,2} & \cdots & \xi_{K,K} \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{\xi}_1. \\ \boldsymbol{\xi}_2. \\ \vdots \\ \boldsymbol{\xi}_K. \end{pmatrix} \quad (30)$$

where $P(S_t = l | S_{t-1} = m) = \xi_{ml}$ for $l = 1, \dots, K$; $m = 1, \dots, K$. Every row of the transition matrix must satisfy $\sum_{l=1}^K \xi_{jl} = 1$ and $\xi_{jl} \geq 0$ for all $j, l = 1, 2, \dots, K$, so it is a probability mass function. It is also assumed that the rows are independent of one another.

As in the case of the MAR processes, the polynomials $\Psi_{S_t}(L^{-1}) = 1 - \sum_{i=1}^{s_{S_t}} \psi_i^{(S_t)} L^{-i}$, $\Phi_{S_t}(L) = 1 - \sum_{i=1}^{r_{S_t}} \phi_i^{(S_t)} L^{-i}$ are causal and noncausal in the lag-

forward and the lag-backward operators respectively. The flexibility of the MSMAR processes is due to the fact that the model can incorporate regime specific orders of causal and noncausal dependence, as well as regime specific parameters of the distribution of errors. The MSMAR process can include a mixture of purely causal with purely noncausal and mixed AR processes at different regimes. In this work the switching is modeled in the correlation structure rather than in the level, although the level can be also incorporated in the model.

In order for the MSMAR process to be regime stationary, the roots of the causal and noncausal polynomials must lie outside of the unit circle.

3.2 Parameter Estimation

Parameter estimation is carried out within Bayesian framework. The joint sampling distribution of the observations and the latent state variables, as a function of the parameters, is called a “complete data likelihood” [1] and, in general, has the following form:

$$p(\mathbf{y}, \mathcal{S} | \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\gamma}) = p(\mathbf{y} | \mathcal{S}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\gamma}) p(\mathcal{S} | \boldsymbol{\xi}) \quad (31)$$

In case of the MSMAR with the Cauchy distributed errors $\{\epsilon_t\}$, the joint density of the time series given a vector of states becomes:

$$p(\mathbf{y} | \mathcal{S}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\gamma}) = \prod_{k=1}^K \prod_{t: S_t=k} \mathcal{C}(\Psi_k(L^{-1})\Phi_k(L)y_t | 0, \gamma_k) \quad (32)$$

Due to the recursive structure of the process, the errors $\{\epsilon_t\}$ can not be computed for the entire time series. Unless the marginal distribution of the observations is known at each time point, the complete data likelihood is not available for the entire vector of observations. A straightforward way to bypass this difficulty, is to approximate the likelihood (32) by conditioning on a vector of the first r_k and the last s_k observations i.e. $\tilde{\mathbf{y}}_k = \{y_1, y_2, \dots, y_{r_k}, y_{T-s_k}, y_{T-s_k+1}, \dots, y_T\}$. Notice that

the the vector $\tilde{\mathbf{y}}_k$ is regime specific. The approximation of the likelihood takes the following form:

$$p(\mathbf{y}|\tilde{\mathbf{y}}, \mathbf{S}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\gamma}) = \prod_{k=1}^K \prod_{\tilde{t}: S_{\tilde{t}}=k} \mathcal{C}(\Psi_k(L^{-1})\Phi_k(L)y_{\tilde{t}}|0, \gamma_k) \quad (33)$$

where $\tilde{t} = \{t : t \neq 1, 2, \dots, r_k, T - s_k, T - s_k + 1\}$

The second factor in eq. (31) is the sampling distribution of the hidden state vector \mathbf{S} . Given the transition matrix $\boldsymbol{\xi}$, $p(\mathbf{S}|\boldsymbol{\xi})$ can be explicitly written as:

$$p(\mathbf{S}|\boldsymbol{\xi}) = \prod_{t=1}^T P(S_t|S_{t-1}, \boldsymbol{\xi})P(S_0) = \prod_{j=1}^K \prod_{m=1}^K \xi_{jm}^{N_{jm}(\mathbf{S})} \quad (34)$$

where N_{jm} is an operator that counts the number of transitions from state j to state m .

The approximate likelihood of (33) and the distribution of the hidden state variables (34) give the following complete-data conditional likelihood:

$$p(\mathbf{y}, \mathbf{S}|\tilde{\mathbf{y}}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\xi}) = \prod_{k=1}^K \prod_{\tilde{t}: S_{\tilde{t}}=k} \mathcal{C}(\Psi_k(L^{-1})\Phi_k(L)y_{\tilde{t}}|0, \gamma_k) \prod_{j=1}^K \prod_{m=1}^K \xi_{jm}^{N_{jm}(\mathbf{S})} \quad (35)$$

A joint posterior distribution of all the parameters and hidden state indicators given the data is quite complicated to sample from directly. However, the problem can be simplified by first sampling the state-specific parameters, given the vector of state indicators, and then sample the state indicators by conditioning on the state-specific parameters.

