Modern Approaches for Nonlinear Data Analysis of Economic and Financial Time Series

Author:
Peter Martey Addo

A dissertation submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the European Doctorate in Economics Erasmus Mundus Mathématiques & informatique (UFR27)

Doctoral Defense
on the
30 May 2014

Doctoral Dissertation Reading Committee:

Philippe de PERETTI (Professor, Université Paris I) (Chair of Jury)
Dominique GUÉGAN (Professor, Université Paris I) (Doctoral Advisor)
Monica BILLIO (Professor, Università Ca’ Foscari Venezia) (Doctoral Advisor)
Massimiliano CAPORIN (Professor, Université de Padova) (Rapporteur)
Michael ROCKINGER (Professor, Université de Lausanne) (Rapporteur)
Modern Approaches for Nonlinear Data Analysis of Economic and Financial Time Series

Author: Peter Martey Addo
Supervisors: Prof. Monica Billio Prof. Dominique Guegan

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the European Doctorate in Economics Erasmus Mundus Mathématiques & informatique (UFR27)

April 2014
“If you’re not prepared to be wrong, you’ll never come up with anything original.”

Ken Robinson
Abstract

This thesis centers on introducing modern non-linear approaches for data analysis in economics and finance with special attention on business cycles and financial crisis. It is now well stated in the statistical and economic literature that major economic variables display non-linear behaviour over the different phases of the business cycle. As such, nonlinear approaches/models are required to capture the features of the data generating mechanism of inherently asymmetric realizations, since linear models are incapable of generating such behavior.

In this respect, the thesis provides an interdisciplinary and open-minded approach to analyzing economic and financial systems in a novel way. The thesis presents approaches that are robust to extreme values, non-stationarity, applicable to both short and long data length, transparent and adaptive to any financial/economic time series. The thesis provides step-by-step procedures in analyzing economic/financial indicators by incorporating concepts based on surrogate data method, wavelets, phase space embedding, 'delay vector variance’ (DVV) method and recurrence plots. The thesis also centers on transparent ways of identifying, dating turning points, evaluating impact of economic and financial crisis. In particular, the thesis also provides a procedure on how to anticipate future crisis and the possible impact of such crisis. The thesis shows that the incorporation of these techniques in learning the structure and interactions within and between economic and financial variables will be very useful in policy-making, since it facilitates the selection of appropriate processing methods, suggested by the data itself.

In addition, a novel procedure to test for linearity and unit root in a nonlinear framework is proposed by introducing a new model – the MT-STAR model – which has similar properties of the ESTAR model but reduces the effects of the identification problem and can also account for asymmetry in the adjustment mechanism towards equilibrium. The asymptotic distributions of the proposed unit root test is non-standard and is derived. The power of the test is evaluated through a simulation study and some empirical illustrations on real exchange rates show its accuracy. Finally, the thesis defines a multivariate Self–Exciting Threshold Autoregressive with eXogenous input (MSETARX) models and present an estimation procedure for the parameters. The modeling procedure for the MSETARX models and problems of estimation are briefly considered.
Résumé

L’axe principal de la thèse est centré sur des approches non-linéaires modernes d’analyse des données économiques et financières, avec une attention particulière sur les cycles économiques et les crises financières. Un consensus dans la littérature statistique et financière s’est établi autour du fait que les variables économiques ont un comportement non-linéaire au cours des différentes phases du cycle économique. En tant que tel, les approches/modèles non-linéaires sont requis pour saisir les caractéristiques du mécanisme de génération des données intrinsèquement asymétriques, que les modèles linéaires sont incapables de reproduire.

À cet égard, la thèse propose une nouvelle approche interdisciplinaire et ouverte à l’analyse des systèmes économiques et financiers. La thèse présente des approches robustes aux valeurs extrêmes et à la non-stationnarité, applicables à la fois pour des petits et de grands échantillons, aussi bien pour des séries temporelles économiques que financières. La thèse fournit des procédures dites étape par étape dans l’analyse des indicateurs économiques et financiers en intégrant des concepts basés sur la méthode de substitution de données, des ondelettes, espace incorporation de phase, la méthode retard vecteur variance (DVV) et des récurrences parcelles. La thèse met aussi en avant des méthodes transparentes d’identification, de datation des points de retournement et de l’évaluation des impacts des crises économiques et financières. En particulier, la thèse fournit également une procédure pour anticiper les crises futures et ses conséquences. L’étude montre que l’intégration de ces techniques dans l’apprentissage de la structure et des interactions au sein et entre les variables économiques et financières sera très utile dans l’élaboration de politiques de crises, car elle facilite le choix des méthodes de traitement appropriées, suggérées par les données.

Acknowledgements

This PhD thesis was written in the period from September 2011 to January 2014 during my studies at the Centre d’Économie de la Sorbonne (CES), Université Paris I - Panthéon Sorbonne and the Scuola Superiore di Economia at Università Ca’Foscari of Venice. I am grateful to European Doctorate in Economics Erasmus Mundus (EDEEM), funded by the European Commission’s Education, Audiovisual and Culture Executive Agency (EACEA), for providing excellent research facilities and a stimulating environment.

A number of people have contributed to the making of this thesis. First of all, I would like to thank my thesis advisors Monica Billio and Dominique Guégan for the excellent support and guidance. I really appreciate their patience, enthusiasm, encouragement and inspiration to me. My sincere thanks also goes to Michael Rockinger, Massimiliano Caporin, and Philippe de Peretti accepting and making time to be a part of my PhD Jury.

I thank the conference participants of CFE-ERCIM 2011 London, ISCEF 2012 Tunis, MAF 2012 Venice, EBIM-EDEEM-Paris1 Doctoral School Jamboree 2012 Paris, CFE-ERCIM 2012 Oviedo, COMPSTAT 2012 Limassol, EBIM-EDEEM Doctoral School Jamboree 2013 Louvain-La-Neuve, and CFE-ERCIM 2013 London. I would also like to thank the participants at the Econometrics Internal Seminar at Center for Operations Research and Econometrics (CORE) for their participation and interest. I am also indebted to a number of people who have given many valuable comments and suggestions, especially Roberto Casarin, Sébastien Van Bellegem, Luc Bauwens, Christian Hafner, Timo Terasvirta, and Yukai Kevin Yang. I am grateful to all anonymous referees and editors, for all my journal submissions, for their careful revision, valuable suggestions, and comments that have significantly improved the original version of the contributions.

My sincere thanks also goes to all EDEEM staff. I would like to thank Bernard Cornet, Jean-Marc Bonnissieux, François Maniquet and Bertrand Wigniolle for their support.

Last but not the least, I would like to thank my family: my parents Ruth Okai Addo and Emmanuel K. Addo, for giving birth to me at the first place and supporting me spiritually throughout my life. I would also like to Raymond Tetteh Addo and Lydia Dede Addo for their love, prayers and support.

On a different note, many people have been a part of my graduate education and I am highly grateful to all of them.

Peter Martey Addo, Paris, April 2014.
# Contents

Abstract ii

Acknowledgements iv

List of Figures ix

List of Tables xiii

1 Introduction 1

1.1 Motivation and Challenges 1

1.1.1 Characterization and Detection of Nonlinear Schemes in economics and financial indicators 2

1.1.2 Financial and Economic Crisis: Detection, Characterization and Turning Point Chronology 2

1.1.3 Nonlinear Models and applications 3

1.2 Contributions 3

1.3 Outline of Thesis 5

2 Nonlinear Dynamics and Wavelets for Business Cycle Analysis 7

2.1 Introduction 7

2.2 Background: Wavelet Analysis and 'Delay Vector Variance' Method 11

2.2.1 Wavelet and Wavelet Analysis 11

2.2.1.1 Choice of Wavelet 12

2.2.2 Surrogate Data Method 13

2.2.3 Optimal Embedding Parameters 15

2.2.4 'Delay Vector Variance’ method 16

2.3 Data Analysis 19

2.4 Conclusion 26

3 Turning point chronology for the Euro-Zone: A Distance Plot Approach 28

3.1 Introduction 28

3.2 Methodology based on recurrence plots 30

3.3 Data Analysis 33

3.4 Conclusion 38
4 Nonlinear dynamics and recurrence plots for detecting financial crisis 42
  4.1 Introduction .......................................................... 42
  4.2 'Delay Vector Variance' Method and Recurrence Plots ............ 44
  4.2.1 Surrogate Data Method and Statistical testing .................. 44
  4.2.2 Optimal Embedding Parameters .................................. 45
  4.2.3 'Delay Vector Variance' method ................................ 47
  4.2.4 Data Analysis based on recurrence plots ....................... 48
  4.3 Data Analysis on S&P 500 and NASDAQ Composite ................. 51
  4.4 Conclusion ........................................................... 60
5 The Univariate MT-STAR Model and a new linearity and unit root test procedure 62
  5.1 Introduction ........................................................... 62
  5.2 Background ........................................................... 64
  5.2.1 Overview of STAR models and Notations ....................... 64
  5.2.2 Drawbacks of the ESTAR model .................................. 65
  5.3 A new Model: The MT-STAR model ................................ 65
  5.3.1 A general MT-STAR(n,p) model .................................. 66
  5.3.2 MT-STAR(1,1) modelling .......................................... 67
  5.4 Linearity and Non Stationarity Tests ................................ 68
  5.4.1 Linearity Test ....................................................... 69
  5.4.2 Test for Non Stationarity in the MT-STAR framework ........... 71
  5.5 Monte Carlo Study ..................................................... 74
  5.5.1 Simulation results .................................................. 76
  5.6 Empirical Illustration ................................................ 77
  5.6.1 Application to the Euro Real Effective Exchange Rate .......... 79
  5.6.2 Application to Five Bilateral Real Exchange Rates relative to the US Dollar .................................................. 80
  5.6.3 Application to Bilateral Real Exchange Rates relative the Euro 83
  5.7 Conclusion ............................................................. 84
6 Multivariate Self-Exciting Threshold Autoregressive Models with eXogenous Input {PRELIMINARY VERSION– please do not quote} 86
  6.1 Introduction ........................................................... 86
  6.2 Multivariate SETARX models ........................................ 88
  6.3 On the Stationarity of MSETARX model ............................. 90
  6.4 Estimation of model parameters ..................................... 92
  6.4.1 Standard LSE Algorithm for Parameter Estimation ............ 92
  6.4.2 Algorithm for Adaptive Parameter Estimation .................. 93
  6.4.3 Simulations .......................................................... 93
  6.5 Concluding remarks ................................................... 95
A Appendix for Chapter 2 97
  A.1 Wavelet Analysis ..................................................... 97
  A.1.1 Notation and Operators .......................................... 97
A.1.2 Continuous Wavelet Transform (CWT) ......................................... 97
A.1.3 Maximal Overlap Discrete Wavelet Transform ............................ 100
A.2 DVV Plots of simulated Processes .................................................. 100

B  Appendix for Chapter 5  ................................................................. 102
B.1 Functional Central Limit Theorem (Recall) ................................. 102
B.2 Proof of Proposition 1 ................................................................. 103
B.3 Proof of Theorem 1 ................................................................. 108
B.4 Proof of Corollary ................................................................. 112

C  Appendix for Chapter 6  ................................................................. 115
C.1 Estimation Results ................................................................. 115

Bibliography  ................................................................. 117
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Complex Morlet wavelet with $f_b = 1$ and $f_c = 0.5$</td>
<td>13</td>
</tr>
<tr>
<td>2.2</td>
<td>Nonlinear and deterministic nature of signals. The first row of Diagrams 4.1(a) and 4.1(b) are DVV plots for a linear benchmark signal: AR(2) signal and a nonlinear benchmark signal: Henon signal, where the red line with crosses denotes the DVV plot for the average of 25 WiAAFT-based surrogates while the blue line denotes that for the original signal. The second row of Diagrams 4.1(a) and 4.1(b) denote the DVV scatter diagrams for those two signals, where error bars denote the standard deviation of the target variances of surrogates.</td>
<td>17</td>
</tr>
<tr>
<td>2.3</td>
<td>US Industrial Production Index (IPI) time series. Diagram 3.1 is the plot of the monthly IPI series for the period: 1919:01 - 2012:07 ($n = 1123$), where the shaded regions corresponds to the US recessions from 1920 published by NBER. Diagram 2.3(b) is the plot of the Spectrum of IPI..</td>
<td>19</td>
</tr>
<tr>
<td>2.4</td>
<td>The optimal embedding parameters obtained via the Differential-Entropy based method using the two types of surrogates are indicated as an open circle in the diagrams with a clear structure. We obtain a lower entropy $R_{ent}(m, \tau)$ with iAAFT surrogates which corresponds to a higher degree of structure. The values will be used in creating delay vectors needed for the DVV analysis.</td>
<td>20</td>
</tr>
<tr>
<td>2.5</td>
<td>This is the DVV analysis with iAAFT surrogates performed on the IPI using the embedding parameters obtained via the differential entropy-based method. We clearly observe a deviation from the bisector on the DVV scatter diagram. The DVV plot also indicates that the process is neither strictly deterministic or strictly stochastic. Thus, the original time series, IPI, exhibits nonlinear dynamics since the surrogates are linear realizations of the original ([119, 120]).</td>
<td>21</td>
</tr>
<tr>
<td>2.6</td>
<td>The optimal embedding parameters obtained via the Differential-Entropy based method using the two types of surrogates are indicated as an open circle in the diagrams with a clear structure. The values of the embedding parameters are $m = 2$ and $\tau = 1$ in both diagrams. This result $\tau = 1$ indicates the presence of time correlations, in the growth rate of IPI, implying a higher degree of structure, thus, a lower amount of disorder.</td>
<td>22</td>
</tr>
<tr>
<td>2.7</td>
<td>Phase Space reconstruction using the embedding parameters $m = 2$ and $\tau = 1$. This represents the embedding of the underlying time series, growth rate of IPI, in phase space. The attractor is clearly visualized.</td>
<td>22</td>
</tr>
<tr>
<td>2.8</td>
<td>The DVV analysis on Growth rate of IPI indicates that it is characterize by nonlinear dynamics.</td>
<td>23</td>
</tr>
</tbody>
</table>
2.9 Coefficients plots obtained from the CWT using complex Morlet wavelet on the growth rate of IPI: First row represents the phase (angle) plot and second row is the corresponding Modulus plot. The colormap ranges from blue to red, and passes through the colors cyan, yellow, and orange. The blue regions on the Angle Coefficient plot corresponds to periods of relative stagnation in the economy from 1920. Thus, we consider only such structures with a minimum of six months as recession in the economy. The corresponding amplitudes can be read from the Modulus plot.

2.10 The IPI growth rate and associated Scalogram from the CWT. The bar with colour scales on the left-hand side of the scalogram plot indicates the percentage of energy for each wavelet coefficient. Higher energy levels can be clearly observed for the Great Depression of the 1930s compared to the period of late-2000s financial crisis, also known as the Global Financial Crisis.

2.11 The pseudo-frequency associated to scale, in Hertz (Hz). The horizontal axis represents the scales and the vertical axis corresponds to the frequency associated to a scale.

2.12 A contour representation of Scalogram, figure 2.10, associated with the US IPI growth rate.

3.1 US Industrial Production Index (IPI) time series. The plot of the monthly IPI series for the period: 1919:01 - 2012:07 ($n = 1123$), where the shaded regions corresponds to the US recessions from 1920 published by NBER.

3.2 The plot of the monthly Euro-zone Industrial Production Index (IPI) time series for the period: 1971:01 - 2011:12 ($n = 492$).

3.3 This is an unthresholded recurrence plot which is sometimes termed distance plot ([74]). It is a matrix plot with calendar time on both the vertical and horizontal axes. The colormap corresponds to the distance to the next recurrence of a state in the time series. This representation shows the distances between states and enhance understanding of the phase space trajectory of US IPI. The existence of butterfly-like structures, of a minimum size of six months, along the main diagonal (bisector) indicates economic crisis. In this case, the start and end of a economic crisis corresponds the start and end on the formation of butterfly-like structure along the main diagonal.

3.4 The distance plot for the Euro-zone IPI and it’s embedded version. The existence of butterfly-like structures, of a minimum size of six months, along the main diagonal (bisector) indicates economic crisis. Embedding the original time series does not yield different results in terms of identification of the turning points associated with the business cycle.

3.5 The shaded regions corresponds to the Butterfly dates for the Euro Area Business Cycles from 1971:01–2011:12. The end of the last recession period that started 2011:07(q3) was not determined considering the sample period.

3.6 The distance plots of US IPI and Euro-zone IPI for the period 1991:01 - 2012:07. The top figure corresponds to the distance plots of US IPI. The second distance plot is for the Euro-zone IPI.

3.7 This is a cross recurrence plot to visualize the simultaneous occurrence of a similar state in both US IPI and Euro IPI.
3.8 The distance plots on monthly IPI series of some European countries: France, Germany and Italy. The data is obtained from [22] and Eurostat, and spans over the period: 1970:01 to 2012:11. 40

4.1 Nonlinear and deterministic nature of signals. The first row of Diagrams 4.1(a) and 4.1(b) are DVV plots for a linear benchmark signal: AR(2) signal and a nonlinear benchmark signal: Henon signal, where the red line with crosses denotes the DVV plot for the average of 25 WiAAFT-based surrogates while the blue line denotes that for the original signal. The second row of Diagrams 4.1(a) and 4.1(b) denote the DVV scatter diagrams for those two signals, where error bars denote the standard deviation of the target variances of surrogates. 48

4.2 The daily adjusted closing price of S&P 500 Index and its associated rate of change for the time period 1990-01-02 to 2012-08-31. Thus, the underlying financial time series is of length 5716 observations. 52

4.3 The daily adjusted closing price of Nasdaq Composite and its associated returns for the time period 1990-01-02 to 2012-08-31. This implies the time series is of length 5716 observations. 52

4.4 The optimal embedding parameters for S&P 500 Index obtained via the Differential-Entropy based method are indicated as an open circle in the diagrams with a clear structure. 53

4.5 The optimal embedding parameters for Nasdaq Composite obtained via the Differential-Entropy based method are indicated as an open circle in the diagrams with a clear structure. 53

4.6 This is the DVV analysis with iAAFT surrogates performed on S&P 500 using the embedding parameters obtained via the differential entropy-based method. We clearly observe a deviation from the bisector on the DVV scatter diagram. The DVV plot also indicates that the process is neither strictly deterministic or strictly stochastic. Thus, the original time series, S&P 500, exhibits nonlinear dynamics since the surrogates are linear realizations of the original ([119, 120]). 54

4.7 This is the DVV analysis with iAAFT surrogates performed on Nasdaq Composite using the embedding parameters obtained via the differential entropy-based method. We clearly observe a deviation from the bisector on the DVV scatter diagram. The DVV plot also indicates that the process is neither strictly deterministic or strictly stochastic. Thus, Nasdaq Composite exhibits nonlinear dynamics since the surrogates are linear realizations of the original. 54

4.8 This is an unthresholded recurrence plot which is sometimes termed distance plot ([74]). It is a matrix plot with calendar time on both the vertical and horizontal axes for the time period 1990-01-02 to 2012-08-31. The colormap corresponds to the distance to the next recurrence of a state in the time series. This representation shows the distances between states and enhance understanding of the phase space trajectory of S&P 500 and embedded version of S&P 500. The existence of butterfly-like structures, of a minimum size of six months, along the main diagonal (bisector) indicates a financial crisis. In this case, the start and end of a financial crisis corresponds the start and end on the formation of butterfly-like structure along the main diagonal. 55
4.9 The distance plot of the Nasdaq Composite and it’s embedded version. It is observed that the Nasdaq was well exposed and hit by the dot-com bubble of the 1990s compared to S&P 500. Nasdaq, on which many dot-coms traded, rose to record highs as indicated by larger distance to next recurrence.

4.10 This is a cross recurrence plot to visualize the simultaneous occurrence of a similar state in both S&P 500 and Nasdaq Composite. The dot-com bubble clearly occurs in both series.

4.11 The distance plot of the growth rate (returns) of S&P 500 and Nasdaq Composite. The growth rate cycle is often subject to very short-term fluctuations mainly due to transitory events making the peaks of such cycles extremely difficult to date. In this case, we are enable to visualize and date these peaks in both underlying series. The peaks dates are recognized by visualizing higher distances to the next recurrence.

4.12 The distance plot on an extended monthly S&P 500 stock price Index: original data range 1990-01-01 to 2012-08-01 with an additional 12 months ahead forecast obtained via the additive nonlinear autoregressive model (4.7). There is no signal of a possible occurrence of financial crisis in about six months from August 2012. The last butterfly-like structure observed started in March, 2012 and ended October 2012.

4.13 The distance plot on an extended monthly S&P 500 price index: original data range 1990-01 to 2008-02 with an additional 3 months ahead forecast obtained via the additive nonlinear autoregressive model (4.7). The region of the two ellipse drawn on the plot shows that crisis that began in July 2007 is a recurrence of the financial crisis that resulted from the collapse of the dot-coms bubble. Thus, the possible impact and duration of the financial crisis that began in July 2007 will be similar to the dot-com bubble.

5.1 Logarithm of euro real effective exchange rates (October, 1980 to October, 2011).

5.2 Normalized real exchange rates relative to the US dollar against UK, Japan, Germany, France and Switzerland (January, 1973 to June, 2008).

5.3 Bilateral real exchange rates relative to the euro (January, 1999 to November, 2011).

6.1 Two-dimensional MSETARX process with six-regimes, delay $d = 6$, autoregressive order $p = 3$ and $\Lambda^{(J)} = 0$ for all $J \in \{1, \cdots, L\}^D$ in equation (6.5). This is a signal section of 500 time samples of the multivariate process.

A.1 DVV analysis on ARIMA and Threshold Autoregressive signals.

A.2 DVV analysis on GARCH and Bilinear signals.
List of Tables

2.1 Results of the non-parametric rank-order test. The null of linearity is rejected as soon as the \textit{Rank-Order} is greater than \textit{Rank-Threshold}. The code $H$ takes the value 0 or 1, where $H = 0$ corresponds to failure of rejecting the null of linearity and $H = 1$ the rejection of linearity for nonlinearity. The number of iAAFT surrogates considered for the DVV-analysis is 25, which is greater than the minimum requirement of 19 surrogates for testing at $\alpha = 0.05$ level of significance. 21

2.2 Results of the non-parametric rank-order test. The null of linearity is rejected as soon as the \textit{Rank-Order} is greater than \textit{Rank-Threshold}. The code $H$ takes the value 0 or 1, where $H = 0$ corresponds to failure of rejecting the null of linearity and $H = 1$ the rejection of linearity for nonlinearity. The number of surrogates considered for the DVV-analysis is 25, which is greater than the minimum requirement of 19 surrogates for testing at $\alpha = 0.05$ level of significance. 23


3.1 Business Cycle Peaks and Troughs in the United States, 1920-2009. The peak and trough dates, in the format YYYY:MM, represent the start and end of “episodes” of some sort. The column \textit{NBER} are the reference chronology for US business cycles published at http://www.nber.org/cycles.html. The results of turning points obtained from the \textit{distance plot} of US IPI is displayed in column labelled \textit{butterfly} dates. 36

3.2 Industrial business cycle dating chronology for the Euro-zone from 1971:01–2011:12. The peak and trough dates, in the format YYYY:MM, represent the start and end of “episodes” of some sort. The end of the last recession period that started 2011:07(q3) was not determined considering the sample period. 37

3.3 Comparison of turning points for the Euro-Zone. The format of the dates “yyyy:mm(qq)” denotes “year:month(quarter)”. The \textit{butterfly} dates correspond to peaks and troughs identified by the \textit{distance plot} when applied to the monthly series of Euro-zone IPI. The turning points obtained in [10] for the monthly Euro-zone IPI series is denoted as [10] dates. The \textit{CEPR} dates are the quarterly turning points determined by the CEPR using the Euro area real GDP. \textit{MU} dates are turning points identified by the Bry-Boschan algorithm when applied to the Moench/Uhlig (MU) monthly series of Euro area real GDP. The end of the last recession period that started 2011:07(q3) was not determined considering the sample period 1971:01–2011:12. 41
4.1 Results of the non-parametric rank-order test. The null of linearity is rejected as soon as the Rank-Order is greater than Rank-Threshold. The code $H$ takes the value 0 or 1, where $H = 0$ corresponds to failure of rejecting the null of linearity and $H = 1$ the rejection of linearity for non-linearity. The number of iAAFT surrogates considered for the DVV-analysis is 25, which is greater than the minimum requirement of 19 surrogates for testing at $\alpha = 0.05$ level of significance.  

4.2 This is the 12 month ahead forecast of monthly S&P 500 stock price Index based on the additive nonlinear autoregressive model (4.7).  

5.1 Asymptotic critical values of ABG statistic. Note: Case 1, Case 2 and Case 3 refer to the underlying model with the raw data, the de-meaned data and the de-trended data, respectively.  

5.2 The size of alternative tests [in %] at nominal level $\alpha = 5\%$.  

5.3 The power of alternative tests against the hypothesis of global MT-STAR stationarity [in %] at nominal level $\alpha = 5\%$: Case 1 with $\sigma = 0.1$.  

5.4 The power of alternative tests against the hypothesis of global MT-STAR stationarity [in %] at nominal level $\alpha = 5\%$: Case 1 with $\sigma = 1$.  

5.5 The power of alternative tests against the hypothesis of global MT-STAR stationarity [in %] at nominal level $\alpha = 5\%$: Case 2 with $\sigma = 0.1$.  

5.6 The power of alternative tests against the hypothesis of global MT-STAR stationarity [in %] at nominal level $\alpha = 5\%$: Case 2 with $\sigma = 1$.  

5.7 The power of alternative tests against the hypothesis of global MT-STAR stationarity [in %] at nominal level $\alpha = 5\%$: Case 3 with $\sigma = 0.1$.  

5.8 The power of alternative tests against the hypothesis of global MT-STAR stationarity [in %] at nominal level $\alpha = 5\%$: Case 3 with $\sigma = 1$.  

A.1 Notations and Operators.
Dedicated to the Victims of Typhoon Haiyan in the Philippines (November, 2013).
Chapter 1

Introduction

1.1 Motivation and Challenges

It is now well stated in the statistical and economic literature that major economic variables display a non-linear behaviour over the different phases of the business cycle. In particular, asymmetry has been recognized as a nonlinear phenomenon in several recent studies investigating various economic and financial time series. Nonlinear models are therefore required to capture the features of the data generating mechanism of inherently asymmetric realizations of some of the macroeconomic business cycle series, since linear models are incapable of generating such behavior. In particular, nonlinearity in an economic indicator conveys information on possible existence of different states of the world or regimes in the economy. In real-world applications of economic time series analysis, the process underlying the generated signal, which is the time series, are a priori unknown. In this respect, the thesis provides an interdisciplinary and open-minded approach to analyzing economic and financial systems in a novel way. The thesis presents approaches that are robust to extreme values, non-stationarity, applicable to both short and long data length, transparent and adaptive to any financial/economic time series.

In this respect, the thesis contributions can be placed under the themes: a) Characterization and Detection of Nonlinear Schemes in economics and financial indicators; b) Financial and Economic Crisis: Detection, Characterization and Turning Point Chronology and c) Nonlinear Models and applications.
1.1.1 Characterization and Detection of Nonlinear Schemes in economics and financial indicators

An unavoidable step in nonlinear modeling is testing linearity against nonlinearity before fitting any of the nonlinear models. However, most linearity testing against nonlinearity in literature, usually require a specification of a stationary nonlinear model under the alternative hypothesis. This makes such tests restrictive to the dynamics of the specified nonlinear model. There is therefore a need to use procedures that tests linearity against any form of nonlinearity in the economic time series.

In line with this objective, we provide a signal modality analysis to distinguish between linearity and nonlinearity by incorporating the surrogate data methodology in to the recently proposed ‘Delay Vector Variance’ (DVV) method. The approach can be applicable to non-stationary data, requires no a priori assumptions on the statistical properties of the data and requires no specification on the exact form (model) of nonlinearity under the alternative hypothesis. Statistical testing of the null of linearity using a non-parametric rank-order test is performed to enhance robust conclusion of results obtained via the DVV-analysis.

We have successfully used this method in understanding exchange rates dynamics, detecting nonlinearity in both business cycle indicators and financial stock series.

1.1.2 Financial and Economic Crisis: Detection, Characterization and Turning Point Chronology

We propose a transparent way of establishing a turning point chronology for the business cycle of any economic system (i.e country or region). Our analysis is achieved by exploiting the concept of recurrence plots, in particular distance plots, to detect and data turning points as well as evaluating the impact of economic crisis. A comprehensive analysis of the feasibility of this approach applied to financial data and economic time series is provided. Our proposed methodology is applicable to both univariate and multivariate time series. We have provided a step-by-step procedure which incorporates concepts on surrogate data method, wavelets, phase space embedding, ‘delay vector variance’ (DVV) method and recurrence plots in order to anticipate future crisis and even the possible impact before they happen. Our findings indicates that the turning points provided by NBER for the US can be obtained with only Industrial Production Index data using the distance plot approach.

The findings from the data analysis with recurrence plots, in this case distance plots, shows that these plots are robust to extreme values, non stationarity and applicable
to both short and long data length. This approach is also replicable and transparent; is adaptive to any time series. In particular, we show that this approach provides a transparent chronology of business cycles since it avoids revision of crisis dates through time. Moreover, our methodology can be used to provide early warning signals of crisis especially when applied to variables which are predictors of such crisis.

1.1.3 Nonlinear Models and applications

We present different statistical models able to take into account certain nonlinearities and their most recent extensions. Special attention is paid to nonlinear models that rely on state-dependent or regime-switching behavior. We compare those nonlinear models developed in univariate and multivariate modeling, both from a theoretical and empirical point of view. The main models that will be developed and extended are the Self-Exciting Threshold Autoregressive (SETAR) models and Smooth Transition Autoregressive (STAR) models.

In this respect, we propose the MT-STAR model which generalizes the T-STAR model proposed in literature by introducing asymmetric adjustment towards equilibrium and presents a new unit root test in a nonlinear framework which contributes to the existing literature in unit root tests. We also derive the limit distributions of this new test and give some empirical applications.

Concerning the multivariate setting, there is work in progress on developing a Multivariate Self-Exciting Threshold Autoregressive (MSETARX) Models with Exogenous inputs. Special attention is paid to establishing a complete modeling strategy for such a multivariate nonlinear model and providing some empirical applications.

1.2 Contributions

The main contributions of this work are to introduce modern non-linear interdisciplinary and open-minded approaches for data analysis in economics and Finance with special attention on business cycles and financial crisis.

In today’s globalized economy, given the increasing importance of central bank’s credibility especially in times of economic crisis, it is reputation that is key. This key policy issue is linked to the continued action via financial stability and monetary policy measures to improve the economic health of the nation or region the bank represents.

Following the recent financial crisis, Analysts could confirm that most conclusion drawn from existing models may not been appropriate for policy-making. These claims are valid
since existing models do not account for the increased interconnectedness, increased risk of recession, and the increased sensitivity of systemic failure. It is therefore essential to conduct research on the identification of risks in the financial system and the detection of early warning signals for systemic risk to enhance better policy-making decisions to prevent and manage such risks. In other words, obtaining early warning signals before economic crisis will aid the central banks in responding in such a way as to ensure stability of the global financial system. Thus, the central banks will be safe from the reputational damage that could take decades to repair.

Asymmetry has been recognized as a nonlinear phenomenon in several recent studies investigating various economic and financial time series. As such, it is therefore necessary to introduce modern approaches for nonlinear data analysis of economic and financial indicators via interdisciplinary research. This approaches need to be robust to extreme values, non stationarity, applicable to both short and long data length, transparent and adaptive to any time series. For instance, a section of the thesis centers on transparent ways of identifying, dating turning points, evaluating impact of economic and financial crisis. In particular, the study also provides a procedure on how to anticipate future crisis and the possible impact of such crisis. In addition, the incorporation of some modern signal processing and machine learning techniques in learning the structure and interactions within and between economic and financial variables will be very useful in policy-making, since it facilitates the selection of appropriate processing methods, suggested by the data itself. Moreover, the identification of transmission channels and how the network structure of the financial system change over time will aid in drawing meaningful conclusions for appropriate policy-making. In this respect, research will help policymakers safeguard the reputation of central banks in today’s globalized economy.

The work in this thesis has contributed in part or full to the following publications:


1.3 Outline of Thesis

This prelude has introduced the idea of introducing modern non-linear approaches for data analysis in economics and Finance with special attention on business cycles and financial crisis. The Chapters of the thesis correspond to the contributions made via publication.

Chapter 2 provides a signal modality analysis to characterize and detect nonlinearity schemes in the US Industrial Production Index time series. The analysis is achieved by using the recently proposed ‘delay vector variance’ (DVV) method, which examines local predictability of a signal in the phase space to detect the presence of determinism and nonlinearity in a time series. Optimal embedding parameters used in the DVV analysis are obtained via a differential entropy based method using Fourier and wavelet-based surrogates. A complex Morlet wavelet is employed to detect and characterize the US business cycle. A comprehensive analysis of the feasibility of this approach is provided. The results coincide with the business cycles peaks and troughs dates published by the National Bureau of Economic Research (NBER).

Chapter 3 proposes a transparent way of establishing a turning point chronology for the Euro-zone business cycle. Our analysis is achieved by exploiting the concept of recurrence plots, in particular distance plots, to characterize and detect turning points of the business cycle. Firstly, we apply the concept of recurrence plots on the US Industrial Production Index (IPI) series: this serves as a benchmark for our analysis since it already exists a reference chronology for the US business cycle, provided by the Dating Committee of the National Bureau of Economic Research (NBER). This concept is then used to construct a turning point chronology for the Euro-zone business cycle. In particular, study shows that this approach permits to detect turning points.
and study the business cycle without \textit{a priori} assumptions on the statistical properties of the underlying economic indicator.

Chapter 4 centers on the detection and characterization of financial crisis. Identification of financial bubbles and crisis is a topic of major concern since it is important to prevent collapses that can severely impact nations and economies. Our analysis deals with the use of the recently proposed ‘delay vector variance’ (DVV) method, which examines local predictability of a signal in the phase space to detect the presence of determinism and nonlinearity in a time series. Optimal embedding parameters used in the DVV analysis are obtained via a differential entropy based method using wavelet-based surrogates. We exploit the concept of recurrence plots to study the stock market to locate hidden patterns, non-stationarity, and to examine the nature of these plots in events of financial crisis. In particular, the recurrence plots are employed to detect and characterize financial cycles. A comprehensive analysis of the feasibility of this approach is provided. We show that our methodology is useful in the diagnosis and detection of financial bubbles, which have significantly impacted economic upheavals in the past few decades.

In Chapter 5, a novel procedure to test for unit root in a nonlinear framework is proposed by first introducing a new model – the MT-STAR model – which has similar properties as the ESTAR model but reduces the effects of the identification problem and can also account for cases where the adjustment mechanism towards equilibrium is not symmetric. The limiting non-standard asymptotic distributions of the proposed unit root test is then derived. Finally, the power of the test is evaluated through a simulation study and some empirical illustrations are given at the end.

