Alberto Chicafo Mulenga

Master of Science

Volatility models and the multivariate links with applications to the economy of Mozambique

Dissertação para obtenção do Grau de Doutor em
Estatística e Gestão de Rísco

Orientadora: Marta Cristina Vieira Faias Mateus,
Professora Auxiliar, NOVA University of Lisbon

Co-orientadores: Pedro José dos Santos Palhinha Mota,
Professor Auxiliar, NOVA University of Lisbon

Joaquim Amaro Graça Pires Faia e Pina Catalão
Lopes, Professor Auxiliar, NOVA University of Lisbon

Júri

Presidente: Prof. Doutor Carlos Manuel Agra Coelho
Arguentes: Profª Doutora Raquel Maria Medeiros Gaspar
Prof. Doutor Gonçalo Manuel Albuquerque Perreira Oliveira de Faria
Vogais: Prof. Doutor Carlos Manuel Agra Coelho
Profª Doutora Marta Cristina Vieira Faias Mateus
Prof. Doutor João Beleza Teixeira Seixas e Sousa

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE **NOVA** DE LISBOA

September, 2019

ProQuest Number: 27841710

ProQuest.

ProQuest 27841710

Published by ProQuest LLC (2020). Copyright of the Dissertation is held by the Author.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

www.manaraa.com

Volatility models and the multivariate links with applications to the economy of Mozambique

Acknowledgements

I am grateful for the opportunity to be a student at this New University of Lisbon. First of all, I would like to express my thanks to my team of supervisors Prof. Dr Marta Faias, Prof. Dr Joaquim P. Pina and Prof. Dr Pedro Mota for having supported me through these years of study. To Professor Doctor Manuel Esquivel for his openness and support in the accompaniment of my teaching and learning process. I thanks to Professor Doctor Bhangy Cassy, who has always advised and encouraged me to proceed with my studies. A thank also to the whole community of the New University of Lisbon, in particular, the Center of Mathematics and Applications (CMA) and the study colleagues Abdul, Ivanilda and Sheyla. Finally thanks to my daughter Fátima Mulenga, for the days when she needed me, I could not be near her.

Lisbon, September 2019

To Fátima.

Abstract

Mozambique, like any other country, needs to know the behavior of the main macroeconomic variables to make better decisions. This study aims to investigate the behavior as well as the relationship between macroeconomic variables and financial time series related to Mozambique. For the study, we use univariate Autoregressive conditional heteroskedasticity models, Vector autoregressive models and multivariate Generalized autoregressive conditional heteroskedasticity models. Overall, the study concludes that: (i) Asymmetry of shocks, volatility and currency-specific behavior affect economic performance, particularly in an open economy, influencing international capital movement and goods and services transactions, (ii) Mozambique's real Gross domestic product plays an important role from cointegration relationships, impulse response functions to forecast error variance decomposition. (iii) The analysis of co-volatility showed the existence of relationships in volatility between different markets which influences the systematic behavior presented by the variables over time. These results contribute and reinforce the existing literature in terms of the choice of the appropriate model, criteria and tests presented that considerably affect the type of results. With this analysis of macroeconomic series related to Mozambique's economy, it somewhat helps the process of policy making that increases Mozambique's sustainable economic growth rate. In this thesis the problem of the normality of the log-returns for stock prices is also addressed for different formulations of price returns, namely intra-day and inter-day log-returns, with and without data trimming and for a large set of companies stock prices.

Keywords: Exchange rate, Goss domestic product, Mozambique, Multivariate time series, Volatility.

Resumo

Moçambique como qualquer outro país precisa de conhecer o comportamento das principais variáveis macroeconómicas para tomar decisões mais adequadas. Este estudo tem como objectivo investigar o comportamento e o relacionamento entre variáveis macroeconómicas e séries temporais financeiras relacionadas com Moçambique. Para a realização do estudo usou-se modelos univariados Autoregressivos com heteroscedasticidade condicional, modelos de Vectores autoregressivos e modelos multivariados Generalizados autoregressivos com heteroscedasticidade condicional. De uma forma geral, do estudo conclui-se que (i) a assimetria dos choques à volatilidade e comportamento específico da moeda afectam o desempenho económico, em particular numa economia aberta, influenciando o movimento internacional de capitais e transações de bens e serviços, (ii) o produto interno bruto real de Moçambique desempenha um papel importante desde as relações de cointegração, funções de resposta aos impulsos até à decomposição da variância do erro de previsão, (iii) a análise da co-volatilidade mostrou a existência de relações na volatilidade entre diferentes mercados o que influencia no comportamento sistemático apresentado pelas variáveis ao longo do tempo. Estes resultados trazem contribuições e reforçam a literatura existente em termos da escolha do modelo apropriado, critérios e testes apresentados que condicionam consideravelmente o tipo de resultados. Com esta análise de séries macroeconómicas relacionadas à economia de Moçambique, de certa forma ajuda o processo de definição de políticas que aumentam a taxa de crescimento económico sustentável de Moçambique. Nesta tese o problema da normalidade dos logarítmos dos retornos para cotações de acções também é abordado, nomeadamente para retornos intradiários ou entre dias consecutivos, com e sem eliminação de dados extremos e com aplicação a um grande número de empresas.

Palavras-chave: Moçambique, Produto interno bruto, Séries temporais multivariadas, Taxa de câmbio, Volatilidade.

Contents

List of Figures

List of Tables

Acronyms

ACF	Autocorrelation Function.
AD	Anderson and Darling.
ADF	Augmented Dickey and Fuller.
AGP	Acordo Geral de Paz.
AIC	Akaike Information Criteria.
APARCH	Asymmetric Power ARCH.
APGARCH	Asymmetric Power GARCH.
AR	Autoregressive.
ARCH	Autoregressive Conditional Heteroskedasticity.
ARIMA	Autoregressive Integrated Moving Average.
ARMA	Autoregressive Moving Average.
BEKK	Baba, Engle, Kraft and Kroner.
BFGS	Broyden Fletcher Goldfarb Shanno.
BHHH	Berndt, Hall, Hall and Hausman.
BIC	Bayesian Information Criteria.
CC	Conditional Correlation.
CCC	Constant Conditional Correlations.
CE	Cointegrating Equation.
CPI	Consumer Price index.
CZA	Coal of South Africa.
DCC	Dynamic Conditional Correlations.
DF	Dickey and Fuller.
DFP	Davidon Fletcher Powell.
DS	Demand Shocks.

DVECH	Diagonal Vector-Half operator.
EGARCH	Exponencial GARCH.
ES	External Shocks.
FEVD	Forecast Error Variance Decomposition.
FPE	Final Prediction Error.
GARCH	Generalized ARCH.
GARCH-M	GARCH-in-mean.
GBP	Great Britain Pound.
GDP	Gross Domestic Product.
GED	Generalized Error Distribution.
GJR-GARCH	Glosten Jagannathan and Runkle - GARCH.
GO-GARCH	Generalized Orthogonal - GARCH.
GPA	General Peace Agreement.
HD	Historical Decomposition.
HQIC	Hannan and Quinn Information Criteria.
IGARCH	Integrated GARCH.
IID	Independent and Identically Distributed.
IRF	Impulse Response Function.
JB	Jarque and Bera.
KPSS	Kwiatkowski, Phillips, Schmidt and Shin.
KS	Kolmogorov and Smirnov.
LJB	Lomnicki, Jarque and Bera.
LM	Lagrange Multiplier.
LR	Likelihood Ratio.
MA	Moving Average.

MAE	Mean Absolute Error.
MAPE	Mean Absolute Percentage Error.
MGARCH	Multivariate GARCH.
MJB	Multivariate Jarque and Bera.
ML	Maximum Likelihood.
MLB	Multivariate Ljung and Box.
MLE	Maximum Likelihood Estimation.
MLM	Multivariate Lagrange Multiplier.
MSE	Mean Squared Error.
MVN	Multivariate Normal.
MZGDP	Mozambique Gross Domestic Product.
MZN	Mozambique New Metical.
NARCH	Nonlinear ARCH.
NIC	News Impact Curve.
NSB	Negative Size Bias.
OECD	Organization for Economic Cooperation and Development.
OLS	Ordinary Least Squares.
PACF	Partial Autocorrelation Function.
PCS	Pearson's Chi-Square.
PDF	Probability Density Function.
PMZGDP	Prices of Mozambique Gross Domestic Product.
PP	Phillips and Perron.
PSB	Positive Size Bias.
QML	Quase Maximum Likelihood.
QMLE	Quase Maximum Likelihood Estimator.
RMSE	Root Mean Squared Error.
RSS	Residual Sum of Squared.
SB	Sign Bias.

SBIC	Schwartz and Bayesian Information Criteria.
SC	Stock Close.
SMA	Structural Moving Average.
SO	Stock Open.
SS	Supply Shocks.
SSR	Sum of Square of Residuals.
SUM	Seemingly Unrelated Regression.
SVAR	Structural Vector Autoregressive.
SVEC	Structural Vector Error Correction.
SW	Shapiro and Wilks.
TD	Trend Deterministic.
TGARCH	Threshold GARCH.
TS-GARCH	Taylor Schwert - GARCH.
VAR	Vector Autoregressive.
VCC	Varying Conditional Correlations.
VEC	Vector Error Correction.
VECH	Vector Half operator.
VECM	Vector Error Correction Model.
VMA	Vector Moving Average.
ZA	Zuid Afrikaanse = South Africa.
ZAGDP	South Africa Gross Domestic Product.
ZAR	Zuid Afrikaanse Republiek.
ZARCH	Zakoian ARCH.
ZGARCH	Zakoian GARCH.

Introduction

The main objective of the modern econometric analysis of time series is to develop models capable of forecasting, interpreting and testing hypotheses concerning economic data. Following Kirchgässner and Wolters [69], time series analysis helps to identify the behavior presented by the data series, which allows the formulation of laws or to explain the observed phenomenon in order to control its behavior and to predict future developments.

There are several models in terms of specification, estimation, and inference, used in such analysis. For example when the analysis is referent to a univariate time series as well as its volatility, the class of autoregressive integrated moving average (ARIMA) and autoregressive conditional heteroskedasticity (ARCH) or Generalized ARCH (GARCH) models are used (Box and Jenkins [22], Bollerslev *et al.* [21] and Brooks [27]). While when the study is carried out for a set of time series, that is multivariate time series with the purpose of knowing the behavior of each series, as well as the interaction of the series and the relationship of their volatilities, the vector autoregressive (VAR) and multivariate GARCH (MGARCH) models, are generally used. The procedures for the study of multivariate data analysis of time series can be found in Bauwens *el al.* [6], Engle and Kroner [43], Faust [49], Lütkepohl [77], among others.

In this context, this study approaches both univariate and multivariate time series models, giving greater focus on the interaction between the series over time as well as the analysis of their volatility. The study is motivated by the need to understand and disclose information related to the behavior of macroeconomic series such as Gross Domestic Product (GDP), inflation, exchange rate and their effects

on short and long terms as well as the results of the dynamics between them in the Mozambican economy without forgetting the interaction between Mozambique and other countries in the world in particular those on the region. Therefore, the basic reference in the subsequent applications is Mozambique.

1.1 The problem

Mozambique achieved independence from Portugal in 1975, however, from 1977 to 1992 was involved in a civil war, as documented in Organization for Economic Cooperation and Development (OECD) [93]. After the signing of the General Peace Agreement (GPA), "in Portuguese, Acordo Geral de Paz (AGP)" of 1992, the country was among the poorest in the world, with low social and economic indicators. Nevertheless, a new era began one of the market economies with free movement of people, goods, and services. Since then, a number of new companies looking to invest appear and on the other hand, old companies try to expand their activities according to their segment of the business. Allied to this, the country is in a situation in which is important to evaluate what are the most appropriate choices in terms of public or private investment, to boost the economic development of the country in the best and more comprehensive way as possible.

It is well known from the economic literature that changes in a macroeconomic variables of a given country are linked to variations in other variables and may also be related to the influence of neighbouring markets. Taking into account the interactions between the different variables, these changes experience higher or lower levels. The fluctuations affect somehow the government policies in decision making. In addition, the lack of accurate information leaves investors unsecured with respect to investment decisions.

Decision making is often supported by the use of quantitative information. Therefore, the main task of econometrics is to summarize information related to a data-set through a model that can help to understand or describe the relationship between variables and to analyse the possible effects of decisions. In this context, we propose to study the behavior of macroeconomic variables as well as their interactions to identify how they influence each other. This understanding of the relationship between macroeconomic variables helps the government, private companies, nongovernmental organizations to make the best investment decisions or not, which develops the economy of the country as a whole.

Thus, the problem in studies can be decomposed in the following points:

1) Even knowing that Mozambique has little systematized information, studies

2

describing the behavior of macroeconomic variables is necessary. For this purpose we use univariate models such as ARIMA or ARCH models, which allows knowing the behavior of the series and make a consequent forecast.

2) Another type of study that somehow helps the decision making process of government policy is multivariate data analysis. This problem is illustrated by VAR models, where several variables are analysed simultaneously and their relationship are studied. Here the decisions are based on information from a set of variables.

3) In the economic and financial context, it is necessary to increase the analysis of the dynamics and volatility between markets, in order to make it possible to understand the links in the short and long term or the influence of changes in policies that regulate the functioning of one market in relation to the others. For this problem, we use MGARCH models, which in the analysis take into account the linkage of the information in a certain market as part of a whole set of markets.

4) The assumption of normality in the statistical analyses is always evoked, in particular, the normality of the log-returns of stock prices is often assumed by the market players in order to validate and use some results. However, several studies regarding different indexes have shown that the normality assumption of the returns usually fails. Rejection of the normality of the series may be due to the type of data, the method used for the test or even the procedure that is chosen. Depending on the test result, the decision may be the most appropriate or not.

Thus, by providing accurate and up-to-date information on the trend of series, relationships between variables and the dynamics between different economic markets, the government as the potential use of information can make strategic decisions that are more sustainable for the country's economic development.

1.2 The objectives

To solve the presented problems four main objectives are formulated:

1. Investigate the behavior of Mozambican New Metical against South African Rand Exchange rate between 2010 to 2014.

2. Analyse the dynamic of the relationship between three macroeconomic variables related to Mozambique economy such as Mozambique real Gross domestic product, Mozambique GDP deflator and the South Africa real GDP, over 1990 to 2012.

3. Analyse the relationship between the Mozambican New Metical against South African Rand and also against Great Britain Pound, plus the quote of the price

3

of Coal of Africa, between 2010-2014.

4. Test the normality assumption for different formulations of log-returns of stock prices in opening and closing using five procedures.

In order to meet the first objective, we inspect how risk is affected by news sign and size within distinct exchange rate trend and by volatility model choice. For that we apply the univariate asymmetric GARCH models to modelling the Mozambican New Metical against South African Rand exchange rate in four sub-periods, where the exchange rate exhibits different behavior and a full sample; identifying the possible sign and size bias in the sub-periods as well as for the full sample and discuss the asymmetric effect of the shocks in volatility over time. This objective is discussed in application of chapter 2.

For the second objective, we use vector error correction (VEC) and Structural vector error correction (SVEC) models to analyse the role of supply, demand and external shocks in an open economy. The study identifies the relationship between aggregate supply and aggregate demand shocks in the presence of external shocks to the economic growth of Mozambique. For this purpose, it is presented the application for VAR models at the end of chapter 3.

For the third objective, we investigate the co-movement of volatilities among the series of Mozambican New Metical against South African Rand exchange rate, Mozambican New Metical against Great Britain Pound exchange rate and the quote of price of Coal for Africa; we assess the quality of the models regarding prediction and discuss the differences in the estimated co-volatility. This is operationalized using Multivariate GARCH models, where we estimate MGARCH models in the class of conditional correlation family for four sub-periods identified between 2010-2014. In the application of chapter 4 we discuss this objective.

Finally, for the fourth objective, we analyse the normality assumption for intraday and inter-day log-returns, where we compare opening and closing prices for a large number of companies quoted in the Nasdaq Composite Index. To perform the normality test we implement five alternative procedures for normality tests: the Pearson's Chi-Square, Kolmogorov-Smirnov, Anderson - Darling's, Shapiro - Wilks's and Jarque-Bera test. Details of the procedures are described in chapter 5.

According to the objectives, the thesis consists essentially of four chapters whose applications are articles published in scientific journals or submitted for publication.

4

1.3 The thesis structure

After the introduction in chapter 1, where we present the problem and the main objectives, the contents of this thesis are structured as follows:

- In chapter 2, we focus on the analysis and application of univariate financial time series. That is, we study the autoregressive conditional heteroskedasticity models, wherein the section 2.1 we present univariate time series concepts and associate tests for normality, stationarity, autocorrelation, unit roots and nonstationary time series. The general formulation and properties of ARCH and GARCH models are presented in section 2.2. In the section 2.3, we present alternative models to modelling volatility in time series, which in the literature are known as extensions of GARCH model proposed to capture certain characteristics presented by the financial time series, such as volatility cluster, leverage effects, asymmetric effect. The univariate procedure of maximum likelihood estimation, forecast and model evaluation are discussed in the section 2.4. The chapter ends with an application in section 2.5, where we apply the univariate asymmetric GARCH model such as Exponential GARCH of Nelson [90], Threshold GARCH of Zakoian [121], GJR-GARCH proposed by Glosten *et al.* [52] and Asymmetric Power ARCH model of Ding *et al.* [34], to analyse the behavior of the Mozambique New Metical against South Africa Rand daily exchange rate, thus, the chapter responds to the first objective of the thesis.

- The chapter 3 is dedicated at the multivariate time series analysis using Vector Autoregressive (VAR) model. In the first section 3.1 of the chapter 3, we introduce conditions for the multivariate stationary process, while the specifications of the VAR model in reduced and structural form are presented in section 3.2. In section 3.3, we present the estimation VAR parameters and they analyse, from model choice, evaluation to analyse by Granger-causality, forecast, impulse response functions and forecast error variance decomposition. The concepts of cointegration and vector error correction model as well as the procedures of Engle and Granger [42] and Johansen [63] for testing cointegration are presented in section 3.4. In section 3.5, we present the structural vector autoregressive and structural vector error correction analysis, where we focus on identifying the sources of shocks in VEC model, impulse response function and forecast error variance decomposition. The final section 3.6 presents the application of vector autoregressive models. Specially we study the dynamics of the Mozambique real Gross Domestic Product uncovering the role of three

aggregate shocks supply, demand and external shocks, taking the South Africa economy as key for measuring the latter shock. Here we use the annual series between 1990 to 2012, to identify the sources of shocks in structural vector error correction system.

- In chapter 4, we study the relations between volatilities and co-volatilities using multivariate GARCH models. After introducing the chapter in section 4.1, the section 4.2 is dedicated to reviewing of MGARCH models in the class of conditional covariance matrix models, which includes VECH and Diagonal VECH models introduced by Bollerslev *et al.* [20], the BEKK model introduced by Engle and Kroner [43]. In section 4.3, we review the category of Conditional Correlations models, which includes the Constant Conditional Correlations GARCH model of Bollerslev [17], the Dynamic Conditional Correlations GARCH model of Engle [40] and the Varying Conditional Correlations GARCH model of Tse and Tsui [117]. In the section 4.4, we present the estimation procedure of MGARCH models. Finally in section 4.5, we present the application of multivariate GARCH models, where we use the Conditional Correlations MGARCH models to analyse the volatility and co-volatilities between the three variables just mentioned.

- In chapter 5, we analyse the assumption of normality of log-returns. Where we introduce its use as well as cases where the data does not follow the normality and its implications. After the introduction in section 5.1, the section 5.2 presents the procedures of five normality tests. Note that, since the tests of normality were carried out in previous chapters, in this chapter will be presented only some variants focused on the returns. Finally, in order to illustrate how procedures work, in section 5.3, we present the application.

- In chapter 6, we present the main results obtained and discussed in the previous chapters as conclusions of the research. We also consider some final remarks related to the perspectives of this kind of investigations.

2

Univariate GARCH modelling

2.1 Univariate time series concepts

Following Kirchgässner and Wolters [69], a time series is defined as a set of quantitative observations collected in chronological order. Often takes over that the observations are collected in equal intervals of time and the time is a discrete variable. The frequency of reference observations in this chapter is daily. The basic objective of the analysis of time series is usually to determine a model that describes the pattern of the time series. This goal includes a description of the actual data, explanation or modelling the data, forecasting or predict the future patterns and control in case of deviations of the regular pattern.

In general, to model a specific time series, it is assumed that the residual component does not influence the other components, that is, the trend, business cycle and seasonal cycle components are not influenced by irregular component, therefore, in the classical time series analysis these components can be represented by deterministic functions of the time series.

Because our focus in this chapter is modelling and discuss properties of GARCH models, we present concepts related to univariate time series.

2.1.1 Stationary and ergodic time series

Let $\{y_t\} = \{...y_{t-1}, y_t, y_{t+1}, ...\}$, where the observations are associated with a time t, thus a random variable called a time series.

According to Kirchgässner and Wolters [69], the concept of ergodicity means

7

that the moments obtained from a sample of observations, when the sample size increases considerably, they converge to the respective population moments. For example, if we assume that, the population moments expectation $E(y_t) = \mu$ and variance $var(y_t) = \sigma^2$ are constant for every t. The y_t process is called ergodic for the mean if the time series average converges to the population average.

$$\lim_{T \to \infty} E\left[\left(\frac{1}{T}\sum_{t=1}^{T} y_t - \mu\right)^2\right] = 0 \tag{2.1}$$

Similarly for variance ergodicity, if the sample variance provides a consistent estimate for the second moment, then the process is said to be ergodic for the second moment.

$$\lim_{T \to \infty} E\left[\left(\frac{1}{T}\sum_{t=1}^{T} (y_t - \mu)^2 - \sigma^2\right)^2\right] = 0 \tag{2.2}$$

More simply, a stochastic process is ergodic if is stationary. There are two kinds of stationarity: the first is associated with the common distribution function that does not change over time, which is said to be *strictly stationarity*. The second form is that in which the process has periods of stationary, therefore, changes over time, so we only consider *weak stationarity* or *stationarity in the second moments*. Consequently, we define stationary for the corresponding moments of the stochastic process $\{y_t\}$:

1. *Mean stationary*: if $E(y_t) = \mu$ is constant for all t.

2. *Variance stationary*: If $var(y_t) = E[(y_t - \mu)^2] = \sigma^2$ is constant and finite for all t.

3. *Covariance stationary*: if $cov(y_t, y_{t-j}) = E[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j$ for all t and any j is only a function of the time distance between the two time periods and does not depend on the actual point in time t.

4. *Weak stationarity*: when it is mean and covariance stationary.

Because in many situations only assume weak stationarity, we most drop the word weak, and we say stationary time series instead of covariance stationary time series. Stationary time series have time invariant first and second moments. According to Tsay[116], when the linear dependence between the values of the time series y_t and its past values y_{t-i} is of interest, the concept of correlation is generalized to autocorrelation. The parameter γ_j is called lag-j autocovariance of y_t and a plot of

ρ_j against j is called the autocorrelation function (ACF). For y_t the ACF is defined by

$$\rho_j = \frac{cov(y_t, y_{t-j})}{\sqrt{var(y_t) \times var(y_{t-j})}} = \frac{\gamma_j}{\gamma_0} \tag{2.3}$$

Summarizing, a stationary time series is defined by its mean, variance and ACF. Therefore, a stationary time series y_t is ergodic if sample moments converge in probability to population moments, that is, $\overline{y} \xrightarrow{p} \mu$, $\widehat{\gamma}_j \xrightarrow{p} \gamma_j$ and $\widehat{\rho}_j \xrightarrow{p} \rho_j$.

2.1.2 Testing for autocorrelation

Following Baltagi [5] and Tsay [116], a time series y_t is said a white noise if $\{y_t\}$ is a sequence of independent and identically distributed (I.I.D) random variables with finite mean, variance and zero autocorrelation, then $\rho_j = 0$ for $j > 0$. In particular, if y_t is normally distributed with mean zero and variance σ^2 is said that the time series is Gaussian white noise. For a white noise time series, all the ACF are zero, except $\rho_0 = 1$. In practice, if all sample ACF is close to zero, then the series is white noise.

According to Zivot and Wang [123], in order to test autocorrelation we test if the series y_t is white noise. To test the null hypothesis of several autocorrelation are zero, Box and Pierce [23], propose the Portmanteau statistic $Q^*(m) = T \sum_{j=1}^{m} \widehat{\rho}_j^2$ for the null hypothesis $H_0 : \forall \rho_j = 0$ against the alternative that $H_1 : \exists \rho_j \neq 0$ for some $j \in \{1, ..., m\}$. Under the null hypothesis $Q^*(m)$ is asymptotically χ^2 distributed with m degrees of freedom.

Ljung and Box [75] proposed the modified Q-statistic of Box and Pierce [23] to increase the power of the test in finite samples.

$$Q(m) = T(T+2) \sum_{j=1}^{m} \frac{\widehat{\rho}_j^2}{T-j} \tag{2.4}$$

where T is the sample size.

We reject the null if $Q(m) > \chi_\alpha^2(m)$. Because most software statistic provide the p-value of $Q(m)$, the decision rule is then reject the null hypothesis if p-value is less than or equal to the significance level α, that is, $RH_0 : p\text{-value} \leq \alpha$.

2.1.3 Testing for normality

According to Tsay [116], Zivot and Wang [123], among various graphic procedures used to check if I.I.D process follows Normal distribution, we can mention the Normal quantile-quantile plot, a scatter-plot of the standardized empirical quantiles of y_t against the quantiles of a standard random variable. If y_t is normally distributed,

then the quantiles will lie on a 45-degree line. In addition to the graphic representation, the Jarque and Bera [62] test is the most common formal statistic test for normality used. The test is based on Skewness and Kurtosis of a random variable that follows the Normal distribution. That is, with $Skew = 0$, $Kurt = 3$. The statistic is computed by:

$$JB = T \times \left(\frac{\widehat{skew}^2}{6} + \frac{(\widehat{kurt} - 3)^2}{24} \right) \tag{2.5}$$

where the sample Skewness and Kurtosis are given by:

$$\widehat{skew} = \frac{1}{(T-1)\sigma^3} \sum_{t=1}^{T} (y_t - \mu)^3 \quad \text{and} \quad \widehat{kurt} = \frac{1}{(T-1)\widehat{\sigma}^4} \sum_{t=1}^{T} (y_t - \widehat{\mu})^4$$

Under the null hypothesis that the series is normally distributed, the statistic can be compared with a Chi-square distribution with 2 degrees freedom. The null hypothesis of normality is rejected if the JB statistic exceeds a critical value from $\chi_\alpha^2(2)$ or p-value is less than the significance level chosen. This and other tests used to diagnose normality of data are described in Mulenga $et\ al.$ [88].

2.1.4 Unit root nonstationary time series

In subsection 2.1.1, we focused on stationary time series. For some series, such as interest rates, exchange rates, prices of an asset, the series tend to be nonstationary. According to Zivot and Wang [123], since the stationary process has time invariant moments, a nonstationary process must have some time dependent moments. The most common forms of nonstationarity are caused by time dependent in mean and variance.

Following Tsay [116], the nonstationary series is called unit root nonstationary time series. The random-walk model, trend stationary and integrate processes are examples of unit root nonstationary time series.

$Random\ walk.$ If y_t can be represent as $y_t = y_{t-1} + u_t$, where $u_t \sim \text{I.I.D}(0, \sigma^2)$, then the series y_t is a random walk process. If the process started at $t=0$, the initial value is $y_0 = \mu$, then $y_1 = \mu + u_1$, $y_2 = \mu + u_1 + u_2$, ..., $y_t = \mu + \sum_{j=1}^{t} u_j$ with $E(y_t) = \mu$ and $var(y_t) = t\sigma^2$ since $u_t \sim \text{I.I.D}(0, \sigma^2)$. Therefore, the variance of y_t is dependent on t. This is a nonstationary time series, however, after first difference of y_t we get u_t which is stationary, see Baltagi [5].

$Trend\ stationary.$ According to Zivot and Wang [123], the series y_t is a trend stationary if it has the form $y_t = TD + u_t$, where TD are deterministic trend terms (constant, trend, seasonal dummies, etc.), that depend on t and u_t is stationary.

Since u_t is stationary y_t never deviates too far away from TD. Consequently, y_t exhibits trend reversion. If TD were known, y_t may be transformed to a stationary process by subtracting the deterministic trend terms: $u_t = y_t - TD$.

Integrated process. The series y_t is an integrated process of order 1, denoted $y_t \sim I(1)$, if it has the form $y_t = y_{t-1} + u_t$ and the first difference of y_t is stationary, $\Delta y_t = u_t$. The $I(1)$ processes are sometimes called difference stationary processes. The order of integration is the minimum number of times the series needs to be differenced to yield a stationary series.

Summarizing, a stationary time series is said integrated of order zero $I(0)$. For the process y_t, if after second differences we get a stationary time series $\Delta^2 y_t = (1-L)^2 y_t$, such process is integrated of order 2. More generally, the process is $I(d)$ if it can be made stable by differencing d times, that is $\Delta^d y_t = (1-L)^d y_t$. There are two principal methods of detecting nonstationarity: (i) visual inspection of the time series graph and its correlogram, (ii) formal statistics tests of unit roots.

There are large number of tests for unit root. For illustration, let us consider a time series y_t in the form: $y_t = \mu + \beta_1 y_{t-1} + u_t$ where $u_t = \rho u_{t-1} + \varepsilon$. Unit roots tests are based on testing the null hypothesis that $H_0 : \rho = 1$ against the alternative $H_1 : \rho < 1$. They are called unit roots because under the null hypothesis the characteristic polynomial has a root equal to unity.

Dickey and Fuller (DF) test. One commonly used test for unit roots is the DF test. In its simplest form we can consider a Autoregressive (AR(1)) process $y_t = \rho y_{t-1} + u_t$ or $\Delta y_t = \delta y_{t-1} + u_t$ where $\delta = \rho - 1$ and u_t is I.I.D sequence of random variables. Under the null hypothesis y_t is nonstationary, consequently the standard t - statistics does not follow t - distribution, not even asymptotically. To test the null hypothesis it is used the statistic $DF_t = \dfrac{\widehat{\rho}}{se(\widehat{\rho})}$. We do not reject the null of unit root at level α if the observed DF_t statistic, is greater than the critical value for *DF test*. For details of the test, see Dickey and Fuller [31] and Enders [38].

According to Zivot and Wang [123], the basic DF test is valid if y_t is an AR(1) model with white noise errors. However, many time series have a more complicated dynamic structure that is not captured by a simple AR(1) model. Said and Dickey [104], augmented the basic AR(1) unit root test to accommodate general Autoregressive Moving Average (ARMA) models, and the test is referred as the Augmented Dickey and Fuller (ADF) test. The ADF test, tests the null hypothesis that a time series $y_t \sim I(1)$ against the alternative that it is $I(0)$, assuming the dynamic in the data have an ARMA structure.

The procedure for ADF test is the same as for the DF test, but here it is applied to the model $\Delta y_t = \mu + \delta t + \rho y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + u_t$, where μ is a constant, δ the

11

coefficient on a time trend, ρ is the coefficient that indicate the presence of unit root in the series, Δy_{t-i} are used to approximate de ARMA structure of the errors and p the lag order of the AR process. Imposing the constraints $\mu = 0$ and $\delta = 0$ correspond to modelling a random walk and using the constraint $\delta = 0$ corresponds to modelling a random walk with a drift. By including lags of the order p the ADF test formulation allows for higher-order autoregressive process. The ADF t-statistic and normalized bias statistic based on the Ordinary Least Squares (OLS) estimates are given by:

$$ADF_t = t_\rho = \frac{\widehat{\rho} - 1}{se(\widehat{\rho})} \quad or \quad ADF_\rho = \frac{T(\widehat{\rho} - 1)}{1 - \widehat{\beta}_1 - ... - \widehat{\beta}_p} \tag{2.6}$$

An important practical issue for the implementation of the ADF test is the specification of the lag length p. If p is too small then the remaining serial correlation in the errors will bias the test. If p is too large then the power of the test will suffer. Ng and Perron [91], suggest the procedure to set an upper bound p_{max} for p, estimate the ADF test regression with $p = p_{max}$. If the absolute value of the t-statistic is greater than 1.6 then set $p = p_{max}$ and perform the unit root test. Otherwise, reduce the lag length by one and repeat the process.

Phillips and Perron [99], developed a number of unit root tests that have become popular in the analysis of financial times series. The **Phillips-Perron (PP) unit root tests** differ from ADF tests mainly in how they deal with serial correlation and heteroskedasticity in the errors. Where the ADF tests use a parametric autoregressive to approximate the ARMA structure of the errors in the test regression, the PP tests ignore any serial correlation in the test regression. That is, the PP tests are not parametric. One advantage of the PP tests over ADF tests is that the PP tests are robust to general forms of heteroskedasticity in the error term u_t. Another advantage is that the user of bandwidth does not have to specify a lag length for the test regression.

The ADF and PP tests are asymptotically equivalent but may differ substantially in finite samples due to the different ways in which they correct for serial correlation in the test regression. In other hand the ADF and PP unit root tests are for the null hypothesis that a time series y_t is $I(1)$. Stationarity tests, are for the null hypothesis that y_t is stationary $I(0)$. The most commonly used stationarity test is KPSS test, proposed by Kwiatkowski *et al.*[71][1].

The KPSS test, in contrary to most unit root tests, the presence of a root unit is not the null hypothesis but the alternative. Additionally, in the KPSS test,

---

[1]KPSS due to authors Kwiatkowski, Phillips, Schmidt and Shin.

the absence of a root unit is not a proof of stationarity but, by design, of trend-stationarity. A major disadvantage for the KPSS test is that it has a high rate of Type I errors, that is, tends to reject the null hypothesis too often. If attempts are made to control these errors (by having larger p-values), then that negatively impacts the test's power. One way to deal with the potential for high Type I errors is to combine the KPSS test with an ADF test. If the result from both tests suggests that the time series in stationary, then it probably is.

2.2 Conditional heteroskedasticity models

In Autoregressive Moving Average models of Box and Jenkins [22], the analysis is concentrated in modelling and predicting the conditional mean equation, where residuals are assumed uncorrelated and homoskedastic, and is rarely concerned with the conditional variance of a time series.

According to Brooks [27], stylized facts of financial times series are: Serial dependence presented in the data; volatility clustering, that is, small or large shocks are followed by small or large shocks; heavy tailed when the distribution of returns shows more leptokurtosis effect than the Normal distribution; leverage effects where the volatility tends to be large for the price falls than for the price rises when the magnitude of the price rise and fall is identical; long memory in volatility; spillover effects which refers to positive or negative effects of those who are not directly involved in it.

According to Tsay [116], volatility has became a very important concept in financial analysis. The term, volatility indicate how much and how quickly the value of an market sector changes, thus, is measured by standard deviation. Most financial time series such has foreign exchange rate and stock prices as well as inflation rates may exhibit some volatility which varies over time, which suggests that their variance may be heteroskedasticity.

Analysis and modelling volatility in financial time series became the object of more interest since the introduction of the Autoregressive conditional heteroskedasticity (ARCH) model in the seminal paper of Engle [39]. Subsequently, were developed several univariate models from class ARCH and GARCH models. For example the Exponential GARCH of Nelson [90], the Threshold GARCH of Zakoian [121], the Asymmetric Power ARCH (APARCH) of Ding *et al.* [34], with the objective of studying the volatility and other characteristics of financial times series. Each model was developed in order to capture a particular characteristic such as volatility clustering, heavy tailed, leverage effects, asymmetry and high persistence in second

13

moments, etc. Details of the specifications and estimation procedure can be found in Bera and Higgins [7], Bollerslev *et al.* [21], Tsay [116], Zivot and Wang [123], among others.

2.2.1 The basic ARCH model

Let r_t denote a stationary times series such as financial returns, then r_t can be expressed as its mean μ, and an autoregressive part to accounting for time persistence plus a white noise u_t if there is no significant autocorrelation in r_t itself, which is called AR(1) process, $r_t = \mu + \phi_1 r_{t-1} + u_t$. To allow for conditional heteroskedasticity, assume that $var_{t-1}(u_t) = \sigma_t^2$, with $var_{t-1}(.)$ denoting the variance conditional on information at time $t-1$, and is modelled in the following way: $\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + ... + \alpha_p u_{t-p}^2$. Since $E(u_t) = 0$ and $var_{t-1}(u_t) = E_{t-1}(u_t^2) = \sigma_t^2$, the variance of residuals can be rewritten as $u_t^2 = \omega + \alpha_1 u_{t-1}^2 + ... + \alpha_p u_{t-p}^2$, which represents an AR(p) process for squared residuals u_t^2.

The joint AR(1) process for returns and AR(p) process for squared residuals, is known as the Autoregressive conditional heteroskedasticity model of Engle [39]. Following Zivot [122], the basic idea of ARCH models is that, the shock of u_t of an asset return is serially uncorrelated but dependent, and the dependence of u_t can be described by a simple quadratic function of its lagged values, which is usually referred to as the ARCH(p) model. An alternative and complete formulation of the ARCH model is:

$$
\begin{aligned}
r_t &= \mu + \phi_1 r_{t-1} + u_t \\
u_t &= z_t \sigma_t \quad where \quad z_t \sim N(0,1) \\
\sigma_t^2 &= \omega + \alpha_1 u_{t-1}^2 + ... + \alpha_p u_{t-p}^2
\end{aligned}
\tag{2.7}
$$

where z_t is a sequence of random variables with mean zero and variance one, $\omega > 0$ and $\alpha_i \geq 0$ for $i > 0$, to ensure that the unconditional variance of u_t exists. In practice the standardized residuals $z_t = u_t/\sigma_t$ is often assumed to follow the standard Normal distribution, standardized t-distribution or a Generalized error distribution.

Properties of ARCH(1) model

Without loss of generality, let a ARCH(1) model be represented by $u_t = z_t \sigma_t = z_t \sqrt{\omega + \alpha_1 u_{t-1}^2}$, where $\{z_t\}_{t=0}^{\infty}$ is a sequence of I.I.D random variables with mean zero and variance one, that is $z_t \sim N(0,1)$. ARCH(1) model has the following properties.

Property 1. The u_t has zero mean.

14

Proof. As σ_t is $\mathscr{F}_{t-1}$ mensurable and z_t is independent of $\mathscr{F}_{t-1}$ and still $E(z_t) = 0$, we will have:

$$
\begin{aligned}
E(u_t) &= E(E_{t-1}(u_t)) = E(E_{t-1}(z_t \sigma_t)) \\
&= E(\sigma_t E_{t-1}(z_t)) = E(\sigma_t E(z_t)) \\
&= E(\sigma_t)E(z_t) = 0, \quad since \quad E(z_t) = 0
\end{aligned}
$$

Property 2. The conditional variance is given by $\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2$.

Proof. As $E(u_t) = 0$, the conditional variance of u_t is:

$$
\begin{aligned}
var_{t-1}(u_t) &= E_{t-1}(u_t^2) = E_{t-1}[z_t^2(\omega + \alpha_1 u_{t-1}^2)] \\
&= E_{t-1}(z_t^2)(\omega + \alpha_1 u_{t-1}^2) \\
&= E(z_t^2)(\omega + \alpha_1 u_{t-1}^2) \\
&= \omega + \alpha_1 u_{t-1}^2 = \sigma_t^2, \quad since \quad E(z_t^2) = var(z_t) = 1
\end{aligned}
$$

Property 3. The unconditional variance is $\overline{\sigma}^2 = \omega/(1 - \alpha_1)$, which only exists if $\omega > 0$ and $|\alpha_1| < 1$.

Proof. As u_t is stationary then $E(u_t^2) = E(u_{t-1}^2) = \sigma^2$, also we have $E(u_t) = 0$, $\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2$ is $\mathscr{F}_{t-1}$ mensurable, z_t is independent of $\mathscr{F}_{t-1}$ and still $E(z_t^2) = var(z_t) = 1$, we have:

$$
\begin{aligned}
var(u_t) = E(u_t^2) &= E[E_{t-1}(u_t^2)] = E[E_{t-1}(z_t^2)(\omega + \alpha_1 u_{t-1}^2)] \\
&= E[E(z_t^2)(\omega + \alpha_1 u_{t-1}^2)] \\
&= E(\omega + \alpha_1 u_{t-1}^2) = \omega + \alpha_1 E(u_{t-1}^2)
\end{aligned}
$$

Therefore, we have $\overline{\sigma}^2 = \omega + \alpha_1 \overline{\sigma}^2$ and $\overline{\sigma}^2 = \omega/(1 - \alpha_1)$. Since the variance of u_t must be positive, we require $0 \leq \alpha_1 < 1$.

Property 4. The auto-covariances are zero. Note that, this properties seems plausible enough for ARCH(1) process, but less for higher order processes.

Proof.

$$
\begin{aligned}
E_{t-1}(u_t u_{t-1}) &= E[u_{t-1} E_{t-1}(u_t)] \\
&= E\left[u_{t-1}\sqrt{\omega + \alpha_1 u_{t-1}^2} E_{t-1}(z_t)\right] \\
&= E\left[u_{t-1}\sqrt{\omega + \alpha_1 u_{t-1}^2} E(z_t)\right] \\
&= E\left[u_{t-1}\sqrt{\omega + \alpha_1 u_{t-1}^2}\right]E(z_t) = 0, \quad since \quad E(z_t) = 0
\end{aligned}
$$

Property 5. Regarding kurtosis, Bera and Higgins [7] show that the process has a heavier tail than the Normal distribution, that is $\dfrac{E(u_t^4)}{var(u_t)^2} = 3\left(\dfrac{1 - \alpha_1^2}{1 - 3\alpha_1^2}\right) > 3$.