Sampling the State-specific Parameters

Given a vector of state indicators \mathbf{S} , we are interested in estimating the following posterior distribution:

$$p(\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\xi}|\mathbf{y}, \mathbf{S}) \propto p(\mathbf{y}|\mathbf{S}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\gamma})p(\mathbf{S}|\boldsymbol{\xi})p(\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\xi}) \quad (36)$$

For computational convenience, a priori, the state-specific parameters $\{\boldsymbol{\psi}_k, \boldsymbol{\phi}_k, \gamma_k, \boldsymbol{\xi}_k\}$ are assumed to be independent within and between the states, so

the joint prior distribution factorizes as $p(\boldsymbol{\psi}, \boldsymbol{\phi}, \gamma, \boldsymbol{\xi}) = \prod_{k=1}^K p(\psi_k)p(\phi_k)p(\gamma_k)p(\boldsymbol{\xi}_k)$. The Cauchy likelihood is not from the exponential family of distributions, so conjugate priors are not available. The choice of the priors for the state-specific parameters $\{\boldsymbol{\psi}, \boldsymbol{\phi}, \gamma\}$ is dictated by the range that these parameters can take. The prior over the causal and noncausal coefficients must ensure stationarity of the MSMAR process, so a reasonable choice is a multivariate normal distribution, $\boldsymbol{\psi}_k \sim \mathcal{N}(\mathbf{0}, \tau_{\psi_k} \mathbf{I})$ and $\boldsymbol{\phi}_k \sim \mathcal{N}(\mathbf{0}, \tau_{\phi_k} \mathbf{I})$. Since the Cauchy scale must be a strictly positive value, the prior for γ_k is chosen to be $\gamma_k \sim \mathcal{G}(\nu_{k1}, \nu_{k2})$ with shape ν_{k1} and rate ν_{k2} .

Essentially, every row of the transition matrix is a probability mass function. Since the rows are mutually independent, a prior over the probability mass functions, that in conjunction with the complete data likelihood, will provide a closed form conditional posterior is $\boldsymbol{\xi}_k. \sim \mathcal{D}(a_{j1}, a_{j2}, \dots, a_{jK}) \equiv \mathcal{D}(\mathbf{a}_k.)$ for $k = 1, 2, \dots, K$.

The complete data likelihood has a convenient structure that by combining with all the prior specifications mentioned above, leads to the following full conditional distributions:

$$p(\boldsymbol{\psi}_k | \cdot) \propto \prod_{\tilde{t}: S_{\tilde{t}}=k} \mathcal{C}(\Psi_k(L^{-1})\Phi_k(L)y_{\tilde{t}}|0, \gamma_k) \mathcal{N}(\boldsymbol{\psi}_k | \mathbf{0}, \tau_{\psi_k} \mathbf{I}) \quad (37)$$

$$p(\boldsymbol{\phi}_k | \cdot) \propto \prod_{\tilde{t}: S_{\tilde{t}}=k} \mathcal{C}(\Phi_k(L^{-1})\Phi_k(L)y_{\tilde{t}}|0, \gamma_k) \mathcal{N}(\boldsymbol{\phi}_k | \mathbf{0}, \tau_{\phi_k} \mathbf{I}) \quad (38)$$

$$p(\gamma_k | \cdot) \propto \prod_{\tilde{t}: S_{\tilde{t}}=k} \mathcal{C}(\Psi_k(L^{-1})\Phi_k(L)y_{\tilde{t}}|0, \gamma_k) \mathcal{G}(\gamma_k | \nu_{k1}, \nu_{k2}) \quad (39)$$

$$\boldsymbol{\xi}_k. \sim \mathcal{D}(a_{k1} + N_{k1}(\mathbf{S}), a_{k2} + N_{k2}(\mathbf{S}), \dots, a_{kK} + N_{kK}(\mathbf{S})) \quad (40)$$

for $k = 1, 2, \dots, K$.

A Metropolis algorithm with normal proposals can be used to sample from the full conditionals (37) - (38). A Metropolis-Hastings algorithm with gamma

proposals can be used to sample from (39). Notice that with the complete data likelihood and the Dirichlet prior, the posterior for the rows of the transition matrix is also Dirichlet, so the rows of the matrix can be sampled directly from the Dirichlet distribution (40).

Sampling the States

Conditioning on a vector of parameters $\{\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\xi}\}^T$, the states \mathcal{S} can be sampled from a distribution estimated by a Forward-Filtering Backward-Smoothing algorithm (FFBS) [1]. The filtering is an inference about the hidden state at time t given observations up to time t . Smoothing yields an inference about the hidden state at time t given all available data. For the MSMAR model, filtering the states at time t is implemented by conditioning on the data up to time $t + s$.

Let $\boldsymbol{\eta} = (\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\xi})^T$, be a vector of all state specific parameters, the maximum noncausal order be $s_{max} = \max\{s_1, s_2, \dots, s_K\}$ and the maximum causal order be $r_{max} = \max\{r_1, r_2, \dots, r_K\}$, the FFBS algorithm for the MSMAR model works as follows:

1. Forward filtering, for $t = 1 + r_{max}, \dots, T$:

- (a) Compute one-step predictive distribution

$$P(S_t = l | \mathbf{y}^{t+s_l-1}) = \sum_{k=1}^K \xi_{kl} P(S_{t-1} = k | \mathbf{y}^{t+s_k-1}, \boldsymbol{\eta}) \quad (41)$$

for $l = 1, 2, \dots, K$

- (b) As the new datum at time $t + s_{max}$ becomes available, the states at time

t are filtered as:

$$P(S_t = l | \mathbf{y}^{t+s_t}, \boldsymbol{\eta}) = \frac{\mathcal{C}(\Psi_l(L^{-1})\Phi_l(L)y_t|0, \gamma_l)P(S_t = l | \mathbf{y}^{t+s_t-1})}{p(y_t | \mathbf{y}^{t+s_{max}-1}, \boldsymbol{\eta})} \quad (42)$$

$$p(y_t | \mathbf{y}^{t+s_{max}-1}, \boldsymbol{\eta}) = \sum_{k=1}^K \mathcal{C}(\Psi_k(L^{-1})\Phi_k(L)y_t|0, \gamma_k)P(S_t = k | \mathbf{y}^{t+s_k-1}) \quad (43)$$

Initialize the filter recursion at:

$$P(S_{r_{max}} = l | \mathbf{y}^{s_{max}-1}) = \sum_{k=1}^K \xi_{kl} P(S_{r_{max}-1} = k) \quad (44)$$

2. To smooth the probabilities and sample the state indicators, the backward-smoothing algorithm runs for $t = T - s_{max} - 1, T - s_{max} - 2, \dots, t$:

(a) Initialize recursion at filtered probability distribution from the previous step $P(S_{T-s_{max}} = l | \mathbf{y}, \boldsymbol{\eta})$, then smooth the state probabilities with:

$$P(S_t = l | \mathbf{y}, \boldsymbol{\eta}) = \sum_{k=1}^K \frac{\xi_{lk} P(S_t = l | \mathbf{y}^t, \boldsymbol{\eta}) P(S_{t+1} = k | \mathbf{y}, \boldsymbol{\eta})}{\sum_{j=1}^K \xi_{jk} P(S_t = j | \mathbf{y}^t, \boldsymbol{\eta})}$$

(b) Sample state indicator:

$$S_t \sim \text{Multinom}(1, \boldsymbol{\theta})$$

$$\boldsymbol{\theta} = (P(S_t = 1 | \mathbf{y}, \boldsymbol{\eta}), \dots, P(S_t = K | \mathbf{y}, \boldsymbol{\eta}))^T$$

Several important points about the FFBS algorithm for the states estimation need to be discussed. Since the filtering recursion starts at $t = 1 + r_{max}$ the states $\{S_1, S_2, \dots, S_{r_{max}}\}$ must be initialized at some values by (44). These values can be resampled from smoothed probabilities obtained later by a backward-smoothing algorithm. The states are not filtered for the time points that are included in the interval between $t = T - s_{max} + 1, \dots, T$, in this work the values of the states at these points are obtained by sampling from the one step predictive distributions

obtained by (41). Since the smoothing algorithm starts at time $t = T - s_{max} - 1$ the probabilities of the state indicators can not be smoothed at the time points included in the interval $t = \{T, T - 1, \dots, T - s_{max} - 1\}$. When the available time series is long, this approximation is reasonable.

The MSMAR process has regime specific causal r_k and noncausal s_k orders that can be regarded as free parameters. In this work the model selection is performed via minimization of the Deviance Information Criterion (DIC), refer to section 2.2 for the description.

3.3 Simulation of the MSMAR processes

The method of simulation of Markov switching MAR processes is similar to the MAR processes described in section 2.3, with the additional sampling of a sequence of the state indicators \mathcal{S} . The method is also based on the partial fraction representation of the mixed AR process [2]. Consider the following reparameterization of the MAR process, $u_t = \Phi(L)y_t$ and $v_t = \Psi(L^{-1})y_t$. The equation (28) can be rewritten as:

$$\Psi(L^{-1})u_t = \epsilon_t$$

$$\Phi(L^1)v_t = \epsilon_t$$

The Mixed AR processes $\{y_t\}$ can be simulated by recovering from the unobserved components u_t and $\{v_t\}$ by using two representations [2],

1. Causal representation

$$y_t = L^s g_1(L)v_t + g_2(L)u_t \tag{45}$$

2. Noncausal representation

$$y_t = f_1(L^{-1})v_t + L^{-r} f_2(L^{-1})u_t \tag{46}$$

In the case of the MSMAR, the process y_t can be recovered conditionally on the simulated sequence of state indicators \mathcal{S} . A sequence of the state indicators

can be simulated by using a transition matrix (30) and the recurrence relation,

$$P(S_{t+1}|S_t = j) = \boldsymbol{\xi}_j.$$

$$S_{t+1} \sim \text{Multinom}(1, \boldsymbol{\xi}_j.)$$

Conditionally on a sequence of state indicators \mathbf{S} , the process y_t can be recovered by using causal representation and noncausal representation:

1. Causal representation

$$y_t = L^{s_t} g_{1S_t}(L)v_t + g_{2S_t}(L)u_t$$

2. Noncausal representation

$$y_t = f_{1S_t}(L^{-1})v_t + L^{-r_{S_t}} f_{2S_t}(L^{-1})u_t$$

In case of the MSMAR(1, 1) process, the simulation can be performed in the following steps:

1. Generate a vector of state indicators S_t and disturbances ϵ_t ,

(a) For $t = 1, 2, \dots, T$

$$\text{sample } (S_t | S_{t-1} = l) \sim \text{Multinom}(1, \boldsymbol{\xi}_l)$$

(b) For $t = 1, 2, \dots, T$

$$\text{sample } (\epsilon_t | S_t = j) \sim \mathcal{C}(0, \gamma_j)$$

2. Generate a sequence of u_t and v_t :

(a) initialize (u_{T+1}) and (v_0)

(b) Compute, for $t = T, T - 1, T - 2, \dots, 1$

$$u_t = \psi^{(s_t)} u_{t+1} + \epsilon_t$$

(c) Compute, for $t = 1, 2, \dots, T$

$$v_t = \phi^{(s_t)} v_{t-1} + \epsilon_t$$

(d) For $t = 1, 2, \dots, T$ recover y_t by:

$$y_t = \frac{1}{1 - \phi^{(s_t)} \psi^{(s_t)}} (u_t + \phi^{(s_t)} v_{t-1})$$

3.4 Simulation Study

This simulation study tries to assess the performance of the MSMAR under the model misspecification. The study was performed on the data that was simulated from the following MSMAR((1, 1), (1, 1)) process:

$$\begin{cases} (1 - 0.87L^{-1})(1 - 0.15L)y_t = \epsilon_t, & \epsilon_t \sim \mathcal{C}(0, 0.2) & \text{if } S_t = 1 \\ (1 - 0.72L^{-1})(1 - 0.25L)y_t = \epsilon_t, & \epsilon_t \sim \mathcal{C}(0, 2.2) & \text{if } S_t = 2 \end{cases} \quad (47)$$

with the transition matrix:

$$\boldsymbol{\xi} = \begin{pmatrix} 0.995 & 0.005 \\ 0.005 & 0.995 \end{pmatrix} \quad (48)$$

Assuming that the true parameters of the process (47) are unknown, the models that were considered for estimating the parameters are:

1. MSMAR((0, 1), (1, 0))

$$\begin{cases} (1 - \phi_{11}L)y_t = \epsilon_t, & \epsilon_t \sim \mathcal{C}(0, 0.2) & \text{if } S_t = 1 \\ (1 - \psi_{21}L^{-1})y_t = \epsilon_t, & \epsilon_t \sim \mathcal{C}(0, 2.2) & \text{if } S_t = 2 \end{cases}$$

2. MSMAR((0, 1), (1, 1))

$$\begin{cases} (1 - \phi_{11}L)y_t = \epsilon_t, & \epsilon_t \sim \mathcal{C}(0, 0.2) & \text{if } S_t = 1 \\ (1 - \psi_{11}L^{-1})(1 - \phi_{21}L^{-1})y_t = \epsilon_t, & \epsilon_t \sim \mathcal{C}(0, 2.2) & \text{if } S_t = 2 \end{cases}$$

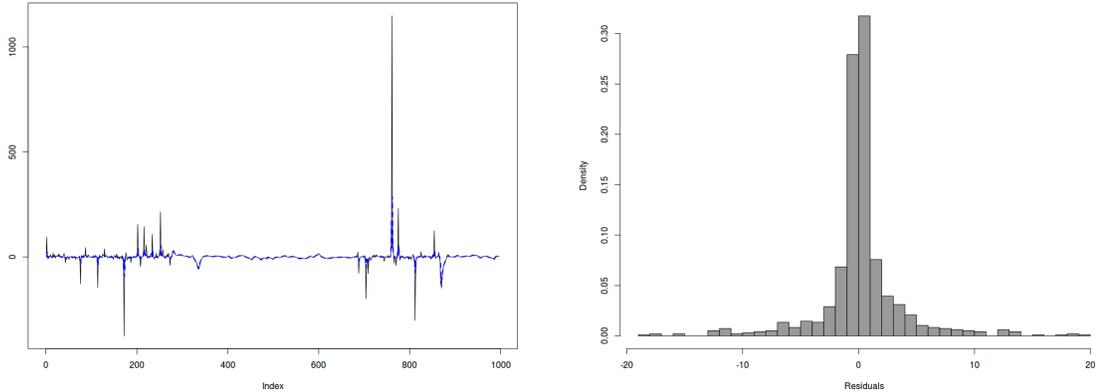
3. MSMAR((1, 0), (1, 1))

$$\begin{cases} (1 - \psi_{11}L)y_t = \epsilon_t, & \epsilon_t \sim \mathcal{C}(0, 0.2) & \text{if } S_t = 1 \\ (1 - \psi_{11}L^{-1})(1 - \phi_{21}L^{-1})y_t = \epsilon_t, & \epsilon_t \sim \mathcal{C}(0, 2.2) & \text{if } S_t = 2 \end{cases}$$

4. MSMAR((1, 1), (1, 1))

$$\begin{cases} (1 - \psi_{11}L^{-1})(1 - \phi_{11}L)y_t = \epsilon_t, & \epsilon_t \sim \mathcal{C}(0, 0.2) & \text{if } S_t = 1 \\ (1 - \psi_{11}L^{-1})(1 - \phi_{21}L^{-1})y_t = \epsilon_t, & \epsilon_t \sim \mathcal{C}(0, 2.2) & \text{if } S_t = 2 \end{cases}$$

The values of the estimated parameters and the corresponding 95% quantile based coverage intervals are reported in table 6. It is interesting to note that all estimated causal and noncausal parameters are very similar in values to each other, even under the model misspecification. The DIC has a minimum value at the “true” model MSMAR((1, 1), (1, 1)). The simulated data and the corresponding fit are shown in figure 10.



(a) Simulated MSMAR((1, 1), (1, 1)) data with the corresponding best fit.

(b) Distribution of the residuals of the best MSMAR fit.

Figure 10: Simulated MSMAR process.

Parameter	MSMAR((0, 1),(1, 0))	MSMAR((0, 1), (1, 1))	MSMAR((1, 0), (1, 1))	MSMAR((1, 1), (1, 1))
ψ_{11}	-	-	0.152 (0.146, 0.161)	0.150 (0.146, 0.156)
ψ_{21}	0.897 (0.888, 0.906)	0.873 (0.863, 0.880)	0.873 (0.863, 0.880)	0.873 (0.863, 0.880)
ϕ_{11}	0.250 (0.244, 0.255)	0.250 (0.245, 0.255)	-	0.250 (0.245, 0.253)
ϕ_{21}	-	0.718 (0.713, 0.725)	0.719 (0.712, 0.726)	0.718 (0.712, 0.726)
γ_1	2.397 (2.108, 2.710)	2.396 (2.113, 2.709)	2.756 (2.436, 3.093)	2.041 (1.805, 2.309)
γ_2	0.612 (0.543, 0.687)	0.230 (0.203, 0.260)	0.232 (0.204, 0.260)	0.230 (0.205, 0.258)
ξ_{11}	0.990 (0.987, 0.994)	0.994 (0.992, 0.996)	0.994 (0.992, 0.996)	0.994 (0.992, 0.996)
ξ_{12}	0.010 (0.006, 0.013)	0.006 (0.004, 0.008)	0.006 (0.004, 0.008)	0.006 (0.004, 0.008)
ξ_{21}	0.009 (0.006, 0.013)	0.005 (0.003, 0.008)	0.005 (0.003, 0.008)	0.005 (0.003, 0.008)
ξ_{22}	0.991 (0.987, 0.994)	0.995 (0.992, 0.997)	0.995 (0.992, 0.997)	0.995 (0.992, 0.997)
$\ln \mathcal{L}(\cdot)$	-2638.831 (-2653.697, -2623.998)	-2172.408 (-2185.118, -2161.695)	-2245.029 (-2257.025, -2234.025)	-2087.864 (-2100.362, -2077.006)
DIC	5394.627	4410.88	4557.134	4243.384

Table 6: Posterior means and 95% quantile based coverage intervals for the parameters of various MSMAR models applied to simulated data.

