Université Paris I - Panthéon Sorbonne
Università Ca'Foscari of Venice

Doctoral Dissertation

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)

Université Paris I - Panthéon Sorbonne
Università Ca'Foscari of Venice

Doctoral Thesis

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 établie 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'indentification, de datation des points de retournement et de lévaluation 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.

En outre, une nouvelle procédure pour tester la linéarité et la racine unitaire dans un cadre non-linéaire est proposé par l'introduction d'un nouveau modèle – le modèle MT-STAR – qui a des propriétés similaires au modèle ESTAR mais réduit les effets des problèmes d'identification et peut aussi représenter l'asymétrie dans le mécanisme d'ajustement vers l'équilibre. Les distributions asymptotiques du test de racine unitaire proposées sont non-standards et sont calculées. La puissance du test est évaluée par simulation et quelques illustrations empiriques sur les taux de change réel montrent son efficacité. Enfin, la thèse développe des modèles multi-variés Self-Exciting Threshold Autoregressive avec des variables exogènes (MSETARX) et présente une méthode d'estimation paramétrique. La modélisation des modèles MSETARX et des problèmes engendrés par son estimation sont brièvement examinés.

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 Bonnisseau, 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

List of Figures

List of Tables

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 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 policy makers 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:

- **Addo, P. M.**, Billio, M., Guégan, D. (2013). *"Nonlinear Dynamics and Recurrence Plots for Detecting Financial Crisis"*, The North-American Journal of Economics and Finance, Volume 26, December 2013, Pages 416–435.

  http://dx.doi.org/10.1016/j.najef.2013.02.014.

- **Addo P.M.**, M. Billio, D. Guégan (2014), *"Turning point chronology for the Euro–Zone: A Distance Plot Approach"*. Journal of Business Cycle Measurement & Analysis (*forthcoming*).

- **Addo, P. M.**, Billio, M., D. Guégan (2014). *"The Univariate MT-STAR Model and a new linearity and unit root test procedure"*. Computational Statistics & Data Analysis (CSDA). http://dx.doi.org/10.1016/j.csda.2013.12.009.

- **Addo, P. M.**, Billio, M., D. Guégan (2014). *"Nonlinear Dynamics and Wavelets for Business Cycle Analysis"*. In Marco Gallegati and Willi Semmler (Eds.), "Wavelets Applications in Economics and Finance: Essays in Honor of James Ramsey", Dynamic Modeling and Econometrics in Economics and Finance. Springer Series (*forthcoming*).

- **Addo, P. M.**, Billio, M., Guégan, D. (2012), *"Understanding exchange rate dynamics"*, In A. Colubi, K. Fokianos, & E. J. Kontoghiorghes (Eds.), Proceedings of the 20th International Conference on Computational Statistics, pp. 1–14. Curran Associates, Inc. ISBN: 978-1-62748-321-6 .

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 *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 were 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 determination 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 resolution 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, ψ, 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. $\psi \in \mathcal{L}^1(\mathbb{R}) \cap \mathcal{L}^2(\mathbb{R})$, to ensure that the Fourier transform of ψ is well-defined and ψ 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 ψ 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 ψ is that wavelet family generated by ψ. 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 ψ 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 ψ.
- Knowing all the inner products $\{\langle f, \tau_t D_s \psi \rangle\}$, the signal should be recoverable.

The wavelet ψ is assumed to satisfy the admissibility condition,

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

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

$$C_{adm_+,\psi} = \int_0^{\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < \infty. \tag{2.2}$$

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

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

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 ψ is $\hat{\psi}(\zeta) = e^{-\pi^2 f_b(\zeta - f_c)^2}$, which is well-defined since $\psi \in \mathcal{L}^1(\mathbb{R})$. It can easily be shown that the Morlet wavelet (2.3) is a modulated Gaussian function and involutive, i.e. $\psi = \tilde{\psi}$. 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

$$\psi(\xi) = \frac{1}{\sqrt{\pi}} e^{i\pi\xi - \xi^2}, \tag{2.4}$$

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).

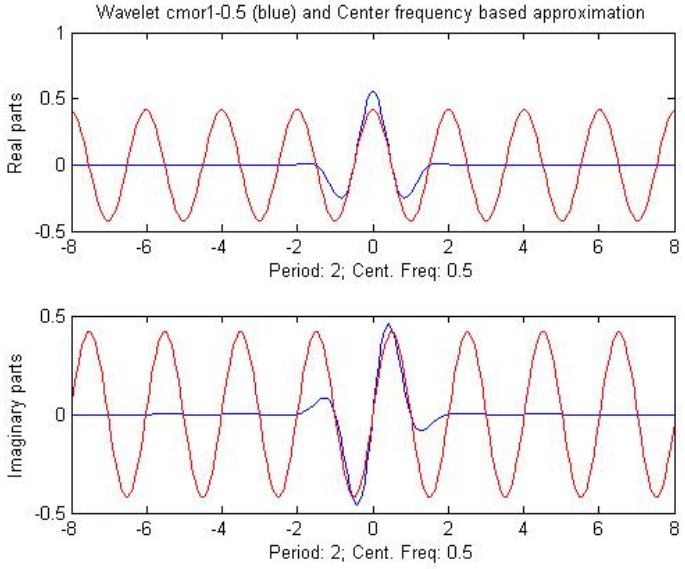


FIGURE 2.1: Complex Morlet wavelet with $f_b = 1$ and $f_c = 0.5$

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 α 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 (DVs) 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

$$R_{ent}(m, \tau) = I(m, \tau) + \frac{m \ln N}{N}, \tag{2.5}$$

where N is the number of delay vectors, which is kept constant for all values of m and τ under consideration,

$$I(m, \tau) = \frac{H(x, m, \tau)}{\langle H(x_{s,i}, m, \tau) \rangle_i} \tag{2.6}$$

where x is the signal, $x_{s,i}$ $i = 1, \cdots, T_s$ 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

$$H(x) = \sum_{j=1}^{T} \ln(T\rho_j) + \ln 2 + C_E \tag{2.7}$$

where T is the number of samples in the data set, ρ_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*, σ^{*2}, which is an inverse measure of the predictability of a time series. The algorithm is summarized below:

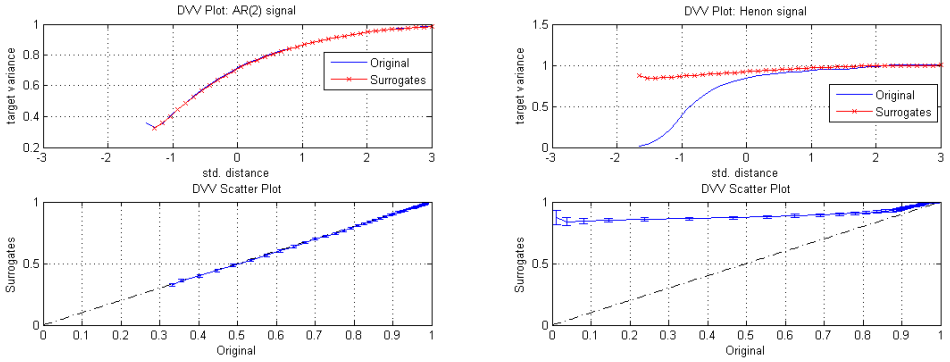
- For an optimal embedding dimension m and time lag τ, generate delay vector (DV): $x(k) = [x_{k-\tau}, \cdots, x_{k-m\tau}]$ and corresponding target x_k
- The mean μ_d and standard deviation, σ_d, are computed over all pairwise distances between DVs, $\|x(i) - x(j)\|$ for $i \neq j$.
- The sets Ω_k 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 ϱ_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 Ω_k, the variance of the corresponding targets, σ_k^2, is computed. The average over all sets Ω_k, normalized by the variance of the time series, σ_x^2, yields the *target variance*, σ^{*2} :

$$\sigma^{*2}(\varrho_d) = \frac{\frac{1}{N}\sum_{k=1}^{N} \sigma_k^2(\varrho_d)}{\sigma_x^2} \tag{2.8}$$

  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, ϱ_d. The minimum *target variance*, $\sigma^{*2}_{min} = \min_{\varrho_d}[\sigma^{*2}(\varrho_d)]$,

[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]



(a) AR(2) signal (b) Henon signal

FIGURE 2.2: 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.

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, σ^{*2}_{min} is inversely related to prevalence of the deterministic component over the stochastic one, lowest σ^{*2}_{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 WiAAFT 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, τ.

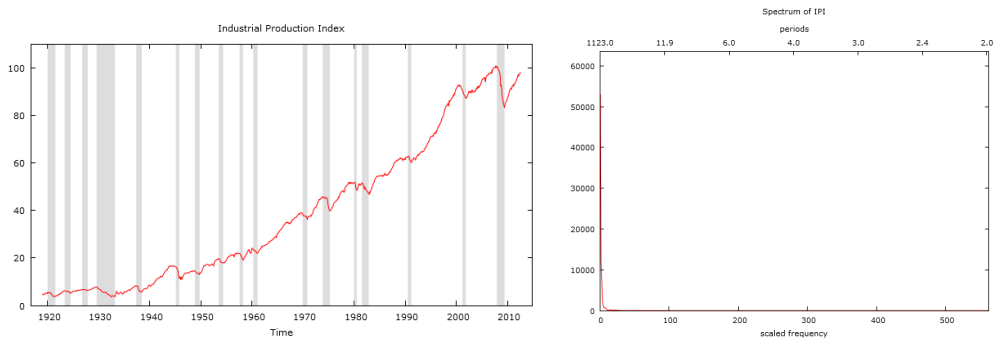
   (b) In order to detect the nonlinear behavior in the underlying time series, we use the DVV method discussed in Section 4.2.3. We are able to generate delay vectors necessary for the DVV analyzes using the (m, τ) obtained in the step (5.1). Unlike classical nonlinearity testing procedures, this non-parametric method is essentially data-driven and carry no *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.



(a) Industrial Production Index with the shaded areas indicating the US recessions.

(b) The Spectrum of the Industrial Production Index (IPI) which can be interpreted as a presence of long memory dynamics

FIGURE 2.3: US Industrial Production Index (IPI) time seris. 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.

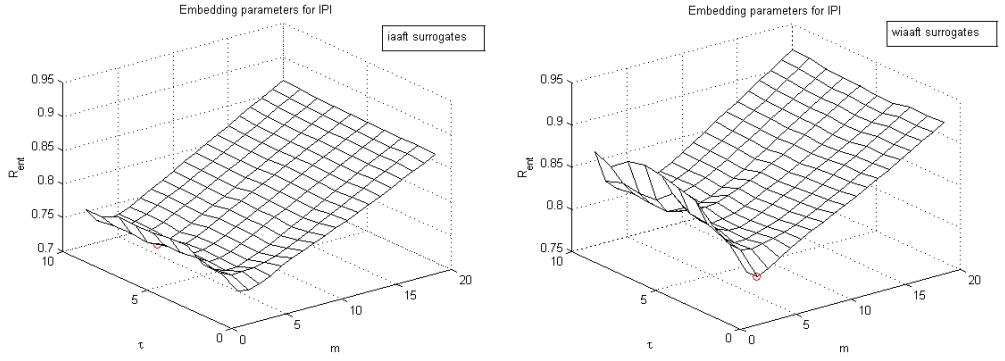
The monthly US Industrial Production Index (IPI) time series[3] spanning over the period January, 1919 to July, 2012 is considered for the data analysis. Figure 3.1 is the plot of the monthly IPI series for the period: 1919:01 - 2012:07, implying 1123 observations, where the shaded regions corresponds 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/

[4]National Bureau of Economic Research: http://www.nber.org/cycles.html

i.e., the embedding dimension, m, and the time lag, τ, for the DVV method with both the iAAFT surrogates and WiAAFT surrogates. We consider two approaches for estimating (m, τ). In the first case, the optimal embedding parameters are estimated using wiAAFT surrogates 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.



(a) Differential-Entropy based method with iAAFT surrogates. The optimal embedding values are $m = 4$ and $\tau = 7$ with *entropy ratio*, $R_{ent}(m, \tau) = 0.7271$.

(b) Differential-Entropy based method with wiAAFT surrogates. The optimal embedding values are $m = 3$ and $\tau = 1$ with $R_{ent}(m, \tau) = 0.7923$.

FIGURE 2.4: 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 is creating delay vectors needed for 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

Data	Code, H	Rank-Order	Rank-Threshold	Decision
IPI	1	26	24.7	Nonlinear Dynamics

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.

could be due to both structural and behavioral changes that can occur in the economy across time. In otherwords, 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]).

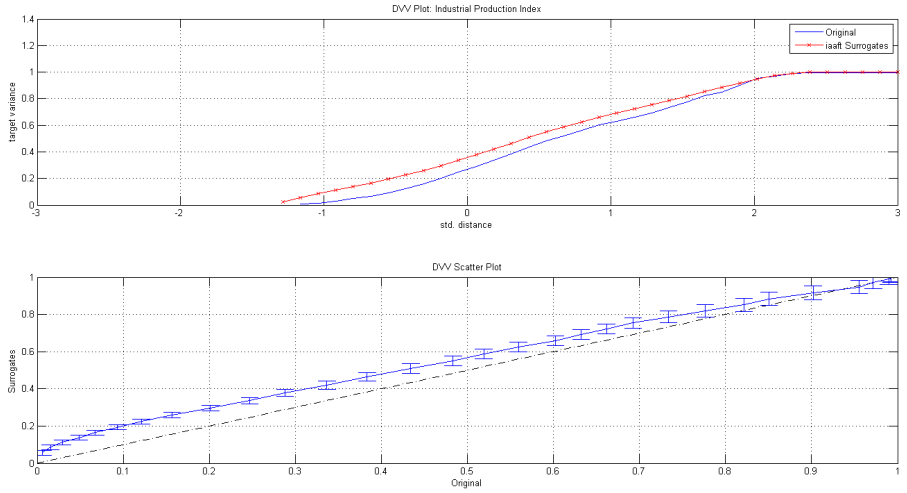
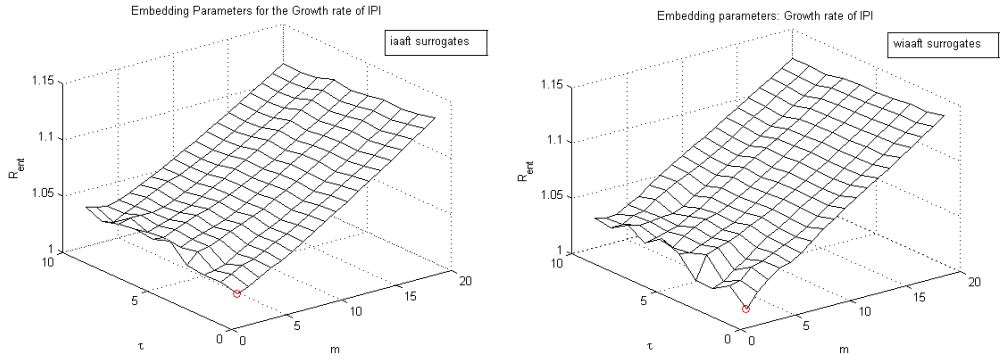


FIGURE 2.5: This is the DVV analysis with iAAFT surrogates performed on the IPI using the embedding parameters obtained via the differential entropy-based method. We clearly observe a deviation from the bisector on the DVV scatter diagram. The DVV plot also indicates that the process is neither strictly 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.



(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.

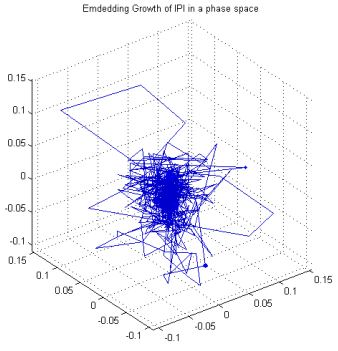
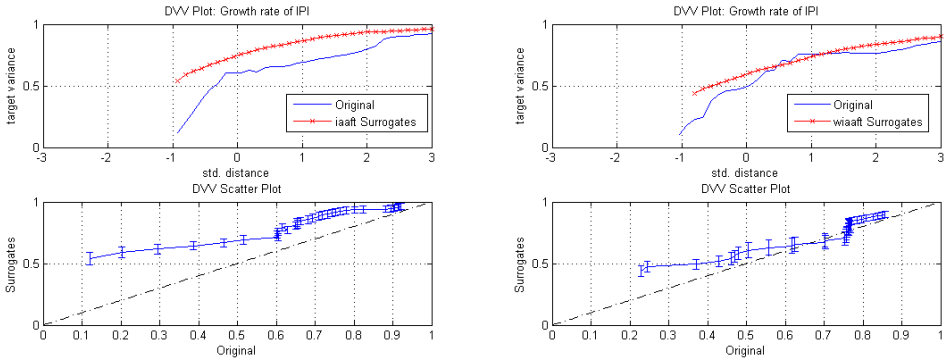


FIGURE 2.7: Phase Space reconstruction using the embedding parameters $m = 2$ and $\tau = 1$. This represents the embedding of the underlying time series, growth rate of IPI, in phase space. The attractor is clearly visualized.



(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.

Data	surrogates	Code, H	Rank-Order	Rank-Threshold	Decision
Growth rate of IPI	wiAAFT	1	25	24.7	Nonlinear Dynamics
Growth rate of IPI	iAAFT	1	26	24.7	Nonlinear Dynamics

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

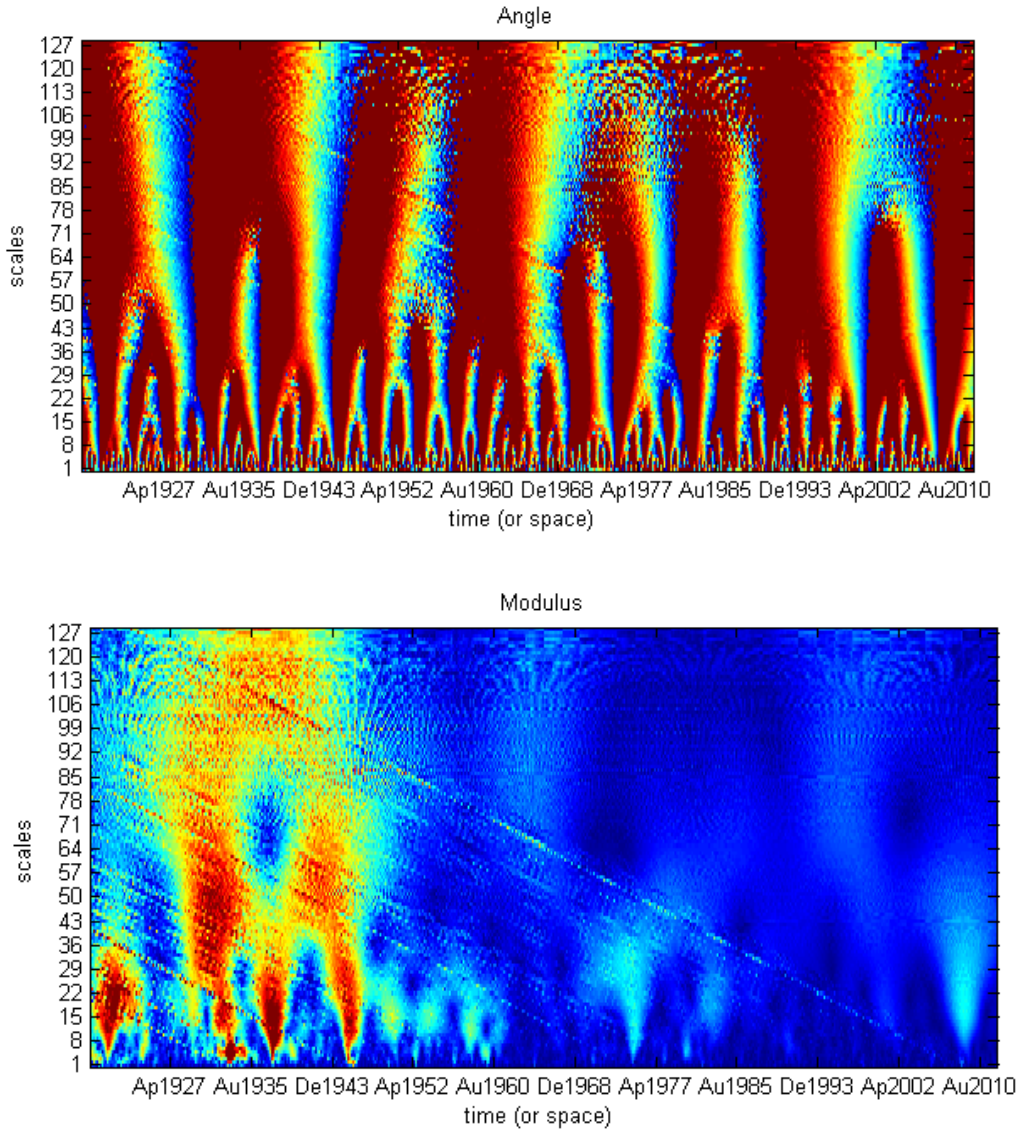


FIGURE 2.9: Coefficients plots obtained from the CWT using complex Morlet wavelet on the growth rate of IPI: First row represents the phase (angle) plot and second row is the corresponding Modulus plot. The colormap ranges from blue to red, and passes through the colors cyan, yellow, and orange. The 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

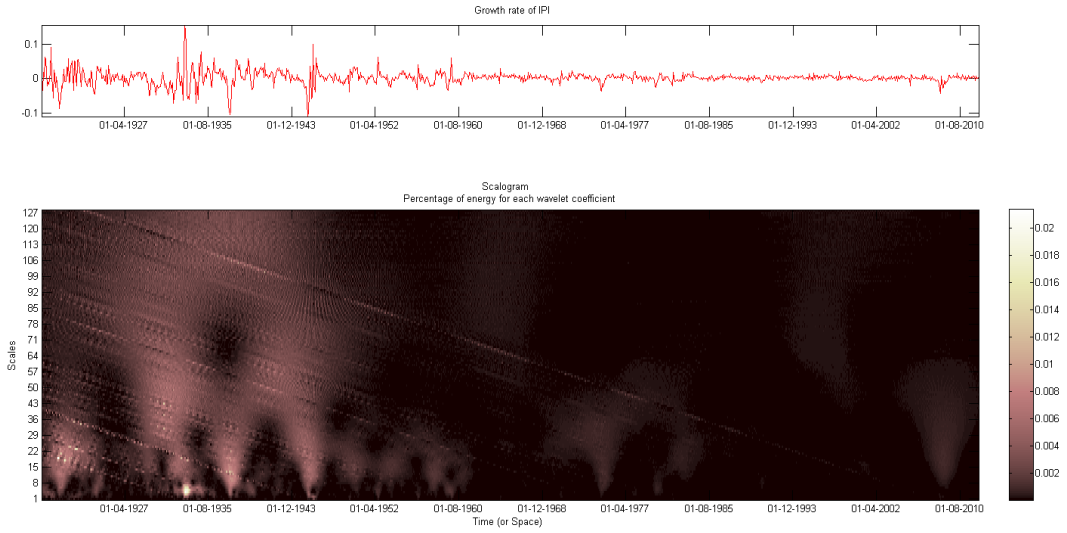


FIGURE 2.10: The IPI growth rate and associated Scalogram from the CWT. The bar with colour scales on the left-hand side of the scalogram plot indicates the percentage of energy for each wavelet coefficient. Higher energy levels can be clearly observed for the Great Depression of the 1930s compared to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

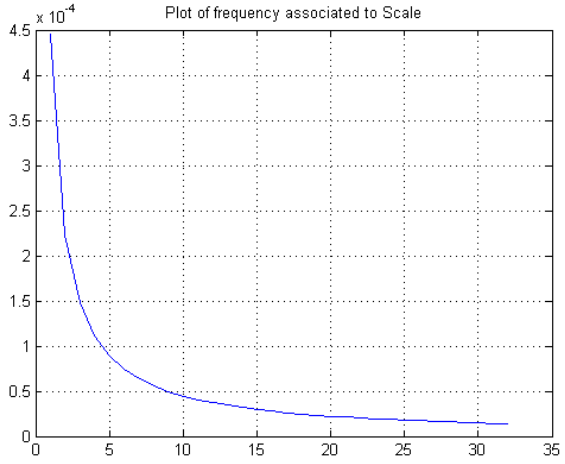


FIGURE 2.11: The pseudo-frequency associated to scale, in Hertz (Hz). The horizontal axis represents the scales and the vertical axis corresponds to the frequency associated to a scale.

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.

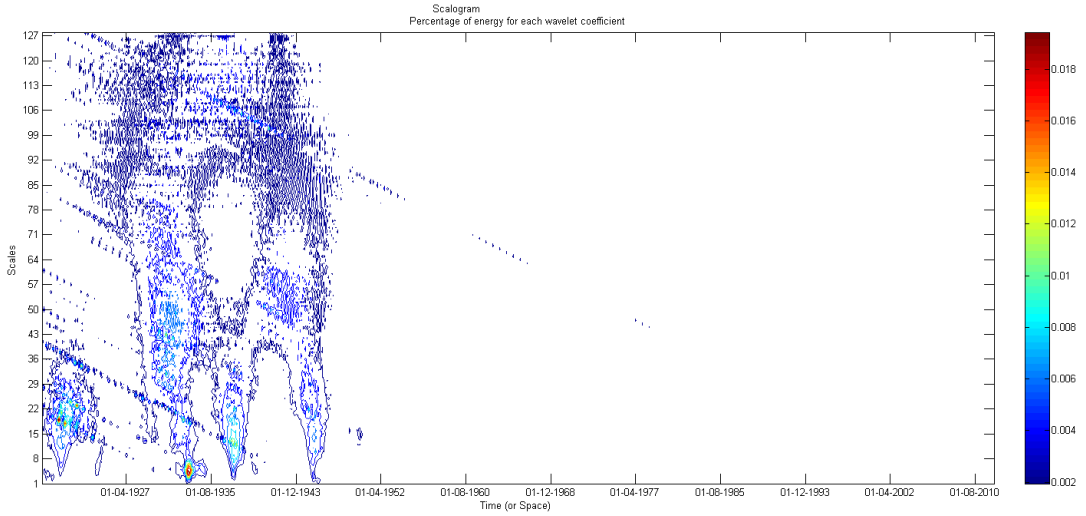


FIGURE 2.12: A contour representation of Scalogram, figure 2.10, associated with the US IPI growth rate.

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 corresponds 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

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

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).

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[2] 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

[2] A comprehensive introduction and bibliography about on recurrence plots is available on http://www.recurrence-plot.tk/ 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 Δt 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 $\vec{x_i} = (u_i, u_{i+\tau}, \cdots, u_{i+(m-1)\tau})$. The coordinates of this vector correspond to the present and lead values of the time

[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 τ are referred to as the embedding dimension and time delay respectively ([5, 96, 127]). The choice of m and τ are usually based on methods for approximating these parameters, such as the false nearest neighbors method (for m) and mutual information (for τ), 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, τ. 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

$$R^x_{i,j} = \begin{cases} 1 & : \|\vec{x_i} - \vec{x_j}\| < \varepsilon \\ 0 & : otherwise \end{cases} \quad \vec{x_i} \in \mathbb{R}^m, i, j = 1, \cdots, \eta, \quad \eta = N - (m-1)\tau, \tag{3.1}$$

where $\|\cdot\|$ is a norm (e.g Euclidean or maximum norm) and ε 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 calender times at which a state of a dynamical system recurs (columns and rows correspond then to a certain pair of calender times) ([6, 96] and the references therein). In general the cut-off distance, ε, has to be chosen as small as possible, but a too small ε 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[5] 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

$$D^x_{i,j} = \|\vec{x_i} - \vec{x_j}\| \tag{3.2}$$

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

[5]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

$$CR^{xy}_{i,j} = \begin{cases} 1 & : \|\vec{x_i} - \vec{y_j}\| < \varepsilon. \\ 0 & : otherwise \end{cases} \tag{3.3}$$

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

$$JR^{xy}_{i,j} = R^{x}_{i,j} \cdot R^{y}_{i,j}, \quad \vec{x_i} \in \mathbb{R}^m, \quad \vec{y_j} \in \mathbb{R}^n, \quad i, j = 1, 2, \cdots, \eta\vartheta. \tag{3.4}$$

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 $D^x_{i,j}$, 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/

[9]National Bureau of Economic Research

[10]The data can be downloaded from Federal Reserve Bank of St. Louis

[11]http://www.nber.org/cycles.html

[12]Source: [22] and Eurostat.

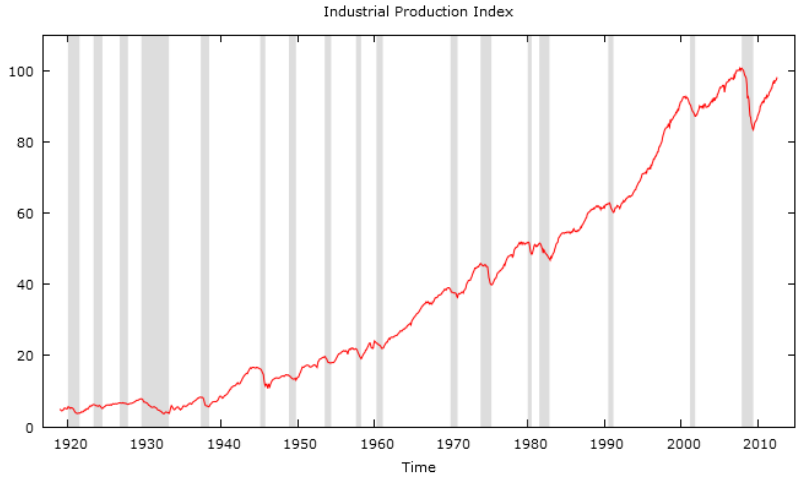


FIGURE 3.1: US Industrial Production Index (IPI) time series. The plot of the monthly IPI series for the period: 1919:01 - 2012:07 ($n = 1123$), where the shaded regions corresponds to the US recessions from 1920 published by NBER.

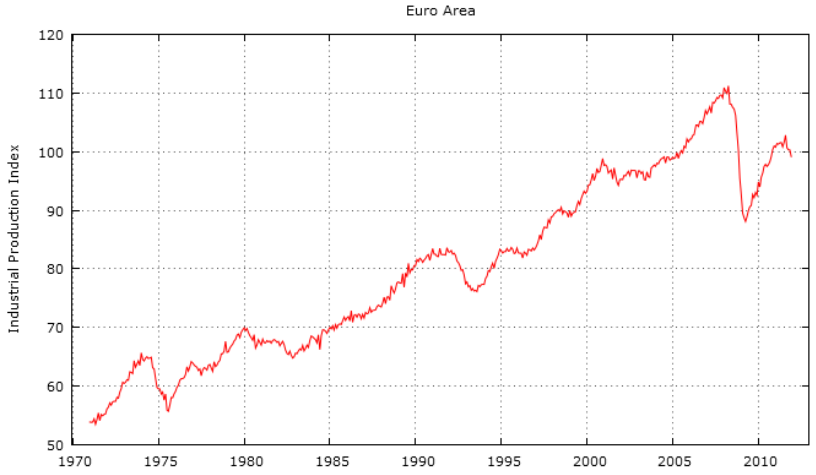


FIGURE 3.2: The plot of the monthly Euro-zone Industrial Production Index (IPI) time series for the period: 1971:01 - 2011:12 ($n = 492$).

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 calender 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 calender date. In other words, fixing a period or date on the horizontal axis, we are able to observe the recurrence of such

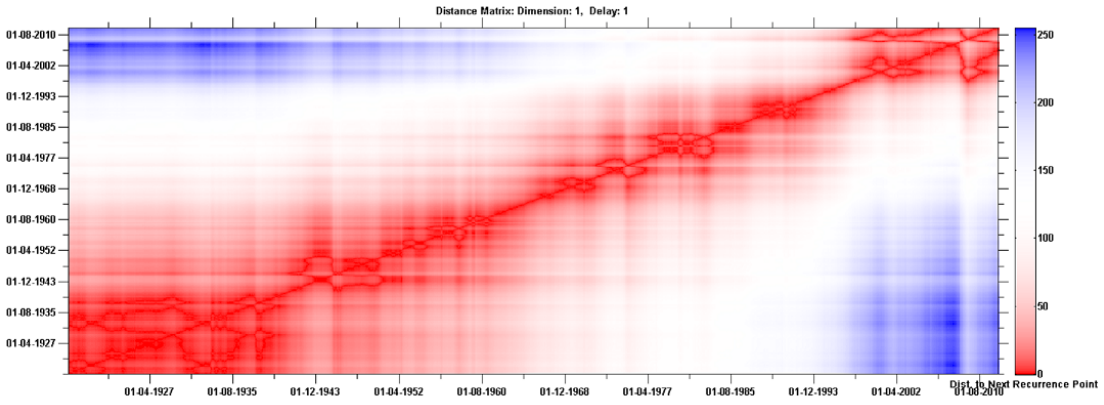


FIGURE 3.3: This is an unthresholded recurrence plot which is sometimes termed *distance plot* ([74]). It is a matrix plot with calender time on both the vertical and horizontal axes. The colormap corresponds to the distance to the next recurrence of a state in the time series. This representation shows 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 calender 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[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 τ 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

[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.

NBER dates		*butterfly* dates	
Peak	Trough	Peak	Trough
1920:01	1921:07	1920:02	1921:08
1923:05	1924:07	1923:05	1924:08
1926:10	1927:11	1926:10	1927:11
1929:08	1933:03	1929:08	1933:04
1937:05	1938:06	1937:05	1938:02
1945:02	1945:10	1945:02	1945:10
1948:11	1949:10	1948:08	1949:10
1953:07	1954:05	1953:07	1954:01
1957:08	1958:04	1957:09	1958:04
1960:04	1961:02	1960:01	1961:03
1969:12	1970:11	1969:10	1970:11
1973:11	1975:04	1974:09	1975:04
1980:01	1980:07	1980:00	1980:08
1981:07	1982:11	1981:08	1983:01
1990:07	1991:03	1990:08	1991:03
2001:03	2001:11	2000:09	2001:11
2007:12	2009:06	2007:12	2009:06

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 [104] 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.

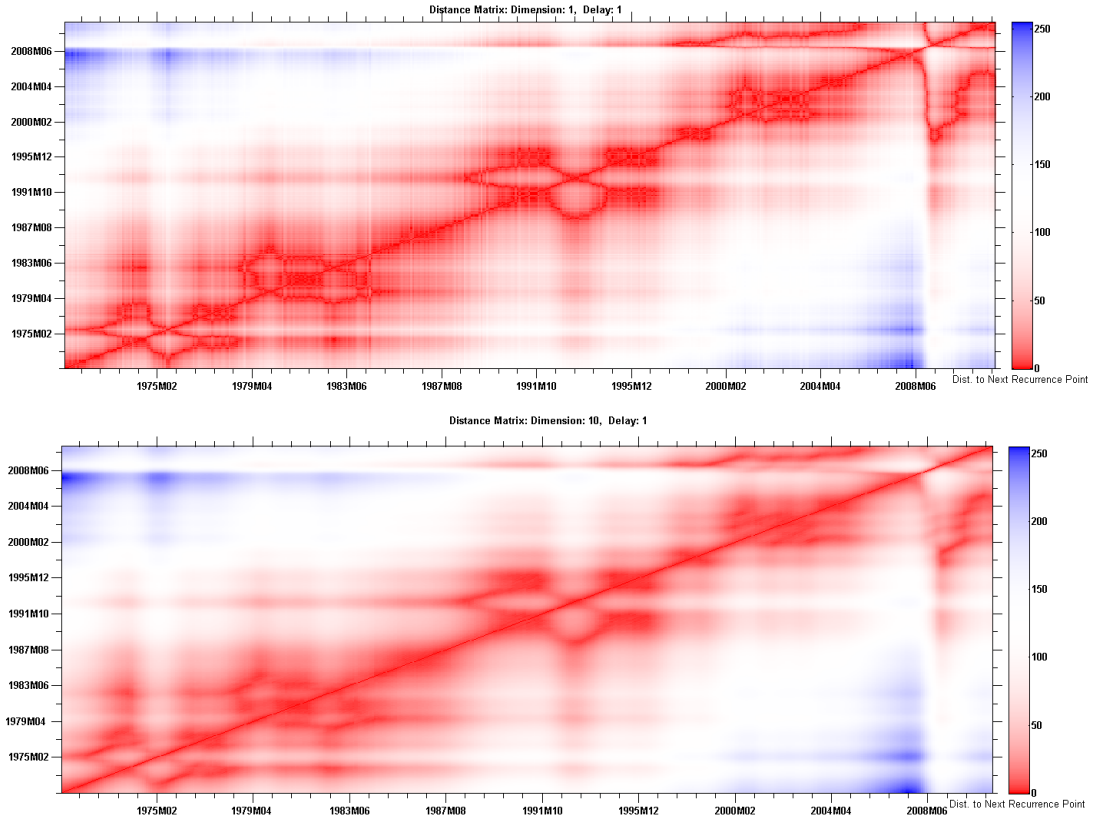


FIGURE 3.4: The *distance plot* for the Euro-zone IPI and it's embedded version. The existence of *butterfly-like* structures, of a minimum size of six months, along the main diagonal (bisector) indicates economic crisis. Embedding the original time series does not yield different results in terms of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

butterfly dates for Euro-Zone	
Peak	Trough
1974:06	1975:08
1980:01	1980:10
1982:05	1982:12
1992:04	1993:06
2000:12	2001:11
2007:12	2009:04
2011:07	

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.

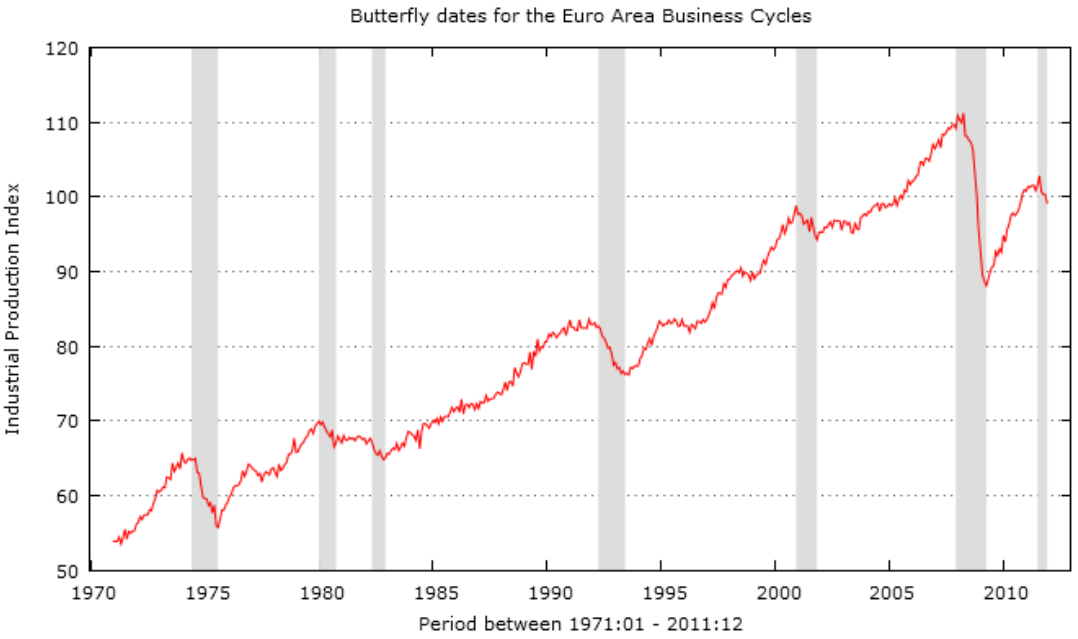


FIGURE 3.5: The shaded regions corresponds to the *Butterfly* dates for the Euro Area Business Cycles from 1971:01–2011:12. The end of the last recession period that started 2011:07(q3) was not determined considering the sample period.

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

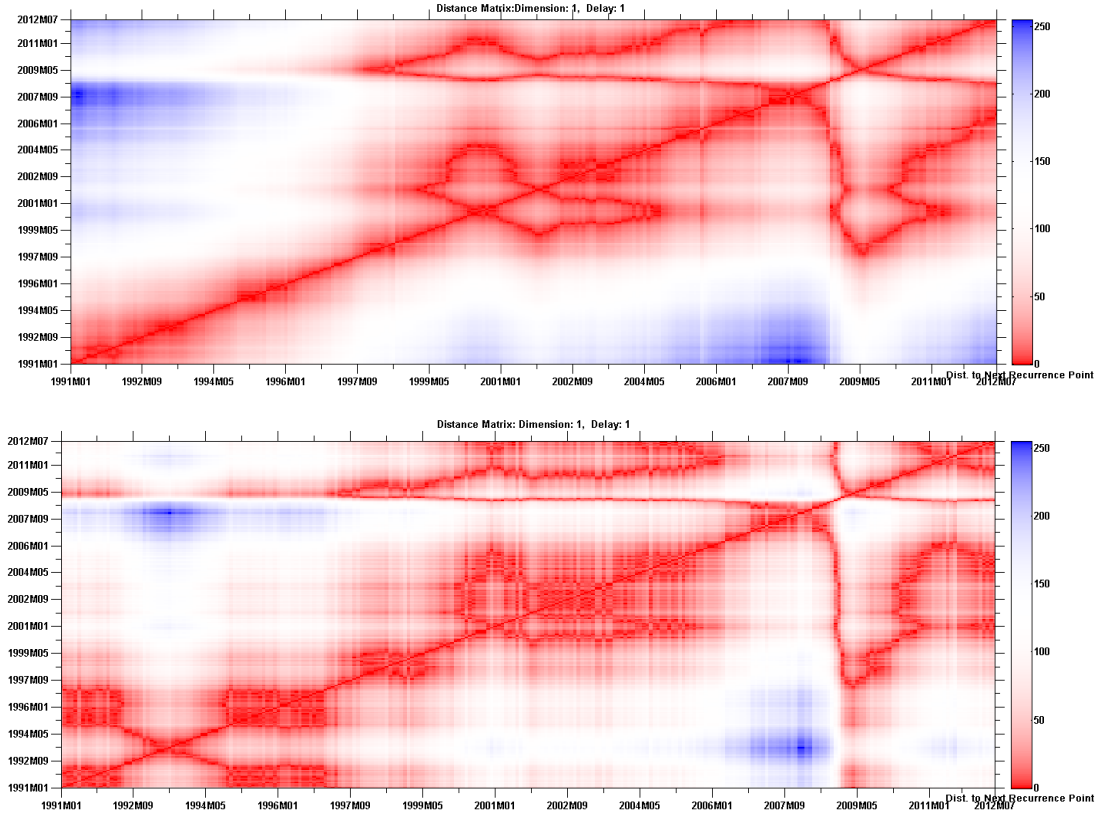


FIGURE 3.6: The *distance plots* of US IPI and Euro-zone IPI for the period 1991:01 - 2012:07. The top figure corresponds to the *distance plots* of US IPI. The second *distance plot* is for the Euro-zone IPI.

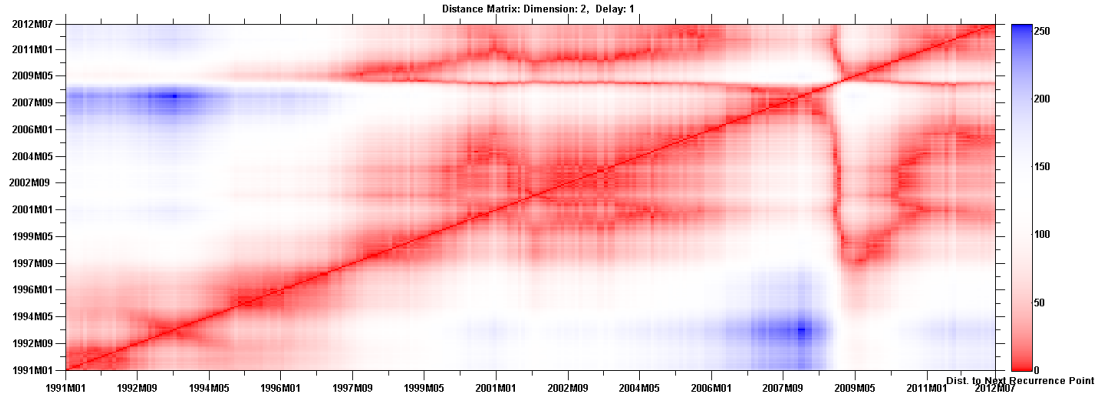
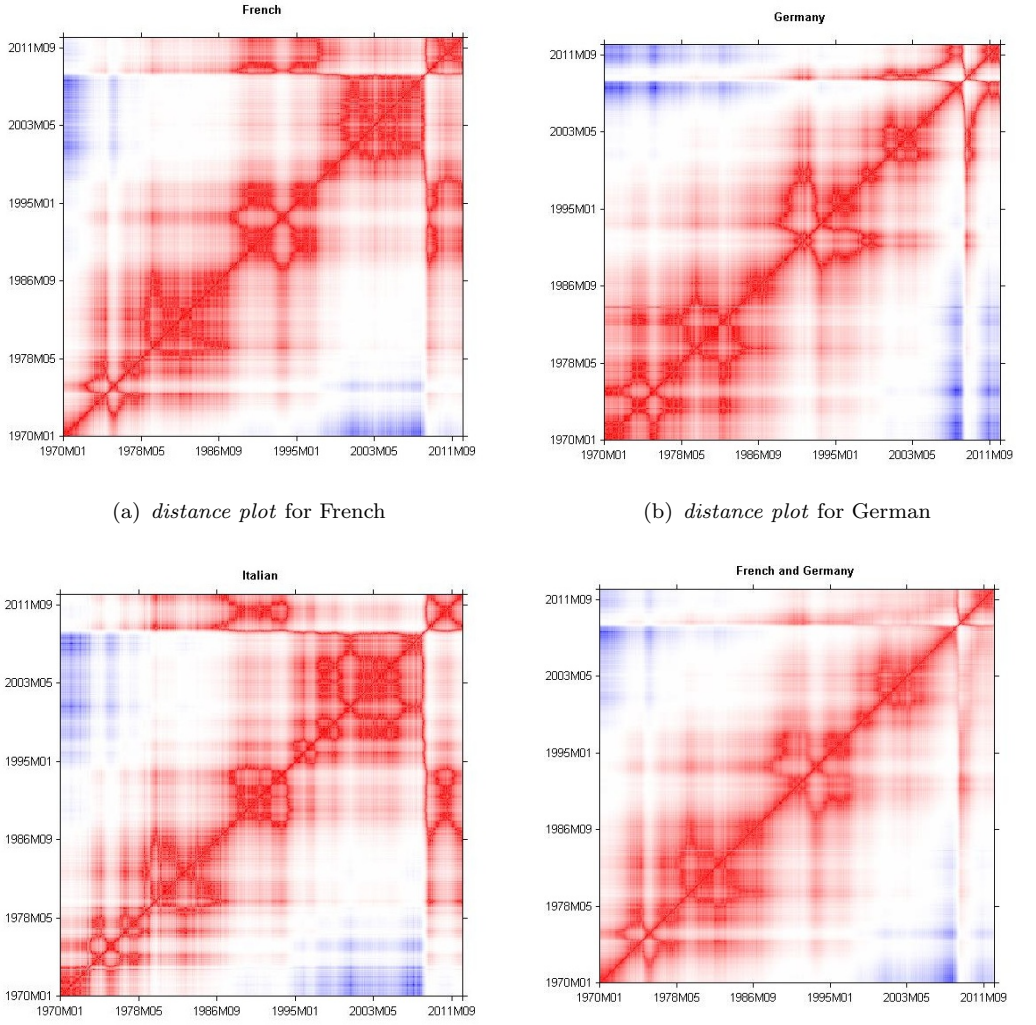


FIGURE 3.7: This is a cross recurrence plot to visualize the simultaneous occurrence of a similar state in both US IPI and Euro IPI.

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



(a) *distance plot* for French

(b) *distance plot* for German

(c) *distance plot* for Italian

(d) Cross recurrence plot for French and German

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.

recurrence plots on multivariate time series[14] (e.g. production, employment, real income and real sales), which in this case will be a vector of indicators.

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

[14]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/

butterfly dates		*[10]* dates		*CEPR* dates		*MU* dates	
Peak	Trough	Peak	Trough	Peak	Trough	Peak	Trough
1974:06(q2)	1975:08(q3)	1974:04(q2)	1975:04(q2)	1974:xx(q3)	1975:xx(q1)	1974:08(q3)	1975:04(q2)
1980:01(q1)	1980:10(q4)	1980:02(q1)	1981:01(q1)	1980:xx(q1)	1982:xx(q3)	1980:03(q1)	1980:09(q3)
1982:05(q2)	1982:12(q4)	1981:10(q4)	1982:12(q4)	–	–	1982:04(q2)	1982:07(q3)
1992:04(q2)	1993:06(q2)	1992:01(q1)	1993:05(q2)	1992:xx(q1)	1993:xx(q3)	1992:02(q1)	1993:01(q1)
2000:12(q4)	2001:11(q4)	2000:12(q4)	2001:12(q4)	–	–	–	–
2007:12(q4)	2009:04(q2)	–	–	2008:xx(q1)	2009:xx(q2)	–	–
2011:07(q3)							

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.

conference participants of CFE-ERCIM 2012 and the participants at the 2013 EBIM-EDEEM Doctoral Jamboree at CORE for their participation and interest.

Chapter 4

Nonlinear dynamics and recurrence plots for detecting financial crisis

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 studys ([20, 72, 87, 125], among others) .This technique is applicable to any time series since it requires no a prior 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/ and [96]

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[2] 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.

[2]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 α 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).

'phase space' by a set of delay vectors[4] (DVs) 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

$$R_{ent}(m, \tau) = I(m, \tau) + \frac{m \ln N}{N}, \tag{4.1}$$

where N is the number of delay vectors, which is kept constant for all values of m and τ under consideration,

$$I(m, \tau) = \frac{H(x, m, \tau)}{\langle H(x_{s,i}, m, \tau) \rangle_i} \tag{4.2}$$

where x is the signal, $x_{s,i}$ $i = 1, \cdots, T_s$ 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

$$H(x) = \sum_{j=1}^{T} \ln(T\rho_j) + \ln 2 + C_E \tag{4.3}$$

where T is the number of samples in the data set, ρ_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 $\tau_{opt} = 1$. 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-anlysis is based on the calculation of the *target variance*, σ^{*2}, which is an inverse measure of the predictability of a time series. The algorithm is summarized below:

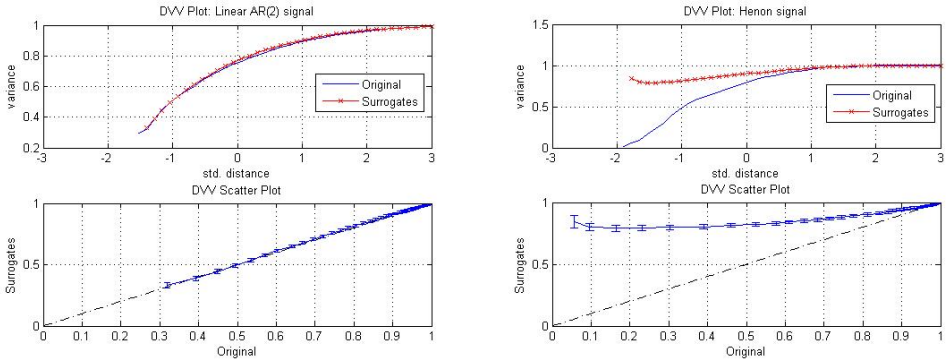
- For an optimal embedding dimension m and time lag τ, generate delay vector (DV): $x(k) = [x_{k-\tau}, \cdots, x_{k-m\tau}]$ and corresponding target x_k
- The mean μ_d and standard deviation, σ_d, are computed over all pairwise distances between DVs, $\|x(i) - x(j)\|$ for $i \neq j$.
- The sets Ω_k 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 ϱ_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 Ω_k, the variance of the corresponding targets, σ_k^2, is computed. The average over all sets Ω_k, normalized by the variance of the time series, σ_x^2, yields the *target variance*, σ^{*2} :

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

  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, ϱ_d. The minimum *target variance*, $\sigma_{min}^{*2} = \min_{\varrho_d}[\sigma^{*2}(\varrho_d)]$, 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, σ_{min}^{*2} is inversely related to prevalence of the deterministic component over the stochastic one, lowest σ_{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]



(a) AR(2) signal (b) Henon signal

FIGURE 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.

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 WiAAFT 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 Appendix A.2 for more on DVV analysis of some simulated process.

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 $\vec{x}(t) = \vec{x_i} \in \mathbb{R}^m$ $(i = 1, \cdots, \eta\ ,\ t = i\Delta t$, where Δt 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 $\vec{x_i} = (u_i, u_{i+\tau}, \cdots, u_{i+(m-1)\tau})$. The coordinates of this vector correspond to the present and lead values of the series. The parameters m and τ 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

$$R_{i,j}(\varepsilon) = \begin{cases} 1 & : \|\vec{x_i} - \vec{x_j}\| < \varepsilon \\ 0 & : otherwise \end{cases} \quad \vec{x_i} \in \mathbb{R}^m, i, j = 1, \cdots, \eta, \quad \eta = N - (m-1)\tau, \tag{4.5}$$

where $\|\cdot\|$ is a norm (e.g Euclidean or maximum norm) and ε 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 calender times at which a state of a dynamical system recurs (columns and rows correspond then to a certain pair of calender 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 $\vec{x_i}$ $(i = 1, \cdots, \eta)$ with each point of the second trajectory $\vec{y_i}$ $(i = 1, \cdots, \vartheta)$ resulting in $\varepsilon \times \vartheta$ array

$$CR_{i,j}(\varepsilon) = \begin{cases} 1 & : \|\vec{x_i} - \vec{y_j}\| < \varepsilon. \\ 0 & : otherwise \end{cases} \tag{4.6}$$

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[7] 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[8] 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, τ, 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 .

[7]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/

[8]The historical financial data can be downloaded from http://quote.yahoo.com/ using the `get.hist.quote` command contained in the `tseries` package in *R*.

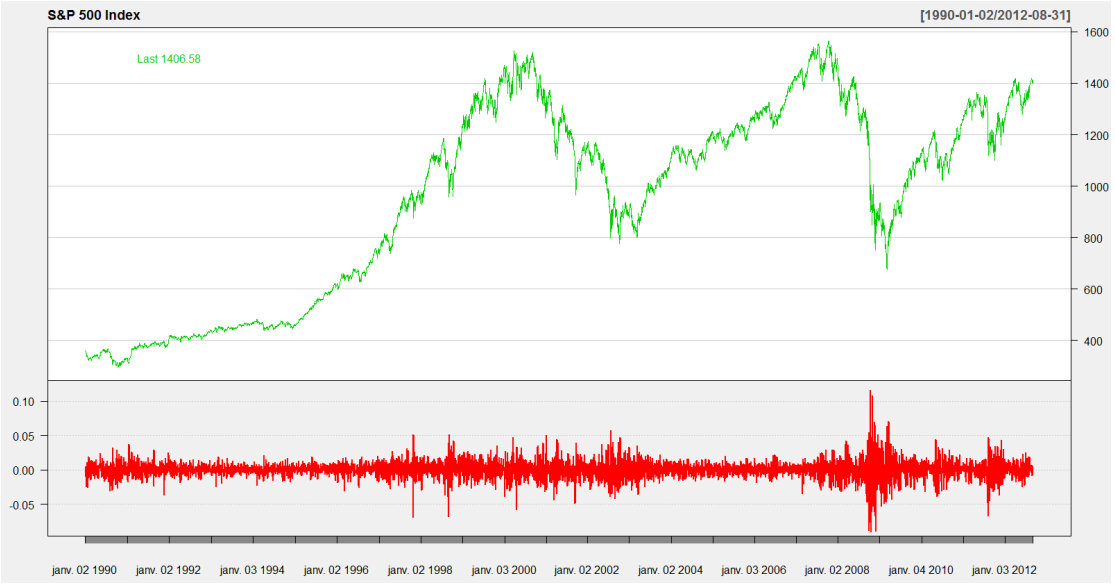


FIGURE 4.2: The daily adjusted closing price of S&P 500 Index and it's associated rate of change for the time period 1990-01-02 to 2012-08-31. Thus, the underlying financial time series is of length 5716 observations.

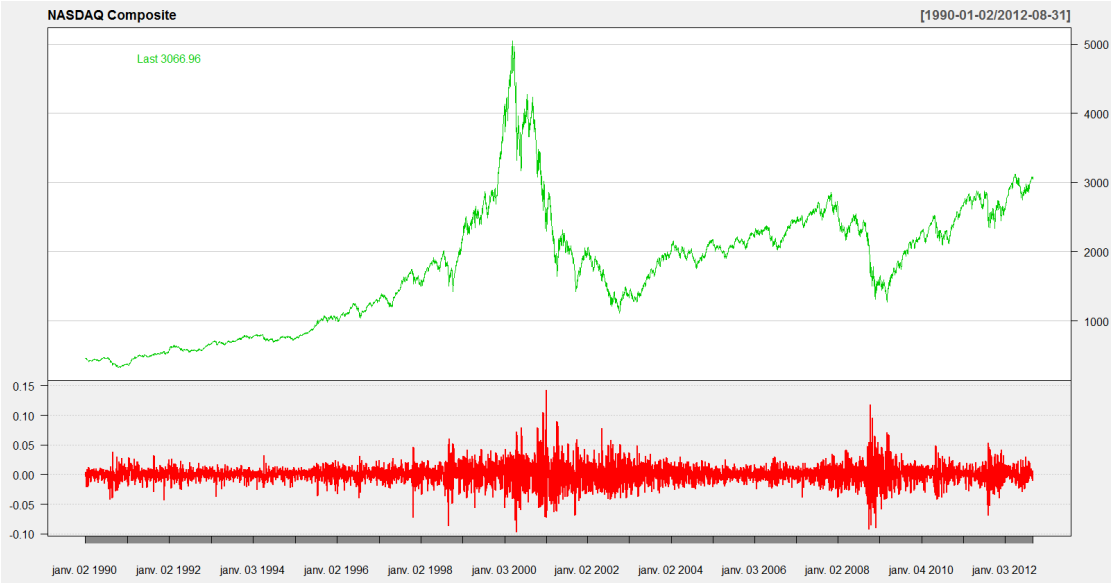
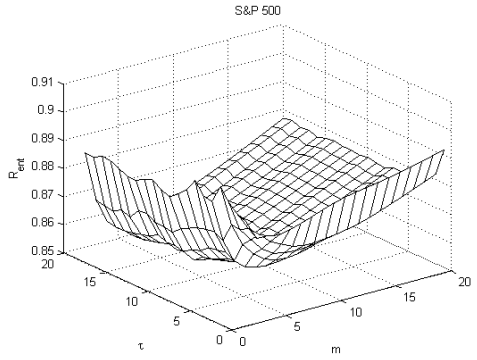


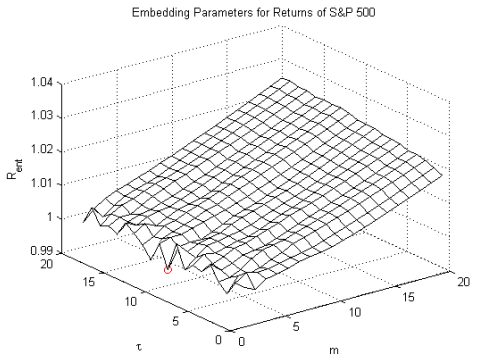
FIGURE 4.3: The daily adjusted closing price of Nasdaq Composite and it's associated returns for the time period 1990-01-02 to 2012-08-31. This implies the time series is of length 5716 observations.

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

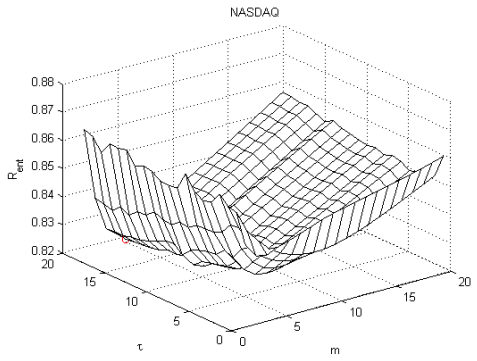


(a) Differential-Entropy based method with wiAAFT surrogates on S&P 500 Index. The optimal embedding values are $m = 6$ and $\tau = 18$ with *entropy ratio*, $R_{ent}(m, \tau) = 0.8526$.

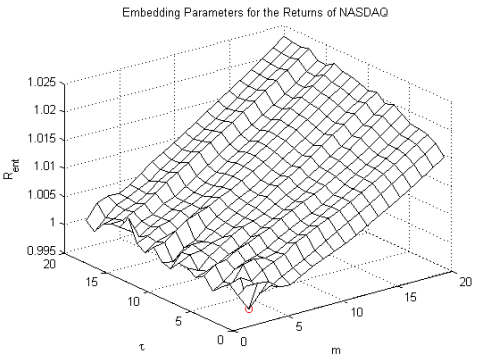


(b) Differential-Entropy based method with wiAAFT surrogates on the returns of S&P 500. The optimal embedding values are $m = 2$ and $\tau = 10$ with *entropy ratio*, $R_{ent}(m, \tau) = 0.9948$.

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 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

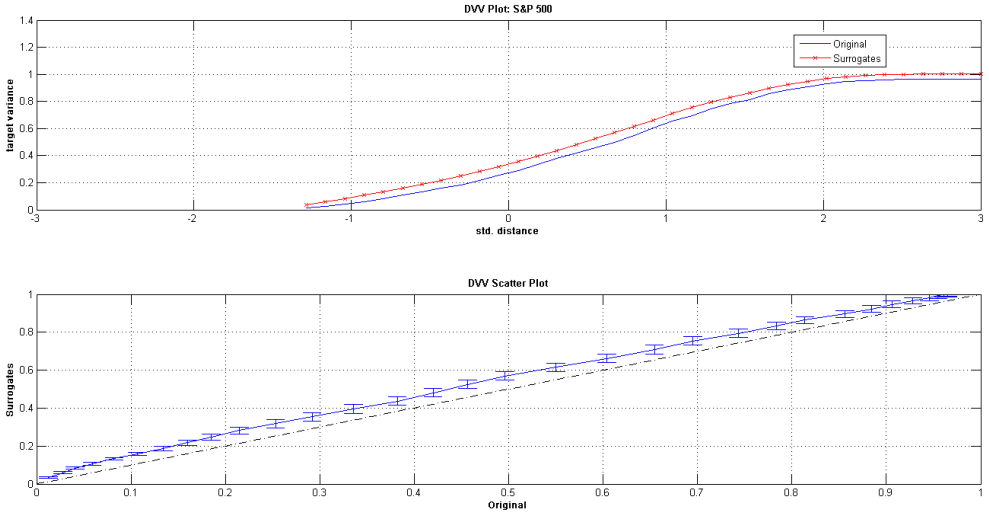


FIGURE 4.6: This is the DVV analysis with iAAFT surrogates performed on S&P 500 using the embedding parameters obtained via the differential entropy-based method. We clearly observe a deviation from the bisector on the DVV scatter diagram. The DVV plot also indicates that the process is neither strictly 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]).

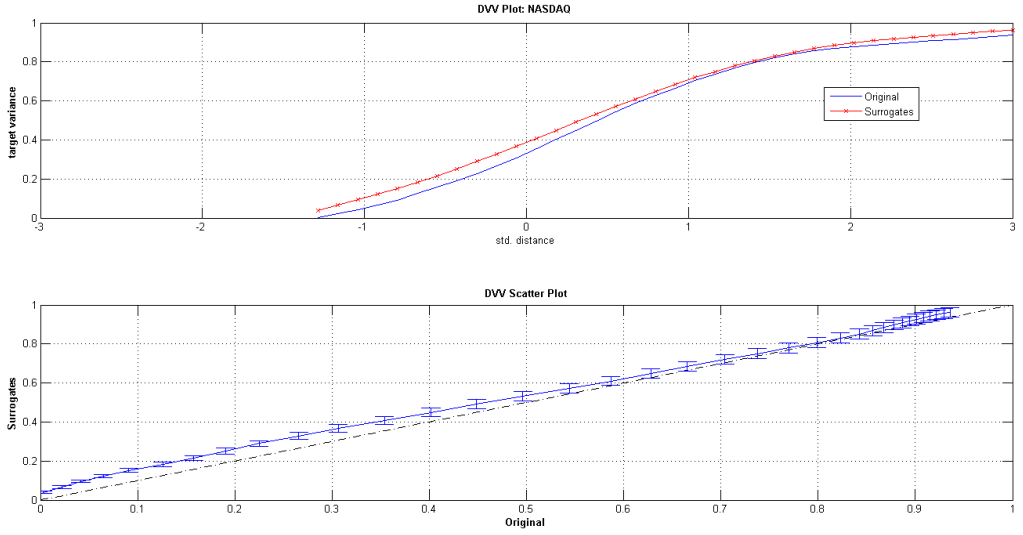


FIGURE 4.7: This is the DVV analysis with iAAFT surrogates performed on Nasdaq Composite using the embedding parameters obtained via the differential entropy-based method. We clearly observe a deviation from the bisector on the DVV scatter diagram. The DVV plot also indicates that the process is neither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

second case, recurrence plots are performed on embedded versions of the underlying financial times using the previously estimated values of embedding parameters.

Data	Code, H	Rank-Order	Rank-Threshold	Decision
S&P 500	1	26	24.7	Nonlinear Dynamics
Nasdaq Composite	1	25	24.7	Nonlinear Dynamics

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.

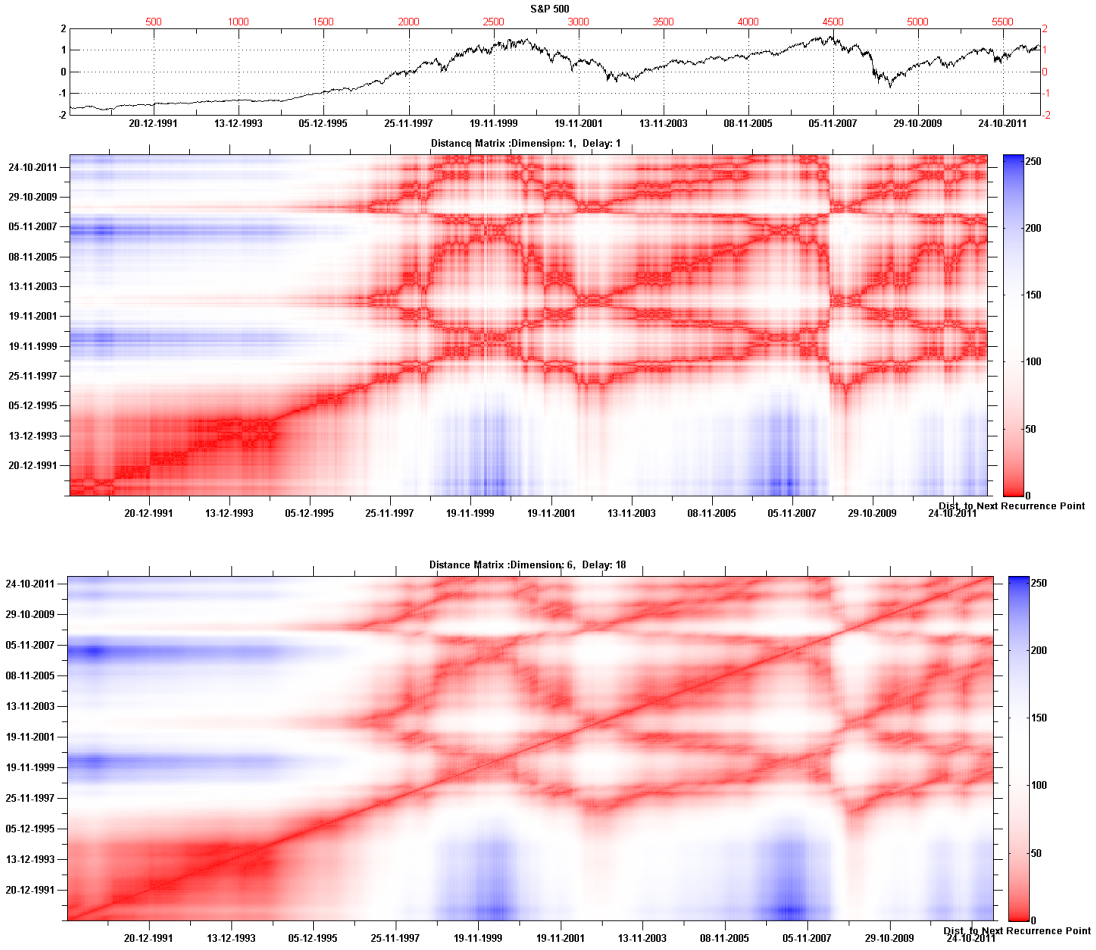


FIGURE 4.8: This is an unthresholded recurrence plot which is sometimes termed *distance plot* ([74]). It is a matrix plot with calender time on both the vertical and horizontal axes for the time period 1990-01-02 to 2012-08-31. The colormap corresponds to the distance to the next recurrence of a state 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 calender 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 it's 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 calender date. Fixing a period or date on the horizontal axis, we are able to observe the recurrence of such events along the calender 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,

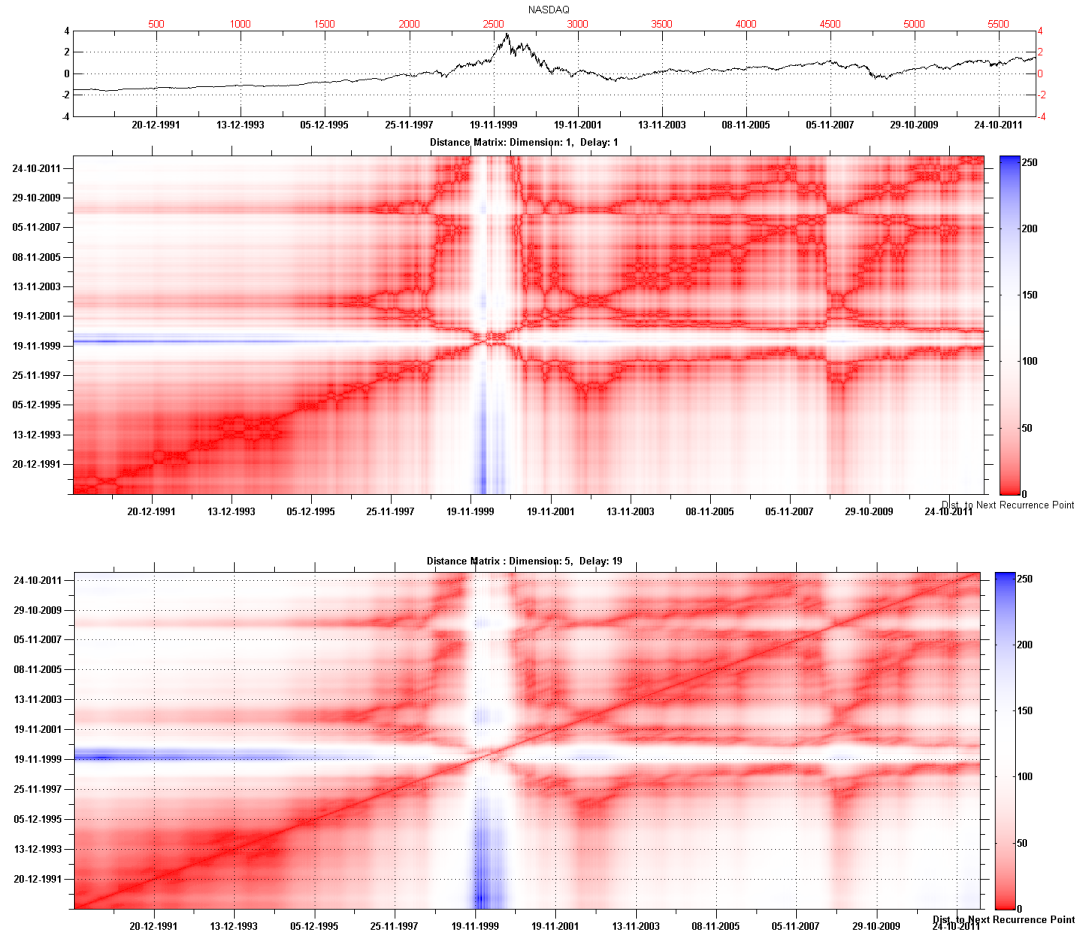


FIGURE 4.9: The *distance plot* of the Nasdaq Composite and it's embedded version. It is observed that the Nasdaq was well exposed and hit by the dot-com bubble of the 1990s compared to S&P 500. Nasdaq, on which many dot-coms traded, rose to record highs as indicated by larger distance to next recurrence.

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 enable to visualize and date these peaks in both returns of the underlying series. The peaks dates are recognised 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 $Y_t = logX_t - logX_{t-1}$.

[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.

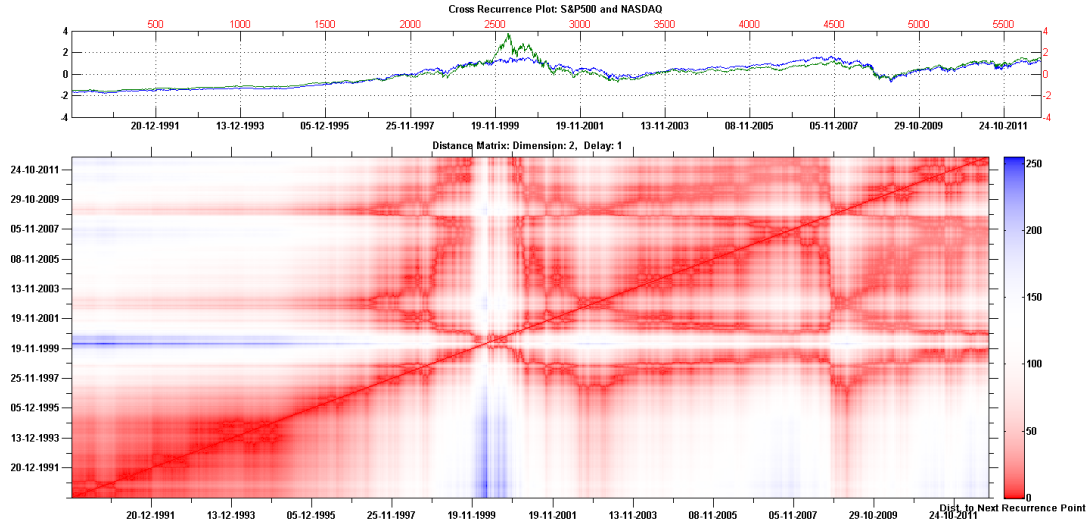


FIGURE 4.10: This is a cross recurrence plot to visualize the simultaneous occurrence of a similar state in both S&P 500 and Nasdaq Composite. The dot-com bubble clearly occurs in both series.

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

$$x_{t+1} = \mu + \sum_{i=1}^{m} g_i(x_{t-(i-1)d}) + \varepsilon_{t+1} \tag{4.7}$$

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, $\varepsilon_{t+1} \sim i.i.dN(0, \sigma^2_{\varepsilon_{t+1}})$ 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 *distance plot* on the extended data with range 1990-01-01 to 2013-08-01 is displayed in Figure 4.12. The last *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 *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

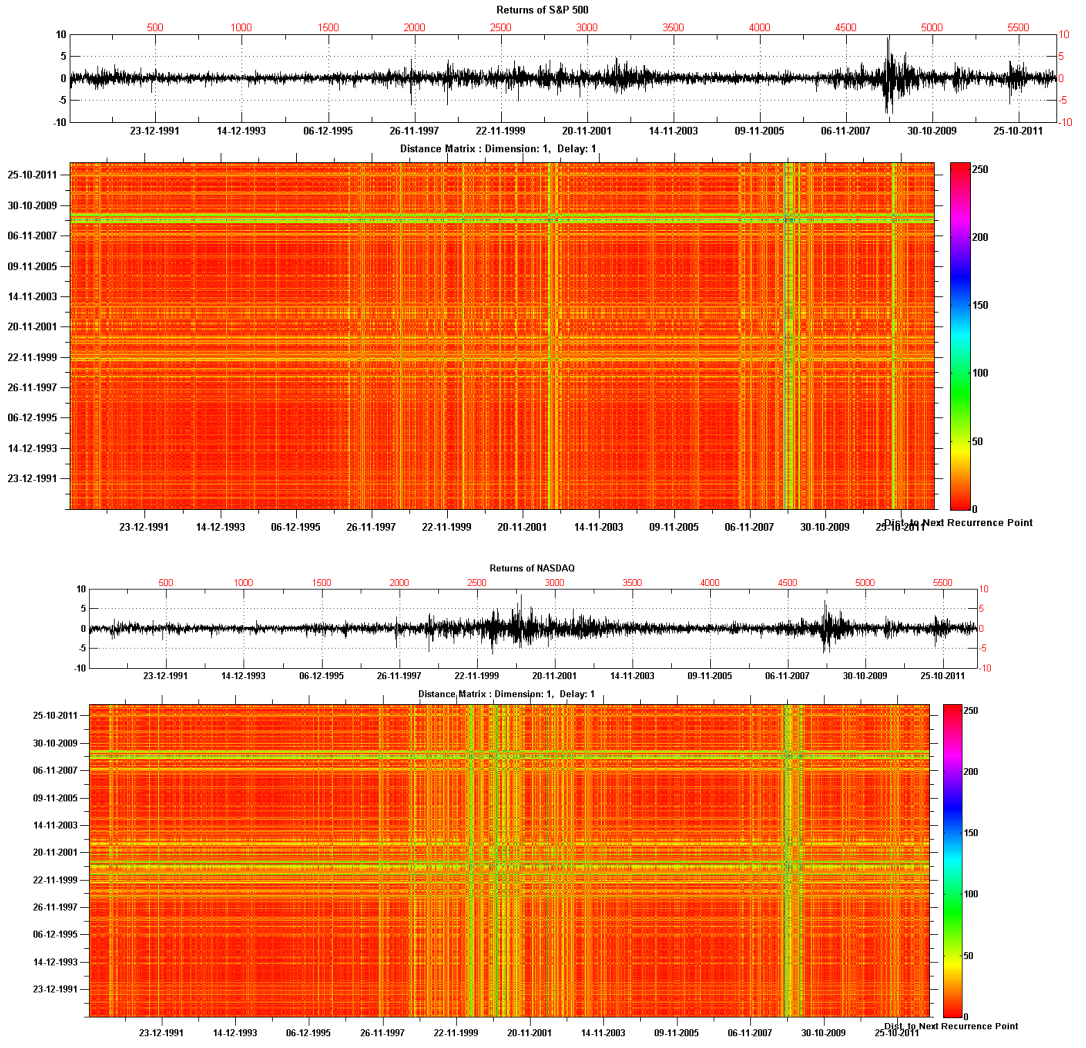


FIGURE 4.11: The *distance plot* of the growth rate (returns) of S&P 500 and Nasdaq Composite. The growth rate cycle is often subject to very short-term fluctuations mainly due to transitory events making the peaks of such cycles extremely difficult to date. In this case, we are enable to visualize and 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.

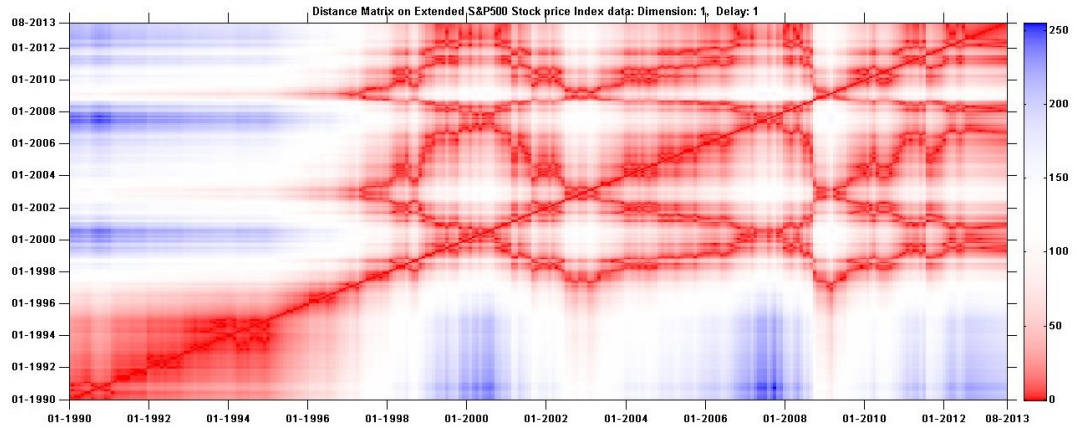


FIGURE 4.12: The *distance plot* on an extended monthly S&P 500 stock price Index: original data range 1990-01-01 to 2012-08-01 with an additional 12 months ahead forecast obtained via the additive nonlinear autoregressive model (4.7). There is no signal of a possible occurrence of financial crisis in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

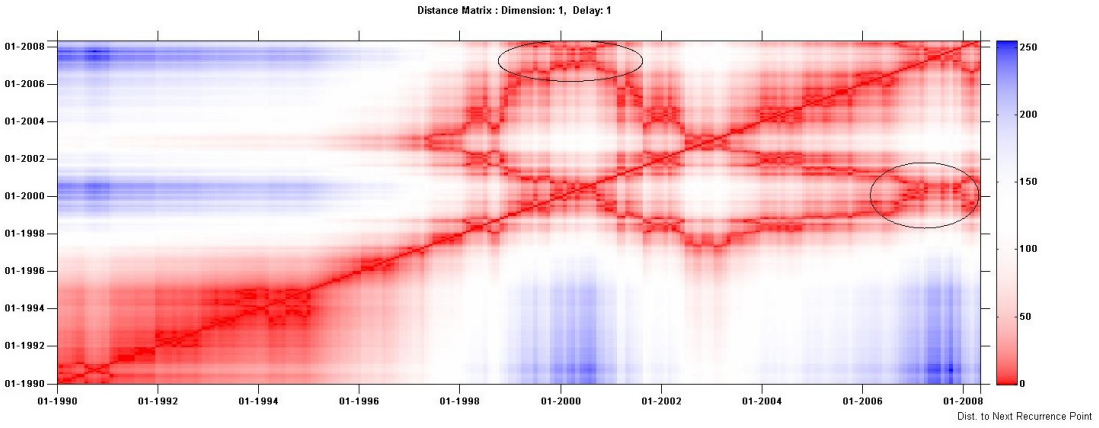


FIGURE 4.13: The *distance plot* on an extended monthly S&P 500 price index: original data range 1990-01 to 2008-02 with an additional 3 months ahead forecast obtained via the additive nonlinear autoregressive model (4.7). The region of the two ellipse drawn on the plot shows that crisis that began in 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,

Month	Forecast value
2012:09	1407.371
2012:10	1402.220
2012:11	1395.267
2012:12	1388.223
2013:01	1381.473
2013:02	1375.089
2013:03	1369.070
2013:04	1363.397
2013:05	1358.050
2013:06	1353.009
2013:07	1348.255
2013:08	1343.769

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

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

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

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

where $\Phi = (\varphi_0, \varphi_1, \cdots, \varphi_p) = (\vartheta_0 - \psi_0, \vartheta_1 - \psi_1, \cdots, \vartheta_p - \psi_p)$. The process $(\varepsilon_t)_t$ 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 $(\varepsilon_t)_t$ 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: γ 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

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

  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 γ and the error term variance σ^2. In the ESTAR setting, very small values of σ^2 involve an unidentified γ, making nearly impossible to obtain a consistent estimate of γ. [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 σ^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 $(y_t)_t$ by the following representation:

$$\begin{cases} y_t = [\Psi\omega_t]\cdot[1 - G_n(y_{t-d},\gamma,c)] + [\Theta\omega_t]\cdot G_n(y_{t-d},\gamma,c) + \varepsilon_t \\ G_n(y_{t-d},\gamma,c) = 1 - (1 + f_n(y_{t-d},\gamma,c))^{-\gamma}, \quad n \in \mathbb{N} \\ f_n(y_{t-d},\gamma,c) = (\prod_{i=1}^{n}(y_{t-d} - c_i))^2 + k(\prod_{i=1}^{n}(y_{t-d} - c_i)),\ c_1 \leq c_2 \leq \cdots \leq c_n, k \in \mathbb{R}, \end{cases} \tag{5.6}$$

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:

$$y_t = [\Psi\omega_t] + [\Phi\omega_t]\cdot G_n(y_{t-d},\gamma,c) + \varepsilon_t, \quad t \geq 1, \tag{5.7}$$

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:

$$\begin{cases} y_t = [\Psi\omega_t] + [\Phi\omega_t]\cdot G(y_{t-d},\gamma,c) + \varepsilon_t \\ G(y_{t-d},\gamma,c) = 1 - (1 + (y_{t-d} - c)^2 + k(y_{t-d} - c))^{-\gamma}. \end{cases} \tag{5.8}$$

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 $\gamma \to \infty$ then $G(\cdot) \to 1 - \mathcal{I}_c$, 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:

- if $\gamma = 0$, $G(y_{t-d}, \gamma, c) = 0$ and we are back to a linear error correction framework where the process $(y_t)_t$ 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 ε_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, $d = 1$,

$$\begin{cases} y_t = \psi y_{t-1} + \varphi y_{t-1} G(y_{t-1}, \gamma, c) + \varepsilon_t \\ G(y_{t-1}, \gamma, c) = 1 - (1 + (y_{t-1} - c)^2 + k(y_{t-1} - c))^{-\gamma} \quad . \end{cases} \tag{5.9}$$

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

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

$$\Delta y_t = \beta y_{t-1} + \varphi y_{t-1} G(y_{t-1}, \gamma, c) + \varepsilon_t, \tag{5.10}$$

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 γ 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 $-2 < \beta < 0$.

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

$$\varepsilon_t = \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + u_t, \tag{5.11}$$

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

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

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 β 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 implys 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:

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

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 Φ 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

$$H_0 : \Phi = 0_{(1\times p)} \quad vs. \quad H_1 : at \quad least \quad one \quad \varphi_i \neq 0, \quad i = 1, \cdots, p \tag{5.14}$$

which is equivalent to test for

$$H_0 : \gamma = 0 \quad vs. \quad H_1 : \gamma > 0 \quad . \tag{5.15}$$

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 Φ 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:

$$G_h(\cdot) = \sum_{n=1}^{h}(-1)^n \frac{\gamma(\gamma+1)(\gamma+2)\ldots(\gamma+n-1)}{n!}[(y_{t-d}-c)^2+k(y_{t-d}-c)]^n+O(.), \quad \gamma > 0. \tag{5.16}$$

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

$$y_t = \sum_{i=1}^{p}\phi_i y_{t-i}+\sum_{j=1}^{p}\delta_{j,0}y_{t-j}+\sum_{j=1}^{p}\delta_{j,1}y_{t-j}y_{t-d}+\sum_{j=1}^{p}\delta_{j,2}y_{t-j}y_{t-d}^2+\cdots+\sum_{j=1}^{p}\delta_{j,2h}y_{t-j}y_{t-d}^{2h}+\xi_t \quad . \tag{5.17}$$

The error terms in the regression (5.17) are now denoted by ξ_t rather than ε_t. After approximating (5.8) by (5.17), to test linearity against nonlinearity, we only need to test the nullity of parameters $\delta_{j,.}$. 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):

$$y_t = \psi_1 y_{t-1} + \varphi_1 y_{t-1}[1 - (1 + (y_{t-1} - c)^2 + k(y_{t-1} - c))^{-\gamma}] + \varepsilon_t, \tag{5.18}$$

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

$$y_t = \psi_1 y_{t-1} + \varphi_1 y_{t-1}\big[\gamma x - \frac{1}{2}\gamma(\gamma+1)x^2 + \frac{1}{6}\gamma(\gamma+1)(\gamma+2)x^3\big] + u_t, \tag{5.19}$$

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

$$y_t = \psi_1 y_{t-1} + \delta_{1,0} y_{t-1} + \delta_{1,1} y_{t-1}^2 + \delta_{1,2} y_{t-1}^3 + \delta_{1,3} y_{t-1}^4 + \delta_{1,4} y_{t-1}^5 + \delta_{1,5} y_{t-1}^6 + \delta_{1,6} y_{t-1}^7 + u_t, \tag{5.20}$$

where

$\delta_{1,0} = \varphi_1[m_1(c^2 - kc) + m_2(c^4 - 2kc^3 + kc^2) + m_3(c^6 - 3kc^5 + 3k^2c^3 + k^3c^3)]$,
$\delta_{1,1} = \varphi_1[m_1(k - 2c) + m_2(6kc^2 - 4c^3 - 2kc) + m_3(-6c^5 + 15kc^4 + 9k^2c^2 + 3k^3c^2)]$,
$\delta_{1,2} = \varphi_1[m_1 + m_2(6c^2 - 6kc + k) + m_3(15c^4 - 30kc^3 - 9k^2c - 3ck^3)]$,
$\delta_{1,3} = \varphi_1[m_2(2k - 4c) + m_3(-20c^3 + 30c^2k + 3k^2 + k^3)]$,
$\delta_{1,4} = \varphi_1[m_2 + m_3(15c^2 - 15kc)]$,
$\delta_{1,5} = \varphi_1[m_3(3k - 6c)]$,
$\delta_{1,6} = \varphi_1[m_3]$,
$m_1 = \gamma$,
$m_2 = -\frac{1}{2}\gamma(\gamma+1)$,
$m_3 = \frac{1}{6}\gamma(\gamma+1)(\gamma+2)$.

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

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

We suggest to use the F-version of the LM test statistics since it has better size properties than the χ^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 $y_{t-1}y_{t-1}^i$ for $i = 1, \cdots, 6$ and compute its sum of squared residuals SSR_1.

3. Under the null (5.21) the test statistic

$$F = \frac{(SSR_0 - SSR_1)/12}{SSR_1/(T-14)}, \tag{5.22}$$

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 $(y_t)_t$ defined in (5.13); a centered process $(z_t)_t = y_t - \mu$; and a demeaned and detrended process, $(w_t)_t = y_t - \mu - \alpha t$. In the following, we restrict $c=0$ and thus the process $(y_t)_t$ becomes:

$$\Delta y_t = \varphi y_{t-1}[1-(1+y_{t-1}^2+ky_{t-1})^{-\gamma}] + \sum_{i=1}^{p}\rho_i \Delta y_{t-i} + u_t, \tag{5.23}$$

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, γ, which is zero under the null and positive under the alternative. Hence we test

$$H_0 : \gamma = 0 \quad vs. \quad H_1 : \gamma > 0. \tag{5.24}$$

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 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

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

by using the statistic

$$\tilde{F}_{NL} = (R(\hat{\beta} - \beta))' \left[\hat{\sigma}_T^2 R (\sum_{t=1}^{T} X_t X_t')^{-1} R' \right]^{-1} R(\hat{\beta} - \beta), \tag{5.27}$$

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

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

Before providing the non-standard limiting distribution of $\tilde{F}_{NL}$ 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 $(\gamma_j)_j = Cov(\Delta y_t, \Delta y_{t-j})$*, then the following results hold:*

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

(b) $\frac{1}{T^{\frac{n+1}{2}}} \sum_{t=1}^{T} y_{t-1}^n u_t \longrightarrow_d [\Psi(1)]^n \sigma^{n+1} \int_0^1 [W(r)]^n dW(r).$

(c) $\frac{1}{T} \sum_{t=1}^{T} \Delta y_{t-i} \cdot \Delta y_{t-j} \longrightarrow_d \gamma_{|i-j|}.$

(d) $\frac{1}{T^{\frac{n}{2}+1}} \sum_{t=1}^{T} y_{t-1}^n \cdot \Delta y_{t-j} \longrightarrow_d 0.$

(e) $\frac{1}{T^{\frac{1}{2}}} \sum_{t=1}^{T} \Delta y_{t-i} u_t \longrightarrow_d N(0, \sigma^2 \gamma_0), \quad i \geq 1.$

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* β *and converges to its true value with rate of convergence*

$$diag(T^2, T^3, T^4, T^{1/2}, T^{1/2}, \cdots, T^{1/2})_{(p+3)\times(p+3)},$$

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

$$\tilde{F}_{NL} \longrightarrow_d v' Q^{-1} v \tag{5.29}$$

with

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

and

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

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* $\psi = 1$

$$\Delta y_t = \delta_{1,2} y_{t-1}^3 + \delta_{1,4} y_{t-1}^5 + \delta_{1,6} y_{t-1}^7 + u_t, \tag{5.30}$$

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

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

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

$$\bar{W}(r) = W(r) - \int_0^1 W(r)dr.$$

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

$$w_t = y_t - \hat{\mu} - \hat{\alpha}t, \tag{5.32}$$

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

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

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

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

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;
4. the [39] unit root test.

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 $1,000,000$.

Fractile (%)	Case 1	Case 2	Case 3
1	4.7217	5.4746	6.6019
5	3.4594	4.1309	5.1323
10	2.8856	3.5145	4.4439

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):

$$y_t = \psi y_{t-1} + \varphi y_{t-1}(1 - (1 + y_{t-1}^2 + k y_{t-1})^{-\gamma}) + \epsilon_t, \tag{5.34}$$

	Case 1					Case 2					Case 3				
T	ABG	KSS	W_{nl}	F_{NL}	DF	ABG	KSS	W_{nl}	F_{NL}	DF	ABG	KSS	W_{nl}	F_{NL}	DF
50	3.812	4.562	4.446	5.146	5.176	3.792	4.986	3.764	4.726	6.408	4.708	5.154	3.852	6.738	7.49
100	3.852	4.812	4.654	5	4.954	3.53	4.732	4.096	4.696	5.636	3.964	4.888	4.114	6.916	6.214
150	3.914	4.918	4.556	4.882	4.95	3.688	4.868	4.624	4.916	5.522	3.656	4.77	4.366	7.254	5.644
200	3.984	5.048	4.688	5.01	5.158	3.682	4.922	4.546	4.836	5.288	3.718	5.066	4.674	7.518	5.67
500	4.466	4.854	4.862	5.094	4.772	4.21	4.88	4.79	4.766	5.07	3.912	4.936	4.91	8.042	5.258
1000	4.718	4.86	4.9	5.094	4.764	4.708	5.068	5.08	4.84	5.218	4.386	5.14	5.062	8.15	5.068
5000	5.084	5.07	5.154	5.168	5.008	5.118	4.896	4.994	4.772	5.108	4.746	4.762	4.78	7.78	5.082

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

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

$\varphi = \{-1.8, -1.5, -1.2, -0.9, -0.6, -0.3, -0.05\}$.
$\gamma = \{0.5, 0.8, 1\}, \quad \psi = 1.$
$\sigma = \{0.1, 1\}, \quad k = \{0, 1, 4\}.$
$T = \{50, 100, 200, 500, 1000\}.$

The values of φ 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 φ 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, γ and σ. The empirical power for each test, we considered, increases as the value of the transition variable γ increases. When the DGP is approximately linear, corresponding to φ very small or large γ, there is no real power gain using a test different than the simple DF. This is not surprising because as γ increases, the model becomes approximately linear. By considering the range of φ values from -0.05 to -1.8, we observe that the empirical power increases with an increase in the magnitude of φ. Thus, the power of the tests decreases when the difference between the regimes is very small, corresponding to a small value of φ. 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 γ, φ, 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 γ and/or φ, 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, φ and γ. 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 φ 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

	Case 1, $\sigma_\epsilon = 0.1$, $\gamma = 0.8$									
	$k = 0$					$k = 4$				
	ABG	KSS	W_{nl}	F_{NL}	DF	ABG	KSS	W_{nl}	F_{NL}	DF
$\varphi = -1.8$										
$T = 50$	18.334	55.466	17.752	18.030	39.908	96.410	66.136	94.048	94.596	99.604
$T = 100$	61.704	96.808	65.500	60.772	97.168	97.300	64.334	95.000	95.584	99.770
$T = 200$	99.884	100.000	99.714	99.540	100.000	97.138	61.974	94.900	95.314	99.832
$T = 500$	100.000	100.000	100.000	100.000	100.000	96.832	62.092	94.900	95.288	99.902
$\varphi = -1.5$										
$T = 50$	14.912	47.168	14.046	14.790	32.562	93.182	55.956	92.008	93.394	98.890
$T = 100$	50.842	94.156	54.800	50.356	93.306	93.994	59.692	92.908	94.176	99.324
$T = 200$	99.412	99.990	99.056	98.564	100.000	94.020	60.214	92.900	93.956	99.474
$T = 500$	100.000	100.000	100.000	100.000	100.000	94.014	62.064	93.066	93.818	99.658
$\varphi = -1.2$										
$T = 50$	12.160	38.162	11.276	12.174	25.236	90.256	53.132	89.566	89.826	98.294
$T = 100$	38.910	88.974	42.518	39.168	84.570	92.148	57.226	92.280	92.232	98.882
$T = 200$	97.212	99.966	96.946	95.624	100.000	92.462	58.334	92.460	92.490	99.138
$T = 500$	100.000	100.000	100.000	100.000	100.000	92.426	58.944	92.138	92.332	99.522
$\varphi = -0.9$										
$T = 50$	9.046	28.252	8.300	9.490	18.950	79.754	43.066	87.028	86.726	97.266
$T = 100$	27.218	79.104	29.582	27.632	68.802	86.276	45.164	88.400	87.768	98.126
$T = 200$	89.042	99.806	90.188	87.080	99.968	89.494	47.618	89.294	88.910	98.756
$T = 500$	100.000	100.000	100.000	100.000	100.000	90.330	50.064	89.350	89.074	99.252
$\varphi = -0.6$										
$T = 50$	7.058	18.916	6.238	7.112	13.698	73.700	41.984	83.470	88.354	95.354
$T = 100$	16.896	61.088	17.960	17.758	45.768	83.424	46.126	90.028	90.358	97.380
$T = 200$	65.764	98.322	70.268	65.594	98.940	89.294	49.134	90.924	90.612	98.216
$T = 500$	100.000	100.000	99.998	100.000	100.000	90.914	50.886	91.002	90.648	98.768
$\varphi = -0.3$										
$T = 50$	5.064	10.838	4.430	5.378	9.808	36.028	20.034	68.686	78.984	71.222
$T = 100$	8.430	31.386	8.484	9.254	21.950	52.336	31.080	91.158	94.718	98.094
$T = 200$	27.872	83.322	31.502	29.494	76.128	75.562	42.312	96.088	96.170	98.932
$T = 500$	99.468	99.992	98.956	98.440	100.000	94.630	54.746	96.356	96.344	99.158
$\varphi = -0.05$										
$T = 50$	3.920	5.560	3.384	4.344	6.972	4.126	2.008	2.036	4.028	4.742
$T = 100$	3.972	7.960	3.730	4.438	8.190	4.442	3.396	4.982	6.722	13.226
$T = 200$	5.326	16.500	5.492	5.944	13.642	6.200	9.336	22.170	23.956	52.726
$T = 500$	18.798	67.800	20.350	19.674	60.666	32.094	43.766	82.326	88.778	99.584

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 $\sigma_\epsilon = 0.1$.

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.

	Case 1, $\sigma_\epsilon = 1$, $\gamma = 0.8$									
	$k = 0$					$k = 4$				
	ABG	KSS	W_{nl}	F_{NL}	DF	ABG	KSS	W_{nl}	F_{NL}	DF
$\varphi = -1.8$										
$T = 50$	100.000	100.000	99.992	99.992	100.000	93.566	55.088	87.526	90.858	99.862
$T = 100$	100.000	100.000	100.000	100.000	100.000	97.410	68.862	95.034	96.090	99.902
$T = 200$	100.000	100.000	100.000	100.000	100.000	98.808	75.950	97.694	98.092	99.954
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.478	75.946	98.814	98.994	99.972
$\varphi = -1.5$										
$T = 50$	99.996	100.000	99.894	99.930	100.000	99.542	90.836	98.960	99.640	99.904
$T = 100$	100.000	100.000	100.000	100.000	100.000	99.706	89.756	99.576	99.756	99.960
$T = 200$	100.000	100.000	100.000	100.000	100.000	99.712	88.240	99.646	99.710	99.946
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.746	86.386	99.652	99.732	99.964
$\varphi = -1.2$										
$T = 50$	99.944	99.986	99.144	99.386	100.000	99.136	89.160	98.296	99.162	99.850
$T = 100$	100.000	100.000	100.000	100.000	100.000	99.160	87.160	99.160	99.220	99.850
$T = 200$	100.000	100.000	100.000	100.000	100.000	99.164	84.692	99.234	99.264	9.878
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.108	80.336	99.158	99.184	99.910
$\varphi = -0.9$										
$T = 50$	98.978	99.720	94.036	94.852	100.000	99.154	91.556	97.110	98.924	99.828
$T = 100$	100.000	100.000	99.992	99.996	100.000	99.346	90.230	99.192	99.346	99.888
$T = 200$	100.000	100.000	100.000	100.000	100.000	99.222	88.098	99.264	99.272	99.878
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.256	84.372	99.358	99.292	99.942
$\varphi = -0.6$										
$T = 50$	84.300	96.182	71.718	72.598	99.884	96.528	94.540	84.400	91.350	99.944
$T = 100$	99.980	99.980	99.316	99.516	100.000	99.764	97.270	99.170	99.738	99.964
$T = 200$	100.000	100.000	100.000	100.000	100.000	99.822	96.976	99.794	99.838	99.974
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.810	95.786	99.820	99.826	99.982
$\varphi = -0.3$										
$T = 50$	30.430	68.462	26.698	26.982	87.154	51.754	73.460	39.262	42.960	97.300
$T = 100$	87.190	96.944	76.336	76.890	99.986	95.204	95.494	83.882	89.856	99.980
$T = 200$	100.000	99.972	99.430	99.650	100.000	99.944	99.162	98.964	99.728	99.988
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.960	99.124	99.956	99.960	99.998
$\varphi = -0.05$										
$T = 50$	4.236	11.708	3.606	4.518	13.928	4.758	12.612	4.004	4.984	17.524
$T = 100$	6.616	24.898	6.690	7.144	29.094	7.658	23.880	7.370	8.032	34.026
$T = 200$	18.366	54.328	19.010	18.806	71.162	19.534	50.618	18.164	19.350	74.406
$T = 500$	82.794	93.840	70.590	72.246	99.978	78.850	92.284	66.512	68.278	99.968

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 τ 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

	Case 2, $\sigma_\epsilon = 0.1$, $\gamma = 0.8$									
	$k = 0$					$k = 4$				
	ABG	KSS	W_{nl}	F_{NL}	DF	ABG	KSS	W_{nl}	F_{NL}	DF
$\varphi = -1.8$										
$T = 50$	15.600	19.974	18.876	16.246	18.214	94.394	51.776	88.536	91.082	95.900
$T = 100$	44.062	61.472	63.272	54.662	49.240	97.236	55.986	94.204	95.232	98.312
$T = 200$	98.404	98.584	99.574	99.080	99.768	97.094	54.808	94.662	95.212	99.122
$T = 500$	100.000	100.000	100.000	100.000	100.000	96.826	56.850	94.786	95.206	99.578
$\varphi = -1.5$										
$T = 50$	13.410	16.710	15.358	13.470	16.218	92.654	45.814	91.134	92.704	98.376
$T = 100$	35.100	51.836	53.052	44.374	39.950	93.790	53.334	92.646	93.936	98.940
$T = 200$	95.170	96.756	98.632	97.400	98.716	93.918	56.534	92.766	93.800	99.206
$T = 500$	100.000	100.000	100.000	100.000	100.000	93.992	59.938	93.008	93.692	99.540
$\varphi = -1.2$										
$T = 50$	11.500	13.782	12.810	11.374	14.652	89.336	43.350	87.584	87.986	97.420
$T = 100$	26.986	41.118	41.406	33.870	31.282	91.936	50.290	91.734	91.876	98.344
$T = 200$	87.114	92.968	95.966	93.106	94.526	92.336	54.394	92.166	92.238	98.770
$T = 500$	100.000	100.000	100.000	100.000	100.000	92.388	56.650	92.040	92.134	99.298
$\varphi = -0.9$										
$T = 50$	9.446	11.056	9.758	8.958	12.984	83.234	47.022	87.330	87.544	96.816
$T = 100$	19.408	29.802	29.692	23.842	23.808	85.500	47.466	88.122	88.254	97.756
$T = 200$	70.276	84.198	88.092	81.948	80.784	88.766	48.604	88.944	89.066	98.464
$T = 500$	100.000	99.994	100.000	100.000	100.000	90.062	49.438	88.986	88.908	98.972
$\varphi = -0.6$										
$T = 50$	7.676	8.884	7.750	7.342	11.132	86.246	53.232	87.758	89.772	94.518
$T = 100$	13.460	19.518	19.134	15.524	17.928	87.190	56.798	90.850	91.018	97.246
$T = 200$	44.526	64.536	67.628	58.218	53.982	89.170	56.676	91.340	91.308	98.114
$T = 500$	99.992	99.888	99.998	99.994	100.000	90.552	54.948	91.100	91.134	98.610
$\varphi = -0.3$										
$T = 50$	6.084	6.652	5.814	5.598	9.018	72.944	29.294	68.744	76.366	46.208
$T = 100$	8.530	11.514	10.532	8.812	12.668	93.124	51.438	93.998	95.718	95.430
$T = 200$	19.488	30.628	31.398	25.016	25.528	95.144	70.954	96.130	96.486	98.778
$T = 500$	94.834	96.096	98.448	97.000	98.800	96.156	72.436	96.550	96.562	99.080
$\varphi = -0.05$										
$T = 50$	4.266	4.706	4.094	4.814	6.650	4.780	3.398	3.998	4.410	6.744
$T = 100$	4.702	5.980	5.284	5.120	7.266	6.890	4.598	7.860	6.898	9.420
$T = 200$	6.044	8.262	7.732	6.596	9.280	18.632	8.750	25.998	21.192	23.202
$T = 500$	14.898	21.824	21.464	16.922	20.176	86.528	46.846	86.200	83.868	94.512

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, $\sigma_\epsilon = 1$, $\gamma = 0.8$									
	$k = 0$					$k = 4$				
	ABG	KSS	W_{nl}	F_{NL}	DF	ABG	KSS	W_{nl}	F_{NL}	DF
$\varphi = -1.8$										
$T = 50$	100.000	99.998	99.992	99.986	100.000	93.380	44.874	83.654	88.498	99.754
$T = 100$	100.000	100.000	100.000	100.000	100.000	97.398	62.134	91.826	94.780	99.856
$T = 200$	100.000	100.000	100.000	100.000	100.000	98.816	71.424	95.450	97.438	99.898
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.490	72.652	97.676	98.872	99.942
$\varphi = -1.5$										
$T = 50$	99.990	99.902	99.876	99.874	100.000	99.514	88.686	98.004	99.588	99.828
$T = 100$	100.000	100.000	100.000	100.000	100.000	99.702	87.570	98.874	99.746	99.902
$T = 200$	100.000	100.000	100.000	100.000	100.000	99.714	85.874	99.226	99.710	99.914
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.746	83.802	99.486	99.722	99.942
$\varphi = -1.2$										
$T = 50$	99.836	99.184	99.044	98.950	99.996	99.072	86.096	97.312	99.036	99.752
$T = 100$	100.000	100.000	100.000	100.000	100.000	99.108	84.272	98.644	99.138	99.768
$T = 200$	100.000	100.000	100.000	100.000	100.000	99.156	81.338	99.032	99.180	99.814
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.092	76.376	99.054	99.154	99.874
$\varphi = -0.9$										
$T = 50$	97.266	94.850	93.590	92.588	99.768	99.106	88.960	96.078	98.622	99.758
$T = 100$	100.000	99.996	99.994	99.988	100.000	99.310	88.526	98.662	99.338	99.838
$T = 200$	100.000	100.000	100.000	100.000	100.000	99.198	85.926	99.140	99.260	99.832
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.230	81.338	99.226	99.284	99.904
$\varphi = -0.6$										
$T = 50$	75.058	74.764	70.874	66.694	92.870	97.524	86.310	86.656	87.924	99.870
$T = 100$	99.918	99.226	99.276	99.048	100.000	99.748	96.276	98.848	99.680	99.926
$T = 200$	100.000	99.998	100.000	100.000	100.000	99.826	96.416	99.600	99.842	99.964
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.806	95.018	99.770	99.828	99.972
$\varphi = -0.3$										
$T = 50$	25.490	31.362	27.454	23.492	41.902	49.620	45.660	42.352	38.732	77.556
$T = 100$	77.288	77.828	75.724	70.570	95.766	95.354	86.178	87.780	86.540	99.846
$T = 200$	99.976	99.222	99.420	99.292	100.000	99.898	97.488	99.132	99.556	99.982
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.958	98.926	99.904	99.960	99.992
$\varphi = -0.05$										
$T = 50$	5.202	5.968	5.042	5.000	7.950	5.372	5.960	5.126	5.302	8.528
$T = 100$	6.796	9.538	8.608	7.118	11.426	7.498	10.272	9.472	7.992	13.206
$T = 200$	14.936	22.336	20.854	16.852	27.400	16.972	22.478	21.824	18.430	32.604
$T = 500$	71.856	71.446	71.420	66.822	94.010	70.022	67.552	68.016	63.664	95.160

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 $q_t \equiv log(\frac{CPI_t^{home}}{CPI_t^{US} S_t})$, 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 ($= -2.7209$) which are significant at 5% and F_{NL} ($= 3.1266$) which is significant at 10%

	Case 3, $\sigma_\epsilon = 0.1$, $\gamma = 0.8$									
	$k = 0$					$k = 4$				
	ABG	KSS	W_{nl}	F_{NL}	DF	ABG	KSS	W_{nl}	F_{NL}	DF
$\varphi = -1.8$										
$T = 50$	12.446	13.014	12.020	16.894	14.740	93.834	46.014	83.366	89.204	94.182
$T = 100$	28.536	36.546	37.420	45.380	30.878	96.958	52.930	92.404	94.550	97.760
$T = 200$	90.210	91.894	94.066	96.320	93.446	97.286	52.762	94.556	95.464	98.938
$T = 500$	100.000	100.000	100.000	100.000	100.000	96.888	54.572	94.660	95.212	99.540
$\varphi = -1.5$										
$T = 50$	10.686	11.236	10.468	15.102	13.340	92.438	42.210	89.200	92.452	98.170
$T = 100$	22.674	29.114	29.690	37.246	25.424	93.826	49.296	92.404	94.006	98.768
$T = 200$	80.418	85.284	88.324	92.218	84.710	94.058	55.022	92.786	94.036	99.150
$T = 500$	100.000	99.998	100.000	100.000	100.000	93.930	59.036	92.878	93.692	99.472
$\varphi = -1.2$										
$T = 50$	9.374	9.718	8.860	13.182	12.284	88.590	39.726	84.208	87.472	96.826
$T = 100$	18.068	22.536	22.664	29.462	21.230	91.646	46.622	90.760	91.672	98.042
$T = 200$	65.410	74.510	78.024	84.288	70.620	92.304	51.602	91.812	92.130	98.474
$T = 500$	100.000	99.990	99.996	100.000	100.000	92.356	55.450	91.928	92.124	99.116
$\varphi = -0.9$										
$T = 50$	7.850	7.904	7.228	11.032	10.674	82.050	44.378	85.398	87.018	96.110
$T = 100$	13.462	16.708	16.370	22.180	17.130	84.650	45.736	87.844	88.530	97.544
$T = 200$	46.000	58.148	61.104	69.876	51.140	88.268	46.856	88.740	89.082	98.180
$T = 500$	100.000	99.906	99.984	99.996	100.000	90.098	48.822	89.260	89.022	98.716
$\varphi = -0.6$										
$T = 50$	6.744	6.776	6.098	9.380	9.292	86.428	51.274	86.564	89.836	92.014
$T = 100$	9.736	11.588	11.024	15.826	13.398	86.278	54.870	89.286	91.138	96.898
$T = 200$	26.740	36.392	37.888	47.072	31.794	88.184	54.616	90.876	91.060	97.722
$T = 500$	99.682	98.686	99.520	99.766	99.924	90.376	53.292	91.000	91.104	98.484
$\varphi = -0.3$										
$T = 50$	5.712	5.542	4.774	7.856	8.082	64.518	23.602	53.268	72.628	33.138
$T = 100$	6.878	8.006	7.476	11.286	9.784	91.166	42.286	88.636	94.912	84.404
$T = 200$	12.894	17.018	17.052	23.682	17.392	94.570	64.922	95.514	96.404	98.416
$T = 500$	76.486	81.586	85.728	91.034	83.026	95.916	70.222	96.424	96.516	98.994
$\varphi = -0.05$										
$T = 50$	4.742	5.066	4.664	6.830	7.018	5.444	3.970	4.142	7.202	6.990
$T = 100$	4.340	5.066	4.664	7.498	6.486	5.288	3.782	5.046	8.358	7.078
$T = 200$	4.772	6.260	5.994	9.592	7.238	10.930	5.336	13.102	19.190	13.246
$T = 500$	9.848	12.548	12.816	18.262	13.318	68.786	24.050	67.674	77.486	73.070

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.

	Case 3, $\sigma_\epsilon = 1$, $\gamma = 0.8$									
	$k = 0$					$k = 4$				
	ABG	KSS	W_{nl}	F_{NL}	DF	ABG	KSS	W_{nl}	F_{NL}	DF
$\varphi = -1.8$										
$T = 50$	99.996	99.938	99.942	99.978	100.000	92.812	38.216	79.406	86.704	99.726
$T = 100$	100.000	100.000	100.000	100.000	100.000	97.166	57.868	89.484	94.046	99.808
$T = 200$	100.000	100.000	100.000	100.000	100.000	98.758	68.608	94.186	97.132	99.920
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.408	71.008	97.132	98.796	99.926
$\varphi = -1.5$										
$T = 50$	99.966	99.448	99.428	99.762	99.996	99.552	87.166	96.700	99.520	99.798
$T = 100$	100.000	100.000	100.000	100.000	100.000	99.634	86.662	98.270	99.662	99.844
$T = 200$	100.000	100.000	100.000	100.000	100.000	99.682	84.584	98.784	99.680	99.886
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.756	82.222	99.306	99.730	99.912
$\varphi = -1.2$										
$T = 50$	99.394	96.852	96.748	98.398	99.912	98.936	83.904	95.560	98.906	99.596
$T = 100$	100.000	100.000	100.000	100.000	100.000	99.148	82.624	98.064	99.228	99.788
$T = 200$	100.000	100.000	100.000	100.000	100.000	99.126	79.486	98.676	99.178	99.798
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.040	74.832	98.884	99.092	99.868
$\varphi = -0.9$										
$T = 50$	92.614	85.202	83.938	89.646	98.232	98.956	85.650	92.534	98.248	99.704
$T = 100$	100.000	99.874	99.902	99.970	100.000	99.208	87.166	97.780	99.262	99.798
$T = 200$	100.000	100.000	100.000	100.000	100.000	99.254	84.946	98.922	99.316	99.830
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.256	80.224	99.178	99.316	99.914
$\varphi = -0.6$										
$T = 50$	60.686	55.322	52.644	61.994	76.888	94.372	74.396	74.958	85.420	99.504
$T = 100$	99.444	96.294	96.400	98.230	99.994	99.714	94.926	97.576	99.636	99.920
$T = 200$	100.000	99.996	99.998	100.000	100.000	99.776	95.654	99.226	99.800	99.946
$T = 500$	100.000	100.000	100.000	100.000	100.000	99.752	94.534	99.650	99.774	99.960
$\varphi = -0.3$										
$T = 50$	18.790	18.622	16.910	22.832	27.428	37.500	29.088	27.768	36.712	57.810
$T = 100$	60.264	56.630	55.172	64.156	80.426	88.940	72.698	74.096	82.858	98.748
$T = 200$	99.740	96.272	97.064	98.604	99.998	99.790	95.084	97.578	99.326	99.984
$T = 500$	100.000	99.998	100.000	100.000	100.000	99.752	94.534	99.650	99.774	99.960
$\varphi = -0.05$										
$T = 50$	5.224	4.968	4.372	7.338	7.458	5.350	4.850	4.462	7.614	7.842
$T = 100$	5.610	6.606	6.196	9.422	8.734	6.298	6.968	6.632	10.212	10.020
$T = 200$	10.222	12.810	12.294	17.846	16.392	11.648	13.642	13.564	19.266	20.046
$T = 500$	53.084	50.890	51.634	61.846	75.572	53.362	47.866	49.350	59.950	79.434

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

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 τ (= 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

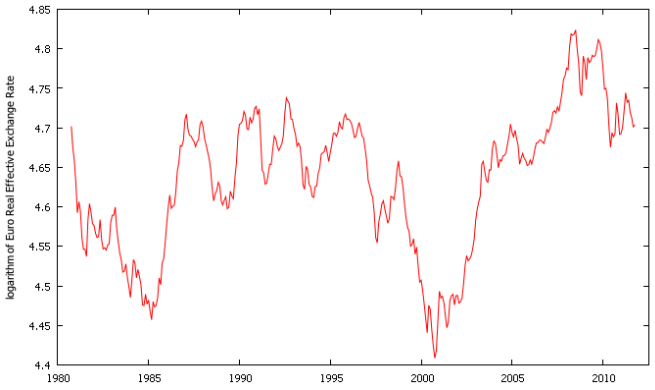


FIGURE 5.1: Logarithm of euro real effective exchange rates (October, 1980 to October, 2011)

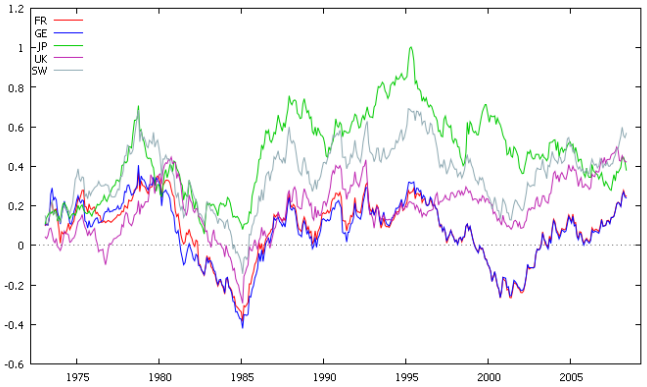


FIGURE 5.2: Normalized real exchange rates relative to the US dollar against UK, Japan, Germany, France and Switzerland (January, 1973 to June, 2008).

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), τ (= 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 τ (= 20.2723), all significant at 1% level.

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

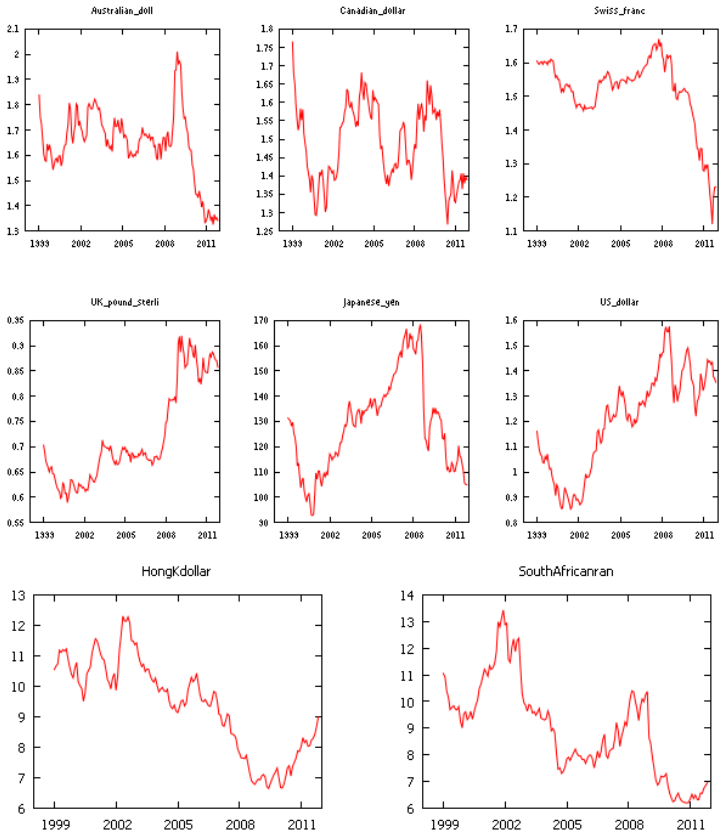


FIGURE 5.3: Bilateral real exchange rates relative to the euro (January, 1999 to November, 2011).

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 $d, p_1, \cdots, p_l \in Z^+$. Any solution of $(y_t)_t$ of

$$y_t + \sum_{j=1}^{l} y_{t,d}^{(j)} \left(a_0^{(j)} + \sum_{i=1}^{p_l} a_i^{(j)} y_{t-i} \right) = \sum_{j=1}^{l} y_{t,d}^{(j)} \varepsilon_t^{(j)} \tag{6.1}$$

where

$$y_{t,d}^{(j)} = \begin{cases} 1; & y_{t-d} \in R_j \\ 0; & y_{t-d} \notin R_j. \end{cases} \tag{6.2}$$

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 $(y_t)_t$ 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 σ-field, denoted as $\mathcal{F}_t$, i.e., $E[\varepsilon_t^{(j)}|\mathcal{F}_{t-1}] = 0$. In this setting, the conditional variance of the process $(\varepsilon_t)_t^{(j)}$ can be a constant, $E[(\varepsilon_t^{(j)})^2|\mathcal{F}_{t-1}] = \sigma^2$ 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 $(\mathbf{f}_t)_t$ 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 $\mathbf{y}_t = (y_{1t}, \cdots, y_{Dt})^T$ such that $L_1, \cdots, L_D \in Z^+$, for each $1 \leq i \leq D$, $(R_j^i)_{j=1,2,\cdots,L_i}$ a disjuction decomposition of the real axis: $R = \bigcup_{j=1}^{L_i} R_j^i \; ; i \in \{1, \cdots, D\}$. Let $L = \max\{L_1, L_2, \cdots, L_D\}$ and $R_j^i = \Phi \; ; j = L_i+1, \cdots, L$. Then any solution $(\mathbf{y}_t)_t$ of

$$\mathbf{y}_t + \sum_{J \in \{1,\cdots,L\}^D} y_{t,d}^{(J)} \left(a_0^{(J)} + \sum_{i=1}^{p_J} A_i^{(J)} \mathbf{y}_{t-i} \right) = \sum_{J \in \{1,\cdots,L\}^D} y_{t,d}^{(J)} \varepsilon_t^{(J)} \tag{6.3}$$

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

$$\left(y_t^{(j_1,\cdots,j_D)} = 1 \right) \Leftrightarrow_{def} \left((\mathbf{y}_{t-d})_i \in R_j^i; j \in (1, \cdots, L)^D; i \in (1, \cdots, D) \right)$$

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

$$\sup_{t \geq 0} E[\|\varepsilon_{t+1}^{(J)}\| | \mathcal{F}_t] = 0 \quad a.s, \quad \sup_{t \geq 0} E[\|\varepsilon_{t+1}^{(J)}\|^2 | \mathcal{F}_t] = \sigma^2 < \infty \quad a.s, \quad \sup_{t \geq 0} E[\|\varepsilon_{t+1}^{(J)}\|^\alpha | \mathcal{F}_t] < +\infty \quad a.s$$

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

Now consider a D-dimensional time series $\mathbf{y}_t = (y_{1t}, \cdots, y_{Dt})^T$ and a κ-dimensional inputs $\mathbf{f}_t = (f_{1t}, \cdots, f_{\kappa t})^T$ such that $L_1, \cdots, L_D \in Z^+$, for each $1 \leq i \leq D$, $(R_j^i)_{j=1,2,\cdots,L_i}$ a disjuction decomposition of the real axis: $R = \bigcup_{j=1}^{L_i} R_j^i \; ; \; i \in \{1, \cdots, D\}$. Let $L = \max\{L_1, L_2, \cdots, L_D\}$ be the maximum of the number of regimes for each subprocess of

$\mathbf{y}_t$ and $R_j^i = \Phi$; $j = L_i + 1, \cdots, L$. Then any solution $(\mathbf{y}_t)_t$ of

$$\begin{cases} \mathbf{y}_t + \sum_{J\in\{1,\cdots,L\}^D} y_{t,d}^{(J)} \left(a_0^{(J)} + \sum_{i=1}^{p_J} A_i^{(J)} \mathbf{y}_{t-i} + \Lambda^{(J)} \mathbf{f}_t \right) = \sum_{J\in\{1,\cdots,L\}^D} y_{t,d}^{(J)} \varepsilon_t^{(J)} \\ \mathbf{f}_t = \sum_{\tau=1}^{q} \Xi_\tau \mathbf{f}_{t-\tau} + \eta_t \end{cases} \tag{6.4}$$

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

$$\mathbf{y}_t + \sum_{J\in\{1,\cdots,L\}^D} y_{t,d}^{(J)} \left(a_0^{(J)} + \sum_{i=1}^{p_J} A_i^{(J)} \mathbf{y}_{t-i} + \Lambda^{(J)} \sum_{\tau=1}^{q} \Xi_\tau \mathbf{f}_{t-\tau} \right) = \sum_{J\in\{1,\cdots,L\}^D} y_{t,d}^{(J)} \omega_t^{(J)} \tag{6.5}$$

where $\omega_t^{(J)} = \varepsilon_t^{(J)} - \Lambda^{(J)} \eta_t$, $a_0^{(J)}$ 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, $\Xi_\tau^{(J)}$ are $\kappa \times \kappa$ coefficient matrices, and $(f_t)_t$ is $\kappa \times 1$ vector. When $\Lambda^{(J)} = \mathbf{0}$ for all $J \in \{1, \cdots, L\}^D$, (6.5) becomes a MSETAR model (6.3).

The representation in equation (6.5) shows that the MSETARX $(L, p_J, q; J \in \{1, \cdots, L\}^D)$ model (6.4) has approximately the same structure as the MSETAR $(L, p_J; J \in \{1, \cdots, L\}^D)$ model (6.3) with exogenous variables or factors $(\mathbf{f}_t)_t$. 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 $(\mathbf{f}_t)_t$ 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_{it}, as in [141].

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

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

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

$$\sup_{t\geq 0} E[\|\omega_{t+1}^{(J)}\|^2 | \mathcal{F}_t] = \breve{\Upsilon}_\omega < \infty \quad a.s \quad .$$

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

$$\mathbf{y}_t = \sum_{J \in \{1,\cdots,L\}^D} y_{t,d}^{(J)} \left(\breve{\Theta}^{(J)}\right)^T \breve{\Phi}_{t-1} + \omega_t \tag{6.6}$$

where $\left(\breve{\Theta}^{(J)}\right)^T = -[a_0^{(J)}, A_1^{(J)}, \cdots, A_p^{(J)}, \Lambda^{(J)}\Xi_1, \Lambda^{(J)}\Xi_2, \cdots, \Lambda^{(J)}\Xi_q]$,
$\breve{\Phi}_t^T = [1, \mathbf{y}_t^T, \mathbf{y}_{t-1}^T, \cdots, \mathbf{y}_{t-p+1}^T, \mathbf{f}_t^T, \mathbf{f}_{t-1}^T, \cdots, \mathbf{f}_{t-q+1}^T]$ and the notation ζ^T denotes the transpose of ζ. 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

$$\mathbf{y}_t = \sum_{J \in \{1,\cdots,L\}^D} y_{t,d}^{(J)} \left(\left(\breve{\Theta}_1^{(J)}\right)^T \breve{\Phi}_{1,t-1} + \Lambda^{(J)} \mathbf{f}_t \right) + \sum_{J \in \{1,\cdots,L\}^D} y_{t,d}^{(J)} \varepsilon_t^{(J)} \tag{6.7}$$

$$\mathbf{f}_t = \left(\breve{\Theta}_2\right)^T \breve{\Phi}_{2,t-1} + \eta_t \tag{6.8}$$

where $\left(\breve{\Theta}_1^{(J)}\right)^T = -[A_1^{(J)}, \cdots, A_p^{(J)}]$, $\left(\breve{\Theta}_2\right)^T = -[\Xi_1, \Xi_2, \cdots, \Xi_q]$,
$\breve{\Phi}_{1,t}^T = [\mathbf{y}_t^T, \mathbf{y}_{t-1}^T, \cdots, \mathbf{y}_{t-p+1}^T]$ and $\breve{\Phi}_{2,t}^T = [\mathbf{f}_t^T, \mathbf{f}_{t-1}^T, \cdots, \mathbf{f}_{t-q+1}^T]$.

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

$$\begin{cases} \mathbf{y}_t = g_1^{(J)}(\mathbf{y}_{t-1}^T, \cdots, \mathbf{y}_{t-p}^T) + g_2^{(J)}(\mathbf{f}_t^T, \cdots, \mathbf{f}_{t-q}^T) + \sum_{J \in \{1,\cdots,L\}^D} y_{t,d}^{(J)} \varepsilon_t^{(J)} \\ \mathbf{f}_t = g_3(\mathbf{f}_{t-1}^T, \cdots, \mathbf{f}_{t-q}^T) + \eta_t \end{cases} \tag{6.9}$$

with $g_1^{(J)}(\mathbf{y}_{t-1}^T, \cdots, \mathbf{y}_{t-p}^T) = \sum_{J \in \{1,\cdots,L\}^D} y_{t,d}^{(J)} \left(\breve{\Theta}_1^{(J)}\right)^T \breve{\Phi}_{1,t-1}$, $g_2^{(J)}(\mathbf{f}_t^T, \cdots, \mathbf{f}_{t-q}^T) = \sum_{J \in \{1,\cdots,L\}^D} y_{t,d}^{(J)} \Lambda^{(J)} \mathbf{f}_t$, and $g_3(\mathbf{f}_{t-1}^T, \cdots, \mathbf{f}_{t-q}^T) = \left(\breve{\Theta}_2\right)^T \breve{\Phi}_{2,t-1}$. The process $\{\mathbf{f}_t, \mathbf{y}_t\}$ of the equation model (6.9) is a Markov process.

Assumption 2. We denote $\mathfrak{y} = (\mathbf{y}_{t-1}^T, \cdots, \mathbf{y}_{t-p}^T)$ and $\mathfrak{f} = (\mathbf{f}_t^T, \cdots, \mathbf{f}_{t-q}^T)$. The multivariate SETARX model (6.9) satisfies the following:

1. The functions $g_1^{(J)}(\mathfrak{y})$, $g_2^{(J)}(\mathfrak{f})$, and $g_3(\mathfrak{f})$ for each $J \in \{1, \cdots, L\}^D$ are nonperiodic and bounded on compact sets, and $g_2^{(J)}(\mathfrak{f}) = O(\|\mathfrak{f}\|^{\gamma_1})$ as $\|\mathfrak{f}\| \to \infty$ for some real γ_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 $\mathscr{A}^{(J)} = [\mathscr{A}_1^{(J)}, \mathscr{A}_2^{(J)}, \cdots, \mathscr{A}_p^{(J)}]$ and $\mathscr{B} = [\mathscr{B}_1, \mathscr{B}_2, \cdots, \mathscr{B}_{q-1}]$, each of which may be the zero matrix, for each $J \in \{1, \cdots, L\}^D$, where $\mathscr{A}_i^{(J)}$ and $\mathscr{B}_\tau$ are matrices of dimension $D \times D$ and $\kappa \times \kappa$ respectively such that $g_1^{(J)}(\mathfrak{y}) = \mathfrak{y}\big(\mathscr{A}^{(J)}\big)^T + o(\|\mathfrak{y}\|)$ and $g_3(\mathfrak{f}) = \mathfrak{f}\big(\mathscr{B}\big)^T + o(\|\mathfrak{f}\|)$ as $\|\mathfrak{y}\|$ and $\|\mathfrak{f}\| \to \infty$. Then the Dp-dimensional square matrix $\mathfrak{A}$ defined by $\mathbf{0}$ if $\mathscr{A}^{(J)} = 0$ and by

$$\mathfrak{A} = \begin{bmatrix} O_D & O_D & \cdots & O_D & \big(\mathscr{A}_1^{(J)}\big)^T \\ I_D & O_D & \cdots & O_D & \big(\mathscr{A}_2^{(J)}\big)^T \\ O_D & I_D & \cdots & O_D & \big(\mathscr{A}_3^{(J)}\big)^T \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O_D & O_D & \cdots & I_D & \big(\mathscr{A}_p^{(J)}\big)^T \end{bmatrix}$$

otherwise, and the κq-dimensional square matrix $\mathfrak{B}$ be defined by

$$\mathfrak{B} = \begin{bmatrix} O_\kappa & O_\kappa & \cdots & O_\kappa & \big(\mathscr{B}_1\big)^T \\ I_\kappa & O_\kappa & \cdots & O_\kappa & \big(\mathscr{B}_2\big)^T \\ O_\kappa & I_\kappa & \cdots & O_\kappa & \big(\mathscr{B}_3\big)^T \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O_\kappa & O_\kappa & \cdots & I_\kappa & \big(\mathscr{B}_q\big)^T \end{bmatrix}$$

satisfy $\varrho(\mathfrak{A}) < 1$ and $\varrho(\mathfrak{B}) < 1$, where ϱ denotes the spectral radius, O_ι denotes the ι-dimensional zero square matrix and I_ι denotes the ι-dimensional unit square matrix.

Lemma 6.1. *Under Assumption B.5, $\{\mathbf{f}_t, \mathbf{y}_t\}$ of the multivariate SETARX model (6.7)-(6.8) represented as a nonlinear ARX model (6.9) is α-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. □

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 the D-cycle) so that $A_1^{(j_{s+1})} = A_1^{(j_1)}$. The process $\{\mathbf{y}_t\}$ of the multivariate SETAR model (6.3) is geometrically ergodic if*

$$\varrho\Big(\prod_{s=1}^{D} -A_1^{(j_s)}\Big) < 1$$

where ϱ 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 $A_{i_s} = -A_1^{(j_s)}$ 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):

$$\mathbf{y}_t = \sum_{J \in \{1,\cdots,L\}^D} y_{t,d}^{(J)} \left(\breve{\Theta}^{(J)}\right)^T \breve{\Phi}_{t-1} + \omega_t \tag{6.10}$$

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

$$\hat{\breve{\Theta}}_{k+1}^{(J)} = \hat{\breve{\Theta}}_k^{(J)} + y_{k+1,d}^{(J)} R_k^{-1} \breve{\Phi}_k \left(\mathbf{y}_{k+1}^T - \breve{\Phi}_k^T \hat{\breve{\Theta}}_k^{(J)}\right) \tag{6.11}$$

$$R_k = \sum_{J\in\{1,\cdots,L\}^D} \sum_{i=0}^{k} y_{k+1,d}^{(J)} \breve{\Phi}_i \breve{\Phi}_i^T \tag{6.12}$$

6.4.2 Algorithm for Adaptive Parameter Estimation

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

$$\breve{\Theta}_k^{(J)} = 0; \quad k \leq p^*$$

$$\breve{\Theta}_{k+1}^{(J)} = \breve{\Theta}_k^{(J)} + y_{k+1,d}^{(J)} \frac{\alpha \breve{\Phi}_k}{s_k^{(J)}} \left(\mathbf{y}_{k+1}^T - \breve{\Phi}_k^T \breve{\Theta}_k^{(J)}\right); \quad k \geq p^*$$

$$r_k^{(J)} = \begin{cases} 1; & k < p^* \\ r_{k-1}^{(J)} + \sum_{J\in\{1,\cdots,L\}^D} y_{k+1,d}^{(J)} \|\breve{\Phi}_k\|^2; & k \geq p^*. \end{cases} \tag{6.13}$$

$$s_k^{(J)} = \begin{cases} 1; & k < p^* \\ s_{k-1}^{(J)} + \mathbf{y}_{k+1}^{(J)} \Big(\max\{\upsilon^{(J)} r_{k-1}^{(J)}, 1\} + \|\mathbf{y}_k\|^2 - s_{k-1}^{(J)} \Big); & k \geq p^*. \end{cases} \tag{6.14}$$

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 $\upsilon^{(J)}$ 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 $d = 6$, $\Lambda^{(J)} = \mathbf{0}$ 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:

   Regime 1 $R_1^i := [-\infty, -0.50) \times [-\infty, 0)$, $\mathscr{A}_1^{(1)} = \begin{pmatrix} -0.02 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$, $\mathscr{A}_2^{(1)} = \begin{pmatrix} 0.53 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$, $\mathscr{A}_3^{(1)} = \begin{pmatrix} 0.00 & 0.53 \\ 0.00 & 0.30 \end{pmatrix}$, $a_0^{(1)} = \begin{pmatrix} 0.74 \\ -0.20 \end{pmatrix}$

   Regime 2 $R_2^i := [-\infty, -0.50) \times [0, \infty)$, $\mathscr{A}_1^{(2)} = \begin{pmatrix} -0.02 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$, $\mathscr{A}_2^{(2)} = \begin{pmatrix} 0.53 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$, $\mathscr{A}_3^{(2)} = \begin{pmatrix} 0.00 & 0.53 \\ 0.00 & 0.30 \end{pmatrix}$, $a_0^{(2)} = \begin{pmatrix} -0.75 \\ -0.20 \end{pmatrix}$

   Regime 3 $R_3^i := [-0.50, 0.50) \times [-\infty, 0)$, $\mathscr{A}_1^{(3)} = \begin{pmatrix} -0.94 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$, $\mathscr{A}_2^{(3)} = \begin{pmatrix} 0.85 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$, $\mathscr{A}_3^{(3)} = \begin{pmatrix} 0.00 & 0.85 \\ 0.00 & 0.30 \end{pmatrix}$, $a_0^{(3)} = \begin{pmatrix} 1.15 \\ -0.20 \end{pmatrix}$

   Regime 4 $R_4^i := [-0.50, 0.50) \times (0.00, \infty)$, $\mathscr{A}_1^{(4)} = \begin{pmatrix} -0.94 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$, $\mathscr{A}_2^{(4)} = \begin{pmatrix} 0.85 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$, $\mathscr{A}_3^{(4)} = \begin{pmatrix} 0.00 & 0.85 \\ 0.00 & 0.30 \end{pmatrix}$, $a_0^{(4)} = \begin{pmatrix} 0.74 \\ 0.20 \end{pmatrix}$

   Regime 5 $R_5^i := [0.50, \infty) \times [-\infty, 0.00)$, $\mathscr{A}_1^{(5)} = \begin{pmatrix} -1.10 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$, $\mathscr{A}_2^{(5)} = \begin{pmatrix} -0.30 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$, $\mathscr{A}_3^{(5)} = \begin{pmatrix} 0.00 & -0.30 \\ 0.00 & 0.30 \end{pmatrix}$, $a_0^{(5)} = \begin{pmatrix} -0.75 \\ 0.20 \end{pmatrix}$

   Regime 6 $R_6^i := [0.50, \infty) \times [0.00, \infty)$, $\mathscr{A}_1^{(6)} = \begin{pmatrix} -1.10 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$, $\mathscr{A}_2^{(6)} = \begin{pmatrix} 0.30 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$, $\mathscr{A}_3^{(6)} = \begin{pmatrix} 0.00 & 0.30 \\ 0.00 & 0.30 \end{pmatrix}$, $a_0^{(6)} = \begin{pmatrix} 1.15 \\ 0.20 \end{pmatrix}$,

   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

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

$$\mathbf{y}_t = \begin{cases} \mathscr{A}_1^{(1)}\mathbf{y}_{t-1} + \Lambda^{(1)}\Xi_1\mathbf{f}_{t-1} + \omega_t; & (y_{t-1})_2 \leq -0.5 \\ \mathscr{A}_1^{(2)}\mathbf{y}_{t-1} + \Lambda^{(2)}\Xi_1\mathbf{f}_{t-1} + \omega_t; & (y_{t-1})_2 \in (-0.5, 0.5] \\ \mathscr{A}_1^{(3)}\mathbf{y}_{t-1} + \Lambda^{(3)}\Xi_1\mathbf{f}_{t-1} + \omega_t; & (y_{t-1})_2 \geq 0.5 \end{cases} \tag{6.15}$$

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

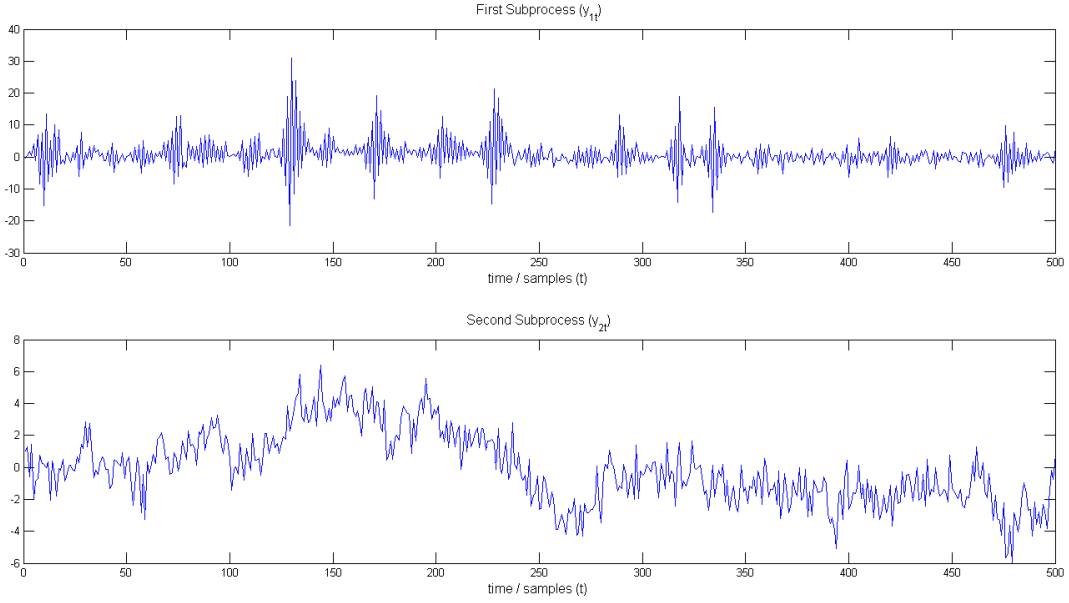


FIGURE 6.1: Two-dimensional MSETARX process with six-regimes, delay $d = 6$, autoregressive order $p = 3$ and $\Lambda^{(J)} = \mathbf{0}$ for all $J \in \{1, \cdots, L\}^D$ in equation (6.5). This is a signal section of 500 time samples of the multivariate process.

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 $d^* = argmax\{\sum_i^D \mathscr{C}^{(i)}(d) \mid d \in \{1, \cdots, d_{max}\}\}$ where $\mathscr{C}^{(i)}(d)$ 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 A

Appendix for Chapter 2

A.1 Wavelet Analysis

A.1.1 Notation and Operators

Consider the mapping

$$\Gamma : \mathcal{L}^2(\mathbb{R}) \longrightarrow \mathcal{L}^2(\mathbb{R})$$

Let $f \in \mathcal{L}^2(\mathbb{R})$, $\alpha, \beta \in \mathbb{R}$ 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.

Operator, Γ	Notation, Γf	Output	Inverse, $\Gamma * f = \Gamma^{-1} f$	Fourier transform, $(\Gamma f)\hat{}$
Dilation	$(D_s f)(t)$	$\frac{1}{s^{1/2}} f(\frac{t}{s})$	$D_{s^{-1}} f$	$D_{s^{-1}} \hat{f}$
Involution	$\tilde{f}$	$\bar{f}(-t)$	$\tilde{f}$	$\bar{\hat{f}}$
Translation	$(\tau_\alpha f)(t)$	$f(t-\alpha)$	$\tau_{-\alpha} f$	$e_{-a} \hat{f}$
Modulation	$(e_\alpha f)(t)$	$e^{i2\pi\alpha t} f(t)$	$e_{-\alpha} f$	$\tau_\alpha \hat{f}$
Reflection	$(Rf)(t)$	$f(-t)$	Rf	$R\hat{f}$

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

due to it's 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 ψ

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

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

$$\begin{aligned} C_\psi : \mathcal{L}^2(\mathbb{R}) &\longrightarrow C_\psi(\mathcal{L}^2(\mathbb{R})) \\ f &\mapsto \langle f, \tau_t D_s \psi \rangle_{\mathcal{L}^2(\mathbb{R})} \end{aligned} \tag{A.2}$$

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

$$(\tau_t D_s \psi)(\xi) = \frac{1}{|s|^{\frac{1}{2}}} \psi\big(\frac{\xi - t}{s}\big) \tag{A.3}$$

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 ψ, the CWT may be thought of in terms of the representation of a signal with respect to the wavelet family generated by ψ, that is, all it's translated and dilated versions. The CWT may be written as

$$(C_\psi f)(t, s) := \langle f, \tau_t D_s \psi \rangle \tag{A.4}$$

For each point (t, s) in the time-scale plan, 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 ψ.

The CWT of a signal f is defined as

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

where $\bar{\psi}(\xi)$ is the complex conjugate of the analyzing wavelet function $\psi(\xi)$. Given that ψ is chosen with enough time-frequency localization[1], the CWT gives a gives a picture of the time-frequency characteristics of the function f over the whole time-scale plane $\mathbb{R} \times (\mathbb{R} \backslash \{0\})$. 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 $TF(\mathbb{R}) := \{\varphi \in \mathcal{L}^2(\mathbb{R} : |\varphi(t)| < \eta(1 + |t|)^{-(1+\varepsilon)} \quad and \quad |\hat{\varphi}(\gamma)| < \eta(1 + |\gamma|)^{-(1+\varepsilon)} for \quad \eta < \infty, \varepsilon > 0\}$

transformation via the relation known as *Calderón's reproducing identity*,

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

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

$$f(\xi) = \frac{1}{C_{adm_+,\psi}} \int_{-\infty}^{\infty} \int_0^{\infty} \langle f, \tau_t D_s \psi \rangle \tau_t D_s \psi(\xi) \frac{1}{s^2} ds dt. \tag{A.7}$$

Let α and β be arbitrary real numbers and f, f_1, and f_2 be arbitrary functions in $\mathcal{L}^2(\mathbb{R})$ The CWT, C_ψ, with respect to ψ 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

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

which is a two-dimensional wavelet energy density function known as the scalogram. The wavelet transform C_ψ 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 :

$$C_\psi = |C_\psi| e^{i\Phi(t,s)} \quad and \quad \Phi(t,s) = \arctan\left(\frac{\mathcal{I}\{C_\psi\}}{\mathcal{R}\{C_\psi\}}\right). \tag{A.9}$$

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:

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

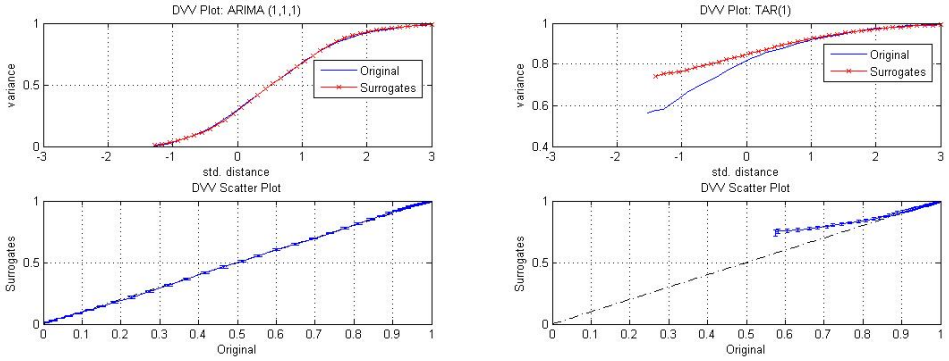
with $S_j(n)$ generated by the recursive relationship

$$S_{j-1}(n) = S_J(n) + D_J(n). \tag{A.11}$$

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.

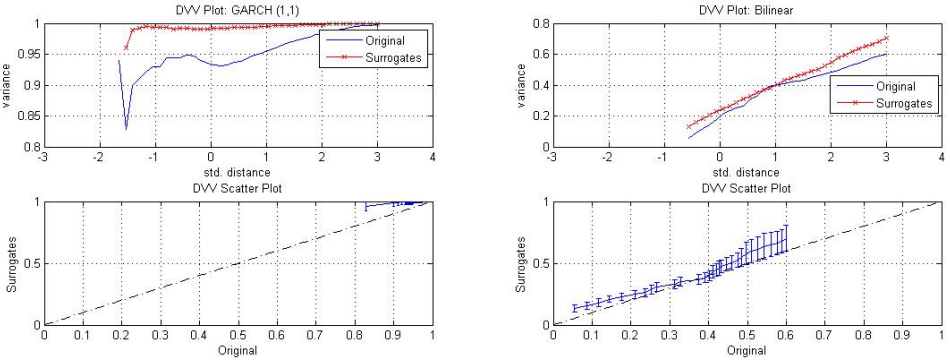
A.2 DVV Plots of simulated Processes

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 heteroskadastic process (GARCH), and a Bilinear process.



(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

$$X_T(r) = \frac{1}{T} \sum_{t=1}^{\lfloor T_r \rfloor} u_t,$$

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 $\lfloor T_r \rfloor$ denotes the largest integer that is less than or equal to T_r. Thus $X_T(r)$ is a step function in r defined as follows:

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

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:

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

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 $(y_t)_t$ 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

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

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 $(y_t)_t$ has only one unit root and all other roots are outside the unit circle, then Δy_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 $\varepsilon_t = \Psi(L) \cdot u_t, \quad u_t \sim i.i.d(0, \sigma^2)$, and

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

The Beveridge-Nelson decomposition implies that:

$$y_t = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t + y_0 = \Psi(1)\Big[u_1 + u_2 + \cdots + u_t\Big] + \eta_t - \eta_0 + y_0, \tag{B.3}$$

where $\Psi(1) = (1 - \alpha_1 - \alpha_2 - \cdots - \alpha_p)^{-1}$ and $(\eta_t)_t$ is a stationary process with zero mean of the form:

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

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

$$Z_T(r) = \frac{1}{T} \sum_{t=1}^{\lfloor T_r \rfloor} \varepsilon_t, \tag{B.4}$$

$$\sqrt{T} Z_T(r) = \sqrt{T} \frac{1}{T} \sum_{t=1}^{\lfloor T_r \rfloor} \varepsilon_t = \sqrt{T} \cdot \frac{1}{T} \Big(\Psi(1) \sum_{t=1}^{\lfloor T_r \rfloor} u_t + \eta_{\lfloor T_r \rfloor} - \eta_0 \Big),$$

$$= \Psi(1) \sqrt{T} X_T(r) + \frac{1}{\sqrt{T}} \Big(\eta_{\lfloor T_r \rfloor} - \eta_0 \Big).$$

From the functional Central Limit theorem and continuous mapping theorem we have:
$\sqrt{T} X_T(\cdot) \to_d \sigma W(\cdot) \quad \Rightarrow \quad \Psi(1) \sqrt{T} X_T(r) \to_d \Psi(1) \cdot \sigma W(r)$, with $(\eta_t)_t$ a zero mean stationary process. Thus we have: $\frac{1}{\sqrt{T}} \Big(\eta_{\lfloor T_r \rfloor} - \eta_0 \Big) \longrightarrow_p 0$ (see [60], example 17.2). It follows that:

$$\sqrt{T} Z_T(\cdot) \to_d \Psi(1) \cdot \sigma W(\cdot) \quad . \tag{B.5}$$

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

$$Z_T(r) = \begin{cases} 0 & if \ 0 \le r < 1/T \\ \varepsilon_1 / T & if \ 1/T \le r < 2/T \\ (\varepsilon_1 + \varepsilon_2)/T & if \ 2/T \le r < 3/T \\ \cdots & \cdots \\ (\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_T)/T & if \ r = 1 \end{cases} \tag{B.6}$$

$$thus \quad \sqrt{T}(Z_T(r) + \frac{y_0}{T}) = \begin{cases} \frac{y_0}{\sqrt{T}} & if \ 0 \le r < 1/T \\ \frac{y_1}{\sqrt{T}} & if \ 1/T \le r < 2/T \\ \frac{y_2}{\sqrt{T}} & if \ 2/T \le r < 3/T \\ \vdots & \vdots \\ \frac{y_T}{\sqrt{T}} & if \ r = 1 \end{cases}$$

The stochastic function $F_T(r) = \left[\sqrt{T}(Z_T(r) + \frac{y_0}{T})\right]^n$ is then expressed in the form:

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

and

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

From (B.5) we have:

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

thus

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

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

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

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

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

$$= \frac{1}{T^{(n+1)/2}} \Psi(1) \sum_{t=1}^{T} y_{t-1}^n u_t + \frac{1}{T^{(n+1)/2}} \sum_{t=1}^{T} y_{t-1}^n \Big(\eta_t - \eta_{t-1}\Big),$$

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

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

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

$$Z_{T\lfloor T_r \rfloor} = \frac{\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{\lfloor T_r \rfloor}}{T} \Rightarrow \Delta Z_{T,\lfloor T_r \rfloor} = Z_{T,\lfloor T_r \rfloor} - Z_{T,\lfloor T_r \rfloor - 1} = \frac{\varepsilon_{\lfloor T_r \rfloor}}{T}. \tag{B.11}$$

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

$$\frac{1}{T^{(n+1)/2}}\sum_{t=1}^{T} y_{t-1}^{n}\varepsilon_t = \sum_{t=1}^{T}\frac{y_{t-1}^{n}}{T^{n/2}}\cdot\sqrt{T}\frac{\varepsilon_t}{T} = \sum_{\lfloor T_r\rfloor=1}^{T} F_{T,\lfloor T_r\rfloor-1}\cdot\sqrt{T}\frac{\varepsilon_{\lfloor T_r\rfloor}}{T} = \sum_{\lfloor T_r\rfloor=1}^{T} F_{T,\lfloor T_r\rfloor-1}\cdot\sqrt{T}\Delta Z_{T,\lfloor T_r\rfloor}.$$

We have the stochastic integral:

$$\sum_{\lfloor T_r\rfloor=1}^{T} F_{T,\lfloor T_r\rfloor-1}\cdot\sqrt{T}\Delta Z_{T,\lfloor T_r\rfloor} = \int_0^1 F_T(r)\cdot\sqrt{T}dZ_T(r) = \int_0^1 F_T(r)\cdot d\sqrt{T}Z_T(r),$$

$$\Rightarrow \quad \frac{1}{T^{(n+1)/2}}\sum_{t=1}^{T} y_{t-1}^{n}\varepsilon_t = \int_0^1 F_T(r)\cdot d\sqrt{T}Z_T(r).$$

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

$$\int_0^1 F_T(r)\cdot d\sqrt{T}Z_T(r) \xrightarrow{d} \int_0^1 [\Psi(1)]^n\sigma^n[W(\cdot)]^n d\Psi(1)\cdot\sigma W(r) = [\Psi(1)]^{n+1}\sigma^{n+1}\int_0^1 [W(r)]^n dW(r),$$

$$\Rightarrow \quad \frac{1}{T^{(n+1)/2}}\sum_{t=1}^{T} y_{t-1}^{n}\varepsilon_t \xrightarrow{d} [\Psi(1)]^{n+1}\sigma^{n+1}\int_0^1 [W(r)]^n dW(r). \tag{B.12}$$

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

$$\frac{1}{T^{(n+1)/2}}\sum_{t=1}^{T} y_{t-1}^{n}(\eta_t-\eta_{t-1}) = \sum_{t=1}^{T}\frac{y_{t-1}^{n}}{T^{n/2}}\cdot\frac{\eta_t-\eta_{t-1}}{T^{1/2}} = \sum_{\lfloor T_r\rfloor=0}^{T-1} F_T(r)\cdot\frac{\eta_{\lfloor T_r\rfloor+1}-\eta_{\lfloor T_r\rfloor}}{\sqrt{T}}$$

From (B.8): $F_T(\cdot) = \left[\sqrt{T}(Z_T(\cdot)+\frac{y_0}{T})\right]^n \longrightarrow_d [\Psi(1)]^n\sigma^n[W(\cdot)]^n$

$$\Rightarrow \quad \sum_{\lfloor T_r\rfloor=0}^{T-1} F_T(r)\cdot\frac{\eta_{\lfloor T_r\rfloor+1}-\eta_{\lfloor T_r\rfloor}}{\sqrt{T}} \xrightarrow{d} [\Psi(1)]^n\sigma^n[W(r)]^n\sum_{\lfloor T_r\rfloor=0}^{T-1}\frac{\eta_{\lfloor T_r\rfloor+1}-\eta_{\lfloor T_r\rfloor}}{\sqrt{T}},$$

$$\Rightarrow \quad \sum_{\lfloor T_r\rfloor=0}^{T-1} F_T(r)\cdot\frac{\eta_{\lfloor T_r\rfloor+1}-\eta_{\lfloor T_r\rfloor}}{\sqrt{T}} \xrightarrow{d} [\Psi(1)]^n\sigma^n[W(r)]^n\left[\sqrt{T}(\frac{1}{T}\sum_{t=1}^{T}\eta_t)-\sqrt{T}(\frac{1}{T}\sum_{t=1}^{T}\eta_{t-1})\right],$$

Where η_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:

$$\left[\sqrt{T}(\frac{1}{T}\sum_{t=1}^{T}\eta_t) - \sqrt{T}(\frac{1}{T}\sum_{t=1}^{T}\eta_{t-1})\right] \xrightarrow{d} 0. \tag{B.13}$$

$$\Rightarrow \quad \sum_{\lfloor T_r\rfloor=0}^{T-1} F_T(r)\cdot\frac{\eta_{\lfloor T_r\rfloor+1}-\eta_{\lfloor T_r\rfloor}}{\sqrt{T}} \xrightarrow{d} 0.$$

It follows that:

$$\frac{1}{T^{(n+1)/2}}\sum_{t=1}^{T} y_{t-1}^{n}(\eta_t-\eta_{t-1}) \xrightarrow{d} 0. \tag{B.14}$$

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

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

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

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

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

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

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

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

$$\Rightarrow \quad \frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-i}\cdot\Delta y_{t-j} \xrightarrow{p} \gamma_{|i-j|}.$$

d.

$$\frac{1}{T^{n/2+1}}\sum_{t=1}^{T} y_{t-1}^{n}\cdot\Delta y_{t-j} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{y_{t-1}^{n}}{T^{n/2}}\cdot\frac{\Delta y_{t-j}}{\sqrt{T}} = \frac{1}{\sqrt{T}}\sum_{\lfloor T_r\rfloor=0}^{T-1} F_T(r)\cdot\frac{\Delta y_{\lfloor T_r\rfloor+1-j}}{\sqrt{T}},$$

where we have $F_T(r) \longrightarrow_d [\Psi(1)]^n\sigma^n[W(r)]^n$.

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

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

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

$$\Rightarrow \quad \frac{1}{T^{n/2+1}}\sum_{t=1}^{T} y_{t-1}^{n}\cdot\Delta y_{t-j} \xrightarrow{d} 0 \quad \Rightarrow \quad \frac{1}{T^{n/2+1}}\sum_{t=1}^{T} y_{t-1}^{n}\cdot\Delta y_{t-j} \xrightarrow{p} 0.$$

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

Notice that u_t is uncorrelated to Δy_{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:

$$E(\Delta y_{t-i} \cdot u_t / \Delta y_{t-i-1} \cdot u_t) = E(\Delta y_{t-i} / \Delta y_{t-i-1} \cdot u_t) \cdot E(u_t / \Delta y_{t-i-1} \cdot u_t)$$

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

Hence Δy_{t-i} is a martingale difference sequence.

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

since Δy_t is a zero mean stationary process

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

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

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

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

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

□

B.3 Proof of Theorem 1

Proof. Let $\tilde{D}_T$ be a 3×3 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 β, and $\hat{\sigma}_T^2$ the variance of Δy_t .

Next, consider the scaling matrix:

$$D_T = \begin{pmatrix} T^2 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & T^3 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & T^4 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & T^{1/2} & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & T^{1/2} & \dots & 0 \\ \dots & & \dots & & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & T^{1/2} \end{pmatrix}_{(p+3)\times(p+3)},$$

where $\tilde{D}_T$ and D_T are diagonal matrices and $\tilde{D}_T R = R D_T$. Therefore, we obtain:

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

$$= (\hat{\beta} - \beta)' R' \tilde{D}_T \Big[\hat{\sigma}_T^2 \tilde{D}_T R (\sum_{t=1}^{T} X_t X_t')^{-1} R' \tilde{D}_T \Big]^{-1} \tilde{D}_T R (\hat{\beta} - \beta),$$

since $\tilde{D}_T$ is symmetric : $\tilde{D}_T = \tilde{D}_T'$

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

and since $\tilde{D}_T R = R D_T$ we have :

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

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

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

with

$$D_T^{-1} = \begin{pmatrix} T^{-2} & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & T^{-3} & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & T^{-4} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & T^{-1/2} & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & T^{-1/2} & \dots & 0 \\ \dots & & \dots & & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & T^{-1/2} \end{pmatrix}_{(p+3)\times(p+3)},$$

where $X_t = (y_{t-1}^3, y_{t-1}^5, y_{t-1}^7, \Delta y_{t-1}, \cdots, \Delta y_{t-p})'_{(p+3)\times 1}$. It follows that: $D_T^{-1}(\sum_{t=1}^T X_t X_t')D_T^{-1} =$

$$\begin{pmatrix} T^{-4}\sum_{t=1}^T y_{t-1}^6 & T^{-5}\sum_{t=1}^T y_{t-1}^8 & T^{-6}\sum_{t=1}^T y_{t-1}^{10} & T^{-5/2}\sum_{t=1}^T y_{t-1}^3\Delta y_{t-1} & \cdots & T^{-5/2}\sum_{t=1}^T y_{t-1}^3\Delta y_{t-p} \\ T^{-5}\sum_{t=1}^T y_{t-1}^8 & T^{-6}\sum_{t=1}^T y_{t-1}^{10} & T^{-7}\sum_{t=1}^T y_{t-1}^{12} & T^{-\frac{7}{2}}\sum_{t=1}^T y_{t-1}^5\Delta y_{t-1} & \cdots & T^{-\frac{7}{2}}\sum_{t=1}^T y_{t-1}^5\Delta y_{t-p} \\ T^{-6}\sum_{t=1}^T y_{t-1}^{10} & T^{-7}\sum_{t=1}^T y_{t-1}^{12} & T^{-8}\sum_{t=1}^T y_{t-1}^{14} & T^{-9/2}\sum_{t=1}^T y_{t-1}^7\Delta y_{t-1} & \cdots & T^{-9/2}\sum_{t=1}^T y_{t-1}^7\Delta y_{t-1} \\ T^{-5/2}\sum_{t=1}^T y_{t-1}^3\Delta y_{t-1} & T^{-\frac{7}{2}}\sum_{t=1}^T y_{t-1}^5\Delta y_{t-1} & T^{-9/2}\sum_{t=1}^T y_{t-1}^7\Delta y_{t-1} & \frac{1}{T}\sum_{t=1}^T(\Delta y_{t-1})^2 & \cdots & \frac{1}{T}\sum_{t=1}^T \Delta y_{t-1}\Delta y_{t-p} \\ \cdots & & \cdots & & & \\ T^{-5/2}\sum_{t=1}^T y_{t-1}^3\Delta y_{t-p} & T^{-\frac{7}{2}}\sum_{t=1}^T y_{t-1}^5\Delta y_{t-p} & T^{-9/2}\sum_{t=1}^T y_{t-1}^7\Delta y_{t-p} & \frac{1}{T}\sum_{t=1}^T \Delta y_{t-1}\Delta y_{t-p} & \cdots & \frac{1}{T}\sum_{t=1}^T(\Delta y_{t-p})^2 \end{pmatrix} \tag{B.15}$$

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: $D_T^{-1}(\sum_{t=1}^T X_t X_t')D_T^{-1} \xrightarrow{d}$

$$\xrightarrow{d} \begin{pmatrix} \Psi^6\sigma^6\int_0^1 W(r)^6dr & \Psi^8\sigma^8\int_0^1 W(r)^8dr & \Psi^{10}\sigma^{10}\int_0^1 W(r)^{10}dr & 0 & 0 & 0 & \dots & 0 \\ \Psi^8\sigma^8\int_0^1 W(r)^8dr & \Psi^{10}\sigma^{10}\int_0^1 W(r)^{10}dr & \Psi^{12}\sigma^{12}\int_0^1 W(r)^{12}dr & 0 & 0 & 0 & \dots & 0 \\ \Psi^{10}\sigma^{10}\int_0^1 W(r)^{10}dr & \Psi^{12}\sigma^{12}\int_0^1 W(r)^{12}dr & \Psi^{14}\sigma^{14}\int_0^1 W(r)^{14}dr & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{p-1} \\ 0 & 0 & 0 & \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{p-2} \\ 0 & 0 & 0 & \gamma_2 & \gamma_1 & \gamma_0 & \dots & \gamma_{p-3} \\ \cdots & \cdots & \cdots & \cdots & & \cdots & \cdots & \\ 0 & 0 & 0 & \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \dots & \gamma_0. \end{pmatrix}.$$

$$\Rightarrow \quad D_T^{-1}(\sum_{t=1}^T X_t X_t')D_T^{-1} \xrightarrow{d} \begin{pmatrix} Q_{3\times 3} & 0 \\ 0 & V_{p\times p} \end{pmatrix}, \tag{B.16}$$

with $Q_{3\times 3} = \begin{pmatrix} \Psi^6\sigma^6\int_0^1 W(r)^6dr & \Psi^8\sigma^8\int_0^1 W(r)^8dr & \Psi^{10}\sigma^{10}\int_0^1 W(r)^{10}dr \\ \Psi^8\sigma^8\int_0^1 W(r)^8dr & \Psi^{10}\sigma^{10}\int_0^1 W(r)^{10}dr & \Psi^{12}\sigma^{12}\int_0^1 W(r)^{12}dr \\ \Psi^{10}\sigma^{10}\int_0^1 W(r)^{10}dr & \Psi^{12}\sigma^{12}\int_0^1 W(r)^{12}dr & \Psi^{14}\sigma^{14}\int_0^1 W(r)^{14}dr \end{pmatrix}$,

and $V_{p\times p} = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{p-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \dots & \gamma_{p-3} \\ \dots & \dots & \dots & & \dots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \dots & \gamma_0 \end{pmatrix}$.

$$\Rightarrow \quad R\Big[D_T^{-1}(\textstyle\sum_{t=1}^T X_t X_t')D_T^{-1}\Big]^{-1}R' \xrightarrow{d} (I_3 \quad 0_{3\times p}) \begin{pmatrix} Q_{3\times 3}^{-1} & 0 \\ 0 & V_{p\times p}^{-1} \end{pmatrix} \begin{pmatrix} I_3 \\ 0_{p\times 3} \end{pmatrix}$$

$$\Rightarrow \quad R\Big(D_T^{-1}(\textstyle\sum_{t=1}^T X_t X_t')D_T^{-1}\Big)^{-1}R' \xrightarrow{d} Q_{3\times 3}^{-1}.$$

$$\Rightarrow \quad \Big[R\Big(D_T^{-1}(\textstyle\sum_{t=1}^T X_t X_t')D_T^{-1}\Big)^{-1}R'\Big]^{-1} \xrightarrow{d} Q_{3\times 3}.$$

we have: $\sum_{t=1}^T X_t u_t = \begin{pmatrix} \sum y_{t-1}^3 u_t \\ \sum y_{t-1}^5 u_t \\ \sum y_{t-1}^7 u_t \\ \sum \Delta y_{t-1} u_t \\ \sum \Delta y_{t-2} u_t \\ \dots \\ \sum \Delta y_{t-p} u_t \end{pmatrix}_{(p+3)\times 1}$. $\Rightarrow D_T^{-1}\sum_{t=1}^T X_t u_t = \begin{pmatrix} \frac{1}{T^2}\sum y_{t-1}^3 u_t \\ \frac{1}{T^3}\sum y_{t-1}^5 u_t \\ \frac{1}{T^4}\sum y_{t-1}^7 u_t \\ \frac{1}{T^{1/2}}\sum \Delta y_{t-1} u_t \\ \frac{1}{T^{1/2}}\sum \Delta y_{t-2} u_t \\ \dots \\ \frac{1}{T^{1/2}}\sum \Delta y_{t-p} u_t \end{pmatrix}$.

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:

$$D_T^{-1}\sum_{t=1}^{T} X_t u_t \quad \xrightarrow{d} \quad \begin{pmatrix} \Psi^3\sigma^4 \int_0^1 W^3 dW \\ \Psi^5\sigma^6 \int_0^1 W^5 dW \\ \Psi^7\sigma^8 \int_0^1 W^7 dW \\ N(0,\sigma^2\gamma_0) \\ \dots \\ N(0,\sigma^2\gamma_0) \end{pmatrix}_{(p+3)\times 1} = \begin{pmatrix} A_{3\times 1} \\ B_{p\times 1} \end{pmatrix}, \tag{B.17}$$

where $A_{3\times 1} = \begin{pmatrix} \Psi^3\sigma^4 \int_0^1 W^3 dW \\ \Psi^5\sigma^6 \int_0^1 W^5 dW \\ \Psi^7\sigma^8 \int_0^1 W^7 dW \end{pmatrix}$ and $B_{p\times 1} = \begin{pmatrix} N(0,\sigma^2\gamma_0) \\ \dots \\ N(0,\sigma^2\gamma_0) \end{pmatrix}$.

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

$$D_T(\hat{\beta}-\beta) \xrightarrow{d} \begin{pmatrix} Q_{3\times 3}^{-1} & 0 \\ 0 & V_{p\times p}^{-1} \end{pmatrix} \begin{pmatrix} A_{3\times 1} \\ B_{p\times 1} \end{pmatrix} = \begin{pmatrix} Q_{3\times 3}^{-1} A_{3\times 1} \\ V_{p\times p}^{-1} B_{p\times 1} \end{pmatrix}_{(p+3)\times 1}.$$

$$\Rightarrow \quad RD_T(\hat{\beta}-\beta) \xrightarrow{d} (I_3 \quad 0_{3\times p}) \begin{pmatrix} Q_{3\times 3}^{-1} A_{3\times 1} \\ V_{p\times p}^{-1} B_{p\times 1} \end{pmatrix}_{(p+3)\times 1}.$$

$$\Rightarrow \quad RD_T(\hat{\beta}-\beta) \xrightarrow{d} Q_{3\times 3}^{-1} A_{3\times 1}. \tag{B.18}$$

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

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

$$\Rightarrow \quad \tilde{F}_{NL} \xrightarrow{d} \frac{1}{\sigma^2}\Big\{(Q_{3\times 3}^{-1}A_{3\times 1})' \cdot Q_{3\times 3} \cdot Q_{3\times 3}^{-1}A_{3\times 1}\Big\}.$$

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

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

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

and

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

□

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:

$$\hat{\boldsymbol{\beta}}_T = \begin{pmatrix} \sum_{t=1}^T y_{t-1}^6 & \sum_{t=1}^T y_{t-1}^8 & \sum_{t=1}^T y_{t-1}^{10} \\ \sum_{t=1}^T y_{t-1}^8 & \sum_{t=1}^T y_{t-1}^{10} & \sum_{t=1}^T y_{t-1}^{12} \\ \sum_{t=1}^T y_{t-1}^{10} & \sum_{t=1}^T y_{t-1}^{12} & \sum_{t=1}^T y_{t-1}^{14} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T y_{t-1}^3 u_t \\ \sum_{t=1}^T y_{t-1}^5 u_t \\ \sum_{t=1}^T y_{t-1}^7 u_t \end{pmatrix}.$$

The following results are needed:

$$\frac{1}{T^4}\sum_{t=1}^T y_{t-1}^6 = \frac{1}{T}\sum_{t=1}^T (\frac{1}{\sqrt{T}} y_{t-1})^6 = \frac{1}{T}\sum_{t=1}^T \int_{\frac{k-1}{T}}^{\frac{k}{T}} (\frac{1}{\sqrt{T}} Y_T(r)dr)^6 \longrightarrow_d \sigma^6 \int_0^1 W(r)^6 dr,$$

where $\frac{k-1}{T} \leq t < \frac{k}{T}$, $Y_T(r) = \frac{S_{k-1}}{\sqrt{T}} = \frac{\sum_{t=1}^{k-1} u_t}{\sqrt{T}}$,
similarly, we have:

$$\frac{1}{T^5}\sum_{t=1}^T y_{t-1}^8 \longrightarrow_d \sigma^8 \int_0^1 W(r)^8 dr,$$

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

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

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

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 $i \in \mathbb{N} > 0$

$$\frac{1}{T^{\frac{(i+1)}{2}}}\sum_{t=1}^T y_{t-1}^i u_t \longrightarrow_d \int_0^1 W^i(r)dW(r) = \sigma^{i+1}\Big\{\frac{1}{i+1}\int_0^1 W(1)^{(i+1)} - \frac{i}{2}\int_0^1 W(r)^{i-1}dr\Big\},$$

and we obtain:

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^3 u_t \longrightarrow_d \int_0^1 W^3(r)dW(r) = \sigma^4\Big\{\frac{1}{4}\int_0^1 W(1)^4 - \frac{3}{2}\int_0^1 W(r)^2 dr\Big\},$$

$$\frac{1}{T^3}\sum_{t=1}^{T} y_{t-1}^5 u_t \longrightarrow_d \int_0^1 W^5(r)dW(r) = \sigma^6\Big\{\frac{1}{6}\int_0^1 W(1)^6 - \frac{5}{2}\int_0^1 W(r)^4 dr\Big\},$$

$$\frac{1}{T^4}\sum_{t=1}^{T} y_{t-1}^7 u_t \longrightarrow_d \int_0^1 W^7(r)dW(r) = \sigma^8\Big\{\frac{1}{8}\int_0^1 W(1)^8 - \frac{7}{2}\int_0^1 W(r)^6 dr\Big\}.$$

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 $\hat{\beta}_T = (\hat{\delta}_1^2, \hat{\delta}_1^4, \hat{\delta}_1^6)'$ and $X_T = (y_{t-1}^3, y_{t-1}^5, y_{t-1}^7)'$. 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

$$D_T(\hat{\beta}_T - \beta) \longrightarrow_L \frac{1}{\sigma^2}(\Gamma Q \Gamma)^{-1}(\Gamma v),$$

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 $\Gamma = diag(1, \sigma^2, \sigma^4)$. Our test statistic has then the following representation:

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

with $R = \mathbb{I}_3$, and has the limiting distribution:

$$\tilde{F}_{NL} \longrightarrow_L ((\Gamma Q \Gamma)^{-1}(\Gamma v))'((\Gamma Q \Gamma)^{-1})^{-1}((\Gamma Q \Gamma)^{-1}(\Gamma v)) = v' Q^{-1} v.$$

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

$$\hat{\sigma}_T^2 = \frac{1}{T-4}\sum_{t=1}^{T}(\Delta y_t - \delta_1^2 y_{t-1}^3 - \delta_1^4 y_{t-1}^5 - \delta_1^6 y_{t-1}^7)^2 \quad \xrightarrow{P} \quad \sigma_T^2.$$

Under the alternative Δy_t and $y_{t-1}^i, \forall i \in \mathbb{N} > 0$ are $I(0)$ and thus it is readily seen that

$$\frac{1}{T}\sum_{t=1}^{T}\Delta y_t = \mathcal{O}_P(1), \quad \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^i = \mathcal{O}_P(1),$$

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

$$\frac{1}{T}\sum_{t=1}^{T} u_t = \mathcal{O}_P(1)$$

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

$$(\mathcal{O}_P(T))^{-1}\mathcal{O}_P(T^2) = (T\mathcal{O}_P(1))^{-1}T^2\mathcal{O}_P(1) = \frac{1}{T}T^2\mathcal{O}_P(1) = T\mathcal{O}_P(1) = \mathcal{O}_P(T).$$

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

□

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 $R_1^i := [-\infty, -0.50) \times [-\infty, 0)$, $\hat{\mathscr{A}}_1^{(1)} = \begin{pmatrix} -0.0278 & -0.0169 \\ 0.0027 & 0.2812 \end{pmatrix}$, $\hat{\mathscr{A}}_2^{(1)} = \begin{pmatrix} 0.5275 & 0.0025 \\ 0.0013 & 0.3073 \end{pmatrix}$, $\hat{\mathscr{A}}_3^{(1)} = \begin{pmatrix} 0.0069 & 0.5419 \\ -0.0005 & 0.3046 \end{pmatrix}$, $\hat{a}_0^{(1)} = \begin{pmatrix} 0.7399 \\ -0.2012 \end{pmatrix}$, (regime time: 10927).

Regime 2 $R_2^i := [-\infty, -0.50) \times [0, \infty)$, $\hat{\mathscr{A}}_1^{(2)} = \begin{pmatrix} -0.0156 & 0.0009 \\ 0.0033 & 0.2935 \end{pmatrix}$, $\hat{\mathscr{A}}_2^{(2)} = \begin{pmatrix} 0.5317 & -0.0051 \\ 0.0043 & 0.3102 \end{pmatrix}$, $\hat{\mathscr{A}}_3^{(2)} = \begin{pmatrix} -0.0012 & 0.5173 \\ -0.0021 & 0.2904 \end{pmatrix}$, $\hat{a}_0^{(2)} = \begin{pmatrix} -0.7404 \\ -0.1951 \end{pmatrix}$, (regime time: 8770).

Regime 3 $R_3^i := [-0.50, 0.50) \times [-\infty, 0)$, $\hat{\mathscr{A}}_1^{(3)} = \begin{pmatrix} -0.9417 & -0.0008 \\ 0.0143 & 0.2859 \end{pmatrix}$, $\hat{\mathscr{A}}_2^{(3)} = \begin{pmatrix} 0.8602 & 0.0040 \\ 0.0211 & 0.3003 \end{pmatrix}$, $\hat{\mathscr{A}}_3^{(3)} = \begin{pmatrix} 0.0067 & 0.8483 \\ -0.0004 & 0.3014 \end{pmatrix}$, $\hat{a}_0^{(3)} = \begin{pmatrix} 1.1337 \\ -0.2408 \end{pmatrix}$, (regime time: 3932).

Regime 4 $R_4^i := [-0.50, 0.50) \times (0.00, \infty)$, $\hat{\mathscr{A}}_1^{(4)} = \begin{pmatrix} -0.9302 & -0.0210 \\ -0.0023 & 0.3142 \end{pmatrix}$, $\hat{\mathscr{A}}_2^{(4)} = \begin{pmatrix} 0.8631 & 0.0033 \\ -0.0135 & 0.3116 \end{pmatrix}$, $\hat{\mathscr{A}}_3^{(4)} = \begin{pmatrix} 0.0066 & 0.8497 \\ 0.0079 & 0.2789 \end{pmatrix}$, $\hat{a}_0^{(4)} = \begin{pmatrix} 0.7101 \\ 0.1960 \end{pmatrix}$, (regime time: 3235).

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

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

Bibliography

[1] Abuaf, N., and Jorion, P. Purchasing power parity in the long run. *Journal of Finance 45* (1990), 157–174.

[2] Addison, P. S. Wavelet transforms and ecg: a review. *Physiological Measurement 26* (2005), 155–199.

[3] Addo, P. M., Billio, M., and Guégan, D. A test for a new modelling : The univariate mt-star model. *Centre d'Economie de la Sorbonne, Université Paris 1 Panthéon-Sorbonne, Working Paper no.*, 11083 (2011).

[4] Addo, P. M., Billio, M., and Guégan, D. Alternative methodology for turning–point detection in business cycle: A wavelet approach. *Centre d'Economie de la Sorbonne, Université Paris 1 Panthéon-Sorbonne, Working Paper no.*, 12023 (2012).

[5] Addo, P. M., Billio, M., and Guégan, D. Understanding exchange rate dynamics. *In A. Colubi, K. Fokianos, & E. J. Kontoghiorghes (Eds.), Proceedings of the 20th International Conference on Computational Statistics* (2012), 1–14.

[6] Addo, P. M., Billio, M., and Guégan, D. Nonlinear dynamics and recurrence plots for detecting financial crisis. *The North American Journal of Economics and Finance. http://dx.doi.org/10.1016/j.najef.2013.02.014* (2013).

[7] Addo, P. M., Billio, M., and Guégan, D. Turning point chronology for the eurozone: A distance plot approach. *Journal of Business Cycle Measurement & Analysis (forthcoming)* (2013).

[8] Addo, P. M., Billio, M., and Guégan, D. The univariate mt-star model and a new linearity and unit root test procedure. *Computational Statistics & Data Analysis. http://dx.doi.org/10.1016/j.csda.2013.12.009* (2014).

[9] Aizenman, J., and Noy, I. Overview of the special issue on international finance in the aftermath of the 2008 global crisis. *The North American Journal of Economics and Finance 23*, 3 (2012), 265–268.

[10] Anas, J., Billio, M., Ferrara, L., and Duca, M. L. A turning point chronology for the euro-zone. *in G L Mazzi and G Savio, Growth and Cycle in the Eurozone* (2007), 261–274.

[11] Anas, J., Billio, M., Ferrara, L., and Mazzi, G. L. A system for dating and detecting turning points in the euro area. *The Manchester Schools 76*, 5 (2008), 549–577.

[12] Anas, J., and Ferrara, L. Detecting cyclical turning points: the abcd approach and two probabilitic indicators. *Journal of Business Cycle Measurement and Analysis 1*, 2 (2006), 1–36.

[13] Arnold, M., and Gunther, R. Adaptive parameter estimation in multivariate self-exciting threshold autoregressive models. *Communications in Statistics: Simulation and Computation 30*, 2 (2001), 257–275.

[14] Artis, M. J., Marcellino, M., and Proietti, T. Dating the euro area business cycle. *CEPR Discussion Papers No 3696 and EUI Working Paper, ECO 2002/24* (2002).

[15] Artis, M. J., Marcellino, M., and Proietti, T. Characterising the business cycle for accession countries. *mimeo* (2003).

[16] Artis, M. J., Marcellino, M., and Proietti, T. Dating the euro area business cycle. *CEPR Discussion Papers*, 3696 (2003).

[17] Ashley, R. A., and Patterson, D. M. Linear versus nonlinear macroeconomics: A statistical test. *International Economic Review*, 30 (1936), 685–704.

[18] Balke, N. Credit and economic activity: Credit regimes and nonlinear propagation of shocks. *Review of Economics and Statistics*, 82 (2000).

[19] Balke, N. S., and Fomby, T. B. Threshold cointegration. *International Economics Reviews 38* (1997), 627–647.

[20] Belaire-Franch, J. Testing for non linearity in an artificial financial market: a recurrence quantification approach. *Journal of Economic Behavior and Organization*, 54 (2004), 483–494.

[21] Bengoechea, P., and Perez-Quiros, G. A useful tool to identify recessions in the euro-area. *Economic Papers*, 215 (2004).

[22] Billio, M., and Casarin, R. Identifying business cycle turning points with sequential monte carlo methods: An online and real-time application to the euro area. *Journal of Forecasting*, 29 (2010), 145–167.

[23] Billio, M., Casarin, R., Ravazzolo, F., and van Dijk, H. K. Combination schemes for turning point predictions. *Quarterly Review of Economics and Finance 52* (2012), 402–412.

[24] Brock, W., LeBaron, B., and Hsieh, D. *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*. Cambridge, Massachusetts: MIT Press, 1991.

[25] Brock, W. A., and Sayers, C. L. Is the business cycle characterised by deterministic chaos. *Journal of Monetary Economics*, 22 (1988), 71–90.

[26] Bruce, L., Koger, C., and Li, J. Dimentionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on Geoscience and Remote Sensing 40* (2002), 2331–2338.

[27] Bry, G., and Boschan, C. *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*. National Bureau of Economic Research, New York, 1971.

[28] Buncic, D. Understanding forecast failure of estar models of real exchange rates. *EERI Research Paper Series, Economics and Econometrics Research Institute (EERI) 18* (2009).

[29] Burns, A. F., and Mitchell, W. C. Measuring business cycles. *NBER* (1946).

[30] Cai, Z., and Masry, E. Nonparametric estimation of additive nonlinear arx time series: Local linear fitting and projections. *Econometric Theory 16*, 4 (2000), 465–501.

[31] Caner, M., and Hansen, B. E. Instrumental variable estimation of a threshold model. *Econometric Theory 20* (2004), 813–843.

[32] Carrasco, M., Bec, F., and Salem, M. B. Detecting mean reversion in real exchange rates from a multiple regime star model. *Annales d'Économie et de Statistique*, 99/100 (2010).

[33] Chan, K. S. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics 21* (1993), 521–533.

[34] Chan, K. S., and Tong, H. On the use of the deterministic lyapounov function for the ergodicity of stochastic difference equations. *Advances in Applied Probability 17* (1985), 666–678.

[35] Chan, K. S., and Tong, H. On estimating thresholds in autoregressive models. *Journal of Time Series Analysis 7* (1986), 197–190.

[36] Chen, R., and Tsay, R. Nonlinear additive arx models. *Journal of the American Statistical Association 88* (1993), 955–967.

[37] Chen, R., and Tsay, R. S. Functional-coefficient autoregressive models. *Journal of the American Statistical Association 88* (1993), 298–308.

[38] Dias, F. C. Nonlinearities over the business cycle: An application of the smooth transition autoregressive model to characterize gdp dynamics for the euro-area and portugal. *Bank of Portugal, Working Paper 9-03* (2003).

[39] Dickey, D. A., and Fuller, W. A. Distribution of the estimators for autoregressive time series. *Journal of the American Statistical Association 74* (1979), 427–431.

[40] Dieu-Hang, T., and Kompas, T. A modification of the estar model and testing for a unit root in a nonlinear framework. Discussion Paper, Crowford school of Economics and Government, Australian National University, September 2010.

[41] Donauer, S., Heinen, F., and Sibbertsen, P. Identification problems in estar models and a new model. Discussion Paper No. dp-444, Universitat Hannover, May 2010.

[42] Durlauf, S. N., and Phillips, P. C. B. Trends versus random walks in time series analvsis. *Econometrica 56* (1988), 1333–1354.

[43] Eckmann, J. P., Kamphorst, S. O., and Ruelle, D. Recurrence plots of dynamical systems. *Europhys Lett*, 5 (1987), 973–977.

[44] Engel, C. Long-run ppp may not hold after all. *Journal of International Economics 57* (2000), 243–73.

[45] Fagan, G., Henry, J., and Mestre, R. An area-wide model for the euro area. *ECB Working Paper*, 42 (2001).

[46] Fan, J., and Yao, Q. *Nonlinear time series: nonparametric and parametric methods*. Springer-Verlag, 2003.

[47] Franses, P. H., and van Dijk, D. *Non-linear time series models in empirical finance*. Cambridge University Press, 2000.

[48] Gallegati, M. Wavelet analysis of stock returns and aggregate economic activity. *Computational Statistics and Data Analysis 52* (2008), 3061–3074.

[49] Gallegati, M., and Gallegati, M. Wavelet variance analysis of output in g-7 countries. *Studies in Nonlinear Dynamics and Econometrics 11*, 3 (2007), 6.

[50] Gautama, T., Mandic, D. P., and Hulle, M. M. V. A differential entropy based method for determining the optimal embedding parameters of a signal. *In Proceedings of ICASSP 2003, Hong Kong IV* (2003), 29–32.

[51] Gautama, T., Mandic, D. P., and Hulle, M. M. V. The delay vector variance method for detecting determinism and nonlinearity in time series. *Physica D 190*, 3–4 (2004), 167–176.

[52] Gautama, T., Mandic, D. P., and Hulle, M. M. V. A novel method for determining the nature of time series. *IEEE Transactions on Biomedical Engineering 51* (2004), 728–736.

[53] Granger, C. W., and Andersen, A. P. *An Introduction to Bilinear Time Series Models*. Gottingen: Vandenhoek and Ruprecht, 1978.

[54] Granger, C. W. J., and Terasvirta, T. *Modelling Nonlinear Economic Relationship*. Oxford University Press,Oxford, 1993.

[55] Guégan, D., and Pham, T. D. Power of score tests against bilinear time series analysis. *Statistica Sinica 2,1* (1992), 157–171.

[56] Guttorp, B., Whitcher, P., and Percival, D. Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research 105* (2000), 14941–14962.

[57] Haggan, V., and Ozaki, T. Modeling nonlinear vibrations using an amplitude-dependent autoregressive time series model. *Biometrika 68* (1981), 189–96.

[58] Haggan, V., and Ozaki, T. Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika 68* (1981), 198–196.

[59] Hamilton, J. D. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57* (1989), 357–384.

[60] Hamilton, J. D. *Time Series Analysis*. Princeton University Press, Princeton, New Jersey, 1994.

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

[62] Hansen, B. E. Inference in tar models. *Studies in Nonlinear Dynamics and Econometrics 2* (1997), 1–14.

[63] Hansen, B. E. Testing for linearity. *Journal of Economic Surveys 13* (1999), 551–576.

[64] Hansen, B. E. Sample splitting and threshold estimation. *Econometrica 68* (2000), 575–603.

[65] Hansen, B. E. Threshold autoregression in economics. *Statistics and Its Interface 4* (2011), 123–127.

[66] Harding, D. Non parametric turning point detection dating rules and the construction of the euro zone chronology. *in G L Mazzi and G Savio (eds), Monographs of Official Statistics: Statistical Methods and Business Cycle Analysis of the Euro Zone* (2004), 122–146.

[67] Harvey, A. C., and Trimbur, T. General model-based filters for extracting trends and cycles in economic time series. *Review of Economics and Statistics 85* (2003), 244–255.

[68] Hastie, T. J., and Tibshirani, R. J. *Generalized Additive Models.* Chapman and Hall, 1990.

[69] Hegger, R., Kantz, H., and Schreiber, T. Practical implementation of nonlinear time series methods: The tisean package. *Chaos 9* (1999), 413–435.

[70] Ho, A., Moody, K., Peng, G., Mietus, C., Larson, J., Levy, M., and Goldberger, D. Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics. *Circulation 96* (1997), 842–848.

[71] Hodrick, R. J., and Prescott, E. C. Postwar u.s. business cycles: an empirical investigation. *Journal of Money, Credit and Banking 29* (1997), 1–16.

[72] Holyst, J. A., Zebrowska, M., and Urbanowicz, K. Observation of deterministic chaos in financial time series by recurrence plots, can one control chaotic economy? *Eur Phys J B*, 20 (2001), 531–535.

[73] Hubrich, K., and Terasvirta, T. Thresholds and smooth transitions in vector autoregressive models. *CREATES Research Paper*, 18 (2013).

[74] Iwanski, J. S., and Bradley, E. Recurrence plots of experimental data: To embed or not to embed? *Chaos 8*, 4 (1998), 861–871.

[75] Jensen, M. An approximate wavelet mle of short and long memory parameters. *Studies in Nonlinear Dynamics and Econometrics 3* (1999), 239–253.

[76] Kantz, H., and Schreiber, T. *Nonlinear Time Series Analysis.* Cambridge: Cambridge University Press, 1997.

[77] Kapetanios, G., Shin, Y., and Snell, A. Testing for a unit root in the nonlinear star framework. *Journal of Econometrics 112* (2003), 359–379.

[78] Kaplan, D. Exceptional events as evidence for determinism. *Physica D 73 (1)* (1994), 38–48.

[79] Keylock, C. J. Constrained surrogate time series with preservation of the mean and variance structure. *Physical Review E 73* (2006), 036707.

[80] Keylock, C. J. Improved preservation of autocorrelative structure in surrogate data using an initial wavelet step. *Nonlinear Processes in Geophysics*, 15 (2008), 435–444.

[81] Keynes, J. M. *The General Theory of Employment, Interest and Money.* Macmillan, London, 1936.

[82] Kim, C. J., and Nelson, C. R. *State-Space Models with Regime-Switching: Classical and Gibbs-Sampling Approaches with Applicationss.* MIT Press, 1999.

[83] Kontolemis, Z. G. Does growth vary over the business cycle? some evidence from the g7 countries. *Economica*, 64 (1997), 441–460.

[84] Kruse, R. A new unit root test against estar based on a class of modified statistics. Discussion Paper No. dp-398, Leibniz University of Hannover, Apr 2008.

[85] Kugiumtzis, D. Test your surrogate data before you test for nonlinearity. *Physics Review E 60* (1999), 2808–2816.

[86] Kumar, P. R., and Varaiya, P. *Stochastic Systems: Estimation, Identification, and Adaptive Control.* Prentice-Hall, 1986.

[87] Kyrtsou, C., and Vorlow, C. E. Complex dynamics in macroeconomics: A noval approach. *In: Diebolt, C., Kyrtsou, C. (Eds.). New Trends in Macroeconomics* (2005), 223–238.

[88] Leistritz, L., Hesse, W., and Arnold, M. Development of interaction measures based on adaptive non-linear time series analysis of biomedical signals. *Biomed Tech 51* (2006), 64–69.

[89] Leonenko, N. N., and Kozachenko, L. F. Sample estimate of the entropy of a random vector. *Problems of Information Transmission 23* (1987), 95–101.

[90] Lothian, J., and Taylor, M. P. Real exchange rate behavior of purchasing power parity under the current float. *Journal of Political Economy 104* (1996), 488–510.

[91] Lutkepohl, H., and Kratzig, H. *Applied Time Series Econometrics*. Cambridge University Press, Cambridge, 2004.

[92] Luukkonen, R., Saikkonen, P., and Terasvirta, T. Testing linearity against smooth transition autoregressive models. *Biometrica*, 75 (1988), 491–499.

[93] Luukkonen, R., Saikkonen, P., and Terasvirta, T. Testing linearity against smooth transition autoregressive models. *Biometrika 75* (1988), 491–499.

[94] Marwan, N. A historical review of recurrence plots. *The European Physical Journal - Special Topics 164*, 1 (2008), 3–12.

[95] Marwan, N., and Kurths, J. Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters A 302*, 5-6 (2002), 299–307.

[96] Marwan, N., Romano, M. C., Thiel, M., and Kurths, J. Recurrence plots for the analysis of complex systems. *Physics Reports 438*, 5–6 (2007), 237–329.

[97] Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., and Kurths, J. Recurrence-plot-based measures of complexity and its application to heart-rate variability. *Physics Review E*, 66 (2002), 026702.

[98] Masry, E., and Tjøstheim, D. Nonparametric estimation and identification of arch nonlinear time series: Strong convergence and asymptotic normality. *Econometric Theory 11* (1995), 258–289.

[99] Masry, E., and Tjøstheim, D. Additive arx time series and projection estimates. *Econometric Theory 13*, 2 (1997), 214–252.

[100] McAdam, P. Us, japan and the euro area: Comparing business-cycle features. *European Central Bank Working Paper*, 283 (2003).

[101] McLeish, D. L. A maximal inequality and dependent strong laws. *The Annals of Probability 5* (1975), 829–839.

[102] Meese, R., and Rogoff, K. Was it real? the exchange rate-interest differential relation over the modern floating rate period. *Journal of Finance 43* (1988), 933–948.

[103] Mitchell, W. C. Business cycles. the problem and its setting. *National Bureau of Economic Research, New York* (1927).

[104] Monch, E., and Uhlig, H. Towards a monthly business cycle chronology for the euro area. *Journal of Business Cycle Measurement and Analysis 1* (2005).

[105] O'Connell, P. G. J. The overvaluation of purchasing power parity. *Journal of International Economics 44* (1998), 1–19.

[106] Park, J. Y., and Phillips, P. C. B. Statistical inference in regressions with integrated processes: Part i. *Econometric Theory 4* (1988), 468–497.

[107] Pascalau, R. Testing for a unit root in the asymmetric nonlinear smooth transition framework. Working Paper, University of Alabama, Department of Economics, Finance and Legal Studies, USA, August 2007.

[108] Pfann, G. A., Schotman, P. C., and Tschernig, R. Nonlinear interest rate dynamics and implications for the term structure. *Journal of Econometrics 74* (1996), 149–176.

[109] Pham, D. T. The mixing property of bilinear and generalized random coefficient autoregressive models. *Stochastic Processes and Their Applications 23* (1986), 291–300.

[110] Phillips, P. C. B., and Perron, P. Testing for a unit root in time series regression. *Biometrika 75* (1988), 335–346.

[111] Pitarakis, J.-Y. A joint test for structural stability and a unit root in autoregressions. *Computational Statistics & Data Analysis. http://dx.doi.org/10.1016/j.csda.2012.07.027* (2012).

[112] Priestley, M. B. *Non-linear and Non-stationary Time Series Analysis*. London: Academic Press, 1988.

[113] Qian, L. On maximum likelihood estimators for a threshold autoregression. *Journal of Statistical Planning and Inference 75* (1998), 21–46.

[114] Ramsey, J. The contributions of wavelets to the analysis of economic and financial data. *Philosophical Transactions of the Royal Society of London A*, 357 (1999), 2593–2606.

[115] Ramsey, J., and Lampart, C. The decomposition of economic relationships by time scale using wavelets: Expenditure and income. *Studies in Nonlinear Dynamics and Econometrics 3* (1998), 23–42.

[116] Romano, M. C., Thiel, M., Kurths, J., Kiss, I. Z., and Hudson, J. Detection of synchronization for non-phase-coherent and non-stationary data. *Europhysics Letters 71*, 3 (2005), 466–472.

[117] Rothman, P., van Dijk, D., and Franses, P. H. A multivariate star analysis of the relationship between money and output. *Econometric Institute Research Report EI*, EI–9945/A (1999).

[118] Said, S. E., and Dickey, W. A. Testing for a unit root in autoregressive moving average models of unknown order. *Biometrika 71* (1984), 599–607.

[119] Schreiber, T., and Schmitz, A. Improved surrogate data for nonlinearity tests. *Physics Review Lett. 77* (1996), 635–638.

[120] Schreiber, T., and Schmitz, A. Surrogate time series. *Physica D 142* (2000), 346–382.

[121] Sichel, D. E. Business cycle asymmetry: a deeper look. *Economic Inquiry XXXI* (1993), 224–236.

[122] Stock, J. H., and Watson, M. W. Testing for common trends. *Journal of the American Statistical Association 83* (1988b), 1097–1107.

[123] Stock, M. W., and Watson, J. H. Forecasting inflation. *Journal of Monetary Economics 44*, 2 (1999), 293–335.

[124] Strikholm, B., and Terasvirta, T. A sequential procedure for determining the number of regimes in a threshold autoregressive model. *Econometrics Journal*, 9 (2006).

[125] Strozzi, F., Gutiérrez, E., Noè, C., Rossi, T., Serati, M., and Zaldívar, J. M. Measuring volatility in the nordic spot electricity market using recurrence quantification analysis. *The European Physical Journal-special Topics*, 164 (2008), 105–115.

[126] Subba, R. T., and Gabr, M. M. *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. No. 24 in Lecture Notes in Statistics. New York: Springer-Verlag, 1984.

[127] Takens, F. Detecting strange attractors in turbulence. *Lecture notes in mathematics* (1981), 366–387.

[128] Taylor, M. P., Peel, D. A., and Sarno, L. Nonlinear mean-reversion in real exchange rates: Toward a solution to the purchasing power parity puzzles. *International Economic Review 42* (2001), 1015–1042.

[129] Teolis, A. *Computational signal processing with wavelets*. Birkhauser, 1998.

[130] Terasvirta, T. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association 89* (1994), 208–218.

[131] Terasvirta, T. Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89 (1994), 208–218.

[132] Terasvirta, T., and Anderson, H. M. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, 7 (1992), 119–136.

[133] Terasvirta, T., and Granger, C. W. J. *Modelling Nonlinear Dynamic Relationships*. Oxford University Press, 1993.

[134] Terasvirta, T., Tjostheim, D., and Granger, C. W. J. *Aspects of modelling nonlinear time series*, vol. 4. In R.F Engle and D.L. McFadden (eds.), Handbook of Econometrics, 1994.

[135] Theiler, J. D., Eubank, J., Longtin, S., Galdrikian, A., and Farmer, B. Testing for nonlinearity in time series: the method of surrogate data. *Physica A 58* (1992), 77–94.

[136] Tjøstheim, D. Estimation in nonlinear time series models. *Stochastic Processes and their Applications 21* (1986), 251–273.

[137] Tjøstheim, D. Non-linear time series and markov chains. *Advances in Applied Probability 22* (1990), 587–611.

[138] Tong, H. On a threshold model. *In Pattern Recognition and Signal Processing, ed. C.H. Chen* (1978).

[139] Tong, H. Threshold models in non-linear time series analysis. *In Lecture notes in statistics 21* (1983).

[140] Tong, H. *Non-Linear Time Series. A Dynamical Systems Approach*. Clarendon Press, Oxford, 1990.

[141] Tsay, R. Testing and modeling multivariate threshold models. *Journal of the American Statistical Association 93*, 443 (1998), 1188–1202.

[142] Tweedie, R. L. Sufficient conditions for ergodicity and recurrence of markov chain on a general state space. *Stochastic Processes and Their Applications 3* (1975), 385–402.

[143] Tweedie, R. L. Sufficient conditions for ergodicity and recurrence of markov chains on a general state space. *Stochastic Processes and Their Applications 3* (1975), 385–403.

[144] Tweedie, R. L. Invariant measures for markov chains with no irreducibility assumptions. *Journal of Applied Probability 25A* (1988), 275–285.

[145] van Dijk, D., Terasvirta, T., and Franses, P. H. Smooth transition autoregressive models -a survey of recent developments. *Econometric Reviews 21* (2002), 1–47.

[146] Vougas, D. V. On unit root testing with smooth transitions. *Computational Statistics & Data Analysis 51*, 2 (2006), 797–800.

[147] Walden, D. B., and Percival, A. T. *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press, 2000.

[148] Weigend, M. C., and Casdagli, A. S. Exploring the continuum between deterministic and stochastic modelling, in time series prediction: Forecasting the future and understanding the past. *Reading, MA: Addison-Wesley* (1994), 347–367.

[149] Wold, H. O. A. A study in the analysis of stationary time series. *Uppsala, Sweden: Almquist and Wiksell* (1938).

[150] Wood, S. N., and Augustin, N. H. Gams with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling*, 157 (2002), 157–177.

[151] Zbilut, J. P., Giuliani, A., and Webber, J. C. L. Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A*, 246 (1998), 122–128.

[152] Zbilut, J. P., and Webber, J. C. L. Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A*, 171 (1992), 199–203.

[153] Zbilut, J. P., and Webber, J. C. L. Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, 76 (1994), 965–973.