Chapter 6 presents a preliminary version of a work in progress on a multivariate nonlinear model. The recent financial crisis of 2007-2009 has lead to a need for regulators and policy makers to understand and track systemic linkages. As the events following the turmoil in financial markets unfolded, it became evident that modern financial systems exhibit a high degree of interdependence and nonlinearity making it difficult in predicting the consequences of such an intertwined system. This study defines a multivariate Self-Exciting Threshold Autoregressive with eXogenous input (MSETARX) models and present an estimation procedure for the parameters. The conditions for stationarity of the nonlinear MSETARX models is provided. The efficiency of an adaptive parameter estimation algorithm and LSE (least squares estimate) algorithm for this class of models is investigated via simulations. The modeling procedure for the MSETARX models and problems of estimation are briefly considered.
Chapter 2

Nonlinear Dynamics and Wavelets for Business Cycle Analysis

2.1 Introduction

In general, performing a nonlinearity analysis in a modeling or signal processing context can lead to a significant improvement of the quality of the results, since it facilitates the selection of appropriate processing methods, suggested by the data itself. In real-world applications of economic time series analysis, the process underlying the generated signal, which is the time series, are *a priori* unknown. These signals usually contain both linear and nonlinear, as well as deterministic and stochastic components, yet it is a common practice to model such processes using suboptimal, but mathematically tractable models. In the field of biomedical signal processing, e.g., the analysis of heart rate variability, electrocardiogram, hand tremor, and electroencephalogram, the presence or absence of nonlinearity often conveys information concerning the health condition of a subject (for an overview 69). In some modern machine learning and signal processing applications, especially biomedical and environmental ones, the information about the linear, nonlinear, deterministic or stochastic nature of a signal conveys important information about the underlying signal generation mechanism. In the analysis of economic indicators, the presence of nonlinearity in the data provides information about both structural and behavioral changes that can occur in the economy across time. In particular, nonlinearity in an economic indicator conveys information on possible existence of different *states of the world* or regimes in the economy. There has been an
increasing concerns on the forecasting performance of some nonlinear models in modeling economic variables. Nonlinear models often provide superior in-sample fit, but rather poor out-of-sample forecast performance (123). In cases where the nonlinearity is spurious or relevant for only a small part of the observations, the use of nonlinear models will lead to forecast failure (130). It is, therefore, essential to investigate the intrinsic dynamical properties of economic time series in terms of its deterministic/stochastic and nonlinear/linear components reveals important information that otherwise remains not clear in using conventional linear methods of time series analysis.

Since the early work by [29], many attempts have been made to measure and forecast business cycles. Many business cycle indicators present asymmetric features that have long been recognized in economics ([81, 103]). Putting it simply, there are sharp retractions during downturns in the economy as opposed to gradual upswings during recoveries ([17, 25, 83, 121]). Asymmetry has been recognized as a nonlinear phenomenon in several recent studies investigating various economic time series. Nonlinear models are therefore required to capture the features of the data generating mechanism of inherently asymmetric realizations of some of the macroeconomic business cycle series, since linear models are incapable of generating such behavior ([38, 54, 131, 132]). Many nonlinear models are only identified when the alternative hypothesis holds (the model is genuinely nonlinear) but not when the null hypothesis is valid. Since the parameters of an unidentified model cannot be estimated consistently, testing linearity before fitting any of these models is an unavoidable step in nonlinear modeling ([92, 145]). However, most linearity testing against nonlinearity in literature, usually require a specification of a stationary nonlinear model under the alternative hypothesis. This makes such tests restrictive to the dynamics of the specified nonlinear model. There is therefore a need to use procedures that tests linearity against any form of nonlinearity in the economic time series.

Several methods for detecting nonlinear nature of a signal have been proposed over the past few years. The classic ones include the 'deterministic versus stochastic' (DVS) plots (148), the Correlation Exponent and 'δ-ε' method (78). For our purpose, it is desirable to have a method which is straightforward to visualize, and which facilitates the analysis of predictability, which is a core notion in online learning. In this paper, we adopt to the recently proposed phase space based 'delay vector variance' (DVV) method (51), for signal characterization, which is more suitable for signal processing application because it examines the nonlinear and deterministic signal behavior at the same time. This method has been used for understanding the dynamics of exchange rates (5), detecting nonlinearity in financial markets (6), qualitative assessment of machine learning algorithms, analysis of functional magnetic resonance imaging (fMRI) data, as well as analyzing nonlinear structures in brain electrical activity and heart
rate variability (HRV) (52). Optimal embedding parameters will be obtained using a
differential entropy based method proposed in 50, which allows for simultaneous deter-
mination of both the embedding dimension and time lag needed for the DVV analysis.
Surrogate generation used in this study will be based on both the Iterative Amplitude
Adjusted Fourier Transform (iAAFT) ([119, 120]) and a recently refined iAAFT with a
wavelet-based approach, denoted WiAAFT (79).

Wavelet analysis has successfully been applied in a great variety of applications like
signal filtering and denoising, data compression, imagine processing and also pattern
recognition. The application of wavelet transform analysis in science and engineering
really began to take off at the beginning of the 1990s, with a rapid growth in the
numbers of researchers turning their attention to wavelet analysis during that decade
(56, 75, 114, 115). The wavelet transforms has the ability to perform local analysis of
a time series revealing how the different periodic components of the time series change
over time. Wavelets are able to locate precisely in time regime shifts, discontinuities, and
isolated shocks to a dynamical system. A wavelet approach has the ability to deal with
non-stationarity of stochastic innovations that are inevitably involved with economic and
financial time series (114). The maximum overlap discrete wavelet transform (MODWT)
has commonly been used by some economists (56, 49, 48, among others). The MODWT
can be seen as a kind of compromise between the discrete wavelet transform (DWT) and
the continuous wavelet transform (CWT); it is a redundant transform, because while it
is efficient with the frequency parameters it is not selective with the time parameters.
The CWT, unlike the DWT, gives us a large freedom in selecting our wavelets and yields
outputs that makes it much easier to interpret. The continuous wavelet transform has
emerged as the most favored tool by researchers as it does not contain the cross terms
inherent in the Wigner-Ville transform and Choi-Williams distribution (CWD) methods
while possessing frequency-dependent windowing which allows for arbitrarily high reso-
lution of the high frequency signal components. Moreover, the time invariance property
of the CWT implies that the wavelet transform of a time-delayed version of a signal is
a time-delayed version of its wavelet transform. This serves as an important property
in terms of pattern recognition. In other words, the identification of the business cycle
turning points for a subset of the entire time series does not change through time for
a given time series. From an economic point of view, this ensures an effective dating
chronology since it avoids revisions through time. This nice property of the CWT is not
readily obtained in the case of DWT and MODWT (2, 147).

Dating is an *ex post* exercise, and in this respect accuracy is an important criterion
since dating is useful for economic decision-making (7). Governments and central banks
are usually very sensitive to indicators showing signs of deterioration in growth to allow
them to adjust their policies sufficiently in advance, avoiding more deterioration or a
recession (11). As such, it will be interesting to choose a wavelet that will provide a better interpretation of the results from an economic point of view and also enhance accurate detection of the dates. In this respect, the choice of the wavelet is important. We are concerned with information about cycles and as such complex wavelets serves as a necessary and better choice. We need complex numbers to gather information about the phase, which, in turn, tells us the position in the cycle of the time-series as a function of frequency and the associated magnitude in this position. This will enable extraction of information about the economy-wide fluctuations in production that occur around a long-term growth trend. Thus, detecting and studying periods of relatively rapid economic growth (an expansion), and periods of relative stagnation or decline (recession) in the economy. There are many continuous wavelets to choose from; however, by far the most popular are the Mexican hat wavelet and the Morlet wavelet. In this work, we employ a complex Morlet wavelet which satisfies these requirements and has optimal joint time-frequency concentration, meaning that it is the wavelet which provides the best possible compromise in these two dimensions.

In this paper, we provide a novel methodology for business cycle modeling which encompasses different existing methods successfully applied in physics and engineering. In particular, we first study the structure of a chosen economic indicator \(^1\) via a phase-space representation using the differential entropy based method with both iAAFT and wavelet-based surrogates. We then use the DVV method to detect the nonlinear nature of the economic indicator using the values of the optimal embedding parameters obtained via the differential entropy method with surrogates. Finally, we perform wavelet analysis with a complex Morlet wavelet using a continuous wavelet transform to discover patterns or hidden information that cannot be captured with traditional methods of business cycle analysis such as spectral analysis. Our results is consistent with business cycle dates published by the National Bureau of Economic Research (NBER). We are able to detect these business cycle dates and study these fluctuations in the economy over frequency and time. This serves as an important finding in terms of forecasting and pattern recognition.

The paper is organized as follows: The concept of wavelet analysis and our choice of analyzing wavelet is presented in Section 2.2.1. Surrogate generation methodology and differential entropy based method for determining optimal embedding parameters of the phase-space representation of time series are then presented in Section 4.2.1 and Section 4.2.2 respectively. Lastly, we present, in Section 4.2.3, an overview of the ‘delay vector variance’ method with illustrations. In Section 2.3, we present a comprehensive

\(^1\)The economic indicator used in the work is the US Industrial Production Index. We remark that our methodology can easily be applied to any known business cycle indicator.
analysis of the feasibility of this approach to analyze the US Business cycle. Section 2.4 concludes.

### 2.2 Background: Wavelet Analysis and 'Delay Vector Variance’ Method

In this section, we present an overview of different existing methods successfully applied in physics and engineering. In particular, we show the usefulness of these methods over other methods and then explain how we merged these methods to business cycle modeling. Our methodology encompasses wavelet analysis, surrogate generation methods, differential entropy method for determining the optimal embedding parameters in phase-space, and the DVV method.

#### 2.2.1 Wavelet and Wavelet Analysis

It is a time-frequency signal analysis method which offers simultaneous interpretation of the signal in both time and frequency allowing local, transient or intermittent components to be elucidated. These components are often not clear due to the averaging inherent within spectral only methods like the fast Fourier transform (FFT).

A wavelet is a function, \( \psi \), which has a small concentrated burst of finite energy in the time domain and exhibits some oscillation in time. This function must be in the space of measurable functions that are absolutely and squared-integrable, i.e. \( \psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \), to ensure that the Fourier transform of \( \psi \) is well-defined and \( \psi \) is a finite energy signal. A single wavelet function generates a family of wavelets by dilating (stretching and contracting) and translating (moving along the time axis) itself over a continuum of dilation and translation values. If \( \psi \) is a wavelet analyzing function then the set \( \{ \tau_t D_s \psi \} \) of all the dilated (by \( s \neq 0 \)) and translated (by \( t \)) versions of \( \psi \) is that wavelet family generated by \( \psi \). Dilation in time by contracting values of scale (\( s > 1 \)) corresponds to stretching dilation in the frequency domain.

The basic concept in wavelet transforms is the projection of data onto a basis of wavelet functions in order to separate large-scale and fine-scale information (26). Thus, the signal is decomposed into a series of shifted and scaled versions of a mother wavelet function to make possible the analysis of the signal at different scales and resolutions. For reconstruction of a signal, it is necessary that \( \psi \) be such that \( \{ \tau_t D_s \psi \} \) span a large enough space of interest.
• Thus, every signal $f$ of interest should be representable as a linear combination of dilated and translated versions of $\psi$.

• Knowing all the inner products $\{(f, \tau_t D_s \psi)\}$, the signal should be recoverable.

The wavelet $\psi$ is assumed to satisfy the admissibility condition,

$$C_{adm, \psi} = \int_{\mathbb{R}} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty, \quad (2.1)$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(\xi)$, $\hat{\psi}(\omega) = \int_{\mathbb{R}} \psi(\xi)e^{-i\omega \xi} d\xi$. The admissibility condition (2.1) implies $\hat{\psi}(0) = \int_{\mathbb{R}} \psi(\xi) d\xi = 0$. For $s$ restricted to $\mathbb{R}^+$, the condition (2.1) becomes

$$C_{adm+, \psi} = \int_{0}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < \infty. \quad (2.2)$$

This means that the wavelet has no zero-frequency component. The value of the admissibility constant, $C_{adm, \psi}$ or $C_{adm+, \psi}$ depends on the chosen wavelet. This property allows for an effective localization in both time and frequency, contrary to the Fourier transform, which decomposes the signal in terms of sines and cosines, i.e. infinite duration waves.

There are essentially two distinct classes of wavelet transforms: the continuous wavelet transform and the discrete wavelet transform. We refer the reader to 2, 147 for a review on wavelet transforms. In this work, we employ a complex wavelet via a continuous wavelet transform (CWT) in order to separate the phase and amplitude information, because the phase information will be useful in detecting and explaining the cycles in the data. We provide in A.1.2 an overview of CWT and its relevance to our work.

### 2.2.1.1 Choice of Wavelet

The Morlet wavelet is the most popular complex wavelet used in practice. A complex Morlet wavelet (129) is defined by

$$\psi(\xi) = \frac{1}{\sqrt{\pi f_b}} e^{i2\pi f_c \xi - \frac{\xi^2}{f_b}} \quad (2.3)$$

depending on two parameters: $f_b$ and $f_c$, which corresponds to a bandwidth parameter and a wavelet center frequency respectively. The Fourier transform of $\psi$ is $\hat{\psi}(\zeta) = e^{-\pi^2 f_b (\zeta - f_c)^2}$, which is well-defined since $\psi \in L^1(\mathbb{R})$. It can easily be shown that the Morlet wavelet (2.3) is a modulated Gaussian function and involutive, i.e. $\psi = \hat{\psi}$. The Fourier transform $\hat{\psi}$ has a maximum value of 1 which occurs at $f_c$, since $\|\psi\|_1 := \int |\psi| = 1$. This wavelet has an optimal joint time-frequency concentration since it has
an exponential decay in both time and frequency domains, meaning that it is the wavelet which provides the best possible compromise in these two dimensions. In addition, it is infinitely regular, complex-valued and yields an exactly reconstruction of the signal after the decomposition via CWT.

In this work, the wavelet that best detects the US business is the complex Morlet wavelet with $f_b = 1$ and $f_c = 0.5$. In this case, the Morlet wavelet becomes

$$\psi(\xi) = \frac{1}{\sqrt{\pi}}e^{i\pi\xi - \xi^2},$$

which we will often refer to as Morlet wavelet. The nature of our choice of wavelet function and the associated center frequency is displayed in figure 2.1. It illustrates the oscillating nature of the wavelet with short duration of the time support. In other words, the wavelet is bounded, centered around the origin, and have time support (respectively frequency support).

![Complex Morlet wavelet with $f_b = 1$ and $f_c = 0.5$.](image)

**Figure 2.1:** Complex Morlet wavelet with $f_b = 1$ and $f_c = 0.5$

### 2.2.2 Surrogate Data Method

Surrogate time series, or 'surrogate' for short, is non-parametric randomized linear version of the original data which preserves the linear properties of the original data. For identification of nonlinear/linear behavior in a given time series, the null hypothesis that the original data conform to a linear Gaussian stochastic process is formulated. An established method for generating constrained surrogates conforming to the properties of a linear Gaussian process is the Iterative Amplitude Adjusted Fourier Transform
(iAAFT), which has become quite popular (85, 119, 120, 129). This type of surrogate time series retains the signal distribution and amplitude spectrum of the original time series, and takes into account a possibly nonlinear and static observation function due to the measurement process. The method uses a fixed point iteration algorithm for achieving this, for the details of which we refer to 119, 120.

Wavelet-based surrogate generation is a fairly new method of constructing surrogate for hypothesis testing of nonlinearity which applies a wavelet decomposition of the time series. The main difference between Fourier transform and wavelet transform is that the former is only localized in frequency, whereas the latter is localized both in time and frequency. The idea of a wavelet representation is an orthogonal decomposition across a hierarchy of temporal and spatial scales by a set of wavelet and scaling functions.

The iAAFT-method has recently been refined using a wavelet-based approach, denoted by WiAAFT (79), that provides for constrained realizations of surrogate data that resembles the original data while preserving the local mean and variance as well as the power spectrum and distribution of the original except for randomizing the nonlinear properties of the signal. The WiAAFT-procedure follows the iAAFT-algorithm but uses the Maximal Overlap Discrete Wavelet Transform (MODWT) where the iAAFT-procedure is applied to each set of wavelet detail coefficients $D_j(n)$ over the dyadic scales $2^{j-1}$ for $j = 1, \cdots, J$, i.e., each set of $D_j(n)$ is considered as a time series of its own. The main difference between iAAFT and wiAAFT algorithms is that the former is designed to produce constrained, linear realizations of a process that can be compared with the original time series on some measure, while the later algorithm restricts the possible class of realizations to those that retain some aspect of the local mean and variance of the original time series (80).

Statistical analysis by the concept of surrogate data tests for a difference between a test statistic computed for the original and linearized versions of the data, i.e., an ensemble of realizations of the null hypothesis linear dynamics. For statistical testing of the null hypothesis of linearity, we follow [135] by using a non-parametric rank-order test. The degree of difference between the original and surrogate data is given by the ranked position of the data asymmetry with respect to the surrogates. For a right-tailed test, we generate at least $N_s = \frac{1}{\alpha} - 1$ surrogates, where $\alpha$ is the level of significance and $N_s$ denotes the number of surrogates. The rank-threshold (or critical value) for right-tailed rank-order test is given by $(1 - \alpha)/(N_s + 1)$. The null of linearity is rejected as soon as the rank-order statistic is greater than the rank-threshold. To achieve a minimal significance requirement of 95% ($\alpha = 0.05$), we need at least 19 surrogates time series for right-tailed tests. Increasing the number of surrogates can increase the discrimination power (119, 120, 135). The concept of surrogate data will be incorporated into the Delay
Vector Variance method (below) to examine the dynamics of an underlying economic indicator.

2.2.3 Optimal Embedding Parameters

In the context of signal processing, an established method for visualizing an attractor of an underlying nonlinear dynamical signal is by means of time delay embedding (69). By time-delay embedding, the original time series \( \{x_k\} \) is represented in the so-called 'phase space' by a set of delay vectors (DV) of a given embedding dimension, \( m \), and time lag, \( \tau : x(k) = [x_{k-\tau}, \cdots, x_{k-m\tau}] \). 50 proposed a differential entropy based method for determining the optimal embedding parameters of a signal. The main advantage of this method is that a single measure is simultaneously used for optimizing both the embedding dimension and time lag. We provide below an overview of the procedure:

The “Entropy Ratio” is defined as

\[
R_{ent}(m, \tau) = I(m, \tau) + \frac{m \ln N}{N},
\]

where \( N \) is the number of delay vectors, which is kept constant for all values of \( m \) and \( \tau \) under consideration,

\[
I(m, \tau) = \frac{H(x, m, \tau)}{\langle H(x_{s,i}, m, \tau) \rangle_i}
\]

where \( x \) is the signal, \( x_{s,i} i = 1, \cdots, T \) surrogates of the signal \( x \), \( \langle \cdot \rangle_i \) denotes the average over \( i \), \( H(x, m, \tau) \) denotes the differential entropies estimated for time delay embedded versions of a time series, \( x \), which an inverse measure of the structure in the phase space. 50 proposed to use the Kozachenko-Leonenko (K-L) estimate (89) of the differential entropy given by

\[
H(x) = \sum_{j=1}^{T} \ln(T \rho_j) + \ln 2 + C_E
\]

where \( T \) is the number of samples in the data set, \( \rho_j \) is the Euclidean distance of the \( j \)-th delay vector to its nearest neighbor, and \( C_E(\approx 0.5772) \) is the Euler constant. This ratio criterion requires a time series to display a clear structure in the phase space. Thus, for time series with no clear structure, the method will not yield a clear minimum, and a different approach needs to be adopted, possibly one that does not rely on a phase space representation. When this method is applied directly to a time series exhibiting strong serial correlations, it yields embedding parameters which have a preference for \( \tau_{opt} = 1 \). In order to ensure robustness of this method to the dimensionality and serial correlations of a time series, 50 suggested to use the iAAFT method for surrogate generation.
since it retains within the surrogate both signal distribution and approximately the autocorrelation structure of the original signal. In this Paper, we opt to use wavelet-based surrogate generation method, WiAAFT by in 79, for reasons already discussed in the previous section.

2.2.4 'Delay Vector Variance' method

The characterization of signal nonlinearities, which emerged in physics in the mid-1990s, have been successfully applied in predicting survival in heart failure cases and also adopted in practical engineering applications (70, 101). The 'delay vector variance' (DVV) method (51) is a recently proposed phase space based method for signal characterization. It is more suitable for signal processing application because it examines the deterministic\(^2\) nature of a time series and when combined with the concept of surrogate data, provides as additional account of the nonlinear behavior of the time series. The DVV-analysis is based on the calculation of the target variance, \(\sigma^*^2\), which is an inverse measure of the predictability of a time series. The algorithm is summarized below:

- For an optimal embedding dimension \(m\) and time lag \(\tau\), generate delay vector (DV): \(x(k) = [x_{k-\tau}, \ldots, x_{k-m\tau}]\) and corresponding target \(x_k\)

- The mean \(\mu_d\) and standard deviation, \(\sigma_d\), are computed over all pairwise distances between DVs, \(|x(i) - x(j)|\) for \(i \neq j\).

- The sets \(\Omega_k\) are generated such that \(\Omega_k = \{x(i) | \|x(k) - x(i)\| \leq \varrho_d\}\), i.e., sets which consist of all DVs that lie closer to \(x(k)\) than a certain distance \(\varrho_d\), taken from the interval \([\min\{0, \mu_d - n_d\sigma_d\}; \mu_d + n_d\sigma_d]\), e.g., uniformly spaced, where \(n_d\) is a parameter controlling the span over which to perform the DVV analysis.

- For every set \(\Omega_k\), the variance of the corresponding targets, \(\sigma_k^2\), is computed. The average over all sets \(\Omega_k\), normalized by the variance of the time series, \(\sigma_x^2\), yields the target variance, \(\sigma^*^2\):

\[
\sigma^*^2(\varrho_d) = \frac{1}{N} \sum_{k=1}^{N} \frac{\sigma_k^2(\varrho_d)}{\sigma_x^2}
\]  

(2.8)

where \(N\) denotes the total number of sets \(\Omega_k(\varrho_d)\)

Graphical representation of DVV-analysis is obtained by plotting \(\sigma^*^2(\varrho_d)\) as function of the standardized distance, \(\varrho_d\). The minimum target variance, \(\sigma_{\min}^*^2 = \min_{\varrho_d} [\sigma^*^2(\varrho_d)]\),

\(^2\)This means that the underlying process that generate the data can theoretically be described precisely by a set of linear or nonlinear equations. Thus, the component of a time series that can be predicted from a number of previous samples. [149]
which corresponds to the lowest point of the curve, is a measure for the amount of noise which is present in the time series. Thus, $\sigma^2_{\text{min}}$ is inversely related to prevalence of the deterministic component over the stochastic one, lowest $\sigma^2_{\text{min}}$ indicating a strong deterministic component. At the extreme right, the DVV plots smoothly converge to unity, as illustrated in Figure 4.1(a) and Figure 4.1(b). The reason behind this is that for maximum spans, all DVs belong to the same set, and the variance of the targets is equal to the variance of the time series.

The analysis addressing the linear or nonlinear nature of the original time series is examined by performing DVV analysis on both the original and a set of WiAATF surrogate time series. Due to the standardization of the distance axis, these plots can be conveniently combined within a scatter diagram, where the horizontal axis corresponds to the DVV plot of the original time series, and the vertical to that of the surrogate time series. If the surrogate time series yield DVV plots similar to that of the original time series, as illustrated by the first row of Figure 4.1(a), the DVV scatter diagram coincides with the bisector line, and the original time series is judged to be linear, as shown in second row of Figure 4.1(a). If not, as illustrated by first row of Figure 4.1(b), the DVV scatter diagram will deviate from the bisector line and the original time series is judged to be nonlinear, as depicted in the second row of Figure 4.1(b). Statistical testing of the null of linearity using a non-parametric rank-order test ([135]) is performed to enhance robust conclusion of results obtained via the DVV-analysis. We refer the reader to A.2 for more on DVV analysis of some simulated process.
We provide below a summary of our methodology which can be characterized in two stages:

1. Stage One: Detection of Nonlinearity in the underlying time series.

   (a) We study the structure of the economic indicator via a phase-space representation using the differential entropy-based method with both iAAFT and wiAAFT surrogates. Embedding parameters that yields lower entropy ratio is selected for the DVV analysis in next step. The main advantage of this differential entropy-based method is that a single measure is simultaneously used to obtain the embedding dimension, \( m \), and time lag, \( \tau \).

   (b) In order to detect the nonlinear behavior in the underlying time series, we use the DVV method discussed in Section 4.2.3. We are able to generate delay vectors necessary for the DVV analyzes using the \((m, \tau)\) obtained in the step (5.1). Unlike classical nonlinearity testing procedures, this non-parametric method is essentially data-driven and carry no \textit{a priori} assumptions about the intrinsic properties or mathematical structure of the underlying time series. In particular, this method provides a straightforward visualization and interpretation of results. With this approach, we are able to obtain important information on the underlying economic indicator, which is essential in choosing the appropriate class of models suggested by the data itself. It is noteworthy that this procedure does not need the underlying time series to be stationary. Statistical testing of the null of linearity using a non-parametric rank-order test ([135]) is performed to enhance robust conclusion of results obtained via the DVV-analysis.

2. Stage Two: Detection and explaining the business cycle.

   The next stage of the methodology deals with the problem of discovering pattern or hidden information that cannot be captured with traditional methods of business cycle analysis such as spectral analysis, which in only localized in frequency. In this work, we perform wavelet analysis using a complex-valued wavelet via a continuous wavelet transform in order to separate the phase and amplitude information. The phase information will be useful in explaining the economy-wide fluctuations in production that occur around a long-term growth trend. Information on the magnitude of such cycles across time will be obtained from the amplitude information.
2.3 Data Analysis

It is now well-known that the United States and all other modern industrial economies experience significant swings in economic activity. In this section, we perform analysis to characterize and detect nonlinear schemes for the US business cycle considering the monthly industrial production. The Industrial Production index is a business cycle indicator that has been widely used in business cycle analysis (see [7, 11, 14, 16, 23]).

Firstly, we characterize the nature of the time series using the DVV method with both iAAFT and WiAAFT surrogates and then employ complex Morlet wavelet to discover the cycles or hidden information in the data. In particular, we show that this new methodology permits to study the dynamics of the underlying economic indicator without a priori assumptions on the statistical properties and also allows for the detection of recessions periods. In addition, we attempt to establish a comparison between the late-2000s financial crisis and the Great Depression of the 1930s.

![Diagram 2.3](image)

**Figure 2.3:** US Industrial Production Index (IPI) time series. Diagram 3.1 is the plot of the monthly IPI series for the period: 1919:01 - 2012:07 ($n = 1123$), where the shaded regions corresponds to the US recessions from 1920 published by NBER. Diagram 2.3(b) is the plot of the Spectrum of IPI.

The monthly US Industrial Production Index (IPI) time series\(^3\) spanning over the period January, 1919 to July, 2012 is considered for the data analysis. Figure 3.1 is the plot of the monthly IPI series for the period: 1919:01 - 2012:07, implying 1123 observations, where the shaded regions corresponds to NBER\(^4\) published dates for US recessions from 1920. Figure 2.3(b) is the plot of the IPI spectrum which can be interpreted as a presence of long memory dynamics in the data.

We now give a comprehensive analysis of the IPI in level. To begin with, we opted for the differential-entropy based method (50) to determine the optimal embedding parameters,

\(^3\)The data can be downloaded from Federal Reserve Bank of St. Louis: [http://research.stlouisfed.org/fred2/](http://research.stlouisfed.org/fred2/)

i.e., the embedding dimension, \( m \), and the time lag, \( \tau \), for the DVV method with both the iAAFT surrogates and WiAAFT surrogates. We consider two approaches for estimating \((m, \tau)\). In the first case, the optimal embedding parameters are estimated using wiAAFT surrogates and are \( m = 3 \) and \( \tau = 1 \) with an entropy ratio, \( R_{ent}(m, \tau) = 0.7923 \), indicated as an open circle in the diagram with a clear structure in Figure 2.4(b). This result indicates the presence of time correlations, in the time series, implying a higher degree of structure, thus, a lower amount of disorder. The second case is by using the iAAFT surrogates of which the estimated values of the optimal embedding parameters are \( m = 4 \) and \( \tau = 7 \) with entropy ratio, \( R_{ent}(m, \tau) = 0.7271 \), which is less than that obtained via wiAAFT surrogates. In selecting the embedding parameters to generate delayed vectors needed to perform the DVV analysis, we choose the estimates with lower entropy ratio implying a higher degree of structure. In this case, \( m = 4 \) and \( \tau = 7 \) is used to generate the delayed vectors needed to perform the DVV analysis.

Based on the optimal embedding parameters \( m = 4 \) and \( \tau = 7 \), we generate delay vectors necessary for the DVV Analysis. The results from the DVV analysis , in figure 2.5, with iAAFT surrogates performed on the IPI indicates a clear deviation from the bisector on the DVV scatter diagram. The DVV plot also shows that the process is neither strictly deterministic or strictly stochastic. Thus, the original time series, IPI, exhibits nonlinear dynamics since the iAAFT surrogates are linear realizations of the original ([119, 120]). Statistical testing of the null of linearity using the non-parametric rank-order test, Table 2.1, indicates that the IPI is nonlinear. Thus, the DVV analysis suggests that the time series under consideration, IPI, behaves more of a nonlinearity with neither a strictly deterministic or strictly component. The nonlinearity in the data
could be due to both structural and behavioral changes that can occur in the economy across time. In other words, the nonlinearity may be as a result of existence of different states of the world or regimes in the economy. Many business cycle indicators present asymmetric features that have long been recognized in economics ([81, 103]). Putting it simply, there are sharp retractions during downturns in the economy as opposed to gradual upswings during recoveries ([17, 25, 83, 121]). Asymmetry has been recognised as a nonlinear phenomenon in several recent studies investigating various economic time series. Nonlinear models are therefore required to capture the features of the data generating mechanism of inherently asymmetric realizations of some of the macroeconomic business cycle series, since linear models are incapable of generating such behaviour ([38, 54, 131, 132]). Thus, some possible class of nonlinear models such as Markov switching models, smooth transition autoregressive (STAR) models, threshold autoregressive models, could capture such nonlinear behavior ([3, 22, 47, 54, 82]).

Table 2.1: Results of the non-parametric rank-order test. The null of linearity is rejected as soon as the Rank-Order is greater than Rank-Threshold. The code $H$ takes the value 0 or 1, where $H = 0$ corresponds to failure of rejecting the null of linearity and $H = 1$ the rejection of linearity for nonlinearity. The number of iAAFT surrogates considered for the DVV-analysis is 25, which is greater than the minimum requirement of 19 surrogates for testing at $\alpha = 0.05$ level of significance.

Figure 2.5: This is the DVV analysis with iAAFT surrogates performed on the IPI using the embedding parameters obtained via the differential entropy-based method. We clearly observe a deviation from the bisector on the DVV scatter diagram. The DVV plot also indicates that the process is neither strictly deterministic or strictly stochastic. Thus, the original time series, IPI, exhibits nonlinear dynamics since the surrogates are linear realizations of the original ([119, 120]).
On understanding the dynamics of the growth rate (suppose we denote the IPI as $X_t$, then the growth rate of IPI defined as $Y_t = \log(X_t) - \log(X_{t-1})$ of IPI to study the business cycle, we obtain the same estimates of embedding parameters $m = 2$ and $\tau = 1$ using the differential entropy-based method with both iAAFT surrogates and WiAAFT surrogates. Using the values of the embedding parameters, we are able to generate the phase space representation as displayed in figure 2.7 and perform the DVV analysis in figure 2.8. The purpose of studying the growth rate of IPI is not to ensure stationarity but to enable a better comparison of IPI dynamics over time. Business cycles are usually measured by considering the growth rate of industrial production index or the growth rate of real gross domestic product.

![Figure 2.6](image1.png)

(a) Differential-Entropy based method with iAAFT surrogates on the growth rate of the IPI

(b) Differential-Entropy based method with wiAAFT surrogates on the growth rate of the IPI

**Figure 2.6:** The optimal embedding parameters obtained via the Differential-Entropy based method using the two types of surrogates are indicated as an open circle in the diagrams with a clear structure. The values of the embedding parameters are $m = 2$ and $\tau = 1$ in both diagrams. This result $\tau = 1$ indicates the presence of time correlations, in the growth rate of IPI, implying a higher degree of structure, thus, a lower amount of disorder.

![Figure 2.7](image2.png)

**Figure 2.7:** Phase Space reconstruction using the embedding parameters $m = 2$ and $\tau = 1$. This represents the embedding of the underlying time series, growth rate of IPI, in phase space. The attractor is clearly visualized.
Chapter 2. Nonlinear Dynamics and Wavelets for Business Cycle Analysis

23

(a) DVV with iAAFT surrogates.  
(b) DVV with wiAAFT surrogates.

Figure 2.8: The DVV analysis on Growth rate of IPI indicates that it is characterize by nonlinear dynamics.

The DVV analysis, in figure 2.8 and the statistical testing of the null of linearity using the non-parametric rank-order test, Table 2.2, suggests that the time series under consideration behaves more of a nonlinear stochastic process than a deterministic one.

Table 2.2: Results of the non-parametric rank-order test. The null of linearity is rejected as soon as the Rank-Order is greater than Rank-Threshold. The code $H$ takes the value 0 or 1, where $H = 0$ corresponds to failure of rejecting the null of linearity and $H = 1$ the rejection of linearity for nonlinearity. The number of surrogates considered for the DVV-analysis is 25, which is greater than the minimum requirement of 19 surrogates for testing at $\alpha = 0.05$ level of significance.

In the following step, we perform CWT on the IPI growth rate using wavelets of the form in equation (2.3) at different bandwidths $f_b$ and center frequency $f_c$. In detecting the recession dates, the wavelet analysis was first performed for the period 1919:02-1940:01 using the US recession dates published by NBER as benchmark. The Morlet wavelet that captures the recession dates in this sample is then chosen as the wavelet to be used for the whole sample period. The Morlet wavelet that best detect cycles and hidden information in the data for the period 1919:02-1940:01, is given in equation (2.4) . The colormap used in the coefficient plots and scalogram plot ranges from blue to red, and passes through the colors cyan, yellow, and orange. The blue root-like structures on the phase-angle plot of the Coefficient plots, in figure 2.9, corresponds to recession periods of the economy. These are the periods where the economy-wide fluctuations in production are below the long-term growth trend.

The detection of the recession dates are represented by blue root-like structures on the angle coefficient plot in figure 2.9. We consider only such structures with a minimum
Figure 2.9: Coefficients plots obtained from the CWT using complex Morlet wavelet on the growth rate of IPI. First row represents the phase (angle) plot and second row is the corresponding Modulus plot. The colormap ranges from blue to red, and passes through the colors cyan, yellow, and orange. The blue regions on the Angle Coefficient plot corresponds to periods of relative stagnation in the economy from 1920. Thus, we consider only such structures with a minimum of six months as recession in the economy. The corresponding amplitudes can be read from the Modulus plot.

of six months\(^5\) as recession in the economy. The corresponding magnitude of these cycles can be read from the modulus plot of the coefficient plot in figure 2.9. The Wall Street Crash of 1929, followed by the Great Depression of the 1930s - the largest and most important economic depression in the 20th century - are well captured on the phase-angle coefficient plot in the figure 2.9 for time period (128 - 235), reported on Table 3.1. The three recessions between 1973 and 1982: the oil crisis - oil prices

\(^5\)This is a known censuring period accepted in Business Cycle literature and National Bureau of Economic Research: http://www.nber.org/cycles.html
Figure 2.10: The IPI growth rate and associated Scalogram from the CWT. The bar with colour scales on the left-hand side of the scalogram plot indicates the percentage of energy for each wavelet coefficient. Higher energy levels can be clearly observed for the Great Depression of the 1930s compared to the period of late-2000s financial crisis, also known as the Global Financial Crisis.

