MODELLING COMOVEMENTS OF SELECTED LARGE CAP CRYPTOCURRENCIES: A SEMI-PARAMETRIC NONCAUSAL VAR APPROACH

MAURI KEMERSLEY RUTTEN HALL

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN ECONOMICS
YORK UNIVERSITY
TORONTO, ONTARIO

MARCH, 2023

Abstract

This dissertation focuses on multivariate mixed causal noncausal models and their application to cryptocurrencies. The empirical application considered in the dissertation focuses on detecting comovements in the US dollar exchange rates of four prominent cryptocurrencies as well as on forecasting for multivariate mixed causal noncausal models. The dissertation explores forecasting methods which can be used when modelling data as mixed causal noncausal multivariate processes.

The dissertation is divided into three chapters. Chapter one is dedicated to the detection of comovements in four prominent cryptocurrency US dollar exchange rates by modelling the exchange rates as mixed causal noncausal processes. The cryptocurrency pairs Bitcoin/Ethereum and Ripple/Stellar are modelled as bivariate mixed causal noncausal processes and then estimated as a single mixed causal noncausal multivariate time series model of dimension four.

Chapter two contains forecasting methods and applications to the cryptocurrencies estimated in chapter one. Nonparametric predictive densities are calculated and a new linear approximation method is introduced and used to calculate one step ahead out of sample forecasts using a mixed causal noncausal vector autoregressive model estimated via the GCov estimator.

Chapter three extends theory for forecasting multivariate mixed causal noncausal processes. Theory pertaining to the calculation of predictive densities for mixed causal noncausal vector autoregressive models with three lags using the latent

causal and noncausal components of the process is discussed. Theory pertaining to the calculation of predictive densities for causal noncausal vector autoregressive models with high dimensional data is also discussed.

Two experiments are conducted via simulation study. Firstly the coverage of the predictive density forecasting method is investigated. Secondly the predictive density forecasting method is compared with the linear approximation forecasting method in terms of their respective Mean Squared Errors.

Dedication

For my father, Tony "Lordstreet" Hall and my mother Mary Dena Hall. I love you both more than words can express. Walk good.

Acknowledgements

I grateful to Professor Joann Jasiak. I could not have completed my study without her help and guidance throughout. I would also like to thank the supervisory committee members Professors Antoine Djogbenou and Razvan Sufana for their support and assistance in completing my dissertation.

I am also deeply indebted to my wife Lisa Rajkumar-Maharaj, my sister Lindsay Hall, my mother Mary Hall and my father Tony Hall for their love and unwavering support.

Table of Contents

Abstract iii

Dedication iv

Acknowledgements v

Table of Contents vi

List of Tables x

List of Figures xii

Introduction 1

1 Detecting Comovements in Cryptocurrency Markets 8

1.1 Introduction 8

1.2 The Causal Noncausal VAR(1) Model 14

1.2.1 Representation Theorem GJ (2017) 15

1.2.2 VAR(1): Identification of the Causal and Noncausal Dimensions 19

1.2.3 Bivariate VAR(1) - Interpretation 20

1.2.4 VAR(p) model 25

1.2.5 Estimation Using Nonlinear Covariances 28

1.2.6 Estimation of Φ_j by The GCov 30

1.2.7 Asymptotic Properties of the GCov Estimator 31

1.2.8 Estimation Procedure, Identification and Estimation 31

1.3 Empirical Analysis of Four Large Cap Cryptocurrencies 32

1.3.1 Cyptocurrencies 33

1.3.2 Bitcoin (BTC) and Ethereum (ETH) 34

1.3.3 VAR(1) For BTC and ETH 39

1.3.4 Impulse Responses 43

1.3.5 VAR(3) Model for BTC and ETH 48

1.3.6 Comparing Goodness of Fit of the Mixed VAR(3) Against Causal VAR(3) for BTC and ETH USD Exchange Rates . . . 55

1.3.7 XRP (Ripple) and XLM (Stellar) . . . 58

1.3.8 VAR(1) for XLM and XRP . . . 63

1.3.9 Bivariate VAR(3) for XRP and XLM . . . 70

1.3.10 Comparing Goodness of Fit of the Mixed VAR(3) Against Causal VAR(3) for XRP and XLM USD Exchange Rates . . . 76

1.3.11 VAR(1) For Bitcoin, Ethereum, Ripple and Stellar . . . 79

1.4 Conclusion . . . 81

2 Forecasting **83**

2.1 Introduction . . . 83

2.2 Predictive Density for Bivariate Noncausal VAR(1) Model . . . 84

2.2.1 Forecasting $y^*_{1,T+1}, y^*_{2,T+1}$ given I_T. . . . 86

2.2.2 Predictive Density of $y_{1,T+1}, y_{2,T+1}$ . . . 89

2.2.3 One Step Ahead Out Of Sample Forecasts: Bivariate Predictive Densities . . . 91

2.3 Linear Approximation Method . . . 95

2.3.1 One Step Ahead Linear Forecasts . . . 99

2.4 Conclusion . . . 106

3 Predictive Densities **107**

3.1 Introduction 107

3.2 Predictive Density for VAR(3) 108

3.3 Predictive Density for VAR(1) with n $>$ 2 114

3.4 Simulation Study 117

3.4.1 Point and Interval Forecast Performance 117

3.4.2 Point and Interval Forecasts from Linear Approximation . . . 120

3.5 Conclusion 121

Conclusion **122**

Bibliography **125**

Appendices **129**

A Summary Sample Statistics 129

B Supplementary Graphs 129

C Finding A^1 and A^2 when the number of noncausal directions is known 136

List of Tables

1.1 BTC and ETH ADF P-Values for Detrended Series 39

1.2 BTC and ETH Excess Kurtosis and Skewness for Detrended Series . 39

1.3 BTC and ETH Normality Tests for VAR(1) Residuals 43

1.4 BTC and ETH Excess Kurtosis and Skewness for VAR(1) Residuals . 43

1.5 XRP and XLM ADF P-Values for Detrended Series 59

1.6 XRP and XLM Excess Kurtosis and Skewness for Detrended Series . 62

1.7 Normality Tests for XRP and XRP VAR(1) Residuals 67

1.8 XRP and XLM Excess Kurtosis and Skewness for VAR(1) Residuals . 67

2.1 Estimation of Noncausal VAR 98

2.2 BTC/ETH VAR(1) MAE 100

2.3 BTC/ETH VAR(3) MAE 101

2.4 XRP/XLM VAR(1) MAE 102

2.5 XRP/XLM VAR(3) MAE 103

2.6 BTC/ETH/XRP/XLM MAE 105

3.1 Available Methods for Multi-dimensional Density Estimation 114

3.2 Mean Squared Forecast Error for each series. 120

A.1 Summary Statistics 129

C.1 Number of Independent Arguments 138

C.2 Minimum Value of H 138

List of Figures

1.1 BTC/USD and ETH/USD Exchange Rates. BTC/USD solid line, ETH/USD dotted line. 35

1.2 BTC and ETH (Adjusted) Detrended by Spline 37

1.3 ACF of Detrended BTC and ETH VAR(1) 38

1.4 ACF of Residuals from VAR(1) for BTC and ETH 40

1.5 ACF of Squared Residuals from VAR(1) for BTC and ETH 41

1.6 Histograms of Residuals from VAR(1) for BTC and ETH 42

1.7 BTC and ETH Plot of Residuals VAR(1) 42

1.8 Noncausal Impulse Response Coefficients for BTC and ETH VAR(1) 45

1.9 BTC and ETH Series and Causal and Noncausal Combinations for VAR(1) 47

1.10 ACF of Residuals from VAR(3) for BTC and ETH 51

1.11 ACF of Squared Residuals from VAR(3) for BTC and ETH 52

1.12 Noncausal Impulse Response Coefficients for BTC and ETH VAR(3) 53

1.13 BTC and ETH Series and Causal Noncausal Combinations for VAR(3) 55

1.14 ACF of Squared Residuals from Causal VAR(3) for BTC and ETH . 58

1.15 XRP/USD and XML/USD Exchange Rates. XRP/USD solid line, XML/USD dotted line. 60

1.16 XLM and XRP (Adjusted) Detrended by Spline 61

1.17 ACF of Adjusted and Detrended XRP and XLM 62

1.18 ACF of Residuals from VAR(1) for XRP and XLM 64

1.19 ACF of Squared Residuals from VAR(1) for XRP and XLM 65

1.20 Histograms of Residuals from VAR(1) for XRP and XLM 66

1.21 XLM and XRP Plot of Residuals for VAR(1) 66

1.22 Noncausal Impulse Response Coefficients for XRP and XLM for VAR(1) 67

1.23 XRP and XLM Series and Causal and Noncausal Combinations for VAR(1) 69

1.24 ACF of VAR(3) Residuals for XRP and XLM 72

1.25 ACF of VAR(3) Squared Residuals for XRP and XLM 73

1.26 XRP and XLM Noncausal Impulse Response Coefficients for VAR(3) 74

1.27 XRP and XLM Series and Causal Noncausal Combinations for VAR(3) 75

1.28 ACF of Causal VAR(3) Squared Residuals for XRP and XLM 77

1.29 BTC, ETH, XRP, and XLM Exchange Rates Spline Detrended (adjusted) 79

2.1 BTC Marginal Predictive Density, Forecast: 9: CI (1.8, 16.1) 92

2.2 ETH Marginal Predictive Density. Forecast 21, CI (0.9, 3.42) 92

2.3 BTC/ETH Joint Density 93

2.4 XRP Marginal Predictive Density Forecast : 0.005 CI (-0.007, 0.013) . 93

2.5 XLM Marginal Predictive Density Forecast XLM: CI (-0.0392, 0.0161) 94

2.6 XRP/XLM Joint Density 94

2.7 BTC and ETH VAR(1) Forecasts 100

2.8 BTC and ETH VAR(3) Forecasts 101

2.9 XRP and XML VAR(1) Forecasts 102

2.10 XRP and XML VAR(3) Forecasts 103

2.11 Forecast Demeaned BTC, ETH, XRP, XLM, one step, out of sample 105

3.1 Marginal Predictive Densities Forecasting the 1000th Observation of Simulated Bivariate Series with t-Distributed Errors (6 Degrees of Freedom) 119

B.1 BTC/ETH Histograms of VAR(3) Residuals 129

B.2 XRP/XLM Histograms of VAR(3) Residuals 130

B.3 BTC/ETH/XRP/XLM Histograms of Residuals from VAR(1) 130

B.4 BTC/ETH/XRP/XLM QQ Plots of Residuals from VAR(1) 131

B.5 BTC/ETH/XRP/XLM QQ Plot of Residuals from VAR(3) 132

B.6 ACF of Demeaned VAR(1) Residuals and Squared Residuals for BTC, ETH, XRP, XLM pairs (n=4) 133

B.7 BTC/ETH/XRP/XLM QQ Plots for Demeaned VAR(1) (n = 4) . . . 134

B.8 BTC ETH XRP XLM Hill Plots of Demeaned Series 135

Introduction

This dissertation compares traditional causal multivariate time series models and mixed causal noncausal time series models in application to modelling cryptocurrency USD exchange rates. We introduce new forecast methods for multivariate mixed causal noncausal models that are easy to implement in practice and extend the existing literature. We provide the predictive density from which we can obtain point forecasts and from which forecast intervals for each variable can be easily inferred. We introduce new theoretical results which extend those presented in Gouriéroux and Jasiak (2016) and provide a simple forecast method easily applicable in practice. The application of causal noncausal models to modelling cryptocurrencies is motivated by their empirical properties, in particular, local explosive patterns such as spikes and bubbles. The traditional stationary causal models are unable to accommodate these patterns.

We explore whether mixed causal noncausal multivariate time series models are preferable to other available models for researchers studying and forecasting cryptocurrency exchange rates. We apply causal noncausal multivariate time series models because these models can accommodate (a) spikes/bubbles (b) conditional heteroskedasticity and (c) are able to detect comovements between time series that

standard causal models cannot.

Research on forecasting large capitalization cryptocurrencies such as Bitcoin and Ethereum has focused primarily on univariate models. Cryptocurrencies such as Bitcoin have been modelled as noncausal univariate processes, for example in Gouriéroux and Hencic (2014). Several papers focus on modelling and forecasting univariate noncausal processes (see Lanne and Luoto (2012), Lanne and Saikkonen (2012), Gouriéroux and Zakoian (2017), Lof and Nyberg (2017) and Fries and Zakoian (2019)). However, since we observe comovements in cryptocurrency US dollar exchange rates we seek to model these processes jointly (i.e. as multivariate processes) instead of univariate processes.

Research on modelling cryptocurrencies as multivariate processes has primarily been focused on causal models, ignoring the possibility of noncausality. For example, Catania et al. (2019) investigate whether combinations of parameter varying multivariate causal models can lead to improved point forecasts and find conflicting evidence. They also find that combinations of time varying parameter causal multivariate models can lead to improved predictive density forecasts relative to standard causal multivariate models, but do not investigate noncausality.

In addition, we model prices rather than returns which separates our approach from the traditional approach as well. We model cryptocurrencies as strictly stationary processes because the assets are mean reverting rather than globally explosive, (i.e. they do not contain a unit root). Our approach relies on the assumption of strict stationarity. This assumption allows us to model prices rather than returns.

Our mixed causal noncausal model is estimated using a semi-parametric estimator introduced by Gouriéroux and Jasiak (2017) which enables the consistent estimation of noncausal models. This allows for the possibility of noncausality in cryptocurrency US dollar exchange rates. Empirical results, especially in finance, suggest that the traditional assumption of causality leads to models which are unable to capture salient features of many observed time series, especially those exhibiting bubble phenomena, where a bubble is defined as locally explosive episodes in a strictly stationary setting (see Hecq and Voisin (2022)). Traditional VAR models estimated under the assumption of Gaussian errors are unable to model the high kurtosis (i.e. fat tails) in the error distributions often observed in financial data due to the assumption of normality of the errors. This dissertation therefore applies a more flexible multivariate time series modelling approach to cryptocurrency exchange rates. Moreover, modelling such time series jointly in the framework of a noncausal Vector Autoregressive Process (VAR) model can accommodate bubble comovements between multiple assets, especially bubble cointegration where 'bubble cointegration' refers to the occurrence of bubbles in assets which move together. Cryptocurrencies often exhibit speculative bubbles, which we define as locally explosive behaviour in strictly stationary processes, and the presence of bubbles suggests noncausal dynamics.

In Chapter 1 we compare the results of the traditional VAR models and mixed causal noncausal models for four cryptocurrencies. The model is estimated by the semi-parametric Generalized Covariance (GCov) Estimator. We analyze the cryptocurrencies Bitcoin (BTC), Ethereum (ETH), Ripple (XRP) and Stellar (XLM) because we believe these two pairs of cryptocurrencies - Bitcoin and Ethereum on

the one hand and Ripple and Stellar on the other - have characteristics in common. Bitcoin and Ethereum are market capitalization leaders while Ripple and Stellar serve similar purposes and share similar underlying technological features.

Evidence that cryptocurrencies exchange rates are not normally distributed suggests that a mixed causal noncausal model may provide a better fit to the data than the traditional Box-Jenkins approach. The time series for the cryptocurrencies under investigation can therefore be decomposed into causal ('regular') and noncausal ('explosive') components. The decomposition of cryptocurrency exchange rates into latent (i.e. unobserved) causal and noncausal components has not been previously undertaken for cryptocurrencies in the literature on multivariate time series modelling.

Showing that the cryptocurrencies under consideration can be modelled as mixed processes, decomposed into causal and noncausal components and that modelling these processes as mixed processes provides a better fit to the data than traditional approaches suggests that mixed causal noncausal models may be a more appropriate choice for modelling cryptocurrencies. In addition, we are able to show that comovements between cryptocurrencies can be empirically detected using mixed causal noncausal models while traditional models fail to detect these comovements.

In Chapter 2 we aim to exploit noncausality for forecasting purposes in the presence of bubble phenomena. Forecasting cryptocurrencies jointly using causal noncausal models allows us to model comovements and bubble cointegration which has not been done previously in the literature. The goal of this chapter is to use the decomposition of multivariate time series processes to forecast the exchange rates of the cryptocurrencies under investigation.

Modelling the noncausal components in particular allows us to find a common bubble component in multivariate series. When series share a common noncausal component we can forecast these series more accurately as a joint process by taking the common bubble co-movements into account which we would otherwise be unable to do using traditional models.

We implement a forecasting method whose purpose is to provide a functional estimator that allows for the use of simulation methods in the creation of predictive densities for univariate processes in a single step. We aim to calculate forecasts and confidence intervals in a bivariate context for the cryptocurrency pairs Bitcoin/Ethereum and Ripple/Stellar.

Our next goal is the calculation of one step ahead forecasts using causal noncausal VAR model estimated via the GCov Estimator. We go on to explore the validity of a linear approximation method, which utilizes a simplified approach relative to the calculation of predictive densities. The linear approximation method combines the linear prediction calculation for casual VAR models with the consistent estimation of the autoregressive coefficient matrix. We implement the linear prediction method because the standard VAR method of forecasting does not produce unbiased forecasts and is not optimal in the context of a mixed causal noncausal VAR. We estimate the autoregressive polynomial as a deterministic function of time using the linear approximation method and calculate one step ahead forecasts using its simple linear forecast formula.

In Chapter 3 we make two theoretical contributions pertaining to the calculation of predictive densities for mixed causal noncausal VAR models. In so

doing we extend existing theory of mixed causal noncausal VAR models.

We obtain theoretical results for the calculation of predictive densities for a bivariate mixed causal noncausal VAR model with three lags using kernel density estimators by outlining how this can be accomplished using the latent causal noncausal components of the process, rather than the observed components of the process. We implement this new method of predictive density computation for multivariate noncausal processes and investigate its properties via simulation study.

We contribute to the theory concerning the calculation of predictive densities for a mixed causal noncausal VAR model at one lag for higher dimensional data (i.e. for dimensions greater than two) by extending it. A difficulty that arises in these cases is that the nonparametric model of the predictive density can exhibit large estimation errors as the dimensions increase. Another difficulty is that the rate of convergence of the Mean Integrated Squared Error (MISE) of the Kernel Density Estimator (KDE) is slow when the number of dimensions is greater than two. To address these issues we incorporate the Regularization Of the Derivative Expectation Operator (RODEO) technique. This technique is used for regularizing the bias reduction associated with a bandwidth reduction to avoid bandwidth over fitting and increases the rate of convergence of the MISE in high dimensional data. Under certain sparsity conditions it avoids the curse of dimensionality and achieves optimal convergence. This allows for the calculation of predictive densities for high dimensional causal noncausal VAR models while circumventing the curse of dimensionality. We also conduct a simulation study testing the validity of the use of predictive densities in the estimation of confidence intervals for the forecasts.

This is a novel undertaking in the sense that neither the theory pertaining to the calculation of predictive densities for a mixed causal noncausal VAR models with three lags using the latent causal noncausal components of the process nor the theory concerning the calculation of predictive densities for causal noncausal VAR models with high dimensional data has been explored previously. The theoretical contribution of this chapter serves as the basis for further research.

Chapter 1

Detecting Comovements in Cryptocurrency Markets

1.1 Introduction

A noncausal autoregressive stochastic process is one whose current value can be modelled as being dependent on future as well as possibly past values of the process (see Breidt et al. (1990), Breidt et al. (1992), Breid et al. (1991), Breidt and Davis (1992b), Brockwell and Davis (1987), Breidt and Davis (1992a), and Hannen (1973)). An autoregressive stochastic process whose current value or values depend only on its past values is known as a causal process. In contrast, a stochastic process whose current value or values depend only on its future values is known as a purely noncausal process. Stochastic processes whose current value or values depend on both future and past values are known as a mixed process.

In this chapter we model comovements between four cryptocurrencies as a mixed causal noncausal VAR(p) process. To estimate this model we apply the Generalized Covariance estimator Gouriéroux and Jasiak (2017) which is a consistent semi-parametric estimator for mixed causal noncausal multivariate non-Gaussian processes.

The assumption of Gaussianity, common in the time series literature, implies the restriction of the parameter space of a given time series model to the causal region. The assumption of Gaussianity implies that forward and backward dynamics of a given stochastic process are not distinguishable and, as such, forward looking dynamics (i.e. those depending on future values of the process) cannot be identified, which implies that for noncausal time series models we must move away from the assumption of normality. Empirical results, especially in finance, suggest that the assumption of causality leads to models which are insufficient for capturing salient features of many observed time series.

In the context of modelling multivariate time series, which have either purely noncausal components or both causal and noncausal components, the representation of the processes becomes relevant. One representation explored in the literature seeks to represent mixed causal noncausal multivariate time series processes in terms of multiplicative autoregressive causal and noncausal polynomials Lanne and Saikkonen (2013) (LS hereafter). An alternative representation is the classical VAR(p) model with the roots of the autoregressive polynomial assumed to be either inside or outside of the unit circle Gouriéroux and Jasiak (2017) Davis and Song (2020). Either representation of the causal noncausal VAR(p) model lends itself to estimation by the semi-parametric GCov estimator which is asymptotically normally distributed.

The advantage of the GCov estimator is that it does not require distributional assumptions on the errors.

This chapter is organized as follows: Section 2 discusses the causal noncausal Vector Autoregressive (VAR) model and the GCov estimator. Section 3 introduces and analyses the cryptocurrencies Bitcoin (BTC), Ethereum (ETH), Ripple (XRP) and Stellar (XLM) via multivariate analysis of their respective USD exchange rates. Section 4 concludes. Appendix A contains summary statistics, Appendix B contains supplementary graphs.

The Vector Autoregression of order p (VAR(p)) model represents the dynamics of a weakly stationary multivariate process $y_t, t = 1, 2, ..., T$ of dimension n:

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + ... + \Phi_p y_{t-p} + e_t \tag{1}$$

where $\Phi_i, i = 1, ..., p$ are $n \times n$ matrices of autoregressive coefficients, e_t is an error vector of length n which follows a weak white noise (i.e. a sequence of uncorrelated random vectors) with mean zero and a positive definite variance matrix Σ. [1]

Under the classical approach y_t is assumed to be causal i.e. past dependent and stationary. This condition implies that the roots of

$$det(Id - \Phi_1 z - \Phi_2 z^2 - ... - \Phi_p z^p) = 0$$

[1] In equation (1) y_t is assumed to have zero mean.

lie outside the unit circle.

This assumption is too stringent for practical applications as it eliminates stationary noncasual or mixed causal noncausal dynamics in non-Gaussian processes. Moreover, estimation methods such as the normality-based Maximum Likelihood (ML) and Ordinary Least Squares (OLS) applied to such processes do not distinguish between the causal and noncausal dynamics of the process due to the lack of identification issue, and therefore yield inconsistent estimates in noncausal and non-Gaussian.

The estimation of multivariate noncausal and mixed processes from a non-Gaussian Maximum Likelihood is complicated in practice due to the presence of a Jacobian which has limited their use in empirical analysis. To eliminate this difficulty LS propose a vector autoregressive VAR model for strictly stationary non-Gaussian time series using a multiplicative polynomial representation:

$$\Pi(L)\Phi(L^{-1})y_t = \varepsilon_t,$$

where $\Pi(L) = Id - \Pi_1 L - ... - \Pi_r L^r$, and, $\Phi(L^{-1}) = Id - \Phi_1 ... - \Phi_s L^{-s}$ are $n \times 1$ autoregressive causal and noncausal polynomials such that $det\Phi(z) \neq 0$ for $|z| \leq 1$ and $det\Pi(z) \neq 0$ for $|z| \leq 1$, and ε_t is a $n \times 1$ sequence of independent and identically distributed (i.i.d.) random vectors with zero mean and finite positive definite variance-covariance matrix.

LS argue that modelling and estimation of non-Gaussian processes using a purely causal VAR model may yield misleading results. In order to estimate the

model the authors rely on the method of Approximate Maximum Likelihood (AML). LS show that under regularity assumptions about the distribution of the error term the maximum likelihood estimator is asymptotically consistent and asymptotically normal. A limitation of the LS paper is the assumption of a known (necessarily symmetric) elliptical error distributions and the fact that their multiplicative polynomial representation with autoregressive orders or r and s may not be unique, and may not always exist for a VAR(p).

Gouriéroux and Jasiak (2017), Davis and Song (2020) and Swensen (2022) consider the classical representation (1) under modified assumptions. More specifically Gouriéroux and Jasiak (2017) assume that the errors e_t follow a sequence of independent and identically distributed (i.i.d) vectors with positive definite variance-covariance matrix Σ and the roots of the autoregressive polynomial lie either outside or inside the unit circle. Both articles discuss the identification and estimation of the causal noncausal VAR(p) models. Davis and Song (2020) rely on the ML estimation which requires a distributional assumption on the error terms involving the risk of mis-specification. Gouriéroux and Jasiak (2017) introduce a semi-parametric estimator called the Generalized Covariance Estimator (the GCov hereafter) for mixed causal noncausal multivariate non-Gaussian processes. The estimator does not require an assumption of a specific parametric error distribution and uses the nonlinear autocovariances for identification of causal noncausal components. Gouriéroux and Jasiak (2017) show that the GCov estimator is consistent and asymptotically normally distributed.

A noncausal VAR model can be useful in modelling dynamic data such as economic and financial time series exhibiting speculative bubbles (see Blanchard and

Watson (1982), Phillips and Yu (015a), Phillips and Yu (015b), Phillips and Shi (2018) and Fries (2019)) since linear noncausal models can accommodate complex nonlinear dynamics in calendar time in the framework of a stationary process. This allows for modelling such time series jointly as a noncausal VAR model can accommodate bubble comovements, especially bubble cointegration.

The standard Box-Jenkins approach involving normality-based ML or OLS estimation is only adequate for causal linear time-series which are Normally distributed, stationary, linear and which can be represented by a moving average with weak white noise errors. This approach is based on the identification and estimation of time series from moments of order up to only two. Consequently, the Box-Jenkins method, which relies on Gaussian maximum likelihood, is unable to accommodate jumps, bubbles and local trends which involve higher moments.

When a Gaussian likelihood is estimated from non-Gaussian processes the resulting Gaussian pseudo-maximum likelihood estimators may be inconsistent when applied to data with noncausal dynamics. Gaussian pseudo-maximum likelihood estimators may be inconsistent when applied to non-Gaussian data and inference based on these estimators can therefore be unreliable (see Gouriéroux and Monfort (2014)). The empirical evidence from the application of multivariate models to the analysis of time series displaying non-Gaussian errors suggests that the estimation of causal AR models using the Gaussian maximum likelihood leads to incorrect inferences; standard errors and test results which are incorrect Gouriéroux and Jasiak (2018).

If the observed time-series is strictly stationary, noncausal and non-Gaussian

then we are able to distinguish between the two-sided (those including the past present and future errors) and one-sided (those including only the current and lagged errors) moving average representations written in terms of i.i.d. errors.

The method of estimation applied in Davis and Song (2020) is the Approximate Maximum Likelihood (AML). This method requires the assumption that the error distribution can be specified parametrically. If the distribution is correctly specified then the estimates stemming from the approximate likelihood method are consistent. If, on the other hand, the distribution is not correctly specified then the approximate likelihood method is unreliable.

This Section presents the causal noncausal VAR model, recalls the representation in terms of purely causal and noncausal components and summarizes the results on the GCov estimator.

1.2 The Causal Noncausal VAR(1) Model

Let us consider a strictly stationary n-dimensional mixed VAR(1) process:

$$Y_t = \Phi Y_{t-1} + \varepsilon_t, \tag{1.2.1}$$

where (ε_t) is a strong multivariate non-Gaussian white noise of dimension n, and Φ is an $n \times n$ matrix. Gouriéroux and Jasiak (2017) assume that (ε_t) is square integrable with zero mean $E(\varepsilon_t) = 0$, and variance-covariance matrix $V(\varepsilon_t) = \Sigma$. [2]

[2]The assumption of square integrability is required for second order identification and also

The eigenvalues of matrix Φ are assumed to be of modulus different from 1 as this ensures that a unique, strictly stationary solution to recursive equation (1.2.1) exists.

Since (ε_t) is not assumed independent of the lagged values of the process $Y_{t-1}, Y_{t-2}...$, the process ϵ_t cannot be interpreted as an innovation.

1.2.1 Representation Theorem GJ (2017)

Gouriéroux and Jasiak (2017) introduce a representation theorem for the causal, noncausal mixed processes that distinguishes their purely causal and noncausal components.

The demonstration in Gouriéroux and Jasiak (2017) shows that in a VAR(1) model if n_1 (resp. $n_2 = n - n_1$) represents the number of eigenvalues of Φ whose modulus is strictly less than 1 (resp. strictly larger than 1), then there exists an invertible $n \times n$ matrix A, and two square matrices: J_1 with dimension $n_1 \times n_1$ and J_2 with dimension $n_2 \times n_2$. The eigenvalues of J_1 (resp. J_2) with their modulus strictly less than 1 (resp. larger than 1) are such that :

$$Y_t = A_1 Y^*_{1,t} + A_2 Y^*_{2,t}, \tag{1.2.2}$$

$$Y^*_{1,t} = J_1 Y^*_{1,t-1} + \varepsilon^*_{1,t}, Y^*_{2,t} = J_2 Y^*_{2,t-1} + \varepsilon^*_{2,t}, \tag{1.2.3}$$

$$\varepsilon^*_{1,t} = A^1 \varepsilon_t,\ \varepsilon^*_{2,t} = A^2 \varepsilon_t, \tag{1.2.4}$$

facilitates the derivation of asymptotic properties of the estimators.

where A_1, A_2 (resp A^1, A^2) are the blocks in the decomposition of matrix A as :

$$A = (A_1, A_2) \text{ [resp. in the decomposition of } A^{-1} \text{ as } A^{-1} = \begin{pmatrix} A^1 \\ A^2 \end{pmatrix}].$$

The matrices J_1 and J_2 are derived from the real Jordan canonical form of Φ such that

$$\Phi = A \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix} A^{-1},$$

where A contains the eigenvectors of Φ as its columns.