15

Proof. As $u_t|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$, then $E_{t-1}(u_t^4) = 3var_{t-1}^2(u_t)$, and we have:

$$
\begin{aligned}
E(u_t^4) &= E[E_{t-1}(u_t^4)] \\
&= E[3var_{t-1}^2(u_t)] \\
&= E[3(\sigma_t^2)^2] = 3E[(\omega + \alpha_1 u_{t-1}^2)^2] \\
&= 3[\omega^2 + 2\omega\alpha_1 E(u_{t-1}^2) + \alpha_1^2 E(u_{t-1}^4)]
\end{aligned}
$$

If u_t is fourth - order stationary with $m_4 = E(u_t^4)$ then we have:

$m_4 = 3(\omega^2 + 2\omega\alpha_1\sigma^2 + \alpha_1^2 m_4) = 3\omega^2\left(1 + 2\frac{\alpha_1}{1-\alpha_1}\right) + 3\alpha_1^2 m_4$, from where we get

$m_4 = \dfrac{3\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}.$

Following Tsay [116], this results has two important implications (a) since $m_4 > 0$, α_1 must satisfy the condition $1 - 3\alpha_1^2 > 0$; that is $0 \leq \alpha_1^2 \leq \frac{1}{3}$ and (b) The unconditional kurtosis of u_t is: $\frac{E(u_t^4)}{[var(u_t)]^2} = 3\frac{\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} \times \frac{(1-\alpha_1)^2}{\omega^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3$.

Thus, the excess kurtosis of u_t is positive and the tail distribution of u_t is heavier than that of a Normal distribution. That is, the shock u_t on ARCH(1) model are more enabled to produce outliers than a Gaussian white noise series.

According to Brooks [27] and Tsay [116], the ARCH model provides a framework for the analysis and development of times series model of volatility. However, ARCH models have some weaknesses:

1. The assumption assumed by the model that the negative and positive shocks have the same effects is not realistic, that is not well sustained because in practice it is known that the shocks have different effects, especially in financial time series.

2. For series with finite fourth moments, the value of α_1^2 of an ARCH(1) model must be between $[0; 1/3]$. That is ARCH(1) is more restrictive.

3. The ARCH models, does not indicate the source of variation of financial time series, only provides the procedures to describe the behavior of the conditional variance associated with those series.

4. ARCH models at some point overestimate the parameters as they respond slowly to large shocks, especially when they are isolated shocks to the return series.

2.2.2 The Generalized ARCH model

According to Bera and Higgins [7], Brooks [27] and Tsay [116], in many applications of the linear ARCH(p) model the number of lags required to capture all of the

dependence on the conditional variance, might be very large. To avoid this problem, Bollerslev [15] and Taylor [115] developed independently a useful extension of the conditional variance function in equation 2.7, which is known as the Generalized ARCH or GARCH(p,q) model, where p is the number of lags for ARCH and q the number of lags for GARCH.

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \quad ; \quad u_t = z_t \sigma_t \tag{2.8}$$

where $z_t \sim N(0,1)$ and the restrictions $\omega > 0$, $\alpha_i \geq 0$ and $\beta_j \geq 0$ are imposed to ensure that the conditional variance is strictly positive. The GARCH model allows the conditional variance to be dependent upon previous own lags. Note that, specification a parsimonious GARCH (p,q) is not easy, in most applications only lower order are used, say GARCH(1,1), GARCH(2,1) or GARCH(1,2). The volatility equation in the simple case, GARCH(1,1) is:

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad and \quad 0 \leq \alpha_1, \beta_1 \leq 1 \tag{2.9}$$

For GARCH(1,1), u_t is covariance stationary if and only if $\alpha_1 + \beta_1 < 1$. According to Brooks [27] in the equation 2.9, the conditional variance σ_t^2 is one period ahead estimate for the variance based on any past information thought relevant. That is, the fitted variance σ_t^2 as a weighted function of a long term average value dependent on ω, information about volatility during the previous period ($\alpha_1 u_{t-1}^2$) and the fitted variance of the model in the previous period ($\beta_1 \sigma_{t-1}^2$).

Properties of GARCH model

The GARCH(p,q) model in equation 2.8 can be written in the following form: $\sigma_t^2 = \omega + \alpha(L)u_{t-1}^2 + \beta(L)\sigma_{t-1}^2$, where $\alpha(L)$ and $\beta(L)$ are polynomials with no common roots and that the roots of the polynomial $\beta(x) = 1$ lie outside the unit circle, this positivity constraint is satisfied if and only if all the coefficients in the infinite power series expansion for $\alpha(x)/(1 - \beta(x))$ are non-negative.

Property 1. *ARMA representation of GARCH model.* Just as an ARCH model can be expressed as an AR model of squared residuals, a GARCH model can be expressed as an ARMA model of squared residuals. Let consider the GARCH(1,1) model in equation 2.9. Since $E_{t-1}(u_t^2) = \sigma_t^2$, the equation 2.9 can be rewritten as: $u_t^2 = \omega + (\alpha_1 + \beta_1)u_{t-1}^2 + \epsilon_t - \beta_1\epsilon_{t-1}$, which is an ARMA(1,1) with $\epsilon_t = u_t^2 - E_{t-1}(u_t^2)$ being the white noise process. The condition for covariance stationary is $\alpha_1 + \beta_1 < 1$ and the unconditional variance is $\overline{\sigma}_t^2 = \dfrac{\omega}{1 - (\alpha_1 + \beta_1)}$.

Property 2. *Volatility clustering.* Let's consider the GARCH(1,1) model in equation 2.9. Usually β_1 is around 0.9 for many weekly or daily financial time series.

17

Following Zivot and Wang [123], this means that, large values of σ_{t-1}^2 will be followed by large value of σ_t^2, and small values of σ_{t-1}^2 will be followed by small values of σ_t^2. This process, is called volatility clustering in financial time series.

Property 3. *The GARCH model have fat tails.* In the financial time series, it is known that the distribution of many high frequency series usually has fatter tails than a Normal distribution. That is, large changes are more often to occur than a Normal distribution would imply. Assuming the fourth order moments exists, the kurtosis implied by a GARCH(1,1) process is greater than 3, the kurtosis of a Normal distribution.

Property 4. *Temporal aggregation produces convergence to normality.* Convergence to normality under temporal aggregation is a key feature of many economic data and is also a property of covariance stationary GARCH process. According to Diebold and Lopez [32], the key is that the law-frequency change is simply the sum of corresponding high-frequency changes. Thus, if a Gaussian central limit theorem can be invoked for sums of GARCH processes, convergence to normality under temporal aggregation can be assumed.

2.3 Extensions of ARCH and GARCH models

Following Brooks [27] and Tsay [116], the basic GARCH model provides a reasonably good model for analysing financial time series and estimating conditional volatility. However, there are some aspects which are not captured, since only squared residuals are used in the conditional variance equation. Signs and size bias of the residuals, asymmetry and leverage effects of the shocks are not taken into account. In this section 2.3, we present some variants of GARCH models that take account of some characteristics exhibited by financial times series.

2.3.1 The Integrated GARCH model

If the AR polynomial of the GARCH representation has a unit root, then we have an Integrated GARCH (IGARCH) model, introduced by Engle and Bollerslev [41]. Thus, the GARCH process is integrated of order one if $1 - \alpha(L) - \beta(L) = 0$ has a root on the unit circle.

Similar to Autoregressive Integrated Moving Average (ARIMA) models, a key feature of IGARCH is that of the impact of past squared shocks $\epsilon_{t-i} = u_{t-i}^2 - \sigma_{t-i}^2$ for $i > 0$ is persistent. An IGARCH(1,1) model can be written as:

$$\sigma_t^2 = \omega + (1 - \beta_1)u_{t-1}^2 + \beta_1\sigma_{t-1}^2 \quad and \quad u_t = z_t\sigma_t \qquad (2.10)$$

18

where $z_t \sim N(0,1)$ and $0 < \beta_1 < 1$.

Following Diebold and Lopez [32], the IGARCH process is potentially important because GARCH roots near to unity are common in high - frequency financial data. According to Tsay [116], the IGARCH process might be caused by occasional level shifts in volatility or to misspecification of the conditional variance function.

2.3.2 The Exponential GARCH model

According to Diebold and Lopez [32], one extension of the GARCH model is Exponential GARCH (EGARCH) model proposed by Nelson [90], which implies that the leverage effects are exponential. For the stationary return series r_t, the EGARCH(p,q) model of Nelson [90], is defined as:

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^{p} \alpha_i \left(\left| \frac{u_{t-i}}{\sigma_{t-i}} \right| - \sqrt{\frac{2}{\pi}} \right) + \sum_{i=1}^{p} \gamma_i \left(\frac{u_{t-i}}{\sigma_{t-i}} \right) + \sum_{j=1}^{q} \beta_j \ln(\sigma_{t-j}^2) \qquad (2.11)$$

where the term $\dfrac{u_{t-i}}{\sigma_{t-i}} = z_{t-i}$ is the standardized residual and $|z_{t-i}|$ the corresponding absolute value, ω is constant, the α_i parameters account for the magnitude or size of news effect, the β_j parameters capture time persistence in conditional volatility and γ_i measures the asymmetric effects of news, implying the so-called leverage effect.

The EGARCH(1,1) model can be written as follows,

$$\ln(\sigma_t^2) = \omega + \alpha_1 \left(\left| \frac{u_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + \gamma_1 \left(\frac{u_{t-1}}{\sigma_{t-1}} \right) + \beta_1 \ln(\sigma_{t-1}^2) \qquad (2.12)$$

Typically γ_1 is negative, positive return shocks generate less volatility than negative return shocks, assuming other factors to remain unchanged. When $\gamma_1 = 0$ good news and bad news of the same magnitude have the same effect on volatility, that is no leverage effect exists. The impact is asymmetric when $\gamma_1 \neq 0$. In the usual case when $\gamma_1 < 0$, the magnitude of the shock can be conditioned on the sign of z_{t-1}. For bad news ($z_{t-1} < 0$), the shock coefficient is measured by the sum $\alpha_1 + \gamma_1$, while for good news ($z_{t-1} > 0$), the magnitude is $\alpha_1 - \gamma_1$.

EGARCH capture the asymmetric reaction of variance towards positive and negative shocks in economy. Thus, the model often provides the best fits when compared to the standard GARCH model. The presence of the asymmetric term is largely responsible for the best fit since many asset return series exhibits leverage effects and the use of standardized shocks in the evolution of the log-variance tend to decrease the effect of large shocks.

According to Hentschel [59] and Nelson [90], due to the volatility specification in terms of the logarithmic transformation, there are no restrictions on the parameters to ensure a positive variance. A sufficient condition for stationarity is $\beta_1 < 1$.

2.3.3 The Threshold GARCH model

The Threshold GARCH (TGARCH) model proposed by Zakoian [121], extends equation 2.8 to allow the conditional standard deviation to depend upon the sign of the lagged innovations. Given a stationary return r_t, the general specification of the TGARCH (p,q) model is:

$$\sigma_t = \omega + \sum_{i=1}^{p} \alpha_i \big| u_{t-i} \big| + \sum_{i=1}^{p} \gamma_i \big| u_{t-i} \big| d_{t-i}(u_{t-i} < 0) + \sum_{j=1}^{q} \beta_j \sigma_{t-j} \qquad (2.13)$$

where d_{t-i} is an indicator function that takes the value 1 if $u_{t-i} < 0$ and zero otherwise. In particular, the TGARCH(1,1) model maybe expressed as:

$$\sigma_t = \omega + \alpha_1 \big| u_{t-1} \big| + \gamma_1 \big| u_{t-1} \big| d_{t-1}(u_{t-1} < 0) + \beta_1 \sigma_{t-1}$$

That is, the conditional standard deviation σ_t, depends on whether u_{t-1} is above or below the threshold value of zero. When $u_{t-1} > 0$ (good news), the total effects are given by $\alpha_1 u_{t-1}$, when $u_{t-1} < 0$ (bad news), the total effects are given by $(\alpha_1 + \gamma_1) u_{t-1}$. Following Ding *et al.* [34] and Hentschel [59], in order to guarantee the positive of σ_t, it is sufficient that $\omega > 0$, $\alpha_1 \geq 0$ and $\gamma_1 < \alpha_1$, but none is needed to ensure a positive variance. However, model stationary requires $\gamma_1^2 < 1 - \alpha_1^2 - \beta_1^2 - 2\alpha_1\beta_1 E(|z_t|)$, where $E(|z_t|) = \sqrt{2/\pi}$, for the Gaussian distribution. So we would expect $\gamma_1 > 0$ when bad news have larger impacts. Due to the Zakoian [121], the TGARCH model is sometimes referred to as ZARCH or ZGARCH model.

2.3.4 The GJR - GARCH model

Another model which account the leverage effect is the GJR-GARCH model, proposed by Glosten *et al.* [52]. The basic idea of the model is closed to the TGARCH model, however, here the model works with conditional variance instead of conditional standard deviation. The generalized form of the GJR-GARCH (p,q) model is given in the following form:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \gamma_i u_{t-i}^2 d_{t-i}(u_{t-i} < 0) + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \qquad (2.14)$$

where d_{t-i} is an indicator function that takes the value 1 if $u_{t-i} < 0$ and zero otherwise.

The GJR-GARCH(1,1) model is $\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \gamma_1 u_{t-1}^2 d_{t-1}(u_{t-1} < 0) + \beta_1 \sigma_{t-1}^2$, where the conditional volatility σ_t^2 is positive when parameters satisfy $\omega > 0$, $\alpha_1 \geq 0$, $\alpha_1 + \gamma_1 \geq 0$ and $\beta_1 \geq 0$, and under normality a GJR-GARCH(1,1) model is covariance stationary if the parameter restrictions are satisfied and $\alpha_1 + \frac{1}{2}\gamma_1 + \beta_1 < 1$.

The GJR-GARCH(1,1) model displays asymmetry when $\gamma_1 \neq 0$, in the sense that positive and negative shocks of equal magnitude have different effect on conditional volatility. Therefore, asymmetry exists if $\gamma_1 > 0$. The conditions for leverage effects are $\alpha_1 < 0$ and $\alpha_1 + \gamma_1 > 0$.

2.3.5 The Asymmetric Power ARCH model

The Asymmetric Power ARCH (APARCH) or APGARCH model of Ding *et al.* [34], parametrizes the non-linearity in the conditional variance using the power δ. This form provides greater flexibility in modelling the memory of volatility while remaining parsimonious. The model can well express the fat tails, excess kurtosis and leverage effects. The general structure of the APGARCH(p,q) process is as follows:

$$\sigma_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i \Big(|u_{t-i}| - \gamma_i u_{t-i} \Big)^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta \tag{2.15}$$

According to Ding *et al.* [34] and Hentschel [59], the usual restrictions on APGARCH parameters are that $\delta > 0$, $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ and $-1 < \gamma_i < 1$. Note that γ_i reflects the leverage effects. For $\gamma_i > 0$ the volatility tends to increase more when returns are negative, as compared to positive returns of the same magnitude.

The residuals u_t follows APGARCH$(1,1)$ model, if we can write $u_t = z_t \sigma_t$ where $z_t \sim N(0,1)$ and $\sigma_t^\delta = \omega + \alpha_1 \Big(|u_{t-1}| - \gamma_1 u_{t-1} \Big)^\delta + \beta_1 \sigma_{t-1}^\delta$.

The Generalized Asymmetric Power ARCH (APGARCH) model includes seven other ARCH extensions as special cases. The ARCH of Engle [39] when $\delta = 2$, $\gamma = 0$ and $\beta = 0$, the GARCH of Bollerslev [15] when $\delta = 2$, and $\gamma = 0$, the TS-GARCH of Taylor [115] or Schwert [105] when $\delta = 1$ and $\gamma = 0$, the GJR - GARCH of Gloslen *et al.* [52] when $\delta = 2$, the TARCH of Zakoian [121] when $\delta = 1$, the NARCH of Higgins and Bera [60] when $\gamma = 0$ and $\beta = 0$ and the log-ARCH of Geweke [51], Milhöj [84] and Pantula [96] when $\delta \to 0$. For details see Ding *et al.* [34].

In the literature, there is a wide variety of ARCH models specifications, so an exhaustive list can be found in Bollerslev [18].

2.3.6 The GARCH-in-mean model

Engle *et al.* [46], introduce the GARCH-in-mean (GARCH-M) model to modelling the relationship between risk and return when risk measured by the conditional variance, varies. The GARCH-M model is useful in financial applications when the conditional variance is related to the mean. In order to model such relations its use

21

the GARCH-M model. A simple GARCH(1,1)-M model can be written as:

$$
\begin{aligned}
r_t &= \mu + \theta\sigma_t + u_t \quad ; \quad u_t = z_t\sigma_t \\
\sigma_t^2 &= \omega + \alpha_1 u_{t-1}^2 + \beta_1\sigma_{t-1}^2
\end{aligned}
\tag{2.16}
$$

where the parameter θ is called the risk premium. If $\theta > 0$ that indicates the returns is positively related to its volatility. The model characterizes the evolution of the mean and the variance of a time series simultaneously. Other specifications of risk premium have also been used in the literature, such $r_t = \mu + \theta\sigma_t^2 + u_t$ and $r_t = \mu + \theta\ln(\sigma_t^2) + u_t$.

2.4 Estimation of GARCH models

2.4.1 Testing residual behavior

In addition to the Autocorrelation, Normality and Augmented Dickey-Fuller tests discussed in section 2.1, before an ARCH model is fitted it is necessary to check whether the residuals present evidence of ARCH effects and how they behave compared with the Normal distribution. It is also important to check the adequacy of the GARCH model specification after estimation. In this subsection, we present some preliminary test for the data series.

Testing for ARCH effects. According to Zivot and Wang [123], before estimating a full ARCH model for the financial time series, it is usually good practice to test for the presence of ARCH effect in the residuals. If there are no ARCH effects in the residuals, then the ARCH model is unnecessary and misspecified.

There are some formal methods to test for ARCH effects, such as McLeod-Li test, proposed by McLeod and Li [83], the Lagrange Multiplier test proposed by Engle [39], the BDS test, suggested by Broock *et al.* [26], etc. We describe the Lagrange Multiplier (LM) test, which is based on the residuals from linear regression $r_t = \mu + \phi_1 r_{t-1} + u_t$ and on auxiliary regression with p lags $\widehat{u}_t^2 = \omega + \alpha_1\widehat{u}_{t-1}^2 + ... + \alpha_p\widehat{u}_{t-p}^2 + \epsilon_t$. Under the null hypothesis, that there is no ARCH effects in squared residuals, $H_0 : \alpha_1 = ... = \alpha_p = 0$, The statistic is $LM = TR^2$, where T is the sample size and R is the sample multiple correlation coefficient from linear regression. The test statistic is asymptotically distributed as Chi-square with p degrees of freedom. We reject the null hypothesis if p-value is less than the significance level chosen α or if $LM \geq \chi_\alpha^2(p)$.

Testing for asymmetric effects. In the basic GARCH model 2.8, since only squared residuals u_{t-i}^2 are used in conditional variance equation, the signs of the residuals or shocks have no effect on conditional volatility. However, a stylized fact of financial volatility is that bad news tends to have a larger impact on volatility

than good news. Based on these facts, the asymmetric news impact on volatility is commonly referred to as the leverage effect.

Follows Zivot [122], a simple diagnostic for possible asymmetric leverage effects is the sample correlation between squared returns and lagged return, $corr(r_t^2, r_{t-1})$. A negative value of this correlation provides evidence for potential leverage effects. In addition to the asymmetric test, Engle and Ng [44], proposed a sign and size bias tests, that is, a negative size bias (NSB), positive size bias (PSB) and joint sign bias (SB) test for asymmetric of volatility processes. The tests are usually applied to the residuals from the asymmetric GARCH models. The sign and size bias test is based on the significance of the coefficient of the following regression:

$$u_t^2 = \phi_0 + \phi_1 S_{t-1}^- + \phi_2 S_{t-1}^- u_{t-1} + \phi_3 S_{t-1}^+ u_{t-1} + \epsilon_t \tag{2.17}$$

where $S_{t-1}^- = 1$ if $u_{t-1} < 0$ and 0 otherwise, and $S_{t-1}^+ = 1 - S_{t-1}^-$.

The SB, NSB and PSB statistics are t-ratios, they follow a standard Normal distribution asymptotically and $\epsilon_t \sim N(0,1)$. The null hypothesis ($H_0 : \phi_1 = \phi_2 = \phi_3 = 0$), can be evaluated by computing TR^2 and the test-statistic has an asymptotic χ^2 distribution with 3 degrees of freedom, where T is the sample size and R is the sample multiple correlation coefficient. A significant value of ϕ_1 indicates the existence of sign bias where positive and negative changes have asymmetric effects on conditional volatility compared to the symmetric GARCH model. On the other hand, the significance of ϕ_2 and ϕ_3 indicate not only sign bias but also if the size of the change is significant.

The presence of leverage effects among the asymmetric models is also examined by testing the null hypothesis that $\gamma = 0$ at a significance level. Rejection of the null hypothesis implies the presence of leverage effects in the returns.

2.4.2 Method of maximum likelihood estimation

If an ARCH effect is found to be significant, the partial autocorrelation function (PACF) of u_t^2 can be used to determine the ARCH order. For an ARCH(p) model, the lag-p sample PACF should not be zero, but $\widehat{\alpha}_i$ should be close to zero for all $i > p$. Making use of this fact to determine the order p, we can consider the following properties: (i) $\widehat{\alpha}_p$ converges to α_p as the sample size T goes to infinity. (ii) $\widehat{\alpha}_l$ converges to zero for all $l > p$ and (iii) the asymptotic variance of $\widehat{\alpha}_t$ is $1/T$ for $l > p$. These results says that, for ARCH(p) series, the sample PACF cuts off at lag-p. Using the plot of PACF, we select p when the lags of sample PACF appear to be large than the approximate two standard error limits $\pm 1.96/\sqrt{T}$.

23

The method of maximum likelihood estimation (MLE) is the most common approach used to estimate the vector of unknown parameters θ of conditional volatility models. According to Lütkepohl and Krätzig [78], the conditions for the inference in maximum likelihood (ML) estimator to hold require that the parameters estimates are maximum likelihood estimates which in turn both the likelihood used in estimation to be correct as well as the specification for the conditional variance. When one specification is used for estimation for example Normal distribution but the data follow a different conditional distribution, these estimators are known as Quase Maximum Likelihood Estimator (QMLE), and the estimated parameters are asymptotically Normal but with a different covariance.

There are three conditional likelihood functions used in ARCH or GARCH estimation. The conditional likelihood under the normality assumption of standardized residuals $z \sim N(0,1)$, under Student t-distribution, when standardized residual follows a heavy-tailed distribution and under specific probability density function, when z_t may assume a generalized error distribution (GED) with probability density function $f(x)$.

a) *Under normality assumption.* Under regularity conditions, the normal likelihood function for the sequence random variables $\mathbf{u}_{T-1}$ containing $u_0, u_1, ..., u_{T-1}$ of an GARCH process is:

$$f(\theta|u_1,...,u_{T-1}) = \prod_{t=1}^{T} (2\pi\sigma_t^2)^{-\frac{1}{2}} exp\Big(-\frac{u_t^2}{2\sigma_t^2} \Big) \tag{2.18}$$

and the normal log-likelihood function is defined as:

$$logL(\theta|u_1,...,u_{T-1}) = -\frac{T}{2}log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}log(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^{T}\frac{u_t^2}{\sigma_t^2} \tag{2.19}$$

According to Tsay [116] and Zivot [122], essentially, the MLE method works by finding the most likely values of the parameters given the historical data and the estimation can be done in the following ways.

1. Specify the appropriate equations for conditional mean and variance.

2. Under the normality assumption and since the term $\log(2\pi)$ does note involve any parameter, the log-likelihood function become:

$$logL = -\frac{1}{2}\sum_{t=1}^{T}log(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^{T}\frac{u_t^2}{\sigma_t^2} \tag{2.20}$$

3. The maximum likelihood estimation is the specification parameter vector θ that maximizes the log-likelihood function.

There are several practical issues to consider in the maximization of $logL$. Starting values for the model parameters need to be chosen. Once the log-likelihood is initialized, it can be maximized using numerical optimization techniques such as Newton-Raphson, BHHH, BFGS and DFP. The most common method is based on a Newton-Raphson iterative optimization routine of the form:

$$\widehat{\theta}_i = \widehat{\theta}_{i-1} + \phi \left(\sum_{i=1}^{T} \frac{\partial logL_t}{\partial \theta} \frac{\partial logL_t}{\partial \theta'} \bigg|_{\theta=\widehat{\theta}_{i-1}} \right)^{-1} \left(\sum_{i=1}^{T} \frac{\partial logL_t}{\partial \theta} \bigg|_{\theta=\widehat{\theta}-1} \right) \quad (2.21)$$

$$with \quad \mathbf{H}(\widehat{\theta}_i) = \sum_{i=1}^{T} \frac{\partial logL_t}{\partial \theta} \frac{\partial logL_t}{\partial \theta'} \quad and \quad \mathbf{S}(\widehat{\theta}_i) = \sum_{i=1}^{T} \frac{\partial logL_t}{\partial \theta}$$

where θ_i denotes the vector of estimated parameters at iteration i, ϕ is a scalar used to modify step-length, $\mathbf{S}(\widehat{\theta}_i)$ and $\mathbf{H}(\widehat{\theta}_i)$ denotes the Gradient or Score vector and Hessian matrix of log-likelihood at iteration i, respectively.

The step-length parameter ϕ_i is chosen such that $log(\theta_i) \geq log(\theta_{i-1})$. For ARCH models, the BHHH algorithm (Berndt et $al.$ [9])[2], is often used to approximate the Hessian matrix using only first derivative information.

According to Brooks [27], by default, many software already has initial values of zero for the conditional variance parameter. Note that, in practice when necessary it is possible to initialize with non-zero values. The specification of the stopping condition depends on the author and can be chosen in two ways: using a convergence criterion associated with a limit value that maximizes $logL$ or by a fixed value that must be reached. In either case, iterations only ends when the difference between two consecutive coefficients estimates is minimum possible. For some software when a stop condition is not indicated, the default value 0.001 is assumed as the condition for stop search the optimal solution. That is, when the difference between two consecutive estimates for all parameters is on the order of 0.1%, assumes that the optimal solution is achieved and the search process ends.

Follows Davidson [30], under regularity conditions the maximum likelihood estimator $\widehat{\theta}$ converges at rate $\sqrt{T}$ and is asymptotically normally distributed, that is: $\sqrt{T}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, \mathbf{\Sigma}^{-1})$, where $\mathbf{\Sigma}$ is the expectation of the outer product of the scores of it $logL(\theta)$,

$$\mathbf{\Sigma} = \frac{1}{T} \sum_{i=1}^{T} E \left[\frac{\partial logL_t}{\partial \theta} \frac{\partial logL_t}{\partial \theta'} \right] \quad (2.22)$$

The log-likelihood function in equation 2.19, is determined under the assumption of conditional normality of innovations presented in equation 2.8, that is $z_t \sim N(0,1)$. Ignoring non normality of innovations z_t we get a misspecification of the log-likelihood

---

[2]BHHH due to Berndt, Hall, Hall and Hausman.

function. Maximizing the misspecified Gaussian log-likelihood function is justified by Quasi Maximum Likelihood (QML) theory. Analytical expressions for the derivatives necessary to implementing the BHHH algorithm or QML inference are given in Bollerslev [15] for the general case of ARMA(p,q) processes with GARCH(q,p) error terms.

b) *Under Student t-distribution.* In order to account the fat-tailed error distribution, Bollerslev [16], proposed fitting a GARCH model with a Student t-distribution for standardized residual. If a random variable u_t has a Student t-distribution with v degrees of freedom, then $var(u_t) = v/(v-2)$ for $v > 2$, and using the relation $z_t = u_t/\sqrt{v/(v-2)}$, the probability density function (pdf) of standardized residual z_t is given by:

$$f(z_t|v) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{(v-2)\pi}}\Big(1 + \frac{z_t^2}{(v-2)}\Big)^{-(v+1)/2} \tag{2.23}$$

where $\Gamma(v) = \int_0^\infty e^{-z}z^{v-1}dz$ is Gamma function and v is parameter that measures the tail thickness. According to Tsay [116], the value of degrees of freedom v varies between 3 and 6, when its chosen a priori.

Using $u_t = z_t\sigma_t$, the conditional log-likelihood function of u_t for specified v is:

$$\begin{aligned} logL(u_t|v) &= T\Big[log\Gamma\Big(\frac{v+1}{2}\Big) - log\Gamma\Big(\frac{v}{2}\Big) - \frac{1}{2}log(\pi(v-2))\Big] \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}\Big[log(\sigma_t^2) + (1+v)log\Big(1 + \frac{u_t^2}{\sigma_t^2(v-2)}\Big)\Big] \end{aligned} \tag{2.24}$$

c) *Under generalized error distribution.* If the standardized residuals follow a GED, Nelson [90], proposed to use this distribution to capture the fat tails usually observed in the distribution of financial time series. If z_t has a GED with mean zero and variance 1, the probability density function of z_t is given by:

$$f(z_t|v) = \frac{v \times exp\Big(-(1/2)|z_t/\lambda|^v\Big)}{\lambda \times 2^{((v+1)/v)}\Gamma(1/v)} \quad where \quad \lambda = \Big[\frac{2^{(-2/v)}\Gamma(1/v)}{\Gamma(3/v)}\Big]^{1/2} \tag{2.25}$$

The parameter v is positive, governing the thickness of the tail behavior of the distribution. When $v = 2$ the distribution reduces to a Gaussian if $v < 2$ the density has ticks tails than the normal density when $v > 2$ the density has thinner tails than the normal. When $v = 1$ the density reduce to the pdf of double exponential distribution.

2.4.3 Model diagnostic, forecasting and evaluation

Model diagnostic

Following Diebold and Lopez [32], after fitting the volatility equation it is necessary to carefully check whether the estimated model is the most appropriate, if necessary the model can be refined by removing the parameters that are not statistically significant.

The most commonly used test of significance for each parameter individually is the called t-Student test, with the null hypothesis. $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$. The test statistic is:

$$t = \frac{\widehat{\beta}}{Se(\widehat{\beta})} \tag{2.26}$$

The test follows a t-distribution with T-p degrees of freedom, where T is the number of observations and p the number of estimated parameters. The null hypothesis is rejected if $\widehat{\beta}$ is too far from zero, that is, if observed statistic in absolute value is greater than the critical value.

The test for the whole model or joint test for significance of all coefficients except the intercept is based on the F-distribution with the null hypothesis $H_0 : \forall \beta_i = 0$ for $i = 1, ..., p$ against the alternative $H_1 : \exists \beta_i \neq 0$. The F - statistic is given by:

$$F = \frac{(SSR_r - SSR_u)/m}{SSR_u/(T-p)} \sim F(m, T-p) \tag{2.27}$$

where SSR_u and SSR_r indicate the sum of squared of residuals for unrestricted and restricted models, T the sample size, p the number of parameters of the unrestricted model and m is the number of parameter restrictions. The null hypothesis is rejected if F-statistic is greater than the correspondent critical value.

By default, for parameter diagnostic is showing the standard errors, p-value and t-statistic from the output of statistical packages.

Forecasting volatility

According to Zivot and Wang [123], one of the objectives of analysing conditional volatility models is to make predictions of the future value of volatility. Since the GARCH model in equation 2.8 has a representation of the ARMA process, the predictions can be made similar to the procedure used in ARMA models. Predicting out-of-sample conditional variance over time using GARCH models can often yield accurate forecasts of future volatility, especially over short horizons.

For simplicity, consider the GARCH(1,1) model and assume that the forecast origin is h. For 1-step ahead forecast, we have: $\sigma_{h+1}^2 = \omega + \alpha_1 u_h^2 + \beta_1 \sigma_h^2$, where

u_h^2 and σ_h^2 are known at the time index h. Therefore, the 1-step ahead forecast is $\sigma_h^2(1) = \omega + \alpha_1 u_h^2 + \beta_1 \sigma_h^2$.

For multi-step ahead forecast, we can use $u_t^2 = \sigma_t^2 z_t^2$ and rewrite the volatility equation 2.9 as $\sigma_{t+1}^2 = \omega + \alpha_1 \sigma_t^2 z_t^2 + \beta_1 \sigma_t^2$. Now adding and subtracting $\alpha_1 \sigma_t^2$, and realized few algebra operations we get $\sigma_{t+1}^2 = \omega + (\alpha_1 + \beta_1)\sigma_t^2 + \alpha_1 \sigma_t^2(z_t^2 - 1)$. When $t=h+1$, the last equation becomes: $\sigma_{h+2}^2 = \omega + (\alpha_1 + \beta_1)\sigma_{h+1}^2 + \alpha_1 \sigma_{h+1}^2(z_{h+1}^2 - 1)$. Since $E(z_{h+1}^2 - 1|\mathscr{F}_h) = 0$, the 2-step ahead volatility forecast at the origin h satisfies the equation $\sigma_h^2(2) = \omega + (\alpha_1 + \beta_1)\sigma_h^2(1)$. In general, we have

$$\sigma_h^2(k) = \omega + \sum_{i=1}^{k-1}(\alpha_1 + \beta_1)^{i-1} + (\alpha_1 + \beta_1)^{k-1}\sigma_h^2(1) \quad ; \quad k \geq 2 \tag{2.28}$$

Notice that, as $k \to \infty$ the volatility forecast in equation 2.28 approaches the unconditional variance $\omega/(1 - \alpha_1 - \beta_1)$, if the GARCH process is stationary, that is $\alpha_1 + \beta_1 < 1$.

Model evaluation and news impact curves

According to Zivot [122], the fit evaluation of a GARCH model can be done using graphical representation or through more formal statistics tests. If the GARCH is correctly specified, the estimated series of standardized residuals $z_t = u_t/\sigma_t$ and their squared z_t^2 should not display serial correlation, conditional heteroskedasticity or any type of nonlinear dependence. The modified Ljung-Box statistics 2.4 can be used to test the null hypothesis of no autocorrelation up to specific lag and Engle's LM statistics described in subsection 2.4.1, can be used to test the null of no remaining ARCH effects. If it is assumed that the errors are Gaussian, then a plot of z_t against time should have approximately 95% of its values between ± 2; a normal qq-plot should look approximately linear; and JB statistics, equation 2.5, should not be too much larger than six.

According to Yu [120], among the measures used to evaluate the accuracy of prediction stands out the root mean square error (RMSE), the mean absolute error (MAE), the U-Thail statistic and mean absolute percentage error (MAPE).

$$MAE = \frac{1}{T}\sum_{i=1}^{T}|\sigma_t^2 - \widehat{\sigma}_t^2| \quad ; \quad RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}\left(\sigma_t^2 - \widehat{\sigma}_t^2\right)^2} \tag{2.29}$$

Measurements of MAE and RMSE are the most popular measures used to check whether a given model makes good predictions. In terms of interpretation, small values or near to zero, indicate the model is well specified and therefore can present

the best estimates within the sample, consequently better predictions.

$$MAPE = \frac{1}{T}\sum_{i=1}^{T}\left|\frac{\sigma_t^2 - \widehat{\sigma}_t^2}{\sigma_t^2}\right| \quad ; \quad U-Thail = \frac{\sum_{i=1}^{T}\left(\sigma_t^2 - \widehat{\sigma}_t^2\right)^2}{\sum_{i=1}^{T}\left(\sigma_{t-1}^2 - \sigma_t^2\right)^2} \tag{2.30}$$

The value of U-Thail is an indication of systematic error since it measures the extent to which the average of the forecast variance and actual series deviates from each other. The U-Thail statistics take values between 0 and 1. Values close to zero, indicate that the model has good performance in forecasting, but values close to one, is not necessarily an indication of bad forecasting performance.

If more than one GARCH models are all capable to modelling leverage effects, such as EGARCH, TGARCH and APGARCH, the choice of appropriate model can be made by using a model selection criterion such as the Bayesian information criterion (BIC), Akaike Information Criterion (AIC). Alternatively, Engle and Ng [44], proposed to use the news impact curve (NIC) introduced by Pagan and Schwert [95], to compare different models.

Following Engle and Ng [44], the news impact curve is a graph of the conditional variance at time *t* as a function of the error term at time *t-1*, assuming that all previews information are constant. If the news impact curve is symmetrical about zero, then the shocks of the same absolute value have the same impact on future conditional variance, otherwise, the shocks of the same magnitude have different effects on future conditional variance.

In next section 2.5, we present an application of univariate GARCH model, especially the asymmetric GARCH models, which is a replication of Faias *et al.* [48] paper, but here we explore a wider range of asymmetric GARCH models and with accrued validation in order to get the best possible model specification.

2.5 Application of univariate GARCH models

Exchange rate uncertainty is key in a globalized economy, both for capital movements and trade balance. Furthermore, as documented, for example in Van der Ploeg [100], volatility can be harmful to macroeconomic performance. A specific notable example is Obeng and Sakyi [92], where exchange rate volatility affects interest rate spreads, key for financing decisions. For another hand, Mozambique currency has gone through distinct periods in terms of the exchange rate, where Mozambique economy has been experiencing an increasing exploration of natural resources, bringing the Dutch disease issue to discussion, which puts accrued care with respect to international operations, either in financing or in trading goods and services.

In this context, for the application of univariate GARCH models, we use Mulenga et al. [87] results which are an extension of Faias *et al.* [48] paper. That is, we use asymmetric GARCH models for daily Mozambique New Metical (MZN) against South Africa Rand (ZAR), a major trading partner, MZN/ZAR exchange rate over January 2010 to December 2014, where four sub-periods have been identified, corresponding to differences in exchange rate behavior. Then, we employ various popular asymmetric GARCH models, which may deliver distinct results, among themselves and according to the sample considered, either the full sample or its subsets. The aim in the application is to identify the possible sign and size effects of news across the distinct exchange rate trends, as well as inspect if and how results that hold for a full heterogeneous sample can be misleading comparing with the results from each particular sub-sample.

The remaining of the section 2.5, is organized as follows: in subsection 2.5.1, we discuss the data and the conditional mean estimation, with particular attention to the statistical properties exhibited; in subsection 2.5.2, we present the asymmetric models, their validation and results; finally, relevant findings are reported in subsection 2.5.3.

2.5.1 Data description

Here, we aim at establishing the statistical properties of the data and those for the residuals from the estimation of the conditional mean process. In Figure 2.1, we display the daily MZN/ZAR exchange rate, y_t, over 2010-2014, a period particularly rich in distinct sub-periods with specific exchange rate behavior.[3] From this plot, we can observe four marked sub-periods: 2010-depreciation, 2011-appreciation, 2012-stability and 2013-2014-stability with some appreciation.

We proceed by checking the above patterns, computing beforehand the percentage change in exchange rate, r_t (standing for returns), as first differences of the log series, $r_t = [\ln(y_t) - \ln(y_{t-1})] \times 100$, which is the object of analysis. Then, we run descriptive statistics and check for stationarity, applying Augmented Dickey-Fuller testing in each sample period. As shown in Table 2.1, the series are, in every sample considered, stationary and from the mean variation in each sub-period, we confirm the above mentioned exchange rate distinct behavior.

Next, we look into volatility and eventual signs of asymmetry. From Table 2.1 one reads that there is excess kurtosis, which signals eventual ARCH effects (Brooks [27]), in each sample period, particularly strong in the depreciation sub-period and less pronounced in the stability sub-period. Following Black [12], we compute, as in

---

[3]Data collected at http://www.oanda.com/currency/historical-rates.
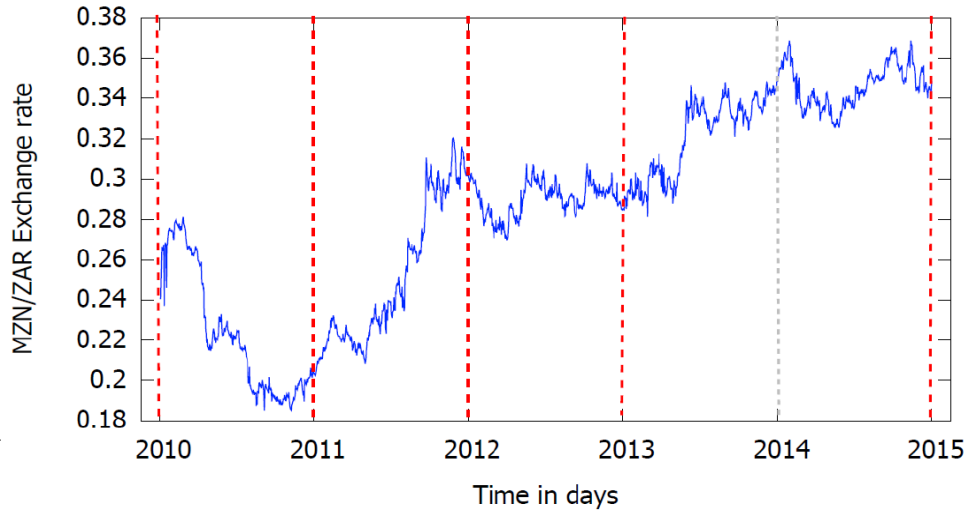
Figure 2.1: MZN/ZAR Exchange rate, 2010-2014

Zivot [122], the correlation between r_t^2 and r_{t-1}, to look for eventual asymmetry in the shocks to the variance. The results show that in periods of appreciation, with or without some stability, there is no asymmetry. Finally, the autocorrelation of absolute returns, $|r_t|$, as in Ding *et al.* [34], suggests the existence of long memory.