Figure 2.11: The pseudo-frequency associated to scale, in Hertz (Hz). The horizontal axis represents the scales and the vertical axis corresponds to the frequency associated to a scale.

The stock prices soared, causing the stock market crash are shown on the blue root-like structures of the phase-angle coefficient plot in the figure 2.9 for the time periods (659-676), (733-739), (751-767). Furthermore, the bursting of dot-com bubble - speculations concerning Internet companies crashed is also detected for the time periods (987-995). The wavelet energy at time periods are displayed on the scalogram in Figure 2.10 and the pseudo-frequency corresponding to scales are displayed in Figure 2.11. This interesting finding provide support for the use of wavelet methodology in business cycle modeling.
In order to compare the late-2000s financial crisis with the Great Depression of the 1930s, we perform the wavelet analysis on the growth rate of the IPI. The IPI growth rate dynamics are well captured by the phase-angle coefficient plot in figure 2.9, where the blue root-like structures correspond to periods of relative stagnation in the economy from 1920. The amplitudes associated with these economic fluctuations can be read from the modulus plot in figure 2.9. Looking at the modulus plot in figure 2.9 and scalograms in figure 2.10, we clearly observe higher amplitude and energy levels on the interval (0.008% - 0.016%) corresponding to the Great depression of the 1930’s compared to the late-2000s financial crisis with energy levels below 0.004%. These results, based on the data set we used, suggests that the intensity of the late-2000s financial crisis, also known as the Global Financial Crisis (GFC) is at the moment not so high as compared to the Great Depression of the 1930s.

### 2.4 Conclusion

In this paper, we have proposed a methodology in business cycle modeling which encompasses different existing methods successfully applied in physics and engineering. Our proposed procedure allows to first study the dynamics of the underlying economic indicator using non-parametric methods which are essentially data-driven and carry no a priori assumptions on the statistical properties, such as possible non-stationarity, or mathematical structure of the time series. We have provided a comprehensive analysis of the feasibility of our approach as essential in selecting the appropriate class of models suggested by the data itself. Finally, we have demonstrated the usefulness of wavelets in discovering patterns or hidden information that vary in nature across time which cannot
Table 2.3: Business Cycle Peaks and Troughs in the United States, 1920-2009. The peak and trough dates, in the format YYYY:MM, represent the start and end of “episodes” of some sort. ( http://www.nber.org/cycles.html ).

<table>
<thead>
<tr>
<th>Peak</th>
<th>Trough</th>
<th>Time Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920:01</td>
<td>1921:07</td>
<td>13-31</td>
</tr>
<tr>
<td>1923:05</td>
<td>1924:07</td>
<td>53-67</td>
</tr>
<tr>
<td>1926:10</td>
<td>1927:11</td>
<td>94-107</td>
</tr>
<tr>
<td>1929:08</td>
<td>1933:03</td>
<td>128-171</td>
</tr>
<tr>
<td>1937:05</td>
<td>1938:06</td>
<td>221-235</td>
</tr>
<tr>
<td>1945:02</td>
<td>1945:10</td>
<td>314-322</td>
</tr>
<tr>
<td>1948:11</td>
<td>1949:10</td>
<td>359-370</td>
</tr>
<tr>
<td>1953:07</td>
<td>1954:05</td>
<td>415-425</td>
</tr>
<tr>
<td>1957:08</td>
<td>1958:04</td>
<td>464-472</td>
</tr>
<tr>
<td>1960:04</td>
<td>1961:02</td>
<td>496-506</td>
</tr>
<tr>
<td>1969:12</td>
<td>1970:11</td>
<td>613-623</td>
</tr>
<tr>
<td>1973:11</td>
<td>1975:04</td>
<td>659-676</td>
</tr>
<tr>
<td>1980:01</td>
<td>1980:07</td>
<td>733-739</td>
</tr>
<tr>
<td>1981:07</td>
<td>1982:11</td>
<td>751-767</td>
</tr>
<tr>
<td>1990:07</td>
<td>1991:03</td>
<td>859-867</td>
</tr>
<tr>
<td>2001:03</td>
<td>2001:11</td>
<td>987-995</td>
</tr>
<tr>
<td>2007:12</td>
<td>2009:06</td>
<td>1068-1086</td>
</tr>
</tbody>
</table>

be captured using traditional methods of business cycle analysis such as the correlation analysis and spectral analysis.

Acknowledgements

We thank the conference participants of ISCEF 2012, CFE–ERCIM 2012, COMPSTAT 2012 and the participants at the Econometrics Internal Seminar at Center for Operations Research and Econometrics (CORE) for their participation and interest. We also would like to thank Sébastien Van Bellegem, Luc Bauwens, Christian Hafner, Timo Terasvirta and Yukai Kevin Yang for their remarks and questions.
Chapter 3

Turning point chronology for the Euro-Zone: A Distance Plot Approach

3.1 Introduction

Since the early work by [29], many attempts have been made to measure and forecast business cycles. The main aim of business cycle analysis is to detect and anticipate economic fluctuations with a particular attention paid to turning points location ([11]). There are a great number of articles on national chronology dating (e.g. Belgium, UK, Canada, Austria and Australia); however, very few countries have official dating procedures (among them the USA (National Bureau of Economic Research, NBER) and Japan (Economic and Social Research Institute)). Since the establishment of the Euro area, several papers have been devoted to its cyclical dating. However, there does not yet exists an official dating for the Euro area. Dating is an ex post exercise, and in this respect accuracy is an important criterion since dating is useful for economic decision-making. Governments and central banks are usually very sensitive to indicators showing signs of deterioration in growth to allow them to adjust their policies sufficiently in advance, avoiding more deterioration or a recession. In this respect, timing is important and the earlier the signal, the better ([10]). It is worth noticing that the announcement of the cycle turning point dates in the United States (US) are often substantially delayed\(^1\). As such, it will be interesting to propose methods that will enhance faster and accurate detection of the dates.

\(^{1}\)For more information on announcement of these dates, see http://www.nber.org/cycles.html
In literature, two types of cycles are usually considered: the classical business cycle and the growth cycle. The growth cycle refers to the deviations from the long-term trend of the series while the classical business cycle refers to fluctuations in the level of the underlying economic indicator. The growth cycle has not been studied much compared to the classical business cycle due to the de-trending problem of this concept ([10, 15]). In this paper, we study the classical business cycle using monthly data to enhance accuracy of the dating. The most commonly dated monthly time series is the industrial production ([10, 14–16]). Our aim is to propose a method with these features:

1. Robustness to extreme values, non stationarity and to any length of data.
2. It must be replicable by anyone. This enhances the transparency of the method.
3. Adaptability of the method to different time series.
4. The chronology would not be revised through time.

Recurrence Plot (RP) is an advanced graphical technique of nonlinear data analysis which reveals all the times when the phase space trajectory of a dynamical system visits roughly the same area in the phase space. The phase space trajectory of a system refers to the path traced in a high-dimensional space representing the time evolution and the dynamics of the system. Although, most applications of this technique have been in physics and biology, it has gained interest in a variety of scientific fields ([94, 96]). Recurrence plots and recurrence quantification analysis ([97, 152]) is gaining some attention also in economics ([6, 20, 72, 87, 125], among others). This technique is applicable to any time series since it requires no a priori assumptions on the statistical properties, such as stationarity, or mathematical structure of the time series. We find this technique as a promising means of analyzing economic data since it is robust against non-stationarity ([96]). In this work, we mainly use distance plots (sometimes called unthresholded recurrence plots) in studying the turning points of business cycles. Whilst the central focus of this paper is the analysis of univariate time series (Industrial Production Index), we also provide the reader with possible extensions beyond the scope of the paper in order to clarify the methodology.

Previous works documenting the Euro-zone business cycle are quite limited by: stationarity of data, choice of filters and usually model-based approaches (see [10, 12, 14, 16, 21, 45, 59, 66, 67, 71, 100, 104]). For instance, the work by [104] first constructs a Euro area monthly real GDP series by interpolation and then applies a modified version of a nonparametric algorithm, the Bry-Boschan procedure ([27]), to determine peaks and

\[\text{http://www.recurrence-plot.tk/} \text{ and [96]}\]
troughs in univariate time series. It is unclear whether information is lost or gained after such interpolation. It is worth noting that the proposed Bry-Boschan procedure by [104] as a turning point selection method is unsuitable when multivariate time series are considered for dating. To our best of knowledge, most of the existing dating methodologies are restrictive in terms of assumptions on model specification and usually not suitable for short and non-stationary data. As such, we propose an alternative methodology using the concept of recurrence plots, which is essentially a data-driven approach with no a priori assumptions on the statistical properties on the underlying economic indicator(s). This concept of recurrence plots can be applicable to both univariate and multivariate time series ([95, 116]).

Our aim is to construct a turning point chronology for the Euro-zone business cycle. Our analysis exploits the concept of recurrence plots on an underlying economic indicator to locate hidden patterns, non-stationarity, and to examine the nature of these plots in events of economic crisis. The qualitative interpretation of our results relates the existence of disruptions (butterfly-like or band structures) to the non-stationarity; rare states (extreme events) or far from the normal; transitions ([6, 43, 94, 96]).

The paper is organized as follows: We provide in Section 3.2, an overview on the concept of recurrence plots and distance plots. In Section 3.3, we present a comprehensive analysis of the feasibility of this approach to analyze both the US and Euro-zone economic cycle. Section 3.4 concludes.

3.2 Methodology based on recurrence plots

The method of recurrence plots (RP) was introduced to visualize the dynamics of phase space\(^3\) trajectories ([43, 94]). It is a graphical technique that depicts the different occasions when a dynamical system visits roughly the same area in the phase space. From Takens’ embedding theorem ([127]), the dynamics can be appropriately presented by a reconstruction of the phase space trajectory\(^4\) \(\vec{x}(t) = \vec{x}_i \in \mathbb{R}^m \ (i = 1, \cdots, \eta, \ t = i\Delta t, \) where \(\Delta t\) is the sampling rate) in the \(m\)-dimensional phase space. For a given one-dimensional time series \(\{u_i\}_{i=1}^{N}\), the phase space vectors \(\vec{x}\) can be reconstructed by embedding the series using Takens’ time delay method \(\vec{x}_i = (u_i, u_{i+\tau}, \cdots, u_{i+(m-1)\tau})\).

\(^3\)Phase spaces are usually high-dimensional which makes it difficult to visualize. [43] introduction of the recurrence plot enables us to investigate the \(m\)-dimensional phase space trajectory through a two-dimensional representation of its recurrences. This is interesting as higher-dimensional phase spaces can only be visualized by projection into the two or three-dimensional sub-spaces.

\(^4\)The state of a system typically changes in time, and, hence, the vector in the phase space describes a trajectory representing the time evolution, the dynamics, of the system. This trajectory is referred to as a phase space trajectory ([76, 127]).
series. The parameters $m$ and $\tau$ are referred to as the embedding dimension and time delay respectively ([5, 96, 127]). The choice of $m$ and $\tau$ are usually based on methods for approximating these parameters, such as the false nearest neighbors method (for $m$) and mutual information (for $\tau$), which ensures the entire covering of all free parameters and avoiding autocorrelated effects ([76]). [50] proposed a differential entropy-based method in which a single measure is simultaneously used to obtain the embedding dimension, $m$, and time lag, $\tau$. We refer the interested reader to [50] and [4, 6] for an overview of this method. We refer to the case for which $m = 1$ and $\tau = 1$ as an unembedded time series. The recurrence plot is the calculation of an $\eta \times \eta$ matrix

$$R^x_{i,j} = \begin{cases} 1 & : \|\vec{x}_i - \vec{x}_j\| < \varepsilon \\ 0 & : \text{otherwise} \end{cases}$$

where $\|\cdot\|$ is a norm (e.g. Euclidean or maximum norm) and $\varepsilon$ is the cut-off distance which defines a region centered at $\vec{x}_i$. If $\vec{x}_j$ falls within this region, the state will be near to $\vec{x}_i$ and is taken to be a recurrence of the state $\vec{x}_i$, which implies $R^x_{i,j} = 1$. The recurrence plot is square matrix plot of the binary values $R^x_{i,j}$, in which the matrix element correspond to those calendar times at which a state of a dynamical system recurs (columns and rows correspond then to a certain pair of calendar times) ([6, 96] and the references therein). In general the cut-off distance, $\varepsilon$, has to be chosen as small as possible, but a too small $\varepsilon$ can lead one to miss some structure, if there is noise distortion ([94]). [153] suggest to set a threshold level equal to the lower 10% of the maximum distance between the embedded vectors. In literature, further variations of the recurrence plots have been proposed for different purposes. In this paper, we make use of a special type of recurrence plot referred to as unthresholded recurrence plots ([74]). This recurrence plot is obtained by plotting a matrix of distances

$$D^x_{i,j} = \|\vec{x}_i - \vec{x}_j\|$$

between the vectors $\vec{x}_i$ and $\vec{x}_j$. The matrix data obtained from the application of equation (3.2) is graphically represented as colors making it easier to read. In otherwords, the unthresholded recurrence plot is a heatmap (colormap) of the distance matrix obtained from equation (3.2). As such, it is appropriate to term the unthresholded recurrence plot as distance plot. In the Section that follows, we will often use the term distance plot and unthresholded recurrence plot interchangeably. Multivariate extensions of recurrence plots have been developed as cross recurrence plots and joint recurrence plots ([95, 116]). In comparing the dynamics of any two time series simultaneously embedded in the same phase space, we employed the cross recurrence plot ([6, 95, 151]). This is

\footnote{For more details on these variations, we refer the reader to http://www.recurrence-plot.tk/variations.php.}
useful in studying the simultaneous occurrence of a state in both series. The cross recurrence plot entails testing for closeness of each point of the first trajectory $\vec{x}_i$ ($i = 1, \cdots, \eta$) with each point of the second trajectory $\vec{y}_j$ ($j = 1, \cdots, \vartheta$) resulting in $\eta \times \vartheta$ array

$$CR_{i,j}^{xy} = \begin{cases} 1 & : \|\vec{x}_i - \vec{y}_j\| < \varepsilon. \\ 0 & : otherwise \end{cases} \quad (3.3)$$

Thus, the purpose of cross recurrence plot (CRP) is to compare the states of two systems, which means that the states represent the same physical system, i.e. the data should have the same unit and the phase space reconstruction must be the same. We employ the CRP in comparing US IPI and Euro area IPI since both time series have the same unit$^6$ and phase space reconstruction ($[6, 95–97]$).

In contrast to cross recurrence plots, joint recurrence plot (JRP) shows all those times at which a recurrence in one dynamical system occurs simultaneously with a recurrence in a second dynamical system. In other words, the JRP is the element-wise product of the recurrence plot of the first system and the recurrence plot of the second system, i.e. for two systems $\vec{x}$ and $\vec{y}$, the joint recurrence plot is

$$JR_{i,j}^{xy} = R_{i,j}^x \cdot R_{i,j}^y, \quad \vec{x}_i \in \mathbb{R}^m, \quad \vec{y}_j \in \mathbb{R}^n, \quad i, j = 1, 2, \cdots, \eta \vartheta. \quad (3.4)$$

JRPs can be computed from more than two systems. The dimension of the considered systems in phase spaces can be different but the data length has to be the same. Joint recurrence plots can be used in order to detect phase synchronization ($[95, 96, 116]$). For example, in the U.S., the NBER uses several indicators (such as Gross Domestic Product, personal income, household employment, Industrial production Index, Gross domestic Income, payroll employment, manufacturing and retail sales) to date recessions. Considering this multivariate time series, in which the seven univariate time series are usually not of the same unit, the joint recurrence plot is more suitable to compare their recurrences, i.e. whether the recurrences of states in processes separately occur at the same times. We remark that the concept of JRP is beyond the scope of the paper and it is only provided in order to clarify the methodology.

Recurrence plots contains typical small-scale$^7$ structures and in particular large-scale structure, also called texture, caused by characteristic behavior of phase space trajectory of systems. These large-scale structures can be visually characterized by homogeneous, periodic, drift or disrupted ($[43, 94, 96, 152]$). The visual appearance of these plots gives hints about the dynamics of the system. The qualitative interpretation of the existence of

$^6$It is same unit since both univariate time series are indexes.

$^7$The small-scale structures as in single dots, diagonal lines and vertical or horizontal lines. These can also be extended clusters resulting from a mixture of the latter.
disruptions (*butterfly-like* or band structures) are as a result of non-stationarity; some states are rare or far from the normal; transitions may have occurred ([43, 94, 96]). We refer to the existence of such *butterfly-like* structures along the main diagonal of the recurrence plot as an indication of economic crisis, where the turning points are identified as the start and end of the formation of these structures. It is worth noting that we mainly use the *distance plots* in identifying the turning points in each univariate time series (principally IPI of U.S. and Euro-zone) and then the cross recurrence plots in studying the simultaneously recurrence of states in both series. The *distance plot* is replicable by definition since it is obtained by plotting a matrix of distances $D_{i,j}$, in equation (3.2), which do not change through time for a given time series. This ensures an effective dating chronology since it avoids revision through time for an underlying economic indicator. The main advantage of recurrence plots is that they provide useful information even for short and non-stationary data, where other methods fail ([96]). The Section that follows will be devoted to dating the Euro-zone business cycle.

### 3.3 Data Analysis

In this Section, we perform analysis to characterize and detect turning points for the Euro-zone economic cycle considering the monthly Industrial Production Index (IPI) series. Firstly, we exploit the concept of recurrence plots\(^8\) on the US IPI series to characterize and detect recessions periods. The essence of starting the analysis with the US data is to serve as a benchmark for our analysis since there already exists reference chronology for the US business cycle, provided by the Dating Committee of the NBER\(^9\). We then use this concept in constructing a monthly turning point chronology for the Euro-zone business cycle. In particular, we show that this approach permits to detect turning points and study the business cycle without *a priori* assumptions on the statistical properties on the underlying economic indicator.

The monthly US Industrial Production Index (IPI) time series\(^10\) spanning over the period January, 1919 to July, 2012 is considered for the data analysis. Figure 3.1 is the plot of the monthly IPI series for the period: 1919:01 - 2012:07, implying 1123 observations, where the shaded regions corresponds to NBER\(^11\) published dates for US recessions from 1920. Figure 3.2 is the plot of monthly Euro-zone\(^12\) IPI series for the period: 1971:01 - 2011:12 ($n = 492$).

---

\(^8\)The *distance plots* are generated with the Cross Recurrence Plot Toolbox in Matlab provided by Norbert Marwan upon request: [http://tocsy.pik-potsdam.de/CRPtoolbox/](http://tocsy.pik-potsdam.de/CRPtoolbox/)

\(^9\)National Bureau of Economic Research

\(^10\)The data can be downloaded from Federal Reserve Bank of St. Louis [http://www.nber.org/cycles.html](http://www.nber.org/cycles.html)

\(^11\)http://www.nber.org/cycles.html

\(^12\)Source: [22] and Eurostat.
We provide in Figure 3.3 the unthresholded recurrence plot which is sometimes termed distance plot ([74]) on the US IPI. It is a matrix plot with both vertical and horizontal axes corresponds to calendar dates. The colormap corresponds to the distance to the next recurrence of a state in the time series. In this paper, we mainly consider the case when the embedding parameters are $m = 1$ and $\tau = 1$ since the results obtained for unembedded and embedded version of the time series do not differ in terms of identification of the turning points associated with the business cycle ([6]). The recurrence plot provides a graphical understanding of the dynamics of phase space trajectories, of the time series, as distances between states are revealed. The recurrence plots allows to study the recurrence of a state at a particular calendar date. In other words, fixing a period or date on the horizontal axis, we are able to observe the recurrence of such
Turning point chronology for the Euro-Zone: A Distance Plot Approach

Figure 3.3: This is an unthresholded recurrence plot which is sometimes termed distance plot ([74]). It is a matrix plot with calendar time on both the vertical and horizontal axes. The colormap corresponds to the distance to the next recurrence of a state in the time series. This representation shows the distances between states and enhance understanding of the phase space trajectory of US IPI. The existence of butterfly-like structures, of a minimum size of six months, along the main diagonal (bisector) indicates economic crisis. In this case, the start and end of a economic crisis corresponds the start and end on the formation of butterfly-like structure along the main diagonal.

Events along the calendar dates on the vertical axis. Intuitively, the plot depicts the different occasions when the economic system, described by IPI, visits roughly the same area in the phase space. The existence of butterfly-like structures along the main diagonal (bisector) of the plot indicates economic crisis as discussed in the previous Section. In this case, the start and end of a declining economic activity corresponds the start and end of the formation of butterfly-like structure along the main diagonal. In this paper, we designate any butterfly-like structure with a minimum size\textsuperscript{13} of six months as economic crisis. Our results on the dating chronology obtained from the distance plot, Figure 3.3, for US IPI are presented in Table 3.1.

Sensitivity analysis on the distance plot with respect to the embedding dimension $m$ and the time delay $\tau$ reveals that for fixed $\tau = 1$ and for any chosen $m > 1$, the turning points identified are same as setting $m = 1$ and $\tau = 1$. In particular, we observe much sharper picture of the butterfly-like structures as the embedding dimension is set close to 1. We noticed that for any given $m$, as the time delay parameter increases ($\tau \geq 1$), the butterfly-like structures grow dimmer making it not easy to clearly identify the peaks and troughs of the business cycle. As such, we recommend that the time delay parameter be set to 1 for better identification of turning points.

The dates from the distance plot, Figure 3.3, for US IPI appears not to differ much from the official business cycle dates published by NBER. Interestingly, the main difference

\textsuperscript{13}The size refers to the length in time from the start and end on the formation of butterfly-like structure along the main diagonal.
occurs in the detection of the peak date of two recession periods: 1973 and 2001 (see Table 3.1). Our findings also indicates that the turning points provided by NBER for the US can be obtained with only IPI data using the distance plot approach. Thus, this results on the preliminary exploratory analysis for the US provides support for the use of our propose procedure to dating the Euro-Zone business cycle.

<table>
<thead>
<tr>
<th>NBER dates</th>
<th>butterfly dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak</td>
<td>Trough</td>
</tr>
<tr>
<td>1920:01</td>
<td>1921:07</td>
</tr>
<tr>
<td>1923:05</td>
<td>1924:07</td>
</tr>
<tr>
<td>1926:10</td>
<td>1927:11</td>
</tr>
<tr>
<td>1929:08</td>
<td>1933:03</td>
</tr>
<tr>
<td>1937:05</td>
<td>1938:06</td>
</tr>
<tr>
<td>1945:02</td>
<td>1945:10</td>
</tr>
<tr>
<td>1948:11</td>
<td>1949:10</td>
</tr>
<tr>
<td>1980:01</td>
<td>1980:07</td>
</tr>
</tbody>
</table>

Table 3.1: Business Cycle Peaks and Troughs in the United States, 1920-2009. The peak and trough dates, in the format YYYY:MM, represent the start and end of “episodes” of some sort. The column NBER are the reference chronology for US business cycles published at http://www.nber.org/cycles.html . The results of turning points obtained from the distance plot of US IPI is displayed in column labelled butterfly dates.

In the following, we focus on detecting and constructing the turning points for the Euro-zone business cycle. We proceed with the same analysis as done on the US data but now on the Euro-zone IPI series. We remark that embedding the original time series does not yield different results in terms of identification of the turning points associated with the business cycle. Our results on the dating chronology of the Euro-zone based on the IPI is presented in Table 3.2 and Figure 3.5 with the distance plot in Figure 3.4. A comparison of turning points for the Euro-Zone considering dates identified by the distance plot and in [10] for the monthly series of Euro-zone IPI is given in Table 3.3. The quarterly turning points provided by the Centre for Economic Policy Research (CEPR) using the Euro area real GDP and the turning points identified by the Bry-Boschan algorithm when applied to the Moench/Uhlig (MU) monthly series of Euro area real GDP is also provided in Table 3.3. Notice that the dates determined by the
distance plot procedure does not differ much from the turning points already provided by CEPR and other two dating chronologies. In particular, our proposed dates comes close to the monthly turning points proposed in [104] within a few months.

**Figure 3.4:** The distance plot for the Euro-zone IPI and it’s embedded version. The existence of butterfly-like structures, of a minimum size of six months, along the main diagonal (bisector) indicates economic crisis. Embedding the original time series does not yield different results in terms of identification of the turning points associated with the business cycle.

<table>
<thead>
<tr>
<th>Year 1</th>
<th>Year 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1974:06</td>
<td>1975:08</td>
</tr>
<tr>
<td>1980:01</td>
<td>1980:10</td>
</tr>
<tr>
<td>1982:05</td>
<td>1982:12</td>
</tr>
<tr>
<td>1992:04</td>
<td>1993:06</td>
</tr>
<tr>
<td>2000:12</td>
<td>2001:11</td>
</tr>
<tr>
<td>2007:12</td>
<td>2009:04</td>
</tr>
<tr>
<td>2011:07</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.2:** Industrial business cycle dating chronology for the Euro-zone from 1971:01–2011:12. The peak and trough dates, in the format YYYY:MM, represent the start and end of “episodes” of some sort. The end of the last recession period that started 2011:07(q3) was not determined considering the sample period.
Chapter 3. Turning point chronology for the Euro-Zone: A Distance Plot Approach

Figure 3.5: The shaded regions corresponds to the Butterfly dates for the Euro Area Business Cycles from 1971:01–2011:12. The end of the last recession period that started 2011:07(q3) was not determined considering the sample period.

In Figure 3.6, we provide the distance plots for the US IPI and Euro-zone IPI for the period 1991:01 - 2012:07. The simultaneous occurrence of economic crisis in both US and Euro-zone for this period can be visualize on the cross recurrence plot displayed in Figure 3.7. These two economic systems were both hit by the crisis resulting from the bust of dot-com bubble and even more on the 2007-2012 global financial crisis, also known as the Global Financial Crisis and 2008 financial crisis. Similar features in the plots for this period suggests strong economic interdependence between the US and Euro-zone economic system.

3.4 Conclusion

In this work, we have demonstrated the usefulness of recurrence plots in identifying, dating and explaining economic crisis. The study provides a preliminary exploratory analysis for the US to validate the procedure to the following application to the Euro-zone. Our findings also indicates that the turning points provided by NBER for the US can be obtained with only IPI data using the distance plots approach. The findings from the data analysis with recurrence plots, in this case distance plots, shows that these plots are robust to extreme values, non stationarity and applicable to both short and long data length. This approach is also replicable and transparent; is adaptive to any time series. In particular, we show that this approach provides a transparent
Figure 3.6: The distance plots of US IPI and Euro-zone IPI for the period 1991:01 - 2012:07. The top figure corresponds to the distance plots of US IPI. The second distance plot is for the Euro-zone IPI.

Figure 3.7: This is a cross recurrence plot to visualize the simultaneous occurrence of a similar state in both US IPI and Euro IPI.

chronology of business cycles since it avoids revision of crisis dates through time. The proposed methodology can easily be used in dating business cycles in country-specific time series for the European economies. Obviously, the nature of the distance plots will depend on the underlying economic indicator(s) of the given country as displayed in Figure 3.8. Finally, a direct extension of this paper will be to consider the use of joint
Chapter 3. Turning point chronology for the Euro-Zone: A Distance Plot Approach

Figure 3.8: The distance plots on monthly IPI series of some European countries: France, Germany and Italy. The data is obtained from [22] and Eurostat, and spans over the period: 1970:01 to 2012:11.

Acknowledgements

The authors are grateful to anonymous referees of JBCMA for their careful revision, valuable suggestions, and comments that have improved this paper. We thank the

\[\text{[14]}\text{The NBER provides a chronology of the classical business cycle for the US, based on the consensus of a set of coincident indicators which are widely accepted among economists and policy makers. see http://www.nber.org/}\]
### Table 3.3: Comparison of turning points for the Euro-Zone. The format of the dates “yyyy:mm(qq)” denotes “year:month(quarter)”. The butterfly dates corresponds to peaks and troughs identified by the distance plot when applied to the monthly series of Euro-zone IPI. The turning points obtained in [10] for the monthly Euro-zone IPI series is denoted as [10] dates. The CEPR dates are the quarterly turning points determined by the CEPR using the Euro area real GDP. MU dates are turning points identified by the Bry-Boschan algorithm when applied to the Moench/Uhlig (MU) monthly series of Euro area real GDP. The end of the last recession period that started 2011:07(q3) was not determined considering the sample period 1971:01–2011:12.

<table>
<thead>
<tr>
<th>butterfly dates</th>
<th>[10] dates</th>
<th>CEPR dates</th>
<th>MU dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak</td>
<td>Trough</td>
<td>Peak</td>
<td>Trough</td>
</tr>
<tr>
<td>1974:06(q2)</td>
<td>1975:08(q3)</td>
<td>1974:04(q2)</td>
<td>1975:04(q2)</td>
</tr>
<tr>
<td>1980:01(q1)</td>
<td>1980:10(q4)</td>
<td>1980:02(q1)</td>
<td>1981:01(q1)</td>
</tr>
<tr>
<td>1982:05(q2)</td>
<td>1982:12(q4)</td>
<td>1981:10(q4)</td>
<td>1982:12(q4)</td>
</tr>
<tr>
<td>1992:04(q2)</td>
<td>1993:06(q2)</td>
<td>1992:01(q1)</td>
<td>1993:05(q2)</td>
</tr>
<tr>
<td>2000:12(q4)</td>
<td>2001:11(q4)</td>
<td>2000:12(q4)</td>
<td>2001:12(q4)</td>
</tr>
<tr>
<td>2007:12(q4)</td>
<td>2009:04(q2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011:07(q3)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

conference participants of CFE-ERCIM 2012 and the participants at the 2013 EBIMEDEEM Doctoral Jamboree at CORE for their participation and interest.
4.1 Introduction

In real-world applications of financial time series analysis, the process underlying the generated signal, which is the time series, are a priori unknown. These signals usually contain both linear and nonlinear, as well as deterministic and stochastic components, yet it is a common practice to model such processes using suboptimal, but mathematically tractable models. In general, performing a nonlinearity analysis in a modeling the time series can lead to a significant improvement of the quality of the results, since it facilitates the selection of appropriate processing methods, suggested by the data itself. It is, therefore, essential to investigate the intrinsic dynamical properties of financial time series in terms of its deterministic/stochastic and nonlinear/linear components reveals important information that otherwise remains not clear in using conventional linear methods of time series analysis. For this purpose, we adopt to the recently proposed phase space based ’delay vector variance’ (DVV) method (51), for signal characterization, which is more suitable for signal processing application because it examines the nonlinear and deterministic signal behavior at the same time. This method has been used for understanding the dynamics of exchange rates (5), qualitative assessment of machine learning algorithms, analysis of functional magnetic resonance imaging (fMRI) data, as well as analysing nonlinear structures in brain electrical activity and heart rate variability (HRV) (52). Optimal embedding parameters will be obtained using a
differential entropy based method proposed in 50, which allows for simultaneous determination of both the embedding dimension and time lag. Surrogate generation used in this study will be based on both the Iterative Amplitude Adjusted Fourier Transform (iAAFT) ([119, 120]) and a recently refined iAAFT with a wavelet-based approach, denoted WiAAFT (79).

Recurrence Plot (RP) is an advanced graphical technique of nonlinear data analysis which reveals all the times when the phase space trajectory of the dynamical system visits roughly the same area in the phase space. Although, most applications of this technique has been in physics and biology, it has gained interest in a variety\(^1\) of scientific fields ([96]). Recurrence plots and recurrence quantification analysis ([97, 152]) is gaining some attention in financial studies ([20, 72, 87, 125], among others). This technique is applicable to any time series since it requires no a priori assumptions on the statistical properties, such stationarity, or mathematical structure of the time series. We find this technique as a promising means of analyzing financial data since it is robust against non-stationarity in the data. In this work, we will mainly use distance plots, which is sometimes called unthresholded recurrence plot in studying the dynamics of stock markets.

In this article, our aims is to characterize and detect nonlinear schemes for the U.S. financial cycle considering both S&P 500 Index and Nasdaq Composite. We refer to fluctuations or swings in financial asset prices over time as financial cycle, which is characterised by periods of financial disruptions and booms. Firstly, our analysis deals with the use of the recently proposed ‘delay vector variance’ (DVV) method, which examines local predictability of a signal in the phase space to detect the presence of determinism and nonlinearity in a time series. Optimal embedding parameters used in the DVV analysis are obtained via a differential entropy based method using wavelet-based surrogates. This enables the characterization of the financial time series without a priori assumption on the dynamics or statistical properties of the series. Secondly, we exploit the concept of recurrence plots on both series to locate hidden patterns, non-stationarity, and to examine the nature of these plots in events of financial crisis. The qualitative interpretation of the existence of disruptions (butterfly-like or band structures) are as a result of nonstationarity; some states are rare (extreme events) or far from the normal; transitions may have occurred ([43, 96]). In particular, we show the usefulness of recurrence plots in the diagnosis and detection of financial bubbles and crisis, which have significantly impacted economic upheavals in the past few decades ([9]).

\(^1\)A comprehensive introduction and bibliography about on recurrence plots is available on [http://www.recurrence-plot.tk/](http://www.recurrence-plot.tk/) and [96]
Chapter 4. Nonlinear dynamics and recurrence plots for detecting financial crisis

The paper is organised as follows: surrogate generation methodology and differential entropy based method for determining optimal embedding parameters of the phase-space representation of time series are presented in section 4.2.1 and section 4.2.2 respectively. In section 4.2.3, an overview of the ‘delay vector variance’ method with illustrations. We provide in section 4.2.4, the concept of recurrence plots. In section 4.3, we present a comprehensive analysis of the feasibility of this approach to analyse the U.S. financial cycle. Section 4.4 concludes.

4.2 ‘Delay Vector Variance’ Method and Recurrence Plots

In this section, we present an overview of different existing methods successfully applied in physics and engineering. In particular, we show the usefulness of these methods over other methods and then explain how we merged these methods to studying the dynamics of financial time series. Our methodology encompasses surrogate generation methods, differential entropy method for determining the optimal embedding parameters in phase-space, the DVV method and data analysis with recurrence plots.

4.2.1 Surrogate Data Method and Statistical testing

Surrogate time series, or ‘surrogate’ for short, is non-parametric randomised linear version of the original data which preserves the linear properties of the original data. For identification of nonlinear/linear behavior in a given time series, the null hypothesis that the original data conform to a linear Gaussian stochastic process is formulated. An established method for generating constrained surrogates conforming to the properties of a linear Gaussian process is the Iterative Amplitude Adjusted Fourier Transform (iAAFT), which has become quite popular (85, 119, 120, 129). This type of surrogate time series retains the signal distribution and amplitude spectrum of the original time series, and takes into account a possibly nonlinear and static observation function due to the measurement process. The method uses a fixed point iteration algorithm for achieving this, for the details of which we refer to 119, 120.

Wavelet-based surrogate generation is a fairly new method of constructing surrogate for hypothesis testing of nonlinearity which applies a wavelet decomposition of the time series. The main difference between Fourier transform and wavelet transform is that the former is only localized in frequency, whereas the latter is localized both in time and frequency. The idea of a wavelet representation is an orthogonal decomposition across a hierarchy of temporal and spatial scales by a set of wavelet and scaling functions.

