Universidade de Lisboa
Instituto Superior de Economia e Gestão

Essays on Econometrics: Nonlinearities and Nonnormalities

Autor: Gabriel Florin Zsurkis

Orientadores:
Prof. Doutor João Carlos Henriques da Costa Nicolau
Prof. Doutor Paulo Manuel Marques Rodrigues

Tese elaborada especialmente para obtenção do grau de Doutor em Matemática Aplicada à Economia e à Gestão

2021

Universidade de Lisboa
Instituto Superior de Economia e Gestão

Essays on Econometrics: Nonlinearities and Nonnormalities

Autor: Gabriel Florin Zsurkis

Orientadores:
Prof. Doutor João Carlos Henriques da Costa Nicolau
Prof. Doutor Paulo Manuel Marques Rodrigues

Tese elaborada especialmente para obtenção do grau de Doutor em Matemática Aplicada à Economia e à Gestão

Júri
Presidente:
Prof. Doutor Nuno João de Oliveira Valério
Professor Catedrático e Presidente do Conselho Científico
Instituto Superior de Economia e Gestão da Universidade de Lisboa

Vogais:
Prof. Doutor João Carlos Henriques da Costa Nicolau
Professor Catedrático
Instituto Superior de Economia e Gestão da Universidade de Lisboa

Prof. Doutor Luís Filipe Farias de Sousa Martins
Professor Associado com agregação
ISCTE Business School do ISCTE - Instituto Universitário de Lisboa

Prof. Doutora Cristina Alexandra de Oliveira Amado
Professora Auxiliar
Escola de Economia e Gestão da Universidade do Minho

Prof. Doutor Paulo João Figueredo Cabral Teles
Professor Auxiliar
Faculdade de Economia da Universidade do Porto

Prof. Doutor Nuno Ricardo Martins Sobreira
Professor Auxiliar
Instituto Superior de Economia e Gestão da Universidade de Lisboa

2021

This work is dedicated to Catarina.

Abstract

This Dissertation consists of three independent papers on econometrics, having in common the fact that each of them proposes a new methodology to deal with issues caused by the departure from linearity and gaussianity assumptions.

We start by introducing a simple and easy to implement procedure to test for multiple structural changes in persistence. An in-depth Monte Carlo analysis shows that the new procedure performs well under various DGPs with persistence changes. The application of the proposed test to OECD countries inflation reveals relevant statistical evidence of breaks in persistence for all countries. Overall, the persistence was high and non-mean-reverting until the early 80's and subsequently decreased, which coincides with the beginning of the Great Moderation.

Then, the second paper introduce a flexible framework able to capture some aspects of the potential nonlinear causal relationships between economic variables. More precisely, the proposed procedure estimates the expected time (ET) an outcome variable takes to cross a fixed threshold given a starting value and conditional on covariates. An application to the economic activity-yield spread relationship for the U.S. suggests that the yield spread may have an important role in stimulating a faster return to desirable growth rates when the economy is in contraction or faces weak growth. Moreover, negative yield spread values in the presence of positive and high industrial production growth rates leads to a quick return to negative growth rates and may trigger a recession.

Finally, the third paper proposes a simple framework that allows us to take into account the magnitude of potential losses incurred throughout the investment horizon, denoted intra-horizon risk, in portfolio optimization. To this end, we introduce a novel nonparametric method to estimate the first passage probability function that only make use of the Markovian property of the returns. An empirical application is provided considering equity, bond and commodity Exchange Traded funds (ETFs). Our results suggest that the proposed framework indicates portfolios with lower expected time to reach the target return than those indicated by the Markowitz' mean-variance approach with similar levels of intra-horizon risk, which may result in higher expected annualized return if the lower threshold that triggers a stop-loss decision is not crossed.

Acknowledgements

I would like to thank to my supervisors, Professors João Nicolau and Paulo M. M. Rodrigues, for their inspiring guidance and great support.

Contents

Abstract

Acknowledgements

Acronyms

Symbols

1 Introduction **1**

2 A Re-Examination of Inflation Persistence Dynamics **5**
2.1 Introduction 5
2.2 Motivation and Proposed Statistic 7
2.2.1 The Test Procedure 7
2.2.2 Unconditional and Conditional Heteroskedasticity 9
2.2.3 Asymptotic Distributions 10
2.3 Monte Carlo Analysis 12
2.3.1 IID Innovations 13
2.3.2 Serially Correlated Errors 19
2.3.3 Heteroskedasticity 22
2.4 Empirical Application 31
2.5 Conclusions 38

3 The Expected Time to Cross a Threshold and its Determinants **39**
3.1 Introduction 39
3.2 The Proposed Methodology 42
3.2.1 The Process and Probabilities of Interest 42
3.2.2 Covariate-Dependent Transition Probabilities 43
3.2.3 Parameter Estimation 44
3.2.4 Covariate-Dependent ET 46
3.3 Monte Carlo Analysis 48
3.4 Empirical Application 51
3.4.1 Data 52
3.4.2 Empirical Results 53
3.5 Conclusions 57

4 The Importance of Intra-Horizon Risk in Portfolio Optimization **59**
4.1 Introduction 59
4.2 The Proposed Approach for Portfolio Selection 61

4.2.1 The Proposed Nonparametric Method to Estimate the First Hitting Time Probabilities 61
4.2.2 The Optimization Problem 66
4.2.3 The Optimization Algorithm 67
4.3 Empirical Application 69
4.4 Conclusions 75

5 Concluding Remarks **77**

A Appendix to "A Re-Examination of Inflation Persistence Dynamics" **79**
A.1 Proof of Main Results 79
A.2 Additional Tables 85

B Appendix to "The Expected Time to Cross a Threshold and its Determinants" **87**

C Appendix to "The Importance of Intra-Horizon Risk in Portfolio Optimization" **91**

Bibliography **102**

List of Figures

2.1 The cosine function for non-integer and integer values of k . . . 9
2.2 Year-on-year growth of the CPI for the G7 countries 35
2.3 Quarterly growth of the CPI for the G7 countries 36

3.1 U.S. Industrial Production (monthly) growth rate and yield spread 53
3.2 Unconditional and conditional estimated expected time curves (ETC) with overlapping block bootstrap 95% CI 56

4.1 Prices of the four ETFs considered 70
4.2 Nonparametric IH and EH risks probabilities 74

List of Tables

2.1 Values for the non-centrality parameter $\tilde{c}$ for different values of k 12
2.2 Critical values 12
2.3 Empirical size and power with *iid* errors 15
2.4 Empirical size and power with *iid* errors - linear trend case . . . 16
2.5 Empirical power when the breaks in persistence are abrupt . . . 17
2.6 Empirical power when the breaks in persistence are abrupt - linear trend case 18
2.7 Finite sample size and power with ε_t autocorrelated 20
2.8 Finite sample sizes and power with ε_t autocorrelated - linear trend case 21
2.9 Empirical size under unconditional variance breaks 23
2.10 Empirical size under unconditional variance breaks - linear trend case 24
2.11 Empirical size and power of Wild bootstrap based statistics under unconditional variance breaks. 25
2.12 Empirical size and power of Wild bootstrap based statistics under unconditional variance breaks - linear trend case 26
2.13 Empirical size assuming *iid* errors in the presence of GARCH effects 28
2.14 Empirical size assuming *iid* errors in the presence of GARCH effects - linear trend case 28
2.15 Empirical size and power in the presence of GARCH effects using EW standard errors and Wild bootstrap based test statistics . . 29
2.16 Empirical size and power in the presence of GARCH effects using EW standard errors and Wild bootstrap based test statistics - linear trend case 30
2.17 Results using year-on-year and quarterly growth of the CPI for the G7 countries 37

3.1 Means and SD of Markov chain parameter estimates 50
3.2 Estimated parameters for $p_i(\mathbf{x})$, $i = 1, ..., r$ 54

4.1 Returns Correlation Matrix 70
4.2 Optimal ETFs allocation suggested by the proposed methodology. 73

A.1 Comparison with the M test of Leybourne et al. (2007) - empirical size 85
A.2 Comparison with the M test of Leybourne et al. (2007) - empirical power 85

Acronyms

AIC	Akaike Information Criterion
ADF	Augmented Dickey-Fuller test
AR(p)	Autoregressive Model of order p
ARMA(p, q)	Autoregressive Moving Average Model of order p,q
BIC	Bayesian Information Criterion
CDF	Cumulative Distribution Function
CPI	Consumer Price Index
DGP	Data Generation Process
DF	Dickey-Fuller test
EH	End-of-Horizon risk
ET	Expected Hitting Time
ETC	Expected Time Curve
EW	Eicker-White robust standard errors
GARCH	Generalized Autoregressive Conditional Heteroskedasticity Model
GBM	Geometric Brownian Motion
GLD	SPDR Gold Shares ETF ticker
GLS	Generalized Least Squares
IH	Intra-Horizon risk
IID	Independent and Identically Distributed
IP	Industrial Production
IJS	iShares S&P Small-Cap 600 Value ETF (IJS)
LQD	iShares iBoxx $ Investment Grade Corporate Bond ETF
MAIC	Modified Akaike Information Criterion
MLE	Maximum Likelihood Estimation
MSE	Mean Square Error
MV	Mean-Variance Optimization
OECD	The Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
OU	Orstein-Uhlenbeck process
QMLE	Quasi-Maximum Likelihood Estimation
SD	Standard Deviation
TA	Threshold Accepting algorithm
TLT	iShares 20+ Year Treasury Bond ETF
VaR	Value-at-Risk
WB	Wild Bootstrap
YS	Yield Spread

Symbols

$E(A\|\mathbf{x})$	The expected time of A given $\mathbf{x}$
$P(A\|\mathbf{x})$	The conditional probability of A given $\mathbf{x}$
Λ	The cumulative density function of the logistic distribution
$\mathcal{I}(A)$	The indicator function, equals 1 if A is true and 0 otherwise
$\stackrel{p}{\to}$	Convergence in probability
$\stackrel{d}{\to}$	Convergence in distribution or weak convergence
$Avar(A)$	The asymptotic variance of A
$Cov(A)$	The covariance of A
$Var(A)$	The variance of A
$\{(y_t);\, t = 0, 1, 2, ...\}$	The stochastic process y_t
$I(\mathrm{k})$	Integrated of order k
$\Rightarrow$	Weak convergence to stochastic integrals
$a := b$	a defines b
$\mathbf{x}$	The line-vector $\mathbf{x}$
$\mathbf{x}'$	The transpose of the line-vector $\mathbf{x}$
$\boldsymbol{\beta}$	The line-vector of parameters $\boldsymbol{\beta}$

Chapter 1

Introduction

This Dissertation introduces new methodologies to deal with some aspects of the departure from linearity and gaussianity assumptions, which nowadays are accepted as common features of economic and financial variables.

For instance, the occurrence of exogenous shocks, such as crises, policy decisions, preference changes or technological advances can lead to abrupt shifts in economic and financial time series, invalidating the use of linear models that assume parameters constancy. Since the presence of structural breaks may affect many or all of the model parameters, leading to inconsistent estimates and poor forecasts if not properly modeled, several statistical procedures were proposed to detect them. Perron (2005) provides an interesting survey. A specific and widely investigated consequence of ignoring structural changes is the misspecification of the economic time series' order of integration. Properly determining whether the stochastic process is difference-stationary or trend-stationary[1] has important implications for economic theory and policy, since accepting the unit root model implies that random shocks have a permanent effect on the economy. In fact, the findings of Nelson and Plosser (1982) that economic time series are $I(1)$ provided support to real business cycle theories and gave rise to interest in cointegrated relationships (see Engle and Granger; 1987).

However, beginning with Perron (1989) and Rappoport and Reichlin (1989), many works allowing for breaks in the deterministic kernel of the data generation process have presented evidences that the majority of shocks to the key economic variables may be transitory and that only few events would have permanent effects (see, for instance, Perron; 1997, Lumsdaine and Papell; 1997 and Papell et al.; 2000). More recently, Prodan (2008) demonstrate that it can be difficult to estimate the number and the magnitude of multiple breaks, especially when the breaks are of opposite sign. In order to circumvent this problem, Enders and Lee (2012) and Rodrigues and Taylor (2012) propose to approximate structural changes of unknown functional form in the deterministic component using Fourier expansions (see Gallant; 1981). This approach reduces the specification problem to the selection of the appropriate frequency components of the Fourier form approximation. Moreover, this type of modelling is also useful in the presence of smooth changes, since it is more consistent with the fact that shifts in economic aggregates are influenced by changes in the behavior of a very large number of agents that may not react simultaneously to a given shock.

---

[1] See Stock (1994) for a more complete description of the usefulness of this analysis.

Another important form of structural changes is characterized by shifts in persistence of a time series, from $I(1)$ to $I(0)$ or vice versa. For example, DeLong and Summers (1988) found that shocks to real output series of the US and European countries were less persistent in the post-World War II years; and Alogoskoufis and Smith (1991) report that inflation persistence has changed over time in the US and that it seems to be a positive function of the degree of monetary and exchange-rate accommodation. Tests for changes in persistence have been developed by, among others, Kim (2000), Leybourne et al. (2003) and Leybourne et al. (2007). However, most of these proposes statistics consider a single shift in persistence and work on testing for multiple changes in persistence is still scant.

The first essay, *A re-examination of inflation persistence dynamics in OECD countries: A new approach*, looks to contribute to this literature by proposing a simple and easy to implement procedure which allows testing for (multiple) persistence changes. Instead of considering two fundamentally different processes (one stationary and the other nonstationary) connected at a point in time, this work proposes a new unit root test that approximates the time-varying autoregressive parameter using a parsimonious cosine function related to the Fourier series. Changes in the autoregressive parameter impact the deterministic terms of the process as well as its unconditional variance. The proposed procedure takes this effect into account by local GLS de-trending the data with a time-varying autoregressive parameter. Unlike Leybourne et al. (2007), estimates for the break dates are not provided. However, as the complete sample is used for estimation and not fractions of the sample as with recursive tests, our approach is expected to have better power performance when shifts in the autoregressive coefficient are the only cause for changes in the trend function. An in-depth Monte Carlo analysis shows that the new procedure has good power and size properties in small samples under various data generating processes with persistence changes. An empirical application to G7 countries' inflation data is also provided. A clear understanding of the inflation-generating process is of great importance for central banks since its evolution influences the monetary policy decisions. For instance, a price stability-oriented monetary policy becomes impossible to implement if the inflation series follows the path of a pure $I(1)$ process. Our results reject the $I(1)$ hypothesis for all countries and suggest that the apparent highly persistent behavior of inflation is probably caused by exogenous events (e.g., crises) and shifts in monetary policy, which have altered its speed of reversion to equilibrium after a shock.

Beyond the limitations reported for the univariate context, linear functions with constant parameters may also have difficulties in properly capturing the relationships between major economic variables, since they typically display nonlinear dynamics (see, for instance, Terasvirta et al.; 2010). Several nonlinear conditional mean models are nowadays available in the literature (see Terasvirta; 2006 for a survey). However, these models usually require a rigid parametric functional form and the estimation of a considerable number of parameters. Thus, it may be useful to consider alternative ways of capturing some aspects of the possible nonlinear relationships between a dependent variable and a set of covariates. A possible approach is to employ binary random variables constructed from underlying continuous stochastic processes in order to focus on particular characteristics such as the frequency and length of some events. For instance, some macroeconometric research considers binary response models to predict a recession dummy indicator using economic and financial variables (see, for instance, Estrella and Mishkin; 1998 and Birchenhall et al.; 1999). Most standard models for binary responses do not incorporate information provided by the past values of the dependent variable, which is crucial in application to typically autocorrelated time series data. The presence of serial dependence can easily be taken into account if the binary variable is treated as a Markov process.

In the second essay (Chapter 2), *The expected time to cross a threshold and its determinants: A simple and flexible framework*, we introduce a flexible framework that relies on the Markov assumption and allows us to obtain covariate-dependent first passage time probabilities without requiring a rigid parametric functional form. The major advantage of the proposed approach is that it provides a simple method to estimate the covariate-dependent expected time (ET) to cross a threshold. Understanding how a set of variables influences the ET a dependent variable takes to cross a fixed threshold (for instance, its mean) given a starting value may provide relevant insights on the nature of the casual relationships between economic variables. For instance, the proposed methodology allows us to infer whether the effects of the explanatory variables on the ET are more (or less) pronounced for a subset of starting values of the outcome variable. It may also be a useful tool to support macroeconomic policy decisions since it estimates how long it will take for an outcome variable to reach a target value given a starting point and specific values of relevant covariates representing economic policy instruments, which is a useful statistic to investigate relevant topics such as dynamic controllability. According to Buiter and Gersovitz (1981), a system is dynamically controllable if a path for the economic instruments exists which is capable of moving the vector with the economic objectives from any initial value to any other target value in pre-assigned finite time. In order to illustrate the potential of the proposed approach, we investigated the effect of the yield spread (YS) on industrial production (IP) for the U.S. economy. The relationship between YS and economic activity has been investigated by an extensive literature. For instance, among others, Harvey (1989), Stock and Watson (1989) and Estrella and Mishkin (1998) found statistical evidence that YS predicts future output growth considering a linear framework of analysis. More recently, Galbraith and Tkacz (2000) found evidence in support of the asymmetric impact of YS on the conditional expectation of output growth using nonlinear threshold autoregressive (TAR) models. Moreover, Estrella and Mishkin (1998) and Kauppi and Saikkonen (2008) show, using probit or logit models, that YS has also been successfully used in predicting recessions. The proposed approach provides additional information about the economic activity-yield spread relationship. Our results suggest that the YS may have an important role in stimulating a faster return to desirable growth rates when the economy is in contraction or faces weak growth. Moreover, the YS value seems also critical when the IP growth rate is larger than average. If YS is negative, the IP growth rate will return quickly to below average values.

This finding may be related to the already mentioned ability of the yield curve inversion (negative YS) to predict recessions.

In the beginnings, most modern financial models were operationalized assuming linear relationships between variables and joint normality on their distributions. Popular examples comprise the Capital Asset Pricing Model (CAPM) and parametric Value at Risk (VaR) models. Furthermore, normal distribution also represents a key assumption of the Markowitz' mean-variance approach (Markowitz; 1952, 1959) for portfolio optimization. But, starting with Mandelbrot (1963), departures from the normal distribution were widely documented and it is nowadays accepted that the distribution of financial variables such as security returns is in general skewed and leptokurtotic (see, for instance Cont; 2001 and Nicolau; 2012). Thus, extreme negative events tend to occur more often than under normality and ignoring this feature may result in a misleading evaluation of the risk. A widely used approach to quantify the market risk is the well-known Value at Risk (VaR), a quantile measure of the profit and loss distribution at the end of a specified trading horizon. However, in a mark-to-market environment where sharp decline in asset values can affect trading strategies, it is critical to also take into the magnitude of the potential losses incurred prior to the final horizon, denoted intra-horizon (IH) by Bakshi and Panayotov (2010). The methodologies proposed to estimate the IH risk are mainly based on first passage probabilities, the probability that an event occurs for the first time within a finite horizon. Bakshi and Panayotov (2010) warns to the limitations of the Geometric Brownian motion framework, which assumes that returns are normally distributed, in quantifying the IH risk. They consider Lévy jump models and show that the presence of extreme events (jumps) tends to amplify this risk. However, analytical expressions for the first passage probability of processes with jumps are generally not available (see, for instance, Kyprianou; 2006). Moreover, choosing the most appropriate one from great number of Lévy-type already proposed in the literature (see, for instance, Madan and Seneta; 1990 and Carr et al.; 2002) may be a very challenging task.

In Chapter 4, *The importance of intra-horizon risk in portfolio optimization*, we introduce a novel nonparametric method to estimate the first hitting time probability function for nonstationary processes. Since the proposed methodology only requires the Markovian property of returns, it is flexible enough to accommodate jumps and other nonlinearities in asset prices. This nonparametric approach is used in a new portfolio optimization problem that allow us to incorporate the intra-horizon risk (IH) in asset allocation decisions. In short, the optimization problem aims to minimize the expected time to achieve a target cumulative return rate given that the probability of breaching a lower threshold is maintained at a level considered acceptable. This probability measures the intra-horizon risk the investor is willing to accept. In order to illustrate the proposed methodology for portfolio selection, an empirical application considering equity, bond and commodity Exchange Traded funds (ETFs). Our results suggest that the proposed framework indicates portfolios with lower expected time to reach the target cumulative return than those indicated by the Markowitz' mean-variance approach with similar levels of intra-horizon risk, which implies a higher annualized return if the lower threshold is not crossed.

Chapter 2

A Re-Examination of Inflation Persistence Dynamics in OECD Countries: A New Approach*

2.1 Introduction

Structural breaks in time series result from the occurrence of exogenous shocks, such as crises or policy decisions, which may have permanent effects on the variables' dynamics. For instance, Perron (1989) showed that the Great Crash of 1929 caused a dramatic decrease in the mean of most aggregate variables of the US economy and that the 1973 oil price shock possibly caused a change in the slope of the output's trend function which was responsible for the subsequent slowdown of its growth rate.

The persistence of time series is frequently assessed through its order of integration, which is determined from the results of a unit root test. However, since the finite sample power performance of traditional unit root tests is not satisfactory when structural breaks are present in the data, this has led to the development of statistics that allow for changes in the deterministic kernel of the data generation process (Perron, 2005, provides an interesting survey). However, most procedures available consider that structural breaks occur instantaneously, which may not be consistent with the fact that shifts in economic aggregates are influenced by changes in the behavior of a very large number of agents that may not react simultaneously to a given shock.

Enders and Lee (2012) and Rodrigues and Taylor (2012) proposed tests that do not require assumptions about the number of breaks and their exact form. To this end Fourier terms have been used to approximate structural changes of unknown functional form in the deterministic component, reducing the specification problem to the selection of the appropriate frequency components of the Fourier approximation. Since from a spectral frequency perspective structural breaks are usually associated to the zero frequency, low-frequency Fourier terms are commonly employed. Interestingly, a small number of low-frequency terms

*Chapter 2, *A Re-Examination of Inflation Persistence Dynamics in OECD Countries: A New Approach*, has been published in co-authorship with João Nicolau and Paulo M. M. Rodrigues in the Oxford Bulletin of Economics and Statistics and is reprinted in this Dissertation with permission from John Wiley & Sons Ltd and the Department of Economics, University of Oxford.

capture a great variety of breaks and sometimes even just a single frequency is sufficient.

In recent literature, the simple stationary/nonstationary ($I(0)/I(1)$) dichotomy which is typically considered when analyzing the properties of time series has been questioned and it has been suggested that certain macroeconomic and financial time series may display changes in persistence over time, i.e., changes from low to high persistence or vice versa. For example, Alogoskoufis and Smith (1991) report that inflation persistence has changed over time in the US and that it seems to be a positive function of the degree of monetary and exchange-rate accommodation. Another example is in DeLong and Summers (1988) who found that shocks to real output series of the US and European countries were less persistent in the post-World War II years.

Several approaches to test for changes in persistence have recently been developed. Most procedures are related either to the residual-based test for stationarity introduced by Kim (2000) or to the sub-sample augmented Dickey-Fuller type tests of Banerjee et al. (1992). The first are based on the ratio of two partial sum processes of the residuals from regressions of the time series of interest on deterministic components before and after a given break date (Busetti and Taylor; 2004 and Harvey et al.; 2006). Since the break date is typically unknown, statistics based on the ratios for all possible break dates, such as the maximum Chow-type test, are considered. Regarding the second class of tests, these are based on the minimum of a sequence of ADF type statistics computed by recursive least squares across changing sub-samples of the data. For instance, Leybourne et al. (2003) extended the work of Banerjee et al. (1992) to allow for local GLS de-trended ADF tests of the null hypothesis of a stable unit root process versus a switch from I(0) to I(1) or vice versa. However, Leybourne et al. (2007) point out that tests for a single change in persistence may not be consistent against processes with multiple changes in persistence. To overcome this drawback, they introduced tests based on doubly-recursive sequences of ADF-type unit root statistics which are valid in the presence of multiple shifts in persistence regardless of the direction of change.[1]

Moreover, a change in persistence, as is typically considered, also originates shifts in the deterministic component of the process (see Section 2.2). The tests proposed in this paper take this effect into account by performing local GLS de-trending of the data with a time-varying autoregressive parameter. Specifically, the tests introduced approximate parameter changes using a single cosine function, but unlike Leybourne et al. (2007) do not provide estimates for the break dates. However, as the complete sample is used for estimation and not fractions of the sample as with recursive tests, it is expected that these approaches have better power performance than the latter when shifts in the autoregressive parameter are the only cause for changes in the trend function.[2]

[1]Phillips et al. (2011) introduced a closely related methodology based on recursive right tailed ADF tests for the detection of speculative bubbles; see also Phillips et al. (2015) and references therein.

[2]Monte Carlo results to support this statement are presented in Tables A.1 and A.2 of Appendix A.

Work on testing for multiple changes in persistence is still scant. This paper looks to contribute to this literature by proposing a simple and easy to implement procedure which allows testing for (multiple) persistence changes.

The remainder of the paper is organized as follows. Section 2.2 introduces the test procedures and derives their asymptotic distributions under the null and local alternative hypotheses. Section 2.3 investigates the finite sample properties of the statistics through Monte Carlo simulations. Specifically, the impact of conditional heteroskedasticity, breaks in the innovation variance and serially correlated errors are examined. Section 2.4 presents an in-depth analysis of inflation data of seven OECD countries: Canada, France, Germany, Italy, Japan, the UK and the US. Section 2.5 concludes, and finally, the Appendix A provides detailed proofs of all results presented throughout the paper.

2.2 Motivation and Proposed Statistic

Consider a model for persistence changes in line with Harvey et al. (2006), i.e.,

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t \tag{2.1}$$

$$u_t = \rho_t u_{t-1} + \varepsilon_t, \tag{2.2}$$

where $\mathbf{x}_t$ is a deterministic kernel which is either a constant or a constant and time trend (i.e. $\mathbf{x}_t := 1$ or $\mathbf{x}_t := [1, t]'$), $\boldsymbol{\beta}$ is the vector of parameters that captures the deterministic structure and (2.2) describes the stochastic behavior of y_t. The innovations ε_t follow a mean zero process satisfying the α-mixing conditions of Phillips and Perron (1988, page 336) with strictly positive and bounded long-run variance $\omega^2 \equiv \lim_{T\to\infty} \frac{1}{T} \mathrm{E}(\sum_{t=1}^{T} \varepsilon_t)^2$.

2.2.1 The Test Procedure

To implement the persistence change test a two-step approach as in Elliott et al. (1996) is employed. First, the time series of interest, y_t, is locally GLS de-trended using $\tilde{\rho}_t := 1 + \frac{\tilde{c}}{T}\cos(k,t)$ where $\tilde{c}$ is fixed and non-positive, k is fixed[3] and

$$\cos(k,t) := \frac{1 + \cos(2\pi kt/T)}{2} = \cos^2(\pi kt/T). \tag{2.3}$$

A simple cosine function with a single frequency is used in order to mimic the pattern of the unknown shifts in the autoregressive parameter ρ_t. The de-trended variable is computed as, $\hat{u}_{\tilde{c},t} = y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}_{\tilde{\mathbf{c}}}$, where $\mathbf{x}_t = 1$ (demeaned case) or $\mathbf{x}_t = [1, t]'$ (linear trend case), and $\hat{\boldsymbol{\beta}}_{\tilde{c}} = \left(\sum_{t=1}^{T} \mathbf{x}_{\tilde{c},t}\mathbf{x}_{\tilde{c},t}{}'\right)^{-1} \sum_{t=1}^{T} \mathbf{x}_{\tilde{c},t} y_{\tilde{c},t}$, with $y_{\tilde{c},1} = y_1$, $\mathbf{x}_{\tilde{c},1} = \mathbf{x}_1$, $y_{\tilde{c},t} := y_t - \tilde{\rho}_t y_{t-1}$ and $\mathbf{x}_{\tilde{c},t} := \mathbf{x}_t - \tilde{\rho}_t \mathbf{x}_{t-1}$, for $t > 1$. As the $\cos(k, t)$ function takes values between 0 and 1, this approach can be seen as local GLS de-trending with time-varying weights.

In the second step, the presence of a unit root in $\hat{u}_{\tilde{c},t}$ is investigated considering, for k known and fixed, the t-statistic on ϕ computed from the test

[3] Note that when $k = 0$, $\tilde{\rho}_t$ corresponds to the representation used by Elliott et al. (1996).

regression

$$\Delta\hat{u}_{\tilde{c},t} = \phi\cos(k,t)\hat{u}_{\tilde{c},t-1} + \sum_{j=1}^{p}\delta_j\Delta\hat{u}_{\tilde{c},t-j} + \eta_t, \tag{2.4}$$

where under the null hypothesis of a unit root, $H_0 : \phi = 0$, and under the alternative $H_A : \phi < 0$. The non-centrality parameter $\tilde{c}$ will assume different values depending on the deterministic component and also on the frequency parameter k considered (see Table 2.1 in Section 2.3)[4]. In (2.4), p denotes the lag truncation order chosen to account for any weak dependence in $\{\varepsilon_t\}$. More generally, when ε_t is a linear process satisfying standard summability and moment conditions, p needs to be such that $1/p + p^3/T \to \infty$ as $T \to \infty$; Said and Dickey (1984) and Chang and Park (2002). Since the difference between the proposed test and DF^{GLS} is only due to the presence of some deterministic terms when k is known, the results of Chang and Park (2002) remain valid in this context.

Changes in persistence impact the conditional and unconditional means as well as the unconditional variance. Hence, both de-trending and testing steps are influenced by the autoregressive parameter ρ_t in (2.2). Moreover, the assumption of parameter constancy ($\rho_t = \rho$), when invalid, seems to favor the null hypothesis of a unit root. Thus, it is important to propose tests that allow for changes in persistence under the alternative hypothesis. Since the number of breaks and its functional form are typically unknown in practice, trigonometric functions have been used to accommodate these features. For instance, Fourier series which are linear combinations of sine and cosine functions are widely applied in this context; Gallant (1981), Enders and Lee (2012) and Rodrigues and Taylor (2012).

In our framework, we use a single factor since the increase in flexibility of the functions employed (i.e. more frequency terms) to describe parameter changes has been associated with a deterioration in the power performance of the tests as a consequence of over-fitting the data; Enders and Lee (2012).

The $\cos(k,t)$ function in (2.3) is crucial to properly approximate, but not to identify persistence change dates. Its shape is entirely determined by the frequency parameter k. Most empirical work using Fourier terms has only considered integer values for k (Enders and Lee; 2012 and Rodrigues and Taylor; 2012), which imply that the starting and ending values of $\cos(k,t)$ are the same. For instance, considering $k = 1$, the resulting function may be useful in cases with two breaks, where an increase in ρ_t somewhere in the middle of the sample is followed by a decrease of similar magnitude later. However, when there is an increase in persistence at an unknown point in time and the parameter does not return to its initial value, a fractional frequency needs to be considered (see Figure 2.1 for illustration).

In economics, it is generally not expected that many breaks in persistence occur given the relatively small sample sizes usually available. Thus, under the alternative hypothesis we allow for a maximum of three periods of persistence change (however this assumption can be easily relaxed if necessary).

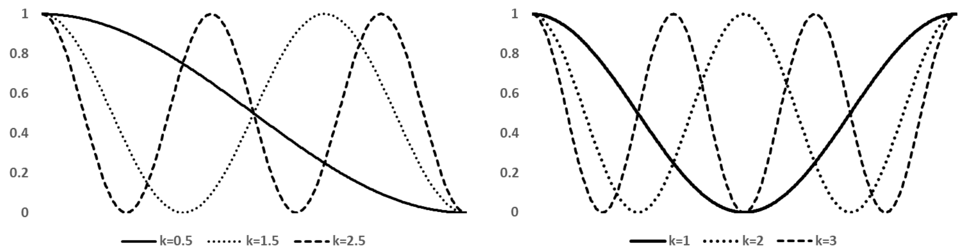
[4] Elliott et al. (1996) showed that there is no uniformly most powerful unit root test and proposed choosing $\tilde{c}$ as the value at which the test is tangent to the power envelope at 50%.

To test the null hypothesis, $H_0 : \phi = 0$, in equation (2.4), when k is unknown (which is the empirically relevant case) the following test statistic is considered,

$$\mathcal{T}_{\hat{k}}^{GLS} := \min_{k \in K} \hat{t}_k^{GLS} = \min_{k \in K} \frac{\sum_{t=2}^{T} \Delta \hat{u}_{\tilde{c},t} \cos(k,t) \hat{u}_{\tilde{c},t-1}}{\left[\hat{\sigma}_k^2 \sum_{t=2}^{T} \cos^2(k,t) \hat{u}_{\tilde{c},t-1}^2\right]^{1/2}}, \tag{2.5}$$

where $K = \{0.5, 1, 1.5, 2, 2.5, 3\}$ and $\hat{\sigma}_k^2$ is the least-squares estimate of $E(\eta_t^2)$ obtained from (2.4) under a fixed k. Note that although a time series is not weakly stationary when $k \neq 0$, it does follow an intrinsically mean-reverting process with some exceptional periods during which ρ_t is close to unity.

FIGURE 2.1: The cosine function for non-integer and integer values of k



2.2.2 Unconditional and Conditional Heteroskedasticity

It is important to examine how the proposed tests perform under breaks in the unconditional or the conditional variance of the error process $\{\varepsilon_t\}$.[5]

Simultaneous increases in persistence and in the innovation variance (two reinforcing effects that cause an increase in σ_y^2) may impact the finite sample performance of the proposed test given that it may be hard to distinguish whether the increase in the unconditional variance of y_t is caused by a true change in persistence or by an exogenous shift in the innovations' variance. Moreover, a large increase in the unconditional variance may cause the process to be confounded more often with a unit root process.

To accommodate conditional/unconditional heteroskedasticity, heteroskedasticity consistent standard errors, as proposed by Eicker-White (EW), are typically employed; Demetrescu (2008) and Phillips (1987). Thus, the proposed test statistic with EW robust standard errors, considering fixed k and no short-run

[5] Hamori and Tokihisa (1997) and Kim et al. (2002) showed that a permanent variance shift causes size distortions in the ADF tests; and although conditional heteroskedasticity does not affect the asymptotic distribution of ADF type test statistics (Phillips; 1987), the presence of ARCH effects does cause size distortions in finite samples; Kim and Schmidt (1993).

dependence in ε_t is,

$$\hat{t}^{GLS}_{k,EW} := \frac{\sum_{t=2}^{T} \Delta\hat{u}_{\tilde{c},t}\cos(k,t)\hat{u}_{\tilde{c},t-1}}{\left(\sum_{t=2}^{T} \cos^2(k,\mathrm{t})\hat{u}^2_{\tilde{c},t-1}\hat{\eta}^2_t\right)^{1/2}}. \tag{2.6}$$

Proposition 1 *Considering data generated from (2.1) and (2.2), and given a fixed k, it follows as $T \to \infty$ that,*

i) under the null hypothesis, $H_0 : \phi = 0$, and considering Assumptions 1 and 2 in Demetrescu (2008), $\hat{t}^{GLS}_{k,EW} - \hat{t}^{GLS}_{k} \xrightarrow{p} 0$;

ii) under the alternative hypothesis, $H_1 : \phi = \frac{c}{T}$, for any fixed non-positive c and $\tilde{c}$, $\hat{t}^{GLS}_{k,EW} - \hat{t}^{GLS}_{k} \xrightarrow{p} 0$;

where " $\xrightarrow{p}$ " represents convergence in probability.

An alternative to the EW approach used in (2.6) also widely applied in the literature to deal with, among other things, (unconditional and conditional) heteroskedasticity of unknown form is the Wild bootstrap; Gonçalves and Kilian (2004) and Cavaliere and Taylor (2008). The approach consists of using the residuals $\hat{\varepsilon}_t$ computed from (2.4) and generating a new unit root process as $\hat{u}^b_t = \hat{u}^b_{t-1} + v^b_t$, where $v^b_t := e_t\hat{\varepsilon}_t$ and e_t is such that any heteroskedasticity in $\hat{\varepsilon}_t$ is preserved in the newly created residuals v^b_t. We consider e_t~i.i.d. $N(0,1)$, but the Rademacher distribution is also frequently used. Next, B bootstrap series of $\hat{u}^b_t$ are generated and in each iteration the bootstrap statistic $\mathcal{T}^{GLS^*}_k$ is computed based on the auxiliary regression

$$\Delta\hat{u}^b_t = \phi^b\cos(k,t)\hat{u}^b_{t-1} + \sum_{j=1}^{p} \delta_j \Delta\hat{u}^b_{t-j} + \eta_t \tag{2.7}$$

where η_t is an error term. The bootstrap p-value is computed as

$$P_b\left(\mathcal{T}^{GLS}_k\right) := \frac{1}{B}\sum_{n=1}^{B} \mathcal{I}\left(\mathcal{T}^{GLS^*}_k > \mathcal{T}^{GLS}_k\right),$$

where B is the number of bootstrap iterations and $\mathcal{I}(.)$ is the indicator function (for more details see e.g. Davidson and Flachaire; 2008).

2.2.3 Asymptotic Distributions

In this section the asymptotic distributions of the proposed tests are derived under the null hypothesis of a unit root and the alternative of local breaks in persistence. Moreover, the test statistics employed in the construction of the asymptotic local power envelope and their asymptotic distributions are also presented.

Theorem 2.2.1 *Under the null hypothesis of a unit root, $H_0 : \phi = 0$ $(c = 0)$, as $T \to \infty$ the limit distribution of the proposed test statistic in (2.5) is,*

$$\mathcal{T}_{\hat{k}}^{GLS_{\mathfrak{v}}} \Rightarrow \min_{k \in K} \left(\frac{\cos(k,1)W_{\mathfrak{v}}^2(1) + \frac{1}{2}(2\pi k)^2 \int_0^1 \cos(2\pi kr)W_{\mathfrak{v}}^2(r)dr - 1}{2\left(\int_0^1 \cos^2(k,r)W_{\mathfrak{v}}^2(r)dr\right)^{1/2}} \right), \mathfrak{v} = \mu, \tau \tag{2.8}$$

where $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ and $\mathcal{T}_{\hat{k}}^{GLS_\tau}$ are computed from local GLS demeaned and local GLS de-trended data, respectively, k is fixed, $W_\mu(r) = W(r)$ is a standard Brownian motion, $\cos(k,r) := \cos^2(\pi kr)$ with $r \in [0,1]$ and

$$\begin{aligned} W_\tau(r) \;\; := \;\; & \sigma W(r) - \sigma r \frac{(1 - \tilde{c}\cos(k,r))W(1) + \tilde{c}^2 \int_0^1 r\cos^2(k,r)W(r)dr}{\int_0^1 \left[1 - 2\tilde{c}r\cos^2(k,r) + r^2\tilde{c}^2\cos(k,r)\right] dr} + \\ & + \sigma r \frac{\tilde{c}k\pi \int_0^1 r\sin(2\pi kr)W(r)dr}{\int_0^1 \left[1 - 2\tilde{c}r\cos^2(k,r) + r^2\tilde{c}^2\cos(k,r)\right] dr}. \end{aligned} \tag{2.9}$$

As in the traditional unit root testing context, local GLS demeaning has no effect on the proposed test's asymptotic distribution (see Appendix for details) and therefore, the asymptotic distribution of $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ is equivalent to that of a test statistic computed from a test regression with no deterministics.