Table 2.1: Descriptive statistics for the data and residuals from mean regression

Statistics	2010-2014	2010	2011	2012	2013-2014
Data					
Mean	0.0236	-0.0544	0.1276	-0.0166	0.0320
ADF z(t)	-21.428 [0.0000]	-15.259 [0.0000]	-12.584 [0.0000]	-11.358 [0.0000]	-18.137 [0.0000]
Kurtosis	16.6216	19.8040	6.5273	5.1835	6.5867
$\text{corr}(r_t^2, r_{t-1})$	-0.1448	-0.2794	0.0019	0.1217	-0.0053
$\text{corr}(r_t, r_{t-1})$	-0.1775	-0.2621	-0.0763	-0.1572	-0.1718
$\text{corr}(|r_t|, |r_{t-1}|)$	0.2491	0.3262	0.1240	0.1464	0.2070
Residuals					
Ljung-Box Q	1.3934 [0.8433]	1.6632 [0.7974]	4.8063 [0.5926]	4.3624 [0.3624]	2.6667 [0.6151]
ARCH-LM	66.957 [0.0000]	42.216 [0.0000]	16.682 [0.0022]	12.258 [0.0155]	42.801 [0.0000]
Observations	1564	312	312	312	625

Note: *p*-value in brackets.

Altogether, the evidence points out to different statistical properties of exchange rate series over marked periods, implying that running a full sample may be misleading. Therefore, we explore the specificity of each sub-period and adjust various

31

volatility models that allow for asymmetry.

Following, before running ARCH models, we first specify the mean equations, presented in general form as ARMA(p,q) processes, identified in a standard way – through autocorrelation functions along with information criteria – employing Ljung-Box test as diagnostic for autocorrelation. Secondly, we perform a diagnostic test for ARCH effects, namely Engle LM statistic.

Specifically, we estimate the equation below,

$$r_t = \mu + \sum_{i=1}^{5} \beta_i d_i + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{j=1}^{q} \theta_j u_{t-j} + u_t \tag{2.31}$$

where μ is a constant, d_i is a dummy variable that takes the value 1 if the day of the week is i and zero, otherwise. $i = 1, ..., 5$ where $i = 1$ for Monday, $i = 2$ for Tuesday, $i = 3$ for Wednesday, $i = 4$ for Thursday and $i = 5$ for Friday and u_t is the residual, also named shock or news. The results for Ljung and Box and ARCH-LM tests, displayed in Table 2.1, ensure a correct specification and, importantly, evidence for the existence of significant ARCH effects in all sample periods.

Summing up, there is statistical evidence for ARCH effects and signs of asymmetry in most of the samples considered. Thereby, can continue the analysis by introducing the process for the conditional variance employing mainstream ARCH models that allow for asymmetry and eventual effects.

2.5.2 Model results

Given the statistical findings from the previous sub-section 2.5.1, we proceed to report the results for most popular asymmetric ARCH models that are able to capture persistence, along with sign and size effects of shocks to the variance. In particular, we estimate four variants of asymmetric GARCH family discussed in section 2.3, namely: The Exponential GARCH (EGARCH) of Nelson [90], the Threshold GARCH (TGARCH) of Zakoian [121], the GJR-GARCH of Glosten *et al.* [52] and the Asymmetric Power ARCH (APARCH) of Ding *et al.* [34].

The specifications estimated and their estimates, with respective inference, are herein analysed.[4] For the conditional mean, the ARMA(1,1) model prevailed in all samples considered, except for the appreciation period, the year 2011, where an ARMA(1,2) is estimated. In these processes, the Monday effect is the strongest.

As to the conditional variance, the GARCH(1,1) specification considered for all asymmetric models: EGARCH(1,1), TGARCH(1,1), GJR-GARCH(1,1) and AP-GARCH(1,1), is used as a standard common window to look into each sample and

---

[4]Computations have been conducted using STATA.

compare across samples. The estimates from these models, presented in Table 2.2, discussed here, are valid given the diagnostic statistics outcome, displayed in Table 2.3.

Table 2.2: Estimates of the asymmetric ARMA-GARCH models

Period	Model	ω	α	γ	β
Full sample	EGARCH	0.0295 [0.0000]	−0.0059 [0.4820]	0.2611 [0.0000]	0.9623 [0.0000]
2010 − 2014	TGARCH	0.0381 [0.0000]	0.1339 [0.0000]	−0.0001 [0.990]	0.8724 [0.0000]
	GJR-GARCH	0.0691 [0.0000]	0.1143 [0.0000]	0.0055 [0.6800]	0.8402 [0.0000]
	APGARCH [Power=1.89]$^{(*)}$	0.0641 [0.0000]	0.1186 [0.0000]	0.0091 [0.7560]	0.8468 [0.0000]
Depreciation	EGARCH	−0.0406 [0.0180]	0.0840 [0.0007]	−0.1590 [0.0000]	0.9453 [0.0000]
2010	TGARCH	0.0922 [0.0070]	0.2316 [0.0000]	−0.0266 [0.3980]	0.7887 [0.0000]
	GJR-GARCH	0.1917 [0.0000]	0.1995 [0.0029]	−0.0332 [0.4919]	0.7507 [0.0000]
	APGARCH [Power=1.95]$^{(*)}$	0.1805 [0.0000]	0.1838 [0.0002]	−0.0176 [0.7785]	0.7563 [0.0000]
Appreciation	EGARCH	0.0128 [0.1050]	0.0010 [0.9700]	0.1421 [0.0000]	0.9965 [0.0000]
2011	TGARCH	0.0105 [0.4330]	0.0760 [0.0030]	0.0130 [0.6340]	0.9335 [0.0000]
	GJR-GARCH	0.0178 [0.1050]	0.0829 [0.0000]	−0.0869 [0.0020]	0.9587 [0.0000]
	APGARCH [Power=3.33]$^{(*)}$	0.0785 [0.1310]	0.0327 [0.4920]	−0.6523 [0.2660]	0.8781 [0.0000]
Stability	EGARCH	0.0199 [0.5340]	0.1734 [0.0000]	0.0320 [0.6860]	0.6112 [0.0110]
2012	TGARCH	0.3930 [0.1030]	−0.0878 [0.0970]	0.1736 [0.0090]	0.6192 [0.0070]
	GJR-GARCH	0.3358 [0.0354]	0.0921 [0.0845]	−0.2196 [0.0035]	0.7007 [0.0000]
	APGARCH [Power=3.35]$^{(*)}$	0.7654 [0.0021]	0.0132 [0.0864]	−0.9418 [0.0000]	0.1962 [0.1906]
Stab-Apprec	EGARCH	0.0075 [0.0930]	0.0094 [0.6030]	0.1688 [0.0000]	0.9876 [0.0000]
2013 − 2014	TGARCH	0.0116 [0.0460]	0.0758 [0.0000]	0.0194 [0.2840]	0.9254 [0.0000]
	GJR-GARCH	0.0138 [0.0150]	0.0587 [0.0000]	0.0232 [0.3200]	0.9176 [0.0000]
	APGARCH [Power=2.16]$^{(*)}$	0.0140 [0.0130]	0.0661 [0.0130]	0.0656 [0.4380]	0.9157 [0.0000]

Note: $^{(*)}$ power is significant at 5%; p-value in brackets.

Regarding the full sample, only the EGARCH model shows significant asymmetry of shocks, but no size effect and the good news have the larger effect on volatility,

contrary to what can be expected from the earlier empirical literature. The remaining TGARCH, GJR-GARCH and APGARCH models show similar results, although there is no asymmetry, there is persistence in volatility and size effects in all models.

Concerning the sub-samples, results differ among them and, importantly, they are often at odds with the full sample findings. This means that overlooking heterogeneous periods, corresponding to alternative currency trend behavior, give misleading results for volatility estimation and assessment of news impact.

Specifically, for the depreciation ("bearish") period, the year 2010, from EGARCH model we get the usual result that negative shocks play the larger role and that size matters, larger shocks affect more the conditional variance. For the APGARCH family, these same results hold although actually without statistical significance. Notice that concordance among the set of models and of the findings against previous literature is verified in this sub-period, and this only.

As to the appreciation ("bullish") period, 2011, the EGARCH shows, interestingly, that positive shocks have the stronger impact, but the size is not significant, thereby large and positive shocks affect exchange rate volatility in the same way. Similar results are also exhibited in the stability-with-appreciation period, 2013-2014. However, it is worth noting that in the appreciation time, GJR-GARCH model points out, contrary to EGARCH, that negative shocks have a larger role. On the other hand, TGARCH and APGARCH (about power 3) indicate that all shocks are equal. Thus, at odds with 2010, the set of models show clear disagreement on which news matter more.

The stability period, 2012, is lively in a variety of results with a noteworthy pattern. To be precise, EGARCH model says that the size of shocks matter, the larger the shock the greater the impact on volatility, no asymmetry. The APGARCH family shows, in particular, the relevance of the sign of news affecting the conditional variance but as power grows the relative role of good-bad news evolves. More precisely, the TGARCH ($\delta = 1$) says that positive shocks are more important, conversely, GJR-GARCH ($\delta = 2$) points out the negative ones and the APGARCH itself (δ estimated, near 3) puts it with greater emphasis.

Looking at Table 2.4, the sign tests confirm in general, the results from estimates. That is, taking sign and size bias, individually or jointly, we obtain the described prominence of negative shocks in "bad"/ depreciation times and positive shocks "good"/appreciation and stability-with-appreciation times, although mild in the appreciation trend.

The news impact curves reinforce the above findings. For the purposes of illustration, Figures 2.2 and 2.3, show the pattern found.

Overall, EGARCH model delivers results on asymmetry and size which are the

Table 2.3: Diagnostic on estimated GARCH models

Period	model	Ljung-Box Q(15)	ARCH-LM(15)
Full sample	EGARCH	17.103[0.313]	21.993[0.108]
2010-2014	TGARCH	22.651[0.092]	23.260[0.078]
	GJR-GARCH	11.963[0.682]	18.047[0.260]
	APGARCH	12.208[0.663]	18.682[0.228]
Depreciation	EGARCH	7.503[0.942]	17.219[0.306]
2010	TGARCH	17.364[0.297]	14.761[0.468]
	GJR-GARCH	9.185[0.867]	11.205[0.738]
	APGARCH	9.098[0.872]	11.246[0.735]
Appreciation	EGARCH	9.476[0.851]	15.627[0.407]
2011	TGARCH	7.897[0.928]	16.011[0.381]
	GJR-GARCH	10.514[0.786]	16.501[0.350]
	APGARCH	18.906[0.218]	16.898[0.325]
Stability	EGARCH	7.846[0.930]	7.582[0.940]
2012	TGARCH	8.030[0.923]	8.463[0.904]
	GJR-GARCH	11.003[0.752]	9.622[0.843]
	APGARCH	7.841[0.930]	7.839[0.930]
Stab-Apprec	EGARCH	20.954[0.138]	18.574[0.234]
2013-2014	TGARCH	21.204[0.130]	19.934[0.174]
	GJR-GARCH	21.065[0.135]	19.113[0.208]
	APGARCH	20.990[0.137]	18.659[0.230]

Note: p-value in brackets.
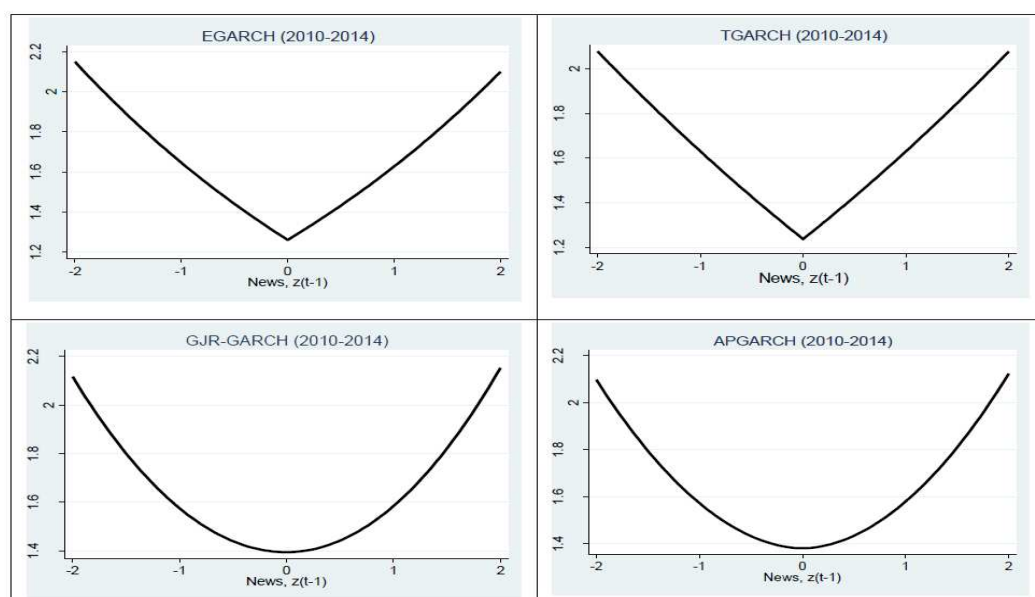
Figure 2.2: News impact curves for 2010-2014

reverse of those exhibited by the remaining APGARCH family models. Disregarding

35

Table 2.4: Test of asymmetries of volatility processes

Period	model	SB	NSB	PSB	Joint test
Full sample	EGARCH	-1.2695[***]	-2.3630[***]	0.6094[***]	118.788[***]
2010-2014	TGARCH	-1.1789[***]	-2.2354[***]	0.6477[***]	112.067[***]
	GJR-GARCH	-1.2634[***]	-2.2878[***]	0.5694[***]	110.973[***]
	APGARCH	-1.2627[***]	-2.2854[***]	0.5713[***]	110.817[***]
Depreciation	EGARCH	-3.4138[**]	-4.4857[***]	0.5485	53.4440[***]
2010	TGARCH	-3.4085[*]	-3.8222[***]	0.9996	39.1550[***]
	GJR-GARCH	-2.6871[*]	-3.5928[***]	0.9240	36.7350[***]
	APGARCH	-2.7230[*]	-3.6220[***]	0.9067	37.1380[***]
Appreciation	EGARCH	-0.4895	-0.3866	0.1878	1.4928
2011	TGARCH	0.2749	-0.2740	0.4091	2.0526
	GJR-GARCH	-0.0288	-0.3529	0.2902	1.5550
	APGARCH	-0.0703	-0.4269	0.2322	1.3373
Stability	EGARCH	1.0662[***]	0.4356	0.8250[***]	15.2390[***]
2012	TGARCH	1.0286[***]	0.4450	0.7779[***]	13.9639[***]
	GJR-GARCH	1.2110[***]	0.3633	0.9705[***]	18.2590[***]
	APGARCH	1.1856[***]	0.3849	0.9870[***]	19.3130[***]
Stabil-Apprec	EGARCH	0.4155	-0.4614[***]	0.7521[***]	24.2112[***]
2013–2014	TGARCH	0.4702[*]	-0.3900[**]	0.7992[***]	24.0240[***]
	GJR-GARCH	0.3497	-0.4429[**]	0.7428[***]	22.8380[***]
	APGARCH	0.3588	-0.4561[***]	0.7395[***]	23.1504[***]

Note: [*],[**] and [***] significance at 10%, 5% and 1%, respectively; p-value in brackets,
SB (sign bias), NSN (negative size bias), PSB (positive size bias)
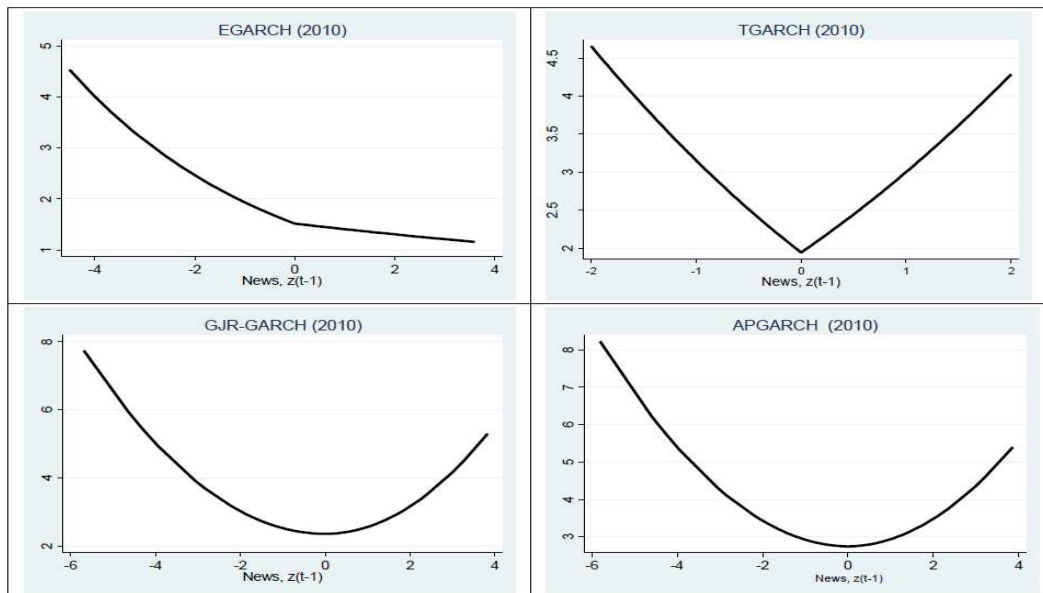
Figure 2.3: News impact curves for 2010

heterogeneity in the sample, as to the currency trend, leads to incorrect conclusions on volatility and type of news impact. Two key novel outcomes are observed: (i)

positive shocks have a larger impact on currency appreciation and stability-with-appreciation times; (ii) the major role for positive versus negative news is reversed, more clearly in the stability time, starting with positive ones and as the power of the APGARCH increases, i.e., going from TGARCH ($\delta = 1$), to GJR-GARCH ($\delta = 2$) and the APGARCH as such (power is estimated), the negative ones emerging as those with larger impact.

Lastly, two broad implications are drawn: (i) role of type and size of shocks are currency trend specific; (ii) model choice is relevant, putting emphasis on selection criteria as well as a robustness analysis is due. Recall that all players in international markets need to assess exchange rate risk, i.e., persistent volatility, and take into account which news has a major effect. Importantly, the central banks also care for this issue in their daily exchange rate management, especially when pressure is stronger, as may be the case in natural resource producing countries.

2.5.3 Conclusions

Exchange rates and especially their volatility are crucial in open economies, particularly in natural resource producing countries, because of their implications for economic performance, as a whole, and for international capital movement and goods and services transactions.

We focus not only on measuring exchange rate risk, that is the volatility that persists, but inspect if news size and sign affect variance in a different way, when considering a full heterogeneous sample according to currency trend behavior and we check whether results are sensitive to model choice. For this purpose we divide the full sample into four periods with distinct currency behavior – 2010-depreciation, 2011-appreciation, 2012-stability and 2013-2014 stability with appreciation – and apply mainstream asymmetric GARCH models, namely Exponential GARCH(1,1), Threshold GARCH(1,1), GJR-GARCH(1,1) and Asymmetric Power GARCH(1,1). The validation of models is done with sign-size bias tests, along with the plots of news impact curves.

Our results bring new contributions and reinforce the existing literature. We find that model choice is key for the conclusions, highlighting the importance of selection criteria and diagnostic testing. Novel contributions come from the pervasive sensitivity of results to currency trends, as well as modelling options in nontrivial ways. Indeed, results from the full heterogeneous sample masks to some extent, the existence of size and sign effects of news. In detail, the evidence suggests that in depreciation times the significance of asymmetry and importance of size effects are stronger, whereas in stability periods exchange rate risk seems to be lower (less

significant ARCH effects, given testing and models estimates).

Likewise, of special interest, we tend to find that positive shocks, i.e., positive news, have a higher effect on volatility, than bad ones, when the currency trend is appreciation and/or stability, while the reverse happens when the odds are of depreciation. These set of results are validated by the sign tests and illustrated in the computed news impact curves. Finally, modelling choice drives results in different ways, with a noteworthy outcome applying to the stability period, 2012, and depreciation one, 2010, where increasing the power of the APGARCH choice – recall that power 1 is TGARCH, 2 is GJR-GARCH – changes the significant sign effect from positive to negative. In sum, when news comes along the trend in exchange rate behavior, risk increases by more than otherwise (likely before the fear of over-reaction, e.g. overvaluation in the face of a Dutch disease), and nonlinearity and memory matter.

All in all, the current study offers key results, useful for policy decision making and with possibly new theoretical underpinnings. Namely, renewed attention should be given to modelling, strengthening the emphasis on selection, diagnostic and need for robustness analysis, along with due care as to the currency trend, which appears relevant for conclusions drawn. Finally, in the policy design, memory and nonlinearity matter and the trend behavior of exchange rate is key for assessing the effects of news on exchange rate risk, is that not always good news are good.

The findings presented here are inconclusive because for more consistent and robust claims it was necessary to compare or do other studies related to Mozambique or countries with similar economies.

Vector Autoregressive Models

3.1 Introduction

Forecasting and measurement of policy effects are key issues in the analysis of macroe-
conomic and financial data. The pursuit of these aims requires the specification and
estimation of models that represent relationships among variables of interest.

In making choices between alternative courses of action, decision makers often
need predictions of economic variables. If time series observations are available for a
variable of interest and the data from the past contain information about the future
development of a variable, it is plausible to use as forecast based on current and
past data. According to Lütkepohl [77], formally, this approach to forecasting is the
major goal of univariate time series analysis, where models of Box and Jenkins [22]
are widely used. However, with economic variables, often the value of one variable is
not only related to its past in time but, it depends on past values of other variables.

Taking into account the relationship between variables, in the literature there
is the concept of simultaneous equations, where the system of equations can be
estimate by indirect least squares, two-stage least squares or instrumental variables,
etc. The common on the most approaches is to separate the system variables in
two groups: the exogenous variables that are not affected by other variables in the
system and endogenous variable that are explained by a model.

Taking into account the relationship between variables Sims [109], criticized the
simultaneous equation estimation procedure where restrictions needed for identifi-
cation and classification of exogenous and endogenous variables in the system and
suggest the Vector Autoregressive (VAR) model for analysis of macroeconomic time

series. Following Stock and Watson [114], these new procedure is a simple technique that allows the systematic capture of dynamics between the various series that are analysed. Following Enders [38], the essential in the VAR model is to treat all variables simultaneously without taking into account the concept of independence or dependence. Thus, in the VAR model all variables are treated as endogenous. In the literature the VAR class models and their applications in macroeconomic and finance, can be see in Enders [38], Hamilton [58], Lütkepohl [77], Pfaff [98], Tsay [116], among others.

Following Engle and Granger [42] and Granger [55], in situations where several variables are driven by common stochastic trend, they have a strong link that may also be of interest from an economic point of view, and that variables are called cointegrating. If cointegrating relations are present in a system of variables, the vector autoregressive form is not the most convenient model setup. In that case it is useful to consider specific parametrizations that support the analysis of the cointegration structure, which are known as vector error correction (VEC) model. According to Lütkepohl [77], the main uses of vector autoregressive models are forecasting and structural analysis. Also follows Stock and Watson [114], the four tasks of multivariate analysis of macroeconomic data are: describe and summarize macroeconomic data, make macroeconomic forecast, quantify what we do or do not know about the true structure of the macro-economy, and advise macroeconomic policy markets.

Recent applications of the VAR models are in authors such as Maysami and Koh [82], where applied VEC model to study the Singapore stock market, and conclude that, the Singapore stock market is interest and exchange rate sensitive, in addition their found that the Singapore stock market is significantly and positively cointegrated with stock markets of Japan and the United States, Chaudhry and Bukhari [29], use the structural VAR model to analysis the impact of macroeconomic shock on Pakistan's textile exports, and their found that, positive shocks in the textile exports of competitor countries lead to temporary decreases followed by eventual increases. The common in most application of vector autoregressive model, is the presence of cointegrating equations which suggests the long-run relationship among the variables in the system.

For this chapter 3, after the introduction we present the review of VAR models in different specifications from the basic VAR, VEC, SVAR to SVEC models. Also, we discuss some specific tests used in the VAR class models as well as the estimation procedures are presented.

3.2 Vector autoregressive process

Let $\mathbf{y}_t = (y_{1t}, ..., y_{kt})'$ denote an $K \times 1$ vector of K stationary time series variables collected in a given period of time t. According to Sims [109], the Vector autoregressive of order p, VAR(p) model, in reduced form can be described as:

$$\mathbf{y}_t = \mathbf{\Phi}_0 + \sum_{i=1}^{p} \mathbf{\Phi}_i \mathbf{y}_{t-i} + \mathbf{u}_t \quad \text{and} \quad \mathbf{u}_t \sim N(0, \mathbf{\Sigma}_u) \tag{3.1}$$

where $\mathbf{\Phi}_0$ is a $K \times 1$ vector of constants, $\mathbf{\Phi}_i$ are $K \times K$ matrices of coefficients, and $\mathbf{u}_t$ is a $K \times 1$ vector white noise process, with the properties: $E(\mathbf{u}_t) = 0$ for all t, $E(\mathbf{u_t u_t'}) = \mathbf{\Sigma}_u$ and $\mathbf{E(u_t u_s')} = \mathbf{0}$ for $t \neq s$. Where the covariance matrix $\mathbf{\Sigma}_u$ is assumed to be positive definite, and $\mathbf{u}_t$ are serially uncorrelated but may be contemporaneously correlated which indicates instantaneous relations between the endogenous variables.

For $p = 1$, the VAR(1) equation is $\mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{u}_t$. For $K=3$, the three-variate VAR(1) model can be written as:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} \phi_{01} \\ \phi_{02} \\ \phi_{03} \end{pmatrix} + \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix}$$

or equivalently, can be written the system in standard form:

$$y_{1t} = \phi_{01} + \phi_{11} y_{1t-1} + \phi_{12} y_{2t-1} + \phi_{13} y_{3t-1} + u_{1t}$$
$$y_{2t} = \phi_{02} + \phi_{21} y_{1t-1} + \phi_{22} y_{2t-1} + \phi_{23} y_{3t-1} + u_{2t}$$
$$y_{3t} = \phi_{03} + \phi_{31} y_{1t-1} + \phi_{32} y_{2t-1} + \phi_{33} y_{3t-1} + u_{3t} \tag{3.2}$$

In addition to the assumption of stationary y_{1t}, y_{2t} and y_{3t}, the residuals u_{1t}, u_{2t} and u_{3t} are uncorrelated white noise error terms, for example $cov(u_{1t}, u_{2t}) = \sigma_{12}$ and 0 otherwise. Notice that each variable is expressed as a linear combination of lagged values of itself and lagged of all other variables in the system.

The coefficients in the system 3.2 can be interpreted as follows: the ϕ_{12} is the link of the variables y_{2t-1} in the variable y_{1t} with presence of y_{1t-1} and y_{3t-1}. Other parameters are interpreted in the same way. If $\phi_{12} = 0$, it means y_{1t} does not depend on y_{2t-1}, and so on.

3.2.1 Reduced and structural forms

For multivariate time series, there are different ways to represent VAR models, here we present the reduced and structural forms. Following Enders [38], in the standard

form 3.2, the concurrent relationship between the series is shown by the off-diagonal elements σ_{ij} of the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{u}_t$. If $\sigma_{ij} = 0$, then there is no concurrent linear relationship between the series. The VAR (p) model in standard form is also called VAR in reduced form 3.1, because it does not show explicitly the concurrent relationship between the series.

According to Stock and Watson [114], the vector autoregressive is in **reduced form** if each variables is expressed as a linear function of its own past and past values of all other variables being considered and a serially uncorrelated error term.

Because in the VAR models all the variables are endogenous, if we multiply the equation 3.1 by $\mathbf{A}$ and written in matrix form we obtain the following equation.

$$\mathbf{A}\mathbf{y}_t = \boldsymbol{\Phi}_0^* + \sum_{i=1}^{p} \boldsymbol{\Phi}_i^* \mathbf{y}_{t-i} + \varepsilon_t \tag{3.3}$$

This is the **structural form**, where the matrix $\mathbf{A}$ reflects the instantaneous relations between the variables in the model, $\boldsymbol{\Phi}_0^* = \mathbf{A}\boldsymbol{\Phi}_0$ is an vector of constants, $\boldsymbol{\Phi}_i^* = \mathbf{A}\boldsymbol{\Phi}_i$ are matrices of parameters and $\varepsilon_t = \mathbf{A}\mathbf{u}_t$ is unobservable white noise vector process with covariance matrices $\boldsymbol{\Sigma}_\varepsilon = \mathbf{A}\boldsymbol{\Sigma}_u \mathbf{A}'$, that is $\varepsilon_t \sim N(0, \boldsymbol{\Sigma}_\varepsilon)$.

The matrix $\mathbf{A}$ usually has ones on its main diagonal, so that the set of equations in 3.2 can be written such that each of the variables appears on the left-hand side of the equations and may depend on contemporaneous values of some or all of the other variables.

According to Kilian [67] and Luetkepohl [76], the parameters of the structural form are not identified without further restrictions. Imposing restrictions on $\mathbf{A}$ and $\boldsymbol{\Sigma}_\varepsilon$ to identify the structural form is a main focus of the structural VAR analysis. Often zero restrictions are placed on some variables to not allowed an instantaneous impact on some other variables. For example, $\mathbf{A}$ may be lower-triangular if there is a recursive relation between the variables. Therefore, the reduced form model is commonly used for two reasons: The first reason is easy in estimation, the second and main reason is that the concurrent correlation cannot be used in forecasting.

3.2.2   Inconvertibility of moving average representation

According to Baltagi [5] and Lütkepohl [77], for VAR(p) process 3.1, under the stability condition can be take the representation.

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i \mathbf{u}_{t-i} \tag{3.4}$$

This form of representation is called the moving average (MA) representation, where $\mathbf{y}_t$ is expressed in terms of past and present innovations vectors $\mathbf{u}_t$ and the mean term $\boldsymbol{\mu}$.

Using the lag operator $\mathbf{L}$, we can consider the structural moving average (SMA) representation for $\mathbf{y}_t$:

$$\mathbf{y}_t = \boldsymbol{\Psi}(\mathbf{L})\mathbf{u}_t \tag{3.5}$$

where $\mathbf{y}_t$ is an $K \times 1$ vector of economic variables, $\mathbf{u}_t$ is an $K \times 1$ vector of residuals and $\boldsymbol{\Psi}(\mathbf{L}) = \boldsymbol{\Psi}_0 + \boldsymbol{\Psi}_1\mathbf{L} + ...$, where $\boldsymbol{\Psi}_i$ is $K \times K$ and we will write $\mathbf{C}_{ij} = |c_{ij}, K|$. The equation 3.5 is called SMA model, since the elements of $\mathbf{u}_t$ are given a structural economic interpretation. Note that the Structural VAR (SVAR) representation is obtained by inverting $\boldsymbol{\Psi}(\mathbf{L})$ from equation 3.5, that is:

$$\boldsymbol{\Psi}(\mathbf{L})^{-1}\mathbf{y}_t = \boldsymbol{\Phi}(\mathbf{L})\mathbf{y}_t = \mathbf{u}_t \tag{3.6}$$

where $\boldsymbol{\Phi}(\mathbf{L}) = \boldsymbol{\Phi}_0 - \sum_{i=1}^{\infty} \boldsymbol{\Phi}_i(\mathbf{L}^i)$, is one sided matrix lag polynomial. In equation 3.6, the exogenous shocks $\mathbf{u}_t$ are written as a distributed lag of current and lagged values of $\mathbf{y}_t$. However, it is not always possible to invert $\boldsymbol{\Psi}(\mathbf{L})$ and move from the SMA representation 3.5 to the SVAR representation 3.6.

Following Granger and Anderson [56], one useful way to discuss the inconvertibility problem is in terms of estimates of $\mathbf{u}_t$ constructed from SVAR using truncated version of $\boldsymbol{\Phi}(\mathbf{L})$, because $\{y_t\}_{t=-\infty}^{T}$ realizations is never available. Consider the estimator $\widehat{\mathbf{u}}_t = \sum_{s=0}^{t-1} \boldsymbol{\Phi}_s\mathbf{y}_{t-s}$ constructed from the truncated realization. If $\widehat{\mathbf{u}}_t$ converges to $\mathbf{u}_t$ in mean square as $t \to \infty$, then the SMA process is said to be invertible and the structural errors can be recovered as one-sided MA of the observed data, at least in large samples.

3.3 Estimation of VAR parameters

The Vector Autoregressive model is a general framework used to describe the dynamic interrelationship between stationary variables. So, in addition to the univariate stationary test discussed in subsection 2.1.1, the first step in VAR analysis should be to determine whether the data are multivariate stationary.

Let $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{kt})'$ denote $K \times 1$ vector of K time series obtained in a given period of time t, where, each variable is a stationary and stable time series.

A multivariate time series $\mathbf{y}_t$, is covariance stationary and ergodic if all time series are stationary and ergodic. That is $E(\mathbf{y}_t) = \boldsymbol{\mu} = (\mu_1, ..., \mu_k)'$, where $\mu_i = E(y_i)$ for $i = 1, ..., k$ and the variance covariance matrix of $\mathbf{y}_t$ is $var(\mathbf{y}_t) = E[(\mathbf{y}_t - \mu)(\mathbf{y}_t - \mu)'] = \boldsymbol{\Gamma}_0$. The matrix $\boldsymbol{\Gamma}_0$ has elements $\gamma_{jt}^0 = cov(y_{it}, y_{jt})$ and the correlation matrix $\mathbf{R}_0$ of $\mathbf{y}_t$ is the $K \times K$ matrix: $corr(\mathbf{y}_t) = \mathbf{D}^{-1}\boldsymbol{\Gamma}_0\mathbf{D}^{-1}$ with $\mathbf{D} = diag(\sigma_{11}, ..., \sigma_{kk})$ the $K \times K$ diagonal matrix with the standard deviation of $y_{ii,t}$ along the diagonal.

The parameters $\mathbf{\Gamma_0}$ and $\mathbf{R_0}$ are estimated from data using the sample moments. $\mathbf{\Gamma_0} = \frac{\mathbf{1}}{\mathbf{T}} \sum_{\mathbf{t=1}}^{\mathbf{T}} (\mathbf{y_t} - \boldsymbol{\mu})(\mathbf{y_t} - \boldsymbol{\mu})' ; \mathbf{R_0} = \overline{\mathbf{D}}^{-1} \overline{\mathbf{\Gamma}}_0 \overline{\mathbf{D}}^{-1} = corr(\mathbf{y_t}), \boldsymbol{\mu} = (\mu_1, ..., \mu_k)'$ where $\mu_i = \frac{1}{T} \sum_{i=1}^{T} y_i$.

3.3.1 Lag length selection for VAR estimation

Before estimation the VAR parameters, is necessary to choose an appropriate lag length order p of VAR. In the literature, there are two popular methods: the Likelihood ratio statistic and using the information criteria.

Likelihood ratio statistic. On Likelihood Ratio (LR) test, the model with a smaller number of lags is treated as a restricted version of a larger dimensional model. Since the two models are nested, under the null hypothesis the restricted model is correct, differences in the likelihood should be small. Let $R(\alpha) = 0$ be a set of restrictions and $\mathrm{L}(\alpha, \Sigma_u)$ the likelihood function, then: $LR = T(\ln|\Sigma_u^{re}| - \ln|\Sigma_u^{un}|)$ converges to $\chi^2(v)$. Where v is the degrees of freedom equal to the difference of dimensions in the two models.

According to Lütkepohl [77], there are four important features of LR test that need to be highlight. First, a test is valid when the set of time series in $\mathbf{y}_t$ are stationary and ergodic, and if the residuals are white noise. Second, it can be computed without explicit distribution assumption on the $y_t's$. What is required is that the residual is a sequence of independent and identically distributed (IID) white noise with bounded fourth moment and that T is sufficiently large, in which case α^{un}, Σ_u^{un}, α^{re}, Σ_u^{re} are pseudo - maximum likelihood estimators. Third, a test is biased in small samples. Hence it is common to use $LR^c = (T - pK)(\ln|\Sigma_u^{re}| - \ln|\Sigma_u^{un}|)$ where pK is the number of estimate parameters in each equation of the unrestricted system. Finally, the LR test is only asymptotically valid.

In practice, an estimate of p can be obtained sequentially, as the next algorithm shows:

1. Choose an upper bound $\overline{p}$;
2. Test VAR($\overline{p} - 1$) against VAR($\overline{p}$) using a LR test. If the null hypothesis is not rejected;
3. Test VAR($\overline{p} - 2$) against VAR($\overline{p} - 1$) using an LR test, and so on, until rejection.

Clearly, $\overline{p}$ depends on the frequency of data. The typical choices are 3, 8 and 18 for annual, quarterly and monthly data respectively.

Using information criteria. According to Zivot and Wang [123], the general approach in using the information criteria is to fit VAR(p) models with orders $p = 0, ..., p_{max}$ and choose the value of p which minimizes some model selection criteria.

The model selection criteria for VAR(p) have the form $IC(p) = \ln|\tilde{\Sigma}(p)| + c_T * \varphi(p, K)$. where $\tilde{\Sigma}(p) = T^{-1}\Sigma_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t'$ is the residual covariance matrix of order p, c_T is a sequence indexed by the sample size T, and $\varphi(K,p)$ is a function of the p which penalizes large VAR(p) models.

The three most common information criteria used are: Akaike Information Criteria (AIC), Schwartz-Bayesian Information Criteria (SBIC), and Hannan-Quinn Information Criteria (HQIC).

$$
\begin{aligned}
AIC &= \ln|\tilde{\Sigma}(p)| + \frac{2}{T}(pK^2) \quad \text{with} \quad c_T = 2/T \\
HQIC &= \ln|\tilde{\Sigma}(p)| + \frac{2\ln(\ln(T))}{T}(pK^2) \quad \text{with} \quad c_T = 2\ln(\ln(T)/T \\
SBIC &= \ln|\tilde{\Sigma}(p)| + \frac{2\ln(T)}{T}(pK^2) \quad \text{with} \quad c_T = 2\ln(T)/T
\end{aligned}
\tag{3.7}
$$

where pK^2 is the freely estimated parameters, K is the number of variables in VAR model, $\tilde{\Sigma}(p)$ the residual variance-covariance matrix.

According to Lütkepohl [77], Kirchgässner and Wolters [69], another criteria order selection is called Final Prediction Errors (FPE), which is based on the minimization of the forecast mean squared error.

$$
FPE = \left(\frac{T + pK + 1}{T - pK - 1}\right)^K \times |\tilde{\Sigma}(p)|
\tag{3.8}
$$

The criteria differ in penalization according to inclusion of additional parameters in the model. The penalization is such that HQIC as well as SBIC consistently determine the order of the true maximal lag in finite sample, while the FPE and the AIC tend to overestimate it. This is summarized in the following relations:

1) $p(SBIC) \le p(HQIC)$;

2) $p(SBIC) \le p(AIC)$ for $T \ge 8$ and

3) $p(HQIC) \le p(AIC)$ for $T \ge 16$.

We notify that, the AIC and FPE gives the same lag length order in small sample.

3.3.2 Methods of estimation VAR parameters

As long as the VAR order p is selected, estimation of the parameters of the model requires that the variables in the set $\mathbf{y}_t = (y_{1t}, ..., y_{kt})'$ are covariance stationary, with their first two moments finite and time invariant. If the variables are not covariance stationary, but their first differences are, they may be modelled with a vector error correction model. In order to find the important inter-relationship among the variables, the OLS or MLE are two approaches used for estimation the Vector autoregressive models.

Ordinary least squares estimation

Lets consider the basic model 3.1, where we assumes that the VAR(p) is variance - covariance stationary and there are no restrictions on the parameters of the model. In Seemingly Unrelated Regression (SUR) system, each equation in the VAR(p) may be written as $\mathbf{y}_i = \mathbf{x}\mathbf{\Phi}_i + \mathbf{u}_i$ for $i = 1, ..., p$, where $\mathbf{y}_i$ is a $(T \times 1)$ vector of observations on the ith equation, $\mathbf{x}$ is a $(T \times K)$ matrix with t^{th} row given by $\mathbf{x}' = (1, y'_{t-1}, ..., y'_{t-p})$, $\mathbf{\Phi}_i$ is a $(K \times 1)$ vector of parameters and $\mathbf{u}_i$ is a $(T \times 1)$ error with covariance matrix $\sigma_i^2 \mathbf{I}_T$.

Given the equation $\mathbf{y}_t = \mathbf{x}\mathbf{\Phi}_i + \mathbf{u}_t$ multiplying by $\mathbf{x}'_t$ and taking expectation we get the equality $E(\mathbf{y_t x'_t}) = \mathbf{\Phi}_i E(\mathbf{x_t x'_t}) + E(\mathbf{u_t x'_t})$. Under the assumption of orthogonality condition $E(\mathbf{x}' \mathbf{u}_t) = 0$ and the matrix $E(\mathbf{x}'_t \mathbf{x}_t) = \Sigma_{t=1}^p \mathbf{x}_t \mathbf{x}'_t$ non-singular, the parameters $\mathbf{\Phi}_i$ can be estimated efficiently by OLS equation by equation.

$$\widehat{\mathbf{\Phi}}_{OLS} = \Big[\sum_{t=1}^T \mathbf{y}_t \mathbf{x}'_t \Big] \Big[\sum_{t=1}^T \mathbf{x}_t \mathbf{x}'_t \Big]^{-1} \tag{3.9}$$

Under standard assumption regarding the behavior of stationary and ergodic VAR models, the vector of coefficients $vec(\widehat{\mathbf{\Phi}})$ is consistent and asymptotically normally distributed with asymptotic covariance matrix $\widehat{avar}(vec(\widehat{\mathbf{\Phi}})) = \widehat{\Sigma} \otimes (\mathbf{x}'\mathbf{x})^{-1}$; where $\widehat{\Sigma} = \dfrac{1}{T-k} \Sigma_{t=1}^T \widehat{\mathbf{u}}_t \widehat{\mathbf{u}}'_t$, the symbol $\otimes$ denote Kronecker multiplication and $\widehat{\mathbf{u}}_t = \mathbf{y}_t - \widehat{\mathbf{\Phi}}'\mathbf{x}$ is the multivariate least squares residual at time t. In these context, if there are no restrictions on the VAR, OLS estimation of the parameters, equation by equation is consistent and efficient.