The first three sections of this overview do not differ from those presented in [4].
The iAAFT-method has recently been refined using a wavelet-based approach, denoted by WiAAFT (79), that provides for constrained realizations\(^3\) of surrogate data that resembles the original data while preserving the local mean and variance as well as the power spectrum and distribution of the original except for randomizing the nonlinear properties of the signal. The WiAAFT-procedure follows the iAAFT-algorithm but uses the Maximal Overlap Discrete Wavelet Transform (MODWT) where the iAAFT-procedure is applied to each set of wavelet detail coefficients \(D_j(n)\) over the dyadic scales \(2^{j-1}\) for \(j = 1, \cdots, J\), i.e., each set of \(D_j(n)\) is considered as a time series of its own. The main difference between iAAFT and wiAAFT algorithms is that the former is designed to produce constrained, linear realisations of a process that can be compared with the original time series on some measure, while the later algorithm restricts the possible class of realizations to those that retain some aspect of the local mean and variance of the original time series (80).

Statistical analysis by the concept of surrogate data tests for a difference between a test statistic computed for the original and linearized versions of the data, i.e., an ensemble of realizations of the null hypothesis linear dynamics. For statistical testing of the null hypothesis of linearity, we follow [135] by using a non-parametric rank-order test. The degree of difference between the original and surrogate data is given by the ranked position of the data asymmetry with respect to the surrogates. For a right-tailed test, we generate at least \(N_s = \frac{1}{\alpha} - 1\) surrogates, where \(\alpha\) is the level of significance and \(N_s\) denotes the number of surrogates. The rank-threshold (or critical value) for right-tailed rank-order test is given by \((1 - \alpha)(N_s + 1)\). The null of linearity is rejected as soon as the rank-order statistic is greater than the rank-threshold. To achieve a minimal significance requirement of 95\% (\(\alpha = 0.05\)), we need at least 19 surrogates time series for right-tailed tests. Increasing the number of surrogates can increase the discrimination power (119, 120, 135). The concept of surrogate data will be incorporated into the Delay Vector Variance method (below) to examine the dynamics of an underlying economic indicator.

### 4.2.2 Optimal Embedding Parameters

In the context of signal processing, an established method for visualizing an attractor of an underlying nonlinear dynamical signal is by means of time delay embedding (69). By time-delay embedding, the original time series \(\{x_k\}\) is represented in the so-called\(^3\)These are surrogate realizations that are generated from the original data to conform to certain properties of the original data, e.g., their linear properties, i.e., mean, standard deviation, distribution, power spectrum and autocorrelation function (119, 120).
Chapter 4. Nonlinear dynamics and recurrence plots for detecting financial crisis

'phase space' by a set of delay vectors\(^4\) (DVVs) of a given embedding dimension, \(m\), and time lag, \(\tau\): \(x(k) = [x_{k-\tau}, \ldots, x_{k-m\tau}]\). 50 proposed a differential entropy based method for determining the optimal embedding parameters of a signal. The main advantage of this method is that a single measure is simultaneously used for optimizing both the embedding dimension and time lag. We provide below an overview of the procedure:

The "Entropy Ratio" is defined as

\[
R_{\text{ent}}(m, \tau) = I(m, \tau) + \frac{m \ln N}{N},
\]

where \(N\) is the number of delay vectors, which is kept constant for all values of \(m\) and \(\tau\) under consideration,

\[
I(m, \tau) = \frac{H(x, m, \tau)}{(H(x_{a,i}, m, \tau))_i},
\]

where \(x\) is the signal, \(x_{a,i}, i = 1, \ldots, T_s\) surrogates of the signal \(x\), \(\langle \cdot \rangle_i\) denotes the average over \(i\), \(H(x, m, \tau)\) denotes the differential entropies estimated for time delay embedded versions of a time series, \(x\), which an inverse measure of the structure in the phase space. 50 proposed to use the Kozachenko-Leonenko (K-L) estimate (89) of the differential entropy given by

\[
H(x) = \sum_{j=1}^{T} \ln(T\rho_j) + \ln 2 + C_E
\]

where \(T\) is the number of samples in the data set, \(\rho_j\) is the Euclidean distance of the \(j\)-th delay vector to its nearest neighbour, and \(C_E(\approx 0.5772)\) is the Euler constant. This ratio criterion requires a time series to display a clear structure in the phase space. Thus, for time series with no clear structure, the method will not yield a clear minimum, and a different approach needs to be adopted, possibly one that does not rely on a phase space representation. When this method is applied directly to a time series exhibiting strong serial correlations, it yields embedding parameters which have a preference for \(\tau_{\text{opt}} = 1\).

In order to ensure robustness of this method to the dimensionality and serial correlations of a time series, 50 suggested to use the iAAFT method for surrogate generation since it retains within the surrogate both signal distribution and approximately the autocorrelation structure of the original signal. In this article, we opt to use wavelet-based surrogate generation method, WiAAFT by in 79, for reasons already discussed in the previous section.

\(^4\)Time delay embedding is an established method for visualising an attractor of the underlying nonlinear dynamical signal, when processing signals with structure (69).
4.2.3 'Delay Vector Variance' method

The Characterization of signal nonlinearities, which emerged in physics in the mid-1990s, have been successfully applied in predicting survival in heart failure cases and also adopted in practical engineering applications (70, 101). The 'delay vector variance' (DVV) method (51) is a recently proposed phase space based method for signal characterization. It is more suitable for signal processing application because it examines the deterministic\(^5\) nature of a time series and when combined with the concept of surrogate data, provides as additional account of the nonlinear behavior of the time series. The DVV-analysis is based on the calculation of the target variance, \(\sigma^*^2\), which is an inverse measure of the predictability of a time series. The algorithm is summarized below:

- For an optimal embedding dimension \(m\) and time lag \(\tau\), generate delay vector (DV): \(x(k) = [x_{k-\tau}, \ldots, x_{k-m\tau}]\) and corresponding target \(x_k\)
- The mean \(\mu_d\) and standard deviation, \(\sigma_d\), are computed over all pairwise distances between DVs, \(\|x(i) - x(j)\|\) for \(i \neq j\).
- The sets \(\Omega_k\) are generated such that \(\Omega_k = \{x(i) | \|x(k) - x(i)\| \leq \varrho_d\}\), i.e., sets which consist of all DVs that lie closer to \(x(k)\) than a certain distance \(\varrho_d\), taken from the interval \([\min\{0, \mu_d - n_d \sigma_d\}; \mu_d + n_d \sigma_d]\), e.g., uniformly spaced, where \(n_d\) is a parameter controlling the span over which to perform the DVV analysis.
- For every set \(\Omega_k\), the variance of the corresponding targets, \(\sigma_k^2\), is computed. The average over all sets \(\Omega_k\), normalized by the variance of the time series, \(\sigma_x^2\), yields the target variance, \(\sigma^*^2\):

\[
\sigma^*^2(\varrho_d) = \frac{1}{N} \sum_{k=1}^{N} \frac{\sigma_k^2(\varrho_d)}{\sigma_x^2} \tag{4.4}
\]

where \(N\) denotes the total number of sets \(\Omega_k(\varrho_d)\)

Graphical representation of DVV-analysis is obtained by plotting \(\sigma^*^2(\varrho_d)\) as function of the standardized distance, \(\varrho_d\). The minimum target variance, \(\sigma^*_{\text{min}}^2 = \min_{\varrho_d} [\sigma^*^2(\varrho_d)]\), which corresponds to the lowest point of the curve, is a measure for the amount of noise which is present in the time series. Thus, \(\sigma^*_{\text{min}}^2\) is inversely related to prevalence of the deterministic component over the stochastic one, lowest \(\sigma^*_{\text{min}}^2\) indicating a strong deterministic component. At the extreme right, the DVV plots smoothly converge to unity, as illustrated in Figure 4.1(a) and Figure 4.1(b). The reason behind this is that

\[^5\]This means that the underlying process that generate the data can theoretically be described precisely by a set of linear or nonlinear equations. Thus, the component of a time series that can be predicted from a number of previous samples. [149]
Chapter 4. Nonlinear dynamics and recurrence plots for detecting financial crisis

4.2.4 Data Analysis based on recurrence plots

The method of recurrence plots (RP) was introduced to visualize the dynamics of phase space trajectories ([43]). It is a graphical technique that depicts the different occasions when a dynamical system visits roughly the same area in the phase space. From Takens’
embedding theorem ([127]), the dynamics can be appropriately presented by a reconstruction of the phase space trajectory \( \vec{x}(t) = \vec{x}_i \in \mathbb{R}^m \) \((i = 1, \ldots, \eta, \ t = i\Delta t, \) where \( \Delta t \) is the sampling rate) in the \( m \)-dimensional phase space. For a given one-dimensional financial time series \( \{u_i\}_{i=1}^N \), the phase space vectors \( \vec{x} \) can be reconstructed by embedding the series using Takens’ time delay method \( \vec{x}_i = (u_i, u_{i+\tau}, \ldots, u_{i+(m-1)\tau}) \). The coordinates of this vector correspond to the present and lead values of the series. The parameters \( m \) and \( \tau \) are referred to as the embedding dimension and time delay respectively. We refer to the case for which \( m = 1 \) and \( \tau = 1 \) as an unembedded time series.

The recurrence plot is the calculation of an \( \eta \times \eta \) matrix

\[
R_{i,j}(\varepsilon) = \begin{cases} 
1 & : \|\vec{x}_i - \vec{x}_j\| < \varepsilon, \ \vec{x}_i \in \mathbb{R}^m, \ i, j = 1, \ldots, \eta, \ \eta = N - (m - 1)\tau, \\
0 & : \text{otherwise}
\end{cases}
\]

where \( \| \cdot \| \) is a norm (e.g. Euclidean or maximum norm) and \( \varepsilon \) is the cut-off distance which defines a region centered at \( \vec{x}_i \). If \( \vec{x}_j \) falls within this region, the state will be near to \( \vec{x}_i \) and is taken to be a recurrence of the state \( \vec{x}_i \), which implies \( R_{i,j} = 1 \). The recurrence plot is square matrix plot of the binary values \( R_{i,j} \), in which the matrix element correspond to those calendar times at which a state of a dynamical system recurs (columns and rows correspond then to a certain pair of calendar times) ([96]). In literature, further variations\(^6\) of the recurrence plots have been proposed for different purposes. In this paper, we make use of a special type of recurrence plot referred to as unthresholded recurrence plots ([74]). This recurrence plot is obtained by plotting a matrix of distances \( D_{i,j} = \|\vec{x}_i - \vec{x}_j\| \) between the vectors \( \vec{x}_i \) and \( \vec{x}_j \). As such, it is appropriate to term the unthresholded recurrence plot as distance plot. In the section that follows, we will often use the term distance plot and recurrence plot interchangeably.

In comparing the dynamics of any two time series simultaneously embedded in the same phase space, we employed the cross recurrence plot ([95, 151]). This is useful in studying the simultaneous occurrence of a state in both series. The cross recurrence plot entails testing for closeness of each point of the first trajectory \( \vec{x}_i \) \((i = 1, \ldots, \eta) \) with each point of the second trajectory \( \vec{y}_i \) \((i = 1, \ldots, \vartheta) \) resulting in \( \varepsilon \times \vartheta \) array

\[
CR_{i,j}(\varepsilon) = \begin{cases} 
1 & : \|\vec{x}_i - \vec{y}_j\| < \varepsilon, \\
0 & : \text{otherwise}
\end{cases}
\]

We provide below a summary of our methodology which can be characterized in two stages:

\(^6\)For more details on these variations, we refer the reader to http://www.recurrence-plot.tk/variations.php.
1. Stage One: Detection of Nonlinearity in the underlying time series.

   (a) Obtain embedded parameters using differential entropy method with wiAAFT surrogates.
   
   (b) Perform Delay vector variance analysis (DVV) on the series using iAAFT surrogates to detect nonlinearity.
   
   (c) Statistical testing is perform to validate the results obtained from DVV analysis.

2. Stage Two: Detection and explaining the financial cycle.

   Data analysis is done based on recurrence plots.
   
   - Underlying financial time series is not embedded before recurrence analysis.
   
   - The series is embedded using the embedding parameter values obtained from the previous stage.

Remark 4.1. [On anticipating crisis] The distance plot, which is obtained by plotting a matrix of distances $D_{i,j}$ in section 4.2.4, uses mainly the original data. In other words, the rare events or crisis are detected based on the data available and the detections dates or periods does not change for any sample period. One implication is that it will be a challenge to use it for forecasting financial crisis directly since additional data will be required. We hereby provide below our proposed procedure to enable forecasting crisis based on distance plots:

1. Use distance plot to detect and explain crisis on any time series ($X_t$).

2. Model the original time series $X_t$ with some modeling approach. The results from the DVV analysis will facilitate the selection of appropriate class of models to use: linear or nonlinear model.

3. from step 2, perform forecast with the model to obtain data points $X_{t+h}$, $h = 1, 2, 3, \cdots, n$. Thus, extending the original data by $n$ data points.

4. Finally, we perform data analysis using recurrence (distance) plot on the extended original data i.e. $(X_1, X_2, \cdots, X_{t+n})$.

As such, in order to anticipate future crisis and even the possible impact before they happen, we recommend the above step-by-step procedure. It is important to note that the interpretation of any rare event detected by distance plots performed on the extended original data is conditional on the chosen forecasting model used in step 2 of the above procedure.
4.3 Data Analysis on S&P 500 and NASDAQ Composite

Identification of financial bubbles and crisis is a topic of major concern since it is important to prevent collapses that can severely impact nations and economies. We refer to fluctuations or swings in financial asset prices over time as financial cycle, which is characterized by periods of financial disruptions and booms. In this section, we perform analysis to characterize and detect nonlinear schemes for the U.S. financial cycle considering both S&P 500 Index and Nasdaq Composite. We consider these two indexes to better capture the U.S. financial cycle since the S&P 500 is regarded as a gauge of the large cap U.S. equities market and the Nasdaq Composite is highly followed in the U.S. as an indicator of the performance of stocks of technology companies and growth companies. Firstly, our analysis deals with the use of the recently proposed ‘delay vector variance’ (DVV) method, which examines local predictability of a signal in the phase space to detect the presence of determinism and nonlinearity in a time series. Optimal embedding parameters used in the DVV analysis are obtained via a differential entropy based method using wavelet-based surrogates. This enables the characterization of the financial time series without a priori assumption on the dynamics or statistical properties of the series. Secondly, we exploit the concept of recurrence plots on both series to locate hidden patterns, non-stationarity, and to examine the nature of these plots in events of financial crisis. In particular, we show the usefulness of recurrence plots in the diagnosis and detection of financial bubbles and crisis, which have significantly impacted economic upheavals in the past few decades.

In this paper, we consider the daily adjusted closing price of S&P 500 Index and Nasdaq Composite spanning over the period 2nd January, 1990 to 31st August, 2012. Figure 4.2 and Figure 4.3 is the plot of the daily adjusted closing price of S&P 500 Index and Nasdaq Composite respectively for the period: 1990-01-02 to 2012-08-31, implying 5716 observations.

To begin with, we opted for the differential-entropy based method (50) to determine the optimal embedding parameters, i.e., the embedding dimension, $m$, and the time lag, $\tau$, for the DVV method with WiAAFT surrogates. The optimal embedding parameters estimated for the S&P 500 in level and it’s returns are $(m = 6, \tau = 18)$ and $(m = 2, \tau = 10)$ respectively. This is displayed in Figure 4.4 as an open circle in the diagram with a clear structure. The embedding parameters obtained for the Nasdaq Composite are $(m = 5, \tau = 19)$ and it’s associated returns as $(m = 3, \tau = 2)$ as showed in Figure 4.5.

---

7The distance plots are generated with the Cross Recurrence Plot Toolbox in Matlab provided by Norbert Marwan upon request: http://tocsy.pik-potsdam.de/CRPtoolbox/

8The historical financial data can be downloaded from http://quote.yahoo.com/ using the get.hist.quote command contained in the tseries package in R.
Figure 4.2: The daily adjusted closing price of S&P 500 Index and its associated rate of change for the time period 1990-01-02 to 2012-08-31. Thus, the underlying financial time series is of length 5716 observations.

Figure 4.3: The daily adjusted closing price of Nasdaq Composite and its associated returns for the time period 1990-01-02 to 2012-08-31. This implies the time series is of length 5716 observations.

We will use the values of the embedding parameters obtained for the level of S&P 500 and Nasdaq Composite to generate delayed vectors needed to perform the DVV analysis.

The results from the DVV analysis with iAAFT surrogates performed on both the S&P 500 and Nasdaq Composite series in level indicates a clear deviation from the bisector on the DVV scatter diagram as shown in Figure 4.6 and Figure 4.7 respectively. The DVV plot also shows that the process is neither strictly deterministic or strictly
Chapter 4. Nonlinear dynamics and recurrence plots for detecting financial crisis

(a) Differential-Entropy based method with wiAAFT surrogates on S&P 500 Index. The wiAAFT surrogates on the returns of S&P 500, optimal embedding values are $m = 6$ and $\tau = 18$ with entropy ratio, $R_{ent}(m, \tau) = 0.8526$.

(b) Differential-Entropy based method with wiAAFT surrogates on the returns of S&P 500. The optimal embedding values are $m = 2$ and $\tau = 10$ with entropy ratio, $R_{ent}(m, \tau) = 0.9948$.

Figure 4.4: The optimal embedding parameters for S&P 500 Index obtained via the Differential-Entropy based method are indicated as an open circle in the diagrams with a clear structure.

(a) Differential-Entropy based method with wiAAFT surrogates on Nasdaq Composite. The wiAAFT surrogates on the returns of Nasdaq optimal embedding values are $m = 5$ and $\tau = 19$ with entropy ratio, $R_{ent}(m, \tau) = 0.8211$.

(b) Differential-Entropy based method with wiAAFT surrogates on the returns of Nasdaq Composite. The optimal embedding values are $m = 3$ and $\tau = 2$ with entropy ratio, $R_{ent}(m, \tau) = 0.9959$.

Figure 4.5: The optimal embedding parameters for Nasdaq Composite obtained via the Differential-Entropy based method are indicated as an open circle in the diagrams with a clear structure.

stochastic. Thus, the two original time series exhibits nonlinear dynamics since the iAAFT surrogates are linear realizations of the original ([119, 120]). Statistical testing of the null of linearity using the non-parametric rank-order test, Table 4.1, indicates that both time series are nonlinear. The DVV analysis suggests that both time series under consideration behaves more of a nonlinearity with neither a strictly deterministic or strictly component.

In the following step, we consider two cases in the identification and characterization of financial crisis using recurrence plots. The first case consist of perform data analysis with recurrence plots without embedding the underlying financial time series. In the
Figure 4.6: This is the DVV analysis with iAAFT surrogates performed on S&P 500 using the embedding parameters obtained via the differential entropy-based method. We clearly observe a deviation from the bisector on the DVV scatter diagram. The DVV plot also indicates that the process is neither strictly deterministic or strictly stochastic. Thus, the original time series, S&P 500, exhibits nonlinear dynamics since the surrogates are linear realizations of the original ([119, 120]).

Figure 4.7: This is the DVV analysis with iAAFT surrogates performed on Nasdaq Composite using the embedding parameters obtained via the differential entropy-based method. We clearly observe a deviation from the bisector on the DVV scatter diagram. The DVV plot also indicates that the process is neither strictly deterministic or strictly stochastic. Thus, Nasdaq Composite exhibits nonlinear dynamics since the surrogates are linear realizations of the original.

In the second case, recurrence plots are performed on embedded versions of the underlying financial times using the previously estimated values of embedding parameters.
Table 4.1: Results of the non-parametric rank-order test. The null of linearity is rejected as soon as the Rank-Order is greater than Rank-Threshold. The code $H$ takes the value 0 or 1, where $H = 0$ corresponds to failure of rejecting the null of linearity and $H = 1$ the rejection of linearity for nonlinearity. The number of iAAFT surrogates considered for the DVV-analysis is 25, which is greater than the minimum requirement of 19 surrogates for testing at $\alpha = 0.05$ level of significance.

Figure 4.8: This is an unthresholded recurrence plot which is sometimes termed distance plot ([74]). It is a matrix plot with calendar time on both the vertical and horizontal axes for the time period 1990-01-02 to 2012-08-31. The colormap corresponds to the distance to the next recurrence of a state in the time series. This representation shows the distances between states and enhance understanding of the phase space trajectory of S&P 500 and embedded version of S&P 500. The existence of butterfly-like structures, of a minimum size of six months, along the main diagonal (bisector) indicates a financial crisis. In this case, the start and end of a financial crisis corresponds the start and end on the formation of butterfly-like structure along the main diagonal.

We provide in Figure 4.8 and Figure 4.9 the unthresholded recurrence plots which is sometimes termed distance plot ([74]) on both underlying financial time series. It is a
matrix plot with both vertical and horizontal axes corresponds to calendar dates. The colormap corresponds to the distance to the next recurrence of a state in the time series. The results obtained for unembedded and embedded version of both time series do not differ in terms of identification of the financial crisis. The recurrence plot shows the distances between states and enhance understanding of the phase space trajectory of the series. The existence of butterfly-like structures along the main diagonal (bisector) indicates a financial crisis. The qualitative interpretation of the existence of disruptions (butterfly-like or band structures) are as a result of nonstationarity; some states are rare or far from the normal; transitions may have occurred ([43, 96]). In this case, the start and end of a declining financial activity corresponds the start and end on the formation of butterfly-like structure along the main diagonal. In this paper, we designate any butterfly-like structure with a minimum size\(^9\) of six months as a financial crisis. A sequential increase in the size of the butterfly-like structures along the main diagonal is an indication of a boom in a bubble cycle. In this way, the start of such a sequential increase in size of butterfly-like structures can be interpreted as the start of a financial bubble. For instance, Figure 4.9 shows that the Nasdaq Composite’s dot-com bubble started from late 1995 and reached its peak in March, 2000. We obtain that the financial crisis that resulted from the collapse of the dot-coms bubble began in March, 2000 and ended in April, 2003. The 2007-2012 global financial crisis, also known as the Global Financial Crisis and 2008 financial crisis, is also well detected on the recurrence plot. We discover a financial crisis\(^10\) that began in July 2007 ended in March 2009. This result is interesting since it confirms that stock prices series tend to shift direction in advance of the business cycle. We detect butterfly-like structure which corresponds to the existence of financial crisis for the period January, 2011 - January, 2012.

The recurrence plots, displayed in this work, allows to study the recurrence of a state at a particular calendar date. Fixing a period or date on the horizontal axis, we are able to observe the recurrence of such events along the calendar dates on the vertical axis. We also find that Nasdaq was well exposed and hit by the dot-com bubble of the 1990s compared to S&P 500. Nasdaq, on which many dot-coms traded, rose to record highs as indicated by larger distance to next recurrence. Similar features displayed in Figure 4.8 and Figure 4.9 suggests strong financial interdependence.

On studying that the simultaneous occurrence of financial bubbles and crisis in both S&P 500 and Nasdaq Composite, we perform a cross recurrence plot on the series without embedding. The results, as seen in Figure 4.10, shows that both financial time series were hit by the dot-com bubble.

\(^9\)The size refers to the length in time from the start and end on the formation of butterfly-like structure along the main diagonal

\(^10\)According to the U.S. National Bureau of Economic Research (NBER), the recession began in December 2007 ended in June 2009,
Chapter 4. Nonlinear dynamics and recurrence plots for detecting financial crisis

Figure 4.9: The distance plot of the Nasdaq Composite and its embedded version. It is observed that the Nasdaq was well exposed and hit by the dot-com bubble of the 1990s compared to S&P 500. Nasdaq, on which many dot-coms traded, rose to record highs as indicated by larger distance to next recurrence.

The returns cycle\(^{11}\) is often subject to very short-term fluctuations mainly due to transitory events making the peaks of such cycles extremely difficult to date. In the recurrence plots, Figure 4.11, we are enabled to visualize and date these peaks in both returns of the underlying series. The peaks dates are recognized by visualising higher distances to the next recurrence.

Finally, to illustrate our proposed procedure in identifying future crisis and even the possible impact before they happen, as discussed in remark 4.1, we consider the monthly S&P 500 Stock Price Index\(^{12}\) from 1990-01-01 to 2012-08-01. It is worth pointing out that the distance plot obtained using the monthly data does not differ from Figure 4.8. The results from the DVV analysis for S&P 500 in Figure 4.6 suggests that the stock

\(^{11}\) This cycle is equivalent to the growth rate cycle used in analyzing economic cycles. Suppose, we denote the financial time series by \(X_t\), then the returns is defined by \(Y_t = \log X_t - \log X_{t-1}\).

\(^{12}\) The data is downloaded from Federal Reserve Economic Data (FRED). The S&P 500 is regarded as a gauge of the large cap U.S. equities market.
price index is characterized by nonlinear dynamics. In this work, we perform a 12 month ahead forecast of the monthly S&P 500 using an additive nonlinear autoregressive model of the form

\[ x_{t+1} = \mu + \sum_{i=1}^{m_t} g_i(x_t - (i-1)d) + \varepsilon_{t+1} \quad (4.7) \]

where \( d \) is time delay, \( m \) is the embedding dimension and \( g_i \) are nonparametric univariate functions of lagged time series values represented by penalized cubic regression splines ([68, 150]). Model (4.7) is also referred to as a nonparametric additive autoregressive model ([68]). We assume that error distribution to be Gaussian, \( \varepsilon_{t+1} \sim i.i.d. \mathcal{N}(0, \sigma_{t+1}^2) \) and apply the model (4.7) with \( m = 2, d = 1 \), to the monthly S&P 500 Index. The details of estimation of this model is well discussed in [150] and [68]. The results of the 12 month ahead forecast based on model (4.7) are reported in Table 4.2. Thus, the original data can now be extended by 12 data points as already presented in remark (4.1). The \textit{distance plot} on the extended data with range 1990-01-01 to 2013-08-01 is displayed in Figure 4.12. The last \textit{butterfly-like} structure observed in Figure 4.12 started in March 2012 and ended October 2012. Based on the forecast values obtained from the model (4.7) for the monthly S&P 500 Index and the \textit{distance plot} in Figure 4.12, we do not anticipate a financial crisis for the United States in about six months from August 2012. In studying the possible impact of a financial crisis after it is identified, we consider the monthly S&P 500 stock price index from January 1990 to February 2008 since a crisis is identified to have began in July 2007. We perform a three month ahead forecast based on model (4.7) and then a recurrence plot analysis on the extended data. In Figure 4.13, the region of the two ellipse drawn on the plot shows that crisis that began in July
Figure 4.11: The distance plot of the growth rate (returns) of S&P 500 and Nasdaq Composite. The growth rate cycle is often subject to very short-term fluctuations mainly due to transitory events making the peaks of such cycles extremely difficult to date. In this case, we are enable to visualize and date these peaks in both underlying series. The peaks dates are recognized by visualizing higher distances to the next recurrence.

2007 is a recurrence of the financial crisis that resulted from the collapse of the dot-coms bubble. Thus, the possible impact and duration of the financial crisis that began in July 2007 will be similar to the dot-com bubble. We observe from Figure 4.8, looking at the “wings” of the butterfly-like structure, that this results on the possible impact of the crisis that began in July 2007 does not differ much from the dot-coms bubble.
Chapter 4. Nonlinear dynamics and recurrence plots for detecting financial crisis

Figure 4.12: The distance plot on an extended monthly S&P 500 stock price Index: original data range 1990-01-01 to 2012-08-01 with an additional 12 months ahead forecast obtained via the additive nonlinear autoregressive model (4.7). There is no signal of a possible occurrence of financial crisis in about six months from August 2012. The last butterfly-like structure observed started in March, 2012 and ended October 2012.

Figure 4.13: The distance plot on an extended monthly S&P 500 price index: original data range 1990-01 to 2008-02 with an additional 3 months ahead forecast obtained via the additive nonlinear autoregressive model (4.7). The region of the two ellipse drawn on the plot shows that crisis that began in July 2007 is a recurrence of the financial crisis that resulted from the collapse of the dot-coms bubble. Thus, the possible impact and duration of the financial crisis that began in July 2007 will be similar to the dot-com bubble.

4.4 Conclusion

In this paper, we have studied the dynamics of two financial time series using non-parametric methods which are essentially data-driven and carry no a priori assumptions on the statistical properties, such as possible non-stationarity, or mathematical structure of the time series. We have provided a comprehensive analysis of the feasibility of our approach as essential in selecting the appropriate class of models suggested by the data itself. Finally, we have demonstrated the usefulness of recurrence plots in identifying,
<table>
<thead>
<tr>
<th>Month</th>
<th>Forecast value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012:09</td>
<td>1407.371</td>
</tr>
<tr>
<td>2012:10</td>
<td>1402.220</td>
</tr>
<tr>
<td>2012:11</td>
<td>1395.267</td>
</tr>
<tr>
<td>2012:12</td>
<td>1388.223</td>
</tr>
<tr>
<td>2013:01</td>
<td>1381.473</td>
</tr>
<tr>
<td>2013:02</td>
<td>1375.089</td>
</tr>
<tr>
<td>2013:03</td>
<td>1369.070</td>
</tr>
<tr>
<td>2013:04</td>
<td>1363.397</td>
</tr>
<tr>
<td>2013:05</td>
<td>1358.050</td>
</tr>
<tr>
<td>2013:06</td>
<td>1353.009</td>
</tr>
<tr>
<td>2013:07</td>
<td>1348.255</td>
</tr>
<tr>
<td>2013:08</td>
<td>1343.769</td>
</tr>
</tbody>
</table>

Table 4.2: This is the 12 month ahead forecast of monthly S&P 500 stock price Index based on the additive nonlinear autoregressive model (4.7).

dating and explaining financial bubbles and crisis. The study provides a proposed outline on how to anticipate these rare events and even their impacted before occurrence. The findings from the data analysis with recurrence plots, in this case distance plots, shows that these plots are robust to extreme values, non stationarity and to the sample; are replicable and transparent; are adaptive to different time series and finally, can provide better chronology of financial cycles since it avoids revision of crisis dates through time.

Acknowledgements

The authors are grateful to anonymous referees of NAJEF for their careful revision, valuable suggestions, and comments that have improved this paper.
Chapter 5

The Univariate MT-STAR Model and a new linearity and unit root test procedure

5.1 Introduction

Smooth transition autoregressive (STAR) models ([35, 130]) and in general nonlinear time series models have been successfully applied in explaining the behavior of different macro-economic time series, such as output, (un)employment, and exchange rates, at different phases of the business cycle. In particular, Exponential Smooth Transition Autoregressive (ESTAR) models have been used for modeling real exchange rates and real interest rates, where the presence of a unit root cannot be rejected using conventional linear unit root tests ([39], [110], [1], [90], [105], [44], [128], [146] and [111]). This has led researchers to develop new testing procedures to detect the presence of nonlinear mean reversion against non-stationarity. Papers that developed such type of tests include [77], [107], [40], [32].

Recently, [77] proposed a test, denoted KSS test, to detect the presence of a particular kind of nonlinear stationary dynamics through the Exponential Smooth Transition Autoregressive (ESTAR) model, which was originally proposed by [58]. Unfortunately, the results of their simulation study indicated that the test does not present completely satisfactory power properties. In this paper, we are mainly concerned with the limitations of the ESTAR model and to develop a new procedure to test for unit roots in a nonlinear framework considering a STAR type model. In particular, we develop a test (ABG test), which enables us to distinguish between a linear non-stationary process and a specific new nonlinear globally stationary STAR process. In view of this objective, we
introduce a new STAR model named MT-STAR model, which has similar properties as the ESTAR model but reduces the effects of the identification problem ([41]) and can also account for asymmetry in the adjustment mechanism towards equilibrium.

The nonlinear structure of the ESTAR model leads to the presence of unidentified parameters for some combinations of the transition parameter and the error term variance. As discussed in the literature ([93], [58], [91]), it is usually difficult to obtain good parameter estimates in ESTAR models, since very small values of the error term variance lead to an unidentified transition parameter, making nearly impossible to obtain consistent estimates of the transition parameter. In a seminal paper, [41] addressed the so-called identification problem of the ESTAR model – the problem of properly distinguishing the transition function in relation to extreme parameter combinations – and proposed an alternative model to the ESTAR model namely the T-STAR model. It is noteworthy that these two competitive models, ESTAR and T-STAR, can be very useful in modeling adjustment processes, which is a growing and relevant part of the econometric literature. However, these two models are limited by the assumption of symmetric adjustment in the transition to equilibrium, which in practice might not be the same for a given degree of positive or negative deviation from equilibrium. The main concern about the symmetric assumption is that if the adjustment towards equilibrium is asymmetric, the alternative hypothesis in the ESTAR model (similarly in the T-STAR) will be mis-specified and tests based on the ESTAR model might not be valid.

Given this limitation, we first introduce a new STAR model, the MT-STAR model, based on a new and more general smooth transition function which nests the T-STAR model by [41] and also accounts for cases where the adjustment towards equilibrium is not necessarily symmetric. Accordingly, the MT-STAR model can be viewed as a modified T-STAR model. For practical purposes, we focus on a particular case of this new transition function and then develop a linearity test and a unit root test for this new STAR model. Regarding the new unit root test, the hypothesis are defined to be linear non-stationary process (unit root) under the null against nonlinear globally stationary MT-STAR process under the alternative. The results indicates that the new test statistic has more sophisticated power properties than some other unit root test proposed by [77], [107] and [40].

Some Monte Carlo experiments allows us to compare finite sample properties of the ABG test and existing alternative unit root tests under a variety of conditions. The new ABG test is correctly sized and quite often superior in terms of power. In particular, it exhibits higher power compared to KSS when the data generating process is nonlinear with asymmetric adjustment to the long-run equilibrium. Finally, the model and the tests are illustrated in an empirical application to monthly real effective exchange rates
for the Euro Zone. The results indicate that the new ABG test shows new evidence on
the stationarity of the EU real effective exchange rates which suggests the validity of
the Purchasing Power Parity (PPP).

The paper is organized as follows. Section 5.2 introduces STAR models and presents
drawbacks of using ESTAR model in modelling adjustment processes. After introducing
the new model in Section 5.3, we discuss in Section 5.4 a linearity and unit root test
associated with this new model. The non-standard limiting distribution of the new
unit root test, ABG test, is derived and consistency of the test is proven. We also
show that the limiting distribution remains unchanged if we account for potential serial
correlation in the innovation terms by augmenting the test regression with lags of the
dependent variable. In Section 5.5, a Monte Carlo study is performed and in Section
5.6 the empirical illustration is discussed. Proofs are given in the appendix.