3.5 Application to Bitcoin/USD Exchange Rate

In this section the MSMAR model is applied to the Bitcoin/USD exchange rate time series and its performance is compared to the MAR models on the basis of the DIC criterion. The Bitcoin/USD exchange rate data is discussed in section 2.5.

Recall that, the series have a global nonlinear trend that can be clearly seen from figure 8a. As in the case of the MAR models, the trend was model by LCLS estimator. The bandwidth h was chosen 40.5 as in the case of the MAR models, to smooth the data just enough to account for the global trend without smoothing the bubble dynamics. The residuals after subtracting the trend, are assumed to follow an MSMAR process, so the full model specification is:

$$u_t - m(t) = y_t$$

$$\Psi_{S_t}(L^{-1})\Phi_{S_t}(L)y_t = \epsilon_t$$

The number of parameters in the MSMAR models grow quickly with increasing

causal-noncausal orders and the number of regimes. Let $p_k = s_k + r_k$ be the total number of the regime-specific causal and noncausal parameters and $m_k = |\boldsymbol{\eta}_k|$ be the cardinality of the set of the parameters of the distribution of the errors. The total number of parameters of the MSMAR processes is $\sum_{k=1}^K (p_k + m_k)$. The MSMAR models are usually over parameterized, so the analysis was performed with two-regime ($K = 2$) models with the most complex model MSMAR((1, 1), (1, 1)). One feature of the MSMAR processes that reduces the space of possible models, given a fixed number of regimes K , is that the MSMAR processes are invariant under the permutation of the regime assignments. For example the MSMAR((0, 1), (1, 1)) is equivalent to the MSMAR((1, 1), (0, 1)). This fact greatly reduces the space of the distinct models to consider, for a fixed number of regimes K . On the other hand, this feature introduces a label-switching phenomena in the MCMC simulations, that makes estimation of the marginal posteriors of the parameters harder, although it does not affect the overall fit.

The models that were fitted to the data were:

1. MSMAR((0, 1), (1, 0))

$$\begin{cases} (1 - \phi_{11}L)y_t = \epsilon_t & \text{if } S_t = 1 \\ (1 - \psi_{21}L^{-1})y_t = \epsilon_t & \text{if } S_t = 2 \end{cases}$$

2. MSMAR((0, 1), (1, 1))

$$\begin{cases} (1 - \phi_{11}L)y_t = \epsilon_t & \text{if } S_t = 1 \\ (1 - \psi_{21}L^{-1})(1 - \phi_{21}L)y_t = \epsilon_t & \text{if } S_t = 2 \end{cases}$$

3. MSMAR((1, 0), (1, 1))

$$\begin{cases} (1 - \psi_{11}L)y_t = \epsilon_t & \text{if } S_t = 1 \\ (1 - \psi_{21}L^{-1})(1 - \phi_{21}L)y_t = \epsilon_t & \text{if } S_t = 2 \end{cases}$$

4. MSMAR((1, 1), (1, 1))

$$\begin{cases} (1 - \psi_{11}L^{-1})(1 - \phi_{11}L)y_t = \epsilon_t & \text{if } S_t = 1 \\ (1 - \psi_{21}L^{-1})(1 - \phi_{21}L)y_t = \epsilon_t & \text{if } S_t = 2 \end{cases}$$

5. MSMAR((2, 1), (2, 1))

$$\begin{cases} (1 - \psi_{11}L^{-1} - \psi_{12}L^{-2})(1 - \phi_{11}L)y_t = \epsilon_t & \text{if } S_t = 1 \\ (1 - \psi_{21}L^{-1} - \psi_{22}L^{-2})(1 - \phi_{21}L)y_t = \epsilon_t & \text{if } S_t = 2 \end{cases}$$

Other second order models were also considered such as MSMAR((1, 2), (1, 2)), MSMAR((2, 1), (1, 2)), MSMAR((2, 2), (1, 1)), MSMAR((2, 2), (1, 2)), MSMAR((2, 0), (0, 2)). The transition matrix of all the models under consideration has the following form:

$$\boldsymbol{\xi} = \begin{pmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{pmatrix} \quad (49)$$

In all models the errors are assumed to follow the Cauchy distribution with a regime specific scale parameter, that is:

$$\begin{cases} \epsilon_t \sim \mathcal{C}(0, \gamma_1) & \text{if } S_t = 1 \\ \epsilon_t \sim \mathcal{C}(0, \gamma_2) & \text{if } S_t = 2 \end{cases}$$

The summaries of the draws of the parameters from the posterior distributions, and the corresponding DIC values are provided in table 7. Since the DIC values of the most second order models were larger than the first order models, we decided not to include the summaries of these models in the table. It can be seen that the DIC criterion favors the MSMAR((2, 1), (2, 1)) model. The final chosen model for the Bitcoin/USD exchange rate analysis is thus the MSMAR((2, 1), (2, 1)). The estimated model has the following form:

$$\begin{cases} (1 - 0.171L^{-1} - 0.027L^{-2})(1 - 0.931L)y_t = \epsilon_t, & \epsilon_t \sim \mathcal{C}(0, 0.054) & \text{if } S_t = 1 \\ (1 - 0.566L^{-1} - 0.232L^{-2})(1 - 0.332L)y_t = \epsilon_t, & \epsilon_t \sim \mathcal{C}(0, 3.422) & \text{if } S_t = 2 \end{cases}$$

with an estimated transition matrix:

$$\hat{\xi} = \begin{pmatrix} 0.992 & 0.008 \\ 0.008 & 0.992 \end{pmatrix} \quad (50)$$

Recall from table 5, the best MAR model's DIC value is 4809.988 which is much larger in comparison to the best MSMAR model's DIC 2942.961, which suggests that the MSMAR model is able to better capture the dynamics of the Bitcoin/USD exchange rate series.