Maximum likelihood estimation

Writing equation 3.1 in a compact form $\mathbf{y}_t = \mathbf{\Phi}\mathbf{x}_t + \mathbf{u}_t$, where $\mathbf{\Phi} = (\mathbf{\Phi_0}, ..., \mathbf{\Phi_p})$ is an $(K \times (pK+1))$ matrix of parameters of the ith equation in VAR, $\mathbf{x}_t = [1, y_{t-1}, ..., y_{t-p}]'$ is an $(pK+1) \times 1$ vector and $\mathbf{u}_t$ is $(K \times 1)$ vector of residuals.

If the residuals are Normal distributed $\mathbf{u}_t \sim N(0, \mathbf{\Sigma}_u)$, the conditional density is multivariate Normal distributed, $\mathbf{y}_t | \mathbf{y}_{t-1}, ..., \mathbf{y}_{t-p} \sim N(\mathbf{\Phi}\mathbf{x}_t, \mathbf{\Sigma}_u)$ and the conditional density of the tth observation is

$$f(\mathbf{y}_t | \mathbf{y}_{t-1}, ..., \mathbf{y}_{t-p}, \Phi) = (2\pi)^{-(\frac{K}{2})} |\mathbf{\Sigma}_u^{-1}|^{1/2} \exp\Big[(-\frac{1}{2}(\mathbf{u}'_t \mathbf{\Sigma}_u^{-1} \mathbf{u}_t))\Big] \tag{3.10}$$

The likelihood function is the product of these densities for $t = 1, ..., T$.

The log – likelihood function is the sum of the log of all these densities

$$\mathrm{Log} L(\Phi, \Sigma_u) = -\frac{TK}{2}\log(2\pi) + \frac{T}{2}\log|\Sigma_u^{-1}| - \frac{1}{2}\sum_{t=1}^T \Big[(\mathbf{y}_t - \Phi\mathbf{x}_t)'\Sigma_u^{-1}(\mathbf{y}_t - \Phi\mathbf{x}_t)\Big] \tag{3.11}$$

The value of $\boldsymbol{\Phi}$ that maximize the log - likelihood is the maximum likelihood estimator of the vector autoregressive coefficients.

$$\widehat{\boldsymbol{\Phi}}_{MLE} = \left[\sum_{t=1}^{T} \mathbf{y}_t \mathbf{x}_t'\right] \left[\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right]^{-1} \tag{3.12}$$

which is just the estimated equation 3.9 coefficients by an OLS regression of the $\mathbf{y}_t$ on $\mathbf{x}_t$. Thus, maximum likelihood estimation of the coefficients for the ith equation of VAR are found by an OLS regression. This means that the ML estimator of the VAR coefficients is equivalent to apply OLS for each equation of the VAR separately. The ML estimator for the variance is $\widehat{\boldsymbol{\Sigma}}_u = \frac{1}{T} \sum_{t=1}^{T} \widehat{\mathbf{u}}_t \widehat{\mathbf{u}}_t'$. Note that the maximum likelihood estimator for the variance is consistent, but is biased in small samples, so it is common to use the variance estimator adjusted by the number of degrees of freedom, $\tilde{\boldsymbol{\Sigma}}_u = \frac{T}{T - pK - 1} \widehat{\boldsymbol{\Sigma}}_u$. These maximum likelihood estimation is necessary and useful for SVAR analysis.

In the estimation process it's important to test if all coefficients are statistically significant for the model. The significance for each coefficient can be done by computing the t-statistic. If the value of t-statistic is greater in absolute value than the critical value (or p-value $< \alpha$), we reject the null hypothesis that the coefficient is not significant.

The significance of all regressor excluding the intercept is tested by F - statistic. Under the null hypothesis that all slope coefficients are zero against that, at least one coefficient is different from zero. We reject the null if observed F-statistic is greater than the critical value, $F_\alpha(K, T - K - 1)$.

3.3.3 Checking the model adequacy

Once a VAR(p) model has been estimated, it is of interest to see whether the residuals are compared with model assumption. That is, one should check for stability of the VAR process, the presence of serial correlation, heteroskedasticity and if the error process is Normally distributed.

Stability condition

Using the lag operator and defined the matrix polynomial in the lag operator $\boldsymbol{\Phi}(\mathbf{L})$ as $\boldsymbol{\Phi}(\mathbf{L}) = \mathbf{I}_k - \boldsymbol{\Phi}_1 \mathbf{L} - \boldsymbol{\Phi}_2 \mathbf{L} = ... = \boldsymbol{\Phi}_p \mathbf{L}^p$, the process 3.1 can be equivalently written as $\boldsymbol{\Phi}(\mathbf{L}) \mathbf{y}_t = \mathbf{u}_t$.

The VAR process is stable if and only if all included variables are stationary, i.e, all roots of the characteristic equation of the lag polynomial are outside the unit

47

circle.

$$\det(\boldsymbol{\Phi}(\mathbf{z})) = \det(\mathbf{I}_k - \boldsymbol{\Phi}_1\mathbf{z} - \boldsymbol{\Phi}_2\mathbf{z}^2 - ... - \boldsymbol{\Phi}_p\mathbf{z}^p) \neq 0 \quad \text{for} \quad |\mathbf{z}| \leq 1 \qquad (3.13)$$

So the necessary and sufficient condition for VAR stability is that all characteristic roots of $\boldsymbol{\Phi}(\mathbf{z}) = 0$ lie outside the unit circle.

Multivariate test for autocorrelation

According to Lütkepohl [77], autocorrelation of the residuals indicates that there is information that has not been accounted for in the model. For testing the lack of serial correlation in the residuals of a VAR(p) model, a Portmanteau test and Lagrange Multiplier (LM) test proposed by Breusch [25] and Godfrey [53] are most commonly applied.

The multivariate Portmanteau test for autocorrelation in a set of residuals is a generalization of the univariate Ljung - Box Portmanteau test for white noise discussed in subsection 2.1.2. The test was proposed by Hosking [61], and it's applied to the residuals of a multivariate regression, such as VAR model with the following null hypotheses, H_0: no residual autocorrelation up to lag h. As a function of h-lags, the test statistic is.

$$MLB(h) = T(T+2)\sum_{i=1}^{h}(T-i)^{-1}tr\left(\widehat{\mathbf{C}}_{\mathbf{0i}}\widehat{\mathbf{C}}_{\mathbf{00}}^{-1}\widehat{\mathbf{C}}_{\mathbf{0i}}'\widehat{\mathbf{C}}_{\mathbf{00}}^{-1}\right) \qquad (3.14)$$

where $\widehat{\mathbf{C}}_{0i} = T^{-1}\sum_{t=i+1}^{T}\widehat{\mathbf{u}}_t\widehat{\mathbf{u}}_{t-i}'$ is the sample auto-covariance matrix of order i. Under the null hypothesis, the multivariate Ljung-Box test $MLB(h)$ is distributed asymptotically as $\chi^2(K^2(h-p))$ where K is the number of time series, p the number of lags in the model estimation and h the number of lags chosen after lag p. Rejection the null indicates that at least one series have autocorrelation in the residuals that is, one series is not white noise.

The Breusch-Godfrey LM-statistic is based upon the following auxiliary regression for residuals: $\widehat{\mathbf{u}}_t = \boldsymbol{\Phi}_1\mathbf{y}_{t-1} + ... + \boldsymbol{\Phi}_p\mathbf{y}_{t-p} + \mathbf{B}_1\widehat{\mathbf{u}}_{t-1} + ... + \mathbf{B}_h\widehat{\mathbf{u}}_{t-h} + \varepsilon_t$. The null hypothesis is $H_0 : \mathbf{B}_1 = ... = \mathbf{B}_h = 0$, against the alternative $H_1 : \exists\mathbf{B}_i \neq 0$ for $i = 1, 2, ..., h$. The test statistic is defined as:

$$LM_h = T\left(K - tr(\tilde{\Sigma}_{re}^{-1} - \tilde{\Sigma}_{un})\right) \qquad (3.15)$$

where $\tilde{\Sigma}_{re}^{-1}$ and $\tilde{\Sigma}_{un}$ are the residual covariance matrix of the restricted and unrestricted models, respectively. The test statistic LM_h is distributed as $\chi^2(hK^2)$. In practice, we can use MLB or LM statistic to check the adequacy of a fitted model.

Multivariate normality test

Testing for normality of log-returns is a common procedure in much applied work and many tests have been proposed. In the literature, there are different procedures for multivariate normality tests, some of them are discussed in Mulenga *et al* [88]. Here we present the procedure suggested by Mardia [80], and by Doornik and Hansen [36], based on multivariate Skewness and Kurtosis tests.

Let K independent variables and $\mathbf{y}_i = (y_1, y_2, ..., y_k)'$ for $i = 1, ..., T$, where $\mathbf{y}$ denote the $T \times K$ matrix of observations. In order to test whether the set of K-variables are multivariate normal distributed with mean $\boldsymbol{\mu} = (\mu_1, ..., \mu_k)'$ and finite covariance $\Sigma > 0$, $MVN(\mu, \Sigma)$. Mardia [80], define the sample measures of multivariate Skewness $b_{M,1}$ and Kurtosis $b_{M,2}$, as:

$$b_{M,1} = \frac{1}{T^2} \sum_{i=1}^{T} \sum_{j=1}^{T} [(\mathbf{y}_i - \overline{\mathbf{y}})' S^{-1} (\mathbf{y}_j - \overline{\mathbf{y}})]^3; \quad b_{M,2} = \frac{1}{T} \sum_{i=1}^{T} [(\mathbf{y}_i - \overline{\mathbf{y}})' S^{-1} (\mathbf{y}_i - \overline{\mathbf{y}})]^2 \quad (3.16)$$

where $\overline{\mathbf{y}}$ and $\mathbf{S}$ be the sample mean vector and the covariance matrix as follows. $\overline{\mathbf{y}} = \frac{1}{T} \sum_{i=1}^{T} \mathbf{y}_i$ and $\mathbf{S} = \frac{1}{T} \sum_{i=1}^{T} (\mathbf{y}_j - \overline{\mathbf{y}})' (\mathbf{y}_i - \overline{\mathbf{y}})$.

The multivariate normality when the population is $N(\mu, \sigma)$, using Mardia measures of multivariate Skewness and Kurtosis is given by:

$$MJB_M = T \left\{ \frac{b_{M,1}}{6} + \frac{(b_{M,2} - K(K+2))^2}{8K(K+2)} \right\} \quad (3.17)$$

The Mardia MJB statistic is asymptotically distributed as $\chi^2_\alpha(K(K+1)(k+2)/6)$ distribution, where K is the number of variables.

An alternative of Jarque-Bera test can be found in Doornik and Hansen [36], Lütkepohl [77] and Srivastava [111]. The Jarque-Bera normality tests for univariate and multivariate series are implemented to the residual of a VAR(p) as well as separate tests for Skewness and Kurtosis. The univariate test is applied to the residual of each equation, while the multivariate version can be computed by using the residuals that are standardized by a Choleski decomposition of the variance-covariance matrix. Note that, the result is dependent upon the ordering of the variables.

Using notation presented in Lütkepohl [77], for orthogonalized VAR residuals, if we call $\widehat{\lambda}_1$ the multivariate skewness, $\widehat{\lambda}_2$ the multivariate kurtosis, the multivariate Jarque - Bera statistics $\widehat{\lambda}_3$ is defined as:

$$\widehat{\lambda}_3 = \widehat{\lambda}_1 + \widehat{\lambda}_2 \quad (3.18)$$

Under the null hypothesis of multivariate Gaussian disturbance, $\widehat{\lambda}_1 = \frac{T \widehat{\mathbf{b}}_1' \widehat{\mathbf{b}}_1}{6} \sim \chi^2(K)$ and $\widehat{\lambda}_2 = \frac{T(\widehat{\mathbf{b}}_2 - 3)'(\widehat{\mathbf{b}}_2 - 3)}{24} \sim \chi^2(K)$, observe that $\mathbf{b}_1$ and $\mathbf{b}_2$ are $K \times 1$

49

vectors of skewness and kurtosis coefficients, that is third and fourth non-central moment vectors of standardized residuals.

$$\widehat{\mathbf{b}}_1 = (\widehat{b}_{11},...,\widehat{b}_{k1})' \qquad ; \qquad \widehat{\mathbf{b}}_{k1} = \frac{1}{T}\sum_{i=1}^{T}\omega_{kt}^3$$

$$\widehat{\mathbf{b}}_2 = (\widehat{b}_{12},...,\widehat{b}_{k2})' \qquad ; \qquad \widehat{\mathbf{b}}_{k2} = \frac{1}{T}\sum_{i=1}^{T}\omega_{kt}^4 \qquad (3.19)$$

where ω_{kt} is orthogonalized VAR residual and the $\widehat{\lambda}_3$ statistic is distributed as $\chi^2(2K)$. The corresponding statistics against the null hypothesis that the residuals from the kth equation come from a univariate normal distribution are.

$$\widehat{\lambda}_{1k} = \frac{T\widehat{b}_{k1}^2}{6} \sim \chi^2(1); \quad \widehat{\lambda}_{2k} = \frac{T(\widehat{b}_{k2}^2 - 3)}{24} \sim \chi^2(1); \quad \widehat{\lambda}_{3k} = \widehat{\lambda}_{1k} + \widehat{\lambda}_{2k} \sim \chi^2(2) \quad (3.20)$$

Note that, this procedure the second is most used in statistical packages such as STATA and EViews.

Since the VAR model is verified that it is adequate to describe the dynamic in the system formed by the variables, follows other purposes of multivariate analysis of time series such as Forecast in-sample or out-of-sample, the Granger causality test, impulse response functions to shocks and forecast error variance decomposition.

3.3.4 Granger Causality

According to Lütkepohl [77], VAR models represent the correlations among a set of variables, they can be used to investigate relationships between the variables of interest. A specific type of relation was pointed out by Granger [54] and is known as Granger – Causality. Let us consider here two variables. By definition y_{2t} Granger - causes y_{1t} if the information of past and present values of y_{2t} help for improving the forecasts of y_{1t}. Then y_{2t} is not Granger – causes y_{1t} if and only if $\mathbf{\Phi}_{12} = 0$, for $i = 1, 2, ..., p$. In other words, all of coefficients on lagged values of y_{2t} are zero in the equation for y_{1t} of the model.

Formally, y_{1t} fails to Granger-cause y_{2t} if for all $s > 0$ the mean squared error of forecast y_{2t+s} based on $(y_{2t}, y_{2t-1},...)$ is the same as the MSE of forecast y_{2t+s} based on $(y_{2t}, y_{2t-1},...)$ and $(y_{1t},...)$, that is $MSE[\hat{E}(y_{2t+s)}|y_{2t}, y_{2t-1},...)] = MSE[\hat{E}(y_{2t+s}|y_{2t}, y_{2t-1},...,y_{1t}, y_{1t-1},...)]$.

In general, among many versions of the Granger - causality test, three situations can be highlighted:

- The first involves two variables and the test check whether a variable influences the forecast of the other by testing the statistical significance of its coefficients in the lags.

- The second situation is that where involve more than two variables. In this case, we test the influence of each individual variable as well as joint influence of a set of independent variables on that considered as dependent.

- The third situation assumes that there is an interaction between all variables included in a system of simultaneous equations such as VAR model estimation. In this case, the test is implemented for all variables according to the number of equations.

Clearly, Granger - causality does not imply true causality. It only implies forecasting ability. The hypothesis that y_{2t} does not Granger cause y_{1t} could be tested simply by running the regression of y_{1t} on lagged values of y_{1t} and y_{2t} and examining whether the coefficient of the latter variable is significantly different from zero.

Test for Granger - Causality: Tests of whether y_{2t} Granger - causes y_{1t} in equation 3.1 or in standard form in equation 3.2 can be implemented using a standard Wald F or χ^2 test. Let's consider a particular autoregressive lag p of y_{1t} as linear function of its own past values, plus past values of y_{2t}. If y_{2t} Granger causes y_{1t}, then same or all the lagged y_{2t} coefficients values have non-zero effects. The null hypothesis of y_{2t} does not Granger causes y_{1t} in the VAR is: $H_0 : \phi_{21} = \phi_{22} = ... = \phi_{2p} = 0$, which can be tested comparing the observed statistical and critical values.

$$F = \frac{(RSS_r - RSS_u)/p}{RSS_u/(T - 2p - 1)} \quad or \quad \chi^2 = \frac{T(RSS_r - RSS_n)}{RSS_u} \tag{3.21}$$

where RSS_r and RSS_u, are the residual sum of squared from restricted and unrestricted regression, $y_{1t} = \mu_1 + \Sigma\phi_{1i}y_{1t-i} + u_{1t}$ and $y_{1t} = \mu_2 + \Sigma\phi_{1i}y_{1t-i} + \Sigma\phi_{2i}y_{2t-i} + u_{2t}$, respectively.

If the statistic F is greater than the 5% critical value for an $F(p, T - 2p - 1)$, we reject the null hypothesis that y_{2t} does not Granger cause y_{1t}. For a regression with fixed regressor and Gaussian distributions, the test statistic F has an exact F distribution, but for a regression with lagged dependent variables as in the Granger-causality regressions, the test is only asymptotically valid. An asymptotically equivalent test is given by χ^2, and we reject the null hypothesis if χ^2 statistic is greater than the 5% critical value for a $\chi^2(p)$. Similarly, the null of y_1 does not Granger - cause y_{2t} can be expressed in the VAR as. $H_0 : \phi_{11} = \phi_{12} = ... = \phi_{1p} = 0$.

An alternative to the Granger-causality test is the procedure proposed by Dolado and Lütkepohl [35]. The test consists in that, for a VAR(p) model one should estimate the VAR(p+1) system to the data and perform a Wald test on the coefficients of the first p lags only.

51

3.3.5 Forecasting in VAR models

According to Zivot and Wang [123], in addition to analysing the relationship between variables, forecasting is one of the objectives of multivariate time series analysis. Forecasting from a VAR model is similar to forecasting from a univariate AR model. That is, forecasting future value of $\mathbf{y}_t$, when the parameters $\mathbf{\Phi}_i$ of the VAR(p) are known and there are no deterministic terms of exogenous variables. The best linear predictor, in terms of minimum mean squared error (MSE), of $\mathbf{y}_{t+1}$ or 1-step forecast based on information available at time T is.

$$\mathbf{y}_{T+1|T} = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_T + ... + \mathbf{\Phi}_p \mathbf{y}_{T-p+1} \qquad (3.22)$$

Forecasts for longer horizon h (h-step) may be obtained recursively as: $\mathbf{y}_{T+h|T} = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{T+h-1|T} + ... + \mathbf{\Phi}_p \mathbf{y}_{T+h-p|T}$, where $\mathbf{y}_{T+j|T} = \mathbf{y}_{T+j}$ for $j \leq 0$. The h-step forecast errors may be expressed as $\mathbf{y}_{T+h} - \mathbf{y}_{T+h|T} = \sum_{s=0}^{h-1} \mathbf{\Psi}_s \mathbf{u}_{T+h-s}$ where the matrices $\mathbf{\Psi}_s$ are determined by recursive substitution with $\mathbf{\Psi}_0 = \mathbf{I}_k$ and $\mathbf{\Phi}_j = 0$ for $j > p$, $\mathbf{\Psi}_s = \sum_{j=1}^{p-1} \mathbf{\Psi}_{s-j} \mathbf{\Phi}_j$.

The forecast are unbiased since all of the forecast errors have expectation zero and the MSE matrix for $\mathbf{y}_{T+h|T}$ is:

$$\mathbf{\Sigma}(h) = MSE(\mathbf{y}_{T+h} - \mathbf{y}_{T+h|T}) = \sum_{s=0}^{h-1} \mathbf{\Psi}_s \mathbf{\Sigma} \mathbf{\Psi}'_s \qquad (3.23)$$

That is $\mathbf{y}_{t+h} - \mathbf{y}_{T+h|T} \sim (0, \mathbf{\Sigma}(h))$.

3.3.6 Impulse response function

From subsection 3.3.3, an VAR(p) model, under the stability condition, equation 3.13 can be written as a linear function of the past innovations, that is a vector moving average (VMA) or MA(∞) process. The coefficients in the matrices $\mathbf{\Psi}_s$ in equation 3.4, quantify the effect of a shock u_{it-s} into the variable y_{it}, that is, the elements ψ_{ij}^s of the matrix $\mathbf{\Psi}_s$ is interpreted as the impulse response. $\dfrac{\partial y_{it+s}}{\partial u_{jt}} = \dfrac{\partial y_{it}}{\partial u_{jt-s}} = \psi_{ij}^s$, for $i, j = 1, ..., T$.

The set of coefficients $\mathbf{\Psi}_{ij} = \psi_{ij}^s$ for $i, j = 1, ..., T$ are called the impulse response function (IRF). It describes the response of y_{it+s} to a one-time impulse in y_{jt} with all other variables dated t or earlier held constant. Moreover, Lütkepohl [77] argue that, the response of variable j to a unity shock in other variable k is plotted graphically to get a visual impression of the dynamic interrelationship within the system. However, this interpretation is only possible if $var(\mathbf{u}_t) = \mathbf{\Sigma}_u$, is an diagonal matrix so that the elements of $\mathbf{u}_t$ are uncorrelated. One way to make the error uncorrelated is estimate

the triangular structural VAR(p) model.

$$
\begin{aligned}
y_{1t} &= \phi_{01} + \phi'_{11}\mathbf{y}_{t-1} + ... + \phi'_{1p}\mathbf{y}_{t-p} + \varepsilon_{1t} \qquad (3.24)\\
y_{2t} &= \phi_{02} + \alpha_{21}y_{1t} + \phi'_{21}\mathbf{y}_{t-1} + ... + \phi'_{2p}\mathbf{y}_{t-p} + \varepsilon_{2t}\\
y_{3t} &= \phi_{03} + \alpha_{31}y_{1t} + \alpha_{32}y_{2t} + \phi'_{31}\mathbf{y}_{t-1} + ... + \phi'_{3p}\mathbf{y}_{t-p} + \varepsilon_{3t}\\
&\;...\\
y_{kt} &= \phi_{0k}k + \alpha_{k1}y_{1t} + ... + \alpha_{kk-1}y_{k-1t} + \phi'_{k1}\mathbf{y}_{t-1} + ... + \phi'_{kp}\mathbf{y}_{t-p} + \varepsilon_{kt}
\end{aligned}
$$

Following Zivot and Wang [123], in matrix form, the triangular structural VAR(p) model is:

$$
\mathbf{A}\mathbf{y}_t = \mathbf{\Phi}_0^* + \sum_{i=1}^{p} \mathbf{\Phi}_i^* \mathbf{y}_{t-i} + \varepsilon_t \qquad (3.25)
$$

where $\mathbf{A}$ is lower triangular matrix with 1's along the diagonal and the triangular structural model 3.25 imposes the recursive causal ordering $y_1 \rightarrow y_2 \rightarrow ... \rightarrow y_k$. These ordering means that the contemporaneous values of the variables on the left of the arrow affect the contemporaneous values of the variables to the right on the arrow but not vice-versa. These contemporaneous effects are captured by the coefficients α_{ij} in 3.25. For example the ordering $y_1 \rightarrow y_2 \rightarrow y_3$ imposes the restrictions y_{1t} affects y_{2t} and y_{3t} but y_{2t} and y_{3t} do not affect y_{1t}; y_{2t} affects y_{3t} but y_{3t} does not affect y_{2t}.

3.3.7 Forecast error variance decomposition

The forecast error variance decomposition (FEVD) gives information for a portion of the variance of the forecast error in predicting y_{it+h} which is due to variability in the structural shocks ε_{ij} between times t and $t+h$. Using the orthogonal shocks ε_t the *h-step* ahead forecast error vector, with known VAR coefficients, may be expressed as $\mathbf{y}_{t+h} - \mathbf{y}_{t+h|t} = \Sigma_{j=0}^{h-1}\mathbf{\Theta}_j\varepsilon_{t+h-j}$. For a particular variable y_{it+h}, the forecast error has the form:

$$
y_{it+h} - y_{it+h|t} = \sum_{j=0}^{h-1}\theta_{i1}^j\varepsilon_{1t+h-j} + ... + \sum_{j=0}^{h-1}\theta_{ik}\varepsilon_{kt+h-j} \qquad (3.26)
$$

Since the structural errors are orthogonal, the variance of the h-step ahead forecast error is: $var(y_{it+h} - y_{it+h|t}) = \sigma_{\varepsilon_1}^2\Sigma_{j=0}^{h-1}(\theta_{i1}^j)^2 + ... + \sigma_{\varepsilon k}^2\sum_{j=0}^{h-1}(\theta_{ik})^2$, where $\sigma_{\varepsilon j}^2 = var(\varepsilon_j)$. The portion of $var(y_{it+h} - y_{it+h|t})$ due to shock ε_j is then

$$
FEVD_{ij}(h) = \frac{\sigma_{\varepsilon j}^2\Sigma_{j=0}^{h-1}(\theta_{ij}^j)^2}{\sigma_{\varepsilon_1}^2\Sigma_{j=0}^{h-1}(\theta_{i1}^j)^2 + ... + \sigma_{\varepsilon k}^2\sum_{j=0}^{h-1}(\theta_{ik}^j)^2} \qquad (3.27)
$$

In a VAR with K variables there are K^2, $FEVD_{ij}(h)$ values which depends on the recursive causal ordering used to identify the structural shocks ε_t and is not

53

unique. It depends on the identification scheme. Thus, different causal ordering will produce different FEVD values.

Traditionally, analysis of impulse response function has been the main objective of interest in the applied SVAR, but not only a natural extension of FEVD concept is the historical decomposition of observed time series.

According to Luektepohl [76], the Historical Decomposition (HD) is another way of looking the contribution of the structural shocks to the observed series. Historical decomposition, tells what portion of the deviation of y_{it} from its unconditional mean is due to the shock ε_j.

The structural shocks push the variables away from their equilibrium values. The interpretation of historical decomposition not need for a reference value that indicates when a shock is influencing the path of the variables. In that case, the reference value is zero, and it is understood that deviations of the shocks below that value are interpreted as negative shocks and deviations above as positive shocks.

3.4 Vector error correction model

According to Johnston and DiNardo [66], when the variables in the VAR are integrated of order one or more, the nonstationary variables suggests cointegrating relations.

Economic theory often suggests that two or more economic variables should be linked more or less closely. Although relationships among the variables are usually assumed to hold only in the long-run, economic forces are expected to act in the direction of eliminating short-run deviations from these long term relationships. Situations where there are r cointegrating vectors, which describe the long-run relationships between variables, the vector error correction model is an appropriate model which is suited for capturing relations between nonstationary variables correctly and can be represent as:

$$\Delta \mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Pi} \mathbf{y}_{t-1} + \sum_{i=1}^{p} \mathbf{\Gamma}_i \Delta \mathbf{y}_{t-i} + \mathbf{u}_t \tag{3.28}$$

where the error correction component $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ and the vector autoregressive component $\mathbf{\Gamma}$ are functions of the $\mathbf{\Phi}'s$. Specifically, $\mathbf{\Gamma}_j = -\sum_{i=j+1}^{p} \mathbf{\Phi}_i$ for $j = 1,...,p-1$, $\mathbf{\Pi} = -(\mathbf{I} - \mathbf{\Phi_1} - ... - \mathbf{\Phi_p}) = -\mathbf{\Phi}(\mathbf{1})$.

Note that, still in equation 3.28, in component $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, $\boldsymbol{\alpha}$ is the speed of adjustment parameter. In addition, a large value of $\boldsymbol{\alpha}$ indicates the mechanism of adjustment has fast convergence to return long-run equilibrium in cases fo deviate from it. The parameter $\boldsymbol{\beta}$ is a matrix of long-run coefficients or cointegrating vector.

Following Johnston and DiNardo [66], the relevant procedure for analyse the relationship between variables nonstationary in level, but linked consists of three steps:

1. Determine the cointegrating rank, that is, the number of cointegrating relations.

2. Estimate the matrices of cointegrating $\boldsymbol{\beta}$ and adjustment $\boldsymbol{\alpha}$ parameters.

3. Estimate the VAR incorporating the cointegrating relations, that is, estimate VEC model.

3.4.1 Cointegration

According to Johansen [63], suppose that, there is long-run relationship between y_t and x_t variables, then a linear combination of y_t and x_t can be directly taken from estimating the following regression $y_t = \beta_0 + \beta_1 x_t + u_t$ and taking the residuals $\widehat{u}_t = y_t - \widehat{\beta}_0 - \widehat{\beta}_1 x_t$. If $\widehat{u}_t \sim I(0)$, then the variables y_t and x_t are said to be cointegrated.

In general let $\mathbf{y}_t = (y_{1t}, ..., y_{kt})'$ variables in a K dimensional process, $\mathbf{y}_t$ are cointegrated of order (d,b), briefly $\mathbf{y}_t \sim CI(d,b)$, if all components of $\mathbf{y}_t$ are $I(d)$ and there exists a linear combination $\mathbf{z}_t = \boldsymbol{\beta}'\mathbf{y}_t$, with $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)' \neq 0$ such that $\mathbf{z}_t$ is $I(d-b)$. For instance, if all components of $\mathbf{y}_t$ are I(1) and $\boldsymbol{\beta}'\mathbf{y}_t$ is stationary I(0), them $\mathbf{y}_t \sim CI(1,1)$. The cointegration rank r is equal to the number of linearly independent cointegration vectors.

Suppose that y_t and x_t are two variables integrated of order 1, that is $y_t \sim I(1)$ and $x_t \sim I(1)$, by saying y_t and x_t are CI(1,1), this means that the regression equation $y_t = \beta x_t + u_t$ makes sense because there is a long-run equilibrium relationship between them. If y_t and x_t are not cointegrated, that is $y_t - \beta x_t = u_t$ is also I(1), they can drift apart from each more and more as time goes on. Thus there is no long-run equilibrium relationship between them. In this case the relationship between y_t and x_t that we obtain by regression y_t on x_t is spurious.

According to Zivot and Wang [123], in many times series regression models is required all variable to be $I(0)$, so that the usual statistical results for the linear regression model are valid. If some or all of the variables in the regression are $I(1)$ that is, in the presence of nonstationary variables, the usual statistical may not be valid. That is, the regression has a high coefficient of determination R^2, t - statistics appear to be significant, but the results are without any economic meaning. The regression output are not consistent that is the regression coefficients are inefficient and the usual tests of statistics inference does not hold.

One important case in which the usual statistical results do not hold is spurious regression when all the regressor are $I(1)$ and not cointegrated. Using the simple equation $y_t = \alpha_0 + \alpha_1 x_t + u_t$ we can consider four situations to choose for regression.

1. If $y_t \sim I(0)$ and $x_t \sim I(0)$, then the classical regression model is appropriate.

2. If $y_t \sim I(i)$ and $x_t \sim I(j)$, $i \neq j$, then the regression are meaningless.

3. If $y_t \sim I(i)$; $x_t \sim I(i)$ and the u_t contains a stochastic trend. Regression $\Delta y_t = \alpha_1 \Delta x_t + \Delta u_t$ is appropriate. Note that, if one of the trends is deterministic and the other is stochastic, first differencing each is not appropriate.

4. If $y_t \sim I(i)$; $x_t \sim I(i)$ and u_i is stationary. Then y_t and x_t are cointegrated.

3.4.2 Testing for cointegration

Let the $(K \times 1)$ vector $\mathbf{y}_t$ be I(1) with $0 < r < K$ cointegration vectors if there exists an $(r \times K)$ matrix $\mathbf{B}'$ such that $\mathbf{B}'\mathbf{y}_t \sim I(0)$. Testing for cointegration may be thought of as testing for the existence of long-run equilibrium among the elements of $\mathbf{y}_t$. Cointegration tests cover two situations: (1) There is at most one cointegration vector and (2) there are possible r, where $0 < r < K$ cointegration vectors. Among many procedures, two approaches will be described here. The Engle and Granger [42] test related to the first situation and the Johansen [63] methodology for the second situation.

Engle and Granger's two-step procedure

Engle and Granger [42], show that if there is a cointegration vector a simple two-step residual-based testing procedure can be employed to test for cointegration. In this case, a long-run equilibrium relationship components of $\mathbf{y}_t$ can be estimated by running $y_{1t} = \beta y_{2t} + u_t$, where $y_{2t} = (y_{2t}, ..., y_{kt})'$ is an $((K-1) \times 1)$ vector.

To test the null hypothesis that the $\mathbf{y}_t$ is not cointegrated, we should test for a unit root in the residuals described in subsection 2.1.4, $u_t \sim I(1)$ vs $u_t \sim I(0)$. The hypothesis are $H_0 : u_t = \beta'y_t \sim I(1)$ (no-cointegration) and $H_1 : u_t = \beta'y_t \sim I(0)$ (cointegration). If the unit root hypothesis is rejected, the hypothesis of no-cointegration is also rejected, thus, the cointegration tests are similar to the unit root tests, the differences are founded in their critical values.

Mackinnon [79], provides tables based on simulation accurately enough for all practical purposes. The potential problems with Engle - Granger procedure is that the cointegrating vector will not involve y_{1t} component. In this case the cointegrating will not be consistently estimated from the OLS regression loading to spurious results.

Estimation of the static model $y_{1t} = \beta y_{2t} + u_t$ is equivalent to omitting the short-term component from the VEC model. If this results in autocorrelation in the residuals, although the results will still hold asymptotically, it might create a severe bias in finite samples. Because of this, it makes sense to estimate the full dynamic model. Since all variables in the VEC are I(0), the model can be consistently estimated using the OLS method, this approach leads to a better performance as it does not push the short-term dynamic into residuals.

Johansen methodology

The Johansen [63] procedure is an alternative approach to test for cointegration which allows to avoid some drawbacks for straightforward test the r unit roots and avoids normalization problem existing in the Engle-Granger approach and test the number of cointegrating relations directly. Because the Johansen procedure is based on ML estimator, that is why, we need the time series to be multivariate normal.

The Johansen test procedure builds on the vector autoregression of order p, here in a dynamic form or VEC model, equation 3.28 updated.

$$\boldsymbol{\Delta}\mathbf{y}_t = \boldsymbol{\Phi}_0 + \boldsymbol{\Pi}\mathbf{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \boldsymbol{\Delta}\mathbf{y}_{t-i} + \mathbf{u}_t \tag{3.29}$$

For $\mathbf{y}_t$ nonstationary I(1) time series components, in order to get a stationary error term $\mathbf{u}_t$, $\boldsymbol{\Pi}\mathbf{y}_{t-1}$ should also be stationary. If the VAR(p) process has unit roots then the coefficient matrix $\boldsymbol{\Pi}$ has reduced rank, $\text{rank}(\boldsymbol{\Pi}) = r < K$. In fact, testing for cointegration is equivalent to checking the rank of the matrix $\boldsymbol{\Pi}$.

1) If $\boldsymbol{\Pi}$ has a full rank then all time series in $\mathbf{y}_t$ are stationary;

2) If the rank of $\boldsymbol{\Pi} = 0$, then there are no cointegrating relationships;

3) If $0 < \text{rank}(\boldsymbol{\Pi}) = r < K$, this implies that $\mathbf{y}_t$ is $I(1)$ with r linearly independent cointegrating vectors and $m = K - r$ non-stationary vectors.

Since $\boldsymbol{\Pi}$ has rank r there exist $K \times r$ matrices with $\text{rank}(\alpha) = \text{rank}(\beta) = r$ such that $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ and $\boldsymbol{\beta}'\mathbf{y}_t$ is stationary, r is the number of cointegrating vector.

The VEC model becomes: $\boldsymbol{\Delta}\mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \boldsymbol{\Gamma}_1 \boldsymbol{\Delta}\mathbf{y}_{t-1} + ... + \boldsymbol{\Gamma}_{p-1}\boldsymbol{\Delta}\mathbf{y}_{t-p+1} + \mathbf{u}_t$, with $\boldsymbol{\beta}'\mathbf{y}_{t-1} \sim I(0)$.

Algorithm for Johansen procedure

a) Specify and estimate VAR(p) model for $\mathbf{y}_t$;

57

b) Determine $\text{rank}(\boldsymbol{\Pi})$ equal to number of cointegrating vectors; the ML estimate for $\boldsymbol{\beta}$ equals the matrix of eigenvectors corresponding to the r largest eigenvalue of $K \times K$ residual matrix;

c) If necessary, impose normalization and identifying restrictions on the cointegrating vectors;

d) Given the normalized cointegrating vectors estimate the resulting cointegrated VEC model by ML.

In general, it is known that for a given r, the ML estimator of $\boldsymbol{\beta}$ defines the combination of $\mathbf{y}_{t-1}$ that yields the r largest canonical correlations of $\boldsymbol{\Delta}\mathbf{y}_t$ with $\mathbf{y}_{t-1}$ after correcting for lagged differences and deterministic variables when present. From these idea Johansen [63], proposes two different likelihood ratio tests of the significance of these canonical correlations and there by the reduced rank of the $\boldsymbol{\Pi}$ matrix: the trace test and maximum eigenvalue test.

a) Trace statistic test. Since the rank of the long-run impact matrix $\boldsymbol{\Pi}$ gives the number of cointegrating relationships in $\mathbf{y}_t$, the Johansen likelihood ratio (LR) statistic for determining the rank of $\boldsymbol{\Pi}$ are based on the estimated eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_k$ of the matrix $\boldsymbol{\Pi}$. The null hypothesis of r cointegrating vectors against the alternative hypothesis of K cointegrating vectors, are $H_0(r_0) : r \leq r_0$ vs $H_1(r_0) : r > r_0$. The LR statistics called trace statistic is $LR_{trace}(r_0) = -T\sum_{i=r_0+1}^{k} log(1-\hat{\lambda})$.

The trace statistic checks whether the smallest $K - r_0$ eigenvalues are statistically different from zero. If $rank(\boldsymbol{\Pi}) = r_0$ then $\hat{\lambda}_{r_0+1}, ..., \hat{\lambda}_k$ should all be close to zero and $LR_{trace}(r_0)$ should be small. If $rank(\boldsymbol{\Pi}) > r_0$ then some of $\hat{\lambda}_{r_0+1}, ..., \hat{\lambda}_k$ will be nonzero, but less than 1 and $LR_{trace}(r_0)$ should be large. The asymptotic null distribution of $LR_{trace}(r_0)$ in not chi-square but instead is a multivariate version of the Dickey-Fuller unit root distribution which depends on the dimension $K - r_0$ and the specification of the deterministic terms.

b) Maximum eigenvalue statistic. The maximum eigenvalue test, tests the null hypothesis of r cointegrating vectors against the alternative hypothesis of $r + 1$ cointegrating vectors, that is $H_0(r_0) : r = r_0$ vs $H_1(r_0) : r = r_0 + 1$. The LR statistic, called maximum eigenvalue statistic is given by $LR_{max}(r_0) = -Tlog(1-\hat{\lambda}_{r_0+1})$.

The asymptotic null distribution of $LR_{max}(r_0)$ is a function of Brownian motion, which depends on the dimension $K - r_0$ and the specification of the deterministic terms. Critical values for $LR_{trace}(r_0)$ and $LR_{max}(r_0)$ statistics are in Osterwald and Lenum [94] for $K - r_0 = 1, ..., 10$.

The algorithm for sequential procedure to determining r cointegrating vectors are:

a) Test $H_0 : r_0 = 0$ vs $H_1 : r_0 > 0$. If we NRH_0 then, there are no cointegrating vectors among K variables in $\mathbf{y}_t$. If we RH_0 then, there is at least 1 cointegrating vector and proceeds to test.

b) Test $H_0 : r_0 = 1$ vs $H_1 : r_0 > 1$. If we NRH_0 then, there is only one cointegrating vector. If we RH_0 then, here is at least two cointegrating vectors. The sequence is continued until we NRH_0.

3.4.3 Estimation vector error correction model

Lets to consider a VEC model without deterministic terms, and update the model in equation 3.30,

$$\Delta\mathbf{y}_t = \mathbf{\Pi}\mathbf{y}_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + ... + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + \mathbf{u}_t \tag{3.30}$$

where $\mathbf{y}_t = (y_{1t}, ..., y_{kt})'$ denote K variable, $\mathrm{rank}(\mathbf{\Pi}) = r$ and $0 < r < K$ so that $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $(K \times r)$ matrices with $\mathrm{rank}(\boldsymbol{\alpha}) = \mathrm{rank}(\boldsymbol{\beta}) = r$. Also we assume $\mathbf{y_t}$ is I(1) process.

Writing the equation 3.30 in compact form we have equation 3.31

$$\mathbf{\Delta Y} = \mathbf{\Pi Y_{-1}} + \mathbf{\Gamma \Delta X} + \mathbf{U} \tag{3.31}$$

where $\mathbf{\Delta Y} = [\Delta y_1, ..., \Delta y_T], \qquad \mathbf{Y}_{-1} = [y_0, ..., y_{T-1}], \qquad \mathbf{\Gamma} = [\Gamma_1, ..., \Gamma_{p-1}]$

$\mathbf{X} = [\Delta X_0, ..., \Delta X_{T-1}] \quad with \quad \mathbf{X}_{t-1} = [\Delta y_{t-1}, ..., \Delta y_{t-p+1}]' \quad and \quad \mathbf{U} = [u_1, ..., u_T]$

According to Breitung *et al.* [24], given a specific matrix $\mathbf{\Pi}$, from equation 3.31 and using OLS estimator, the matrix of parameters $\mathbf{\Gamma}$ is given by

$$\widehat{\mathbf{\Gamma}} = (\mathbf{\Delta Y} - \mathbf{\Pi Y_{-1}})\mathbf{X}'(\mathbf{XX}')^{-1} \tag{3.32}$$

substituting in 3.31, after some algebra operations we get.

$$\mathbf{\Delta YM} = \mathbf{\Pi Y_{-1}M} + \widehat{\mathbf{U}} \quad where \quad \mathbf{M} = \mathbf{I} - \mathbf{X}'(\mathbf{XX}')^{-1}\mathbf{X} \tag{3.33}$$

For any integer number $0 < r < K$ an estimator $\widehat{\mathbf{\Pi}}$ of $\mathbf{\Pi}$ with $\mathrm{rank}(\mathbf{\Pi}) = r$ can be obtained by reduce rank regression based on the model 3.33. Following Johansen [65], the estimator is determined defining first the matrices.

$$\mathbf{S_{00}} = \mathbf{T^{-1}\Delta YM\Delta Y'}; \quad \mathbf{S_{01}} = \mathbf{T^{-1}\Delta YMY'_{-1}}; \quad \mathbf{S_{11}} = \mathbf{T^{-1}Y_{-1}MY'_{-1}}$$

next, is necessary to solve the generalized eigenvalue problem given by the equality $\det(\lambda \mathbf{S_{11}} - \mathbf{S'_{01}S_{00}^{-1}S_{01}}) = \mathbf{0}$.