By premultiplying both sides of equation (1.2.1) and (1.2.4) by matrix A^{-1} we can decompose Y_t into causal and noncausal components as follows :

$$Y_t^* = \begin{pmatrix} Y_{1,t}^* \\ Y_{2,t}^* \end{pmatrix} \equiv A^{-1} Y_t, \varepsilon_t^* = \begin{pmatrix} \varepsilon_{1,t}^* \\ \varepsilon_{2,t}^* \end{pmatrix} \equiv A^{-1} \varepsilon_t.$$

We get :

$$Y_t^* = \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix} Y_{t-1}^* + \varepsilon_t^*,$$

$$Y_{j,t}^* = J_j Y_{j,t-1}^* + \varepsilon_{j,t}^*, j = 1, 2,$$

In addition, equation $Y_t = AY_t^*$, is equivalent to :

$$Y_t = A_1 Y^*_{1,t} + A_2 Y^*_{2,t},$$

which is the decomposition given in equation (1.2.2).

Given that the eigenvalues of J_1 have modulus strictly less than 1, the recursive equation below is causal :

$$Y^*_{1,t} = J_1 Y^*_{1,t-1} + \varepsilon^*_{1,t},$$

and recursive backward substitutions can be used derive the causal one-sided moving average representation of $Y^*_{1,t}$ given by the expression below (where L is the lag operator) :

$$Y^*_{1,t} = \Sigma_{h=0}^{\infty} J_1^h \varepsilon^*_{1,t-h} = (Id - J_1 L)^{-1} \varepsilon^*_{1,t}, \tag{1.2.5}$$

and where

$$(Id - J_1 L)^{-1} \equiv \Sigma_{h=0}^{\infty} J_1^h L^h. \tag{1.2.6}$$

The second recursive equation is noncausal : $Y^*_{2,t} = J_2 Y^*_{2,t-1} + \varepsilon^*_{2,t}$ can, using recursive substitution, be written thus :

$$Y_{2,t}^{*} = J_2^{-1} Y_{2,t+1}^{*} - J_2^{-1} \varepsilon_{2,t+1}^{*} = (Id - J_2 L)^{-1} \varepsilon_{2,t}^{*}, \tag{1.2.7}$$

where :

$$(Id - J_2 L)^{-1} \equiv -\Sigma_{h=1}^{\infty} J_2^{-h} L^{-h}. \tag{1.2.8}$$

It follows that:

1. There exists a strong (i.i.d) two-sided moving average representation of the solution of (1.2.1). Processes $(Y_{1,t}^{*})$ and $(Y_{2,t}^{*})$ are purely causal and noncausal processes, respectively. They can be interpreted as the causal and noncausal components of process (Y_t);
2. These components are deterministic functions of (Y_t) since : $Y_{j,t}^{*} = A^j Y_t, j = 1, 2$.

In the case of a VAR(1) we can always factor the matrix lag polynomial into a product of a causal matrix lag polynomial and a noncausal matrix lag polynomial:

$$(Id - \Phi L)^{-1} = \Theta_1(L)\Theta_2(L^{-1}),$$

where the roots of $det\, \Theta_1(z) = 0$ and $det\, \Theta_2(z) = 0$ are of modulus greater than 1.

It is possible to represent $(Id - \Phi L)^{-1}$ as a product of a causal matrix lag polynomial and a noncausal matrix lag polynomial because the process is assumed to

be a VAR(1) process. However, such a decomposition does not, in general, exist when the autoregressive order of the process is strictly greater than 1 unless it is written as an augmented VAR(1) process in special cases Lanne and Saikkonen (2013).

This means that modelling a VAR(p) model where $p > 1$ is not always possible when representing $(Id - \Phi L)^{-1}$ as a product of a causal matrix lag polynomial and a noncausal matrix lag polynomial.

1.2.2 VAR(1): Identification of the Causal and Noncausal Dimensions

Gouriéroux and Jasiak (2017) show that second-order identification (the identification of a multivariate VAR(1) process from moments up to order two) is, in fact, almost feasible. This finding contradicts the generally held belief that second-order identification of non-Gaussian linear processes is not viable. However, in general for $p \geq 1$ this is not the case and the higher order moments are required for identification.

Let n_1 and n_2 denote the causal and noncausal dimensions respectively. A mixed model VAR(1) with characteristics $[(n_1, A_1, J_1), (n_2, A_2, J_2)]$ cannot be distinguished at second order from a mixed VAR(1) model with characteristics $[(n_2, A_2, J_2^{-1}), (n_1, A_1, J_1^{-1})]$. In this case, one is able to substitute the eigenvalues in place of their reciprocals without changing the representation, provided this is done for all causal (resp. noncausal) components simultaneously.

For a mixed VAR(1) process of n components, second-order identification

the product $n_1 \cdot n_2$ is feasible. If the observed VAR process is of dimension n, the researcher is able to identify the pairs of representations corresponding to causal noncausal dimensions at second-order : $\{(n,0),(0,n)\},\{(n-1,1),(1,n-1)\},\ldots,$ and the associated causal noncausal directions and dynamics. However, one cannot distinguish the causal noncausal dimensions within each pair without using nonlinear autocorrelations.

In other cases, higher order moments are needed to identify the dynamics, i.e. the expectations of nonlinear functions of the process.

1.2.3 Bivariate VAR(1) - Interpretation

When matrix Φ is triangular, then depending on the eigenvalues, all components of y_t do not always contribute to both the explosive (i.e. noncausal) and regular (causal) dynamics.

Suppose that matrix Φ has the following Jordan decomposition:

$$\Phi = AJA^{-1}$$

where J is the 2 by 2 matrix of eigenvalues A is the 2 by 2 matrix with columns containing the eigenvectors of Φ. Suppose also that $\phi_{12} = 0$ or $\phi_{21} = 0$, so that matrix Φ is upper or lower triangular. It is known that for any $n \times n$ triangular matrix the following properties hold:

1) The eigenvalues of an upper or lower triangular matrix are the diagonal

entries of the matrix

2) For any triangular matrix, a vector with all elements zero, except the first one is an eigenvector. There is a second eigenvector with all elements zero, except the first two, etc.

Therefore, a triangular 2 by 2 matrix Φ has a triangular matrix A, with a triangular inverse A^{-1}. It follows that the past values of one component of y_t do not contribute to either the explosive dynamics y^*_{2t}, or the regular dynamics $y^*_{1,t}$.

Let the matrix J be written as

$$J = \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix}, \text{ where } J_1 < 1 < J_2. \text{ Then matrix } A \text{ has entries}$$

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

and its inverse is

$$A^{-1} = \begin{pmatrix} a^{11} & a^{12} \\ a^{21} & a^{22} \end{pmatrix}$$

Accordingly, we have row vectors $A^1 = [a^{11}\ a^{12}]$ and $A^2 = [a^{21}\ a^{22}]$ corresponding to the latent components y^*_{1t} and y^*_{2t} with regular and explosive dynamics, respectively.

Example 1: Upper triangular Φ

Suppose the element $\phi_{21} = 0$ in matrix

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$$

which makes an upper triangular matrix

$$\Phi_U = \begin{pmatrix} \phi_{11} & \phi_{12} \\ 0 & \phi_{22} \end{pmatrix}$$

If $J_1 = \phi_{11}$, $J_2 = \phi_{22}$ so that $J_2 > J_1$, we get $a^{21} = 0$

$$A^{-1} = \begin{pmatrix} a^{11} & a^{12} \\ 0 & a^{22} \end{pmatrix}$$

Then both y_{1t} and y_{2t} contribute to the regular component y^*_{1t}, but process y_{1t} does not contribute to the explosive $y^*_{2t} = y_{2t}$:

$$y^*_{1,t} = a^{11}y_1 + a^{12}y_2 = \sum_{j=0}^{+\infty} \lambda_1 \epsilon^*_{1,t-j}, \tag{1.2.9}$$

with $\epsilon^*_{1,t}$ as the causal innovation $\epsilon^*_{1,t} = a^{11}\epsilon_{1,t} + a^{12}\epsilon_{2,t}$, and

$$y^*_{2,t} = a^{22}y_{2,t} = -\sum_{j=0}^{+\infty} [\lambda_2^{-j-1}\, a^{22}\epsilon_{2,t+j+1}]. \tag{1.2.10}$$

The noncausal innovation $\epsilon^*_{2,t} = a^{22}\epsilon_{2,t}$ is a function of ϵ_2 only. We observe that $\underline{y_{1,T}}$ affects only the error term associated with $y_{1,T+1}$, i.e. the non-explosive error.

If $J_1 = \phi_{22}$, $J_2 = \phi_{11}$ so that $J_2 > J_1$, we get $a^{11} = 0$

$$A^{-1} = \begin{pmatrix} 0 & a^{12} \\ a^{21} & a^{22} \end{pmatrix}$$

In this case process y_{1t} is explosive and does not contribute to the regular component $y^*_{1t} = y_{2t}$, while both y_{1t} and y_{2t} contribute to the explosive component y^*_{2t}.

Example 2: Lower triangular Φ

Suppose the element $\phi_{12} = 0$ in matrix

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$$

which makes a lower triangular matrix

$$\Phi_L = \begin{pmatrix} \phi_{11} & 0 \\ \phi_{21} & \phi_{22} \end{pmatrix}$$

Then, if $J_1 = \phi_{11}$, $J_2 = \phi_{22}$ so that $J_2 > J_1$, we get $a^{12} = 0$

$$A^{-1} = \begin{pmatrix} a^{11} & 0 \\ a^{21} & a^{22} \end{pmatrix}$$

Process y_{2t} does not contribute to the regular component $y^*_{1t} = y_{1t}$, but both processes contribute to the explosive y^*_{2t}.

$$y^*_{1,t} = a^{11} y_1 = \sum_{j=0}^{+\infty} \lambda_1 \epsilon^*_{1,t-j}, \tag{1.2.11}$$

where $\epsilon^*_{1,t}$ is the causal innovation $\epsilon^*_{1,t} = a^{11}\epsilon_{1,t}$ and a function of $\epsilon_{1,t}$ only. The explosive component is

$$y^*_{2,t} = a^{21} y_{1,t} + a^{22} y_{2,t} = -\sum_{j=0}^{+\infty} [\lambda_2^{-j-1} \left(a^{21}\epsilon_{1,t+j+1} + a^{22}\epsilon_{2,t+j+1}\right)]. \tag{1.2.12}$$

If $J_1 = \phi_{22}$, $J_2 = \phi_{11}$ so that $J_2 > J_1$, we get $a^{22} = 0$

$$A^{-1} = \begin{pmatrix} a^{11} & a^{12} \\ a^{21} & 0 \end{pmatrix}$$

In this case process y_{2t} does not contribute to the explosive component $y^*_{2t} = y_{1t}$, while both y_{1t} and y_{2t} contribute to the regular component y^*_{1t}.

Independence

Note that the independence between y_1 and y_2 arises when $\phi_{12} = \phi_{21} = 0$ and the joint density of errors $g(\epsilon_{1,t}, \epsilon_{2,t}) = g_1(\epsilon_{1,t})g_2(\epsilon_{2,t}), \forall t$.

1.2.4 VAR(p) model

Consider the VAR(p) process:

$$Y_t = \Phi_1 Y_{t-1} + ... + \Phi_p Y_{t-p} + \epsilon_t \,, \tag{1.2.13}$$

where (ϵ_t) is a sequence of independent and identically distributed (i.i.d.) random vectors of dimension n with variance-covariance matrix Σ. We can write this model as a VAR(1) model for X_t where X_t is obtained by stacking the current and lagged values of Y_t

$$X_t = (Y_t', Y_{t-1}', ..., Y_{t-p+1}')' \,, \tag{1.2.14}$$

and so we have

$$X_t = \Psi X_{t-1} + u_t. \tag{1.2.15}$$

The autoregressive polynomial Ψ can be written as

$$\Psi = \begin{bmatrix} \Phi_1 & ... & ... & \Phi_p \\ Id & 0 & ... & 0 \\ 0 & Id & ... & 0 \\ 0 & ... & Id & 0 \end{bmatrix}. \tag{1.2.16}$$

and

$$u_t = \begin{bmatrix} \epsilon_{1,t} \\ \vdots \\ \epsilon_{1,n} \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{1.2.17}$$

By the representation theorem given in section (2.2) Ψ can also be written in Jordan canonical form:

$$\Psi = B \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix} B^{-1}. \tag{1.2.18}$$

Similar to Y_t in the VAR(1) case, X_t can be written as the sum of causal and noncausal components as follows

$$X_t = B_1 X_{1,t}^* + B_2 X_{2,t}^*,$$

where

$$X_{1,t}^* = J_1 X_{1,t-1}^* + u_{1,t}^*,$$

$$X_{2,t}^* = J_2 X_{2,t-1}^* + u_{2,t}^*,$$

and

$$X_{1,t}^* = B^1 X_t,$$

$$X_{2,t}^* = B^2 X_t.$$

As in the VAR(1) case, the causal and noncausal errors are deterministic functions of the of the process u_t,

$$u_{1,t}^* = B^1 u_t,$$

$$u_{2,t}^* = B^2 u_t.$$

Errors $u_{1,t}^*$ and $u_{2,t}^*$ satisfy $n(p-1)$ linearly independent and deterministic relationships since they both depend on ϵ_t and dim $u_{1,t}^*$ + dim $u_{2,t}^* = n_1 + n_2 = np$ and np is greater than dim $\epsilon_t = n$ whenever $p > 1$.

Extending results other than those previously obtained for the mixed VAR(1) process to a VAR(p) must be done with caution. For example, the causal and noncausal matrix polynomials $\Phi(L)$ and $\Phi(L^{-1})$ respectively cannot be decomposed into a multiplicative form such as $\Theta_1^*(L)\Theta_2^*(L^{-1})$ in the general VAR(p) case (where $p > 1$).

1.2.5 Estimation Using Nonlinear Covariances

The semi-parametric estimation method applicable to mixed causal noncausal multivariate processes is called the Generalized Covariance Estimator and was introduced by Gouriéroux and Jasiak (2017).

It follows from the nonlinear identification result in Ming-Chung and Kung-Sik (2007) that there exist nonlinear covariance based conditions that can be used to identify causal and noncausal components of a given series provided the error terms ε_t are serially independent. The nonlinear covariance based conditions, for example, could be the covariances between nonlinear transforms of the error terms defined for a given set of functions as:

$$c_{j,k}(h,\Phi) = Cov[a_j(Y_t - \Phi Y_{t-1}), a_k(Y_t - \Phi Y_{t-h-1})],$$

$$j,k = 1,...,K,$$

$$h = 1,..,H,$$

for a given set of functions $a_k, k = 1,...K$.

Let us denote by $\Psi_l(Y_t,\phi)$, $l = 1,...,KH$, function $a_k(Y_{t-h} - \Phi Y_{t-h-1})$, where $\phi = vec\Phi\prime$. For each covariance $c_{kl} = Cov[\Psi_k(Y_t,\phi), \Psi_l(Y_t,\phi)], k,l = 1,...,KH$, we can write its sample counterpart:

$$\hat{\gamma}_{k,l,T} = \widehat{Cov}[\Psi_k(Y_t,\phi), \Psi_l(Y_t,\phi)],$$

$$k, l = 1, ..., KH.$$

We can then define a covariance estimator as follows:

The Covariance estimator $\tilde{\phi_T}$ of $\phi = vec\Phi\prime$ minimizes the following objective function:

$$\tilde{\phi}_T = \hat{\gamma}_T\prime(\phi)\Omega\hat{\gamma}_T(\phi),$$

with respect to ϕ where $\hat{\gamma_T}(\phi)$ denotes the vector obtained by stacking $\hat{\gamma}_{k,l,T}(\phi)$ and Ω is a $(KH \times KH)$ positive definite weighting matrix.

The selection of the function a_j determines the semi-parametric efficiency bound of the covariance estimator and the asymptotic efficiency of a Covariance estimator based on a given set of covariances depends on the matrix of weights Ω the estimator, in general, asymptotically semi-parametrically efficient.

One can then search for optimal weights Ω that ensure asymptotic semi-parametric efficiency and the associated estimator is then called the Generalized Covariance (GCov) estimator.

The definition of the Generalized Covariance estimator is similar to the definition of a Generalized Method of Moments (GMM) estimator since by analogy, we can obtain a consistent covariance estimator with a simple weighting scheme such as an identity matrix (although that first step estimator may not be fully semi-parametrically efficient). The differences lie in 1) the use of the central moments only in the GCov approach and 2) reduced dimension of the objective function to be minimized.

1.2.6 Estimation of Φ_j by The GCov

The restriction of zero covariance between the causal and noncausal components of the VAR process for negative lags of h, based only on second-order moments does not allow us to identify the mixed models $(n_1, n-n_1)$ and $(n-n_1, n_1)$. Provided the error terms ε_t are not serially dependent, nonlinear covariance-based conditions are necessary.

The choice of nonlinear covariances is a problem similar to the problem that arises when choosing instruments in a GMM setting. One can choose a combination of quadratic and linear transformations to capture the absence of leverage effect at lag $h, h \leq 0$ for example, but several other options exist and will depend on the application at hand. Note that the GMM estimator is not available for the models with noncausal components. The error term cannot be interpreted as an innovation and no instruments are available, see Lanne and Saikkonen (2011b) and Lanne and Saikkonen (2011a) for further discussion on the issue.

We consider the functions $a_k, k = 1, ..., K$, and sample autocorrelations $\hat{\rho}_{j,k}(h, \Phi) = Corr[a_j(Y_t - \Phi Y_{t-1}), a_k(Y_{t-h} - \Phi Y_{t-h-1})]$, $\Gamma(0)^{-1}\Gamma(1)$. Then the GCov estimator can be represented as a weighted covariance estimator:

$$\hat{\Phi}_T = \underset{\Phi}{\operatorname{argmin}} \sum_{j=1}^{K} \sum_{k=1}^{K} [\sum_{h=1}^{H} \hat{\rho}_{j,k}^2(h, \Phi)] \tag{1.2.19}$$

where H is the highest selected lag and the theoretical auto correlations $\rho_{j,k}$ are replaced by their sample counterparts $\hat{\rho}_{j,k}$, seeGouriéroux and Jasiak (2017).

The efficiency of the GCov estimator depends on the choice of functions $a_k, k = 1, ..., K$ and the maximum lag H.

1.2.7 Asymptotic Properties of the GCov Estimator

Gouriéroux and Jasiak (2017) show that GCov estimator $\hat{\phi}$ of $vec(\Phi')$ is consistent and asymptotically normal with asymptotic variance given by:

$$(D'\Sigma^{-1}D)^{-1} = V_{asy}[\sqrt{T}(\hat{\phi_T} - \phi)].$$

The rows of matrix D are:

$$D_{k,l} = -\frac{\partial}{\partial \phi'}\widehat{Cov}[\Psi_k(Y_t, \phi), \Psi_l(Y_t, \phi)].$$

The elements of matrix Σ are:

$$\sigma_{(k,l),(k',l')} = Cov_{asy}(\sqrt{T}\widehat{Cov}[\Psi_k, \Psi_l], \sqrt{T}\widehat{Cov}[\Psi_{k'}, \Psi_{l'}])$$

where $\Psi_i = \Psi_i(Y_t, \phi) = a_k(Y_{t-h} - \Phi Y_{t-h-1})$ for $i = (k, l, k', l')$.

1.2.8 Estimation Procedure, Identification and Estimation

In practice, the estimation of a (bivariate) VAR(p) can be accomplished using the GCov estimator along the following steps:

1. Estimate $\Phi_1, ..., \Phi_p$ for a given autoregressive order p using the GCov estimator.

This can be done using linear and nonlinear functions of $\epsilon_t(\phi) = Y_t - \Phi_1 Y_{t-1} - ... - \Phi_p Y_{t-p}$.

2. Using the p estimated autoregressive coefficients $\widehat{\Phi_1}, ..., \widehat{\Phi_p}$ compute $\widehat{\Psi}$ and derive the Jordan canonical form of $\widehat{\Psi}$ The decomposition will yield $\widehat{n_1}$ and $\widehat{J_i}$ and $\widehat{B_i}$ for i = {1,2}.

3. Compute the residuals $\widehat{\epsilon_t} = Y_t - \widehat{\Phi_1} Y_{t-1} - ... - \widehat{\Phi_p} Y_{t-p}$ and compute their nonlinear autocorrelation function. If the residual autocorrelations are still significant at some lags, then re-estimate the model increasing p and repeat until the residual autocorrelations are no longer significant.

Finding A^1 and A^2 when the number of noncausal directions is known shows that if n_1, n_2 are known the vectors or matrices A^1 and A^2 can be estimated directly.

1.3 Empirical Analysis of Four Large Cap Cryptocurrencies

In 2018 Bitcoin (BTC) was worth over half the market capitalization of all cryptocurrencies and Ethereum (ETH) had the second largest market capitalization. Ripple (XRP)and Stellar (XLM) were not leaders in terms of market capitalization. We believe these two pairs (BTC and on the one hand and ETH and XRP and XLM on the other) have characteristics in common - Bitcoin and Ethereum are market capitalization leaders while Ripple and Stellar serve similar purposes and share similar underlying technological features. We will search to detect their comovements.

In this chapter we endeavour to contribute to knowledge of cryptocurrencies markets by examining the pairs Bitcoin and Ethereum and Ripple and Stellar with the aim of modelling comovements between each of these pairs of cryptocurrencies. In addition, we model all four cryptocurrencies as a single multivariate process. In particular, we aim to ascertain the nature of their respective processes and to exploit noncausality for forecasting purposes in the presence of bubble phenomena. The presence of bubbles suggest noncausal dynamics, especially due to the speculative character of these digital assets resulting in speculative bubbles.

We decompose the process into casual and noncausal components. The purpose of this decomposition is to pursue the modelling of bubble phenomena (i.e. locally explosive behaviour in strictly stationary time series) often found in financial time series. Modelling the noncausal components in particular is what allows us to find a common bubble component multivariate series. When series share a common noncausal component we can potentially forecast these series more accurately as a joint process by taking the common bubble co-movements into account than we would be able to otherwise.

1.3.1 Cyptocurrencies

Let us consider the US dollar exchange rates of the following cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Ripple (XRP) and Stellar (XLM) observed at daily frequency obtained from the Bitfinex exchange (www.bitfinex.com). The data display short lived local trends which suggest the presence of noncausal dynamics.

1.3.2 Bitcoin (BTC) and Ethereum (ETH)

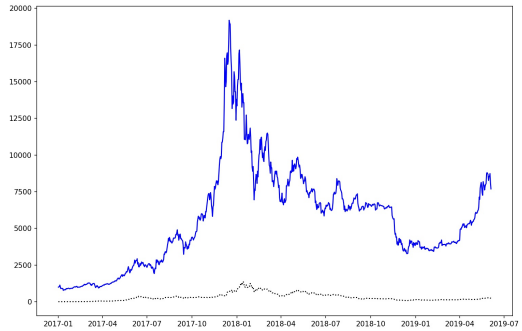
The sample of Bitcoin and Ethereum exchange rates against the US Dollar (USD) (BTC and ETH hereafter) consists of $T = 885$ observations on daily closing rates collected between January, 01 2017 and June, 04 2019.

Figure 1.1a displays the daily BTC/USD and ETH/USD exchange rates over the entire sampling period of 885 days. Both Bitcoin and Ethereum experienced a large increase in value relative to the US dollar since early 2017 but had, as of 2018, lost a large proportion of their respective US dollar values since their peak in late 2017. In addition, both exchange rates show evidence of bubble phenomena, i.e. explosive trends characterized by periods of explosive increases in level followed by rapid decreases in level or vice versa.

Figure 1.1b displays BTC/USD and ETH/USD exchange rates with medians subtracted. The BTC/USD exchange rate is divided by a factor of ten for comparison and further modelling and, along with the original ETH/USD exchange rate, it is hereafter referred to as the adjusted series.

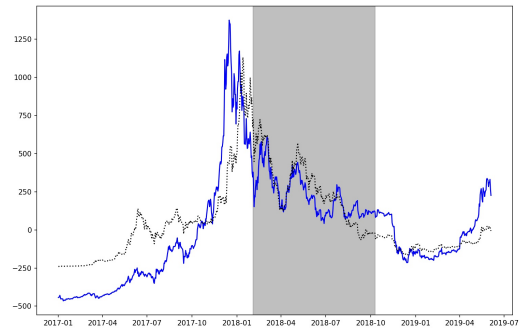
In Figure 1.1c we show in grey a sub-sample of $T = 250$ observations over the period February, 02 2018 and October, 10 2018 selected for further analysis of the series. This sub-sample is shown in Figure 1.1d again to document the comovements between the series and appears to be nonstationary.

We chose this sub-sample wherein $T = 250$ because it displayed many small bubbles as the larger late 2017 bubble is bursting as the price of cryptos decreased. This sub-sample shown in Figures 1.1b and 1.1d is then detrended with Python



(a) 2017-01-01 to 2019-06-04

(b) Adjusted 2017-01-01 to 2019-06-04



(c) Sub-sample 2018-02-03 to 2018-10-10 in grey

(d) Sub-sample 2018-02-03 to 2018-10-10

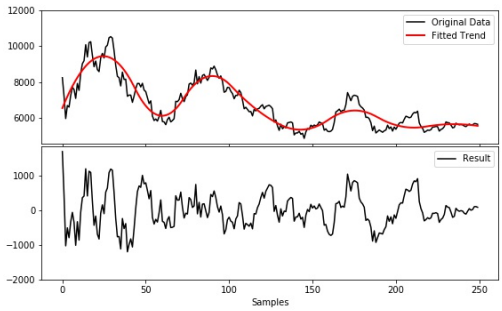
Figure 1.1: BTC/USD and ETH/USD Exchange Rates. BTC/USD solid line, ETH/USD dotted line.

package Obspy [3] using a spline of order 2 with a knot every 30 observations. We detrend the subsample using splines instead of the Hodrick–Prescott filter because the Hodrick–Prescott filter is parametric and destroys the serial dependence pattern Hecq and Voisin (2022). Moreover, one cannot predict it or extrapolate from data detrended using this method. As such we use splines because they are non-parametric. The original and detrended data for BTC and ETH exchange rates can be seen in Figures 1.2. Figure 1.2c displays the detrended, adjusted BTC/USD exchange rate as a solid line and the detrended ETH/USD exchange rate as a dotted line.

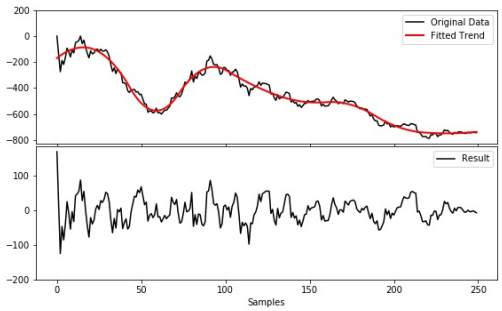
In addition, the ACF in Figure 1.3 of the detrended data is consistent with a stationary process with finite range of serial dependence. The shaded region in Figure 1.3 is the asymptotically valid confidence interval at 95%. We observe that the first four auto and cross correlations are statistically significant.

The unadjusted series are non-normally distributed, as evidenced by their excess kurtosis shown in Summary Sample Statistics in Table A.1.