Figure 11 shows the data with the regime assignments. It is clear from the plot, that one regime corresponds to relatively linear constant trends, and the other regime corresponds to the explosive behavior. The MSMAR((2, 1), (2, 1)) suggests that there are two regimes with the second order noncausal and first order causal dependence. The MSMAR model as well as MAR model still was not able to completely remove the autocorrelation in the series. As was discussed in section 2.5, the series may include seasonal components which are not modeled by the MSMAR in this work.

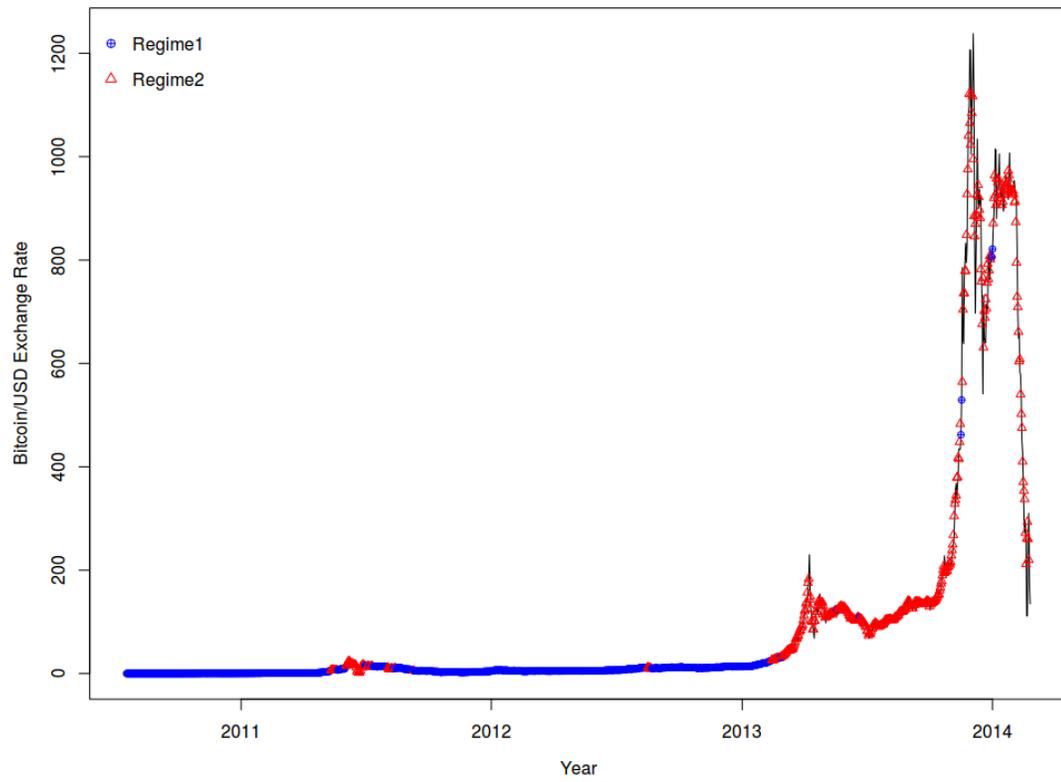
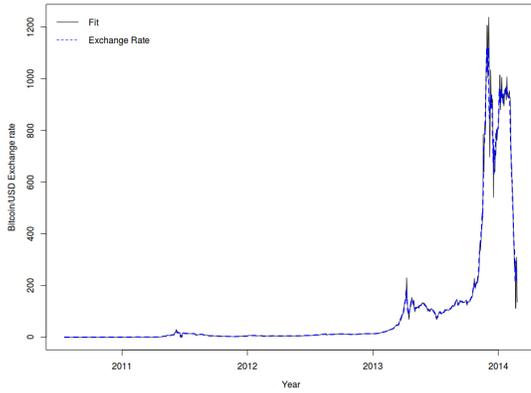
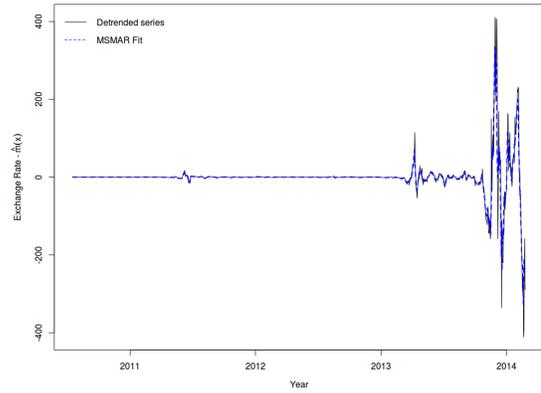


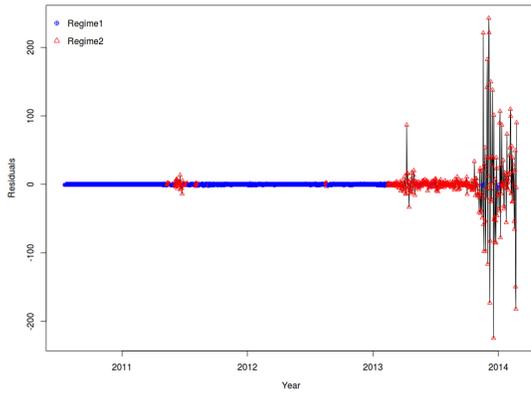
Figure 11: Posterior modes for the regimes of the Bitcoin/USD exchange rate data



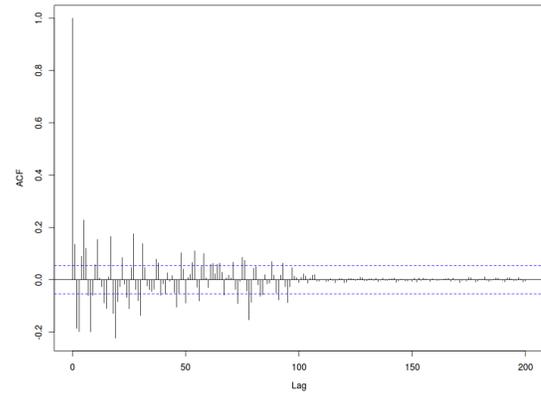
(a) MSMAR((2, 1), (2, 1)) fit added to the nonparametric estimate, $\hat{u}_t = \hat{m}(x) + \hat{y}_t$.