5.2 Background

5.2.1 Overview of STAR models and Notations

Let \((y_t)_{t \in \mathbb{N}}\) be a univariate stochastic process. \((y_t)_{t \in \mathbb{N}}\) is called STAR\((p)\), \(p \geq 1\), if

\[
y_t = [\Psi \omega_t] \cdot [1 - G(y_{t-d}, \gamma, c)] + [\Theta \omega_t] \cdot G(y_{t-d}, \gamma, c) + \varepsilon_t, \quad (5.1)
\]

where \(d \leq p\), \(\Psi = (\psi_0, \psi_1, \cdots, \psi_p)\), \(\Theta = (\vartheta_0, \vartheta_1, \cdots, \vartheta_p)\), and \(\omega_t = (y_{t-1}, \cdots, y_{t-p})'\). The equation (5.1) can be reparametrized as:

\[
y_t = [\Psi \omega_t] + [\Phi \omega_t] \cdot G(y_{t-d}, \gamma, c) + \varepsilon_t, \quad t \geq 1, \quad (5.2)
\]

where \(\Phi = (\varphi_0, \varphi_1, \cdots, \varphi_p) = (\vartheta_0 - \psi_0, \vartheta_1 - \psi_1, \cdots, \vartheta_p - \psi_p)\). The process \((\varepsilon_t)\) is assumed to be a martingale difference sequence with respect to the history of the time series up to time \(t - 1\), denoted as \(\Omega_{t-1} = \{y_{t-1}, \cdots, y_{t-p}\}\), i.e., \(E[\varepsilon_t | \Omega_{t-1}] = 0\). For computational purposes, we restrict the conditional variance of the process \((\varepsilon_t)\) to be constant, \(E[\varepsilon_t^2 | \Omega_{t-1}] = \sigma^2\). However, this restriction could be relaxed by allowing for potential asymmetric autoregressive conditional heteroscedasticity.

The transition function \(G(\cdot, \gamma, c) : \mathbb{R} \to [0, 1]\), which models the regime-switching behavior, depends on three parameters: \(\gamma\) which controls the degree of nonlinearity, the threshold \(c\) and the delay \(d\) which can be chosen to maximize the goodness of fit over \(d = \{1, 2, \cdots, d_{max}\}\) ([77]). In practice this last parameter is often chosen to be equal to 1, therefore \(y_{t-d} = y_{t-1}\) in (5.1) and (5.2). In \(\Psi(\cdot)\) and \(\Phi(\cdot)\), the parameter \(p\) is
generally determined using the Akaike Information Criterion (AIC). In literature, two main functions are used:

- The Logistic function

\[ G(y_{t-d}, \gamma, c) = \left(1 + \exp\{\gamma(y_{t-d} - c)\}\right)^{-1}, c \in \mathbb{R}, \gamma > 0, \tag{5.3} \]

and the resultant model is called the logistic STAR [LSTAR] model.

- The Exponential function

\[ G(y_{t-d}, \gamma, c) = 1 - \exp\{-\gamma(y_{t-d} - c)^2\}, c \in \mathbb{R}, \gamma > 0, \tag{5.4} \]

and the resultant model is called the ESTAR model.

### 5.2.2 Drawbacks of the ESTAR model

The nonlinear structure of the ESTAR model leads to the presence of unidentified parameters, occurring for some combinations of the transition parameter \( \gamma \) and the error term variance \( \sigma^2 \). In the ESTAR setting, very small values of \( \sigma^2 \) involve an unidentified \( \gamma \), making nearly impossible to obtain a consistent estimate of \( \gamma \). [41] addressed the so-called identification problem of the ESTAR (i.e. the problem of properly distinguishing the transition function in relation to extreme parameter combinations) by showing that the variance of the conditional maximum likelihood estimator \( \hat{\gamma} \) tends to infinity as \( \sigma^2 \) vanishes (see Lemma 2.4 of [41]). The authors proposed a new type of nonlinear model formulation named T-STAR using an alternative transition function to (5.4) given by

\[ T(y_{t-d}, \gamma, c) = 1 - (1 + (y_{t-d} - c)^2)^{-\gamma}, \gamma > 0, \tag{5.5} \]

which shares same properties of (5.4) and also reduces the identification problem associated with ESTAR models. In addition the authors proposed a linearity and a unit root test for the new model. In this paper, we extend this last work by first introducing a new model which takes into account asymmetric adjustments, and then developing a linearity and a unit root test for our model.

### 5.3 A new Model: The MT-STAR model

In this section, we extend the work of [41] by introducing the possibility of asymmetric adjustment towards the equilibrium. In particular, we introduce a more general smooth
transition function which nests the T-STAR model and also accounts for cases where the adjustment towards the equilibrium is not necessarily symmetric.

5.3.1 A general MT-STAR(n,p) model

We define the general univariate MT-STAR(n,p) model of order p for a process \((y_t)\) by the following representation:

\[
\begin{cases}
y_t = [\Psi \omega_t] \cdot [1 - G_n(y_{t-d}, \gamma, c)] + [\Theta \omega_t] \cdot G_n(y_{t-d}, \gamma, c) + \varepsilon_t \\
G_n(y_{t-d}, \gamma, c) = 1 - (1 + f_n(y_{t-d}, \gamma, c))^{-\gamma}, \quad n \in \mathbb{N} \\
f_n(y_{t-d}, \gamma, c) = (\prod_{i=1}^n (y_{t-d} - c_i))^2 + k(\prod_{i=1}^n (y_{t-d} - c_i)), \quad c_1 \leq c_2 \leq \cdots \leq c_n, k \in \mathbb{R},
\end{cases}
\]

(5.6)

where \(G_n(\cdot, \gamma, c)\) is the general nth order smooth transition function, \(n\) is the degree of the polynomial of the transition function, \(d \leq p\), \(\Psi = (\psi_0, \psi_1, \cdots, \psi_p)\), \(\Theta = (\vartheta_0, \vartheta_1, \cdots, \vartheta_p)\), and \(\omega_t = (1, y_{t-1}, \cdots, y_{t-p})'\). It can also can be reparametrized as follows:

\[
y_t = [\Psi \omega_t] + [\Phi \omega_t] \cdot G_n(y_{t-d}, \gamma, c) + \varepsilon_t, \quad t \geq 1,
\]

(5.7)

where \(\Phi = (\varphi_0, \varphi_1, \cdots, \varphi_p) = (\vartheta_0 - \psi_0, \vartheta_1 - \psi_1, \cdots, \vartheta_p - \psi_p)\). Assuming \(n = 1\) in equations (5.7) and denoting \(G(\cdot, \gamma, c) = G_1(\cdot, \gamma, c)\), we obtain:

\[
\begin{cases}
y_t = [\Psi \omega_t] + [\Phi \omega_t] \cdot G(y_{t-d}, \gamma, c) + \varepsilon_t \\
G(y_{t-d}, \gamma, c) = 1 - (1 + (y_{t-d} - c)^2 + k(y_{t-d} - c))^{-\gamma}.
\end{cases}
\]

(5.8)

and we can observe that the MT-STAR model (5.8) incorporates the component \(k(y_{t-d} - c)\) to the transition function (5.5), which makes the functional form of \(G(y_{t-d}, \gamma, c)\) not necessarily symmetric:

- For \(k = 0\), the function \(G(y_{t-d}, \gamma, c)\) in (5.8) induces a non monotonic change which is symmetric around \(y_{t-d} = c\). Suppose \(\gamma \rightarrow \infty\) then \(G(\cdot) \rightarrow 1 - \mathcal{I}_c\), where \(\mathcal{I}_c\) is the indicator function at \(c\), which corresponds to a single abrupt break only at \(y_{t-d} = c\). This generates the same behavior as the transition function \(T(y_{t-d}, \gamma, c)\) defined in (5.5) proposed by [41].

- For \(k \neq 0\), the transition function \(G(y_{t-d}, \gamma, c)\) is asymmetric around the equilibrium \(c\) of \(y_t\).

Moreover, the transition function \(G(y_{t-d}, \gamma, c)\) has the following interesting properties:
Chapter 5. The Univariate MT-STAR Model and a new linearity and unit root test procedure

- if $\gamma = 0$, $G(y_{t-d}, \gamma, c) = 0$ and we are back to a linear error correction framework where the process $(y_t)$ follows an autoregressive (AR) model in our representation.

- If $\gamma > 0$, $G(y_{t-d}, \gamma, c)$ approximates to 0 when $y_{t-d}$ is near $c$ and approximates to 1 when $y_{t-d}$ approaches $\pm \infty$. Thus $G(y_{t-d}, \gamma, c)$ has the same properties of the function $T(y_{t-d}, \gamma, c)$ defined in (5.5) eventhough $G(y_{t-d}, \gamma, c)$ is asymmetric around the equilibrium $c$ of $y_t$.

Hence, by not imposing the assumption of symmetry, the function $G(y_{t-d}, \gamma, c)$ in the MT-STAR model (5.8) is more general than the transition function $T(y_{t-d}, \gamma, c)$ in the T-STAR model of [41]. Accordingly in modeling nonlinear adjustment processes without a priori knowledge on their symmetry, the MT-STAR model would be more appropriate than the T-STAR and the ESTAR models whose definition appear too restrictive. The MT-STAR model can therefore be seen as a modified version of the T-STAR and also as an alternative model to the M-ESTAR model (see [40]), applicable to the same situations. The existence and uniqueness of a stationary distribution for the process $\{y_t, t \geq 1\}$ in (5.8) is guaranteed by the geometric ergodicity (see [142], [136], [34], [37], [46]) as long as the density function of $\varepsilon_t$ is positive everywhere on the real line and satisfies the condition

$$|\psi_i + \varphi_i| < 1 \quad \forall i \in \{1, 2, \cdots, p\}.$$ 

5.3.2 MT-STAR(1,1) Modelling

For application reasons, we restrict to the univariate MT-STAR (1,1) model with delay, $d = 1$,

$$\begin{cases} y_t = \psi y_{t-1} + \varphi y_{t-1} G(y_{t-1}, \gamma, c) + \varepsilon_t \\ G(y_{t-1}, \gamma, c) = 1 - (1 + (y_{t-1} - c)^2 + k(y_{t-1} - c))^{-\gamma} \end{cases}$$

This process $(y_t)$ is geometrically ergodic if $|\psi + \varphi| < 1$. Consider now the following cases with respect of the properties of the process $(\varepsilon_t)$:

1. If $(\varepsilon_t)$ is a strong white noise $(0, \sigma^2)$, the model (5.9) can be reparameterized as

$$\Delta y_t = \beta y_{t-1} + \varphi y_{t-1} G(y_{t-1}, \gamma, c) + \varepsilon_t,$$

where $\beta = \psi - 1$, $\Delta y_t = y_t - y_{t-1}$ and it is geometrically ergodic when $-2 < \beta + \varphi < 0$. If the smoothness parameter $\gamma$ approaches zero, the MT-STAR model becomes a linear AR(1) model, i.e. $\Delta y_t = \beta y_{t-1} + \varepsilon_t$, that is stationary if $-2 < \beta < 0$. 

Chapter 5. The Univariate MT-STAR Model and a new linearity and unit root test procedure

2. If the process \((\varepsilon_t)_t\) follows the form:

\[
\varepsilon_t = \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + u_t, \quad (5.11)
\]

with \((u_t)_t\) a strong white noise \((0, \sigma^2)\), then the model \((5.10)\) becomes

\[
\Delta y_t = \beta y_{t-1} + \varphi y_{t-1} G(y_{t-1}, \gamma, c) + \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + u_t. \quad (5.12)
\]

The model equation \((5.12)\) is interesting for testing purposes ([39] and [118]) and it includes \((5.10)\) as a particular case. The effective determination of the speed of adjustment in the MT-STAR model arises for \(\gamma > 0\). The adjustment coefficient changes smoothly from \(\beta\) when \(y_{t-1}\) is at equilibrium to \(\beta + \varphi\) when \(y_{t-1}\) is far from \(c\). The adjustment speed depends thus on the magnitude of deviations from equilibrium. This makes sense for many economic models where the underlying process tends to display mean reverting behavior for large deviations from equilibrium but might follow a unit root or even explosive behavior when the deviations are small. In the setting of the MT-STAR, it implies that when \(\beta \geq 0\), we must have \(\varphi < 0\) and \(-2 < \beta + \varphi < 0\) since under these conditions the process displays a unit root or an explosive behavior for some deviations from equilibrium and mean reversion for large deviations.

Additionally, following the current practice in the literature (e.g., [19] for threshold autoregressive models and [77] for ESTAR models) one can impose \(\beta = 0\) which implies that the process \((y_t)_t\) follows a unit root process in the first regime. When \(\beta = 0\), the MT-STAR model \((5.12)\) becomes:

\[
\Delta y_t = \varphi y_{t-1} [1 - (1 + (y_{t-1} - c)^2 + k(y_{t-1} - c))^{-\gamma}] + \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + u_t, \quad (5.13)
\]

which is globally stationary for \(-2 < \varphi < 0\) with the lag-polynomial \(\alpha(L) = 1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_p L^p\) having no roots inside the unit circle and is locally non-stationary when \(y_{t-1} = c\), since it contains a unit root.

5.4 Linearity and Non Stationarity Tests

In this section, we develop a testing procedure based on the MT-STAR model which does not impose the restriction of symmetric adjustment to the equilibrium. This allows us to overcome difficulties of the standard linear DF test and the KSS test based on the nonlinear ESTAR framework when the true process is stationary but in a nonlinear and
asymmetric way. In this respect, we first develop a test for linearity and then proceed to propose a unit root test for this model.

### 5.4.1 Linearity Test

Testing linearity against STAR modelling constitutes a first step towards building STAR models. Thus, testing linearity is important as a preliminary modelling step. The null hypothesis of linearity can be expressed as null on $\Phi$ parameters in model (5.8). Many nonlinear models are only identified when the alternative hypothesis holds (the model is genuinely nonlinear) but not when the null hypothesis is valid. Since the parameters of an unidentified model cannot be estimated consistently, testing linearity before fitting any of these models is a necessary step in nonlinear modelling.

The procedure is based on model equation (5.8). Under the linearity, we have only one regime, and no transition between two regimes. We can thus test for

$$H_0: \Phi = 0 \quad \text{vs.} \quad H_1: \text{at least one } \varphi_i \neq 0, \ i = 1, \cdots, p$$

which is equivalent to test for

$$H_0: \gamma = 0 \quad \text{vs.} \quad H_1: \gamma > 0 .$$

The MT-STAR model (5.8) reduces under the null to a linear AR($p$) model for any of the hypothesis (5.14) or (5.15). When $\gamma = 0$, $H_1$ is not identified, and thus the vector $\Phi$ and $c$ can take any value without changing the value of the likelihood function and vice versa. We thus approximate $G$ in (5.8) as follows:

$$G_h(\cdot) = \sum_{n=1}^{h} (-1)^n \frac{\gamma(\gamma+1)(\gamma+2)\cdots(\gamma+n-1)}{n!} [(y_{t-d}-c)^2 + k(y_{t-d}-c)]^n + O(.), \ \gamma > 0 .$$

By expanding terms $[(y_{t-d}-c)^2 + k(y_{t-d}-c)]^n, n = 1, \cdots, h$, and making some rearrangements, we obtain the auxiliary regression model for $y_t$ for a fixed $d \leq p$ and $h \in \mathbb{N}$,

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-d-i} + \sum_{j=1}^{p} \delta_{j,0} y_{t-j} + \sum_{j=1}^{p} \delta_{j,1} y_{t-j} y_{t-d} + \sum_{j=1}^{p} \delta_{j,2} y_{t-j}^2 y_{t-d} + \cdots + \sum_{j=1}^{p} \delta_{j,2h} y_{t-j}^{2h} y_{t-d} + \xi_t .$$

The error terms in the regression (5.17) are now denoted by $\xi_t$ rather than $\varepsilon_t$. After approximating (5.8) by (5.17), to test linearity against nonlinearity, we only need to test the nullity of parameters $\delta_{j,\cdot}$. We do this using a simple ($F$) of Fisher test and the properties of the error term under the null and thus the asymptotic distribution
of the classical $F$ test remain unaffected. We illustrate in detail this result for the MT-STAR(1, 1) model defined in (5.9):

$$y_t = \psi_1 y_{t-1} + \varphi_1 y_{t-1} \left[ 1 - \left( 1 + (y_{t-1} - c)^2 + k(y_{t-1} - c) \right)^{-\gamma} \right] + \epsilon_t,$$  

(5.18)

where $c \neq 0$. To test linearity for process $(y_t)_t$, we approximate $G$ in (5.18) by $G_3$ given by (5.16) with $d = 1$ and $h = 3$. The process $(y_t)_t$ in (5.18) becomes

$$y_t = \psi_1 y_{t-1} + \varphi_1 y_{t-1} \left[ \gamma x - \frac{1}{2} \gamma (\gamma + 1)x^2 + \frac{1}{6} \gamma (\gamma + 1)(\gamma + 2)x^3 \right] + \epsilon_t,$$  

(5.19)

where $x = [(y_{t-1} - c)^2 + k(y_{t-1} - c)]$. The auxiliary regression in (5.19) is given by

$$y_t = \psi_1 y_{t-1} + \delta_{1,0} y_{t-1} + \delta_{1,1} y_{t-1}^2 + \delta_{1,2} y_{t-1}^3 + \delta_{1,3} y_{t-1}^4 + \delta_{1,4} y_{t-1}^5 + \delta_{1,5} y_{t-1}^6 + \delta_{1,6} y_{t-1}^7 + \epsilon_t,$$  

(5.20)

where

$$\delta_{1,0} = \varphi_1 \left[ m_1 (c^2 - kc) + m_2 (c^4 - 2kc^3 + kc^2) + m_3 (c^6 - 3kc^5 + 3k^2c^3 + k^3c^3) \right],$$

$$\delta_{1,1} = \varphi_1 \left[ m_1 (k - 2c) + m_2 (6kc^2 - 4c^3 - 2kc) + m_3 (-6c^5 + 15kc^4 + 9k^2c^2 + 3k^3c^2) \right],$$

$$\delta_{1,2} = \varphi_1 \left[ m_1 + m_2 (6c^2 - 6kc + k) + m_3 (15c^4 - 30kc^3 - 9k^2c - 3k^3c) \right],$$

$$\delta_{1,3} = \varphi_1 \left[ m_2 (2k - 4c) + m_3 (-20c^3 + 30c^2k + 3k^2 + k^3) \right],$$

$$\delta_{1,4} = \varphi_1 \left[ m_2 + m_3 (15c^2 - 15kc) \right],$$

$$\delta_{1,5} = \varphi_1 \left[ m_3 (3k - 6c) \right],$$

$$\delta_{1,6} = \varphi_1 \left[ m_3 \right],$$

$$m_1 = \gamma,$$

$$m_2 = -\frac{1}{2} \gamma (\gamma + 1),$$

$$m_3 = \frac{1}{6} \gamma (\gamma + 1)(\gamma + 2).$$

Now to test linearity against nonlinear MT-STAR, we test for

$$H_0 : \delta_{1,1} = \cdots = \delta_{1,6} = 0 \quad \text{vs.} \quad H_1 : \text{at least one } \delta_{1,i} \neq 0, \quad i = 1, \cdots, 6. \quad (5.21)$$

We suggest to use the $F$-version of the LM test statistics since it has better size properties than the $\chi^2$ variant, which may be heavily oversized in small samples (145). This can be performed with the following steps:

1. Estimate the model under the null hypothesis of linearity by regressing $y_t$ on $y_{t-1}$ and compute its sum of squared residuals $SSR_0$.

2. Estimate the auxiliary regression of $y_t$ on $y_{t-1}$ and $y_{t-1}^i$ for $i = 1, \cdots, 6$ and compute its sum of squared residuals $SSR_1$.  

Chapter 5. The Univariate MT-STAR Model and a new linearity and unit root test procedure
3. Under the null (5.21) the test statistic

\[ F = \left( \frac{SSR_0 - SSR_1}{12} \right) / \left( \frac{SSR_1}{(T-14)} \right), \]  

(5.22)
is approximately distributed as \( F \) of Fisher with 12 numerator degrees of freedom and \( T - 14 \) denominator degrees of freedom.

Remark 5.1. It is noteworthy to underline that if we approximate the transition function \( G \) in (5.16) by setting \( h = 1 \) and letting \( k = 0 \):

- if \( c = 0 \), we obtain the auxiliary regression for the development of the KSS test by [77] as a special case;
- if \( c \neq 0 \), we obtain the auxiliary regression used to drive the modified Wald form test statistic of [84].

5.4.2 Test for Non Stationarity in the MT-STAR framework

We propose a testing procedure for the null hypothesis of a linear unit root process against a nonlinear globally stationary MT-STAR process, defined in (5.13) and containing a partial unit root in one regime. This unit root test is proposed with the same approach as in [77] and [84] for the STAR framework.

We distinguish three cases for testing procedure: a process \((y_t)\) defined in (5.13); a centered process \((z_t) = y_t - \mu\); and a demeaned and detrended process, \((w_t) = y_t - \mu - \alpha t\).

In the following, we restrict \( c = 0 \) and thus the process \((y_t)\) becomes:

\[ \Delta y_t = \varphi y_{t-1}[1 - (1 + y_{t-1}^2 + ky_{t-1})^{-\gamma}] + \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + u_t, \]  

(5.23)

which is a globally stationary MT-STAR process provided that \(-2 < \varphi < 0\) with the polynomial in \( L \), \( \alpha(L) = 1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_p L^p \) having no roots inside the unit circle (we assume this condition hereafter). Our test directly focuses on the parameter, \( \gamma \), which is zero under the null and positive under the alternative. Hence we test

\[ H_0 : \gamma = 0 \quad \text{vs.} \quad H_1 : \gamma > 0. \]  

(5.24)

To avoid the presence of nuisance parameters under the null hypothesis, we approximate the smooth transition function \( G(y_{t-1}, \gamma, c) \) in model (5.23) by (5.16) with \( d = 1 \) and \( h = 3 \). We detail now the three cases.
1. We assume that \((y_t)_t\) is a zero mean process. According to the auxiliary regression is:

\[
\Delta y_t = \delta_{1,2} y_{t-1}^3 + \delta_{1,4} y_{t-1}^5 + \delta_{1,6} y_{t-1}^7 + \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + u_t, \tag{5.25}
\]

\(u_t \sim i.i.d(0, \sigma^2)\). Suppose that the initial sample size is \(T + p + 1\) such that there are \(T\) observations in the regression. We test the hypothesis

\[H_0 : R\beta = r \quad vs. \quad H_1 : R\beta \neq r\tag{5.26}\]

where

\[
R = \begin{pmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\end{pmatrix}_{3 \times (p+3)} = (I_3 \ 0_{3 \times p}), \quad \beta = \begin{pmatrix}
\delta_{1,2} \\
\delta_{1,4} \\
\delta_{1,6} \\
\rho_1 \\
\rho_2 \\
\vdots \\
\rho_p \\
\end{pmatrix}_{(p+3) \times 1} \quad \text{and} \quad r = \begin{pmatrix}
0 \\
0 \\
\end{pmatrix}_{3 \times 1}.
\]

by using the statistic

\[
\hat{F}_{NL} = (R(\hat{\beta} - \beta))^\prime \left[ \hat{\sigma}^2 T R(\sum_{t=1}^{T} X_t X_t^\prime)^{-1} R \right]^{-1} R(\hat{\beta} - \beta), \tag{5.27}
\]

where

\[
X_t = \begin{pmatrix}
y^3_{t-1} \\
y^5_{t-1} \\
y^7_{t-1} \\
\Delta y_{t-1} \\
\Delta y_{t-2} \\
\vdots \\
\Delta y_{t-p} \\
\end{pmatrix}_{(p+3) \times 1}, \quad \hat{\beta} \quad \text{is the OLS estimator for the parameter} \ \beta, \quad \text{and} \quad \hat{\sigma}^2_T \quad \text{is the variance of} \ \Delta y_t \ \text{in} \ (5.25). \quad \text{Under the null of} \ \beta = 0 \ \text{and statistic} \ (5.27) \ \text{becomes}
\]

\[
\hat{F}_{NL} = \frac{\hat{\beta}^\prime [Var(\hat{\beta})]^{-1} \hat{\beta}}{3}. \tag{5.28}
\]

Before providing the non-standard limiting distribution of \(\hat{F}_{NL}\) test statistic (5.28), we need to introduce a technical result.

**Proposition 5.2.** If \((y_t)_t\) is a linear single unit root process of the form:

\[
\Delta y_t = \rho_1 \Delta y_{t-1} + \rho_2 \Delta y_{t-2} + \cdots + \rho_p \Delta y_{t-p} + u_t, \quad u_t \sim i.i.d(0, \sigma^2),
\]
with \((\gamma_j)_j = \text{Cov}(\Delta y_t, \Delta y_{t-j})\), then the following results hold:

\[
\begin{align*}
(a) \quad & \frac{1}{T} \sum_{t=1}^{T} y_{t-1}^{n_1} \rightarrow_d [\Psi(1)]^{\gamma^n} \int_0^1 [W(r)]^n dr. \\
(b) \quad & \frac{1}{T} \sum_{t=1}^{T} y_{t-1}^{n_1} u_t \rightarrow_d [\Psi(1)]^{\gamma n_1} \int_0^1 [W(r)]^{n_1} dW(r). \\
(c) \quad & \frac{1}{T} \sum_{t=1}^{T} \Delta y_{t-i} \cdot \Delta y_{t-j} \rightarrow_{d} \gamma_{[i-j]}. \\
(d) \quad & \frac{1}{T} \sum_{t=1}^{T} y_{t-1}^{n_1} \cdot \Delta y_{t-j} \rightarrow_d 0. \\
(e) \quad & \frac{1}{T} \sum_{t=1}^{T} \Delta y_{t-i} u_t \rightarrow_d N(0, \sigma^2 \gamma_0), \quad i \geq 1.
\end{align*}
\]

with \(\Psi(1) = (1 - \rho_1 - \rho_2 - \cdots - \rho_p)^{-1}\). The proof is provided proposition is provided in the Appendix.

**Theorem 5.3.** Let be the process \((y_t)_t\) defined in (5.25), and \(y_1, \cdots, y_T\) a sample of size \(T\). If \(\hat{\beta}\) is a consistent estimator of \(\beta\) and converges to its true value with rate of convergence

\[
\text{diag}(T^2, T^3, T^1/2, T^1/2, \cdots, T^1/2)_{(p+3) \times (p+3)},
\]

then the \(\tilde{F}_{NL}\) test statistic given in (5.28) has the following asymptotic distribution:

\[
\tilde{F}_{NL} \rightarrow_d v'Q^{-1}v
\]

with

\[
v = \begin{bmatrix}
\frac{1}{3} W(1)^4 - \frac{3}{2} \int_0^1 W(r)^2 dr \\
\frac{1}{6} W(1)^6 - \frac{5}{2} \int_0^1 W(r)^4 dr \\
\frac{1}{8} W(1)^8 - \frac{7}{2} \int_0^1 W(r)^6 dr
\end{bmatrix}
\]

and

\[
Q = \begin{bmatrix}
\int_0^1 W(r)^6 dr & \int_0^1 W(r)^8 dr & \int_0^1 W(r)^{10} dr \\
\int_0^1 W(r)^8 dr & \int_0^1 W(r)^{10} dr & \int_0^1 W(r)^{12} dr \\
\int_0^1 W(r)^{10} dr & \int_0^1 W(r)^{12} dr & \int_0^1 W(r)^{14} dr
\end{bmatrix}
\]

where \(W(r)\) is the standard Brownian motion defined on \(r \in [0, 1]\). Under the alternative the test is consistent.

The proof is given in the Appendix.

**Corollary 5.4.** If the process \((y_t)_t\) in (5.10) is restricted to \(c = 0\) and \(\psi = 1\)

\[
\Delta y_t = \delta_{1,2} y_{t-1}^3 + \delta_{1,4} y_{t-1}^5 + \delta_{1,6} y_{t-1}^7 + u_t,
\]

with \(u_t \sim i.i.d(0, \sigma^2)\), then the test statistic \(\tilde{F}_{NL}\) (5.28) computed for \(\beta = (\delta_{1,2}, \delta_{1,4}, \delta_{1,6})'\) has the same asymptotic distribution given in Theorem 5.3.
2. We consider now the de-meaned process \((z_t)_t:\)

\[ z_t = y_t - \frac{1}{T} \sum_{k=1}^{T} y_k, \]

and the following process

\[ \Delta z_t = \varphi z_t[1 - (1 + (z_t)^2 + k z_t)^{-\gamma}] + u_t, \quad u_t \sim i.i.d(0, \sigma^2). \quad (5.31) \]

For this model, the asymptotic distribution of \(\tilde{F}_{NL}\) test statistic (5.28) has the same form as in Theorem 5.3 replacing the Brownian motion \(W(r)\) by the de-meaned Brownian motion

\[ \tilde{W}(r) = W(r) - \int_0^1 W(r) dr. \]

3. We conclude with the de-meaned and de-trended process \((w_t)_t,\) defined as

\[ w_t = y_t - \hat{\mu} - \hat{\alpha} t, \quad (5.32) \]

where \(\hat{\mu}\) and \(\hat{\alpha}\) are the ordinary least squares estimators of \(y_t = \mu + \alpha t + \epsilon_t,\) and the auxiliary regression

\[ \Delta w_t = \varphi w_t[1 - (1 + (w_t)^2 + k w_t)^{-\gamma}] + u_t. \quad (5.33) \]

Again the asymptotic distribution of the test statistic \(\tilde{F}_{NL}\) given in (5.28) is the same as in the Theorem 5.3 replacing the Brownian motion \(W(r)\) by the de-meaned and de-trended Brownian motion

\[ \tilde{W}(r) = W(r) + (6r - 4) \int_0^1 W(r) dr + (6 - 12r) \int_0^1 r W(r) dr. \]

Derivations of \(\tilde{W}(r)\) are given in the proof of Theorem 5.1 of [122] and in [106]; the result can also be derived from Theorem 2.1 of [42]. As [106] demonstrate, \(\tilde{W}(r)\) can be thought as a de-trended Brownian motion, i.e. the residual of the projection of \(W\) onto \((1, t).\) The theoretical power of the test based on \(\tilde{F}_{NL}\) is not known, thus in the next section we investigate its empirical size and power by some Monte Carlo experiments.

### 5.5 Monte Carlo Study

In this section, we carry out a Monte Carlo simulation to study the size and power properties of the new test procedure (labelled ABG test) for finite sample sizes. In particular,
we compare the test with some existing unit root tests in the STAR framework:

1. the KSS test proposed by [77];

2. the $F_{NL}$ test proposed by [107] to test for a unit root in the asymmetric nonlinear smooth transition framework;

3. the $W_{nl}$ test of [40]. The M-ESTAR model is a modification of the ESTAR to account for cases where the adjustments to equilibrium are not necessarily symmetric;


We consider three different cases: Case 1 with the raw data, Case 2 with the de-meaned data and Case 3 with the de-trended data. In Table 5.1, we provide the asymptotic critical values for the ABG test based on the distribution given in Theorem 5.3 for the three cases. We set the sample size to $T = 10,000$ and the number of replications to 1,000,000.

<table>
<thead>
<tr>
<th>Fractile (%)</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.7217</td>
<td>5.4746</td>
<td>6.6019</td>
</tr>
<tr>
<td>5</td>
<td>3.4594</td>
<td>4.1309</td>
<td>5.1323</td>
</tr>
<tr>
<td>10</td>
<td>2.8856</td>
<td>3.5145</td>
<td>4.4439</td>
</tr>
</tbody>
</table>

Table 5.1: Asymptotic critical values of ABG statistic. Note: Case 1, Case 2 and Case 3 refer to the underlying model with the raw data, the de-meaned data and the de-trended data, respectively.

We calibrate each test, assuming a random walk under the null. Table 5.2 presents the size of the alternative tests for different sample sizes at nominal level $\alpha = 5\%$ using 50,000 replications. The proposed ABG test appears to be properly sized for a nominal level of 1% for any sample size considered. However, for a nominal level of 5% or 10% there exist some slight distortions for small samples sizes, $T < 500$, but the test approaches its nominal level as the sample size increases. This does not invalidate the use of the proposed test in small samples since a small size of the test implies that the real type I error is smaller than the nominal error and thus we are willing to accept. The results for $F_{NL}$ test for Case 3 test indicates that in small samples it tends to over-reject the null hypothesis when the true process has a unit root.

We evaluate the empirical powers of the five alternative tests. We generate a dataset according to the following Data Generating Process (DGP):

$$y_t = \psi y_{t-1} + \varphi y_{t-1}(1 - (1 + y_{t-1}^2 + ky_{t-1})^{-\gamma}) + \epsilon_t, \quad (5.34)$$
Chapter 5. The Univariate MT-STAR Model and a new linearity and unit root test procedure

<table>
<thead>
<tr>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>3.812</td>
<td>4.562</td>
</tr>
<tr>
<td>100</td>
<td>3.852</td>
<td>4.812</td>
</tr>
<tr>
<td>150</td>
<td>3.914</td>
<td>4.918</td>
</tr>
<tr>
<td>200</td>
<td>3.984</td>
<td>5.048</td>
</tr>
<tr>
<td>500</td>
<td>4.466</td>
<td>4.854</td>
</tr>
<tr>
<td>1000</td>
<td>4.718</td>
<td>4.86</td>
</tr>
<tr>
<td>5000</td>
<td>5.084</td>
<td>5.07</td>
</tr>
</tbody>
</table>

Table 5.2: The size of alternative tests [in %] at nominal level $\alpha = 5\%$

where $\varepsilon_t \sim N(0, \sigma^2)$. We choose a broad range of parameter values and sample sizes:

$\varphi = \{-1.8, -1.5, -1.2, -0.9, -0.6, -0.3, -0.05\}.$

$\gamma = \{0.5, 0.8, 1\}$, $\psi = 1$. 

$\sigma = \{0.1, 1\}$, $k = \{0, 4\}$. 

$T = \{50, 100, 200, 500, 1000\}$.

The values of $\varphi$ are chosen to satisfy the condition $-2 < \varphi < 0$ for global stationarity. The values of $k$ are considered to account for both symmetric and asymmetric adjustments. Thus, $k$ is the measure of the magnitude of asymmetry in the adjustment process. In particular, the value of $k = 4$ is considered to illustrate the cases in which the DGP is highly asymmetric. The parameter $\varphi$ indicates the difference between the regimes. As this parameter approaches zero, the DGP approximates a linear process and it is expected that the DF test performs very well. We examine the power of the various tests considering two cases for error variances $\sigma = \{0.1, 1\}$. We report simulation results for the Case 1, Case 2 and Case 3 at $\alpha = 5\%$, $\gamma = 0.8$, $T = \{50, 100, 200, 500\}$ for $\sigma = \{0.1, 1\}$, $k = \{0, 4\}$.

5.5.1 Simulation results

In general, the empirical power increases with the sample size $T$ for any value of $k$, $\gamma$ and $\sigma$. The empirical power for each test, we considered, increases as the value of the transition variable $\gamma$ increases. When the DGP is approximately linear, corresponding to $\varphi$ very small or large $\gamma$, there is no real power gain using a test different than the simple DF. This is not surprising because as $\gamma$ increases, the model becomes approximately linear. By considering the range of $\varphi$ values from $-0.05$ to $-1.8$, we observe that the empirical power increases with an increase in the magnitude of $\varphi$. Thus, the power of the tests decreases when the difference between the regimes is very small, corresponding to a small value of $\varphi$. The results indicate a good overall performance of the unit root test in all sample sizes considered, especially as sample size increases. The ability to distinguish between a unit root process and a globally stationary MT-STAR model increases when
either the difference between the regimes becomes larger or even more when the sample size increases.

A general finding is that our suggested ABG test is relatively more powerful for both symmetric \((k = 0)\) and asymmetric \((k \neq 0)\) DGP when \(\sigma = 1\) regardless of the values of \(\gamma, \varphi,\) and \(T\). We refer to Table 5.4, Table 5.6 and Table 5.8. Interestingly, in the region of the null where the series is relatively more persistent, corresponding to relatively small values of \(\gamma\) and/or \(\varphi\), the ABG test performs the best relative to the KSS test for \(T \geq 500\) and \(\sigma = 1\). From Table 5.4, Table 5.6 and Table 5.8, we obtain that for any value of \(k\), when \(T = 500\) and \(\varphi = -0.05\), the KSS test exhibits lower power relative to the ABG test. Considering that most economic time series are likely to be highly persistent or close to unit root, this might be a useful finding at least empirically ([77]).