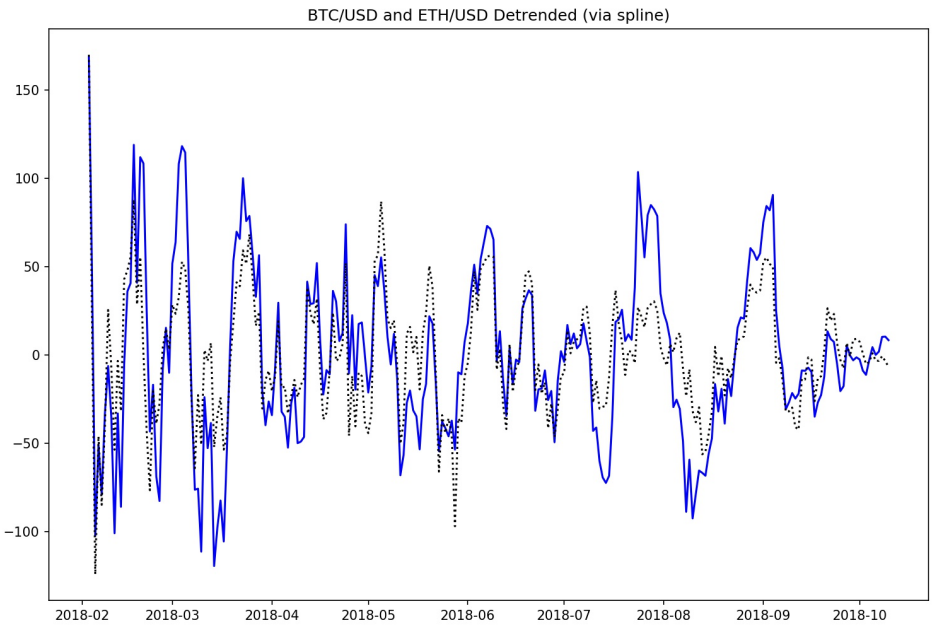
[3] Specifically, from Obspy we import the spline package from obspy.signal.detrend



(a) BTC (Adjusted) Detrended by Spline



(b) ETH (Adjusted) Detrended by Spline



(c) BTC/USD: solid line, ETH/USD: dotted line

Figure 1.2: BTC and ETH (Adjusted) Detrended by Spline

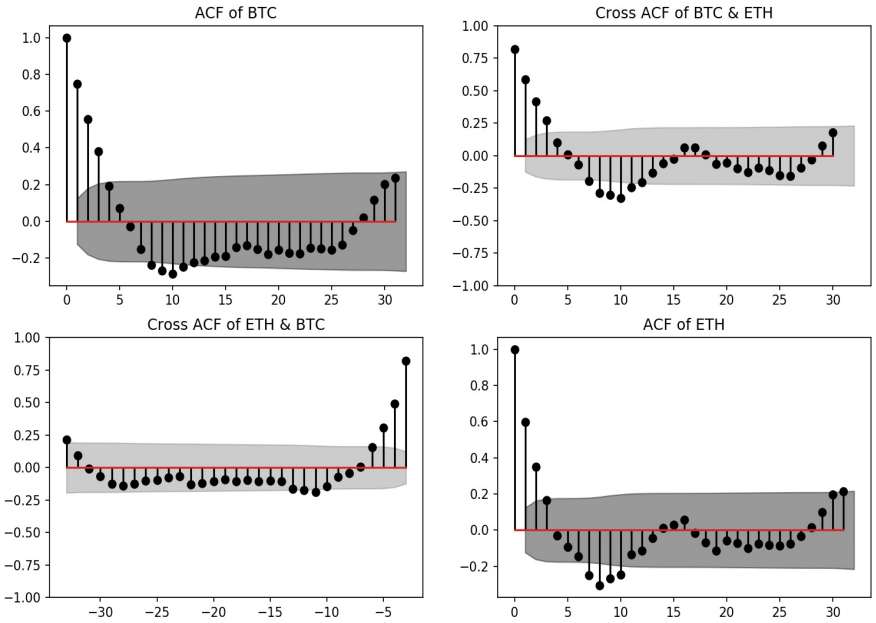


Figure 1.3: ACF of Detrended BTC and ETH VAR(1)

By applying the Augmented Dickey-Fuller test to the detrended data we find that the resulting process is stationary as all p-values are near zero (for models with and without a constant, constant, with a linear trend and with both linear and quadratic trends) as shown in Table 1.1. The detrended data is not normally distributed, with non-zero excess kurtosis and skewness as documented in Table 1.2 below.

In addition, the Hill Plots for BTC and ETH, B.8a and B.8b respectively in Supplementary Graphs show tails with exponent below 2.

	No Constant	Constant	Linear Trend	Quad and Linear Trend
BTC	0.0	0.0	0.0	0.0
ETH	0.0	0.0	0.0	0.0

Table 1.1: BTC and ETH ADF P-Values for Detrended Series

	Excess Kurtosis	Skew
BTC	0.245	0.31
ETH	2.15	0.338

Table 1.2: BTC and ETH Excess Kurtosis and Skewness for Detrended Series

1.3.3 VAR(1) For BTC and ETH

By setting H equal to 11 in the objective function (1.2.19):

$$\rho(\Phi, y, ...) = \sum_{j=1}^{K}\sum_{k=1}^{K}[\sum_{h=1}^{H}\hat{\rho}_{j,k}^{2}(h, \Phi)],$$

using the second moments of the residuals as the nonlinear function and minimizing this function with respect to Φ we obtain the following estimated autoregressive matrix:

$$\hat{\Phi}_{GCOV_{BTC/ETH}} = \begin{bmatrix} 0.12 & 1.18 \\ -0.56 & 2.08 \end{bmatrix}.$$

The eigenvalues for this matrix are 0.55 and 1.6 respectively. These values are consistent with a mixed causal noncausal process. The standard errors for the first row are 0.059, 0.093 respectively and the standard errors for the second row are 0.064 and 0.1 respectively. The coefficients are statistically significant based on the standard Wald test.

Next, we compute the residuals and show the ACF of the residuals in Figure 1.4 and squared residuals in Figure 1.5 of the VAR(1) model. We find that most serial correlation has been removed but there still exists evidence of slight correlation at lags 1 and 2, especially in the squared residuals.

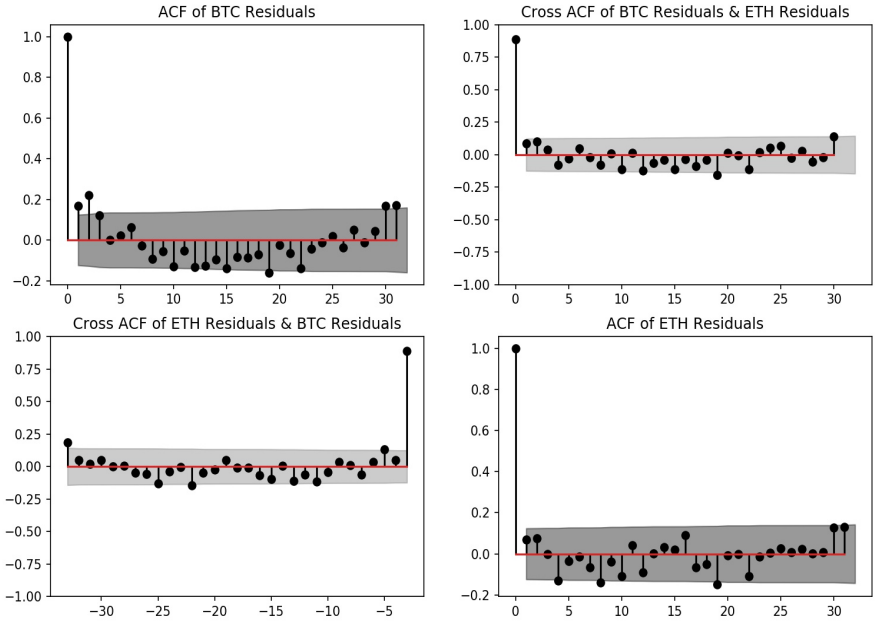


Figure 1.4: ACF of Residuals from VAR(1) for BTC and ETH

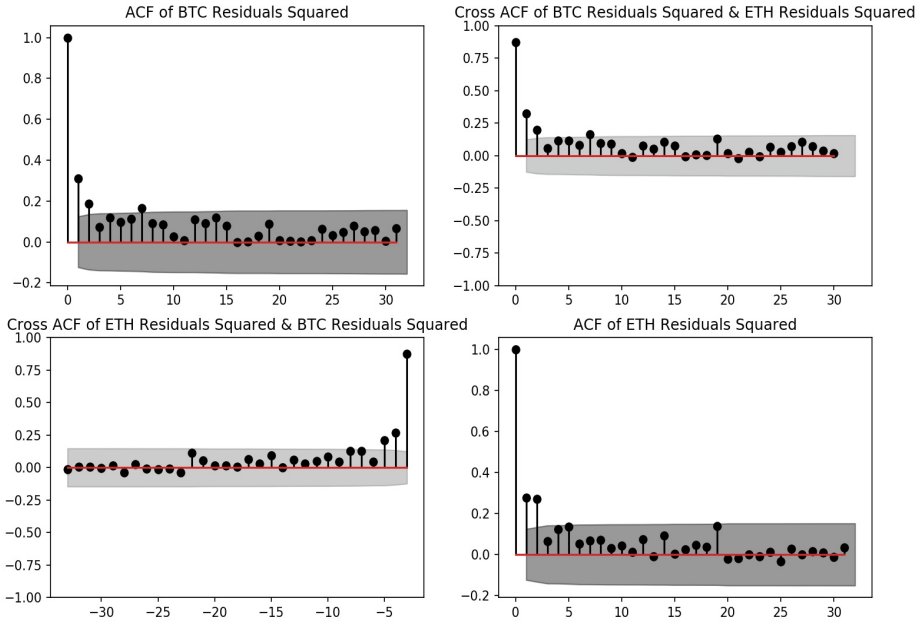


Figure 1.5: ACF of Squared Residuals from VAR(1) for BTC and ETH

Let us now look at the histograms and QQ plots shown in Figures 1.6 and 1.7 respectively of the residuals for the VAR(1) BTC and ETH providing evidence of their non-Gaussian distributions.The residual densities have long left tails. Their densities are difficult to specify parametrically, with departures from normality.

The residual variance covariance matrix from the mixed VAR(1) estimated via the GCov for BTC/ETH is

$$\hat{\Sigma}_{BTC/ETH} = \begin{bmatrix} 1002.37 & 676.4 \\ 676.4 & 650.1 \end{bmatrix}$$

The marginal densities are shown in 1.6. Both densities display nonnormality and long tails.

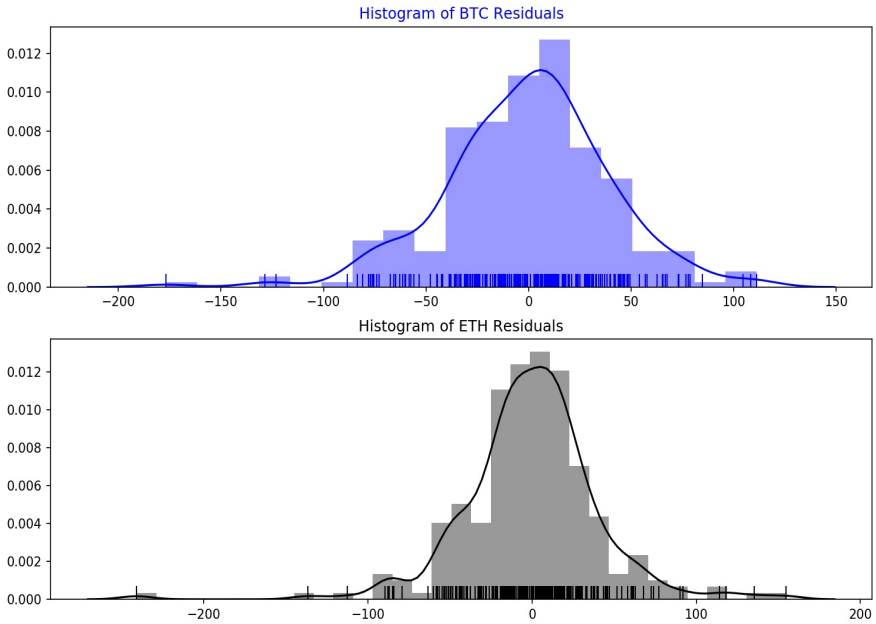
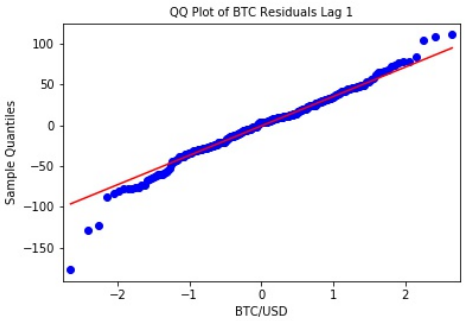
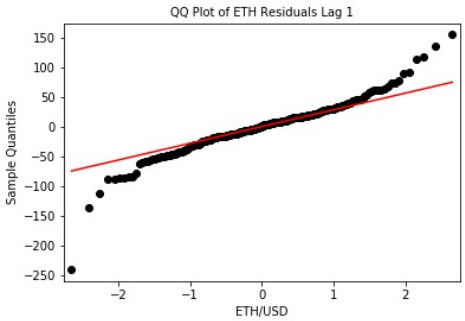


Figure 1.6: Histograms of Residuals from VAR(1) for BTC and ETH



(a) BTC QQ Plot of Residuals VAR(1)



(b) ETH QQ Plot of Residuals VAR(1)

Figure 1.7: BTC and ETH Plot of Residuals VAR(1)

In order to further investigate the normality of the residuals we employ a battery of statistical tests: JB - Jarque-Bera, KS - Komolgorov and Smirnoff, DP - D'Agostino and Pearson, Sh - Shapiro whose test statistics and p-values can be seen in Table 1.3 and excess kurtosis and skewness where 'p' stands for 'p-value'.

	JB	JB-p	KS	KS-p	DP	DP-p	Sh	Sh-p
BTC	35.03	0.0	0.514	0.0	0.97	0.0	19.2	0.0
ETH	312.5	0.0	0.485	0.0	48.04	0.0	0.93	0.0

Table 1.3: BTC and ETH Normality Tests for VAR(1) Residuals

The non-normality is evidenced by the skewness and the excess kurtosis as seen in Table 1.4.

	Excess Kurtosis	Skew
BTC	1.63	-0.42
ETH	5.38	-0.53

Table 1.4: BTC and ETH Excess Kurtosis and Skewness for VAR(1) Residuals

1.3.4 Impulse Responses

Following Davis and Song (2020) the impulse response coefficients for the causal and noncausal components are computed for the bi-variate process (at lag one) as follows:

For a single lag we have

$$Y_t = \sum_{i=-\infty}^{\infty} AF_i A^{-1} Z_{t-i}$$

where Z_{t-i} is the error term and

$$F_i = \begin{cases} \begin{pmatrix} J_1^i & 0 \\ 0 & 0 \end{pmatrix}, & i \geq 0, \\ \begin{pmatrix} 0 & 0 \\ 0 & -J_2^i \end{pmatrix}, & i < -1. \end{cases}$$

then the upper left $(n \times m)$ matrix (the entire matrix in the single lag case) is denoted by M_i and these are the impulse response coefficients of this linear filter. In addition we use the Cholesky decomposition of the covariance matrix $\Sigma_0^* = P_L P_L'$ and derive new (de-correlated) impulse response coefficients

$$\{M_i P_L; i = ..., -1, 0, 1, ...\}.$$

The Cholesky decomposition is used in order to avoid having correlated errors. The impulse response coefficients are plotted in Figure 1.8. For a purely causal process, the coefficients would be zero for all negative lags, but as shown below, the four curves are non-zero for $t < 0$ which implies noncausality.

This the coefficients graphed in cases for which $t > 0$ represent the causal response of the de-correlated noise to a one time positive shock, while the coefficients graphs for which $t < 0$ represent the noncausal response of the de-correlated noise to a one time positive shock.

In Figure 1.8 the first row contains impulse response coefficients for BTC

and the second row contains impulse response coefficients for ETH. The first column (consisting of M_{11} and M_{21}) contains the first component of the transformed noise while the second column (consisting of M_{12} and M_{22}) contains the second component of the transformed noise.

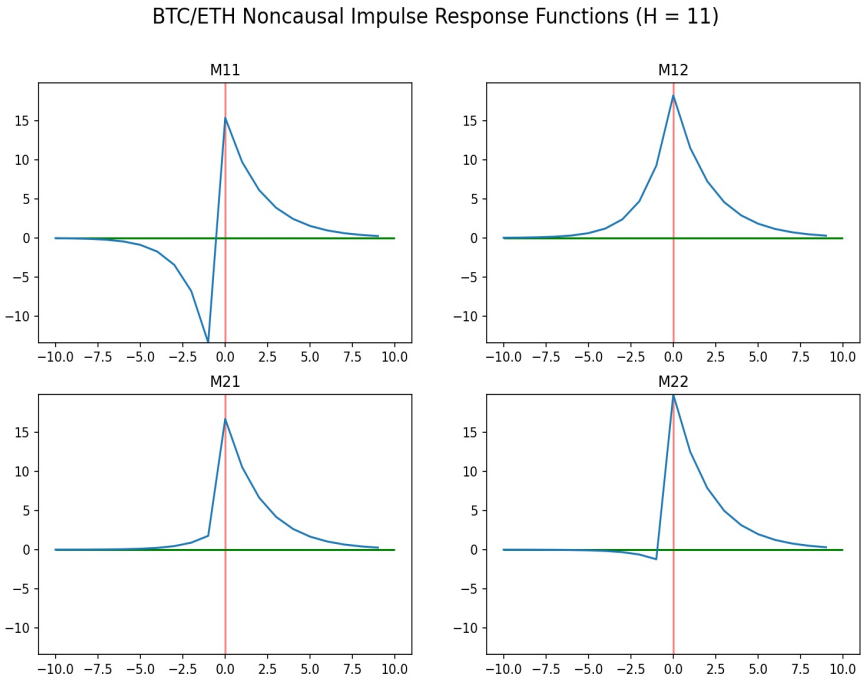


Figure 1.8: Noncausal Impulse Response Coefficients for BTC and ETH VAR(1)

Since the first column of 1.8 contains the impulse response coefficients for BTC and ETH we can see that a large positive one time shock to the first component of the de-correlated noise will cause a large negative effect in the noncausal(i.e. explosive or bubble) part the BTC and a positive effect in the causal (i.e. regular) part of the BTC. In contrast the aforementioned shock will have a small positive impact on the noncausal part of ETH but a large positive effect on the causal part of ETH.

Since the second column of Figure 1.8 contains the impulse response coefficients for BTC and ETH in response to a shock to the second component of the de-correlated noise we can see that large positive one time shock to the second component of the transformed noise causes a positive response in the noncausal part as well as the causal part of BTC. In contrast the shock has minimal impact on the noncausal part of ETH while having a large positive impact on the causal part of ETH.

We further interpret this to mean that a one time positive shock will unambiguously lead to a positive effect on the causal part of BTC since the response is positive for the causal part of BTC through either the first or second component of the de-correlated noise (i.e. the graphs on the first row have positive coefficients for $t > 0$) while the impact of the shock on the noncausal part of BTC is less clear since the causal response is both positive and negative. In contrast, for ETH, the response of the noncausal part to a shock is unambiguously positive but minimal while the response of the causal part is unambiguously positive. Both the causal and noncausal components of the impulse responses dissipate within ten periods. The causal coefficients of the impulse response functions are larger in absolute value than their noncausal coefficients.

Having decomposed the autoregressive coefficient matrix $\hat{\Phi} = \hat{A}\hat{J}\hat{A}^{-1}$ (i.e. into Jordan normal form) we can use blocks of the matrix $\hat{A}^{-1}$ to obtain the causal noncausal components of the process and causal noncausal components of the residuals of both processes.

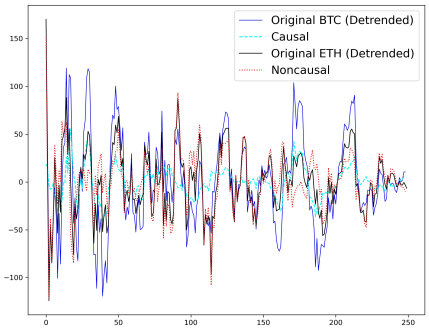
In order to calculate the causal and noncausal components of the process

(which are deterministic functions of the estimated errors) we use blocks of of A^{-1},

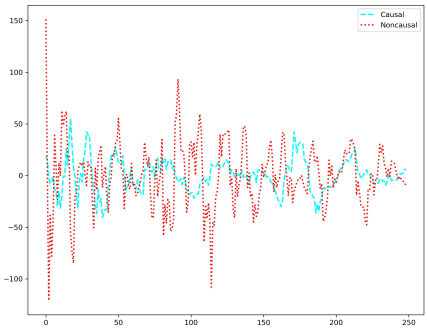
$$\hat{Y}^*_{1,t} = \hat{A}^1\hat{\epsilon}_t,\ \hat{Y}^*_{2,t} = \hat{A}^2\hat{\epsilon}_t,$$

$$\text{where } \hat{A}^{-1} = \begin{pmatrix} \hat{A}^1 \\ \hat{A}^2 \end{pmatrix}.$$

First we show the two series of exchange rates along with their causal and noncausal components representing the 'regular' and 'explosive' common dynamics in Figure1.9a. Next we show only the causal and noncausal components in Figure 1.9b.



(a) Detrended Series and Causal Noncausal Linear Combinations



(b) Causal and Noncausal Linear Combinations Only

Figure 1.9: BTC and ETH Series and Causal and Noncausal Combinations for VAR(1)

Figure 1.9a shows the graph of the causal and noncausal components of the multivariate process for the sub-sample of BTC and ETH without the original series. We see that the causal component of the model is more smooth compared to the noncausal component. Figure 1.9b shows that the noncausal component displays more volatility than the common causal component. This is expected since the noncausal component represents the common bubble or explosive local trend in the

series.

We can also decompose the residuals into causal and noncausal components using the matrix $\hat{A}^{-1}$ as follows:

$$\hat{\epsilon}^*_{1,t} = \hat{A}^1\hat{\epsilon}_t \ , \ \hat{\epsilon}^*_{2,t} = \hat{A}^2\hat{\epsilon}_t$$

The fact that most of the volatility observed in the process can be attributed to the noncausal component, both in the process itself and in the residuals of the process, is consistent with the idea that the noncausal component of the process can be thought of as a "bubble" component in nonlinear processes.

Since the process shows autocorrelation in the squared residuals at lag 1 we increase the number of lags in the VAR model to remove the correlation in the squared residuals. This autocorrelation appears to be removed by lag 3, i.e. when the VAR(3) model is fitted to the time series.

1.3.5 VAR(3) Model for BTC and ETH

By setting H in the objective function (1.2.19)

$$\rho(\Phi, y, ...) = \sum_{j=1}^{K}\sum_{k=1}^{K}[\sum_{h=1}^{H}\hat{\rho}^2_{j,k}(h, \Phi)]$$

equal to 11 and minimizing with respect to Φ we obtain the estimated autoregressive coefficient matrices at lags one, two and three, given below.

LAG 1 MATRIX:

$$\hat{\Phi}_{GCOV_{BTC/ETH}} = \begin{bmatrix} -0.792 & 2.059 \\ -1.268 & 2.06 \end{bmatrix}$$

LAG 2 MATRIX:

$$\hat{\Phi}_{GCOV_{BTC/ETH}} = \begin{bmatrix} 1.717 & -1.439 \\ -1.268 & 2.06 \end{bmatrix}$$

LAG 3 MATRIX:

$$\hat{\Phi}_{GCOV_{BTC/ETH}} = \begin{bmatrix} -0.497 & 0.242 \\ 0.087 & -0.099 \end{bmatrix}$$

AUGMENTED MATRIX:

$$\hat{\Psi}_{GCov_{BTC/ETH}} = \begin{bmatrix} -0.792 & 2.059 & 1.717 & -1.439 & -0.497 & 0.242 \\ -1.268 & 2.06 & -1.268 & 2.06 & 0.087 & -0.099 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

The VAR(p) is stationary if the augmented matrix are inside/outside the unit circle. The eigenvalues for the augmented matrix for the VAR(3) estimated

above are as follows. Two eigenvalues lie outside the unit circle and have the values 1.42 and -1.079.There are two real valued eigenvalues 0.4 and -0.09 and two complex eigenvalues 0.576+0.4i and 0.576+0.4i, both are of modulus 0.7.

The standard errors for the coefficients of the first matrix are 0.098, 0.125 for the first row and 0.09, 0.12 for the second row. The standard errors for the second matrix are 0.10, and 0.09 for the first row and 0.11, 0.1 for the second row. The standard errors for the third matrix are 0.087 and 0.078 for the first row and 0.072, 0.068 for the second row, all coefficients for the first two lags are statistically significant according to the standard Wald test. For the third lag the coefficients pertaining to ETH are not statistically significant while those pertaining to BTC are statistically significant.

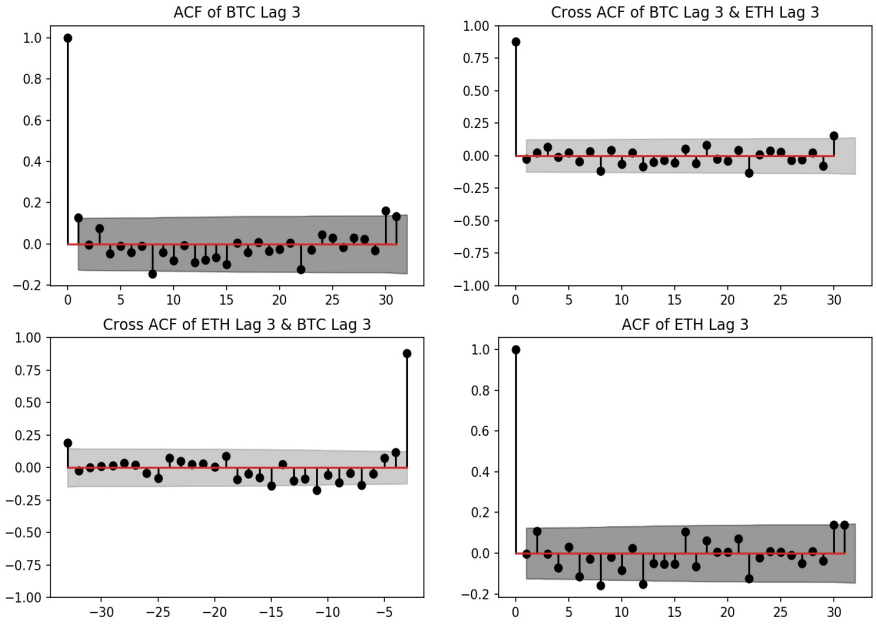


Figure 1.10: ACF of Residuals from VAR(3) for BTC and ETH

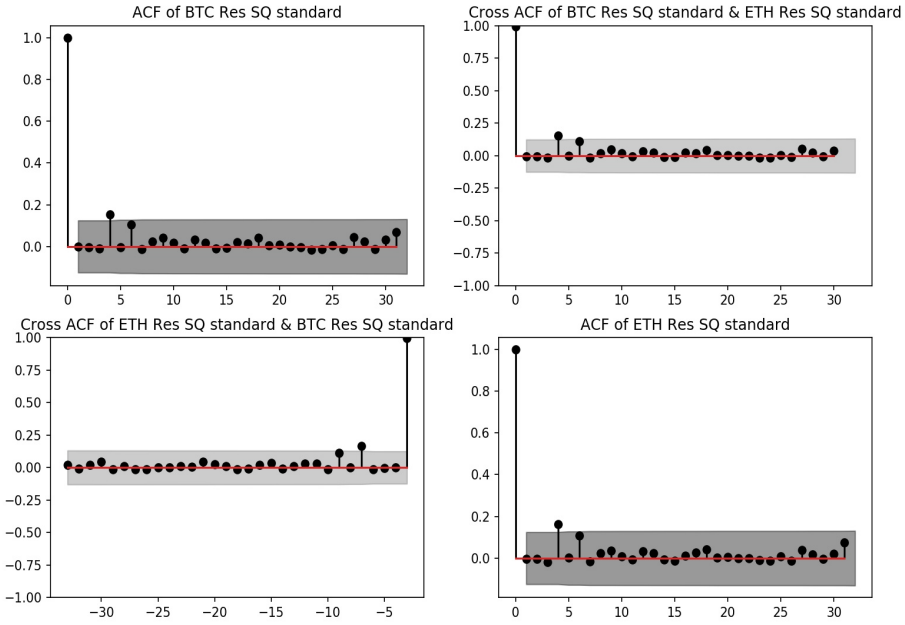


Figure 1.11: ACF of Squared Residuals from VAR(3) for BTC and ETH

Figure 1.10 shows that there remains no statistically significant serial correlation in the residuals. Figure 1.11 also shows that there does not remain much statistically significant serial correlation in the sequential residuals. The VAR(3) models provides a good fit to the data.

The histograms of VAR(3) residuals for BTC and ETH are in Figures B.1 Appendix B. Both series display large tails confirming nonnormality. This is confirmed by the QQ plots in Figures B.5a and B.5b.

The graphs of histograms and QQ plots of the residuals are consistent with the residuals having 'fat tails' and thus indicate non-normality. The following

statistics present further evidence for the non-normality of the residuals.

The Impulse response function for the VAR(3) for BTC/ETH can be found in 1.12.

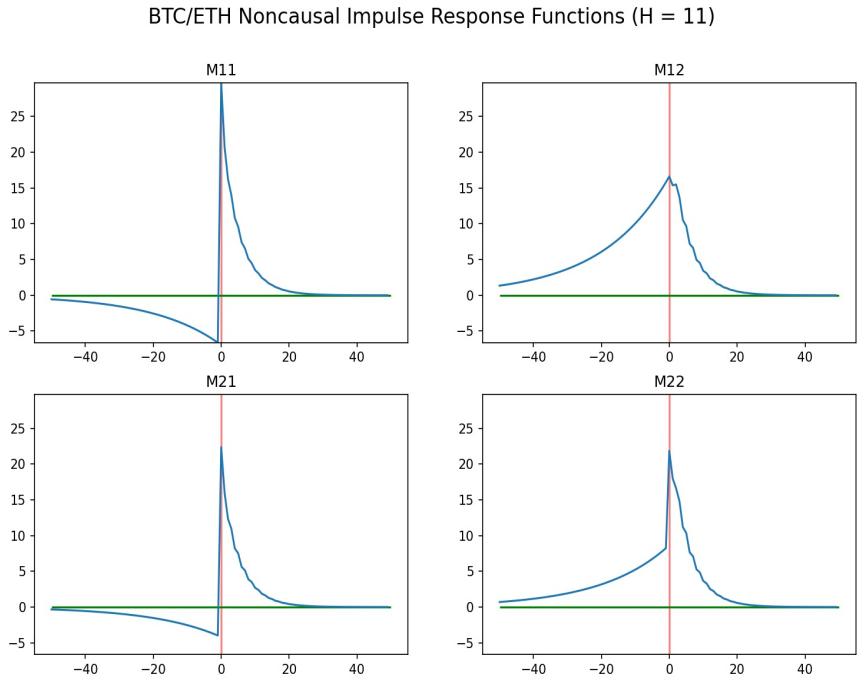


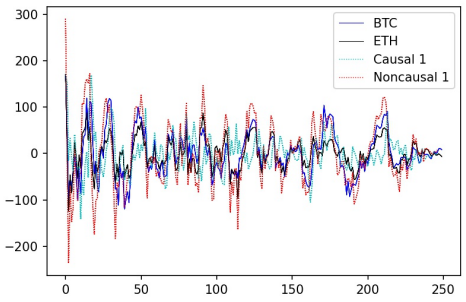
Figure 1.12: Noncausal Impulse Response Coefficients for BTC and ETH VAR(3)

In Figure 1.12 we observe a one time positive shock to the first component of the de-correlated noise results in a negative effect on the noncausal part of BTC and a positive effect on the causal part of BTC. The shock has the same effect on ETH. A positive one time shock in the second component of the de-correlated noise results in a positive effect on the noncausal and noncausal parts for both BTC and ETH.