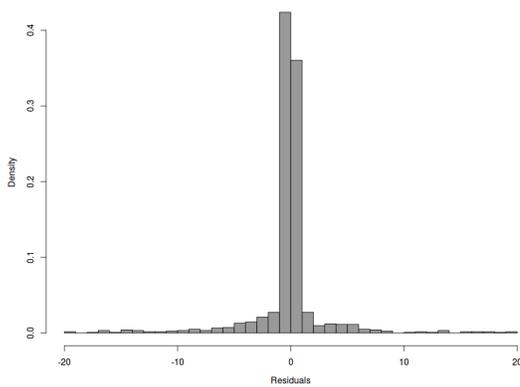
(b) A fit of the MSMAR((2, 1), (2, 1)) to the detrended series.



(c) Residuals of the MSMAR((2, 1), (2, 1)) with the corresponding regimes.



(d) Tn ACF of the residuals of the MSMAR((2, 1), (2, 1)) fit.



(e) A histogram of the residuals of the MSMAR((2, 1), (2, 1)) fit.

Figure 12: Visual summaries of the fit of the best MSMAR model.

Parameter	MSMAR((0, 1),(1, 0))	MSMAR((0, 1), (1, 1))	MSMAR((1, 0), (1, 1))	MSMAR((1, 1), (1, 1))	MSMAR((2, 1), (2, 1))
ψ_{11}	-	-	0.944 (0.930, 0.955)	0.098 (0.028, 0.138)	0.171 (0.159, 0.183)
ψ_{12}	-	-	-	-	-0.027 (-0.034, -0.021)
ψ_{21}	0.930 (0.868, 0.950)	0.876 (0.858, 0.898)	0.876 (0.858, 0.898)	0.877 (0.860, 0.900)	0.566 (0.511, 0.617)
ψ_{22}	-	-	-	-	0.232 (0.177, 0.288)
ϕ_{11}	0.996 (0.98, 1.007)	0.998 (0.985, 1.008)	-	0.954 (0.939, 1.001)	0.931 (0.919, 0.943)
ϕ_{12}	-	-	-	-	-
ϕ_{21}	-	0.154 (0.107, 0.189)	0.156 (0.113, 0.191)	0.154 (0.109, 0.188)	0.332 (0.295, 0.361)
ϕ_{22}	-	-	-	-	-
γ_1	0.056 (0.050, 0.062)	0.056 (0.051, 0.063)	0.054 (0.049, 0.060)	0.053 (0.047, 0.059)	0.054 (0.049, 0.060)
γ_2	3.733 (3.132, 4.408)	3.781 (3.202, 4.431)	3.726 (3.182, 4.341)	3.464 (2.932, 4.077)	3.422 (2.923, 3.989)
ξ_{11}	0.993 (0.990, 0.995)	0.993 (0.990, 0.995)	0.993 (0.991, 0.995)	0.992 (0.990, 0.995)	0.992 (0.990, 0.995)
ξ_{12}	0.007 (0.005, 0.010)	0.007 (0.005, 0.010)	0.007 (0.005, 0.009)	0.008 (0.005, 0.010)	0.008 (0.005, 0.010)
ξ_{21}	0.008 (0.005, 0.011)	0.008 (0.005, 0.011)	0.007 (0.005, 0.010)	0.008 (0.005, 0.011)	0.008 (0.006, 0.011)
ξ_{22}	0.992 (0.989, 0.995)	0.992 (0.989, 0.995)	0.993 (0.990, 0.995)	0.992 (0.989, 0.995)	0.992 (0.989, 0.994)
$\ln \mathcal{L}(\cdot)$	-1418.694 (-1437.565, -1400.374)	-1412.527 (-1431.022, -1393.422)	-1413.169 (-1429.123, -1397.54)	-1415.619 (-1432.895, -1398.716)	-1390.996 (-1408.616, -1373.359)
<i>DIC</i>	3015.076	3007.705	2958.686	2981.338	2942.961

Table 7: Posterior means and 95% quantiled based coverage intervals for the parameters of various MSMAR models in application to the Bitcoin/USD exchange rate series.

List of References

- [1] S. Frühwirth-Schnatter, *Finite mixture and Markov switching models*. Springer Science & Business Media, 2006.
- [2] C. Gourieroux and J. Jasiak, “Filtering, prediction and simulation methods for noncausal processes,” *Journal of Time Series Analysis*, 2015.
- [3] M. Lanne, J. Luoto, and P. Saikkonen, “Optimal forecasting of noncausal autoregressive time series,” *International Journal of Forecasting*, vol. 28, no. 3, pp. 623–631, 2012.
- [4] M. Lanne, A. Luoma, and J. Luoto, “Bayesian model selection and forecasting in noncausal autoregressive models,” *Journal of Applied Econometrics*, vol. 27, no. 5, pp. 812–830, 2012.

CHAPTER 4

Conclusion and Future Research Directions

In modeling the bubble phenomena, the Voigt distribution seems promising as an alternative to the Cauchy distribution. The DIC criterion suggests that the MARV processes are able to slightly better model the bubble phenomena than the MARC in the bitcoin data over the period between 02/20/2013 - 07/20/2013. On the other hand the MARC process slightly outperformed MARV in application to a longer series of the bitcoin data. The next step in developing the MARV processes would be to compare the performance of the MARV to MARC on a different dataset, for example the U.S inflation, which was analyzed in [1]. It would be also interesting to compare the performance of the MARV with the mixed causal-noncausal processes with t-distributed errors.

