ESSAYS ON TIME SERIES ANALYSIS OF FORECASTING, STRUCTURAL BREAKS, AND CONVERGENCE.

by

Harun Özkan

Economics, İstanbul Bilgi University, 2016

Submitted to the Graduate School of Social Sciences
in partial fulfillment of the requirements for the degree of
PhD in Economics

Graduate Program in Economics
İstanbul Bilgi University
2016

Essays on Time Series analysis of Forecasting, Structural breaks, and Convergence

Öngörü, Yapısal Kırılma ve Yakınsama'nın Zaman Serisi İncelemesi Üzerine Denemeler

Harun Özkan

111806005

Tez Danışmanı	: Prof. Dr. Ege Yazgan
Jüri Üyesi	: Prof. Dr. Thanasis Stengos
Jüri Üyesi	: Prof. Dr. Ramazan Gençay
Jüri Üyesi	: Doç. Dr. Cenktan Özyıldırım
Jüri Üyesi	: Yrd. Doç. Dr. Serda Selin Öztürk

Tezin Onaylandığı Tarih : 24/09/2016

Toplam Sayfa Sayısı : 117

Anahtar Kelimeler (Türkçe)
1) Öngörü
2) Yapısal kırılma
3) Markov Rejim Değişimi
4) Kesirli Tümleşme
5) GSYH Yakınsaması

Anahtar Kelimeler (İngilizce)
1) Forecasting
2) Structural Breaks
3) Markov Regime Switching
4) Fractional Integration
5) GDP Convergence

ACKNOWLEDGEMENTS

Firstly, I would like to express my genuine gratitude to my advisor Prof. M. Ege Yazgan for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, it is my privilege to thank Prof. Thanasis Stengos and Prof. Ramazan Gençay for their precious insights they have provided with me to widen my research from various perspectives.

I would certainly remiss not to mention the rest of my thesis committee: Assoc. Prof. Serda Selin Öztürk and Assoc. Prof. Cenk Yıldırım, for their insightful comments and encouragement.

I thank my fellow colleagues in İstanbul Bilgi University for the stimulating discussions and for all the fun we have had. In particular, I thank Fuat Can Beylunioğlu for his valuable contributions to the research compiled in the Chapter 4 of this thesis.

I would also like to thank my parents and my friends for their wise counsel and sympathetic ear.

Abstract

ESSAYS ON TIME SERIES ANALYSIS OF FORECASTING, STRUCTURAL BREAKS, AND CONVERGENCE.

This thesis consists of four essays.

The first essay tries to investigate whether monetary policy regime changes affect the success of forecasting inflation. The forecasting performance of some linear and nonlinear univariate models are analyzed for 14 different countries that have adopted inflation-targeting (IT) monetary regimes at some point in their economic history. The results show that forecasting performance is generally superior under an IT monetary regime compared to NIT periods. In more than half of the countries covered in this study, superior forecasting accuracy can be achieved in IT periods regardless of the model used. In contrast, among most of the remaining countries, the results remain ambiguous, and the evidence on the superiority of NIT is limited to very few countries.

In the second essay, a new and powerful method for detecting and testing of structural breaks in mean by using wavelets is exposed. Wavelet transformation decompose the variance of a process into its additive low and high frequency components. If an identically and independently process decomposed through one-scale wavelet transformation, variance of the wavelet coefficients (high-frequency) and scaling coefficients (low-frequency) will be assigned equal weights. If the structural change is in mean, the sum of squared scaling coefficients will absorb more variation leading to unequal weights between the variances of the wavelet and scaling coefficients. We use this feature of the wavelet decomposition to design statistical test for the change in the mean of an independently distributed process. We establish the limiting null distribution of our test, demonstrate that it has good empirical size and substantive power against

the existing alternatives.

The third essay proposes a specific general Markov-regime switching estimation both in the long memory parameter d and the mean of a time series. Following Tsay and Härdle (2009) we employ Viterbi algorithm, which combines the Viterbi procedures, in two state Markov-switching parameter estimation. It is well-known that existence of mean break and long memory in time series can be easily confused with each other in most cases. Thus, we aim at observing the deviation and interaction of mean and d estimates for different cases. A Monte Carlo experiment reveals that the finite sample performance of the proposed algorithm for a simple mixture model of Markov-switching mean and d changes with respect to the fractional integrating parameters and the mean values for the two regimes.

In the fourth and final essay, we examine the convergence hypothesis using a long memory framework that allows for structural breaks and does not rely on a benchmark country using both univariate and multivariate estimates of the long memory parameter d. Using per capita GDP gaps, we confirm the findings of non-stationarity and long memory behavior that have been found previously in the literature using univariate tests. However, the support for these findings is much weaker when using a multivariate framework, in which case we find more evidence of stationary behavior. Based on these results, we also investigate club formation, something that would suggest the presence of conditional convergence. We describe a club formation methodology using the sequential testing criteria that we have employed in our analysis as the basis for forming clusters or clubs of countries with similar convergence characteristics.

ÖZET

ESSAYS ON TIME SERIES ANALYSIS OF FORECASTING, STRUCTURAL BREAKS, AND CONVERGENCE

Bu tez dört ayrı denemeden oluşuyor.

Birinci denemede para politikası rejiminin enflasyonun ekonometrik kestirmini ekliyip etkilemediği incelemeye konu edilmiştir. Enflasyon hedeflemesi (IT) politikasına bir vakit geçmiş olan 14 ülke için doğrusal ve doğrusal olmayan tek–değişkenli modeller ile kestirim performansı sınanmıştır. Çalışmanın bulguları incelemeye konu olan ülkelerin yaklaşık yarısı için enflasyon hedeflemesi uygulanan dönemlerde uygulanmayan (NIT) dönemlere göre sözü edilen modellerin genel anlamda daha başarılı performans sergilediğine işaret etmektedir. Diğer ülkeler içinse, genel olarak, bir politika döneminin diğerine üstünlüğüne işaret etmediği bulgulanmıştır.

İkinci denemede ortalamadaki yapısal kırılmaların keşfi ve testi için dalgacıklar (wavelets) kullanılarak yeni ve güçlü bir test önerilemektedir. Özdeş ve bağımsız bir süreç bir basamak dalgacık dönüşümü ile ayrıştırmaya tâbi tutulursa dalgacık (yüksek frekanslı) ve ölçekleme (yüksek frekanslı) katsayılarının eşit ağırlıkta dağılması beklenir. Ortalamada kırılma var ise ölçekleme katsayılarının kare toplamları lehine denge bozulacaktır. Dalgazık dönüşümünün bu özelliği kullanılarak yapısal kırılma için bir test tasarlanmış, test istatistiğinin asimptotik özellikleri, dağılımı türetilmiş, Monte Carlo çalışması ile ampirik olarak alternatif testlere göre gücü kıyaslamalı olarak gösterilmiştir.

Üçüncü denemede zaman serisinin uzun hafıza parametresi d' katsayısında ve ortalamasında özel bir Markov-rejim geçiş modeli tâhmin yöntemi önerilmekte. Tsay

ve Hardle(2009)'un önerisini izleyerek Viterbi algoritmasına dayalı iki rejimli bir modelin parametrelerinin tâhmini ele alınmıştır. Ortalama ve uzun hafıza parametrelerinin bir çok testte çok kolayca birbiri ile karıştırıldığı literatürde iyi bilinen bir özelliktir. Çalışmada bu iki parametrenin birbiri ile karşılıklı etkileşimi ve sapması gözlenmektedir. Bu algoritmaya dayanarak yapılan Monte Carlo deneyi algoritmanın sonlu örnek performansının kesirli tümleşme parametresinin yüksekliğince ve rejimlerin ortalama değerinin farkının yüksekliğince değiştiğini göstermektedir.

Dördüncü ve sonuncu deneme, yapısal kırılmaya izin veren ve kıstas bir ülkeye dayanmayan uzun hafıza yaklaşımı çerçevesinde tek-değişkenli ve çok-değişkenli kesirli tümleşme parametresi tahminleri ile yakınsama hipotezini incelemeyi amaçlıyor. Kişi başı GSYİH farklarını kullanarak yaptığımız uygulama, literatürde daha önce tek-değişkenli testlerle bulgulanan durağansızlık ve uzun hafıza özellikleri ile örtüşmektedir. Ne var ki, çok-değişkenli tâhminle elde edilen sonuçlar çoğunlukla durağanlığa işaret etmekte ve sözü edilen bulgularla daha az uyum içndedir. Ayrıca, bu bulgulara dayanarak, koşullu yakınsama hipotezinin öngördüğü yakınsama kulüplerinin (ülke öbeklerinin) oluşumu inceleniyor. Bu amaçla sıralı test ölçütlerine dayanan bir kulüpler oluşturma yöntemi öneriliyor.

Contents

List of Figures

List of Tables

LIST OF SYMBOLS

$\mathcal{P}$	State transition probability matrix of a hidden Markov model
Θ	Parameter set
d	Fractional integration parameter
L	Lag operator
$\mathcal{F}_t$	$\sigma-$field generated by the fields $\mathcal{F}_{0\leq s\leq t}$
$\mathbb{E}$	Expectation operator
$\mathbb{P}$	Probability function
$I(0)$	Integrated of order zero: stationary (time series).
$I(1)$	Integrated of order one.
$o_p(x)$	Little-O of x in probability.
$\mathcal{O}(n)$	Set of functions that have complexity not greater than order of n as n tends to infinity.

LIST OF ACRONYMS/ABBREVIATIONS

ACF	*Autocorrelation Function*
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
AR	Autoregressive
ARFIMA	Autoregressive Fractionally Integrated Moving Average
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ARNN	Autoregressive Neural Network
BIC	Bayesian Information Criteria
BJCR algorithm	Bahl–Cocke–Jelinek–Ravi algorithm.
CDF	Cumulative Distribution Function
CUSUM	Cumulative Sum of Squared Errors
DGP	Data Generating Process
DWT	Discrete Wavelet Transform
FBM	Fractional Brownian Motion
DM	Diebold–Mariano (test)
EM	Expectation Maximization
ELW	Exact Local Whittle
FELW Whittle	Fully Extended Local Whittle
FD	Fractionally Differenced
FFT	Fast Fourier Transform
FTSE	The Financial Times Stock Exchange
FGN	Fractional Gaussian Noise
GDP	Gross domestic Product
HAC	Heteroskedasticity and Autoregression Consitent
IT	Inflation Targeting
LSTAR	Logistic Smooth Transition Autoregressive
MA	Moving Average
MCMC	Monte Carlo Markov Chain

MLW	Multivariate Local Whittle
MODWT	Maximal Overlap Discrete Wavelet Transform
MOSUM	Moving Sum of Squared Errors
MS	Markov (regime) Switching
NIT	Non Inflation Targeting
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary Least Squares
PWT	Penn World Table
RW	Random Walk
SETAR	Self-exciting thershold Autoregressive

Preface

This thesis comprises of four essays. The title of the first essay is "Comparison of inflation forecasting performances under inflation targeting and non-inflation targeting regimes". A version of this essay is published in Empirical Economics in volume 48, no. 2, with the title "Is forecasting inflation easier under inflation targeting?", which is a co-authored paper with M. Ege Yazgan. Second essay has the title "A Test of structural change of unknown location with wavelets" and, likewise, a condensed version is published in Finance Research Letters, No 12 as a joint work with M. Ege Yazgan. Third essay, "Markov regime switching in mean and in fractional integration parameter" is a slightly extended version of a forthcoming paper in Communications in Statistics - Simulation and Computation, with the same title and as a joint work with T. Stengos and M. Ege Yazgan. Fourth and final essay is based on a joint working paper with T. Stengos and M. Ege Yazgan which is, as of now, in the process of referee revision in a journal.

The figures and tables of the essays are mostly presented at the ends of the essays and the reader is notified for their relevant places in the text.

The thesis heavily relies on computations. The computational works are overwhelmingly carried out in R. The maximum likelihood function in Essay 3 and Multivariate Local Whitttle (MLW) are re-coded in C due to computational time performance concerns. Since the replication codes are way too long to include in appendix, they are available upon request by e-mail from the author at `harunozkan@gmail.com`.

Introduction

This thesis consists of four separate papers. They are placed in chronological order with respect to the author's involvement in them. Essays 2 and 3 offer new methodological and computational approaches while Essays 1 and 4 are applied pieces of works trying to contribute to two different economic questions. Although the four essays are separate pieces, Essays 1, 3, and 4 are connected with a sheer thread since they revolve around structural break, regime switching, and fractional integration approaches to time series, their interaction and their computational traits. Essay 1, on the other hand, has a separate object from others.

Essay 1 is an application of forecasting with univariate models to monthly inflation of some economies to understand whether inflation targeting (IT) policies made it easy to predict inflation rate. Inflation targeting (IT) has been adopted by several industrialized and emerging market economies, and it appears to have been successful in terms of stabilizing both inflation and the real economy. Forecasting inflation constitutes an important part of this monetary-policy strategy and directly influences its ultimate. The linear models used in Essay 1 are random walk (RW), autoregressive (AR), and autoregressive moving-average (ARMA) models; the nonlinear ones are logistic smooth transition autoregressive (LSTAR), self-exciting threshold autoregressive (SETAR), markov-switching autoregressive (MS-AR), and autoregressive neural networks (ARNN) models.

The primary goal of Essay 2 is to test for structural breaks in the mean of an independently distributed process at an unknown location. We use features of Haar wavelet decomposition to design a statistical test for the change in the mean of an independently distributed process. Although, the primary focus of this essay is on structural breaks in *mean* of an independently distributed time series, our framework can be generalized to structural breaks in variance, and structural break in stationary and non-stationary time series. We construct our statistical test of no structural break under the null hypothesis. We derive its null distribution and demonstrate that it is asymptotically normally distributed.By a periodic function, in our Monte Carlo simulations, we allow for abrupt as well as gradual structural breaks. Besides nature of the break, we also consider location, amplitude of the break, and size of the time series in our computations.

Essay 3, departs from the observation by Diebold and Inoue (2001) that a mixture model of latent Markov-switching mean can generate long memory dependence. In other words, structural change and long memory may be easily confused in estimation. To this aim, we borrow the MS-ARFIMA model of Tsay and Härdle (2009) with slight modifications and, in addition, allow for fractional integration parameter (d) be regime switching. With a Monte Carlo experiment we observe the interactions and entangling of long memory dependence on the estimates of the latent regime parameters of mean in the MS–ARFIMA framework. For the estimation of path of hidden states we employ Viterbi decoding algorithm along with maximum likelihood function of the model.

Essay 4 tries to examine the evidence of long memory type (absolute) convergence among countries in terms of per capita GDP. It extends a previous study (Stengos and Yazgan, 2014b) to proceed to investigate the possibility of club formation, a factor that would suggest the presence of conditional convergence. In that case, initial conditions would partly determine at least the long-run outcomes, and if countries with similar starting points exhibit similar long-run economic behavior (convergence clubs). Club or cluster formation has recently become a very active area of research, as there are many different ways in which one can explore their presence and/or absence. In Essay 4, we will present a methodology on club formation based on the testing criteria and will employ tools from graphing theory to provide evidence for the existence of such clubs in our group of countries.

1. Essay 1: Comparison of inflation forecasting performances under inflation targeting and non-inflation targeting regimes.

Inflation targeting (IT) has been adopted by several industrialized and emerging market economies, and it appears to have been successful in terms of stabilizing both inflation and the real economy (Svensson, 2010). IT is a monetary-policy strategy that is characterized by (1) an announced numerical inflation target, (2) a particular implementation of a monetary policy that has been called forecast targeting and to a considerable extent relies on an inflation forecast, and (3) a high degree of transparency and accountability. Hence, forecasting inflation constitutes an important part of this monetary-policy strategy and directly influences its ultimate success[1] .

Despite its importance, to the best of our knowledge, no study has analyzed the relative performance of inflation forecasts in IT periods compared to non-IT (NIT) periods, i.e., periods in which an alternative monetary policy regime has been implemented. In this study, we fill this gap by systematically analyzing the predictive performance of several time-series models for a group of countries in the IT and NIT periods of their economic history. In general, the empirical evidence that is presented in this essay supports the notion that IT provides a more suitable environment in which to forecast inflation.

The performance of inflation forecasting in different time periods has already been analyzed for the US. The evidence that has been gleaned from US data suggests that the success of inflation predictions seems to differ in the different monetary-policy regimes that have been implemented in different periods of time. In some periods, US inflation appears to be more predictable in this sense: the forecasts that are generated by multivariate models are more accurate than the forecasts that are based on simple "naïve" models, such as random walk. Whereas virtually no model seems to improve

[1]In addition to the special role given to IT, the prominence of inflation forecasting has been raised by the recent formalization of the New Keynesian optimal policy. The New Keynesian model has been used to demonstrate that the optimal choice of policy will depend on the optimal forecasts (see Svensson, 2005; Faust and Wright, 2012).

upon the "naïve" models in other periods (see Stock and Watson, 2007, 2009; Rossi and Sekhposyan, 2010; D'Agostino et al., 2006, 2011), a recent paper D'Agostino and Surico (2011) provides evidence that a policy regime that successfully stabilizes inflation in the US makes it harder to improve upon those forecasts that are based on "naïve" models.

Indeed, the US has not yet adopted all of the explicit characteristics IT, and although America seems to be taking steps in that direction, it can be classified as a non-targeter country. Therefore, the time periods considered in the above studies do not match the division that we use in this study. In contrast, this strand of literature methodologically focuses on the effect of a period of time on the forecasting performance of time-series models that are measured relative to a "naïve model" such as random walk. However, in this essay, we use a class of linear and nonlinear time-series models and focus on the effect of a period of time on the forecasting performance of each of these models. We conclude that an IT period increases the likelihood of "correct" forecasts.

Since its inception in the early 1990s in New Zealand, Canada, the U.K., and Sweden, the success of IT has been questioned. Ball and Sheridan (2004) showed that the available evidence for a group of developed economies does not lend credence to the belief that adopting an inflation-targeting regime (IT) was instrumental in reducing inflation and inflation volatility. Lin and Ye (2007) showed that inflation targeting has no significant effects on either inflation or inflation variability in seven industrial countries. In contrast, Gonçalves and Salles (2008) extended Ball and Sheridan's analysis to emerging market economies, and they found that compared to non-targeters, developing countries that adopted the IT regime experienced greater declines not only in inflation but also in growth volatility. Recently, the findings of de Mendonça and de Guimarães e Souza (2012) suggested that although IT is successful in developing economies in terms of both reducing inflation volatility and driving inflation down to internationally acceptable levels, the adoption of IT does not appear to represent an advantageous strategy in advanced economies. These findings are consistent with the previous literature[2] .

For the group of countries considered IT seems to be successful in terms of reducing and stabilizing inflation, which is suggested by the descriptive statistics presented in the following section. Hence, our evidence can also be interpreted in the following manner: any policy regime that successfully stabilizes inflation (i.e., an IT policy regime) makes inflation easier to forecast[3] .

The rest of the chapter is organized as follows: the following section describes the data and forecasting models and discusses the methodology used to compare the accuracy of the forecasts that are made with these models in IT and NIT regimes. Section 3 illustrates the results, and section 4 concludes.

1.1. Data and Forecasting Models

To compare forecasting performance of the time series models, presented below, we use consumer price index (CPI) inflation data obtained from International Financial Statistics for 14 countries. The time span of the inflation data, the adoption date of IT, mean and coefficient of variation of inflation regarding IT and NIT periods for each countries are displayed in Table 1.1 below.

As can be observed in the table, the mean of inflation in IT regimes is lower than the same mean in NIT regimes for all 14 of the countries that are examined. However, the results for volatility as measured by CVs are ambiguous, and as a result, it is not possible to assert that IT is successful in reducing volatility[4] . We use seasonally adjusted data to estimate time-series models. Seasonal adjustment is performed by

[2]See Brito and Bystedt (2010) for a counter argument that claims that there is no evidence that an inflation-targeting regime (IT) improves economic performance, as measured by the behavior of inflation and output growth in developing countries.

[3]This interpretation is not directly in contrast with D'Agostino and Surico's results, which are mentioned above. These results provide evidence that a policy regime that successfully stabilizes inflation in the US makes it harder to improve upon the forecasts that are based on "naïve" models. However, the evidence that we provide here can be interpreted thusly: a policy regime that successfully stabilizes inflation (i.e., an IT regime) makes it easier to forecast inflation irrespective of the underlying model that is used for forecasting.

[4]The effect of IT regimes on inflation volatility will be studied in future research.

Table 1.1: Inflation in countries under IT and NIT regimes

Country	NIT period	IT period	μ_{NIT}	μ_{IT}	CV_{NIT}	CV_{IT}
Canada	1957M02 - 1991M01	1991M02 - 2010M01	0.004	0.002	1.0000	2.0000
Chile	1973M02 - 1989M12	1990M01 - 2010M01	0.039	0.006	2.2564	1.1667
Colombia	1957M02 - 1999M12	2000M01 - 2010M01	0.014	0.005	1.4286	1.0000
Hungary	1976M02 - 2000M12	2001M01 - 2010M01	0.011	0.004	1.2727	1.5000
Israel	1975M05 - 1991M12	1992M01 - 2010M01	0.042	0.004	1.8333	1.5000
S. Korea	1970M02 - 1999M12	2000M01 - 2010M01	0.008	0.003	1.1250	1.6667
Mexico	1980M01 - 1998M12	2000M01 - 2010M01	0.018	0.005	2.5556	0.8000
Norway	1957M02 - 2001M02	2001M03 - 2010M01	0.004	0.002	1.5000	2.5000
Poland	1988M02 - 1997M12	1998M01 - 2010M01	0.047	0.003	1.6809	1.6667
S. Africa	1957M02 - 1999M12	2000M01 - 2010M01	0.007	0.005	1.0000	1.2000
Sweden	1957M02 - 1992M12	1993M01 - 2010M01	0.005	0.001	1.2000	4.0000
Thailand	1965M02 - 1999M12	2000M01 - 2010M01	0.005	0.002	1.4000	3.0000
Turkey	1983M03 - 2000M12	2001M01 - 2010M01	0.038	0.014	1.7368	1.1429
UK	1957M02 - 1991M12	1992M01 - 2010M01	0.006	0.002	1.1667	2.0000

Note: Inflation is computed using the log-difference of the CPI index; μ and CV refer to the mean and coefficient of variation, respectively, for the IT and NIT periods.

the X12-ARIMA filtering methodology of the U.S. Census Bureau. Therefore, we forecast seasonally adjusted inflation figures, and to evaluate their success, we first “deseasonalize” them using the estimated additive seasonal adjustment factors of X12-ARIMA. Then, we compare these forecasts with the actual figures.

In our out-of-sample forecasting exercise, we concentrate exclusively on univariate models, and we consider three types of linear univariate models and four types of nonlinear univariate models. The linear models are random walk (RW), autoregressive (AR), and autoregressive moving-average (ARMA) models; the nonlinear ones are logistic smooth transition autoregressive (LSTAR), self-exciting threshold autoregressive (SETAR), markov-switching autoregressive (MS-AR), and autoregressive neural networks (ARNN) models.

Let $\hat{y}_{t+h|t}$ be the forecast of y_t that is generated at time t for the time $t+h$ $(h \geq 1)$ by any forecasting model. In the RW model, $\hat{y}_{t+h|t}$ is equal to the the value of y_t at time t.

The ARMA model is

$$y_t = \alpha + \sum_{i=1}^{p} \phi_{1,i} y_{t-i} + \sum_{i=1}^{q} \phi_{2,i} \varepsilon_{t-i} + \varepsilon_t. \tag{1.1}$$

where p and q are selected to minimize Akaike Information Criterion (AIC) and with a maximum lag of 24. After estimating the parameters of equation (1.1) one can easily produce h-step ($h \geq 1$) forecasts by the following recursive equation:

$$\hat{y}_{t+h|t} = \alpha + \sum_{i=1}^{p} \hat{\phi}_{1,i} y_{t+h-i} + \sum_{i=1}^{q} \hat{\phi}_{2,i} \varepsilon_{t+h-i}\,. \tag{1.2}$$

When $h > 1$, to obtain forecasts we iterate on a one-period forecasting model, by feeding the previous period forecasts as regressors into the model. That means when $h > p$ and $h > q$, $y_{t+h-i|t}$ is replaced by $\hat{y}_{t+h-i|t}$ and ε_{t+h-i} by $\hat{\varepsilon}_{t+h-i|t} = 0$. An obvious alternative to iterating forward on a single-period model would be to tailor the forecasting model directly to the forecast horizon, i.e., estimate the following equation by using the data up to t.

$$y_t = \alpha + \sum_{i=0}^{p} \phi_{1,i} y_{t-i-h} + \sum_{i=0}^{q} \phi_{2,i} \varepsilon_{t-i-h} + \varepsilon_t, \tag{1.3}$$

for $h \geq 1$. We use the fitted values of this regression to directly produce h-step ahead forecast [5] .

Because it is a special case of ARMA, the estimation and forecasts of the AR model can be obtained by simply setting $q = 0$ in (1.1) and (1.3).

[5] Deciding whether the direct or the iterated approach is better is an empirical matter because it involves a trade off between the estimation efficiency and the robustness-to-model misspecification; see Elliott and Timmermann (2008). Marcellino et al. (2006) address these points empirically using a dataset of 170 US monthly macroeconomic time series. They find that the iterated approach generates the lowest MSE-values, particularly if lengthy lags of the variables are included in the forecasting models and if the forecast horizon is long.

The LSTAR model is

$$y_t = \left(\alpha_1 + \sum_{i=1}^{p} \phi_{1,i}\, y_{t-i}\right) + d_t \left(\alpha_2 + \sum_{i=1}^{q} \phi_{2,i}\, y_{t-i}\right) + \varepsilon_t, \tag{1.4}$$

where $d_t = (1 + \exp\{-\gamma(y_{t-1} - c)\})^{-1}$. Whereas ε_t are regarded as normally distributed i.i.d. variables with zero mean, α_1, α_2, $\phi_{1,i}$, $\phi_{2,i}$, γ and c are simultaneously estimated by maximum likelihood.

In the LSTAR model, the direct forecast can be obtained in the same manner as with ARMA, which is also the case for all of the subsequent nonlinear models[6] , but it is not possible to apply any iterative scheme to obtain forecasts that are multiple steps in advance, as in the linear models. This impossibility follows from the general fact that the conditional expectation of a nonlinear function is not necessarily equal to a function of that conditional expectation. In addition, one cannot iteratively derive the forecasts for the time steps $h > 1$ by plugging in the previous forecasts (see, for example, Kock and Teräsvirta, 2011)[7] . Therefore, we use the Monte Carlo integration scheme suggested by Lin and Granger (1994) to numerically calculate the conditional expectations, and we then produce the forecasts iteratively. Some computational details about the algorithmic steps of this Monte Carlo scheme are presented in Figure 1.1.

When $|\gamma| \to \infty$ LSTAR model approaches two-regime SETAR model, which is also included in our forecasting models. Alike LSTAR and most nonlinear models, in forecasting with SETAR, it is not possible to use simple iterative scheme to generate multi period forecasts. In this case, we employ a version of the Normal Forecasting Error (NFE) method suggested by Al-Qassam and Lane (1989) to generate multistep forecasts[8] . NFE is an explicit form recursive approximation to calculate higher step

[6]This process involves replacing y_t with y_{t+h} on the left-hand side of equation (4) and running the regression using data up to time t to fitted values for corresponding forecasts .

[7]Indeed, d_t is convex in y_{t-1} whenever $y_{t-1} < c$ and $-d_t$ is convex whenever $y_{t-1} > c$. Therefore, by Jensen's inequality, naïve estimation underestimates d_t if $y_{t-1} < c$ and it overestimates d_t if $y_{t-1} > c$.

[8]A detailed exposition of approaches for forecasting from a SETAR model can be found in van Dijk et al. (2003)

Figure 1.1: Monte Carlo scheme of Granger and Ding (1994).

A version of Monte Carlo approach, which was first suggested by Lin and Granger (1994), is adopted for numerical computation of the multi-step iterative forecasts within LSTAR and ARNN models. The main advantage of this choice is computational speed and accuracy against the alternative approach of numerical integration: as the forecasting steps get higher, numerical integration becomes significantly slower.

Computing more than one step forecasts via Monte Carlo framework for nonlinear models in general consists of the following steps:

- Step 1: Compute $\hat{y}_{t+1|t}$ by directly plugging in $y_t, y_{t-1}, \ldots$ into the estimated equation.
- Step 2: Generate n normal random variates with a mean of zero and a variance of $\hat{\sigma}^2$ to form a vector of simulated $\varepsilon_{t+1|t}$ values.
- Step 3: Compute simulated $y_{t+2|t}$s by plugging in the simulated values of $\varepsilon_{t+1|t}$ along with $y_{t+1|t}$ and $y_t, y_{t-1}, \ldots$ $n-$times.
- Step 4: Compute the Monte Carlo estimation of $y_{t+2|t}$ which is $\hat{y}_{t+2|t}$.
- Step 5: Repeat Steps 2, 3, and 4 to increase t for getting the higher step forecasts until the end of the forecast horizon.

Notice that in order to apply a Monte Carlo scheme for forecasting it is necessary to assume a probability distribution for the error terms $\{\varepsilon_t\}$. Here, in all models $\{\varepsilon_t\}$ is assumed to be i.i.d. $N(0, \sigma^2)$ for all t.

forecasts under normality assumption of error terms and is shown by De Gooijer and De Bruin (1998) to perform reasonably accurate compared with numerical integration and Monte Carlo method alternatives.

The two-regime MS-AR model that we consider here is as follows:

$$y_t = \alpha_s + \sum_{i=1}^{p} \phi_{s,i} y_{t-i} + \varepsilon_t, \tag{1.5}$$

where s_t is a two-state discrete Markov chain with $S = \{1, 2\}$ and $\varepsilon_t \sim$ i.i.d. $N(0, \sigma^2)$. We estimate MS-AR using the maximum likelihood algorithm expectation-maximization.

Although MS-AR models may encompass complex dynamics, point forecasting is less complicated in comparison to other non-linear models. The h-step forecasts from the MS-AR model is

$$\begin{aligned} \hat{y}_{t+h|t} = {} & P\left(s_{t+h} = 1 \mid y_t, ..., y_0\right) \left(\alpha_{s=1} + \sum_{i=1}^{p} \hat{\phi}_{s=1,i} y_{t+h-i}\right) \\ & + P\left(s_{t+h} = 2 \mid y_t, ..., y_0\right) \left(\alpha_{s=2} + \sum_{i=1}^{p} \hat{\phi}_{s=2,i} y_{t+h-i}\right), \end{aligned} \tag{1.6}$$

where $P\left(s_{t+h} = i \mid y_t, ..., y_0\right)$ is the ith element of the column vector $dfracbfP^h \hat{\xi}_{t|t}$. In addition, $\hat{\xi}_{t|t}$ represents the filtered probabilities vector and $\dfrac{bf}{P}^h$ is the constant transition probabilities matrix (see, for example, Hamilton, 1994). Hence, multistep forecasts can be obtained iteratively by plugging in $1, 2, 3, \ldots$-period forecasts that are similar to the iterative forecasting method of AR processes.

ARNN, which is the autoregressive single-hidden-layer feed-forward neural network model[9] that is suggested in Teräsvirta (2006), is defined as follows:

$$y_t = \alpha + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{h} \lambda_j d \left(\sum_{i=1}^{p} \gamma_i y_{t-i} - c\right) + \varepsilon_t, \tag{1.7}$$

[9] See Franses and Dijk (2000) for a review of feed-forward-type neural network models

where d is the logistic function, which is defined above as $d(x) = (1 + \exp\{-x\})^{-1}$. In general, the estimation of an ARNN model may be computationally challenging. Here, we follow the QuickNet method, which is a type of "relaxed greedy algorithm"; it was originally suggested by White (2006). In contrast, the forecasting procedure for ARNN is identical to the procedure for LSTAR.

To obtain pseudo-out-of-sample forecasts for a given horizon h, the models are estimated by running regressions with data that were collected no later than the date $t_0 < T$, where t_0 refers to the date when the estimation is initialized and T refers to the final date in our data. The first h-horizon forecast is obtained using the coefficient estimates from the initial regression. Next, the time subscript is advanced, and the procedure is repeated for $t_0 + 1$, $t_0 + 2$,...,$T - h$ to obtain $N_f = T - t_0 - h - 1$ distinct h-step forecasts. In our applications, N_f differs between 17 and 40 for different values of h, which are between 1 and 24 for each of our countries in both the IT and NIT periods. Therefore, t_0 is defined to meet these requirements for each of the countries and as indicated in Table 1.1 above.

In particular, there are 40 distinct point forecasts for $h = 1$, 39 distinct point forecasts for $h = 2$ and so on.

For each of these h-step forecasts, we calculate N_f forecast errors for each of the above models. Then, we calculate the models' out-of-sample mean-squared errors (MSE) for both the IT (MSE_{IT}) and NIT (MSE_{NIT}) periods.

1.2. Results

Figure 1.2 displays the out-of-sample forecasting MSE ratios; these ratios were computed separately as

$$MSE_R = \frac{MSE_{NIT}}{MSE_{IT}}$$

for seven time-series models, for each horizon from 1 to 24, and for all 14 of the studied countries[10] . In these figures, the MSE_Rs are plotted against the forecast horizon h, which is placed on the horizontal axis. Hence, the higher MSE_R plots are placed above the horizontal "1" line, which indicates the superior forecast accuracy that is achieved in an IT regime for the corresponding h.

[Figure 1.2 is here.]

The information given in these figures are summarized in Table 2 below.

[Table 2 is here]

As can be observed in Table 2, whereas all of the models have superior forecasting power in the IT periods of Colombia, Israel, South Korea, Mexico, Poland, Sweden, and Turkey, all of the models appear to be superior in the NIT periods of Norway and South Africa in terms of forecast accuracy. All models, except the RW model which is ambiguous in general, also better forecast in Canada. In Chile, all of the models have better forecast accuracy in the IT periods except for ARMA and SETAR. The results for Thailand and the United Kingdom are ambiguous for all of the models (except for ARMA). For Hungary, although the results vary across the forecasting models, the majority of the results remain ambiguous. Hence, the overall forecast accuracy in IT periods appears to be superior to the forecast accuracy in NIT periods. For half of the countries covered in this study, the IT periods provide better forecast accuracy irrespective of the model used. In contrast, NIT provides better forecast accuracy in two or three countries irrespective of which forecasting model is employed.

We also use Diebold and Mariano (1995) (DM test) to test the statistical significance of the results that were obtained above. Although the DM test is frequently used

[10] For the sake of brevity, we only provide the results of the iterative forecasts. The results that were obtained with direct forecasts are qualitatively similar and available upon request.

to assess the relative accuracy of forecasts that are derived from two competing models, we use the DM test to compare forecasts that were derived from two different periods using the same model[11] . In our case, the DM test is used to compare IT forecasts using the NIT forecasts as benchmarks. The null hypothesis is that the IT forecast is no better than the benchmark forecast (ie. the NIT forecasts), against the alternative of the superiority of IT forecasts over the benchmark forecast. We use MSEs as the loss functions in our DM tests[12] .

[Table 3 is here]

The $p-$values of the DM tests are illustrated in Table 3. In most cases where IT is already established as the better period for forecasting in Table 2, we are able to reject the null hypothesis of equal forecast accuracy. Therefore, the superiority of IT forecasts is confirmed by these DM tests. The cases for which IT is found to be superior are indicated by the emboldened numbers in the table.

[Tables 4 and 5 are here.]

[11] As long as we assume the same variance for both periods, the DM test is still valid in this case. However, one may object to this assumption by indicating that IT can reduce the variance of the inflation.

[12] Monthly inflation forecasts are scaled by 100. As a caveat, one should keep in mind that comparing two MSE series via Diebold-Mariano statistics is a scale-dependent process, i.e., the statistics change under the multiplication of two series by a constant. Here, we scale the monthly inflation figures by 100 to express them in terms of monthly percentages. See Clark and West (2006) for a detailed discussion on the effects of scaling the out-of-sample MSE-based tests.

Table 1.2: The overall relative forecasting performances of models in IT and NIT periods in terms of their MSE measures for the horizons 1-24.

Countries	**Superior forecast accuracy in IT periods**	**Superior forecast accuracy in NIT periods**	**Ambiguous**
Canada		AR, ARMA, SETAR, LSTAR, ARNN, MS	RW
Chile	RW, AR, LSTAR, ARNN, MS		ARMA, SETAR
Colombia	RW, AR, SETAR, LSTAR, ARNN, MS		
Hungary	RW	ARNN, MS	AR, ARMA, SETAR, LSTAR
Israel	RW, AR, ARMA, SETAR, LSTAR, ARNN, MS		
South Korea	RW, AR, ARMA, SETAR, LSTAR, ARNN, MS		
Mexico	RW, AR, ARMA, SETAR, LSTAR, ARNN, MS		
Norway		RW, AR, ARMA, SETAR, LSTAR, ARNN, MS	
Poland	RW, AR, ARMA, SETAR, LSTAR, ARNN, MS		
South Africa		RW, AR, ARMA, SETAR, LSTAR, ARNN, MS	
Sweden	RW, AR, ARMA, SETAR, LSTAR, ARNN, MS		
Thailand			RW, AR, ARMA, SETAR, LSTAR, ARNN, MS
Turkey	RW, AR, ARMA, ARNN, SETAR, LSTAR, MS		
United Kingdom		ARMA	RW, AR, SETAR, LSTAR, ARNN, MS

Table 1.3: The overall relative forecasting performances of models in IT and NIT periods in terms of their MSE measures for the horizons 1-24.

	h	Canada	Chile	Colombia	Hungary	Israel	S. Korea	Mexico	Norway	Poland	S. Africa	Sweden	Thailand	Turkey	UK
RW	1	0.371	**0.000**	**0.000**	**0.000**	**0.000**	**0.045**	**0.000**	0.952	**0.000**	0.876	**0.031**	0.960	**0.000**	0.178
	2	0.334	**0.000**	**0.000**	**0.001**	**0.000**	**0.040**	**0.000**	0.967	**0.000**	0.886	**0.042**	0.956	**0.000**	0.192
	3	0.371	**0.000**	**0.000**	**0.002**	**0.000**	**0.039**	**0.000**	0.961	**0.000**	0.921	0.060	0.966	**0.000**	0.215
	12	0.370	**0.000**	**0.000**	**0.017**	**0.000**	0.055	**0.000**	0.935	**0.000**	0.957	0.208	0.959	**0.000**	0.224
	24	0.497	**0.000**	**0.000**	0.073	**0.000**	0.110	**0.000**	0.731	**0.000**	0.884	0.523	0.994	**0.000**	0.611
AR	1	0.999	**0.000**	**0.000**	0.289	**0.000**	**0.001**	**0.000**	0.997	**0.000**	0.539	**0.005**	0.970	0.184	0.626
	2	0.999	**0.000**	**0.000**	0.203	**0.000**	**0.008**	**0.000**	0.997	**0.000**	0.488	**0.005**	0.976	0.280	0.618
	3	0.999	**0.000**	**0.000**	0.229	**0.000**	**0.021**	**0.000**	0.995	**0.000**	0.490	**0.005**	0.970	0.291	0.599
	12	0.993	**0.000**	**0.000**	0.265	**0.000**	0.127	**0.000**	0.985	**0.000**	0.404	**0.002**	0.926	0.209	0.442
	24	0.980	**0.037**	**0.000**	0.994	**0.001**	0.161	**0.001**	0.738	**0.000**	0.601	0.074	0.304	**0.006**	0.609
ARMA	1	0.998	0.065	**0.000**	0.473	**0.000**	**0.000**	**0.000**	0.999	**0.000**	0.540	**0.003**	0.952	**0.000**	0.591
	2	0.999	0.114	**0.000**	0.425	**0.000**	0.091	**0.000**	0.999	**0.000**	0.491	**0.003**	0.966	0.287	0.582
	3	0.999	0.181	**0.001**	0.484	**0.000**	0.114	**0.000**	0.997	**0.000**	0.501	**0.003**	0.946	**0.002**	0.554
	12	0.981	0.242	**0.000**	0.308	**0.000**	**0.023**	**0.000**	0.989	**0.000**	0.416	**0.001**	0.887	**0.020**	0.478
	24	0.966	0.503	**0.000**	0.991	**0.001**	0.244	**0.000**	0.600	**0.000**	0.663	0.140	1.000	**0.003**	0.606

Table1.3 – continued from previous page

M	h	Canada	Chile	Colombia	Hungary	Israel	S. Korea	Mexico	Norway	Poland	S. Africa	Sweden	Thailand	Turkey	UK
SETAR	1	0.994	0.197	**0.000**	**0.004**	**0.000**	**0.045**	**0.000**	0.983	**0.000**	0.617	**0.004**	0.988	**0.000**	0.908
	2	0.992	0.240	**0.000**	0.893	**0.000**	0.132	**0.000**	0.983	**0.000**	0.712	**0.004**	0.971	**0.000**	0.883
	3	0.990	0.344	**0.000**	0.893	**0.000**	**0.032**	**0.000**	0.983	**0.000**	0.710	**0.004**	0.973	**0.000**	0.927
	12	0.954	0.963	**0.000**	**0.001**	**0.005**	0.237	**0.000**	0.946	**0.000**	0.561	**0.003**	0.943	**0.000**	0.807
	24	0.722	0.096	**0.000**	0.069	**0.021**	0.208	**0.000**	0.795	**0.000**	0.639	0.064	0.941	**0.000**	0.849
LSTAR	1	0.998	**0.000**	**0.000**	0.099	**0.001**	**0.001**	**0.000**	0.955	**0.000**	0.164	**0.018**	0.980	**0.022**	0.855
	2	0.998	**0.000**	**0.000**	0.129	**0.002**	**0.003**	**0.000**	0.957	**0.000**	0.173	**0.019**	0.980	**0.006**	0.848
	3	0.998	**0.001**	**0.000**	0.118	**0.001**	**0.008**	**0.000**	0.935	**0.000**	0.201	**0.019**	0.978	**0.001**	0.824
	12	0.984	**0.014**	**0.000**	0.216	**0.034**	**0.037**	**0.000**	0.928	**0.000**	0.215	**0.040**	0.948	**0.001**	0.715
	24	0.877	0.650	**0.000**	0.763	**0.017**	**0.046**	**0.000**	0.772	**0.000**	0.287	0.069	0.872	**0.001**	0.834
ARNN	1	0.992	**0.000**	**0.000**	0.233	**0.000**	**0.000**	**0.000**	0.998	**0.000**	0.528	**0.003**	0.937	**0.000**	0.625
	2	0.999	**0.013**	**0.004**	0.098	**0.000**	**0.020**	**0.000**	1.000	**0.000**	0.520	**0.003**	0.931	**0.000**	0.641
	3	0.998	**0.000**	**0.006**	0.728	**0.000**	**0.036**	**0.000**	0.995	**0.000**	0.468	**0.003**	0.928	**0.000**	0.462
	12	0.990	**0.000**	**0.002**	0.563	**0.000**	0.090	**0.000**	0.989	**0.000**	0.393	**0.001**	0.971	0.093	0.614
	24	0.945	0.198	**0.000**	0.875	**0.000**	0.142	**0.045**	0.983	**0.000**	0.542	0.092	1.000	0.602	0.631

Table1.3 – continued from previous page

M	h	Canada	Chile	Colombia	Hungary	Israel	S. Korea	Mexico	Norway	Poland	S. Africa	Sweden	Thailand	Turkey	UK
MS-AR	1	0.998	0.110	**0.000**	**0.029**	**0.000**	**0.008**	**0.007**	0.998	**0.000**	0.560	**0.021**	0.993	**0.000**	0.696
	2	0.998	**0.050**	**0.001**	**0.027**	**0.000**	**0.003**	**0.012**	0.998	**0.000**	0.589	**0.034**	0.995	**0.003**	0.706
	3	0.998	0.100	**0.010**	**0.020**	**0.000**	**0.007**	**0.010**	0.993	**0.000**	0.623	**0.033**	0.994	**0.021**	0.763
	12	0.999	0.058	**0.000**	0.140	**0.000**	0.691	**0.009**	0.973	**0.000**	0.575	0.097	0.967	**0.019**	0.775
	24	0.993	0.415	**0.000**	0.651	**0.001**	**0.019**	**0.004**	0.991	**0.001**	0.618	0.331	0.986	0.147	0.912

Note: Emboldened items refer to the cases where the null hypothesis is rejected at 5 % level of significance.

Table 1.4: The p-values of the DM tests where the null hypothesis is that IT forecasts are no better than NIT forecasts.

	h	Canada	Chile	Colombia	Hungary	Israel	S. Korea	Mexico	Norway	Poland	S. Africa	Sweden	Thailand	Turkey	UK
AR	1	0.366	0.682	0.997	0.981	0.395	0.906	1.000	**0.000**	1.000	0.280	0.423	0.054	1.000	0.614
	2	0.229	0.903	1.000	0.993	0.739	0.984	1.000	**0.005**	0.999	0.187	0.085	**0.006**	1.000	0.681
	3	0.259	0.858	1.000	0.998	0.987	0.675	1.000	**0.000**	0.998	0.153	**0.002**	**0.006**	1.000	0.659
	12	**0.031**	0.413	0.984	0.815	0.389	0.500	0.999	**0.002**	0.503	0.064	0.115	**0.005**	0.999	0.155
	24	0.243	0.947	0.659	0.918	0.222	0.517	0.964	**0.018**	0.480	**0.031**	**0.013**	**0.001**	0.998	0.156
ARMA	1	0.366	0.714	0.997	0.981	0.400	0.905	1.000	**0.000**	1.000	0.280	0.422	0.054	1.000	0.614
	2	0.229	0.921	1.000	0.993	0.741	0.984	1.000	**0.005**	0.999	0.187	0.085	**0.006**	1.000	0.681
	3	0.259	0.883	1.000	0.998	0.987	0.673	1.000	**0.000**	0.995	0.153	**0.002**	**0.006**	1.000	0.659
	12	**0.031**	0.580	0.984	0.815	0.423	0.499	0.999	**0.002**	0.552	0.064	0.115	**0.005**	0.999	0.155
	24	0.243	0.983	0.660	0.918	0.261	0.516	0.963	**0.018**	0.485	**0.031**	**0.013**	**0.001**	0.998	0.156
SETAR	1	0.366	0.575	0.997	0.981	0.380	0.906	1.000	**0.000**	1.000	0.281	0.420	0.054	1.000	0.613
	2	0.228	0.915	1.000	0.993	0.754	0.984	1.000	**0.005**	1.000	0.187	0.084	**0.006**	1.000	0.679
	3	0.258	0.886	1.000	0.998	0.988	0.674	1.000	**0.000**	0.999	0.153	**0.002**	**0.006**	1.000	0.655
	12	**0.031**	0.580	0.983	0.809	0.467	0.499	0.998	**0.002**	0.925	0.064	0.115	**0.005**	0.999	0.155
	24	0.242	0.983	0.655	0.916	0.309	0.518	0.890	**0.018**	0.838	**0.031**	**0.013**	**0.001**	0.997	0.157

Table1.4 – continued from previous page

M	h	Canada	Chile	Colombia	Hungary	Israel	S. Korea	Mexico	Norway	Poland	S. Africa	Sweden	Thailand	Turkey	UK
LSTAR	1	0.366	0.721	0.997	0.981	0.396	0.906	1.000	**0.000**	1.000	0.281	0.419	0.054	1.000	0.614
	2	0.228	0.923	1.000	0.993	0.750	0.984	1.000	**0.005**	1.000	0.187	0.085	**0.006**	1.000	0.677
	3	0.257	0.885	1.000	0.998	0.989	0.672	1.000	**0.000**	0.999	0.152	**0.002**	**0.006**	1.000	0.652
	12	**0.031**	0.562	0.984	0.803	0.286	0.500	0.996	**0.002**	0.902	0.063	0.115	**0.005**	0.999	0.155
	24	0.243	0.965	0.664	0.900	**0.028**	0.518	0.966	**0.018**	0.899	**0.031**	**0.013**	**0.001**	0.998	0.157
ARNN	1	0.366	0.650	0.997	0.981	0.365	0.906	1.000	**0.000**	1.000	0.280	0.423	0.054	1.000	0.615
	2	0.228	0.899	1.000	0.993	0.705	0.984	1.000	**0.005**	0.999	0.187	0.085	**0.006**	1.000	0.681
	3	0.259	0.843	1.000	0.998	0.984	0.674	1.000	**0.000**	0.997	0.153	**0.002**	**0.006**	1.000	0.659
	12	**0.031**	0.482	0.985	0.814	0.079	0.501	0.998	**0.002**	1.000	0.064	0.115	0.005	0.999	0.155
	24	0.243	0.970	0.674	0.917	0.027	0.518	0.886	**0.018**	1.000	**0.031**	**0.013**	**0.001**	0.997	0.156
MS-AR	1	0.365	0.722	0.996	0.981	0.412	0.896	0.361	**0.000**	1.000	0.281	0.421	0.054	1.000	0.609
	2	0.228	0.921	1.000	0.993	0.756	0.984	1.000	**0.005**	1.000	0.187	0.085	**0.006**	1.000	0.677
	3	0.258	0.886	1.000	0.998	0.989	0.668	0.899	**0.000**	1.000	0.153	**0.002**	**0.006**	1.000	0.656
	12	**0.031**	0.582	0.983	0.812	0.391	0.490	1.000	**0.002**	0.969	0.064	0.115	**0.005**	0.999	0.155
	24	0.242	0.983	0.652	0.917	0.178	0.517	1.000	**0.018**	0.902	**0.031**	**0.013**	**0.001**	0.998	0.157

Note: Emboldened items refer to the cases where the null hypothesis is rejected at 5 % level of significance.

Table 1.5: The $p-$values of the Diebold-Mariano statistics for IT period. (Benchmark Model: Random Walk)

	h	Canada	Chile	Colombia	Hungary	Israel	S. Korea	Mexico	Norway	Poland	S. Africa	Sweden	Thailand	Turkey	UK
AR	1	0.143	0.948	0.992	0.138	**0.043**	0.847	0.968	**0.004**	0.692	0.623	**0.008**	0.237	0.076	0.083
	2	0.157	0.507	0.995	0.210	**0.001**	0.592	0.827	**0.013**	0.154	0.591	**0.000**	**0.020**	**0.009**	0.095
	3	**0.024**	0.354	0.958	0.767	0.058	**0.049**	0.884	**0.010**	0.092	0.743	**0.001**	**0.008**	**0.001**	0.076
	12	0.053	**0.036**	0.231	0.649	0.436	0.289	0.957	0.147	0.677	0.999	**0.042**	0.087	0.134	0.061
	24	0.157	**0.013**	**0.009**	0.172	0.857	0.273	0.925	0.160	0.399	0.909	0.586	**0.034**	0.390	0.280
ARMA	1	0.144	0.956	0.994	0.124	0.104	0.836	0.976	**0.005**	0.813	0.586	**0.007**	0.235	0.497	0.088
	2	0.153	0.540	0.997	0.195	**0.006**	0.621	0.889	**0.017**	0.298	0.560	**0.000**	**0.019**	0.111	0.101
	3	**0.024**	0.388	0.975	0.746	0.235	**0.070**	0.941	**0.010**	0.200	0.717	**0.000**	**0.007**	0.057	0.081
	12	0.054	**0.043**	0.239	0.631	0.752	0.296	0.968	0.151	0.821	0.999	**0.040**	0.088	0.883	0.067
	24	0.155	**0.003**	0.008	0.147	0.883	0.291	0.950	0.163	0.644	0.902	0.625	**0.034**	0.851	0.283
SETAR	1	0.127	0.948	0.987	0.229	**0.036**	0.816	0.994	**0.009**	0.719	0.729	**0.008**	0.224	0.080	0.119
	2	0.172	0.498	0.991	0.887	**0.001**	0.400	0.942	0.032	0.164	0.760	**0.000**	**0.016**	**0.009**	0.241
	3	**0.036**	0.351	0.937	1.000	**0.044**	**0.027**	0.964	0.014	0.118	0.769	0.001	0.003	0.002	0.070
	12	**0.056**	**0.033**	0.302	1.000	0.423	0.402	0.969	0.048	0.762	1.000	**0.047**	0.141	0.419	0.065
	24	0.131	**0.047**	**0.007**	1.000	0.855	0.436	0.953	0.401	0.538	0.941	0.496	0.029	0.571	0.248

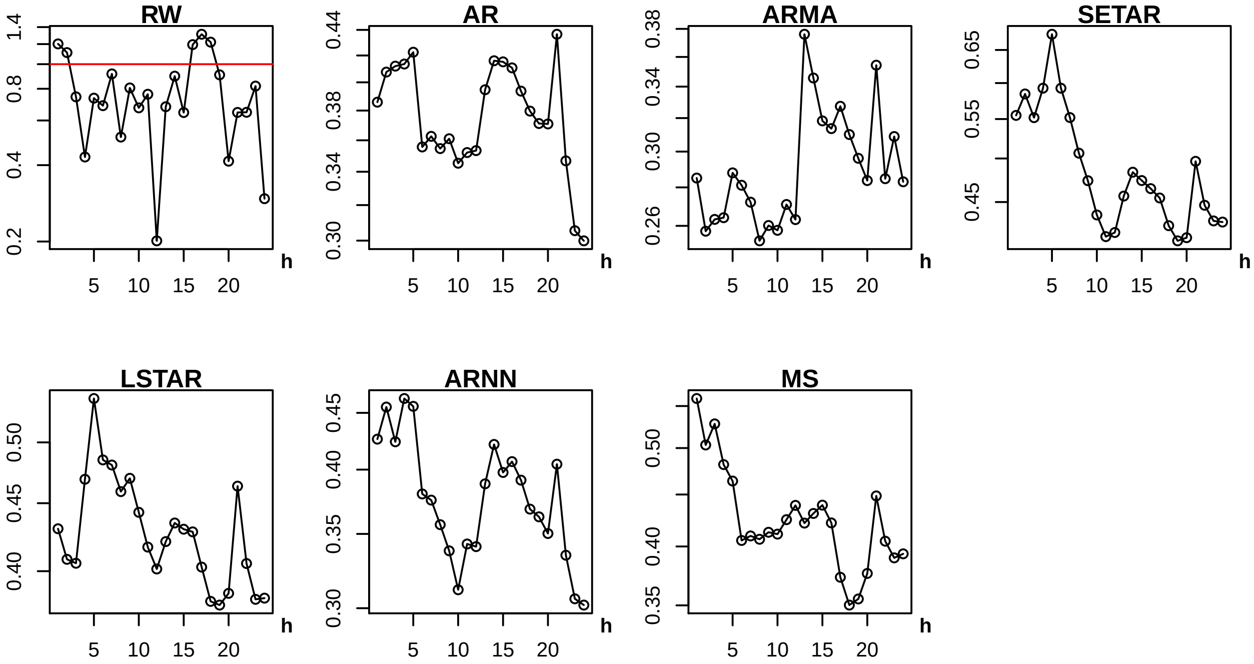
Table1.5 – continued from previous page

M	*h*	Canada	Chile	Colombia	Hungary	Israel	S. Korea	Mexico	Norway	Poland	S. Africa	Sweden	Thailand	Turkey	UK
LSTAR	1	0.127	0.974	0.993	0.292	**0.042**	0.728	0.995	**0.001**	0.869	0.727	**0.007**	0.224	0.161	0.091
	2	0.159	0.668	0.998	0.786	**0.002**	0.537	0.968	**0.007**	0.266	0.933	**0.000**	0.297	**0.036**	0.113
	3	**0.027**	0.461	0.992	0.834	0.060	**0.046**	0.992	**0.039**	0.184	0.832	**0.000**	**0.003**	**0.040**	0.059
	12	0.054	0.104	0.384	0.382	0.607	0.202	0.999	0.214	0.852	0.998	**0.039**	0.079	0.927	0.055
	24	0.106	**0.001**	**0.008**	0.666	0.863	0.334	0.989	0.097	0.754	0.967	0.625	0.941	0.859	0.267
ARNN	1	0.157	0.953	0.996	0.266	**0.040**	0.808	0.979	**0.005**	0.655	0.589	**0.008**	0.231	0.448	0.081
	2	0.167	0.516	0.998	0.478	**0.001**	0.466	0.897	**0.021**	0.134	0.582	**0.000**	**0.016**	0.101	0.091
	3	**0.027**	0.364	0.989	0.917	**0.044**	**0.037**	0.946	**0.015**	0.081	0.757	**0.001**	**0.005**	0.055	0.070
	12	0.058	**0.034**	0.374	0.649	0.377	0.291	0.985	0.128	0.676	1.000	**0.045**	0.104	0.956	0.059
	24	0.145	**0.024**	0.006	0.232	0.868	0.293	0.953	0.220	0.408	0.918	0.527	0.051	0.999	0.291
MS-AR	1	0.120	0.945	0.990	0.141	**0.008**	0.783	0.958	**0.011**	0.679	0.418	**0.006**	0.230	0.156	0.052
	2	0.148	0.541	0.995	0.083	**0.000**	0.691	0.898	**0.022**	0.083	0.767	**0.000**	**0.028**	0.115	0.072
	3	**0.026**	0.420	0.972	0.754	**0.010**	0.137	0.983	**0.019**	0.104	0.889	**0.000**	**0.009**	**0.036**	0.062
	12	0.058	0.069	0.440	0.717	0.392	0.182	0.994	0.093	0.836	0.999	**0.041**	0.096	0.945	**0.049**
	24	0.156	**0.020**	**0.010**	0.259	0.862	0.666	0.995	0.191	0.634	0.916	0.598	**0.023**	0.904	0.275

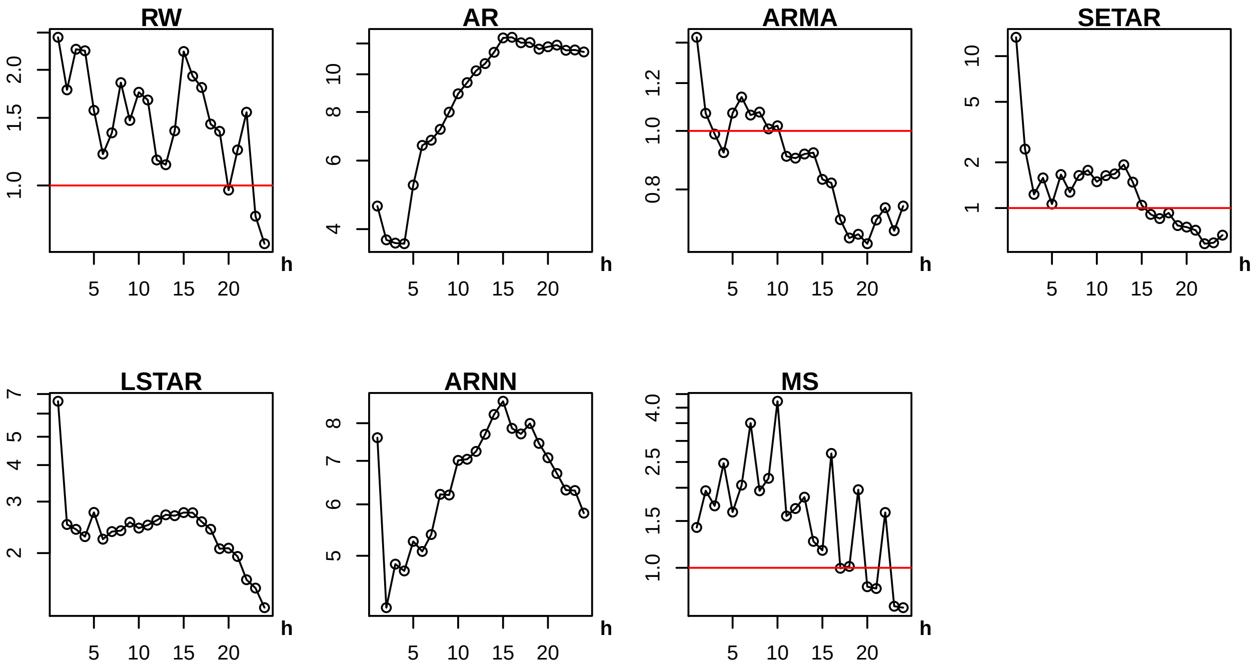
Note: Emboldened items refer to the cases where the null hypothesis is rejected at 5 % level of significance.

Figure 1.2: MSE ratios computed for 14 IT countries.

Canada

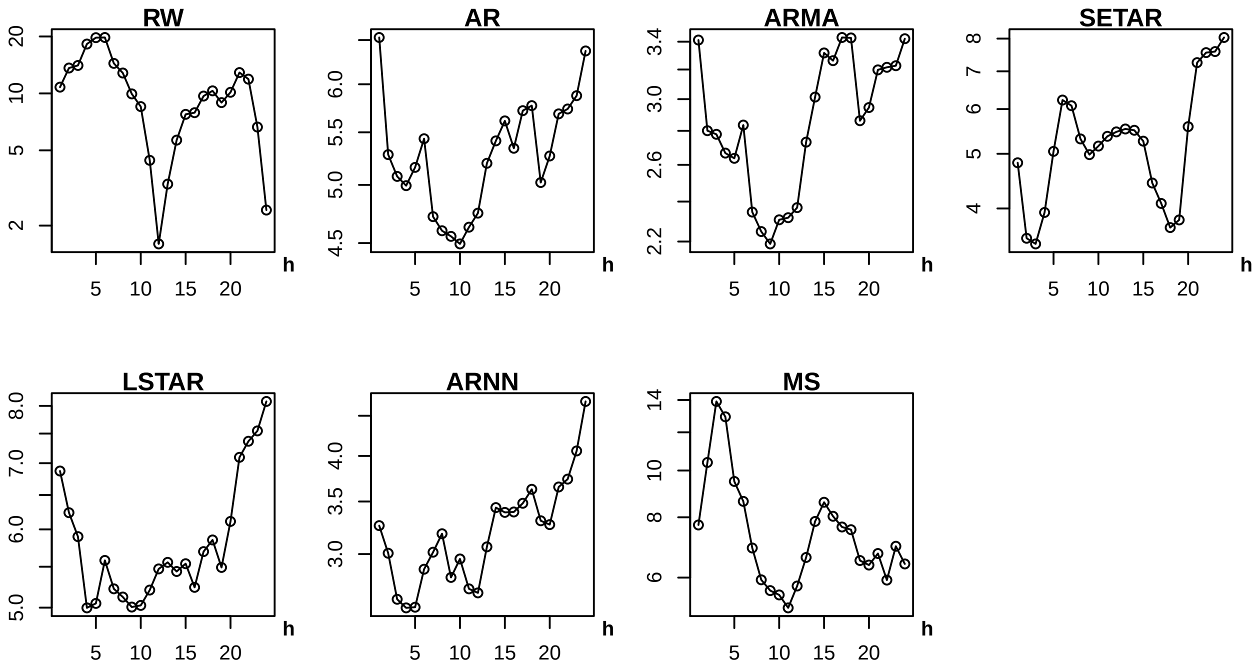


Chile

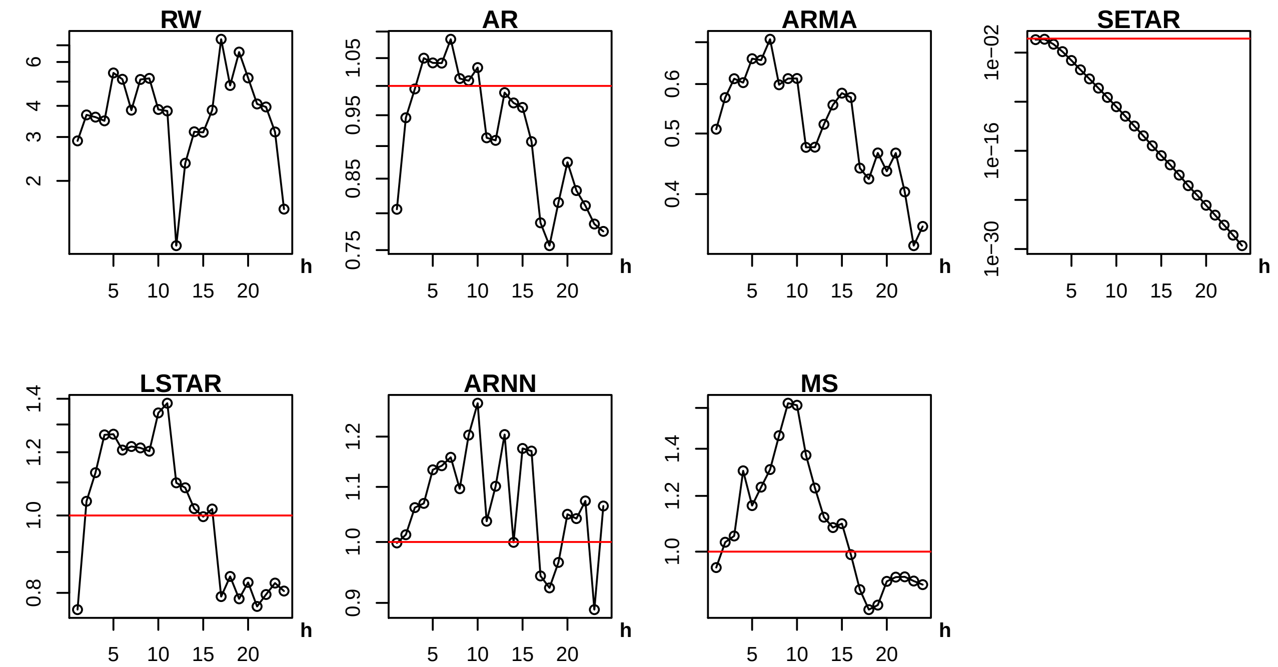


Notes: The $x-$ axis is h whihc denotes the forecast horizon. The $y-$ axis is the ratio of mean squared error (MSE) of the forecasting model in NIT period to mean squared error (MSE) of the forecasting model in IT period.

Colombia

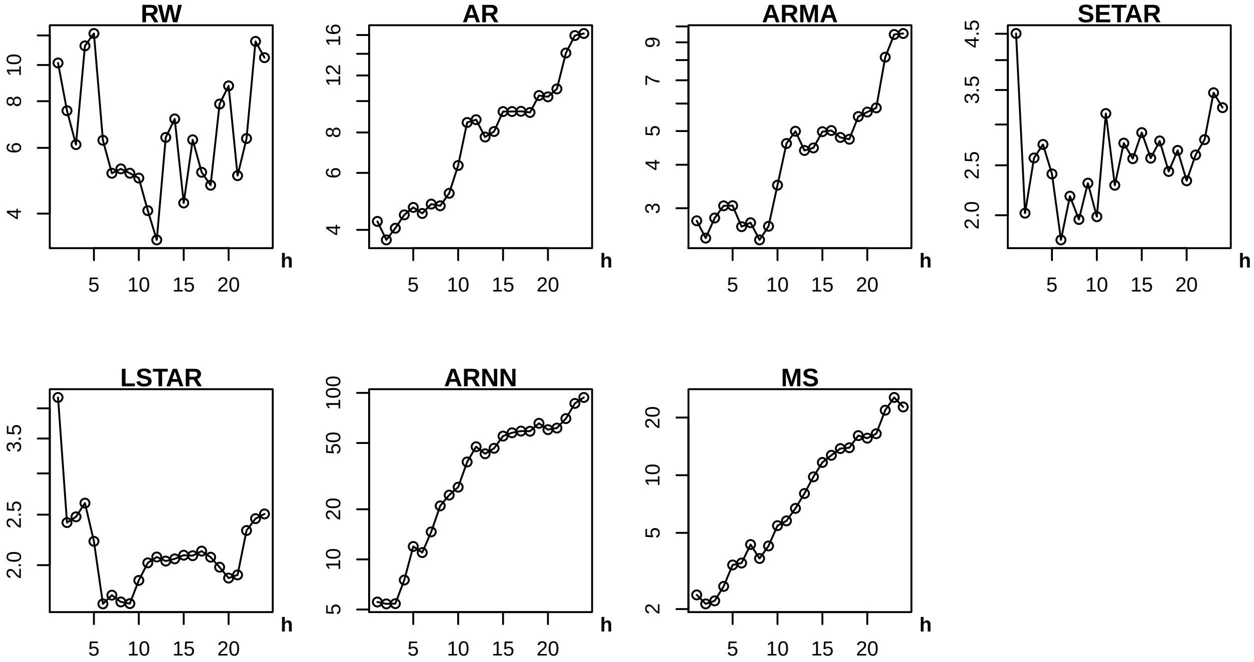


Hungary

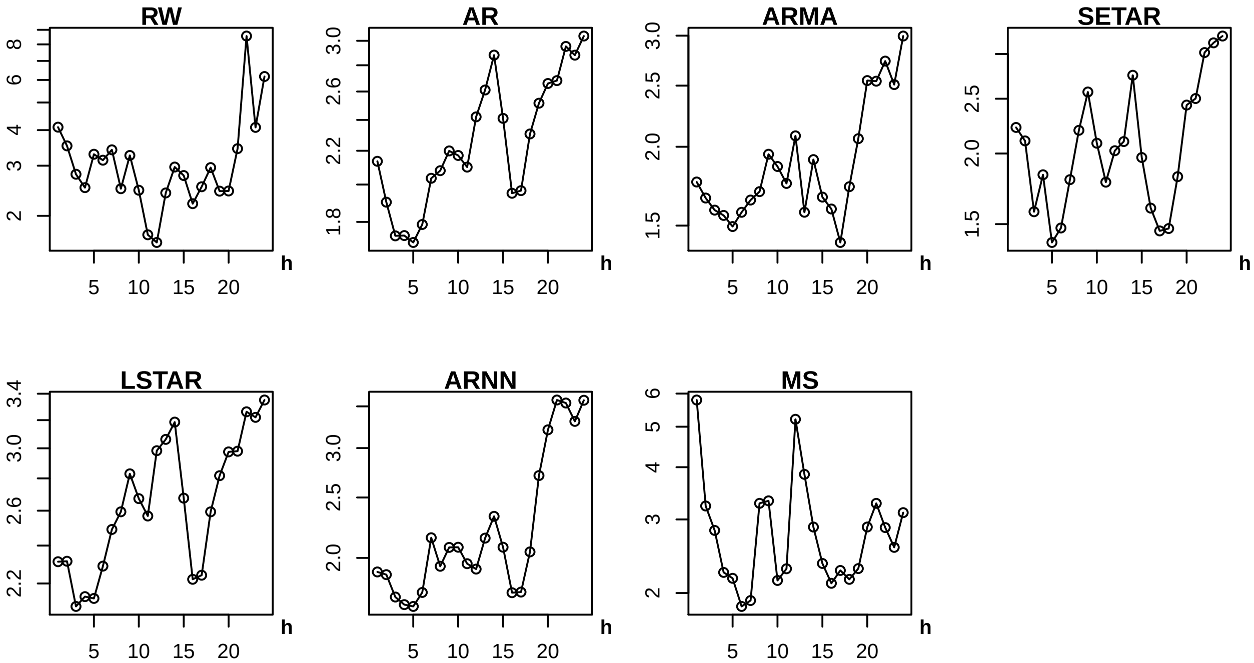


Notes: The $x-$ axis is h whihc denotes the forecast horizon. The $y-$ axis is the ratio of mean squared error (MSE) of the forecasting model in NIT period to mean squared error (MSE) of the forecasting model in IT period.

Israel

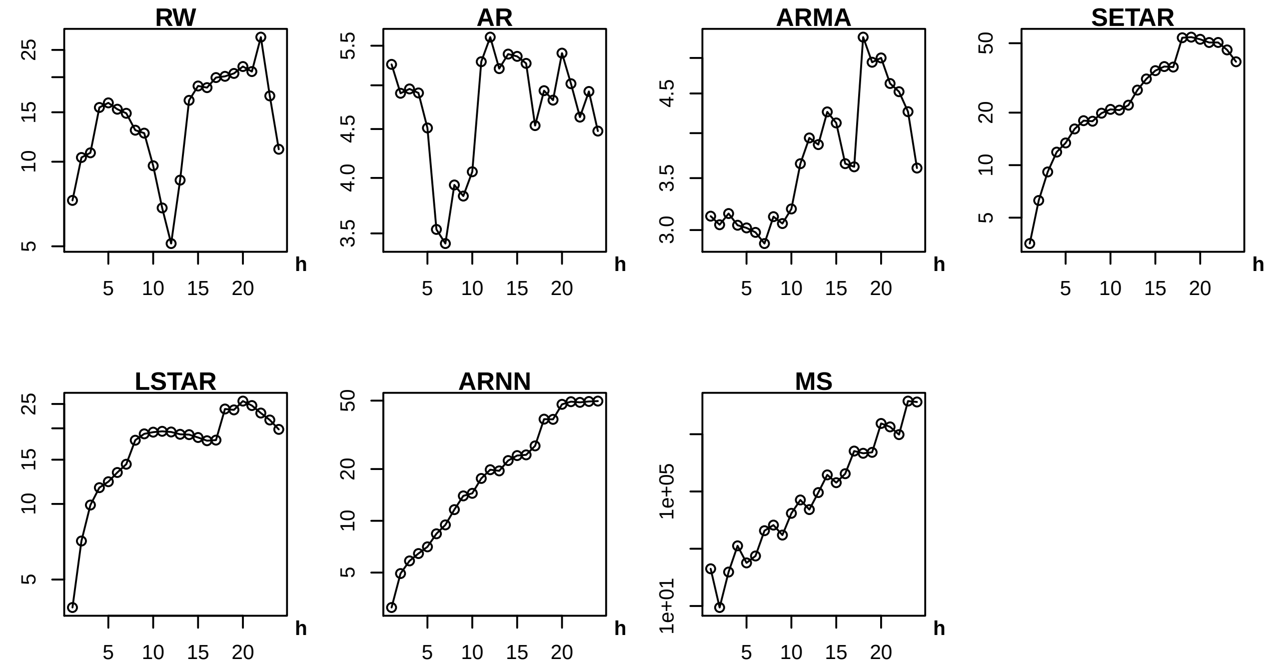


South Korea

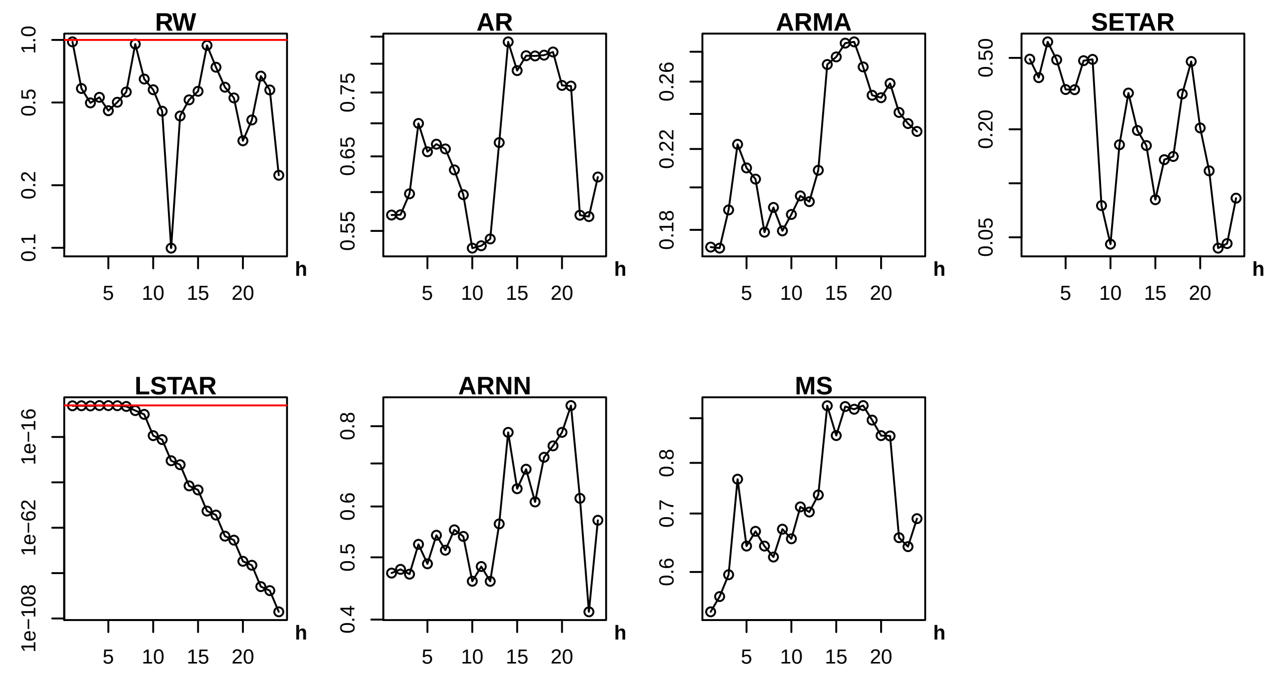


Notes: The $x-$ axis is h whihc denotes the forecast horizon. The $y-$ axis is the ratio of mean squared error (MSE) of the forecasting model in NIT period to mean squared error (MSE) of the forecasting model in IT period.

Mexico

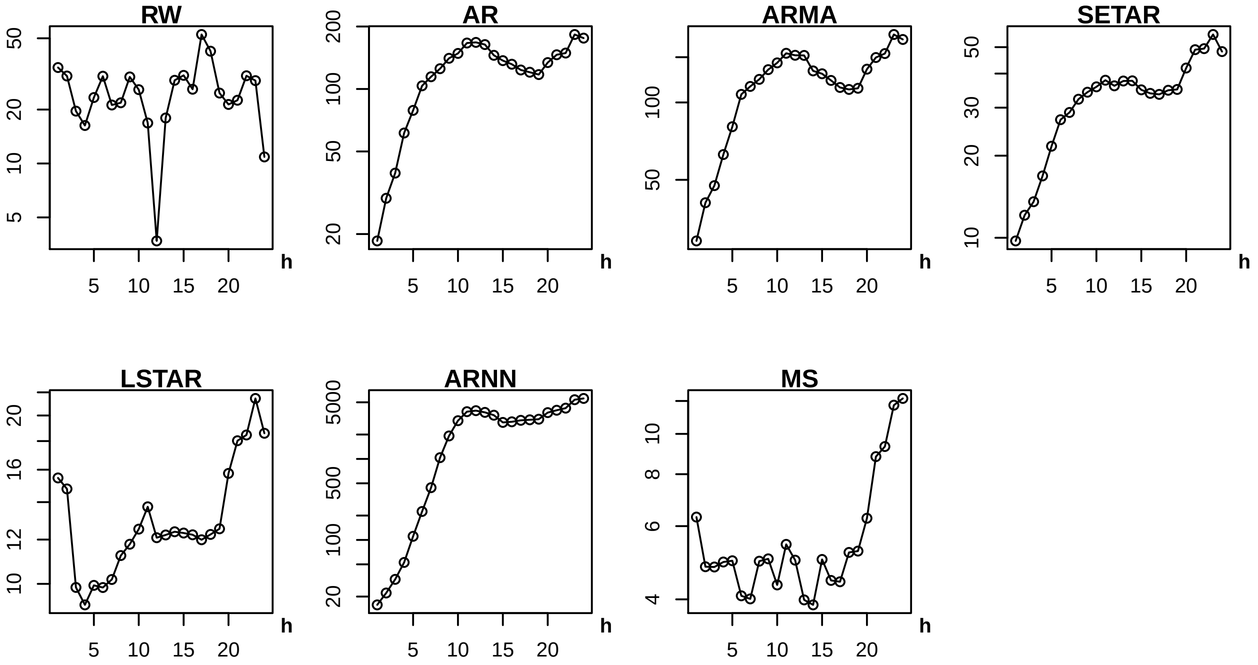


Norway

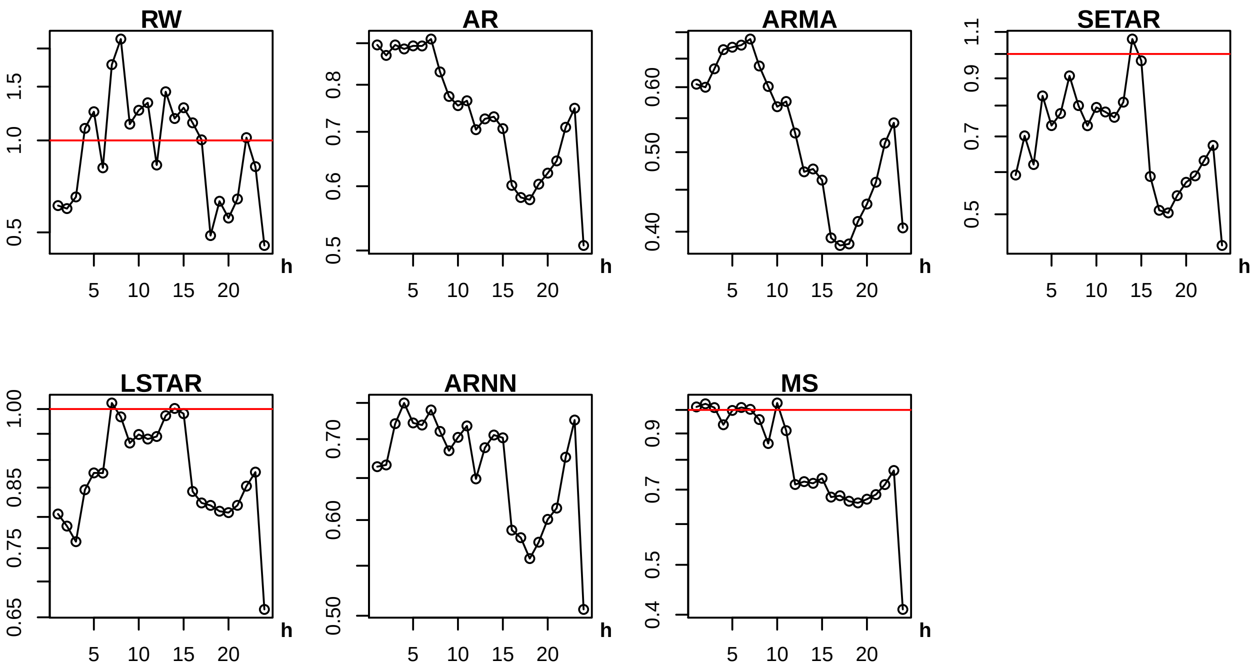


Notes: The $x-$ axis is h whihc denotes the forecast horizon. The $y-$ axis is the ratio of mean squared error (MSE) of the forecasting model in NIT period to mean squared error (MSE) of the forecasting model in IT period.

Poland

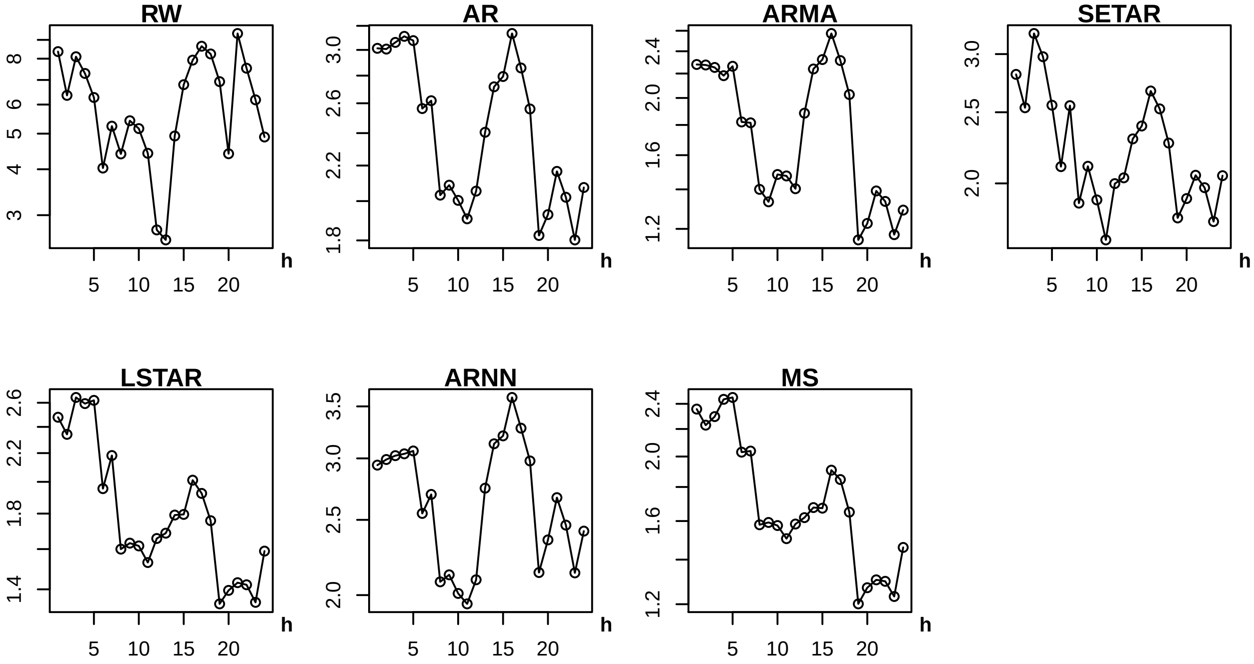


South Africa

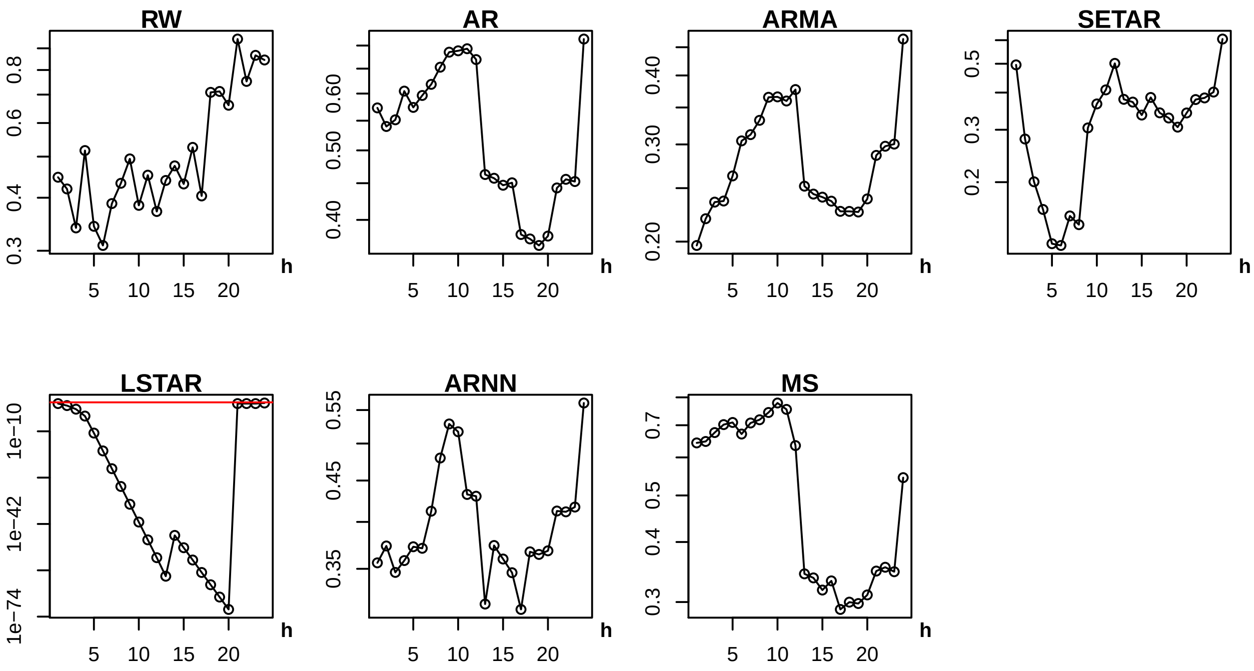


Notes: The $x-$ axis is h whihc denotes the forecast horizon. The $y-$ axis is the ratio of mean squared error (MSE) of the forecasting model in NIT period to mean squared error (MSE) of the forecasting model in IT period.

Sweden

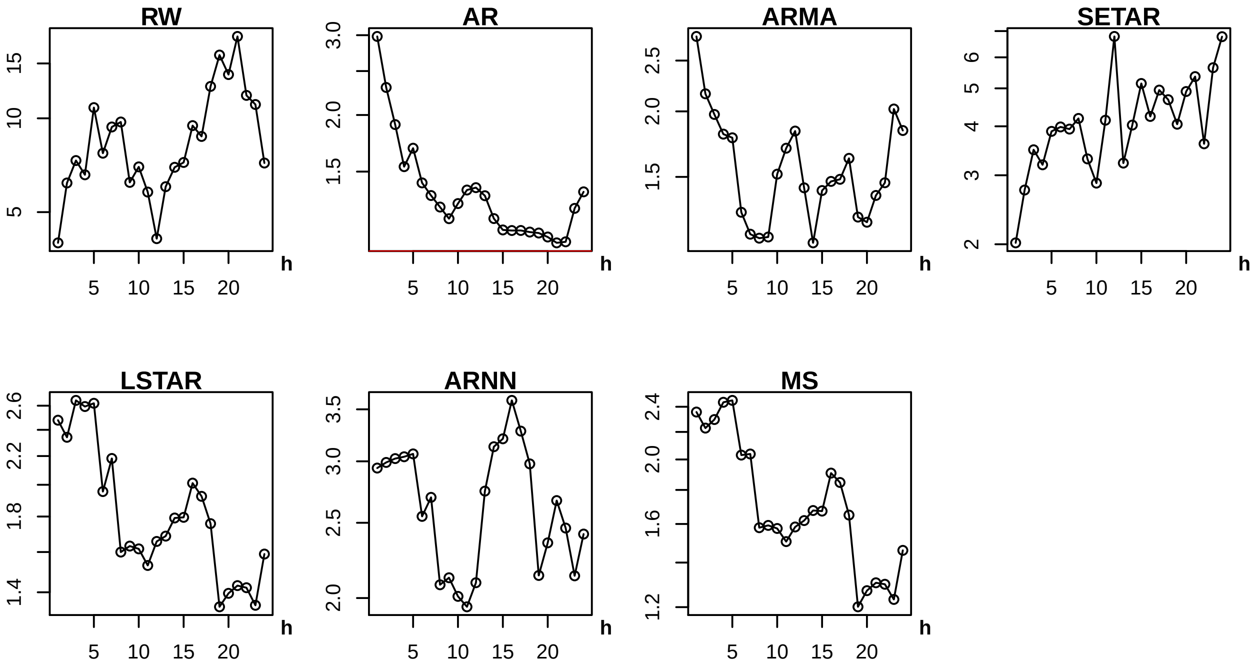


Thailand

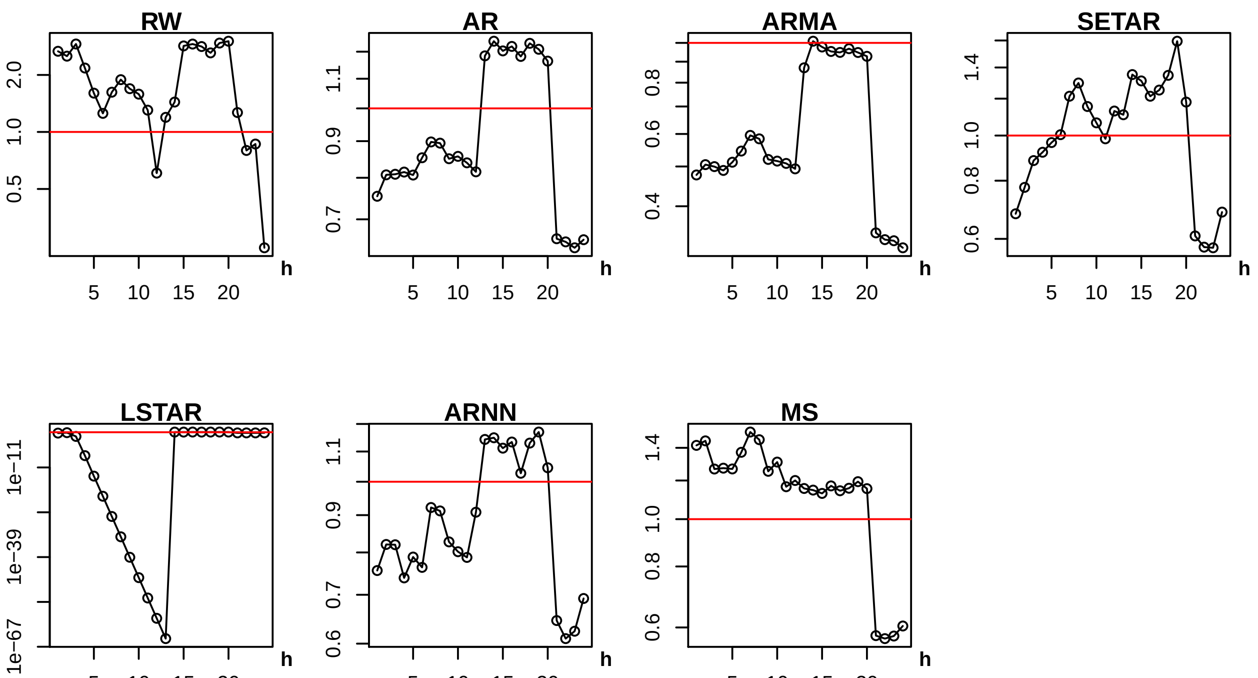


Notes: The $x-$ axis is h whihc denotes the forecast horizon. The $y-$ axis is the ratio of mean squared error (MSE) of the forecasting model in NIT period to mean squared error (MSE) of the forecasting model in IT period.

Turkey



United Kingdom



Notes: The $x-$ axis is h whihc denotes the forecast horizon. The $y-$ axis is the ratio of mean squared error (MSE) of the forecasting model in NIT period to mean squared error (MSE) of the forecasting model in IT period.

2. Essay 2: A Test of structural change of unknown location with wavelets

The primary goal of this essay is to test for structural breaks in the mean of an independently distributed process at an unknown location. A Haar wavelet decomposition additively splits the data into its local weighted averages, ie. the scaling coefficients, and local weighted differences, ie., wavelet coefficients. For a detailed account of Haar wavelets see Gençay et al. (2002) and Percival and Walden (2006). If the process has a constant mean, the variance of the wavelet and scaling coefficients are of equal magnitude. If, however, there is change in the mean of the process, the variances of the wavelet and scaling coefficients will diverge with more allocation to the variance of the scaling coefficients. We use this feature of the wavelet decomposition to design a statistical test for the change in the mean of an independently distributed process.[13] It is through these weighted local differences in moving windows, we construct our statistical test of no structural break under the null hypothesis. We derive its null distribution and demonstrate that it is asymptotically normally distributed.

Since these weighted averages and weighted differences are calculated locally at a given time window and on a moving window scheme, any significant change between such consecutive averages is indicative of a structural break. It is through these weighted local differences, we construct our statistical test of no structural break under the null hypothesis. We derive its null distribution and demonstrate that it is asymptotically normally distributed.

The length of the moving window is determined by the length of the wavelet filter. If the length of the filter is two, such as the Haar filter, localized differences amount to the differences between two consecutive observations. If the length of the filter is four, it is the weighted difference between the last two to the first two observations in

[13] Although, the primary focus of this essay is on structural breaks in *mean* of an independently distributed time series, our framework can be generalized to structural breaks in variance, and structural break in stationary and non-stationary time series.

a window of four observations. The longer the filter, it captures the local structural features of the data more accurately but due to its length boundary issues surface. In our test, we use Haar filter which has a length two and is a good compromise between localization in a local time window but at the same time not costly in terms of boundary treatments.

Structural breaks can be in permanent or in temporary nature. If a structural break is permanent, a change in mean or variance, is permanent to the indefinite future. In temporary breaks, the mean or the variance may shift away from their null values but they revert back to such values after some time. Whether such breaks are in temporary or in permanent nature, they may yield their presence abruptly or gradually. To capture such possibilities, in our Monte Carlo simulations, we model break locations through sinusoidals to allow for abrupt as well as gradual structural breaks. The reason why we primarily focus on smooth multiple structural breaks is twofold. The first is that most economic and financial data exhibit gradual structural changes in a time window and the most abrupt ones are exceptions rather than the rule. The second is that our framework also allows for the abrupt changes and is an encompassing framework.[14] Our Monte Carlo simulations indicate minimal empirical size distortions relative to their nominal ones and significant power improvements in comparison to existing structural break tests.

2.1. Structural break tests.

The literature on structural change tests is extensive. Several tests for structural breaks have been proposed in the literature. Chow (1960) derived F-tests for structural breaks with a known break point. Brown et al. (1975) developed CUSUM and CUSUM squared tests that are also applicable when the time of the break is unknown. More recently, contributions by Ploberger et al. (1989), Hansen (2002), Andrews (1993), Inclan and Tiao (1994), Andrews et al. (1996) and Chu et al. (1996) have extended tests for the presence of breaks to account for heteroskedasticity. Methods for estimating the

[14] The usefulness of modelling structural breaks using this framework was previously emphasized by Ludlow and Enders (2000), Becker et al. (2004), Becker et al. (2006), Ashley and Patterson (2010) and Stengos and Yazgan (2014a,b).

number and timing of multiple break points, as in Bai and Perron (1998, 2003a) and Altissimo and Corradi (2003), have also been developed.[15] These tests are formulated in the context of linear regression models and they use the estimated residuals to detect the departure from the constancy of parameters.

The machinery of our wavelet framework can be described in the following way. The wavelet decomposition yield equal number of localized weighted differences (wavelet coefficients) as the number of data points. We square each wavelet coefficient to obtain its magnitude and take their sample average. We expect this sample average to be equal to the half of the variance of the process under the null when there is no structural change.[16] Whereas the sample average of the squared wavelet coefficients should be significantly smaller than the null average if there are one or more structural breaks. We center and standardize the sample average of squared wavelet coefficients to obtain the null distribution of no structural change.

The competing tests such as cumulative sum (CUSUM), moving sum (MOSUM), approach the problem through cumulative sums of either recursive residuals (one step ahead prediction errors) or OLS residuals. They start from a given window to calculate the sequence of cumulated sums by increasing the window length one by one[17] and reject the null of no structural change when properly standardized supremum of these cumulated sums crosses the critical lines, or when the maximum cumulated sums is sufficiently large.[18] In a Sup-F type test[19] , it is the split of the data into two or more

[15]Bai and Perron (2003a,b) provide heteroskedasticity and serial correlation corrected versions of these tests.

[16]The average of the squared wavelet and the average of squared scaling coefficients equal to the overall variance in a one-level wavelet decomposition.

[17]MOSUM test uses moving window by keeping its size constant.

[18]While the CUSUMs of the recursive residuals, properly standardized, tend in distribution to a standard Wiener process (Sen (1982) and Krämer et al. (1988)), Ploberger et al. (1989) show that the OLS-based CUSUMs tend in distribution to a Brownian bridge. Since OLS residuals sum to zero when there is an intercept in the regression, one therefore cannot expect their cumulated sum to drift off after a structural change, as often happens with recursive residuals, and which provides the rationale for the standard CUSUM test. Since, no matter how large a structural shift has occurred, the cumulated OLS residuals will eventually return to the origin, Ploberger et al. (1989) provide critical lines that are parallel to the horizontal axis (of the plot of cumulated residuals) unlike positively and negatively sloped critical lines of conventional CUSUM test.

[19]The Sup-F test for structural breaks is first put forward by Andrews (1993) and further generalized by Bai and Perron (1998).

blocks, and calculation of the unrestricted and restricted residual sum of squares for all possible sub-samples, for a given length, is necessary to find the supremum on an F-test.

In the presence of multiple structural breaks, MOSUM and CUSUM type tests yield smaller maximum value of their cumulated sums of residuals (CSR). A similar argument applies to the Sup-F test where the difference between sums of squares residuals (SSRs) of the restricted and unrestricted models become small in the presence of multiple structural breaks. These tests rely on larger CSRs (or larger differences between restricted and unrestricted SSRs) to detect structural change and smaller CSRs do not yield high power. The reason for smaller CSR is that OLS estimation goes through the average of such multiple structural breaks (in particular, when such breaks are reverting), which yields smaller residuals underestimating structural change locations. Ours, on the other hand, always operates in a local time window with excellent frequency localization features, the imbalance between squared wavelet and scaling coefficients are preserved, such that the power of our test is not compromised in the presence of multiple structural breaks.

Several recent papers have successfully have demonstrated the usefulness of wavelets in econometric hypothesis testing framework. Recently, Fan and Gençay (2010) propose a unified wavelet spectral approach to unit root testing by providing a spectral interpretation of existing Von Neumann unit root tests. Xue et al. (2010) propose wavelet-based jump tests to detect jump arrival times in high frequency financial time series data. These wavelet-based unit root, cointegration and jump tests have desirable empirical size and higher power relative to the existing tests. Gençay and Gradojevic (2011) utilize wavelets for errors-in-variables estimation.

The outline of this essay is as follows. Section 2 introduces the wavelet-based structural change test and its limiting null distribution. The Monte Carlo simulations are described in Section 3. We conclude afterwards.

2.2. The Wavelet test for structural change

Let a time series $\{y_t\}_{t=1}^{T}$ evolve according to the following data generation process (DGP)

$$y_t = \boldsymbol{\mu}_t + \varepsilon_t \tag{2.1}$$

where ε_t are normally, identically and independently distributed with mean zero and variance σ^2 for $t = 1, \ldots, T$. The structural changes in the mean of y_t occurring at (unknown) dates t_1, t_2,. . .,t_k can be formulated as follows

$$\boldsymbol{\mu}_t = \begin{cases} \mu_1 & \text{for} \quad 1 \leq t \leq t_1, \\ \mu_2 & \text{for} \quad t_1 < t \leq t_2, \\ \vdots & \vdots \\ \mu_z & \text{for} \quad t_{z-1} < t \leq T. \end{cases} \tag{2.2}$$

Under the null hypothesis, it is assumed that there is no structural change in the mean, $H_0 : \mu_1 = \mu_2 = \ldots = \mu_z = \mu$. Under the alternative hypothesis (H_1), there may be one or more break in mean[20] at unknown locations, so that there exists at least one $i \in \{1, \ldots, z\}$ such that $\mu_i \neq \mu$.

2.3. Wavelet and scaling coefficients

Consider the unit scale Haar maximum overlapping discrete wavelet transformation (MODWT) of $\{y_t\}_{t=1}^{T}$ where T is the number of observations. The wavelet and scaling coefficients for this transformation are given by

$$W_t = \frac{1}{2}(y_t - y_{t-1}), \qquad t = (1, 2, \ldots, T, \quad \text{mod } T) \tag{2.3}$$

$$V_t = \frac{1}{2}(y_t + y_{t-1}), \qquad t = (1, 2, \ldots, T, \quad \text{mod } T) \tag{2.4}$$

[20] We use the terminology of structural change in mean or break in mean interchangeably.

Where we assume that the boundary condition is circular.[21] As we approach to the end of the sample, if enough data is not available for a given filter length, we start utilizing data from the beginning of the sample to complete the analysis. This type of boundary treatment is innocuous given that we use Haar filter which has two coefficients and the data is assumed to be stationary.[22]

The wavelet coefficients, $\{W_t\}_{t=1}^T$, capture the behaviour of $\{y_t\}$ in the high frequency band $[\frac{1}{2}, 1]$, while the scaling coefficients, $\{V_t\}_{t=1}^T$, capture the behaviour of $\{y_t\}$ in the low frequency band $[0, \frac{1}{2}]$. Accordingly, the variance (energy) of $\{y_t\}$ is given by the sum of the energies of $\{W_t\}_{t=1}^T$ and $\{V_t\}_{t=1}^T$ where

$$\sum_{t=1}^{T} y_t^2 = \sum_{t=1}^{T} W_t^2 + \sum_{t=1}^{T} V_t^2. \tag{2.5}$$

Equation 2.5 is known as analysis of variance (ANOVA) based on MODWT.

2.4. Statistical properties of wavelet tests

We propose test statistics based on the square averages of the wavelet and scaling coefficients. Consider the variables δ_m^2 and κ_m^2

$$\delta_m^2 = \frac{1}{m} \sum_{t=j}^{m+j} W_t \qquad \kappa_m^2 = \frac{1}{m} \sum_{t=j}^{m+j} V_t^2 \tag{2.6}$$

for $j \in 1, 2, 3, \ldots, T-m$, where m is an arbitrary length of the window where the structural change is sought after. In Equation (2.6), δ_m^2 and κ_m^2 are defined to be average of squared scaling and wavelet coefficients over an interval in the data, re-

[21] $a \mod b$ denotes the congruent modulo b of a where a and b are integers.

[22] The circular boundary treatment may be problematic for trending or nonstationary data processes so that going back to the beginning to complete the analysis may induce superficial structural change due to the level change between the beginning and the end of the sample. The longer the filter, it will require more coefficients and therefore the analysis will be more prone to the boundary coefficients. On the other hand, longer filters may have better localization properties and may be more powerful in identifying the local structural variations the data.

spectively. Notice that, when $m = T$, δ_T^2 and κ_T^2 are reduced to the averages of the squared wavelet and scaling coefficients for the whole sample. In the remainder of this essay, we will explore the statistical properties of δ_T^2 under the null hypothesis of $\mu_1 = \mu_2 = \ldots = \mu_z = \mu$ in Equation (2.2).[23] The following proposition establishes the expected value and variance of δ_T^2 under the null hypothesis of no structural change.

Proposition 2.1. *Under* H_0 , $\mathbb{E}\{\delta_T^2\} = \dfrac{\sigma^2}{2}$. $\quad Var(\delta_T^2) \cong \dfrac{3\sigma^4}{T}$ *for large* T.

Proof. The proof is in Appendix. □

Let $s^2 = (1/T)\sum_{t=1}^{T}(y_t - \bar{y})^2$ be a consistent estimator of σ^2. Accordingly, we center δ_T^2 and work with $\delta_T^2 - \frac{1}{2}s^2$ which has zero expectation under H_0. In the following proposition, we derive the variance of $\delta_T^2 - \frac{1}{2}s^2$.

Proposition 2.2. *Under* H_0, $Var\left(\delta_T^2 - \dfrac{s^2}{2}\right) \cong \dfrac{\sigma^4}{4T}$ *for large* T.

Proof. The proof is in Appendix. □

By normalizing $\delta_T^2 - \dfrac{s^2}{2}$ with its standard deviation, we propose the test statistic $\widetilde{GYO}_W$.

$$\widetilde{GYO}_W = \frac{\left(\delta_T^2 - \dfrac{s^2}{2}\right)}{\sqrt{s^4/4T}} = \sqrt{T}\left(2\frac{\delta_T^2}{s^2} - 1\right) \tag{2.7}$$

The asymptotic distribution of $\widetilde{GYO}_W$ is given in the following proposition:

Proposition 2.3. *Under* H_0 , $\widetilde{GYO}_W \xrightarrow{(d)} N(0,1)$ *as* $T \to \infty$ *where* $\xrightarrow{(d)}$ *denotes convergence in distribution.*

Proof. The proof is in Appendix. □

[23] We will leave the study of κ_m^2 for future research.

The following proposition states the expected value of the test statistics under the alternative hypothesis.

Proposition 2.4. *Under H_1, assuming a single break, let t_1 be the last point after which the process y_t assumes the new mean μ_2 and let $q = t_1/T$ be the fraction of the data with μ_1 where $q \in \{\frac{1}{T}, \frac{2}{T}, \ldots \frac{T-1}{T}\}$. Then,*

$$\mathbb{E}\left\{\widetilde{GYO}_W\right\} = \sqrt{T}\left\{\frac{\sigma^2 + \dfrac{|\mu_1 - \mu_2|^2}{2T}}{\sigma^2 + q(1-q)\,|\mu_1 - \mu_2|^2} - 1\right\}. \tag{2.8}$$

Proof. The proof is in Appendix. □

Corollary 2.5. *For a given break location q and sample size T, $\mathbb{E}\left\{\widetilde{GYO}_W\right\} \to 0$ as $|\mu_1 - \mu_2| \to 0$.*

That is as the break disappears the test statistics return its value under the null as expected. Notice that, for all $q \in \{\frac{1}{T}, \frac{2}{T}, \ldots, \frac{T-1}{T}\}$ and $T \in \{3, 4, \ldots\}$, it is true that $q(1-q) > 1/(2T)$ and the quotient at the right-hand-side of Equation 2.8 is always smaller than the denominator and, accordingly, $\mathbb{E}\left\{\widetilde{GYO}_W\right\} < 0$. Hence,

Corollary 2.6. *For a given break location q and size $(|\mu_1 - \mu_2|)$, $\mathbb{E}\left\{\widetilde{GYO}_W\right\} \to -\infty$ as $T \to \infty$ and $\mathbb{E}\left\{\widetilde{GYO}_W\right\} \in (-\infty; 0)$ for $T > 2$*

Corollary 2.7. *For a given break location q and sample size T, as $|\mu_1 - \mu_2| \to \infty$, $\mathbb{E}\left\{\widetilde{GYO}_W\right\}$ decreases and approaches to $-\sqrt{T}\left[1 - \frac{1}{2Tq(1-q)}\right]$.*

Therefore, a ccording to Corollary 2.6, our test statistic lives in $(-\infty; 0]$ and increases its power as T increases for a given location and size of the break.

In Figure 2.1 we illustrate the behaviour of the expected value of the test statistics, under the alternative, for given break sizes. The graph on panel (a) illustrates the case for a small break whereas the panel (b) refers to a larger break. Given the size of the break, it is apparent that increase in q or T dramatically carries $\mathbb{E}\left\{\widetilde{GYO_W}\right\}$ away from zero to negative region. As T and q increases, the larger the break size, the higher the rate of decrease in the value of the test statistics. This indicates the fact that the power is higher when the size of the break is larger. Notice also that the behaviour of the test statistics is symmetrical around $q = 0.5$ (i.e. it assumes the same value when we depart the same distance from $q = 0.5$ to both direction), where the test attains highest power for a given T and q.

[Insert Figure 2.1 here]

In Figure 2.2, we depict the test statistics for given location of the break. The panel (a) shows the case for small q, i.e. the break is at the beginning of the data, whereas the left panel is for a break at the centre of the data. As T and $|\mu_1 - \mu_2|$ increases $\mathbb{E}\left\{\widetilde{GYO_W}\right\}$ decreases and approaches to $-\sqrt{T}$ for both values of q shown in Figure 2.2. However, the expected value of our test statistics decreases faster when $q = 0.5$ compared to $q = 0.05$, for given magnitudes of $|\mu_1 - \mu_2|$ and T, as shown in the panel (b). The behaviour of the test statistics is symmetrical across $q = 0.5$ in this case too, as established in previous (Figure 2.1).

[Insert Figure 2.2 here]

In the following proposition we generalize Proposition 2.4 to multiple breaks:

Proposition 2.8. *Under H_1, suppose there are z mean breaks, let t_i be the last points after which the process y_t assumes the new mean μ_{i+1} such that $i \in \{0, 1, \ldots, z\}$, $t_0 = 0$, $q_i = (t_i - t_{i-1})/T$, $i \in \{1, 2, \ldots, z\}$ be the fractions of the data with constant means.*

Then,

$$\mathbb{E}\left\{\widetilde{GYO}_W\right\} = \sqrt{T}\left\{\frac{\sigma^2 + (2T)^{-1}\sum_{i=0}^{z-1}|\mu_{i+1}-\mu_i|^2}{\sigma^2 + \sum_{i=1}^{z} q_i\mu_i^2 - \sum_{i=1}^{z}\sum_{j=1}^{z} q_i q_j \mu_i \mu_j} - 1\right\}.$$

Proof. The proof is in Appendix. □

As in the case of single break, the break locations t_i and the distribution of μ_i determine the expected value of $\widetilde{GYO}_W$ under the alternative hypothesis.

Suppose that μ_j is different from the rest such that $\mu_l = \mu$, $l \neq j$ and $\mu_j = \mu + \epsilon$ for a non-zero ϵ. Then,

$$\sum_{i=1}^{z} q_i\mu_i^2 = \mu^2\sum_{i=1}^{z} q_i + 2q_j\mu\epsilon + q_j\epsilon^2 = \mu^2 + 2q_j\mu_j\epsilon + q_j\epsilon^2 > \mu^2 + 2q_j\mu\epsilon + q_j^2\epsilon^2 = (\mu + q_j\epsilon)^2$$
$$= \left(\sum_{i=1}^{z} \mu q_i + q_j\epsilon\right)^2 = \left(\sum_{i=1}^{z} \mu_i q_i\right)^2 = \sum_{i=1}^{z}\sum_{j=1}^{z} q_j\mu_j q_i\mu_i$$

since $0 < q_j < 1$ by assumption. Hence the test statistics approaches to $-\infty$ as $T \to \infty$.

2.5. Monte Carlo simulations

In the Monte Carlo simulations, our framework allows for single and multiple structural breaks. Following Becker et al. (2004), Becker et al. (2006) and Ludlow and Enders (2000)[24] , we let structural breaks to be smooth or abrupt. Since a Fourier expansion is capable of approximating absolutely integrable functions to any desired degree of accuracy, smooth or abrupt breaks can be approximated with such sinusoidals by the appropriate choice of the frequency mix. Therefore we can capture the structural

[24]See also Ashley and Patterson (2010) and Stengos and Yazgan (2014a,b).

breaks in the mean of Equation (2.1) via the following function.[25]

$$\boldsymbol{\mu}_t \cong \rho + \alpha \sum_{i=1}^{n} \left\{ (2i-1)^{-1} \sin \left[\frac{2\pi(2i-1)kt}{T} \right] \right\} \tag{2.9}$$

where n is the number of frequencies included in the approximation, k represents a particular frequency and α indicate their size (amplitude). The frequency coefficient, k, alone, determines how many breaks and whether these breaks are in temporary or permanent nature. Moreover if a single frequency is used ($n = 1$), the transitions tend to be smooth whereas higher n values facilitate abrupt temporary or permanent breaks.

In Figure 2.3, sample paths of y_t are displayed for permanent and temporary breaks. In the top panel when $k = 0.9$, we observe a single permanent structural break (permanent in the sense that the series do not return their initial mean level). In the second panel when $k = 1.4$, the series are subject to temporary single breaks. The remaining panels depict multiple breaks both in temporary and abrupt nature. Moreover, Figure 2.3 also illustrates sample paths of y_t for higher $n = 1$ values where breaks happen more abruptly. When n assumes a value as high as 4 (the third column) breaks become significantly abrupt. For higher values of n such as 128 the occurrence of structural breaks become visibly sudden.

[Insert Figure 2.3 here]

We compare the performance of $\widetilde{GYO}_W$ test with three well known structural break tests statistics: Sup-F test, OLS-based cumulative CUSUM test, and MOSUM test. The Sup-F test is first proposed by Andrews (1993) and generalized by Bai and Perron (1998). It can be obtained by calculating a series of F statistics over all potential change points in the data and taking their supremum. The Sup-F is a regression based test. It uses the null hypothesis of no break in the regression equation at time i against the alternative of two different regressions over the intervals $(1, i)$ and $(i+1, T)$, where

[25] We approximate the discrete square function in Equation (2.2) with the continuous one in Equation (2.9).

i is a point in $(T_0,\, T - T_0)$. T_0 is generally set to $h \times T$, h is a bandwidth parameter $h \in (0,\, 1)$. Hansen (1997) provided an approximation for the asymptotic $p-$values.

The OLS-based CUSUM test statistics of Ploberger and Kramer (1992) is given by[26]

$$S_C(t) = \frac{1}{\hat{\sigma}\sqrt{T}} \sum_{i=1}^{\lfloor Tt \rfloor} \hat{u}_i; \quad 0 \leq t \leq 1 \tag{2.10}$$

where $\hat{u}$ are the OLS residuals from the model under the null and $\hat{\sigma}$ is the standard deviation of the estimated residuals. Instead of using cumulative sums of the same residuals, another possibility to detect a structural change is to analyze moving sums of residuals. OLS-based MOSUM test of Chu et al. (1995) considers this possibility. This test does not contain the sum of all residuals up to a certain time t but the sum of a fixed number of residuals in a data window whose size is determined by the bandwidth parameter h. The test statistic is computed as

$$S_M(t) = \frac{1}{\hat{\sigma}\sqrt{T}} \sum_{i=\lfloor N_T t \rfloor + 1}^{\lfloor N_T t \rfloor + \lfloor Th \rfloor} \hat{u}_i \qquad (0 \leq t \leq 1 - h) \tag{2.11}$$

where $N_T = (T - \lfloor Th \rfloor)/(1 - h)$.

We reject the "no structural break" null hypothesis when $S_C(t)$ is greater than the critical boundary curve $\lambda_{CUSUM}(t, \alpha)$ for at least one t. The same rejection criteria applies to MOSUM test for the statistic $S_M(t)$ and the boundary $\lambda_{MOSUM}(t, h, \alpha)$. Boundaries $\lambda_{CUSUM}(t, \alpha)$ and $\lambda_{MOSUM}(t, h, \alpha)$ are calculated according to Zeileis (2005).

As for the wavelet test, it "deposits" the structural breaks into the scaling coefficient variance in a local window. As the number of structural breaks increase, the larger percentage of the variance is allocated to the scaling coefficients. This, in turn, makes the contribution of the wavelet variance marginal relative to the overall variance. We should keep in mind that the sum of the variance of the wavelet and scaling coef-

[26] $\lfloor Tt \rfloor$ is the nearest integer to Tt smaller than or equal to Tt.

ficients is equal to the overall variance. Since the scaling coefficient variance becomes the driver of the overall variance, in the presence of multiple structural breaks, we use the marginalization of the wavelet coefficient variance as a test for structural break.

There may be a comparison between the MOSUM test and the wavelet test because they both operate in local windows. One possible drawback of the MOSUM test is that it assigns equal weight to each observation in its local window. The wavelet test, on the other hand, uses time-frequency optimized weights to sum the information.

The empirical size and power properties of the wavelet coefficient test statistics, $\widetilde{GYO}_W$, is calculated for Equation (2.9) under varying parameter sets of α, k, and T by assuming $\rho = 0$. The experiment is first run for smooth structural breaks by setting $n = 1$, then carry on with abrupt breaks for higher values of n up to 1024. Since $\widetilde{GYO}_W$ test is a left-tailed test, we use one-sided $p-$ values in our Monte Carlo simulations. α values are allowed to vary in the range of $0.2, 0.4, \ldots, 2$. The k parameter varies between 0 and 5 by gradual increments of $\Delta k = 0.05$ between 0.0 and 2.0, and $\Delta k = 0.1$ between 2.1 and 5.0. We consider the cases when data length, T, is 50, 100, and 200. We run $10,000$ replications for each combination of α, T, and k. In each combination we use the same pseudo-random generated i.i.d standard normal series to simulate ε_t. The rejection frequencies are calculated at both 1, 5, 10 percent level of significance. We report size corrected empirical power calculations.[27]

[Insert Table 2.1 and Figure 2.4 here]

For the case of smooth breaks ($n = 1$), the results of the Monte Carlo experiments are illustrated both in Table 2.1 and Figure 2.4. In Table 2.1, we provide the size adjusted rejection frequencies of Sup-F, OLS-based CUSUM, OLS-based MOSUM, and $\widetilde{GYO}_W$ test statistics for $T = 50$. The values in the first panel, where $k = 0$ corresponds to case of pure white-noise process, are the size values of the tests. The remaining values, reported in the following panels, are the size corrected power of the

[27]For the sake of brevity, a subset of Monte Carlo simulation results are reported.

tests statistics. Each panel corresponds to selected α values for different ks.

In Figure 2.4, we plot the size corrected rejection frequencies of the tests statistics for a continuum of k values when $T = 50$.[28] Similar to Table 2.1, each panel plot corresponds to selected α values. The values at the intersection of the $y-$axis are the size values of the tests. As k varies in the $x-$axis each line is the size adjusted power of that test statistics.

The information contained in the table and figure clearly points out that, for all parameters, at the region where $k > 1$, $\widetilde{GYO}_W$ has more empirical power than other tests. On the other hand, for small values of αs and in a small region in which k is smaller or slightly higher than 1, Sup-F and, to a small extent, MOSUM test perform better than $\widetilde{GYO}_W$ although they loose the power quickly as k increases. This implies that our competitor do relatively better in their power when there is a single break. Otherwise, our test has higher relative power for higher k values which is when there are multiple structural breaks. As the number of smooth structural breaks in the data increases, $\widetilde{GYO}_W$ performs much more better as a structural break test.

In Figure 2.5, to question the performance of our test in a larger time horizon, we set $T = 200$. We obtain similar results to those of reported above. We observe that wavelet test and Sup-F and MOSUM tests have similar performance, in particular with higher value of α. In the bottom panel, this region is where k is close to zero. These results indicate that, our test has similar power properties with their competitors even in the case of single breaks. On the other hand, when multiple breaks spanned in a larger time horizon, the others, especially MOSUM tests do not loose their power as quickly when k increases. MOSUM test keeps its high power up to very large number of ks, i.e. for reasonably large number of breaks. It can be concluded that, when T is large, generally speaking MOSUM and our test performs similarly. This results make sense since both tests make use of local information.

[Insert Figure 2.5 here]

[28]The size correction is for 5 percent level of significance.

We also consider the size and power properties of the tests when the breaks are in abrupt nature. In Figure 2.6 and Figure 2.7, we illustrate the results when $n = 128$.[29] Overall, in the case of abrupt breaks, our test not only preserves its empirical size but also increases its power compared to smooth case. In the case of single breaks, except for small values of α and for small T, our test starts competing for power with the Sup-F test and others. Even for small T and when there are multiple breaks, our test appears to be the best performer. Similar to the smooth break case, when T gets larger, the other tests also become more powerful when multiple breaks, and our test ceases to be the best performer except for cases with large number of structural breaks.

[Insert Figure 2.6 and Figure 2.7 here]

[29] As mentioned above we increased n up to 1024 and obtained qualitatively similar results.

Table 2.1: Size corrected powers of the $\widetilde{GYO}_W$, Sup-F, CUSUM, and MOSUM tests for smooth breaks

	$\widetilde{GYO}_W$						Sup-F						CUSUM						MOSUM					
T	50		100		200		50		100		200		50		100		200		50		100		200	
Significance	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
(Size) k=0.0	0.01	0.04	0.01	0.06	0.01	0.05	0.01	0.06	0.01	0.06	0.01	0.05	0.00	0.03	0.00	0.03	0.01	0.04	0.00	0.01	0.00	0.01	0.00	0.02
	$\alpha = 0.8$																							
k=0.6	0.71	0.88	0.94	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.57	0.83	1.00	1.00	1.00	1.00
k=1.0	0.73	0.89	0.95	0.99	1.00	1.00	0.83	0.95	1.00	1.00	1.00	1.00	0.83	0.95	1.00	1.00	1.00	1.00	0.57	0.84	0.99	1.00	1.00	1.00
k=1.2	0.79	0.92	0.97	0.99	1.00	1.00	0.39	0.79	0.98	1.00	1.00	1.00	0.18	0.68	0.94	1.00	1.00	1.00	0.61	0.86	1.00	1.00	1.00	1.00
k=1.6	0.71	0.87	0.95	0.98	1.00	1.00	0.18	0.45	0.71	0.89	0.99	1.00	0.03	0.22	0.41	0.77	0.93	1.00	0.44	0.72	0.97	1.00	1.00	1.00
k=2.0	0.73	0.91	0.95	0.98	1.00	1.00	0.13	0.34	0.63	0.81	0.98	1.00	0.04	0.20	0.36	0.68	0.88	0.99	0.41	0.74	0.97	1.00	1.00	1.00
k=4.0	0.65	0.84	0.95	0.99	1.00	1.00	0.00	0.02	0.03	0.10	0.13	0.36	0.00	0.02	0.01	0.08	0.05	0.24	0.05	0.15	0.21	0.51	0.75	0.98
k=5.0	0.65	0.84	0.95	0.99	1.00	1.00	0.00	0.02	0.03	0.10	0.13	0.36	0.00	0.02	0.01	0.08	0.05	0.24	0.05	0.15	0.21	0.51	0.75	0.98
	$\alpha = 1.2$																							
k=0.6	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.99	1.00	1.00	1.00	1.00
k=1.0	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.90	0.99	1.00	1.00	1.00	1.00
k=1.2	1.00	1.00	1.00	1.00	1.00	1.00	0.76	0.99	1.00	1.00	1.00	1.00	0.41	0.98	1.00	1.00	1.00	1.00	0.91	0.99	1.00	1.00	1.00	1.00
k=1.6	0.99	1.00	1.00	1.00	1.00	1.00	0.35	0.73	0.97	1.00	1.00	1.00	0.06	0.39	0.79	0.98	1.00	1.00	0.73	0.97	1.00	1.00	1.00	1.00
k=2.0	1.00	1.00	1.00	1.00	1.00	1.00	0.23	0.58	0.93	0.99	1.00	1.00	0.05	0.30	0.72	0.93	1.00	1.00	0.72	0.96	1.00	1.00	1.00	1.00
k=4.0	0.99	1.00	1.00	1.00	1.00	1.00	0.00	0.01	0.03	0.11	0.27	0.59	0.00	0.01	0.01	0.07	0.07	0.39	0.04	0.16	0.31	0.71	0.99	1.00
k=5.0	0.65	0.84	0.95	0.99	1.00	1.00	0.00	0.02	0.03	0.10	0.13	0.36	0.00	0.02	0.01	0.08	0.05	0.24	0.05	0.15	0.21	0.51	0.75	0.98

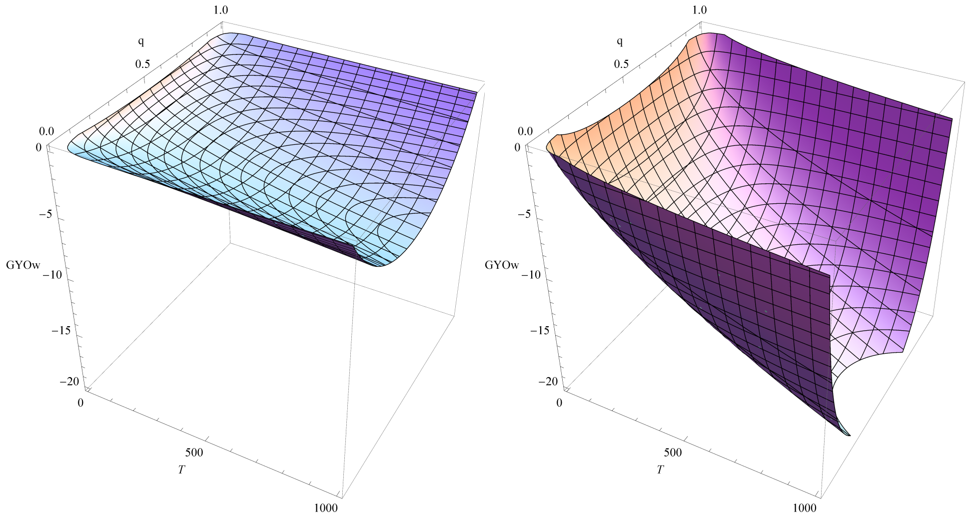
Continued on next page

Table 2.1 continued

	$\widetilde{GYO}_W$						Sup-F						CUSUM						MOSUM					
T	50		100		200		50		100		200		50		100		200		50		100		200	
	$\alpha = 1.6$																							
k=0.6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
k=1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00
k=1.2	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00	0.63	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
k=1.6	1.00	1.00	1.00	1.00	1.00	1.00	0.48	0.87	1.00	1.00	1.00	1.00	0.08	0.52	0.96	1.00	1.00	1.00	0.90	1.00	1.00	1.00	1.00	1.00
k=2.0	1.00	1.00	1.00	1.00	1.00	1.00	0.34	0.77	0.99	1.00	1.00	1.00	0.05	0.39	0.90	1.00	1.00	1.00	0.91	1.00	1.00	1.00	1.00	1.00
k=4.0	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.02	0.10	0.38	0.76	0.00	0.00	0.00	0.05	0.07	0.51	0.02	0.12	0.35	0.82	1.00	1.00
k=5.0	0.65	0.84	0.95	0.99	1.00	1.00	0.00	0.02	0.03	0.10	0.13	0.36	0.00	0.02	0.01	0.08	0.05	0.24	0.05	0.15	0.21	0.51	0.75	0.98
	$\alpha = 2.0$																							
k=0.6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
k=1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
k=1.2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.79	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
k=1.6	1.00	1.00	1.00	1.00	1.00	1.00	0.62	0.94	1.00	1.00	1.00	1.00	0.09	0.65	0.99	1.00	1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00
k=2.0	1.00	1.00	1.00	1.00	1.00	1.00	0.44	0.87	1.00	1.00	1.00	1.00	0.04	0.46	0.98	1.00	1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00
k=4.0	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.02	0.08	0.48	0.88	0.00	0.00	0.00	0.03	0.05	0.58	0.01	0.07	0.33	0.88	1.00	1.00
k=5.0	0.65	0.84	0.95	0.99	1.00	1.00	0.00	0.02	0.03	0.10	0.13	0.36	0.00	0.02	0.01	0.08	0.05	0.24	0.05	0.15	0.21	0.51	0.75	0.98

Notes: The numbers represent the fraction of the cases that H_0 is rejected in $10,000$ replications, for each test (size corrected powers). The DGP is $y_t = \boldsymbol{\mu}_t + \varepsilon_t$, where $\boldsymbol{\mu}_t$ is given in Equation (2.9) with $\rho = 0$, $n = 1$, and ε_t is $iid \sim N(0,1)$. The bandwidth parameter, h, is taken as 0.15 for Sup-F and MOSUM test. Sup-F test, due to Andrews (1993), does not use the heteroskedasticity and autocorrelation consistent (HAC) kernel.

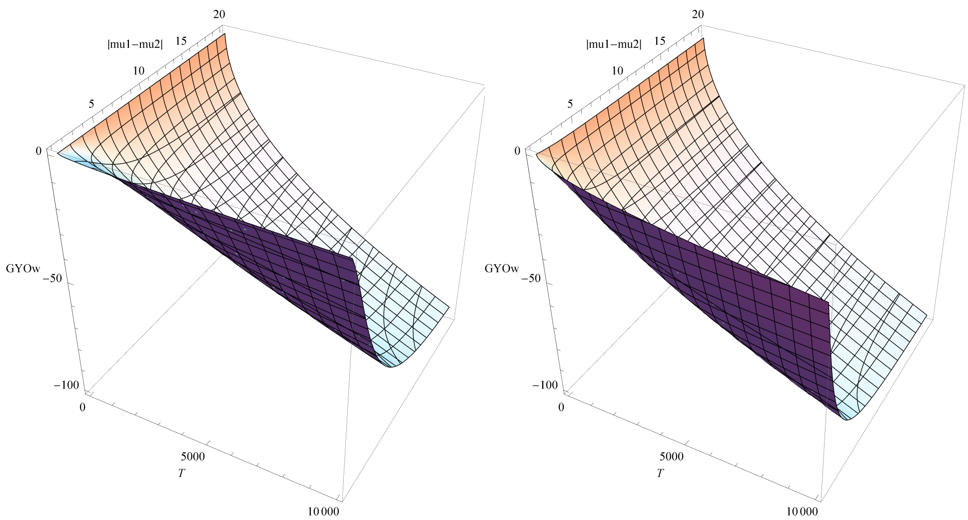
Figure 2.1: Behaviour of $\mathbb{E}\{GYO_W\}$ for a given size.



(a) $|\mu_1 - \mu_2| = 1$ (b) $|\mu_1 - \mu_2| = 3$

On panel (a) the mean difference (the size of the break) is 1 whereas on panel (b) the size of the break is 3 . $T \in \mathbb{N}$ is the length of the series. q is the exact location of the break expressed as percentage of the length of the data, ie, $q = \tau/T$ where $\tau \in \{1, 2, \ldots, T\}$. GYO_w is the expectation of the test statistic defined in Equation 2.8. The plots are produced by Mathematica®.

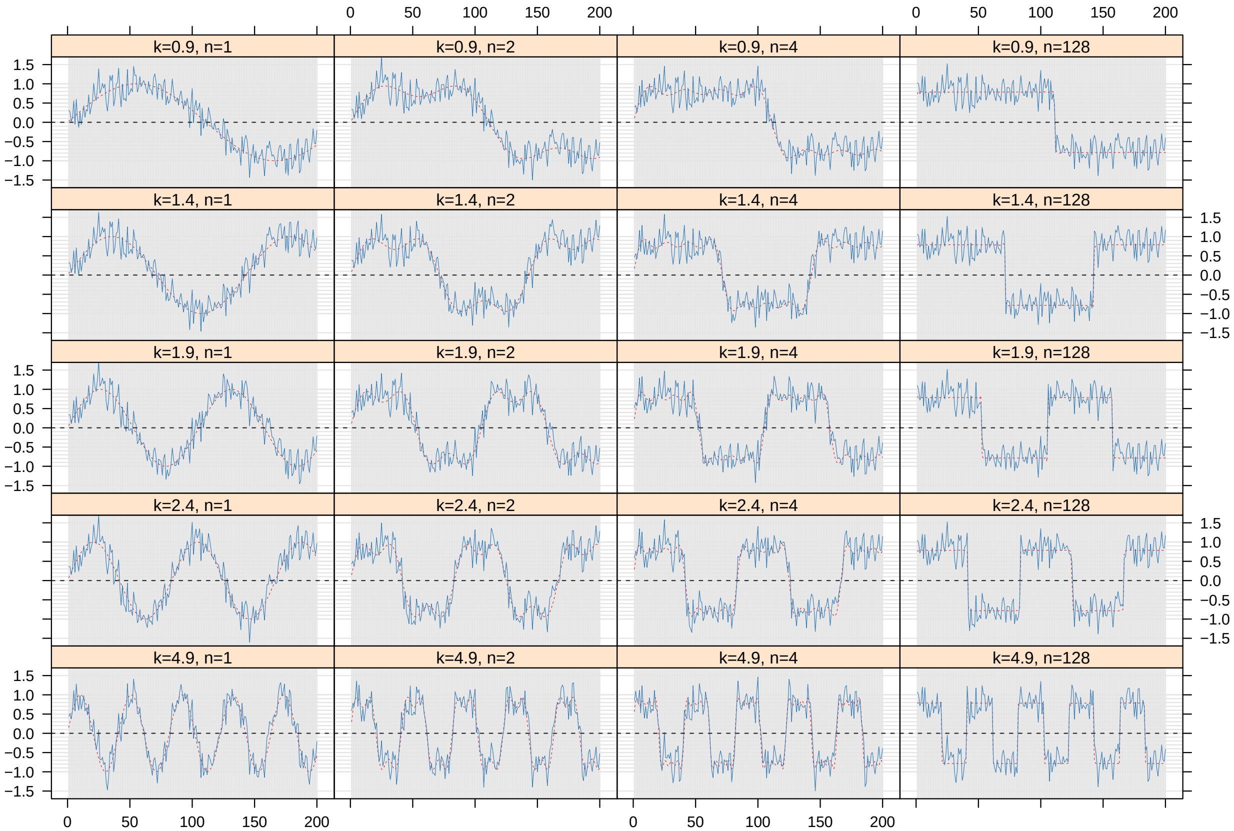
Figure 2.2: Behaviour of $\mathbb{E}\{GYO_W\}$ for a given location.



(a) $q = 0.05$ (b) $q = 0.5$

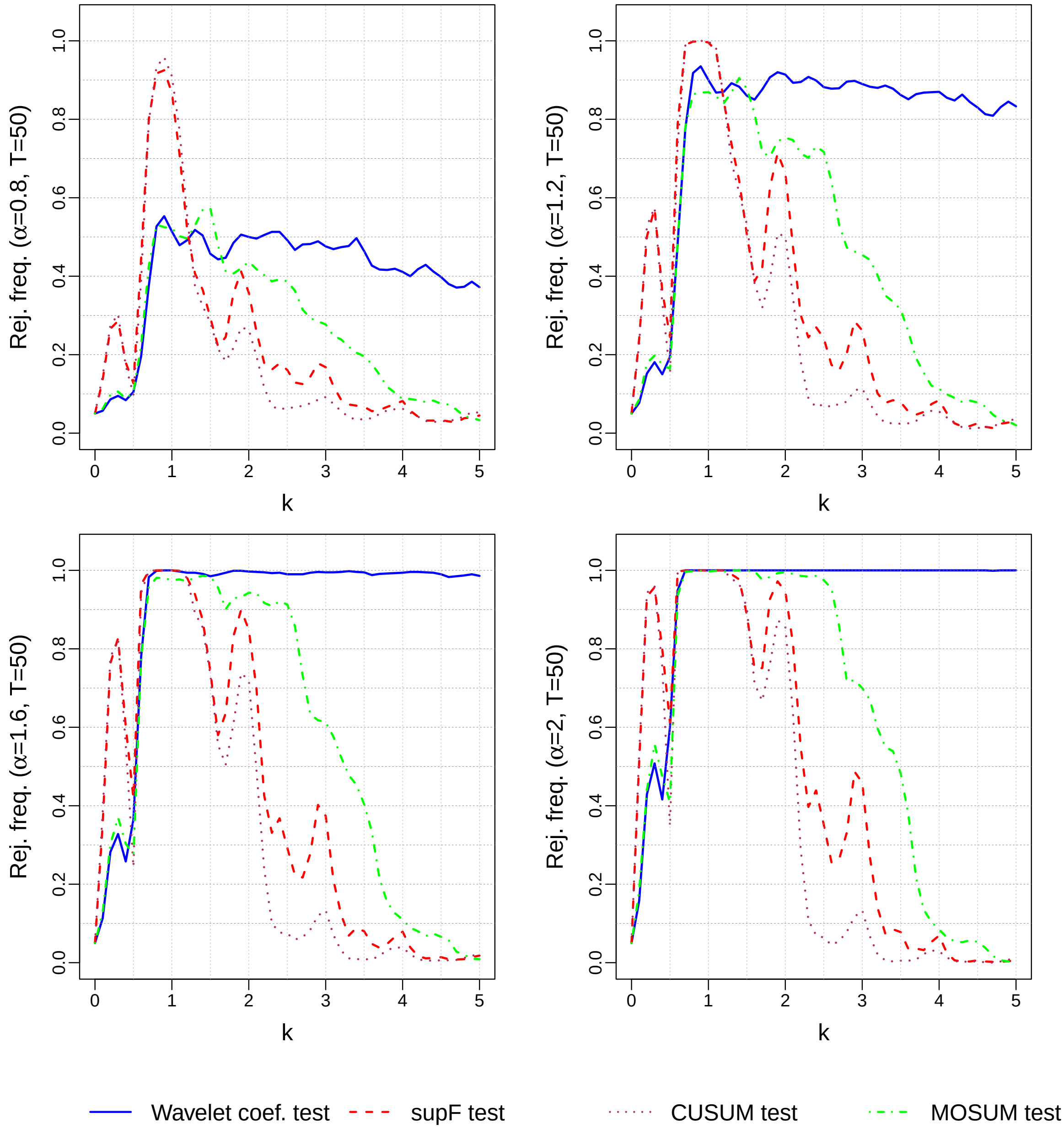
$T \in \mathbb{N}$ is the length of the series, q is the exact location of the break expressed as percentage of the length of the data, ie, $q = \tau/T$ where $\tau \in \{1, 2, \ldots, T\}$. $|mu1 - mu2|$ is the size of the break. GYO_w is the expectation of the test statistic defined in Equation 2.8. The plots are produced by Mathematica®.

Figure 2.3: Sample paths of smooth/abrupt and permanent/temporary breaks



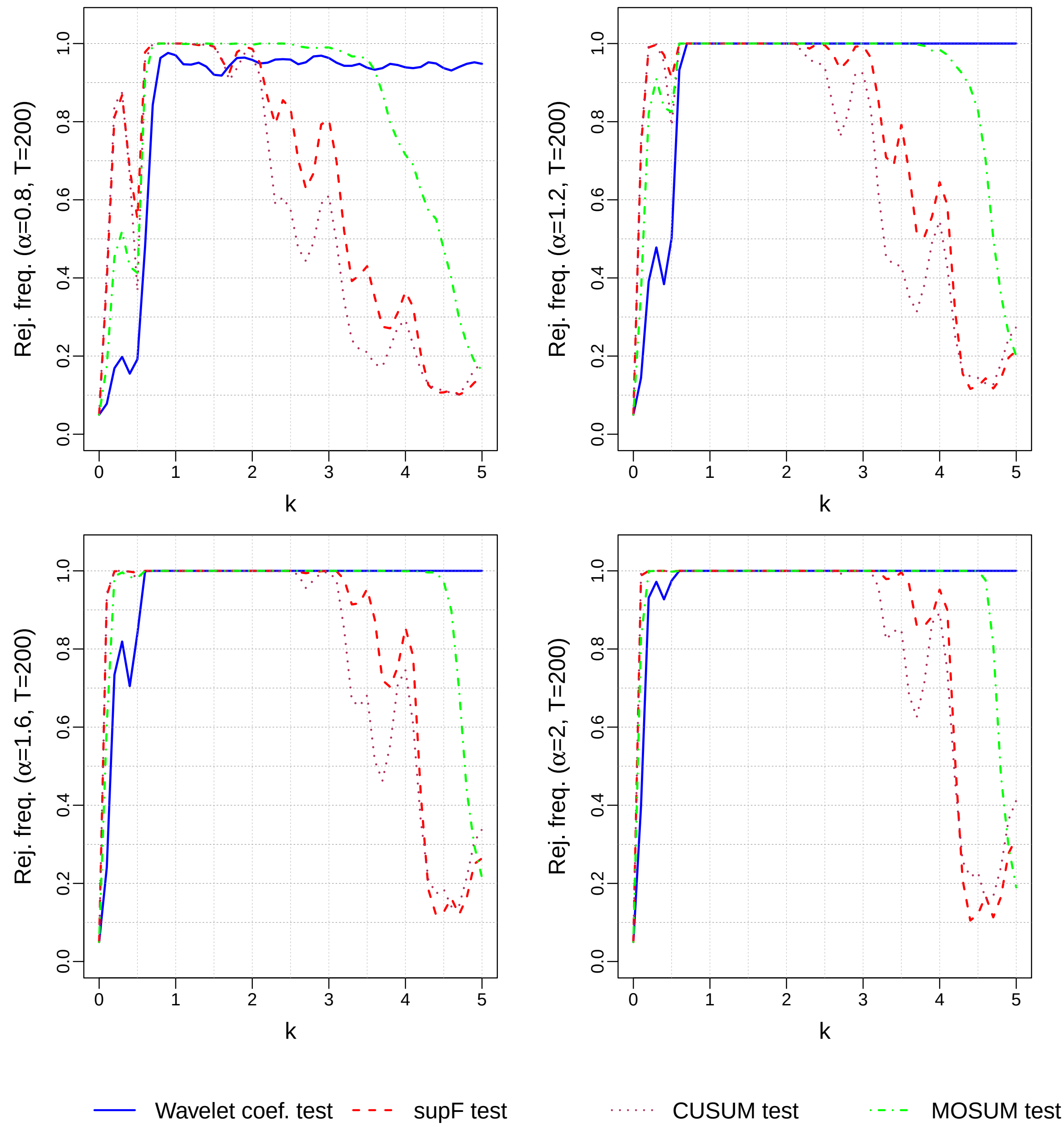
Notes: The dashed line is $\boldsymbol{\mu}_t$, as given in Equation (2.9) with $\alpha = 1$ and $\rho = 0$, and the solid line is $y_t = \boldsymbol{\mu}_t + \varepsilon_t$, where ε_t is $iid \sim N(0, 0.25)$, and $T = 200$.

Figure 2.4: Size corrected powers of the $\widetilde{GYO}_W$, Sup-F, CUSUM, and MOSUM tests for smooth breaks ($n = 1, T = 50$).



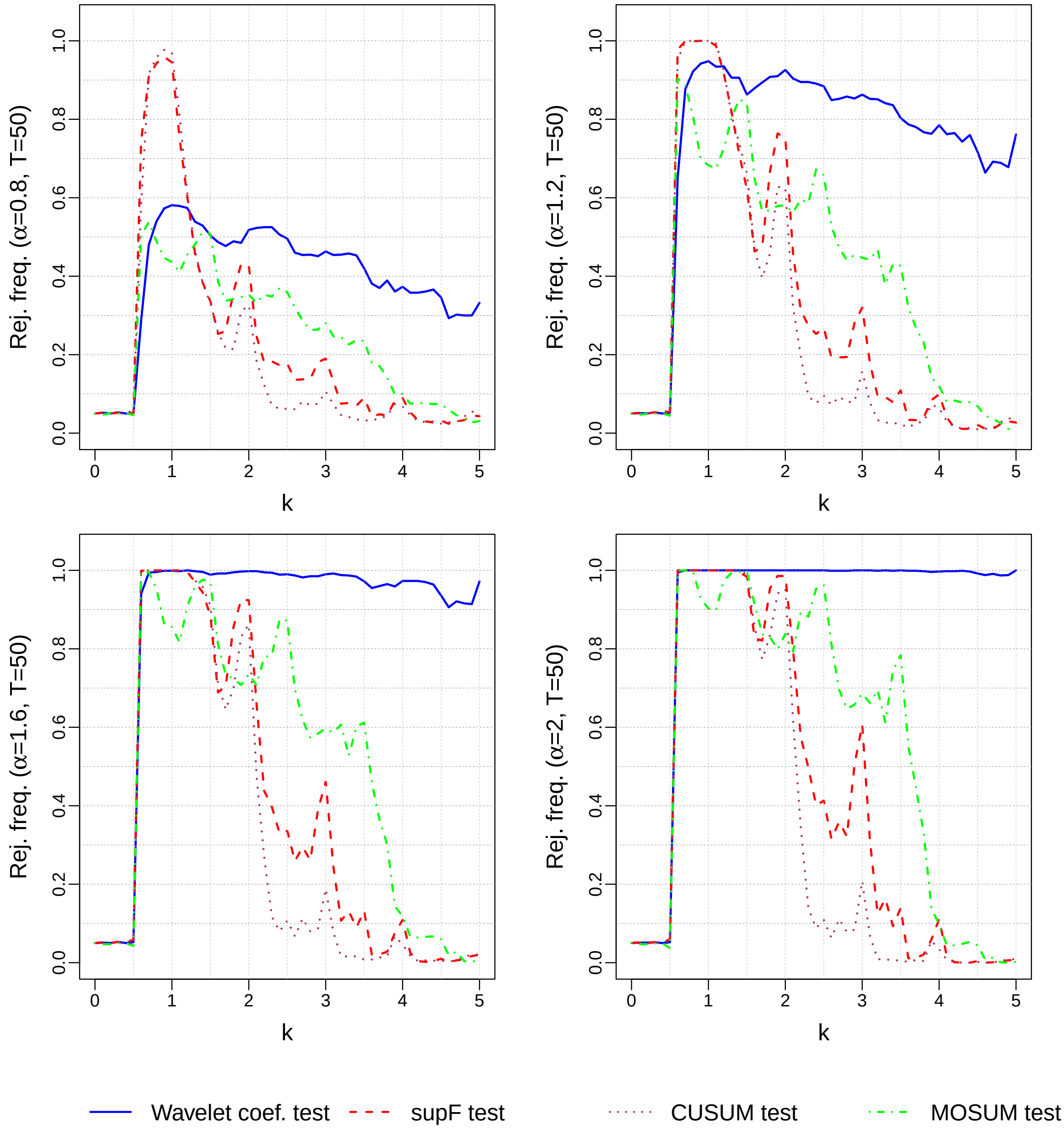
Notes: $x-$axis are the k values in Equation (2.9) and at the $y-$axis are size corrected (at the 5% level) empirical powers. The DGP is $y_t = \boldsymbol{\mu}_t + \varepsilon_t$, where $\boldsymbol{\mu}_t$ is given in Equation (2.9) with $\rho = 0$, and ε_t is $iid \sim N(0, 1)$. The bandwidth parameter, h, is taken as 0.15 for Sup-F and MOSUM test. Sup-F test, due to Andrews (1993), does not use the heteroskedasticity and autocorrelation consistent (HAC) kernel.

Figure 2.5: Size corrected powers of the $\widetilde{GYO}_W$, Sup-F, CUSUM, and MOSUM tests for smooth breaks ($n = 1, T = 200$).



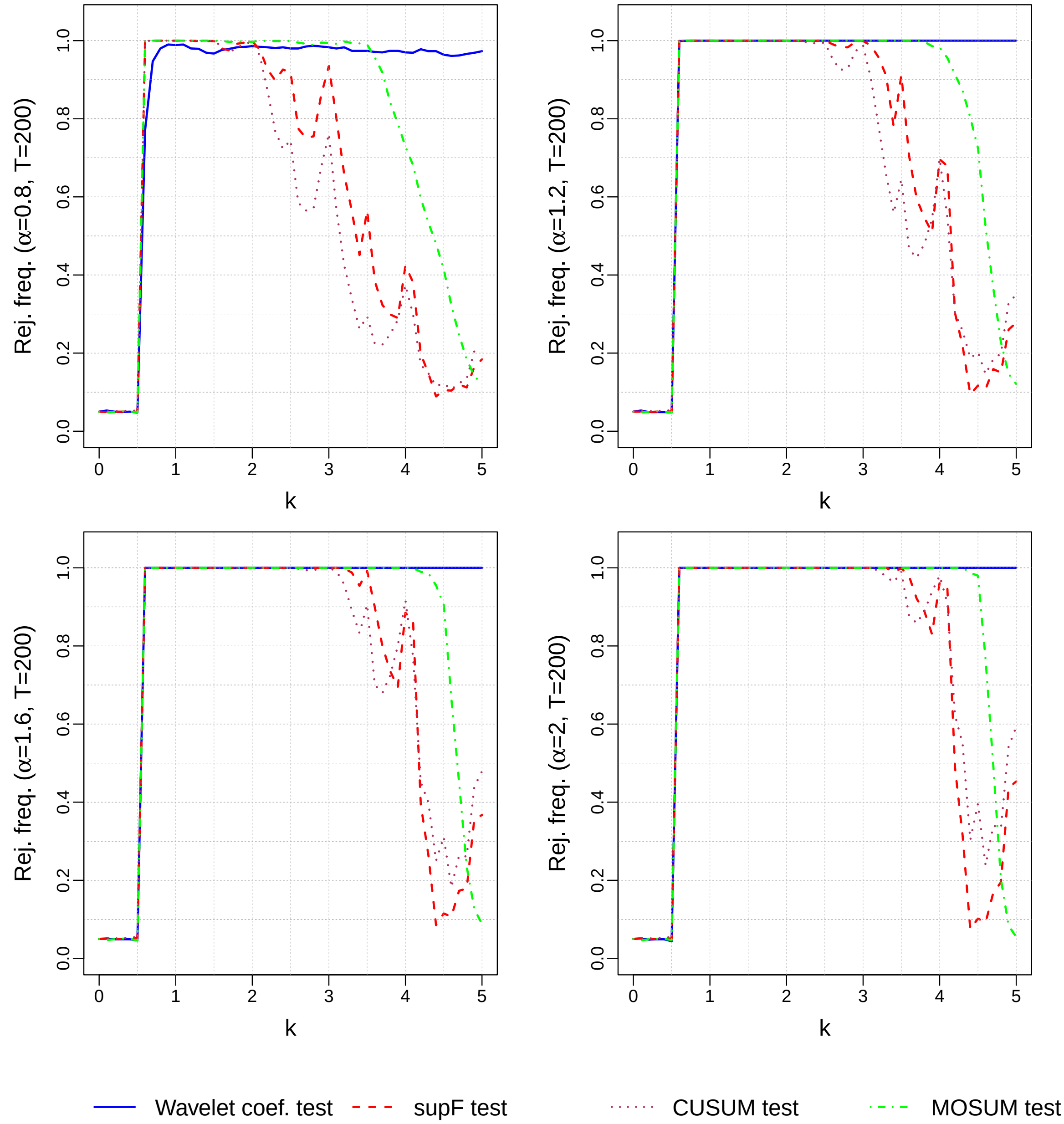
Notes: x−axis are the k values in Equation (2.9) and at the y−axis are size corrected (at the 5% level) empirical powers. The DGP is $y_t = \boldsymbol{\mu}_t + \varepsilon_t$, where $\boldsymbol{\mu}_t$ is given in Equation (2.9) with $\rho = 0$, and ε_t is $iid \sim N(0, 1)$. The bandwidth parameter, h, is taken as 0.15 for Sup-F and MOSUM test. Sup-F test, due to Andrews (1993), does not use the heteroskedasticity and autocorrelation consistent (HAC) kernel.

Figure 2.6: Size corrected powers of the $\widetilde{GYO}_W$, Sup-F, CUSUM, and MOSUM tests for abrupt breaks ($n = 128, T = 50$).



Notes: $x-$axis are the k values in Equation (2.9) and at the $y-$axis are size corrected (at the 5% level) empirical powers. The DGP is $y_t = \boldsymbol{\mu}_t + \varepsilon_t$, where $\boldsymbol{\mu}_t$ is given in Equation (2.9) with $\rho = 0$, and ε_t is $iid \sim N(0,1)$. The bandwidth parameter, h, is taken as 0.15 for Sup-F and MOSUM test. Sup-F test, due to Andrews (1993), does not use the heteroskedasticity and autocorrelation consistent (HAC) kernel.

Figure 2.7: Size corrected powers of the $\widetilde{GYO}_W$, Sup-F, CUSUM, and MOSUM tests for abrupt breaks $(n = 128, T = 200)$.



Notes: $x-$axis are the k values in Equation (2.9) and at the $y-$axis are size corrected (at the 5% level) empirical powers. The DGP is $y_t = \boldsymbol{\mu}_t + \varepsilon_t$, where $\boldsymbol{\mu}_t$ is given in Equation (2.9) with $\rho = 0$, and ε_t is $iid \sim N(0, 1)$. The bandwidth parameter, h, is taken as 0.15 for Sup-F and MOSUM test. Sup-F test, due to Andrews (1993), does not use the heteroskedasticity and autocorrelation consistent (HAC) kernel.

3. Essay 3: Markov regime switching in mean and in fractional integration parameter

The use of fractional or long memory methods has been extensive in econometrics as they have been found to be quite effective in describing the behavior of many macroeconomic and financial data (Lobato and Velasco, 2000; Ding et al., 1993). In particular, recent findings suggest that long memory phenomenon is observed in LIBOR (Cajueiro and Tabak, 2005b, 2007a), interest rate (Cajueiro and Tabak, 2007b), trading volume (Lobato and Velasco, 2000; Lux and Kaizoji, 2007), and volatility of returns (Lobato and Velasco, 2000; Cajueiro and Tabak, 2005c; Granger and Ding, 1994) -albeit controversial findings in terms of long memory behavior are present for the stock returns(Cajueiro and Tabak, 2005a; Limam, 2003; Willinger et al., 1999).

Similarly, since the seminal work of Hamilton (1989) the Markov–switching model has been a popular vehicle to analyze economic phenomena that are likely to obey regime changes. Recently, Tsay and Härdle (2009) (hereafter TH) has combined the two approaches in a unified framework by introducing a Markov-switching-ARFIMA (MS-ARFIMA) process which extends the hidden Markov model with a latent state variable, allowing for the different regimes to have different degrees of long memory. Recent papers in the literature have been concerned with changes in the persistence of a univariate time series, considering primarily a shift from a unit root process $[I(1)]$ to a stationary process $[I(0)]$ or vice versa at some unknown date over the sample under consideration. In that strand of the literature the analysis centers on the properties of estimators (and tests) in these extreme cases, see Perron (2006) for a survey of testing procedures and Chong (2001) and Kejriwal and Perron (2010) for some recent results on the properties of the break estimators. These models however deal with the extreme dichotomy of $[I(1)]$ versus $[I(0)]$ and do not allow for long memory and fractional integration.

Models that allow for different long memory regimes have been used in the literature but the regime switching is forced by an observable state variable as opposed to the latent nature of the state variable in the MS–ARFIMA model, see Haldrup and Nielsen (2006). The main motivation behind the MS–ARFIMA model of TH has been the observation by Diebold and Inoue (2001) that a mixture model of latent Markov-switching mean can generate long memory dependence. In other words, structural change and long memory may be easily confused in estimation. Hence, the main emphasis of the TH approach has been to disentangle the impact of long memory dependence on the estimates of the latent regime parameters in the MS–ARFIMA framework. It is worth noting that in this context the direct application of the EM algorithm used by Hamilton (1989) and Hamilton (1990) is not applicable due to the non-Markovian nature of the model. One of the main contributions of TH is the use of the Viterbi algorithm to estimate the MS–ARFIMA model. This algorithm is capable of tackling the hidden Markov process observed in a general ARFIMA framework, something that is not generally possible with the EM algorithm. The algorithm discussed above is not the only choice practically available in estimation of the parameters of Markov regime switching model. Other widely used algorithms are, for example, the forward-backward algorithm, the Baum-Welch algorithm and the BCJR algorithm. The forward-backward algorithm is an inference algorithm for hidden Markov models which computes the posterior marginals of all hidden state variables given a sequence of observations. The Baum-Welch algorithm is a special forward-backward algorithm and also relies on EM algorithm. The BCJR algorithm, also called Maximum posteriori probability (MAP) decoder, named after Bahl et al. (2006), relies on maximization of posteriori probabilities. These algorithm has several modified and tailored versions[30] .

However, the TH analysis did not consider regime switching in the long memory parameter but only in the mean both in their simulations and their empirical application. Yet long memory parameter regime switching may have similar contamination effects on the estimation of the mean parameters as it would be the case in the opposite case considered by TH. This possibility in fact was indicated by Diebold and Inoue (2001) concern mentioned above. In this essay we explicitly consider the case

[30]Examining the performance of these other algorithms is beyond the scope of the present note and is left for future research.

where the long memory parameter may be subject to regime switching, while the mean can be unchanging or regime switching itself. By allowing the direct impact of long memory regime switching on the mean parameters (whether in regime switching mode or not) would allow us to asses the possible contamination and impact that such long memory structural break may have on the mean estimates. We conduct a Monte Carlo simulation that considers contamination to be going both ways between mean and long memory parameter breaks. Our results suggest that in addition to the findings by TH that only considered breaks in the mean parameter, breaks in the long memory parameter can have similar effects on the (in sample) fitting ability of the model irrespective of the presence of breaks in the mean parameter, confirming the contamination concern raised by Diebold and Inoue (2001).

The rest of the essay is organized as follows. In the next section we present the model that we analyze as well as the Viterbi algorithm we use following TH. We then proceed to present our simulations that analyze the possible contamination that could run from long memory parameter breaks to mean parameter breaks and vice-versa. Finally we conclude.

3.1. Viterbi maximum likelihood EM algorithm

Consider the fractionally integrated process y_t defined as

$$(1-L)^d y_t = \mu + \varepsilon_t \tag{3.1}$$

where ε_t is white noise, L is the lag operator, d is the fractional integration parameter and μ is real valued drift.

Let $y_1 = 0$ and $\hat{y} = \phi_{t1} y_{t-1} + \ldots + \phi_{t1} y_1$ be the one-step predictors of the process

y_t. The coefficients has the following recursive structure:

$$\begin{aligned}
\phi_{tt} &= \left[\gamma(t) - \sum_{i=1}^{t-1}\gamma(t-i)\right]\nu_{t-1} \\
\phi_{tj} &= \phi_{tj} - \phi_{tt}\phi_{t-1t-j}, \qquad j = 1,\ldots,T-1 \\
\nu_{tj} &= \nu_{t-1}(1-\phi_{tt}^2), \qquad t = 1,\ldots,T-1
\end{aligned} \tag{3.2}$$

where $\gamma(t)$ is the autocovariance function of order t, $\nu_0 = \gamma_0$. In a fractionally integrated processes such as specified in Equation 3.1, we have

$$\gamma(t) = \frac{\Gamma(1-2d)\Gamma(d-t)}{\Gamma(d)\Gamma(1-d)\Gamma(1-d-t)}$$

where $\Gamma(x)$ is the gamma function.

Defining the prediction error $e_t = y_t - \hat{y}_t$, then $e_t = Ly$ where

$$L = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ -\phi_{11} & 1 & 0 & \ldots & 0 \\ -\phi_{22} & -\phi_{21} & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi_{T-1T-1} & -\phi_{T-1T-2} & -\phi_{T-1T-3} & \ldots & 1 \end{bmatrix} \tag{3.3}$$

Setting $\Gamma_\theta = LDL'$, where D is diagonal matrix with $diag(\nu_0,\ldots,\nu_{T-1})$, we have $\det\Gamma_\theta = \prod_{j=1}^{n}\nu_{j-1}$. Consequently, $Y'\Gamma_\Theta^{-1}Y = e'^{-1}e$, where $Y = (y_1,\ldots,y_T)$ and $e_t = y_t - \hat{y}_t$. Then, the log-likelihood function may be written as

$$\mathcal{L}(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\log\nu_{t-1} - \frac{1}{2}\sum_{t=1}^{T}\frac{e_t^2}{\nu_{t-1}} \tag{3.4}$$

for the model specified in Equation 3.1.

Now, we consider a 2−state homogeneous Markov chain S_t taking values 1 or 2. Let $S_{t t=1}^{T}$ be the latent sample path of the Markov chain. At each time point t, S_t can assume only an integer value of 1 or 2, and its transition probability matrix is

$$\mathcal{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \tag{3.5}$$

where $p_{ij} = \mathbb{P}(S_t = j \mid S_{t-1} = i)$ and $p_{i1} + p_{i2} = 1$ for $i = 1, 2$.

We specify the corresponding regime switching fractionally integrated process as

$$(1 - L)^{d^{s_t}} y_t = \mu^{s_t} + \varepsilon_t^{s_t} \tag{3.6}$$

with the unobserved state vector $S = (s_t)_{1 \leq t \leq T}$.

In this study, we are going to employ the Viterbi algorithm to estimate the unobserved state vector.S. The Viterbi algorithm is forward decoding procedure widely used in signal processing problems with Hidden Markov Models (HMM) specification. The main idea behind this algorithm is to “decode” the sequence of states s_t in the Markov chain iteratively starting from time 1 to T. Given the previous state, s_{t-1}, and the unconditional likelihood function for the observations up to t, $(y_s)_{0 \leq s \leq t}$ and a parameter vectors ξ_1 and ξ_2, it enables us to choose the most probable state at t, i.e., s_t.

In particular, specifiying the unconditional likelihood functions for both states at t as

$$l_t^j = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(y_t - \hat{y}_t^{(j)})^2}{2\sigma_j^2}} \tag{3.7}$$

where $\hat{y}_t^{(j)} = \hat{y}_t^{(1)}(\xi_j,\ (y_u)_{1 \leq u \leq t})$ and $\xi_j = (\mu_j, d_j, \sigma_j,)$, $j = 1, 2$ and it can be calculated by $Ly + y$. Given the likelihood function,, the parameter vectors ξ^j, and the transition probabilities P, by Viterbi specification we can iteratively estimate the path of the states s_t as follows:

$$\mathcal{S}_t = \underset{s_t \in \{1,2\}}{\arg\max}(l_t^{s_t} P(s_t|\mathcal{S}_{t-1})) \tag{3.8}$$

given s_0 and the initial probabilities $\pi_j = \mathbb{P}(s_0 = j)$, and $\hat{y_0}$, where $\mathcal{S}_\sqcup$ is the Viterbi decoded state at t.

Then, the estimation of parameter vectors ξ_j and the transition probabilities p_{ij} is obtained by maximixzing the following log-likelihood function:

$$\mathcal{L}(\xi_1, \xi_2, y) = -\frac{1}{2}\sum_{t=1}^{T} \log \sigma_j - \frac{1}{2}\sum_{t=1}^{T} \frac{e_t^2}{\sigma_j} + \sum_{t=1}^{T} \mathbb{P}(\mathcal{S}_t|\mathcal{S}_{t-1}) \tag{3.9}$$

The computational complexity of this maximization is $\mathcal{O}(n^2)$. In comparison, the EM algorithm of Hamilton (1989), which is exhaustively used in estimation of Markov regime switching models, cannot be employed in this case (when d is allowed to be state depended) for one main reason: the number of possible state paths the EM algorithm is going to account for is n^T (since $n = 2$ in our case, it is 2^T, hence, implying a computational complexity of $\mathcal{O}(2^T)$. Thus, EM is computationally much more demanding in comparison to Viterbi algorithm, which is $\mathcal{O}(n^2)$ in computational complexity. Furthermore, Tsay (2009) argues, that the model we examine above "... cannot be written in a state-space form due to the presence of a fractional differencing parameter, implying that we cannot apply the EM algorithm considered in Hamilton (1990) ... because the non-Markovian nature of the model prevents us from using the results in (4.2) of Hamilton (1990)" (Tsay, 2009, page. 3).

TH made a similar argument that the EM algorithm proposed by Hamilton (1989) is unstable for estimating the Markov regime switching d. We confirmed that by conducting a small scale computation exercise using the EM algorithm[31] .

[31]However, , we made a naïve experiment to estimate d_{s_t} by Hamilton (1989) method, for a series generated by the process $(1-L)^{d^{s_t}}\ y_t = \mu^{s_t} + \varepsilon_t$ with the true values of $d_1 = 0, d_2 = 0.5, \mu_1 = 0, \mu_2 = 0$

In the context of the present model, even though the log-likelihood function given by Equation 3.9, is well-defined, the properties of the estimators of the long-memory parameter and the mean level under regime switching are not derived. We conjecture that an analytical derivation of the asymptotic distribution of these estimators might entail an analytical framework such as Hansen (2000).

We now proceed with the application of the Viterbi algorithm in a Monte Carlo simulation where we will study breaks in both long memory and mean parameters.

3.2. The Monte Carlo experiment

The Monte Carlo experiments that we conduct in this essay is to consider the simplest possible framework of analysis an MS-ARFIMA $(0, d, 0)$ to do the analysis as we are interested in isolating the effects of spillovers between the regime shifts of the two simple parameters in the model, μ and d. We carry out the Monte Carlo experiment for the Cartesian product space $\mu_1 \times \mu_2 \times d_1 \times d_2$ of true μ_1, μ_2, d_1, d_2 vales such that $\mu_i \in \{0, 0.5, 1.0\}$, $d_i \in \{-0.50, 0.00, 0.25, 0.50, 0.75, 1.00\}$ for $i = 1, 2$.

First, we generated the series y_t

$$(1 - L)^{d^{s_t}} y_t = \mu^{s_t} + \varepsilon_t$$

where $s_t = 1$ if $t \leq T/2$ and 2 otherwise; ε_t are *iid* and distributed $N(0, 0.25)$; T is 256. d_1, d_2, μ_1, and μ_2 are selected such that $d_i \in \{-0.50, 0.00, 0.25, 0.50, 0.75, 1.00\}$, $\mu_i \in \{0.0, 0.5, 1.0\}$ for $i = 1, 2$.

Then, we obtain the estimates of $\widehat{d_1}, \widehat{d_2}, \widehat{\mu_2}, \widehat{\mu_2}$, and $\widehat{\sigma_{\varepsilon t}}$ values by the procedure described above. We conduct 1000 replications for the (d_1, d_2, μ_1, μ_2) quadruple.

---

and $\varepsilon_t \sim iidN(0, 0.25)$. We obtained $\widehat{d_1} = 0.142, \widehat{d_2} = 0.411, \widehat{\mu_2} = -0.12, \widehat{\mu_2} = 0.42$, and $\sigma^2_{\varepsilon_t} = 0.32$. From a computational perspective, it also took much longer to compute the estimators than the Viterbi algorithm did; with a 2.7 GHz i7 single thread processor it takes 1.8 times longer time than that of the Viterbi algorithm computation.

The results are presented in Tables 1 to 4 and Figure 1. Table 1 simply presents the average parameter estimates over the total number of replications, Table 2 presents the root mean square errors for the parameter estimates, Table 3 the mean absolute biases and finally Table 4 the in sample root mean squared errors of the model fit. The results of Table 4 are best seen in Figure 1 and it becomes clear that whether there is a break in the mean (different values of μ_1, μ_2) or not the model fit seems to be unaffected (as measured by the root mean squared error). In other words we observe that structural breaks in the long memory parameter produce similar in sample estimation patterns irrespective of the presence of breaks in the mean parameter. Hence, these results confirm the original concern raised by Diebold and Inoue (2001), something that was not possible to be seen in the work of TH who only considered in his simulation study breaks in the mean but not in the long memory parameter d. The upshot of our analysis is that one has to be careful in claiming the occurrence of structural breaks in such models as the presence of such breaks may be difficult to disentangle. For that purpose one would need to develop more appropriate joint testing procedures that would be able to discern the nature of possible such breaks. The results also confirm that as d, the long memory parameter gets close to unity, the estimates of the mean level parameter μ tend to become inconsistent as expected in the presence of unit roots.

Table 3.1: True parameters and estimated parameters by the Monte Carlo study.

d	$(\mu_1, \mu_2) = (0, 0)$				$(\mu_1, \mu_2) = (0, 0.5)$				$(\mu_1, \mu_2) = (0, 1)$			
	d_1	d_2	μ_1	μ_2	d_1	d_2	μ_1	μ_2	d_1	d_2	μ_1	μ_2
(-0.5,-0.5)	0.080	-0.344	-0.083	0.011	-0.505	-0.335	-0.042	0.563	-0.472	-0.363	0.032	0.918
(-0.5, 0)	-0.566	-0.156	-0.065	0.149	-0.407	-0.161	0.099	0.532	-0.172	0.068	0.015	0.905
(-0.5, 0.25)	-0.518	-0.235	-0.095	0.158	-0.455	-0.161	0.099	0.545	-0.472	0.078	0.015	0.905
(-0.5, 0.5)	-0.580	0.530	-0.135	0.003	-0.627	0.445	-0.014	0.524	-0.595	0.388	0.013	0.896
(-0.5, 0.75)	-0.510	0.730	-0.091	0.003	-0.627	0.725	-0.014	0.558	-0.595	0.788	0.010	0.913
(-0.5, 1)	-0.415	0.906	0.047	0.172	-0.338	1.080	0.161	0.717	-0.466	0.844	-0.017	1.036
(0.0, 0.0)	-0.096	-0.017	-0.240	-0.024	-0.256	0.141	-0.079	0.562	-0.109	0.095	-0.057	0.880
(0.0, 0.25)	-0.106	-0.252	-0.014	0.008	0.144	0.254	-0.079	0.562	0.110	0.245	-0.055	0.895
(0.0, 0.5)	0.028	0.425	-0.082	-0.099	0.015	0.496	-0.111	0.594	0.108	0.553	-0.004	0.927
(0.0, 0.75)	0.017	0.785	-0.082	-0.099	0.115	0.725	-0.111	0.524	0.095	0.714	-0.074	0.927
(0.0, 1.0)	0.058	1.008	0.127	-0.071	0.282	0.963	-0.175	0.573	0.041	0.971	-0.094	1.387
(0.5, 0.5)	0.524	0.518	-0.136	0.056	0.542	0.536	-0.074	0.529	0.448	0.549	0.089	1.207
(0.5, 0.75)	0.556	0.768	-0.106	0.046	0.552	0.752	-0.074	0.519	0.476	0.746	0.029	1.207
(0.5, 1)	0.465	0.934	-0.140	0.213	0.528	1.036	-0.071	0.528	0.513	1.132	-0.013	1.155
(.25,1)	0.235	0.964	-0.119	0.115	0.215	1.014	-0.073	0.582	0.234	1.175	-0.053	1.154
(.75,1)	0.765	0.957	-0.136	0.103	0.765	1.217	-0.071	0.608	0.713	1.107	-0.031	1.168
(1, 1)	1.149	1.188	-0.015	-0.232	0.826	1.004	0.082	0.681	0.950	1.198	-0.139	0.996

Table 3.2: Root mean squared errors for the parameters

d	$(\mu_1, \mu_2) = (0, 0)$				$(\mu_1, \mu_2) = (0, 0.5)$				$(\mu_1, \mu_2) = (0, 1)$			
	d_1	d_2	μ_1	μ_2	d_1	d_2	μ_1	μ_2	d_1	d_2	μ_1	μ_2
(-0.5,-0.5)	0.580	0.156	0.083	0.011	0.005	0.165	0.042	0.063	0.028	0.137	0.032	0.082
(-0.5, 0)	0.066	0.156	0.065	0.149	0.093	0.161	0.099	0.032	0.328	0.068	0.015	0.095
(-0.5, 0.25)	0.018	0.485	0.095	0.158	0.045	0.411	0.099	0.045	0.028	0.172	0.015	0.095
(-0.5, 0.5)	0.080	0.030	0.135	0.003	0.127	0.055	0.014	0.024	0.095	0.112	0.013	0.104
(-0.5, 0.75)	0.010	0.020	0.091	0.003	0.127	0.025	0.014	0.058	0.095	0.038	0.010	0.087
(-0.5, 1)	0.085	0.094	0.047	0.172	0.162	0.080	0.161	0.217	0.034	0.156	0.017	0.036
(0.0, 0.0)	0.096	0.017	0.240	0.024	0.256	0.141	0.079	0.062	0.109	0.095	0.057	0.120
(0.0, 0.25)	0.106	0.502	0.014	0.008	0.144	0.004	0.079	0.062	0.110	0.005	0.055	0.105
(0.0, 0.5)	0.028	0.075	0.082	0.099	0.015	0.004	0.111	0.094	0.108	0.053	0.004	0.073
(0.0, 0.75)	0.017	0.035	0.082	0.099	0.115	0.025	0.111	0.024	0.095	0.036	0.074	0.073
(0.0, 1.0)	0.058	0.008	0.127	0.071	0.282	0.037	0.175	0.073	0.041	0.029	0.094	0.387
(0.5, 0.5)	0.024	0.018	0.136	0.056	0.042	0.036	0.074	0.029	0.052	0.049	0.089	0.207
(0.5, 0.75)	0.056	0.018	0.106	0.046	0.052	0.002	0.074	0.019	0.024	0.004	0.029	0.207
(0.5, 1)	0.035	0.066	0.140	0.213	0.028	0.036	0.071	0.028	0.013	0.132	0.013	0.155
(.25,1)	0.015	0.036	0.119	0.115	0.035	0.014	0.073	0.082	0.016	0.175	0.053	0.154
(.75,1)	0.015	0.043	0.136	0.103	0.015	0.217	0.071	0.108	0.037	0.107	0.031	0.168
(1, 1)	0.149	0.188	0.015	0.232	0.174	0.004	0.082	0.181	0.050	0.198	0.139	0.004

Note: Root mean sqared errors are computed as $\sqrt{N^{(-1)} \sum_i^N (\widehat{x}_i - x)^2}$ where x may be d_1, d_2, μ_1, μ_2.

Table 3.3: Absolute estimation errors

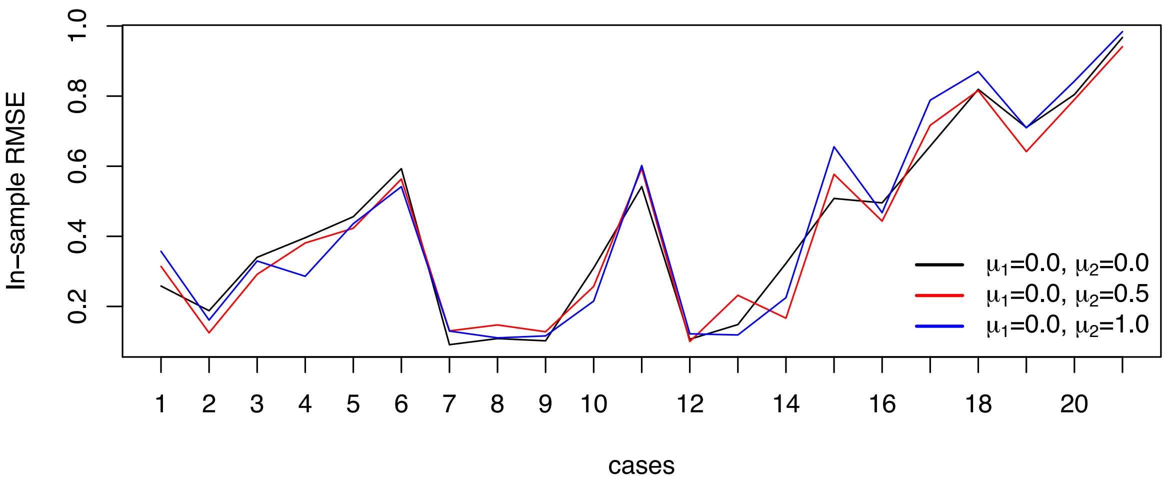
d	$(\mu_1, \mu_2) = (0, 0)$				$(\mu_1, \mu_2) = (0, 0.5)$				$(\mu_1, \mu_2) = (0, 1)$			
	d_1	d_2	μ_1	μ_2	d_1	d_2	μ_1	μ_2	d_1	d_2	μ_1	μ_2
(-0.5,-0.5)	0.580	0.158	0.130	0.101	0.025	0.166	0.107	0.119	0.035	0.139	0.104	0.132
(-0.5, 0)	0.070	0.157	0.119	0.179	0.096	0.163	0.141	0.107	0.328	0.072	0.097	0.138
(-0.5, 0.25)	0.029	0.485	0.139	0.188	0.053	0.412	0.140	0.112	0.035	0.173	0.103	0.139
(-0.5, 0.5)	0.084	0.036	0.170	0.101	0.129	0.058	0.100	0.102	0.098	0.116	0.098	0.145
(-0.5, 0.75)	0.025	0.032	0.134	0.102	0.129	0.034	0.099	0.114	0.098	0.045	0.100	0.133
(-0.5, 1)	0.088	0.096	0.111	0.199	0.163	0.083	0.188	0.239	0.040	0.158	0.102	0.107
(0.0, 0.0)	0.101	0.032	0.260	0.104	0.257	0.143	0.129	0.115	0.111	0.098	0.116	0.157
(0.0, 0.25)	0.108	0.502	0.101	0.100	0.146	0.028	0.128	0.114	0.113	0.022	0.115	0.146
(0.0, 0.5)	0.037	0.081	0.127	0.139	0.030	0.021	0.150	0.135	0.110	0.057	0.098	0.125
(0.0, 0.75)	0.031	0.041	0.130	0.140	0.117	0.045	0.151	0.105	0.098	0.042	0.124	0.123
(0.0, 1.0)	0.062	0.025	0.161	0.125	0.283	0.050	0.202	0.124	0.047	0.039	0.136	0.400
(0.5, 0.5)	0.035	0.029	0.168	0.114	0.048	0.043	0.127	0.107	0.059	0.053	0.133	0.229
(0.5, 0.75)	0.062	0.028	0.147	0.112	0.057	0.023	0.126	0.104	0.035	0.021	0.106	0.229
(0.5, 1)	0.042	0.070	0.172	0.232	0.038	0.047	0.122	0.104	0.030	0.135	0.097	0.185
(.25,1)	0.027	0.043	0.156	0.153	0.041	0.026	0.124	0.130	0.025	0.177	0.111	0.184
(.75,1)	0.029	0.050	0.168	0.141	0.026	0.218	0.125	0.148	0.043	0.110	0.103	0.197
(1, 1)	0.151	0.190	0.097	0.252	0.175	0.021	0.124	0.208	0.053	0.199	0.171	0.097

Note: Absolute estimation errors computed as $N^{(-1)} \sum_i^N |\widehat{x_i} - x|$ where x is d_1, d_2, μ_1, μ_2.

Table 3.4: In-sample mean squared errors

Cases	(d_1, d_2)	$(\mu_1, \mu_2) = (0, 0)$	$(\mu_1, \mu_2) = (0, 0.5)$	$(\mu_1, \mu_2) = (0, 1)$
1	(-0.5,-0.5)	0.2582	0.3142	0.3573
2	(-0.5, 0)	0.1879	0.1247	0.1611
3	(-0.5, 0.25)	0.3402	0.2915	0.3296
4	(-0.5, 0.5)	0.3960	0.3810	0.2860
5	(-0.5, 0.75)	0.4560	0.4230	0.4360
6	(-0.5, 1)	0.5930	0.5634	0.5417
7	(0.0, 0.0)	0.0909	0.1302	0.1297
8	(0.0, 0.25)	0.1083	0.1473	0.1104
9	(0.0, 0.5)	0.1020	0.1278	0.1160
10	(0.0, 0.75)	0.3099	0.2571	0.2152
11	(0.0, 1.0)	0.5420	0.5930	0.6020
12	(0.25, 0.25)	0.1064	0.1000	0.1219
13	(0.25, 0.5)	0.1483	0.2318	0.1186
14	(0.25, 0.75)	0.3227	0.1665	0.2251
15	(0.25, 1.0)	0.5081	0.5770	0.6555
16	(0.5,0.5)	0.4953	0.4434	0.4674
17	(0.5, 0.75)	0.6573	0.7166	0.7883
18	(0.5, 1)	0.8193	0.8158	0.8699
19	(0.75, 0.75)	0.7105	0.6416	0.7095
20	(0.75, 1.0)	0.8044	0.7900	0.8426
1	(1, 1)	0.9674	0.9410	0.9839

Figure 3.1: Average in-sample RMSE of 1000 replications for different (d_1, d_2, μ_1, μ_2) quadruple cases.



4. Essay 4: Testing the persistence in convergence with bi-variate and multivariate d estimators.

In a recent paper Stengos and Yazgan (2014a)[32] use a long memory analytical framework to examine the convergence hypothesis based on the estimation of d, the parameter that describes the underlying (long-memory) process and determines the speed of convergence of output (GDP per capita) gaps between different economies. The main finding of that paper is that although the long memory framework of analysis is much richer than a simple $I(1)/I(0)$ alternative, which produces a simple absolute divergence and rapid convergence dichotomy, the latter seems to be sufficient to capture the behavior of the gaps in per capita GDP levels and growth rates. The former produces a pattern of divergence whereas the latter produces a pattern of rapid convergence. Overall, it was found that any evidence of mean reversion and long memory was not strong enough, which is in contrast to some previous work in the literature that also uses a long-memory framework to analyze convergence, see Dufrénot et al. (2012). However, all previous research has relied on the univariate estimation of the long-memory parameter d, without accounting for possible correlations among the different output gaps (country differences), an issue that we want to address in this essay by employing a multivariate estimation and testing framework, following Shimotsu (2007). Using the latter methodology and in contrast to the results obtained using univariate estimation, we find evidence of mean reversion and slow (stationary) convergence. This evidence suggests that the overwhelming evidence in favor of divergence found in the literature may be partly explained by the use of methods that do not allow for interdependence among the persistence parameter. The latter acts as a moderating mechanism against divergence, as output pairs that appear to follow a non-stationary trajectory and non-convergence may individually be pulled back towards stationarity and convergence by their dependence on pairs that are stationary and convergent. This observation confirms the usefulness of multivariate long memory methods to address the issue of convergence, as they utilize more information than their univariate

[32] see also Stengos and Yazgan (2014b)

counterparts.

Using the results from the above analysis, we proceed to further examine the evidence of long memory type (absolute) convergence that we discovered. We proceed to investigate the possibility of club formation, a factor that would suggest the presence of conditional convergence. In that case, initial conditions would partly determine at least the long-run outcomes, and if countries with similar starting points exhibit similar long-run economic behavior, one could speak of convergence clubs. Club formation has recently become a very active area of research, as there are many different ways in which one can explore their presence (or absence). We will present a methodology on club formation based on the testing criteria that we have followed in our analysis thus far, and we will employ results from graphing theory to provide evidence for the existence of such clubs in our group of countries.

The remainder of the essay is organized as follows. The next section presents the methodology. We then proceed to present the data and results of the different tests that we apply to the per capita output gaps. We then proceed to a more detailed discussion of the methodology that we employ in the convergence club analysis. The final section concludes.

4.1. Testing framework with long memory.

The simple univariate pair-wise difference between the log of per capita income of country i and j at time t is defined as

$$Z_t = Y_t^i - Y_t^j = \beta(t) + U_t \quad U_t \sim I(d),\ i = 1, ..., N,\ i \neq j,\ t = 1, ..., T \tag{4.1}$$

The process Z_t is described as $(1-L)^d Z_t = \varepsilon_t$, where L is the lag operator and ε_t is the disturbance term. The fractional integration parameter is given by d under the assumption that the process is invertible ($d > -0.5$). The $\beta(t)$ function is a deterministic function of the time trend t and can be linear, as in $\beta(t) = \beta_0 + \beta_1 t$. Alternately, as in

Stengos and Yazgan (2014a), it can be defined in a way that admits structural breaks.

$$\beta(t) = \beta_0 + \beta_1 \sin\left(\frac{2\pi kt}{T}\right) + \beta_2 \cos\left(\frac{2\pi kt}{T}\right) \tag{4.2}$$

This functional form allows for the presence of (smooth) structural breaks. Note here that different values of k will have different implications for the permanent or transitory nature of the breaks. If k is an integer, temporary breaks will result, whereas fractional frequencies would imply permanent breaks because the function would not complete a full oscillation. One advantage of adopting this specification for structural breaks is that it does not require any prior knowledge of the dates on which those breaks occur. On the contrary, it assumes that breaks happen smoothly instead of abruptly, something that would make their detection more difficult.

In a multivariate setting, the long memory process underlying Equation (1) can be expressed as

$$\begin{pmatrix} (1-L)^{d_1} & & 0 \\ & \ddots & \\ 0 & & (1-L)^{d_q} \end{pmatrix} \begin{pmatrix} Z_{1,t} \\ \\ Z_{q,t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1,t} \\ \\ \varepsilon_{q,t} \end{pmatrix}, \qquad -\frac{1}{2} < d_1, ..., d_q < \frac{1}{2}, \tag{4.3}$$

where $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{qt})'$ is a covariance stationary process whose spectral density $f_\varepsilon(\omega_j)$ is bounded and bounded away from zero at zero frequency $\omega_j = 0$ (see (Shimotsu, 2007)). In the multivariate setting, the elements of the $q-$dimensional vector $\mathbf{Z}_t$ are interdependent and correlated with each other, in contrast with the univariate analysis.[33]

Following Stengos and Yazgan (2014a), one can distinguish between different convergence cases that are implied by different values of d. We follow that approach, which allows for a much richer classification of convergence types whereby one can distinguish between rapid convergence, stationary convergence and mean reverting non-

[33] As will be clear below in the present application, q is equal to the number of pair-wise differences between the log of per capita incomes of the countries in our data set.

stationary convergence, where initial differences either decay rapidly and play no role, or linger and have a lasting influence on the present, or fall somewhere in between them.

As in Stengos and Yazgan (2014a), we will concentrate on the estimated values of d and provide tests of convergence based on these estimates. In the next section we will elaborate on the different testing strategies that we will adopt.

4.2. Testing for convergence.

We will consider two types of tests on the estimated ds. The first test is based on the estimation of ds using the multivariate approach illustrated in Equation (3). In this approach, the long memory parameters $\mathbf{d} = (d_1, ..., d_q)^{'}$ are jointly estimated by the semiparametric estimator of Shimotsu (2007), which uses only Fourier frequencies in the neighborhood of the origin. Let $I_Z(\omega_j)$ denote the periodogram of a series Z_t based on a discrete Fourier transform $W_Z(\omega_j)$ at frequency $\omega_j = \frac{2\pi j}{T}$ for $j = 0, ...T-1$, such that $I_Z(\omega_j) = W_Z(\omega_j)W_Z^*(\omega_j)$ with $W_Z^*(\omega_j)$ are the complex conjugate of $W_Z(\omega_j)$,defined as $W_Z(\omega_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} Z_t e^{it\omega_j}$. As shown by Shimotsu (2007), the periodogram $I_Z(\omega_j)$ can be used to define a multivariate estimator of d obtained by minimizing an appropriate likelihood function(Shimotsu, 2007, p. 281). This estimator is asymptotically normally distributed with a variance-covariance matrix that is positively related to covariances among the long memory parameters $\mathbf{d} = (d_1, ..., d_q)^{'}$, i.e., variances are increasing with correlations among $\mathbf{d}$.

The second type of test is based on standard univariate estimation approaches that were used in Stengos and Yazgan (2014a). Based on the estimates of the ds (either multivariate or univariate), we perform the following tests on each of them

Test 1: $H_0^1 : d = 0$ against $H_A^1 : d > 0$ (rapid convergence against long memory convergence)

Test 2: $H_0^2 : d = 0.5$ against $H_A^2 : d < 0.5$ (limit stationary long memory convergence against stationary convergence)

Test 3: $H_0^3 : d = 0.5$ against $H_A^3 : d > 0.5$ (limit stationary long memory convergence against non-stationary mean reverting convergence)

Test 4: $H_0^4 : d = 1$ against $H_A^5 : d < 1$ (non-convergence against non-stationary mean reverting convergence)

Test 5: $H_0^5 : d = 1$ against $H_A^5 : d > 1$ (non-convergence against stochastic divergence)

We calculated the critical values as described below and used them to perform the tests. Then, we compare the test results obtained by the multivariate estimator with those obtained by univariate estimators, as in Stengos and Yazgan (2014a). The univariate estimators covered include the Exact Local Whittle estimator of Shimotsu and Phillips (2005, 2006), Two Stage Feasible Exact Local Whittle of Shimotsu (2010), Fully Extended Local Whittle estimator of Abadir et al. (2007) and prior de-trending versions (see Shimotsu, 2010) of the 2 latter estimators. In all, we make use of four univariate and one multivariate test.

4.2.1. De-trending for structural breaks

To control for structural breaks, we "de-trend" data by estimating $\beta_0, \beta_1, \beta_2$, and k in Equation (2) with the nonlinear least squares. Then, we subtract $\beta(t)$ function, estimated as such, from the data series U_t, before the estimation of $\mathbf{d}$ and conducting the tests.

4.2.2. Monte Carlo based critical values.

We conduct Monte Carlo simulations to compute the critical values of the statistic corresponding to each of the above tests under the null hypothesis under consideration.

The test statistic is computed as

$$\frac{\sqrt{\upsilon}(\widehat{d_q} - d_0)}{\sigma(\widehat{d_q})} \tag{4.4}$$

where υ is the bandwidth parameter, d_0 is the value of d under the null hypothesis, $\widehat{d_q}$ is the estimate of d, and $\sigma(\widehat{d_q})$ is its standard error defined in (Shimotsu, 2007, p. 283). For other estimates of d using for example the univariate Whittle estimators reported in Stengos and Yazgan (2014a) one applies the correspondingly appropriate standard error variance implied by the method used. For the simulations of the critical values, we consider 50,000 iterations. For each iteration, we generate a series from $U_t = Z_t \sim I(d)$ for different values of d corresponding to the different null hypotheses listed above. In the simulations, we assume that the data is already de-trended. De-trending for structural breaks after estimating the $\beta(t)-$function avoids the problem of having to rely on specific values of the $\beta-$parameters to obtain critical values in the simulations. Hence, the test results will avoid possible misspecification due to the reliance on "incorrect" β parameter values [34] In other words, we do not rely on a specific $\beta(t)-$function with particular parametric values of the $\beta-$ parameters to obtain the critical values of the various test statistics. As mentioned above, we de-trend the data by estimating $\beta_0, \beta_1, \beta_2$, and k in Equation (2) using the non-linear least square method.

In Table 1 we provide critical values at the 5 and 10 percent significance levels for $T = 100$, 200, and 500, along with those of the univariate Whittle estimators reported in Stengos and Yazgan (2014a). These critical values are then used in the empirical analysis that follows.

Table 4.1 here.

[34] Ashley and Patterson (2010) suggest isolating and separately examing both a local mean (i.e., a non-linear trend or the realization of a stochastic trend) and its deviations as a modelling strategy that would complement the estimation of a fractionally integrated model.

4.3. Data.

We update our Maddison data set that was used in Stengos and Yazgan (2014a), and also include data from the Penn World Tables (PWT) in our analysis as an additional source. The Maddison data consist of annual GDP per capita data covering the period from 1950 to 2010 for 141 countries[35] and PWT data of annual GDP per capita for the period from 1950 to 2011 for 74 countries. The country coverage of both data sets are illustrated in Table 2.[36] Hence our sample corresponds to $T = 60$ and $N = 141$ for Maddison and to $T = 61$ and $N = 74$ for PWT.

We first investigate the convergence of GDP per capita for all of the 141 and 74 countries taken together as a group and then separately as belonging to different groups from different continents (the Middle East and Central Asia, Europe, AsiaPacific, Sub-Saharan Africa and the Western Hemisphere, and for developing countries taken separately as a single group)[37] . These groups of countries are listed in Table 2 below.

Table 4.2.

In addition to these geographical groups, we will also consider other categories based on levels of economic development, such as emerging markets, the Group of Seven (G7) and the OECD. Emerging markets are grouped according to both FTSE and S& P classifications. Moreover, we also use groupings based on data availability. Countries whose data are available from 1830, 1850, 1860, 1900, and 1930 onwards are

[35]The data come from the Maddison Project (Bolt and van Zanden, 2013). Some countries are missing observations at the end of the period in the final two years. The data are available at http://www.ggdc.net/maddison/maddison-project/home.htm, and they include all possible countries available.

[36]The PWT data come from Feenstra et al. (2013). We use PPP converted GDP per Capita G-K Methods in USD Dollars. Some countries have some missing observations at the beginning of the period, and 53 of these 74 countries are missing no observations. The remaining countries have some missing observations, but no country is missing more than 9. The data are available at http://www.rug.nl/research/ggdc/data/penn-world-table

[37]This classification is based on the usual classification made by the International Monetary Fund's regional economic outlook documents.

taken as a group in the Maddison data set. These country groups are presented in Table 4.3.

Table 4.3.

4.4. Empirical Findings.

Following Pesaran (2007), to analyze output per capita convergence across 141 and 74 countries, we apply the five tests discussed above (each corresponding to a convergence classification) to all possible pairs of $Z_t = Y_t^i - Y_t^j, i = 1, 2, ..., N-1$, and $j = i+1, 2, ..., N$ in a sequential manner. Hence, we examine all $N(N-1)/2 = 9,870$ and $N(N-1)/2 = 2,701$ output gaps for Maddison and PWT deta sets, respectively. For each test, to obtain evidence on the specific type of convergence of which the rejection (non-rejection) of the test indicates, we expect the fraction of output gap pairs for which the null hypothesis is rejected to be greater (smaller) than the size of the test applied to the individual output gap pairs. For example, if one tests for mean reverting convergence by using Test 4 above for a group of N countries, Pesaran's approach requires all $N(N-1)/2$ pairs be subjected to unit root testing and a fraction of rejection among them to exceed the size of tests.

In particular, Pesaran (2007) showed that, if a group of N countries are non-convergent, the rejection rate of the null hypothesis of unit root ($H_0 : Z_t \sim I(d = 1)$) calculated by $N(N-1)/2$ tests is equal to the nominal size of the individual tests, i.e. the probability of Type 1 error. More specifically, it is shown that under the null hypothesis of N countries being non-convergent, the rejection rate of individual tests converges to the nominal size, α as N and $T \to \infty$, even though individual tests are not independent cross- sectionally. Similarly, any weak correlation among the different pairs over time would not matter asymptotically as in the case of approach in (Shimotsu, 2007, p. 283). Thus, in order to reject convergence of N countries, it is enough to show that the proportion of rejections over $N(N-1)/2$ tests is larger than the significance level of individual tests. In that case for example, if the significance level is

5%, the proportion of rejections must exceed 0.05[38] . Even though this approximation may not yield an accurate estimate of the true size of the test, reporting the rejection rates would give us an idea of the evidence in favour or against the particular test in question when comparing the different estimators of d used. To summarize, rejection rates higher than a given significance level in a given application would imply evidence against the convergence hypothesis. On the other hand, rejection rates lower or close to the employed significance level will provide evidence for the non-rejection (validity) of the convergence hypothesis. Hence, in Table 4 below, rejection frequencies that greatly exceed a nominal size of 0.05 would be taken as evidence against the null. Conversely, rejection frequencies that are below the nominal size value will be taken as evidence in favor of the null.

The five tests are applied in sequential order in the sense that we continue to apply them until we find evidence in favour of some type of convergence, if there is any. The column denoted by ALL of Tables 4 and 5 summarizes the results of the four tests applied to all $9,870$ (Maddison) and $2,701$ (PWT) GDP per capita gap pairs at the 5 significance level based on critical values computed for $T = 100$. The table shows the rejection frequencies of the five tests defined above that are obtained from de-trended series using multivariate and univariate estimators, as described above.

Table 4.5 and Table 4.6.

As shown in Table 4, all of the test results belonging to Test 1 report strong rejection of the null hypothesis of rapid convergence against the alternative of long memory. The evidence from Test 2, however, suggests that all of the tests calculated based on multivariate and univariate estimators find evidence in favor of the null hypothesis of a limit stationary long memory process. Test 3 registers very high rejection rates that conclusively indicate evidence in favor of limit stationary convergence and

[38]Clearly in small samples there will be significant size distortions and in applications, the power of the tests used relative to size distortions should be given attention. The above results rest on asymptotic arguments that the rejection rate would converge to α in the limit and that of course would not be the case if N and $T$are relatively small in a given application.

non-convergence. The fact that the rejection rates of Test 4, obtained from univariate estimators, are slightly above the 5 percent significance level, constitutes weak evidence in favor of the alternative hypothesis of mean reverting non-stationary convergence, given possible size distortions due to the sequential nature of our testing procedure (although the different tests are assumed to be independent, there may still be size distortions).[39] However, the results associated with multivariate estimator of Shimotsu (2007) conclusively indicate mean reverting convergence with a much larger rejection rate. These results also hold for all of the grouping varieties considered in Table 4.

The evidence presented in Table 5, obtained from the smaller dataset of PWT, generally confirms the results obtained from the Maddison dataset, although some evidence on stationary convergence is visible when Test 2 and multivariate estimators are considered for the Western hemisphere in particular. The test based on multivariate estimators also shows that the European and Middle-East and Asian countries also display weak evidence of stationary convergence.[40]

We present the group results in Table 6 for the data from the Maddison data set only. Although the country groups whose data are available from 1830 and 1850 onwards provide evidence on stationary convergence, for the remaining groups, the evidence is somewhat weaker.

Table 4.7

The evidence is contrary to the previous findings in Stengos and Yazgan (2014a), where using only univariate statistics provided strong evidence in favor of (absolute) non-convergence, which is also confirmed here but not for the multivariate statistic. The difference in the evidence found using the latter as opposed to the former can be

[39] The results obtained in Stengos and Yazgan (2014a) are similar but weaker with slightly smaller rejection ratios.

[40] Although we stopped the sequence of testing at Test 2, relying on this evidence for the Western hemisphere, we continued for the remaining two groups to guard against possible size distortions in these tests.

explained by the fact that correlations among the estimates of the d's result in the standardized multivariate test statistics to account for the interdependence among the different pairs. In that case, the variances of high d's are mitigated by the presence of negative covariances with other pairs that result in smaller variances overall for the test statistics. Hence, pairs that appear to suggest non-stationary behavior and non-convergence on their own may be pulled back towards stationarity and convergence by their dependence on pairs that are stationary and convergent. In that case, the test statistics of the multivariate test may take "larger" absolute values on average than their univariate counterparts. This evidence in favor of the convergence hypothesis is all the more remarkable in that it is obtained without relying on a benchmark country and allowing for the presence of structural breaks. To summarize, contrary to previous evidence, such as in Stengos and Yazgan (2014a) and Dufrénot et al. (2012), which relied on univariate statistics, using a multivariate approach to estimate the long memory coefficients results in evidence that points towards a mean reverting process for per capita output gaps. These results hold for all different groups of countries, the Middle East and Central Asia, G7, S&P, FTSE, and OECD groups. For Europe and two small country groups whose data dated back to 1830 and 1850, considerable evidence on stationary convergence is present. For Asia and the Pacific, the Western Hemisphere and for three relatively small groups of countries whose data is available from 1860, 1900 and 1930, the evidence on stationary convergence is present but weaker, leaving mean reverting convergence as a second possibility.

It is worth noting that as in Stengos and Yazgan (2014a), the results are based on pair wise comparisons for all possible pairs within a group, as opposed to relying on a benchmark or group leader, as in Dufrénot et al. (2012). Using a benchmark results in differences in output gaps are to be expected, whereas these differences are smoothed out if gaps are only constructed as a difference of individual countries from the leader in the group. This is certainly true for the univariate tests, all of which point towards a long memory non-stationary behavior in the transitional dynamics of the output gaps. The evidence from the multivariate test, however, points towards mean reverting convergence irrespective of the absence of a benchmark leader country due to the greater interdependence between the different pairs captured by this test, which was ignored entirely by its univariate counterparts. We will now proceed to further analyze the evidence found above by exploring the possibility of conditional convergence and club formation.

4.5. Convergence Clubs: A Maximal Clique Method

The above analysis largely implies that the dominant form convergence is of a non-stationary mean reverting nature, which seems to hold unconditionally for all countries. However, the analysis also indicates that stationary convergence, another stronger form of convergence, is also present in smaller group of countries forming convergence clubs. These results indicate the presence of conditional convergence because the differences among some groups of countries show high persistence that can only be corrected in the very long run, indicating the presence of cross-country structural heterogeneity. If initial conditions determine, at least partly, long-run outcomes, and countries with similar initial conditions exhibit similar long-run outcomes, then one can speak of convergence clubs (Durlauf et al., 2005).

The issues of the definition of convergence clubs and their clustering have been widely discussed in the economic growth literature where the evidence on convergence clubs is usually provided on the basis of the convergence of various a-priori defined homogeneous country groups, which were assumed to share the same initial conditions.[41]

In this sense, the above pair-wise method can also be used for testing convergence among a-priorily determined a group of N countries. In fact, Pesaran (2007) also considered different initial set of countries based on geographic characteristics for his pairwise method, but found no evidence on convergence clubs. However, Phillips and Sul (2007) developed an algorithm that classifies groups endogenously rather than

[41]Baumol (1986) for example grouped countries with respect to political regimes (OECD membership, command economies and middle income countries), Chatterji (1992) allowed for clustering that based on initial income per capita levels and tested convergence cross-sectionally, while Durlauf and Johnson (1995) grouped countries using a regression tree method based on different variables such as initial income levels and literacy rates that determined the different "nodes" of the regression tree that defined the country clubs with the common initial conditions and literacy charactheristics. Hausmann et al. (2005), by considering a priori grouping criteria such as initial incomes, found some evidence on convergence clubs by using time series methods.

using a-priori criteria. Similar to Phillips and Sul (2007), in this section, we also attempt to develop an endogenous clustering algorithm to determine the formation of convergence clubs using the pair-wise framework. As stated by Pesaran, "in principle, the convergence results from the analysis of pair-wise output gaps can be used to form "convergence clubs", but special care must be taken in addressing the specification search bias that such a strategy would entail." (Pesaran, 2007, p. 314).

Our approach attempts to pin down the convergence clubs for each type of convergence considered above. Among all pair-wise test results, we search for sets of countries that would yield the desired test result when subjected to the given test. For example, to obtain rapidly converging clubs, we search for sets of countries that would provide a rejection rate below 5 percent for Test 1 considered above. In other words, sets indicating non-rejection of the null of $H_0^1 : d = 0$ form rapidly converging clubs. To begin this search, we use all the pair-wise test results obtained from the all-country analysis. This problem can be solved by using an algorithm designed to find the maximal complete subgraph, or the maximal clique in graph theory terminology.

Consider the above four tests in terms of their implications for convergence types in terms of our sequential procedure. While non-rejection of the null of $H_0^1 : d = 0$ implies unconditionally rapid convergence, the non-rejection of $H_0^3 : d = 0.5$ implies the possibility of mean reverting convergence, provided that the non-rejection of the null of Test 2 has already been obtained in favor of limit stationarity. However, the rejection of the null of Test 2 and 4 provides evidence in favor of the alternatives, $H_A^2 : d < 0.5$ and $H_A^4 : d < 1$ implying stationary and mean reverting convergence. The maximal clique method that we present in this subsection, combines the maximal clique algorithm of graph theory with the previously described pairwise convergence tests of $H_0 : Z_t \sim I(d)$. Rather than testing a priory grouped country clusters, the method explores all convergent groups in a list of N countries that was previously subjected to pairwise convergence tests. In this sense, the method is an endogenous extension of Pesaran (2007) similar to the one by Hobijn and Franses (2000)

The method consists of two steps. First, all possible pairwise differences of N

countries, included in the initial set of of countries, are subjected to one of the tests outlined above. If the rejection rate obtained from $N(N-1)/2$ tests falls below (above) the significance level, that would be evidence in favor of the convergence type implied the alternative (null) hypothesis and the list of N countries will be taken to form a convergence group. If this club involves all examined countries, then all countries are said to be convergent and we do not go any further in seeking out the presence of convergence clubs. However, as shown above, and as in Pesaran (2007), Dufrénot et al. (2012) and Stengos and Yazgan (2014a) it is very unlikely to examine all countries as a group and find evidence for a stronger form of convergence for all. Nevertheless, if a subgroup of countries is found convergent via pairwise method, then it can be said that this subgroup constitutes a convergence club. The main challenge, as indicated above, is to find a method to determine this subgroup rather than relying on a-priori classifications. In the second step we undertake this task.

Suppose $\mathcal{W}$ denotes the set of all countries. Hence by definition, the cardinality of $\mathcal{W}$ is equal to N; (let $\#(\cdot)$ be the cardinality operator, then $\#(\mathcal{W}) = N$). Further, suppose that $\mathcal{E}$ is a subset of $\mathcal{W}$. In this case, in order $\mathcal{E}$ to be a convergence club, all (or a certain percentage determined by the significance level) binary combinations obtained with elements of $\mathcal{E}$ should satisfy the pairwise tests. Hence, since $\#(\mathcal{E}) = M < N$ the rejection rate obtained via all $M(M-1)/2$ pairs should fall below (or above) the significance level. Therefore in the second step, from the $N(N-1)/2$ test results, the objective is to find a class of subsets $\mathcal{G}$ for which all subsets, e.g. $\mathcal{E}$ satisfy pairwise convergence property. Let G denotes the class of all subsets satisfying the desired pairwise property. Then,

$$\mathcal{G} := \{\mathcal{E} : for\ \gamma\ (or\ less)\ percent\ i, j \in \mathcal{E},\ i \neq j,\ t(Z^{ij}) = 1\}$$

where $Z^{ij} = Y_i - Y_j$, $t(\cdot)$ is the test result of the series in the bracelet and takes the value of 1 for convergent pair, i, j and 0 otherwise. γ is the desired rejection rate.[42]

[42]The description above represents the type of convergence by the null hypothesis. In that case, γ denotes the acceptable number of proportions of rejections, under the null. Asymptotically γ should be equal to the significance level applied to all individual tests. Similarly, one could express the type of convergence by the validity of the alternative. In that case, a high proportions of rejections of the null

Hence, the problem can be expressed as

$$\arg\max_{\mathcal{G}} \{\, \#\mathcal{E} : \;\; \mathcal{E} \in \mathcal{G}.\}$$

In graph theory terms, countries become vertices, the test result (rejecting or not rejecting pairs) of a pair become edges, and as such the set of all vertices and edges constitutes an undirected graph. If an undirected graph has edges between all vertices then the graph is said to be complete. If there is a subset of an undirected graph having all properties of a complete graph, the subset is so called a clique. Therefore, in our case, all convergence clubs of a country list can be expressed as cliques. Solving the problem defined above is known as finding maximal cliques.

Pairwise test results form an undirected graph and accordingly, countries and test results determine the vertices and edges respectively. Hence, the problem becomes to find a subgraph that has edges between each vertices, or in other words, a maximal clique. Figures 1 and 2 present a graphical presentation of the concepts introduced above.

Finding a maximal clique can be too hard from a computational point of view. The computational complexity of solution to maximal clique problem is known as NP-Complete whose brute-force solution requires $2^N - \binom{N}{2} - N - 1$ trials. First, Bron and Kerbosch (1973) developed an algorithm to solve the problem in exponential time. In the recent literature, various planar graph algorithms have been developed that enables the problem to be solved in polynomial time. In this study, we will employ the branch and bound algorithm proposed by Konc and Janezic (2007) which is an improved branch-and-bound algorithm that ends in polynomial time.

We should note that, the maximal clique method is not a conclusive technique. In other words, it does not cluster the country list into subgroups, but finds club(s) having

would correspond as evidence against the null in favour of the alternative, depending on the power of the test employed with γ asymptotically approaching unity. However, below we will adopt somewhat different γ levels, taking into considerations all possible size distortions as well as weaknesses in the power of the tests under consideration.

a maximum number of elements. Hence we offer the following clustering algorithm to detect convergence clubs.

(i) Apply the desired test to all Z_{ij} such that $i, j \in \mathcal{W}$ and $i \neq j$

(ii) Test the null hypothesis. If the desired convergence type embedded in the null (alternative) hypothesis, the resulting variable takes the value of 0 (1) or 0 otherwise.

(iii) Construct adjacency matrix from the resulting variable values obtained in (2).

(iv) Find maximal clique(s) from the adjacency matrix via the algorithm proposed by Konc and Janezic (2007). If more than one clique is detected, proceed to the next step. If only one clique is detected jump to (6)

(v) Select one of the detected cliques randomly and proceed to the next step.

(vi) The group of countries in the clique is labeled as a convergence club. Eliminate respective rows and columns of the countries from the adjacency matrix. And step back to (5). Stop if all the rows and columns are eliminated from adjacency matrix.

To illustrate our approach, we applied our procedure to the results we have already obtained. Because the approach requires a considerably large amount of computational time, we restrict our universe $\mathcal{W}$ to a pre-selected group. Also note that as will be clear below, this does not place any restrictions on the main message presented below. The pre-selected groups are illustrated in Table 7, which presents the results of the search for different type of convergences.

Table 7

In Table 7, in the search for convergence clubs, when the type of convergence depends on the non-rejection of the null hypothesis (rapid convergence), we took a 10 percent rejection rate as the benchmark (we hereby attempt to compensate for possible over-rejection displayed by the different tests that we use). In other words, any club producing a 10 percent rejection rate or less is taken as evidence for the validity of the null of rapid convergence. However, we set a 50 percent benchmark rejection rate for the cases where the type of convergence depends on the rejection of the null hypothesis (stationary and mean reverting convergence), accounting for all the possible size distortions due to the sequential nature of our testing procedure.[43]

For the case of rapid convergence, there is no evidence for convergence clubs because the set of 7 and 6 convergent pairs in Europe or in the G7 extended by the group of emerging countries (represented by S&P) are not able to form a convergence club with at least three countries. For the same country groups, 2 convergence pairs are able to form a convergent club of 3 in the case of stationary convergence. However, as shown in Table 7, there are many convergence clubs with different numbers of countries for the case of mean reverting convergence. Note also that these clubs are not required to be disjoint sets. The maximum size of these clubs is 12, and there are 6 different convergent clubs, each containing 12 countries. Figure 3 below illustrates one of these clubs with its associated universe.

Figure 3

We repeat the same analysis using the country groups based data availability as universes in which to search for convergence clubs. The results are illustrated in Table 8. We observe a similar pattern here. While there is little evidence for convergence clubs in rapid convergence, there is ample evidence for convergent clubs in the case of stationary convergence. We also provide a sample from rapidly converging clubs in Figure 4.

Table 8 and Figure 4

[43]Because the significance level of individual tests is adopted as 5 %, in principle, these benchmarks should be equal asymptotically to 5 % and 100 % respectively, if the tests suffer from no size distortions and are consistent. As mentioned above, taking into account the effect of all possible size distortions on the Type I error and allowing for weak test power we made the choices of 10% and 50% respectively. Clearly, these choices are arbitrary and need to be further investigated within the context of a full analysis of the clustering method. We leave this issue for further research.

Table 4.1: Empirical critical values of Test1, 2, 3, and 4 for T = 100, 200, and 500.

	Test 1			Test 2			Test 3			Test 4			Test 5		
CV	95%			05%			95%			05%			95%		
T	100	200	500	100	200	500	100	200	500	100	200	500	100	200	500
FELW	2.207	1.990	1.812	-2.216	-2.076	-1.930	2.254	2.071	2.084	-2.253	-2.153	-2.000	2.180	1.974	1.798
FELWd	1.945	1.765	1.624	-3.079	-2.668	-2.355	2.149	1.939	1.938	-2.647	-2.235	-2.042	2.175	1.975	1.795
2FELW	2.206	1.990	1.812	-2.216	-2.076	-1.930	2.253	2.018	1.795	-2.312	-2.153	-2.000	2.180	1.974	1.798
2FELWd	1.944	1.765	1.624	-3.079	-2.668	-2.354	2.153	1.878	1.670	-2.534	-2.235	-2.042	2.175	1.975	1.795
MLW	2.02	1.985	1.924	-2.879	-2.268	-2.082	2.013	1.912	1.785	-2.267	-2.125	-1.942	2.004	1.854	1.536

Notes: FELW: Fully Extended Local Whittle, 2FELW:2-Stage Feasible Exact Local Whittle estimator, 2FELWd: 2-Stage Feasible Exact Local Whittle estimator with detrending, FELWd: Fully Extended Local Whittle with detrending; MLW: Multivariate Local Whittle Estimator. Simulations are carried out assuming $\upsilon = T^{0.6}$ for all Whittle estimators.

Table 4.2: Countries and group of countries belonging to Maddison and PWT datasets.

	Maddison and PWT	**Only PWT**	**Only Maddison**
Middle-East and Central Asia	Egypt, Iran, Islamic Republic of, Jordan, Morocco, Pakistan		Afghanistan, Bahrain, Iraq, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syrian Arab Republic (Syria), United Arab Emirates, Yemen, Palestinian Territory, Occupied, Algeria, Djibouti, Libya, Mauritania, Somalia, Sudan, Tunisia
Europe	Austria, Belgium, Switzerland, Germany, Denmark, Spain, Finland, France, United Kingdom, Greece, Ireland, Israel, Italy, Netherlands, Norway, Portugal, Sweden, Turkey	Cyprus, Iceland, Luxembourg, Malta	Albania, Bulgaria, Czechoslovakia, Hungary, Poland, Romania, Yugoslavia, Croatia, Macedonia, Republic of, Slovenia
Asia and Pacific	Australia, Bangladesh, China, India, Japan, Korea, Republic of, Sri Lanka, Malaysia, New Zealand, Philippines, Thailand, Taiwan, Republic of China		Indonesia, Myanmar, Hong Kong, Special Administrative Region of China, Nepal, Singapore, Cambodia, Lao PDR, Mongolia, Korea, Democratic People's Republic of, Viet Nam
Sub-Saharan Africa	Benin, Burkina Faso, Congo, Democratic Republic of the, Ethiopia, Ghana, Guinea, Kenya, Mauritius, Malawi, Nigeria, Uganda, South Africa, Zambia	Zimbabwe	Angola, Botswana, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo (Brazzaville), Côte d'Ivoire, Equatorial Guinea, Gabon, Gambia, Guinea-Bissau, Lesotho, Liberia, Madagascar, Mali, Mozambique, Namibia, Niger, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Swaziland, Tanzania, United Republic of, Togo
Western Hemisphere	Argentina, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, Guatemala, Honduras, Jamaica, Mexico, Panama, Peru, Paraguay, El Salvador, Trinidad and Tobago, Uruguay, United States of America, Venezuela (Bolivarian Republic of)		Cuba, Haiti, Nicaragua, Puerto Rico

Table 4.3: Country Groups based on Economic Characteristics and Data Availability

1830	Italy, Sweden, UK, USA, Denmark, France, Netherlands, Norway, Australia
1850	1830 + Belgium, Germany, Greece, Spain
1860	1850 + Finland, Switzerland
1900	1860 + Austria, Portugal, New Zealand, Canada, Brazil, Chile, Colombia, Peru, Uruguay, Venezuela, Japan, Sri Lanka, Argentina, Mexico, Ecuador, India
1930	1900 + Ireland, Turkey, Costa Rica, Guatemala, South Africa
G7	France, Germany, Italy, UK, Canada, USA, Japan
FTSE	Hungary, Poland, Brazil, Mexico, Thailand, Taiwan, Malaysia, Turkey, South Africa
S&P	Brazil, Hungary, Poland, Chile, Colombia, Mexico, Peru, China, India, Philippines, Thailand, Taiwan, Malaysia, Turkey, Egypt, Morocco, South Africa
OECD	Austria, Belgium, Denmark, Finland, France, Germany, Italy, Netherlands, Norway, Sweden, Switzerland, UK, Ireland, Greece, Portugal, Spain, Australia, New Zealand, Canada, USA, Hungary, Poland, Chile, Mexico, Japan, South Korea, Singapore, Israel, Turkey

Table 4.4: Rejection frequencies of Test 1, Test 2, Test 3, and Test 4 for Maddison Data ($\times 10^{-3}$).

		ALL	EUR	WHE	MEA	AAP	SSA	G7	S&P	FTSE	OECD
Test 1	FELW	998	1000	997	996	1000	997	1000	1000	1000	1000
	FELWd	998	1000	997	996	1000	998	1000	1000	1000	1000
	2FELW	998	1000	997	996	1000	997	1000	1000	1000	1000
	2FELWd	998	1000	997	996	1000	998	1000	1000	1000	1000
	MLW	996	1000	1000	1000	978	998	1000	1000	1000	1000
Test 2	FELW	015	011	007	018	039	017	0	0	0	045
	FELWd	014	011	007	014	035	014	0	0	0	045
	2FELW	014	011	013	014	039	016	0	0	0	045
	2FELWd	013	011	013	011	035	013	0	0	0	045
	MLW	034	032	007	014	056	020	0	0	0	0
Test 3	FELW	974	981	993	971	944	973	857	1000	1000	955
	FELWd	974	981	993	967	944	974	857	1000	1000	955
	2FELW	974	979	987	975	944	973	857	1000	1000	955
	2FELWd	975	979	987	975	944	974	857	1000	1000	955
	MLW	903	921	940	899	905	950	1000	938	875	1000
Test 4	FELW	140	119	100	174	121	121	286	188	375	273
	FELWd	132	108	090	174	117	114	286	188	375	227
	2FELW	140	122	097	174	117	121	286	188	375	273
	2FELWd	135	116	093	174	117	115	286	188	375	227
	MLW	473	418	503	576	468	362	286	625	750	409

Notes: The abbreviations used in the table are as follows: ALL (All countries), AAP (Asian and Pacific countries), MEA (Middle-East and Asian countries), EUR (European countries), SSA (Sub-Saharan countries), WHE (Western-hemisphere countries), G7: Group of 7 countries, OECD: OECD countries, FTSE: Financial Times emerging market country group,S&P: Standart and Poors country group, FELW: Fully Extended Local Whittle, 2FELW:2-Stage Feasible Exact Local Whittle estimator, 2FELWd: 2-Stage Feasible Exact Local Whittle estimator with detrending, FELWd: Fully Extended Local Whittle with detrending; MLW: Multivariate Local Whittle Estimator. Simulations are carried out by assuming $\upsilon = T^{6}$ for all Whittle estimators..

Table 4.5: Rejection frequencies of Tests 1, 2, 3, and 4 for PWT data ($\times 10^{-3}$).

		ALL	EUR	WHE	MEA	AAP	SSA	G7	S&P	FTSE	OECD
Test 1	FELW	995	996	1000	1000	1000	995	1000	1000	1000	1000
	FELWd	995	996	1000	1000	1000	995	1000	1000	1000	1000
	2FELW	996	996	1000	1000	1000	995	1000	1000	1000	1000
	2FELWd	996	996	1000	1000	1000	995	1000	1000	1000	1000
	MLW	973	926	1000	970	1000	1000	1000	1000	1000	1000
Test 2	FELW	024	035	0	061	022	019	0	0	0	0
	FELWd	024	035	0	045	022	014	0	0	0	0
	2FELW	023	035	0	061	022	019	0	0	0	0
	2FELWd	023	035	0	045	022	014	0	0	0	0
	MLW	076	108	200	106	011	052	0	0	0	0
Test 3	FELW	937	931	-	864	945	919	1000	1000	1000	1000
	FELWd	941	939	-	864	934	924	1000	1000	1000	1000
	2FELW	940	935	-	864	945	933	1000	1000	1000	1000
	2FELWd	942	939	-	864	934	933	1000	1000	1000	1000
	MLW	807	814	-	758	868	810	800	667	750	810
Test 4	FELW	216	216	-	288	187	195	0	250	250	190
	FELWd	207	212	-	258	176	176	0	250	250	190
	2FELW	212	203	-	273	187	186	0	333	500	238
	2FELWd	209	203	-	258	187	181	0	333	500	238
	MLW	721	714	-	773	868	733	800	750	750	810

Notes: The abbreviations used in the table are as follows: ALL (All countries), AAP (Asian and Pacific countries), MEA (Middle-East and Asian countries), EUR (European countries), SSA (Sub-Saharan countries), WHE (Western-hemisphere countries), G7: Group of 7 countries, OECD: OECD countries, FTSE: Financial Times emerging market country group,S&P: Standart and Poors country group, FELW: Fully Extended Local Whittle, 2FELW:2-Stage Feasible Exact Local Whittle estimator, 2FELWd: 2-Stage Feasible Exact Local Whittle estimator with detrending, FELWd: Fully Extended Local Whittle with detrending; MLW: Multivariate Local Whittle Estimator. Simulations are carried out by assuming $\upsilon = T^{6}$ for all Whittle estimators.

Table 4.6: Rejection frequencies of Tests 1 and 2 for group of countries having available data since some selected years between 1830 and 1930 according to Maddison's data ($\times 10^{-3}$).

		1830	1850	1860	1900	1930
Test 1	FELW	916	846	876	940	946
	FELWd	972	846	876	942	949
	2FELW	1000	846	876	940	946
	2FELWd	1000	846	876	942	950
	MLW	1000	987	952	953	954
Test 2	FELW	222	372	210	120	127
	FELWd	222	372	181	110	121
	2FELW	111	308	210	120	135
	2FELWd	111	308	181	110	124
	MLW	222	294	133	140	141

Notes: FELW: Fully Extended Local Whittle, 2FELW:2-Stage Feasible Exact Local Whittle estimator, 2FELWd: 2-Stage Feasible Exact Local Whittle estimator with detrending, FELWd: Fully Extended Local Whittle with detrending; MLW: Multivariate Local Whittle Estimator. Simulations are carried out by assuming $\upsilon = T^{6}$ for all Whittle estimators.

Table 4.7: Convergence Clubs

Test Type	Country Groups	Club Sizes										
		# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	# 10	# 11	# 12
Rapid Conv. ($H_0^1 : d = 0$)	G7 + Europe	1	-	-	-	-	-	-	-	-	-	-
	Europe + Emerging	7	-	-	-	-	-	-	-	-	-	-
	G7 + Emerging	6	-	-	-	-	-	-	-	-	-	-
Stationary Conv. ($H_A^2 : d < 0.5$)	G7 + Europe	-	-	-	-	-	-	-	-	-	-	-
	Europe + Emerging	2	1	-	-	-	-	-	-	-	-	-
	G7 + Emerging	2	1	-	-	-	-	-	-	-	-	-
Mean Reverting Conv. ($H_A^4 : d < 1$)	G7 + Europe	38	726	3246	3737	4577	6185	3639	731	67	-	-
	Europe + Emerging	96	2971	18474	25220	41378	84168	72565	24080	6925	1397	6
	G7 + Emerging	66	310	1739	3353	2533	2586	3265	1605	243	23	-

Table 4.8: Convergence Clubs

Club Type	Years	Club Sizes			
		# 2	# 3	# 4	# 5
Rapid Conv.	1830	-	-	-	-
	1850	12	2	-	-
	1860	14	2	-	-
	1900	29	3	-	-
	1930	37	3	-	-
Stationary Conv.	1830	-	-	-	-
	1850	12	23	49	25
	1860	14	30	72	28
	1900	29	90	381	208
	1930	38	139	649	299

Figure 4.1: A sample undirected graph

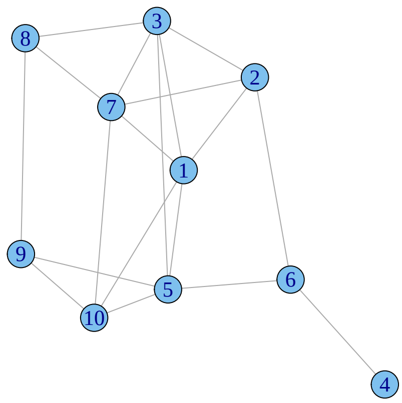


Figure 4.2: A sample maximum clique

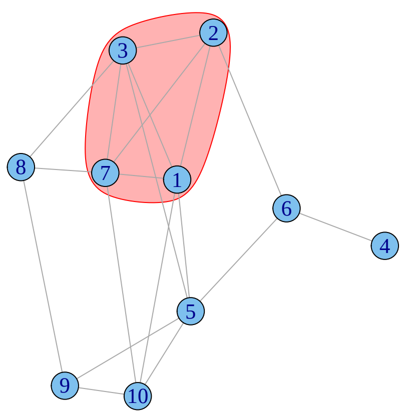


Figure 4.3: A Club of Mean Reverting Convergence (Europe + Emerging Markets)

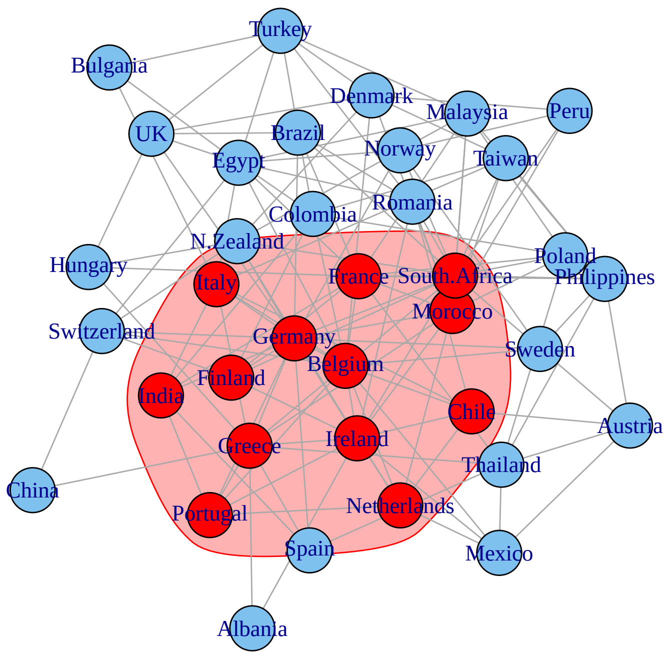
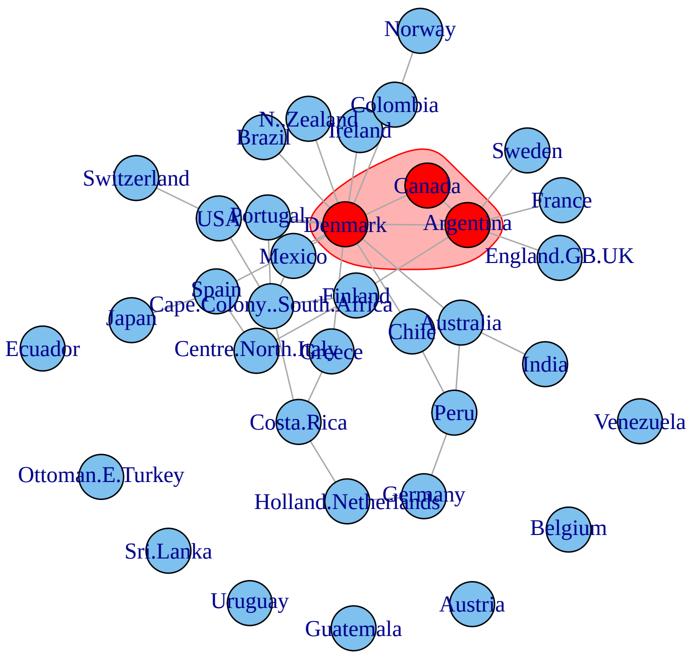


Figure 4.4: A Rapidly Converging Club (1930)



5. CONCLUSION

Several different conclusions can be formulated based on the findings of the essays in this study.

Essay 1 is tried to investigate whether (IT) policies made it easy to predict inflation rate. We have found clear evidence that forecasts are superior in IT periods compared to NIT periods, which was the main goal of this essay.In more than half of the countries covered in this study, an IT period provides better forecasting accuracy regardless of the model used. In contrast, among the remaining countries, whereas the results remain ambiguous for most cases , the evidence on the superiority of NIT is limited to very few countries.

Additionally, we test the relative forecast accuracy of the utilized models in each period against the benchmark of the random walk model. The null hypothesis is that the forecasts of the model under consideration are no better than those of the random walk model; the alternative is that random walk forecasts more accurately than the model under consideration. Therefore, we cannot assert that forecasts from a particular model are superior to random-walk forecasts in most countries and horizons. Furthermore, comparing the results in IT and NIT periods does not lead us to conclude that one or more models overwhelmingly provide better forecasts in one period in comparison to other periods.

In Essay 2, we have developed a structural change test based on maximum overlap wavelet transformation, derived its limiting null distribution, and demonstrated its empirical size and power properties against a number of alternative tests. Our results indicate that the wavelet based test($\widetilde{GYO_W}$) has higher power when there are multiple structural breaks at unknown locations. In general, the power of the competing tests start to deteriorate after single break irrespective of whether the break is abrupt or smooth. On the contrary, the wavelet tests is uniformly keeps a high level of power at all multiple break levels. On the other hand, when we deal with a larger data set, MOSUM test attains the same power as the wavelet test. We attribute this to the fact that both, MOSUM and wavelet tests, use local information in a given window.

Essay 3 conducts a general Markov-regime switching estimation both in the long memory parameter d and the mean of a time series. It is suggested in the literature that mean and long memory parameter breaks in time series can be easily confused with each other and our results generally confirm the above assertion, since we observe that structural breaks in the long memory parameter produce similar in sample estimation patterns irrespective of the presence of breaks in the mean parameter. Our results are complementary to those of Tsay and Härdle (2009) who only considers breaks in the mean parameter in his simulation study. Our results call for the development for joint tests that would be able to discern the nature of possible such breaks.

However, given the results of our analysis there are many outstanding issues that we leave for future research. First, a more comprehensive Monte Carlo simulation study would need to allow for a more general MS-ARFIMA (p, d, q)structure. Furthermore, one would need to investigate the asymptotic properties of the estimators of μ and $d$we obtain under regime switching. Even though the likelihood function is well defined, in the context of unknown breaks, we conjecture that one would need to use a framework that is used in threshold models with unknown thresholds, see Hansen (2000). An investigation of such a problem could be a promising item in a future research agenda.

In, we use a long memory framework of analysis that does not rely on a benchmark country but allows for the presence of structural breaks to estimate the time series properties of output gaps for counties in the post-World War II period and, as such, provide evidence on the convergence hypothesis. The focus of the essay is first the estimation of d, the parameter that determines the speed of convergence between different economies. We estimate d using a multivariate estimator from Shimotsu (2007) and a number of other univariate methods found in recent studies.

The main finding of Essay 4 is that, for per capita GDP gaps, the long memory parameter d takes values in the range from $0.5 < d < 1$ (mean reverting convergence), and the range from $0 < d < 0.5$ (stationary convergence) seems be a possibility, especially for some country groups when using the multivariate approach. The difference between univariate and multivariate procedures in these findings can be explained by the fact that correlations among the estimates of the $d's$ in the multivariate test statistics account for the interdependence among the different pairs. In that case, any divergent behavior is mitigated by the presence of convergent pairs that act as stabilizing factors for the group as a whole.

Using the results from the above analysis, we proceed to further examine the evidence we found for long memory type (absolute) convergence. We then investigated the possibility of club formation, something that would suggest the presence of conditional convergence and offer a club formation methodology based on maximal cliques in conjunction with sequential testing criteria. This would make a very fruitful area of new research since –to the best of our knowledge– the properties of club formation mechanisms have not been explored yet in the literature. This research will shed light on issues regarding possible mis-classifications and incorrect club membership which need to be addressed in order to guide practitioners who will be applying these methods to explore convergence and club formation.

Bibliography

Abadir, K., W. Distaso, and L. Giraitis (2007). Nonstationarity-extended local whittle estimation. *Journal of Econometrics 141*, 1353–1384.

Al-Qassam, M. S. and J. A. Lane (1989). Forecasting exponential autoregressive models of order 1. *Journal of Time Series Analysis 91*(2), 95–113.

Altissimo, F. and V. Corradi (2003). Strong rules for detecting the number of breaks in a time series. *Journal of Econometrics 117*(2), 207–244.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), pp. 821–856.

Andrews, D. W. K., I. Lee, and W. Ploberger (1996). Optimal changepoint tests for normal linear regression. *Journal of Econometrics 70*(1), 9–38.

Ashley, R. A. and D. M. Patterson (2010). Apparent long memory in time series as an artifact of a time-varying mean: considering alternatives to the fractionally integrated model. *Macroeconomic Dynamics 14*(S1), 59–87.

Bahl, L., J. Cocke, F. Jelinek, and J. Raviv (2006). Optimal decoding of linear codes for minimizing symbol error rate (corresp.). *IEEE Trans. Inf. Theor. 20*(2), 284–287.

Bai, J. and P. Perron (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica 66*(1), pp. 47–78.

Bai, J. and P. Perron (2003a). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*(2), 1–22.

Bai, J. and P. Perron (2003b). Critical values for multiple structural change tests. *The Econometrics Journal 6*(3), 72–78.

Ball, L. M. and N. Sheridan (2004). Does inflation targeting matter? In *The Inflation-Targeting Debate*, NBER Chapters, pp. 249–282. National Bureau of Economic Research Inc.

Baumol, W. J. (1986). Productivity growth, convergence, and welfare: what the long-run data show. *The American Economic Review 76*, 1072–1085.

Becker, R., W. Enders, and S. Hurn (2004). A general test for time dependence in parameters. *Journal of Applied Econometrics 19*(7), 899–906.

Becker, R., W. Enders, and J. Lee (2006). A Stationarity Test in the Presence of an Unknown Number of Smooth Breaks. *Journal of Time Series Analysis 27*(3), 381–409.

Bolt, J. and J. L. van Zanden (2013). The first update of the maddison project; re-estimating growth before 1820. Technical report, Maddison-Project Working Paper WP-4, University of Groningen.

Brito, R. D. and B. Bystedt (2010). Inflation targeting in emerging economies: Panel evidence. *Journal of Development Economics 91*(2), 198–210.

Brown, R. L., J. Durbin, and J. M. Evans (1975). Techniques for Testing the Constancy of Regression Relationships over Time. *Journal of the Royal Statistical Society. Series B (Methodological) 37*(2), pp. 149–192.

Cajueiro, D. O. and B. M. Tabak (2005a). Possible causes of long-range dependence in the brazilian stock market. *Physica A: Statistical Mechanics and its Applications 345*(3), 635–645.

Cajueiro, D. O. and B. M. Tabak (2005b). Ranking efficiency for emerging equity markets ii. *Chaos, Solitons & Fractals 23*(2), 671–675.

Cajueiro, D. O. and B. M. Tabak (2005c). Testing for time-varying long-range dependence in volatility for emerging markets. *Physica A: Statistical Mechanics and its Applications 346*(3), 577–588.

Cajueiro, D. O. and B. M. Tabak (2007a). Long-range dependence and multifractality in the term structure of libor interest rates. *Physica A: Statistical Mechanics and its Applications 373*, 603–614.

Cajueiro, D. O. and B. M. Tabak (2007b). Time-varying long-range dependence in us interest rates. *Chaos, Solitons & Fractals 34*(2), 360–367.

Chatterji, M. (1992). Convergence clubs and endogenous growth. *Oxford Review of economic policy 8*(4), 57–69.

Chong, T. T.-L. (2001). Structural change in ar (1) models. *Econometric Theory 17*(01), 87–155.

Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica 28*(3), pp. 591–605.

Chu, C.-S. J., K. Hornik, and C.-M. Kuan (1995). MOSUM Tests for Parameter Constancy. *Biometrika 82*(3), pp. 603–617.

Chu, C.-S. J., M. Stinchcombe, and H. White (1996). Monitoring Structural Change. *Econometrica 64*(5), pp. 1045–1065.

Clark, T. E. and K. D. West (2006). Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics 135*(1-2), 155 – 186.

Conniffe, D. and J. E. Spencer (2001). When Moments of Ratios Are Ratios of Moments. *Journal of the Royal Statistical Society. Series D (The Statistician) 50*(2), pp. 161–168.

D'Agostino, A., L. Gambetti, and D. Giannone (2011). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics 26*(7), 82–101.

D'Agostino, A., D. Giannone, and P. Surico (2006). (Un)Predictability and macroeconomic stability. Working Paper Series 605, European Central Bank.

D'Agostino, A. and P. Surico (2011). A century of inflation forecasts. CEPR Discussion Papers 8292, C.E.P.R. Discussion Papers.

De Gooijer, J. G. and P. T. De Bruin (1998). On forecasting SETAR processes. *Statistics & Probability Letters 37*(1), 7–14.

de Mendonça, H. F. and G. J. de Guimarães e Souza (2012). Is inflation targeting a good remedy to control inflation? *Journal of Development Economics 98*(2), 178 – 191.

Diebold, F. X. and A. Inoue (2001). Long memory and regime switching. *Journal of econometrics 105*(1), 131–159.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–63.

Ding, Z., C. W. Granger, and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance 1*(1), 83–106.

Dufrénot, G., V. Mignon, and T. Naccache (2012). Testing catching-up between the developing countries:"growth resistance" and sometimes "growth tragedy". *Bulletin of Economic Research 64*(4), 470–508.

Durlauf, S. N. and P. A. Johnson (1995). Multiple regimes and cross-country growth behaviour. *Journal of Applied Econometrics 10*(4), 365–384.

Durlauf, S. N., P. A. Johnson, and J. R. Temple (2005). Growth econometrics. *Handbook of economic growth 1*, 555–677.

Elliott, G. and A. Timmermann (2008). Economic forecasting. *Journal of Economic Literature 46*(1), 3–56.

Fan, Y. and R. Gençay (2010). Unit root tests with wavelets. *Econometric Theory 26*, 1305–1331.

Faust, J. and J. H. Wright (2012). Forecasting inflation. Unpublished manuscript.

Feenstra, R. C., R. Inklaar, and M. Timmer (2013). The next generation of the penn world table. Technical report, National Bureau of Economic Research.

Franses, P. and D. Dijk (2000). *Nonlinear time series models in empirical finance.* Cambridge University Press.

Gençay, R., F. Selçuk, and B. Whitcher (2002). An introduction to wavelets and other filtering methods in finance and economics. *Waves in Random Media 12*(3), 399–399.

Gençay, R. and N. Gradojevic (2011). Errors-in-variables estimation with wavelets. *Journal of Statistical Computation and Simulation 11*(8), 1545–1564.

Gonçalves, C. E. S. and J. a. M. Salles (2008). Inflation targeting in emerging economies: What do the data say? *Journal of Development Economics 85*(1-2), 312–318.

Granger, C. W. and Z. Ding (1994). Stylized facts on the temporal and distributional properties of daily data from speculative markets. *UCSD Department of Economics Discussion Paper*, 94–19.

Haldrup, N. and M. O. Nielsen (2006). A regime switching long memory model for electricity prices. *Journal of econometrics 135*(1), 349–376.

Hamilton, J. (1994). *Time series analysis*. Princeton University Press.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica 57*(2), 357–84.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of econometrics 45*(1), 39–70.

Hansen, B. E. (1997). Approximate Asymptotic p Values for Structural-Change Tests. *Journal of Business & Economic Statistics 15*(1), 60–67.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica 68*(3), 575–603.

Hansen, B. E. (2002). Tests for Parameter Instability in Regressions with I(1) Processes. *Journal of Business & Economic Statistics 20*(1), 45–59.

Hausmann, R., L. Pritchett, and D. Rodrik (2005). Growth accelerations. *Journal of Economic Growth 10*(4), 303–329.

Hobijn, B. and P. H. Franses (2000). Asymptotically perfect and relative convergence of productivity. *Journal of Applied Econometrics 15*(1), 59–81.

Inclan, C. and G. C. Tiao (1994). Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance. *Journal of the American Statistical Association 89*(427), pp. 913–923.

Kejriwal, M. and P. Perron (2010). Testing for multiple structural changes in cointegrated regression models. *Journal of Business & Economic Statistics 28*(4), 503–522.

Kock, A. B. and T. Teräsvirta (2011). Forecasting with nonlinear time series models. In M. P. Clements and D. F. Hendry (Eds.), *Oxford Handbook on Economic Forecasting*, pp. 61–87. Oxford University Press.

Krämer, W., W. Ploberger, and R. Alt (1988). Testing for structural change in dynamic models. *Econometrica 56*(6), 1355–69.

Limam, I. (2003). Is long memory a property of thin stock markets? international evidence using arab countries. *Review of Middle East Economics and Finance 1*(3), 251–266.

Lin, J. L. and C. W. J. Granger (1994). Forecasting from non-linear models in practice. *Journal of Forecasting 13*, 1–9.

Lin, S. and H. Ye (2007). Does inflation targeting really make a difference? Evaluating the treatment effect of inflation targeting in seven industrial countries. *Journal of Monetary Economics 54*(8), 2521–2533.

Lobato, I. N. and C. Velasco (2000). Long memory in stock-market trading volume. *Journal of Business & Economic Statistics 18*(4), 410–427.

Ludlow, J. and W. Enders (2000). Estimating non-linear ARMA models using Fourier coefficients. *International Journal of Forecasting 16*(3), 333–347.

Lux, T. and T. Kaizoji (2007). Forecasting volatility and volume in the tokyo stock market: Long memory, fractality and regime switching. *Journal of Economic Dynamics and Control 31*(6), 1808–1843.

Marcellino, M., J. H. Stock, and M. W. Watson (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics 135*(1-2), 499–526.

Percival, D. B. and A. T. Walden (2006). *Wavelet methods for time series analysis*, Volume 4. Cambridge university press.

Perron, P. (2006). Dealing with structural breaks. *Palgrave handbook of econometrics 1*, 278–352.

Pesaran, H. M. (2007). A pair-wise approach to testing for output and growth convergence. *Journal of Econometrics 138*(1), 312–355.

Phillips, P. C. and D. Sul (2007). Transition modeling and econometric convergence tests. *Econometrica 75*(6), 1771–1855.

Ploberger, W. and W. Kramer (1992). The CUSUM Test with OLS Residuals. *Econometrica 60*(2), 271–85.

Ploberger, W., W. Krämer, and K. Kontrus (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics 40*(2), 307–318.

Rossi, B. and T. Sekhposyan (2010). Have economic models' forecasting performance for US output growth and inflation changed over time, and when? *International Journal of Forecasting 26*(4), 808–835.

Sen, P. K. (1982). Invariance principles for recursive residuals. *The Annals of Statistics 10*, 307–312.

Shimotsu, K. (2007). Gaussian semiparametric estimation of multivariate fractionally integrated processes. *Journal of Econometrics 137*(2), 277–310.

Shimotsu, K. (2010). Exact local whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory 26*, 501–540.

Shimotsu, K. and P. Phillips (2005). Exact local whittle estimation of fractional integration. *Annals of Statistics 33*, 1890–1933.

Shimotsu, K. and P. Phillips (2006). Local whittle estimation of fractional integration and some of its variants. *Journal of Econometrics 130*, 209–233.

Stengos, T. and M. E. Yazgan (2014a). Persistence in convergence. *Macroeconomic Dynamics 18*(04), 753–782.

Stengos, T. and M. E. Yazgan (2014b). Persistence in real exchange rate convergence. *Studies in Nonlinear Dynamics & Econometrics 18*(1), 73–88.

Stock, J. H. and M. W. Watson (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking 39*(1), 3–33.

Stock, J. H. and M. W. Watson (2009). Phillips curve inflation forecasts. In J. Fuhrer, Y. K. Kodrzycki, J. Little, and G. P. Olivei (Eds.), *Understanding inflation and the implications for monetary policy: a Phillips curve retrospective*. The MIT Press.

Svensson, L. E. (2010). Inflation targeting. In B. M. Friedman and M. Woodford (Eds.), *Handbook of Monetary Economics*, Volume 3 of *Handbook of Monetary Economics*, Chapter 22, pp. 1237–1302. Elsevier.

Svensson, L. E. O. (2005). Monetary Policy with Judgment: Forecast Targeting. *International Journal of Central Banking 1*(1).

Teräsvirta, T. (2006). Forecasting economic variables with nonlinear models. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1 of *Handbook of Economic Forecasting*, Chapter 8, pp. 413–457. Elsevier.

Tsay, W. (2009). Analysing inflation by the arfima model with markov-switching fractional differencing parameter. Technical report, Working Paper, The Institute of Economics, Academia Sinica.

Tsay, W.-J. and W. K. Härdle (2009). A generalized arfima process with markov-switching fractional differencing parameter. *Journal of Statistical Computation and Simulation 79*(5), 731–745.

van Dijk, D., P. H. Franses, M. P. Clements, and J. Smith (2003). On SETAR nonlinearity and forecasting. *Journal of Forecasting 22*(5), 359–375.

White, H. (2001). *Asymptotic theory for econometricians*. Academic Press.

White, H. (2006). Approximate nonlinear forecasting methods. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, Chapter 9, pp. 459–512. Elsevier.

Williams, J. D. (1941). Moments of the Ratio of the Mean Square Successive Difference to the Mean Square Difference in Samples From a Normal Universe. *The Annals of Mathematical Statistics 12*(2), pp. 239–241.

Willinger, W., M. S. Taqqu, and V. Teverovsky (1999). Stock market prices and long-range dependence. *Finance and stochastics 3*(1), 1–13.

Xue, Y., S. Fagan, and R. Gençay (2010). Testing for Jump Arrivals in Financial Time Series. Technical report, Department of Economics, Simon Fraser University.

Zeileis, A. (2005). A unified approach to structural change tests based on ML scores, F statistics, and OLS residuals. *Econometric Reviews 24*(4), 445–466.

Appendix A: Proofs for essay 2

Proposition 2.1. *Under* H_0 , $\mathbb{E}\{\delta_T^2\} = \frac{\sigma^2}{2}$. $Var(\delta_T^2) \cong \frac{3\sigma^4}{T}$ *for large* T.

Proof. Without any loss of generality, we can set $\mu = 0$ in the H_0 since both σ^2 and s^2 are both invariant under mean change. For the remaining part of the proof we assume that $t \in \{1, 2, \ldots, T, \mod T\}$. By expanding δ^2 as the sum of W_i^2, we have

$$\begin{aligned}
\mathbb{E}\delta_T^2 &= \mathbb{E}\left\{\frac{1}{T}\sum_{t=1}^{T} W_t^2\right\} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}W_t^2 = \frac{1}{T}T\mathbb{E}(W_1^2) = \mathbb{E}(W_1^2) \\
&= \mathbb{E}\left\{\left(\frac{y_1 - y_0}{2}\right)^2\right\} = \frac{1}{4}\mathbb{E}\{(y_1^2 + y_0^2 - 2y_1y_0)\} \\
&= \frac{1}{4}\left\{\mathbb{E}(y_1^2) + \mathbb{E}(y_0^2) - 2\mathbb{E}(y_1y_0)\right\} = \frac{1}{4}\left\{2\sigma^2\right\} = \frac{1}{2}\sigma^2
\end{aligned}$$

As for the variance,

$$\begin{aligned}
\mathbb{E}\delta_T^2 = Var(\delta_T^2) &= Var\left(T^{-1}\sum_{t=j}^{T+j-1} W_t^2\right) \\
&= T^{-2}\sum_{t=j}^{T+j-1} Var(W_t^2) + 2T^{-2}\sum_{t=j}^{T+j-1}\sum_{s=t+1}^{T+j-1} Cov(W_t^2, W_s^2).
\end{aligned}$$

We also have that $Cov(W_t, W_{t+k}) = 0$ for $k \geq 2$ (see Proposition A.1). Then,

$$Var(\delta_T^2) = T^{-2}\sum_{t=j}^{T+j-1} Var(W_t^2) + 2T^{-2}\sum_{t=j+1}^{T+j-1} Cov(W_{t-1}^2, W_t^2).$$

Writing the probabilistic expression for $Cov(W_{t-1}^2, W_t^2)$:

$$\begin{aligned}
Cov(W_{t-1}^2, W_t^2) &= \mathbb{E}\{(W_{t-1}^2 - \mathbb{E}W_{t-1}^2)(W_t^2 - \mathbb{E}W_t^2)\} \\
&= \mathbb{E}\{W_{t-1}^2 W_t^2\} - \mathbb{E}W_{t-1}^2\mathbb{E}W_t^2.
\end{aligned}$$

where $\mathbb{E}W_t^2$, the unconditional mean of W_t^2, is $\sigma^2/2$, $\forall t$. Then, expanding $\mathbb{E}\{W_{t-1}^2 W_t^2\}$:

$$\begin{aligned}
\mathbb{E}\{W_{t-1}^2 W_t^2\} &= \mathbb{E}\{(\frac{y_{t-1}-y_{t-2}}{2})^2(\frac{y_t-y_{t-1}}{2})^2\} \\
&= 2^{-4}\mathbb{E}\{(y_{t-1}^2 - 2y_{t-1}y_{t-2} + y_{t-2}^2)(y_t^2 - 2y_t y_{t-1} + y_{t-1}^2)\} \\
&= 2^{-4}\mathbb{E}\{y_{t-1}^2 y_t^2 - 2y_{t-1}^3 y_t + y_{t-1}^4 - 2y_{t-1}y_{t-2}y_t^2 + 4y_{t-1}^2 y_t y_{t-2} - 2y_{t-1}^3 y_{t-2} \\
&\quad + y_{t-2}^2 y_t^2 - 2y_{t-1}y_{t-2}^2 y_t + y_{t-2}^2 y_{t-1}^2\}
\end{aligned}$$

since y_t and y_{t-1} are independent $\forall t$. Also we know that under H_0, first four raw moments of y_t are

$$\mathbb{E}\{y_t\} = 0, \quad \mathbb{E}\{y_t^2\} = \sigma^2, \quad \mathbb{E}\{y_t^3\} = 0, \quad \mathbb{E}\{y_t^4\} = 3\sigma^4, \quad \forall t.$$

Then,

$$\begin{aligned}
2^4\mathbb{E}\{W_{t-1}^2 W_t^2\} &= 3(\sigma^2)^2 - 3\sigma^4. \\
&= 6\sigma^4 \Rightarrow \\
\mathbb{E}\{W_{t-1}^2 W_t^2\} &= \frac{3}{8}\sigma^4.
\end{aligned}$$

Henceforth,

$$Cov(W_{t-1}^2, W_t^2) = \frac{3}{8}\sigma^4 - \frac{1}{4}\sigma^4 = \frac{1}{8}\sigma^4.$$

and

$$\begin{aligned}
Var(\delta_T^2) &= T^{-2}T\,Var(W_t^2) + 2T^{-2}(T-1)\,Cov(W_{t-1}^2, W_t^2) \\
&= \frac{\sigma^4}{T}(\frac{1}{2} + \frac{T-1}{4T}) \quad \longrightarrow \quad \frac{3\sigma^4}{4T} \text{ as } T \text{ gets larger}
\end{aligned}$$

since $Var(W_t^2) = \dfrac{\sigma^4}{2}$ (see Proposition A.2). □

Lemma A.1. *Under H_0, $Cov(W_t^2, W_{t+k}^2) = 0$ for $k \geq 2$.*

Proof.

$$\begin{aligned}
Cov(W_t, W_{t+k}) &= \mathbb{E}\{(W_t^2 - \mathbb{E}W_t^2)(W_{t+k}^2 - \mathbb{E}W_{t+k}^2)\} \\
&= \mathbb{E}\{W_t^2 W_{t+k}^2\} - \mathbb{E}W_t^2 \mathbb{E}W_{t+k}^2 \\
&= \mathbb{E}\{(\frac{y_t - y_{t-1}}{2})^2 (\frac{y_{t+k} - y_{t+k-1}}{2})^2\} - (\frac{\sigma^2}{2}\frac{\sigma^2}{2}) \\
&= 2^{-4}\mathbb{E}\{(y_t^2 - 2y_t y_{t-1} + y_{t-1}^2)(y_{t+k}^2 - 2y_{t+k}y_{t+k-1} + y_{t+k-1}^2)\} - \frac{\sigma^4}{4} \\
&= 2^{-4}\mathbb{E}\{y_t^2 y_{t+k}^2 - 2y_t^2 y_{t+k} y_{t+k-1} + y_t^2 y_{t+k-1}^2 \\
&\quad -2y_t y_{t-1} y_{t+k}^2 + 4y_t y_{t-1} y_{t+k} y_{t+k-1} - 2y_t y_{t-1} y_{t+k-1}^2 \\
&\quad +y_{t-1}^2 y_{t+k}^2 - 2y_{t-1}^2 y_{t+k} y_{t+k-1} + y_{t-1}^2 y_{t+k-1}^2\} - \frac{\sigma^4}{4}\,.
\end{aligned}$$

But, since y_s and y_r are independent whenever $s \neq r$, we have $\mathbb{E}\{y_s^m y_r^n\} = \mathbb{E}\{y_s^m\}\mathbb{E}\{y_r^n\}$ $\forall m \in \mathbb{R}, \forall n \in \mathbb{R}$. Moreover, we know that $\mathbb{E}\{y_s\} = 0$ and $\mathbb{E}\{y_s^2\} = \sigma^2,\ \forall s \in \{1, \ldots, T\}$. Then, it follows that

$$2^4\, Cov(W_t, W_{t+k}) = \sigma^2\sigma^2 + \sigma^2\sigma^2 + \sigma^2\sigma^2 + \sigma^2\sigma^2 - 4\sigma^4 = 0$$

□

Lemma A.2. *Under H_0, $Var(W_t^2) = \dfrac{\sigma^4}{2}$.*

Proof. Since $y_t - y_{t-1} = u_t - u_{t-1}$, μ in $W_t = \dfrac{u_t - u_{t-1}}{2}$ is eliminated. So,

$$\begin{aligned}
Var(W_t^2) &= Var((\frac{u_t - u_{t-1}}{2})^2) \qquad \Rightarrow \\
2^4 Var(W_t^2) &= Var(u_t^2 - 2u_t u_{t-1} + u_{t-1}^2) \\
&= Var(u_t^2) + Var(u_{t-1}^2) + Var(-2u_t u_{t-1}) \qquad \text{(A.1)} \\
&\quad +2Cov(u_t^2, -2u_t u_{t-1}) + 2Cov(u_{t-1}^2, -2u_t u_{t-1}) + 2Cov(u_t^2, u_{t-1}^2) \\
&= 2Var(u_t^2) + 4Var(u_t u_{t-1}) - 8Cov(u_t^2, u_t u_{t-1}) + 2Cov(u_t^2, u_{t-1}^2).
\end{aligned}$$

since $Var(u_t^2) = Var(u_{t-1}^2)$ and $Cov(u_t^2, u_t u_{t-1}) = Cov(u_{t-1}^2, u_t u_{t-1})$. Also, we have

$$\begin{aligned}
Var(u_t^2) = Var(u_{t-1}^2) &= \mathbb{E}\{(u_t^2)^2\} - \mathbb{E}\{u_t^2\}\mathbb{E}\{u_t^2\} \\
&= 3\sigma^4 - (\sigma^2)^2 = 2\sigma^4, \\
Var(u_t u_{t-1}) &= \mathbb{E}\{(u_t^2 u_{t-1}^2)\} - \mathbb{E}\{u_t u_{t-1}\}\mathbb{E}\{u_t u_{t-1}\} \\
&= \mathbb{E}\{u_t^2\}\mathbb{E}\{u_{t-1}^2\} - 0 = \sigma^4, \\
Cov(u_t^2, u_t u_{t-1}) &= \mathbb{E}\{(u_t^2 u_t u_{t-1})\} - \mathbb{E}\{u_t^2\}\mathbb{E}\{u_t u_{t-1}\} \\
&= \mathbb{E}\{u_{t-1}\}(\mathbb{E}\{u_t^3\} - \mathbb{E}\{u_t^2\}\mathbb{E}\{u_t\}) \\
&= 0 \times (0 - \sigma^2 \times 0) = 0, \\
Cov(u_t^2, u_{t-1}^2) &= \mathbb{E}\{(u_t^2 u_{t-1}^2)\} - \mathbb{E}\{u_t^2\}\mathbb{E}\{u_{t-1}^2\} \\
&= \mathbb{E}\{u_t^2\}\mathbb{E}\{u_{t-1}^2\} - \mathbb{E}\{u_t^2\}\mathbb{E}\{u_{t-1}^2\} = 0.
\end{aligned}$$

since $\mathbb{E}\{u_t\} = \mathbb{E}\{u_{t-1}\} = 0$ and $\mathbb{E}\{u_t^3\} = 0$.

Then, by making necessary substitutions in Equation (A.1),

$$\begin{aligned}
2^4 Var(W_t^2) &= 2 \times 2\sigma^4 + 4 \times \sigma^4 - 8 \times 0 + 2 \times 0 \Rightarrow \\
Var(W_t^2) &= \frac{8\sigma^4}{16} = \frac{\sigma^4}{2}.
\end{aligned}$$

□

Proposition 2.2. *Under H_0, $Var\left(\delta_T^2 - \frac{s^2}{2}\right) \cong \frac{\sigma^4}{4T}$ for large T.*

Proof. By rewriting $\delta_T^2 - \frac{s^2}{2}$ we have sum of first-order serially correlated variables:

$$\begin{aligned}
\delta_T^2 - \frac{s^2}{2} &= \frac{1}{T}\sum_{t=1}^{T} W_t^2 - \frac{1}{2T}\sum_{t=1}^{T}(y_t-\bar{y})^2 = \frac{1}{4T}\left\{\sum_{t=1}^{T}(y_t-y_{t-1})^2 - 2\sum_{t=1}^{T}(y_t-\bar{y})^2\right\} \\
&= \frac{1}{4T}\left\{\sum_{t=1}^{T}\left(y_t^2+y_{t-1}^2-2y_ty_{t-1}-2y_t^2+4\bar{y}y_t-2\bar{y}^2\right)\right\} \\
&= \frac{1}{4T}\left(\sum_{t=1}^{T}y_t^2+\sum_{t=1}^{T}y_{t-1}^2-2\sum_{t=1}^{T}y_ty_{t-1}-2\sum_{t=1}^{T}y_t^2+4\sum_{t=1}^{T}\bar{y}y_t-2\sum_{t=1}^{T}\bar{y}^2\right) \\
&= \frac{1}{4T}\left(2\sum_{t=1}^{T}y_t^2-2\sum_{t=1}^{T}y_ty_{t-1}-2\sum_{t=1}^{T}y_t^2+4\sum_{t=1}^{T}\bar{y}y_t-2\sum_{t=1}^{T}\bar{y}^2\right) \qquad \text{(A.2)} \\
&= \frac{1}{2T}\left(-\sum_{t=1}^{T}y_ty_{t-1}+2\sum_{t=1}^{T}\bar{y}y_t-T\bar{y}^2\right) = \frac{1}{2T}\left(-\sum_{t=1}^{T}y_ty_{t-1}+2T\bar{y}^2-T\bar{y}^2\right) \\
&= \frac{1}{2T}\left(-\sum_{t=1}^{T}y_ty_{t-1}+\sum_{t=1}^{T}y_t\bar{y}\right) \\
&= -\frac{1}{2T}\sum_{t=1}^{T}y_t\left(y_{t-1}-\bar{y}\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
Var\left[\delta_T^2-\frac{s^2}{2}\right] &= \frac{1}{4T^2}Var\left(\sum_{t=1}^{T}y_ty_{t-1}-T\bar{y}^2\right) \\
&= \frac{1}{4T^2}\left(Var\left(\sum_{t=1}^{T}y_ty_{t-1}\right)+Var\left(T\bar{y}^2\right)-2Cov\left(T\bar{y}^2,\sum_{t=1}^{T}y_ty_{t-1}\right)\right).
\end{aligned} \qquad \text{(A.3)}$$

Below, we will expand the three terms in the parenthesis of Equation (A.3) in (a), (b), and (c) respectively. Before proceeding, notice that both δ_T^2 and $\frac{s^2}{2}$ are invariant under μ. So is $\delta_T^2 - \frac{s^2}{2}$ and accordingly, we set $\mu = 0$ without loss of generality.

(a)We have that $Var\left(\sum_{t=1}^{T}y_ty_{t-1}\right) = \left(\sum_{t=1}^{T}Var\left(y_ty_{t-1}\right)\right)+\sum_{t=1}^{T}Cov\left(y_ty_{t-1},\sum_{s\neq t}^{T}y_sy_{s-1}\right)$. But, covariance terms in the summation operator are zero, as we set μ to zero above, and

$$Var\left(\sum_{t=1}^{T} y_t y_{t-1}\right) = \sum_{t=1}^{T} Var\left(y_t y_{t-1}\right) = T\sigma^4.$$

since $Cov(y_t y_{t-1}, y_s y_{s-1}) = 0$ whenever $s \neq t$.
(b)We have $Var\left(T\bar{y}^2\right) = \frac{1}{T^2} Var\left(\left(\sum_{t=1}^{T} y_t\right)^2\right)$. By substituting $\left(\sum_{t=1}^{T} y_t\right)^2 = \sum_{t=1}^{T} y_t^2 + \sum_{t=1}^{T}\sum_{s\neq t} y_s y_t$,

$$Var\left(T\bar{y}^2\right) = \frac{1}{T^2}\left[Var\left(\sum_{t=1}^{T} y_t^2\right) + Var\left(\sum_{t=1}^{T}\sum_{s\neq t} y_s y_t\right) + 2Cov\left(\sum_{t=1}^{T} y_t^2, \sum_{t=1}^{T}\sum_{s\neq t} y_s y_t\right)\right]. \tag{A.4}$$

$Cov\left(\sum_{t=1}^{T} y_t^2, \sum_{t=1}^{T}\sum_{s\neq t} y_s y_t\right)$ is zero since $\mathbb{E}\left(\sum_{t=1}^{T} y_t^2 \times \sum_{t=1}^{T}\sum_{s\neq t} y_s y_t\right)$ and $\mathbb{E}\left(\sum_{t=1}^{T}\sum_{s\neq t} y_s y_t\right)$ are both zero.
We need to derive $Var\left(\sum_{t=1}^{T} y_t^2\right)$ term in Equation (A.4). Since y_t are independent and identical,

$$Var\left[\sum_{t=1}^{T} y_t^2\right] = \left[\sum_{t=1}^{T} Var(y_t^2)\right] = T \times Var(y_t^2), \qquad t \in \{1, 2, \ldots, T\}.$$

y_t^2 is square integrable since $y_t \in L^4$, in other words, fourth (central) moment of normal random variables exists. Then, writing down the probabilistic expression for $Var(y_t^2)$,

$$\begin{aligned} Var(y_t^2) &= \mathbb{E}\left[\left(y_t^2\right)^2\right] - \left\{\mathbb{E}\left[y_t^2\right]\right\}^2 \\ &= \mathbb{E}\left[y_t^4\right] - \left\{\mathbb{E}\left[y_t^2\right]\right\}^2. \end{aligned}$$

Plugging in the second and fourth moments of a normal random variate with zero mean and σ^2 variance parameter, such that, $\mathbb{E}\left[y_t^2\right] = \sigma^2$ and $\mathbb{E}\left[y_t^4\right] = 3\sigma^4$, we have $Var(y_t^2) = 3\sigma^4 - (\sigma^2)^2 = 2\sigma^4$. Then

$$Var\left[\sum_{t=1}^{T} y_t^2\right] = 2T\sigma^4.$$

Next, we need to obtain variance of $\sum_{t=1}^{T}\sum_{s\neq t} y_s y_t$ in terms of σ. Below we do it:

$$Var\left(\sum_{t=1}^{T}\sum_{s\neq t} y_s y_t\right) = \sum_{t=1}^{T}\sum_{s\neq t} Var(y_s y_t) + Cov\left(\sum_{t=1}^{T}\sum_{s\neq t} y_s y_t, \sum_{u=1}^{T}\sum_{\substack{v\neq u\\ v\neq s}} y_u y_v\right). \quad \text{(A.5)}$$

Regardless of the independence between the pairs (y_s, y_t) and (y_u, y_v), we have

$$Cov\left(\sum_{t=1}^{T}\sum_{s\neq t} y_s y_t, \sum_{u=1}^{T}\sum_{\substack{v\neq u\\ v\neq s}} y_u y_v\right) = \sum_{t=1}^{T}\sum_{u=1}^{T}\sum_{s\neq t}\sum_{\substack{v\neq u\\ v\neq s}} Cov(y_s y_t, y_u y_v).$$

$Cov(y_t y_s, y_u y_v)$ is non-zero if and only if $\{t, s\} = \{u, v\}$. Otherwise, if $\{t, s\} \neq \{u, v\}$, at least one of the variables t, s, u, or v will be different from three others by assumption. Without loss of generality, we may suppose t is different from others, and

$$\begin{aligned} Cov(y_t y_s, y_u y_v) &= \mathbb{E}(y_t y_s y_u y_v) - \mathbb{E}(y_t y_s)\mathbb{E}(y_u y_v) \\ &= \mathbb{E}(y_t)\mathbb{E}(y_s y_v y_u) - \mathbb{E}(y_t)\mathbb{E}(y_s)\mathbb{E}(y_u y_v) = 0 \end{aligned}$$

However, $\{t, s\} = \{u, v\}$ equality holds if and only if $t = v$ and $s = u$. Notice that, $t = s = u = v$ case is ruled out by the inequalities $t \neq s$ or $u \neq v$. Further, $t = u$ and $s = v$ case contradicts with $v \neq s$.

Then, sum of the covariances of all pairs for which $\{t, s\} = \{u, v\}$ holds is

$$\sum_{t=1}^{T}\sum_{s\neq t} Cov(y_s y_t, y_t y_s)$$

which is equivalent to $\sum_{t=1}^{T}\sum_{s\neq t} Var(y_s y_t)$. Then,

$$Cov\left(\sum_{t=1}^{T}\sum_{s\neq t} y_s y_t, \sum_{u=1}^{T}\sum_{\substack{v\neq u\\ v\neq s}} y_u y_v\right) = \sum_{t=1}^{T}\sum_{s\neq t} Var(y_s y_t)$$

and, by Equation(A.5),

$$Var\left(\sum_{t=1}^{T}\sum_{s\neq t} y_s y_t\right) = \sum_{t=1}^{T}\sum_{s\neq t} Var(y_s y_t) + \sum_{t=1}^{T}\sum_{s\neq t} Var(y_s y_t)$$
$$= T(T-1)\sigma^4 + T(T-1)\sigma^4 = (2T^2 - 2T)\sigma^4. \qquad (A.6)$$

Hence, by Equations (A.4) and (A.6) we have

$$Var\left(T\bar{y}^2\right) = \frac{1}{T^2}\left[(2T^2-2T)\sigma^4 + 2T\sigma^4\right] = 2\sigma^4.$$

(c)Rewriting,

$$Cov\left(T\bar{y}^2, \sum_{t=1}^{T} y_t y_{t-1}\right) = \frac{1}{T}Cov\left(\left(\sum_{t=1}^{T} y_t\right)^2, \sum_{t=1}^{T} y_t y_{t-1}\right) \qquad (A.7)$$

and we have $\left(\sum_{t=1}^{T} y_t\right)^2 = \sum_{t=1}^{T} y_t^2 + \sum_{t=1}^{T}\left(y_t \sum_{s\neq t}^{T} y_s\right)$. Then,

$$Cov\left(T\bar{y}^2, \sum_{t=1}^{T} y_t y_{t-1}\right) = \frac{1}{T}Cov\left(\sum_{t=1}^{T} y_t^2, \sum_{t=1}^{T} y_t y_{t-1}\right) + \frac{1}{T}Cov\left(\sum_{t=1}^{T} y_t \sum_{s\neq t}^{T} y_s, \sum_{t=1}^{T} y_t y_{t-1}\right)$$

$Cov\left(\sum_{t=1}^{T} y_t^2, \sum_{t=1}^{T} y_t y_{t-1}\right)$ is zero since third raw moment of y_t is zero as we have set μ to zero above. However,

$$Cov\left(\sum_{t=1}^{T} y_t y_{t-1}, \sum_{t=1}^{T}\left(y_t \sum_{s\neq t}^{T} y_s\right)\right) = Var\left(\sum_{t=1}^{T} y_t y_{t-1}\right) = T\sigma^4,$$

since $y_t y_{t-1}$ terms appear on the both terms inside the covariance operator. On the other hand, $Cov(y_t y_{t-1}, y_s y_r)$ are zero unless $s = t$ and $r = t-1$ condition is satisfied. Then, from Equation (A.7) it follows that $Cov\left(T\bar{y}^2, \sum_{t=1}^{T} y_t y_{t-1}\right) = \sigma^4$.

In conclusion of (a), (b) and (c), we have

$$Var\left[\delta_T^2 - \frac{s^2}{2}\right] = \frac{1}{4T^2}\left(T\sigma^4 + 2\sigma^4 - 2\sigma^4\right) = \frac{\sigma^4}{4T}.$$

□

Corollary A.3. *Let* $\{X_t\}$ *be a martingale difference sequence such that* $\mathbb{E}\{|X_t|^{2+\gamma}\} < \infty$ *for some* $\gamma > 0$ *and all* t*. If* $Var(X_t) > 0$ *and* $T^{-1}\sum_{t=1}^{T} X_t^2 - Var(X_t) \xrightarrow{p} 0$ *then* $\frac{\sum_{t=1}^{T} X_t}{\sqrt{T}Var(X_t)} \xrightarrow{(d)} N(0,1)$ *as* $N \to \infty$*.*

Proof. For proof, see White (2001). □

Proposition 2.3. *Under* H_0 *,* $\widetilde{GYO}_W \xrightarrow{(d)} N(0,1)$ *as* $T \to \infty$*.*

Proof. In equation A.2, we obtained the identity

$$\delta_T^2 - \frac{s^2}{2} = -\frac{1}{2T}\sum_{t=1}^{T} y_{t-1}\left(y_t - \bar{y}\right)$$

Define a new series such that $Z_T = T(\delta_T^2 - \frac{s^2}{2})$. Hence,

$$Z_T = \frac{1}{2}\sum_{t=1}^{T} y_{t-1}\left(\bar{y} - y_t\right). \tag{A.8}$$

Z_T is a sum and define the increments of Z_T as $\Delta Z_T := Z_T - Z_{T-1} = y_{T-1}\left(\bar{y} - y_T\right)$. Let $(\Omega, \mathbb{P}, \mathcal{F})$ be the complete probability space on which ΔZ_T lives. Further, let $\mathcal{F}_T$ be the $\sigma-$field generated by $(\Delta Z_1, \Delta Z_2, , \ldots, \Delta Z_T)$, i.e., ΔZ_T is adapted to $\mathcal{F}_T$. Then ΔZ_T is an $\mathcal{F}_T$-martingale difference sequence:

$$\begin{aligned} \mathbb{E}\{\Delta Z_T|\mathcal{F}_{T-1}\} &= \frac{1}{2}\mathbb{E}\{y_{T-1}\left(\bar{y} - y_T\right)|\mathcal{F}_{T-1}\} \\ &= \frac{1}{2}y_{T-1}\left(\bar{y} - \mathbb{E}\{y_T|\mathcal{F}_{T-1}\}\right) = \frac{1}{2}y_{T-1}\left(\mu - \mu\right) = 0. \end{aligned} \tag{A.9}$$

On the other hand, ΔZ_T have finite moments of order greater than 2:

$$\begin{aligned}\mathbb{E}\{|\Delta Z_T|^{2+\gamma}\} &\le \mathbb{E}\{|y_{T-1}|^{2+\gamma}|(y_T-\bar{y})|^{2+\gamma}\}\\ &= \mathbb{E}\{|y_{T-1}|^{2+\gamma}\}\mathbb{E}\{|(y_T-\bar{y})|^{2+\gamma}\}.\end{aligned}$$

The first and second terms inside the expectation operators above are normal random variates. Therefore, since the moments for y_{T-1} and $(y_T-\bar{y})$ exist provided $Var(\Delta z_T)$ is finite, then the moments of all orders for $y_{T-1}(y_T-\bar{y})$ exists as well.

Lastly, by using Chebishev inequality

$$\mathbb{P}\{\omega : |T^{-1}\sum_{t=1}^{T}\Delta Z_t^2 - Var(\Delta Z_t)| > \epsilon\} \le \frac{\mathbb{E}\{|T^{-1}\sum_{t=1}^{T}\Delta Z_t^2 - Var(\Delta Z_t)|\}}{\epsilon}$$
$$\le \left|\frac{\mathbb{E}\{T^{-1}\sum_{t=1}^{T}\Delta Z_t^2 - Var(\Delta Z_t)\}}{\epsilon}\right| = \left|\frac{\mathbb{E}\{T^{-1}\sum_{t=1}^{T}\Delta Z_t^2\} - \mathbb{E}\{Var(\Delta Z_t)\}}{\epsilon}\right|$$

for all $\epsilon > 0$. But, for large T, $\mathbb{E}\{T^{-1}\sum_{t=1}^{T}\Delta Z_t^2\} = \mathbb{E}\{Var(\Delta Z_t)\}$ and we have

$$\mathbb{P}\{\omega : |T^{-1}\sum_{t=1}^{T}\Delta Z_t^2 - Var(\Delta Z_t)| > \epsilon\} = 0 \tag{A.10}$$

for all $\epsilon > 0$. Combining these results with Corollary A.3 we have

$$\sqrt{T}\ \overline{\Delta Z_t} = \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\Delta Z_t = \frac{1}{\sqrt{T}}Z_T \xrightarrow{(d)} N(0,\sigma^2/2).$$

Then, by Equation (A.9) we have

$$\sqrt{T}\left(\frac{\delta_T^2 - s^2/2}{\sigma^2/2}\right) \xrightarrow{(d)} N(0,1). \tag{A.11}$$

Now, we are going to have $s^2/2$ in the denominator of Equation (A.11) instead of $\sigma^2/2$, such that,

$$\sqrt{T}\left(\frac{\delta_T^2 - s^2/2}{s^2/2}\right) = \sqrt{T}\left(\frac{\delta_T^2 - s^2/2}{\sigma^2/2} \times \frac{\sigma^2}{s^2}\right). \tag{A.12}$$

Conniffe and Spencer (2001) shows that $\dfrac{\delta_T^2 - s^2/2}{\sigma^2/2}$ is independent of s^2. Also, we know that $\sqrt{T}\,\dfrac{\sigma^2}{s^2} = 1 + o_P(1)$ [44] . These results sum up to

$$\sqrt{T}\left(\frac{\delta_T^2 - s^2/2}{s^2/2}\right) = \sqrt{T}\left(\frac{\delta_T^2 - s^2/2}{\sigma^2/2}\right) + o_P(1) \tag{A.13}$$

which converges in distribution to $N(0,1)$ by Equation (A.11). □

Proposition 2.8. *Under H_1, assuming a single break, let t_1 be the last point after which the process y_t assumes the new mean μ_2 and let $q = t_1/T$ be the fraction of the data with μ_1 where $q \in \{\frac{1}{T}, \frac{2}{T}, \dots \frac{T-1}{T}\}$. Then,*

$$\mathbb{E}\left\{\widetilde{GYO}_W\right\} = \sqrt{T}\left\{\frac{\sigma^2 + \dfrac{|\mu_1 - \mu_2|^2}{2T}}{\sigma^2 + q(1-q)\,|\mu_1 - \mu_2|^2} - 1\right\}.$$

Proof. The proof consists of three parts. Assuming y_t follows the dynamics defined in Equation (2.2) first we will derive expectation of unconditional variance of y_t, i.e. $\mathbb{E}(s^2)$ and second, $\mathbb{E}(\delta_T^2)$. Finally, we will use Willams (1942)'s result that any moment of the ratio of (δ_T^2) to s^2 is the ratio of the moments of δ_T^2 to s^2 to obtain $\mathbb{E}\left\{\dfrac{\delta_T^2}{s^2}\right\}$.

(a) We have

$$s^2 = \mathbb{E}\left\{(y_t - \mathbb{E}y_t)^2\right\} = \mathbb{E}\left\{y_t^2\right\} - \left(\mathbb{E}\left(y_t\right)\right)^2 \tag{A.14}$$

[44] $x_t = o_p(1)$ denotes $\lim_{t\to\infty} \mathbb{P}\{|x_t| \geq \epsilon\} = 0,\ \forall \epsilon > 0$.

by definition, and,unconditional expectation of y_t and y_t^2 are sums of conditional expectations weighted with probabilities:

$$\mathbb{E}\left(y_t\right) = \sum_{i=1}^{z} \mathbb{P}\{t \in (t_{i-1}, t_i]\} \times \mathbb{E}\left[y_t \,\middle|\, t \in (t_{i-1}, t_i]\right] = \sum_{i=1}^{z} q_i \mu_i$$

and

$$\mathbb{E}\left(y_t^2\right) = \sum_{i=1}^{z} \mathbb{P}\{t \in (t_{i-1}, t_i]\} \times \mathbb{E}\left[y_t^2 \,\middle|\, t \in (t_{i-1}, t_i]\right] = \sum_{i=1}^{z} q_i(\mu_i^2 + \sigma^2) = \sigma^2 + \sum_{i=1}^{z} q_i \mu_i^2.$$

Then, substituting $\mathbb{E}\left(y_t\right)$ and $\mathbb{E}\{y_t^2\}$ into Equation(A.14):

$$s^2 = \sigma^2 + \sum_{i=1}^{z} q_i \mu_i^2 - \left(\sum_{i=1}^{z} q_i \mu_i\right)^2 = \sigma^2 + \sum_{i=1}^{z} q_i \mu_i^2 - \sum_{i=1}^{z}\sum_{j=1}^{z} q_i q_j \mu_i \mu_j.$$

(b) In this part we will obtain $\mathbb{E}\delta_T^2$ under H_1.

$$\begin{aligned}
\mathbb{E}\delta_T^2 &= \mathbb{E}\left\{(4T)^{-1} \sum_{j=1}^{T} \left(y_t - y_{j-1}\right)^2\right\} \\
&= (4T)^{-1}\mathbb{E}\left\{\sum_{i=0}^{z-1} \sum_{t=t_i+1}^{t_{i+1}} \left(y_t - y_{j-1}\right)^2 + \sum_{i=0}^{z-1} \left(y_{t_i+1} - y_{t_i}\right)^2\right\}
\end{aligned}$$

Since expectation is a linear operator we can distribute the expectations inside the sums and we have that $\mathbb{E}\left\{\left(y_t - y_{t-1}\right)^2\right\} = \mathbb{E}\left\{\left(\varepsilon_t - \varepsilon_{t-1}\right)^2\right\} = 2\sigma^2$ whenever $t \notin \{t_0 + 1, t_1 + 1, \ldots, t_{z-1} + 1\}$. On the other hand, when $t \in \{t_0 + 1, t_1 + 1, \ldots, t_{z-1} + 1\}$

$$\begin{aligned}
\mathbb{E}\left\{\left(y_{t_i+1} - y_{t_i}\right)^2\right\} &= \mathbb{E}\left\{\left(\,|\mu_{i+1} - \mu_i| + \varepsilon_j - \varepsilon_{j-1}\right)^2\right\} \\
&= |\mu_{i+1} - \mu_i|^2 + \mathbb{E}\left\{\left(\varepsilon_j - \varepsilon_{j-1}\right)^2\right\} \qquad \text{(using } \mathbb{E}\varepsilon_j = 0\text{)} \\
&= |\mu_{i+1} - \mu_i|^2 + 2\sigma^2.
\end{aligned}$$

Then,

$$\mathbb{E}\delta_T^2 = (4T)^{-1}\left(2\sigma^2(T-z) + 2\sum_{i=0}^{z-1}|\mu_{i+1}-\mu_i|^2\right)$$
$$= \frac{\sigma^2(T-z)}{2T} + \frac{\sum_{i=0}^{z-1}|\mu_{i+1}-\mu_i|^2}{2T} = \frac{\sigma^2}{2} + \frac{\sum_{i=0}^{z-1}|\mu_{i+1}-\mu_i|^2 - z\sigma^2}{2T} \cong \frac{\sigma^2}{2}, \quad \text{for large } T.$$

(c) In the final part of the proof, we will combine the results in (a) and (b) by making use of a particular conclusion of the result of Williams (1941) such that

$$\mathbb{E}\left\{\frac{\delta_T^2}{s^2/2}\right\} = \frac{\mathbb{E}\left\{\delta_T^2\right\}}{\mathbb{E}\left\{s^2/2\right\}}. \tag{A.15}$$

Then we have

$$\mathbb{E}\left\{\frac{\delta_T^2}{s^2/2} - 1\right\} = \frac{\sigma^2}{\sigma^2 + \sum_{i=1}^{z} q_i\mu_i^2 - \sum_{i=1}^{z}\sum_{j=1}^{z} q_i q_j \mu_i \mu_j} - 1$$

as $T \to \infty$ and we are done. □