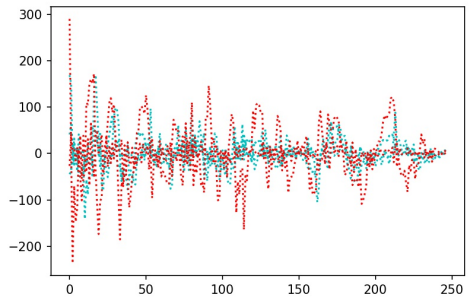
We interpret this to mean that a one time positive shock will unambiguously lead to a positive effect on the causal part of BTC since the response is positive for the causal part of BTC through either the first or second component of the de-correlated noise (i.e. the graphs on the first row have positive coefficients for $t > 0$) while the impact of the shock on the noncausal part of BTC is less clear since the causal response is both positive and negative. However we can see in the bivariate VAR(3), as compared with the bivarate VAR(1) model impulse response coefficients, that the positive noncausal response is larger than the negative noncausal response.

For ETH, the causal response is unambiguously positive as was the case for the bivariate VAR(1) model but the noncausal impulse response coefficients are now both negative and positive whereas in the VAR(1) case they were positive only. However the positive impulse response coefficients are larger.

In addition, can see that the impulse responses are smaller compared with those observed for the VAR(1) model shown in Figure 1.8 except that time taken for the impulse response coefficients to decay to zero is longer in the case of the bivariate VAR(3) model for BTC and ETH. Below we see the decomposition of the VAR(3) into its latent causal and noncausal components.



(a) BTC ETH Series and Causal Noncausal Combinations



(b) BTC ETH Series Causal Noncausal Combinations

Figure 1.13: BTC and ETH Series and Causal Noncausal Combinations for VAR(3)

In Figure 1.13 the original data and the highest variance latent causal and noncausal components are graphed. Panel (a) contains both the original series and the causal and noncausal combinations while Panel (b) contains only the causal and noncausal combinations. The Causal combinations are graphed as solid blue lines while the noncausal combinations are graphed as dotted red lines.

The Noncausal 1 component is the most explosive stationary combination of the two processes while the Causal 1 is the highest variance regular combination.

1.3.6 Comparing Goodness of Fit of the Mixed VAR(3) Against Causal VAR(3) for BTC and ETH USD Exchange Rates

Let us compare the fit of the mixed VAR(3) with a pure causal VAR(3) for BTC and ETH USD exchange rates.

The ACF in Figure 1.14 shows the serial dependence in the squared residuals of a linear and causal VAR(3) model estimated by the OLS estimator on the same sample. Note that the OLS is equivalent to MLE based on the Normal density, but our data has been shown not to be Normally distributed. By definition the VAR estimated by OLS has eigenvalues less than 1.

The mixed causal noncausal VAR is able to capture serial dependence in the data that a standard linear causal VAR model is not able to capture. That is to say, we observe that the autocorrelation of the squared residuals at lags one to three are statistically significant for the causal VAR(3) while they were not for the mixed VAR(3).

The OLS VAR(3) coefficients are as follows:

LAG 1 MATRIX:

$$\hat{\Phi}_{OLS_{BTC/ETH}} = \begin{bmatrix} 0.867 & -0.166 \\ -0.066 & 0.677 \end{bmatrix}$$

LAG 2 MATRIX:

$$\hat{\Phi}_{OLS_{BTC/ETH}} = \begin{bmatrix} 0.108 & 0.0574 \\ 0.143 & 0.029 \end{bmatrix}$$

LAG 3 MATRIX:

$$\hat{\Phi}_{OLS_{BTC/ETH}} = \begin{bmatrix} -0.172 & 0.006 \\ -0.057 & -0.104 \end{bmatrix}$$

$$\hat{\Psi}_{OLS_{BTC/ETH}} = \begin{bmatrix} 0.867 & -0.166 & 0.108 & 0.0574 & -0.172 & 0.006 \\ -0.066 & 0.677 & 0.143 & 0.029 & -0.057 & -0.104 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

The eigenvalues for this augmented matrix are, 0.943, 0.456+0.43i, 0.456-0.43i (modulus 0.627), 0.463, -0.55, -0.23.

The OLS VAR(3) shows statistically significant coefficients for BTC at time t with BTC at time $t-1$ and ETH at time t with ETH at time $t-1$. No other coefficients are statistically significant. The results show no evidence of a feedback effect or comovements.

The correlation matrices for the mixed and causal VAR models are shown below. We can see they two models have similar correlations between their respective residuals.

$$Corr_{[OLS_{BTC/ETH}]} = \begin{bmatrix} 1 & 0.857 \\ 0.857 & 1 \end{bmatrix}$$

$$Corr_{GCov_{BTC/ETH}} = \begin{bmatrix} 1 & 0.889 \\ 0.889 & 1 \end{bmatrix}$$

We can see the correlations between residuals are similar between the two models but the OLS model fails to capture the comovements and feedback effects that the mixed causal noncausal model captures. Both models show strong correlation but the OLS model fails to capture cross correlation. This is because the OLS model assumes all eigenvalues lie within the unit circle and is therefore misspecified.

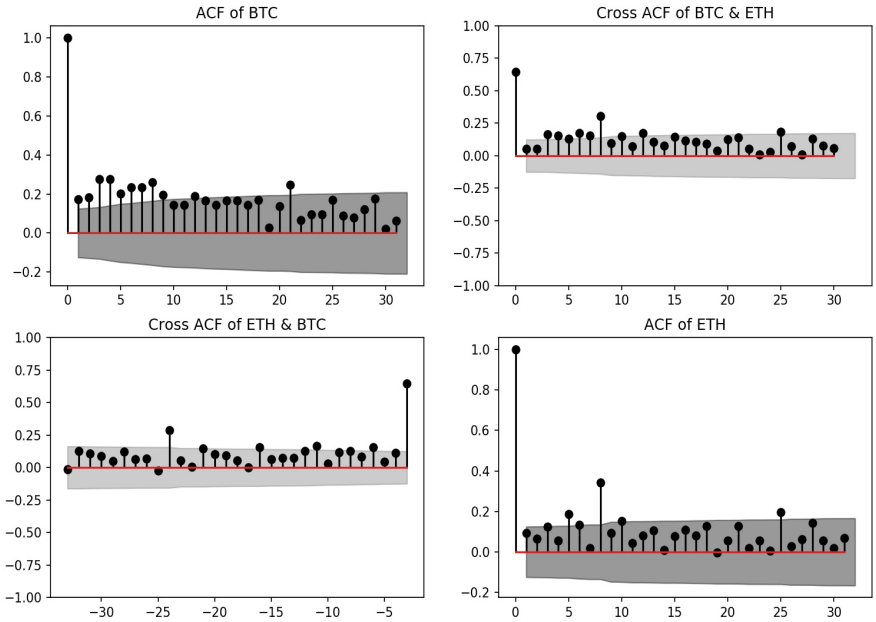


Figure 1.14: ACF of Squared Residuals from Causal VAR(3) for BTC and ETH

1.3.7 XRP (Ripple) and XLM (Stellar)

Figure 1.15a shows the USD exchange rates for Ripple and Stellar for the full sample of 882 daily exchange rate observations between 2017 01 and 2019 06, referred to as XRP and XLM hereafter. Figure 1.15b shows the same two time series

with their medians subtracted (referred to as the adjusted series henceforth) and Figure 1.15c shows the sub-sample of $T = 250$ observations between 2018 03 25 and 2018 11 29 used for analysis in the context of the full sample in grey and Figure 1.15d shows the adjusted sub-sample.

The summary statistics for the two series are given in Summary Sample Statistics, Table A.1. We find that the series are not normally distributed as indicated by excess kurtosis for example. In addition the series display features indicating nonstationarity and the potential presence of bubble phenomina.

The XRP and XLM exchange rate data is then detrended using a spline of order three and with a knot at every 25 observations using Python package Obspy [4]. Figures 1.16a and 1.16b show the original and detrended series for XRP and XLM respectively.

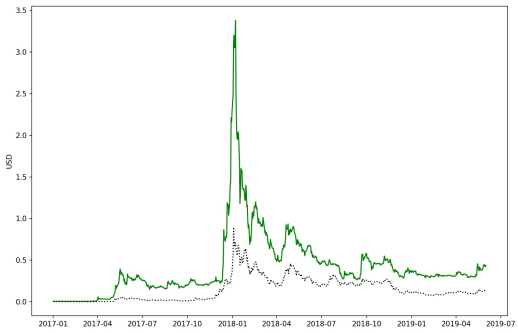
Figure 1.16c shows the adjusted and detrended sub-sample for XLM and XRP with XRP as the solid line and XLM as the dotted line.

By applying the Augmented Dickey-Fuller test to the detrended data we find that the resulting process is stationary as all p-values are near zero (for models with and without a constant, constant, with a linear trend and with both linear and quadratic trends) as shown in 1.5. The detrended data is not normally distributed, with non-zero excess kurtosis and skewness as documented in Table 1.6 below.

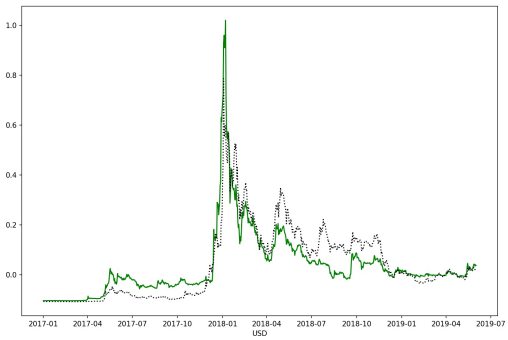
	No Constant	Constant	Linear Trend	Quad and Linear Trend
XRP	0.0	0.004	0.023	0.074
XLM	0.0	0.16	0.073	0.019

Table 1.5: XRP and XLM ADF P-Values for Detrended Series

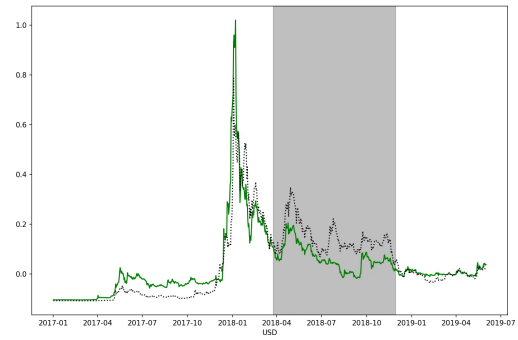
[4]Specifically, from Obspy we import the spline package from obspy.signal.detrend



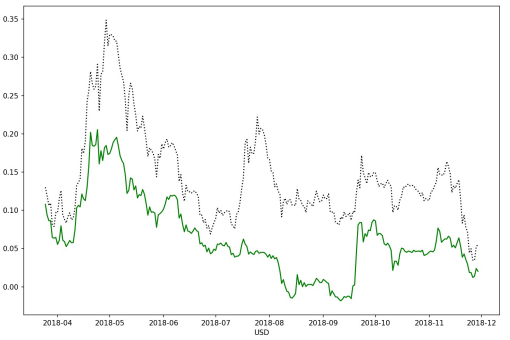
(a) 2017-01-01 to 2019-06-04



(b) Adjusted 2017-01-01 to 2019-06-04

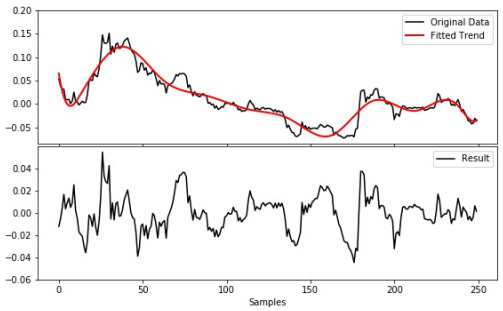


(c) Sub-sample 2018-03-25 to 2018-11-29 in grey

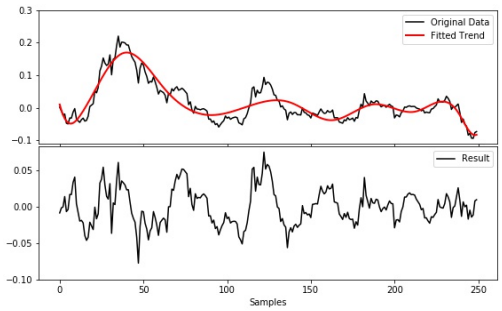


(d) Sub-sample 2018-03-25 to 2018-11-29

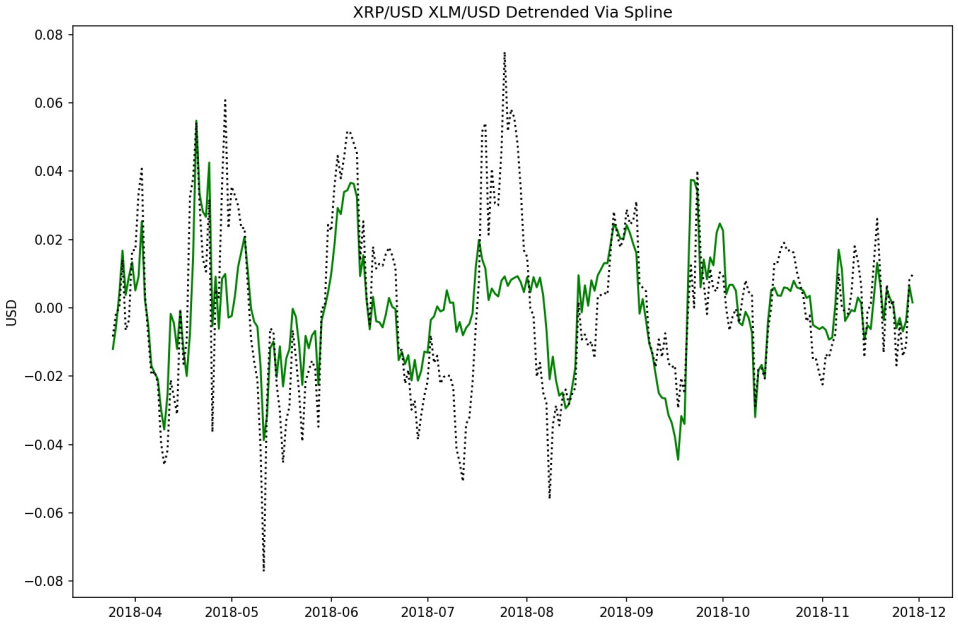
Figure 1.15: XRP/USD and XML/USD Exchange Rates. XRP/USD solid line, XML/USD dotted line.



(a) XRP (Adjusted) Detrended by Spline



(b) XLM (Adjusted) Detrended by Spline



(c) XRP/USD: solid line, XLM/USD: dotted line

Figure 1.16: XLM and XRP (Adjusted) Detrended by Spline

	Excess Kurtosis	Skew
XRP	0.48	0.102
XLM	0.017	0.28

Table 1.6: XRP and XLM Excess Kurtosis and Skewness for Detrended Series

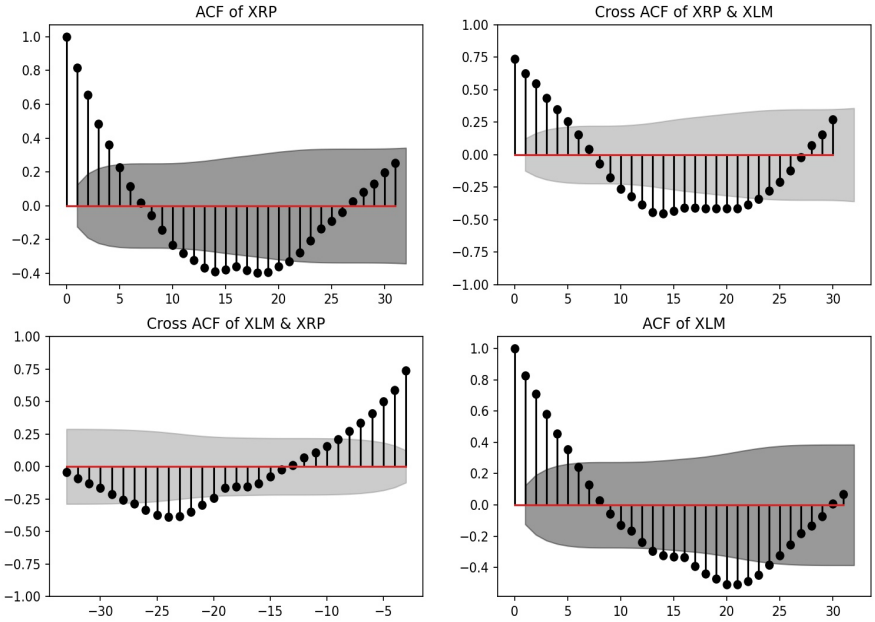


Figure 1.17: ACF of Adjusted and Detrended XRP and XLM

Figure 1.17 shows the ACF of the adjusted and detrended series. The ACF indicates that the adjusted, detrended series is consistent with a stationary process.

In addition, the Hill Plots for XRP and XLM, B.8c and B.8d respectively in Supplementary Graphs show tails with exponent below 2.

1.3.8 VAR(1) for XLM and XRP

By setting H in the objective function (1.2.19) equal to 6

$$\sum_{j=1}^{K}\sum_{k=1}^{K}[\sum_{h=1}^{H}\hat{\rho}_{j,k}^{2}(h,\Phi)]$$

and minimizing with respect to Φ using the second moments as the nonlinear function of the residuals we obtain the following estimated autoregressive coefficient matrix

$$\hat{\Phi}_{GCOV_{XRP/XLM}} = \begin{bmatrix} 0.95 & 0.213 \\ 0.303 & 1.04 \end{bmatrix}.$$

The eigenvalues for this matrix are 0.745 and 1.26 . These values are consistent with a mixed causal noncausal process since one eigenvalue lies within the unit circle while the other lies outside the unit circle. The standard errors for the first row are 0.05 and 0.03 respectively while the standard errors for the second row are 0.094 and 0.067 respectively, statistically significant according to the standard Wald test.

The ACF of the residuals and squared residuals of the VAR(1) model (Figures 1.18 and 1.19 respectively) show that most serial correlation has been removed but also that there still exists evidence of some correlation at lags 1 and 2 in the squared residuals.

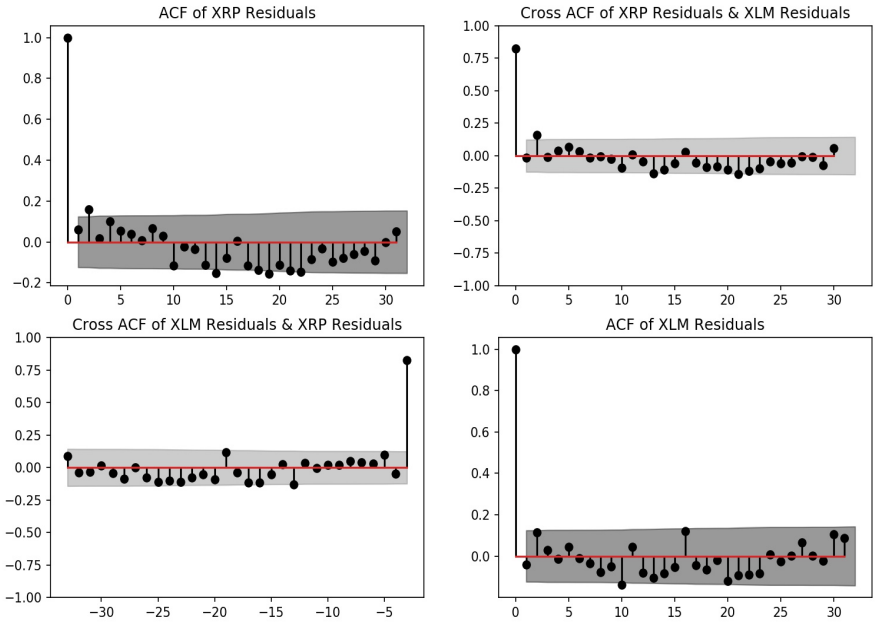


Figure 1.18: ACF of Residuals from VAR(1) for XRP and XLM

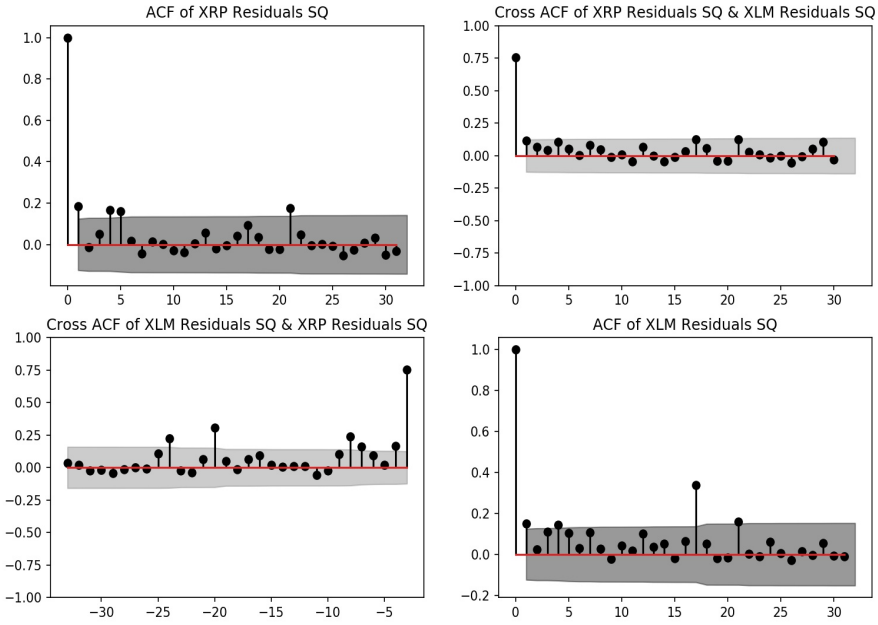


Figure 1.19: ACF of Squared Residuals from VAR(1) for XRP and XLM

The residual variance covariance matrix from the mixed VAR(1) estimated via the GCov is

$$\hat{\Sigma}_{XRP/XLM} = \begin{bmatrix} 0.000133 & 0.000163 \\ 0.000163 & 0.00029 \end{bmatrix}$$

In Figure 1.20 we see the histograms of the residuals from XRP and XLM respectively and in Figure 1.21 we see the QQ plots for the residuals from the VAR(1) of XRP and XLM. These indicate that the distribution of the residuals is non-Gaussian. This makes the data difficult to model using standard distributional

assumptions such as a t-distributed errors.

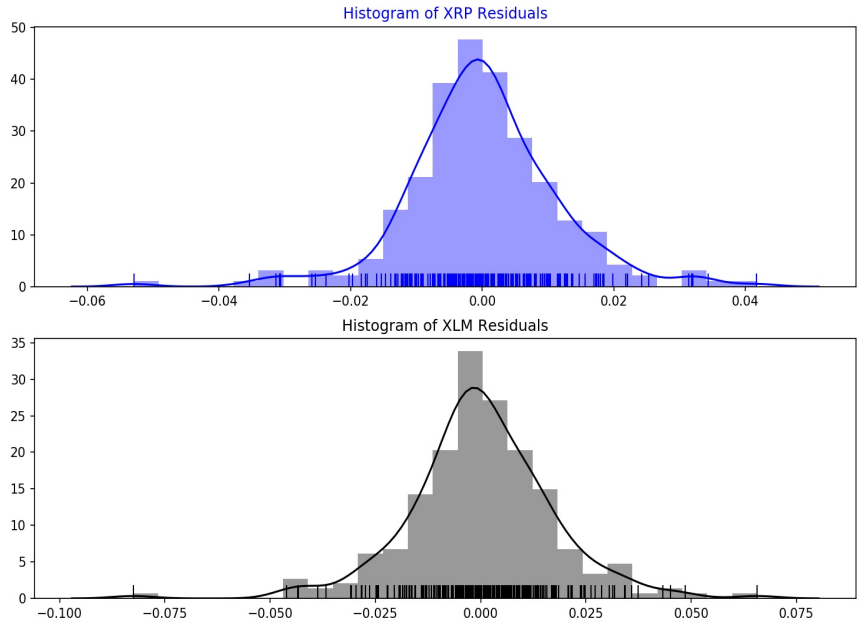
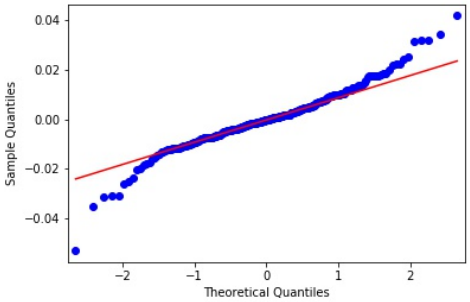
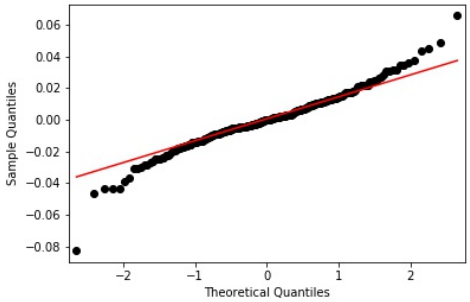


Figure 1.20: Histograms of Residuals from VAR(1) for XRP and XLM



(a) XRP QQ Plot of Residuals VAR(1)



(b) XLM QQ Plot of Residuals VAR(1)

Figure 1.21: XLM and XRP Plot of Residuals for VAR(1)

Table 1.7 shows the results of the normality tests applies to the residuals.

We observe the departures for non-normality are due to heavy tails displayed in the QQ plots. Table 1.8 shows excess kurtosis and skewness for the residuals of the VAR(1) for XRP and XLM.

	JB	JB-p	KS	KS-p	DP	DP-p	Sh	Sh-p
XRP	75.03	0.0	0.48	0.0	21.2	0.0	0.96	0.0
XLM	79.03	0.0	0.514	0.0	19.2	0.0	0.97	0.0

Table 1.7: Normality Tests for XRP and XRP VAR(1) Residuals

	Excess Kurtosis	Skew
XRP	2.74	-0.12
XLM	2.76	-0.2

Table 1.8: XRP and XLM Excess Kurtosis and Skewness for VAR(1) Residuals

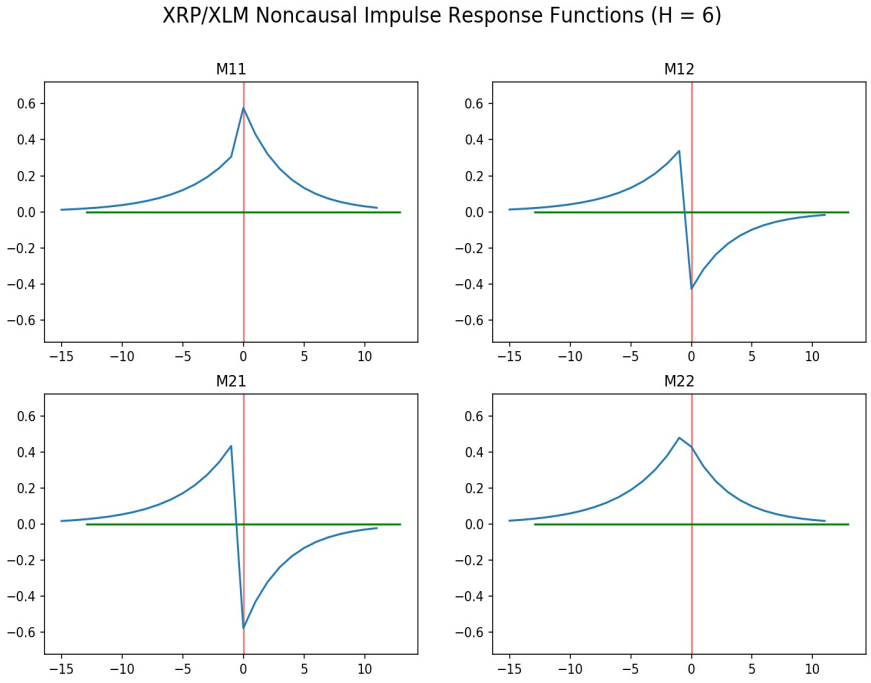


Figure 1.22: Noncausal Impulse Response Coefficients for XRP and XLM for VAR(1)

These results are similar to those obtained for Bitcoin and Ethereum in that the tests indicate nonnormality and heavy tails.

Next, we plot the impulse response coefficients for this model where there is a large positive shock at time zero (where the Cholesky decomposition is used to de-correlate the noise components).

In Figure 1.22 the first row contains impulse response coefficients for XRP and the second row contains impulse response coefficients for XLM. The first column (consisting of M_{11} and M_{21}) contains the first component of the transformed noise while the second column (consisting of M_{12} and M_{22}) contains the second component of the transformed noise.