Theorem 2.2.2 *Under the local alternative hypothesis, $H_A : \phi = \frac{c}{T} < 0$, the limit distribution of the proposed statistic is*

$$\mathcal{T}_{\hat{k}}^{GLS_{\mathfrak{v}}} \Rightarrow \min_{k \in K} \left(\frac{\cos(k,1)J_{\mathfrak{v},c}^2(1) + \frac{1}{2}(2\pi k)^2 \int_0^1 \cos(2\pi kr)J_{\mathfrak{v},c}^2(r)dr - 1}{2\left(\int_0^1 \cos^2(k,r)J_{\mathfrak{v},c}^2(r)dr\right)^{1/2}} \right), \mathfrak{v} = \mu, \tau \tag{2.10}$$

where $k \in K$, $r \in (0,1)$, $J_{\mu,c}(r) = J_c(r)$ is a standard OU process and $J_{\tau,c}(r)$ is a local GLS de-trended Ornstein-Uhlenbeck (OU) process.

In order to construct the asymptotic power envelope, an asymptotically equivalent test to the infeasible most powerful invariant likelihood ratio statistic proposed by Elliott et al. (1996) will be used. Specifically, for a given $\tilde{c}$ and k it has the form:

$$P_{\tilde{c}} := \frac{\sum_{t=1}^T \hat{\varepsilon}_{\tilde{c},t}^2 - \tilde{\rho}_t \sum_{t=1}^T \hat{\varepsilon}_{0,t}^2}{\hat{\sigma}^2} \tag{2.11}$$

where $\hat{\sigma}^2 := T^{-1} \sum_{t=1}^T \hat{\varepsilon}_{\bar{c},t}^2$, $\hat{\varepsilon}_{0,t}$ and $\hat{\varepsilon}_{\tilde{c},t}$ are the residuals of the model defined by (2.1) and (2.2) considering $\tilde{c} = 0$ and $\tilde{c} < 0$, respectively, and $\tilde{\rho}_t := 1 + \frac{\tilde{c}}{T}\cos(k,t)$.

Theorem 2.2.3 *Under $H_0 : \phi = 0$ $(c = 0)$ and i.i.d. innovations the asymptotic distribution of the test statistic presented in (2.11) is*

$$P_{\tilde{c}}^\mu \Rightarrow \tilde{c}^2 \int_0^1 \cos^2(r,k)J_c^2(r) - \tilde{c}\cos(T,k)J_c^2(1)$$

and

$$P_{\tilde{c}}^{\tau} \Rightarrow \tilde{c}^2 \int_0^1 \cos^2(r,k) J_{\tau,c}^2(r) + (1 - \tilde{c}\cos(T,k)) J_{\tau,c}^2(1),$$

where $P_{\tilde{c}}^{\mu}$ and $P_{\tilde{c}}^{\tau}$ are demeaned and de-trended test statistics, respectively, $J_c(r)$ is a standard Ornstein-Uhlenbeck (OU) process and $J_{\tau,c}(r)$ is a local GLS de-trended OU process.

2.3 Monte Carlo Analysis

This section investigates the finite sample properties of the tests previously introduced under the null and alternative hypotheses. All simulations are performed in Gauss 10.

Table 2.1 presents values for $\tilde{c}$, given a specific k, for which the power of the test is tangent to the power envelope at 50%, and Table 2.2 provides the necessary critical values for $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ and $\mathcal{T}_{\hat{k}}^{GLS_\tau}$ considering $T \in \{150, 250, 500, 1000\}$ and $k \in \{0.5, 1, 1.5, 2, 2.5, 3\}$.

TABLE 2.1: Values for the non-centrality parameter $\tilde{c}$ for different values of k

k	0	0.5	1	1.5	2	2.5	3
$\mathbf{x}_t = 1$	-7.0	-15.6	-11.8	-12.7	-10.7	-11.2	-10.2
$\mathbf{x}_t = [1, t]'$	-13.5	-25.4	-25.8	-26.1	-22.2	-23.3	-20.2

Notes: Values computed based on 100,000 replications and T=1000.

TABLE 2.2: Critical values

	$\mathcal{T}_{\hat{k}}^{GLS_\mu}$			$\mathcal{T}_{\hat{k}}^{GLS_\tau}$		
T	1%	5%	10%	1%	5%	10%
150	-3.266	-2.695	-2.403	-4.092	-3.589	-3.336
250	-3.192	-2.629	-2.346	-4.008	-3.517	-3.268
500	-3.152	-2.592	-2.303	-3.958	-3.467	-3.215
1000	-3.133	-2.574	-2.285	-3.935	-3.438	-3.189

Notes: The reported critical values are based on 100,000 simulations.

In what follows, the finite sample performance of the proposed tests will be compared to that of DF^{GLS_ς}, with $\varsigma = \mu, \tau$. We investigate how the tests perform under iid innovations, in the presence of autocorrelation and under conditional and unconditional heteroskedasticity. For the Wild bootstrap approach 1000 Monte Carlo and bootstrap replications were used, whereas for all the other simulations 10,000 Monte Carlo replications were considered.

2.3.1 IID Innovations

Two data generation processes (DGPs) are considered: the first (henceforth DGP1) is,

$$y_t = \rho_t y_{t-1} + \varepsilon_t \tag{2.12}$$

with $\rho_t = 1+\phi\cos(k,t)$, $\cos(k,t) := (1+\cos(2\pi kt/T))/2$ and $\phi \in \{-0.1, -0.2, 0\}$; and the second (DGP2) is,

$$y_t = \begin{cases} \rho_1 y_{t-1} + \varepsilon_t & for \quad t = 1, ..., \lfloor \tau_1 T \rfloor \\ \rho_2 y_{t-1} + \varepsilon_t & for \quad t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor \\ \rho_3 y_{t-1} + \varepsilon_t & for \quad t = \lfloor \tau_2 T \rfloor + 1, ..., T, \end{cases} \tag{2.13}$$

where $\tau_1 \leq \tau_2$, $\tau_1 \in \{0.3, 0.4, 0.6, 0.8\}$, $\tau_2 \in \{0.3, 0.6, 0.7, 0.8\}$, $\rho_1 \in \{0.8,\ 0.9\}$, $\rho_2 \in \{0.99,\ 1\}$ and $\rho_3 \in \{0.8, 0.9, 0.99,\ 1\}$ are used to investigate the finite sample power properties, and $\rho_1 = \rho_2 = \rho_3 = 1$ ($\phi = 0$) for the finite sample size of the tests. In both cases, $\varepsilon_t \sim N(0,1)$ and $y_1 = \varepsilon_1 \sim N(0,1)$ is used.

DGP1 implies that the transition from regimes with $\phi < 0$ ($\rho_t < 1$) to regimes with $\phi = 0$ ($\rho_t = 1$) is smooth, since ρ_t is a function of ϕ and the time-varying weights defined by $\cos(k,t)$ in (2.3). This cosine function is approximately 1 for the first values of t, so that ρ_t is smaller at the beginning of the sample. Table 2.3 presents the rejection rates of $\mathcal{T}_{\hat{k}}^{GLS_\mu}$. The results show that the empirical size is close to the nominal 5% significance level for all cases considered. In contrast to DF^{GLS_μ} and when there are breaks in persistence, $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ provides power gains for almost all of the simulation parameters used (the only exception is when $k = 0.5$). As expected, the proposed test displays more difficulties in rejecting the null hypothesis when $\phi = -0.1$, since ρ_t is already large before persistence increases. However, also in this case, there are significant power gains which further increase with the sample size (the empirical power of $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ is close to 100% for $T = 500$).

For $\phi = -0.1$, the de-trended version of the proposed test, $\mathcal{T}_{\hat{k}}^{GLS_\tau}$, only presents power gains for all values of k when $T = 500$. Nonetheless, there are some relevant positive differences relatively to the DF^{GLS_τ} for $k > 0.5$ even when $T = 250$. For $\phi = -0.2$, relevant power gains are observed even when $T = 150$ and $k > 0.5$ (see Table 2.4).

In DGP2, to save space, we only allow for a maximum of two abrupt changes which result in a single period of higher persistence. When $\rho_2 = \rho_3$, the break divides the process into two regimes, with a larger ρ_t in the last sub-period. If $\rho_1 = \rho_3$ the sub-period of higher persistence occurs in the middle of the sample and ρ_t returns to the value assumed at the beginning of the process. The results in Table 2.5 show, as expected, that the power of $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ and DF^{GLS_μ} is lower when $\rho_2 = 1$ in a larger percentage of the sample. For instance, if a persistence change occurs at $\tau_1 = \tau_2 = 0.6$ the time series behaves as a random walk over the last 40% of the sample. When the sample size is moderate ($T = 250$) and a change occurs, $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ displays significant power gains even when $\rho_1 = 0.9$ and $\rho_2 = \rho_3 = 1$. When $\rho_2 = \rho_3$, $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ only provides power gains relative to DF^{GLS_μ} when $T \geq 250$. The differences between the rejection rates of these two tests are maximized when the process exhibits higher persistence for a relevant portion of the sample ($\tau_1 = \tau_2 = 0.6$). The same holds when $\rho_1 = \rho_3$, where power gains are observed when $T = 150$ and $\tau_1 = 0.3$ and $\tau_2 = 0.7$.

When the statistics are de-trended, and the changes in persistence are abrupt and $\rho_2 = \rho_3$, the advantages of using $\mathcal{T}_{\hat{k}}^{GLS_\tau}$ only become clear when the sample size is large. On the other hand, when $\rho_1 = \rho_3$, there are positive differences relative to the DF^{GLS_τ} for all cases (see Table 2.6).

Note that the trigonometric function used considers that persistence is always lower at the beginning of the sample. The symmetric cases can be investigated using the time series in reverse chronological order. This transformation alters the asymptotic distributions of the statistics (see (2.17)). Since the critical values are very close to those obtained using the normal chronological order, reversing the time series when the process starts with a period of higher persistence leads to results similar to those discussed in this section and are therefore omitted.

TABLE 2.3: Empirical size and power with *iid* errors

DGP: $y_t = y_{t-1} + \phi((1+\cos(2\pi kt/T))/2)y_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim N(0,1)$

	DF^{GLS_μ}			$\mathcal{T}_{\hat{k}}^{GLS_\mu}$			$DF^{GLS_\mu^*}$			$\mathcal{T}_{\hat{k}}^{GLS_\mu^*}$		
						$T = 150$						
k/ϕ	0	-0.1	-0.2	0	-0.1	-0.2	0	-0.1	-0.2	0	-0.1	-0.2
0	0.052	0.925	0.999	0.051	0.707	0.962	0.053	0.918	0.996	0.048	0.679	0.952
0.5	0.052	0.317	0.523	0.051	0.286	0.654	0.053	0.330	0.547	0.048	0.268	0.640
1.0	0.052	0.356	0.601	0.051	0.375	0.769	0.053	0.349	0.619	0.048	0.366	0.748
1.5	0.052	0.336	0.574	0.051	0.354	0.734	0.053	0.336	0.574	0.048	0.337	0.725
2.0	0.052	0.377	0.658	0.051	0.428	0.811	0.053	0.377	0.659	0.048	0.419	0.803
2.5	0.052	0.361	0.635	0.051	0.420	0.796	0.053	0.365	0.624	0.048	0.400	0.762
3.0	0.052	0.405	0.725	0.051	0.461	0.843	0.053	0.423	0.726	0.048	0.465	0.820
						$T = 250$						
0	0.052	0.996	1.000	0.052	0.942	0.996	0.040	0.935	0.998	0.040	0.935	0.998
0.5	0.052	0.484	0.674	0.052	0.557	0.942	0.040	0.477	0.663	0.040	0.532	0.930
1.0	0.052	0.555	0.765	0.052	0.689	0.967	0.040	0.549	0.762	0.040	0.685	0.961
1.5	0.052	0.524	0.753	0.052	0.657	0.958	0.040	0.526	0.754	0.040	0.628	0.958
2.0	0.052	0.601	0.854	0.052	0.740	0.974	0.040	0.578	0.851	0.040	0.720	0.971
2.5	0.052	0.588	0.839	0.052	0.717	0.967	0.040	0.577	0.832	0.040	0.703	0.955
3.0	0.052	0.666	0.918	0.052	0.774	0.976	0.040	0.661	0.919	0.040	0.752	0.970
						$T = 500$						
0	0.050	1.000	1.000	0.052	0.999	1.000	0.041	1.000	1.000	0.041	1.000	1.000
0.5	0.050	0.683	0.818	0.052	0.961	0.999	0.041	0.684	0.818	0.041	0.950	0.998
1.0	0.050	0.774	0.911	0.052	0.982	1.000	0.041	0.758	0.909	0.041	0.978	1.000
1.5	0.050	0.761	0.913	0.052	0.977	1.000	0.041	0.778	0.923	0.041	0.966	1.000
2.0	0.050	0.859	0.976	0.052	0.988	1.000	0.041	0.852	0.968	0.041	0.984	0.999
2.5	0.050	0.848	0.968	0.052	0.983	1.000	0.041	0.844	0.972	0.041	0.972	1.000
3.0	0.050	0.922	0.995	0.052	0.989	1.000	0.041	0.915	0.992	0.041	0.985	0.999

Notes: $DF^{GLS_\mu^*}$ and $\mathcal{T}_{\hat{k}}^{GLS_\mu^*}$ are Wild bootstrap based test statistics.

Table 2.4: Empirical size and power with *iid* errors - linear trend case

DGP: $y_t = y_{t-1} + \phi((1 + \cos(2\pi kt/T))/2)y_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim N(0,1)$

	DF^{GLS_τ}			$\mathcal{T}_{\hat{k}}^{GLS_\tau}$			$DF^{GLS_\tau^*}$			$\mathcal{T}_{\hat{k}}^{GLS_\tau^*}$		
						$T = 150$						
k/ϕ	0	−0.1	−0.2	0	−0.1	−0.2	0	−0.1	−0.2	0	−0.1	−0.2
0	0.050	0.594	0.991	0.048	0.364	0.859	0.053	0.584	0.988	0.046	0.340	0.839
0.5	0.050	0.184	0.400	0.048	0.153	0.373	0.053	0.195	0.415	0.046	0.146	0.351
1.0	0.050	0.119	0.280	0.048	0.134	0.396	0.053	0.120	0.274	0.046	0.110	0.348
1.5	0.050	0.126	0.298	0.048	0.130	0.376	0.053	0.132	0.297	0.046	0.125	0.338
2.0	0.050	0.134	0.314	0.048	0.147	0.439	0.053	0.133	0.307	0.046	0.142	0.406
2.5	0.050	0.130	0.307	0.048	0.147	0.434	0.053	0.129	0.324	0.046	0.130	0.415
3.0	0.050	0.146	0.348	0.048	0.171	0.481	0.053	0.149	0.343	0.046	0.155	0.446
						$T = 250$						
0	0.048	0.960	1.000	0.051	0.725	0.994	0.048	0.969	1.000	0.056	0.706	0.994
0.5	0.048	0.333	0.604	0.051	0.290	0.759	0.048	0.356	0.629	0.056	0.309	0.737
1.0	0.048	0.230	0.478	0.051	0.298	0.819	0.048	0.223	0.456	0.056	0.300	0.787
1.5	0.048	0.246	0.504	0.051	0.280	0.797	0.048	0.259	0.517	0.056	0.264	0.789
2.0	0.048	0.253	0.537	0.051	0.336	0.850	0.048	0.264	0.544	0.056	0.325	0.824
2.5	0.048	0.244	0.542	0.051	0.325	0.833	0.048	0.243	0.527	0.056	0.309	0.818
3.0	0.048	0.282	0.598	0.051	0.369	0.872	0.048	0.275	0.602	0.056	0.361	0.844
						$T = 500$						
0	0.050	1.000	1.000	0.048	0.995	1.000	0.049	1.000	1.000	0.046	0.987	1.000
0.5	0.050	0.607	0.806	0.048	0.764	0.999	0.049	0.612	0.808	0.046	0.749	0.997
1.0	0.050	0.486	0.728	0.048	0.827	0.999	0.049	0.490	0.727	0.046	0.807	1.000
1.5	0.050	0.509	0.758	0.048	0.806	0.999	0.049	0.517	0.767	0.046	0.782	0.998
2.0	0.050	0.535	0.811	0.048	0.859	1.000	0.049	0.540	0.793	0.046	0.851	1.000
2.5	0.050	0.540	0.828	0.048	0.847	1.000	0.049	0.550	0.817	0.046	0.837	1.000
3.0	0.050	0.595	0.884	0.048	0.879	1.000	0.049	0.597	0.882	0.046	0.871	0.999

Notes: $DF^{GLS_\tau^*}$ and $\mathcal{T}_{\hat{k}}^{GLS_\tau^*}$ are Wild bootstrap based test statistics.

TABLE 2.5: Empirical power when the breaks in persistence are abrupt

$$\text{DGP: } y_t = \begin{cases} \rho_1 y_{t-1} + \varepsilon_t & for \quad t = 1, ..., \lfloor \tau_1 T \rfloor \\ \rho_2 y_{t-1} + \varepsilon_t & for \quad t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor \\ \rho_3 y_{t-1} + \varepsilon_t & for \quad t = \lfloor \tau_2 T \rfloor + 1, ..., T, \end{cases} \qquad \varepsilon_t \sim N(0,1)$$

	DF^{GLS_μ}		$\mathcal{T}_{\hat{k}}^{GLS_\mu}$		DF^{GLS_μ}		$\mathcal{T}_{\hat{k}}^{GLS_\mu}$		DF^{GLS_μ}		$\mathcal{T}_{\hat{k}}^{GLS_\mu}$	
	CV	WB	CV	WB	CV	WB	CV	WB	CV	WB	CV	WB
						$T = 150$						
(ρ_1, ρ_2, ρ_3)	$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.3)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.6, 0.6)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.8, 0.8)$			
$(0.8, 0.99, 0.99)$	0.233	0.252	0.287	0.287	0.515	0.541	0.724	0.714	0.783	0.792	0.908	0.889
$(0.8, 1, 1)$	0.148	0.165	0.205	0.211	0.408	0.418	0.661	0.655	0.718	0.723	0.897	0.876
$(0.9, 0.99, 0.99)$	0.194	0.207	0.188	0.191	0.398	0.419	0.373	0.384	0.629	0.650	0.553	0.526
$(0.9, 1, 1)$	0.124	0.135	0.133	0.133	0.314	0.308	0.315	0.317	0.561	0.576	0.525	0.502
(ρ_1, ρ_2, ρ_3)	$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.6)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.4, 0.7)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.7)$			
$(0.8, 0.99, 0.8)$	0.752	0.768	0.880	0.866	0.743	0.746	0.920	0.896	0.607	0.623	0.865	0.838
$(0.8, 1, 0.8)$	0.667	0.676	0.876	0.863	0.658	0.675	0.924	0.905	0.501	0.509	0.873	0.850
$(0.9, 0.99, 0.9)$	0.569	0.578	0.538	0.520	0.568	0.573	0.574	0.544	0.454	0.457	0.500	0.473
$(0.9, 1, 0.9)$	0.485	0.487	0.534	0.517	0.491	0.498	0.581	0.561	0.362	0.360	0.513	0.490
						$T = 250$						
(ρ_1, ρ_2, ρ_3)	$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.3)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.6, 0.6)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.8, 0.8)$			
$(0.8, 0.99, 0.99)$	0.342	0.345	0.475	0.452	0.640	0.628	0.940	0.933	0.863	0.873	0.992	0.992
$(0.8, 1, 1)$	0.170	0.157	0.286	0.286	0.449	0.448	0.878	0.856	0.763	0.777	0.989	0.989
$(0.9, 0.99, 0.99)$	0.305	0.316	0.332	0.327	0.569	0.557	0.686	0.644	0.798	0.809	0.857	0.842
$(0.9, 1, 1)$	0.149	0.125	0.189	0.180	0.392	0.402	0.573	0.531	0.687	0.698	0.834	0.815
(ρ_1, ρ_2, ρ_3)	$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.6)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.4, 0.7)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.7)$			
$(0.8, 0.99, 0.8)$	0.856	0.853	0.982	0.981	0.857	0.855	0.990	0.990	0.745	0.732	0.981	0.977
$(0.8, 1, 0.8)$	0.733	0.729	0.981	0.980	0.737	0.730	0.991	0.987	0.570	0.577	0.983	0.979
$(0.9, 0.99, 0.9)$	0.773	0.769	0.831	0.813	0.775	0.777	0.868	0.852	0.659	0.655	0.804	0.793
$(0.9, 1, 0.9)$	0.635	0.638	0.822	0.810	0.633	0.631	0.879	0.864	0.481	0.475	0.820	0.804
						$T = 500$						
(ρ_1, ρ_2, ρ_3)	$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.3)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.6, 0.6)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.8, 0.8)$			
$(0.8, 0.99, 0.99)$	0.578	0.569	0.752	0.753	0.824	0.816	0.998	0.998	0.948	0.934	1.000	1.000
$(0.8, 1, 1)$	0.177	0.161	0.369	0.355	0.476	0.472	0.959	0.957	0.801	0.800	1.000	1.000
$(0.9, 0.99, 0.99)$	0.550	0.549	0.644	0.657	0.793	0.789	0.978	0.973	0.928	0.921	0.997	0.997
$(0.9, 1, 1)$	0.165	0.147	0.291	0.269	0.440	0.440	0.881	0.877	0.761	0.756	0.995	0.993
(ρ_1, ρ_2, ρ_3)	$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.6)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.4, 0.7)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.7)$			
$(0.8, 0.99, 0.8)$	0.961	0.965	1.000	1.000	0.962	0.955	1.000	1.000	0.910	0.909	1.000	1.000
$(0.8, 1, 0.8)$	0.784	0.770	1.000	1.000	0.792	0.785	1.000	1.000	0.615	0.615	1.000	1.000
$(0.9, 0.99, 0.9)$	0.941	0.945	0.995	0.994	0.943	0.935	0.997	0.996	0.878	0.866	0.992	0.990
$(0.9, 1, 0.9)$	0.727	0.723	0.992	0.988	0.733	0.730	0.998	0.997	0.560	0.554	0.993	0.993

Table 2.6: Empirical power when the breaks in persistence are abrupt - linear trend case

$$\text{DGP: } y_t = \begin{cases} \rho_1 y_{t-1} + \varepsilon_t & for \quad t = 1, ..., \lfloor \tau_1 T \rfloor \\ \rho_2 y_{t-1} + \varepsilon_t & for \quad t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor \\ \rho_3 y_{t-1} + \varepsilon_t & for \quad t = \lfloor \tau_2 T \rfloor + 1, ..., T, \end{cases} \qquad \varepsilon_t \sim N(0,1)$$

	DF^{GLS_τ}		$\mathcal{T}_{\hat{k}}^{GLS_\tau}$		DF^{GLS_τ}		$\mathcal{T}_{\hat{k}}^{GLS_\tau}$		DF^{GLS_τ}		$\mathcal{T}_{\hat{k}}^{GLS_\tau}$	
	CV	WB	CV	WB	CV	WB	CV	WB	CV	WB	CV	WB
						$T = 150$						
(ρ_1, ρ_2, ρ_3)	$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.3)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.6, 0.6)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.8, 0.8)$			
$(0.8, 0.99, 0.99)$	0.185	0.206	0.211	0.199	0.442	0.460	0.440	0.429	0.712	0.728	0.661	0.658
$(0.8, 1, 1)$	0.143	0.156	0.168	0.154	0.360	0.369	0.386	0.372	0.646	0.660	0.637	0.620
$(0.9, 0.99, 0.99)$	0.129	0.147	0.124	0.100	0.244	0.253	0.186	0.176	0.375	0.381	0.259	0.234
$(0.9, 1, 1)$	0.105	0.119	0.103	0.091	0.198	0.203	0.157	0.155	0.337	0.346	0.241	0.215
(ρ_1, ρ_2, ρ_3)	$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.6)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.4, 0.7)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.7)$			
$(0.8, 0.99, 0.8)$	0.449	0.450	0.615	0.584	0.461	0.475	0.657	0.604	0.304	0.320	0.531	0.494
$(0.8, 1, 0.8)$	0.376	0.380	0.610	0.575	0.393	0.390	0.651	0.598	0.241	0.248	0.530	0.468
$(0.9, 0.99, 0.9)$	0.228	0.222	0.232	0.208	0.240	0.230	0.230	0.196	0.166	0.175	0.181	0.153
$(0.9, 1, 0.9)$	0.187	0.184	0.231	0.208	0.198	0.185	0.214	0.181	0.125	0.131	0.172	0.135
						$T = 250$						
(ρ_1, ρ_2, ρ_3)	$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.3)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.6, 0.6)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.8, 0.8)$			
$(0.8, 0.99, 0.99)$	0.256	0.236	0.383	0.372	0.596	0.614	0.810	0.791	0.846	0.851	0.956	0.952
$(0.8, 1, 1)$	0.174	0.164	0.274	0.257	0.445	0.457	0.713	0.696	0.749	0.760	0.946	0.942
$(0.9, 0.99, 0.99)$	0.195	0.171	0.208	0.215	0.436	0.449	0.381	0.391	0.673	0.694	0.545	0.535
$(0.9, 1, 1)$	0.134	0.132	0.151	0.148	0.313	0.340	0.300	0.321	0.576	0.597	0.506	0.500
(ρ_1, ρ_2, ρ_3)	$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.6)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.4, 0.7)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.7)$			
$(0.8, 0.99, 0.8)$	0.617	0.611	0.929	0.906	0.632	0.621	0.964	0.951	0.456	0.437	0.902	0.868
$(0.8, 1, 0.8)$	0.483	0.463	0.920	0.900	0.499	0.496	0.964	0.943	0.317	0.297	0.897	0.857
$(0.9, 0.99, 0.9)$	0.433	0.416	0.487	0.480	0.452	0.451	0.516	0.495	0.315	0.287	0.414	0.410
$(0.9, 1, 0.9)$	0.326	0.311	0.480	0.463	0.345	0.326	0.506	0.492	0.211	0.203	0.410	0.393
						$T = 500$						
(ρ_1, ρ_2, ρ_3)	$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.3)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.6, 0.6)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.8, 0.8)$			
$(0.8, 0.99, 0.99)$	0.412	0.396	0.680	0.678	0.799	0.791	0.991	0.989	0.951	0.938	1.000	1.000
$(0.8, 1, 1)$	0.200	0.192	0.413	0.404	0.509	0.517	0.920	0.912	0.812	0.815	1.000	1.000
$(0.9, 0.99, 0.99)$	0.360	0.339	0.474	0.466	0.723	0.720	0.863	0.855	0.910	0.904	0.968	0.958
$(0.9, 1, 1)$	0.169	0.164	0.267	0.256	0.434	0.440	0.697	0.675	0.749	0.750	0.951	0.940
(ρ_1, ρ_2, ρ_3)	$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.6)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.4, 0.7)$				$(\bar{\tau}_1, \bar{\tau}_2) = (0.3, 0.7)$			
$(0.8, 0.99, 0.8)$	0.819	0.813	0.999	0.999	0.841	0.833	1.000	1.000	0.699	0.685	0.999	0.999
$(0.8, 1, 0.8)$	0.563	0.580	0.998	0.997	0.582	0.572	1.000	1.000	0.386	0.381	0.997	0.992
$(0.9, 0.99, 0.9)$	0.740	0.732	0.941	0.937	0.757	0.745	0.963	0.950	0.607	0.608	0.913	0.904
$(0.9, 1, 0.9)$	0.479	0.489	0.924	0.917	0.497	0.494	0.966	0.945	0.313	0.309	0.901	0.878

2.3.2 Serially Correlated Errors

To investigate the finite sample properties of the tests in the presence of autocorrelation it is assumed that the error process of (2.12) is an ARMA, such as,

$$\varepsilon_t = \delta\varepsilon_{t-1} + \theta e_{t-1} + e_t, \qquad e_t \sim N(0,1), \tag{2.14}$$

with $y_1 = \varepsilon_1 \sim N(0,1)$, $\delta \in \{0, 0.3, 0.6\}$ and $\theta \in \{-0.8, -0.4, 0, 0.4, 0.8\}$. The lags of the augmented test regression are chosen using the MAIC information criteria proposed by Ng and Perron (2001).

Table 2.7 summarizes the finite sample results for the demeaned tests. Although we also performed simulations for $T = 150$ and $T = 500$, for the sake of space only the results for $T = 250$ are reported as the conclusions are qualitatively the same. The $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ test displays good finite sample size performance. Its empirical size only exceeds the nominal 5% significance level when $\theta < 0$, but even in these cases the results are close to or slightly lower than those of the DF^{GLS_μ}.

To evaluate the power performance we considered $\phi = -0.1$ and that the persistence change is smooth and approximated by cosine functions. The sign of θ also affects the power properties of the proposed test. When $\theta \geq 0$ and $k > 0.5$ there are relevant power gains even when $T = 150$. For larger samples, the power differences relative to DF^{GLS_μ} are more pronounced and occur for all values of k. All tests have more difficulties in rejecting the null hypothesis when $\theta = -0.8$ and $\delta = 0$.

Simulation results for the de-trended statistics are presented in Table 2.8. Similarly to the demeaned case, also the empirical size of $\mathcal{T}_{\hat{k}}^{GLS_\tau}$ exceeds the nominal 5% level less than the DF^{GLS_τ} when $\theta < 0$, and the two tests are more undersized when $\theta > 0$. This problem is attenuated as the sample size increases. The results show that the power of the tests considered is low when a linear trend is used and ρ_t is large before the increase in persistence. For $T = 500$ the results improve and the superiority of the proposed test becomes evident especially when $\theta \geq 0$.

Table 2.7: Finite sample size and power with ε_t autocorrelated

DGP:		$y_t = y_{t-1} + \phi((1+\cos(2\pi kt/T))/2)y_{t-1} + \varepsilon_t$ $\varepsilon_t = \delta\varepsilon_{t-1} + \theta e_{t-1} + e_t,\ e_t \sim N(0,1)$					
$T = 250$			DF^{GLS_μ}			$\mathcal{T}_{\hat{k}}^{GLS_\mu}$	
(ϕ, k)	θ / δ	0	0.3	0.6	0	0.3	0.6
(0, 0)	−0.8	0.078	0.091	0.104	0.063	0.074	0.091
	−0.4	0.052	0.053	0.029	0.055	0.056	0.034
	0	0.044	0.044	0.043	0.045	0.046	0.044
	0.4	0.043	0.040	0.039	0.046	0.043	0.044
	0.8	0.039	0.039	0.040	0.045	0.044	0.046
(−0.1, 0.5)	−0.8	0.235	0.317	0.426	0.164	0.266	0.433
	−0.4	0.361	0.395	0.289	0.378	0.452	0.282
	0	0.392	0.375	0.351	0.439	0.426	0.411
	0.4	0.356	0.335	0.312	0.403	0.379	0.343
	0.8	0.289	0.283	0.262	0.331	0.322	0.301
(−0.1, 1.0)	−0.8	0.290	0.389	0.512	0.226	0.350	0.550
	−0.4	0.434	0.480	0.366	0.518	0.604	0.496
	0	0.473	0.463	0.445	0.605	0.603	0.606
	0.4	0.437	0.420	0.400	0.588	0.576	0.553
	0.8	0.362	0.359	0.342	0.529	0.524	0.509
(−0.1, 1.5)	−0.8	0.275	0.364	0.492	0.204	0.330	0.523
	−0.4	0.401	0.441	0.316	0.483	0.570	0.436
	0	0.436	0.418	0.393	0.562	0.561	0.545
	0.4	0.392	0.373	0.344	0.533	0.517	0.486
	0.8	0.321	0.313	0.293	0.481	0.469	0.449
(−0.1, 2.0)	−0.8	0.323	0.432	0.576	0.249	0.380	0.592
	−0.4	0.487	0.536	0.405	0.582	0.672	0.584
	0	0.518	0.509	0.498	0.668	0.677	0.676
	0.4	0.489	0.474	0.452	0.658	0.648	0.634
	0.8	0.423	0.419	0.410	0.608	0.605	0.591
(−0.1, 2.5)	−0.8	0.308	0.411	0.553	0.243	0.378	0.584
	−0.4	0.466	0.512	0.364	0.557	0.645	0.540
	0	0.493	0.478	0.462	0.641	0.641	0.637
	0.4	0.455	0.439	0.415	0.623	0.610	0.587
	0.8	0.388	0.384	0.369	0.570	0.567	0.549
(−0.1, 3)	−0.8	0.355	0.476	0.642	0.269	0.415	0.624
	−0.4	0.547	0.608	0.464	0.618	0.704	0.635
	0	0.585	0.585	0.589	0.706	0.717	0.714
	0.4	0.560	0.553	0.550	0.696	0.689	0.675
	0.8	0.508	0.512	0.515	0.651	0.648	0.633

TABLE 2.8: Finite sample sizes and power with ε_t autocorrelated - linear trend case

DGP:		$y_t = y_{t-1} + \phi((1+\cos(2\pi kt/T))/2)y_{t-1} + \varepsilon_t$ $\varepsilon_t = \delta\varepsilon_{t-1} + \theta e_{t-1} + e_t,\ e_t \sim N(0,1)$					
$T = 250$			DF^{GLS_τ}			$\mathcal{T}_{\hat{k}}^{GLS_\tau}$	
(ϕ, k)	θ / δ	0	0.3	0.6	0	0.3	0.6
	−0.8	0.066	0.085	0.107	0.029	0.047	0.093
	−0.4	0.046	0.046	0.013	0.050	0.054	0.022
$(0, 0)$	0	0.033	0.032	0.034	0.039	0.040	0.043
	0.4	0.030	0.027	0.026	0.037	0.033	0.033
	0.8	0.022	0.022	0.023	0.035	0.033	0.034
	−0.8	0.169	0.229	0.315	0.050	0.093	0.227
	−0.4	0.218	0.254	0.104	0.170	0.228	0.076
$(-0.1, 0.5)$	0	0.231	0.204	0.186	0.196	0.185	0.176
	0.4	0.188	0.161	0.138	0.162	0.136	0.116
	0.8	0.118	0.109	0.099	0.112	0.107	0.102
	−0.8	0.117	0.162	0.235	0.054	0.106	0.247
	−0.4	0.151	0.172	0.067	0.201	0.254	0.123
$(-0.1, 1.0)$	0	0.155	0.138	0.132	0.227	0.218	0.211
	0.4	0.128	0.112	0.097	0.201	0.179	0.160
	0.8	0.078	0.075	0.069	0.159	0.154	0.145
	−0.8	0.133	0.181	0.245	0.055	0.106	0.240
	−0.4	0.160	0.182	0.071	0.190	0.234	0.101
$(-0.1, 1.5)$	0	0.167	0.146	0.133	0.206	0.195	0.184
	0.4	0.130	0.111	0.092	0.174	0.152	0.138
	0.8	0.079	0.073	0.065	0.137	0.134	0.126
	−0.8	0.148	0.201	0.278	0.068	0.128	0.278
	−0.4	0.175	0.201	0.078	0.242	0.295	0.168
$(-0.1, 2.0)$	0	0.181	0.161	0.151	0.268	0.258	0.248
	0.4	0.144	0.127	0.109	0.241	0.219	0.196
	0.8	0.094	0.088	0.082	0.201	0.192	0.178
	−0.8	0.151	0.203	0.275	0.072	0.137	0.278
	−0.4	0.172	0.194	0.068	0.228	0.287	0.157
$(-0.1, 2.5)$	0	0.17	0.151	0.141	0.252	0.246	0.236
	0.4	0.135	0.117	0.099	0.230	0.211	0.190
	0.8	0.089	0.081	0.070	0.187	0.183	0.171
	−0.8	0.177	0.238	0.327	0.080	0.151	0.313
	−0.4	0.209	0.232	0.088	0.277	0.334	0.207
$(-0.1, 3)$	0	0.202	0.186	0.179	0.305	0.293	0.289
	0.4	0.168	0.146	0.134	0.276	0.257	0.237
	0.8	0.113	0.108	0.105	0.234	0.226	0.213

2.3.3 Heteroskedasticity

Unconditional Heteroskedasticity

To evaluate the impact of unconditional heteroscedasticity on the performance of the tests we consider DGP1 with the following specification for the innovation process,

$$\varepsilon_t \sim \begin{cases} N(0,\sigma_1^2) & for \quad t = 1, ..., \lfloor \bar{\tau}_1 T \rfloor \\ N(0,\sigma_2^2) & for \quad t = \lfloor \bar{\tau}_1 T \rfloor + 1, ..., \lfloor \bar{\tau}_2 T \rfloor \\ N(0,\sigma_3^2) & for \quad t = \lfloor \bar{\tau}_2 T \rfloor + 1, ..., T, \end{cases} \tag{2.15}$$

and $y_1 \sim N(0, \sigma_1^2)$.

It is important to infer how changes in the innovation variance affect the finite sample properties of the proposed test. To save space, only results based on DGP1 with $\phi = -0.1$ are reported. Two cases were considered: in the first, a single break is allowed that either increases or decreases the innovation variance; and in the second, two breaks in variance are allowed, the first causing an increase in the innovation variance and the second a reduction to the value it assumed before the occurrence of any break. The values considered in the simulations allow for large changes in the unconditional variance.

Size results for demeaned data are reported in Table 2.9. The results show that the proposed test reveals in some cases finite sample size distortions. However, the rejection rates are, overall, not too far from those of the DF^{GLS_μ} test. Over-rejections are more severe when the innovation variance faces a large increase (e.g. $\sigma_2^2 = 4$), even if a negative variation of the same magnitude occurs later. The largest empirical size distortions are observed when breaks occur closer to the end of the sample. For decreases in volatility, there are also small over-rejections of the null, especially when these occur at the beginning of the sample ($\bar{\tau}_1 = 0.3$). Table 2.10 presents the rejection rates for the linear trend case. Although the empirical size of the proposed test shows the same patterns reported for the demeaned case the rejection rates under the null are larger.