Let $\lambda_1 \geq ... \geq \lambda_k$ the ordered eigenvalue with corresponding matrix of eigenvectors $\mathbf{V} = [b_1, ..., b_k]$ satisfying the relation $\lambda_i \mathbf{S_{11}} b_i = \mathbf{S'_{01}} \mathbf{S_{00}^{-1}} \mathbf{S_{01}} b_i$ and normalized such that $\mathbf{V'S_{11}V} = \mathbf{I_K}$. The reduced rank estimator of $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta'}$ is then obtained by choosing $\widehat{\boldsymbol{\beta}}$.

Let $\widehat{\boldsymbol{\beta}} = [b_1, ..., b_k]$, replacing $\boldsymbol{\alpha}\widehat{\boldsymbol{\beta}}$ in model 3.33 we have $\mathbf{\Delta YM} = \boldsymbol{\alpha}\widehat{\boldsymbol{\beta}}\mathbf{Y_{-1}M} + \widehat{\mathbf{U}}$, now using OLS estimator we have $\widehat{\boldsymbol{\alpha}} = \mathbf{\Delta YMY'_{-1}}\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}'\mathbf{Y_{-1}MY'_{-1}}\widehat{\boldsymbol{\beta}})^{-1}$.

Known the values of matrices of cointegration $\widehat{\boldsymbol{\beta}}$ and adjustment $\widehat{\boldsymbol{\alpha}}$ parameters, the corresponding estimator of $\mathbf{\Pi}$ is $\widehat{\mathbf{\Pi}} = \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'$. Using equation 3.32, we determine the estimator of $\mathbf{\Gamma}$, that is: $\widehat{\mathbf{\Gamma}} = (\mathbf{\Delta Y} - \widehat{\mathbf{\Pi}}\mathbf{Y_{-1}})\mathbf{X'}(\mathbf{XX'})^{-1}$.

Under Gaussian assumption the ML estimators are consistent and jointly asymptotically normal.

$$\sqrt{T}vec(\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}) \xrightarrow{d} N(0, \Sigma_{\mathbf{\Gamma}}) \text{ and } \sqrt{T}vec(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}) \xrightarrow{d} N(0, \Sigma_{\mathbf{\Pi}})$$

3.5 Structural VAR and SVEC analysis

In the previews sections, we discusses the VAR and VEC models which analyse the relation between a set of variables. However, there are some disadvantages to point, for example, the many number of parameters, difficult to interpret due to complex interactions between variables in the model, the impulse response are generally not unique and interpretation is often not clear which set of impulse responses actually reflects at a given point in the system. Due to this difficult, the VAR require reduced form or structural restrictions to identify the relevant innovations and impulse response.

In the literature there are different possible restrictions proposed, known as structural VAR (SVAR) models, described by Bernanke [8], Blanchard and Quah [14] and Sims [110] or Structural VEC (SVEC) models describe in Breitung *et al.* [24] and King *et al.* [68].

According to Breitung *et al.* [24], Zivot and Wang [123], the general structural models include deterministic terms and exogenous variables. It should be noted that for the present study, both the deterministic term and the exogenous variables will not be included in the model, as they do not influence the responses to impulses or stochastic shocks respectively.

From VAR model 3.1 we obtain the corresponding SVAR model, equation 3.34 multiplying the equation 3.1 by matrix $\mathbf{A}$, assuming that the matrix is nonsingular.

$$\mathbf{A}\mathbf{y}_t = \mathbf{\Phi}_1^* \mathbf{y}_{t-1} + ... + \mathbf{\Phi}_p^* y_{t-p} + \boldsymbol{\varepsilon}_t \tag{3.34}$$

where $\mathbf{\Phi}_i^* = \mathbf{A}\mathbf{\Phi}_i$ are structural form parameter matrices and $\boldsymbol{\varepsilon}_t = \mathbf{A}\mathbf{u}_t$ is a structural error term with mean zero and covariance matrix Σ_ε, that is, $\boldsymbol{\varepsilon} \sim N(0, \Sigma_\varepsilon)$, note that

$\sum_{\varepsilon} = \mathbf{A} \sum_u \mathbf{A}'$ and the inverse of matrix $\mathbf{A}$ allows modelling instantaneous relations among the variables in the system.

Following Kilian [67], Zivot and Wang [123], the three main types of structural analysis are (1) Granger - causality tests, (2) impulse response function and (3) forecast error variance decomposition, as well as historical decomposition.

According to Lüktepohl [77], the SVAR or SVEC models instead of identifying the coefficients, the identification is realized in residuals of system, which are interpreted as exogenous shocks. The goal of this identification is to obtain impulse response functions and variance decomposition with a "more" causal meaning. Here the problem is to know the effect of a unit shock of one variable onto another variable while imposing a restriction. In the literature the terms short – run and long - run are used, meaning that the effect of the shock lasts a short or a long time on the other variables, respectively.

3.5.1 Identification structural shocks

According to Breitung $et\ al.$ [24], if the process $\mathbf{y}_t$ is I(0), the effects of shocks in the variables of a given system are most seen in its Wold moving average (MA) representation, $\mathbf{y}_t = \mathbf{u}_t + \mathbf{\Psi}_1 \mathbf{u}_{t-1} + \mathbf{\Psi}_2 \mathbf{u}_{t-2} + ...$, where $\mathbf{\Psi}_0 = \mathbf{I}_K$, $\mathbf{\Psi}_s = \sum_{j=1}^{s} \mathbf{\Psi}_{s-j} \mathbf{\Phi}_j$ and $\mathbf{\Phi}_j$ is $K \times K$ matrix of coefficients of VAR(p) model, equation 3.1. The coefficients ψ_{ij} of the matrices $\mathbf{\Psi}_s$ is interpreted as the response to the impulse of structural shocks in the system. Note that, obtain forecast error impulse responses is feasible if Σ_u is diagonal, otherwise, the forecast error impulse responses are not valid, because the underlying shocks are not likely to occur in isolation if the components of $\mathbf{u}_t$ are instantaneously correlated.

In practice, one way to obtain orthogonal innovations is to use Choleski decomposition of the white noise covariance matrix $\Sigma_u = \mathbf{PP}'$ such that $\mathbf{P}$ is a lower triangular matrix with positive elements on the main diagonal.

According to Breitung $et\ al.$ [24] and Lütkepohl [77] there are different procedures used to non-sample information in specifying and then obtaining unique impulse responses. The three most used are associated with the so-called $\mathbf{A}$ - model, $\mathbf{B}$ - model and $\mathbf{AB}$ - model (see also Amisano and Giannini [1]).

The A - model

Lets the structural errors from equation 3.34 have the following MA representation

$$\mathbf{y}_t = \mathbf{\Theta_0}\varepsilon_t + \mathbf{\Theta_1}\varepsilon_{t-1} + \mathbf{\Theta_2}\varepsilon_{t-2} + ..., \tag{3.35}$$

where the matrices $\mathbf{\Theta_j} = \mathbf{\Psi_j}\mathbf{A^{-1}}$ for $j = 1, 2, ...$; represent the response to structural shocks $\boldsymbol{\varepsilon}_t$ assumed to be mutually uncorrelated. The relation between structural innovations and residuals from the VAR model 3.1 are $\boldsymbol{\varepsilon}_t = \mathbf{A}\mathbf{u}_t$, which is called the $\mathbf{A}$ - model. For this model, the identifying restrictions are imposed on $\mathbf{A}$ matrix such that, the product $\mathbf{A}\mathbf{u}_t = \boldsymbol{\varepsilon}_t$ has a diagonal covariance matrix. If it is plausible that $\mathbf{A}$ has a one in the main diagonal, they are necessary $K(K-1)/2$ restrictions on $\mathbf{A}$ to ensure just identified shocks, consequently identified impulse response.

Summarizing, according to Breitung *et al.* [24] and Lütkepohl [77], in $\mathbf{A}$ - model, the set of innovations $\boldsymbol{\varepsilon}_t$ is modelled in the system of equations as a function of the residuals such that $\mathbf{A}\mathbf{u}_t = \boldsymbol{\varepsilon}_t$ and the linear restrictions on $\mathbf{A}$ can be write in explicit form as $vec(\mathbf{A}) = R_{\mathbf{A}}\gamma\mathbf{A} + \mathbf{r_A}$, where $\gamma\mathbf{A}$ represent all unrestrited elements of $\mathbf{A}$, $\mathbf{R_A}$ is a suitable matrix with zero and one elements and $\mathbf{r_A}$ is the vector of normalized constants.

The B - model

In SVAR model equation 3.34, because the shocks are not observed, some assumptions are need to identify them. For instance we required that the structural shocks should be orthogonal, and that the structural shock $\boldsymbol{\varepsilon}_t$ are assumed to be related to the model residuals by linear relations $\mathbf{u}_t = \mathbf{B}\boldsymbol{\varepsilon}_t$, where $\mathbf{B}$ is a $K \times K$ matrix. Note that we assume that for $\mathbf{A}\mathbf{u}_t = \mathbf{B}\boldsymbol{\varepsilon}_t$ and the matrix $\mathbf{A} = \mathbf{I}_K$.

Using the relation between model residuals and structural innovations, one can write equation 3.34 as follows:

$$\mathbf{A}\mathbf{y}_t = \mathbf{\Phi}_1^*\mathbf{y}_{t-1} + ... + \mathbf{\Phi}_p^* y_{t-p} + \mathbf{B}\boldsymbol{\varepsilon}_t \qquad (3.36)$$

where $\boldsymbol{\varepsilon}_t \sim N(0, \mathbf{I}_K)$, $\mathbf{\Phi}^*$ are $K \times K$ coefficient matrices, $\mathbf{B}$ is $K \times K$ matrix such that $\mathbf{\Sigma}_u = \mathbf{B}\mathbf{\Sigma}_\varepsilon\mathbf{B}'$. Because the covariance matrix of $\boldsymbol{\varepsilon}_t$ are normalized with one in diagonal, the covariance matrix of residuals became $\sum_{\mathbf{u}} = \mathbf{B}\mathbf{B}'$, where $\mathbf{B}$ is lower triangular chosen by Choleski decomposition.

According to Lütkepohl and Krätzig [78], the equality $\mathbf{u}_t = \mathbf{B}\boldsymbol{\varepsilon}_t$ where the structural shock $\boldsymbol{\varepsilon}_t \sim N(0, \mathbf{I}_K)$ is the so called $\mathbf{B}$-model. Due to the symmetry of the covariance matrix of $\sum_{\mathbf{u}}$, we specify only $K(K+1)/2$ different equations and need additional $K(K-1)/2$ restrictions to be imposed to identify all K^2 elements of $\mathbf{B}$.

In summary, for $\mathbf{B}$-model, the identification consists to exclude some linear combinations of the structural shocks by imposing restriction of the form $vec(\mathbf{B}) = \mathbf{R_B}\boldsymbol{\gamma_B} + \mathbf{r_B}$, where $\boldsymbol{\gamma_B}$ contains unrestricted elements of $\mathbf{B}$, $\mathbf{R_B}$ are the restricted matrices with 0 or 1 elements, and $\mathbf{r_B}$ is a vector of constant values normalized.

The AB model

According to Lütkepohl [77], situation where is possible to consider both types of restrictions, that is, from equation 3.36, $\mathbf{A}\mathbf{u}_t = \mathbf{B}\boldsymbol{\varepsilon}_t$ where $\boldsymbol{\varepsilon}_t \sim N(0, \mathbf{I}_K)$, we are facing a model known as **AB** - model which relates the reduced form errors $\mathbf{u}_t$ to the underlying structural shocks $\boldsymbol{\varepsilon}_t$.

Identification the structural shocks can be done, imposing restrictions on $\mathbf{A}$ and $\mathbf{B}$ parameter matrices. Here for K-variables $K(K-1)/2$ restrictions are necessary for orthogonalizing the shocks because there are $K(K-1)/2$ different instantaneous covariances. Detail of the procedures can be found in Breitung *et al.* [24] and Galí [50].

For this model, from a previous relation between residuals of the model and structural shock, we get $\mathbf{u}_t = \mathbf{A}^{-1}\mathbf{B}\boldsymbol{\varepsilon}_t$, and the corresponding covariance matrix is $\sum_u = \mathbf{A}^{-1}\mathbf{B}\mathbf{B}'\mathbf{A}^{-1\prime}$. Thus, we have $K(K+1)/2$ equations. Meaning that we need K^2 elements for each matrix $\mathbf{A}$ and $\mathbf{B}$, additionally $2K^2 - \frac{1}{2}K(K+1)$ restrictions to identify all $2K^2$ elements of $\mathbf{A}$ and $\mathbf{B}$ at least locally. Even considering that the matrix $\mathbf{A}$ has elements one on main diagonal, $2K^2 - K - \frac{1}{2}K(K+1)$ restrictions are needed for identification of the **AB** - model.

In order to avoid the large number of restrictions in most applications is used the special cases just discussed. The $\mathbf{A}$ - model where $\mathbf{B} = \mathbf{I}_K$ or the $\mathbf{B}$-model with $\mathbf{A} = \mathbf{I}_K$. According to Breitung *et al.* [24] and Lütkepohl [77], considering the general case **AB**-model, the constraints are normally normalized and they can be written in the form of linear equations.

$$vec(\mathbf{A}) = \mathbf{R_A}\boldsymbol{\gamma_A} + \mathbf{r_A} \quad \text{and} \quad vec(\mathbf{B}) = \mathbf{R_B}\boldsymbol{\gamma_B} + \mathbf{r_B} \tag{3.37}$$

where $\mathbf{R_A}$ and $\mathbf{R_B}$ are possible restricted matrices with zero and ones, $\boldsymbol{\gamma_A}$ and $\boldsymbol{\gamma_B}$ are unrestricted vectors with free parameters and finally $\mathbf{r_A}$ and $\mathbf{r_B}$ are vectors of fixed parameters.

Identification SVEC model

If all or some variables in the system are integrated with r cointegrating vectors, which describe the long-run relationships between variables, the VEC model, equation 3.30 is an appropriate model and the corresponding SVEC model is given by,

$$\mathbf{A}\Delta\mathbf{y}_t = \boldsymbol{\Pi}^*\mathbf{y}_{t-1} + \boldsymbol{\Gamma}_1^*\Delta\mathbf{y}_{t-1} + ... + \boldsymbol{\Gamma}_{p-1}^*\Delta\mathbf{y}_{t-p+1} + \mathbf{B}\boldsymbol{\varepsilon_t} \tag{3.38}$$

where $\boldsymbol{\Pi}^*$ and $\boldsymbol{\Gamma}_i^*$ are structural form parameters matrices and $\boldsymbol{\varepsilon}_t$ is a structural error term, $\mathbf{B}$ is a $K \times K$ matrix of impact of shocks and matrix $\mathbf{A}$ models the instantaneous relations among the variables in the system.

According to Pfaff [98], in order to identify the structural form parameters, we need to considers the Beveridge-Nelson (Beveridge and Nelson [10]) moving average representation of the variables $\mathbf{y}_t$ if they adhere to the VEC model process as in equation 3.30.

$$\mathbf{y}_t = \underbrace{\mathbf{\Xi} \sum_{i=1}^{t} \mathbf{u}_i}_{I(1)} + \underbrace{\sum_{j=0}^{\infty} \mathbf{\Xi}_j^* \mathbf{u}_{t-j}}_{I(0)} + \mathbf{y}_0^* \tag{3.39}$$

where the matrix $\mathbf{\Xi}$ is defined as $\mathbf{\Xi} = \boldsymbol{\beta}_\perp \left(\boldsymbol{\alpha}_\perp' (\mathbf{I}_k - \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i) \boldsymbol{\beta}_\perp \right)^{-1} \boldsymbol{\alpha}_\perp'$, with $\boldsymbol{\alpha}_\perp$ and $\boldsymbol{\beta}_\perp$ indicate orthogonal components of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively. The first term I(1) is the common trend that drives the system of $\mathbf{y}_t$, the second term I(0) it bounded by the infinite sum because the series of $\mathbf{\Xi}_j^*$ converges to zero as $j \to \infty$ and $\mathbf{y}_0^*$ contains all initial values.

According to Breitung *et al.* [24] and Pfaff [98], since the long-run effects of shocks is captured on the common trend, this is the centre of modelling Structural VEC model.

Replacing $\mathbf{u}_t$ in equation 3.39 by $\mathbf{A}^{-1}\mathbf{B}\boldsymbol{\varepsilon}_t$ and then, assuming that $\boldsymbol{\varepsilon}_t \sim N(0, \mathbf{I}_K)$ and the matrix $\mathbf{A} = \mathbf{I}_k$, then the impact of the short-run or transitory orthogonal shocks are obtained by $\mathbf{\Xi_j^* B}$ and the long-run by $\mathbf{\Xi B}$ in the same way as in the stationary VAR model.

For example in $\mathbf{B}$ - model using the relation $\mathbf{u}_t = \mathbf{B}\boldsymbol{\varepsilon}_t$, is necessary at least $K(K-1)/2$ restrictions to identify $\mathbf{B}$. But if rank$(\mathbf{\Xi B}) = K-r$, where r is the rank of the system $\mathbf{y}_t$, means that r columns have zeros in the matrix of long-run, which corresponds to r structural innovations having transitory shocks, the remainder $m = K - r$ structural innovations have permanent shocks. To identify the permanent shocks exactly we need $m \times (m-1)/2$ additional restrictions. Similarly $r(r-1)/2$ additional contemporaneous restrictions needed to identify the transitory shocks. Continuing with the assumption that $\mathbf{A} = \mathbf{I}_K$, we have just enough restrictions to identify $\mathbf{B}$ and $\mathbf{\Xi B}$. In this case we have a model exactly identified, if we impose additional restrictions on the parameters, so it would be an overidentified model, and the overidentified restriction could be tested.

3.5.2   Estimation SVAR and SVEC models

From subsection 3.5.1 concerning the identification of structural shocks and assuming that the $\mathbf{A}$-model and $\mathbf{B}$-model are special cases for $\mathbf{AB}$-model, we consider the general case.

Given the following SVAR model in compact form

$$\mathbf{A}\mathbf{Y}_t = \mathbf{A}\mathbf{\Phi}\mathbf{Y}_{t-1} + \mathbf{B}\boldsymbol{\varepsilon}_t \tag{3.40}$$

where $\mathbf{Y}_{t-1} = (y_{t-1}, ..., y_{t-p})'$; $\mathbf{\Phi} = (\Phi_1, ..., \Phi_p)$ and $\boldsymbol{\varepsilon}_t$ is assumed to be Normal distribution with mean zero and covariance matrix $\mathbf{I}_K$, that is $\boldsymbol{\varepsilon}_t \sim N(0, \mathbf{I}_K)$.

Using the VAR model 3.1, the residuals form corresponding to SVAR model 3.40 have the form $\mathbf{u}_t = \mathbf{A}^{-1}\mathbf{B}\boldsymbol{\varepsilon}_t$.

The log-likelihood function for a sample of observations $y_1, ..., y_T$ is given by

$$
\begin{aligned}
\ln l(\mathbf{\Phi}, \mathbf{A}, \mathbf{B}) &= -\frac{KT}{2}\ln(2\pi) - \frac{T}{2}\ln|\mathbf{A}^{-1}\mathbf{B}\mathbf{B}'\mathbf{A}'^{-1}| \\
&\quad - \frac{1}{2}\mathrm{tr}\{(\mathbf{Y} - \mathbf{\Phi}\mathbf{X})'[\mathbf{A}^{-1}\mathbf{B}\mathbf{B}'\mathbf{A}'^{-1}]^{-1}(\mathbf{Y} - \mathbf{\Phi}\mathbf{X})\} \\
&= \mathrm{const} + \frac{T}{2}\ln|\mathbf{A}|^2 - \frac{T}{2}\ln|\mathbf{B}|^2 \\
&\quad - \frac{1}{2}\mathrm{tr}\{\mathbf{A}'\mathbf{B}'^{-1}\mathbf{B}^{-1}\mathbf{A}(\mathbf{Y} - \mathbf{\Phi}\mathbf{X})(\mathbf{Y} - \mathbf{\Phi}\mathbf{X})'\}
\end{aligned} \tag{3.41}
$$

where, $\mathbf{Y} := [y_1, ..., y_T]$, $\mathbf{X} := [\mathbf{Y_0}, ..., \mathbf{Y_{T-1}}]$ and the matrix rules are
$|\mathbf{A}^{-1}\mathbf{B}\mathbf{B}'(\mathbf{A}^{-1})'| = |\mathbf{A}^{-1}|^2|\mathbf{B}|^2 = |\mathbf{A}|^{-2}|\mathbf{B}|^2$ and $\mathrm{tr}(VW) = \mathrm{tr}(WV)$

According to Lütkepohl [77], if there is no restrictions in reduced form VAR parameters $\mathbf{\Phi}$, knowing $\mathbf{A}$ and $\mathbf{B}$, the log-likelihood function in 3.41 is maximized with respect to $\mathbf{\Phi}$ by $\hat{\mathbf{\Phi}} = \mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}$. Next substitute $\mathbf{\Phi}$ with $\hat{\mathbf{\Phi}}$ in 3.41 to get the concentrated log-likelihood.

$$\ln l_c(\mathbf{A}, \mathbf{B}) = \mathrm{const} + \frac{T}{2}\ln|\mathbf{A}|^2 - \frac{T}{2}\ln|\mathbf{B}|^2 - \frac{T}{2}\mathrm{tr}(\mathbf{A}'\mathbf{B}'^{-1}\mathbf{B}^{-1}\mathbf{A}\tilde{\Sigma}_u) \tag{3.42}$$

where $\tilde{\Sigma}_u = T^{-1}(\mathbf{Y} - \mathbf{\Phi}\mathbf{X})(\mathbf{Y} - \mathbf{\Phi}\mathbf{X})'$ is the covariance matrix of VAR residuals.

According to Breitung *et al.* [24], maximization of the concentrated log-likelihood function 3.42 with respect to matrices $\mathbf{A}$ and $\mathbf{B}$, subject to the structural restrictions 3.37, can be done by iterative methods because a closed form solution is not available. In this context Amisano and Giannini [1], suggest to use a scoring algorithm to estimate $\boldsymbol{\gamma_A}$ and $\boldsymbol{\gamma_B}$ parameters iteratively.

$$\begin{bmatrix} \tilde{\gamma}_A \\ \tilde{\gamma}_B \end{bmatrix}_{i+1} = \begin{bmatrix} \tilde{\gamma}_A \\ \tilde{\gamma}_B \end{bmatrix}_i + \ell\, I\left(\begin{bmatrix} \tilde{\gamma}_A \\ \tilde{\gamma}_B \end{bmatrix}_i\right)^{-1} \mathfrak{S}\left(\begin{bmatrix} \tilde{\gamma}_A \\ \tilde{\gamma}_B \end{bmatrix}\right) \tag{3.43}$$

where ℓ is the step length, I(*) denotes de information matrix of the free parameters $\boldsymbol{\gamma}_A$ and $\boldsymbol{\gamma}_B$, $\mathfrak{S}(*)$ is the score vector and the subscript refers to the iteration number from which the signified estimator is obtained.

For specific SVAR model, with initial values, iteration stops when the relative change in log-likelihood and the change in the parameters are no longer significant

according to a certain chosen criterion. The ML estimator obtained with this procedure is asymptotically efficient and normally distributed, where the asymptotic covariance matrix is estimated by the inverse of the information matrix, That is, the ML estimator for $\sum_u$ is given by $\widetilde{\sum}_u^r = \widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{B}}\widetilde{\mathbf{B}}'\widetilde{\mathbf{A}}'^{-1}$, where $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ are estimators of $\mathbf{A}$ and $\mathbf{B}$ respectively.

We notice that the asymptotic covariance matrix $\tilde{\sum}_u^r$ only corresponds to the reduced form estimate $\tilde{\sum}_u$ if the SVAR is exactly identified. When we have over-identifying restrictions imposed on $\mathbf{A}$ and $\mathbf{B}$ (or $\mathbf{B}$) the estimator differ from $\tilde{\sum}_u$. In this case, the LR statistic is used to check validity of over-identifying restrictions,

$$LR = T(\ln|\tilde{\mathbf{\Sigma}}_u^r| - \ln|\tilde{\mathbf{\Sigma}}_u|) \tag{3.44}$$

under the null hypothesis that the restrictions are valid, the statistic LR is asymptotic χ^2 distribution with degree of freedom equal to the number of over-identifying restrictions. The null hypothesis is rejected if the LR statistics is greater then the corresponding critical value.

According to Breitung $et\ al.$ [24], the concentrated likelihood function 3.42 can be used to estimate the structural parameters $\mathbf{A}$ and $\mathbf{B}$ for SVEC models. If there is no restrictions imposed on short-run parameters, the estimated covariance matrix $\widetilde{\sum}_u$ represents the residual covariance matrix from a reduced rank regression. If the short-run parameters are restricted or restrictions are imposed on the cointegration vectors, $\sum_u$ may be estimated using other estimator instead of the ML estimator.

In practice, if long-run constraints are considered, then the constraints are expressed as linear restrictions and maximization of concentrated likelihood function in equation 3.42 can be done using the scoring algorithm suggested by Amisano and Giannini [1].

In particular, if we consider that $\mathbf{A} = \mathbf{I}_K$, restrictions may be imposed such that some long-run shocks have no impact on the system variables and can be written in implicit form as $R_{\mathbf{\Xi}}vec(\mathbf{\Xi B}) = 0$, where $R_{\mathbf{\Xi}}$ is an appropriate restriction matrix. Using the rules of the operator vec, we can rewrite the restrictions as $R_{\mathbf{\Xi}}(I_K \otimes \mathbf{\Xi})vec(\mathbf{B}) = R_{\mathbf{B},\ell}vec(\mathbf{B}) = 0$, substitute $\mathbf{\Xi}$ by the estimator $\widehat{\mathbf{\Xi}}$ obtained from reduced form we have $\widehat{R}_{\mathbf{B},\ell} = R_{\mathbf{\Xi}}(\mathbf{I}_K \otimes \widehat{\mathbf{\Xi}})$.

This implicit constraint is transformed into the explicit form and used in the maximization procedure of the SVEC model in combination with contemporaneously constraints on elements of matrix $\mathbf{B}$ of the form $vec(\mathbf{B}) = R_{\mathbf{B}}\boldsymbol{\gamma}_{\mathbf{B}}$. The procedure works without problem even when there are more constraints on structural parameters than are required for accurate identification. In these cases, over-identification is tested by equation 3.44.

3.5.3 Structural statistical inferences

As in the VAR models, some analyses performed after the parameter estimation of structural models consists of a representation of impulse response functions and consequent interpretation, forecast error variance decomposition as well as historical decomposition.

From subsection 3.3.6 it is known that given a system of stationary variables $\mathbf{y}_t \sim I(0)$, the set of residuals from the regression equations can be represented as a moving average model where the corresponding coefficients represent the response to the impulse of structural shocks in the system. Since residuals exhibit changes, these changes are reflected in the variables involved in the system. Which causes the effect of the impulses of the $\mathbf{y}_t$ components associated with the residuals to be transitory and disappear over time.

If SVEC model is applied in the study, the impulse response functions are computed based on the VAR model with cointegrated variables, or a level version of VEC model, which implies that the impulse responses $\boldsymbol{\Psi}_s$ may not converges to zero as $s \to \infty$, consequently, some shocks will have permanent effects.

According to Bruggemann [28] inference on impulse response functions based on bootstrap methods sometimes they have better small sample interval properties compared with intervals based on asymptotic theory. Because in application we use bootstrap confidence interval we will mention three types of bootstrap confidence intervals.

a) **Standard percentile interval**. The method is based on the standard percentile interval. In tribute to the author in the literature, the method is known as Efron interval, (see Efron and Tibshinani [37]).

b) **Hall's percentile interval**. The method is based on the confidence interval of the empirical distribution. Using the same justification as above the method is called Hall interval, (see Hall [57]).

c) **Hall's studentized interval**. The method uses a studentized statistics to construct the bootstrap confidence interval. The procedure is known as studentized Hall, (see Hall [57]).

The procedure and significance of the forecast error variance decomposition, for investigating the impacts of shocks on the SVEC system is similar to that presented for VAR models in the subsection 3.3.7.

67

3.6 Application of vector autoregressive models

3.6.1 Introduction

In this application, as described in Mulenga [85] we analyse the dynamics of Mozambique real Gross Domestic Product (GDP), GDP deflator and the South Africa real GDP, which here play the role of external shocks. Mozambique and South Africa are neighbor countries with a continuous relationship based on the use of labor force, infrastructures, resource exploration, cross-border trade and so on. Because of this relationship and influence in the Mozambique economy, we chose South Africa real GDP to serve as external shocks.

According to Mussagy and Allaro [89], the economic crises, changes in institutional arrangement, policy changes and regime shifts among others can influence macroeconomic analysis of time series data. In particular, Mozambique since 1975 had three changes in format and currency designation, beginning with the Metical of Republic Popular of Mozambique in 1980, when the former Portuguese Shield was abandoned, followed by the Metical of Republic of Mozambique in 1990 and finally the New Metical of Mozambique in 2006, when three zeros were withdrawn in the currency as a way to reduce inflation. On the other hand, the exchange law has also been changed to adapt the standards of operations in different regimes from the popular to the current market economy of a free circulation of people, goods and services. For example, the exchange law of January 3/96 was changed to March law 11/2009. Given these constant changes, the macroeconomic series tend to exhibit fluctuations and present cuts that create difficulty in finding long series for econometric studies like this one. Therefore and due to the absence of long periods of the series, we only obtained the annual series between 1990 to 2012.

For our tri-variable analysis, we use instead of structural vector autoregressive model, the structural vector error correction analysis to take into account the vector error correction and identify contemporaneous and long-run effects of shocks. The identification is based on economic relationship between variables, inspired in Blanchard and Quah [14] analysis of aggregate supply and aggregate demand shocks, interpreted as permanent and transitory shocks. In this application, we use both short and long-run restrictions as Galí [50] and we interpret the aggregate supply, demand and external shocks splitting them into permanent and transitory. Regarding the estimation procedure, we use one stage Johansen approach (see Johansen [65]) to estimate the SVEC model, where we placed two restrictions on the short-run effects of shocks and two for long-run relation. In practice, we assume that prices are not affected by contemporaneous supply shocks nor by external shocks. For long-run we

consider that only an aggregate supply shock has permanent effects.

The application contribution relies on identifying aggregate supply and aggregate demand shocks in the presence of external shocks to the economic growth of Mozambique. That is, the investigation identifies the sources of shocks in macroeconomic variables, that somehow helps the process of defining policies that can increase the rate of sustainable economic growth. Therefore, we consider three sources of shocks that are aggregate supply, aggregate demand and external shocks and we inspect whether there are permanent or transitory shocks.

The section is organized as follows: after the introduction in subsection 3.6.1, subsection 3.6.2 presents the data and statistical properties of the series, subsection 3.6.3 presents the vector error correction model specification and SVEC analysis to identification of shocks from the aggregate supply, aggregate demand and external activity. The subsection 3.6.4 reports the empirical results from both VEC and SVEC analysis and, finally, the subsection 3.6.5 presents the main conclusions.

3.6.2 Data and statistical properties

For our study, we use three macroeconomic variables, two related to Mozambique economy and one is referred to the neighbor country of South Africa, that is, the Mozambique real Gross Domestic Product (MZGDP), representing the supply, the Mozambique GDP deflator (PMZGDP), represent the demand and the South Africa real Gross Domestic Product (ZAGDP), which represent the external shocks for Mozambique economy. As explained in introduction, the application use annual series from 1990 through 2012, collected from the World Bank database [1].

Generally the observations of the time series are not stationary, and do not present a behavior that can be described by a linear function. In such cases, transformations may be useful to smooth the variability over time of a single series. In particular, we take logarithm to the series. For visualization of the behavior of the data series, the Figure 3.1, presents the series.

The time series behavior shows nonstationarity for the series in level. So we apply a formal unit root test for the series with the null hypothesis that the series has a unit root, and include three lags in Dickey and Fuller regression to eliminate serial correlation in the error term. From Augmented Dickey and Fuller, we have p-value equal to 0.9478, 0.9448 and 0.9478 for Mozambique real GDP, Mozambique GDP deflator and South Africa real GDP, respectively, that confirm the nonstationarity of the series in level. Then we transform in stationary series making first differences, as indicated by Augmented Dickey and Fuller p-value, equal to 0.0082 for Mozambique

---

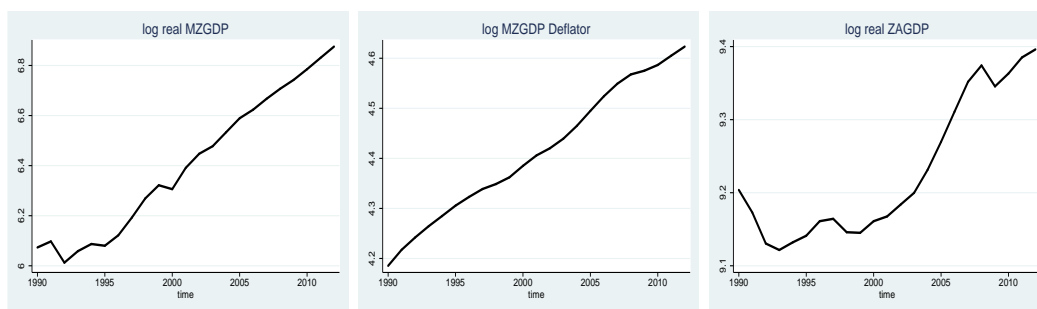[1]https://datacatalog.worldbank.org/search/indicators.

Figure 3.1: The behavior of the series of Mozambique log real GDP, Mozambique log GDP deflator and South Africa log real GDP, 1990-2012

real GDP, 0.0000 for Mozambique GDP deflator and 0.0289 for South Africa real GDP, all less than 0.05.

Table 3.1: Correlations between pairs of log differences of the series

Variable (i,j)	$MZGDP_{t-1}$	$MZGDP_t$	$MZGDP_{t+1}$
$PMZGDP_t$	-0.1886	-0.2862	-0.2828
$ZAGDP_t$	0.1616	0.3629	0.3314
Variable (i,j)	$PMZGDP_{t-1}$	$PMZGDP_t$	$PMZGDP_{t+1}$
$ZAGDP_t$	0.0390	0.2563	0.2029

To verify if there is any relationship between the variables over time, we calculate the Pearson linear correlation for all pairs of log differences of the series with lead-lag of up to one period, as the Table 3.1 illustrates. The results indicate that all correlations are below $2/\sqrt{T}$ that is, 0.43 meaning a weak relation between the series. In particular, the correlations between Mozambique real GDP (output) and Mozambique GDP deflator (prices) are all negative over time indicating that the series develops in the opposite directions. While the external economic activity represented by South Africa real GDP has a positive relationship with Mozambique output and prices. That is, there are positive co-movements of the series over time, which suggests a long-term equilibrium relationship[2].

3.6.3 Structural VEC model specification and testing

Before estimating the parameters of the VEC or SVEC it is necessary to verify if the required conditions for the adequacy of the selected model are satisfied. Once the parameters of the model has been estimated, is necessary to check if the model is an adequate description of the data. For this purpose, we perform the test for stability

---

[2]Correlations computed by STATA.

condition of the system, the residual autocorrelation test, the residual normality test and finally the cointegration test.

Regarding multivariate tests we use the multivariate Portmanteau test for autocorrelation proposed by Hosking [61], as described in Lütkepohl [77], which is a generalization of the univariate Ljung-Box test. For multivariate normality test, we use the statistic proposed by Doornik and Hansen [36], which is a generalization of Jarque and Bera normality test for univariate case. The test is based on sample measures of multivariate Skewness and Kurtosis. In this case, before estimation the VEC and SVEC models, it is necessary to know the number of cointegrating vectors of the series. In this application we use the Johansen [63] procedure which uses likelihood ratio (LR) tests to determine the number r_0 of cointegrating vectors, as the trace test. We reject the null hypothesis of r_0 cointegrating vectors if the LR statistics is greater than the critical value.

Next we unfold an important task in the multivariate analysis of time series or in the study of the VAR models, that of verifying if there is any relationship between the pairs of variables using the Granger - causality test (Granger [54]). In our study we test Granger-causality between a pair of variables using the Dolado and Lütkepohl [35] procedure, since our series are integrated and eventually cointegrated. That is, for a VAR(p) model we fit VAR(p+1) system to the data and perform a Wald test on the coefficients of the first p lags only.

Because in our system of variables, the Augmented Dickey and Fuller test indicate that the series are nonstationary individually in level, the VAR analyse is not adequate. Therefore, situations where there are r cointegrating vectors, which describe the long-run relationships between variables, the VEC model, equation 3.28, is an appropriate.

According to Breitung *et al.* [24], Zivot and Wang [123], the general SVEC model include deterministic terms and exogenous variables. However, for our present purpose the deterministic terms are of no importance because they are do not shape the impulses in the system. Similarly, exogenous variables under the control of some policy markets may not react to stochastic shocks, therefore they are ignored.

Hence, instead of considering a complete model we will consider the simplified one.

$$\mathbf{A}\Delta\mathbf{y}_t = \mathbf{\Pi}^*\mathbf{y}_{t-1} + \mathbf{\Gamma}_1^*\Delta\mathbf{y}_{t-1} + ... + \mathbf{\Gamma}_{p-1}^*\Delta\mathbf{y}_{t-p+1} + \mathbf{B}\boldsymbol{\varepsilon_t} \qquad (3.45)$$

where $\mathbf{\Pi}^*$ and $\mathbf{\Gamma}_i^*$ are structural form parameters matrices and $\boldsymbol{\varepsilon}_t$ is a structural error term, $\mathbf{B}$ is a $K \times K$ of impact of shocks and matrix $\mathbf{A}$ models the instantaneous relations among the variables in the system.

Using our SVEC model with three variables, that is, the Output, Prices and

External activity, we assume that all variables are integrated of order one, with rank $r = 2$, which suggests two transitory shocks and one permanent shock. The permanent shock we identify by difference $m = K - r = 3 - 2 = 1$, which corresponds to one column with unrestricted elements on the long-run matrix and remaining two columns with hard zeros. The two restrictions mean that the demand and external shocks have no long term effects on variables, while the supply shock is allowed to have on all variables. Because the total number of restrictions is $K(K-1)/2 = 3(3-1)/2 = 3$, we replace the third restriction in $\mathbf{B}$ and we need $r(r-1)/2 = 1$ additional contemporaneous restriction to identify the transitory shocks. So the two restrictions on the short-run matrix are that the first and third shocks have no instantaneous effect on prices, that is, $\mathbf{b}_{21} = \mathbf{b}_{23} = 0$. In general we impose two long-run independent restrictions and additional two contemporaneous restrictions, to identify the supply shocks, demand shocks, and external shocks to be traced in an impulse response analysis, as illustrated ahead in the application.

$$\text{Short}-\text{run} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ 0 & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \quad \text{and} \quad \text{Long}-\text{run} = \begin{bmatrix} c_{11} & 0 & 0 \\ c_{21} & 0 & 0 \\ c_{31} & 0 & 0 \end{bmatrix}$$

As described in subsection 3.5.3, given $\mathbf{y}_t \sim I(0)$, the set of residuals from the regression equations can be represented as MA model where the coefficients are known as response to the impulse of structural shocks in the system.

For our study, we are interested in having accumulated effects, so we add the long-run effects of the matrices for all considered periods and because we apply the SVEC model, the impulse response functions are computed based on the VAR model with cointegrated variables, or a level version of VEC model, which implies that the impulse responses $\boldsymbol{\Psi}_s$ may not converges to zero as $s \to \infty$, consequently, some shocks will have permanent effects.

In practice, obtain forecast error impulse responses based on matrices related to long-run and short-run is appropriate if the covariance matrix of residuals Σ_u is diagonal, otherwise, the forecast error impulse responses are not valid, due to underlying shocks that are not likely to occur in isolation if the components of $\mathbf{u}_t$ have instantaneously correlations. To avoid this problem, orthogonal shocks are preferred for impulse response analysis.

Following Breitung *et al.* [24], one way to obtain orthogonal shocks is to use Choleski decomposition of the covariance matrix Σ_u. If $\mathbf{B}$ is a lower triangular matrix such that $\Sigma_u = \mathbf{B}\mathbf{B}'$ the orthogonal shocks are given by $\mathbf{u}_t = \mathbf{B}\boldsymbol{\varepsilon}_t$, where the forecast errors $(\mathbf{u}_t)$ are linear functions of the structural innovations $(\boldsymbol{\varepsilon}_t)$, assuming $\boldsymbol{\varepsilon}_t \sim N(0, \mathbf{I}_K)$.