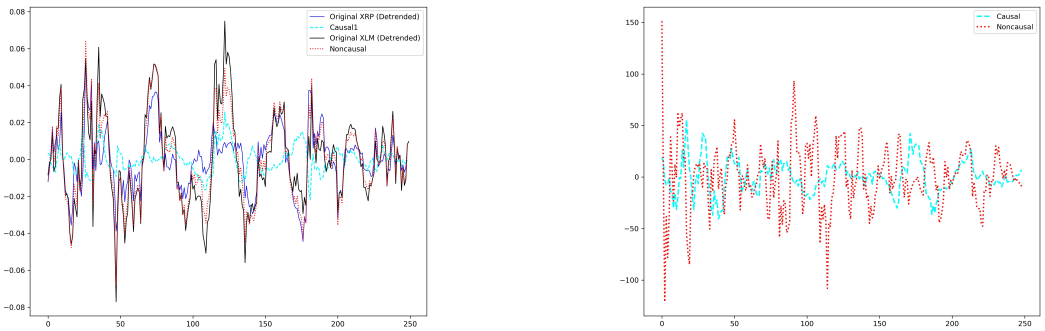
Since the first column of 1.22 contains the impulse response coefficients for XRP and XLM we can see that a large positive one time shock to the first component of the de-correlated noise will cause a large positive effect on both the noncausal and causal part of XRP. In contrast the aforementioned shock will have a large positive impact on the noncausal part of XLM but a large negative effect on the causal part of XLM.

Since the second column of Figure 1.22 contains the impulse response coefficients for XRP and XLM in response to a shock to the second component of the de-correlated noise we can see that large positive one time shock to the second component of the transformed noise causes a positive response in the noncausal part and a negative effect on the causal part of XRP. In contrast the shock has a positive impact on both the causal and noncausal parts of XLM.

We interpret this to mean that a one time positive shock will unambiguously lead to a positive effect on the noncausal part of XRP since the response is positive for the noncausal part of XRP through either the first or second component of the de-correlated noise (i.e. the graphs on the first row have positive coefficients for $t < 0$) while the impact of the shock on the causal part of XRP is less clear since the causal response is both positive and negative.

Similarly, for XLM, the response of the noncausal part to a shock is unambiguously positive while the response of the causal part is less clear.

Both the causal and noncausal components of the impulse responses dissipate within ten periods. The causal coefficients of the impulse response functions are larger in absolute value than their noncausal coefficients.



(a) Detrended Series and Causal Noncausal Linear Combinations

(b) Causal and Noncausal Linear Combinations Only

Figure 1.23: XRP and XLM Series and Causal and Noncausal Combinations for VAR(1)

Figure 1.23a displays the original detrended series along with the causal and noncausal components. Figure 1.23a displays common causal and noncausal

components for XRP and XLM. Similar to the BTC and ETH case, the causal component displays less volatility than does the noncausal component, as the last one depicts the explosive feature of the bi-variate series.

Similar to the BTC and ETH case, the noncausal residuals display higher levels of volatility relative to the causal residuals since the noncausal combinations represent the explosive or bubble components of the process.

1.3.9 Bivariate VAR(3) for XRP and XLM

By setting H in the objective function equal to 6

$$\sum_{j=1}^{K}\sum_{k=1}^{K}[\sum_{h=1}^{H}\hat{\rho}_{j,k}^{2}(h,\Phi)]$$

(see Equation ((1.2.19)) and minimizing with respect to Φ we obtain the following estimated autoregressive coefficient matrices at lags one, two and three given below.

LAG 1 MATRIX:

$$\hat{\Phi}_{GCOV_{XRP/XLM}} = \begin{bmatrix} 1.52 & 0.04 \\ 1.70 & 0.67 \end{bmatrix}$$

LAG 2 MATRIX:

$$\hat{\Phi}_{GCOV_{XRP/XLM}} = \begin{bmatrix} -2.19 & 1.61 \\ -4 & 2.97 \end{bmatrix}$$

LAG 3 MATRIX:

$$\hat{\Phi}_{GCOV_{XRP/XLM}} = \begin{bmatrix} 1.66 & -1.35 \\ 3.33 & -2.53 \end{bmatrix}$$

AUGMENTED MATRIX:

$$\hat{\Psi}_{GCOV_{XRP/XLM}} = \begin{bmatrix} 1.52 & 0.04 & -2.19 & 1.61 & 1.66 & -1.35 \\ 1.7 & 0.67 & -4.01 & 2.97 & 3.33 & -2.53 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

All coefficients are statistically significant according to the standard Wald test.

The augmented matrix $\hat{\Psi}_{GCOV_{XRP/XLM}}$ has the following eigenvalues, with one above unity in modulus: 1.8, 0.75, -0.64+0.2i and -0.64-0.2i (of modulus 0.67) 0.46+0.52i 0.46-0.52i (of modulus 0.69) which is consistent with a mixed causal noncausal process.

The ACFs for the residuals and squared residuals from the VAR(3) model for XRP and XLM an are shown in Figures 1.24 and 1.25 respectively. These indicates that the noncausal VAR(3) has captured the dependencies between the series and is thus an improvement over the mixed bivariate VAR(1) model.

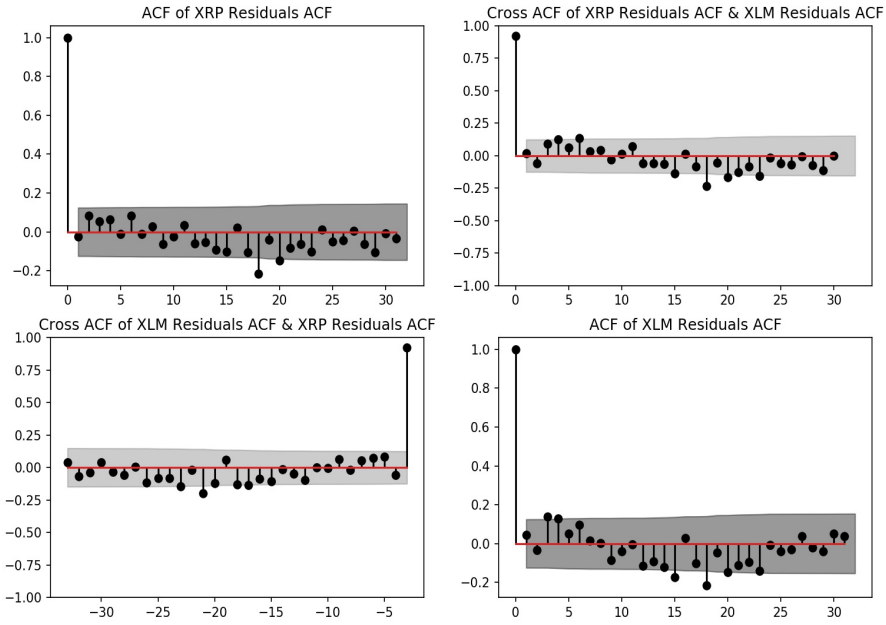


Figure 1.24: ACF of VAR(3) Residuals for XRP and XLM

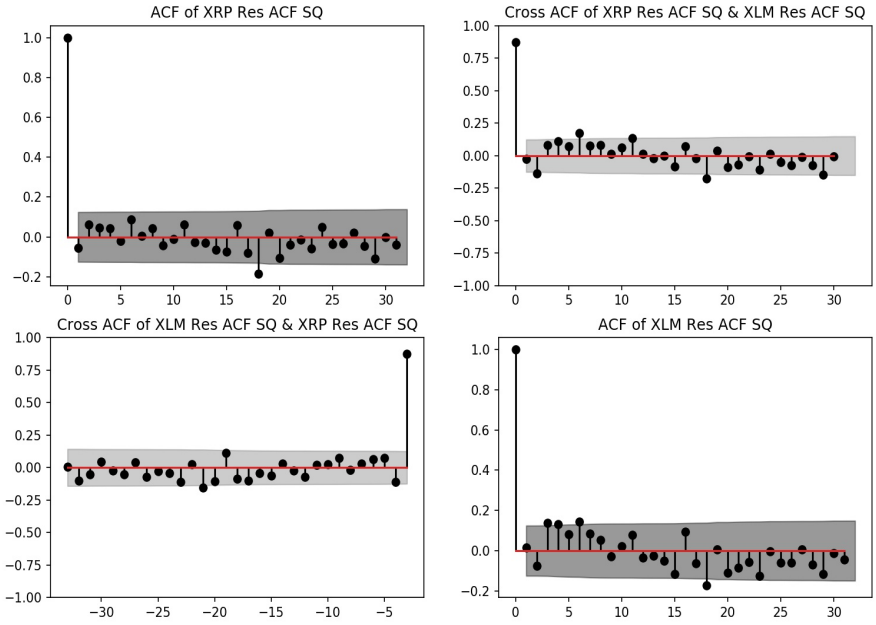


Figure 1.25: ACF of VAR(3) Squared Residuals for XRP and XLM

The histograms of VAR(3) residuals for XRP and XLM are in Figures B.2 Appendix B. Both series display large tails. This is confirmed by the QQ plots in Figures B.5c and B.5d.

The graphs of histograms and QQ plots of the residuals for XRP and XLM are consistent with the residuals having 'fat tails' and thus indicate non-normality.

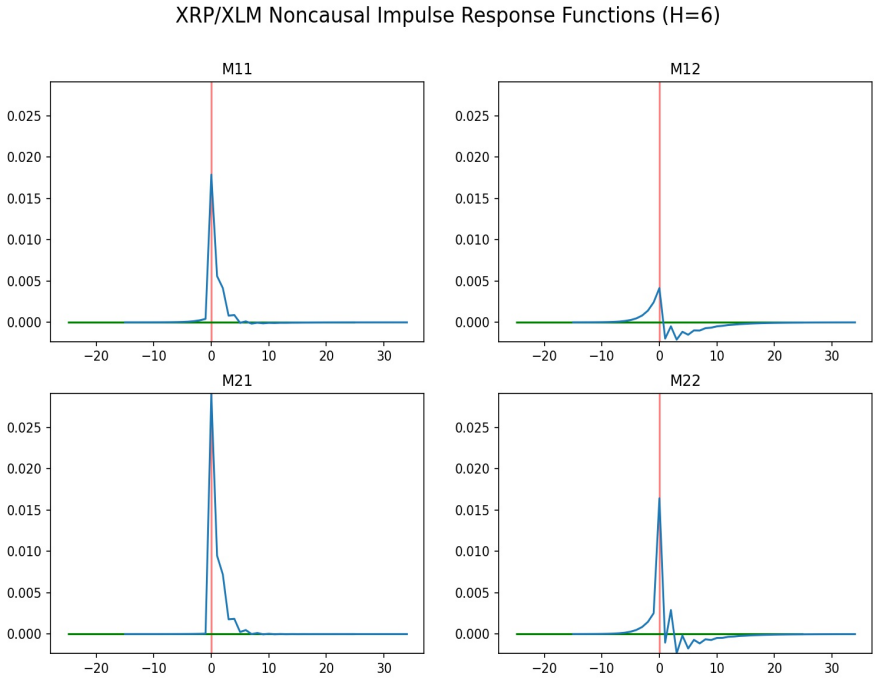


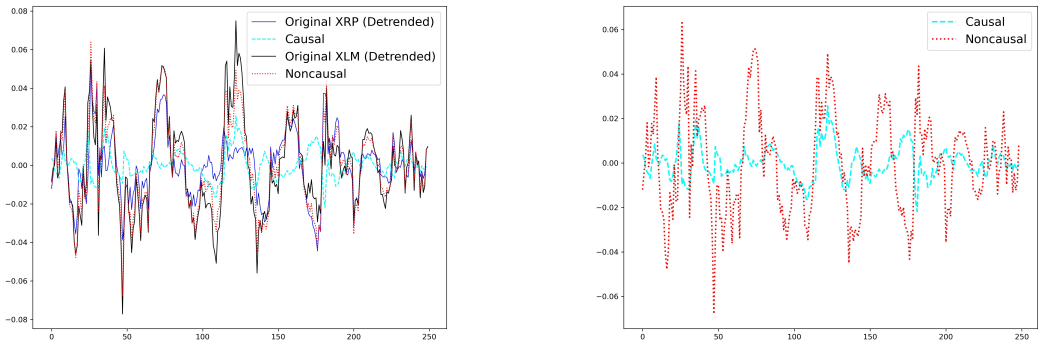
Figure 1.26: XRP and XLM Noncausal Impulse Response Coefficients for VAR(3)

From Figure 1.26 we can see that the impact of a one time shock to the causal part of XRP whether through the first of second component is positive since the impact is either large positive as in panel $M11$ or small negative as in panel $M12$ while the impact on the noncausal part is positive since though the first component of the de-correlated noise the impact is minimal ($M21$ for $t < 0$) while through the second component of the de-correlated noise the effect is positive ($M12$ for $t < 0$).

For XLM the impact on the noncausal part is positive since through the first component of the de-correlated noise the impact is either minimal as can be seen in $M21$ or positive as can be seen in $M22$. The impact on the causal part of XLM is interpreted as being positive since the impact on the causal part is either

large positive as can be seen in $M21$ or minimally negative as shown in $M22$.

In contrast to the VAR(1) bivariate model the noncausal components shown the impulse response coefficients in the VAR(3) are less prominent for the first component (i.e. the first column) but are more prominent for the second component of the de-correlated noise (i.e. the second column).



(a) XRP and XLM Series and Causal Noncausal Combinations

(b) XRP and XLM Series Causal Noncausal Combinations

Figure 1.27: XRP and XLM Series and Causal Noncausal Combinations for VAR(3)

Figure 1.27 above displays the causal and noncausal components of the XRP/XLM pair of cryptocurrencies.

1.3.10 Comparing Goodness of Fit of the Mixed VAR(3) Against Causal VAR(3) for XRP and XLM USD Exchange Rates

Let us compare the VAR(3) estimated in a mixed causal noncausal context using the GCov estimator with the results obtained from a causal VAR(3) estimated via OLS on the XRP and XLM data. Note that the OLS is equivalent to MLE based on the Normal density, but our data has been shown not to be Normally distributed. By definition the VAR estimated by OLS has eigenvalues less than 1.

The ACF of the squared residuals in Figure 1.28 shows that the causal VAR(3) fails to remove serial correlation from the squared residuals. The mixed causal noncausal VAR(3) model, in contrast is able to remove these nonlinear autocorrelations in a linear stationary model.

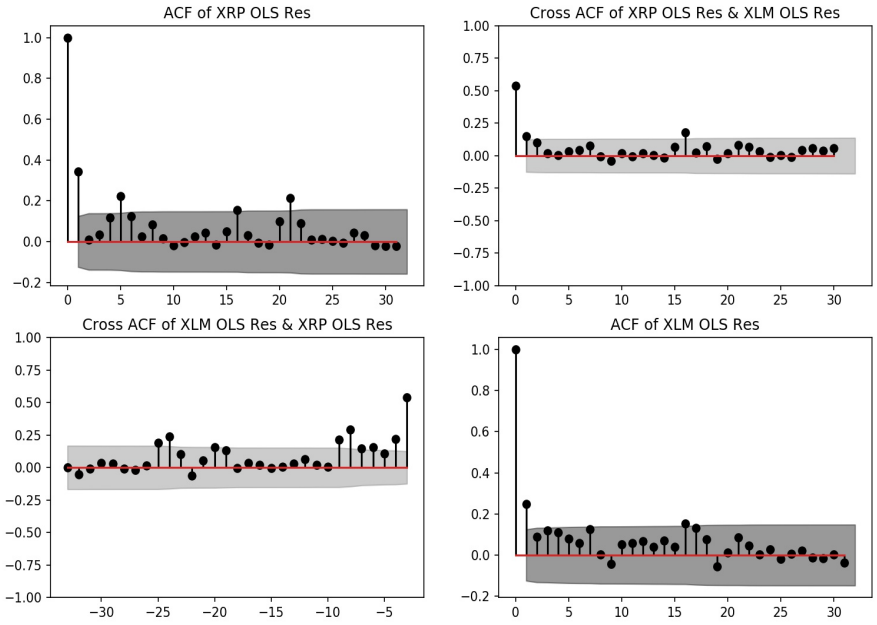


Figure 1.28: ACF of Causal VAR(3) Squared Residuals for XRP and XLM

The mixed causal noncausal VAR is able to capture serial dependence in the data that a standard linear causal VAR model is not able to capture. That is to say, we observe that the autocorrelation of the squared residuals at lags one to three are statistically significant for the causal VAR(3) while they were not for the mixed VAR(3).

The OLS VAR(3) coefficients are as follows:

LAG 1 MATRIX:

$$\hat{\Phi}_{OLS_{XRP/XLM}} = \begin{bmatrix} 0.867 & -0.166 \\ -0.066 & 0.677 \end{bmatrix}$$

LAG 2 MATRIX:

$$\hat{\Phi}_{OLS_{XRP/XLM}} = \begin{bmatrix} 0.108 & 0.0574 \\ 0.143 & 0.029 \end{bmatrix}$$

LAG 3 MATRIX:

$$\hat{\Phi}_{OLS_{XRP/XLM}} = \begin{bmatrix} -0.172 & 0.006 \\ -0.057 & -0.104 \end{bmatrix}$$

The OLS VAR(3) shows statistically significant coefficients for XRP_t with XRP_{t-1}, XRP_t with XLM_{t-2} andXLM_t with XLM_{t-1} at time $t-1$. No other coefficients Ripple are statistically significant.

The correlation matrix is shown below which shows that both the mixed and causal VAR models have similar correlations in their respective residuals.

$$Corr_{[OLS_{XRP/XLM}]} = \begin{bmatrix} 1 & 0.74 \\ 0.74 & 1 \end{bmatrix}$$

$$Corr_{GCov_{XRP/XLM}} = \begin{bmatrix} 1 & 0.82 \\ 0.82 & 1 \end{bmatrix}$$

As was the case with the BTC/ETH pair, the causal OLS model shows strong correlation in the residuals but fails to capture the cross correlation due to misspecification of the model due to the assumption of causality.

1.3.11 VAR(1) For Bitcoin, Ethereum, Ripple and Stellar

We now consider a noncausal VAR(1) of all four cryptocurrencies using 200 observations recorded between March 5th 2018 and October 10th 2018 in which the values of BTC and ETH have been divided by a factor of 1000 in order to adjust the data to a common range of values. The data has been demeaned and scaled in order to perform the estimation and can be seen in Figure 1.29.

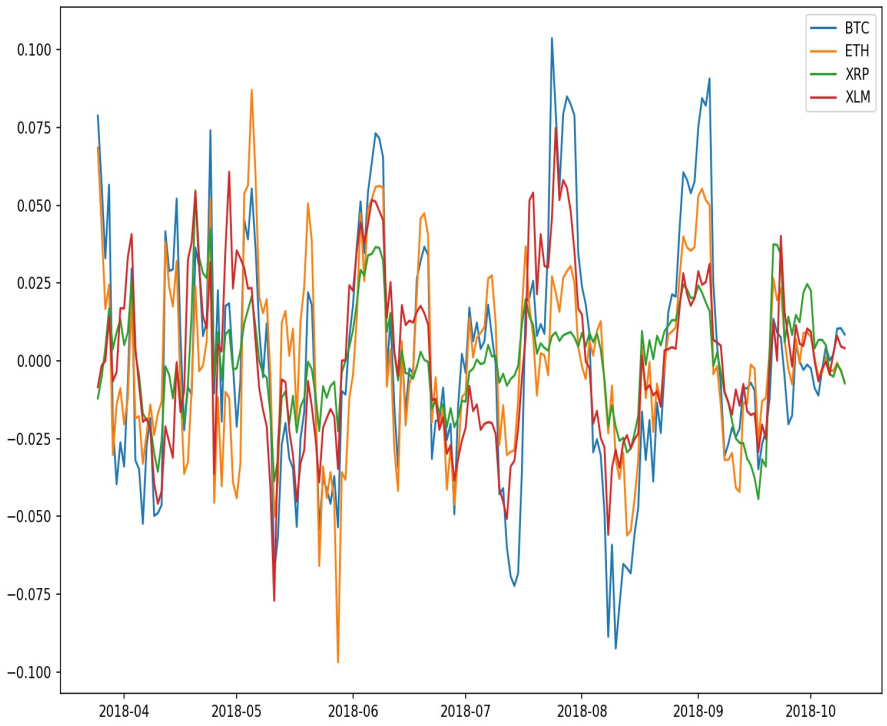


Figure 1.29: BTC, ETH, XRP, and XLM Exchange Rates Spline Detrended (adjusted)

Setting H equal to 14 in the objective function

$$\rho(\Phi, y, ...) = \sum_{j=1}^{K}\sum_{k=1}^{K}[\sum_{h=1}^{H}\hat{\rho}_{j,k}^{2}(h, \Phi)]$$

(see Equation (1.2.19)) and minimizing with respect to Φ using the second moments as the nonlinear function of the residuals we obtain the following estimated autoregressive matrix,

$$\hat{\Phi}_{GCOV_{BTC/ETH/XRP/XLM}} = \begin{bmatrix} 0.69 & 0.075 & 0.099 & 0.56 \\ -0.094 & 0.918 & -0.15 & 0.588 \\ -0.2 & 0.0979 & 0.995 & 0.295 \\ 0.306 & -0.326 & -0.0115 & 1.068 \end{bmatrix}.$$

The eigenvalues of the autoregressive matrix given above are as follows: 1.16, 0.79, 0.79+0.23i, 0.79-0.23i with one eigenvalue outside the unit circle and three eigenvalues inside the unit circle (the complex eigenvalues have modulus 0.832). This result that implies a mixed process (i.e. a process containing both causal and noncausal components).

The histograms of the residuals and QQ plots of the residuals, both in Supplementary Graphs, Figures B.3 and B.4 respectively show large tails consistent with a non-normal distribution of the VAR(1) residuals. The Jarque-Bera and Shapiro Wilk test statistics both indicate that the residuals for BTC, ETH, XRP and XLM are not normally distributed.

Located in Supplementary Graphs Figures B.6a and B.6c display the ACFs for the residuals and squared residuals for BTC and ETH respectively, while Figures B.6b and B.6d display the ACFs for the residuals and squared residuals for XRP and XLM.

We observe here that the model removes the serial correlation in the residuals and the squared residuals. This implies that the model provides a good fit to the data.

1.4 Conclusion

In this chapter we examined US dollar exchange rates for the following cryptocurrencies: Bitcoin, Ethereum, Ripple and Stellar. We modelled these cryptocurrency exchange rates as bivariate VAR(1) and VAR(3) mixed processes for the pairs Bitcoin/Ethereum and Ripple/Stellar and as a VAR(1) mixed process for the four cryptocurrency exchange rates (i.e. a VAR(1) of dimension four).

Using mixed causal noncausal modelling has allowed us to decompose the processes into their causal (i.e. 'regular') and noncausal (i.e. 'explosive') components allowing us to observe latent joint bubble dynamics within these processes.

We compare the results from the OLS VAR modelling with the mixed causal noncausal models and find that modelling these processes as multivariate processes containing both causal and noncausal components enables us to detect nonlinear dependencies within and between the series as well as comovements between the processes using a mixed linear VAR model (applying semi-parametric estimation)

while standard linear causal VARs cannot.

Chapter 2

Forecasting

2.1 Introduction

The goal of this chapter is to use the decomposition of multivariate time series processes to forecast processes that display empirical evidence of what are commonly called bubbles, but which we define as locally explosive behaviour in strictly stationary processes.

We derive the formulas proposed in Gouriéroux and Jasiak (2016) to calculate predictive densities for mixed causal noncausal VAR(1) models which we use to forecast one step ahead and calculate confidence intervals. The model is estimated using the GCov estimator. We then use a linear approximation method to calculate one step ahead out of sample forecasts of the cryptocurrency pairs Bitcoin/Ethereum (BTC/ETH) and Ripple/Stellar (XPR/XLM).

We use a kernel estimator of the predictive densities which was first introduced in Gouriéroux and Jasiak (2016). Since the semi-parametric density estimator can be computationally demanding we consider a simplified approach that combines the linear prediction method for causal VAR models with a consistent estimation of the autoregressive coefficient matrix.

2.2 Predictive Density for Bivariate Noncausal VAR(1) Model

This section derives the formulas proposed in Gouriéroux and Jasiak (2016) in an alternative way, using the latent causal and noncausal components of a given series as opposed to the observed components.

Bivariate VAR(1): Eigenvalues of Φ inside and outside the unit circle

Let us consider a bivariate VAR(1). The latent causal and noncausal components satisfy:

$$\begin{aligned} y_{1,t}^* &= \lambda_1 y_{1,t-1}^* + \epsilon_{1,t}^*, \\ y_{2,t}^* &= \lambda_2 y_{2,t-1}^* + \epsilon_{2,t}^*, \end{aligned}$$

where $|\lambda_1| < 1$ and $|\lambda_2| > 1$. $\epsilon_t^* = (\epsilon_{1,t}^*, \epsilon_{2,t}^*)$ are assumed to be i.i.d. sequences. $\epsilon_{1,t}^*$ and $\epsilon_{2,t}^*$ can be correlated.

The causal part has the causal MA(∞) representation $y^*_{1,t} = \sum_{i=0}^{\infty} \lambda^i \epsilon^*_{1,t-i}$, which is a function of $\epsilon^*_{1,t}, \epsilon^*_{1,t-1}, ...$, and hence a function of $\epsilon^*_t, \epsilon^*_{t-1}, ...$

The noncausal part has the stationary representation:

$$y^*_{2,t} = \frac{1}{\lambda_2} y^*_{2,t+1} - \frac{1}{\lambda_2} \epsilon^*_{2,t+1},$$

and a forward-looking MA(∞) representation $y^*_{2,t} = -\frac{1}{\lambda_2} \sum_{i=0}^{\infty} (\frac{1}{\lambda_2})^i \epsilon^*_{2,t+1+i}$,

which is a function of $\epsilon^*_{2,t+1}, \epsilon^*_{2,t+2}, ...$, hence a function of $\epsilon^*_{t+1}, \epsilon^*_{t+2}, ...$

Moreover, the knowledge of $y^*_{2,1}, y^*_{2,2}, ..., y^*_{2,T}$ is equivalent to the knowledge of

$$\epsilon^*_{2,2}, \epsilon^*_{2,3}, ..., \epsilon^*_{2,T-1}, y^*_{2,T}.$$

At time T, the information contains the shocks between dates 2 and $T-1$ and their independent summaries of past before $t = 1$ and future after $t = T + 1$:

$$I_T = \begin{pmatrix} y^*_{1,1} & y^*_{1,2} & \cdots & y^*_{1,T} \\ y^*_{2,1} & y^*_{2,2} & \cdots & y^*_{2,T} \end{pmatrix} \iff \begin{pmatrix} y^*_{1,1} & \epsilon^*_{1,2} & \cdots & \epsilon^*_{1,T-1} & \epsilon^*_{1,T} & \\ & \epsilon^*_{2,2} & \cdots & \epsilon^*_{2,T-1} & \epsilon^*_{2,T} & y^*_{2,T} \end{pmatrix}$$

Variable $y^*_{1,T}$ is the state variable that summarizes the effects of past $\epsilon^*_1, \epsilon^*_0, ...$ Variable $y^*_{2,T}$ is the state variable that summarizes the effect of future $\epsilon^*_{T+1}, \cdots, \epsilon^*_{T+h}...$ on $\underline{y_T}^*$

Therefore $\underline{\epsilon_T}^* = (\epsilon^*_1, \cdots, \epsilon^*_T)$ is independent of $y^*_{2,T}$.

2.2.1 Forecasting $y^*_{1,T+1}, y^*_{2,T+1}$ given I_T.

We have:

$$y^*_{1,T+1} = \lambda_1 y^*_{1,T} + \epsilon^*_{1,T+1},$$

where $\lambda_1 y^*_{1,T} \in (\underline{\epsilon_T}^*) \subset I_T = (\underline{\epsilon_T}^*, y^*_{2,T})$.

Also, we have

$$y^*_{2,T+1} = \frac{1}{\lambda_2} y^*_{2,T+2} - \frac{1}{\lambda_2} \epsilon^*_{2,T+2},$$

where $\frac{1}{\lambda_2} y^*_{2,T+1}$ is a function of $\epsilon^*_{T+2}, \epsilon^*_{T+3}$....

Therefore $(\epsilon^*_{1,T+1}, y^*_{2,T}, y^*_{2,T+1})$ is independent of $\underline{\epsilon_T}^* = (\epsilon^*_1, \cdots, \epsilon^*_T)$.

It follows that the following densities are equal:

$$l(\epsilon^*_{1,T+1}, y^*_{2,T}, y^*_{2,T+1} | \underline{\epsilon_T}^*) = l(\epsilon^*_{1,T+1}, y^*_{2,T}, y^*_{2,T+1})$$

and hence

$$\begin{aligned} l(\epsilon^*_{1,T+1}, y^*_{2,T+1} | I_T) &= l(\epsilon^*_{1,T+1},\, y^*_{2,T+1} | y^*_{2,T},\, \underline{\epsilon_T}^*) \\ &= l(\epsilon^*_{1,T+1},\, y^*_{2,T+1} | y^*_{2,T}). \end{aligned}$$

By applying the Bayes formula, we rewrite this equality as:

$$l(\epsilon^*_{1,T+1}, y^*_{2,T+1}|y^*_{2,T}) = \frac{l_2(y^*_{2,T+1})}{l_2(y^*_{2,T})}\, l_1(\epsilon^*_{1,T+1},\, y^*_{2,T}|y^*_{2,T+1}), \tag{2.2.1}$$

where l_2 is the stationary density of $y^*_{2,t}$. The relevant variables are $\epsilon^*_{1,T+1}, y^*_{2,T} = \frac{1}{\lambda_2} y^*_{2,T+1} - \frac{1}{\lambda_2}\epsilon^*_{2,T+1}$ with independence between $y^*_{2,T+1}$ and the set $(\epsilon^*_{1,T+1}, \epsilon^*_{2,T+1})$. Let the density of $(\epsilon^*_{1,t}, \epsilon^*_{2,t})$ be denoted by $g^*(\epsilon^*_1, \epsilon^*_2)$. We have

$$\begin{pmatrix} \epsilon^*_{1,T+1} \\ \epsilon^*_{2,T+1} \end{pmatrix} = \begin{pmatrix} \epsilon^*_{1,T+1} \\ y^*_{2,T+1} - \lambda_2 y^*_{2,T} \end{pmatrix}.$$

If $y^*_{2,T+1}$ is considered fixed, we change the variables $\begin{pmatrix} \epsilon^*_{1,T+1} \\ \epsilon^*_{2,T+1} \end{pmatrix}$ to $\begin{pmatrix} \epsilon^*_{1,T+1} \\ -\lambda_2 y^*_{2,T} \end{pmatrix}$ with a Jacobian equal to $|\lambda_2|$: $|det[\frac{\partial \epsilon^*_{2,T+1}}{\partial y^*_{2,T}}]| = |\lambda_2|$.