The new ABG test performs better for the de-meaned and/or de-trended data than KSS, \(F_{NL}\) and \(W_{nl}\) tests, in terms of power, for any values of \(k\) and for all sample sizes. For instance, for \(k = 4\), when the DGP is highly asymmetric, the ABG test is more powerful than the other tests for any sample size. This result supports our prior arguments that the KSS test might be unable to detect the presence of a globally stationary process if the adjustment is asymmetric.

We also examine the behavior of the ABG unit root test in presence of very small variances for the innovation term, say \(\sigma = 0.1\). In this case the power of all the tests considered deteriorates compared to the case of the white noise distribution. However, for sample size \(T \geq 200\), we get nearly the same power properties for all the tests for any given values of \(k, \varphi\) and \(\gamma\). In particular, for small values of error variances, \(\sigma = 0.1\), and different values of \(k\) (from 0 to 4), which corresponds to the degree of asymmetry, the power of KSS test deteriorates as \(k\) increases. In Table 5.3, Table 5.5 and Table 5.7, for any \(\varphi\) and \(T\), the KSS test records lower power compared to the ABG, \(F_{NL}\) and \(W_{nl}\) tests. In particular, for \(k = 4\), the ABG, \(F_{NL}\) and \(W_{nl}\) tests exhibit good power properties for all values of \(T\).

### 5.6 Empirical Illustration

We now consider a real exercise to compare the accuracy of the different tests on real data and we focus on the analysis of the Purchasing Power Parity (PPP). In this exercise, we consider three datasets: monthly real effective exchange rate CPI deflated of the euro over the period from October 1980 to October 2011; eight bilateral real exchange rates relative to the euro over the period from January 1999 to November 2011; and five normalized real exchange rates relative to the US dollar over the period from January 1973 to June 2008. Unit root tests have become a very popular tool in the literature
concerned with the analysis of the validity of the Purchasing Power Parity (PPP), which is certainly one of the most important parities in international macroeconomics. The finding of a unit root in real exchange rates by [102] shifted the modelling approach for real exchange rates to non-linear models.

The Purchasing Power Parity (PPP) holds if and only if the real exchange rates are stationary. As such, real exchange rates should not behave like a unit root process but rather be non-linear and globally stationary processes to support PPP. Accordingly to test the unit root hypothesis means to test the non-validity of the PPP theory. Since linear unit root tests of [39] and [110] often fail to reject the null hypothesis of non-stationarity when applied to real exchange rates data, researchers tend to use nonlinear unit root tests when the model that is assumed under the alternative is congruent with economic models of financial markets.

<table>
<thead>
<tr>
<th>$\phi = -1.8$</th>
<th>$\phi = -1.5$</th>
<th>$\phi = -1.2$</th>
<th>$\phi = -0.9$</th>
<th>$\phi = -0.6$</th>
<th>$\phi = -0.5$</th>
<th>$\phi = -0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k = 0$</td>
<td>$k = 0$</td>
<td>$k = 0$</td>
<td>$k = 0$</td>
<td>$k = 0$</td>
<td>$k = 0$</td>
<td>$k = 0$</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>61.704</td>
<td>45.000</td>
<td>45.062</td>
<td>45.000</td>
<td>45.056</td>
<td>45.016</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
</tr>
</tbody>
</table>

Table 5.3: The power of alternative tests against the hypothesis of global MT-STAR stationarity [in %] at nominal level $\alpha = 5\%$: Case 1 with $\sigma = 0.1$. 

Chapter 5. The Univariate MT-STAR Model and a new linearity and unit root test procedure

<table>
<thead>
<tr>
<th>$\phi = -1.8$</th>
<th>$\phi = -1.5$</th>
<th>$\phi = -1.2$</th>
<th>$\phi = -0.9$</th>
<th>$\phi = -0.6$</th>
<th>$\phi = -0.5$</th>
<th>$\phi = -0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 100$</td>
<td>50.842</td>
<td>50.842</td>
<td>50.842</td>
<td>50.842</td>
<td>50.842</td>
<td>50.842</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\phi = -1.8$</th>
<th>$\phi = -1.5$</th>
<th>$\phi = -1.2$</th>
<th>$\phi = -0.9$</th>
<th>$\phi = -0.6$</th>
<th>$\phi = -0.5$</th>
<th>$\phi = -0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 100$</td>
<td>38.910</td>
<td>38.910</td>
<td>38.910</td>
<td>38.910</td>
<td>38.910</td>
<td>38.910</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>97.212</td>
<td>97.212</td>
<td>97.212</td>
<td>97.212</td>
<td>97.212</td>
<td>97.212</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\phi = -1.8$</th>
<th>$\phi = -1.5$</th>
<th>$\phi = -1.2$</th>
<th>$\phi = -0.9$</th>
<th>$\phi = -0.6$</th>
<th>$\phi = -0.5$</th>
<th>$\phi = -0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 50$</td>
<td>7.058</td>
<td>7.058</td>
<td>7.058</td>
<td>7.058</td>
<td>7.058</td>
<td>7.058</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>65.764</td>
<td>65.764</td>
<td>65.764</td>
<td>65.764</td>
<td>65.764</td>
<td>65.764</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\phi = -1.8$</th>
<th>$\phi = -1.5$</th>
<th>$\phi = -1.2$</th>
<th>$\phi = -0.9$</th>
<th>$\phi = -0.6$</th>
<th>$\phi = -0.5$</th>
<th>$\phi = -0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 50$</td>
<td>5.064</td>
<td>5.064</td>
<td>5.064</td>
<td>5.064</td>
<td>5.064</td>
<td>5.064</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>8.430</td>
<td>8.430</td>
<td>8.430</td>
<td>8.430</td>
<td>8.430</td>
<td>8.430</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>27.872</td>
<td>27.872</td>
<td>27.872</td>
<td>27.872</td>
<td>27.872</td>
<td>27.872</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>18.798</td>
<td>18.798</td>
<td>18.798</td>
<td>18.798</td>
<td>18.798</td>
<td>18.798</td>
</tr>
</tbody>
</table>
Chapter 5. *The Univariate MT-STAR Model and a new linearity and unit root test procedure* 79

<table>
<thead>
<tr>
<th>$k = 0$</th>
<th>$k = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varphi = -1.8$</td>
<td>$\varphi = -1.8$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>100.000</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>100.000</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>100.000</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
</tr>
<tr>
<td>$\varphi = -1.5$</td>
<td>$\varphi = -1.5$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>99.996</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>100.000</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>100.000</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
</tr>
<tr>
<td>$\varphi = -1.2$</td>
<td>$\varphi = -1.2$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>99.444</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>100.000</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>100.000</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
</tr>
<tr>
<td>$\varphi = -0.9$</td>
<td>$\varphi = -0.9$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>98.978</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>100.000</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>100.000</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
</tr>
<tr>
<td>$\varphi = -0.6$</td>
<td>$\varphi = -0.6$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>84.300</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>100.000</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>100.000</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
</tr>
<tr>
<td>$\varphi = -0.3$</td>
<td>$\varphi = -0.3$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>30.430</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>87.190</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>100.000</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
</tr>
<tr>
<td>$\varphi = -0.05$</td>
<td>$\varphi = -0.05$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>4.236</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>6.616</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>18.366</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>82.794</td>
</tr>
</tbody>
</table>

Table 5.4: The power of alternative tests against the hypothesis of global MT-STAR stationarity [in %] at nominal level $\alpha = 5\%$: Case 1 with $\sigma_{\epsilon} = 1$.

5.6.1 Application to the Euro Real Effective Exchange Rate

We apply the new ABG unit root test and the others tests ($KSS, W_{nl}, F_{NL}$, and $DF$) to the monthly real effective exchange rate CPI deflated time series for the euro. The data spans from 1980:10 to 2011:10 implying 373 observations. The log of the time series is depicted in Figure 5.1. We can observe that no linear trend appears in the data but the mean appears to be highly significant. Accordingly, we de-mean the data in a first step. In the next step we estimate the regressions with a lag length ($\hat{p} = 1$) chosen accordingly to the Bayesian information criterion (BIC). We obtain that the KSS test ($= -2.6763$) and the $\tau$ test ($= 7.4801$) fail to reject the null hypothesis of unit root at the 10% level suggesting that the PPP does not hold. Furthermore, the DF unit root test against linear alternatives does not provide any evidence against the null hypothesis. However, the ABG test ($= 11.2695$), the $W_{nl}$ ($= 14.5695$) and the $F_{NL}$ ($= 11.2245$) reject the null hypothesis at 1% level of significance. This yields new evidence on the stationarity of the
Chapter 5. *The Univariate MT-STAR Model and a new linearity and unit root test procedure*

80

Table 5.5: The power of alternative tests against the hypothesis of global MT-STAR stationarity [in %] at nominal level $\alpha = 5\%$: Case 2 with $\sigma_\epsilon = 0.1$.

<table>
<thead>
<tr>
<th>$k = 0$</th>
<th>$k = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi = -1.8$</td>
<td>$\phi = -1.8$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>$T = 50$</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>$T = 100$</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>$T = 200$</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>$T = 500$</td>
</tr>
<tr>
<td>$\phi = -1.5$</td>
<td>$\phi = -1.5$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>$T = 50$</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>$T = 100$</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>$T = 200$</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>$T = 500$</td>
</tr>
<tr>
<td>$\phi = -1.2$</td>
<td>$\phi = -1.2$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>$T = 50$</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>$T = 100$</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>$T = 200$</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>$T = 500$</td>
</tr>
<tr>
<td>$\phi = -0.9$</td>
<td>$\phi = -0.9$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>$T = 50$</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>$T = 100$</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>$T = 200$</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>$T = 500$</td>
</tr>
<tr>
<td>$\phi = -0.6$</td>
<td>$\phi = -0.6$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>$T = 50$</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>$T = 100$</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>$T = 200$</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>$T = 500$</td>
</tr>
<tr>
<td>$\phi = -0.3$</td>
<td>$\phi = -0.3$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>$T = 50$</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>$T = 100$</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>$T = 200$</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>$T = 500$</td>
</tr>
<tr>
<td>$\phi = -0.05$</td>
<td>$\phi = -0.05$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>$T = 50$</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>$T = 100$</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>$T = 200$</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>$T = 500$</td>
</tr>
</tbody>
</table>

Table 5.5: The power of alternative tests against the hypothesis of global MT-STAR stationarity [in %] at nominal level $\alpha = 5\%$: Case 2 with $\sigma_\epsilon = 0.1$.

euro real effective exchange rate which suggests the validity of PPP. It is worth noting that the time series has possibly an asymmetric adjustment to equilibrium since the only STAR type test that rejects the null hypothesis accounts for such adjustment. The false non-rejections of the null by the KSS test leading to reject a nonlinear adjustment process for real exchange rates could be due to extremely small error variances of the process. We conclude that nonlinearities with potential asymmetric adjustment to the long-run equilibrium are relevant for the data.

5.6.2 Application to Five Bilateral Real Exchange Rates relative to the US Dollar

For this application, we consider our proposed test procedure compared with other tests previously discussed on the same dataset used by [28]. Our objective is to verify if indeed
Case 2, $\sigma_\epsilon = 1, \gamma = 0.8$

<table>
<thead>
<tr>
<th>$\varphi = -1.8$</th>
<th>$\varphi = -1.5$</th>
<th>$\varphi = -1.2$</th>
<th>$\varphi = -0.9$</th>
<th>$\varphi = -0.6$</th>
<th>$\varphi = -0.3$</th>
<th>$\varphi = -0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 50$</td>
<td>100.000</td>
<td>99.998</td>
<td>99.992</td>
<td>99.986</td>
<td>100.000</td>
<td>93.380</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>97.398</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>98.816</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>99.490</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\varphi = 0.8$</th>
<th>$\varphi = 0.5$</th>
<th>$\varphi = 0$</th>
<th>$\varphi = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 50$</td>
<td>99.990</td>
<td>99.902</td>
<td>99.876</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\varphi = 0.8$</th>
<th>$\varphi = 0$</th>
<th>$\varphi = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 50$</td>
<td>99.836</td>
<td>99.184</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>99.836</td>
<td>99.184</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>99.836</td>
<td>99.184</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>99.836</td>
<td>99.184</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\varphi = 0.5$</th>
<th>$\varphi = 0$</th>
<th>$\varphi = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 50$</td>
<td>97.266</td>
<td>94.850</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>97.266</td>
<td>94.850</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>97.266</td>
<td>94.850</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>97.266</td>
<td>94.850</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\varphi = 0$</th>
<th>$\varphi = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 50$</td>
<td>5.202</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>6.796</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>14.936</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>71.856</td>
</tr>
</tbody>
</table>

Table 5.6: The Power of Alternative Tests against the hypothesis of global MT-STAR stationarity [in %] at nominal level $\alpha = 5\%$: Case 2 with $\sigma_\epsilon = 1$. 

the ESTAR model is appropriate for modelling the real exchange rates considered by [28]. The data consists of five real exchange rates relative to the US dollar corresponding to UK, Japan, Germany, France and Switzerland, from January 1973 to June 2008. These real exchange rates are constructed in the standard way as $q_t \equiv \log(CPI_{home}^{t}/CPI_{US}^{t}S_t)$, where $S_t$ is the home currency price of one US dollar. We employ our linearity test on each real exchange rates. We obtain that the only exchange rate that exhibits clear nonlinearity is the German real exchange rate series. This result shows that the real exchange rates considered in this exercise might follow a nonlinear model like the ESTAR model. In order to test for unit root, we first consider the underlying time series without de-meaning it and then perform the unit root testing procedures. We obtain ABG (= 5.0121) and KSS (= -2.7887) for the French series, which are significant at 1% and 5% level suggesting that PPP holds. For the German series, ABG (= 4.2385), KSS (= -2.7299) which are significant at 5% and $F_{NL} (= 3.1266)$ which is significant at 10%
Chapter 5. The Univariate MT-STAR Model and a new linearity and unit root test procedure

Case 3, $\sigma_\epsilon = 0.1, \gamma = 0.8$

<table>
<thead>
<tr>
<th>$k = 0$</th>
<th>$k = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varphi = -1.8$</td>
<td>$\varphi = -1.0$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>12.446</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>28.536</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>90.210</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
</tr>
<tr>
<td>$\varphi = -1.5$</td>
<td>$\varphi = -1.2$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>10.686</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>22.674</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>80.418</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
</tr>
<tr>
<td>$\varphi = -0.9$</td>
<td>$\varphi = -0.6$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>7.850</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>13.462</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>46.000</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>100.000</td>
</tr>
<tr>
<td>$\varphi = -0.5$</td>
<td>$\varphi = -0.05$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>5.712</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>9.736</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>26.740</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>99.682</td>
</tr>
<tr>
<td>$\varphi = -0.05$</td>
<td>$\varphi = -0.05$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>4.742</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>4.340</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>4.772</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>9.848</td>
</tr>
</tbody>
</table>

Table 5.7: The power of alternative tests against the hypothesis of global MT-STAR stationarity [in %] at nominal level $\alpha = 5\%$: Case 3 with $\sigma_\epsilon = 0.1$.

level. Furthermore, no test statistic fails to reject the non-validity of PPP at 5% for the Japanese, UK, and Swiss real exchange rates series. As one can see from Figure 5.2, no linear trend appears for the different series but the means appear to be highly significant. Thus we de-mean each series and perform the unit root tests. Results show that no test statistic provides support that the French, German, Japanese and Swiss real exchange rates series are nonlinear globally stationary process. However, for the UK series we obtain ABG (= 6.1335), which is significant at 1% level suggesting PPP holds. These findings provide no support for the use of the ESTAR model to forecast such real exchange rates, particularly for Japan, UK and Switzerland. Hence, our findings are consistent with the results of [28] on no forecast gained by ESTAR model over linear autoregressive model.
5.6.3 Application to Bilateral Real Exchange Rates relative the Euro

We now consider eight bilateral monthly exchange rates of Australian dollar, Canadian dollar, Swiss franc, UK pound sterling, Japanese yen, US dollar, Hong Kong dollar and South African rand relative to the euro. Our data is taken from the European Central Bank and spans from 1999:01 to 2011:11 implying 155 observations (see Figure 5.3). We obtain that linearity is rejected for the Canadian and UK exchange rates based on our linearity test procedure. By performing the unit root tests with no transformation on the exchange rates, we obtain $ABG = 4.5540$, $W_{nl} = 5.0981$, $F_{NL} = 4.1139$, and $\tau = 10.1964$, which are significant at 5% level suggesting that the PPP holds for Canadian exchange rates. We obtain evidence of possible asymmetric adjustment to the long-run equilibrium for the UK exchange rate series since $ABG = 5.0491$ and $F_{NL} = 3.2379$ are significant at 1% and 10% respectively. The bilateral exchange rates series for Australia, Switzerland, Japan, US, Hong Kong, South Africa, behave like linear
5.7 Conclusion

This paper extends the work of [41] by introducing the possibility of asymmetric adjustment to the equilibrium. Based on the work of [130] and [77], the present paper proposes a new unit root test, called ABG test, that has power against nonlinear but globally stationary alternatives, where the adjustment is smooth over time. The paper derives the unit root processes and might not present better forecast ability when modelled with ESTAR model over linear autoregressive models. Furthermore, when we perform the unit root tests on the de-meaned or de-trended real exchange rates series depending on the evolution of the time series, we obtain no evidence that PPP holds for Australia, Switzerland, Japan, US and Hong Kong. However, for the case of South Africa, we obtain $F_{NL} (= 4.0637)$, and for the UK series the $W_{nl} (= 5.9599)$, $\tau (= 11.9197)$, KSS ($= -3.2675$) which are significant at 10% level. We still obtain evidence of PPP on Canadian series with ABG ($= 6.3594$), $W_{nl} (= 10.1362)$, KSS ($= -4.3833$), $F_{nl}$ ($= 6.8746$) and $\tau (= 20.2723)$, all significant at 1% level.
asymptotic limit distributions of this new test and some empirical application show that it should be used together with tests proposed by [130] and [77] to distinguish whether the adjustment to the long-run equilibrium is either symmetric or asymmetric over time. We also assess the power performance of the ABG test by Monte Carlo experiments. A possible extension would be to establish the theoretical power of the test by allowing the alternative to be contiguous to the null hypothesis (see [55]).

Acknowledgements

The authors are grateful to anonymous referees and Associate Editor of CSDA for their careful revision, valuable suggestions, and comments that have significantly improved the original version of this contribution. We thank the conference participants of CFE–ERCIM 2011 and MAF 2012 for their participation and interest.
Chapter 6

Multivariate Self-Exciting Threshold Autoregressive Models with eXogenous Input

{PRELIMINARY VERSION–please do not quote}

6.1 Introduction

Recently there has been considerable interest in nonlinear time series analysis ([8, 24, 65, 112, 133, 134, 140], and references therein), due primarily to the various limitations encountered with linear time series models in real applications. Many nonlinear time series models have been introduced in the literature and illustrated to be useful in some applications ([53, 57, 112, 126, 139, 140]). For instance, [138, 140] proposed the threshold autoregressive (TAR) model and showed its usefulness in describing the asymmetric limit cycle of the annual sunspot number. Let \((\Omega, \mathcal{F}, P)\) be a probability space, \(R = \bigcup_{j=1}^{l} R_j,\)

\(R_j = (r_{j-1}, r_j], -\infty = r_0 < r_1 < \cdots < r_l = \infty\) a disjunctive decomposition of the real axis. Let \(d, p_1, \cdots, p_l \in \mathbb{Z}^+.\) Any solution of \((y_t)_t\) of

\[
y_t + \sum_{j=1}^{l} y_{t,d}^{(j)} \left( a_0^{(j)} + \sum_{i=1}^{p_l} a_i^{(j)} y_{t-i} \right) = \sum_{j=1}^{l} y_{t,d}^{(j)} \triangleq t
\]

\( (6.1) \)
where

$$y_{t,d}^{(j)} = \begin{cases} 1; & y_{t-d} \in R_j \\ 0; & y_{t-d} \notin R_j \end{cases}$$

is a univariate Self–Exciting Threshold Autoregressive process denoted by SETAR \((l, p_1, \cdots, p_l)\) with delay \(d\) (see [139, 140] and the references therein). The process \((y_t)\) is assumed to be ergodic and its stationary distribution has a finite second moment. The process \((\varepsilon_t)_{t}^{(j)}\) in model equation (6.1) for each regime \(j\) is assumed to be a martingale difference sequence with respect to an increasing sequence of \(\sigma\)-field, denoted as \(F_t\), i.e.,

\[E[\varepsilon_t^{(j)} | F_{t-1}] = 0.\]

In this setting, the conditional variance of the process \((\varepsilon_t)_{t}^{(j)}\) can be a constant, \(E[(\varepsilon_t^{(j)})^2 | F_{t-1}] = \sigma^2\) or allowed for possibly asymmetric autoregressive conditional heteroscedasticity. The model equation (6.1) is nonlinear in time when the number of regimes \(l > 1\) and is a piecewise linear model in the threshold space \(y_{t-d}\).

Thus SETAR model (6.1) adopts a piecewise linear setting in such a fashion that regime switches are triggered by an observed variable crossing an unknown threshold. For a review on the asymptotic theory and inference for the SETAR model (6.1), see [33, 62–64, 113, 140]. Despite the simplicity of SETAR models, they have been shown to be able to capture economically interesting asymmetries, regime changes (such as periods of low/high stock market valuations, recessions/expansions, periods of low/high interest rates, etc), and empirically observed nonlinear dynamics relevant to economic data. For instance, [108] used a single–threshold SETAR model in describing the dynamic behaviour of the three–month US T-bill interest rate.

In analysing multivariate relationships between economic variables, the linear Vector Autoregression (VAR) models have gain popularity for empirical macroeconomic modelling, policy analysis and forecasting. However, the inability of these linear models to capture non-linear dynamics such as regime switching and asymmetric responses to shocks, has gained attention in macroeconomic research. For example, a significant number of empirical studies document asymmetries in the effects of monetary policy on output growth ([117] and reference therein). In this respect, the interest in nonlinear ARX time series and regression models has been increasing in econometrics as in other disciplines ([36, 73, 133] and references therein). In this work, we consider the introduction of an exogenous input \((f_t)\) as an extension of the Multivariate SETAR model formulation and has a structural form of a nonlinear bivariate ARX model ([99]). Unlike the multivariate threshold model proposed in [141], we allow the possibility of the threshold variable to also be a multivariate process. In this case, the regime of the whole system is not necessarily determined by a single stationary subprocess. In otherwords, there exists thresholds for all subprocess of the multivariate process.

A short overview of the paper is as follows. In Section 6.2 we define the multivariate
SETAR process with exogenous input denoted MSETARX model as an extension of the multivariate SETAR model. In Section 6.3 we find conditions for stationarity of the MSETARX models, whereas Section 6.4 is used to present the LSE (least squares estimate) algorithm and an adaptive parameter estimation algorithm (13, 88) based on the stochastic gradient principles for linear systems shown to be suitable for nonlinear systems. The performance of the proposed algorithms for estimating the parameters of Multivariate SETARX models is evaluated via simulations in Section 6.4.3. In Section 6.5, the modeling procedure for the MSETARX models and problems of estimation are briefly considered.

### 6.2 Multivariate SETARX models

Consider a $D$-dimensional time series $y_t = (y_{t1}, \cdots, y_{tD})^T$ such that $L_1, \cdots, L_D \in \mathbb{Z}^+$, for each $1 \leq i \leq D$, $(R_j^i)_{j=1,2,\ldots,L_i}$ a disjunction decomposition of the real axis: $R = \bigcup_{j=1}^{L_j} R_j^i ; i \in \{1,\cdots,D\}$. Let $L = \max\{L_1, L_2, \cdots, L_D\}$ and $R_j^i = \Phi : j = L_i + 1, \cdots, L_i$. Then any solution $(y_t)_t$ of

$$y_t + \sum_{J \in \{1,\cdots,L\}^D} y_{t,d}^{(J)} \left( a_0^{(J)} + \sum_{i=1}^{P_j} A_i^{(J)} y_{t-i} \right) = \sum_{J \in \{1,\cdots,L\}^D} y_{t,d}^{(J)} \varepsilon_t^{(J)} \quad (6.3)$$

is called a multivariate SETAR process denoted MSETAR $(L, p_j; J \in \{1,\cdots,L\}^D)$, where $y_{t,d}^{(J)} : \{1,\cdots,L\}^D \mapsto \{0,1\}$ is the indicator variable defined by the following relation:

$$(y_{t,d}^{(J_1,\cdots,J_D)}) = 1 \Leftrightarrow \text{def} \quad (y_{t-d})_t \in R_j^i ; j \in \{1,\cdots,L\}^D; i \in \{1,\cdots,D\}$$

and $\{\varepsilon_t^{(J)}, \mathcal{F}_t\}$ be a sequence of martingale difference with respect to an increasing sequence of $\sigma$-field $\{\mathcal{F}_t\}$ such that

$$\sup_{t \geq 0} E[||\varepsilon_t^{(J)}|||\mathcal{F}_t] = 0 \quad a.s., \quad \sup_{t \geq 0} E[||\varepsilon_t^{(J)}||^2|\mathcal{F}_t] = \sigma^2 < \infty \quad a.s., \quad \sup_{t \geq 0} E[||\varepsilon_t^{(J)}||^\alpha|\mathcal{F}_t] < +\infty \quad a.s,$$

for some $\alpha > 2$ and $||.||$ be a matrix norm.

Now consider a $D$-dimensional time series $y_t = (y_{t1}, \cdots, y_{tD})^T$ and a $\kappa$-dimensional inputs $f_t = (f_{t1}, \cdots, f_{t\kappa})^T$ such that $L_1, \cdots, L_D \in \mathbb{Z}^+$, for each $1 \leq i \leq D$, $(R_j^i)_{j=1,2,\ldots,L_i}$ a disjunction decomposition of the real axis: $R = \bigcup_{j=1}^{L_j} R_j^i ; i \in \{1,\cdots,D\}$. Let $L = \max\{L_1, L_2, \cdots, L_D\}$ be the maximum of the number of regimes for each subprocess of
\( y_t \) and \( R^i_j = \Phi : j = L_i + 1, \ldots, L \). Then any solution \( (y_t)_t \) of

\[
\begin{align*}
    y_t + \sum_{J \in \{\ldots, L\}^D} y_{t,d}^{(J)} \left( a^{(J)}_0 + \sum_{i=1}^{p_J} A_i^{(J)} y_{t-i} + \Lambda^{(J)} \sum_{\tau=1}^{q} \Xi_{\tau} f_{t-\tau} \right) = \sum_{J \in \{\ldots, L\}^D} y_{t,d}^J \epsilon_t^{(J)} \\
    f_t = \sum_{\tau=1}^{\tau} \Xi_{\tau} f_{t-\tau} + \eta_t
\end{align*}
\]

is called a multivariate SETAR process with exogenous input denoted MSETARX \((L, p_J, q; J \in \{1, \ldots, L\}^D)\). The variables \((y_t)_t\) and \((f_t)_t\) in model (6.4) are endogenous and exogenous, respectively, and the econometrics significance of estimating the relationship between \((y_t)_t\) and \((f_t)_t\) is well known. The model equation (6.4) can be rewritten as

\[
\begin{align*}
    y_t + \sum_{J \in \{\ldots, L\}^D} y_{t,d}^{(J)} \left( a^{(J)}_0 + \sum_{i=1}^{p_J} A_i^{(J)} y_{t-i} + \Lambda^{(J)} \sum_{\tau=1}^{q} \Xi_{\tau} f_{t-\tau} \right) = \sum_{J \in \{\ldots, L\}^D} y_{t,d}^J \omega_t^{(J)}
\end{align*}
\]

where \( \omega_t^{(J)} = \epsilon_t^{(J)} - \Lambda^{(J)} \eta_t \), \( a^{(J)}_0 \) and \( \omega_t^{(J)} \) are \( D \times 1 \) vectors, \( A_i^{(J)} \) are \( D \times D \) coefficient matrices, \( \Lambda^{(J)} \) are \( D \times \kappa \) coefficient matrices, \( \Xi_{\tau} \) are \( \kappa \times \kappa \) coefficient matrices, and \( (f_t)_t \) is \( \kappa \times 1 \) vector. When \( \Lambda^{(J)} = 0 \) for all \( J \in \{1, \ldots, L\}^D \), (6.5) becomes a MSETAR model (6.3).

The representation in equation (6.5) shows that the MSETARX \((L, p_J, q; J \in \{1, \ldots, L\}^D)\) model (6.4) has approximately the same structure as the MSETAR \((L, p_J; J \in \{1, \ldots, L\}^D)\) model (6.3) with exogenous variables or factors \((f_t)_t\). For simplicity, we assume the exogenous inputs enter the model in a linear autoregressive fashion. It is worth pointing out that the dynamics of process \((f_t)_t\) could be captured by suitable linear/nonlinear model, principal components, and among other model specifications. Unlike the multivariate threshold model in [141], the threshold space is of dimension equal to the dimension of the multivariate process. Thus there exists thresholds for all subprocess of the multivariate process (6.5). In this case, the regime of the whole system is not necessarily determined by a single stationary subprocess, say \( y_{1t} \), as in [141].

Assumption 1. Let \( \{\epsilon_t^{(J)}, \mathcal{F}_t\} \) and \( \{\eta_t, \mathcal{F}_t\}; \forall J \in \{1, \ldots, L\}^D \) be two independent sequence of martingale difference with respect to an increasing sequence of \( \sigma \)-field \( \{\mathcal{F}_t\} \) such that

\[
\sup_{t \geq 0} E[\|\epsilon_{t+1}^{(J)}\|^2 | \mathcal{F}_t] = \bar{\Upsilon}_\epsilon < \infty \ a.s \quad \text{and} \quad \sup_{t \geq 0} E[\|\eta_{t+1}\|^2 | \mathcal{F}_t] = \bar{\Upsilon}_\eta < \infty \ a.s
\]

This ensures that \( \{\omega_t^{(J), \mathcal{F}_t}\} \) is a sequence of martingale difference with respect to an increasing sequence of \( \sigma \)-field \( \{\mathcal{F}_t\} \) where

\[
\sup_{t \geq 0} E[\|\omega_{t+1}^{(J)}\|^2 | \mathcal{F}_t] = \bar{\Upsilon}_\omega < \infty \ a.s
\]
Simple orthogonality assumptions on the errors \( \omega_t^{(J)} \) are insufficient to identify nonlinear models ([31]) and as such it is important that Assumption B.4 holds.

Let \( p = \max\{p_J | J \in \{1, \cdots, L\}^D \} \) and \( q \) be the model orders for model (6.5). Now, suppose that \( \omega_t^{(J)} \), and \( p \) be regime independent. We can rewrite model equation (6.5) as

\[
y_t = \sum_{J \in \{1, \cdots, L\}^D} y_{t,d}^{(J)} \left( \tilde{\Theta}_1^{(J)} \right)^T \tilde{\Phi}_{1,t-1} + \omega_t
\]

(6.6)

where \( \left( \tilde{\Theta}_1^{(J)} \right)^T = -[a_0^{(J)}, A_1^{(J)}, \cdots, A_p^{(J)}, \Lambda^{(J)} \Xi_1, \Lambda^{(J)} \Xi_2, \cdots, \Lambda^{(J)} \Xi_q] \), \( \tilde{\Phi}_t^T = [y_t^T, y_{t-1}^T, \cdots, y_{t-p+1}^T, f_t^T, f_{t-1}^T, \cdots, f_{t-q+1}^T] \) and the notation \( \zeta^T \) denotes the transpose of \( \zeta \). We remark that the MSETARX model with the representation (6.6) permits us to make use of the [13] proposed adaptive parameter estimation algorithm for the MSETAR model (6.3).

### 6.3 On the Stationarity of MSETARX model

In this section, we establish the conditions for the existence of a solution for the model equation (6.4). Let \( p = \max\{p_J | J \in \{1, \cdots, L\}^D \} \) and \( q \) be the model orders for model (6.4). Now, suppose that \( p \) and \( q \) be regime independent and \( a_0^{(J)} = 0 \) for each \( J \).

We rewrite model equation (6.4) in the form

\[
y_t = \sum_{J \in \{1, \cdots, L\}^D} y_{t,d}^{(J)} \left( \tilde{\Theta}_1^{(J)} \right)^T \tilde{\Phi}_{1,t-1} + \Lambda^{(J)} f_t + \sum_{J \in \{1, \cdots, L\}^D} y_{t,d}^{(J)} \varepsilon_t^{(J)}
\]

(6.7)

\[
f_t = \left( \tilde{\Theta}_2 \right)^T \tilde{\Phi}_{2,t-1} + \eta_t
\]

(6.8)

where \( \left( \tilde{\Theta}_1^{(J)} \right)^T = -[A_1^{(J)}, \cdots, A_p^{(J)}], \left( \tilde{\Theta}_2 \right)^T = -[\Xi_1, \Xi_2, \cdots, \Xi_q] \), \( \tilde{\Phi}_{1,t}^T = [y_t^T, y_{t-1}^T, \cdots, y_{t-p+1}^T] \) and \( \tilde{\Phi}_{2,t}^T = [f_t^T, f_{t-1}^T, \cdots, f_{t-q+1}^T] \).

The equation model (6.7)-(6.8) can be represented as a nonlinear ARX model ([99]) of the form:

\[
\begin{cases}
\dot{y}_t = g_1^{(J)}(y_{t-1}^T, \cdots, y_{t-p}^T) + g_2^{(J)}(f_t^T, \cdots, f_{t-q}^T) + \sum_{J \in \{1, \cdots, L\}^D} y_{t,d}^{(J)} \varepsilon_t^{(J)} \\
\dot{f}_t = g_3(f_t^T, \cdots, f_{t-q}^T) + \eta_t
\end{cases}
\]

(6.9)

with \( g_1^{(J)}(y_{t-1}^T, \cdots, y_{t-p}^T) = \sum_{J \in \{1, \cdots, L\}^D} y_{t,d}^{(J)} \left( \tilde{\Theta}_1^{(J)} \right)^T \tilde{\Phi}_{1,t-1}, \ g_2^{(J)}(f_t^T, \cdots, f_{t-q}^T) = \sum_{J \in \{1, \cdots, L\}^D} y_{t,d}^{(J)} \Lambda^{(J)} f_t, \ g_3(f_t^T, \cdots, f_{t-q}^T) = \left( \tilde{\Theta}_2 \right)^T \tilde{\Phi}_{2,t-1}. \) The process \( \{f_t, y_t\} \) of the equation model (6.9) is a Markov process.
Chapter 6. Multivariate Self-Exciting Threshold Autoregressive Models with eXogenous Input

Assumption 2. We denote \( \eta = (y_{t-1}^T, \ldots, y_{t-p}^T) \) and \( f = (f_1^T, \ldots, f_{t-q}^T) \). The multivariate SETARX model (6.9) satisfies the following:

1. The functions \( g_1^J(\eta), g_2^J(f), \) and \( g_3(f) \) for each \( J \in \{1, \ldots, L\}^D \) are nonperiodic and bounded on compact sets, and \( g_2^J(f) = O(\|f\|^\gamma_1) \) as \( \|f\| \to \infty \) for some real \( \gamma_1 \).

2. Assumption B.4 holds, the sup \( t \geq 0 E[\|\eta_{t+1}\|^{\max(1, \gamma_1+\gamma_2)}|\mathcal{F}_t] < \infty \) for some \( \gamma_2 > 0 \).

3. There exist \( \mathcal{A}^{(J)} = [\mathcal{A}_1^{(J)}, \mathcal{A}_2^{(J)}, \ldots, \mathcal{A}_p^{(J)}] \) and \( \mathcal{B} = [\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{q-1}] \), each of which may be the zero matrix, for each \( J \in \{1, \ldots, L\}^D \), where \( \mathcal{A}_i^{(J)} \) and \( \mathcal{B}_i \) are matrices of dimension \( D \times D \) and \( \kappa \times \kappa \) respectively such that \( g_1^J(\eta) = \eta(\mathcal{A}^{(J)})^T + o(\|\eta\|) \) and \( g_3(f) = f(\mathcal{B})^T + o(\|\eta\|) \) as \( \|\eta\| \) and \( \|f\| \to \infty \). Then the \( Dp \)-dimensional square matrix \( \mathfrak{A} \) defined by \( \mathfrak{A} \) if \( \mathcal{A}^{(J)} = 0 \) and by

\[
\mathfrak{A} = \begin{bmatrix}
O_D & O_D & \cdots & O_D & (\mathcal{A}_1^{(J)})^T \\
I_D & O_D & \cdots & O_D & (\mathcal{A}_2^{(J)})^T \\
& O_D & I_D & \cdots & (\mathcal{A}_3^{(J)})^T \\
& & \ddots & \ddots & \ddots \\
& & & O_D & I_D & \cdots & (\mathcal{A}_p^{(J)})^T
\end{bmatrix}
\]

otherwise, and the \( \kappa q \)-dimensional square matrix \( \mathfrak{B} \) be defined by

\[
\mathfrak{B} = \begin{bmatrix}
O_\kappa & O_\kappa & \cdots & O_\kappa & (\mathcal{B}_1)^T \\
I_\kappa & O_\kappa & \cdots & O_\kappa & (\mathcal{B}_2)^T \\
& O_\kappa & I_\kappa & \cdots & (\mathcal{B}_3)^T \\
& & \ddots & \ddots & \ddots \\
& & & O_\kappa & I_\kappa & \cdots & (\mathcal{B}_q)^T
\end{bmatrix}
\]

satisfy \( \rho(\mathfrak{A}) < 1 \) and \( \rho(\mathfrak{B}) < 1 \), where \( \rho \) denotes the spectral radius, \( O_\kappa \) denotes the \( \kappa \)-dimensional zero square matrix and \( I_\kappa \) denotes the \( \kappa \)-dimensional unit square matrix.

Lemma 6.1. Under Assumption B.5, \( \{f_t, y_t\} \) of the multivariate SETARX model (6.7)-(6.8) represented as a nonlinear ARX model (6.9) is \( \alpha \)-mixing with mixing coefficient \( \alpha(k) \sim e^{-\beta k} \) for some \( \beta > 0 \).

Proof. The result is known as in Lemma 3.1 in [99] and thus we do not provide the proof since it is roughly same. We refer the interested reader to remarks after Assumption 3.3 and Lemma 3.1 in [99] and the references therein. \( \square \)
Remark 6.2. Lemma 6.1 provides sufficient conditions for the multivariate SETARX process (6.9) to be stationary ([99, 109, 137]). The proof of this Lemma as in Lemma 3.1 in [99] implies geometric ergodicity and stronger conclusion of absolute regularity with an exponentially decreasing rate ([109, 137, 143, 144]).

Lemma 6.3. Let \( a_0^{(J)} = 0 \) for each \( J \) in model (6.3) and \( p = 1 \). Assume that there is a \( D \)-cycle of indexes \( j_1 \to j_2 \to j_3 \to \cdots \to j_D \to j_1 \) with the notation \( A_1^{(j_s)} \) corresponding to \( A_1^{(j_s)} \mod \text{the } D \)-cycle so that \( A_1^{(j_{s+1})} = A_1^{(j_1)} \). The process \( \{y_t\} \) of the multivariate SETAR model (6.3) is geometrically ergodic if

\[
\varrho \left( \prod_{s=1}^{D} -A_1^{(j_s)} \right) < 1
\]

where \( \varrho \) denotes the spectral radius and the product notation \( \prod_{s=m}^{n} A^{(j_s)} = A^{(j_m)} \cdots A^{(j_{m+1})} A^{(j_m)} \) is interpreted as the identity matrix if \( n = m - 1 \).

Proof. The result about geometric ergodicity follows from Theorem 4.5 and equation model (4.12) in [137] with \( A_i = -A_1^{(j_s)} \) and \( k = D \). □

6.4 Estimation of model parameters

In this section, we assume that assumption B.5 and Lemma 6.3 are satisfied. We also assume the model orders \( p, q, d, \) and \( L \) of model (6.4)-(6.5)-(6.6) are known. Let model (6.4) be represented as a MSETAR \( (L,p_J; J \in \{1, \cdots, L\}^D) \) model (6.3) with exogenous variables or factors as in model (6.5)-(6.6). We propose to use estimation procedures based on the standard LSE approach and the concept of self-tuning regulators used in the study of adaptive control of stochastic linear systems (see [86]). [13] has shown that algorithms for estimation of parameters based on the stochastic gradient principles for linear systems are also suitable for nonlinear systems. Alternatively, following [30], one can use local linear fitting plus the projection method to estimate components \( g_1^{(J)}(\cdot) \) and \( g_2^{(J)}(\cdot) \) of model equation (6.9). The function \( g_3(\cdot) \) can then be estimated directly using a standard approach or by kernel-type estimation ([98]).

6.4.1 Standard LSE Algorithm for Parameter Estimation

Consider the MSETARX \( (L,p_J,q; J \in \{1, \cdots, L\}^D) \) model in equation (6.6):

\[
y_t = \sum_{J \in \{1, \cdots, L\}^D} y_{t,J}^{(J)} \left( \hat{\Theta}^{(J)} \right)^T \hat{\Phi}_{t-1} + \omega_t \tag{6.10}
\]
where \( \left( \hat{\Theta}^{(J)} \right)^T = -[a_0^{(J)}, A_1^{(J)}, \ldots, A_p^{(J)}, \Lambda^{(J)} \Xi_1, \Lambda^{(J)} \Xi_2, \ldots, \Lambda^{(J)} \Xi_q], \)
\( \hat{\Phi}^T_t = [1, y_t^T, y_{t-1}^T, \ldots, y_{t-p+1}^T, f_t^T, f_{t-1}^T, \ldots, f_{t-q+1}^T] \) with the autoregressive orders \( p, q, \) delay \( d, \) and thresholds known. Then the LSE is \( \hat{\Theta}^{(J)} = \sum_{J \in \{1, \ldots, L\}^D} y_t^{(J)} \left( \hat{\Phi}^T_{t-1} \hat{\Phi}_{t-1} \right)^{-1} \hat{\Phi}^T_{t-1} y_t. \) Following [86] presentation of the stochastic gradient algorithm for ARX systems, the true parameter \( \hat{\Theta}^{(J)} \) can also be estimated by the LSE using the recursion,

\[
\hat{\Theta}^{(J)}_{k+1} = \hat{\Theta}^{(J)}_k + y_{k+1,d} R_k^{-1} \hat{\Phi}_k (y_{k+1} - \hat{\Phi}^T_k \hat{\Theta}^{(J)}_k) \quad (6.11)
\]

\[
R_k = \sum_{J \in \{1, \ldots, L\}^D} \sum_{i=0}^k y_{k+1,d} \hat{\Phi}_i \hat{\Phi}^T_i \quad (6.12)
\]

### 6.4.2 Algorithm for Adaptive Parameter Estimation

Let \( 0 < \alpha \leq 1, 0 < v^{(J)} \leq 1, \) \( p^* = \max\{p, d, q\}, \) and \( \hat{\Theta} \) be the coefficients of the MSETARX \( (L, p, q; J \in \{1, \ldots, L\}^D) \) model in equation (6.6).

\[
\hat{\Theta}^{(J)}_k = 0; \quad k \leq p^*
\]

\[
\hat{\Theta}^{(J)}_{k+1} = \hat{\Theta}^{(J)}_k + y_{k+1,d} \frac{\alpha \hat{\Phi}_k}{s_k^{(J)}} (y_{k+1} - \hat{\Phi}^T_k \hat{\Theta}^{(J)}_k); \quad k \geq p^*
\]

\[
r_k^{(J)} = \begin{cases} 
1; & k < p^* \\
\frac{r_{k-1}^{(J)}}{r_{k-1}^{(J)} + \sum_{J \in \{1, \ldots, L\}^D} y_{k+1,d} \hat{\Phi}_k \hat{\Phi}^T_k}; & k \geq p^*. 
\end{cases} \quad (6.13)
\]

\[
s_k^{(J)} = \begin{cases} 
1; & k < p^* \\
s_{k-1}^{(J)} + y_{k+1} \left[ \max\{v^{(J)} r_{k-1}^{(J)}, 1\} + \|y_k\|^2 - s_{k-1}^{(J)} \right]; & k \geq p^*. 
\end{cases} \quad (6.14)
\]

This algorithm 6.4.2 corresponds to the adaptive parameter estimation algorithm proposed by [13], with the control sequence being \( (s_k^{(J)})^{-1} \) instead of \( (r_k^{(J)})^{-1}. \) The simulation results presented by the authors showed that as the control sequence \( (r_k^{(J)})^{-1} \) becomes large, a further progress towards the true coefficients is prevented or slowed down since this control sequence which weight the prediction error decrease too fast. The relaxed control sequence \( (s_k^{(J)})^{-1} \) have similar properties as \( (r_k^{(J)})^{-1} \) with the convergence spend decreased by the factors \( v^{(J)} \) and in particular, improves the estimation accuracy ([13]). This algorithm was applied in [88] for the analysis of biomedical signals.

### 6.4.3 Simulations

In this section, we carry out a simulation exercise to study the performance of the parameter estimation algorithm presented in Section 6.4.1&6.4.2 on MSETARX models. In
this respect, we consider two data generating process (DGP) according to the following:

1. Consider a simulated 50,000 points of a two-dimensional MSETARX process with six-regimes, delay \(d = 6\), \(\Lambda^{(J)} = 0\) for all \(J \in \{1, \cdots, L\}^D\) in equation (6.5), standard normal noise \(N(0, 1)\) added to all regimes and autoregressive order \(p = 3\) defined by:

   \[
   \text{Regime 1 } R_1^r := [-\infty, -0.50) \times [-\infty, 0), \quad \varphi_1^{(1)} = \begin{pmatrix} -0.02 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}, \quad \varphi_2^{(1)} = \begin{pmatrix} 0.53 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}, \\
   \varphi_3^{(1)} = \begin{pmatrix} 0.00 & 0.53 \\ 0.00 & 0.30 \end{pmatrix}, \quad a_0^{(1)} = \begin{pmatrix} 0.74 \\ -0.20 \end{pmatrix}
   \]

   \[
   \text{Regime 2 } R_2^r := [-\infty, -0.50) \times [0, \infty), \quad \varphi_1^{(2)} = \begin{pmatrix} -0.02 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}, \quad \varphi_2^{(2)} = \begin{pmatrix} 0.53 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}, \\
   \varphi_3^{(2)} = \begin{pmatrix} 0.00 & 0.53 \\ 0.00 & 0.30 \end{pmatrix}, \quad a_0^{(2)} = \begin{pmatrix} -0.75 \\ -0.20 \end{pmatrix}
   \]

   \[
   \text{Regime 3 } R_3^r := [-0.50, 0.50) \times [-\infty, 0), \quad \varphi_1^{(3)} = \begin{pmatrix} -0.94 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}, \quad \varphi_2^{(3)} = \begin{pmatrix} 0.85 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}, \\
   \varphi_3^{(3)} = \begin{pmatrix} 0.00 & 0.85 \\ 0.00 & 0.30 \end{pmatrix}, \quad a_0^{(3)} = \begin{pmatrix} 1.15 \\ -0.20 \end{pmatrix}
   \]

   \[
   \text{Regime 4 } R_4^r := [-0.50, 0.50) \times (0.00, \infty), \quad \varphi_1^{(4)} = \begin{pmatrix} -0.94 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}, \quad \varphi_2^{(4)} = \begin{pmatrix} 0.00 & 0.30 \\ 0.20 \end{pmatrix}, \\
   \varphi_3^{(4)} = \begin{pmatrix} 0.00 & 0.85 \\ 0.00 & 0.30 \end{pmatrix}, \quad a_0^{(4)} = \begin{pmatrix} 0.74 \\ 0.00 \end{pmatrix}
   \]

   \[
   \text{Regime 5 } R_5^r := [0.50, \infty) \times [-\infty, 0.00), \quad \varphi_1^{(5)} = \begin{pmatrix} -1.10 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}, \quad \varphi_2^{(5)} = \begin{pmatrix} -0.30 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}, \\
   \varphi_3^{(5)} = \begin{pmatrix} 0.00 & -0.30 \\ 0.00 & 0.30 \end{pmatrix}, \quad a_0^{(5)} = \begin{pmatrix} -0.75 \\ 0.20 \end{pmatrix}
   \]

   \[
   \text{Regime 6 } R_6^r := [0.50, \infty) \times [0.00, \infty), \quad \varphi_1^{(6)} = \begin{pmatrix} -1.10 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}, \quad \varphi_2^{(6)} = \begin{pmatrix} 0.30 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}, \\
   \varphi_3^{(6)} = \begin{pmatrix} 0.00 & 0.30 \\ 0.00 & 0.30 \end{pmatrix}, \quad a_0^{(6)} = \begin{pmatrix} 1.15 \\ 0.20 \end{pmatrix}
   \]

   and a signal section is shown in Figure 6.1. In Appendix C.1, autoregressive coefficient estimates obtained via the LSE algorithm is provided.

2. Consider a three-regime \((L = 3)\) bivariate \((D = 2)\) MSETARX \((L, p_j, q; J \in \{1, \cdots, L\}^D)\) model with a bivariate exogenous input \((\kappa = 2)\), model orders be
Chapter 6. Multivariate Self-Exciting Threshold Autoregressive Models with eXogenous Input

unit \( (p = \max\{p_J|J \in \{1, \cdots, L\}^D\} = 1, q = 1) \) and delay \( d = 1 \):

\[
y_t = \begin{cases} 
\mathcal{A}_1^{(1)} y_{t-1} + \Lambda^{(1)} \Xi_1 f_{t-1} + \omega_t; & (y_{t-1})_2 \leq -0.5 \\
\mathcal{A}_1^{(2)} y_{t-1} + \Lambda^{(2)} \Xi_1 f_{t-1} + \omega_t; & (y_{t-1})_2 \in (-0.5, 0.5] \\
\mathcal{A}_1^{(3)} y_{t-1} + \Lambda^{(3)} \Xi_1 f_{t-1} + \omega_t; & (y_{t-1})_2 \geq 0.5 
\end{cases}
\]  

\( (6.15) \)

where \( \mathcal{A}_1^{(1)} = \begin{pmatrix} -0.3 & 0.6 \\ -0.7 & 0.4 \end{pmatrix}, \mathcal{A}_1^{(2)} = \begin{pmatrix} 1.5 & -1 \\ 0.2 & 0.3 \end{pmatrix}, \mathcal{A}_1^{(3)} = \begin{pmatrix} 0.3 & -0.1 \\ 0.2 & 0.6 \end{pmatrix}, \Xi_1 = \begin{pmatrix} 0.5 & 0 \\ 0.3 & 0 \end{pmatrix}, \Lambda^{(1)} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0 \end{pmatrix}, \Lambda^{(2)} = \begin{pmatrix} 0.3 & 0 \\ 0 & 0.2 \end{pmatrix}, \Lambda^{(3)} = \begin{pmatrix} 0.8 & 0 \\ 0 & 0 \end{pmatrix} \). It is worth noting that the multivariate process \( y_t \) is unstable in the inner regime and only the second subprocess determines the current regime.

\[\text{Figure 6.1: Two-dimensional MSETARX process with six-regimes, delay } d = 6, \text{ autoregressive order } p = 3 \text{ and } \Lambda^{(J)} = \mathbf{0} \text{ for all } J \in \{1, \cdots, L\}^D \text{ in equation (6.5). This is a signal section of 500 time samples of the multivariate process.}\]

6.5 Concluding remarks

The recent financial crisis of 2007-2009 has lead to a need for regulators and policy makers to understand and track systemic linkages. As the events following the turmoil in financial markets unfolded, it became evident that modern financial systems exhibit a high degree of interdependence and nonlinearity making it difficult in predicting the consequences of such an intertwined system. In this study, we define a nonlinear
multivariate SETARX model useful in modeling economic relationships and to capture non-linear dynamics such as regime switching and asymmetric responses to shocks. We then present an estimation procedure for the parameters.

In general, testing linearity is the first step of a proper modelling strategy of nonlinear models as it is possible that a linear model could adequately capture the relationship considered. Nonlinear models are usually not identified when the underlying process is linear ([8, 63, 73, 134, 141]). The proposed test statistic for detecting threshold nonlinearity in vectors time series and the procedure for building multivariate threshold models discussed in [141] could be performed on each subprocess in the MSETARX model setting. In this case, [13] suggests a reasonable choice of the delay to be \(d^* = \arg\max \{\sum_i \mathcal{C}(i)(d) \mid d \in \{1, \cdots, d_{\text{max}}\}\} \) where \(\mathcal{C}(i)(d)\) is the value of the test statistic ([141]) for each subprocess \(i\). One could apply the Wald test procedure used in [18], which is a generalisation of [61] approach, to test linearity. Another possibility of testing linear VAR model against a MSETARX model would be to generalise the approach the approach of [124] to multivariate models.

After the parameter estimation of model (6.5), it is necessary to evaluate the model by appropriate misspecification tests before putting it into practice. The general purpose is to find out if the assumptions made in the estimation step appear satisfied ([62, 64, 124, 141]). For more details about modelling strategies and issues of vector threshold autoregressive models, we refer interested readers to [65, 141]. This model could be very useful in studying huge data sets such as the analysis of high-frequency financial data.

Many problems remain open for the multivariate SETARX models. For example, establishing a testing procedure in determining the number of regimes and the specification of the threshold space will required a careful investigation.

**Acknowledgement**

This research is supported by the Erasmus Mundus Fellowship. We are grateful to Lutz Leistritz for his support.
Appendix for Chapter 2

A.1 Wavelet Analysis

A.1.1 Notation and Operators

Consider the mapping

\[ \Gamma : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R}) \]

Let \( f \in L^2(\mathbb{R}) \), \( \alpha, \beta \in \mathbb{R} \) and \( s \in \mathbb{R}^+ \) where \( \mathbb{R}^+ := \{ t \in \mathbb{R} : t > 0 \} \). Unless otherwise stated, the complex conjugate of \( z \in \mathbb{C} \) is denoted \( \bar{z} \) and the magnitude of \( z \) is denoted \( |z| \). The symbol \( i \) will represent the square root of \(-1\), i.e., \( i^2 = -1 \). We present in Table A.1 some notations and operators that will be often referred to in this manuscript.

<table>
<thead>
<tr>
<th>Operator, ( \Gamma )</th>
<th>Notation, ( \Gamma f )</th>
<th>Output</th>
<th>Inverse, ( \Gamma * f = \Gamma^{-1}f )</th>
<th>Fourier transform, ( (\Gamma f)\hat{\hphantom{1}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dilation</td>
<td>( (Dsf)(t) )</td>
<td>( \frac{1}{\sqrt{s}}f\left(\frac{t}{s}\right) )</td>
<td>( D_{s^{-1}}f )</td>
<td>( D_{s^{-1}}f )</td>
</tr>
<tr>
<td>Involution</td>
<td>( \hat{f} )</td>
<td>( \hat{f}(-t) )</td>
<td>( \hat{f} )</td>
<td>( \hat{f} )</td>
</tr>
<tr>
<td>Translation</td>
<td>( (\tau_{\alpha}f)(t) )</td>
<td>( f(t - \alpha) )</td>
<td>( \tau_{-\alpha}f )</td>
<td>( e^{-\alpha}\hat{f} )</td>
</tr>
<tr>
<td>Modulation</td>
<td>( (e_{\alpha}f)(t) )</td>
<td>( e^{i2\pi\alpha t}f(t) )</td>
<td>( e_{-\alpha}f )</td>
<td>( \tau_{\alpha}\hat{f} )</td>
</tr>
<tr>
<td>Reflection</td>
<td>( (Rf)(t) )</td>
<td>( f(-t) )</td>
<td>( Rf )</td>
<td>( R\hat{f} )</td>
</tr>
</tbody>
</table>

Table A.1: Notations and Operators

A.1.2 Continuous Wavelet Transform (CWT)

The continuous wavelet transform (CWT) differs from the more traditional short time Fourier transform (STFT) by allowing arbitrarily high localization in time of high frequency signal features. The CWT permits for the isolation of the high frequency features
Appendix A. *Appendix for Chapter 2* 98

due to its variable window width related to the scale of observation. In particular, the CWT is not limited to using sinusoidal analyzing functions but allows for a large selection of localized waveforms that can be employed as long as they satisfy predefined mathematical criteria (described below).

Let $\mathcal{H}$ be a Hilbert space, the CWT may be described as a mapping parameterized by a function $\psi$

$$C_\psi : \mathcal{H} \rightarrow C_\psi(\mathcal{H}).$$ (A.1)

The CWT of a one-dimensional function $f \in L^2(\mathbb{R})$ is given by

$$C_\psi : L^2(\mathbb{R}) \rightarrow C_\psi(L^2(\mathbb{R}))$$
$$f \mapsto \langle f, \tau_t \Delta_s \psi \rangle _{L^2(\mathbb{R})}$$ (A.2)

where $\tau_t \Delta_s \psi$ is a dilated (by $s$) and translated (by $t$) version of $\psi$ given as

$$(\tau_t \Delta_s \psi)(\xi) = \frac{1}{|s|^\frac{1}{2}} \psi\left(\frac{\xi - t}{s}\right)$$ (A.3)

Thus, the CWT of one-dimensional signal $f$ is a two-dimensional function of the real variables time $t$, and scale $s \neq 0$. For a given $\psi$, the CWT may be thought of in terms of the representation of a signal with respect to the wavelet family generated by $\psi$, that is, all its translated and dilated versions. The CWT may be written as

$$(C_\psi f)(t,s) := \langle f, \tau_t \Delta_s \psi \rangle$$ (A.4)

For each point $(t,s)$ in the time-scale plane, the wavelet transform assigns a (complex) numerical value to a signal $f$ which describes how much $f$ like a translated by $t$ and scaled by $s$ version of $\psi$.

The CWT of a signal $f$ is defined as

$$(C_\psi f)(t,s) = \frac{1}{|s|^\frac{1}{2}} \int_\mathbb{R} f(\xi) \bar{\psi}(\frac{\xi - t}{s}) d\xi$$ (A.5)

where $\bar{\psi}(\xi)$ is the complex conjugate of the analyzing wavelet function $\psi(\xi)$. Given that $\psi$ is chosen with enough time-frequency localization\(^1\), the CWT gives a picture of the time-frequency characteristics of the function $f$ over the whole time-scale plane $\mathbb{R} \times (\mathbb{R} \setminus \{0\})$. When $C_{adm,\psi} < \infty$, it is possible to find the inverse continuous

\(^{1}\)The time-frequency concentrated functions, denoted $TF(\mathbb{R})$, is a space of complex-valued finite energy functions defined on the real line that decay faster than $\frac{1}{t}$ simultaneously in the time and frequency domains. This is defined explicitly as $TF(\mathbb{R}) := \{ \varphi \in L^2(\mathbb{R}) : |\varphi(t)| < \eta(1 + |t|)^{(1+\varepsilon)} \text{ and } |\hat{\varphi}(\gamma)| < \eta(1 + |\gamma|)^{-(1+\varepsilon)} \text{ for } \eta < \infty, \varepsilon > 0 \}$
transformation via the relation known as *Calderón’s reproducing identity*,

\[ f(\xi) = \frac{1}{C_{adm,\psi}} \int_{\mathbb{R}^2} \langle f, \tau_1 D_s \psi \rangle \tau_1 D_s \psi(\xi) \frac{1}{s^2} ds dt. \] (A.6)

and if \( s \) restricted in \( \mathbb{R}^+ \), then the *Calderón’s reproducing identity* takes the form

\[ f(\xi) = \frac{1}{C_{adm+,\psi}} \int_{-\infty}^{\infty} \int_{0}^{\infty} \langle f, \tau_1 D_s \psi \rangle \tau_1 D_s \psi(\xi) \frac{1}{s^2} ds dt. \] (A.7)

Let \( \alpha \) and \( \beta \) be arbitrary real numbers and \( f, f_1, \) and \( f_2 \) be arbitrary functions in \( \mathcal{L}^2(\mathbb{R}) \).

The CWT, \( C_\psi \), with respect to \( \psi \) satisfies the following conditions:

1. **Linearity**
   - \( (C_{\psi}(\alpha f_1 + \beta f_2))(t, s) = \alpha (C_{\psi} f_1)(t, s) + \beta (C_{\psi} f_2)(t, s) \)

2. **Time Invariance**
   - \( (C_{\psi}(\tau_\beta f))(t, s) = (C_{\psi} f)(t - \beta, s) \)

3. **Dilation**
   - \( (C_{\psi}(D_\alpha f))(t, s) = (C_{\psi} f)(\alpha t, \alpha^{-1} s) \)

4. **Negative Scales**
   - \( C_{\psi} f(t, -s) = (C_{\psi} R f)(-t, s) \)

The time invariance property of the CWT implies that the wavelet transform of a time-delayed version of a signal is a time-delayed version of its wavelet transform. This serves as an important property in terms of pattern recognition. This nice property is not readily obtained in the case of Discrete wavelet transforms (2, 147).

The contribution to the signal energy at the specific scale \( s \) and location \( t \) is given by

\[ \mathcal{E}(t, s) = |C_{\psi}|^2 \] (A.8)

which is a two-dimensional wavelet energy density function known as the scalogram.

The wavelet transform \( C_\psi \) corresponding to a complex wavelet is also complex valued. The transform can be separated into two categories:

- Real part \( \mathcal{R}\{C_\psi\} \) and Imaginary part \( \mathcal{I}\{C_\psi\} \)
- Modulus (or Amplitude), \( |C_\psi| \) and phase (or phase-angle), \( \Phi(t, s) \),
which can be obtained using the relation:

\[ C_\psi = |C_\psi|e^{i\phi(t,s)} \quad \text{and} \quad \phi(t,s) = \arctan\left( \frac{\text{I}\{C_\psi\}}{\text{R}\{C_\psi\}} \right). \]  
(A.9)

### A.1.3 Maximal Overlap Discrete Wavelet Transform

The Maximal Overlap Discrete Wavelet Transform (MODWT), also related to notions of 'cycle spinning' and 'wavelet frames', is with the basic idea of downsampling values removed from discrete wavelet transform. The MODWT unlike the conventional discrete wavelet transform (DWT), is non-orthogonal and highly redundant, and is defined naturally for all sample sizes, \( N \) (147). Given an integer \( J \) such that \( 2^J < N \), where \( N \) is the number of data points, the original time series represented by the vector \( X(n) \), where \( n = 1, 2, \cdots, N \), can be decomposed on a hierarchy of time scales by details, \( D_j(n) \), and a smooth part, \( S_j(n) \), that shift along with \( X \):

\[ X(n) = S_J(n) + \sum_{j=1}^{J} D_j(n) \]  
(A.10)

with \( S_j(n) \) generated by the recursive relationship

\[ S_{j-1}(n) = S_j(n) + D_j(n). \]  
(A.11)

The MODWT details \( D_j(n) \) represent changes on a scale of \( \tau = 2^{j-1} \), while the \( S_j(n) \) represents the smooth or approximation wavelet averages on a scale of \( \tau_J = 2^{J-1} \). 49 employed this wavelet transform to investigate the issue of moderation of volatility in G-7 economies and also to detect the importance of the various explanations of the moderation.

### A.2 DVV Plots of simulated Processes

We provide the structure of the DVV analysis on some simulated processes such as: a Threshold autoregressive process (TAR), linear autoregressive integrated moving average (ARIMA) signal, a Generalized autoregressive conditional heteroskedastic process (GARCH), and a Bilinear process.
Appendix A. Appendix for Chapter 2

(a) DVV analysis on ARIMA(1,1,1) signal  
(b) DVV analysis on TAR(1) signal

**Figure A.1:** DVV analysis on ARIMA and Threshold Autoregressive signals

(a) DVV analysis on GARCH(1,1) signal  
(b) DVV analysis on Bilinear signal

**Figure A.2:** DVV analysis on GARCH and Bilinear signals
Appendix B

Appendix for Chapter 5

B.1 Functional Central Limit Theorem (Recall)

Suppose \( u_t \sim i.i.d(0, \sigma^2), \) \( t = 1, 2, \cdots, T \) and

\[
X_T(r) = \frac{1}{r} \sum_{t=1}^{\lfloor Tr \rfloor} u_t,
\]

where \( X_T(r) \) is a variable constructed from the sample mean of the first \( r^{th} \) fraction of random variables \( u_t, \) \( r \in [0, 1] \) and where \( \lfloor Tr \rfloor \) denotes the largest integer that is less than or equal to \( Tr. \) Thus \( X_T(r) \) is a step function in \( r \) defined as follows:

\[
X_T(r) = \begin{cases} 
0 & \text{if } 0 \leq r < 1/T \\
\frac{u_1}{T} & \text{if } 1/T \leq r < 2/T \\
\frac{(u_1 + u_2)}{T} & \text{if } 2/T \leq r < 3/T \\
\cdots & \cdots \\
\frac{(u_1 + u_2 + \cdots + u_T)}{T} & \text{if } r = 1
\end{cases}
\]

Then the sequence of stochastic functions \( \left\{ \frac{\sqrt{T}X_T(\cdot)}{\sigma} \right\}_{T=1}^\infty \) converges in distribution to a standard Brownian motion \( W(\cdot), \frac{\sqrt{T}X_T(\cdot)}{\sigma} \xrightarrow{d} W(\cdot). \)
B.2 Proof of Proposition 1

Proof. We want to prove (a), (b), (c), (d), (e) of proposition 1. First let us introduce the AR\((p+1)\) process:

\[
y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \cdots + \phi_p y_{t-p} + \phi_{p+1} y_{t-p-1} + u_t \quad \text{where } u_t \sim i.i.d(0, \sigma^2).
\]

This process can be written in the form:

\[
y_t = \rho y_{t-1} + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta y_{t-2} + \cdots + \alpha_p \Delta y_{t-p} + u_t,
\]

where: \( \rho = \phi_1 + \phi_2 + \cdots + \phi_{p+1} \), \( \alpha_i = - (\phi_{i+1} + \phi_{i+2} + \cdots + \phi_{i+p+1}) \).

Suppose now that \((y_t)\) contains a single unit root, that is one of the roots is 1 in the equation:

\[
1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_{p+1} z^{p+1} = 0,
\]

thus

\[
1 - \phi_1 - \cdots - \phi_{p+1} = 0 \quad \text{and} \quad \phi_1 + \phi_2 + \cdots + \phi_{p+1} = 1 \quad \text{implies that } \rho = 1.
\]

Under the null hypothesis of a unit root we have:

\[
\Delta y_t = \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta y_{t-2} + \cdots + \alpha_p \Delta y_{t-p} + u_t,
\]

thus

\[
(1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p) \Delta y_t = u_t. \tag{B.1}
\]

a. Suppose that the process \((y_t)\) has only one unit root and all other roots are outside the unit circle, then \(\Delta y_t\) is stationary. From equation (B.1) we have:

\[
\Delta y_t = (1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p)^{-1} u_t = \Psi(L) \cdot u_t,
\]

where \( \Psi(L) = (1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p)^{-1} \).

Next, define \( \varepsilon_t = \Delta y_t \forall t \), it means that \( \varepsilon_t = \Psi(L) \cdot u_t \), \( u_t \sim i.i.d(0, \sigma^2) \), and

\[
y_t = y_0 + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t. \tag{B.2}
\]

The Beveridge-Nelson decomposition implies that:

\[
y_t = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t + y_0 = \Psi(1) \left[ u_1 + u_2 + \cdots + u_t \right] + \eta_t - \eta_0 + y_0, \tag{B.3}
\]
where \( \Psi(1) = (1 - \alpha_1 - \alpha_2 - \cdots - \alpha_p)^{-1} \) and \((\eta_t)\) is a stationary process with zero mean of the form:

\[
\eta_t = \sum_{j=0}^{\infty} \lambda_j u_{t-j} \quad \text{where} \quad \sum_{j=0}^{\infty} |\lambda_j| < \infty.
\]

We construct a variable \( Z_T(r) \) from the sample mean of the first \( r \)th fraction of random variables \( \varepsilon_t \), \( r \in [0, 1] \):

\[
Z_T(r) = \frac{1}{T} \sum_{t=1}^{[Tr]} \varepsilon_t,
\]

(B.4)

\[
\sqrt{T} Z_T(r) = \frac{1}{T} \sum_{t=1}^{[Tr]} \varepsilon_t = \frac{1}{T} \left( \Psi(1) \sum_{t=1}^{[Tr]} u_t + \eta_{[Tr]} - \eta_0 \right),
\]

(B.5)

From the functional Central Limit theorem and continuous mapping theorem we have:

\[
\sqrt{T} X_T(\cdot) \to_d \sigma W(\cdot) \quad \Rightarrow \quad \Psi(1) \sqrt{T} X_T(r) \to_d \Psi(1) \cdot \sigma W(r), \quad \text{with} \quad (\eta_t) \quad \text{a zero mean stationary process. Thus we have:} \quad \frac{1}{\sqrt{T}} \left( \eta_{[Tr]} - \eta_0 \right) \to_p 0 \quad \text{(see [60], example 17.2). It follows that:}
\]

\[
\sqrt{T} Z_T(\cdot) \to_d \Psi(1) \cdot \sigma W(\cdot).
\]

The stochastic function \( Z_T(r) \) in (B.4) can be expressed in the form of a step function in \( r \) as follows:

\[
Z_T(r) = \begin{cases} 
0 & \text{if } 0 \leq r < 1/T \\
\varepsilon_1/T & \text{if } 1/T \leq r < 2/T \\
(\varepsilon_1 + \varepsilon_2)/T & \text{if } 2/T \leq r < 3/T \\
\vdots & \\
(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_T)/T & \text{if } r = 1
\end{cases}
\]

(B.6)

Thus

\[
\sqrt{T} (Z_T(r) + \frac{\eta_0}{T}) = \begin{cases} 
\frac{\eta_0}{\sqrt{T}} & \text{if } 0 \leq r < 1/T \\
\frac{\eta_1}{\sqrt{T}} & \text{if } 1/T \leq r < 2/T \\
\frac{\eta_2}{\sqrt{T}} & \text{if } 2/T \leq r < 3/T \\
\vdots & \\
\frac{\eta_T}{\sqrt{T}} & \text{if } r = 1
\end{cases}
\]
The stochastic function \( F_T(r) = \left[ \sqrt{T}(Z_T(r) + \frac{y_0}{T}) \right]^n \) is then expressed in the form:

\[
F_T(r) = \begin{cases} 
\frac{y_0^2}{T^2} & \text{if } 0 \leq r < 1/T \\
\frac{y_1^2}{T^2} & \text{if } 1/T \leq r < 2/T \\
\frac{y_2^2}{T^2} & \text{if } 2/T \leq r < 3/T \\
\vdots & \vdots \\
\frac{y_T^2}{T^2} & \text{if } r = 1 
\end{cases} \tag{B.7}
\]

and

\[
\int_0^1 F_T(r) dr = \frac{y_0^2}{T^2} \frac{1}{T} + \frac{y_1^2}{T^2} \frac{1}{T} + \frac{y_2^2}{T^2} \frac{1}{T} + \cdots + \frac{y_{T-1}^2}{T^2} \frac{1}{T} = \frac{1}{T^{(n/2)+1}} \sum_{t=1}^T y_{t-1}^n.
\]

From (B.5) we have:

\[
F_T(\cdot) = \left[ \sqrt{T}(Z_T(\cdot) + \frac{y_0}{T}) \right]^n \rightarrow_d [\Psi(1)]^n \sigma^n [W(\cdot)]^n, \tag{B.8}
\]

thus

\[
\int_0^1 F_T(r) dr \rightarrow_d [\Psi(1)]^n \sigma^n \int_0^1 [W(r)]^n dr = \frac{1}{T^{(n/2)+1}} \sum_{t=1}^T y_{t-1}^n \rightarrow_d [\Psi(1)]^n \sigma^n \int_0^1 [W(r)]^n dr.
\]

b. Using the Beveridge-Nelson decomposition in (B.3):

\[
y_t = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t + y_0 = \Psi(1) \left[ u_1 + u_2 + \cdots + u_t \right] + \eta_t - \eta_0 + y_0.
\]

\[
\Rightarrow \Delta y_t = y_t - y_{t-1} = \varepsilon_t = \Psi(1) u_t + \eta_t - \eta_{t-1}, \tag{B.9}
\]

\[
\frac{1}{T^{(n+1)/2}} \sum_{t=1}^T y_{t-1}^n \varepsilon_t = \frac{1}{T^{(n+1)/2}} \sum_{t=1}^T y_{t-1}^n \left( \Psi(1) u_t + \eta_t - \eta_{t-1} \right),
\]

\[
= \frac{1}{T^{(n+1)/2}} \Psi(1) \sum_{t=1}^T y_{t-1}^n u_t + \frac{1}{T^{(n+1)/2}} \sum_{t=1}^T y_{t-1}^n \left( \eta_t - \eta_{t-1} \right),
\]

\[
\Rightarrow \frac{1}{T^{(n+1)/2}} \Psi(1) \sum_{t=1}^T y_{t-1}^n u_t = \frac{1}{T^{(n+1)/2}} \sum_{t=1}^T y_{t-1}^n \varepsilon_t - \frac{1}{T^{(n+1)/2}} \sum_{t=1}^T y_{t-1}^n \left( \eta_t - \eta_{t-1} \right). \tag{B.10}
\]

By denoting \( F_T(r) = F_{T[T_r]} \), from the function \( F_T(r) \) in (B.7) we have: \( F_{T[T_r]} = \frac{y_{T[T_r]}^2}{T^{n/2}} \).

By denoting \( Z_T(r) = Z_{T[T_r]} \), from the function \( F_T(r) \) in (B.6) we have

\[
Z_{T[T_r]} = \frac{\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{T[T_r]}}{T} \Rightarrow \Delta Z_{T[T_r]} = Z_{T[T_r]} - Z_{T[T_r]-1} = \frac{\varepsilon_{T[T_r]}}{T}. \tag{B.11}
\]
Appendix B. Appendix for Chapter 5

Now let us consider the first element of the RHS of (B.10):

$$
\frac{1}{T(n+1)^2} \sum_{t=1}^{T} y_{t-1}^{\sigma_{T}} = \mathbb{E}[Z_{T}(\cdot) + \eta_{T}]^{n} \rightarrow_{d} [\Psi(1)]^{n} \sigma_{n}[W(\cdot)]^{n}
$$

From (B.6) and (B.8), and by convergence to stochastic integrals we have:

$$
\int_{0}^{1} F_{T}(r) \cdot d\sqrt{T}Z_{T}(r) \rightarrow \int_{0}^{1} [\Psi(1)]^{n} \sigma_{n}[W(\cdot)]^{n} d\Psi(1) = [\Psi(1)]^{n+1} \sigma^{n+1} \int_{0}^{1} [W(\cdot)]^{n} dW(r),
$$

Next, consider the second element on RHS of (B.10):

$$
\frac{1}{T(n+1)^2} \sum_{t=1}^{T} y_{t-1}(\eta_{t} - \eta_{t-1}) = \mathbb{E}[Z_{T}(\cdot) + \eta_{T}]^{n} \rightarrow_{d} [\Psi(1)]^{n+1} \sigma^{n+1} \int_{0}^{1} [W(\cdot)]^{n} dW(r).
$$

Where $\eta_{t}$ is a stationary process with zero mean of the form: $\eta_{t} = \sum_{j=0}^{\infty} \lambda_{j} u_{t-j}$ where $\sum_{j=0}^{\infty} |\lambda_{j}| < \infty$. Applying the central limit theorem for stationary process we have:

$$
\left[\sqrt{T}\left(\frac{1}{T} \sum_{t=1}^{T} \eta_{t}\right) - \sqrt{T}\left(\frac{1}{T} \sum_{t=1}^{T} \eta_{t-1}\right)\right] \rightarrow_{d} 0.
$$

It follows that:

$$
\frac{1}{T(n+1)^2} \sum_{t=1}^{T} y_{t-1}(\eta_{t} - \eta_{t-1}) \rightarrow 0.
$$
Appendix B. Appendix for Chapter 5

From (B.10), (B.12), and (B.14) we obtain:

\[
\frac{1}{T^{(n+1)/2}} \Psi(1) \sum_{t=1}^{T} y^R_{t-1} u_t \xrightarrow{d} [\Psi(1)]^{n+1} \sigma^{n+1} \int_0^T [W(r)]^n dW(r).
\]

\[
\Rightarrow \quad \frac{1}{T^{(n+1)/2}} \sum_{t=1}^{T} y^R_{t-1} u_t \xrightarrow{d} [\Psi(1)]^{n} \sigma^{n+1} \int_0^T [W(r)]^n dW(r).
\]

c. From (B.9) : \( \Delta y_t = y_t - y_{t-1} = \varepsilon_t = \Psi(1)u_t + \eta_t - \eta_{t-1} \) where \( \eta_t \) is a zero mean stationary process, \( u_t \sim i.i.d(0, \sigma^2) \)

\[
\Rightarrow \quad E(\Delta y_t) = E[\Psi(1)u_t + \eta_t - \eta_{t-1}] = \Psi(1)E(u_t) + E(\eta_t) - E(\eta_{t-1}) = 0.
\]

\[
\Rightarrow \quad \frac{1}{T} \sum_{t=1}^{T} \Delta y_{t-i} : \Delta y_{t-j} = \frac{1}{T} \sum_{t=1}^{T} [\Delta y_{t-i} - E(\Delta y_{t-i})][\Delta y_{t-j} - E(\Delta y_{t-j})].
\]

\( y_t \) has a single unit root \( \Rightarrow \Delta y_t \) is stationary with: \( E(\Delta y_t) = 0, \text{Var}(\Delta y_t) \) is finite and \( Cov(\Delta y_t, \Delta y_{t-s}) = \gamma_s \) only depends on \( s \). By the law of large numbers:

\[
\frac{1}{T} \sum_{t=1}^{T} [\Delta y_{t-i} - E(\Delta y_{t-i})][\Delta y_{t-j} - E(\Delta y_{t-j})] \xrightarrow{p} Cov(\Delta y_{t-i}, \Delta y_{t-j}) = \gamma_{i-j}.
\]

\[
\Rightarrow \quad \frac{1}{T} \sum_{t=1}^{T} \Delta y_{t-i} : \Delta y_{t-j} \xrightarrow{p} \gamma_{i-j}.
\]

d.

\[
\frac{1}{T^{n/2+1}} \sum_{t=1}^{T} y^R_{t-1} \cdot \Delta y_{t-j} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y^R_{t-1} \frac{\Delta y_{t-j}}{\sqrt{T}} = \frac{1}{\sqrt{T}} \sum_{[T_r]=0}^{T-1} F_T(r) \cdot \frac{\Delta y_{[T_r]+1-j}}{\sqrt{T}},
\]

where we have \( F_T(r) \rightarrow d [\Psi(1)]^{n} \sigma [W(r)]^n \).

\[
\Rightarrow \quad \frac{1}{T^{n/2+1}} \sum_{t=1}^{T} y^R_{t-1} \cdot \Delta y_{t-j} \xrightarrow{d} [\Psi(1)]^{n} \sigma [W(r)]^n \frac{1}{T} \sum_{[T_r]=0}^{T-1} \Delta y_{[T_r]+1-j}.
\]

\[
\Rightarrow \quad \frac{1}{T^{n/2+1}} \sum_{t=1}^{T} y^R_{t-1} \cdot \Delta y_{t-j} \xrightarrow{d} [\Psi(1)]^{n} \sigma [W(r)]^n \frac{1}{T} \sum_{t=1}^{T} \Delta y_{t-j},
\]

and \( \Delta y_t \) is a zero mean stationary process. By applying the law of large number for a covariance stationary process it follows that: \( \frac{1}{T} \sum_{t=1}^{T} \Delta y_{t-j} \xrightarrow{L} 0 \)

\[
\Rightarrow \quad \frac{1}{T^{n/2+1}} \sum_{t=1}^{T} y^R_{t-1} \cdot \Delta y_{t-j} \xrightarrow{0} \Rightarrow \quad \frac{1}{T^{n/2+1}} \sum_{t=1}^{T} y^R_{t-1} \cdot \Delta y_{t-j} \xrightarrow{p} 0.
\]
Appendix B. Appendix for Chapter 5

108

e. Consider the sequence: \( \Delta y_{t-i} \cdot u_t \forall i \geq 1 \)

Notice that \( u_t \) is uncorrelated to \( \Delta y_{t-i} \) i.e. \( E(\Delta y_{t-i} \cdot u_t) = E(\Delta y_{t-i}) \cdot E(u_t) = 0 \),
then

the conditional mean is:

\[
E(\Delta y_{t-i} / \Delta y_{t-i-1} \cdot u_t) = E(\Delta y_{t-i} / \Delta y_{t-i-1} \cdot u_t) \cdot E(u_t / \Delta y_{t-i-1} \cdot u_t) = 0.
\]

Hence \( \Delta y_{t-i} \) is a martingale difference sequence.

\[
Var(\Delta y_{t-i} \cdot u_t) = E[(\Delta y_{t-i} \cdot u_t)^2] = E[(\Delta y_{t-i})^2 \cdot u_t^2] = \sigma^2 E[(\Delta y_{t-i})^2],
\]

since \( \Delta y_t \) is a zero mean stationary process

\[
\Rightarrow E[(\Delta y_{t-i})^2] = Var(\Delta y_{t-i}) = Var(\Delta y_t) = \gamma_0.
\]

\[
\Rightarrow Var(\Delta y_{t-i} \cdot u_t) = \sigma^2 \gamma_0.
\]

Thus, the martingale difference sequence \( \Delta y_{t-i} \cdot u_t \) satisfies the usual central limit theorem:

\[
\sqrt{T} \left( \frac{1}{T} \sum_{t=1}^{T} \Delta y_{t-i} u_t \right) \xrightarrow{d} N(0, \sigma^2 \gamma_0).
\]

\[
\Rightarrow \frac{1}{T^{1/2}} \sum_{t=1}^{T} \Delta y_{t-i} u_t \xrightarrow{d} N(0, \sigma^2 \gamma_0).
\]

\[\square\]

B.3 Proof of Theorem 1

Proof. Let \( \tilde{D}_T \) be a \( 3 \times 3 \) diagonal matrix of the form:

\[
\tilde{D}_T = \begin{pmatrix}
T^2 & 0 & 0 \\
0 & T^3 & 0 \\
0 & 0 & T^4
\end{pmatrix},
\]

\[
R = \begin{pmatrix}
1 & 0 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & 0 & \cdots & 0 \\
0 & 0 & 1 & 0 & \cdots & 0
\end{pmatrix}_{3 \times (p+3)}
\]

\[
\beta = (\delta_{1.2}, \delta_{1.4}, \delta_{1.6}, \rho_1, \rho_2, \cdots, \rho_p)'_{(p+3) \times 1}
\]

and \( X_t = (y_{t-1}^3, y_{t-1}^5, y_{t-1}^7, \Delta y_{t-1}, \cdots, \Delta y_{t-p})'_{(p+3) \times 1} \) with \( \hat{\beta} \) being the OLS estimator for the parameter \( \beta \), and \( \sigma_T^2 \) the variance of \( \Delta y_t \).
Next, consider the scaling matrix:

\[
D_T = \begin{pmatrix}
T^2 & 0 & 0 & 0 & 0 & \ldots & 0 \\
0 & T^3 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & T^4 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & T^{1/2} & 0 & \ldots & 0 \\
0 & 0 & 0 & 0 & T^{1/2} & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & 0 & 0 & 0 & 0 & \ldots & T^{1/2}
\end{pmatrix}_{(p+3) \times (p+3)}
\]

where \( \hat{D}_T \) and \( D_T \) are diagonal matrices and \( \hat{D}_T R = RD_T \). Therefore, we obtain:

\[
\hat{F}_{NL} = (\hat{\beta} - \beta)'R' \hat{D}_T \cdot \hat{D}_T^{-1} \left[ \hat{\sigma}_T^2 R \left( \sum_{t=1}^{T} X_t X_t' \right)^{-1} R' \hat{D}_T \right]^{-1} \hat{D}_T R (\hat{\beta} - \beta)
\]

since \( \hat{D}_T \) is symmetric : \( \hat{D}_T = \hat{D}_T' \)

\[
\Rightarrow \quad \hat{F}_{NL} = (\hat{\beta} - \beta)' (\hat{D}_T R)' \left[ \hat{\sigma}_T^2 \hat{D}_T R \left( \sum_{t=1}^{T} X_t X_t' \right)^{-1} (\hat{D}_T R)' \right]^{-1} \hat{D}_T R (\hat{\beta} - \beta).
\]

and since \( \hat{D}_T R = RD_T \) we have:

\[
\hat{F}_{NL} = (\hat{\beta} - \beta)' (RD_T)' \left[ \hat{\sigma}_T^2 RD_T \left( \sum_{t=1}^{T} X_t X_t' \right)^{-1} (RD_T)' \right]^{-1} RD_T (\hat{\beta} - \beta).
\]

\[
\Rightarrow \quad \hat{F}_{NL} = [(RD_T)(\hat{\beta} - \beta)]' \left[ \hat{\sigma}_T^2 RD_T \left( \sum_{t=1}^{T} X_t X_t' \right)^{-1} D_T R' \right]^{-1} RD_T (\hat{\beta} - \beta)
\]

\[
\Rightarrow \quad \hat{F}_{NL} = [(RD_T)(\hat{\beta} - \beta)]' \left\{ \hat{\sigma}_T^2 R \left[ D_T^{-1} \left( \sum_{t=1}^{T} X_t X_t' \right) D_T^{-1} \right]^{-1} R' \right\}^{-1} RD_T (\hat{\beta} - \beta)
\]

with

\[
D_T^{-1} = \begin{pmatrix}
T^{-2} & 0 & 0 & 0 & 0 & \ldots & 0 \\
0 & T^{-3} & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & T^{-4} & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & T^{-1/2} & 0 & \ldots & 0 \\
0 & 0 & 0 & 0 & T^{-1/2} & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & 0 & 0 & 0 & 0 & \ldots & T^{-1/2}
\end{pmatrix}_{(p+3) \times (p+3)}
\]
where \( X_t = (y_{t-1}^3, y_{t-1}^5, y_{t-1}^7, \ldots, \Delta y_{t-p})'_{(p+3)\times 1} \). It follows that: \( D_T^{-1}(\sum_{t=1}^{T} X_t X_t') D_T^{-1} = \)

\[
\begin{pmatrix}
\sum_{t=1}^{T} y_{t-1}^3 & \sum_{t=1}^{T} y_{t-1}^5 & \sum_{t=1}^{T} y_{t-1}^7 & \cdots & \sum_{t=1}^{T} \Delta y_{t-p}
\end{pmatrix}_{(p+3)\times 1}.
\]

Under the null hypothesis, \((y_t)_t\) is a unit root process that satisfies the condition of proposition 1. Hence applying the proposition, it follows that: \( D_T^{-1}(\sum_{t=1}^{T} X_t X_t') D_T^{-1} \overset{d}{\to} \)

\[
\begin{pmatrix}
\Psi_6 \sigma^6 \int_0^T W(r)^6 dr & \Psi_8 \sigma^8 \int_0^T W(r)^8 dr & \Psi_{10} \sigma^{10} \int_0^T W(r)^{10} dr & 0 & 0 & 0 & \cdots & 0 \\
\Psi_6 \sigma^6 \int_0^T W(r)^8 dr & \Psi_8 \sigma^8 \int_0^T W(r)^{10} dr & \Psi_{10} \sigma^{10} \int_0^T W(r)^{12} dr & 0 & 0 & 0 & \cdots & 0 \\
\Psi_{10} \sigma^{10} \int_0^T W(r)^{10} dr & \Psi_{12} \sigma^{12} \int_0^T W(r)^{12} dr & \Psi_{14} \sigma^{14} \int_0^T W(r)^{14} dr & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & 0
\end{pmatrix}
\]

\[
\Rightarrow D_T^{-1}(\sum_{t=1}^{T} X_t X_t') D_T^{-1} \overset{d}{\to} \begin{pmatrix} Q_{3 \times 3}^{-1} & 0 \\ 0 & V_{p \times p}^{-1} \end{pmatrix},
\]

with \(Q_{3 \times 3} = \)

\[
\begin{pmatrix}
\Psi_6 \sigma^6 \int_0^T W(r)^6 dr & \Psi_8 \sigma^8 \int_0^T W(r)^8 dr & \Psi_{10} \sigma^{10} \int_0^T W(r)^{10} dr \\
\Psi_8 \sigma^8 \int_0^T W(r)^8 dr & \Psi_{10} \sigma^{10} \int_0^T W(r)^{10} dr & \Psi_{12} \sigma^{12} \int_0^T W(r)^{12} dr \\
\Psi_{10} \sigma^{10} \int_0^T W(r)^{10} dr & \Psi_{12} \sigma^{12} \int_0^T W(r)^{12} dr & \Psi_{14} \sigma^{14} \int_0^T W(r)^{14} dr
\end{pmatrix},
\]

and \(V_{p \times p} = \)

\[
\begin{pmatrix}
\gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} \\
\gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\
\gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{p-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_{0}
\end{pmatrix}.
\]

\[
\Rightarrow R \left( D_T^{-1}(\sum_{t=1}^{T} X_t X_t') D_T^{-1} \right)^{-1} R' \overset{d}{\to} (I_3)_{3 \times 3} (Q_{3 \times 3}^{-1} 0) (I_3)_{3 \times 3}^{-1} (V_{p \times p})_{p \times 3}^{-1}.
\]

\[
\Rightarrow R \left( D_T^{-1}(\sum_{t=1}^{T} X_t X_t') D_T^{-1} \right)^{-1} R' \overset{d}{\to} Q_{3 \times 3}^{-1}.
\]

\[
\Rightarrow \left[ R \left( D_T^{-1}(\sum_{t=1}^{T} X_t X_t') D_T^{-1} \right)^{-1} R \right]^{-1} \overset{d}{\to} Q_{3 \times 3}.
\]

we have: \(\sum_{t=1}^{T} X_t u_t = \sum_{t=1}^{T} \Delta y_{t-1} u_t \). \(\Rightarrow D_T^{-1}(\sum_{t=1}^{T} X_t u_t) = \frac{1}{T} \sum_{t=1}^{T} \Delta y_{t-1} u_t \).
We have showed that the inner product $\sum_{t=1}^{T} X_t X_t'$ of the regressor matrix including the additional regressors is asymptotically block diagonal. Applying the proposition 1, under the null hypothesis of a unit root, it follows that:

$$D_T^{-1} \sum_{t=1}^{T} X_t u_t \xrightarrow{d} \begin{pmatrix} 0 & 0 & N(0, \sigma^2 \gamma_0) \\ 0 & 0 & N(0, \sigma^2 \gamma_0) \\ \vdots & \vdots & \cdots \end{pmatrix}_{(p+3)\times 1} = \begin{pmatrix} A_{3\times 1} \\ B_{p\times 1} \end{pmatrix},$$

where $A_{3\times 1} = \begin{pmatrix} \Psi^3 \sigma^4 f_0^1 W^3 dW \\ \Psi^5 \sigma^6 f_0^1 W^5 dW \\ \Psi^7 \sigma^8 f_0^1 W^7 dW \end{pmatrix}$ and $B_{p\times 1} = \begin{pmatrix} N(0, \sigma^2 \gamma_0) \\ \cdots \\ N(0, \sigma^2 \gamma_0) \end{pmatrix}$.

Using $D_T(\hat{\beta} - \beta) = \left[D_T^{-1}(\sum_{t=1}^{T} X_t X_t') D_T^{-1}\right]^{-1} D_T^{-1}(\sum_{t=1}^{T} X_t u_t)$ and from B.16 and B.17 we have:

$$D_T(\hat{\beta} - \beta) \xrightarrow{d} \begin{pmatrix} 0 & 0 & N(0, \sigma^2 \gamma_0) \\ 0 & 0 & N(0, \sigma^2 \gamma_0) \\ \vdots & \vdots & \cdots \end{pmatrix}_{(p+3)\times 1}.$$

Using $\left[R\left(D_T^{-1}(\sum_{t=1}^{T} X_t X_t') D_T^{-1}\right)^{-1} R'\right]^{-1} d \xrightarrow{} Q_{3\times 3}$ and B.18 we have:

$$\tilde{F}_{NL} = \left[(RD_T)(\hat{\beta} - \beta)\right]' \left[\sigma^2 R\left(D_T^{-1}(\sum_{t=1}^{T} X_t X_t') D_T^{-1}\right)^{-1} R'\right]^{-1} RD_T(\hat{\beta} - \beta).$$

$$\Rightarrow \tilde{F}_{NL} \xrightarrow{d} \frac{1}{\sigma^2} \left\{ (Q_{3\times 3}^{-1} A_{3\times 1})' Q_{3\times 3} \cdot Q_{3\times 3}^{-1} A_{3\times 1} \right\}.$$ (since $Q_{3\times 3}$ is symmetric).

It follows that: $\tilde{F}_{NL} \xrightarrow{d} v' Q^{-1} v$ with,

$$v = \begin{bmatrix} \frac{1}{4} W(1)^4 - \frac{3}{2} f_0^1 W(r)^2 dr \\ \frac{1}{6} W(1)^6 - \frac{2}{3} f_0^1 W(r)^4 dr \\ \frac{1}{8} W(1)^8 - \frac{7}{4} f_0^1 W(r)^6 dr \end{bmatrix}.$$
and

\[ Q = \begin{bmatrix} f_0^1 W(r)^6 dr & f_0^1 W(r)^8 dr & f_0^1 W(r)^{10} dr \\ f_0^1 W(r)^8 dr & f_0^1 W(r)^{10} dr & f_0^1 W(r)^{12} dr \\ f_0^1 W(r)^{10} dr & f_0^1 W(r)^{12} dr & f_0^1 W(r)^{14} dr \end{bmatrix}. \]

\[ \Delta \in \mathbb{N} > 0 \]

### B.4 Proof of Corollary

**Proof.** We consider the asymptotic behavior of the least squares estimator \( \hat{\beta} = (\hat{\delta}_1^2, \hat{\delta}_1^4, \hat{\delta}_1^6)' \) under the null \( \Delta y_t = u_t \) and the OLS estimator \( \hat{\beta} \) can be written as:

\[ \hat{\beta}_T = \left( \begin{array}{c} \sum_{t=1}^{T} y_{t-1}^6 \\ \sum_{t=1}^{T} y_{t-1}^8 \\ \sum_{t=1}^{T} y_{t-1}^{10} \end{array} \right)^{-1} \left( \begin{array}{c} \sum_{t=1}^{T} y_{t-1}^3 u_t \\ \sum_{t=1}^{T} y_{t-1}^5 u_t \\ \sum_{t=1}^{T} y_{t-1}^7 u_t \end{array} \right). \]

The following results are needed:

\[ \frac{1}{T^4} \sum_{t=1}^{T} y_{t-1}^6 = \frac{1}{T} \sum_{t=1}^{T} \left( \frac{1}{\sqrt{T}} y_{t-1} \right)^6 = \frac{1}{T} \sum_{t=1}^{T} \int_{k-1}^{k} \left( \frac{1}{\sqrt{T}} Y_T(r) dr \right)^6 \rightarrow_d \sigma^6 \int_0^1 W(r)^6 dr, \]

where \( k-1 \leq t < k \), \( Y_T(r) = \frac{S_{k-1}}{\sqrt{T}} = \frac{\sum_{i=1}^{k-1} u_i}{\sqrt{T}} \),

similarly, we have:

\[ \frac{1}{T^6} \sum_{t=1}^{T} y_{t-1}^8 \rightarrow_d \sigma^8 \int_0^1 W(r)^8 dr, \]

\[ \frac{1}{T^6} \sum_{t=1}^{T} y_{t-1}^{10} \rightarrow_d \sigma^{10} \int_0^1 W(r)^{10} dr, \]

\[ \frac{1}{T^7} \sum_{t=1}^{T} y_{t-1}^{12} \rightarrow_d \sigma^{12} \int_0^1 W(r)^{12} dr, \]

\[ \frac{1}{T^8} \sum_{t=1}^{T} y_{t-1}^{14} \rightarrow_d \sigma^{14} \int_0^1 W(r)^{14} dr. \]

Now by using the continuous mapping theorem, Itô’s formula, and the weak convergence of stochastic integrals it is possible to obtain a general result for \( i \in \mathbb{N} > 0 \)

\[ \frac{1}{T^{i+1}} \sum_{t=1}^{T} y_{t-1}^i u_t \rightarrow_d \int_0^1 W^i(r) dW(r) = \sigma^{i+1} \left\{ \frac{1}{i+1} \int_0^1 W(1)^{(i+1)} - \frac{1}{2} \int_0^1 W(r)^{i-1} dr \right\}. \]
Appendix B. Appendix for Chapter 5

and we obtain:

\[
\frac{1}{T^2} \sum_{t=1}^{T} y_{t-1}^3 u_t \rightarrow_d \int_0^1 W^3(r) dW(r) = \sigma_1^4 \left\{ \frac{1}{4} \int_0^1 W(1)^4 - \frac{3}{2} \int_0^1 W(r)^2 dr \right\},
\]

\[
\frac{1}{T^3} \sum_{t=1}^{T} y_{t-1}^5 u_t \rightarrow_d \int_0^1 W^5(r) dW(r) = \sigma_1^6 \left\{ \frac{1}{6} \int_0^1 W(1)^6 - \frac{5}{2} \int_0^1 W(r)^4 dr \right\},
\]

\[
\frac{1}{T^4} \sum_{t=1}^{T} y_{t-1}^7 u_t \rightarrow_d \int_0^1 W^7(r) dW(r) = \sigma_1^8 \left\{ \frac{1}{8} \int_0^1 W(1)^8 - \frac{7}{2} \int_0^1 W(r)^6 dr \right\}.
\]

It is straightforward that the estimators have different convergence rates. Thus, the least squares estimators need to be scaled using the following scaling matrix: \( D_T = diag(T^2, T^3, T^4) \). Denote \( \hat{\beta}_T = (\hat{\beta}_T^1, \hat{\beta}_T^2, \hat{\beta}_T^3)' \) and \( X_T = (y_{t-1}^2, y_{t-1}^5, y_{t-1}^7)' \). Then, we have that:

\[
D_T(\hat{\beta}_T - \beta) = [D_T^{-1}(\sum_{t=1}^{T} X_t X_t')D_T^{-1}]^{-1}[D_T^{-1}(\sum_{t=1}^{T} X_t u_t)].
\]

After some algebra one gets that

\[
D_T(\hat{\beta}_T - \beta) \rightarrow_L \frac{1}{\sigma_1^2} (\Gamma Q \Gamma)^{-1} (\Gamma v),
\]

where

\[
v = \begin{bmatrix}
\frac{1}{4} W(1)^4 - \frac{3}{2} \int_0^1 W(r)^2 dr \\
\frac{1}{6} W(1)^6 - \frac{5}{2} \int_0^1 W(r)^4 dr \\
\frac{1}{8} W(1)^8 - \frac{7}{2} \int_0^1 W(r)^6 dr
\end{bmatrix}
\]

and

\[
Q = \begin{bmatrix}
\int_0^1 W(r)^6 dr & \int_0^1 W(r)^8 dr & \int_0^1 W(r)^{10} dr \\
\int_0^1 W(r)^8 dr & \int_0^1 W(r)^{10} dr & \int_0^1 W(r)^{12} dr \\
\int_0^1 W(r)^{10} dr & \int_0^1 W(r)^{12} dr & \int_0^1 W(r)^{14} dr
\end{bmatrix}
\]

and \( \Gamma = diag(1, \sigma_1^2, \sigma_4^2) \). Our test statistic has then the following representation:

\[
\hat{F}_{NL} = (\hat{\beta}_T - \beta)' (RD_T)' [\hat{\sigma}_T^2 D_T R (\sum_{t=1}^{T} X_t X_t')^{-1} D_T R']^{-1} RD_T (\hat{\beta}_T - \beta),
\]

with \( R = I_3 \), and has the limiting distribution:

\[
\hat{F}_{NL} \rightarrow_L ((\Gamma Q \Gamma)^{-1} (\Gamma v))' ((\Gamma Q \Gamma)^{-1})^{-1} ((\Gamma Q \Gamma)^{-1} (\Gamma v)) = v'Q^{-1}v.
\]

By the law of large numbers it is easy to show that under the null as \( T \rightarrow \infty \)

\[
\hat{\sigma}_T^2 = \frac{1}{T-4} \sum_{t=1}^{T} (\Delta y_t - \delta_1^3 y_{t-1}^3 - \delta_1^4 y_{t-1}^5 - \delta_1^6 y_{t-1}^7)^2 \xrightarrow{P} \sigma_T^2.
\]
Under the alternative $\Delta y_t$ and $y_{i-1}^i, \forall i \in \mathbb{N} > 0$ are $I(0)$ and thus it is readily seen that

$$\frac{1}{T} \sum_{t=1}^{T} \Delta y_t = O_P(1),$$

$$\frac{1}{T} \sum_{t=1}^{T} y_{i-1}^i = O_P(1),$$

are bounded in probability. Furthermore the innovation process $(u_t)_t$ is by assumption $I(0)$ and thus

$$\frac{1}{T} \sum_{t=1}^{T} u_t = O_P(1)$$

as well. For the OLS estimate $\hat{\beta}$ we have

$$(O_P(T))^{-1} O_P(T^2) = (T O_P(1))^{-1} T^2 O_P(1) = \frac{1}{T} T^2 O_P(1) = T O_P(1) = O_P(T).$$

Hence, the $\tilde{F}_{NL}$ statistic diverges to infinity at the rate $O_P(D_T)$. 

$\square$
Appendix C

Appendix for Chapter 6

C.1 Estimation Results

We provide below the estimation of parameters obtained via LSE algorithm in Section 6.4.1 on the first simulated process in Section 6.4.3. The regime time corresponds to the number of temporal samples, where the multivariate process stayed in each regime.

Regime 1 \( R_1 := (-\infty, -0.50) \times (-\infty, 0) \), \( \hat{\varphi}(1) = \begin{pmatrix} -0.0278 & -0.0169 \\ 0.0027 & 0.2812 \end{pmatrix} \), \( \hat{\varphi}(2) = \begin{pmatrix} 0.5275 & 0.0025 \\ 0.0013 & 0.3073 \end{pmatrix} \), \( \hat{a}(1) = \begin{pmatrix} 0.7399 \\ -0.2012 \end{pmatrix} \), (regime time: 10927).

Regime 2 \( R_2 := (-\infty, -0.50) \times [0, \infty) \), \( \hat{\varphi}(2) = \begin{pmatrix} -0.0156 & 0.0009 \\ 0.0033 & 0.2935 \end{pmatrix} \), \( \hat{\varphi}(3) = \begin{pmatrix} 0.5317 & -0.0051 \\ 0.0043 & 0.3102 \end{pmatrix} \), \( \hat{a}(2) = \begin{pmatrix} -0.7404 \\ -0.1951 \end{pmatrix} \), (regime time: 8770).

Regime 3 \( R_3 := [-0.50, 0.50) \times (-\infty, 0) \), \( \hat{\varphi}(3) = \begin{pmatrix} -0.9417 & -0.0008 \\ 0.0143 & 0.2859 \end{pmatrix} \), \( \hat{\varphi}(4) = \begin{pmatrix} 0.8602 & 0.0040 \\ 0.0211 & 0.3003 \end{pmatrix} \), \( \hat{a}(3) = \begin{pmatrix} 1.1337 \\ -0.2408 \end{pmatrix} \), (regime time: 3932).

Regime 4 \( R_4 := [-0.50, 0.50) \times (0.00, \infty) \), \( \hat{\varphi}(4) = \begin{pmatrix} -0.9302 & -0.0210 \\ -0.0023 & 0.3142 \end{pmatrix} \), \( \hat{\varphi}(5) = \begin{pmatrix} 0.8631 & 0.0033 \\ -0.0135 & 0.3116 \end{pmatrix} \), \( \hat{a}(4) = \begin{pmatrix} 0.0066 & 0.8497 \\ 0.0079 & 0.2789 \end{pmatrix} \), \( \hat{a}(5) = \begin{pmatrix} 0.7101 \\ 0.1960 \end{pmatrix} \), (regime time: 3235).
**Appendix C. Appendix for Chapter 6**

**Regime 5**

$R_5^i := [0.50, \infty) \times [-\infty, 0.00)$, $\hat{A}^{(5)}_1 = \begin{pmatrix} -1.1008 & -0.0056 \\ 0.0032 & 0.2923 \end{pmatrix}$, $\hat{A}^{(5)}_2 = \begin{pmatrix} -0.2918 & -0.0012 \\ 0.0019 & 0.3106 \end{pmatrix}$, $\hat{A}^{(5)}_3 = \begin{pmatrix} 0.0066 & -0.2971 \\ 0.0029 & 0.2958 \end{pmatrix}$, $\hat{a}^{(5)}_0 = \begin{pmatrix} -0.7595 \\ 0.1927 \end{pmatrix}$, (regime time: 9697).

**Regime 6**

$R_6^i := [0.50, \infty) \times [0.00, \infty)$, $\hat{A}^{(6)}_1 = \begin{pmatrix} -1.0995 & 0.0087 \\ 0.0013 & 0.3147 \end{pmatrix}$, $\hat{A}^{(6)}_2 = \begin{pmatrix} 0.3011 & -0.0150 \\ 0.0029 & 0.2904 \end{pmatrix}$, $\hat{A}^{(6)}_3 = \begin{pmatrix} -0.0004 & 0.3033 \\ 0.0026 & 0.2996 \end{pmatrix}$, $\hat{a}^{(6)}_0 = \begin{pmatrix} 1.1508 \\ 0.1942 \end{pmatrix}$, (regime time: 13433).
Bibliography


[61] Hansen, B. E. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64 (1996), 413–430.