Hence, the MA representation is $\mathbf{y}_t = \boldsymbol{\Theta}_0\varepsilon_1 + \boldsymbol{\Theta}_1\varepsilon_{t-1} + ...$, where $\boldsymbol{\Theta}_i = \boldsymbol{\Psi}_i\mathbf{B}$ for $i = 1, 2, ...$, here $\boldsymbol{\Theta}_0 = \mathbf{B}$ is lower triangular matrix with 1's along the diagonal and the triangular structural model, equation 3.39, imposes the recursive causal ordering $y_{1t} \to y_{2t} \to ... \to y_{kt}$. This ordering means that, an ε shock instantaneous in variable y_{1t} produce effects for all variables, while a shock in y_{2t} cannot have impact in y_{1t} but only on other variables, and so on. If there is no order imposed by the relations between the variables, Sims [110] suggests trying different orthogonal orders in the identification until finding the one that is most appropriate. Regarding impulse response statistical inference, we based on Bootstrap Efron Percentile and construct 95% confidence intervals.

The forecast error variance decomposition, are another tool for investigating the impacts of shocks in SVEC system. The FEVD gives information for a portion of the variance of the forecast error in predicting y_{it+h} which is due to variability in the structural shocks between times t and t+h. Using the orthogonal shocks the h-step ahead forecast error vector, with known SVEC coefficients, may be expressed as $\mathbf{y}_{t+h} - \mathbf{y}_{t+h|t} = \Sigma_{j=0}^{h-1}\boldsymbol{\Theta}_j\varepsilon_{t+h-j}$. Since the structural errors are orthogonal, the variance of the h-step ahead forecast error is: $var(y_{it+h} - y_{it+h|t}) = \sigma_{\varepsilon_1}^2 \Sigma_{j=0}^{h-1}(\theta_{i1}^j)^2 + ... + \sigma_{\varepsilon k}^2 \Sigma_{j=0}^{h-1}(\theta_{ik}^j)^2$, where $\sigma_{\varepsilon_j}^2 = var(\varepsilon_j)$ and the proportion of $var(y_{it+h} - y_{it+h|t})$ due to shock ε_j is then $FEVD_{ij}(h) = var(\varepsilon_j)/var(y_{it+h} - y_{it+h|t})$.

According to Breitung *et al.* [24], Zivot and Wang [123], when we want to make an identification using Choleski decomposition, it is important to take into account the order in which the variables are arranged in the system $\mathbf{y}_t$, because this order influences significantly the results. In particular, impulse response functions as well as forecast error variance decomposition are affected by the identification choice.

3.6.4 Empirical results

From data description, the ADF test found that the series are nonstationary in level individually but stationary after first differences, therefore, in this case, the good choice for analyze the set of these three variables is using the vector error correction model. From the literature theory, if two or more variables are individually nonstationary they may be cointegrated. That is, two or more variables may have common underlying stochastic trend along which they move together on a nonstationary path. As the vector error correction model fits a vector autoregressive in which some variables in the vector autoregressive system are cointegrated, VEC should be a good model.

In order to use the VEC model, first we apply the Johansen test for cointegration to verify the relationship between variables over time, and then to determine how

73

many cointegration equations exists. For that we run Johansen trace tests.

Table 3.2: Johansen trace test for cointegration, sample 1992–2012

maximum rank	number of parameters	log likelihood	eigenvalue	trace statistics	critical value 5%
0	12	194.509	–	62.332	29.68
1	17	217.321	0.886	16.706	15.41
2	20	225.101	0.523	1.147*	3.76
3	21	225.695	0.053	–	–

The results in Table 3.2 show that for rank equal to 2, the trace statistic of 1.147 is less than its correspondent critical value 3.76 at the 5% level. So we cannot reject the null hypothesis that there are two cointegrating equations and assume 2 cointegrating vectors. Using this evidence, we estimate the VEC parameters with two cointegrating vectors.

Because we are interested in cointegrating relations, we report in Table 3.3, the adjustment parameters and cointegration equations coefficients, where we notice that only the coefficients for log South Africa GDP are not significant, while all other adjustment α are statistically significant.

Table 3.3: Cointegrating and adjustment parameters (*p*-value in brackets)

	Real MZGDP	MZGDP Deflator	Real ZAGDP
Cointegrating parameters			
CE1 β_1	1.0000[0.0000]	–	-2.7120[0.0000]
CE2 β_2	–	1.0000[0.0000]	-1.1480[0.0000]
Adjustment parameters			
α_1	-0.7200[0.0000]	0.1310[0.0000]	0.0940[0.5160]
α_2	2.3210[0.0000]	-0.3060[0.0020]	0.1200[0.9780]

With respect to results in Table 3.3, using the adjustment parameters, it can be said that the Mozambique real GDP equation reverts to equilibrium regarding the long run relation between home and foreign real GDP, but deviates with respect to the second cointegration relation. For GDP deflator, the convergence result holds for the relation between prices and foreign output, but not for the first cointegrating vector. About foreign real GDP, we observe divergence with respect to deviations from equilibrium in both cointegration relations. Finally notice that, the Mozambique

74

real GDP has the fastest velocity to return to equilibrium in cases of deviation from the first equilibrium relation compared to other equations.

In order to assess the behavior of the cointegrating vectors we plot them in Figure 3.2, in-sample predicted values, where the graphics show oscillation above and below the horizontal line zero, which indicates a stationary process.
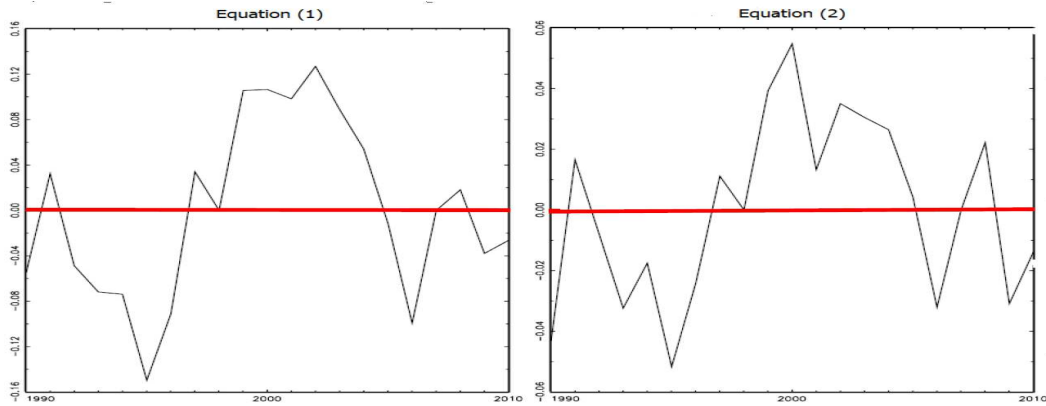

Figure 3.2: Plot of the estimated error correction terms: CE1 and CE2

To verify the suitability of the model, we apply the multivariate Portmanteau statistic for the presence of autocorrelation up to 12 lags indicate $\widehat{LB}$ equal to 69.0855 with p-value equal to 0.9700 greater than 0.05, suggesting no autocorrelation in residuals thus, we cannot reject the null hypothesis that there is no autocorrelation at lag 12 for instance. The statistics for normality of residuals indicate that the individual series, as well as the joint residuals, follow a Normal distribution. For illustration we report a multivariate version of Doornik and Hansen statistics for joint cases where it was obtained Skewness equal to 3.1269 with associated p-value equal to 0.3725, Kurtosis 1.7736 with p-value 0.6207 and the joint test statistic 4.9005 with p-value equal to 0.5566, therefore obtaining all p-value greater than 0.05.

In order to investigate the relationship between variables and determine which variable helps to improve the forecast of another over time we use Dolado and Lütkepohl procedure, that is, we run the VAR(2) model and perform a Wald test on the coefficients of the first lag only.

The results for Granger - causality test reported in Table 3.4, where it can be observed that, the Mozambique GDP deflator and South Africa real GDP variables both individually and together are helping to improve the forecast of Mozambique real GDP, that is, we reject the null hypothesis that there is no Granger-causality between Mozambique GDP deflator, South Africa real GDP variables and Mozambique real GDP. While for Mozambique GDP deflator and South Africa real GDP equations, because the p-value is greater then 0.05, we cannot reject the null hypothesis.

Table 3.4: Granger - causality Wald statistics

Equation variable	Excluded variable	χ^2	df	p-value
Real MZGDP	MZGDP Deflator	5.74	1	0.0166
	Real ZAGDP	6.27	1	0.0123
	ALL	6.55	2	0.0378
MZGDP Deflator	Real MZGDP	0.18	1	0.6687
	Real ZAGDP	0.84	1	0.3582
	ALL	1.03	2	0.5987
Real ZAGDP	Real MZGDP	0.85	1	0.3564
	MZGDP Deflator	2.45	1	0.1173
	ALL	3.23	2	0.1994

As the VEC model proved to be adequate, we continued to structural vector error correction analysis, where we use the model to identify the supply shocks, demand shocks and external shocks related to the impulse response of shocks imposing restrictions on the matrix of contemporaneous effects and long-run effects of shocks. With restrictions on matrix $\mathbf{B}$, we assume that, the contemporaneous of aggregate supply shocks and external shocks do not affect GDP deflator variable. The two constraints placed on the long-run matrix are that aggregate demand shocks and external shocks do not have permanent effects on the variables, which means that only aggregate supply shock has permanent effects. After the identification process, we estimate SVEC B-model with short and long-run restrictions. The results indicate that the SVEC model is over-identified with LR statistic $\chi^2(1) = 2.2016$ with associated p - value of 0.1379, greater than 0.05, thus, the identification is valid.

We estimate the structural decomposition by computing the contemporaneously effects matrix $\mathbf{B}$, as well as the identified total long-run effects matrix $\mathbf{\Xi B}$. Based on bootstrap in particular Efron interval, we obtained the estimates of matrices of long-run and short-run, where the values of t-statistics are in brackets. As explained in the specification and identification of the model, with 2 cointegrated vectors, we have two columns with zero for long-run matrix as observed.

$$
\text{Short} - \text{run} = \begin{bmatrix} \underset{[3.7850]}{0.0160} & \underset{[0.6383]}{0.0039} & \underset{[2.4447]}{0.0094} \\ 0 & \underset{[4.1757]}{0.0029} & 0 \\ \underset{[-3.1299]}{-0.0062} & \underset{[3.2668]}{0.0092} & \underset{[2.7511]}{0.0034} \end{bmatrix} ; \quad \text{Long} - \text{run} = \begin{bmatrix} \underset{[2.4565]}{0.0113} & 0 & 0 \\ \underset{[2.5263]}{0.0048} & 0 & 0 \\ \underset{[2.4706]}{0.0042} & 0 & 0 \end{bmatrix}
$$

These estimates suggest that supply shocks have a significant, positive long-run effect on output, prices and South Africa real GDP, whereas the long-run effect of demand shocks and external shocks are restricted to zero with respect to output and

equal to zero for other variables. Regarding the contemporaneous impact of shocks, most coefficients are statistically significant, only the lagged value of demand shocks in the Mozambique real GDP equation is not significant[3].

From estimated matrices, we can observe that Mozambique real GDP and Mozambique GDP deflator variables are positively related with own and with other contemporaneous lagged shocks in the system, while the South Africa real GDP variable is negative affected by lagged contemporaneous of supply shocks and positive with contemporaneous demand shocks and own lagged shocks. In long-run impact matrix, it's clear that, each variable is only positively related with lagged of supply shocks.

To investigate the plausibility of the identification scheme on the structural VEC model and learn about the effects of the different shocks in variables, we consider two tasks, the impulse response functions and forecast error variance decomposition analysis, of the structural shocks. Therefore, in our analysis we investigate the impulse response functions, and the identification suggests they show concordance with the theory regarding the long-run relationships between the variables in terms of the direction and magnitude of the impact. The Figure 3.3 within a horizon of eight periods, shows the impulse response functions with 95% confidence intervals based on bootstrap replications of Mozambique real GDP (Output), Mozambique GDP deflator (Prices) and South Africa real GDP (ZA Output) for each of the identified structural shocks, supply, demand and external shocks, respectively.

The impulse responses in Figure 3.3 are consistent with theoretical priors, i.e., are in the expected direction. The aggregate supply shocks have a positive and permanent impact on Mozambique real GDP, while for Mozambique GDP deflator and South Africa real GDP the impact of Aggregate supply shocks is negative and transitory, as expected. Positive aggregate demand shocks have a positive and permanent impact on Mozambique real GDP, on Mozambique GDP deflator, as well as on South Africa real GDP. Positive external shocks have a transitory impact on Mozambique real GDP, that is, the shocks cause a temporary fluctuation in output around the third period, followed by the expansionary impact. While for Mozambique GDP deflator and South Africa real GDP the external shocks causes a permanent and positive impact.

Moving to the second task in analysing the SVEC model, we report in Table 3.5 the proportion of forecast error variance decomposition of Mozambique real GDP, Mozambique GDP deflator and South Africa real GDP account by the three shocks, i.e., the supply shocks (SS), demand shocks (DS) and external shocks (ES) over five horizons.

---

[3]Model parameters, impulse response functions and forecast error variance decomposition computed by JMulti.
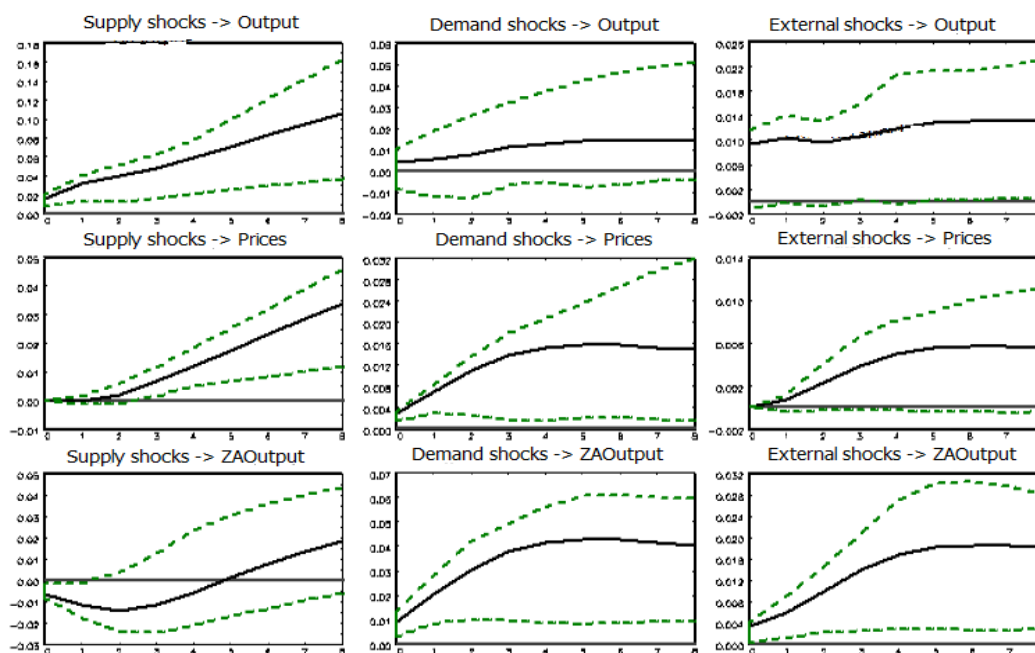
Figure 3.3: SVEC accumulated impulse response functions

Table 3.5: SVEC forecast error variance decomposition of real MZGDP, real MZGDP deflator and real ZAGDP

	Real MZGDP			MZGDP Deflator			Real ZAGDP		
lag	SS	DS	ES	SS	DS	ES	SS	DS	ES
1	71.0	4.0	25.0	00.0	100.0	0.0	28.0	63.0	09.0
2	83.0	3.0	14.0	00.0	98.0	2.0	24.0	70.0	06.0
4	84.0	4.0	12.0	30.0	63.0	7.0	17.0	73.0	10.0
6	87.0	4.0	09.0	59.0	36.0	5.0	27.0	76.0	10.0
8	90.0	3.0	07.0	71.0	26.0	3.0	35.0	56.0	09.0

On the variance decomposition for Mozambique real GDP, it may be noted that the long-run development of Mozambique real GDP is dominated by the history of supply shocks, start with a percentage of 71 at the first period until 90 percent at the end the eighth period. The long-run contribution of the demand shocks is very small ranging from 3 to 4 percent. Regarding Mozambique GDP deflator, the corresponding variance decomposition accounted by the long-run three structural shocks, show that the long-run development is totally dominated by the history of Mozambique demand shocks in the first period, followed by decreases over time until 26% at the eighth period. Second, the percentage forecast error variance decomposition due to supply shocks increases over time, starting with zero until 71% in the eighth period. For South Africa real GDP, the variance decomposition, in contrary to other past two situations, here the variance in long-run development is

dominated by the demand shocks and supply shocks, and small in proportion due to own variability about 10%.

3.6.5 Conclusions

In this application, we analyse the role of aggregate supply, aggregate demand and external shocks in an open economy. Specifically, we study the dynamic of the Mozambique real GDP with respect to real GDP deflator and external shocks represented by South Africa real GDP. The ADF test indicates that the three variables are stationary after first differences, and the Johansen test for cointegration indicates two cointegrating vectors where the first equation is composite by Mozambique real GDP and South African real GDP, the second pair is composite by Mozambique real GDP deflator and South Africa real GDP.

From the adjustment parameters, it can be said that the error correction mechanism for output and prices equations are opposite, since in the output equation, the adjustment to equilibrium deviations start with an negative feedback followed by an positive feedback, while in the prices equation, the adjustment start with an positive feedback followed by a negative feedback. In addition, the output equation have faster convergence to return to long-run equilibrium in cases of deviations from this, compared to prices and external activity equations. From long-run matrix, we notice that there are two transitory shocks and one permanent shock, and that all variables in the system are affected by the structural supply shock.

Also from our results, the impulse response functions display the expected direction, for illustration, the aggregate supply shocks have a positive and permanent impact on Mozambique real GDP, while for Mozambique real GDP deflator and South Africa real GDP, the impact of aggregate supply shocks is negative and transitory, as expected. With respect to forecasting error variance decomposition, we notice that the long-run development of real MZGDP is dominated by its own supply shocks, the variance of real MZGDP deflator is dominated also by own demand shocks. However, this predominance decreases over time giving rise to the influence of supply shocks. Regarding the variance of real ZAGDP, it is mainly attributed to demand and supply shocks.

Summarizing, we analyse the dynamic relations between three macroeconomic time series related to the Mozambique economy and conclude that, the Mozambique real GDP plays an important role in both cointegration relations, in the analysis of impulse response function as well as the forecast error variance decomposition. The knowledge of the relations among variables, allows the government to make decisions taking into account these relationships. The study contributes to the

analysis of macroeconomic data of the Mozambique economy in the presence of an external variable, which helps the process of defining policies that increase the rate of sustainable economic growth of Mozambique.

Multivariate GARCH models

4.1 Introduction to MGARCH models

Starting from Chapter 2, in the literature it's known that the ARCH model 2.7, the Generalized ARCH, equation 2.8 and their extensions in section 2.3, were developed for studying the volatility of univariate financial time series and are capable to capture particular characteristics, including asymmetry and high persistence in second moments. However, when the goal is to analyse the co-movement of volatilities of two or more financial time series, the univariate models are not adequate, in such cases, the multivariate GARCH (MGARCH) models are more appropriate.

According to Bauwens *et al.* [6], one of the applications of MGARCH models is the analysis of volatility relationship between several financial time series, or investigate the relationship among markets. Early articles on multivariate GARCH models include Engle *et al.* [45], Bollerslev *et al.* [20], Diebold and Nerlove [33]. The first Multivariate GARCH model were that of Bollerslev *et al.* [20], called VECH model, then many MGARCH models ware developed, such as BEKK-GARCH model of Engle and Kroner [43], the Dynamic Constant Correlation (DCC) GARCH model of Engle [40], the Varying Conditional Correlation (VCC) of Tse and Tsui [117], to understand and predict the co-movements of multivariate time series.

Let $\mathbf{r}_t$ denote a $K \times 1$ vector of log returns at time t. The multivariate GARCH model is defined as $\mathbf{r}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_i \mathbf{r}_{t-i} + \mathbf{u}_t$ with $\mathbf{u}_t = \mathbf{H}_t^{1/2} \mathbf{z}_t$ or $\mathbf{u}_t \sim N(0, \mathbf{H}_t)$ and $\mathbf{z}_t \sim N(0, \mathbf{I}_K)$.

According to Lütkepohl [77], the simpler multivariate VAR-ARCH(p) model can

81

be represented in the form:

$$vech(\mathbf{H}_t) = \mathbf{\Omega} + \mathbf{A}_1 vech(\mathbf{u}_{t-1}\mathbf{u}'_{t-1}) + ... + \mathbf{A}_p vech(\mathbf{u}_{t-p}\mathbf{u}'_{t-p}) \qquad (4.1)$$

where $vech(.)$ is an operator that stacks the columns of the lower triangular part of its argument square matrix and:

- $\boldsymbol{\mu}$ is a $K \times 1$ vector of constants;

- $\boldsymbol{\Omega}$ is a $(K(K+1)/2) \times 1$ vector of constants;

- $\boldsymbol{\Phi}_i$ and $\mathbf{A}_i$ are $K \times K$ matrices of parameters for mean and variance equations;

- $\mathbf{u}_t$ is a $K \times 1$ vector of error of K returns at time t, with $E[\mathbf{u}_t] = 0$, $cov[\mathbf{u}_t] = \mathbf{H}_t$;

- $\mathbf{H}_t$ is a $K \times K$ matrix of conditional variances and covariances of $\mathbf{u}_t$ at time t;

- $\mathbf{H}_t^{1/2}$ is any $K \times K$ positive definite matrix at time t such that $\mathbf{H}_t$ is the conditional variance matrix of $\mathbf{u}_t$; $\mathbf{H}_t^{1/2}$ may be obtained by a Choleski factorization of $\mathbf{H}_t$;

- $\mathbf{z}_t$ is a $K \times 1$ random vector of IID errors such that $E[\mathbf{z}_t] = 0$ and $var[\mathbf{z}_t] = \mathbf{I}_K$.

The parameter $\boldsymbol{\mu}$ is a vector of constants μ_i, which represents the constant in the conditional mean equation. For now, it is not important to address them. It is also assumed that the residuals $\mathbf{u}_t$ are not linearly correlated over time, but there has nonlinear serial dependence between residuals.

Following Bauwens *et al.* [6], the main problem in the MGARCH models is the size of the conditional covariance matrices $\mathbf{H}_t$, as it increases with the number of variables in the system. For a small number of variables (K small), this problem is not observed, becoming more evident when the number of variables increases, as in the estimation process is such that the matrix $\mathbf{H}_t$ must be inverted which makes the process less flexible making it difficult to obtain a parsimonious model as desired. Another important estimation requirement is that the matrix of covariances must be positive definite or invertible. In order to overcome this problem several specifications are proposed in the literature.

According to Silvennoinen and Terasvirta [108], the different specifications of MGARCH models can be divided into four categories: (i) models of the conditional covariance matrix, In which the matrix $\mathbf{H}_t$ is modelled directly, featuring the VECH and BEKK models; (ii) the factor models similar to the principal component analysis, the category includes, for example, Factor-ARCH and GO-ARCH; (iii) the models of conditional variances and correlations, where the matrix $\mathbf{H}_t$ is decomposed into

matrix of conditional correlation $\mathbf{R}_t$ and diagonal matrix of conditional variances $\mathbf{D}_t$; (iv) and finally, the nonparametric and semiparametric models, which is an alternative to parametric estimation of the conditional variance structure.

4.2 Conditional covariance MGARCH models

4.2.1 VECH model

In the class of Conditional covariance matrix, the first multivariate GARCH model was introduced by Bollerslev *et al.* [20], which is called VECH model. It is much general compared to the subsequent formulations. In the VECH model, every conditional variance and covariance is a function of all lagged conditional variances and covariances, as well as lagged squared residuals and cross-product of residuals and lagged values of the elements of $\mathbf{H}_t$. The VECH model can be express as:

$$vech(\mathbf{H}_t) = \mathbf{\Omega} + \sum_{i=1}^{p} \mathbf{A}_i vech(\mathbf{u}_{t-i}\mathbf{u}'_{t-i}) + \sum_{j=1}^{q} \mathbf{B}_j vech(\mathbf{H}_{t-j}) \tag{4.2}$$

where $\mathbf{H}_t$ is the covariance matrix of the residuals, $\mathbf{\Omega}$ is an $(K(K+1)/2) \times 1$ vector of constants , $\mathbf{A}_i$ and $\mathbf{B}_j$ are $K(K+1)/2 \times K(K+1)/2$ parameter matrices, $\mathbf{u}_{t-i}$ is an $K \times 1$ vector and K is the number of variables. Note that, from equation 4.2, it can be seen that VECH(p,q) model is multivariate extension of the GARCH(p,q) model, that is VAR-MGARCH(p,q).

The representation in 4.2 is very general and flexible but there is a disadvantage that only a sufficient condition for $\mathbf{H}_t$ to be positive definite for all t is not restrictive. Furthermore the number of parameters is $(p+q) \times (K(K+1)/2)^2 + K(K+1)/2$ which is excessively large as N increases. For example, if $p = q = 1$ and $K = 2$, the number of parameters is 21, if $K = 3$ it is 78. This may cause computational difficulties, which implies that in practice the VECH model is used only in the bivariate case.

4.2.2 Diagonal VECH model

To improve parsimony Bollerslev *et al.* [20], proposed diagonal VECH (p,q) or DVECH (p,q) model. Where the coefficient matrices $\mathbf{A}_i, i = 1, ..., p$ and $\mathbf{B}_j, j = 1, .., q$ are assumed diagonal. In this case, the conditional covariance between $u_{i,t}$ and $u_{j,t}$, that is $h_{ij,t}$ depends only on lagged cross-products of the two shocks involved and lagged values of the covariance itself. The DVECH(p,q) model is given by:

$$\mathbf{H}_t = \mathbf{\Omega} + \sum_{i=1}^{p} \mathbf{A}_i^* \odot (\mathbf{u}_{t-i}\mathbf{u}'_{t-i}) + \sum_{j=1}^{q} \mathbf{B}_j^* \odot \mathbf{H}_{t-j} \tag{4.3}$$

where $\odot$ denotes the Hadamard product, that is, element-by-element product; $\boldsymbol{\Omega}$, is the $K \times K$ symmetric matrix of constant, $\mathbf{A}_i^* = diag(vech(\mathbf{A}_i))$, $\mathbf{B}_j^* = diag(vech(\mathbf{B}_j))$ and $\mathbf{A}_i$, $\mathbf{B}_j$ are $K \times K$ symmetric matrices with elements α_{ij} and β_{ij} respectively. Each covariance equation in 4.3 can be specified in scalar form as: $h_{ij,t} = \omega_{ij} + \alpha_{ij}u_{i,t-1}u_{j,t-1} + \beta_{ij}h_{ij,t-1}$.

The matrix of residual covariances $\mathbf{H}_t$ is positive definite for all t and the number of parameters is reduced to $(p+q+1)(K(K+1)/2)$, as no interaction is allowed between the different conditional variances and covariances. For example, if $p = q = 1$ and $K = 2$ then the number of parameters for DVECH(1,1) model is 9, and if $K = 3$ it is 18.

According to Bauwens *et al.* [6], Engle and Kroner [43] and Ledoit *et al.* [72], assuming that $\mathbf{H}_t$ is positive definite, the DVECH(p,q) model from equation 4.3 is ensured to be covariance stationary if $w_{ii} > 0$, for $\forall i = 1, ..., K$. All elements of $\mathbf{A}_i, i = 1, ..., p$ and $\mathbf{B}_j, j = 1, ..., q$, all eigenvalues of $\sum_{i=1}^{p} \mathbf{A}_i + \sum_{j=1}^{q} \mathbf{B}_j$ are less than one in modulus.

4.2.3 BEKK model

According to Bauwens *et al.* [6], because it is difficult to guarantee the positivity of $\mathbf{H}_t$ in the VECH representation without imposing strong restrictions on the parameters, based on earlier work by Baba *et al.* [4], Engle and Kroner [43], propose a new parametrization for $\mathbf{H}_t$ and call BEKK (p,q) model, therefore, the acronym BEKK refers to the initial letter of the authors Baba, Engle, Kraft and Kroner, who developed the multivariate GARCH model in a preliminary version.

$$\mathbf{H}_t = \boldsymbol{\Omega}\boldsymbol{\Omega}' + \sum_{i=1}^{p}\sum_{k=1}^{K}\left(\mathbf{A}'_{ki}\mathbf{u}_{t-i}\mathbf{u}'_{t-i}\mathbf{A}_{ki}\right) + \sum_{j=1}^{q}\sum_{k=1}^{K}\left(\mathbf{B}'_{kj}\mathbf{H}_{t-j}\mathbf{B}_{kj}\right) \tag{4.4}$$

where $\boldsymbol{\Omega}$ is a lower triangular matrix with $(K(K+1)/2)$ parameters, $\mathbf{A}_i$ and $\mathbf{B}_j$ denote $K \times K$ matrices with K^2 parameters each. The BEKK model is covariance stationary if and only if the eigenvalues of $\sum_{i=1}^{p}\sum_{k=1}^{K}\mathbf{A}_{ki} \otimes \mathbf{A}_{ki} + \sum_{j=1}^{q}\sum_{k=1}^{K}\mathbf{B}_{kj} \otimes \mathbf{B}_{kj}$ are less than one in modulus, where $\otimes$ denotes the Kronecker product of two matrices.

Even restricting to the model of the first order, empirical applications often involve the highly simplification version. For example, when both $\mathbf{A}$ and $\mathbf{B}$ are assumed to be diagonal matrices, the model is called Diagonal BEKK, proposed by Bollerslev *et al.* [20]. The main advantage in Diagonal BEKK model is that the number of parameters to be estimated decrease and the new parametrization guarantees that $\mathbf{H}_t$ be positive definite, if $\boldsymbol{\Omega}\boldsymbol{\Omega}'$ is a positive definite matrix.

Note that for VECH, DVECH and BEKK models the conditional covariance matrices $\mathbf{H}_t$ are imposed to be positive definite for each t. These requires complicated restrictions on the off-diagonal elements.

4.3 Conditional correlations MGARCH models

Following Silvennoinen and Terasvirta [108], the models in the class of Conditional Correlations (CC), are based on the decomposition of the conditional covariance matrix $\mathbf{H}_t$, into a matrix of conditional correlations $\mathbf{R}_t$, and a diagonal matrix of conditional variances $\mathbf{D}_t$, that is:

$$\mathbf{r}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_i \mathbf{r}_{t-i} + \mathbf{u}_t \quad ; \quad \mathbf{u}_t = \mathbf{H}_t^{1/2} z_t \quad \text{where} \quad \mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t \qquad (4.5)$$

The class of CC-MGARCH models are considered to be the simple in terms of representation and estimation, as the number of parameters to be estimate increases gradually as the number of variables increases. In this category the covariance between variables $h_{ij,t} = \rho_{ij,t}\sqrt{h_{ii,t} \times h_{jj}}$, where the variances h_{ii} and h_{jj} are obtained by a univariate GARCH process discussed in chapter 2. The correlations ρ_{ij} show how the series of residuals move together. The CC - MGARCH models, include Constant Conditional Correlation (CCC), Dynamic Conditional Correlation (DCC) and Varying Conditional Correlation (VCC) models.

4.3.1 Constant conditional correlation

The Constant Conditional Correlation (CCC) GARCH model was proposed by Bollerslev [17], assuming that the conditional correlations between the elements of $\mathbf{u}_t$ are time-invariant. By definition a process $\mathbf{u}_t$ is called CCC - GARCH(p,q) model if the matrix of variances $\mathbf{D}_t = diag(h_{11,t}^{1/2}, ..., h_{kk,t}^{1/2})$, and the correlation matrix $\mathbf{R} = \rho_{ij}$ is symmetric positive definite with $\rho_{ii} = 1$, for $\forall i$.

$$\mathbf{H}_t = \boldsymbol{\Omega} + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{u}_{t-i}^2 + \sum_{j=1}^{q} \mathbf{B}_j \mathbf{H}_{t-j} \qquad (4.6)$$

where $\boldsymbol{\Omega}$ is a $K \times 1$ vector with positive coefficients $\mathbf{A}_i$ and $\mathbf{B}_j$ are $K \times K$ matrices with non-negative coefficients.

The individual conditional variances and conditional covariances are:

$$\begin{aligned} h_{ii,t} &= \omega_{ii} + \alpha_{ii} u_{ii,t-1}^2 + \beta_{ii} h_{ii,t-1} \quad ; \quad i = 1, ..., K \\ h_{ij,t} &= \rho_{ij}\sqrt{h_{ii,t} h_{jj,t}} \quad ; \quad \text{for} \quad i \neq j \end{aligned} \qquad (4.7)$$

85

4.3.2   Dynamic conditional correlation

The assumption that the conditional correlations are constant may seem unrealistic in many empirical applications. Engle [40] and Tse and Tsui [117] propose a generalization of the CCC model of Bollerslev [17] by making the conditional correlation matrix time-dependent. The model is then called a Dynamic Conditional Correlation (DCC) model. Here we will just focus on DCC-GARCH model proposed by Engle [40], where the matrix of variances $\mathbf{D}_t$ are diagonal with variances h_{ii} described by a univariate GARCH(1,1) process.

$$\mathbf{D}_t = diag(h_{11,t}^{1/2},...,h_{kk,t}^{1/2}) \quad \text{where} \quad h_{ii,t} = \omega_{ii} + \alpha_{ii}u_{ii,t-1}^2 + \beta_{ii}h_{ii,t-1} \qquad (4.8)$$

From equation 4.5, where $\mathbf{r}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_i\mathbf{r}_{t-i} + \mathbf{u}_t$, $\mathbf{u}_t = \mathbf{H}_t^{1/2}\mathbf{z}_t$ and $\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t$, The process $\mathbf{u}_t$ is called DCC - GARCH model if the conditional correlation matrix $\mathbf{R}_t$ is defined as:

$$\mathbf{R}_t = diag(q_{11,t}^{1/2},...,q_{kk,t}^{1/2})\mathbf{Q}_t diag(q_{11,t}^{1/2},...,q_{kk,t}^{1/2}) \qquad (4.9)$$

and $\mathbf{Q}_t = (q_{ii,t})$ is the $K \times K$ symmetric positive definite matrix which has the form:

$$\mathbf{Q}_t = (1 - \lambda_1 - \lambda_2)\overline{\mathbf{Q}} + \lambda_1 z_{t-1}z'_{t-1} + \lambda_2\mathbf{Q}_{t-1} \qquad (4.10)$$

note that $z_{i,t}$ is a standardized residuals, λ_1 and λ_2 are adjustment parameter, they are non-negative and satisfy $\lambda_1 + \lambda_2 < 1$ and $\overline{\mathbf{Q}} = Cov(z_t z'_t)$ is the unconditional variance matrix of standardized residuals.

According to Bauwens *et al.* [6], the main drawback of the model is that all conditional correlations follow the same dynamic structure. The number of parameters to be estimated is given by $(K + 1)(K + 4)/2$, relatively smaller compared to the number obtained in the BEKK model for cases where K is small. When the number of variables increases to continue to have parsimonious results and decrease the complexity in estimation DCC parameters, the process is performed in two steps.

4.3.3   Varying Conditional Correlation

The Varying Conditional Correlation (VCC) model proposed by Tse and Tsui [117] uses a nonlinear combination of univariate GARCH model with time-varying cross equation weights to model the conditional covariance matrix of errors. The VCC - GARCH model decomposes the conditional variance and covariances $\mathbf{H}_t$ into conditional variance matrix $\mathbf{D}_t$ and the conditional correlation matrix $\mathbf{R}_t$, where $\mathbf{D}_t = diag(h_{11,t}^{1/2},...,h_{kk,t}^{1/2})$. The variance $h_{ii,t}$ follows a univariate GARCH(1,1) model, $h_{ii,t} = \omega_{ii} + \alpha_{ii}u_{ii,t-1}^2 + \beta_{ii}h_{ii,t-1}$ and the conditional covariance following a dynamic

process, $h_{ij,t} = \rho_{ij,t}\sqrt{h_{ii,t}h_{jj,t}}$. Because the correlations $\rho_{ij,t}$ vary with time, this model is known as the VCC - GARCH model.

The time varying conditional correlation matrix $\mathbf{R}_t$ is generated by the equation:

$$\mathbf{R}_t = (1 - \lambda_1 - \lambda_2)\mathbf{R} + \lambda_1 \mathbf{R}_{t-1} + \lambda_2 \mathbf{\Psi}_{t-1} \tag{4.11}$$

where $\mathbf{R} = \{\rho_{ij}\}$ is a $K \times K$ positive definite parameter matrix with $\rho_{ii} = 1$ and $\mathbf{\Psi}_{t-1}$ is a $K \times K$ matrix whose elements are functions of the lagged observations of $\mathbf{u}_t$. The parameters λ_1 and λ_2 are assumed to be non-negative with the additional constraint that $\lambda_1 + \lambda_2 < 1$. Thus, $\mathbf{R}_t$ is a weighted average of $\mathbf{R}$, $\mathbf{R}_{t-1}$ and $\mathbf{\Psi}_{t-1}$. Hence, if $\mathbf{\Psi}_{t-1}$ is a well-defined correlation matrix, that is, positive definite with unit diagonal elements, $\mathbf{R}_t$ will also be a well-defined correlation matrix.

The sample correlation matrix of $\{u_{t-1}, ..., u_{t-m}\}$, denoted by $\mathbf{\Psi}_t = \{\psi_{ij,t}\}$, is computed as follows:

$$\psi_{ij,t} = \frac{\sum_{k=1}^{m} u_{i,t-k}u_{j,t-k}}{\sqrt{(\sum_{k=1}^{m} u_{i,t-k}^2)(\sum_{k=1}^{m} u_{j,t-k}^2)}} \quad ; \quad 1 \le i < j \le K \tag{4.12}$$

Note that, for DCC or VCC, if $\lambda_1 = \lambda_2 = 0$, the model reduces to CCC model.

4.4 Estimation of multivariate GARCH models

4.4.1 Testing multivariate residual behavior

According to Bauwens *et al.* [6], before an MGARCH model is fitted it is necessary to ensure that the series are stationary. For these proposes its common to use the Augmented Dickey and Fuller test or another one such as Phillips and Perron test or KPSS[1] stationarity test. Its necessary also to check whether the residuals present evidence of univariate and multivariate ARCH effects and how they behave compared with the Normal distribution, as well as also important to check the adequacy of the MGARCH model specification after estimation.

Because the univariate tests are considered in chapter 2, section 2.1, in this chapter, we focus on the multivariate tests described in chapter 3, section 3.3, which include the multivariate Ljung - Box test proposed by Hosking [61], which is a generalization of the univariate Ljung - Box Portmanteau (or Q) test for a white noise. This diagnostic check is a first step, required to ensure the consistent estimation of the mean process and thereby obtain the appropriately estimated error terms. Next it is applied the multivariate Lagrange multiplier test for ARCH effects in joint residuals and the multivariate normality test, which is a multivariate version

---

[1]Due to Kwiatkowski, Philips, Schmidt and Shin [71].

of Jarque and Bera [62] normality test. In addition to the formal testing, three graphical representations are also used to verify the adequacy of the model. One plot for each equation includes a residual plot, an empirical distribution plot of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the residuals and their squares.

Because in the chapters before 1-3, the multivariate ARCH effect test was not discussed, it is now appropriate to present this test.

Multivariate test for Heteroskedasticity

According to Engle [39] and Lütkepohl [77], the Multivariate Lagrange Multiplier (MLM) test is a standard tool to detect multivariate ARCH effect. Lets consider $\mathbf{u}_t = (u_{1t}, ..., u_{kt})$ a vector of residuals from a fitted VAR. The multivariate *LM* test for ARCH is based on the following auxiliary regression:

$$vech(\mathbf{u}_t \mathbf{u}_t') = \boldsymbol{\beta}_0 + \mathbf{B}_1 vech(\mathbf{u}_{t-1} \mathbf{u}_{t-1}') + ... + \mathbf{B}_q vech(\mathbf{u}_{t-q} \mathbf{u}_{t-q}') + \mathbf{v}_t \qquad (4.13)$$

where $\mathbf{v}_t$ assigns a spherical error process and *vech* is the columns stacking operator for symmetric matrices that stacks the columns from the main diagonal.

The dimension of $\boldsymbol{\beta}_0$ is $\frac{1}{2}K(K+1)$ and for matrices $\mathbf{B}_i$ is $\frac{1}{2}K(K+1) \times \frac{1}{2}K(K+1)$. The null hypothesis is $H_0 : \mathbf{B}_1 = ... = \mathbf{B}_q = 0$ against $H_1 : \exists \mathbf{B}_i \neq 0$ for $i = 1, ..., q$. Denoting $\widehat{\boldsymbol{\Omega}}_{vech}$ the resulting residual covariance matrix and corresponding matrix obtained for *p=0* by $\widehat{\boldsymbol{\Omega}}_0$, the test statistics is defined as:

$$MLM_{ARCH}(q) = \frac{1}{2}TK(K+1) - Ttr(\widehat{\boldsymbol{\Omega}}_{vech}\widehat{\boldsymbol{\Omega}}_0^{-1}) \qquad (4.14)$$

Under the null hypothesis, the *MLM*-statistic is asymptotically distributed as $\chi^2(qK^2(K+1)^2/4)$, where K is the number of series of residuals, q is the number of lags in auxiliary regression. The null hypothesis is rejected if $MLM(q) > \chi^2_\alpha(.)$, which means that, at least one series of residual has ARCH effect. The *MLM* test is also useful to check for remaining ARCH in the residual of a fitted GARCH model. In that case, the residual $\mathbf{u}_t = (u_{1t}, ..., u_{kt})$ is replaced by standardized residuals $\mathbf{z}_t = (z_{1t}, ..., z_{kt})$ and each $z_{it} = u_{it}/\sqrt{h_{it}}$.