Table 2.11 presents the results for the tests when the Wild bootstrap is used and $T = 250$. The empirical size of $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ is close to the nominal 5% significance level for all cases investigated. Regarding the power properties, comparing the results in Table 2.11 with those in Table 2.3, where the innovation variance is constant, we observe that the power loss is greater for an increase in volatility, e.g. when $(\sigma_1, \sigma_2, \sigma_3) = (1, 2, 2)$, with $k = 0.5$ or $\bar{\tau}_1 = 0.7$. When a decrease in the unconditional variance is observed, the power loss is greater for $k > 0.5$ and $\bar{\tau}_1 = 0.3$. Finally, the occurrence of two breaks in the unconditional variance causes more severe power losses when $k = 1$ and the magnitude of the changes is larger ($\sigma_2 = 2$).

The superior power performance of $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ relatively to DF^{GLS_μ} also depends on k, $\bar{\tau}_1$ and σ. For instance, when the innovation variance increases, the rejection rates of $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ are greater for $k > 0.5$ and the difference relative to DF^{GLS_μ} reaches its maximum for $\bar{\tau}_1 = 0.3$. When the variance decreases, $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ shows better power properties for all k and the superiority relative to DF^{GLS_μ} is more prominent for $\bar{\tau}_1 = 0.7$. On the other hand, if two breaks in the unconditional variance occur, the positive difference in the percentage of rejection relative to the DF^{GLS_μ} test is greater when $k = 1$ and the change in the innovation variance is smaller ($\sigma_2 = 1.5$).

Table 2.12 presents the results for the linear trend case with $T = 250$. Power losses (when compared to Table 2.4) caused by the occurrence of breaks in the variance are similar to those reported for the demeaned case. Regarding the comparison with the DF^{GLS_τ} test, we see that power gains are small when $\phi = -0.1$. In this case, the superior performance of $\mathcal{T}_{\hat{k}}^{GLS_\tau}$ only becomes clear for larger samples ($T = 500$).

TABLE 2.9: Empirical size under unconditional variance breaks

DGP: $y_t = y_{t-1} + \varepsilon_t$,

$$\varepsilon_t \sim \begin{cases} N(0, \sigma_1^2) & for \quad t = 1, ..., \lfloor \tau_1 T \rfloor \\ N(0, \sigma_2^2) & for \quad t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor \\ N(0, \sigma_3^2) & for \quad t = \lfloor \tau_2 T \rfloor + 1, ..., T, \end{cases}$$

	DF^{GLS_μ}	$\mathcal{T}_{\hat{k}}^{GLS_\mu}$	DF^{GLS_μ}	$\mathcal{T}_{\hat{k}}^{GLS_\mu}$	DF^{GLS_μ}	$\mathcal{T}_{\hat{k}}^{GLS_\mu}$
$T = 150$						
$(\sigma_1, \sigma_2, \sigma_3)$	$(\tau_1 = 0.3, \tau_2 = 0.3)$		$(\tau_1 = 0.5, \tau_2 = 0.5)$		$(\tau_1 = 0.7, \tau_2 = 0.7)$	
$(1, 2, 2)$	0.068	0.077	0.070	0.082	0.069	0.083
$(2, 1, 1)$	0.059	0.071	0.062	0.073	0.057	0.052
$(\sigma_1, \sigma_2, \sigma_3)$	$(\tau_1 = 0.3, \tau_2 = 0.6)$		$(\tau_1 = 0.4, \tau_2 = 0.7)$		$(\tau_1 = 0.3, \tau_2 = 0.7)$	
$(1, 1.5, 1)$	0.058	0.059	0.058	0.062	0.060	0.066
$(1, 2, 1)$	0.068	0.078	0.065	0.079	0.071	0.084
$T = 250$						
$(\sigma_1, \sigma_2, \sigma_3)$	$(\tau_1 = 0.3, \tau_2 = 0.3)$		$(\tau_1 = 0.5, \tau_2 = 0.5)$		$(\tau_1 = 0.7, \tau_2 = 0.7)$	
$(1, 2, 2)$	0.069	0.077	0.069	0.080	0.072	0.088
$(2, 1, 1)$	0.063	0.070	0.061	0.069	0.056	0.058
$(\sigma_1, \sigma_2, \sigma_3)$	$(\tau_1 = 0.3, \tau_2 = 0.6)$		$(\tau_1 = 0.4, \tau_2 = 0.7)$		$(\tau_1 = 0.3, \tau_2 = 0.7)$	
$(1, 1.5, 1)$	0.056	0.059	0.059	0.058	0.060	0.063
$(1, 2, 1)$	0.069	0.072	0.069	0.078	0.069	0.080
$T = 500$						
$(\sigma_1, \sigma_2, \sigma_3)$	$(\tau_1 = 0.3, \tau_2 = 0.3)$		$(\tau_1 = 0.5, \tau_2 = 0.5)$		$(\tau_1 = 0.7, \tau_2 = 0.7)$	
$(1, 2, 2)$	0.065	0.075	0.071	0.083	0.068	0.089
$(2, 1, 1)$	0.058	0.066	0.057	0.068	0.057	0.060
$(\sigma_1, \sigma_2, \sigma_3)$	$(\tau_1 = 0.3, \tau_2 = 0.6)$		$(\tau_1 = 0.4, \tau_2 = 0.7)$		$(\tau_1 = 0.3, \tau_2 = 0.7)$	
$(1, 1.5, 1)$	0.051	0.058	0.058	0.061	0.054	0.059
$(1, 2, 1)$	0.061	0.072	0.070	0.078	0.063	0.073

TABLE 2.10: Empirical size under unconditional variance breaks - linear trend case

DGP: $y_t = y_{t-1} + \varepsilon_t$,

$$\varepsilon_t \sim \begin{cases} N(0,\sigma_1^2) & for \quad t = 1, ..., \lfloor \tau_1 T \rfloor \\ N(0,\sigma_2^2) & for \quad t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor \\ N(0,\sigma_3^2) & for \quad t = \lfloor \tau_2 T \rfloor + 1, ..., T, \end{cases}$$

	DF^{GLS_τ}	$\mathcal{T}_{\hat{k}}^{GLS_\tau}$	DF^{GLS_τ}	$\mathcal{T}_{\hat{k}}^{GLS_\tau}$	DF^{GLS_τ}	$\mathcal{T}_{\hat{k}}^{GLS_\tau}$
			$T = 150$			
$(\sigma_1, \sigma_2, \sigma_3)$	$(\tau_1 = 0.3, \tau_2 = 0.3)$		$(\tau_1 = 0.5, \tau_2 = 0.5)$		$(\tau_1 = 0.7, \tau_2 = 0.7)$	
$(1, 2, 2)$	0.074	0.074	0.077	0.088	0.081	0.107
$(2, 1, 1)$	0.074	0.083	0.076	0.079	0.070	0.066
$(\sigma_1, \sigma_2, \sigma_3)$	$(\tau_1 = 0.3, \tau_2 = 0.6)$		$(\tau_1 = 0.4, \tau_2 = 0.7)$		$(\tau_1 = 0.3, \tau_2 = 0.7)$	
$(1, 1.5, 1)$	0.061	0.063	0.059	0.065	0.058	0.065
$(1, 2, 1)$	0.074	0.096	0.070	0.102	0.069	0.095
			$T = 250$			
$(\sigma_1, \sigma_2, \sigma_3)$	$(\tau_1 = 0.3, \tau_2 = 0.3)$		$(\tau_1 = 0.5, \tau_2 = 0.5)$		$(\tau_1 = 0.7, \tau_2 = 0.7)$	
$(1, 2, 2)$	0.063	0.073	0.065	0.089	0.068	0.108
$(2, 1, 1)$	0.059	0.077	0.058	0.075	0.052	0.063
$(\sigma_1, \sigma_2, \sigma_3)$	$(\tau_1 = 0.3, \tau_2 = 0.6)$		$(\tau_1 = 0.4, \tau_2 = 0.7)$		$(\tau_1 = 0.3, \tau_2 = 0.7)$	
$(1, 1.5, 1)$	0.058	0.063	0.054	0.062	0.058	0.062
$(1, 2, 1)$	0.069	0.091	0.065	0.101	0.071	0.093
			$T = 500$			
$(\sigma_1, \sigma_2, \sigma_3)$	$(\tau_1 = 0.3, \tau_2 = 0.3)$		$(\tau_1 = 0.5, \tau_2 = 0.5)$		$(\tau_1 = 0.7, \tau_2 = 0.7)$	
$(1, 2, 2)$	0.065	0.078	0.066	0.088	0.067	0.105
$(2, 1, 1)$	0.059	0.075	0.056	0.070	0.054	0.062
$(\sigma_1, \sigma_2, \sigma_3)$	$(\tau_1 = 0.3, \tau_2 = 0.6)$		$(\tau_1 = 0.4, \tau_2 = 0.7)$		$(\tau_1 = 0.3, \tau_2 = 0.7)$	
$(1, 1.5, 1)$	0.058	0.062	0.062	0.069	0.059	0.064
$(1, 2, 1)$	0.071	0.089	0.077	0.103	0.074	0.093

TABLE 2.11: Empirical size and power of Wild bootstrap based statistics under unconditional variance breaks.

DGP: $y_t = y_{t-1} + \phi((1+\cos(2\pi kt/T))/2)y_{t-1} + \varepsilon_t,$

$$\varepsilon_t \sim \begin{cases} N(0,\sigma_1^2) & for \quad t = 1, ..., \lfloor \tau_1 T \rfloor \\ N(0,\sigma_2^2) & for \quad t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor \\ N(0,\sigma_3^2) & for \quad t = \lfloor \tau_2 T \rfloor + 1, ..., T, \end{cases}$$

	$DF^{GLS_{\mu^*}}$			$\mathcal{T}_{\hat{k}}^{GLS_\mu^*}$		
$T = 250$	$\sigma_1 = 1$, $\sigma_2 = 2$ and $\sigma_3 = 2$					
$(\phi, k)/(\bar{\tau}_1,\ \bar{\tau}_2)$	$(0.3, 0.3)$	$(0.5, 0.5)$	$(0.7, 0.7)$	$(0.3, 0.3)$	$(0.5, 0.5)$	$(0.7, 0.7)$
$(0, 0)$	0.051	0.054	0.055	0.058	0.046	0.044
$(-0.1, 0.5)$	0.372	0.312	0.313	0.323	0.276	0.276
$(-0.1, 1.0)$	0.428	0.575	0.650	0.590	0.654	0.676
$(-0.1, 1.5)$	0.519	0.552	0.463	0.581	0.571	0.423
$(-0.1, 2.0)$	0.604	0.538	0.534	0.689	0.649	0.640
$(-0.1, 2.5)$	0.569	0.510	0.570	0.652	0.597	0.577
$(-0.1, 3.0)$	0.607	0.641	0.589	0.736	0.716	0.654
	$\sigma_1 = 2$, $\sigma_2 = 2$ and $\sigma_3 = 1$					
$(0, 0)$	0.057	0.053	0.041	0.052	0.037	0.049
$(-0.1, 0.5)$	0.624	0.683	0.651	0.683	0.699	0.671
$(-0.1, 1.0)$	0.582	0.452	0.463	0.606	0.552	0.576
$(-0.1, 1.5)$	0.490	0.491	0.583	0.564	0.549	0.685
$(-0.1, 2.0)$	0.486	0.573	0.562	0.591	0.685	0.688
$(-0.1, 2.5)$	0.532	0.613	0.599	0.601	0.670	0.712
$(-0.1, 3.0)$	0.597	0.606	0.670	0.638	0.658	0.733
	$\sigma_1 = 1$, $\sigma_2 = 1.5$ and $\sigma_3 = 1$					
$(\phi, k)/\bar{\tau}$	$(0.3, 0.6)$	$(0.4, 0.7)$	$(0.3, 0.7)$	$(0.3, 0.6)$	$(0.4, 0.7)$	$(0.3, 0.7)$
$(0, 0)$	0.046	0.052	0.056	0.042	0.052	0.050
$(-0.1, 0.5)$	0.505	0.460	0.486	0.520	0.455	0.482
$(-0.1, 1.0)$	0.422	0.458	0.424	0.569	0.588	0.548
$(-0.1, 1.5)$	0.528	0.610	0.566	0.663	0.685	0.671
$(-0.1, 2.0)$	0.667	0.616	0.629	0.754	0.692	0.696
$(-0.1, 2.5)$	0.574	0.550	0.580	0.695	0.679	0.700
$(-0.1, 3.0)$	0.640	0.653	0.667	0.729	0.747	0.749
	$\sigma_1 = 1$, $\sigma_2 = 2$ and $\sigma_3 = 1$					
$(0, 0)$	0.052	0.056	0.052	0.059	0.058	0.057
$(-0.1, 0.5)$	0.505	0.434	0.461	0.509	0.387	0.436
$(-0.1, 1.0)$	0.352	0.401	0.339	0.442	0.482	0.444
$(-0.1, 1.5)$	0.520	0.647	0.581	0.625	0.684	0.667
$(-0.1, 2.0)$	0.707	0.592	0.604	0.727	0.643	0.648
$(-0.1, 2.5)$	0.558	0.514	0.573	0.653	0.618	0.650
$(-0.1, 3.0)$	0.602	0.617	0.649	0.677	0.710	0.740

TABLE 2.12: Empirical size and power of Wild bootstrap based statistics under unconditional variance breaks - linear trend case

DGP: $y_t = y_{t-1} + \phi((1+\cos(2\pi kt/T))/2)y_{t-1} + \varepsilon_t,$

$$\varepsilon_t \sim \begin{cases} N(0,\sigma_1^2) & for \quad t = 1, ..., \lfloor \tau_1 T \rfloor \\ N(0,\sigma_2^2) & for \quad t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor \\ N(0,\sigma_3^2) & for \quad t = \lfloor \tau_2 T \rfloor + 1, ..., T, \end{cases}$$

	$DF^{GLS_\tau^*}$			$\mathcal{T}_{\hat{k}}^{GLS_\tau^*}$		
$T = 250$	$\sigma_1 = 1,\ \sigma_2 = 2$ and $\sigma_3 = 2$					
$(\phi, k)/(\bar{\tau}_1,\ \bar{\tau}_2)$	$(0.3, 0.3)$	$(0.5, 0.5)$	$(0.7, 0.7)$	$(0.3, 0.3)$	$(0.5, 0.5)$	$(0.7, 0.7)$
$(0, 0)$	0.045	0.054	0.051	0.050	0.050	0.060
$(-0.1, 0.5)$	0.259	0.230	0.200	0.188	0.178	0.176
$(-0.1, 1.0)$	0.175	0.262	0.332	0.234	0.285	0.341
$(-0.1, 1.5)$	0.260	0.317	0.244	0.254	0.287	0.222
$(-0.1, 2.0)$	0.294	0.258	0.236	0.279	0.263	0.256
$(-0.1, 2.5)$	0.276	0.241	0.322	0.274	0.242	0.265
$(-0.1, 3.0)$	0.266	0.315	0.273	0.311	0.319	0.269
	$\sigma_1 = 2,\ \sigma_2 = 2$ and $\sigma_3 = 1$					
$(0, 0)$	0.044	0.044	0.051	0.057	0.049	0.046
$(-0.1, 0.5)$	0.470	0.520	0.493	0.391	0.386	0.352
$(-0.1, 1.0)$	0.264	0.202	0.171	0.302	0.228	0.203
$(-0.1, 1.5)$	0.245	0.199	0.259	0.234	0.192	0.290
$(-0.1, 2.0)$	0.213	0.250	0.255	0.218	0.262	0.264
$(-0.1, 2.5)$	0.223	0.269	0.226	0.235	0.279	0.244
$(-0.1, 3.0)$	0.245	0.230	0.288	0.279	0.262	0.304
	$\sigma_1 = 1,\ \sigma_2 = 1.5$ and $\sigma_3 = 1$					
$(\phi, k)/\bar{\tau}$	$(0.3, 0.6)$	$(0.4, 0.7)$	$(0.3, 0.7)$	$(0.3, 0.6)$	$(0.4, 0.7)$	$(0.3, 0.7)$
$(0, 0)$	0.051	0.049	0.049	0.053	0.051	0.050
$(-0.1, 0.5)$	0.365	0.317	0.326	0.290	0.254	0.262
$(-0.1, 1.0)$	0.163	0.180	0.156	0.184	0.208	0.181
$(-0.1, 1.5)$	0.231	0.270	0.249	0.249	0.284	0.278
$(-0.1, 2.0)$	0.312	0.268	0.266	0.320	0.289	0.282
$(-0.1, 2.5)$	0.238	0.218	0.235	0.300	0.247	0.262
$(-0.1, 3.0)$	0.261	0.263	0.283	0.327	0.313	0.353
	$\sigma_1 = 1,\ \sigma_2 = 2$ and $\sigma_3 = 1$					
$(0, 0)$	0.053	0.053	0.054	0.043	0.050	0.049
$(-0.1, 0.5)$	0.361	0.302	0.317	0.267	0.208	0.245
$(-0.1, 1.0)$	0.126	0.146	0.124	0.126	0.128	0.124
$(-0.1, 1.5)$	0.220	0.278	0.254	0.226	0.278	0.252
$(-0.1, 2.0)$	0.321	0.267	0.253	0.302	0.279	0.255
$(-0.1, 2.5)$	0.228	0.210	0.226	0.276	0.194	0.237
$(-0.1, 3.0)$	0.240	0.253	0.278	0.276	0.270	0.316

Conditional Heteroskedasticity

Finally, to investigate the finite sample distortions caused by the existence of conditional heteroskedasticity we consider that the innovations, ε_t, in (2.12) follow a GARCH(1,1) process, such that $\varepsilon_t = e_t\sqrt{h_t}$ with $h_t = \omega + \zeta\varepsilon_{t-1}^2 + \xi h_{t-1}$, $y_1 = \varepsilon_1 \sim N(0,1)$, $h_1 = 1$, $e_t \sim N(0,1)$, $\zeta \in \{0.7, 0.8, 0.9\}$, $\xi \in \{0, 0.05, 0.1, 0.2\}$ and $\omega = 1 - \zeta - \xi$, implying an unconditional variance of unity.

Table 2.13 reports the empirical size for $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ (results for $\mathcal{T}_{\hat{k}}^{GLS_\tau}$ are provided in Table 2.14). Results show that this test suffers relevant size distortions in the presence of conditional heteroskedasticity, especially when ζ is large. Its empirical size is slightly larger than that of DF^{GLS_μ} and considerably exceeds the nominal 5% level for all parameter values considered. However, the results using the EW version of the test in (2.6) and the Wild bootstrap approach presented in Table 2.15 for $T = 250$ (see Table 2.16 for results for $\mathcal{T}_{\hat{k}}^{GLS_\tau}$), show that the empirical sizes are close to the nominal 5% significance level for both demeaned and de-trended cases.

Under the alternative hypothesis relevant power gains are observed relatively to DF^{GLS} when EW is considered. When the Wild bootstrap technique is employed, the superiority relatively to DF^{GLS} is less pronounced for all parameter configurations considered.

Table 2.13: Empirical size assuming *iid* errors in the presence of GARCH effects

DGP:
$$y_t = y_{t-1} + \varepsilon_t$$
$$\varepsilon_t = e_t\sqrt{h_t},\ h_t = \omega + \zeta\varepsilon_{t-1}^2 + \xi h_{t-1}$$
$$\omega = 1 - \zeta - \xi,\ e_t \sim N(0,1)$$

	DF^{GLS_τ}			$\mathcal{T}_{\hat{k}}^{GLS_\tau}$		
			$T = 150$			
ξ/ζ	0.7	0.8	0.9	0.7	0.8	0.9
0	0.063	0.065	0.071	0.083	0.093	0.106
0.05	0.065	0.071	0.079	0.086	0.099	0.114
0.1	0.068	0.075	-	0.089	0.102	-
0.2	0.077	-	-	0.102	-	-
			$T = 250$			
0	0.066	0.073	0.079	0.085	0.094	0.109
0.05	0.068	0.074	0.079	0.088	0.099	0.114
0.1	0.069	0.076	-	0.092	0.104	-
0.2	0.076	-	-	0.102	-	-
			$T = 500$			
0	0.061	0.067	0.071	0.071	0.085	0.096
0.05	0.064	0.068	0.074	0.078	0.091	0.104
0.1	0.066	0.071	-	0.086	0.098	-
0.2	0.069	-	-	0.096	-	-

Table 2.14: Empirical size assuming *iid* errors in the presence of GARCH effects - linear trend case

DGP:
$$y_t = y_{t-1} + \varepsilon_t$$
$$\varepsilon_t = e_t\sqrt{h_t},\ h_t = \omega + \zeta\varepsilon_{t-1}^2 + \xi h_{t-1}$$
$$\omega = 1 - \zeta - \xi,\ e_t \sim N(0,1)$$

	DF^{GLS_τ}			$\mathcal{T}_{\hat{k}}^{GLS_\tau}$		
			$T = 150$			
ξ/ζ	0.7	0.8	0.9	0.7	0.8	0.9
0	0.079	0.091	0.102	0.127	0.146	0.168
0.05	0.084	0.097	0.108	0.130	0.153	0.183
0.1	0.087	0.102	-	0.138	0.164	-
0.2	0.098	-	-	0.157	-	-
			$T = 250$			
0	0.075	0.084	0.099	0.120	0.142	0.168
0.05	0.076	0.089	0.106	0.126	0.149	0.178
0.1	0.081	0.097	-	0.133	0.161	-
0.2	0.093	-	-	0.157	-	-
			$T = 500$			
0	0.070	0.078	0.090	0.103	0.126	0.158
0.05	0.072	0.084	0.100	0.110	0.138	0.173
0.1	0.076	0.091	-	0.120	0.153	-
0.2	0.089	-	-	0.147	-	-

TABLE 2.15: Empirical size and power in the presence of GARCH effects using EW standard errors and Wild bootstrap based test statistics

DGP:

$$y_t = y_{t-1} + \phi((1+\cos(2\pi kt/T))/2)y_{t-1} + \varepsilon_t,$$
$$\varepsilon_t = e_t\sqrt{h_t},\ h_t = \omega + \zeta\varepsilon_{t-1}^2 + \xi h_{t-1}$$
$$\omega = 1 - \zeta - \xi,\ e_t \sim N(0,1)$$

$T = 250$		$DF^{GLS_\mu^{EW}}$			$\mathcal{T}_{\hat{k}}^{GLS_\mu^{EW}}$			$DF^{GLS_\mu^*}$			$\mathcal{T}_{\hat{k}}^{GLS_\mu^*}$		
	ξ/ζ	0.7	0.8	0.9	0.7	0.8	0.9	0.7	0.8	0.9	0.7	0.8	0.9
$\phi = 0$	0	0.051	0.049	0.049	0.058	0.054	0.052	0.046	0.055	0.056	0.06	0.064	0.067
	0.05	0.05	0.05	0.048	0.056	0.055	0.05	0.051	0.055	0.055	0.057	0.063	0.064
	0.1	0.05	0.048	-	0.057	0.052	-	0.052	0.051	-	0.058	0.068	-
	0.2	0.052	-	-	0.053	-	-	0.053	-	-	0.058	-	-
$\phi = -0.1$													
$k = 0.5$	0	0.396	0.37	0.336	0.434	0.399	0.348	0.442	0.436	0.413	0.468	0.459	0.428
	0.05	0.386	0.352	0.299	0.422	0.379	0.306	0.442	0.427	0.405	0.473	0.453	0.4
	0.1	0.375	0.335	-	0.407	0.352	-	0.446	0.426	-	0.465	0.44	
	0.2	0.34	-	-	0.36	-	-	0.445	-	-	0.458		
$k = 1$	0	0.454	0.422	0.382	0.571	0.531	0.483	0.531	0.518	0.491	0.556	0.528	0.487
	0.05	0.438	0.406	0.351	0.557	0.513	0.432	0.525	0.506	0.469	0.541	0.515	0.448
	0.1	0.426	0.385	-	0.543	0.486	-	0.516	0.499	-	0.533	0.492	-
	0.2	0.39	-	-	0.496	-	-	0.497	-	-	0.513	-	-
$k = 1.5$	0	0.429	0.396	0.359	0.52	0.482	0.432	0.523	0.518	0.488	0.551	0.526	0.494
	0.05	0.415	0.38	0.325	0.505	0.46	0.391	0.523	0.514	0.469	0.544	0.515	0.447
	0.1	0.4	0.36	-	0.488	0.437	-	0.526	0.514	-	0.525	0.493	-
	0.2	0.366	-	-	0.449	-	-	0.524	-	-	0.485	-	-
$k = 2$	0	0.494	0.459	0.4138	0.613	0.581	0.532	0.556	0.55	0.534	0.637	0.612	0.554
	0.05	0.482	0.44	0.3748	0.603	0.56	0.484	0.553	0.543	0.511	0.627	0.592	0.496
	0.1	0.467	0.418	-	0.589	0.536	-	0.544	0.532	-	0.619	0.551	-
	0.2	0.428	-	-	0.545	-	-	0.538	-	-	0.573	-	-
$k = 2.5$	0	0.475	0.435	0.3921	0.603	0.566	0.514	0.544	0.542	0.52	0.621	0.599	0.555
	0.05	0.459	0.417	0.3579	0.589	0.546	0.462	0.546	0.535	0.495	0.618	0.587	0.516
	0.1	0.444	0.395	-	0.576	0.518	-	0.546	0.524	-	0.597	0.569	-
	0.2	0.404	-	-	0.531	-	-	0.519	-	-	0.574	-	-
$k = 3$	0	0.533	0.491	0.441	0.66	0.625	0.569	0.613	0.585	0.56	0.684	0.643	0.575
	0.05	0.518	0.469	0.398	0.647	0.603	0.517	0.595	0.582	0.527	0.663	0.621	0.519
	0.1	0.498	0.443	-	0.631	0.577	-	0.596	0.562	-	0.65	0.598	-
	0.2	0.45	-	-	0.587	-	-	0.571	-	-	0.613	-	-

Notes: $DF^{GLS_\mu^{EW}}$ and $\mathcal{T}_{\hat{k}}^{GLS_\mu^{EW}}$ are EW based test statistics and, $DF^{GLS_\mu^*}$ and $\mathcal{T}_{\hat{k}}^{GLS_\mu^*}$ are Wild bootstrap based test statistics.

Table 2.16: Empirical size and power in the presence of GARCH effects using EW standard errors and Wild bootstrap based test statistics - linear trend case

DGP: $y_t = y_{t-1} + \phi((1+\cos(2\pi kt/T))/2)y_{t-1} + \varepsilon_t$, $\varepsilon_t = e_t\sqrt{h_t}$, $h_t = \omega + \zeta\varepsilon_{t-1}^2 + \xi h_{t-1}$, $\omega = 1-\zeta-\xi$, $e_t \sim N(0,1)$

$T = 250$		$DF^{GLS_\tau^{EW}}$			$\mathcal{T}_{\hat{k}}^{GLS_\tau^{EW}}$			$DF^{GLS_\tau^*}$			$\mathcal{T}_{\hat{k}}^{GLS_\tau^*}$		
	ξ/ζ	0.7	0.8	0.9	0.7	0.8	0.9	0.7	0.8	0.9	0.7	0.8	0.9
$\phi = 0$	0	0.041	0.037	0.035	0.049	0.044	0.040	0.049	0.051	0.057	0.064	0.069	0.074
	0.05	0.039	0.035	0.033	0.049	0.043	0.040	0.049	0.055	0.061	0.068	0.068	0.076
	0.1	0.038	0.033	-	0.047	0.042	-	0.058	0.056	-	0.068	0.075	-
	0.2	0.027	-	-	0.044	-	-	0.058	-	-	0.070	-	-
$\phi = -0.1$													
$k = 0.5$	0	0.229	0.200	0.170	0.225	0.201	0.175	0.321	0.323	0.311	0.297	0.290	0.275
	0.05	0.219	0.186	0.150	0.216	0.191	0.156	0.323	0.318	0.306	0.296	0.283	0.26
	0.1	0.207	0.174	-	0.207	0.180	-	0.322	0.307	-	0.295	0.282	-
	0.2	0.169	-	-	0.184	-	-	0.308	-	-	0.281	-	-
$k = 1$	0	0.168	0.147	0.126	0.227	0.210	0.188	0.217	0.214	0.219	0.251	0.234	0.224
	0.05	0.159	0.139	0.114	0.219	0.201	0.174	0.218	0.22	0.216	0.245	0.227	0.208
	0.1	0.153	0.127	-	0.213	0.193	-	0.219	0.218	-	0.237	0.22	-
	0.2	0.125	-	-	0.196	-	-	0.226	-	-	0.219	-	-
$k = 1.5$	0	0.169	0.148	0.126	0.215	0.196	0.176	0.238	0.240	0.241	0.246	0.24	0.232
	0.05	0.161	0.138	0.114	0.208	0.187	0.162	0.249	0.243	0.238	0.252	0.242	0.218
	0.1	0.151	0.131	-	0.200	0.181	-	0.246	0.242	-	0.247	0.237	-
	0.2	0.124	-	-	0.185	-	-	0.243	-	-	0.239	-	-
$k = 2$	0	0.189	0.165	0.139	0.283	0.262	0.244	0.250	0.249	0.251	0.265	0.259	0.246
	0.05	0.180	0.155	0.128	0.276	0.252	0.227	0.249	0.249	0.249	0.274	0.258	0.254
	0.1	0.168	0.145	-	0.265	0.242	-	0.257	0.257	-	0.277	0.259	-
	0.2	0.138	-	-	0.247	-	-	0.262	-	-	0.260	-	-
$k = 2.5$	0	0.186	0.164	0.142	0.282	0.261	0.242	0.250	0.247	0.251	0.286	0.267	0.246
	0.05	0.177	0.153	0.127	0.274	0.252	0.226	0.246	0.249	0.245	0.279	0.257	0.228
	0.1	0.167	0.142	-	0.264	0.242	-	0.250	0.257	-	0.267	0.243	-
	0.2	0.137	-	-	0.245	-	-	0.255	-	-	0.259	-	-
$k = 3$	0	0.203	0.182	0.156	0.332	0.310	0.289	0.281	0.284	0.279	0.319	0.297	0.285
	0.05	0.193	0.169	0.139	0.324	0.299	0.268	0.285	0.289	0.285	0.313	0.295	0.260
	0.1	0.184	0.157	-	0.315	0.287	-	0.287	0.291	-	0.306	0.283	-
	0.2	0.148	-	-	0.288	-	-	0.287	-	-	0.293	-	-

Notes: $DF^{GLS_\mu^{EW}}$ and $\mathcal{T}_{\hat{k}}^{GLS_\mu^{EW}}$ are EW based test statistics and, $DF^{GLS_\mu^*}$ and $\mathcal{T}_{\hat{k}}^{GLS_\mu^*}$ are Wild bootstrap based test statistics.

2.4 Empirical Application

The very strong persistence of inflation has been seen as a "stylized fact" of developed economies. However, in recent years, substantial evidence has emerged suggesting that this conclusion may result from changes in the time series properties of inflation that have not been taken into account. For instance, the inflation target and the willingness to stabilize inflation may vary over time causing inflation series not to return to a constant mean in a linear AR framework.

An important branch of literature uses AR model-based measures, such as, the largest autoregressive root (LAAR) and the sum of the autoregressive coefficients (SARC) to investigate inflation persistence; Levin and Piger (2003) and Taylor (2000). In general, these works conclude that, when the deterministic terms or the autoregressive parameters are allowed to change, inflation persistence is lower and far from that of a random walk process. There is some consensus that inflation persistence decreased since the 1980s (possibly due to preferences for price stability). For time-variation in the autoregressive parameters researchers have tended to rely on split samples or rolling regressions. An alternative approach also used is to estimate the path of the time-varying autoregressive parameter using a state-space model and the Kalman filter; Beechey and Österholm (2012).

Several works have analyzed the order of integration as a measure of inflation persistence and employed stationarity or unit root tests in the analysis. Testing procedures that allow the order of integration to endogenously change over time have also been considered in this context. For instance, Harvey et al. (2006) found evidence, using persistence change tests, that CPI inflation in the US suffered a shift in persistence from $I(1)$ to $I(0)$. Halunga et al. (2009) applied the same tests to UK and US inflation data and reported similar results. To circumvent the single change in persistence, which is a limitation of these tests, the sample is partitioned when a break is found and the test re-applied on each sub-sample. The findings achieved using this approach indicate a first change from $I(0)$ to $I(1)$ in the early 1970s and a subsequent reversion to $I(0)$ in the early 1980s, suggesting that the nonstationary dynamics of inflation lasted only about ten years.

Most of the recently proposed methodologies provide strong evidence against the statement that inflation dynamics is described by a pure $I(1)$ process. In this section, we apply the proposed tests, allowing for multiple changes in persistence, to inflation data for several developed economies. In the empirical analysis we allow for a maximum of three breaks, which seems adequate given the available sample sizes.

In our analysis, we consider inflation persistence as the speed at which inflation converges to equilibrium after a shock. Thus, the parameter $\phi_t := \phi\cos(k,t)$ in the test regression $\Delta\hat{\pi}_t = \phi_t\hat{\pi}_{t-1} + \sum_{j=1}^{p}\delta_j\Delta\hat{\pi}_{t-j} + \varepsilon_t$, where $\hat{\pi}_t$ is the locally GLS demeaned inflation time series, is a reasonable indicator of the persistence dynamics and its statistical significance is tested using the procedure introduced in Section 2.2, i.e.,

$$\mathcal{T}_{\hat{k}}^{GLS_\mu} = \min_{k\in K} \hat{t}_k^{GLS_\mu},\ K = \{0.5, 1, 1.5, 2, 2.5, 3\}\,. \tag{2.16}$$

However, as the timing of the breaks influences the power performance of the tests, we also considered the time series in reverse chronological order. The reversed test,

$$\mathcal{T}_{\hat{k},r}^{GLS_\mu} := \min_{k\in K} \hat{t}_{k,r}^{GLS} = \min_{k\in K} \frac{\sum_{t=2}^{T} \Delta\hat{\pi}_{t,r} \cos(k,t)\hat{\pi}_{t-1,r}}{\left[\hat{\sigma}_k^2 \sum_{t=2}^{T} \cos^2(k,t)\hat{\pi}_{t-1,r}^2\right]^{1/2}},$$

where $\hat{\pi}_{1,r} = \hat{\pi}_T, \hat{\pi}_{2,r} = \hat{\pi}_{T-1}, ..., \hat{\pi}_{T,r} = \hat{\pi}_1$, has the following asymptotic distribution,

$$\mathcal{T}_{\hat{k},r}^{GLS_\mu} \Rightarrow \min_{k\in K} \frac{-\cos(k,0)W(1)^2 + \frac{1}{2}(2\pi k)^2 \int_0^1 \cos(2\pi kr)W(r)^2 dr - 1}{2\left(\int_0^1 \cos^2(k,r)W(r)^2 dr\right)^{1/2}}\,. \tag{2.17}$$

The values of $\tilde{c}$ for a given $k \in K = \{0.5, 1, 1.5, 2, 2.5, 3\}$ and the critical values of the $\mathcal{T}_{\hat{k},r}^{GLS_\mu}$ test are very close to those obtained using the normal chronological order (see Tables 2.1 and 2.2) and are omitted for the sake of space, but can be obtained from the authors.

Our sample consists of quarterly CPI data for the G7 countries obtained from OECD Statistics for the period from 1955Q1 to 2018Q2. The quarterly CPI is then used to compute the year-on-year and the quarterly growth rates of the CPI data.[6] Note that when the quarterly growth rate is stationary, the year-on-year growth of the CPI introduces a non-invertible moving average component in the resulting time series, which is responsible for a loss of power of the *ADF* tests. However, since the year-on-year growth of CPI is relevant for monetary policy decisions (inflation targeting regimes typically observe its evolution) we will also consider this definition.

Figures 2.2 and 2.3 display, respectively, the year-on-year and the quarterly growth rates of the CPI time series for the G7 countries. Simple visual inspection suggests that the 1970s was the period with the highest inflation rates for almost all countries. Over these years, which have as a milestone the collapse of Bretton Woods, the world economy faced a period of turbulence in which the option for a highly accommodative monetary policy seems to have triggered an unusual increase in inflation persistence only attenuated in the early 80's with the shift to a more restrictive monetary policy. This option, characterized by the introduction of inflation targeting as a framework for monetary policy, contributed to a long period of low and stable inflation in developed countries. Since the 2008 global financial crisis, price stability is once more a concern but this time because of the risk that inflation could remain too low for too long. In order to avoid deflation and to bring inflation back to a desirable level, central

[6]According to Hassler and Demetrescu (2005), the power performance of the *ADF* test crucially depends on whether inflation is assumed to be equal to the year-on-year growth or to the quarterly growth of CPI.

banks implemented an expansionary monetary policy, which may have had an impact on inflation persistence. Thus, for the sample period considered in this work, two breaks in persistence may have occurred associated with periods of turbulence and relevant changes in monetary policy.

In the analysis we also consider the unit root test proposed by Rodrigues and Taylor (2012) to infer if allowing for structural breaks in the intercept is sufficient to gather evidence against the null hypothesis. Moreover, Wild bootstrap p-values of the proposed test are also computed in order to prevent that the results are influenced by breaks in the unconditional variance (or even by the presence of GARCH effects).

Table 2.17 reports the results for the year-on-year growth rate of CPI. Considering the minimum between the proposed test computed in normal and reverse chronological order (the critical values used are in the note to Table 2.17), the null hypothesis of a unit root is rejected at the 5% significance level for all countries except Japan. For five of the seven countries, the stronger rejections occur when the reverse chronological order is considered. The estimated $\hat{k}$ parameter equals one for all cases, suggesting that the period of stronger persistence occurred somewhere around the middle of the sample. France and Germany are the two exceptions for which using the time series in normal chronological order leads to stronger rejections of the null hypothesis. Here, two breaks in persistence seem to have occurred. One before the middle of the sample and the other at the end of the sample, suggesting that the 2008/2009 financial crisis originated a statistically significant change in inflation persistence.

Table 2.17 also reports results for the quarterly growth rate of CPI. When the minimum between $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ and $\mathcal{T}_{\hat{k},r}^{GLS_\mu}$ is considered, there is statistical evidence against the unit root hypothesis for all countries at the 5% significance level. In terms of the frequency parameter, $\hat{k}$, associated with the smallest test statistic, there are some differences relative to the conclusions drawn above for the year-on-year growth rate. The selected $\hat{k}$ is the same for France, Italy and the US but differs for the other four countries. For instance, for Japan, the quarterly growth rate of CPI had three periods of stronger persistence, one of them at the beginning of the sample.

The proposed cosine term can only provide an approximation of the breaks in persistence and, for the chosen values of k, it suggests that the persistence parameter only assumes two values: one close to unity when the function reaches its minimum value and a lower persistence value when the function reaches its maximum. Furthermore, considering $\hat{k} = 3$ implies that there were a maximum of three periods of shorter duration with higher persistence in inflation. Given the sample length available this number may be large in some cases. Lastly, simulations show that the power gains achieved by the new test are maximized when k is smaller. Even with these limitations, we found relevant evidence against the unit root hypothesis in the inflation rate series for several industrial countries.