Then, we get the conditional density:

$$l(\epsilon^*_{1,T+1}, y^*_{2,T}|y^*_{2,T+1}) = |\lambda_2| g^*(\epsilon^*_{1,T+1},\, y^*_{2,T+1} - \lambda_2 y^*_{2,T}).$$

It follows that the conditional density is:

$$l_1(y^*_{1,T+1}, y^*_{2,T}|y^*_{2,T+1}, \underline{\epsilon_T}^*) = |\lambda_2| g(y^*_{1,T+1} - \lambda_1 y^*_{1,T},\, y^*_{2,T+1} - \lambda_2 y^*_{2,T}). \tag{2.2.2}$$

Let us introduce the marginal densities $l_2(y^*_{2,T+1})$ and $l_2(y^*_{2,T})$. The generic values of variables $y^*_{1,T+1}, y^*_{2,T+1}$ are hence denoted by y^*_1, y^*_2. The predictive density of

$y^*_{1,T+1}, y^*_{2,T+1}$ is:

$$l(y^*_1, y^*_2|I_T) = \frac{l_2(y^*_2)}{l_2(y^*_{2,T})} |\lambda_2| \, g^*(y^*_1 - \lambda_1 y^*_{1,T},\, y^*_2 - \lambda_2 y^*_{2,T}) \tag{2.2.3}$$

The predictive density depends on unknown scalar parameters λ_1, λ_2 and functional parameters l_2, g that can be estimated. The marginal density $l_2(y^*_2)$ can be approximated by a kernel estimator $\hat{l}_2(y^*_2)$:

$$\hat{l}_2(y^*_2) = \frac{1}{T}\frac{1}{h_2}\sum_{t=1}^{T} K\left(\frac{y^*_2 - y^*_{2,t}}{h_2}\right),$$

where h_2 is a bandwidth. Then, the joint density $g^*(\epsilon^*_1, \epsilon^*_2)$ can be approximated by

$$\hat{g}^*(\epsilon^*_1, \epsilon^*_2) = \frac{1}{T}\frac{1}{h_{11}h_{12}}\sum_{t=1}^{T} K\left(\frac{\epsilon^*_1 - \hat{\epsilon}^*_{1,t}}{h_{11}}\right) K\left(\frac{\epsilon^*_2 - \hat{\epsilon}^*_{2,t}}{h_{12}}\right).$$

where $\hat{\epsilon}^*_{1,t}$ and $\hat{\epsilon}^*_{2,t}$ are the residuals $\hat{\epsilon}^*_t = \hat{A}^{-1}\hat{\epsilon}_t = \hat{A}^{-1}(y_t - \hat{\Phi} y_{t-1})$ and h_{11}, h_{12} are two bandwidths adjusted for the variability of $\hat{\epsilon}^*_{1,t}$ and $\hat{\epsilon}^*_{2,t}$, respectively.

Then the estimator of the predictive density $l(y^*_1, y^*_2|I_T)$ is

$$\hat{l}(y^*_1, y^*_2|I_T) = \frac{\frac{1}{T}\frac{1}{h_2}\sum_{t=1}^{T} K\left(\frac{y^*_2 - y^*_{2,t}}{h_2}\right)}{\frac{1}{T}\frac{1}{h_2}\sum_{t=1}^{T} K\left(\frac{y^*_{2,T} - y^*_{2,t}}{h_2}\right)} |\hat{\lambda}_2| \hat{g}^*(y^*_1 - \hat{\lambda}_1 y^*_{1,T},\, y^*_2 - \hat{\lambda}_2 y^*_{2,T})$$

$$= \frac{\frac{1}{T}\frac{1}{h_2}\sum_{t=1}^{T} K\left(\frac{y_2^* - y_{2,t}^*}{h_2}\right)}{\frac{1}{T}\frac{1}{h_2}\sum_{t=1}^{T} K\left(\frac{y_{2,T}^* - y_{2,t}^*}{h_2}\right)} |\hat{\lambda}_2| \frac{1}{T}\frac{1}{h_{11}h_{12}} \sum_{t=1}^{T} \left[K\left(\frac{y_1^* - \hat{\lambda}_1 y_{1,T}^* - \hat{\epsilon}_{1,t}^*}{h_{11}}\right) K\left(\frac{y_2^* - \hat{\lambda}_2 y_{2,T}^* - \hat{\epsilon}_{2,t}^*}{h_{12}}\right)\right], \quad (2.2.4)$$

where $\hat{\lambda}_i, i = 1, 2$ are consistent estimators of $\lambda_i = 1, 2$ in:

$$\hat{\Phi} = \hat{A} \begin{pmatrix} \hat{\lambda}_1 & 0 \\ 0 & \hat{\lambda}_2 \end{pmatrix} \hat{A}^{-1},$$

and $y_t^* = \hat{A}^{-1} y_t$ and $\hat{\epsilon}_t^* = \hat{A}^{-1}\hat{\epsilon}_t$.

2.2.2 Predictive Density of $y_{1,T+1}, y_{2,T+1}$

From the formula of predictive density $l(y_{1,T+1}^*, y_{2,T+1}^* | y_{1,T}^*, y_{2,T}^*)$ we obtain the predictive density $l(y_{1,T+1}, y_{2,T+1} | y_{1,T}, y_{2,T})$ of $y_{1,T+1}, y_{2,T+1}$ as follows.

Let the coefficients of the inverse of A be denoted by

$$\hat{A}^{-1} = \begin{pmatrix} \hat{a}^{11} & \hat{a}^{12} \\ \hat{a}^{21} & \hat{a}^{22} \end{pmatrix}.$$

Then, we can compute:

$$
\begin{aligned}
y^*_{1,t} &= \hat{a}^{11} y_{1,t} + \hat{a}^{12} y_{2,t} \\
y^*_{2,t} &= \hat{a}^{21} y_{1,t} + \hat{a}^{22} y_{2,t}
\end{aligned}
$$

and have

$$
\begin{aligned}
y^*_{1} &= \hat{a}^{11} y_{1} + \hat{a}^{12} y_{2} \\
y^*_{2} &= \hat{a}^{21} y_{1} + \hat{a}^{22} y_{2}
\end{aligned}
$$

Next, we replace the starred variables in the formula (2.2.3). In addition, the transformation requires including the Jacobian $\frac{1}{|det(\hat{A})|}$ as an additional factor on the right hand side of the formula:

$$
l(y_1, y_2 | I_T) = l(y^*_1, y^*_2) \tfrac{1}{|det(\hat{A})|},
$$

where

$$
\begin{aligned}
l(y_1, y_2 | I_T) \quad = \quad & \frac{\frac{1}{T}\frac{1}{h_2}\sum_{t=1}^{T} K\left(\frac{\hat{a}^{21} y_1 + \hat{a}^{22} y_2 - (\hat{a}^{21} y_{1,t} + \hat{a}^{22} y_{2,t})}{h_2}\right)}{\frac{1}{T}\frac{1}{h_2}\sum_{t=1}^{T} K\left(\frac{(\hat{a}^{21}(y_{1,T} + \hat{a}^{22} y_{2,T}) - (\hat{a}^{21} y_{1,t} + \hat{a}^{22} y_{2,t})}{h_2}\right)} \\
& \cdot |\hat{\lambda}_2| \frac{1}{T} \frac{1}{h_{11} h_{12}} \sum_{t=1}^{T} K\left(\frac{\hat{a}^{11} y_1 + \hat{a}^{12} y_2 - \hat{\lambda}_1(\hat{a}^{11} y_{1,T} + \hat{a}^{12} y_{2,T}) - (\hat{a}^{11}\hat{\epsilon}_{1,t} + \hat{a}^{12}\hat{\epsilon}_{2,t})}{h_{11}}\right) \\
& \cdot K\left(\frac{\hat{a}^{21} y_1 + \hat{a}^{22} y_2 - \hat{\lambda}_2(\hat{a}^{21} y_{1,T} + \hat{a}^{22} y_{2,T}) - (\hat{a}^{21}\hat{\epsilon}_{1,t} + \hat{a}^{22}\hat{\epsilon}_{2,t})}{h_{12}}\right).
\end{aligned}
$$

2.2.3 One Step Ahead Out Of Sample Forecasts: Bivariate Predictive Densities

The marginal densities are obtained by integrating out the predictive density

$$l(y_1|I_T) = \int l(y_1, y_2|I_T)dy_2$$

and

$$l(y_2|I_T) = \int l(y_1, y_2|I_T)dy_1$$

using the formulas presented in Predictive Density for Bivariate Noncausal VAR(1) Model, which are kernel estimates of multivariate forecasts introduced in Gouriéroux and Jasiak (2016). In the estimation of the kernel densities the bandwidth parameters h_{11}, h_{12} and h_2 in (2.2.5) are set as follows: for BTC and ETH $h_{11} = 4$, and $h_{12} = h_2 = 1$, while for XRP and XLM $h_11 = h_{12} = h_2 = 1$.

We obtain the following forecasts and confidence intervals (denoted as CI) for the cryptocurrency pairs BTC/ETH and XRP/XLM:

One Step Ahead Predictive Densities for BTC and ETH:

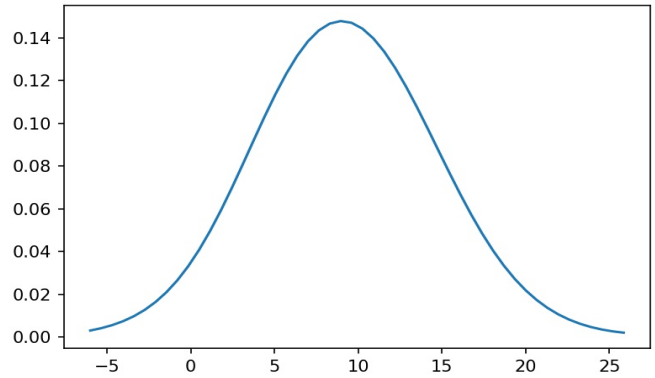


Figure 2.1: BTC Marginal Predictive Density, Forecast: 9: CI (1.8, 16.1)

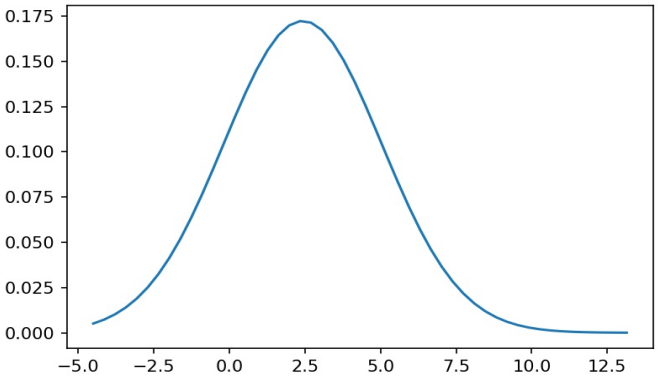


Figure 2.2: ETH Marginal Predictive Density. Forecast 21, CI (0.9, 3.42)

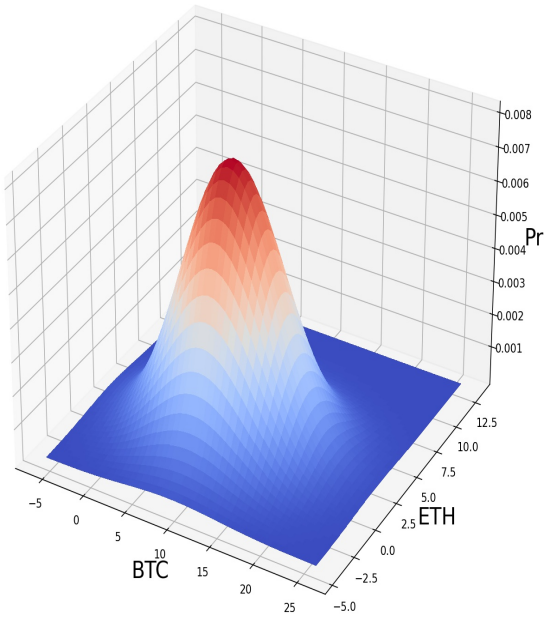


Figure 2.3: BTC/ETH Joint Density

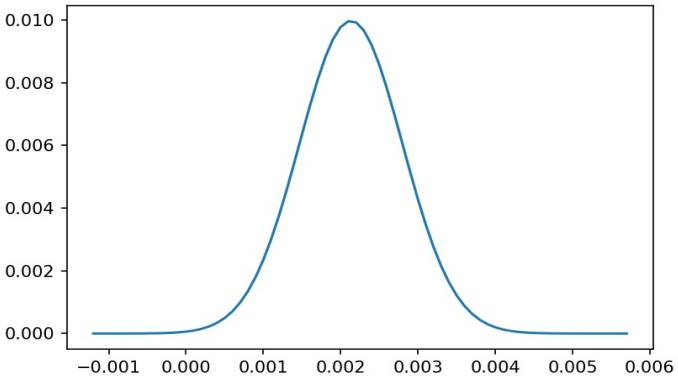


Figure 2.4: XRP Marginal Predictive Density Forecast : 0.005 CI (-0.007, 0.013)

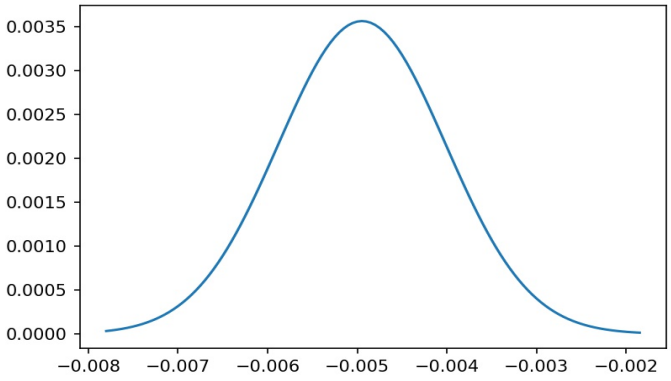


Figure 2.5: XLM Marginal Predictive Density Forecast XLM: CI (-0.0392, 0.0161)

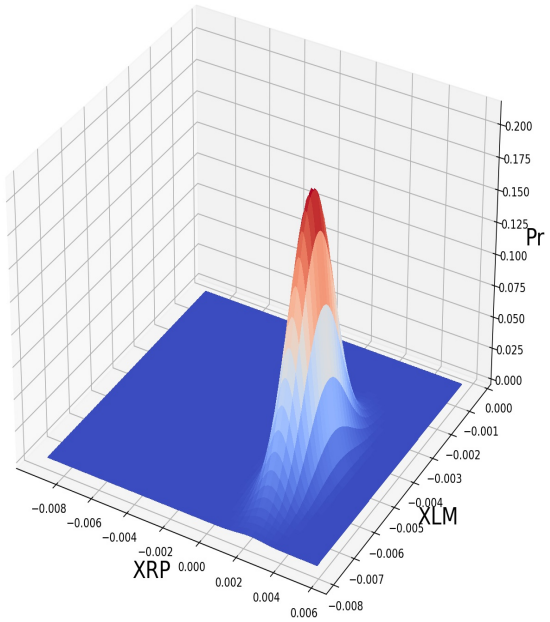


Figure 2.6: XRP/XLM Joint Density

The predictive densities are centred at $y_{1,T+1}$ and $y_{2,T+1}$, respectively. Therefore their modes can be considered as point forecasts $\hat{y}_{1,T+1}$ $\hat{y}_{2,T+1}$. The prediction interval is obtained for the 10th and 90th percentiles. The predictive intervals contain the true values of $y_{1,T+1}$ and $y_{2,T+1}$.

The confidence intervals contain the true values of the processes under consideration and we can see the densities are peaked.

2.3 Linear Approximation Method

The semi-parametric density estimator can be computationally demanding. In this section, we consider a simplified approach that combines the linear prediction method for causal VAR models with the consistent estimation of the autoregressive coefficient matrix.

Let us briefly recall the traditional causal VAR(1) model for the process Y_t of dimension n:

$$Y_t = \Phi Y_{t-1} + \epsilon_t \tag{2.3.5}$$

where $\epsilon_t, t = 1, ..., T$ is a i.i.d. Gaussian multivariate white noise with mean vector 0 and variance matrix Σ. We also assume that ϵ_t is uncorrelated with Y_{t-1} : $Cov(\epsilon_t, Y_{t-h}) = 0,\ \forall h \geq 1$, and Y_t is stationary. The last statement implies that matrix Φ has all eigenvalues of modulus less than 1. From the i.i.d. condition on the noise, it follows that $E(\epsilon_t|\underline{\epsilon}_{t-1}) = 0$. Also, the process being weakly stationary has a

one-sided (past dependent) moving average representation $Y_t = \sum_{h=0}^{\infty} \Phi^h \epsilon_{t-h}$. Then, there is an equivalence of information sets $(\underline{\epsilon}_{t-1}) = (\underline{\epsilon}_{t-1}, \underline{Y}_{t-1})$ and we have:

$$E(\epsilon_t|\underline{\epsilon}_{t-1}) = E(\epsilon_t|\underline{\epsilon}_{t-1}, \underline{Y}_{t-1}) = 0 => E(\epsilon_t|\underline{Y}_{t-1}) = 0.$$

For one-step ahead out-of-sample predictions of Y_{T+1} made at time T, we use the prediction formula $\hat{Y}_{T+1} = \Phi Y_T$, which is an unbiased forecast because in the formula:

$$E(Y_{T+1}|\underline{Y}_T) = \Phi Y_T + E(\epsilon_{T+1}|\underline{Y}_T) \tag{2.3.6}$$

where the last term is 0. Then, the classical normality-based forecast intervals can be calculated. Because of the eigenvalues of matrix Φ being inside the unit circle, the forecast converges to the marginal zero mean of Y_T when the forecast horizon increases and the forecast error variance converges to the marginal variance of Y_T.

In the context of a mixed VAR model, the last term of formula (2.3.6) is not equal to zero. This is because, in the strictly stationary mixed VAR process where $\tilde{\epsilon}_t, t = 1, ..., T$ is an i.i.d. non-Gaussian multivariate white noise with mean vector 0 and variance matrix Ω, the condition of zero correlation between $\tilde{\epsilon}_t$ and Y_{t-1} does not hold and $\tilde{\epsilon}_t$ does not satisfy the martingale difference sequence properties (unlike the causal error). More specifically,

$$E(\tilde{\epsilon}_t|\underline{\tilde{\epsilon}}_{t-1}) = 0,$$

while

$$E(\tilde{\epsilon}_t|\underline{\tilde{\epsilon}}_{t-1}, \underline{Y}_{t-1}) \neq 0 \text{ and } E(\tilde{\epsilon}_t|\underline{Y}_{t-1}) \neq 0.$$

Therefore, the linear forecast formula (2.3.6) does not provide an unbiased forecast and is not optimal in the context of the mixed VAR. However, one can hope that at horizon 1 the forecast error is not too large and this shortcoming is offset by the simplicity of the formula.

In practice, matrix Φ is unknown and replaced by an estimator. We know that the OLS-based or equivalently the Gaussian maximum likelihood method always estimate the Φ matrices with eigenvalues inside the unit circle. Therefore, in application to a model where the true matrix Φ has eigenvalues outside the unit circle, the OLS and Gaussian ML are inconsistent with respect to the true Φ_0. Nevertheless, these estimators converge as T $\to \infty$ to an autoregressive matrix of a linear approximation to the non-linear mixed VAR process, which approximates "arbitrarily well" the nonlinear dynamics, according to the Wold theorem.

In contrast, the GCov estimator of the autoregressive matrix is consistent in application to either the causal or noncausal VAR model. It provides a reliable estimate of matrix Φ regardless of whether its roots are located inside or outside the unit circle.

So far we know the following forecast properties given in Table 2.1 below:

Method	Estimator of Φ	Property of forecast	Prediction interval
OLS + linear forecast	inconsistent	"optimal" linear	incorrect
GCov + linear forecast	consistent	unknown	unknown
GCov + nonlinear forecast	consistent	optimal	correct

Table 2.1: Estimation of Noncausal VAR

In this section we explore the use of the linear approximation to the process and its simple forecast formula with the matrix Φ consistently estimated by the GCov. We estimate $\Phi(t)$ as a deterministic function of time in a possibly non-stationary and nonlinear model

$$Y_t = \Phi(t)Y_{t-1} + \xi_t;$$

where ξ_t is a multivariate i.i.d. error with mean 0 and variance $\Sigma(t)$ and the eigenvalues of $\Phi(t)$ are assumed different from 1. To implement this approach we use the rolling estimation with a fixed window and estimate the VAR(1) model locally from the GCov estimator, providing the value of $\hat{\Phi}(t)$ with eigenvalues different from 1. Next we calculate the linear forecasts one step ahead at each T_r where T_r denotes the last point in the window $r = 1, ..., R$ so that T/R is the window length:

$$\hat{Y}_{T_R+1} = \hat{\Phi}(T_R)Y_{T_R}. \tag{2.3.7}$$

This is a linear forecasting procedure applied locally by rolling. In this context, the rolling is equivalent to the use of a rectangular kernel in the semi-parametric estimation, under the assumption that Y_t is locally stationary [see Dahlhaus et al.

(2019)]. It provides a local linear approximation of the future value of the model at the end of each window.

2.3.1 One Step Ahead Linear Forecasts

In our application, we use the rolling approach with the window of one period ahead and estimate the VAR(1) model over that window from the GCov estimator.

We first calculate the linear forecasts of bivariate VAR(1) and VAR(3) models of BTC and ETH 20 periods ahead. We choose H in (1.2.19) such that it minimizes the Mean Absolute Error (MAE) between the forecast and the actual value. The one step ahead out of sample predictions of the bivariate VAR(1) models of BTC/ETH are plotted in Figure 2.7 with the Mean Absolute Errors for the forecasts given in Table 2.2. The one step ahead out of sample predictions of the bivariate VAR(3) models of BTC/ETH are plotted in Figure 2.8 with the Mean Absolute Errors for the forecasts given in Table 2.3.

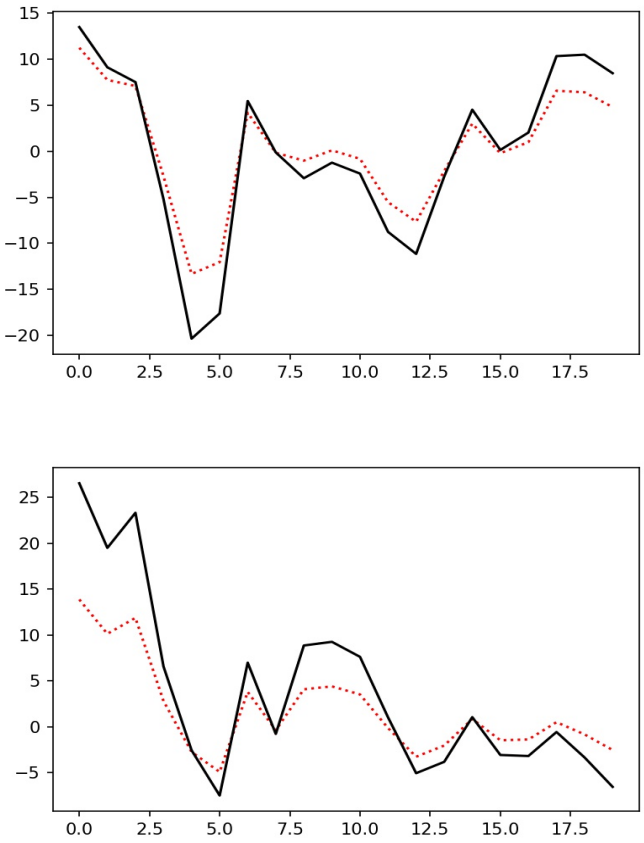


Figure 2.7: BTC and ETH VAR(1) Forecasts

BTC	ETH
2.35	3.66

Table 2.2: BTC/ETH VAR(1) MAE

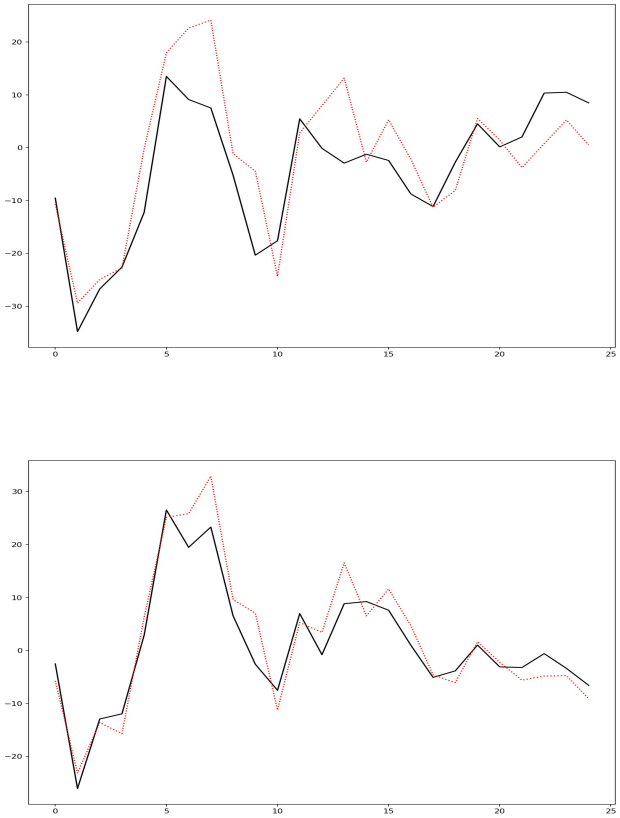


Figure 2.8: BTC and ETH VAR(3) Forecasts

BTC	ETH
6.43	3.46

Table 2.3: BTC/ETH VAR(3) MAE

We then calculate the linear forecasts of bivariate VAR(1) and VAR(3) models of XRP and XLM 20 periods ahead. We choose H in (1.2.19) such that it minimizes the Mean Absolute Error (MAE). The one step ahead out of sample predictions of the bivariate VAR(1) models of XRP/XLM are plotted in Figure 2.9 with the Mean Absolute Errors for the forecasts given in Table 2.4. The one step

ahead out of sample predictions of the bivariate VAR(3) models of XRP/XLM are plotted in Figures 2.10 with the Mean Absolute Errors for the forecasts given in Table 2.5.

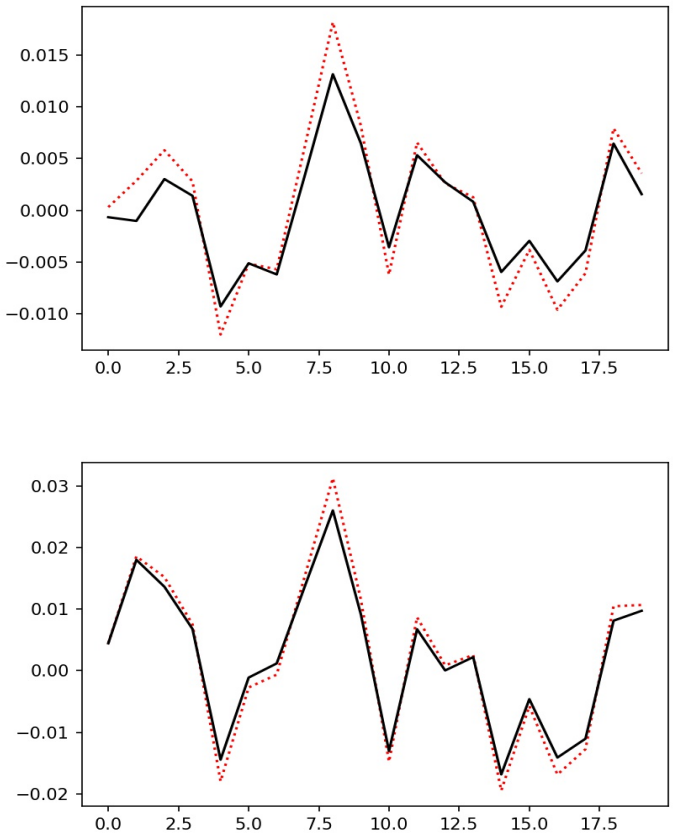


Figure 2.9: XRP and XML VAR(1) Forecasts

XRP	XLM
0.0019	0.0018

Table 2.4: XRP/XLM VAR(1) MAE

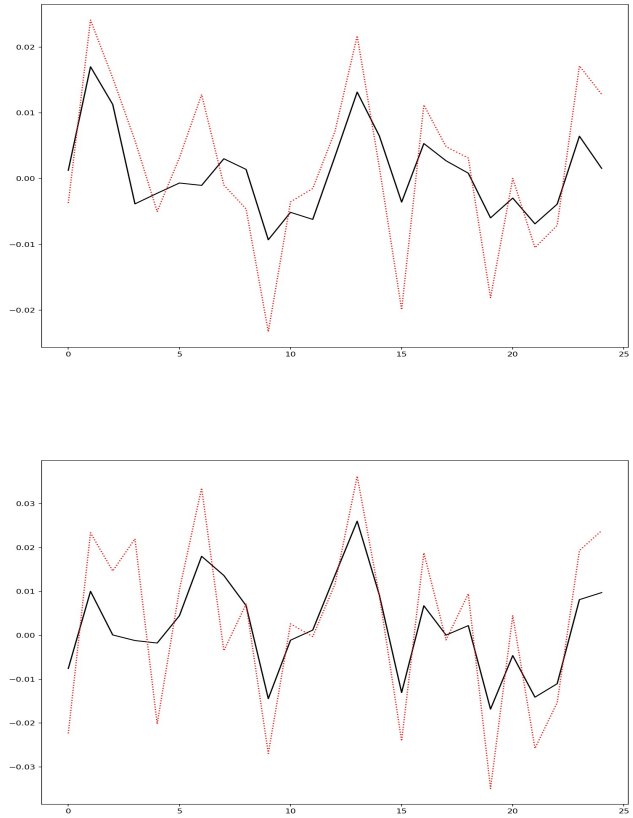


Figure 2.10: XRP and XML VAR(3) Forecasts

XRP	XLM
0.0035	0.0034

Table 2.5: XRP/XLM VAR(3) MAE

Next we calculate the linear forecasts of the VAR(1) containing all four cryptocurrencies. The data is demeaned but not detrended and is the closing daily USD exchange between 2018-06-01 and 2018-08-30. We look at the one-step ahead forecast for 40 periods on the demeaned data for the four cryptocurrency exchange rates (against the US dollar). The model is a VAR(1) with $n = 4$ since all four

cryptocurrencies are used. We use a rolling window of one period ahead. The ACFs and QQ plots for the demeaned series can be found in Supplementary Graphs The forecasts and the original data are shown in Figure 2.11.

From the Hill Plots in Figure B.8 Supplementary Graphs we can see that the demeaned daily closing USD exchange rates display heavy tails, but the tails are not extremely heavy.

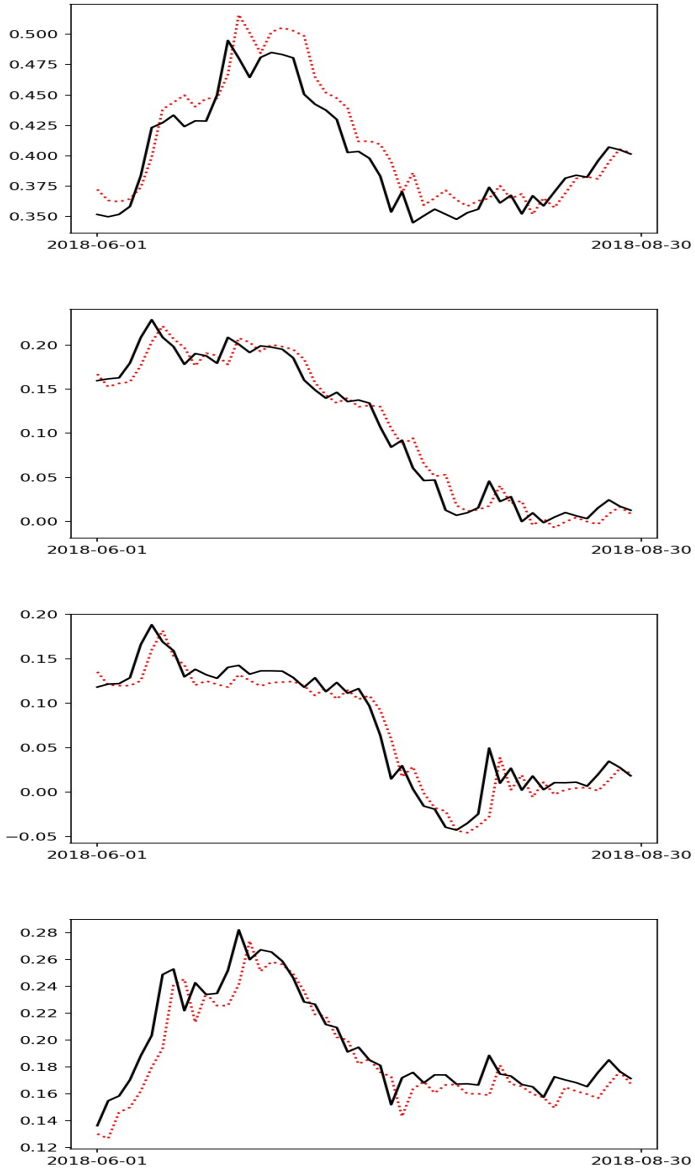


Figure 2.11: Forecast Demeaned BTC, ETH, XRP, XLM, one step, out of sample

BTC	ETH	XRP	XLM
0.24	0.013	0.074	0.012

Table 2.6: BTC/ETH/XRP/XLM MAE

2.4 Conclusion

In this chapter we used the decomposition of multivariate time series processes into its causal and noncausal components for the purposes of forecasting the cryptocurrency US dollar exchange rates for the pairs Bitcoin/Ethereum (BTC/ETH) and Ripple/Stellar (XPR/XLM) since these processes exhibit evidence of locally explosive behaviour in strictly stationary processes. We modelled these pairs as bivariate series as well as a single series of dimension four in which are cryptocurrency exchange rates were modelled as a single VAR(1).

We derived formulas proposed in Gouriéroux and Jasiak (2016) using the latent causal and noncausal components to calculate predictive densities for mixed causal noncausal VAR(1) models which we use to forecast one step ahead forecasts. We calculated predictive densities for the cryptocurrency exchange rates under investigation using a kernel estimator of the predictive densities and, in addition, calculated one step ahead out of sample forecasts of the cryptocurrency exchange rates using a linear approximation method which is computationally less demanding than the predictive density approach. We calculate the Mean Absolute Error for the forecasts and find that the forecasting methods yield reasonable results.

Chapter 3

Predictive Densities

3.1 Introduction

In this chapter we present theoretical results showing how to calculate a predictive density for a mixed causal noncausal bivariate VAR(3) model. We then show how to calculate a predictive density for a VAR(1) model with more than two dimensions, i.e. for $n > 2$ where n is the number of series included in the mixed causal noncausal VAR model. Finally, we conduct a simulation study to investigate the confidence intervals obtained for forecasts determined from the quantiles of marginal predictive densities.

We conduct two experiments via simulation study. Firstly we conduct a simulation study whose aim to is test the coverage of the predictive density method of forecasting which finds that the 90 percent confidence intervals estimated using the predictive density forecasting method contains the true value of the simulated

process being forecast in at least 90 percent of the trials.

In our second experiment we use simulated data to compare the accuracy of the predictive density estimation method compared to the linear approximation method. We find that the simulation method outperforms the linear approximation method in terms of the mean squared error of the forecast but note that the linear approximation method is less computationally demanding.

3.2 Predictive Density for VAR(3)

The VAR(3) model for the bi-variate series y_t is

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \Phi_3 y_{t-3} + \epsilon_t \tag{3.2.1}$$

where the bi-variate process $\epsilon_t, t = 1, ..., T$ is i.i.d. with mean vector 0 and variance-covariance matrix Σ.

We define the process X_t:

$$X_t = \begin{bmatrix} y_t' \\ y_{t-1}' \\ y_{t-2}' \end{bmatrix}$$

and rewrite the model (1) as

$$X_t = \Psi X_{t-1} + u_t,$$

where

$$\Psi = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi'_{11} & \phi'_{12} & \phi''_{11} & \phi''_{12} \\ \phi_{21} & \phi_{22} & \phi'_{21} & \phi'_{22} & \phi''_{21} & \phi''_{22} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

is the 6×6 matrix of autoregressive coefficients and error u_t is defined as

$$u_t = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Suppose the matrix $\Psi = B \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix} B^{-1}$

where J_1 is a 4×4 matrix of eigenvalues of modulus less than 1 and J_2 is a 2×2 matrix of eigenvalues of modulus greater than 1, see Section 1.3.5.

Then $X_t = B_1 X_{1t}^* + B_2 X_{2t}^*$ where

$$X_{1t}^* = B^1 X_t$$

is of dimension $4 \times T$ and

$$X_{2t}^* = B^2 X_t$$

is of dimension $2 \times T$ using the notation of section 2.4.

The error vector $u_{1t}^* = B^1 u_t$ is of dimension $4 \times T$: .

$$u_{1t}^* = \begin{bmatrix} b^{11} & b^{12} & b^{13} & b^{14} & b^{15} & b^{16} \\ b^{21} & b^{22} & b^{23} & b^{24} & b^{25} & b^{26} \\ b^{31} & b^{32} & b^{33} & b^{34} & b^{35} & b^{36} \\ b^{41} & b^{42} & b^{43} & b^{44} & b^{45} & b^{46} \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} b^{11}\epsilon_{1t} + b^{12}\epsilon_{2t} \\ b^{21}\epsilon_{1t} + b^{22}\epsilon_{2t} \\ b^{31}\epsilon_{1t} + b^{32}\epsilon_{2t} \\ b^{41}\epsilon_{1t} + b^{42}\epsilon_{2t} \end{bmatrix}$$

and is a function of ϵ_{1t}, ϵ_{2t}.

The error vector $u_{2t}^* = B^2 u_t$ is of dimension $2 \times T$:

$$u_{2t}^* = \begin{bmatrix} b^{51}\epsilon_{1t} + b^{52}\epsilon_{2t} \\ b^{61}\epsilon_{1t} + b^{62}\epsilon_{2t} \end{bmatrix}$$

We have:

$$X^*_{1t} = J_1 X^*_{t-1} + u^*_{1t}$$

$$X^*_{2t} = J_2 X^*_{t-2} + u^*_2 t$$

Let the future values of $X^*_{2,T+1}$ be denoted by x^*_2. Then the predictive density can be written as

$$l(y|I_T)\frac{l_2(x^*_2)}{l_2(X^*_{2,T})}|detJ_2|g^*(X^*_{1,T+1} - J_1 X^*_{1,T}, X^*_{2,T+1} - J_2 X^*_{1,T})$$

with densities

$$l_2\left(B^2\begin{bmatrix} y_1 \\ y_2 \\ y_{1,T} \\ y_{2,T} \\ y_{1,T-1} \\ y_{2,T-1} \end{bmatrix}\right)$$

and

$$l_2\left(B^2\begin{bmatrix} y_{1,T} \\ y_{2,T} \\ y_{1,T-1} \\ y_{2,T-1} \\ y_{1,T-2} \\ y_{2,T-2} \end{bmatrix}\right)$$

which are the bivariate densities of the noncausal linear combination. The density $g^*(\epsilon^*_{1,T}, \epsilon^*_{2,T})$ is bi-variate and involves the following transformations.

Let us use the generic notation $x^* = [X_{1,T+1}, X_{2,T+1}]$.

The predictive density can be written as

$$l(x^*|I_T) = \frac{l_2(x^*_2)}{l_2(x^*_{2,T})}|detJ_2|g^*(x^*_1 - J_1x_{1,T}, x^*_2 - J_2x_{2,T})$$

where

$$g^*(u^*_t) = g^*(B^{-1}u_t) = |detB^{-1}| \cdot g(\epsilon_t) = \frac{1}{|det(B)|}g(\epsilon_t),$$

since $detB^{-1} = \frac{1}{det(B)}$.

Because we have the density of observed variables

$$\frac{1}{detB}l(y|I_T) = \frac{l_2\left(B^2\begin{bmatrix} y \\ y_T \\ y_{T-1} \end{bmatrix}\right)}{l_2\left(B^2\begin{bmatrix} y_T \\ y_{T-1} \\ y_{T-2} \end{bmatrix}\right)} \cdot \frac{1}{det(B)}g(\epsilon_1, \epsilon_2)$$

as the determinants cancel out on both sides.

The bivariate error density can be estimated from

$$\hat{g}(\epsilon_1, \epsilon_2) = \frac{1}{T}\frac{1}{h_1 h_2}\sum_{t=1}^{T} K\left(\frac{\epsilon_1 - (y_{1t} - \phi_{11}y_{1,t-1} - \phi_{21}y_{2,t-1})}{h_1}\right) K\left(\frac{\epsilon_2 - (y_{2t} - \phi_{12}y_{1,t-1} - \phi_{22}y_{2,t-1})}{h_2}\right)$$

and the second component density is

$$\hat{l}_2(x_2) = \frac{1}{T}\frac{1}{h_{11} \cdot h_{22}}\sum_{t=1}^{T} K\left(\frac{x_{21} - x_{21,t}}{h_{11}}\right) \cdot K\left(\frac{x_{22} - x_{22,t}}{h_{22}}\right)$$

where h_{11} and h_{12} are the bandwidths adjusted for the variation of the two components of bivariate series $x^*_{2,t}$, which are denoted by $x^*_{21,t}$ and $x^*_{22,t}$ in the formula above.

3.3 Predictive Density for VAR(1) with n > 2

Suppose that $n = 4$. In the cryptocurrency model estimated in the section VAR(1) For Bitcoin, Ethereum, Ripple and Stellar we found one eigenvalue outside the unit circle. Then, the density is l_2 is unidimensional, and can be easily estimated using a Kernel estimator. However, the densities $g(\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4)$ is of a series of dimension 4 and is difficult to estimate using a standard kernel approach.

When n is large the nonparametric model of $g(\epsilon_t)$ becomes ineffective in practice due to large estimation errors. For example, the rate of convergence of the Mean Integrated Squared Error (MISE) of the kernel density estimator (KDE) is slow for $d > 2$.

Modl (2020) reports the following methods listed in Table 3.1 available for multidimensional density estimation.

Method	R-package
RODEO	lpderodeo
KDE with normal Kernel	base, KS
MASS procedure for KDE	2d
normal mixture	mvmesh and grDevices
nearest neighbour	KDE with uniform Kernal adaptive bandwidth
decision tree	detpack
nine compute	Kdevine
smoothing spline density	gss

Table 3.1: Available Methods for Multi-dimensional Density Estimation

The regularization of the derivative expectation operator (RODEO) developed by Lafferty and Wasserman (2008) is a technique for regularizing the bias reduction associated

with a bandwidth reduction to avoid bandwidth over fitting. This increases the rate of the MISE (Mean Integrated Squared Error) convergence in high dimensional data. When the unknown density function satisfies some suitably defined sparsity conditions, the RODEO avoids the curse of dimensionality and achieves an optimal convergence.

The assumption is, for a multivariate density g we can write:

$$g(x_1, ..., x_n) = g(R_R) \cdot b(x_1, ..., x_n),$$

where R_R is a subset of 1, ...,n and b is a baseline density. For example, g can be either the subset of causal or noncausal errors $e_t = y_t - \Phi y_{t-1}$. If the function b is a multivariate baseline density such as $b(x) =$ uniform, then $g(x) = g(x_R)$ and the density of all errors is approximated from the density of the causal or noncausal errors only.

For the bandwidth matrix $H = diag(h_1, ..., h_n)$ we have

$$\hat{f}_H(x) = \frac{1}{Tdet(H)} \sum_{t=1}^{T} K(H^{-1}(x - x_i)) = \frac{1}{T} \sum_{t=1}^{T} \prod_{j=1}^{h} \frac{1}{h_j} K\left(\frac{x_i - x_{i,j}}{h_j}\right).$$

The kernel density estimator starts with a large bandwidth which gets reduced each time $\frac{\partial \hat{f}_H(x)}{\partial h_j}$ is large.

If a Normal baseline used, with diagonal Σ, the density estimator is

$$\bar{f}_H(x) = \frac{\hat{b}(x) \sum_{i=1}^{T} K_H(x_t - x)}{T \int K_H(u - x) \hat{b}(u) du},$$

where $\hat{b(x)}$ is a parametric density estimate at point x estimated by maximum likelihood.

Suppose $b(x)$ is a multivariate normal density with diagonal variance-covariance matrix. We use the Gaussian kernel as follows

$$\bar{f}_H(x) = \frac{1}{T}\sum_{t=1}^{T}\prod_{j=1}^{h} K\left(\frac{x_{ij}-x_j}{h_j}\right)\cdot\sqrt{\frac{|H-\hat{\epsilon}|}{|\hat{\epsilon}|}}exp\left\{-\frac{(x-\hat{\mu})'(\widehat{\Sigma}^{-1}-(H+\hat{\Sigma})^{-1})(x-\hat{\mu})}{2}\right\},$$

where $\hat{\mu}$ and $\hat{\Sigma}$ are MLE for normal distributions, see Lui et al. (2007) p.5.

In a multivariate VAR(1) model the statistical inference can be made without estimating the univariate error density. More specifically, we can calculate the probability of a component taking values less than a given constant z as follows

$$P[Y_{1,T+1} < z|Y_T] = E[\mathbb{1}_{Y,T+1} < z|Y_T] =$$

$$\int_w ... \int \mathbb{1}_{[y_1<z]}\frac{l_2(A^2y)}{l_2(A^2y_T)}|det(J_2)|g(y-\Phi y_T)dy =$$

$$\frac{|detJ_2|}{l_2(A^2y_T)}\int ... \int [\mathbb{1}_{[y_1<z]}l_2(A^2y)]g(y-\Phi y_T)dy =$$

$$\frac{|detJ_2|}{l_2(A^2y_T)}\int ... \int \left[\mathbb{1}_{[\epsilon_1+\Phi_1 y_T<z]}l_2(A^2(\epsilon+\Phi y_T))\right]g(\epsilon)d\epsilon =$$

$$\frac{|detJ_2|}{l_2(A^2y_T)}E_\epsilon\left[\mathbb{1}_{[\epsilon_1<z-\Phi_1 y_T]}l_2[A^2(\epsilon+\Phi y_T)]\right] =$$

$$\frac{|detJ_2|}{\hat{l}_2(A^2 y_T)} \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_{[\hat{\epsilon}_{1t} < z - \Phi_1 y_T]} \hat{l}_2[A^2(\hat{\epsilon}_t + \Phi y_T)],$$

which requires knowledge of $\hat{l}_2$ only.

3.4 Simulation Study

3.4.1 Point and Interval Forecast Performance

We conduct simulation experiments to assess the performance of forecasting bivariate causal noncausal series using predictive densities defined in the previous chapter. Our experiments are designed as follows. We simulate two mixed causal noncausal bivariate processes with $T = 1000$ observations and t-distributed errors with 4 and 6 degrees of freedom, respectively.

In both cases the simulated series has the following autoregressive matrix

$$\begin{pmatrix} 0.7 & -1.3 \\ 0 & 2 \end{pmatrix}.$$

This matrix has eigenvalues 0.7 and 2, with one eigenvalue outside the unit circle and one within the unit circle which is consistent with a mixed causal noncausal process. We consider two experiments generating errors with different distributions for each experiment. The errors in the bivariate VAR models are i.i.d and t(4) and t(6) distributed. Matrix Σ is the variance-covariance matrix of errors and it is assumed to be an identity matrix.

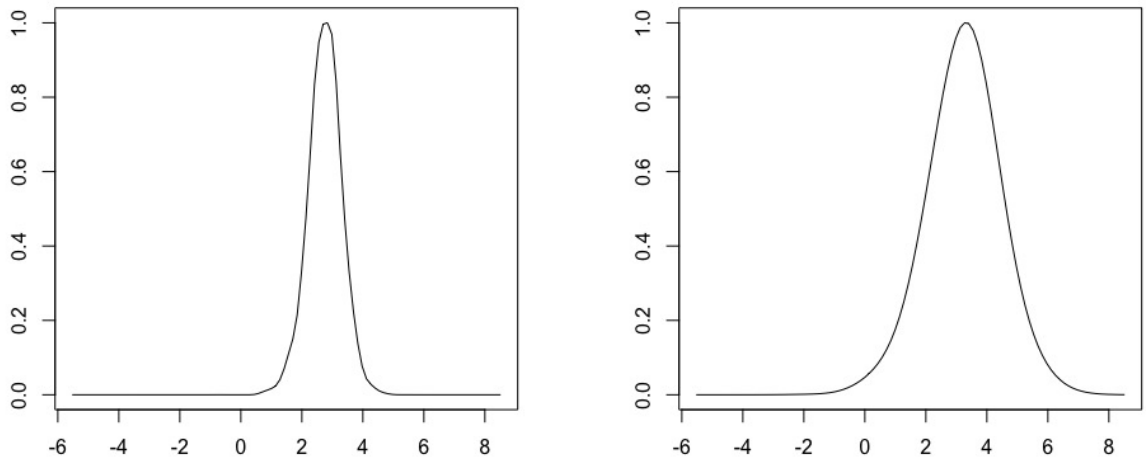
When generating the paths we simulate $T = 1200$ observations and remove 100 observations from the beginning and end of the simulated series.

In each case the simulated error series is divided by its standard deviation in order to normalize the variance between simulated data with t-distributed errors with different degrees of freedom (4 and 6 degrees of freedom for the two series respectively).

The bandwidths h_1 and h_2 in 2.2.4 are selected using a rule-of-thumb for choosing the bandwidth of a Gaussian kernel density estimator. The default is 0.9 times the minimum of the standard deviation of the residuals times the interquartile range divided by 1.34 times negative one-fifth power of the sample size Silverman (1986).

The bandwidths h_{12} and h_{22} in the kernel density estimator are set equal to the standard deviation of the residuals.

We then calculate the bivariate predictive density based forecast of 1000th observation, calculate the 90th percentile confidence interval based on the marginal densities of the predictive density and record whether or not the true value lies in the aforementioned confidence interval. We repeat this process 10 thousand times for the bivariate series simulated with t-distributed errors with 4 degrees of freedom and for the bivariate series simulated with t-distributed errors with 6 degrees of freedom using the same autoregressive matrix. Each time we compute the 90 percent prediction interval based on the predictive densities for each marginal density of series 1 and 2. An example of marginal predictive densities is given below in Figure 3.1.



(a) Marginal Density Series 1

(b) Marginal Density Series 2

Figure 3.1: Marginal Predictive Densities Forecasting the 1000th Observation of Simulated Bivariate Series with t-Distributed Errors (6 Degrees of Freedom)

In the first experiment, in which the simulated errors are t-distributed with 4 degrees of freedom, we find that 90th percentile confidence interval contains the true value of the final (1000th) observation of the series in 97.2 percent of trials for the first series and 91.6 percent of the time for the second series.

In the second experiment, in which the simulated errors are t-distributed with 6 degrees of freedom, we find that the 90th percentile confidence interval contains the true value of the final (1000th) observation of the series in 97.2 percent of trials for the first series and 91.7 percent of the time for the second series.

Thus in both cases the true value being forecast lies within the 90th percentile confidence interval in over 90 percent of the simulated trials.

3.4.2 Point and Interval Forecasts from Linear Approximation

Next we investigate the performance of the linear approximation method for point prediction. We compare the Mean Squared Errors (MSE) of the forecasts for the linear approximation method against the forecasts obtained using predictive densities as shown in Table 3.2.

D of F	MSE Linear Approximation	MSE Predictive Density
4	MSE(y_1) = 0.98, MSE(y_2) = 2.85	MSE(y_1) = 0.4, MSE(y_2) = 1.71
6	MSE(y_1) = 0.88, MSE(y_2) = 2.44	MSE(y_1) = 0.57, MSE(y_2) = 1.85

Table 3.2: Mean Squared Forecast Error for each series.

The MSE are calculated from 10000 replications of the same dynamic model and reported below along with the MSE based on the point forecasts obtained using predictive densities. The point forecasts for the predictive density method are obtained by selecting the prediction with the highest probability mode assigned by the predictive density.

The linear approximation method is simpler because there is no need to estimate bandwidths and is computationally easier. [1]

We do not compare this approach with the traditional forecast based on the VAR because we have already done so in chapter 2, more specifically we conclude that the OLS (Gaussian MLE) produces spurious results. The only approach comparable is the univariate "no change" forecast, which does not account comovements due to the fact that

[1]Note that the linear approximation method has no confidence interval yet defined.

it is univariate. The "no change" forecast takes the prior observation as the prediction for the next observation.

Table 3.2 shows that the MSE from the linear approximation are bigger than those from the predictive density method. Nonetheless the linear approximation method is computationally easier. The results therefore suggest that the predictive density approach outperforms the linear approximation approach in terms of the mean squared forecast error since the mean squared error of the forecasts for the predictions obtain using the predictive density are lower than those obtained using the linear approximation method.

3.5 Conclusion

In this chapter we presented theoretical results showing how to calculate a predictive density for a mixed causal noncausal bivariate VAR(3) model. We also presented theoretical results showing how the calculation of a predictive density for a VAR(1) model with more than two dimensions can be achieved.

We conducted two experiments via a simulation study. We first assessed the coverage of the predictive density forecasting method and found that the predictive density estimation method is valid in terms of calculating 90 percent confidence intervals (i.e. the true value being forecast is contained in the 90 percent confidence interval in over 90 percent of the trials). We then compared the accuracy of the predictive density estimation method relative to the linear approximation method and found that the predictive density forecasting method gives more accurate forecasts than the linear approximation method as measured by the mean squared error. We note, however, that the simple linear approximation forecast method has the benefit of being computationally less demanding than the predictive density forecast method.

Conclusion

This dissertation compares traditional causal multivariate time series models and mixed causal noncausal time series models in application to modelling cryptocurrency USD exchange rates and introduces a new linear approximation forecast method applicable to multivariate mixed causal noncausal models. We also investigate the forecasting of cryptocurrency US dollar exchange rates in the context multivariate causal noncausal models. To that end, in addition to the introduction of a linear forecasting method, we provide the predictive density from which we can obtain forecasts and from which forecast intervals for each variable can be easily inferred. We provide new theoretical results by extending theory presented in Gouriéroux and Jasiak (2016) .

In chapter 1 we conclude that, for the series under investigation (namely Bitcoin (BTC), Ethereum (ETH), Ripple (XRP) and Stellar (XLM)), mixed causal noncausal models capture salient dynamics of these time series, including nonlinear dynamics within and comovements between the series. The research shows that mixed multivariate causal noncausal model, estimated semi-parametrically, is able to detect comovements where traditional OLS based VAR models cannot.

By modelling these processes as mixed (i.e. containing both causal and noncausal components) we capture nonlinear dependencies not accounted for by standard VAR

models. We are able to show that comovements between cryptocurrencies can be empirically detected using mixed causal noncausal models while traditional models fail to detect these comovements.

In Chapter 2 we forecast cryptocurrencies jointly as bivariate processes using their noncausal properties in the presence of bubble phenomena. If series share a common noncausal component we can forecast these series jointly, taking the common bubble co-movements into account. We calculated forecasts and confidence intervals for the cryptocurrency pairs Bitcoin/Ethereum and Ripple/Stellar. We obtain one step ahead forecasts using a causal noncausal VAR model estimated via the GCov Estimator and examined the validity of a simple linear approximation procedure. The linear approximation method combines the linear prediction method for casual VAR models with the consistent estimation of the autoregressive coefficient matrix.

In Chapter 3 we present two theoretical results pertaining to the calculation of predictive densities for mixed causal noncausal VAR models and conduct a simulation study in order to examine the performance of the two prediction methods; one using a predictive densities using a kernel density estimator and the other using a linear approximation forecasting method.

We obtain theoretical results for the calculation of predictive densities for a bivariate mixed causal noncausal VAR model with three lags using kernel density estimators, outlining how this can be accomplished using the latent causal noncausal components of the process rather than the observed components of the process. We implement this new method of predictive density computation for multivariate noncausal processes and investigate its properties via simulation study.

In our simulation study we investigate the validity of the 90 percent confidence

intervals obtained when calculating predictive densities and find that the true value is contained within the estimated confidence interval in over 90 percent of the trials. We compare the Mean Square Error for the simple linear approximation method of one step ahead forecasting and the method using predictive densities and find that the predictive density method outperforms the simple linear approximation method.

Bibliography

Blanchard, O. and Watson, M. (1982). *Bubble, Rational Expectations and Financial Market.* Lexington Books, Lexington, MA.

Breid, F., Davis, R. A., Lh, K.-S., and Rosenblatt, M. (1991). Maximum Likelihood Estimation for Noncausal Autoregressive Processes. *Journal of Multivariate Analysis*, 36(2):175–198.

Breidt, F. and Davis, R. (1992a). Time-reversibility, Identifiability, and Independence of Innovations for Stationary Time Series. *J. of Time Series Analysis*, (13):377–390.

Breidt, F., Davis, R., and Dunsmuir, W. (1992). *On Backcasting in Linear Time Series Models, New Directions in Time Series Analysis, Part I.* Springer-Verlag.

Breidt, F., Davis, R., and Lii, K. (1990). Nonminimum Phase Non-Gaussian Autoregressive Processes. *Proc. Natl. Acad. Sci.*, (87):179–181.

Breidt, J. and Davis, A. (1992b). Time-Reversibility, Identifiability and Independence of Innovations for Stationary Time-Series. *Journal of Time Series Analysis*, 13:377–390.

Brockwell, P. and Davis, R. (1987). *Time Series: Theory and Methods.* Springer.

Cambanis, S. and Fakhre-Zakeri, I. (1994). On Predictions of Heavy Tailed Autoregressive Sequences: Forward Versus Reversed Time. *Theory Prob. Appl*, (39):217–233.

Catania, L., Grassi, S., and Ravazzolo, F. (2019). Forecasting cryptocurrencies under model and parameter instability. *International Journal of Forecasting*, 35(2):485–501.

Dahlhaus, R., Richter, S., and Wu, W. B. (2019). Towards a General Theory for Nonlinear Locally Stationary Processes. *Bernoulli*, 25(2):1013–1044.

Davis, R. A. and Song, L. (2020). Noncausal Vector AR Processes with Application to Economic Time Series. *Journal of Econometrics*, 216(1):246–267. Annals Issue in Honor of George Tiao: Statistical Learning for Dependent Data.

Fries, S. (2019). Conditional Moments of Noncausal α-Stable Markov Processes and the Prediction of Bubble Burst Odd. *arXiv*. preprint: 1805.05397, revised December 2019.

Fries, S. and Zakoian, J. M. (2019). Mixed Causal-Noncausal AR Processes and the Modelling of Explosive Bubbles. *Econometric Theory*, (35):1–37.

Gourieroux, C. and Monfort, A. (2015). Pricing with Finite Dimensional Dependence. *Journal of Econometrics*, 187(2):408–417.

Gouriéroux, C. and Hencic, A. (2014). *Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rate.*, volume 583. Springer.

Gouriéroux, C. and Jasiak, J. (2016). Filtering, Prediction and Simulation Methods for Noncausal Processes. *Journal of Time Series Analysis*, 37:405–430.

Gouriéroux, C. and Jasiak, J. (2017). Noncausal Vector Autoregressive Process: Representation, Identification and Semi-parametric Estimation. *Journal of Econometrics*, 200:118–134.

Gouriéroux, C. and Jasiak, J. (2018). Misspecification of Noncausal Order in Autoregressive Processes. *Journal of Econometrics*, 205:226–248.

Gouriéroux, C. and Monfort, A. (2014). Revisiting Identification and Estimation in Structural VARMA Models. *CREST*, pages 2014–30.

Gouriéroux, C. and Monfort, A. (204). Infrequent Extreme Risks. *The Geneva Papers on Risk and Insurance Theory*, (29):5–22.

Gouriéroux, C. and Zakoian, J. M. (2017). Local Explosion Modelling by Noncausal Cauchy Autoregressive Process. *Journal of the Royal Statistical Society*, (79):737–756.

Hannen, J. (1973). The Asymptotic Theory of Linear Time-Series Models. *Journal of Applied Probability*, 10(1):130–145.

Hecq, A. and Voisin, E. (2022). Predicting Crashes in Oil Prices During the COVID-19 Pandemic with Mixed Causal-Noncausal Models. *Advances in Econometrics in honour of Joon y. Park*, forthcoming.

Lafferty, J. and Wasserman, L. (2008). Rodeo: Sparse, Greedy Nonparametric Regression. *The Annals of Statistics*, 36(1):28–63.

Lanne, M., L. A. and Luoto, J. (2012). Bayesian Model Selection and Forecasting in Noncausal Autoregressive Models. *Journal of Applied Econometrics*, (27):812–830.

Lanne, M., L. J. and Saikkonen, P. (2012). Optimal Forecasting of Noncausal Autoregressive Time Series. *International Journal of Forecasting*, (28):623–631.

Lanne, M. and Saikkonen, P. (2011a). GMM Estimators with Non-Causal Instruments. *Oxford Bulletin on Economics and Statistics*, 71:581–591.

Lanne, M. and Saikkonen, P. (2011b). Noncausal Autoregressions for Economic Time Series. *Journal of Time Series Econometrics*, 3(3):1941–1928.

Lanne, M. and Saikkonen, P. (2013). Noncausal Vector Autoregression. *Econometric Theory*, 3(3):447–481.

Liu, H., Lafferty, J., and Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. 10:2295–2328.

Lof, M. and Nyberg, H. (2017). Noncausality and the Commodity Currency Hypothesis. *Energy Economics*, (65):424–433.

Lui, H., Lafferty, J., and Wasserman, L. (2007). Nonparametric density estimation in high dimensions using the rodeo. *Procedings from the Eleventh International Conference on Artificial Intelligence and Statistics, PMLR*, pages 283–290.

Ming-Chung, L. and Kung-Sik, C. (2007). Multivariate Reduced-Rank Nonlinear Time Series Modelling. *Statistica Sinica*, 71(1):139–159.

Modl, L. (2020). The RODEO Approach for Nonparametric Density Estimation. Master's thesis, Humboldt Univeristy.

Phillips, P. and Shi, S. (2018). Financial Bubble Implosion and Reverse Regression. *Econometric Theory*, (34):705–753.

Phillips, P., S. S. and Yu, J. (2015a). Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P500. *International Economic Revie*, (56):1043–1075.

Phillips, P., S. S. and Yu, J. (2015b). Testing for Multiple Bubbles: Limit Theory of Real Time Detectors. *International Economic Review*, (56):1079–1134.

Rubin, D. (1988). Using the SIR Algorithm to Simulate Posterior Distribution", in Bayesian Statistics. *Bayesian Statistics*.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London; New York.

Swensen, A. (2022). On Causal and Non-Causal Cointegrated Vector Autoregressive Time Series. *Journal of Time Series Analysis*, (43):178–196.

Appendices:

A Summary Sample Statistics

	BTC	ETH	XRP	XLM
T	250	250	250	250
Mean	5613	317.1	0.41	0.15
Std	3460	263.71	0.4	0.145
Excess Kurtosis	1.52	-0.81	14.74	1.89
Skew	0.96	1.33	3.16	1.27
Min	778.6	8.2	0.005	0.0017
Max	19187	1380	3.38	0.9

Table A.1: Summary Statistics

B Supplementary Graphs

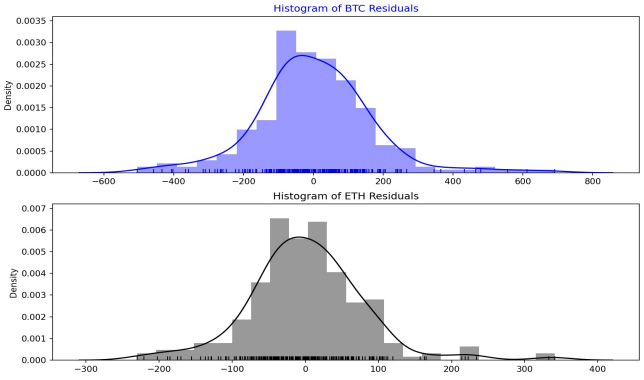


Figure B.1: BTC/ETH Histograms of VAR(3) Residuals

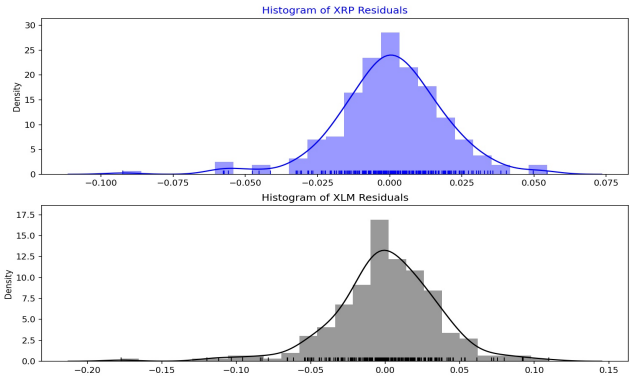
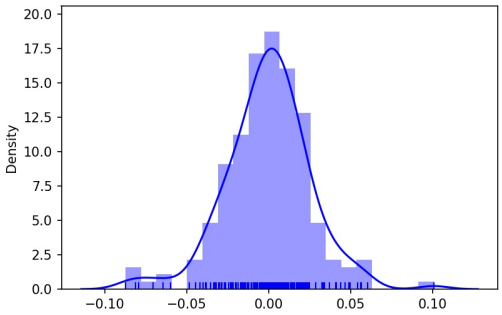
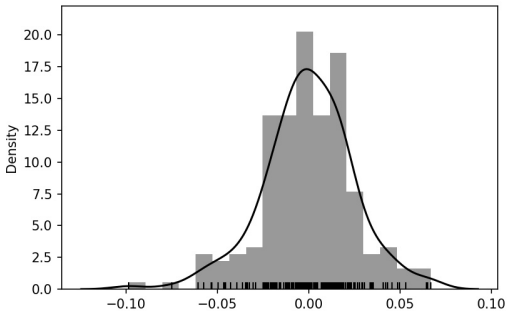


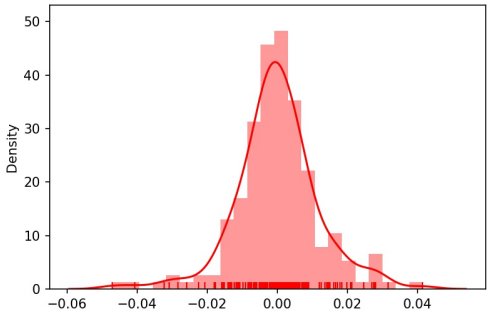
Figure B.2: XRP/XLM Histograms of VAR(3) Residuals



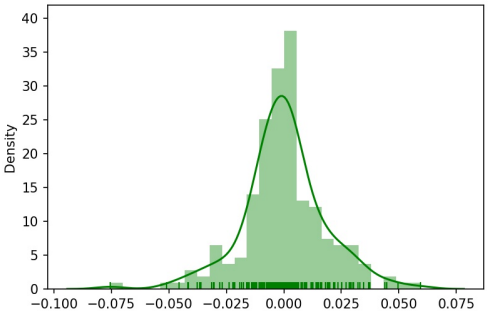
(a) BTC Residuals



(b) ETH Residuals

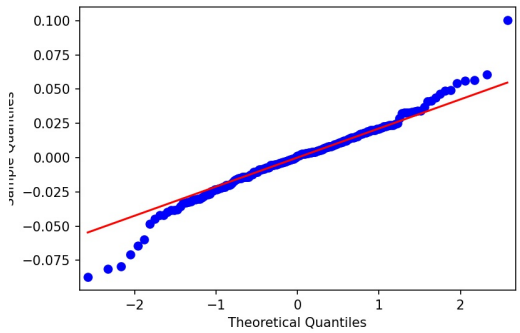


(c) XRP Residuals

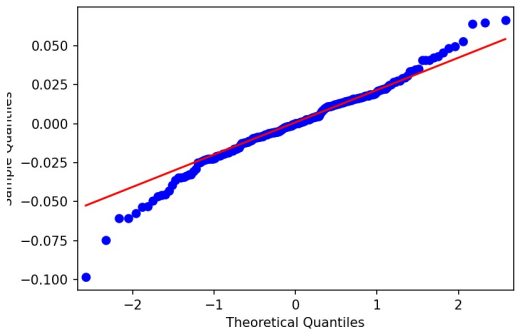


(d) XLM Residuals

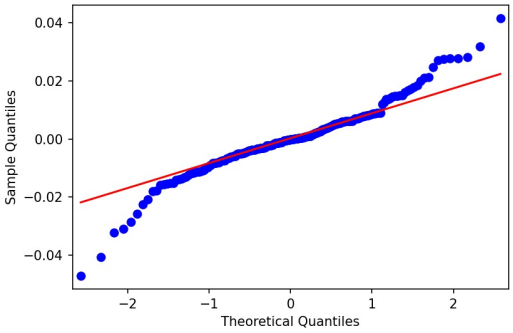
Figure B.3: BTC/ETH/XRP/XLM Histograms of Residuals from VAR(1)



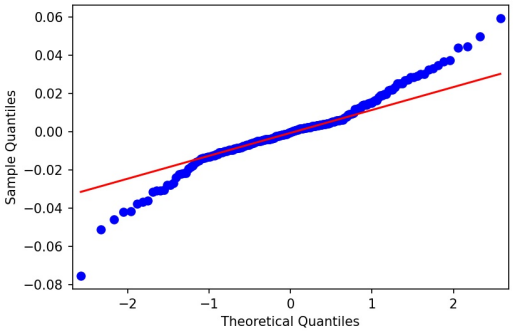
(a) BTC QQ Plot



(b) ETH QQ Plot

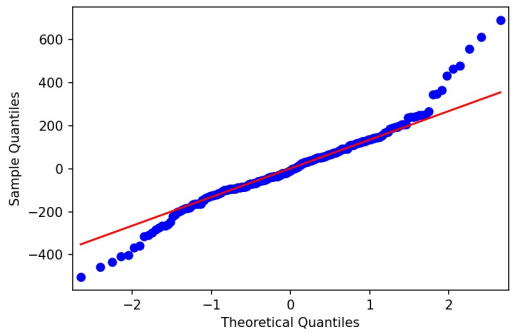


(c) XRP QQ Plot

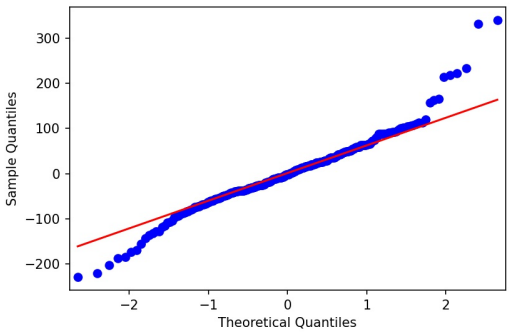


(d) XLM QQ Plot

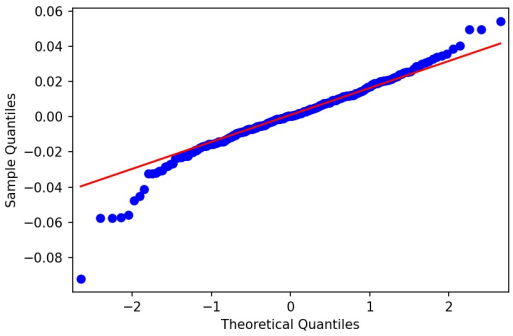
Figure B.4: BTC/ETH/XRP/XLM QQ Plots of Residuals from VAR(1)



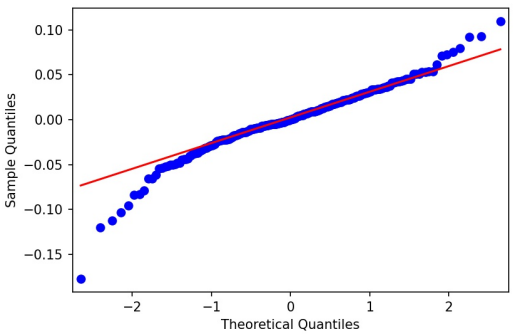
(a) BTC Residuals



(b) ETH Residuals

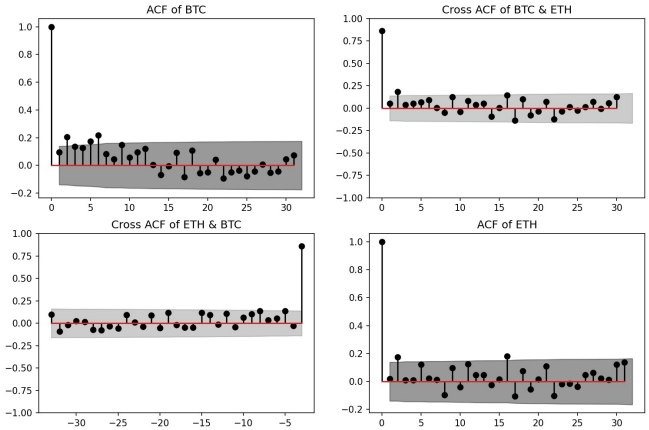


(c) XRP Residuals

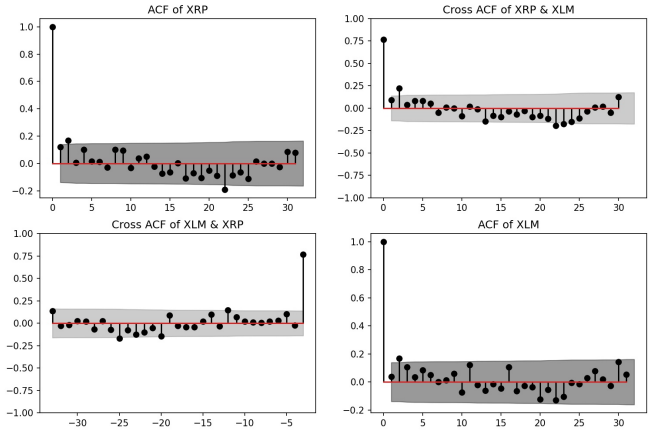


(d) XLM Residuals

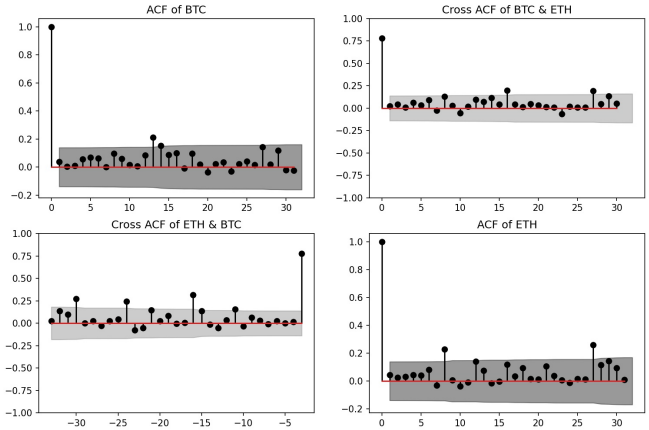
Figure B.5: BTC/ETH/XRP/XLM QQ Plot of Residuals from VAR(3)



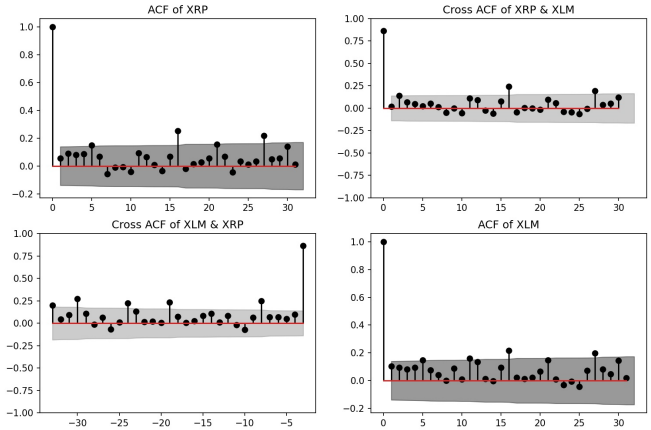
(a) BTC/ETH Residuals ACF



(b) XRP/XLM Residuals ACF

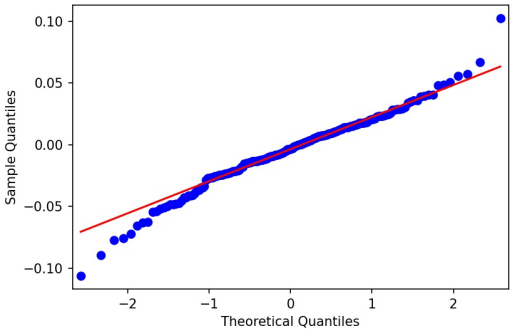


(c) BTC/ETH Squared Residuals ACF

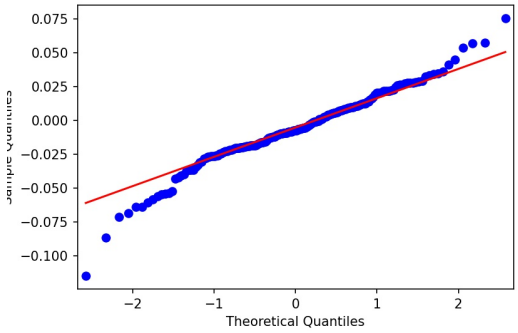


(d) XRP/XLM Squared Residuals ACF

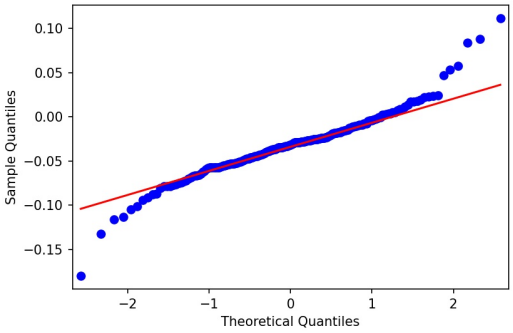
Figure B.6: ACF of Demeaned VAR(1) Residuals and Squared Residuals for BTC, ETH, XRP, XLM pairs (n=4)



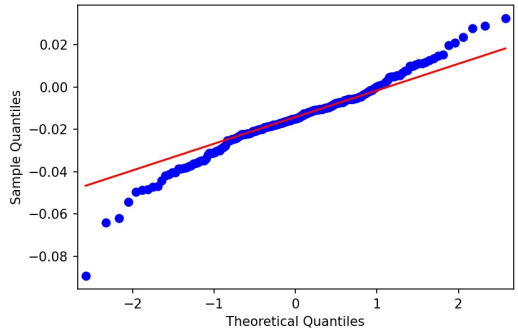
(a) BTC Demeaned VAR(1) Residuals QQ Plot



(b) ETH Demeaned VAR(1) Residuals QQ Plot

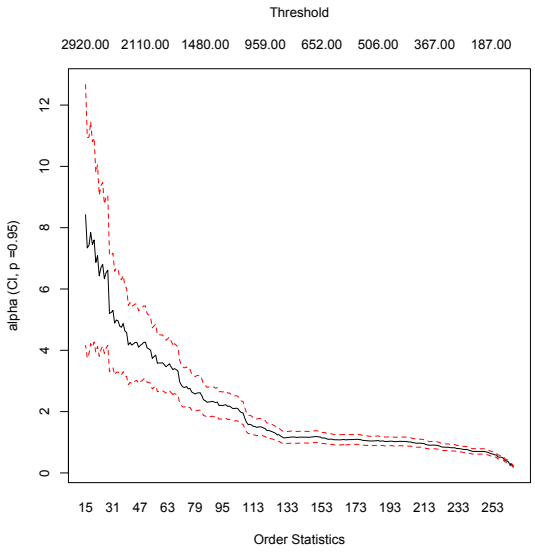


(c) XRP Demeaned VAR(1) Residuals QQ Plot

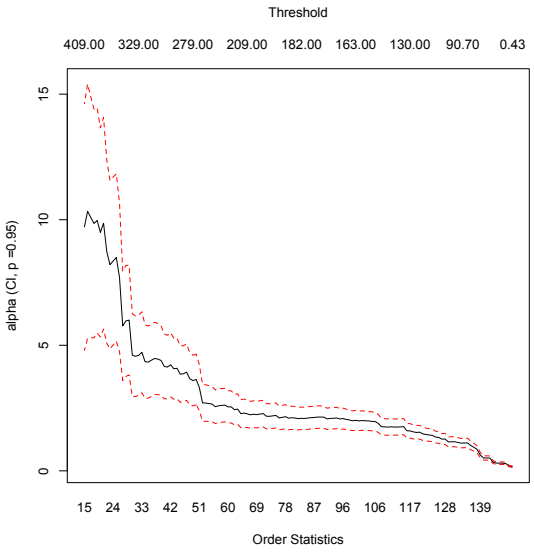


(d) XLM Demeaned VAR(1) Residuals QQ Plot

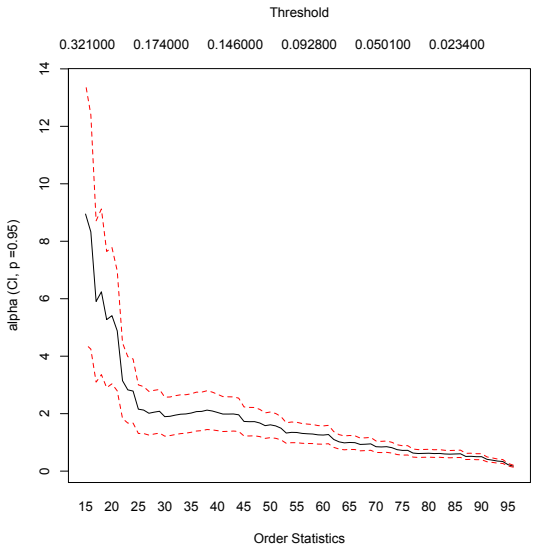
Figure B.7: BTC/ETH/XRP/XLM QQ Plots for Demeaned VAR(1) (n = 4)



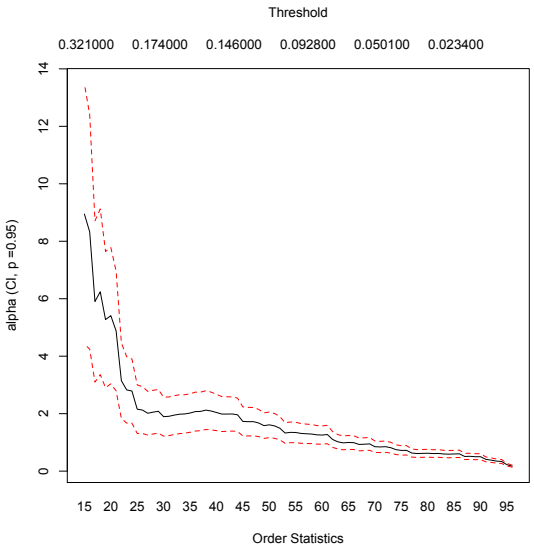
(a) BTC Hill Plot



(b) ETH Hill Plot



(c) XRP Hill Plot



(d) XRP Hill Plot

Figure B.8: BTC ETH XRP XLM Hill Plots of Demeaned Series

C Finding A^1 and A^2 when the number of non-causal directions is known

A simplification arises when the causal/noncausal dimensions $n_1, n_2 = n - n_1$ are known. Then the causal/noncausal directions can be estimated by minimizing the following objective function of the constrained optimization:

$$(\hat{A}^1, \hat{A}^2) = \arg \min_{A^1, A^2}{}^{-H}_{h=0} ||A^1 \hat{\Gamma}(h)(A^2)'||^2, \tag{C.1}$$

$$\text{s.t.} \quad A^1 \hat{\Gamma}(0)(A^1)' = Id_{n_1}, A^2 \hat{\Gamma}(0)(A^2)' = Id_{n-n_1},$$

where $||C||^2 = Tr(CC')$, $\hat{\Gamma}(h)$ is the empirical counterpart of $\Gamma(h) = Cov(Y_t, y_{t-h})$, and $H, H > 0$ is sufficiently large. This standardization of matrix A^{-1} is different from the one where the rows of A^j form orthonormal vectors as in $A^j(A^j)' = Id, j = 1, 2$, available from standard Jordan decomposition software.

The global optimization (C.1) is similar to optimizations in canonical correlation analysis, providing the least correlated linear transformations A^1, A^2, at any (negative) lag. It is possible to replace the constrained optimization (C.1) by two subsequent ones, in order to reduce the dimensions of the arguments in the objective function Gourieroux and Monfort (2015). For example, when $n_1 = n_2 = 2$, consistent estimators of the causal and noncausal directions can be derived in two optimizations, first with respect to the first rows of A^1 and A^2, then with respect to their second rows as follows :

First optimization :

$$(\hat{a}_1^1, \hat{a}_1^2) = \arg \min_{a_1^1, a_1^2}^{-H} {}_{h=0}(a_1^1 \hat{\Gamma}(h) a_1^{2'})^2,$$

$$\text{s.t. } a_1^1 \hat{\Gamma}(0)(a_1^1)' = 1, a_1^2 \hat{\Gamma}(0)(a_1^2)' = 1,$$

where a_1^1 and a_1^2 have dimensions $(1, n)$.

Second optimization :

The next optimization is with respect to the second rows of A^1, A^2, denoted by a_2^1, a_2^2, given the first rows obtained in the first stage. These estimators are :

$$(\hat{a}_2^1, \hat{a}_2^2) = \arg \min_{a_2^1, a_2^2}^{-H} {}_{h=0} || \begin{pmatrix} \hat{a}_1^1 \\ a_2^1 \end{pmatrix} \hat{\Gamma}(h) \begin{pmatrix} \hat{a}_1^2 \\ a_2^2 \end{pmatrix}' ||^2 \tag{C.2}$$

,

$$\text{s.t. } \hat{a}_2^1 \hat{\Gamma}(0)(\hat{a}_2^1)' = 1, \hat{a}_2^2 \hat{\Gamma}(0)(\hat{a}_2^2)' = 1,$$

$$\hat{a}_2^1 \hat{\Gamma}(0)(\hat{a}_1^1)' = 0, \hat{a}_2^2 \hat{\Gamma}(0)(\hat{a}_1^2)' = 0.$$

The constrained optimization (C.1) may be preferred if the dimension of the arguments is not too high. More specifically, optimization (C.1) involves n^2 arguments, which are the elements of matrix A^{-1}, subject to $\frac{n_1(n_1+1)}{2} + \frac{(n-n_1)(n-n_1+1)}{2}$ constraints. Then, the number of functionally independent arguments is equal to :

$$2n_1(n-n_1)+\frac{n_1(n_1-1)}{2}+\frac{(n-n_1)(n-n_1-1)}{2}.$$

Table C.1 shows the number of independent arguments.

Size n	Causal dim n_1 =1	Causal dim n_1=2
2	2	/
3	5	/
4	9	10
5	14	16

Table C.1: Number of Independent Arguments

For each lag h, we have $n_1(n-n_1)$ different elements to be minimized inside the norm $||\ ||^2$, and there are $(H+1)n_1(n-n_1)$ such elements. Thus there exists the smallest value of H to select for a given size n and causal dimension n_1 ensuring that the above optimization has a unique solution. The order condition, that is this minimum lag H, is given in Table C.2.

Size n	Causal dim n_1=1	Causal dim n_1= 2
2	1	/
3	2	/
4	2	2
5	3	2

Table C.2: Minimum Value of H

The empirical autocovariances $\hat{\Gamma}(h), h=0,\ldots,-H$ converge a.s. to their theoretical counterpart $\Gamma(h)$, when the number of observations T tends to infinity. Thus

the solutions $\hat{A}^1, \hat{A}^2$ of the constrained optimization (C.1) converge a.s. to the solutions of the associated limiting minimization where the empirical $\hat{\Gamma}(h)$ are replaced by their theoretical counterparts $\Gamma(h)$. If n_1 is equal to the true dimension $n_{1,0}$, we know that the limiting objective function is minimized for the true causal and noncausal directions. Thus $\hat{A}^j$ is a consistent estimator of $A_0^j, j = 1, 2$. If n_1 is equal to $n - n_{1,0}$, the minimized limiting objective function is also equal to zero. If n_1 is different from $n_{1,0}$ and $n - n_{1,0}$, the minimized asymptotic objective function is generically strictly positive.