4.4.2 Maximum likelihood estimation for MGARCH models

The method of maximum likelihood estimation (MLE) is the most common approach used in the estimation of the vector of unknown parameters θ of both univariate and multivariate GARCH models, under the assumption of independently and identically distributed normally standardized residuals. However, the assumption of normality

of the residuals is often violated and difficult to justify in many financial data. This has motivated the use of alternative parametric distributional assumptions such as the standardized Student t distribution with $v \geq 2$ degrees of freedom or the generalized error distribution. An alternative method to the MLE based on the Normal density is the quasi-maximum likelihood estimator proposed by Bollerslev and Wooldridge [19]. Because our application is related to the conditional variance and correlation GARCH models, we describe de MLE only for these category of multivariate GARCH models.

From section 4.3, the conditional variance and correlation models are defined as $\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$ for $\mathbf{u}_t = \mathbf{H}_t^{1/2} \mathbf{z}_t$ and $\mathbf{z}_t \overset{IID}{\sim} N(0, \mathbf{I}_K)$.

When the standardized errors $\mathbf{z}_t$, are multivariate Gaussian distributed, the joint distribution of $\mathbf{z}_t = (z_{1t}, z_{2t}, ..., z_{kt})$, with $E(z_t) = 0$ and $E(z_t' z_t) = I$, for $t = 1, ..., T$, is given by

$$f(\mathbf{z}_t) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{K/2}} \exp \left\{ -\frac{1}{2} \mathbf{z}_t' \mathbf{z}_t \right\} \tag{4.15}$$

The corresponding likelihood function for $\mathbf{z}_t = \mathbf{H}_t^{-1/2} \mathbf{u}_t$ is:

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{K/2} |\mathbf{H}_t|^{1/2}} \exp \left\{ -\frac{1}{2} \mathbf{u}_t' |\mathbf{H}_t|^{-1} \mathbf{u}_t \right\}$$

Taking logarithm of the likelihood function where $\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$ and using a linear transformation we get the log-likelihood.

$$\begin{aligned}
\log(L(\boldsymbol{\theta})) &= -\frac{1}{2} \sum_{t=1}^{T} \left(K \log(2\pi) + \log(|\mathbf{H_t}|) + \mathbf{u}_t' \mathbf{H}_t^{-1} \mathbf{u}_t \right) \\
&= -\frac{1}{2} \sum_{t=1}^{T} \left(K \log(2\pi) + \log(|\mathbf{D}_t \mathbf{R}_t \mathbf{D}_t|) + \mathbf{u}_t' \mathbf{D}_t^{-1} \mathbf{R}_t^{-1} \mathbf{D}_t^{-1} \mathbf{u}_t \right) \quad (4.16) \\
&= -\frac{1}{2} \sum_{t=1}^{T} \left(K \log(2\pi) + 2\log(|\mathbf{D}_t|) + \log(|\mathbf{R}_t|) + \mathbf{u}_t' \mathbf{D}_t^{-1} \mathbf{R}_t^{-1} \mathbf{D}_t^{-1} \mathbf{u}_t \right)
\end{aligned}$$

According to Engle and Sheppard [47], the estimation of the log-likelihood for the DCC model is made in two steps. In the first step $\mathbf{R}_t$ is replaced by the identity matrix $\mathbf{I}_K$, which results in the quasi-likelihood function:

$$\log(\mathbf{L_1}(\phi)) = \sum_{i=1}^{K} \left(-\frac{1}{2} \sum_{t=1}^{T} \left[\log(h_{it}) + \frac{u_{it}^2}{h_{it}} \right] + \text{constant} \right) \tag{4.17}$$

From this first step, the parameter vector $\phi = (\phi_1, ..., \phi_K)$ is estimated. When ϕ is estimated, the conditional variance h_{it} is also estimated for each serie, $\mathbf{z}_t = \mathbf{D}_t^{-1/2} \mathbf{u}_t$ and then $\overline{\mathbf{Q}} = E(z_t z_t')$ can be also estimated. After the first step only λ_1 and λ_2 are

unknown. These parameters are estimated in the second step using the correctly specified log-likelihood and the quasi - likelihood function given by:

$$\log(\mathbf{L_2}(\psi)) = -\frac{1}{2}\sum_{t=1}^{T}\Big(K\log(2\pi) + 2\log(|\mathbf{D}_t|) + \log(|\mathbf{R}_t|) + z_t'\mathbf{R}_t^{-1}z_t\Big)$$

Since $\mathbf{D}_t$ is constant when conditioning on the parameters from step one, we can exclude the constant terms and maximize:

$$\log(\mathbf{L_2^*}(\psi)) = -\frac{1}{2}\sum_{t=1}^{T}\Big(\log(|\mathbf{R}_t|) + z_t'\mathbf{R}_t^{-1}z_t\Big) \tag{4.18}$$

A full maximum likelihood estimator is consistent as well as the two-step procedure estimator, providing very similar results.

4.4.3 Forecast and evaluation MGARCH models

After fitting the MGARCH model, one might be interested in determining the forecast of the conditional covariance matrix $\mathbf{H}_{t+k}$. As the representation of the conditional covariance matrix $\mathbf{H}_t$ varies with the model specified, the forecast will depend on the underlying structure. Forecasting the covariance matrix $\mathbf{H}_{t+h} = \mathbf{D}_{t+h}\mathbf{R}_{t+h}\mathbf{D}_{t+h}$ using conditional correlation models, namely DCC and VCC, requires a two-step procedure. In the first step, the forecast is for conditional variance matrix $\mathbf{D}_{t+h}$ and in the second step is for the conditional correlation matrix $\mathbf{R}_{t+h}$.

According to Engle and Sheppard [47], the h-step ahead forecast of $\mathbf{D}_{t+h}$ of a standard GARCH(1,1) is given by:

$$E(\mathbf{D}_{t+h}|\mathscr{F}_t) = diag(\sqrt{E(h_{1,t+h}|\mathscr{F}_t)},...,\sqrt{E(h_{k,t+h}|\mathscr{F}_t)}) \tag{4.19}$$

where

$E(h_{i,t+h}|\mathscr{F}_t) = \sum_{i=1}^{h-2}\omega(\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)E(h_{i,t+1}|\mathscr{F}_t)$ and
$E(h_{i,t+1}|\mathscr{F}_t) = \omega + \alpha_1 u_{i,t}^2 + \beta_1 h_{i,t}$.
For nonlinear processes,
$\mathbf{Q}_{t+h} = (1 - \lambda_1 - \lambda_2)\overline{\mathbf{Q}} + \lambda_1(z_{t+h-1}z_{t+h-1}') + \lambda_2\mathbf{Q}_{t+h-1}$, we have
$E(z_{t+h-1}z_{t+h-1}'|\mathscr{F}_t) = E(\mathbf{R}_{t+h-1}|\mathscr{F}_t) = E(\mathbf{Q}_{t+h-1}^{*-1}\mathbf{Q}_{t+h-1}\mathbf{Q}_{t+h-1}^{*-1}|\mathscr{F}_t)$.

Since $E(\mathbf{Q}_{t+h-1}^{*-1}\mathbf{Q}_{t+h-1}\mathbf{Q}_{t+h-1}^{*-1}|\mathscr{F}_t)$ is unknown, the h-step ahead forecast of the correlation cannot be directly obtained. However, there exist two methods that approximate this forecast.

The first method assumes the approximation $E(z_{t+1}z_{t+1}'|\mathscr{F}_t) \approx \mathbf{Q}_{t+1}$ in order to have h-step ahead forecast of Q, that is:

$$E(\mathbf{Q}_{t+h}|\mathscr{F}_t) = \sum_{i=0}^{h-2}(1 - \lambda_1 - \lambda_2)\overline{\mathbf{Q}}(\lambda_1 + \lambda_2)^i + (\lambda_1 + \lambda_2)^{h-1}\mathbf{Q}_{t+1}$$

and $\mathbf{R}_{t+h} = \mathbf{Q}_{t+h}^{*-1}\mathbf{Q}_{t+h}\mathbf{Q}_{t+h}^{*-1}$, where $\mathbf{Q}_{t+h}^*$ is a diagonal matrix with the square root of the diagonal elements of $\mathbf{Q}_{t+h}$.

The second method assumes $\overline{\mathbf{Q}} \approx \overline{\mathbf{R}}$ and $E(\mathbf{Q}_{t+1}|\mathcal{F}_t) = E(\mathbf{R}_{t+1}|\mathcal{F}_t)$. Using this approximation, we can forecast $\mathbf{R}_{t+h}$ directly using the relationship:

$$E(\mathbf{R}_{t+h}|\mathcal{F}_t) = \sum_{i=0}^{h-2}(1-\lambda_1-\lambda_2)\overline{\mathbf{R}}(\lambda_1+\lambda_2)^i + (\lambda_1+\lambda_2)^{h-1}\mathbf{R}_{t+1} \quad \text{and} \quad (4.20)$$

$$E(\mathbf{H}_{t+h}|\mathcal{F}_t) = E(\mathbf{D}_{t+h}|\mathcal{F}_t)E(\mathbf{R}_{t+h}|\mathcal{F}_t)E(\mathbf{D}_{t+h}|\mathcal{F}_t) = \mathbf{H}_{t+h} = \mathbf{D}_{t+h}\mathbf{R}_{t+h}\mathbf{D}_{t+h}$$
(4.21)

Evaluation the model. We conduct the evaluation of model performance by assessing in-sample forecast accuracy measure either by Mean Absolute Error (MAE) or by Root Mean Squared Errors (RMSE). In particular, upon deciding the model as to constant versus non-constant correlation, we inspect the performance of the model over the full sample and in each sub-sample regarding the variance, the covariance and the correlation forecasts.

$$MAE_{ij} = \frac{1}{T}\sum_{k=1}^{T}|\rho_{ik} - \hat{\rho}_{ik}| \quad ; \quad RMSE_{ij} = \sqrt{\frac{1}{T}\sum_{k=1}^{T}(\rho_{ik} - \hat{\rho}_{ik})^2} \qquad (4.22)$$

where we use $\rho_{ij,t} = cor(r_{it}, r_{jt})$ as the proxy for realized correlation. If the evaluation regards the performance of variance or covariance forecast, we use r_{it}^2 as the proxy for realized variance and $cov_{ij,t} = cov(r_{it}, r_{jt})$ as the proxy for realized covariance, and $\hat{\rho}_{ij}$, $\hat{h}_{ii}$ and $c\hat{o}v_{ij,t}$ are the estimated conditional correlation, variance and covariance, respectively.

4.5 Application for multivariate GARCH models

For multivariate GARCH model application, we inspect the eventual empirical relation between exchange rates and commodity prices across alternative exchange rates behavior, that its appreciation, depreciation and stability. Namely, we apply the Constant Conditional Correlation GARCH model of Bollerslev [17], the Dynamic Conditional Correlation GARCH model of Engle [40] and the Varying Conditional Correlation GARCH model of Tse and Tsui [117], to Mozambique related data.

It should be noted that this application is an extension of the results presented in the XI workshop on Statistics, Mathematics and Computation in Portalegre 2017, whose final article was submitted for publication, Mulenga *et al.* [86].

4.5.1   Data description

Mozambique is an emerging economy where resources are little explored. The country is economically linked to neighbor countries from workforce to the development of infrastructure such as port, roads and airports which are shared. The South African economy is linked to the mining resources, and Coal measured in Great Britain Pound is one of them. On the other hand, Mozambique has a growing natural resources activity, which actually raised the discussion on Dutch disease. Altogether, the eventual link among exchange rates and Coal is inspected, in particular considering different periods distinguished by specific behavior of the exchange rate, namely, appreciation, depreciation and stability.

For this study we use a multivariate GARCH model with three variables to analyse the co-movement of volatilities and relationship between the Coal of Africa (CZA), collected from finance.yahoo.com[2], the Mozambican New Metical (MZN) against South African Rand (ZAR) and Great Britain Pound (GBP), collected from OANDA.FX rates[3].

The daily time series span from January 4, 2010 to December 30, 2014. Note that holidays and weekends were removed, remaining a sample of 1294 observations.

The Figure 4.1, shows the behavior of CZA prices and the exchange rates. The graph of CZA displays a continuing decline, the MZN/ZAR exchange rate suggests four periods, the period of depreciation 2010, appreciation 2011, followed by stability in 2012. Between 2013-2014 the exchange rate exhibits stability with some appreciation. For MZN/GBP exchange rate we observe three periods, depreciation in 2010, appreciation in 2011 and depreciation with relative appreciation between 2012-2014.
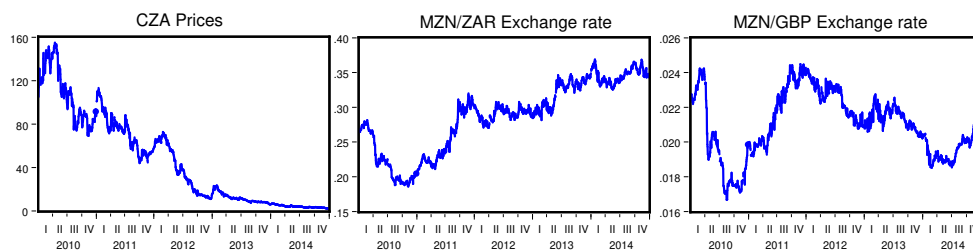

Figure 4.1: Behavior of CZA prices, MZN/ZAR and MZN/GBP exchange rates, period 2010 - 2014

As usual the series are not stationary, therefore, we transform them into the corresponding returns using the formula $r_{it} = [ln(y_{it}) - ln(y_{it-1})] \times 100$. Then

---

[2]http://finance.yahoo.com/q/hp?=CZA.L.

[3]http://www.oanda.com/currency/historical-rates.

the returns, although stationary where confirmed by Augmented Dickey-Fuller test discussed in section 2.1, exhibit some persistence and the residuals from the mean process show deviations from normality and the presence of ARCH effects. After fitting a VAR(1) model and obtaining the corresponding residuals, the results from the multivariate check are presented in Table 4.1.

Table 4.1: Multivariate tests for residuals of the returns, p-value in brackets

Statistic	2010-2014	2010	2011	2012	2013-2014
MLB Q(12)	137.37 [0.0000]	159.62 [0.0000]	192.83 [0.0000]	215.77 [0.0000]	339.37 [0.0000]
MJB Stat	2238.17 [0.0000]	604.64 [0.0000]	153.56 [0.0000]	055.91 [0.0000]	273.83 [0.0000]
MLM-ARCH	476.77 [0.0000]	181.73 [0.1376]*	338.23 [0.0000]	192.61 [0.0510]*	303.36 [0.0000]

* The test statistics show ARCH effects for 14% and 6%. However, results equation-by-equation indicate that some ARCH effects may be present.

In particular, the multivariate Ljung - Box test for autocorrelation rejected the null hypothesis of no autocorrelation up to 12 lags, for all sub-samples. However, on an equation-by-equation basis the autocorrelation functions are well-behaved, i.e., largely inside the 95 per cent bands. So, for the sake of a parsimonious model, a VAR(1) is employed. The MJB statistic showed that the joint residuals are not multivariate normal. Finally, the test for ARCH effects failed to reject the null hypothesis for 2010 and 2012, but rejected for the whole sample and for the remaining periods. That is, the joint residuals have ARCH effects, so we can setup the MGARCH models.

4.5.2 Results for conditional correlation MGARCH models

Our multivariate GARCH modelling is concentrated on correlation, possibly varying, among the above series. Namely, the models estimated are the CCC, DCC and VCC, using VAR(1)-MGARCH(1,1). The conditional distribution of the error term was assumed to be multivariate normal. Starting with GARCH(1,1) properties, all α_i and β_j coefficients are positive and in most pairs $\beta > \alpha$ as it was expected for GARCH model. In terms of significance only four of 27 parameters are not statistically significant at 5%.

Regarding to the GARCH model properties, we can see from Table 4.2, all models has α close to 0.05, indicating that the markets are relatively stable. The persistence of exchange rates are higher than that of coal price in DCC and VCC models. But when using CCC levels of persistence are much closer, which indicate

that the changes in the exchange rate are related to the changes of time. In other side comparing the series, the coal prices have lower persistence in volatility.

In general, the univariate GARCH(1,1) exhibits high persistence, keeping $\alpha + \beta < 1$. In Table 4.2, as an example, we present estimates for CCC, DCC and VCC models, for the period 2010-2014.

Table 4.2: Estimates for DCC, VCC and CCC - GARCH(1,1) models, for period 2010-2014, p-value in parenthesis

Return from	ω	α	β	$\alpha + \beta$
		DCC model		
CZA	3.9005 (0.0020)	0.0601 (0.0030)	0.7016 (0.0000)	0.7617
MZN/ZAR	0.0400 (0.0040)	0.0549 (0.0000)	0.9146 (0.0000)	0.9695
MZN/GBP	0.0453 (0.0000)	0.0568 (0.0000)	0.9010 (0.0000)	0.9578
		VCC model		
CZA	3.6944 (0.0040)	0.0536 (0.0040)	0.7175 (0.0080)	0.7711
MZN/ZAR	0.0284 (0.0050)	0.0410 (0.0000)	0.9365 (0.0080)	0.9775
MZN/GBP	0.0381 (0.0000)	0.0430 (0.0000)	0.9187 (0.0000)	0.9617
		CCC model		
CZA	3.6995 (0.0040)	0.0525 (0.0050)	0.9185 (0.0000)	0.7710
MZN/ZAR	0.0313 (0.0090)	0.0342 (0.0000)	0.9405 (0.0000)	0.9745
MZN/GBP	0.0456 (0.0010)	0.0380 (0.0000)	0.9153 (0.0000)	0.9533

The estimated conditional variance, shown in Figure 4.2, assumes a time dependent conditional correlation matrix (DCC) besides the case where it is time invariant (CCC). From the depicted behavior we can observe that the volatility of Coal prices is different from that of exchange rates.

Note that the differences of estimated conditional variance are in the behavior and the scale selected, in the two models. Still in Figure 4.2, exchange rates volatility are similar, in both series. And it can be observed that the behavior of the conditional variance estimated for the CCC model is similar for all three variables considered. This similarity can be justified by the assumption that the correlation between the residuals is time invariant imposed.

To verify the eventual relationship between commodity prices and the exchange rates, we use the estimated conditional covariance between Coal of Africa prices and MZN/GBP exchange rate using both the time invariant CCC model and the VCC model with a variable matrix of conditional correlation. The Figure 4.3, illustrates the fluctuations in co-movement of volatility for the periods 2010, 2011 and 2012, corresponding to depreciation, appreciation and stability respectively of the
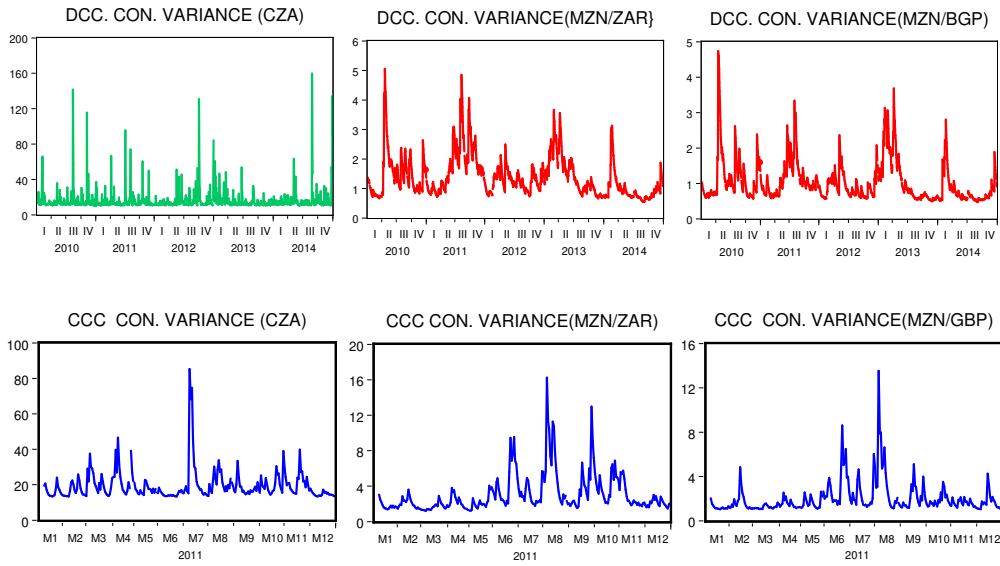
Figure 4.2: DCC and CCC conditional variance for full sample and the year 2011

MZN/ZAR exchange rates. Taking as an example the VCC model over these three periods, it is observed that the conditional covariance behavior changes from one sub-period to another, in addition to being associated with different magnitude of scale.

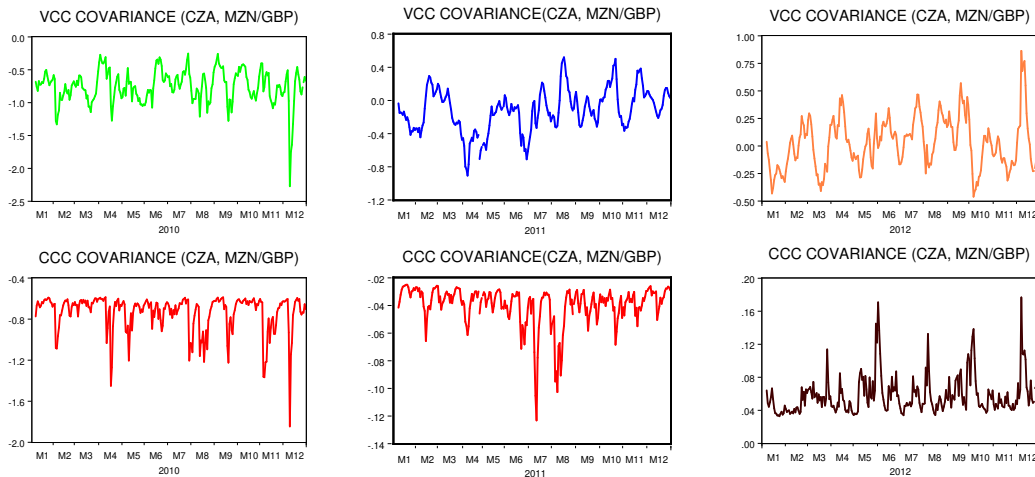Figure 4.3: VCC and CCC conditional covariance between CZA and MZN/GBP exchange rate

For VCC model, in the depreciation period, the conditional covariance is negative, for appreciation and stability periods it is around zero with some positive observations. However, for CCC the estimated conditional covariance for depreciation and appreciation period is negative, and different from that of stability period

95

which is positive. This leads us to say that the link between commodity prices and exchange rates is not always the same, particularly across different marked exchange rate behavior.

Considering the conditional covariance from DCC model, the Figure 4.4 show that, the graphics for pairs between CZA prices and exchange rates are similar with oscillation around zero, while the estimated covariance between exchange rates display different behavior with only positive values in sample where the highest peaks occur in the first two quarters of 2013 followed by values less then 3 for subsequent quarters.
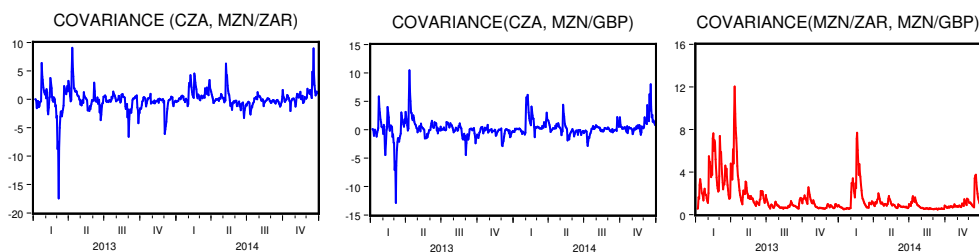
Figure 4.4: DCC conditional covariance between CZA, MZN/ZAR and MZN/GBP series, 2013-2014

Regarding the adjustment parameter, they are positive in all cases, statistically significant at 5%, and the constraint $\lambda_1 + \lambda_2 < 1$ is satisfied. The Wald test for the null hypothesis of $\lambda_1 = \lambda_2 = 0$ in the DCC and VCC leads to reject the null hypothesis. That is, the parameters are jointly significant, therefore, the DCC and VCC models does not reduce to CCC model. Hence, the assumption of time-invariant conditional correlation assumed in the CCC model is too restrictive for the data.

In Figure 4.5, as an illustration of correlations behavior, the case of DCC model over 2010-2014 is presented.
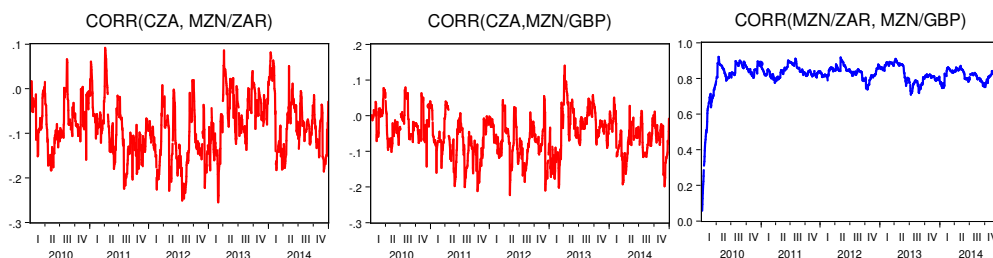

Figure 4.5: DCC conditional correlation between CZA, MZN/ZAR and MZN/GBP series, 2010-2014

The conditional correlation between CZA prices and the MZN/ZAR exchange rate is negative, weak and most of the time not significant at 5%. For CZA and

MZN/GBP exchange rate is negative with most cases non-significant, except for the period 2013-2014 where the correlation is positive, as shown in Table 4.3. Differently, the correlation between exchange rates is positive, strong and significant at any level of significance.

Table 4.3: Pairwise conditional correlation from DCC, VCC and CCC - GARCH(1,1) models

Period	Model	(CZA, MZAR)	(CZA, MGBP)	(MZAR, MGBP)
2010-2014	DCC	-0.0805 (0.0600)	-0.0023 (0.9560)	0.8691 (0.0000)
Full sample	VCC	-0.0954 (0.0040)	-0.0030 (0.9290)	0.8815 (0.0000)
	CCC	-0.0920 (0.0010)	-0.0029 (0.9160)	0.8566 (0.0000)
2010	DCC	-0.2070 (0.0230)	-0.1517 (0.0350)	0.9059 (0.0000)
Appreciation	VCC	-0.2378 (0.0010)	-0.1446 (0.0530)	0.8947 (0.0000)
	CCC	-0.1994 (0.0010)	-0.1109 (0.0074)	0.8827 (0.0000)
2011	DCC	-0.1308 (0.0810)	-0.0203 (0.7900)	0.8737 (0.0000)
Depreciation	VCC	-0.1413 (0.0560)	-0.0254 (0.7360)	0.8875 (0.0000)
	CCC	-0.1451 (0.0200)	-0.0400 (0.5310)	0.8805 (0.0000)
2012	DCC	-0.1376 (0.0610)	0.0226 (0.7630)	0.8827 (0.0000)
Stability	VCC	-0.1358 (0.0510)	-0.0010 (0.9920)	0.8956 (0.0000)
	CCC	-0.1352 (0.0290)	0.0041 (0.9480)	0.8883 (0.0000)
2013-2014	DCC	-0.0055 (0.9400)	0.0159 (0.8240)	0.8685 (0.0000)
Stability -	VCC	-0.0128 (0.8280)	0.0298 (0.6130)	0.8834 (0.0000)
Appreciation	CCC	-0.0110 (0.8040)	0.0286 (0.5150)	0.8493 (0.0000)

p-value in parenthesis, MZAR=MZN/ZAR and MGBP=MZN/GBP exchange rate.

4.5.3 Forecast evaluation

For an assessment of model fit, applied to the DCC-GARCH (1,1) model, this subsection reports in-sample forecast performance computing mean absolute error for conditional variance and covariance, shown in Table 4.4. A comparison of the MAE points out that DCC model performs better in the forecast of conditional variance for the MZN/GBP exchange rate in all periods except in 2012, the period of stability. The best performance for conditional covariance was that of the covariance between CZA prices and MZN/GBP exchange rate.

Table 4.4: DCC mean absolute error for conditional variance and covariance predictions

Conditional variance	2010-2014	2010	2011	2012	2013-2014
var(CZA)	19.3620	42.4487	26.5154	31.9470	33.0938
var(MZN/ZAR)	1.5722	2.8948	4.6514	2.3620	2.6875
var(MZN/GBP)	1.3414	2.5558	2.4554	2.5046	2.4960
Cond. covariance	2010-2014	2010	2011	2012	2013-2014
cov(CZA,MZN/ZAR)	0.3299	1.3295	0.8879	0.9056	1.0705
cov(CZA,MZN/GBP)	0.0187	0.9170	0.4984	0.6061	0.9352
cov(MZAR,MGBP)*	1.2153	2.3998	3.4487	2.3505	2.2685

* MZAR, MGBP indicate MAN/ZAR and MZN/GBP exchange rate

4.5.4   Conclusions

We investigate the co-movement of three financial time series, the Coal of South Africa prices, MZN/ZAR and MZN/GBP exchange rates, using the CCC, DCC and VCC models. The results indicate that: (i) the individual series of returns are stationary but show some persistence, do not follow the Normal distribution and exhibit ARCH effects; (ii) the behavior of variance and covariance varies across different marked periods, depending on the type of systematic behavior of the exchange rate; (iii) there is a negative correlation between CZA prices and exchange rates, while the correlation between exchange rates is positive and significant at 5% level.

Altogether, the study concludes for the existence of volatility links among commodity prices and exchange rates, possibly with time-varying correlations, and, interestingly, depending on the specific systematic behavior of exchange rates.

Normality of log-returns

5.1 Introduction

Checking the normality is one of the assumptions required in many studies using parametric statistical methods. In this line of ideas, the normality of the log-return of stock prices is often assumed by the market players in order to use some useful results. The simplicity of the Black-Scholes formula (see Black and Scholes [13]) for pricing European options is one of the reasons why the geometric Brownian motion is one of the most popular processes used in mathematical finance. However, for this process is assumed that the log-return of the prices are normal distributed random variables, a condition that many times seems to be not true in practice.

In this chapter, we test the normality assumption for different formulations of price returns, namely intra-day and inter-day log-returns, for a large set of companies data and using five alternative goodness-of-fit tests. We implement the Pearson's Chi-Square (PCS), Kolmogorov-Smirnov (KS), Anderson-Darling (AD), Shapiro-Wilks (SW) and Jarque-Bera (JB) tests for normality and we consider the log-returns of consecutive closing prices, consecutive closing and opening prices and also opening and closing prices from the same day. We also repeat the tests when we remove some of the higher and lower log-returns from the samples.

Importantly, when analysing cointegration, particularly using Johansen approach in Johansen [63, 64], normality is also required. Furthermore, all the above holds for the various financial time series, namely for exchange rate analysis. The normality tests presented in the application of this chapter are described in Mulenga *et al* [88].

The chapter is organized as follows: in section 5.2 we introduce the tests for

the less familiarized readers; in section 5.3 we present and discuss the results of the application of the tests and in subsection 5.3.3 we conclude with some remarks.

5.2   Statistical tests for normality

We apply several goodness-of-fit tests to investigate the normality of the log-returns of stock prices. We choose to use several goodness-of-fit tests from different kinds since all the tests have some advantages and disadvantages. From the "area tests"kind we choose to implement the Pearson's Chi-square test that compares the real number of observations with the expected number of observations in each class, we can say that compares the data histogram with the histogram of the distribution being tested. Both the Kolmogorov-Smirnov and the Anderson-Darling tests uses the cumulative distribution function and the empirical distribution function and are based in a measure of the discrepancy between those two functions and therefore are considered in the class of "distance tests". Some advantages of this kind of tests is that they are easy to compute, they are more powerful than the Chi-Square test, over a wide range of alternatives, and they provide consistent tests.

The Shapiro-Wilks (or Shapiro-Francia) is a test based in the regression between the order statistics of the sample and the mean value of the order statistics from the tested distribution. This test for normality has higher power than the previous ones. Finally, the Jarque-Bera test, is based in the Lagrange multiplier test and computes the sample skewness and kurtosis, to tests if they match with the ones from a normal distribution. For a discussion regarding the power comparison of the different tests we can cite, among others, Razali and Yap [101] and Yap and Sim [119].

5.2.1   Pearson's Chi-Square test

The Pearson's Chi-Square test (first introduced in Pearson [97]), compares the frequency observed in a sample with a particular theoretical distribution, that is, for a number k of classes $C_1, ..., C_k$, mutually exclusive and of total probability, the number of observations O_i in class $i, i = 1, ..., k$ is compared with the expect number of observations E_i, in that same class provided the null hypothesis is true.

The Pearsons's chi-square statistic is,

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{5.1}$$

having a χ^2 (Chi-square) distribution with $k - p - 1$ degrees of freedom, where k is the number of classes and p is the number of estimated parameters. The null

hypothesis, of the sample be from a particular distribution, is rejected if the observed value of the statistics is greater than the critical value obtained from the Chi-square distribution.

5.2.2 Kolmogorov-Smirnov's test

The Kolmogorov-Smirnov statistic (introduced in Kolmogorov [70]) allows us to test if a sample of observations is from some completely specified continuous distribution, F_0, by means of comparing a particular kind of distance between the empirical and the theoretical cumulative distribution function.

The test statistic, D_n, is given by,

$$D_n = \max_{1 \leq i \leq n} \left[\max \left\{ F_0(X_{i:n}) - F_n(X_{i-1:n}), F_n(X_{i:n}) - F_0(X_{i:n}) \right\} \right] \qquad (5.2)$$

where F_0 is the distribution function of the theoretical distribution being tested and F_n is the empirical distribution function. Critical values for this statistic can be found in Birnbaum [11] or Massey [81], however, when some parameters of the distribution, considered in the null hypothesis, have to be estimated from the sample, then the commonly tabulated critical points can led to conservative results. In this situation, Lilliefors's critical values (that can be found in Lilliefors [74]) should be used.

5.2.3 Anderson-Darling's test

The Anderson-Darling test is the third goodness-of-fitness test that we use to test for normality. More information about this test can be found in Anderson and Darling [2, 3], but again, it compares the observed cumulative distribution function to the expected cumulative distribution function as the Kolmogorov-Smirnov test.

The statistic A_n^2, for the Anderson-Darling's test, is defined by,

$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) \left[\ln(F_0(X_{i:n})) + \ln(1 - F_0(X_{n-i+1:n})) \right], \qquad (5.3)$$

where, as before, F_0 denotes the distribution function assumed in the null hypothesis.

In Anderson and Darling [3], asymptotic critical points for significance levels of $1\%, 5\%$ and 10% are presented and more extensive tables of critical points obtained from Monte Carlo simulation can be found in Lewis [73]. When the distribution to be tested is normal or exponential and the distribution parameters are unknown and needed to be estimated, we can find the critical values for the Anderson-Darling's statistics in Stephens [112, 113].

101

5.2.4   Shapiro - Wilk's test

In the Shapiro-Wilk's test, presented in the Shapiro and Wilks [107], the test statistic W_n was constructed through the regression of the order sample statistics against the expected normal order statistics and is a suitable test when the location and scale parameters are unknown.

The test statistic, W_n, is defined by,

$$W_n = \frac{\left(\sum_{i=1}^{n} a_i X_{i:n}\right)^2}{\sum_{i=1}^{n}(X_{i:n} - \bar{X})^2} \tag{5.4}$$

where the coefficients $a_i = \mathbf{a}^T$ are given by

$$\mathbf{a}^T = (a_1, a_2, ..., a_n) = \frac{\mathbf{m}^T \mathbf{V}^{-1}}{\left(\mathbf{m}^T \mathbf{V}^{-1} \mathbf{V}^{-1} \mathbf{m}\right)^{\frac{1}{2}}} \tag{5.5}$$

with

$$\mathbf{m}^T = (m_1, m_2, ..., m_n), \quad \mathbf{V} = [v_{ij}]_{n \times n} \tag{5.6}$$

represents the vector of expected values of the standard normal order statistics and the corresponding covariance matrix, respectively.

The values for $\mathbf{a}$ and the percentage points of W_n are known up to sample sizes of $n = 50$ and can be found in the original paper. For samples of larger dimension an extension of the Shapiro - Wilk's test can be found in Royston [102] or in alternative, the Shapiro-Francia statistic (with simpler coefficients and about the same overall power) introduced in Shapiro and Francia [106] can be used.

Percentage points for the Shapiro-Francia statistic can be found in Shapiro and Francia [106] for sample sizes $n = 35, 50, 51(2)99$ and for samples of larger dimension in Royston [103].

Small values of the statistic are the significant ones, i.e. indicate non-normality.

5.2.5   Jarque-Bera's test

The last test that we implement is the Jarque-Bera's test, described in Jarque and Bera [62]. The Lagrange multipliers method was used to derived an asymptotic efficient test where the skewness and kurtosis of the sample data are compared to the ones of the normal distribution.

The test statistic, JB_n, is defined by,

$$JB_n = n\left(\frac{\widehat{skew}^2}{6} + \frac{(\widehat{kurt} - 3)^2}{24}\right) \tag{5.7}$$

where $\widehat{skew}$ and $\widehat{kurt}$ are the sample estimators of the skewness and kurtosis, as presented before in chapter 2. The large values of the statistic are the significant ones and should be compared with the right tail critical values of a Chi-square distribution with 2 degrees of freedom.

5.3 Application of normality tests

5.3.1 Data description

The data considered in this study, are daily prices from companies belonging to the Nasdaq Composite Index and are obtained from the yahoo finance site[1]. For instance, the data from the Microsoft Corporation company (MSFT) can be obtained through the link [2].

All the testing and analysis is performed using the Wolfram Mathematica 10.2 software, see Mathematica [118].

We selected the daily prices from years 2005 to 2016 and in each year we only consider the companies that have more than 220 transactions days and transaction volume of at least 50000 units per day, the total number of those companies is represent by K (in each year) and from those we take count of how many have their prices failing the normality assumption.

5.3.2 Results and discussion

In the following, let SO_i and SC_i be the stock opening and closing prices at day i, respectively. Using the log-returns of closing daily prices, namely,

$$\log\left(\frac{SC_{i+1}}{SC_i}\right) \tag{5.8}$$

and all the normality tests for a level of significance of 1%, we get the results presented in Table 5.1.

We can observe that we have normality rejection percentages, ranging from 44% to near 65% with the Pearson's Chi-square test (the one with the smaller rejection rates) to more than 90% if we consider the Jarque-Bera's test (the one with the higher rejection rates). Notice that as expected, the year of 2008 (the year of the financial crisis) is the one with higher normality rejection rate and this for all tests being considered.

---

[1]https://finance.yahoo.com/.

[2]https://finance.yahoo.com/quote/MSFT/history?period1=1104537600period2=1483142400 interval=1dfilter=historyfrequency=1d.

Table 5.1: Total (and percentage) of samples rejected by each test for normality at 1% level of significance ($\log(SC_{i+1}/SC_i)$)

Year	K	PCS (%)	KS (%)	AD (%)	SW (%)	JB (%)
2005	901	494(54.83)	658(73.03)	763(84.68)	800(88.80)	814(90.34)
2006	1018	546(53.63)	747(73.34)	850(83.50)	909(89.29)	925(90.86)
2007	1161	691(59.52)	923(79.50)	1050(90.44)	1085(93.45)	1095(94.32)
2008	1180	765(64.83)	1010(85.60)	1114(94.41)	1142(96.78)	1135(96.19)
2009	1200	642(53.50)	934(77.83)	1090(90.83)	1105(92.08)	1104(92.00)
2010	1260	557(44.21)	812(64.44)	1000(79.37)	1028(81.59)	1047(83.10)
2011	1374	686(49.93)	1007(73.29)	1208(87.92)	1267(92.21)	1287(93.67)
2012	1399	723(51.68)	974(69.62)	1129(80.70)	1198(85.63)	1218(87.06)
2013	1586	989(62.36)	1253(79.00)	1400(88.27)	1475(93.00)	1486(93.69)
2014	1441	784(54.41)	1072(74.39)	1224(84.94)	1291(89.60)	1320(91.60)
2015	1433	775(54.08)	1042(72.71)	1203(83.95)	1275(88.97)	1301(90.79)
2016	1443	902(62.51)	1182(81.91)	1318(91.34)	1361(94.32)	1379(95.56)

The year of 2010 is the one with smaller rejection rate and also for all the tests, from 2008 to 2010 we have a decreasing of the normality rejection rate from near 65% to a little more than 44% for the Pearson's Chi-square test and from near 96% to near 83% for the Jarque-Bera's test. For the Kolmogorov-Smirnov's, Anderson-Darling's or Shapiro-Wilk's tests we have decreasing rejection rates from 85.6% to 64.44%, 94.41% to 79.37% and 96.78% to 81.59%, respectively.

Next, we consider closing and opening prices from the same day and the corresponding log-returns,

$$\log\left(\frac{SC_i}{SO_i}\right) \tag{5.9}$$

For the same level of significance of 1%, as previously, we get the results presented in Table 5.2.

The main difference from Table 5.1 to Table 5.2 is the overall decreasing of the normality rejection rate in all years and for all considered tests. That decreasing is more notorious in the years of 2005-2006 (with changes as higher as 30% for the Pearson's Chi-square test) and less obvious in the crisis year of 2008 (decreasing at most 10%). Again, the year of 2008 (followed by the 2016 year) is the one with higher percentage of companies with log-returns of stock prices far away from normality and, as before, the year of 2010 is the one with the smaller percentages of normality rejection, decreasing from 54.41% in 2008 to 23.73% for the Pearson's Chi-square test in 2010 and from 92.88% to 71% for the Jarque-Bera's test, the tests with smaller/higher rejection rates, respectively.