The proposed test provides statistical evidence of breaks in persistence for several countries considering both the year-on-year and the quarterly growth rate of CPI. When the year-on-year growth CPI is considered, $\hat{k} = 1$ is selected for most countries employing the minimum between the proposed test computed in normal and reverse chronological order, which seems to suggest that the difficulty in rejecting the $I(1)$ hypothesis is largely due to the Great Inflation of the 1970s (beyond the limitations of available unit root tests such as, for instance, low power in small samples). Regarding the last ten years, the plausible change in persistence after the crisis seems not to have such a large influence, since the proposed unit root test continues to provide strong rejections of the unit root hypothesis.

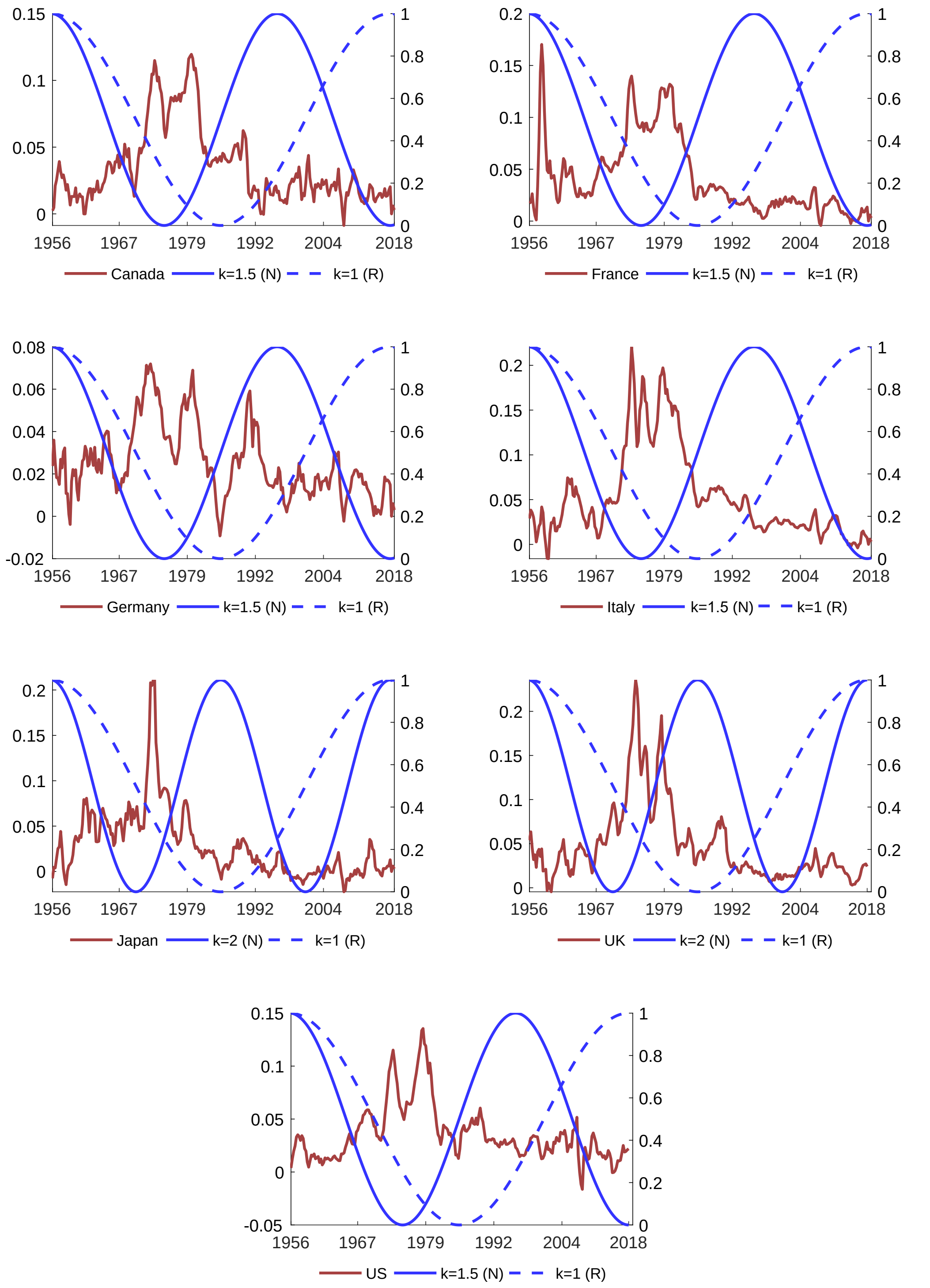


FIGURE 2.2: Year-on-year growth of the CPI for the G7 countries

Notes: Year-on-year growth of the CPI means percentage change relative to the same quarter of the previous year.

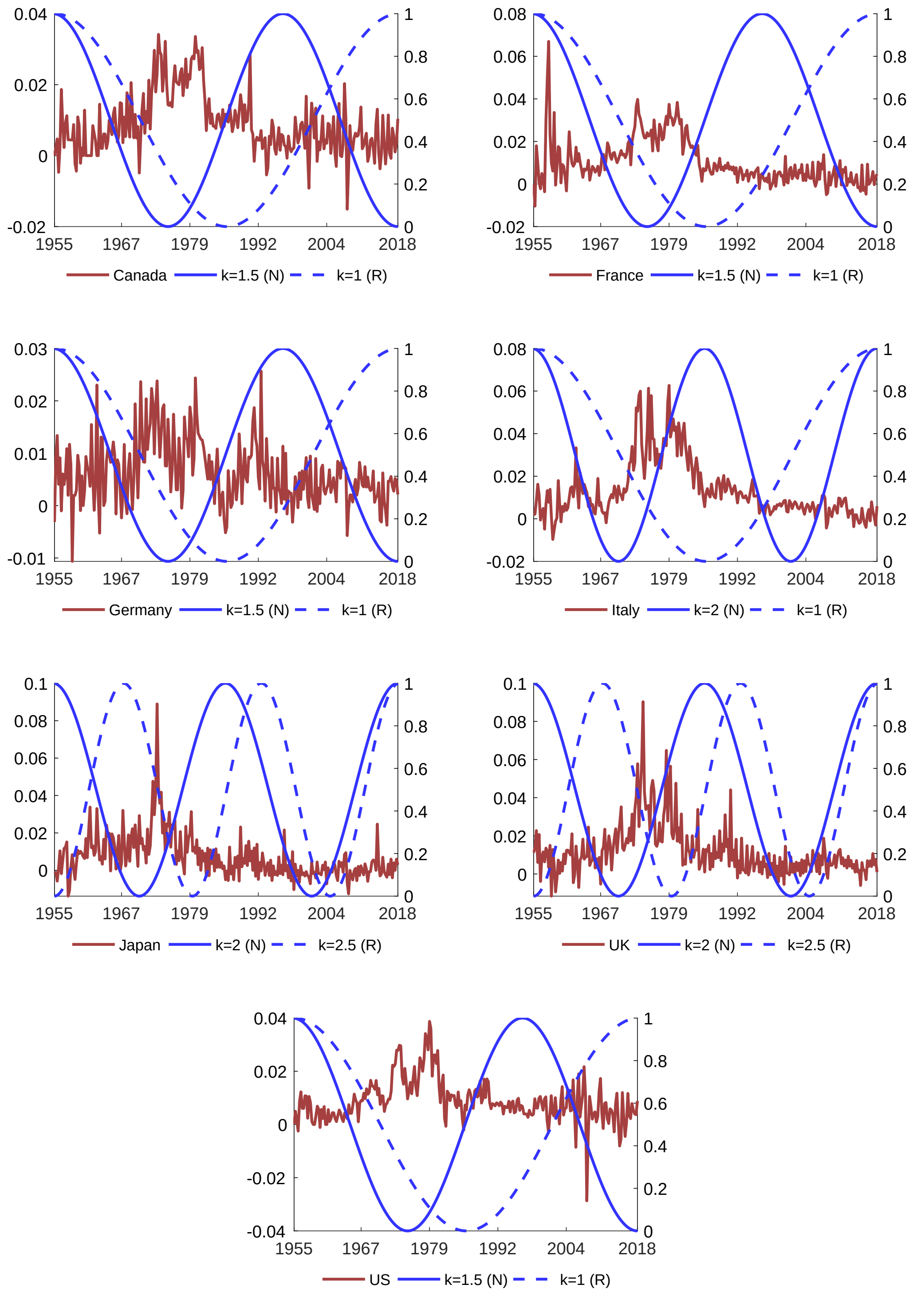


FIGURE 2.3: Quarterly growth of the CPI for the G7 countries

Notes: Quarterly growth of the CPI means percentage change relative to the previous quarter.

Table 2.17: Results using year-on-year and quarterly growth of the CPI for the G7 countries

			Year-on-year growth			
Country	DF^{GLS_μ}	$\hat{k}$	$\mathcal{T}_{\hat{k}}^{GLS_\mu}$	WB p-value	k	$t_\alpha^{ERS_f^\mu}$
Canada	−1.040	1.5	−2.900 [b]	0.028	1	−2.276
France	−1.348	1.5	−3.494 [a]	0.016	1	−2.724
Germany	−2.481 [b]	1.5	−3.692 [a]	0.002	1	−3.040 [c]
Italy	−1.719 [c]	1.5	−2.280	0.161	1	−3.109 [c]
Japan	−1.320	2	−2.002	0.230	1	−2.378 [c]
United Kingdom	−1.392	2	−2.446 [c]	0.111	1	−1.951
United States	−1.065	1.5	−2.031	0.235	1	−2.343
			reverse chronological order			
Canada	−1.348	1	−3.748 [a]	0.000	1	−2.432
France	−1.109	1	−3.108 [b]	0.045	1	−2.381
Germany	−2.051 [b]	1	−3.583 [a]	0.007	1	−3.104 [c]
Italy	−1.432	1	−3.174 [b]	0.033	1	−3.114 [c]
Japan	−1.647	1	−2.586 [c]	0.091	1	−2.805
United Kingdom	−1.278	1	−2.894 [b]	0.054	1	−2.068
United States	−1.471	1	−4.096 [a]	0.002	1	−2.533
			Quarterly growth			
Country	DF^{GLS^μ}	$\hat{k}$	$\mathcal{T}_{\hat{k}}^{GLS_\mu}$	WB p-value	k	$t_\alpha^{ERS_f^\mu}$
Canada	−1.499	1.5	−3.021 [b]	0.016	1	−2.790
France	−1.303	1.5	−3.313 [a]	0.034	1	−2.713
Germany	−0.901	1.5	−1.706	0.325	1	−1.905
Italy	−1.375	2	−2.243	0.170	1	−2.487
Japan	−1.355	2	−1.967	0.219	1	−2.561
United Kingdom	−1.383	2	−3.178 [b]	0.029	1	−2.108
United States	−1.417	1.5	−2.528 [c]	0.235	1	−2.460
			reverse chronological order			
Canada	−1.745 [b]	1	−2.697 [b]	0.044	1	−1.887
France	−1.005	1	−3.013 [b]	0.046	1	−2.046
Germany	−1.635	1	−3.125 [b]	0.017	1	−3.156 [b]
Italy	−1.591	1	−3.205 [a]	0.037	1	−2.447
Japan	−1.654	2.5	−3.016 [b]	0.036	1	−3.055 [c]
United Kingdom	−0.988	2.5	−2.516 [c]	0.124	1	−2.079
United States	−1.798 [b]	1	−3.411 [a]	0.007	1	−2.329

Notes: (1) a, b and c denote significance at the 1%, 5% and 10% significance levels, respectively; (2) $t_\alpha^{ERS_f^\mu}$ corresponds to the test proposed by Rodrigues and Taylor (2012); (3) the lags of the unit root tests were chosen using the MAIC information criterion; (4) the values of the critical values of $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ with reverse chronological order and T=250 are -3.198, -2.634 and -2.352 for 1%, 5% and 10% significance levels, respectively; (5) the critical values for the minimum between $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ with normal and reverse chronological order and T=250 are -3.382, -2.839 and -2.568 for 1%, 5% and 10% significance levels, respectively.

2.5 Conclusions

In this paper we propose a simple approach to detect potential persistence changes and allow for the possibility of the occurrence of up to three breaks in persistence (note that more breaks can be allowed for if deemed necessary). The unknown shape and timing of the breaks are approximated using a cosine term. The test procedure is based on a one-sided t-statistic where this statistic is minimized over a set of values chosen *a priori* for the frequency parameter k.

We find via Monte Carlo simulations that our proposed test has interesting power performance when compared to e.g. the local GLS de-trended unit root tests when breaks in persistence are present, as well as others.

In addition to the DGPs suggested by the specification of the alternative hypothesis, which implies smooth breaks in persistence, we also investigated the power properties of the proposed test when abrupt breaks in ρ_t in (2.2) occur and results remain favourable. The power gains relative to the DF^{GLS} test are even greater if increases in persistence induce temporary nonstationary behaviour. Moreover, we also performed simulations to investigate the effects of unconditional and conditional heteroskedasticity. Although our proposed test shows some size distortions when the homoskedasticity assumption does not hold, this problem is attenuated using the Wild bootstrap which produces empirical sizes close to the nominal 5% level.

Applications of the proposed test to G7 countries' inflation data provided relevant statistical evidence of breaks in persistence. When the year-on-year growth of CPI in reverse chronological order is used, the null hypothesis of a unit root is rejected for all countries. Comparing these results with those obtained in normal chronological order suggests that the evidence of nonstationarity of the inflation series previously reported in the literature is possibly due to the occurrence of a period of stronger persistence in the first half of the sample.

This paper alerts to the consequences of ignoring the occurrence of breaks in persistence. Most of the work on unit root testing that employed Fourier series to approximate smooth structural breaks has focused on changes in the deterministic parameters only. However, changes in the behaviour of economic and financial variables caused by, for instance, exogenous events, shifts in monetary policy or improvements in the available technology may have altered not only the equilibrium value but also the speed of reversion to equilibrium after a shock. As previously mentioned, changes in persistence affect the conditional mean, the unconditional mean and the unconditional variance of the process. Thus, it is possible that some of the evidence in the literature regarding the occurrence of structural breaks in volatility may have been influenced by persistence changes; Sensier and van Dijk (2003) and McConnell and Perez-Quiros (2000).

Chapter 3

The Expected Time to Cross a Threshold and its Determinants: A Simple and Flexible Framework†

3.1 Introduction

The first hitting time or first passage time, i.e., the time a variable takes to reach a certain value, is a fundamental concept in stochastic analysis and represents an important modeling tool in fields, such as finance, biology and life sciences.

Although there is a large literature in economics and finance addressing this topic (see, for instance, Durbin; 1971, Lo et al.; 2002 or Giesecke; 2006), first-hitting time densities are mostly obtained for Wiener diffusion processes under the assumption of continuous-time, due to the tractability offered by the Itô calculus. However, this approach often requires strong computational efforts and closed form solutions are known only for some standard continuous-time models.

Since most economic and financial data is only available in discrete time, researchers usually opt for modeling duration time as a stochastic process instead of defining duration as the first time a stochastic process crosses a given threshold. Thus, continuous-time based first passage time densities and duration models (typically built in discrete time) have the same objective, namely, to characterize the length of time that separates different stochastic events. In fact, as illustrated by Whitmore (1986), duration models can be seen as reduced form representations of first passage time densities.

Most of the existing duration analysis literature is based on the specification of the hazard function, that is, on the conditional probability of exiting the initial state within a short interval having survived up to the starting time of that interval. Thus, the hazard function specification emphasizes the conditional probabilities[1]. Since a duration process can intuitively be associated with a dynamic sequence of conditional probabilities, the hazard-based approach is

†Chapter 3, *The Expected Time to Cross a Threshold and its Determinants: A Simple and Flexible Framework*, has been published (with a different empirical application) in co-authorship with João Nicolau and Paulo M. M. Rodrigues in the Journal of Economic Dynamics and Control 122, 104047, 2021 and is reprinted in this Dissertation with permission from Elsevier.

[1]Note that, for any hazard function specification there is a mathematically equivalent representation in terms of a probability distribution; see Kiefer (1988).

a convenient way to interpret duration data and can be sufficiently flexible to handle relevant issues such as the presence of censored observations and time-varying covariates. For instance, parametric hazard models have been used in labor economics to examine duration dependence and the determinants of unemployment exit probabilities (see, for instance, Meyer; 1990, McCall; 1994 and Sueyoshi; 1995); another example is its application to firm survival (see, for instance, Audretsch and Mahmood; 1995 and Mata and Portugal; 2002). In addition, this methodology has also been employed (without covariates) to investigate the presence of duration dependence in economic cycles (see, for instance, Sichel; 1991 and Ohn et al.; 2004).

A closely related approach to model duration dependence is to treat the occurrence of a given event as a random variable which follows a point process[2]. Let $\{t_i\}_{i\in\{1,2,\dots\}}$, with $0 \leq t_i \leq t_{i+1}$, be a sequence of non-negative random variables representing the times at which the events occur. The sequence $\{t_i\}$ is called a point process. A complete description of such processes is formulated in terms of the conditional intensity function which can be associated, roughly speaking, with the probability per unit of time, to observe an event in the next instant[3]. Thus, different parameterizations of this function result in different point process models. Existing models can be grouped into two classes. The first, formulated in calendar time, considers that the marginal effects of an event that has occurred in the past is independent of the intervening history; and the second class, focuses on the intervals between events and assumes that the duration between successive events depends on the number of intervening events. The autoregressive conditional dynamic (ACD) model proposed by Engle and Russell (1998) is an important model of this class[4].

In this paper, we focus on first hitting time processes. Thus, unlike point process models, we are not interested in the actual sequence $\{t_i\}$, but only in the random variable associated with the time at which the event occurs for the first time (t_1). As already mentioned, the first hitting time problems were mostly addressed in continuous time invoking Wiener processes, which involve complex mathematical concepts and can result in models which are difficult to estimate.

Nicolau (2017) introduces an intuitive and easy to implement framework for estimating the first passage time probability function in a discrete time context. The main contribution of the present work is the introduction of a novel approach to estimate transition probabilities allowing for covariates and which generalizes the framework of Nicolau (2017). To this end, we adapt the approach proposed by Islam and Chowdhury (2006) to estimate covariate-dependent Markov models of any order to the present context.

---

[2]For an introduction to point processes see, for instance, Cox and Isham (1980).

[3]The conditional intensity function can be seen as a counterpart to the hazard function.

[4]The ACD model and its extensions have become a leading tool in modeling irregularly spaced high-frequency financial data, which are characterized by the occurrence of strong clustering structures in the waiting times between consecutive events.

Understanding how a set of covariates influences the time a dependent variable takes to cross a fixed threshold may provide relevant insights on the potential causal relationships between economic variables. The proposed covariate-dependent expected time (ET) to cross a threshold estimator may also be a useful tool to support macroeconomic policy decisions, where there are desirable values or even formal targets for some key variables, such as, output growth, inflation, or unemployment. Thus, it is important to assess the effectiveness of the instruments (or covariates) in driving the outcome variable towards some preassigned values. In practice, the impact of the covariates may not be symmetric and may also depend on the distance between the starting point and the target value. Consider, for instance, the connection between monetary policy and real economic growth. Since both negative and above-trend growth rates are undesirable, monetary policy plays a key role in fostering a healthy level of economic growth. To this end, a tight monetary policy is adopted when the rapid economic growth causes inflationary pressures and an easy one is implemented in a recession in order to boost a rapid economic recovery. However, it has been noticed in the literature that the responsiveness of the real economy to monetary shocks is different during recessions and expansions (see, for instance, Florio; 2004 and references therein). The framework introduced in this paper allows us to investigate these possible nonlinear dynamics by estimating the ET conditional on different starting values. If, for a given starting value, changes in covariates are reflected in changes in the ET estimation, it suggests that the chosen covariates affect the movement towards a specified threshold in that specific situation. When other starting points are considered the conclusions may, however, differ.

To further illustrate the usefulness of the approach introduced in this paper an application to the industrial production (IP) - yield spread (YS) relationship is provided. Until the 2008-2010 financial crisis, short-term nominal interest rates were the primary monetary policy instrument used to achieve price stability, which is a key objective of central banks. For instance, in response to the financial crisis, central banks cut nominal interest rates in order to stimulate economic growth. However, as short-term interest rates in recent years have been close to their zero lower bound and economic growth remained low, unconventional monetary policies such as quantitative easing have been employed to reduce long-term interest rates and spur aggregate demand. In our analysis we will use the sovereign yield curve slope as a proxy for the monetary policy stance since it captures both conventional monetary policy and unconventional measures such as asset purchases and forward-guidance; see Saldías (2017). Thus, the focus of the empirical application is to investigate whether YS[5] influences the ET the IP growth rate takes to return to its stationary mean starting from a specific value.

The remainder of the paper is organized as follows. Section 3.2 introduces the proposed methodology to estimate the conditional ET to cross a threshold (given a specific starting point). Section 3.3 investigates the finite sample properties of the parameter estimates that describe the relationship between the ET

[5]YS is computed as the difference between the 10-years government bond yields and the 3-month T-bill rate.

and the covariates. Section 3.4 presents an empirical application to the U.S. economy, where we infer how YS influences the ET that the IP growth rate takes to return to its mean. Section 3.5 concludes and a Technical Appendix B collects the detailed proofs of the results presented in the paper.

3.2 The Proposed Methodology

3.2.1 The Process and Probabilities of Interest

Let $\mathbf{x}_t$ be a $k \times 1$ vector of covariates $\{(y_t, \mathbf{x}_t)\}$ be a vector of discrete-time processes with state space $\Re^{k+1}$ characterized by the following Assumption.

Assumption 1

(A) $y_t|\boldsymbol{x}_t$ *is a Markov process of order* r*;*

(B) $\{(y_t, \boldsymbol{x}_t)\}$ *is a jointly stationary vector stochastic process.*

Let A be a measurable set of range D of the process of interest, and define the first hitting time of A as $T_A := \inf\{t > 0 : y_t \in A\}$. There is a σ-finite measure $m(y)$ such that $m(A) > 0$ implies $E(T_A|X_0 = a) < \infty$ for every $a \in D \backslash \overline{A}$, where $\overline{A}$ is the closure of set A. Assumption 1.A ensures that the process $\{y_t\}$ is positive Harris recurrent, that is, if the process starts from a level a not belonging to the generic set A, it will visit A as $T \to \infty$ almost surely an infinite number of times (see Meyn et al.; 2009, chapter 9).

Consider the first hitting time $T_{z_1} = \inf\{t > 0 : y_t \geq z_1\}$ and that the process starts at z_0, with $z_0 < z_1$. The case $z_0 > z_1$ with $T_{z_1} = \inf\{t > 0 : y_t \leq z_1\}$ is almost analogous[6]. The distribution of T_{z_1} is usually difficult to derive, especially for non-linear processes. Thus, we consider a simple semi-parametric method to estimate these quantities. First, we define the following binary variable:

$$S_t := S_t(z_0, z_1) = \begin{cases} 0 & \text{if } \; y_t < z_1, \; y_{t-1} < z_1, ..., y_{t-k+1} < z_1, y_{t-k} \leq z_0, \\ 1 & \text{otherwise,} \end{cases} \tag{3.1}$$

where $k \geq 0$ and $S_0 = 0$ if $y_0 = z_0$ (note that z_0 is the starting value of the process). Then, the probability that y_t crosses the threshold z_1 for the first time starting fom z_0 is,

$$P(T_{z_1} = t) = P(S_t = 1, S_{t-1} = 0, S_{t-2} = 0, ..., S_1 = 0|S_0 = 0),$$

which is equivalent to

$$P(T_{z_1} = t) = \left(1 - p_t\right) \prod_{i=1}^{t-1} p_i \tag{3.2}$$

where $p_i := P(S_i = 0|S_{i-1} = 0, S_{i-2} = 0, ..., S_0 = 0)$ (see Appendix for details) .

[6] In practice, we can easily transform a $z_0 > z_1$ case into $z_0 < z_1$ by replacing z_0, z_1 and y_t by $-z_0$, $-z_1$ and $-y_t$, respectively.

Proposition 2 *Considering that Assumption 1.A holds and that $y_t|\mathbf{x}_t$ is a Markov process of order r, then $S_t|\mathbf{x}_t$ is also an rth order two-state Markov chain.*

Since in view of the Markovian property if $t > r$ then $p_t(\mathbf{x}) = p_r(\mathbf{x})$, from Proposition 1 and expression (3.2) it follows that,

$$P(T_{z_1} = t|\mathbf{x}) = \begin{cases} [1 - p_t(\mathbf{x})] \prod_{i=1}^{t-1} p_i(\mathbf{x}) & \text{for } \ t \leq r, \\ \left\{ [1 - p_r(\mathbf{x})] \prod_{i=1}^{r-1} p_i(\mathbf{x}) \right\} p_r(\mathbf{x})^{t-r} & \text{for } \ t > r, \end{cases} \tag{3.3}$$

where $p_i(\mathbf{x}) = P(S_t = 0|S_{t-1} = 0, ..., S_{t-i} = 0|\mathbf{x})$ for $1 \leq i \leq r$.

3.2.2 Covariate-Dependent Transition Probabilities

When Assumption 1.A and Proposition 2 hold, we can treat $S_t|\mathbf{x}_t$ as a Markov chain with state space $\{0, 1\}$ and use standard Markov chain inference to estimate the covariate-dependent transition probabilities.

For instance, if r=1, the transition probability matrix is

$$P(\mathbf{x}) = \begin{bmatrix} \pi_{00}(\mathbf{x}) & \pi_{01}(\mathbf{x}) \\ \pi_{10}(\mathbf{x}) & \pi_{11}(\mathbf{x}) \end{bmatrix}$$

with

$$p_1(\mathbf{x}) := \pi_{00}(\mathbf{x}) = P(S_t = 0|S_{t-1} = 0; \mathbf{x}) = \frac{\exp^{\mathbf{x}'\boldsymbol{\beta}_1}}{1 + \exp^{\mathbf{x}'\boldsymbol{\beta}_1}} =: \Lambda(\mathbf{x}'\boldsymbol{\beta}_1)$$

where $\boldsymbol{\beta}_1$ is a vector of parameters and $\Lambda(\mathbf{x}'\boldsymbol{\beta}_1)$ is the cumulative distribution function of the logistic distribution.

The generalization to higher order Markov chains is achieved in a straightforward way by extending the first order Markov chain model. Note that the transition probabilities of the rth order model can be arranged in a $2^r \times 2$ matrix of which we only need a line; see Islam and Chowdhury (2006). As an illustration, consider a matrix with the outcomes of an rth order Markov chain S_t, i.e.,

$$\begin{array}{cccccc} m & \{S_{t-r} & S_{t-(r-1)} & \dots & S_{t-1}\} \\ 1 & 0 & 0 & \dots & 0 \\ 2 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 2^r - 1 & 1 & 1 & \dots & 0 \\ 2^r & 1 & 1 & \dots & 1 \end{array} \left[\begin{array}{cc} \multicolumn{2}{c}{S_t} \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \\ 0 & 1 \end{array} \right],$$

where m is an index that identifies each of the possible outcomes of $\{S_{t-1}, S_{t-2}, ..., S_{t-r}\}$. For instance, $m = 1$ corresponds to the outcomes $S_{t-1} = 0, S_{t-2} = 0, ..., S_{t-r} = 0$.

As shown in (3.3), the covariate-dependent transition probabilities $p_1(\mathbf{x})$, ..., $p_r(\mathbf{x})$ are needed to obtain the probability function for $T_{z1}|\mathbf{x}$. Hence, if $S_t|\mathbf{x}$ is an rth order Markov chain, we can define,

$$p_r(\mathbf{x}) = P(S_t = 0|S_{t-1} = 0, ..., S_{t-r} = 0; \mathbf{x}) = \Lambda(\mathbf{x}'\boldsymbol{\beta}_r). \quad (3.4)$$

Moreover, for $j = 1, ..., r-1$,

$$p_j(\mathbf{x}) = P(S_t = 0|S_{t-1} = 0, ..., S_{t-j} = 0\,; \mathbf{x}) = \Lambda(\mathbf{x}'\boldsymbol{\beta}_j). \quad (3.5)$$

Then, for an rth order Markov chain, the log-likelihood function of a Markov chain of order $j < r$ will be used to estimate the parameter vector $\boldsymbol{\beta}_j$ and, consequently, obtain the probabilities $p_1(\mathbf{x})$, ..., $p_{r-1}(\mathbf{x})$.

Despite similarities between (3.4) and (3.5) and the standard logit model for binary responses, the proposed approach is less restrictive. Firstly, the focus here is on the transition probabilities between states and not on the conditional probability of success. Moreover, no specific functional form for the underlying latent variable model is assumed. In fact, the only assumption we make on the data generating process of y_t is Assumption 1.A.

3.2.3 Parameter Estimation

For an ith order Markov chain the log-likelihood function can be expressed as the sum of 2^i components, where each represents a particular outcome of $\{S_{t-1}, S_{t-2}, ..., S_{t-i}\}$; see the Appendix for details. Thus, we can maximize individually the part of the log-likelihood function which corresponds to S_{t-1}=S_{t-2}=,...,=$S_{t-i} = 0$, considering for observation t that,

$$\begin{aligned} \ln L_i &= \ln f(S_t|S_{t-1} = 0, ..., S_{t-i} = 0; \mathbf{x}_t; \boldsymbol{\beta}_i) \\ &= \delta_i\, S_t \ln\Big(1 - \Lambda(\mathbf{x}_t'\boldsymbol{\beta}_i)\Big) + \delta_i(1 - S_t)\ln\Big(\Lambda(\mathbf{x}_t'\boldsymbol{\beta}_i)\Big), \end{aligned} \quad (3.6)$$

where $f(.)$ is a conditional density function, and δ_i is an indicator function which is equal to one when $S_{t-1} = S_{t-1} =, ..., = S_{t-i} = 0$ and zero otherwise. Note that when $\delta_i = 1$ (3.6) corresponds to the conditional log-likelihood function of the well-known logit model (see, for instance, Hayashi; 2000).

Consistency and Asymptotic Normality of the Parameter Estimators

Since estimation of the transition probabilities and consequently of the ET to cross a threshold only depends on $\boldsymbol{\beta}_i$, $1 \leq i \leq r$, it is crucial that consistent estimates of these coefficient vectors are obtained.

Theorem 3.2.1 *Consistency of conditional QMLE without compactness*
Let $\{S_t, \mathbf{x}_t\}$ *be jointly stationary with conditional density* $f(S_t|S_{t-1} = 0, ..., S_{t-i} = 0; \mathbf{x}_t; \boldsymbol{\beta}_i)$ *and*

$$\hat{\boldsymbol{\beta}}_i = \underset{\boldsymbol{\beta}_i \in B_i}{argmax}\, \frac{1}{T}\sum_{t=1}^{T} \ln f(S_t|S_{t-1} = 0, ..., S_{t-i} = 0; \mathbf{x}_t; \boldsymbol{\beta}_i) \quad (3.7)$$

the quasi-ML estimator. Moreover, consider that,

(1) *the true parameter vector* $\boldsymbol{\beta}_i$ *is an element of the interior of a convex parameter space* $\boldsymbol{B}_i \subset \mathbb{R}^p$*, where* p *is the dimension of* $\boldsymbol{\beta}_i$*;*

(2) $\ln f(S_t|S_{t-1}=0,...,S_{t-i}=0;\mathbf{x}_t;\boldsymbol{\beta}_i)$ *is concave in* $\boldsymbol{\beta}_i$ *for all* $\{S_t,\mathbf{x}_t\}$ *and measurable for all* $\boldsymbol{\beta}_i$ *in* $\boldsymbol{B}_i$*;*

(3) $P[f(S_t|S_{t-1}=0,...,S_{t-i}=0;\mathbf{x}_t;\boldsymbol{\beta}_i) \neq f(S_t|S_{t-1}=0...=S_{t-i}=0;\mathbf{x}_t;\boldsymbol{\beta}_{i,0})] > 0$ *for all* $\boldsymbol{\beta}_i \neq \boldsymbol{\beta}_{i,0}$*;*

(4) $E[\,|\ln f(S_t|S_{t-1}=0,...,S_{t-i}=0;\mathbf{x}_t;\boldsymbol{\beta}_i)|\,]$ *exists and is finite for all* $\boldsymbol{\beta}_i$ *in* $\boldsymbol{B}_i$*.*

Then, as $T\to\infty$*,* $\hat{\boldsymbol{\beta}}_i$ *exists with probability 1 and* $\hat{\boldsymbol{\beta}}_i \xrightarrow{p} \boldsymbol{\beta}_i$ *.*

The first and second order derivatives of the logistic cumulative density function are, $\Lambda(v)' = \Lambda(v) - (1-\Lambda(v))$ and $\Lambda(v)'' = [1-2\Lambda(v)]\Lambda(v)[1-\Lambda(v)]$, respectively. Thus, the score and Hessian for observation t are, respectively,

$$s(\mathbf{w}_t;\boldsymbol{\beta}_i) = \frac{\partial \ln L}{\partial \boldsymbol{\beta}_i} = [S_t - \Lambda(\mathbf{x}_t'\boldsymbol{\beta}_i)]\mathbf{x}_t; \tag{3.8}$$

$$\boldsymbol{H}(\mathbf{w}_t;\boldsymbol{\beta}_i) = \frac{\partial s(\mathbf{w}_t;\boldsymbol{\beta}_i)}{\partial \boldsymbol{\beta}_i'} = -\Lambda(\mathbf{x}_t'\boldsymbol{\beta}_i)[1-\Lambda(\boldsymbol{x}_t'\boldsymbol{\beta}_i)]\mathbf{x}_t\mathbf{x}_t', \tag{3.9}$$

where $\mathbf{w}_t := (S_t, \mathbf{x}_t')'$. Since $\mathbf{x}_t\mathbf{x}_t'$ is positive definite, $\boldsymbol{H}(\mathbf{w}_t;\boldsymbol{\beta}_i)$ is negative semidefinite and the log-likelihood function is concave, therefore condition (2) of Theorem 1 holds. The last two conditions of Theorem 3.2.1 are satisfied under the non-singularity of $\mathrm{E}(\mathbf{x}_t\mathbf{x}_t')$ (see Appendix for details).

Theorem 3.2.2 *Asymptotic normality of conditional MLE*
Let $\mathbf{w}_t = (S_t,\mathbf{x}_t')'$ *be jointly stationary and* $\hat{\boldsymbol{\beta}}_i \xrightarrow{p} \boldsymbol{\beta}_i$*. In addition, consider that*

(1) $\boldsymbol{\beta}_i$ *is in the interior of* $\boldsymbol{B}_i$ *(identification);*

(2) $f(S_t|S_{t-1}=0...=S_{t-i}=0;\mathbf{x}_t;\boldsymbol{\beta}_i)$ *is twice continuously differentiable in* $\boldsymbol{\beta}_i$ *for all* $\mathbf{w}_t$*;*

(3) $E[s(\mathbf{w}_t;\boldsymbol{\beta}_{0,r})]=0$ *and* -$E[\boldsymbol{H}(\mathbf{w}_t;\boldsymbol{\beta}_r)]$ $=E[s(\mathbf{w}_t;\boldsymbol{\beta}_{0,r})s(\mathbf{w}_t;\boldsymbol{\beta}_{0,r})']$ *where* $s(\boldsymbol{w}_t;\boldsymbol{\beta}_i)$ *and* $\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\beta}_i)$ *are as defined in (3.8) and (3.9) (local dominance condition on the Hessian);*

(4) *for some neighborhood* $\mathscr{N}$ *of* $\boldsymbol{\beta}_i$*,*

$$E[\ \sup_{\boldsymbol{\beta}_i\in B} ||\boldsymbol{H}(\mathbf{w}_t;\boldsymbol{\beta}_i)||\ ] < \infty,$$

so that for any consistent estimator $\tilde{\boldsymbol{\beta}}_i$*,* $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{H}(\mathbf{w}_t;\tilde{\boldsymbol{\beta}}_i) \xrightarrow{p} E[\boldsymbol{H}(\mathbf{w}_t;\boldsymbol{\beta}_i)]$*;*

(5) $E[\boldsymbol{H}(\mathbf{w}_t;\boldsymbol{\beta}_i)]$ *is nonsingular.*

Thus, if conditions (1) - (5) hold, $\hat{\boldsymbol{\beta}}_i$ is asymptotically normal with

$$Avar\,(\hat{\boldsymbol{\beta}}_i) = \left(E[\boldsymbol{H}(\mathbf{w}_t;\boldsymbol{\beta}_i)]\right)^{-1}\,\boldsymbol{\Sigma}_i\,\left(E[\mathbf{H}(\mathbf{w}_t;\boldsymbol{\beta}_i)]\right)^{-1}$$

where $\boldsymbol{\Sigma}_i$ is the long-run variance of $\{s(\mathbf{w}_t;\boldsymbol{\beta}_i)\}$.

Assuming that $\hat{\boldsymbol{\Sigma}}_i$ is a consistent estimator of $\boldsymbol{\Sigma}_i$, then a consistent estimator of the asymptotic variance of $\hat{\boldsymbol{\beta}}_i$ is

$$\widehat{Avar\,(\hat{\boldsymbol{\beta}}_i)} = \left\{\frac{1}{T}\sum_{t=1}^{T}\mathbf{H}(\mathbf{w}_t;\hat{\boldsymbol{\beta}}_i)\right\}^{-1}\,\hat{\boldsymbol{\Sigma}}_i\,\left\{\frac{1}{T}\sum_{t=1}^{T}\mathbf{H}(\mathbf{w}_t;\hat{\boldsymbol{\beta}}_i)\right\}^{-1}. \qquad (3.10)$$

Theorem 3.2.3 *Considering the results of Theorems 3.2.1 and 3.2.2 it follows, as $T \to \infty$, that*

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) \xrightarrow{d} N(0, Avar\,(\hat{\boldsymbol{\beta}}_i)).$$

Theorem 3.2.4 *Let Assumption 1(B) hold. From application of the Delta method it follows that, as $T \to \infty$*

$$\sqrt{T}\left(p_i(\mathbf{x}) - \Lambda(\mathbf{x}'\boldsymbol{\beta}_i)\right) \xrightarrow{d} N\left(0, [\Lambda(\mathbf{x}'\boldsymbol{\beta}_i)]^2\,\mathbf{x}'\,\mathrm{Avar}\,(\hat{\boldsymbol{\beta}}_i)\,\mathbf{x}\right)$$

The positive Harris recurrence of $S_t|\mathbf{x}_t$ is crucial to ensure that the process moves from one state to another an infinite number of times as $T \to \infty$. This prevents, for example, from having too many zeros in the sequence of S_t (i.e., that y_t crosses z_1 too few times), which results in inaccurate estimates of $\boldsymbol{\beta}_i$ and $p_i(\mathbf{x}) \to 1$.

3.2.4 Covariate-Dependent ET

The covariate-dependent ET to cross z_1 when the process y_t starts at z_0 is,

$$E(T_{z_1}|\mathbf{x}) = \sum_{t=1}^{\infty} t\; P(T_{z_1} = t|\mathbf{x}). \qquad (3.11)$$

If S_t is a first order Markov Chain, i.e. $r = 1$, then $P(S_t = 0|S_{t-1} = 0;\mathbf{x}) = p_1(\mathbf{x})$ and

$$E(T_{z_1}|\mathbf{x}) = [1 - p_1(\mathbf{x})]\sum_{t=1}^{\infty} t\,p_1(\mathbf{x})^{t-1} = \Big[1 - p_1(\mathbf{x})\Big]^{-1}.$$

Theorem 3.2.5 *Let $\widehat{E(T_{z_1}|\mathbf{x})} = \Big[1 - \widehat{p_1(\mathbf{x})}\Big]^{-1}$. For $r = 1$,*

$\widehat{E(T_{z_1}|\mathbf{x})} \xrightarrow{p} E(T_{z_1}|\mathbf{x})$ *and*
$\sqrt{T}\left(\widehat{E(T_{z_1}|\mathbf{x})} - E(T_{z_1}|\mathbf{x})\right) \xrightarrow{d} N\left(0, \left[\mathbf{x}\,\exp(\mathbf{x}'\boldsymbol{\beta}_1)\right]'\mathrm{Avar}\left(\widehat{p_1(\mathbf{x})}\right)\left[\mathbf{x}\,\exp(\mathbf{x}'\boldsymbol{\beta}_1)\right]\right),$

where $Avar\left(\widehat{p_1(\mathbf{x})}\right) := [\Lambda(\mathbf{x}'\boldsymbol{\beta}_1)]^2\,\mathbf{x}'\,Avar\,(\hat{\boldsymbol{\beta}}_1)\,\mathbf{x}$, and $0 < p_1 < 1$.

Using (3.3) and (3.11) we have that

$$E(T_{z_1}|\mathbf{x}) = \sum_{t=1}^{r} t\,[1-p_t(\mathbf{x})]\prod_{j=1}^{t-1} p_j(\mathbf{x}) + \left\{[1-p_r(\mathbf{x})]\prod_{j=1}^{r-1} p_j(\mathbf{x})\right\}\sum_{t=r+1}^{\infty} t\; p_r(\mathbf{x})^{t-r}$$
$$= \sum_{t=1}^{r} t\,[1-p_t(\mathbf{x})]\prod_{j=1}^{t-1} p_j(\mathbf{x}) + \left\{[1-p_r(\mathbf{x})]\prod_{j=1}^{r-1} p_j(\mathbf{x})\right\}\frac{p_r(\mathbf{x})[1+r-r\,p_r(\mathbf{x})]}{[1-p_r(\mathbf{x})]^2}. \tag{3.12}$$

where, by convention, $\prod_{j=1}^{b} p_j(\mathbf{x}) = 1$ if $b < 1$.

By the continuous mapping theorem, if $\hat{\boldsymbol{\beta}}_i$ is consistent then $\widehat{E(T_{z_1}|\mathbf{x})}$ will also be a consistent estimator of $E(T_{z_1}|\mathbf{x})$ since it is a continuous function of $\hat{\boldsymbol{\beta}}_i$.

From (3.12) it follows that for $r \geq 1$ we have,

$$r = 1 \Rightarrow E(T_{z_1}|\mathbf{x}) = \frac{1}{1-p_1(\mathbf{x})},$$
$$r = 2 \Rightarrow E(T_{z_1}|\mathbf{x}) = \frac{1+p_1(\mathbf{x})-p_2(\mathbf{x})}{1-p_2(\mathbf{x})},$$
$$r = 3 \Rightarrow E(T_{z_1}|\mathbf{x}) = \frac{p_1(\mathbf{x})\big(p_2(\mathbf{x})+1\big)-p_3(\mathbf{x})\big(p_1(\mathbf{x})+1\big)+1}{1-p_3(\mathbf{x})},\ \text{etc.}$$

Therefore, as already stated, it is critical to have consistent estimates of $\boldsymbol{\beta}_i$, $1 \leq i \leq r$. In small samples, S_t may not move from one state to another a sufficient number of times when r is relatively large and estimating these parameters may be problematic. In practice, the choice of r depends on the sample size, the level of persistence, the starting point z_0 and the threshold z_1. We suggest that r is chosen based on two indicators: the residuals of the regression of y_t on its own r lags and $\mathbf{x}_t$, and the statistical significance of $\boldsymbol{\beta}_r$. It is also possible to estimate r using some information criteria such as BIC (Bayesian information criteria); see, for instance, Katz (1981) and Raftery (1985). However, this approach is cumbersome, since it requires estimating the entire transition probability matrix for several Markov chains of different orders, while we are only interested in the probability in (3.4).

As is evident from (3.12), an exact asymptotic expression for the distribution of $\widehat{E(T_{z_1}|\mathbf{x})}$ is difficult to obtain since it is a complex non-linear function of $\hat{\boldsymbol{\beta}}_i$. However, advances in computing have made resampling techniques, in particular bootstrapping approaches, a valuable tool for the estimation of standard errors and for the construction of confidence intervals.

In this work, suitable bootstrap methods, which allow for serial dependence, are applied. Many different bootstrap techniques for dependent data have been proposed (see, for instance, MacKinnon; 2007, Section 6 for a brief overview). A widely used approach in this context is the block bootstrap algorithm (Härdle et al.; 2003). The block bootstrap consists in dividing the time series into several blocks of b consecutive observations in order to preserve the original structure within a block, and to re-sample the blocks, which may be overlapping or non-overlapping and of fixed or of variable length, as in e.g. the stationary block bootstrap proposed by Politis and Romano (1994).

Lahiri (2003, Chapter 5) compares the performance of four block bootstrap approaches[7] and shows that, in terms of their MSEs, the overlapping block bootstrap outperforms the non-overlapping and the stationary block bootstrap procedures. This conclusion is valid if the block length increases as the sample size T increases at a rate not slower than the optimal rate $\kappa T^{1/3}$, where κ is constant.

Thus, in what follows we will employ the overlapping block bootstrap, also known as "blocks of blocks" bootstrap, proposed by Politis and Romano (1992a).

Defining $\boldsymbol{Z}_t \equiv (y_t, \mathbf{x}_t)$, we construct $T-b+1$ overlapping blocks as

$$\boldsymbol{Z}_1, ..., \boldsymbol{Z}_b, \quad \boldsymbol{Z}_2, ..., \boldsymbol{Z}_{b+1}, ..., \boldsymbol{Z}_{T-b+1}, ..., \boldsymbol{Z}_T, \tag{3.13}$$

which are re-sampled in the usual way, using an iid random variable on $\{1, 2, ..., T-b+1\}$. The block bootstrap algorithm consists of the following steps:

- *Step 1*: Choose the block length b. In the empirical application, we opt for $b = T^{1/3}$;
- *Step 2*: Resample the blocks as illustrated in (3.13) and generate the bootstrap sample $(y_t^*, \mathbf{x}_t^*)$;
- *Step 3*: Build the process S_t^* in (3.1) using y_t^* and estimate the covariate-dependent probabilities in (3.4) and (3.5) ;
- *Step 4*: Compute $\widehat{E(T_{z_1}|\mathbf{x})}^*$.
- *Step 5*: Repeat Steps 1 to 4 a B number of times, where B is the number of bootstrap simulations, and compute the empirical distribution of $\widehat{E(T_{z_1}|\mathbf{x})}^*$ and respective confidence intervals.

3.3 Monte Carlo Analysis

This section investigates the finite sample properties of the parameter estimates $\hat{\boldsymbol{\beta}}_r$. We generate the $S_t|\mathbf{x}_t$ process by simulating two-state Markov chains of orders $r = 1, 2, ..., 5$.

In order to simplify the simulation exercise but without loss of generality we make some simplifying assumptions about the data generation process (DGP) of $S_t|\mathbf{x}_t$. In specific, we assume that $p_r(\mathbf{x})$ is covariate-dependent and the remaining probabilities are constant and equal to 0.5. In practice, one would expect that all transition probabilities depend on covariates. As stated in Subsection 3.2.3, these assumptions have no effect on the consistency of $\hat{\boldsymbol{\beta}}_r$ since the part of the log-likelihood which corresponds to $S_{t-1} = 0, ..., S_{t-r} = 0$ is maximized individually.

[7] In addition to the overlapping, non-overlapping and stationary block bootstraps, Lahiri (2003) also considers the circular block bootstrap proposed by Politis and Romano (1992b).

As an illustration, consider a second order ($r = 2$) Markov chain. In this case, the transition probabilities matrix is completely defined by the probabilities

- $p_2(\mathbf{x}) = P(S_t = 0|S_{t-1} = 0, S_{t-2} = 0; \mathbf{x}) = \Lambda(\mathbf{x}'\boldsymbol{\beta}_2)$,
- $P(S_t = 0|S_{t-1} = 0, S_{t-2} = 1; \mathbf{x}) = 0.5$,
- $P(S_t = 0|S_{t-1} = 1, S_{t-2} = 0; \mathbf{x}) = 0.5$,
- $P(S_t = 0|S_{t-1} = 1, S_{t-2} = 1; \mathbf{x}) = 0.5$,

and our interest centers exclusively on $p_2(\mathbf{x})$, which can be estimated by maximizing the log-likelihood function in (3.6) for $i = r = 2$. However, it is noteworthy that, since the DGP is a second order Markov chain, $p_1(\mathbf{x}) = P(S_t|S_{t-1}; \mathbf{x})$, also needed to compute the ET, will depend on the first two probabilities presented above, that is, on $p_2(\mathbf{x})$ and $P(S_t = 0|S_{t-1} = 0, S_{t-2} = 1; \mathbf{x})$.

The DGP also considers that $\mathbf{x}_t := \big(1, x_{2t}\big)'$ and $\boldsymbol{\beta}_2 = \big(\beta_{2,1}, \beta_{2,2}\big)'$ in (3.4), where $x_{2t} \sim N(0, 1)$. Therefore, as $T \to \infty$,

$$
\begin{aligned}
\frac{\sum_{t=1}^{T} p_2(\mathbf{x}_t)}{T} &\xrightarrow{p} E\big(p_2(\mathbf{x}_t)\big) \\
&= \int_{-\infty}^{+\infty} P(S_t = 0|S_{t-1}, ..., S_{t-r} = 0; \mathbf{x}_t) f(x_{2t}) dx_{2t} \\
&= \int_{-\infty}^{+\infty} \frac{\exp^{\mathbf{x}_t'\boldsymbol{\beta}_2}}{1 + \exp^{\mathbf{x}_t'\boldsymbol{\beta}_2}} \frac{1}{\sqrt{2\pi}} \exp^{-\frac{x_{2t}^2}{2}} dx_{2t}.
\end{aligned}
$$

We will investigate two cases, $\boldsymbol{\beta}_2 = (0.0, 3)'$ and $\boldsymbol{\beta}_2 = (2.3, 3)'$ which imply that $\frac{\sum_{t=1}^{T} p_r(\mathbf{x}_t)}{T} \to 0.5$ and $\frac{\sum_{t=1}^{T} p_r(\mathbf{x}_t)}{T} \to 0.75$, respectively. The second case is particularly relevant since we are interested in $z_1 = \bar{y}$ and the macroeconomic variables tend to exhibit some persistence (or slow mean reversion after a shock), which results in higher values of $E(T_{z_1}|\mathbf{x})$.

As the order of the Markov chain is unknown in practice, for each Markov chain of order r generated in the simulations, we estimate first up to fifth order Markov chains.

Table 3.1 summarizes the Monte Carlo results for $T \in \{500, 1000, 2000\}$. When $i \geq r$, the $\boldsymbol{\beta}_2$ parameters seem to be consistently estimated even for $T = 500$. As expected, the mean of the parameter estimates is closer to the "true" parameter values and the standard deviation of the estimates reduces in all cases when the sample size increases. Although MLE seems to produce consistent estimates of $\boldsymbol{\beta}_2$ even when Markov chains of higher order than the one considered in the DGP is estimated ($i > r$), these estimates are less accurate since

$$p_{r-1}(\mathbf{x})p_{r-2}(\mathbf{x})...p_1(\mathbf{x}) = P(S_{t-1} = 0, ..., S_{t-r} = 0|\mathbf{x})$$

lowers as r increases and there are less cases with $\{S_{t-1} = 0, ..., S_{t-r} = 0\}$. Thus, $\boldsymbol{\beta}_r$ will be estimated using a smaller number of observations, since $\delta_i = 1$ in expression (3.6) occurs less often.

TABLE 3.1: Means and SD of Markov chain parameter estimates

		$T = 500$				$T = 1000$				$T = 2000$			
		$\beta_{i,1} = 0.0$		$\beta_{i,2} = 3.0$		$\beta_{i,1} = 0.0$		$\beta_{i,2} = 3.0$		$\beta_{i,1} = 0.0$		$\beta_{i,1} = 3.0$	
r	i	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
	1	-0.005	0.196	3.076	0.407	-0.002	0.137	3.035	0.278	0.001	0.099	3.019	0.193
	2	-0.014	0.284	3.157	0.621	-0.008	0.196	3.077	0.413	-0.001	0.138	3.037	0.281
1	3	-0.032	0.448	3.362	1.043	-0.017	0.283	3.160	0.629	-0.007	0.197	3.076	0.411
	4	-0.054	0.673	3.541	1.391	-0.031	0.434	3.332	0.998	-0.018	0.289	3.158	0.630
	5	0.008	0.799	3.202	1.382	-0.033	0.624	3.353	1.178	-0.036	0.453	3.365	1.114
	1	0.008	0.137	0.800	0.169	0.004	0.098	0.791	0.118	0.003	0.069	0.786	0.083
	2	-0.010	0.287	3.169	0.627	-0.006	0.197	3.079	0.410	-0.003	0.137	3.036	0.279
2	3	-0.023	0.441	3.353	1.030	-0.015	0.288	3.169	0.636	-0.007	0.197	3.075	0.406
	4	-0.039	0.667	3.507	1.371	-0.031	0.448	3.354	1.013	-0.014	0.289	3.164	0.623
	5	0.007	0.812	3.225	1.397	-0.029	0.622	3.358	1.165	-0.035	0.452	3.390	1.138
	1	0.015	0.127	0.366	0.142	0.007	0.091	0.359	0.100	0.005	0.066	0.358	0.071
	2	0.008	0.194	0.815	0.246	0.003	0.136	0.797	0.168	0.003	0.097	0.791	0.118
3	3	-0.026	0.439	3.354	1.037	-0.014	0.287	3.158	0.624	-0.008	0.193	3.071	0.403
	4	-0.039	0.647	3.520	1.380	-0.030	0.440	3.331	0.990	-0.015	0.282	3.158	0.621
	5	0.008	0.802	3.220	1.384	-0.033	0.617	3.342	1.179	-0.039	0.446	3.371	1.095
	1	0.020	0.128	0.186	0.137	0.010	0.089	0.179	0.094	0.006	0.065	0.175	0.066
	2	0.022	0.184	0.381	0.207	0.011	0.127	0.364	0.141	0.005	0.092	0.360	0.099
4	3	0.016	0.281	0.851	0.360	0.010	0.192	0.812	0.243	0.003	0.135	0.799	0.167
	4	-0.042	0.659	3.504	1.369	-0.028	0.441	3.336	0.984	-0.018	0.281	3.165	0.622
	5	-0.007	0.798	3.250	1.402	-0.031	0.620	3.374	1.167	-0.045	0.455	3.377	1.119
	1	0.029	0.127	0.097	0.133	0.016	0.090	0.093	0.092	0.009	0.065	0.091	0.064
	2	0.036	0.179	0.195	0.197	0.021	0.127	0.184	0.135	0.009	0.091	0.179	0.093
5	3	0.042	0.260	0.398	0.305	0.023	0.185	0.374	0.204	0.008	0.129	0.364	0.140
	4	0.026	0.424	0.921	0.609	0.017	0.277	0.844	0.358	0.004	0.192	0.817	0.247
	5	0.004	0.791	3.295	1.404	-0.030	0.601	3.353	1.169	-0.041	0.445	3.363	1.093
		$\beta_{i,1} = 2.3$		$\beta_{i,2} = 3.0$		$\beta_{i,1} = 2.3$		$\beta_{i,2} = 3.0$		$\beta_{i,1} = 2.3$		$\beta_{i,1} = 3.0$	
	1	2.340	0.289	3.067	0.384	2.322	0.201	3.035	0.263	2.312	0.139	3.017	0.185
	2	2.351	0.342	3.090	0.458	2.327	0.236	3.047	0.310	2.314	0.162	3.023	0.216
1	3	2.370	0.409	3.129	0.553	2.335	0.277	3.064	0.365	2.319	0.189	3.031	0.254
	4	2.397	0.499	3.179	0.686	2.345	0.325	3.084	0.429	2.325	0.222	3.041	0.295
	5	2.430	0.609	3.247	0.846	2.363	0.387	3.119	0.515	2.334	0.259	3.057	0.346
	1	0.824	0.195	0.993	0.185	0.816	0.136	0.980	0.129	0.812	0.096	0.976	0.091
	2	2.358	0.389	3.109	0.525	2.330	0.264	3.055	0.350	2.317	0.183	3.029	0.243
2	3	2.385	0.470	3.162	0.646	2.341	0.310	3.075	0.412	2.323	0.215	3.039	0.284
	4	2.420	0.580	3.233	0.812	2.358	0.369	3.105	0.487	2.330	0.252	3.052	0.332
	5	2.470	0.720	3.331	1.000	2.379	0.446	3.147	0.595	2.338	0.296	3.071	0.391
	1	0.433	0.178	0.492	0.149	0.423	0.126	0.483	0.103	0.422	0.088	0.483	0.073
	2	0.828	0.263	1.005	0.250	0.818	0.186	0.987	0.176	0.816	0.13	0.981	0.123
3	3	2.416	0.566	3.215	0.778	2.357	0.369	3.105	0.488	2.328	0.252	3.049	0.332
	4	2.460	0.689	3.301	0.963	2.383	0.444	3.152	0.591	2.338	0.295	3.068	0.391
	5	2.501	0.814	3.389	1.117	2.410	0.541	3.208	0.727	2.348	0.345	3.093	0.462
	1	0.247	0.168	0.263	0.136	0.236	0.119	0.256	0.098	0.228	0.084	0.253	0.068
	2	0.444	0.250	0.507	0.207	0.434	0.175	0.495	0.147	0.423	0.124	0.486	0.101
4	3	0.849	0.378	1.052	0.367	0.833	0.261	1.012	0.247	0.817	0.182	0.989	0.168
	4	2.506	0.838	3.409	1.159	2.414	0.548	3.211	0.740	2.351	0.361	3.101	0.482
	5	2.515	0.934	3.452	1.265	2.444	0.650	3.277	0.881	2.369	0.433	3.140	0.583
	1	0.151	0.156	0.144	0.133	0.136	0.111	0.138	0.094	0.126	0.078	0.133	0.066
	2	0.267	0.231	0.277	0.195	0.247	0.166	0.263	0.135	0.232	0.118	0.256	0.095
5	3	0.471	0.345	0.539	0.301	0.445	0.244	0.507	0.203	0.426	0.174	0.491	0.141
	4	0.894	0.553	1.142	0.567	0.850	0.371	1.048	0.353	0.821	0.253	1.004	0.234
	5	2.469	1.004	3.379	1.281	2.477	0.764	3.341	1.007	2.392	0.530	3.200	0.722

Notes: r refers to the Markov chain order considered in the data generating process and $i = 1, .., 5$ is the order of the Markov chain used to estimate $\boldsymbol{\beta}_i$. All results presented are based on 10000 Monte Carlo simulations.

3.4 Empirical Application

It is nowadays widely accepted that the relation between major economic variables is nonlinear (see, for instance, Terasvirta et al.; 2010). However, since most existing nonlinear models assume a parametric functional form and require the estimation of a considerable numbers of parameters, alternative ways to capture some aspects of the nonlinear relationships have been developed. For instance, in order to cater for the possibility that the yield curve predicts more accurately when drastic changes in output occur, some authors considered a binary dependent variable which equals one when the National Bureau of Economic Research (NBER) dates a recessions and zero otherwise (see, for instance, Estrella and Hardouvelis; 1991 and Estrella and Mishkin; 1998). Then, discrete choice models, such as logit or probit, are employed to estimate the effect of YS on the probability of a recession.

Most standard models for binary responses do not take the dynamic structure of the data into account, which is crucial for applications to time series data. A relevant exception is the parametric (linear) dynamic probit model which includes lags of the binary response variable in the probit function; see e.g. Kauppi and Saikkonen (2008) and Antunes et al. (2018).

An alternative approach to incorporate information provided by the past values of the dependent variable is to assume the Markovian property. In this context, the interpretation is in terms of transitional rather than marginal probabilities (see, for instance, Azzalini; 1994) and the focus is on the estimation of the probabilities of transitions between states. The framework that we introduce in this paper, which also relies on the Markov assumption, allows us to obtain covariate-dependent transition probabilities without requiring a rigid parametric functional form[8] or the estimation of a large number of parameters. Moreover, instead of simply indicating the presence or not of a recession, the binary variable is given by (3.1). Since the proposed approach allow us to consider different threshold z_1 and starting z_0 values, it may be useful to capture additional information about possible nonlinear relationships between a dependent variable and a set of covariates.

However, the major advantage of our approach is that it provides a simple method to estimate the covariate-dependent expected time (ET) to cross a threshold, which may be a useful reduced-form tool to investigate relevant topics such as dynamic controllability. Roughly speaking, Buiter and Gersovitz (1981) define that a system is dynamically controllable if a path for the economic instruments exists which is capable of moving the vector with the economic objectives from any initial value to any other target value in pre-assigned finite time. They argue that this ability to achieve a vector of target values is relevant for economic policy even if these target values cannot be maintained. Thus, by choosing a target value z_1 and calculating covariate-dependent ET for different starting points z_0 we may gather evidence about the effectiveness of the covariates in driving a dependent variable towards z_1.

[8]The only parametric assumption is that the covariates influence the transition probabilities via a logistic function, as defined in (3.4) and (3.5).

This section provides an empirical application to the economic activity-yield spread relationship. The link between YS and economic growth is related to monetary policy, which influences the shape of the yield curve - the representation of several yields or interest rates across different contract lengths - over the business cycles. For instance, monetary policy influences directly, through the use of open market operations, the level of short-term maturity yields. Central banks have a reference interest rate as an important pro-cyclical instrument to pursue their objectives (e.g., price stability). They will lower short run yields in recessions in an attempt to stimulate the economy and will do the opposite when there are inflationary pressures. However, as short-term interest rates have been close to their zero lower bound in recent years, central banks also began to influence long-term interest rates using unconventional monetary policy operations such as, quantitative easing, in order to stimulate aggregate demand and avoid a scenario of low growth and deflation.

Therefore, central banks' monetary policies exert a strong influence on YS. When the economy is in recession, monetary policy actions that have led YS to positive values will promote a faster economic recovery. On the other hand, if the economy is in expansion and the inflation rate suggests that we are facing an over heated economy, monetary policy actions can be taken in order to reduce the YS and slow down economic growth; for instance, by increasing the reference interest rates.

We consider YS as a proxy for the monetary policy stance and investigate how this variable affects the ET that IP growth rate takes to return to its mean after an exogenous shock. In practice, this will be done considering several starting values z_0, each of which correspond to a different S_t process, and by estimating the vector of parameters $\boldsymbol{\beta}_i$ that indicates how YS influences the ET for IP growth rate to return to its mean.

3.4.1 Data

The proposed methodology is applied to U.S. data. The IP index was selected as an indicator of economic activity due to its higher (monthly) frequency and faster availability relatively to GDP, the most commonly used measure of economic activity. IP seems an adequate choice since, at least for the more industrialized countries, the value added by industrial production represents a substantial share of GDP. Moreover, the IP index exhibits more cyclical fluctuations than the financial index. It is expected that the larger number of observations available and the greater cyclical variability of the IP index will have a positive impact on parameter estimation. We consider the monthly seasonally adjusted U.S. IP index obtained from OECD's Main Economic Indicators Publication, for the period from December 1964 to February 2019 (651 observations).

YS, which is the difference between the long-term[9] and the short-term interest rates[10], used as a proxy for the monetary policy stance, is obtained from

[9]Long-term interest rates are computed using government securities with outstanding maturities of 10 years.

[10]Short-term interest rates are either the three month interbank offer rate associated to loans provided and taken among banks for any excess or shortage of liquidity over several

OECD's Monthly Monetary and Financial Statistics. Figure 3.1 graphically presents both time series.

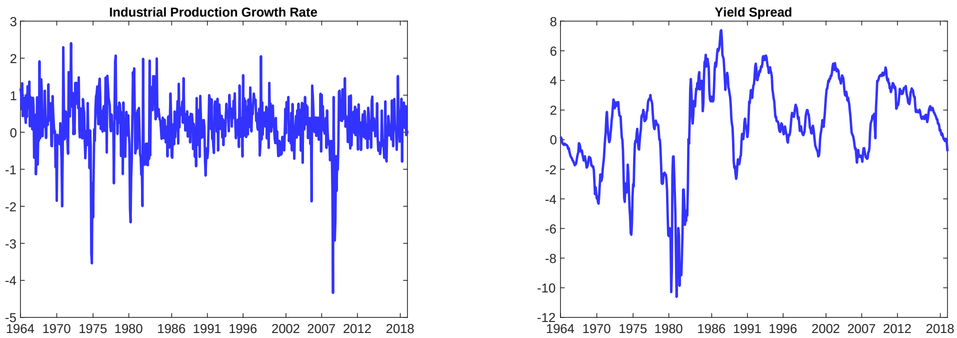


FIGURE 3.1: U.S. Industrial Production (monthly) growth rate and yield spread

3.4.2 Empirical Results

Let y_t be the monthly IP growth rate and x_t YS. We consider the empirical mean of the IP growth rate as the threshold, $z_1 = \overline{y}$, and estimate $\boldsymbol{\beta}_i$ $(1 \leq i \leq r)$ for the covariate-dependent probabilities defined in (3.4) and (3.5) considering several starting values z_0, each corresponding to a different S_t process as defined in (3.1). The $\boldsymbol{\beta}_i$ parameters are crucial to properly estimate the impact of the covariates on the ET for y_t to cross z_1 when it starts from z_0; see (3.12).

YS has been identified as an important leading indicator in the literature. For instance, Estrella and Mishkin (1998) conclude that the steepness of the yield curve is an accurate predictor of real activity, especially between two and six quarters ahead. Thus, we estimated the probabilities in (3.4) and (3.5) using k-periods lagged YS as covariates, with $k = 1, ..., 6$, considering Markov chains of different orders and starting values z_0. Since, overall the 3-month lagged YS provides stronger statistical evidence, we will use this variable as a covariate. In other words, we will consider that $\mathbf{x}_t = (1, x_{t-3})'$ and $\boldsymbol{\beta}_i = (\beta_{i,1}, \beta_{i,2})$, with $i = 1, ..., r$.

The Markov chain order r was chosen based on the analysis of the residuals of a regression of y_t on its own lags and YS. Additionally, we also take into account the statistical significance of $\beta_{i,1}$ and $\beta_{i,2}$ for several starting values z_0. Thus, a third order Markov process (i.e. $r = 3$) has been considered. For $r > 3$, $\hat{\beta}_{r,2}$ is not statistically significant and the standard error of $\hat{\beta}_{r,1}$ also becomes substantially higher for almost all starting values z_0 considered.

months or the rate associated with Treasury bills, certificates of deposit or comparable instruments, each of three month maturity.

TABLE 3.2: Estimated parameters for $p_i(\mathbf{x})$, $i = 1, ..., r$

		$i=1$				$i=2$				$i=3$			
		$\hat{\beta}_{1,1}$		$\hat{\beta}_{1,2}$		$\hat{\beta}_{2,1}$		$\hat{\beta}_{2,2}$		$\hat{\beta}_{3,1}$		$\hat{\beta}_{3,2}$	
α	z_0	coef.	prob.	coef.	prob.	coef.	prob.	coef.	prob.	coef.	prob.	coef.	prob.
-1.50	-0.886	0.761	0.011	-0.171	0.049	0.957	0.008	-0.176	0.072	0.998	0.013	-0.131	0.162
-1.25	-0.705	0.960	0.000	-0.183	0.018	1.193	0.000	-0.199	0.033	1.019	0.003	-0.225	0.033
-1.00	-0.524	0.755	0.000	-0.167	0.006	0.944	0.000	-0.118	0.068	0.903	0.001	-0.201	0.027
-0.75	-0.343	0.485	0.003	-0.170	0.001	0.811	0.000	-0.137	0.031	0.964	0.000	-0.190	0.033
-0.50	-0.162	0.509	0.000	-0.148	0.001	0.833	0.000	-0.133	0.018	1.026	0.000	-0.191	0.014
-0.25	0.020	0.319	0.007	-0.101	0.000	0.864	0.000	-0.140	0.009	0.996	0.000	-0.211	0.005
0.25	0.382	0.177	0.102	0.102	0.015	0.282	0.071	0.132	0.020	0.231	0.192	0.254	0.009
0.50	0.563	0.448	0.002	0.061	0.125	0.489	0.006	0.152	0.015	0.438	0.045	0.270	0.003
0.75	0.744	0.595	0.000	0.048	0.227	0.633	0.002	0.138	0.050	0.490	0.032	0.209	0.020
1.00	0.925	0.809	0.000	0.073	0.191	0.742	0.001	0.178	0.042	0.812	0.003	0.184	0.063
1.25	1.106	1.051	0.000	0.110	0.154	0.972	0.000	0.167	0.089	1.092	0.001	0.134	0.167
1.50	1.287	0.958	0.000	0.042	0.363	0.955	0.002	0.147	0.157	0.991	0.005	0.125	0.219

Notes: $p_i(\mathbf{x}) := P(S_t = 0|S_{t-1} = 0, ..., S_{t-i} = 0; \mathbf{x})$ as defined in (3.5) and $z_0 = \overline{y} + \alpha\hat{\sigma}_y$ ($\overline{y}$ and $\hat{\sigma}_y$ are, respectively, the sample mean and standard deviation of y_t).

Table 3.2 shows that the coefficients of YS ($\hat{\beta}_{i2}, i = 1, 2, 3$) are negative when $z_0 < \overline{y}$ and positive when $z_0 > \overline{y}$. For the first case the YS coefficients are statistically significant at the 10% significance level for almost all starting values considered. When $z_0 > z_1$, significance is only observed for $i = 2, 3$ and if the starting values are not particularly large.

We computed the unconditional (proposed by Nicolau; 2017) and covariate dependent ET time curves[11] (ETC) and their 95% confidence intervals using the overlapping block bootstrap described in Subsection 3.2.4 with 999 bootstrap replications and block length equal to $T^{1/3}$.

Figure 3.2 shows the estimated (unconditional and covariate dependent) ETCs and their 95% confidence bounds considering 24 different starting points z_0, equally spaced in the interval $\left(\overline{y} - 1.5\hat{\sigma}_y, \overline{y} + 1.5\hat{\sigma}_y\right)$, with $\overline{y}$ and $\hat{\sigma}_y$ the sample mean and standard deviation of y_t, respectively, and six different values for the explanatory variable z_{t-3}.

The bootstrap-based ETC estimates presented in Figure 3.2 seem to confirm the results in Table 3.2. That is, when the IP growth rate is below its mean, $z_0 < \bar{y}$, the β_{i2} estimates are negative. As a consequence, the ET to reach $z_1 \geq \bar{y}$ is relatively low if YS is positive. Therefore, the IP growth rate easily recovers from low values when YS is positive; however, recovery is slower when YS is negative. For example, consider the case where YS is equal to 3.9% (see Figure 3.2 panel F). If the initial value of the IP growth rate is negative, recovery is fast and takes around 2 months on average to reach $\bar{y}$. On the other hand, if YS is negative and equal to -1.9%, recovery is much slower and can take on average four months (see Figure 3.2 panel A).

[11] As in Nicolau (2017), we call ET curve to the graphical representation of the ET estimates for different starting values z_0, but same threshold $z_1 = \bar{y}$.

When the IP growth rate is above its mean, $z_0 > \bar{y}$, the estimates of β_{i2} become positive and the opposite interpretation applies. Thus, in this case, the ET to decrease to $z_1 \leq \bar{y}$ is delayed if YS is positive (prosperity tends to last) and accelerated if YS is negative. For example, if the YS is equal to 3.9% (see Figure 3.2 panel F), the IP growth rate will remain above its mean value for about 4 months while the same only happens for two months on average if YS is negative and equal to -1.9% (see Figure 3.2 panel A).

Hence, we can conclude that there is statistical evidence that YS influences the ET the IP growth rate takes to return to its mean, which is graphically reflected in the asymmetric shapes of the conditional ET curves for high and low YS values. For instance, panels A, E and F of Figure 3.2 illustrate this feature particularly well.

The relationship between YS and economic activity has been investigated by an extensive literature. Harvey (1989), Stock and Watson (1989), Estrella and Hardouvelis (1991) and Estrella and Mishkin (1998), among others, found statistical evidence that YS predicts future output growth. Most of this research is based on a linear framework of analysis (OLS regressions), considering an appropriate lead-lag relationship. However, there are some exceptions to this practice. For instance, Galbraith and Tkacz (2000) used the linearity tests against TAR models suggested by Hansen (1996) and found evidence in support of the asymmetric impact of YS on the conditional expectation of output growth. When YS is above a specific threshold value, the additional effect of a large positive spread becomes small and statistically insignificant.

Moreover, YS has also been successfully used in predicting recessions (see, for instance, Estrella and Hardouvelis; 1991, Estrella and Mishkin; 1998 and Kauppi and Saikkonen; 2008). Most of this evidence was obtained using probit or logit models where the dependent variable used was a recession indicator, which equals 1 when the NBER dates a recession and zero otherwise.

The proposed methodology is somehow related with this strand of the literature since it considers a logit specification for the transition probabilities in (3.4) and (3.5). However, it is much more flexible and can provide additional information about the economic activity-yield spread relationship. We considered $z_1 = \bar{y}$ in order to illustrate the proposed approach, but other threshold values such as $z_1 = 0$ could also have been considered. With $z_1 = \bar{y}$, the covariate-dependent ET curve may provide important insights about the predictive content of YS. For instance, let us consider that $y_{t-1} > \bar{y}$. If YS is positively related to future economic activity, a negative YS increases the probability that the dependent variable will cross $\bar{y}$ in period t and consequently ET will be lower.

A visual informal analysis of the ET curves presented in Figure 3.2 suggests some interesting facts. First, for $z_0 > \bar{y}$, the ET estimates change very little when YS increases from -1.9 to -0.6 (see panels A and B) and from 0 to 1 (see panels C and D). It seems that a large YS value is necessary to increase the ET substantially (see panels E and F). On the other hand, if $z_0 < \bar{y}$, the decrease in the ET estimates when YS increases seems clearer.

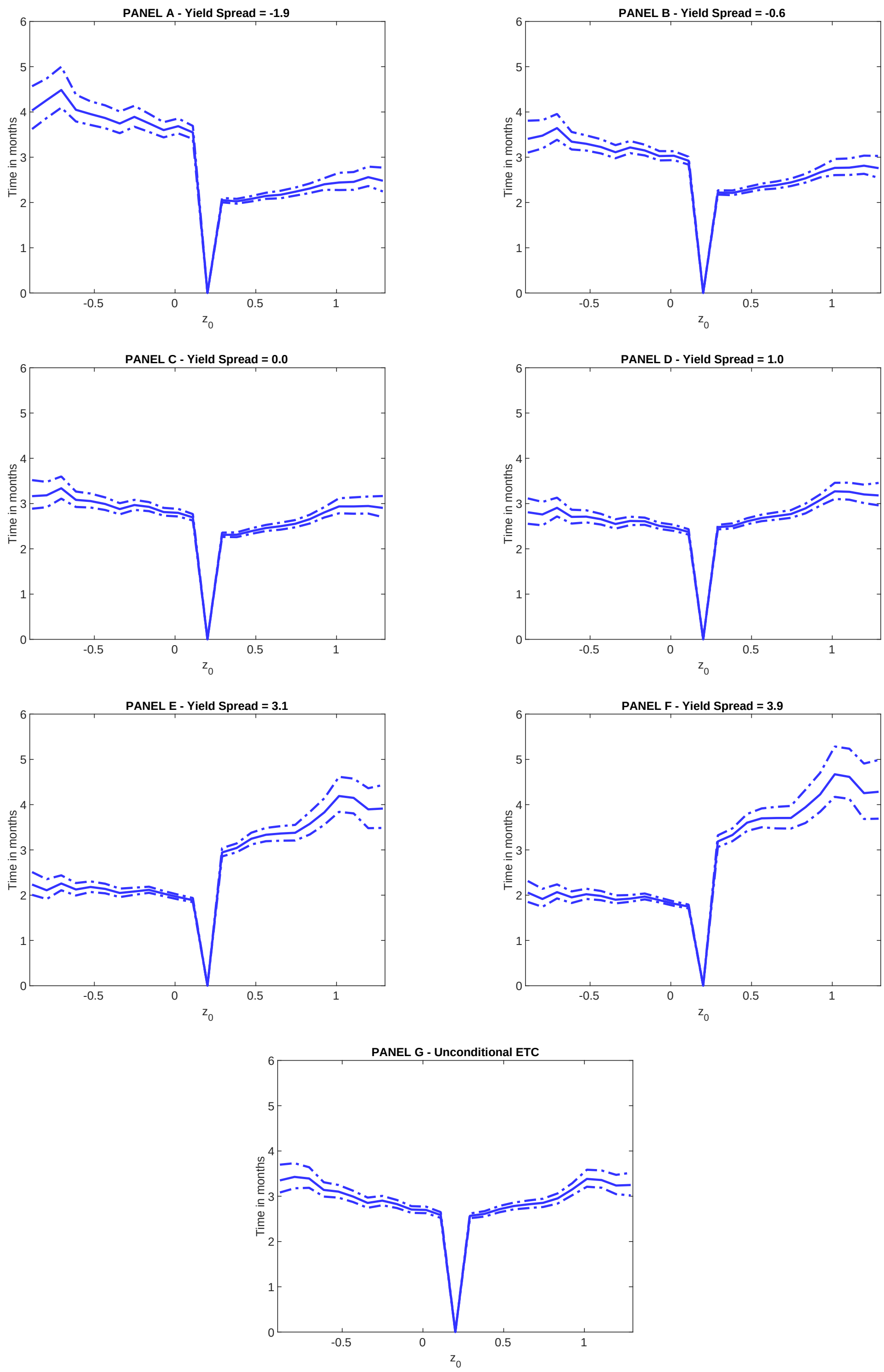


FIGURE 3.2: Unconditional and conditional estimated expected time curves (ETC) with overlapping block bootstrap 95% CI

Notes: For the conditional case the following six values for the covariate were considered: $(\bar{x} - \hat{\sigma}_x) = -1.9$, $Q_1 = -0.6$, 0 , $\bar{x}$=1 , $Q_3 = 3.1$, $(\bar{x} - \hat{\sigma}_x) = 3.9$; where Q_1 and Q_3 are the first and third quartiles of x_t, respectively.

3.5 Conclusions

In this paper we propose a simple and easy to implement approach to investigate the effect of covariates on the expected time (ET) to cross a threshold given a specific starting point. In order to estimate the parameters that describe the relationship between the ET and the covariates, we adapt the procedure to estimate Markov models of any order proposed by Islam and Chowdhury (2006). We confirm via Monte Carlo simulations that the relevant parameters, $\boldsymbol{\beta}_r$, are consistently estimated even when the sample size is relatively small ($T = 500$).

However, since the expression for ET in (3.12) is a highly nonlinear function of $\boldsymbol{\beta}_i$, with $i = 1, .., r$, we consider an overlapping block-bootstrap procedure to obtain the standard errors of the ET estimates and to construct relevant confidence intervals. Existing literature on the topic suggests that this block-bootstrap variant is a good choice to resample dependent data. We used it to obtain confidence intervals for the ET that the U.S. IP growth rate takes to revert to its mean given a starting point and a particular YS value. Figure 3.2 shows that the width of the confidence intervals is relatively narrow even when the starting value z_0 is far (in absolute value) from the threshold value z_1 (which results in less accurate estimates of $\boldsymbol{\beta}_i$).

The empirical application to the U.S. economy shows that there is statistical evidence supporting the influence of YS on the ET for the IP growth rate to return to its mean. Namely, a high YS reduces this ET if the IP growth rate starts by assuming below average values. Thus, monetary policies that result in higher YS could play an important role in stimulating a faster return to desirable growth rates in periods of weak growth or contraction. Moreover, the YS value seems also critical when the IP growth rate is larger than average. If YS is negative, the IP growth rate will return quickly to below average values. This finding may be related to the widely documented ability of the yield curve inversion (negative YS) to predict recessions (see, for instance, Estrella and Mishkin; 1998).

The application of the proposed methodology to the economic activity-yield spread relationship illustrates that it provides insights that may be relevant to policymakers. Thus, the proposed approach can be adapted to support a wide range of economic decisions since it provides a flexible and easy to implement framework that allows us to infer about the nonlinear relationship between a dependent variable associated with an economic objective and a set of relevant covariates associated with economic policy instruments.

Chapter 4

The Importance of Intra-Horizon Risk in Portfolio Optimization

4.1 Introduction

Investors always face a trade-off between risk and return when commit capital to the financial market. Thus, the portfolio optimization theory is crucial to support efficient asset allocation decisions.

One of the most influential methodologies dealing with this problem was proposed by Markowitz (1952, 1959). Markowitz's approach associates profits and risk to the expectation and variance of returns, giving rise to the mean-variance (MV) portfolio optimization models[1] and to the concept of efficient frontier, the curve that represents the efficient portfolio for a given risk level.

Despite of being a common workhorse model for portfolio selection, the traditional MV approach has not been free of criticism. One of its disadvantages is the fact that it is a one-period model which does not accommodate uncertainty, since the estimated expected returns and respective covariance matrix are constant over the investment horizon considered and, therefore, no portfolio re-allocation is possible when relevant information arises. Extensions of Markowitz' MV approach to a multi-period framework were proposed by, for instance, Merton (1971) and Li and Ng (2000).

Another important drawback of the mean-variance approach is that it only focuses on the first two statistical moments. Since non-normality is a stylized feature of asset returns[2], describing the portfolio using only its first two moments may be inappropriate, causing the MV criterion to fail in selecting the optimum portfolio.

In response to this limitation, alternative asymmetric risk measures which penalize downside deviations more than upside deviations have been proposed to replace the variance. Markowitz (1959) also proposed the semi-variance as a risk measure however the computational resources required to approximate the resulting solution were eventually a practical restriction at the time. Roy

[1] In fact, two optimization problems can be considered: either to minimize the variance for a given expected return or to maximize the expected returns for a given variance acceptable to the investors.

[2] Empirical evidence suggests (see, for instance, Cont; 2001) that skewness and kurtosis of returns exhibit pervasive behaviors. More recently, Nicolau and Rodrigues (2019) provide evidence on the heavy-tailedness of the exchange rate returns series distribution.

(1952) introduces the safety-first criterion, which is an alternative risk measure that seeks to minimize the probability of the portfolio's return falling below a predefined threshold. More recently, this topic is addressed by Ogryczak and Ruszczynski (1999) and Ogryczak and Ruszczynski (2001). Another important extension was the introduction of the Value-at-Risk (VaR) and related risk measures, concerned with extremely unfavorable results, as criteria for optimal portfolio selection (see, for instance, Basak and Shapiro; 2001 and Campbell et al.; 2001). These developments were motivated by the importance of the VaR as financial risk measure, and for regulatory purposes. For instance, the Basel Committee on Banking Supervision stipulated the VaR as the preferred approach to measure market risk.

All works reported so far focus on the value of risk measures at the end of a specified trading horizon and ignore the dynamic path of possible losses. Alternative approaches that take into account the magnitude of potential losses incurred over the trading horizon, denoted intra-horizon (IH) risk, were introduced by Kritzman and Rich (2002), Boudoukh et al. (2004) and Bakshi and Panayotov (2010). The IH risk is critical in a mark-to-market environment since sharp declines in asset values can affect trading strategies.

The methodologies proposed to capture the IH risk are mainly based on first passage probabilities, the probability that an event occurs for the first time within a finite horizon. In order to compute this statistic, asset price dynamics was typically modeled using the Geometric Brownian motion framework. A relevant exception is Bakshi and Panayotov (2010). This work warns of the importance of considering models that allow for sudden large changes in asset prices. Thus, they consider processes with jumps (Lévy jump models) for the underlying asset returns in order to compute the first-passage probability and show that the existence of such jumps tends to amplify the IH risk. However, results also reveal large variations in risk measures across different jump models, indicating model risk.

To the best of our knowledge Gupta et al. (2016) are the first that incorporate IH risk into asset allocation optimization. This component of the market risk is crucial in practice, since portfolio managers typically have thresholds that impose a stop-loss decision when broken. Moreover, the IH risk increases as the investment horizon increases (see, for instance, Kritzman and Rich; 2002 and Gupta et al.; 2016). Thus, in this context, the optimization problem has to incorporate a constraint regarding the probability of breaching some pre-specified drawdown threshold somewhere during the investment horizon.

In this work, we present a framework to formalize portfolio selection problems that allows us to incorporate the portfolio management issues referred to above. To this end, we introduce a novel nonparametric approach to estimate the first hitting time probabilities, which relies on the Markovian property of returns. Thus, stationary Markov chains are used to estimate all the relevant probabilities and expected moments. This method remains valid even when the underlying price process exhibits an upward stochastic trend, as generally happens with security prices. Since the approach is nonparametric, the widely documented model uncertainty problem in portfolio analysis (see, for instance Avramov and Zhou; 2010, for a literature review) is avoided.

In short, the proposed optimization problem aims to minimize the expected time to achieve a target return rate given that the probability of crossing a lower threshold, which triggers a stop-loss decision, is maintained at a level considered acceptable. This probability is used to manage the IH risk an investor is willing to accept.

The remainder of the paper is organized as follows. Section 4.2 introduces the proposed nonparametric approach and formalizes the optimization problem. Section 4.3 presents an empirical application of the proposed methodology to portfolio selection. In order to illustrate its potential, we considered the allocation of wealth across four Exchange Traded Funds (ETFs) from three different asset classes: equities, bonds and commodities. Section 4.4 concludes the paper.

4.2 The Proposed Approach for Portfolio Selection

Assume that an asset manager is responsible for an initial wealth and seeks to invest it in a number of different securities (stock, bonds, etc.) in order to obtain a target wealth in the shortest period of time. However, his mandate has an important restriction regarding the risk that can be taken: the portfolio value cannot fall below a defined threshold in a given time horizon.

In this context, the first time a portfolio value, say y_t, crosses some predefined threshold value is crucial in practice. When the target value of the portfolio is first attained, the asset manager's mandate is successfully completed. On the other hand, when the portfolio value falls below the lower threshold for the first time, the investment process may stop in order to avoid greater losses.

Let us define the first time that y_t crosses $y^+ = (1+r^+)x_0$ and $y^- = (1+r^-)x_0$ as

$$T^+ := T_{(1+r^+)x_0} = \inf\{t > 0 : y_t \geq y^+\} \tag{4.1}$$

and

$$T^- := T_{(1+r^-)x_0} = \inf\{t > 0 : y_t \leq y^-\}, \tag{4.2}$$

respectively, supposing that the process y_t starts at value x_0; r^- and r^+ are such that $y^- < x_0 < y^+$.

4.2.1 The Proposed Nonparametric Method to Estimate the First Hitting Time Probabilities

In principle, the assumption that y_t crosses the threshold y^+ (y^-) an infinite number of times over time as $T \to \infty$ is required in order to estimate the first passage time distributions of T^+ and T^- and all the relevant probabilities and expected moments in a nonparametric way. As shown by Nicolau (2017), this aspect follows from a stationary assumption and, in particular, from a positive Harris recurrence of y. However, since the focus here is the price of a security, which may exhibit a strong positive trend very much like a random walk with drift, y_t will not cross a fixed threshold enough times to allow for suitable estimation of the probabilities of interest. Nevertheless, in this nonstationary context, the event "y that increases (decreases) $r\%$" may actually occur an infinite number of times as time goes to infinity. Thus, the rationale behind the methodology here introduced is that this type of events (y increases or decreases $r\%$) may be modeled through stationary Markov chains, which still enables us to estimate the first passage time distributions of T^+ and T^-.

Assumption A1 (Conditional Homogeneity).
$P(T_{(1+r)x} \leq t|x_0 = x) = P(T_{(1+r)z} \leq t|x_0 = z)$ *for any value* x *and* z.

In general, Assumption A1 is plausible if, for instance, y is the price of a security traded in an efficient market. In fact, if prices fully reflect all known information, as implied by the Efficient Markets Hypothesis proposed by Fama (1970), it is not possible for investors to purchase undervalued stocks or sell stocks for inflated prices. Thus, in this context, the time that y takes to increase $r^+ 100\%$ is independent of the value at which the stock is currently traded, that is, it must be independent of value that x_0 assumes. It can be proved that the well-known Geometric Brownian motion provides an example of a process that satisfies Assumption A1.

Closed-form solutions to the first-passage probability are only available for few specifications. One of them is the Geometric Brownian motion, widely used in modeling market prices due to its mathematical tractability (see, for instance, Karlin and Taylor; 1975). However, there are nowadays significant empirical evidences that the distribution of equity returns is in general skewed and leptokurtotic (see, for instance, Mandelbrot; 1963 and Cont; 2001). For instance, asset prices often suddenly change by a very large amount (see, for instance, Eraker et al.; 2003) and this feature is difficult to reproduce within continuous-path models based on Brownian motions (see, for instance, Huang; 1985 and Johannes; 2004), which motivated the use of Lévy models to capture the short-run behavior of security prices. Beyond the lack of analytical explicitness (see, for instance, Kyprianou; 2006), choosing the most appropriate one from the great number of Lévy-type models proposed in the literature (see, for instance, Madan and Seneta; 1990 and Carr et al.; 2002) may also be a very challenging task[3]. Therefore, the nonparametric method we introduce to estimate the first hitting time probability function may be a useful approach, since it is flexible enough to accommodate jumps and other nonnormalities in asset prices.

The strategy to estimate the first hitting time probabilities of T^+ and T^- is based on two steps. Let us focus on the T^+ statistic in (4.1). First, the event increase of $r^+ 100\,\%$ is translated into an auxiliary process. Second, we extract from this auxiliary process all the relevant information regarding the estimation of the parameters of interest.

[3]For instance, Bakshi and Panayotov (2010) reveal large variations in the risk measures estimates across different Lévy jump models.

The Auxiliary Processes

The algorithm to build the auxiliary process is the following. Let $y^+ = x_0(1+r^+)$, with $r^+ > 0$, were $r^+\,100\%$ may be understood as growth rate (in percentage):

1. Set $i = 1$;

2. Let $x_0 = y_i$ (y_1 is the first observation of y);

3. Define

$$a_t^{(i)} := \begin{cases} 1 & \text{if } \; y_t < y^+,\; y_{t-1} < y^+, ..., y_{t-k+1} < y^+, y_{t-k} \le x_0, \\ 0 & \text{otherwise,} \end{cases}$$

4. Run $a_t^{(i)}$ from $t = i$ to $t = n$;

5. Set $i \to i + 1$ and return to step 2 until $i = n - 1$;

6. Stack all observations $a_t^{(i)}$, $i = 1, 2, ...$ in a vector. We designate this process as S_t^+.

The algorithm is easily adapted to the case $r^- < 0$. The resulting auxiliary process, S_t^-, can be obtained by running the algorithm above by replacing x_0, y^+ and y_t by $-x_0$, y^- and $-y_t$, respectively. Then, the probability that y_t crosses the threshold y^τ for the first time at time t, starting from x_0, is

$$P(T^\tau = t) = P(S_t^\tau = 0, S_{t-1}^\tau = 1, S_{t-2}^\tau = 1, ..., S_1^\tau = 1, S_0^\tau = 1),$$

which is equivalent to

$$P(T^\tau = t) = \big(1 - p_t\big) \prod_{i=1}^{t-1} p_i\,, \quad \tau = +, - \tag{4.3}$$

where $p_i := P(S_i^\tau = 1 | S_{i-1}^\tau = 1, S_{i-2}^\tau = 1, ..., S_0^\tau = 1)$, $x_0 < x_1$ for S_t^+ and $x_0 > x_1$ for S_t^-.

Assumption A2 (Markovian Property). *The processes S_t^+ and S_t^- are stationary discrete-time Markov processes of finite order.*

Therefore, the strategy is to treat S_t^+ and S_t^- as Markov chains with state space $\{0, 1\}$ and estimate the transition probabilities between the two states. A consequence of Assumption A2 is that event "y increases or decreases $r100\%$, $|r| < \infty$" may occur an infinite number of times as $T \to \infty$. If, for instance, we set r to be negative (positive) and y exhibits a strong positive (negative) trend over time, S_t will be formed mostly by ones and no relevant information can be extracted from this sequence regarding the transition probabilities. That is, the probability of transition from the state with $S_t = 1$ to the state with $S_t = 0$ will tend to zero.

In view of the Markovian property, if $t > k$ then $p_i = p_k$, where k is the Markov chain' order. Then,

$$P(T^\tau = t) = \begin{cases} (1-p_t)\displaystyle\prod_{i=1}^{t-1} p_i & \text{for } \ t \leq k, \\ \left((1-p_k)\displaystyle\prod_{i=1}^{k-1} p_i\right) p_r^{t-k} & \text{for } \ t > k. \end{cases} \tag{4.4}$$

Consequently, we have

$$\begin{aligned} E(T^\tau) &= \sum_{t=1}^{k} t\,(1-p_t)\prod_{j=1}^{t-1} p_j + \left\{(1-p_k)\prod_{j=1}^{k-1} p_j\right\} \sum_{t=k+1}^{\infty} t\, p_k^{t-k} \\ &= \sum_{t=1}^{k} t\,(1-p_t)\prod_{j=1}^{t-1} p_j + \prod_{j=1}^{k-1} p_j \frac{p_k(1+k-k\,p_k)}{1-p_k}. \end{aligned} \tag{4.5}$$

Therefore, in order to obtain estimate of (4.4) and (4.5) we only need the probabilities

$$p_j = P(S_i^\tau = 1 | S_{i-1}^\tau = 1, S_{i-2}^\tau = 1, ..., S_{t-j}^\tau = 1), \tag{4.6}$$

with $1 \leq j \leq r$, which can be easily estimated from standard Markov chain inference. For instance, if $k = 1$, the maximum likelihood estimate is $\hat{p}_1 = \eta_{11}/\eta_1$ where η_{11} is the number of transitions of type $\{S_{t-1}^\tau = 1, S_t^\tau = 1\}$ and η_1 counts the number of cases for which $\{S_{t-1}^\tau = 1\}$. As another example, when $k = 3$, $\hat{p}_3 = \eta_{1111}/\eta_{111}$ where η_{1111} and η_{111} are the number of transitions of type $\{S_{t-3}^\tau = 1, S_{t-2}^\tau = 1, S_{t-1}^\tau = 1, S_t^\tau = 1\}$ and $\{S_{t-3}^\tau = 1, S_{t-2}^\tau = 1, S_{t-1}^\tau = 1\}$, respectively.

Estimation of the Markov Order k

There are several approaches to estimate the order of the Markov chain (see Katz; 1981 and Zhao et al.; 2001). Most of them rely on Akaike's information criterion and the Bayesian information criterion. These methods are not only cumbersome, as they require building several Markov chains of a different order, but they are probably also inefficient because our object of interest centers exclusively on the probabilities

$$p_k := P\left(S_t^\tau = 1 |\, S_{t-1}^\tau = 1, S_{t-2}^\tau = 1, ..., S_{t-k}^\tau = 1\right)$$

and not on the entire Markov chain. We propose a new efficient and straightforward method below to estimate the order k defined in the previous equation. We consider the auxiliary regression

$$\mathcal{I}_{\{S_t^\tau=1\}} = \alpha \mathcal{I}_{\left\{S_{t-1}^\tau=1,...,S_{t-(k-1)}^\tau=1\right\}} + \beta \mathcal{I}_{\left\{S_{t-1}^\tau=1,...,S_{t-k}^\tau=1\right\}} + \varepsilon_t, \quad t = k+1, k+2, ..., n \tag{4.7}$$

where $\mathcal{I}_{\{A\}}$ is a dummy variable equal to 1 if A is true and the error term ε_t is not correlated with $\mathcal{I}_{\left\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-(k-1)}=1\right\}}$ and $\mathcal{I}_{\left\{S^{\tau}_{t}=1,S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-k}=1\right\}}$. Both the independent and dependent variables only take on zero and ones. In the Appendix C we prove that

$$\beta = \frac{p_k - p_{k-1}}{1 - p_{k-1}} \tag{4.8}$$

where $p_k = P\left(S^{\tau}_t = 1 |\, S^{\tau}_{t-1} = 1, \ldots, S^{\tau}_{t-k} = 1\right)$.This means that we can assess whether p_k is equal to p_{k-1} by a standard t-test using the ratio $\hat{\beta}/\hat{\sigma}_{\beta}$. If the hypothesis $\beta = 0$ is not rejected then p_k is equal to p_{k-1} and the true order k^* is lower than k. For example, for $k = 2$ to simplify, if $P\left(S^{\tau}_t = 1 |\, S^{\tau}_{t-1} = 1, S^{\tau}_{t-2} = 1\right)$ $= P\left(S^{\tau}_t = 1 |\, S^{\tau}_{t-1} = 1\right)$ then the event $\{S^{\tau}_t = 1\}$ is independent of $\left\{S^{\tau}_{t-2} = 1\right\}$ given $\left\{S^{\tau}_{t-1} = 1\right\}$ and the order is lower than $k = 2$.

To illustrate the methodology on how to estimate the true order k^* we may start with say $k^{\max} = 3$. Then:

1. Run the regression $\mathcal{I}_{\{S^{\tau}_t=1\}} = \alpha\mathcal{I}_{\left\{S^{\tau}_{t-1}=1,S^{\tau}_{t-2}=1\right\}} + \beta\mathcal{I}_{\left\{S^{\tau}_{t-1}=1,S^{\tau}_{t-2}=1,S^{\tau}_{t-3}=1\right\}}$ $+ \varepsilon_t$ and test $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$.

2. If $\beta = 0$ is rejected the order is estimated as $k^* = 3$

3. If $\beta = 0$ is not rejected run $\mathcal{I}_{\{S^{\tau}_t=1\}} = \alpha\mathcal{I}_{\left\{S^{\tau}_{t-1}=1\right\}} + \beta\mathcal{I}_{\left\{S^{\tau}_{t-1}=1,S^{\tau}_{t-2}=1\right\}}$ and test $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$.

4. If $\beta = 0$ is rejected the order is estimated as $k^* = 2$

5. If $\beta = 0$ is not rejected the order is estimated as $k^* = 1$ (we assume the S_t is time-dependent, so the Markov chain is at least a first order Markov process; otherwise we could run $\mathcal{I}_{\{S^{\tau}_t=1\}} = \alpha + \beta\mathcal{I}_{\left\{S^{\tau}_{t-1}=1\right\}}$)

It can be proved that the β in (4.8) can also be obtained from the auxiliary regression

$$\begin{aligned} \mathcal{I}_{\{S^{\tau}_t=1\}} &= \theta_1\mathcal{I}_{\left\{S^{\tau}_{t-1}=1\right\}} + \theta_2{}_{\left\{S^{\tau}_{t-1}=1,S^{\tau}_{t-2}=1\right\}} + \ldots + \theta_{k-1}\mathcal{I}_{\left\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-(k-1)}=1\right\}} \\ &+ \beta\mathcal{I}_{\left\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-k}=1\right\}} + \varepsilon_t, \qquad t = k+1, k+2, \ldots, n. \end{aligned} \tag{4.9}$$

That is, the β estimate is numerically identical in both regression (4.7) and (4.9). Note also that, in the case of the regression with only one "explanatory " variable

$$\mathcal{I}_{\{S^{\tau}_t=1\}} = \theta\mathcal{I}_{\left\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-k}=1\right\}} + \varepsilon_t, \qquad t = k+1, k+2, \ldots, n \tag{4.10}$$

the parameter θ represents p_k (the proof is available in the Appendix C) .

4.2.2 The Optimization Problem

In order to emphasize the allocation problem let us redefine (4.1) and (4.2) as functions of the vector of portfolio weights $\boldsymbol{\omega} = (\omega_1, ..., \omega_m)'$:

$$T^+(\boldsymbol{\omega}) = \inf\{t > 0 : y_t(\boldsymbol{\omega}) > (1 + r^+)\, x_0\},$$

and

$$T^-(\boldsymbol{\omega}) = \inf\{t > 0 : y_t(\boldsymbol{\omega}) < (1 + r^-)\, x_0\},$$

where x_0 is the initial wealth available, r^+ is the desired portfolio's cumulative return, r^- is the cumulative loss rate that defines a lower threshold, and

$$y_t(\boldsymbol{\omega}) = \prod_{i=1}^{t} \left(1 + \sum_{j=1}^{m} \omega_j r_{ji}\right) x_0,$$

with r_{ji} being the simple (one-period) return of asset j at time i.

Suppose that three external parameters are given to a portfolio manager: the cumulative return target r^+, a maximum drawdown threshold that equals $(1 + r^-)x_0$ and triggers a stop-loss decision and a maximum number of days, say N, to achieve the desired portfolio's return r^+. When r^+ is reached or the maximum drawdown threshold is broken, the portfolio is terminated. Thus, the asset manager faces two risks. First, the risk that the stop-loss decision is made before reaching r^+, called IH risk. Second, the risk that the return target is never met within the investment horizon $[0, N]$.

As stated by Kritzman and Rich (2002) and Gupta et al. (2016), unlike the risk of not reaching the target return, which diminishes with time, the IH risk increases quite dramatically as the investment horizon increases and may lead to premature termination of the investment. Then it is important to achieve the target return r^+ as fast as possible.

Thus, we propose the following optimization:

$$\min_{\{\omega_1,\ldots,\omega_m\}} E\Big(T^+(\boldsymbol{\omega})\Big) \tag{4.11}$$

subject to the constraints

$$P\Big(T^-(\boldsymbol{\omega}) \leq N\Big) < p_0, \tag{4.12}$$

$$\boldsymbol{\omega}'\mathbf{1} = 1,$$

where $\mathbf{1}$ is a m-dimensional vector of ones, and N and p_0 are parameters that allow as to manage the intra-horizon risk. More precisely, the probability p_0 in (4.12) that a cumulative loss rate of more than $r^-\,100\%$ occurs in a time horizon of N trading days is a proxy of the portfolio manager's risk tolerance.

Therefore, the objective is to minimize the expected time to achieve a cumulative return target in (4.11) subject to the constraint in (4.12), which allows us to manage the downside risk by choosing the parameters p_0, r^- and N considered more appropriate given the asset manager' degree of risk aversion.

As an extreme example of risk aversion, when $p_0 \approx 0$, the weight vector $\boldsymbol{\omega}$ must be selected such that the probability of facing a cumulative loss rate of $r^-100\%$ at any time until the end of the investment horizon N is practically zero.

4.2.3 The Optimization Algorithm

The objective function to be minimized has the form

$$\Psi(\boldsymbol{\omega}) = E\Big(T^+(\boldsymbol{\omega})\Big) + \gamma\, \mathcal{I}\Big(p_0(\boldsymbol{\omega})\Big), \tag{4.13}$$

where

$$\mathcal{I}\Big(p_0(\boldsymbol{\omega})\Big) := \begin{cases} 0 \text{ if } P\Big(T^-(\boldsymbol{\omega}) \le N\Big) = \sum_{t=1}^{N} P\Big(T^-(\boldsymbol{\omega}) = t\Big) < p_0, \\ 1 \text{ otherwise,} \end{cases} \tag{4.14}$$

and γ assumes a high positive value (for instance, 1,000,000) in order to discard solutions that do not respect the constraint in (4.12).

As is evident from (4.13) and (4.14), classical optimization methods (for instance, linear and quadratic programming) based on exploiting the derivatives of the objective function are not applicable here since, for a particular vector $\boldsymbol{\omega}$, we have to iteratively:

1. Build the auxiliary processes S^+ and S^-;
2. Estimate the Markov chain orders for S^+ and S^-;
3. Obtain the probability functions for the first-passage-time processes T^+ and T^-;
4. Compute $E(T^+)$ and $P(T^-(\boldsymbol{\omega}) \le N)$.

Thus, in order to minimize the objective function in (4.13), heuristic iterative stochastic search methods should be considered. We will employ the Threshold Accepting (TA) algorithm introduced by Dueck and Scheuer (1990). It was one of first heuristic approaches applied to portfolio selection problems; see Dueck and Winker (1992). Roughly speaking, this heuristic algorithm consists in starting with a randomly chosen feasible solution and successively picking (randomly) new solutions. Each of the new solutions, called neighbor solutions, are evaluated. If the new solution is better or as long as its deviation from the previous solution does not exceed certain thresholds, even though it is worse, it is accepted. Then, in order to implement this method we need to define, in addition to the objective function and constraints handling functions, a neighborhood function and the thresholds.

The Neighborhood Function and the Threshold Sequence

Since several specifications can be considered for the neighborhood function and for the thresholds, many variations of the TA algorithm are possible. We will use the variant proposed in Gilli and Schumann (2010b) and Gilli et al. (2011), since they have shown that it performs well in portfolio optimization problems.

The neighborhood algorithm in Gilli et al. (2011), which defines how we move from a solution to the next has the following steps:

Neighborhood Algorithm

1. randomly select $j_1 \in \{\text{assets with weight} > w^{min}\}$;
2. randomly select $j_2 \in \{\text{assets with weight} < w^{max}\}$;
3. set ϵ, determined by a draw of a uniformly distributed over [0, 0.5%] random variable;
4. $\omega_{j_1} = \omega_{j_1} - \epsilon$;
5. $\omega_{j_2} = \omega_{j_2} + \epsilon$;

where w^{min} and w^{max} are such that $w^{min} < w_j < w^{max}$ for $j = 1, ..., m$.

Regarding the threshold sequence, it consists in an ordered vector of positive numbers that decrease to zero or at least become very small. In order to compute the threshold sequence, consider the following algorithm, also from Gilli et al. (2011):

Threshold Sequence Algorithm

1. set the number of thresholds n_{rounds} and the number of random steps n_{deltas};
2. randomly generate feasible current solution, say, x^c ;
3. **for** $i = 1 : n_{\text{deltas}}$: **do**
generate $x^n \in \mathcal{N}(x^c)$ and compute $\Delta_i = |\Psi(x^n) - \Psi(x^c)|$, where $N(x^c)$ is the neighbor to the current solution defined using the neighborhood algorithm above; then, set $x^c = x_n$;
end for;
4. compute the empirical distribution CDF of Δ_i, $i = 1, ..., n_{\text{deltas}}$;
5. compute the threshold sequence $\tau_k = \text{CDF}^{-1}\left(\frac{n_{\text{rounds}} - k}{n_{\text{rounds}}}\right)$, with $k = 1, ..., n_{\text{rounds}}$; thus, this algorithm uses n_{rounds} equidistant quantiles .

Finally, the optimization algorithm, which employs the two algorithms presented above:

Threshold Accepting (TA) Algorithm

1. use the threshold sequence algorithm to construct the threshold sequence τ;
2. randomly generate feasible current solution, say, x^c ;
3. set $x^* = x^c$;
4. **for** $r = 1 : n_{\text{rounds}}$ **do**

for $i = 1 : n_{\text{steps}}$, where n_{steps} is the number od steps per threshold, **do**
generate $x^n \in \mathcal{N}(x^c)$ and compute $\Delta = \Psi(x^n) - \Psi(x^c)$
if $\Delta < \tau_r$ **then** $x^c = x^n$;
if $\Psi(x^c) \leq \Psi(x^*)$ **then** $x^c = x^n$;
end for;
end for;
5. return x^*.

In the empirical application, we choose $n_{\text{rounds}} = 10$, since Gilli and Schumann (2010b) report that the performance of the algorithm stays roughly the same for more than 10 thresholds. Moreover, we consider $n_{\text{steps}} = 1000$ and $n_{\text{deltas}} = 2000$. Then, the number of iterations is given by $n_{\text{rounds}} \times n_{\text{steps}} = 10000$. The TA algorithm was run in `R`, using the `TAopt:Optimisation with Threshold Accepting` code contained in the `NMOF: Numerical Methods and Optimization in Finance` package (see Schumann; 2011–2020). In order to initiate the algorithm, we provide random values for the vector of weights $\boldsymbol{\omega}$, since finding a good solution should not depend on "good" starting values (Gilli and Schumann; 2010a).

4.3 Empirical Application

In order to illustrate the proposed methodology for portfolio selection, we will consider the following four Exchange Traded Funds (ETFs): an US small-cap value ETF (ticker: IJS); a commodity ETF (gold, ticker: GLD); an US treasury bond ETF (ticker: TLT) and an investment Grade Corporate Bond ETF (ticker: LQD). We consider daily data (adjusted close prices) from January 2, 2009 through December 28, 2018 (2515 observations). The data source is Yahoo! Finance.

Figure 4.1 graphically displays the prices of the four ETFs considered. As contemporaneous correlation between asset classes has increased in recent years[4], it is nowadays more challenging to construct diversified portfolios, that is, portfolios in which positive performance of some investments neutralizes the negative performance of others. In order to reduce the sensitivity to adverse market swings, we considered four ETFs from three different asset classes: equities, bonds and commodities. Since bond and equity markets tend to move in opposite directions, downward movements in one will possibly be compensated by positive movements in the other. Table 4.1 shows that the correlations between the selected ETFs are not high in absolute value. Low or negative correlation between portfolio components, have been pointed out by portfolio managers, as a way of limiting risk.

[4]For instance, the correlations between stock returns are high (and positive) even when very different sectors (or countries) are considered.

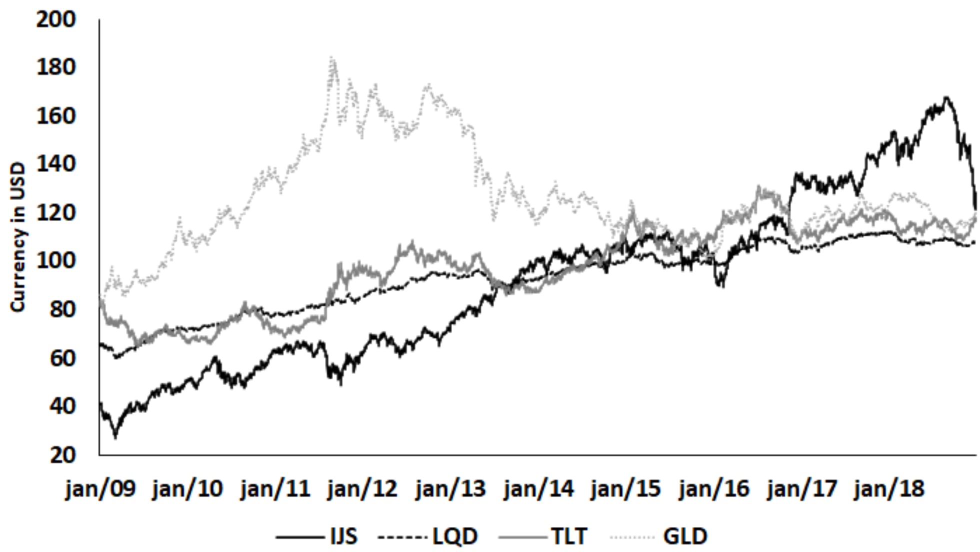


FIGURE 4.1: Prices of the four ETFs considered

TABLE 4.1: Returns Correlation Matrix

	IJS	GLD	TLT	LQD
IJS	1.00	-	-	-
GLD	0.02	1	-	-
TLT	0.22	-0.01	1	-
LQD	0.17	-0.43	0.68	1

In this section, we compare the portfolios selected using the proposed methodology with those that are obtained employing the workhorse model for portfolio optimization: mean-variance optimization (Markowitz; 1952), whose objective function to be minimized can be written as

$$\Phi = (1 - \lambda)\, \boldsymbol{\omega}' \, \boldsymbol{\Sigma} \, \boldsymbol{\omega} - \lambda \, \boldsymbol{\omega}' \boldsymbol{\mu}, \tag{4.15}$$

where $\boldsymbol{\mu}$ is the vector of assets' expected returns and $\boldsymbol{\Sigma}$ is the variance-covariance matrix of the assets' returns. Since $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown, their sample estimates are used in practice. The parameter $\lambda \in [0, 1]$ is a measure of risk aversion. If $\lambda = 0$, the portfolio with lower variance is selected; on the other hand, for $\lambda = 1$, we maximize return. The complete mean-variance efficient frontier[5] can be obtained by varying λ between 0 and 1.

As suggested by Gupta et al. (2016), we assume that the investment manager faces two risks. The risk of breaching the maximum accepted drawdown $y^- =$

[5]The set of optimal portfolios that offer the highest expected return for a defined level of risk.

$x_0(1 + r^-)$, called IH risk, and the risk of not achieving the desired return in the investment horizon $[0, N]$, called end-of-horizon risk. This last definition differs from that of Gupta et al. (2016) since we assume that the portfolio is terminated the first time the desired return r^+ is reached and not in N. Since the IH risk increases with time, the portfolio manager has incentives to close the investment as soon as possible. Moreover, there is a maximum horizon of N for the investment process.

Unlike the mean-variance (MV) and related optimization methods based on end-of-horizon measures, the optimization problem in (4.11) and (4.12) can incorporate both intra-horizon and end-of-horizon risks. Roughly speaking, its objective is to minimize the expected time to achieve an r^+ cumulative return for the portfolio while the probability of crossing a lower threshold is maintained below some predefined value. The probability $P\Big(T^-(\boldsymbol{\omega}) \leq N\Big) < p_0$ in (4.12) allows us take into account all the asset prices' path when the portfolio composition is decided and, therefore, provides a simple and flexible way to manage the IH risk.

Table 4.2 presents the chosen portfolios considering the optimization problem in (4.11) and (2.8) with $r^+ = 10\%$, $r^- = -5\%$, $N = 250$ and several values of p_0. The solutions to this optimization problem were obtained using the Threshold Accepting (TA) algorithm presented in Subsection 4.2.2, where it is also detailed how the parameters for this heuristic search method are chosen. We considered $w^{min} = 0$ and $w^{max} = 1$ in the Neighborhood Algorithm, that is, short-selling is forbidden. In order to allow short selling assets, we only have to choose a negative value for w^{min} and rerun the optimization algorithm.

We also present the portfolios associated with four points of the standard mean-variance (MV) approach' efficient frontier and some summary statistics. For instance, all the four ETFs series exhibit higher kurtosis and lower skewness than implied by the normal curve, which is line with the empirical findings reported in the literature (see, for instance, Cont; 2001 and Hong and Stein; 2003).

Regarding the asset allocations presented in Table 4.2, the MV approach proposes to invest all the wealth in the equity ETF when $\lambda \geq 0.61$ (moderate and high risk tolerance) in (4.15). In a similar way, if the risk tolerance is high the investor will accept a higher p_0 in (4.12), which results in an increase in the proportion of wealth invested in the equity ETF and a decrease in the bonds and commodity ETFs.

On the other hand, in an extreme risk aversion context, the minimum variance portfolio ($\lambda = 0$) suggests investing almost 90% of the initial wealth x_0 in the Investment Grade Corporate Bond ETF (LQD) and the proposed nonparametric method with $p_0 = 2.5\%$ recommends investing approximately 75% in the two bond ETFs (TLT and LQD) and the rest in the equity ETF (IJS). The sample statistics for the daily log-returns of the portfolios show that the volatility is lower for the MV approach when compared with the proposed optimization approach. However, our proposed method with $p_0 = 2.5\%$ results in a log-returns series with higher mean, lower kurtosis and less negative skewness. As stated by Maringer (2008), the fact that investors are strictly risk adverse does not necessarily imply an investment with lower volatility since they may be more sensitive towards skewness and kurtosis. In fact, the prospect theory introduced by Kahneman and Tversky (1979) argue that decisions are merely driven by loss aversion and the prospect of ending up with a lower than the current wealth. Consequently, it assumes that investors overweight tail events, meaning that they either overestimate the likelihood or the magnitude of losses. Therefore, in this context investors will be particularly concerned with the higher moments of their portfolios.

Table 4.2 also presents the expected time to cross $y^+ = (1 + r^+)x_0$ in the proposed nonparametric framework. As expected, the average number of days to achieve the cumulative return target of 10% is negatively correlated with the risk level the investor is willing to accept. For instance, when the risk tolerance is high, it takes on average 155 and 158 days for $\lambda = 1$ and $p_0 = 15\%$, respectively, to attain the desired cumulative return.

Following Gupta et al. (2016), we consider that a more appropriate measure of the portfolio risk requires taking into account both IH and end-of-horizon (EH) risks. Thus, $P\big(T^- \leq N\big)$, which gives us the probability that the lower threshold $y^- = x_0(1+r^-)$ is surpassed[6] somewhere during the investment horizon $[0, N]$, is used as a *proxy* for the IH risk . Moreover, we consider $P(T^+ > N)$ as a *proxy* for the EH risk.

Figure 4.2 presents, for the portfolios in Table 4.2, the IH and total (IH+EH) risks considering the proposed nonparametric framework with $N = 250$. Since the optimization method presented in (4.11) and (4.12) intends to incorporate the IH risk in the portfolio allocation decision[7], we will first focus on the IH risk estimates in Panel A of Figure 4.2. As expected, a higher risk aversion (λ in (4.15) and p_0 in (4.12) assuming low values) results in portfolios with lower probability of breaching the maximum accepted intra-horizon drawdown $y^- = 0.95x_0$ ($r^- = -5\%$) and higher expected time to attain the 10% cumulative return. Moreover, considering the proposed optimization method with $p_0 =$ 2.5% or $\lambda = 0$ results in almost the same IH risk (2.3% versus 2.5%), but our method has associated a much lower $E(T^+)$. On the other hand, if the investor wants a higher return, $p_0 = 15\%$ in (4.12) results in a portfolio with a similar $E(T^+)$ to that obtained choosing $\lambda = 0.75$ in (4.15), but with a much lower IH probability (15% versus 27%).

Panel B of Figure 4.2 suggests that the EH risk is extremely relevant in the presence of a maximum investment horizon N and its importance more pronounced for the MV optimization. For instance, $\lambda = 0$ in (4.15) chooses a portfolio with very low IH risk, but with the highest total risk (EH+IH risks) when compared to the alternatives. Summing up, the portfolios with better $E(T^+)$ - total risk trade-off are those obtained choosing $p_0 = 7.5\%, 12.5\%$ and 15%.

[6] Crossing the maximum accepted drawdown threshold $y^- = x_0(1+r^-)$ implies a stop-loss decision.

[7] Since γ in (4.13) assumes an extremely high positive value, the optimization method will propose solutions that respect the constraint in (4.12).

TABLE 4.2: Optimal ETFs allocation suggested by the proposed methodology.

Parameters	Portfolio				Log-returns				
p_0	GLD	IJS	LQD	TLT	Ann. mean	Ann. s.d.	Skew.	Kurt.	$E(T^+)$
2.5%	0.000	0.256	0.428	0.316	5.79%	6.97%	-0.226	5.361	314
5.0%	0.000	0.460	0.143	0.397	6.75%	8.60%	-0.156	5.204	215
7.5%	0.000	0.589	0.000	0.411	7.44%	10.17%	-0.132	5.149	183
10.0%	0.000	0.675	0.000	0.325	8.15%	11.65%	-0.121	5.237	170
12.5%	0.000	0.730	0.000	0.270	8.61%	12.83%	-0.122	5.405	164
15.0%	0.000	0.773	0.000	0.227	9.02%	13.99%	-0.123	5.613	158
MV									
$\lambda = 0.00$	0.042	0.066	0.892	0.000	5.29%	5.49%	-0.370	7.089	457
$\lambda = 0.25$	0.000	0.274	0.726	0.000	6.57%	6.61%	-0.284	6.959	314
$\lambda = 0.50$	0.000	0.699	0.262	0.038	9.09%	13.71%	-0.174	6.149	185
$\lambda = 0.75$	0.000	1.000	0.000	0.000	11.30%	21.71%	-0.114	7.325	155
Individual									
GLD	1.000	0.000	0.000	0.000	3.37%	16.40%	-0.496	7.890	992
IJS	0.000	1.000	0.000	0.000	11.30%	21.71%	-0.114	7.325	155
LQD	0.000	0.000	1.000	0.000	5.01%	5.75%	-0.409	6.608	495
TLT	0.000	0.000	0.000	1.000	3.39%	14.40%	-0.131	4.427	311

Notes: $r^+ = 10\%$, $r^- = 5\%$ and $N = 250$ in (4.11) and (4.12).

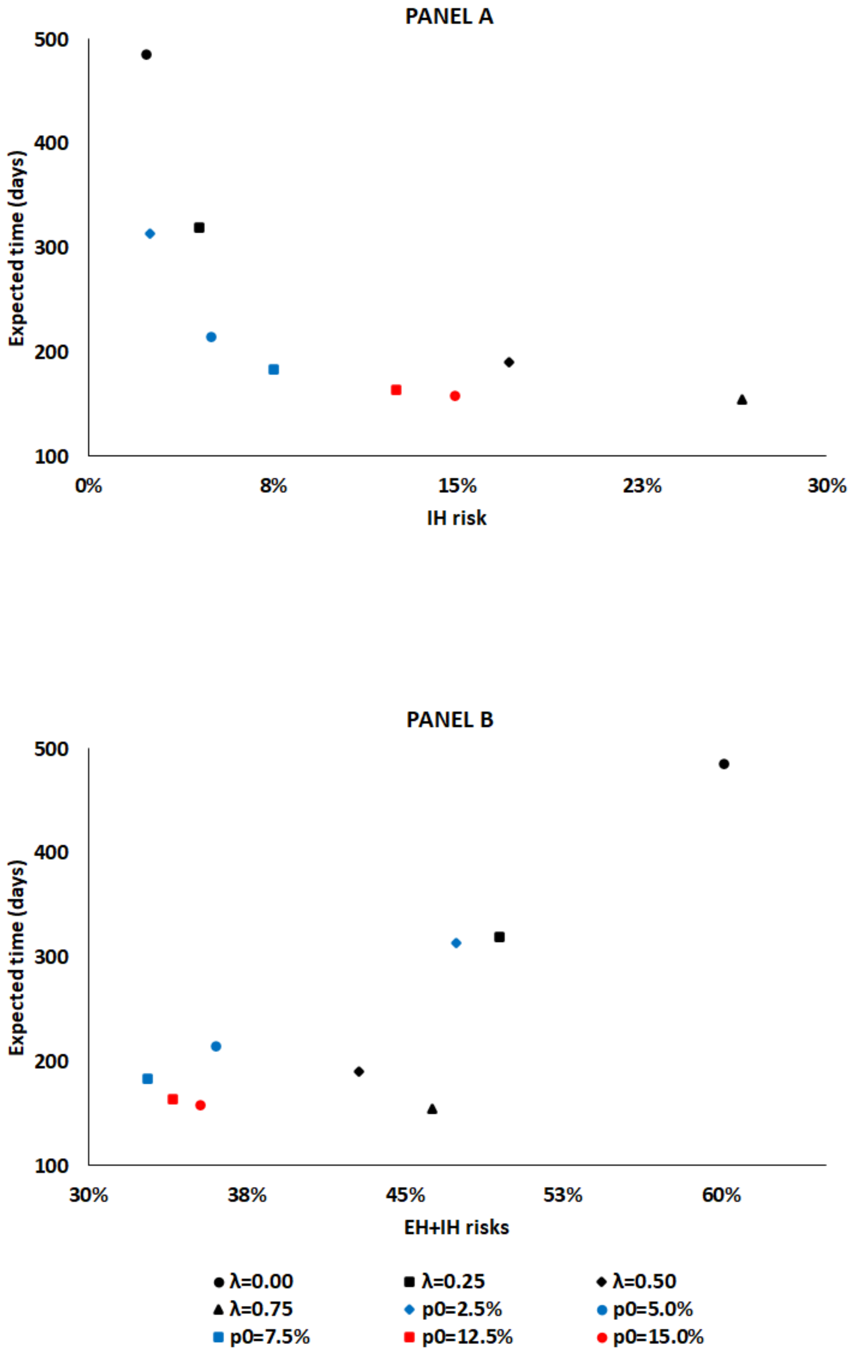


FIGURE 4.2: Nonparametric IH and EH risks probabilities

Notes: Nonparametric IH and EH risks probabilities were obtained using (4.4) and (4.5) with a maximum intra-horizon drawdown of -5% ($r^- = -5\%$) and a target cumulative return $r^+ = 10\%$.

4.4 Conclusions

In this paper we introduce simple nonparametric framework that intends to incorporate the intra-horizon risk in portfolio optimization. The proposed approach is based on first-hitting probabilities, then taking into account the entire asset price' path, which allows us to manage the probability of breaching the maximum accepted drawdown during the investment horizon. Given the reported difficulties of the Geometric Brownian Motion model in describing some financial data' observed features (see for instance, Cont; 2001 and Bakshi and Panayotov; 2010), several alternative processes implying more flexible distributions for the returns have been considered. However, most of them do not have closed-form analytical expressions for the first passage probability, which makes it difficult to find a tractable way to estimate this function. In order to overcome these limitations, we introduce a novel nonparametric method to estimate the first-hitting time.

The proposed optimization problem aims to minimize the expected time for a given target return to be reached, subject to a downside constraint: the probability of crossing a drawdown threshold should be lower than some predefined value that reflects the investment manager' degree of risk aversion. Thus, the proposed nonparametric method is employed to estimate the first passage probability functions of two stochastic processes: one representing the first time the target return is crossed and the other representing the first time the drawdown threshold is breached.

Heuristic methods are needed to solve the proposed optimization problem. We use the Threshold Accepting (TA) algorithm in the provided empirical application. This heuristic method was already applied to portfolio selection problems by, among others, Gilli and Schumann (2010a) and Gilli et al. (2011). Considering the proposed framework and assuming low risk tolerance results in portfolios with intra-horizon risk comparable to that of the Markowitz' mean-variance approach which minimizes the variance. However, the expected time to reach the target return of these portfolios is much lower, which implies a higher annualized return if the lower threshold is not crossed.

It is noteworthy that, for the parameters values considered in the empirical application, a cumulative return target of 10%, maximum drawdown of -5% and a maximum of 250 days to achieve the desired portfolio's cumulative return, the end-of-horizon risk is in most cases higher than intra-horizon risk.

Summing up, the proposed nonparametric framework provides a flexible way to incorporate the intra-horizon risk in the portfolio selection process. The provided empirical application suggests that using the optimization problem we introduce results in portfolios with lower expected time to cross the target return than those indicated by the Markowitz' mean-variance approach, for similar levels of intra-horizon risk .

Chapter 5

Concluding Remarks

This Thesis contributes to the existing time series econometrics literature by developing new techniques which address some relevant aspects of the economic and financial variables' dynamics, such as structural breaks, nonlinear relationships and non-normality.

The first paper, *A re-examination of inflation persistence dynamics in OECD countries: A new approach*, proposes a simple approach to detect multiple structural breaks in persistence that has two main advantages relative to the existing literature. First, the parameter changes are approximated by a single cosine function that allows to consider up to three breaks[1] in persistence. Then, the complete sample is used for estimation and not fractions of the sample as with recursive tests, which may have positive effects on power performance. Second, this procedure is, to the best of our knowledge, the first that also takes into account the effect of the shifts in the autoregressive parameters on the deterministic component of the process. An in-depth Monte Carlo analysis shows that the new procedure good power and size properties in small samples. Finally, empirical applications of the proposed test to G7 countries' inflation data provided relevant statistical evidence of breaks in persistence.

In the second paper, *The expected time to cross a threshold and its determinants: A simple and flexible framework*, we introduce a flexible framework to estimate the expected time (ET) an outcome variable takes to cross a threshold conditional on covariates. The major advantage of this methodology is that it allows us to capture nonlinear interactions without requiring the specification of a rigid parametric functional form. More precisely, computing the expected time (ET) for several starting values can provided important insights about the nature of the relationships between economic variables. The results of the empirical application suggest that an expansionary monetary policy may have an important role in stimulating a faster return of the industrial production to desirable growth rates. On the other hand, a contractionary monetary policy may precipitate the return to negative growth rates and trigger a recession.

Finally, the third paper, *The importance of intra-horizon risk in portfolio optimization*, presents a new portfolio optimization problem that takes into account the magnitude of potential losses incurred throughout the entire investment horizon. In order to operationalize this asset allocation framework, a novel nonparametric method to estimate the first passage probability function is introduced. Since, unlike the existing alternatives, this methodology provides

[1]Note that more breaks can be allowed for if deemed necessary.

a very tractable way to obtain the probabilities of interest and may accommodate jumps and other nonlinearities, we believe that it represents a relevant contribution to the literature. The results of the empirical application suggest that the proposed framework indicates portfolios able to achieve higher annualized return by the Markowitz' mean-variance approach with similar levels of intra-horizon risk.

Appendix A

Appendix to “A Re-Examination of Inflation Persistence Dynamics”

A.1 Proof of Main Results

Proof of Theorem 2.2.1

Consider first limit results for the local GLS demeaned and local GLS de-trended data.

Case A: Local GLS Demeaning

Under local GLS demeaning we estimate the parameter vector $\boldsymbol{\beta}$ in (2.1) using $\mathbf{x}_t = 1$. Hence, consider $y_{\tilde{c},1} := y_1$, $y_{\tilde{c},t} := y_t - \tilde{\rho}_t y_{t-1}$, $\mathbf{x}_{\tilde{c},1} := \mathbf{x}_1$, $\mathbf{x}_{\tilde{c},t} := \mathbf{x}_t - \tilde{\rho}_t \mathbf{x}_{t-1}$, and compute the OLS estimates as,

$$\hat{\boldsymbol{\beta}}_{\tilde{c}} = \left[\sum_{t=1}^{T} \mathbf{x}_{\tilde{c},t}\mathbf{x}'_{\tilde{c},t},\right]^{-1} \left[\sum_{t=1}^{T} \mathbf{x}_{\tilde{c},t} y_{\tilde{c},t}\right]. \tag{A.1}$$

Thus, the local GLS demeaned data is, $\hat{u}_{\tilde{c},t} = y_t - \mathbf{x}'_t \hat{\boldsymbol{\beta}}_{\tilde{c}} = u_t - \mathbf{x}'_t(\hat{\boldsymbol{\beta}}_{\tilde{c}} - \boldsymbol{\beta})$ or equivalently,

$$\hat{u}_{\tilde{c},t} = u_t - \mathbf{x}_t \left[\sum_{t=1}^{T} \mathbf{x}_{\tilde{c},t}\mathbf{x}'_{\tilde{c},t},\right]^{-1} \left[\sum_{t=1}^{T} \mathbf{x}_{\tilde{c},t} u_{\tilde{c},t}\right].$$

Since,

$$\sum_{t=1}^{T} \mathbf{x}_{\tilde{c},t}\mathbf{x}'_{\tilde{c},t}, = 1 + \left(\frac{\tilde{c}}{T}\right)^2 \sum_{t=2}^{T} \cos^2(\mathrm{k},\mathrm{t}) = 1 + o(1).$$

and

$$\sum_{t=1}^{T} \mathbf{x}_{\tilde{c},t} u_{\tilde{c},t} \quad = \quad u_1 - \left(\frac{\tilde{c}}{T}\right) \sum_{t=2}^{T} \cos(k,t)\Delta u_t + \left(\frac{\tilde{c}}{T}\right)^2 \sum_{t=2}^{T} \cos^2(k,t) u_{t-1}$$

we establish that,

$$\frac{1}{\sqrt{T}}\left[\sum_{t=1}^{T}\mathbf{x}_{\tilde{c},t}\mathbf{x}_{\tilde{c},t}',\right]^{-1}\sum_{t=1}^{T}\mathbf{x}_{\tilde{c},t}u_{\tilde{c},t}\rightarrow 0.$$

Since $\frac{1}{T}\sum_{t=2}^{T}\cos^2(k,t)\rightarrow\int_0^1\cos^2(k,r)dr$ then $\left(\frac{\tilde{c}}{T}\right)^2\sum_{t=2}^{T}\cos^2(k,t)=o(1)$.

Case B: Local GLS De-trending

For local GLS de-trending consider again (A.1) but with the denominator and numerator given as $D_T\sum_{t=1}^{T}\mathbf{x}_{\tilde{c},t}\mathbf{x}_{\tilde{c},t}{}'D_T$ and $D_T\sum_{t=1}^{T}\mathbf{x}_{\tilde{c},t}u_{\tilde{c},t}$, respectively, where
$D_T := \operatorname{diag}\left(1, T^{-1/2}\right)$.

Hence,

$$\begin{aligned} D_T\sum_{t=1}^{T}\mathbf{x}_{\tilde{c},t}\mathbf{x}_{\tilde{c},t}{}'D_T &= D_T\mathbf{x}_1\mathbf{x}_1{}'D_T + D_T\sum_{t=2}^{T}\mathbf{x}_{\tilde{c},t}\mathbf{x}_{\tilde{c},t}{}'D_T \\ &= \begin{bmatrix} 1 & T^{-1/2} \\ T^{-1/2} & T^{-1}\end{bmatrix} + D_T\begin{bmatrix}\Xi_1 & \Xi_2 \\ \Xi_2 & \Xi_3\end{bmatrix}D_T \end{aligned}$$

$$\Rightarrow \begin{bmatrix} 1 & 0 \\ 0 & \int_0^1\left[1-2\tilde{c}r\cos^2(k,r)+r^2\tilde{c}^2\cos^2(k,r)\right]dr\end{bmatrix}, \tag{A.2}$$

where, $\mathbf{x}_1 = (1,1)'$ and $\mathbf{x}_{\tilde{c},t} = (-\tilde{c}\cos(k,t)T^{-1}, 1-(t-1)\tilde{c}\cos(k,t)T^{-1})'$ for $t>1$, with $\Xi_1 := \left(\frac{\tilde{c}}{T}\right)^2\sum_{t=2}^{T}\cos^2(k,t)$, $\Xi_2 := \left(1-\frac{(t-1)\tilde{c}\cos(k,t)}{T}\right)\frac{\tilde{c}}{T}\cos(k,t)$, and $\Xi_3 := \left(1-\frac{(t-1)\tilde{c}\cos(k,t)}{T}\right)^2$.

Moreover, note that,

$$\begin{aligned} \frac{1}{T}\sum_{t=2}^{T}\cos(k,t) &\Rightarrow \int_0^1\cos(k,r)dr; \\ \frac{1}{T^2}\sum_{t=2}^{T}t\cos(k,t) &\Rightarrow \int_0^1 r\cos(k,r)dr; \\ \frac{1}{T^3}\sum_{t=2}^{T}t^2\cos^2(k,t) &\Rightarrow \int_0^1 r^2\cos^2(k,r)dr; \\ T^{-5/2}\sum_{t=3}^{T}t\cos(k,t)u_{t-1} &\Rightarrow \int_0^1 r\cos(k,r)W(r), \qquad 0\le r\le 1. \end{aligned}$$

Finally,

$$D_T \sum_{t=1}^{T} \mathbf{x}_{\tilde{c},t} u_{\tilde{c},t} = \begin{bmatrix} \Xi_4 \\ \Xi_5 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 0 \\ \sigma W(1)(1 - \tilde{c}\cos(k, T)) - \sigma \pi k \tilde{c} \int_0^1 r\sin(2\pi kt/T)W(r)dr \\ +\sigma \tilde{c}^2 \int_0^1 r\cos(k, r)W(r)dr \end{bmatrix}$$

where Ξ_4 is defined as in (A.2) and since $u_{\tilde{c},t} = \Delta u_{\tilde{c},t} - \tilde{c}\cos(k, t)T^{-1}$,

$$\begin{aligned}
\Xi_5 =& u_1 + u_T - u_1 - \frac{\tilde{c}}{T} \sum_{t=2}^{T} \cos(k, t) u_{t-1} - \frac{\tilde{c}}{T} \sum_{t=2}^{T} (t-1)\cos(k, t)\Delta u_t + \\
& \left(\frac{\tilde{c}}{T}\right)^2 \sum_{t=2}^{T} (t-1)\cos(k, t)^2 u_{t-1} \\
=& (1 - \tilde{c}\cos(k, T))u_T + 2\tilde{c}\cos(k, 2)u_1 T^{-1} - \frac{\pi k}{T}\frac{\tilde{c}}{T} \sum_{t=3}^{T} t\sin(2\pi \mathrm{kt/T}) u_{t-1} + \\
& \frac{\tilde{c}}{T} \sum_{t=3}^{T} \cos(k, t-1) u_{t-1} - \frac{\tilde{c}}{T} \sum_{t=2}^{T} \cos(k, t) u_{t-1} + \left(\frac{\tilde{c}}{T}\right)^2 \sum_{t=2}^{T} t\cos(k, t)^2 u_{t-1} - \\
& \left(\frac{\tilde{c}}{T}\right)^2 \sum_{t=2}^{T} \cos(k, t)^2 u_{t-1} + \frac{\tilde{c}}{T} \sum_{t=2}^{T} \cos(k, t)\Delta u_t.
\end{aligned}$$

It follows from the FCLT and CMT that

$$T^{-1/2}\hat{u}_{[Tr]} = T^{-1/2} u_{[Tr]} - T^{-1/2} \mathbf{x}'_{[Tr]} \left[D_T \sum_{t=1}^{T} \mathbf{x}_{\tilde{c},t} \mathbf{x}_{\tilde{c},t}{}' D_T \right]^{-1} \left[D_T \sum_{t=1}^{T} \mathbf{x}_{\tilde{c},t} u_{\tilde{c},t} \right]$$

$$\begin{aligned}
&\Rightarrow \sigma W(r) - \sigma r \left[\frac{(1 - \tilde{c}\cos(k, r))W(1) + \tilde{c}^2 \int_0^1 r\cos^2(k, r)W(r)dr}{\int_0^1 \left[1 - 2\tilde{c}r\cos^2(k, r) + r^2\tilde{c}^2\cos(k, r)\right] dr} \right] + \\
&+ \sigma r \left[\frac{\tilde{c}k\pi \int_0^1 r\sin(2\pi \mathrm{kr})W(r)dr}{\int_0^1 \left[1 - 2\tilde{c}r\cos^2(k, r) + r^2\tilde{c}^2\cos(k, r)\right] dr} \right] =: \sigma W_\tau(r),
\end{aligned} \tag{A.3}$$

where $[Tr]$ refers to the integer closest to Tr.

Thus from the results for Case A and for Case B above, we can now state the limit results for the test statistic. The OLS t-statistics to test $H_0 : \phi = 0$, computed from a test regression as in (2.4) based on locally GLS demeaned (μ) or de-trended (τ) data, i.e., $\hat{u}^{\mathfrak{v}}_{\tilde{c},t} = y_t - \mathbf{x}'_t \hat{\boldsymbol{\beta}}_{\mathfrak{v},\tilde{c}}$, $\mathfrak{v} = \mu$ or τ, is,

$$\hat{t}_k^{GLS_{\mathfrak{v}}} := \frac{\sum_{t=2}^{T} \Delta \hat{u}^{\mathfrak{v}}_{\tilde{c},t} \cos(k,t) \hat{u}^{\mathfrak{v}}_{\tilde{c},t-1}}{\left[\hat{\sigma}_k^2 \sum_{t=2}^{T} \cos^2(k, \mathrm{t}) \hat{u}^{\mathfrak{v}2}_{\tilde{c},t-1}\right]^{1/2}}, \text{ with } \mathfrak{v} = \mu,\ \tau. \tag{A.4}$$

Considering that $\hat{u}^{\mathfrak{v}}_{\tilde{c},t} = \Delta\hat{u}^{\mathfrak{v}}_{\tilde{c},t} + \hat{u}^{\mathfrak{v}}_{\tilde{c},t-1}$, squaring both sides and multiplying by $\cos(k,t)$ leads to $\cos(k,t)\hat{u}^{\mathfrak{v}2}_{\tilde{c},t} = \cos(k,t)\left[(\Delta\hat{u}_{\tilde{c},t})^{\mathfrak{v}2} + 2\Delta\hat{u}^{\mathfrak{v}}_{\tilde{c},t}\hat{u}^{\mathfrak{v}}_{\tilde{c},t-1} + \hat{u}^{\mathfrak{v}2}_{\tilde{c},t-1}\right]$. Summing over t and rearranging gives,

$$\sum_{t=2}^{T}\Delta\hat{u}^{\mathfrak{v}}_{\tilde{c},t}\mathrm{cos}(k,t)\hat{u}^{\mathfrak{v}}_{\tilde{c},t-1} =$$

$$= \frac{1}{2}\left[\sum_{t=2}^{T}\cos(k,t)\hat{u}^{\mathfrak{v}2}_{\tilde{c},t} - \sum_{t=2}^{T}\cos(k,t)\hat{u}^{\mathfrak{v}2}_{\tilde{c},t-1} - \sum_{t=2}^{T}\cos(k,t)(\Delta\hat{u}_{\tilde{c},t})^{\mathfrak{v}2}\right]. \tag{A.5}$$

Since under the null $\Delta\hat{u}^{\mathfrak{v}}_{\tilde{c},t} = \hat{\varepsilon}_t$, it follows that,

$$\sum_{t=2}^{T}\Delta\hat{u}^{\mathfrak{v}}_{\tilde{c},t}\mathrm{cos}(k,t)\hat{u}^{\mathfrak{v}}_{\tilde{c},t-1} =$$

$$= \frac{1}{2}\left[\cos(k,T)\hat{u}^{\mathfrak{v}2}_{\tilde{c},T} - \cos(k,2)\hat{u}^{\mathfrak{v}2}_{\tilde{c},1} - \sum_{t=3}^{T}\Delta\mathrm{cos}(k,t)\hat{u}^{\mathfrak{v}2}_{\tilde{c},t-1} - \sum_{t=2}^{T}\cos(k,t)\hat{\varepsilon}^{2}_{t}\right]. \tag{A.6}$$

In what follows the following limit results will prove useful. In specific, as $T \to \infty$,

$$\begin{aligned}
\cos(k,T)\frac{1}{T}\hat{u}^{2}_{\tilde{c},T} &\Rightarrow \sigma^2\mathrm{cos}(k,1)W_{\mathfrak{v}}(1)^2; && \text{(A.7)}\\
\cos(k,2)\frac{1}{T}\hat{u}^{2}_{\tilde{c},1} &\Rightarrow \sigma^2\mathrm{cos}(k,0)W_{\mathfrak{v}}(0)^2 = 0; && \text{(A.8)}\\
\frac{1}{T}\sum_{t=3}^{T}\Delta\mathrm{cos}(k,t)\hat{u}^{2}_{\tilde{c},t-1} &\Rightarrow \frac{\sigma^2}{2}(2\pi k)^2\int_0^1 \cos(2\pi kr)W_{\mathfrak{v}}(r)^2dr; && \text{(A.9)}\\
\frac{1}{T}\sum_{t=2}^{T}\cos(k,t)(\Delta\hat{u}_{\tilde{c},t-1})^2 &\to \sigma^2\int_0^1 \cos(k,r)dr = \frac{\sigma^2}{2}. && \text{(A.10)}
\end{aligned}$$

The result in (A.9) is obtained given that $\Delta\mathrm{cos}(k,t) = -\frac{1}{2}(2\pi k/T)\mathrm{sin}(2\pi kt/T) + o(1)$ (see Enders and Lee, 2012), and Lemma A.1 in (Bierens; 1997)). Recall that $\cos(k,t) := \frac{1}{2}(1+\cos(2\pi kt/T))$. Hence, for the numerator of (A.4) we establish as $T \to \infty$ that,

$$\frac{1}{T}\sum_{t=2}^{T}\Delta\hat{u}^{\mathfrak{v}}_{\tilde{c},t}\mathrm{cos}(k,t)\hat{u}^{\mathfrak{v}}_{\tilde{c},t-1}$$

$$\Rightarrow \frac{\sigma^2}{2}\left\{\cos(k,1)W_{\mathfrak{v}}(1)^2 + \frac{1}{2}(2\pi k)^2\int_0^1 \cos(2\pi kr)W_{\mathfrak{v}}(r)^2dr - 1\right\} \tag{A.11}$$

and for the denominator it follows from the continuous mapping theorem that,

$$\frac{1}{T^2}\sum_{t=2}^{T}\cos^2(k,t)\hat{u}^{\mathfrak{v}2}_{\tilde{c},t-1} \Rightarrow \sigma^2\int_0^1 \cos^2(k,r)W_{\mathfrak{v}}(r)^2dr. \tag{A.12}$$

Thus, from (A.11) and (A.12) it follows, under joint convergence, that,

$$\hat{t}_k^{GLS_\mathfrak{v}} \Rightarrow \frac{\cos(k,1)W_\mathfrak{v}(1)^2 + \frac{1}{2}(2\pi k)^2 \int_0^1 \cos(2\pi kr)W_\mathfrak{v}(r)^2 dr - 1}{2\left(\int_0^1 \cos^2(k,r)W_\mathfrak{v}(r)^2 dr\right)^{1/2}},$$

where k is a fixed value and $\mathfrak{v} = \mu$ or τ depending on whether local GLS demeaning or local GLS de-trending is used, respectively. ■

Proof of Theorem 2.2.3

An extension of the FCLT to near integrated process, $\rho_t = 1 - \frac{c}{T}$, states that, $\frac{1}{\sqrt{T}}u_{[Tr]} \Rightarrow \sigma J_c(r),\ 0 \leq r \leq 1$, where J_c is a standard OU process(see Phillips; 1987). The $P_{\tilde{c}}$ test statistic is given by

$$P_{\tilde{c}} = \frac{\sum_{t=1}^T \hat{\varepsilon}_{\tilde{c},t}^2 - \left[1 + \frac{\tilde{c}}{T}\cos(k,t)\right] \sum_{t=1}^T \hat{\varepsilon}_{0,t}^2}{\hat{\sigma}^2},$$

where $\hat{\varepsilon}_{0,t}$ is the residual term under $H_0 : \tilde{\rho}_t = 0$ and $\hat{\varepsilon}_{\tilde{c},t}$ is the residual term under $H_1 : \tilde{\rho}_t := 1 + \frac{\tilde{c}}{T}$ for a given $\tilde{c}$. The null hypothesis is rejected for small values of this statistic. Note that, in the case of demeaning,

$$\hat{\varepsilon}_{\tilde{c},t} = y_t - \left(1 + \frac{\tilde{c}}{T}\cos(\mathrm{k},\mathrm{t})\right) y_{t-1} - \beta_1 \left(1 - \left(1 + \frac{\tilde{c}}{T}\cos(\mathrm{k},\mathrm{t})\right)\right)$$

$$\to \Delta u_t - \frac{\tilde{c}}{T}\cos(\mathrm{k},\mathrm{t})u_{t-1},$$

$$\hat{\varepsilon}_{\tilde{c},t}^2 \to \Delta u_t^2 - \frac{2\tilde{c}}{T}\Delta u_t \cos(\mathrm{k},\mathrm{t})u_{t-1} + \left(\frac{\tilde{c}}{T}\right)^2 \cos^2(k,t)u_{t-1}^2,$$

$$\hat{\varepsilon}_{0,t}^2 = (\Delta u_t)^2,$$

Putting these results together we have that,

$$\sum_{t=1}^T \hat{\varepsilon}_{\tilde{c},t}^2 - \left[1 + \frac{\tilde{c}}{T}\cos(\mathrm{k},\mathrm{t})\right] \sum_{t=1}^T \hat{\varepsilon}_{0,t}^2 = -\frac{2\tilde{c}}{T}\sum_{t=2}^T \Delta u_t \cos(k,t)u_{t-1}$$

$$-\frac{\tilde{c}}{T}\sum_{t=2}^T \cos(k,t)\hat{\varepsilon}_{0,t}^2 + \left(\frac{\tilde{c}}{T}\right)^2 \sum_{t=2}^T \cos^2(k,t)u_{t-1}^2,$$

and given that

$$-\frac{2\tilde{c}}{T}\sum_{t=2}^T \Delta u_t \cos(\mathrm{k},\mathrm{t})u_{t-1} \Rightarrow \tilde{c}\left[\sigma^2 \int_0^1 \cos(\mathrm{k},\mathrm{r})dr - \sigma^2\cos(\mathrm{k},\mathrm{T})J_c^2(1)\right],$$

$$\left(\frac{\tilde{c}}{T}\right)^2 \sum_{t=2}^T \cos^2(\mathrm{k},\mathrm{t})u_{t-1}^2 \Rightarrow \tilde{c}^2\sigma^2 \int_0^1 \cos^2(\mathrm{k},\mathrm{r})J_c^2(r),$$

$$\frac{\tilde{c}}{T}\sum_{t=2}^{T}\cos_j(k,t)\hat{\varepsilon}_{0,t}^2 \Rightarrow \tilde{c}\sigma^2\int_0^1 \cos(k,r)dr,$$

the asymptotic distribution of $P_{\tilde{c}}$ is $P_{\tilde{c}} \Rightarrow \tilde{c}^2\int_0^1 \cos^2(\mathrm{k},\mathrm{r})J_c^2(r) - \tilde{c}\cos(\mathrm{k},\mathrm{T})J_c^2(1)$. Finally, when de-trending is considered $\hat{\varepsilon}_{\tilde{c},t} = y_t - \rho_t y_{t-1} - \beta_1(1-\rho_t) - \beta_2(t - \rho_t\,(t-1))$. Thus, using the FCLT result presented previously it follows that,

$$P_{\tilde{c}} \Rightarrow \tilde{c}^2\int_0^1 \cos^2(\mathrm{k},\mathrm{r})\left[J_c^\tau(r)\right]^2 + (1-\tilde{c}\cos(\mathrm{k},\mathrm{T})\left[J_c^\tau(1)\right]^2,$$

where J_c^τ is the local GLS de-trended OU process. ■

Proof of Proposition 1

The proposed test statistic with Eicker-White standard errors is defined as,

$$\hat{t}_{k,W}^{GLS} := \frac{\sum_{t=2}^{T}\Delta\hat{u}_t\cos(k,t)\hat{u}_{t-1}}{\left(\sum_{t=2}^{T}\cos^2(k,\mathrm{t})\hat{u}_{t-1}^2\hat{\eta}_t^2\right)^{1/2}}.$$

As the numerator is the same as in equation (2.5), we only need to examine the denominator. Hence, under the null hypothesis, considering

$$\frac{1}{T^2}\sum_{t=2}^{T}\cos^2(\mathrm{k},\mathrm{t})\hat{u}_{t-1}^2\hat{\eta}_t^2 = \frac{1}{T^2}\sum_{t=2}^{T}\cos^2(\mathrm{k},\mathrm{t})\hat{u}_{t-1}^2\sigma^2 + \frac{1}{T^2}\sum_{t=2}^{T}\cos^2(\mathrm{k},\mathrm{t})\hat{u}_{t-1}^2\left(\hat{\eta}_t^2-\sigma^2\right) \tag{A.13}$$

and noting that, $\sigma^2\frac{1}{T^2}\sum_{t=2}^{T}\cos^2(\mathrm{k},\mathrm{t})u_{t-1}^2 \Rightarrow \sigma^2\int_0^1\cos^2(\mathrm{k},\mathrm{r})W(r)^2dr$, we only need to prove that the second term in (A.13) is $o_p(1)$. Thus, from the result in Demetrescu (2008), $\frac{1}{T^2}\sum_{t=2}^{T}\hat{u}_{t-1}^2\left(\hat{\eta}_t^2-\sigma^2\right)\xrightarrow{p}0$, it follows that,
$\frac{1}{T^2}\sum_{t=2}^{T}\cos^2(\mathrm{k},\mathrm{t})\hat{u}_{t-1}^2\left(\hat{\eta}_t^2-\sigma^2\right) \leq \frac{1}{T^2}\sum_{t=2}^{T}\hat{u}_{t-1}^2\left(\hat{\eta}_t^2-\sigma^2\right)$, and therefore
$\frac{1}{T^2}\sum_{t=2}^{T}\cos^2(\mathrm{k},\mathrm{t})\hat{u}_{t-1}^2\left(\hat{\eta}_t^2-\sigma^2\right)$ is also $o_p(1)$ since $\cos^2(k,t)\leq 1$ for fixed $k>0$.

We also need to show that, $\frac{1}{T^2}\sum_{t=2}^{T}\hat{u}_{t-1}^2\cos^2(\mathrm{k},\mathrm{t})\left(\hat{\eta}_t^2-\sigma^2\right)\xrightarrow{p}0$ is still true when the process is near integrated. That is, when $u_t = (1+\tilde{c}\cos(k,t)/T)u_{t-1} + \varepsilon_t$. Since, $u_0 = 0$, we have

$$u_t = \sum_{i=0}^{t-1}\left(1-\frac{c}{T}\cos(k,\mathrm{i})\right)^i\varepsilon_{t-i};$$

$$\left(1-\frac{c}{T}\cos(k,i)\right)^i = 1-\frac{c}{T}\,i\cos(k,i)+O(T^{-1}),$$

and

$$u_t = \sum_{i=0}^{t-1}\varepsilon_{t-i} - \frac{c}{T}\sum_{i=0}^{t-1}i\cos(k,\mathrm{i})\,\varepsilon_{t-i} + O(T^{-0.5}).$$

Since $i/T = O(1)$, the result can be derived in the same way as under the null of a unit root. ■

A.2 Additional Tables

TABLE A.1: Comparison with the M test of Leybourne et al. (2007) - empirical size

DGP:	$y_t = y_{t-1} + u_t$ $u_t = \phi u_{t-1} + e_t,\ e_t \sim N(0,1)$			
	$T = 200$		$T = 400$	
ϕ	M	$\mathcal{T}_{\hat{k}}^{GLS_\mu,nr}$	M	$\mathcal{T}_{\hat{k}}^{GLS_\mu,nr}$
0.4	0.105	0.077	0.080	0.044
0.6	0.105	0.077	0.082	0.044

Notes: $\mathcal{T}_{\hat{k}}^{GLS_\mu,nr}$ is the minimum between $\mathcal{T}_{\hat{k}}^{GLS_\mu}$ with normal and reverse chronological order.

TABLE A.2: Comparison with the M test of Leybourne et al. (2007) - empirical power

DGP:

$$\begin{cases} y_t - d\,y_{\lfloor\tau_0 T\rfloor} = \rho_1(y_{t-1} - d\,y_{\lfloor\tau_0 T\rfloor}) + e_t & for \quad t = 1, ..., \lfloor\tau_1 T\rfloor \\ y_t - d\,y_{\lfloor\tau_1 T\rfloor} = \rho_2(y_{t-1} - d\,y_{\lfloor\tau_1 T\rfloor}) + e_t & for \quad t = \lfloor\tau_1 T\rfloor + 1, ..., \lfloor\tau_2 T\rfloor \\ y_t - d\,y_{\lfloor\tau_2 T\rfloor} = \rho_3(y_{t-1} - d\,y_{\lfloor\tau_2 T\rfloor}) + e_t & for \quad t = \lfloor\tau_2 T\rfloor + 1, ..., T. \end{cases}$$

				$T = 200$							
				$d = 1$				$d = 0$			
				$\alpha = 0.8$		$\alpha = 0.9$		$\alpha = 0.8$		$\alpha = 0.9$	
$(\tau_1\,,\,\tau_2)$	ρ_1	ρ_2	ρ_3	M	$\mathcal{T}_{\hat{k}}^{GLS_\mu,nr}$	M	$\mathcal{T}_{\hat{k}}^{GLS_\mu,nr}$	M	$\mathcal{T}_{\hat{k}}^{GLS_\mu,nr}$	M	$\mathcal{T}_{\hat{k}}^{GLS_\mu,nr}$
(0.30 , 0.00)	1	α	-	0.801	0.773	0.254	0.394	0.687	0.886	0.185	0.586
(0.50 , 0.00)	1	α	-	0.499	0.458	0.157	0.221	0.334	0.589	0.096	0.384
(0.30 , 0.00)	α	1	-	0.168	0.187	0.076	0.119	0.168	0.187	0.076	0.119
(0.50 , 0.00)	α	1	-	0.450	0.504	0.137	0.240	0.450	0.504	0.137	0.240
(0.25 , 0.75)	1	α	1	0.597	0.233	0.190	0.164	0.462	0.462	0.140	0.310
(0.35 , 0.75)	1	α	1	0.438	0.167	0.149	0.122	0.297	0.328	0.097	0.225
(0.25 , 0.50)	α	1	α	0.399	0.384	0.142	0.231	0.296	0.919	0.100	0.576
(0.35 , 0.50)	α	1	α	0.629	0.589	0.211	0.367	0.636	0.986	0.196	0.741
				$T = 400$							
(0.30 , 0.00)	1	α	-	1.000	0.937	0.813	0.713	1.000	0.978	0.687	0.877
(0.50 , 0.00)	1	α	-	0.991	0.663	0.511	0.415	0.963	0.724	0.334	0.599
(0.30 , 0.00)	α	1	-	0.634	0.276	0.179	0.191	0.634	0.276	0.179	0.191
(0.50 , 0.00)	α	1	-	0.988	0.715	0.468	0.505	0.988	0.715	0.468	0.505
(0.25 , 0.75)	1	α	1	0.995	0.257	0.601	0.206	0.984	0.715	0.468	0.472
(0.35 , 0.75)	1	α	1	0.957	0.518	0.451	0.148	0.886	0.423	0.294	0.340
(0.25 , 0.50)	α	1	α	0.998	0.746	0.404	0.377	0.896	0.997	0.302	0.933
(0.35 , 0.50)	α	1	α	0.813	0.713	0.637	0.570	0.995	1.000	0.642	0.992

Notes: $\lfloor\tau_i T\rfloor$ denotes the integer part of $\tau_i T$ and $\tau_0 = 0$; d is a dummy variable that equals 1 for the data generating processes considered in Leybourne et al. (2007), and 0 otherwise.

Appendix B

Appendix to "The Expected Time to Cross a Threshold and its Determinants"

Proof of main results

The Markov chain's log likelihood function in (3.6)

For an ith order Markov chain, the log-likelihood function can be expressed as the sum of i^2 components. As an illustration, consider a second order Markov chain ($i = 2$) and define:

$$\begin{aligned}
&\delta_{000} = 1 \text{ if } \{S_t = 0, S_{t-1} = 0, S_{t-2} = 0\};\\
&\delta_{100} = 1 \text{ if } \{S_t = 0, S_{t-1} = 0, S_{t-2} = 1\};\\
&\delta_{010} = 1 \text{ if } \{S_t = 0, S_{t-1} = 1, S_{t-2} = 0\};\\
&\delta_{110} = 1 \text{ if } \{S_t = 0, S_{t-1} = 1, S_{t-2} = 1\};\\
&p_{000}(\mathbf{x}) := p_2(\mathbf{x}) = P(S_t = 0|S_{t-1} = 0, S_{t-1} = 0);\\
&p_{100}(\mathbf{x}) = P(S_t = 0|S_{t-1} = 0, S_{t-1} = 1);\\
&p_{010}(\mathbf{x}) = P(S_t = 0|S_{t-1} = 1, S_{t-1} = 0);\\
&p_{110}(\mathbf{x}) = P(S_t = 0|S_{t-1} = 1, S_{t-1} = 1).
\end{aligned}$$

The log-likelihood for observation t can be expressed as

$$\ln L = L_1 + L_2 + L_3 + L_4,$$

where

$$\begin{aligned}
L_1 &= \delta_{001} \ln(1 - p_{000}(\mathbf{x})) + \delta_{000} \ln(p_{000}(\mathbf{x}));\\
L_2 &= \delta_{101} \ln(1 - p_{100}(\mathbf{x})) + \delta_{100} \ln(p_{100}(\mathbf{x}));\\
L_3 &= \delta_{011} \ln(1 - p_{010}(\mathbf{x})) + \delta_{010} \ln(p_{010}(\mathbf{x}));\\
L_4 &= \delta_{111} \ln(1 - p_{110}(\mathbf{x})) + \delta_{110} \ln(p_{110}(\mathbf{x})).
\end{aligned}$$

■

Proof of Proposition 2

Consider the probability $P(S_t = 0|S_{t-1} = 0, S_{t-2} = 0, ..., S_0 = 0; \mathbf{x})$. The results for other cases are similar. The event $\{S_{t-1} = 0, S_{t-2} = 0, ..., S_0 = 0\}$ represents $\{y_{t-1} < z_1, y_{t-2} < z_1, ..., y_1 < z_1, y_0 \leq z_0\}$. Therefore,

$$P(S_t = 0|S_{t-1} = 0, S_{t-2} = 0, ..., S_0 = 0; \mathbf{x}) \equiv$$
$$\equiv P(y_t < z_1|y_{t-1} < z_1, y_{t-2} < z_1, ..., y_0 \leq z_0; \mathbf{x})$$

and since y_t is an rth order Markov process,

$$\begin{aligned}
&P(S_t = 0|S_{t-1} = 0, S_{t-2} = 0, ..., S_0 = 0; \mathbf{x})\\
&= P(S_t = 0|S_{t-1} = 0, S_{t-2} = 0, ..., S_{t-r} = 0; \mathbf{x})\\
&= P(y_t < z_1|y_{t-1} < z_1, y_{t-2} < z_1, ..., y_{t-r+1} \leq z_1, y_{t-r} \leq z_0; \mathbf{x}).
\end{aligned}$$

■

Proof of Theorem 3.2.1

For condition (3) (conditional density identification) note that

$$E[(\mathbf{x}_t\boldsymbol{\beta}_{i,0} - \mathbf{x}_t\boldsymbol{\beta}_i)^2] = E[\{\mathbf{x}_t(\boldsymbol{\beta}_{i,0} - \boldsymbol{\beta}_i)\}^2] = (\boldsymbol{\beta}_{i,0} - \boldsymbol{\beta}_i)'E(\mathbf{x}_t\mathbf{x}_t')(\boldsymbol{\beta}_{i,0} - \boldsymbol{\beta}_i) > 0 \ ,$$

where $\boldsymbol{\beta}_i$ is the true parameter vector and $\boldsymbol{\beta}_{i,0}$ a parameter vector such that $\boldsymbol{\beta}_{i,0} \neq \boldsymbol{\beta}_i$. Hence, $\mathbf{x}_t\boldsymbol{\beta}_i \neq \mathbf{x}_t\boldsymbol{\beta}_{i,0}$ with positive probability and since $\Lambda(v)$ is strictly monotonic, we have $\Lambda(\mathbf{x}_t\boldsymbol{\beta}_i) \neq \Lambda(\mathbf{x}_t\boldsymbol{\beta}_{i,0})$ when $\mathbf{x}_t\boldsymbol{\beta}_i \neq \mathbf{x}_t\boldsymbol{\beta}_{i,0}$.

Condition (4) holds if $E[|\log(f(S_t|S_{t-1} = 0... = S_{t-r} = 0; \mathbf{x}_t; \boldsymbol{\beta}_i)|] < \infty$ for all $\boldsymbol{\beta}_i$.

For the logistic function, it is easy to verify that

$$|\ln \Lambda(v)| \leq |\ln \Lambda(0)| + |v|.$$

Furthermore, note that

$$\begin{aligned}
&|\ln f(S_t|S_{t-1} = 0, ..., S_{t-r} = 0; \mathbf{x}_t; \boldsymbol{\beta}_i)|\\
&\leq |S_t| \ \ln(\Lambda(\mathbf{x}_t\boldsymbol{\beta}_i))| + |1 - S_t| \ |\ln(1 - \Lambda(\mathbf{x}_t\boldsymbol{\beta}_i))|\\
&\leq |\ln(\Lambda(\mathbf{x}_t\boldsymbol{\beta}_i))| + |\ln(\Lambda(-\mathbf{x}_t\boldsymbol{\beta}_i))|\\
&(\text{since } |S_t| \leq \ 1 \text{ and } |1 - S_t| \leq 1)\\
&\leq 2[\,|\ln(\Lambda(0)| + ||\mathbf{x}_t|| \times ||\boldsymbol{\beta}_i||\,]\\
&(\text{due to the Cauchy-Schwartz inequality})
\end{aligned}$$

The nonsingularity of $\mathrm{E}(\mathbf{x}_t\mathbf{x}_t')$ implies $\mathrm{E}(x_{it}^2) < \infty$ for all i and, therefore, $\mathrm{E}(||\mathbf{x}_t^2||) < \infty$ and $\mathrm{E}(||\mathbf{x}_t||) < \infty$. Thus, the nonsingularity of $\mathrm{E}(\mathbf{x}_t\mathbf{x}_t')$ ensures that the logit ML estimator is consistent. ■

Proof of Theorem 3.2.2

Condition (1) is satisfied for the logit model if the compact parameter space $\boldsymbol{B}_i$ is taken to be $\mathbb{R}^p$. Condition (2) is obviously satisfied. To check condition (3) note that since $\mathrm{E}[S_t|S_{t-1}=0,...,S_{t-r}=0;\mathbf{x}_t]=\Lambda(\mathbf{x}_t\boldsymbol{\beta}_i)$, we have $\mathrm{E}[s(\mathbf{w}_t;\boldsymbol{\beta}_i)|\mathbf{x}_t]=0$ and, by the Law of Total Expectations $\mathrm{E}[s(\mathbf{w}_t;\boldsymbol{\beta}_i)]=0$.

In order to derive the conditional information matrix, note that, using the standard rules of differentiation we have that,

$$\mathrm{E}\left[\frac{\partial^2 \ln L}{\partial\boldsymbol{\beta}_i\partial\boldsymbol{\beta}_i'}\right]=-\mathrm{E}\left[\frac{\partial \ln L}{\partial\boldsymbol{\beta}_i}\ \frac{\partial \ln L}{\partial\boldsymbol{\beta}_i'}\right]+\mathrm{E}\left[\frac{1}{\ln L}\ \frac{\partial^2 \ln L}{\partial\boldsymbol{\beta}_i\partial\boldsymbol{\beta}_i'}\right],$$

where it is easy to verify that the second term on the right-hand side is zero. Thus, the following relationship between the expected value of the Hessian matrix and the expected outer product of the scores holds:

$$-\mathrm{E}[\boldsymbol{H}(\mathbf{w}_t;\boldsymbol{\beta}_i)]=\mathrm{E}[\mathbf{s}(\mathbf{w}_t;\boldsymbol{\beta}_i)\ \mathbf{s}(\mathbf{w}_t;\boldsymbol{\beta}_i)'],$$

where $s(\mathbf{w}_t;\boldsymbol{\beta}_i)$ and $\boldsymbol{H}(\mathbf{w}_t;\boldsymbol{\beta}_i)$ are the functions defined in (3.8) and (3.9),respectively. Regarding the local dominance of the Hessian - condition (4) -, since $\Lambda(\mathbf{x}_t'\boldsymbol{\beta}_i)[1-\Lambda(\mathbf{x}_t'\boldsymbol{\beta}_i)]<1$, we have $||\boldsymbol{H}(\mathbf{w}_t;\boldsymbol{\beta}_i)||\leq||\mathbf{x}_t\mathbf{x}_t'||$ for all $\boldsymbol{\beta}_i$. It can be shown that $\mathrm{E}[||\mathbf{x}_t\mathbf{x}_t'||]<\infty$ if $\mathrm{E}[\mathbf{x}_t\mathbf{x}_t']$ is nonsingular (and hence finite). Finally, condition (5) also requires that $\mathrm{E}[\mathbf{x}_t\mathbf{x}_t']$ is nonsingular. ■

Proof of Theorem 3.2.4

Under Assumption 1.B, the joint stationarity of $\{y_t,\mathbf{x}_t\}$ implies the joint stationarity of $\{S_t,\mathbf{x}_t\}$, given the measurability of (3.1). ■

Appendix C

Appendix to "The Importance of Intra-Horizon Risk in Portfolio Optimization"

Proof for equation (4.8)
Let us prove equation (4.8) in four steps.

Step 1 - some notations. First, let us define

$$\begin{aligned} \pi_{k+1} &= P\left(S_t^\tau = 1, ..., S_{t-k}^\tau = 1\right) \;\; (\; k+1 \text{ variables involved}) && \text{(C.1)} \\ p_k &= P\left(S_t^\tau = 1 \middle| \, S_{t-1}^\tau = 1, ..., S_{t-k}^\tau = 1\right), && \text{(C.2)} \end{aligned}$$

with S_t^τ defined in 4.2.1.

By stationarity these probabilities are independent of t. It follows that

$$\pi_{k+1} = p_k p_{k-1} ... p_1 \pi_1 = p_k \pi_k. \tag{C.3}$$

Step 2 - We derive the relation $\pi_k - \pi_1 \pi_{k-1} = \alpha\left(\pi_{k-1} - \pi_{k-1}^2\right) + \beta\left(\pi_k - \pi_{k-1}\pi_k\right).$

Pre-multiplying equation (4.7)

$$\mathcal{I}_{\{S_t^\tau = 1\}} = \alpha \mathcal{I}_{\left\{S_{t-1}^\tau = 1, ..., S_{t-(k-1)}^\tau = 1\right\}} + \beta \mathcal{I}_{\left\{S_{t-1}^\tau = 1, ..., S_{t-k}^\tau = 1\right\}} + \varepsilon_t, \tag{C.4}$$

by $\mathcal{I}_{\left\{S_{t-1}^\tau = 1, ..., S_{t-(k-1)}^\tau = 1\right\}}$ allows us to obtain

$$\begin{aligned} Cov\Big(\mathcal{I}_{\{S_t^\tau=1\}}, \mathcal{I}_{\{S_{t-1}^\tau=1,...,S_{t-(k-1)}^\tau=1\}}\Big) &= \alpha Var\Big(\mathcal{I}_{\{S_{t-1}^\tau=1,...,S_{t-(k-1)}^\tau=1\}}\Big) && \text{(C.5)} \\ &+ \beta Cov\Big(\mathcal{I}_{\{S_{t-1}^\tau=1,...,S_{t-(k-1)}^\tau=1\}}, \mathcal{I}_{\{S_{t-1}^\tau=1,...,S_{t-k}^\tau=1\}}\Big) \end{aligned}$$

Both the left and right side of the previous equations can be expressed in terms of $\{\pi_i; i = 1, 2, ..., k\}$. In fact,

$$\begin{aligned} & Cov\Big(\mathcal{I}_{\{S_t^\tau=1\}}, \mathcal{I}_{\{S_{t-1}^\tau=1,...,S_{t-(k-1)}^\tau=1\}}\Big) \\ = \; & E\Big(\mathcal{I}_{\{S_t^\tau=1\}} \mathcal{I}_{\{S_{t-1}^\tau=1,...,S_{t-(k-1)}^\tau=1\}}\Big) - E\Big(\mathcal{I}_{\{S_t^\tau=1\}}\Big) E\Big(\mathcal{I}_{\{S_{t-1}^\tau=1,...,S_{t-(k-1)}^\tau=1\}}\Big) \\ = \; & E\Big(\mathcal{I}_{\{S_t^\tau=1,...,S_{t-(k-1)}^\tau=1\}}\Big) - E\Big(\mathcal{I}_{\{S_t^\tau=1\}}\Big) E\Big(\mathcal{I}_{\{S_{t-1}^\tau=1,...,S_{t-(k-1)}^\tau=1\}}\Big) \\ = \; & P\big(S_t^\tau = 1, ..., S_{t-(k-1)}^\tau = 1\big) - P\big(S_t^\tau = 1\big) P\big(S_{t-1}^\tau = 1, ..., S_{t-(k-1)}^\tau = 1\big) \\ = \; & \pi_k - \pi_1 \pi_{k-1} \end{aligned}$$

Regarding the right side of equation (C.5), we have

$$\begin{aligned} Var\left(\mathcal{I}_{\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-(k-1)}=1\}}\right) &= \pi_{k-1}-\pi^2_{k-1} \\ Cov\left(\mathcal{I}_{\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-(k-1)}=1\}},\mathcal{I}_{\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-k}=1\}}\right) &= \pi_k-\pi_{k-1}\pi_k \end{aligned}$$

Therefore, equation (C.5) can be stated as

$$\pi_k-\pi_1\pi_{k-1}=\alpha\left(\pi_{k-1}-\pi^2_{k-1}\right)+\beta\left(\pi_k-\pi_{k-1}\pi_k\right). \tag{C.6}$$

Step 3 - We derive the relation $\pi_{k+1}-\pi_1\pi_k=\alpha\left(\pi_k-\pi_{k-1}\pi_k\right)+\beta\left(\pi_k-\pi^2_k\right)$.
Pre-multiplying (C.4) by $\mathcal{I}_{\left\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-k}=1\right\}}$ allows us to obtain

$$\begin{aligned} &Cov\left(\mathcal{I}_{\{S^{\tau}_{t}=1\}},\mathcal{I}_{\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-k}=1\}}\right) \\ &=\alpha\,Cov\left(\mathcal{I}_{\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-(k-1)}=1\}},\mathcal{I}_{\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-k}=1\}}\right)+\beta\,Var\left(\mathcal{I}_{\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-k}=1\}}\right). \end{aligned} \tag{C.7}$$

Again, this expression can be expressed in terms of $\{\pi_i; i=1,2,...,k\}$ as follows:

$$\begin{aligned} Cov\left(\mathcal{I}_{\{S^{\tau}_{t}=1\}},\mathcal{I}_{\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-k}=1\}}\right), &= \pi_{k+1}-\pi_1\pi_k, \\ Cov\left(\mathcal{I}_{\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-(k-1)}=1\}}\,\mathcal{I}_{\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-k}=1\}}\right) &= \pi_k-\pi_{k-1}\pi_k, \\ Var\left(\mathcal{I}_{\{S^{\tau}_{t-1}=1,\ldots,S^{\tau}_{t-k}=1\}}\right) &= \pi_k-\pi^2_k \end{aligned}$$

Therefore, equation (C.7) can be stated as

$$\pi_{k+1}-\pi_1\pi_k=\alpha\left(\pi_k-\pi_{k-1}\pi_k\right)+\beta\left(\pi_k-\pi^2_k\right) \tag{C.8}$$

Step 4 - Formula for β
Equations (C.6) and (C.8) form the system

$$\begin{cases} \pi_k-\pi_1\pi_{k-1}=\alpha\left(\pi_{k-1}-\pi^2_{k-1}\right)+\beta\left(\pi_k-\pi_{k-1}\pi_k\right) \\ \pi_{k+1}-\pi_1\pi_k=\alpha\left(\pi_k-\pi_{k-1}\pi_k\right)+\beta\left(\pi_k-\pi^2_k\right) \end{cases}$$

which can be solved with respect to α and β. Thus,

$$\beta=\frac{\pi^2_k-\pi_{k+1}\pi_{k-1}}{\pi_k\left(\pi_k-\pi_{k-1}\right)}. \tag{C.9}$$

Using the relation (C.3), $\pi_{k+1}=p_kp_{k-1}...p_1\pi_1$, equation (C.9) can be further simplified as

$$\beta=\frac{p_k-p_{k-1}}{1-p_{k-1}}.$$

■

Proof for equation (4.10)
The OLS estimate of θ is

$$\hat{\theta} = \frac{\sum_t^n \mathcal{I}_{\left\{S_{t-1}^{\tau}=1,\ldots,S_{t-k}^{\tau}=1\right\}} \mathcal{I}_{\left\{S_t^{\tau}=1\right\}}}{\sum_t^n \mathcal{I}^2_{\left\{S_{t-1}^{\tau}=1,\ldots,S_{t-k}^{\tau}=1\right\}}}.$$

Under certain conditions of regularity, the numerator converges to

$$\begin{aligned}\frac{\sum_t^n \mathcal{I}_{\left\{S_{t-1}^{\tau}=1,\ldots,S_{t-k}=1\right\}} \mathcal{I}_{\left\{S_t=1\right\}}}{n} &= \frac{\sum_t^n \mathcal{I}_{\left\{S_t=1,S_{t-1}=1,\ldots,S_{t-k}=1\right\}}}{n} \\ &\xrightarrow{p} P\left(S_t^{\tau}=1, S_{t-1}^{\tau}=1, \ldots, S_{t-k}^{\tau}=1\right)\end{aligned}$$

and the denominator,

$$\frac{\sum_t^n \mathcal{I}^2_{\left\{S_{t-1}^{\tau}=1,\ldots,S_{t-k}^{\tau}=1\right\}}}{n} = \frac{\sum_t^n \mathcal{I}_{\left\{S_{t-1}^{\tau}=1,\ldots,S_{t-k}^{\tau}=1\right\}}}{n} \xrightarrow{p} P\left(S_{t-1}^{\tau}=1, \ldots, S_{t-k}^{\tau}=1\right).$$

Therefore

$$\hat{\theta} \xrightarrow{p} \frac{P\left(S_t^{\tau}=1, S_{t-1}^{\tau}=1, \ldots, S_{t-k}^{\tau}=1\right)}{P\left(S_{t-1}^{\tau}=1, \ldots, S_{t-k}^{\tau}=1\right)} = P\left(S_t^{\tau}=1 \middle| S_{t-1}^{\tau}=1, \ldots, S_{t-k}^{\tau}=1\right).$$

■

Bibliography

Alogoskoufis, G. S. and Smith, R. (1991). The Phillips Curve, The persistence of Inflation, and the Lucas Critique: Evidence from Exchange-Rate Regimes, *The American Economic Review* **81**(5): 1254–1275.

Antunes, A., Bonfim, D., Monteiro, N. and Rodrigues, P. M. M. (2018). Forecasting banking crises with dynamic panel probit models, *International Journal of Forecasting* **34**(2): 249 – 275.

Audretsch, D. B. and Mahmood, T. (1995). New firm survival: New results using a hazard function, *The Review of Economics and Statistics* **77**(1): 97–103.

Avramov, D. and Zhou, G. (2010). Bayesian portfolio analysis, *Annual Review of Financial Economics* **2**(1): 25–47.

Azzalini, A. (1994). Logistic regression for autocorrelated data with application to repeated measures, *Biometrika* **81**(4): 767–775.

Bakshi, G. and Panayotov, G. (2010). First-passage probability, jump models, and intra-horizon risk, *Journal of Financial Economics* **95**(1): 20 – 40.

Banerjee, A., Lumsdaine, R. L. and Stock, J. H. (1992). Recursive and sequential tests of the unit-root and trend-break hypotheses: Theory and international evidence, *Journal of Business & Economic Statistics* **10**(3): 271–287.

Basak, S. and Shapiro, A. (2001). Value-at-risk-based risk management: Optimal policies and asset prices, *The Review of Financial Studies* **14**(2): 371–405.

Beechey, M. and Österholm, P. (2012). The Rise and Fall of U.S. Inflation Persistence, *International Journal of Central Banking* **8**(3): 55–86.

Bierens, H. J. (1997). Testing the unit root with drift hypothesis against nonlinear trend stationarity, with an application to the us price level and interest rate, *Journal of Econometrics* **81**(1): 29 – 64.

Birchenhall, C. R., Jessen, H., Osborn, D. R. and Simpson, P. (1999). Predicting U.S. business-cycle regimes, *Journal of Business & Economic Statistics* **17**(3): 313–323.

Boudoukh, J., Richardson, M. P., Stanton, R. H. and Whitelaw, R. F. (2004). Maxvar - long horizon value at risk in a mark-to-market environment., *Journal of Investment Management* **2**(3).

Buiter, W. H. and Gersovitz, M. (1981). Issues in controllability and the theory of economic policy, *Journal of Public Economics* **15**(1): 33 – 43.

Busetti, F. and Taylor, A. (2004). Tests of stationarity against a change in persistence, *Journal of Econometrics* **123**(1): 33 – 66.

Campbell, R., Huisman, R. and Koedijk, K. (2001). Optimal portfolio selection in a value-at-risk framework, *Journal of Banking & Finance* **25**(9): 1789 – 1804.

Carr, P., Geman, H., Madan, D. and Yor, M. (2002). The fine structure of asset returns: An empirical investigation, *The Journal of Business* **75**.

Cavaliere, G. and Taylor, A. R. (2008). Testing for a change in persistence in the presence of non-stationary volatility, *Journal of Econometrics* **147**(1): 84 – 98.

Chang, Y. and Park, J. Y. (2002). On the asymptotics of adf tests for unit roots, *Econometric Reviews* **21**(4): 431–447.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues, *Quantitative Finance* **1**(2): 223–236.

Cox, D. and Isham, V. (1980). *Point Processes*, Routledge.

Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last, *Journal of Econometrics* **146**(1): 162 – 169.

DeLong, J. B. and Summers, L. (1988). How does macroeconomic policy affect output?, *Brookings Papers on Economic Activity* **19**(2): 433–494.

Demetrescu, M. (2008). On the Dickey-Fuller test with White standard errors, *Statistical Papers* **51**(1): 11.

Dueck, G. and Scheuer, T. (1990). Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing, *Journal of Computational Physics* **90**(1): 161 – 175.

Dueck, G. and Winker, P. (1992). New concepts and algorithms for portfolio choice, *Applied Stochastic Models and Data Analysis* **8**(3): 159–178.

Durbin, J. (1971). Boundary-crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov-Smirnov test, *Journal of Applied Probability* **8**(3): 431–453.

Elliott, G., Rothenberg, T. J. and Stock, J. H. (1996). Efficient tests for an autoregressive unit root, *Econometrica* **64**(4): 813–836.

Enders, W. and Lee, J. (2012). A Unit Root Test Using a Fourier Series to Approximate Smooth Breaks, *Oxford Bulletin of Economics and Statistics* **74**(4): 574–599.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing, *Econometrica* **55**(2): 251–276.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica* **66**(5): 1127–1162.

Eraker, B., Johannes, M. and Polson, N. (2003). The impact of jumps in volatility and returns, *The Journal of Finance* **58**(3): 1269–1300.

Estrella, A. and Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity, *The Journal of Finance* **46**(2): 555–576.

Estrella, A. and Mishkin, F. S. (1998). Predicting U.S. recessions: Financial variables as leading indicators, *The Review of Economics and Statistics* **80**(1): 45–61.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work, *The Journal of Finance* **25**(2): 383–417.

Florio, A. (2004). The asymmetric effects of monetary policy, *Journal of Economic Surveys* **18**(3): 409–426.

Galbraith, J. W. and Tkacz, G. (2000). Testing for asymmetry in the link between the yield spread and output in the g-7 countries, *Journal of International Money and Finance* **19**(5): 657 – 672.

Gallant, A. (1981). On the bias in flexible functional forms and an essentially unbiased form: The fourier flexible form, *Journal of Econometrics* **15**(2): 211 – 245.

Giesecke, K. (2006). Default and information, *Journal of Economic Dynamics and Control* **30**(11): 2281 – 2303.

Gilli, M., Maringer, D. and Schumann, E. (2011). *Numerical Methods and Optimization in Finance*, Academic Press.

Gilli, M. and Schumann, E. (2010a). A note on "good starting values" in numerical optimisation, *Working Papers 044*, COMISEF.

Gilli, M. and Schumann, E. (2010b). Distributed optimization of portfolio's Omega, *Parallel Computing* **36**: 381–389.

Gonçalves, S. and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* **123**(1): 89 – 120.

Gupta, P., Skallsjo, S. and Li, B. (2016). *Multi-Asset Investing: A Practitioner's Framework*, John Wiley & Sons, Ltd.

Halunga, A. G., Osborn, D. R. and Sensier, M. (2009). Changes in the order of integration of US and UK inflation, *Economics Letters* **102**(1): 30 – 32.

Hamori, S. and Tokihisa, A. (1997). Testing for a unit root in the presence of a variance shift, *Economics Letters* **57**(3): 245 – 253.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* **64**(2): 413–430.

Harvey, C. R. (1989). Forecasts of economic growth from the bond and stock markets, *Financial Analysts Journal* **45**(5): 38–45.

Harvey, D. I., Leybourne, S. J. and Taylor, A. R. (2006). Modified tests for a change in persistence, *Journal of Econometrics* **134**(2): 441 – 469.

Hassler, U. and Demetrescu, M. (2005). Spurious Persistence and Unit Roots due to Seasonal Differencing: The Case of Inflation Rates, *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)* **225**(4): 413–426.

Hayashi, F. (2000). *Econometrics*, Princeton University Press.

Hong, H. and Stein, J. C. (2003). Differences of opinion, short-sales constraints and market crashes, *Review of Financial Studies* **16**(Summer): 487–525.

Härdle, W., Horowitz, J. and Kreiss, J.-P. (2003). Bootstrap methods for time series, *International Statistical Review* **71**(2): 435–459.

Huang, C.-F. (1985). Information structure and equilibrium asset prices, *Journal of Economic Theory* **35**(1): 33 – 71.

Islam, M. A. and Chowdhury, R. I. (2006). A higher order Markov model for analyzing covariate dependence, *Applied Mathematical Modelling* **30**(6): 477 – 488.

Johannes, M. (2004). The statistical and economic role of jumps in continuous-time interest rate models, *The Journal of Finance* **59**(1): 227–260.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk, *Econometrica* **47**(2): 263–291.

Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes*, second edition edn, Academic Press, Boston.

Katz, R. W. (1981). On some criteria for estimating the order of a markov chain, *Technometrics* **23**(3): 243–249.

Kauppi, H. and Saikkonen, P. (2008). Predicting U.S. recessions with dynamic binary response models, *The Review of Economics and Statistics* **90**(4): 777–791.

Kiefer, N. M. (1988). Economic duration data and hazard functions, *Journal of Economic Literature* **26**(2): 646–679.

Kim, J.-Y. (2000). Detection of change in persistence of a linear time series, *Journal of Econometrics* **95**(1): 97 – 116.

Kim, K. and Schmidt, P. (1993). Unit root tests with conditional heteroskedasticity, *Journal of Econometrics* **59**(3): 287 – 300.

Kim, T.-H., Leybourne, S. and Newbold, P. (2002). Unit root tests with a break in innovation variance, *Journal of Econometrics* **109**(2): 365 – 387.

Kritzman, M. and Rich, D. (2002). The mismeasurement of risk, *Financial Analysts Journal* **58**(3): 91–99.

Kyprianou, A. E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*, Springer-Verlag Berlin Heidelberg.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*, Springer-Verlag New York.

Levin, A. and Piger, J. (2003). Is inflation persistence intrinsic in industrial economies?, *Working Papers 2002-023*, Federal Reserve Bank of St. Louis.

Leybourne, S., Kim, T.-H., Smith, V. and Newbold, P. (2003). Tests for a change in persistence against the null of difference-stationarity, *The Econometrics Journal* **6**(2): 291–311.

Leybourne, S., Kim, T.-H. and Taylor, A. R. (2007). Detecting Multiple Changes in Persistence, *Studies in Nonlinear Dynamics & Econometrics* **11**(3): 1–34.

Li, D. and Ng, W.-L. (2000). Optimal dynamic portfolio selection: Multiperiod mean-variance formulation, *Mathematical Finance* **10**(3): 387–406.

Lo, A. W., MacKinlay, A. and Zhang, J. (2002). Econometric models of limit-order executions, *Journal of Financial Economics* **65**(1): 31 – 71.

Lumsdaine, R. L. and Papell, D. H. (1997). Multiple trend breaks and the unit-root hypothesis, *The Review of Economics and Statistics* **79**(2): 212–218.

MacKinnon, J. G. (2007). Bootstrap Hypothesis Testing, *Working Paper 1127*, Economics Department, Queen's University.

Madan, D. and Seneta, E. (1990). The variance gamma (vg) model for share market returns, *The Journal of Business* **63**: 511–24.

Mandelbrot, B. (1963). The variation of certain speculative prices, *The Journal of Business* **36**(4): 394–419.

Maringer, D. (2008). Risk preferences and loss aversion in portfolio optimization, *in* E. Kontoghiorghes, B. Rustem and P. Winker (eds), *Computational Methods in Financial Engineering: Essays in Honour of Manfred Gilli*, Springer.

Markowitz, H. (1952). Portfolio Selection, *The Journal of Finance* **7**(1): 77–91.

Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments*, Yale University Press.

Mata, J. and Portugal, P. (2002). The survival of new domestic and foreign-owned firms, *Strategic Management Journal* **23**(4): 323–343.

McCall, B. P. (1994). Testing the proportional hazards assumption in the presence of unmeasured heterogeneity, *Journal of Applied Econometrics* **9**(3): 321–334.

McConnell, M. M. and Perez-Quiros, G. (2000). Output Fluctuations in the United States: What Has Changed Since the Early 1980's?, *The American Economic Review* **90**(5): 1464–1476.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* **3**(4): 373 – 413.

Meyer, B. D. (1990). Unemployment insurance and unemployment spells, *Econometrica* **58**(4): 757–782.

Meyn, S., Tweedie, R. L. and Glynn, P. W. (2009). *Markov Chains and Stochastic Stability*, Cambridge Mathematical Library, 2 edn, Cambridge University Press.

Nelson, C. R. and Plosser, C. R. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications, *Journal of Monetary Economics* **10**(2): 139 – 162.

Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power, *Econometrica* **69**(6): 1519–1554.

Nicolau, J. (2012). *Modelação de Séries Temporais Financeiras*, Vol. Coleção Económicas - 2ª Série, Almedina.

Nicolau, J. (2017). A simple nonparametric method to estimate the expected time to cross a threshold, *Statistics & Probability Letters* **123**: 146 – 152.

Nicolau, J. and Rodrigues, P. M. M. (2019). A new regression-based tail index estimator, *The Review of Economics and Statistics* **101**(4): 667–680.

Ogryczak, W. and Ruszczynski, A. (1999). From stochastic dominance to mean-risk models: Semideviations as risk measures, *European Journal of Operational Research* **116**(1): 33 – 50.

Ogryczak, W. and Ruszczynski, A. (2001). On consistency of stochastic dominance and mean-semideviation models, *Mathematical Programming* **89**(232): 217.

Ohn, J., Taylor, L. W. and Pagan, A. (2004). Testing for duration dependence in economic cycles, *The Econometrics Journal* **7**(2): 528–549.

Papell, D. H., Murray, C. J. and Ghiblawi, H. (2000). The structure of unemployment, *The Review of Economics and Statistics* **82**(2): 309–315.

Perron, P. (1989). The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, *Econometrica* **57**(6): 1361–1401.

Perron, P. (1997). Further evidence on breaking trend functions in macroeconomic variables, *Journal of Econometrics* **80**(2): 355 – 385.

Perron, P. (2005). Dealing with Structural Breaks, *Boston University - Department of Economics - Working Papers Series WP2005-017*, Boston University - Department of Economics.

Phillips, P. C. B. (1987). Time Series Regression with a Unit Root, *Econometrica* **55**(2): 277–301.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression, *Biometrika* **75**(2): 335–346.

Phillips, P. C. B., Shi, S. and Yu, J. (2015). Testing for Multiple Bubbles: Limit theory of Real Time Detectors, *International Economic Review* **56**(4): 1079–1134.

Phillips, P. C. B., Wu, Y. and Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?, *International Economic Review* **52**(1): 201–226.

Politis, D. N. and Romano, J. P. (1992a). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation, *The Annals of Statistics* **20**(4): 1985–2007.

Politis, D. N. and Romano, J. P. (1994). The Stationary Bootstrap, *Journal of the American Statistical Association* **89**(428): 1303–1313.

Politis, D. and Romano, J. (1992b). *Exploring the Limits of Bootstrap*, Wiley, New York, chapter A Circular Block Resampling Procedure for Stationary Data, pp. 263–270.

Prodan, R. (2008). Potential pitfalls in determining multiple structural changes with an application to purchasing power parity, *Journal of Business & Economic Statistics* **26**(1): 50–65.

Raftery, A. E. (1985). A model for high-order markov chains, *Journal of the Royal Statistical Society. Series B (Methodological)* **47**(3): 528–539.

Rappoport, P. and Reichlin, L. (1989). Segmented trends and non-stationary time series, *The Economic Journal* **99**(395): 168–177.

Rodrigues, P. M. M. and Taylor, A. M. R. (2012). The Flexible Fourier Form and Local Generalised Least Squares De-trended Unit Root Tests, *Oxford Bulletin of Economics and Statistics* **74**(5): 736–759.

Roy, A. D. (1952). Safety first and the holding of assets, *Econometrica* **20**(3): 431–449.

Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika* **71**(3): 599–607.

Saldías, M. (2017). The nonlinear interaction between monetary policy and financial stress, *IMF Working Papers* **17**: 1.

Schumann, E. (2011–2020). *Numerical Methods and Optimization in Finance (NMOF) Manual. Package version 2.1-0)*.

Sensier, M. and van Dijk, D. (2003). Testing for Volatility Changes in US Macroeconomic Time Series, *Centre for Growth and Business Cycle Research Discussion Paper Series 36*, Economics, The Univeristy of Manchester.

Sichel, D. E. (1991). Business cycle duration dependence: A parametric approach, *The Review of Economics and Statistics* **73**(2): 254–260.

Stock, J. (1994). *Unit Roots, Structural Breaks, and Trends*, Elsevier, Amsterdam, pp. 2740–2843.

Stock, J. H. and Watson, M. W. (1989). New Indexes of Coincident and Leading Economic Indicators, *NBER Macroeconomics Annual* **4**: 351–394.

Sueyoshi, G. T. (1995). A class of binary response models for grouped duration data, *Journal of Applied Econometrics* **10**(4): 411–431.

Taylor, J. B. (2000). Low inflation, pass-through, and the pricing power of firms, *European Economic Review* **44**(7): 1389 – 1408.

Terasvirta, T. (2006). Chapter 8 Forecasting economic variables with nonlinear models, *in* G. Elliott, C. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Vol. 1, Elsevier, pp. 413 – 457.

Terasvirta, T., Tjostheim, D. and Granger, C. (2010). *Modelling Nonlinear Economic Time Series*, Oxford University Press.

Whitmore, G. A. (1986). First-passage-time models for duration data: Regression structures and competing risks, *Journal of the Royal Statistical Society. Series D (The Statistician)* **35**(2): 207–219.

Zhao, L., Dorea, C. and Gonçalves, C. (2001). On determination of the order of a markov chain, *Statistical Inference for Stochastic Processes* **4**: 273–282.