This work proposed the Markov-switching extension of the mixed causal-noncausal processes and it was shown that it is able to model the bitcoin data much better than the MAR processes, based on the DIC criterion. One of the directions in exploring the Markov switching causal-noncausal processes would be to consider a mixture of t-distributions. A possible advantage in using a mixture of t-distributions is that by allowing the degrees-of-freedom to depend on the regime, it may be more flexible in modeling complex variation in the data.

The MSMAR model, considered in this work, has also several limitations. The simulation study revealed that the transition matrix is underestimated in the settings of a frequently changing regimes. A possible explanation for this behavior is that the model does not include a state specific intercept, so under some parameter values the regimes are not well separated and the model can not identify the regimes correctly. In other words, if the true data generating process

has the regimes that are changing very often, the MSMAR may not be able to correctly identify the state indicators. Another limitation is that the MSMAR model does not completely remove the autocorrelation in the bitcoin data, this may be due to the seasonality in the series, so an addition of an AR(5) may be needed. There is also a limitation in the current implementation of the MCMC sampling algorithm. The MCMC algorithm used in this work does not take into account the label switching phenomena. One of the possible solutions to this problem is to use a random permutation algorithm described in [2].

Yet another direction to take in exploring the mixed causal-noncausal processes is an incorporation of the trend in the model, rather than as a data preprocessing tool.

One of the most important uses of time series analysis is forecasting. The next step in extending the MSMAR processes is to develop the forecasting method.

List of References

- [1] M. Lanne and J. Luoto, “A noncausal autoregressive model with time-varying parameters: An application to us inflation,” 2013.
- [2] S. Frühwirth-Schnatter, *Finite mixture and Markov switching models*. Springer Science & Business Media, 2006.

BIBLIOGRAPHY

- Abramowitz, M. and Stegun, I. A., *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Corporation, 1964, vol. 55.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M., *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- Breidt, F. J. and Davis, R. A., “Time-reversibility, identifiability and independence of innovations for stationary time series,” *Journal of Time Series Analysis*, vol. 13, no. 5, pp. 377–390, 1992.
- Davis, R. A., Lii, K.-S., and Politis, D. N., “Maximum likelihood estimation for noncausal autoregressive processes,” in *Selected Works of Murray Rosenblatt*. Springer, 2011, pp. 396–419.
- Frühwirth-Schnatter, S., *Finite mixture and Markov switching models*. Springer Science & Business Media, 2006.
- Garber, P. M., “Famous first bubbles,” *The Journal of Economic Perspectives*, vol. 4, no. 2, pp. 35–54, 1990.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B., “Bayesian data analysis. texts in statistical science series,” 2004.
- Gourieroux, C. and Jasiak, J., “Filtering, prediction and simulation methods for noncausal processes,” *Journal of Time Series Analysis*, 2015.
- Gouriéroux, C. and Zakoian, J.-M., “Local explosion modelling by noncausal process,” 2016.
- Gouriéroux, C., Zakoian, J.-M., *et al.*, “Explosive bubble modelling by noncausal process,” Tech. Rep., 2013.
- Härdle, W., Müller, M., Sperlich, S., and Werwatz, A., *Nonparametric and semi-parametric models*. Springer Science & Business Media, 2012.
- Hecq, A., Lieb, L., and Telg, S., “Identification of mixed causal-noncausal models in finite samples,” *Annals of Economics and Statistics/Annales d’Économie et de Statistique*, no. 123/124, pp. 307–331, 2016.
- Hencic, A. and Gouriéroux, C., “Noncausal autoregressive model in application to bitcoin/usd exchange rates,” in *Econometrics of Risk*. Springer, 2015, pp. 17–40.

- Henderson, D. J. and Parmeter, C. F., *Applied nonparametric econometrics*. Cambridge University Press, 2015.
- Kuntz, M., “A new implementation of the humlicek algorithm for the calculation of the voigt profile function,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 57, no. 6, pp. 819–824, 1997.
- Lanne, M. and Luoto, J., “A noncausal autoregressive model with time-varying parameters: An application to us inflation,” 2013.
- Letchworth, K. L. and Benner, D. C., “Rapid and accurate calculation of the voigt function,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 107, no. 1, pp. 173–192, 2007.
- Lether, F. and Wenston, P., “The numerical computation of the voigt function by a corrected midpoint quadrature rule for (-),” *Journal of computational and applied mathematics*, vol. 34, no. 1, pp. 75–92, 1991.
- Nov, Y. and Nov, O., “Living in a bubble? toward a unified bubble theory,” *International Journal of General Systems*, vol. 37, no. 5, pp. 627–635, 2008.
- Ouyang, Y. and Yin, H., “Time series prediction with a non-causal neural network,” in *Computational Intelligence for Financial Engineering & Economics (CIFER), 2104 IEEE Conference on*. IEEE, 2014, pp. 25–31.
- Poppe, G. and Wijers, C., “More efficient computation of the complex error function,” *ACM Transactions on Mathematical Software (TOMS)*, vol. 16, no. 1, pp. 38–46, 1990.
- Schreier, F., “Comments on a common misunderstanding about the voigt line profile,” *Journal of the Atmospheric Sciences*, vol. 66, no. 6, pp. 1860–1864, 2009.
- Schreier, F., “Optimized implementations of rational approximations for the voigt and complex error function,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 112, no. 6, pp. 1010–1025, 2011.
- Siegel, J. J., “What is an asset price bubble? an operational definition,” *European financial management*, vol. 9, no. 1, pp. 11–24, 2003.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A., “The deviance information criterion: 12 years on,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 76, no. 3, pp. 485–493, 2014.