Finally, we select the closing prices at day i and opening prices at day $i+1$ and

Table 5.2: Total (and percentage) of samples rejected by each test for normality at 1% level of significance ($\log(SC_i/SO_i)$)

Year	K	PCS (%)	KS(%)	AD (%)	SW (%)	JB(%)
2005	901	218(24.20)	411(45.62)	551(61.15)	635(70.48)	672(74.58)
2006	1018	256(25.15)	460(45.19)	607(59.63)	697(68.47)	739(72.60)
2007	1161	449(38.67)	745(64.17)	924(79.59)	980(84.41)	996(85.79)
2008	1180	642(54.41)	909(77.03)	1058(89.66)	1095(92.80)	1096(92.88)
2009	1200	504(42.00)	808(67.33)	980(81.67)	1023(85.25)	1028(85.67)
2010	1260	299(23.73)	530(42.06)	727(57.70)	845(67.06)	895(71.03)
2011	1374	471(34.28)	800(58.22)	1034(75.25)	1141(83.04)	1179(85.81)
2012	1399	457(32.67)	752(53.75)	937(67.00)	1066(76.20)	1100(78.63)
2013	1586	664(41.87)	1021(64.38)	1197(75.47)	1314(82.85)	1365(86.07)
2014	1441	510(35.39)	834(57.88)	1046(72.59)	1156(80.22)	1187(82.37)
2015	1433	563(39.29)	902(62.94)	1123(78.37)	1220(85.14)	1255(87.58)
2016	1443	643(44.56)	960(66.53)	1180(81.77)	1251(86.69)	1288(89.26)

the corresponding log-returns,

$$\log\left(\frac{SO_{i+1}}{SC_i}\right) \tag{5.10}$$

Again the five normality tests are executed for the same 1% significance level and the results can be found in Table 5.3.

Table 5.3: Total (and percentage) of samples rejected by each test for normality at 1% level of significance ($\log(SO_{i+1}/SC_i)$)

Year	K	PCS(%)	KS (%)	AD (%)	SW (%)	JB (%)
2005	901	820(91.01)	838(93.01)	851(94.45)	869(96.45)	865(96.00)
2006	1018	955(93.81)	964(94.70)	982(96.46)	999(98.13)	1000(98.23)
2007	1161	1098(94.57)	1132(97.50)	1149(98.97)	1153(99.31)	1156(99.57)
2008	1180	1147(97.20)	1172(99.32)	1179(99.92)	1179(99.92)	1178(99.83)
2009	1200	1012(84.33)	1116(93.00)	1164(97.00)	1168(97.33)	1168(97.33)
2010	1260	1027(81.51)	1121(88.97)	1190(94.44)	1229(97.54)	1236(98.10)
2011	1374	1248(90.83)	1344(97.82)	1370(99.71)	1368(99.56)	1367(99.49)
2012	1399	1243(88.85)	1329(95.00)	1360(97.21)	1362(97.36)	1370(97.93)
2013	1586	1459(91.99)	1526(96.22)	1549(97.67)	1556(98.11)	1553(97.92)
2014	1441	1340(92.99)	1382(95.91)	1408(97.71)	1422(98.68)	1423(98.75)
2015	1433	1342(93.65)	1396(97.42)	1418(98.95)	1422(99.23)	1421(99.16)
2016	1443	1393(96.54)	1425(98.75)	1438(99.65)	1442(99.93)	1441(99.86)

From the observation of Table 5.3 is obvious that for this data the results are of almost complete rejection of the normality assumption. In fact, even for the most

favorable test in not rejecting the normality assumption (the Pearson's Chi-square test), we have more than 80% of rejection rate, in the most favorable year of 2010.

Although being test depending, the overall result is that the normality assumption is largely rejected for all tests in most years but in higher percentages when we consider prices from one trading day to another (inter-day prices) as we observe in Tables 5.1 or 5.3 and lower percentages when we consider the opening and closing prices from the same day (intra-day prices, as in Table 5.2). This emphasises the idea that the non-normality of the log-returns of the stock prices seems to be more dependent from what happens when the markets are closed (see Table 5.3).

Since the rejection rates are higher when we apply the Jarque-Bera's test and because this particular test is more sensitive to extremal data it seems reasonable that some of the reason for the normality assumption failure is due to that specific observations. In fact, we observe that if we remove some of the more extreme observations from the log-returns data and we perform the same normality tests we get very different results.

Repeating the normality tests for trimmed samples, with the higher five and lower five log-returns removed, we get the results presented in Tables 5.4, 5.5 and 5.6. Notice that, when we remove 10 observations from the sample with approximately 250 entries, we are removing about 4% of the observations.

Table 5.4: Total (and percentage) of samples rejected by each test for normality at 1% level of significance, when the data is trimmed ($\log(SC_{i+1}/SC_i)$)

Year	K	PCS (%)	KS (%)	AD (%)	SW (%)	JB (%)
2005	901	113(12.54)	184(20.42)	187(20.75)	194(21.53)	86(09.54)
2006	1018	110(10.81)	198(19.45)	218(21.41)	216(21.22)	85(08.35)
2007	1161	152(13.09)	244(21.02)	291(25.06)	265(22.83)	120(10.34)
2008	1180	225(19.07)	356(30.17)	472(40.00)	411(34.83)	238(20.17)
2009	1200	196(16.33)	310(25.83)	369(30.75)	311(25.92)	144(12.00)
2010	1260	161(12.78)	218(17.30)	221(17.54)	176(13.97)	72(05.71)
2011	1374	172(12.52)	251(18.27)	289(21.03)	218(15.87)	87(06.33)
2012	1399	193(13.80)	270(19.30)	256(18.30)	243(17.37)	122(08.72)
2013	1586	310(19.55)	408(25.73)	419(26.42)	408(25.73)	224(14.12)
2014	1441	199(13.81)	276(19.15)	292(20.26)	275(19.08)	141(09.78)
2015	1433	189(13.19)	258(18.00)	245(17.10)	243(16.96)	120(08.37)
2016	1443	293(20.30)	348(24.12)	349(24.19)	311(21.55)	167(11.57)

Again, we can observe that the normality assumption is more times rejected when we consider inter-day prices, just as before (see Tables 5.4 and 5.6). However, in this new framework, the normality assumption is not rejected so many times as before.

106

Table 5.5: Total (and percentage) of samples rejected by each test for normality at 1% level of significance, when the data is trimmed ($\log(SC_i/SO_i)$)

Year	K	PCS (%)	KS(%)	AD (%)	SW(%)	JB (%)
2005	901	79(8.77)	113(12.54)	124(13.76)	132(14.65)	37(4.11)
2006	1018	91(8.94)	134(13.16)	128(12.57)	131(12.87)	49(4.81)
2007	1161	124(10.68)	203(17.48)	235(20.24)	199(17.14)	88(7.58)
2008	1180	178(15.08)	316(26.78)	423(35.85)	373(31.61)	213(18.05)
2009	1200	188(15.67)	284(23.67)	308(25.67)	261(21.75)	102(8.50)
2010	1260	129(10.24)	147(11.67)	132(10.48)	124(9.84)	37(2.94)
2011	1374	150(10.92)	209(15.21)	214(15.58)	180(13.10)	62(4.51)
2012	1399	179(12.79)	225(16.08)	192(13.72)	187(13.37)	62(4.43)
2013	1586	253(15.95)	330(20.81)	299(18.85)	288(18.16)	126(7.94)
2014	1441	165(11.45)	222(15.41)	213(14.78)	196(13.60)	89(6.18)
2015	1433	179(12.49)	243(16.96)	219(15.28)	208(14.52)	69(4.82)
2016	1443	252(17.46)	290(20.10)	279(19.33)	245(16.98)	100(6.93)

Table 5.6: Total (and percentage) of samples rejected by each test for normality at 1% level of significance, when the data is trimmed ($\log(SO_{i+1}/SC_i)$)

Year	K	PCS(%)	KS(%)	AD(%)	SW(%)	JB(%)
2005	901	500(55.49)	493(54.72)	442(49.06)	468(51.94)	276(30.63)
2006	1018	555(54.52)	549(53.93)	515(50.59)	535(52.55)	341(33.50)
2007	1161	555(47.80)	678(58.40)	694(59.78)	656(56.50)	516(44.44)
2008	1180	592(50.17)	784(66.44)	912(77.29)	831(70.42)	611(51.78)
2009	1200	477(39.75)	621(51.75)	608(50.67)	567(47.25)	336(28.00)
2010	1260	499(39.60)	522(41.43)	425(33.73)	392(31.11)	279(22.14)
2011	1374	594(43.23)	795(57.86)	931(67.76)	824(59.97)	561(40.83)
2012	1399	618(44.17)	686(49.04)	563(40.24)	524(37.46)	362(25.88)
2013	1586	806(50.82)	881(55.55)	780(49.18)	734(46.28)	487(30.71)
2014	1441	627(43.51)	699(48.51)	682(47.33)	678(47.05)	561(38.93)
2015	1433	596(41.59)	679(47.38)	665(46.41)	676(47.17)	521(36.36)
2016	1443	734(50.87)	861(59.67)	864(59.88)	885(61.33)	708(49.06)

In fact, the percentage of normality rejections decreases substantially, in the case of Shapiro-Wilk's and Jarque-Bera's tests, from values above the 80% to values below 25% for the $\log(SC_{i+1}/SC_i)$ data (Table 5.1 vs Table 5.4) and from values above 90% to values ranging 40%-50% for the $\log(SO_{i+1}/SC_i)$ data (Table 5.3 vs Table 5.6). The same reduction in the rejection rate is also observed for the remaining tests and for all the years.

If we compare the normality rejection percentage for intra-day prices, that is,

$(\log(SC_i/SO_i))$, the decreasing is from more than 75% to less than 20% (Shapiro-Wilk's, Tables 5.2 and 5.5). The reduction is also evident for the other tests and, in particular, notice that the Jarque-Bera's test in the previous framework gives the higher rates of rejection but for the trimmed data with the more extreme observations removed, is the test with lower rates of rejection. This outcome was somewhat expected because as already said, the more extreme observations will condition the skewness and the kurtosis of the distribution, meaning that some tests supposed to be more affected by those parameters are indeed strongly affected.

In this framework is also obvious that the departure from normality is stronger and seems more dependent to what happens during the closing periods of the financial markets, see Tables 5.4 and 5.6. Even in this set up of trimmed data, we observe that the 2008 year (the year of the financial crisis) and for almost all tests, is again the year with the higher rates of normality rejection, as expected.

5.3.3 Final remarks

The results of the data testing allows us to discuss two questions: The normality assumption of the log-returns of the stock prices is still a reasonable assumption? There are differences in the normality assumption of the log-returns in intra-day and inter-day prices? Regarding the first question we can observe that when we remove some of the more extreme observations the normality assumption is reasonable for most of the companies stock prices. That is, some trimming ensures normality, validating the applications of several models, of pricing and multivariate time series, including volatility analysis.

On the other hand, it seems reasonable to say that what affects more strongly the non-normality of the logarithm of returns for the stock prices are the "things"that happens when the markets are closed. The focus and results in the application are also relevant, with due differences, for all kinds of financial times series, e.g., bonds and currencies, validating pricing and forecasting models, including volatility analysis.

Conclusions and Recommendations

6.1 Conclusions

Mozambique like any other country has the National Institute of Statistics, which in collaboration with other institutions, have the task of collecting data, process and disseminate the information for use by either the government or other organizations. Collecting data, processing and providing appropriate information to users involves knowing the sampling techniques, the most appropriate models as well as a solid knowledge of statistical data analysis as a whole.

Considering the need for knowledge of procedures, data analysis and processing techniques in order to provide accurate information to users, especially information for macroeconomic indicators, we propose modelling and analysis of univariate and multivariate volatility applied to macroeconomic variables of Mozambique such as Exchange rate, real GDP and GDP deflator or related to the country such as prices of Coal of Africa and South Africa real GDP.

In order to understand and disseminate information related to the behavior of macroeconomic time series, we identify three tasks (i) investigate the behavior of exchange rate, (ii) analyse of the dynamic interplay of macroeconomic time series, and (iii) analyse the relationship between co-volatilities of returns. These tasks correspond to the same number of problems, which we solved by modelling the series using different specifications such as ARCH, GARCH, VAR, SVEC and MGARCH models presented in previews chapters to analyse the dynamics between the series or their volatilities.

In the study of the exchange rate behavior, we take Mozambique daily New Metical exchange rate against South Africa Rand, over 2010-2014. Out of modelling full sample, we identified four sub-periods with distinct currency behavior (Figure 2.1), – 2010 - depreciation, 2011 - appreciation, 2012 - stability and 2013-2014 stability with appreciation – and apply four popular asymmetric GARCH models described in section 2.3. The validation of models is done with sign and size bias tests, along with the plots of news impact curves.

The dynamics between variables was made, verifying the role of Mozambique real GDP, GDP deflator and external shocks represented by South Africa real GDP. Next, we use the SVEC model 3.38 to take into account the vector error correction and identify contemporaneous and long-run effects of shocks, where for identification we assume that prices are not affected by contemporaneous supply shocks nor by external shocks. For long-run we consider that only an aggregate supply shock has permanent effects.

The analysis of the relationship and co-volatilities between variables was performed taking the prices of Coal of South Africa, the Mozambique New Metical against South Africa Rand, and Mozambique New Metical against Great Britain Pound exchange rates, were used multivariate GARCH models in the class of conditional correlation, as discussed in section 4.3.

Because the assumption of normality of returns is required in many studies, we verify the normality of log-returns of stock prices using five statistical tests described in section 5.2.

Summarizing, from the study, one can present the following conclusions:

(a) For MZN/ZAR exchange rate series, data description show that the series exhibit different behavior in the study period and in periods of appreciation with or without some stability, there is no asymmetry. Results from the full heterogeneous sample masks to some extent, the existence of size and sign effects of news. Comparing estimates from different behavior suggests that in depreciation times the significance of asymmetry and importance of size effects are stronger, than in the stability periods where the exchange rate risk is lower, in addition, the positive news have a higher effect on volatility, than bad news, when the currency trend is appreciation or stability.

(b) Results from tri-variable analysis indicates stationary series after first differences, and the Johansen test for cointegration identified two cointegrating vectors where the first equation is composite by Mozambique real GDP and South African real GDP, the second pair is composite by Mozambique real GDP deflator and South Africa real GDP (Table 3.2 and 3.3). Regarding impulse response functions the aggregate supply shocks have a positive and permanent impact on Mozambique real

GDP, while for Mozambique real GDP deflator and South Africa real GDP, the impact of aggregate supply shocks is negative and transitory, which confirm the existing economic theory in the literature (Figure 3.3). With respect to forecasting error variance decomposition, we notice that the long-run development of both Mozambique real GDP and real GDP deflator are dominated by its own supply and demand shocks respectively. However, for GDP deflator this predominance decreases over time giving rise to the influence of supply shocks. Regarding the variance of South Africa real GDP, it is mainly attributed to demand and supply shocks (Table 3.5).

c) The results from multivariate GARCH models indicate that the series exhibit ARCH effects and the behavior of variance and covariance vary across different market periods, depending on the behavior of the exchange rate. The univariate GARCH(1,1) parameter estimates indicates the existence of hight persistence, since $\alpha + \beta \approx 0.9$, (Table 4.2). Regarding relationship, the results of this sample indicate a positive correlation between exchange rates, while the relations between exchange rates and Coal prices are negative (Table 4.3). Also from the sample, the Wald test indicate that DCC and VCC models do not reduce to CCC model, suggesting the assumption of time-invariant too restrictive for the data analysed.

d) The study of normality of stock prices shows that most tests reject the hypothesis of normality of log-returns of stock prices. However, removing some of the extreme observations the normality assumption is reasonable for most of the stocks prices.

Among several findings found throughout the study, the following can be underlined:

- In the policy design, memory and nonlinearity matter and the trend behavior of the exchange rate is key for assessing the effects of news on exchange rate risk, because not always good news is good.

- The Mozambique real GDP plays an important role in both cointegration relations, in the analysis of impulse response function as well as the forecast error variance decomposition.

- The volatility links between commodity prices and exchange rates is not always the some, particularly across different marked exchange rate behavior.

- The rejection rate of the normality assumption of the log-returns of stock prices increases when the stock price of returns are based on consecutive days and when some extreme observations are considered.

The analyse of exchange rate behavior, bring new contributions and reinforce the existing literature, that the model choice is key for the quality of results, highlighting the importance of selection criteria and diagnostic testing. Novel contributions come from the pervasive sense of results to currency trends, as well as modelling options in nontrivial ways. The knowledge of the relations among variables, allows the government to make decisions taking into account these relationships, therefore, the study contributes to the analysis of macroeconomic data of the Mozambique economy in the presence of an external variable, which helps the process of defining policies that increase the rate of sustainable economic growth of Mozambique. The study also reinforces the idea that the success in applying a statistical test depends on the preparation and type of data available.

6.2 Recommendations

After analysing some macroeconomic time series related to the Mozambique economy using different econometric models, from this experience we recommend that:

- In order to provide adequate information to users such as government, other organizations or companies, it is necessary to conduct macro-econometric studies using appropriate models that can capture the behavior of the variables under study. Only in this way the forecasts are valid.

- The National Institute of Statistics in collaboration with other research institutions such as higher education, ministries should promote studies, especially econometric studies with a view to improving the quality of economic information, consequently developing the country.

- Even though this study contributes by presenting a range of models used to study variables related to Mozambique economy, macroeconomic studies still can be performed using similar or other variables related to Mozambique or between the countries of the region to know the influence of countries economies.

Bibliography

[1] G. Amisano and C. Giannini. *Topics in structural VAR Econometrics*. Second Edition. Springer-Verlag Berlin, 1997. DOI: 10.1007/978-3-642-60623-6.

[2] T. Anderson and D. Darling. "Asymptotic Theory of Certain Goodness-of-Fit Criteria Based on Stochastic Processes." In: *Annals of Mathematical Statistics* 23.2 (1952), pp. 193–212.

[3] T. Anderson and D. Darling. "A Test of Goodness of Fit." In: *Journal of the American Statistical Association* 49.268 (1954), pp. 765–769.

[4] Y. Baba, R. F. Engle, D. Kraft, and K. Kronner. "Multivariate simultaneous generalized ARCH." In: *Department of Economics: University of California, San Diego* (1990).

[5] B. H. Baltagi. *Econometrics*. Fourth Edition. Springer-Verlag, 2008. ISBN: 978-3-540-76515-8. DOI: 10.1007/978-3-540-76516-5.

[6] L. Bauwens, S. Laurent, and J. V. K. Rombouts. "Multivariate GARCH models: A survey." In: *Journal of Applied Econometrics* 21 (2006), pp. 79–109. DOI: 10.1002/jae.842.

[7] A. K. Bera and M. L. Higgins. "ARCH models: Properties, estimation and testing." In: *Journal of Economic Surveys* 7.4 (1993), pp. 307–366. DOI: 10.1111/j.1467-6419.1993.tb00170.x.

[8] B. S. Bernanke. "Alternative explanations of the money-income correlation." In: *Carnegie-Rochester Conference Series on Public Policy* 25 (1986), pp. 49–99. URL: https://doi.org/10.1016/0167-2231(86)90037-0.

[9] E. Berndt, B. Hall, R. Hall, and J. Hausman. "Estimation and inference in nonlinear structural models." In: *Annals of Economic and Social Measurement* 3.4 (1974), pp. 653–665. URL: http://www.nber.org/books/aesm74-4.

[10] S. Beveridge and C. R. Nelson. "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the Business Cycle." In: *Journal of Monetary Economy* 7.2 (1981), pp. 151 –174.

[11]   Z. W. Birnbaum. "Numerical Tabulation of the Distribution of Kolmogorov's Statistic for Finite Sample Size." In: *Journal of the American Statistical Association* 47.259 (1952), pp. 425–441.

[12]   F. Black. "Studies of stock price volatility changes." In: *Proceedings of the 1976 meeting of the Business and Economic Statistics Section.* Ed. by T. G. Anderson and T. Bollerslev. Vol. 1. American Statistical Association, 1976, pp. 177–181.

[13]   F. Black and M. Scholes. "The Pricing of Options and Corporate Liabilities." In: *Journal of Political Economy* 81.3 (1973), pp. 637–654.

[14]   O. J. Blanchard and D. Quah. "The dynamic effects of aggregate demand and supply disturbances." In: *American Economic Review* 79.4 (1989), pp. 655 –673. URL: http://www.jstor.org/stable/1827924.

[15]   T. Bollerslev. "Generalized autoregressive conditional heteroskedasticity." In: *Journal of Econometrics* 31.3 (1986), pp. 307–327. URL: https://doi.org/10.1016/0304-4076(86)90063-1.

[16]   T. Bollerslev. "A conditional heteroskedasticity time series for speculative prices rates of return." In: *Review of Economics and Statistics* 69.3 (1987), pp. 542–547. DOI: 10.2307/1925546.

[17]   T. Bollerslev. "Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model." In: *Review of Economics and Statistics* 72.3 (1990), pp. 498–505. DOI: 10.2307/2109358.

[18]   T. Bollerslev. "Glossary to ARCH (GARCH)." In: *Center for Research in Econometric Analysis of Time Series* 49 (2008), pp. 01–46. URL: http://ssm.com/abstract=1263250.

[19]   T. Bollerslev and J. W. Wooldridge. "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariance." In: *Econometric Reviews* 11.2 (1992), pp. 143 –172. DOI: 10.1080/07474939208800229.

[20]   T. Bollerslev, R. F. Engle, and J. M. Wooldridge. "A capital asset pricing model with time varying covariance." In: *Journal of Political Economy* 96.1 (1988), pp. 116 –131. DOI: 10.1086/261527.

[21]   T. Bollerslev, R. F. Engle, and D. B. Nelson. "ARCH models." In: *Handbook of Econometrics* 4 (1994), pp. 2959–3038.

[22]   G. E. P. Box and G. M. Jenkins. *Time series analysis: Forecasting and control.* Holden-Day San Francisco, 1970.

[23] G. E. P. Box and D. A. Pierce. "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models." In: *Journal of the American Statistical Association* 65.332 (1970), pp. 1509–1526.

[24] J. Breitung, R. Bruggemann, and H. Lutkepohl. "Structural vector autoregressive modeling and impulse responses." In: *Applied Times Series Econometrics.* Ed. by H. Lutkepohl and M. Kratzig. 2004. Chap. 4, pp. 159 –196.

[25] T. S. Breusch. "Testing for autocorrelation in dynamic linear models." In: *Australian Economic Papers* 17 (1978), pp. 334–355.

[26] W. A. Broock, J. A. Scheinkman, W. D. Dechert, and B. LeBaron. "A test for independence based on the correlation dimension." In: *Econometric Reviews* 15.3 (1996), pp. 197–235. DOI: 10.1080/07474939608800353.

[27] C. Brooks. *Introduction econometrics for finance.* Second Edition. Cambridge Press, 2008. ISBN: 978-0-521-69468-1.

[28] R. Bruggemann. "Finite sample properties of impulse response intervals in SVECMs with long-run identifications restrictions." In: *Economic Risk, Discussion paper 021* (2006).

[29] A. Chaudhry and S. K. H. Bukhari. "A structural VAR analysis of the impact of macroeconomic shocks on Pakistan's textile exports." In: *Economic modelling* 32 (2013), pp. 302 –315.

[30] J. Davidson. *Econometric theory.* First Edition. Published by Oxford (GB): Blackwell, 2000. ISBN: 10:0631215840.

[31] D. A. Dickey and W. A. Fuller. "Distribution of the estimators for autoregressive time series with a unit root." In: *Journal of the American Statistical Association* 74.366 (1979), pp. 427–431. DOI: 10.2307/2286348.

[32] F. X. Diebold and J. A. Lopez. "Modeling Volatility Dynamics." In: *Macroeconometrics: Developments, Tensions, and Prospects.* Ed. by K. D. Hoover. Dordrecht: Springer Netherlands, 1995, pp. 427–472. ISBN: 978-94-011-0669-6.

[33] F. X. Diebold and M. Nerlove. "The dynamics of exchange rate volatility: A multivariate latent factor ARCH model." In: *Journal of Applied Econometrics* 4.1 (1989), pp. 1–21.

[34] Z. Ding, C. W. J. Granger, and R. F. Engle. "A long memory property of stock market returns and a new model." In: *Journal of Empirical Finance* 1 (1993), pp. 83–106. URL: https://doi.org/10.1016/0927-5398(93)90006-D.

[35]   J. Dolado and H. Lutkepohl. "making Wald tests work for coitengrated VAR systems." In: *Econometric Reviews* 15.4 (1996), pp. 369–396. DOI: 10.1080/07474939608800362.

[36]   J. A. Doornik and H. Hansen. "An Omnibus test for univariate and multivariate normality." In: *Working paper* (1994).

[37]   B. Efron and R. Tibshirani. *An introduction to the bootstrap.* Chapman & Hall, 1993. ISBN: 0-412-04231-2.

[38]   W. Enders. *Applied econometric time series.* 3rd edition. John Wiley & Sans, 2010. ISBN: 978-0-470-50539-7.

[39]   R. F. Engle. "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation." In: *Econometrica* 50.4 (1982), pp. 987–1007. DOI: 10.2307/1912773.

[40]   R. F. Engle. "Dynamic conditional correlation: A simple class of multivariate GARCH models." In: *Journal of Business and Economic Statistics* 20.3 (2002), pp. 339–350. DOI: 10.1198/073500102288618487.

[41]   R. F. Engle and T. Bollerslev. "Modelling the persistence of conditional variances." In: *Econometric Reviews* 5.1 (1986), pp. 1–50. DOI: 10.1080/07474938608800108.

[42]   R. F. Engle and C. W. J. Granger. "Co-integration and error correction: Representation, estimation and testing." In: *Econometrica* 55.2 (1987), pp. 251–276. DOI: 10.2307/1913236.

[43]   R. F. Engle and K. Kroner. "Multivariate simultaneous Generalized ARCH." In: *Econometric Theory* 11.1 (1995), pp. 122–150. URL: http://www.jstor.org/stable/3532933.

[44]   R. F. Engle and V. Ng. "Measuring and testing the impact of news on volatility." In: *Journal of Finance* 48 (1993), pp. 1749–1778. URL: https://doi.org/10.1111/j.1540-6261.1993.tb05127.x.

[45]   R. F. Engle, C. W. J. Granger, and D. Kraft. "Combining competing forecasts of inflation using a bivariate ARCH model." In: *Journal of Economic Dynamics and Control* 8.2 (1984), pp. 151–165. URL: https://doi.org/10.1016/0165-1889(84)90031-9.

[46]   R. F. Engle, D. M. Lilien, and R. P. Robins. "Estimating time varying risk premia in the term structure: The ARCH-M model." In: *Econometrica* 55.2 (1987), pp. 391–407. DOI: 10.2307/1913242.

[47] R. F. Engle and K. Sheppard. *Theoretical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH.* Working Paper 8554. National Bureau of Economic Research, 2001. DOI: 10.3386/w8554.

[48] M. Faias, P. Mota, A. Mulenga, and J. P. Pina. "Asymmetry of ARCH and natural resources disease or virtue: Mozambique experience." In: *Proceeding of the International Conference on Numerical Analysis and Applied Mathematics.* Vol. 1738. 1. Rhodes, Greece, 2016, pp. 19004–01–19004–04. DOI: 10.1063/1.4951971.

[49] J. Faust. "The robustness of identified VAR conclusions about money." In: *Carnegie Rochester Conference series on Public Policy* 49 (1998), pp. 207–244. DOI: 10.1016/s0167-2231(99)00009-3.

[50] J. Galí. "How well does the IS-LM model fit postwar U.S.Data?" In: *The Quarterly Journal of Economics* 107.2 (1992), pp. 709–738.

[51] J. Geweke. "Modeling the persistence of conditional variances: A comment." In: *Econometric Review* 5 (1986), pp. 57–61.

[52] L. R. Glosten, R. Jagannathan, and D. E. Runkle. "On the relation between the expected value and the volatility of the nominal excess return on stocks." In: *Journal of Finance* 48.3 (1993), pp. 1779–1801.

[53] L. G. Godfrey. "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables." In: *Econometrica* 46 (1978), pp. 1293–1301. DOI: 10.2307/1913829.

[54] C. W. J. Granger. "Investigate causal relations by econometric models and cross-spectral methods." In: *Econometrica* 37.3 (1969), pp. 424–438. DOI: 10.2307/1912791.

[55] C. W. J. Granger. "Some properties of time series data and their use in econometric model specification." In: *Journal of Econometrics* 16 (1981), pp. 121–130. DOI: 10.1016/0304-4076(81)90079-8.

[56] C. W. J. Granger and A. P. Anderson. *An introduction to bilinear time series models.* 1st ed. Vandenhoeck und Ruprecht, 1978. ISBN: 10:3525112394.

[57] P. Hall. *The bootstrap and edgeworth expansion.* Springer Series in Statistics, 1992. ISBN: 978-0-3879-4508-8.

[58] J. D. Hamilton. *Time series analysis.* Princeton University Press, 1994. ISBN: 9780691042893.

[59] L. Hentschel. "All in the family: Nesting symmetric and asymmetric GARCH models." In: *Journal of Financial Economics* 39.1 (1995), pp. 71–104.

[60] M. L. Higgins and A. K. Bera. "A class of nonlinear ARCH models." In: *International Economic Review* 33 (1992), pp. 137–158.

[61] J. R. M. Hosking. "The multivariate Portmanteau statistics." In: *Journal of the American Statistical Association* 75.371 (1980), pp. 602–608. DOI: 10.1080/01621459.1980.10477520.

[62] C. M. Jarque and A. K. Bera. "A test for normality of observations and regression residuals." In: *International Statistical Review / Revue Internationale de Statistique* 55.2 (1987), pp. 163–172. DOI: 10.2307/1403192.

[63] S. Johansen. "Statistical analysis of co-integration vectors." In: *Journal of Economic Dynamics and Control* 12.2–3 (1988), pp. 231–254.

[64] S. Johansen. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." In: *Econometrica* 59.6 (1991), pp. 1551–1580.

[65] S. Johansen. *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford university Press, Oxford, 1995. ISBN: 9780198774501.

[66] J. Johnston and J. DiNardo. *Econometric methods.* 4th edition. McGraw-Hill, 1997. ISBN: 0-07-913121-2.

[67] L. Kilian. "Structural vector autoregressions." In: *CEPR Discussion Paper Series* 8515 (2011), pp. 1–54.

[68] R. G. King, C. I. Plosser, J. H. Stock, and M. W. Watson. "Stochastic trends and economic fluctuations." In: *American Economic Review* 81.4 (1991), pp. 819–840.

[69] G. Kirchgassner and J. Wolters. *Introduction to modern time series analysis.* Springer-Verlag Berlin Heidelberg, 2007. ISBN: 978-3-540-73291-4. DOI: 10.1007/978-3-540-73291-4.

[70] A. N. Kolmogorov. "Sulla determinazione empirica di una legge di distribuzione (On the empirical determination of a distribution law)." In: *Giornale dell'Istituto Italiano degli Attuari* 4 (1933), pp. 83–91.

[71] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin. "Testing the Null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" In: *Journal of Econometrics* 46.1–3 (1992), pp. 159–175. DOI: 10.1016/0304-4076(92)90104-Y.

[72] O. Ledoit, P. Santa-clara, and M. Wolf. "Flexible multivariate GARCH modeling with an application to international stock markets." In: *The Review of Economic and Statistics* 85.3 (2003), pp. 735–747.

[73] A. A. Lewis. "Distribution of the Anderson-Darling Statistic." In: *Annals of Mathematical Statistics* 32.4 (1961), pp. 1118–1124.

[74] H. W. Lilliefors. "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown." In: *Journal of the American Statistical Association* 62.318 (1967), pp. 399–402.

[75] G. M. Ljung and G. E. Box. "On a measure of a lack of fit in time series models." In: *Biometrika* 65.2 (1978), pp. 297–303.

[76] H. Luetkepohl. "Vector autoregressive models." In: *EUI Working Papers* 30 (2011), pp. 1–33.

[77] H. Lutkepohl. *New introduction to multiple time series analyses*. Springer-Berlin Heidelberg New York, 2005. ISBN: 3-540-40172-5.

[78] M. Lutkepohl and M. Kratzig. *Applied time series econometrics*. Cambridge University Press, 2004. URL: http://www.cambridge.org/9780521839198.

[79] J. G. MacKnnon. "Critical values for co-integration tests." In: *Long-Run Economic Relationships*. Ed. by R. F. Engle and C. W. J. Granger. Oxford University Press, 1991, pp. 267–276.

[80] K. V. Mardia. "Measures of multivariate skewness and kurtosis with applications." In: *Biometrika* 57.3 (1970), pp. 519–530. DOI: 102307/2334770.

[81] F. J. Massey. "The Kolmogorov-Smirnov Test for Goodness of Fit." In: *Journal of the American Statistical Association* 46.253 (1951), pp. 68–78.

[82] R. C. Maysami and T. S. Koh. "A vector error correction model of the Singapore stock market." In: *International Review of Economics & Finance* 9.1 (2000), pp. 79–96. URL: https://doi.org/10.1016/S1059-0560(99)00042-8.

[83] A. I. McLeod and W. K. Li. "Diagnostic checking ARMA time series models using squared residual autocorrelation." In: *Journal of Time Series Analysis* 4.4 (1983), pp. 269–273. DOI: 10.1111/j.1467-9892.1983.tb00373.x.

[84] A. Milhoj. "A multiplicative parameterization of ARCH models." In: *Working paper. Department of Statistics, University of Copenhagem* (1987).

[85] A. Mulenga. "The role of supply, demand and external shocks in an open economy: The case of Mozambique." In: *International Journal of Computational Economics and Econometrics* (2019), (submitted).

[86] A. Mulenga, M. Faias, P. Mota, and J. P. Pina. "The link between exchange rates and commodity prices is not always the same: Lessons from Mozambique." In: *Contribution to Statistics*. Springer series, 2017, (submitted).

[87]   A. Mulenga, M. Faias, P. Mota, and J. P. Pina. "Exchange rate volatility: an asymmetric tale from Mozambique." In: *African Finance Journal* (2019), (submitted, September).

[88]   A. Mulenga, M. Faias, P. Mota, and J. P. Pina. "What happens when the stock markets are closed?" In: *Electronic Journal of Applied Statistical Analysis* 12.02 (2019), pp. 405–415. DOI: 10.1285/i20705948v12n2p405.

[89]   H. I. Mussagy and H. B. Allaro. "Structural change in Mozambique: Economic performance before and after the civil war." In: *Journal of Economic and Sustainable development* 7.11 (2016), pp. 119 –125.

[90]   D. B. Nelson. "Conditional heteroskedasticity in asset returns: A new approach." In: *Econometrica* 59 (1991), pp. 347–370. URL: http://www.jstor.org/stable/2938260.

[91]   S. Ng and P. Perron. "Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag." In: *Journal of the American Statistical Association* 90.429 (1995), pp. 268–281. DOI: 10.2307/2291151.

[92]   S. K. Obeng and D. Sakyi. "Macroeconomic determinants of interest rate spreads in Ghana." In: *African Journal of Economic and Management Studies* 8.1 (2017), pp. 76 –88.

[93]   OECD. *OECD Investiment policy reviews: Mozambique. Executive summary and recommendations.* 2013. URL: http://dx.doi.org/10.1787/9789264203310-en.

[94]   M. Osterwald-Lenum. "A note with quantiles of the asymptotic distribution of the maximum likelihood co-integration rank test statistics." In: *Oxford Bulletin of Economics and Statistics* 54.3 (1992), pp. 461–472. URL: https://doi.org/10.1111/j.1468-0084.1992.tb00013x.

[95]   A. R. Pagan and G. W. Schwert. "Alternative models for conditional stock volatility." In: *Journal of Econometrics* 45.1 (1990), pp. 267–290. URL: http://dx.doi.org/10.1016/0304-4076(90)90101-X.

[96]   S. G. Pantula. "Modeling the persistence of conditional variances: A comment." In: *Econometrics Review* 5 (1986), pp. 71–74.

[97]   K. Pearson. "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling." In: *Philosophical Magazine Series 5* 50.302 (1900), pp. 157–175.

[98]   B. Pfaff.   "VAR, SVAR and SVECM models: implementation within R Package vars." In: *Journal of Statistical Software* 25.4 (2008), pp. 1–32. URL: http://www.jstatsoft.org/.

[99]   P. C. B. Phillips and P. Perron.   "Testing for a unit root in time series regression." In: *Biometrika* 75.2 (1988), pp. 335–346.

[100]   F. V. der Ploeg.   "Natural resources: Curse of Blessing?" In: *Journal of Economic Literature* 49.2 (2011), pp. 366–420. DOI: 10.1257/jel.49.2.366.

[101]   N. M. Razali and B. Yap.   "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests." In: *Journal of Statistical Modeling and Analytics* 2 (Jan. 2011), pp. 21–33.

[102]   J. P. Royston. "An Extension of Shapiro and Wilk's W Test for Normality to Large Samples." In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 31.2 (1982), pp. 115–124.

[103]   J. Royston. "A Simple Method for Evaluating the Shapiro-Francia W' Test of Non-Normality." In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 32.3 (1983), pp. 297–300.

[104]   E. Said and D. A. Dickey. "Testing for unit roots in autoregressive moving average models of unknown order." In: *Biometrika* 71.3 (1984), pp. 599–607.

[105]   G. M. Schwert. "Stork returns and real activity: A century of evidence." In: *Journal of Finance* 45.4 (1990), pp. 1237–1257.

[106]   S. S. Shapiro and R. S. Francia. "An Approximate Analysis of Variance Test for Normality." In: *Journal of the American Statistical Association* 67.337 (1972), pp. 215–216.

[107]   S. Shapiro and M. Wilk. "An Analysis of Variance Test for Normality (Complete Samples)." In: *Biometrika* 52.3/4 (1965), pp. 591–611.

[108]   A. Silvennoinen and T. Teräsvirta. *Multivariate GARCH models.* SSE/EFI Working Paper Series in Economics and Finance 669. Stockholm School of Economics, 2008. URL: https://EconPapers.repec.org/RePEc:hhs:hastef:0669.

[109]   C. A. Sims. "Macroeconomic and reality." In: *Econometrica* 48.1 (1980), pp. 01–48. DOI: 10.2307/1912017.

[110]   C. A. Sims. "An autoregressive index model for U.S. 1948–1975." In: *Large-Scale Macro-Econometric Models.* Ed. by J.Kmenta and B.Ramsey. North-Holland, Amsterdam, 1981, pp. 283–327.

[111] M. S. Srivastava. "A measure of skewness and kurtosis and a graphical method for assessing multivariate normality." In: *Statistics & Probability Letters* 2.5 (1984), pp. 263–267. URL: https://doi.org/10.1016/0167-7152(84)90062-2.

[112] M. Stephens. "EDF Statistics for Goodness of Fit and Some Comparisons." In: *Journal of the American Statistical Association* 69.347 (1974), pp. 730–737.

[113] M. Stephens. "Asymptotic Results for Goodness-of-Fit Statistics with Unknown Parameters." In: *Annals of Statistics* 4.2 (1976), pp. 357–369.

[114] J. H. Stock and M. M. Watson. "Vector autoregressions." In: *Journal of Economic Perspectives* 15.4 (2001), pp. 101–115.

[115] S. J. Taylor. "Forecasting the volatility of currency exchange rates." In: *International Journal of Forecasting* 3.1 (1987), pp. 159–170.

[116] R. S. Tsay. *Analysis of financial time series.* 2nd edition. John Wiley & Sans, 2005. ISBN: 13 978-0-471-69074-0.

[117] Y. K. Tse and A. K. C. Tsui. "A multivariate generalized autoregressive conditional heteroskedasticity model with time-varying correlations." In: *Journal of Business and Economic Statistics* 20.3 (2002), pp. 351–362.

[118] I. Wolfram Research. *Mathematica, Version 10.2.* Champaign, IL. 2015.

[119] B. W. Yap and C. M. Sim. "Comparisons of various types of normality tests." In: *Journal of Statistical Computation and Simulation* 81.12 (2011), pp. 2141–2155. DOI: 10.1080/00949655.2010.520163.

[120] J. Yu. "Forecasting volatility in The New Zealand stock market." In: *Applied Financial Economics* 12.3 (2002), pp. 193–202. URL: https://doi.org/10.1080/09603100110090118.

[121] J. M. Zakoian. "Threshold heteroskedastic models." In: *Journal of Economic Dynamics and Control* 18.5 (1994), pp. 931–955. URL: https://EconPapers.repec.org/RePEc:eee:dyncon:v:18:y:1994:i:5:p:931-955.

[122] E. Zivot. "Practical issues in the analysis of univariate GARCH models." In: *Handbook of Financial Time Series* (2009), pp. 113–155. URL: https://link.springer.com/chapter/10.1007/978-3-540-71297-8_5.

[123] E. Zivot and J. Wang. *Modeling financial times series with S-PLUS.* 2nd edition. Springer-Verlag New York, 2006. ISBN: 978-0-387-27965-7. URL: http://www.springer.com/978-0-387-27965-7.

Volatility models and the multivariate links with applications to the economy of Mozambique

Alberto Mulenga

2019

Alberto Chicafo Mulenga

Master of Science

Volatility models and the multivariate links with applications to the economy of Mozambique

Dissertação para obtenção do Grau de Doutor em

Estatística e Gestão de Rísco

September, 2019

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Alberto Chicafo Mulenga

Master of Science

Volatility models and the multivariate links with applications to the economy of Mozambique

Dissertação para obtenção do Grau de Doutor em

Estatística e Gestão de Rísco

September, 2019

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA