Università Degli Studi di Siena

Ph.D. in Economics, XXIII cycle

(A.A. 2007/2008 - A.A. 2010/2011)

Doctoral Dissertation

Three Essays on Nonlinear Time-series Econometrics

Novella Maugeri

Ph.D. Coordinator

Prof. Ugo Pagano

Supervisor

Prof. Marco Paolo Tucci

Doctoral Defense Committee

Dr. Stelios Bekiros

Prof. Giampiero Gallo

Prof. Massimiliano Marcellino

The undersigned Novella Maugeri, in her quality of doctoral candidate for a Ph.D. degree in Economics granted by the *Università Degli Studi di Siena* confirms that the research exposed in this dissertation is original and that it has not been and it will not be used to pursue or attain any other academic degree of any level at any other academic institution, be it foreign or Italian.

Acknowledgments

This thesis is the result of four years of work, each of these years characterized by different people I met on the way and different environments I have worked in. As a consequence, also these acknowledgments go through these four years and mention all the deserving people I found on the way.
There have been times where I have find myself quite lost, especially during the early years of my Ph.D. However, in the end I can say that I finally *found myself* (research-wise), and this is especially thanks to the people I met at the Center of Economic Studies at the University of Leuven. In particular, I want to thank Professor Paul De Grauwe for inviting me to spend a visiting research period there, and Pablo Rovira Kaltwasser and Leonardo Iania for the long times spent with me talking about research and many other things. A special thanks goes in particular to Sophie Bereau, for listening to my preliminary ideas and ultimately for giving me *the sparkle* for starting this Ph.D. project. As a matter of fact, this dissertation is the product of my efforts combined with two very important ingredients: this *sparkle*, and the careful guidance, the infinite patience and academic experience of my supervisor, Professor Marco Paolo Tucci. I am such a lucky person that his support and guidance have never lacked me throughout these difficult years, and this is really a rare and precious gift I received: Thank you so much Professor!
Thanks a lot also to the person who made me fall in love with economics for the first time, with his amazing teaching skills and with his enthusiasm, Professor Domenico Tosato. Another professor who deserves my gratitude is Guido Erreygers, for reminding me to constantly believe in myself despite everything that hits me and pushes me away from my goals. A nostalgic thought goes also to Professor Fernando Vianello, who made me understand that the important questions are always the ones that provoke more questions rather than easy answers (*I miss you a lot*).

There have been also been times when I was convinced that Siena was not the right place for me to continue with research. However, I soon realized that there was something within me that had to change, and not with the place I worked and lived in. For making me realize this very important truth I have to thank windy Rotterdam and all the nice people I have met at the Tinbergen Institute. For giving me the strength of continuing after the storm, I have to thank Liliana's wisdom, the peace and beauty of Siena, and the beauty of your soul Vince, my Love. Without you my life would be so different right now and I couldn't even imagine these four years: thank you for letting me in your hearth, though we both know there was only a tiny-tiny hole to enter.

All my gratitude to my Ph.D. colleagues, who have started this crazy journey with me in Siena, and who are now shining bright in different parts of the world: *Grazie* to Burak, Guy, Angela and Gerrit, Sandra and Ana, Alesia, Simon and Youssef...*what if we never met??* I don't even want to think about it.

At a different stages of the journey, I have also been lucky to meet my *alter ego*, Nove, and strong sweet Natalie: somebody says we have become too old to find more true friends, but you are the proof that this is not true! Thank you.

Last but not least, I have to thank all my friends who were with me before this Ph.D. experience, and who have remained there all the times I needed it. The list is long and in no way it can be exhaustive but I will try.

Thank you *Gina*. Despite the fact that your life has changed so much in the last few years please leave always a little spot for me as you have done so far, I'll make sure I fit!

A big thanks to *Marchetto*, for your patience and support expressed especially in the long nice chats we had on Skype. My hope is that we can always be so close, despite being geographically so far away.

Thank you Sara: with your special way of expressing your love and care for me you helped me growing and become a better person. Remember, wherever you will end-up I will always be there for you if you let me.

Thanks also to my *little genius* Noemi: I don't know how you do it, but you always manage to make me feel better!

Thank you Michela, for the profound insights you always give me about love, life and everything that goes in between.

To Andrea and all my fellow dancers, thank you for making me discover another part of me who was hidden inside: she will never be forgotten, so will you.

Finally, I want to say thank you to my wonderful brother *Leo* and to my family, for "it has been much more Heaven than it was Hell". Davvero, Grazie.

Contents

Introduction **1**

1 How Rational Are Rational Expectations? New Evidence from Well Known Survey Data **6**

1.1 Introduction 7
1.2 The Data 10
1.3 Assessing Theories of Rationality 14
1.3.1 Adaptive Expectations and Sticky Information Diffusion 15
1.3.2 'Strong' Rationality 16
1.3.3 'Weak Rationality' 17
1.3.4 *Strategic Stickiness*: 'Weak Rationality' and Asymmetric Adjustment 18
1.4 Results 19
1.4.1 Adaptive Expectations and Sticky Information Diffusion 20
1.5 Concluding Remarks 35
1.6 Appendix 37
1.6.1 The European Commission Consumers Survey and the Carlson-Parking's (1975) Method 37
1.6.2 Replicating VDF (2000) 38

2 Some Pitfalls in Smooth Transition Models Estimation: A Monte Carlo Study **44**

2.1 Introduction 45
2.2 Smooth Transition Models 48

2.3 Concentrating the Sum of Squares Function 50
2.4 Monte Carlo Design 53
2.5 Experiment I: STECM Estimation 56
2.6 Experiment II: STAR Estimation 62
2.7 Investigating of the Shape of the Loglikelihood Function 65
2.8 Concluding Remarks 70
2.9 Appendix: Detailed Descriptive Statistics 72

3 Forecasting Italian GDP Growth through the Term Spread: How Many Stars to the STAR Model? 77

3.1 Introduction 78
3.2 Background 82
3.3 Methodology 85
3.3.1 Estimating the STAR model 85
3.3.2 The Forecasting Exercise 88
3.4 The Dataset and the Chosen Time Sample 91
3.5 Results 97
3.5.1 Forecasting with the STAR model 98
3.5.2 Forecasting with multiple models: a 'Horse Race' 101
3.6 Concluding Remarks 108
3.7 Appendix: Detailed Tables for the Forecasting Exercise 110

Bibliography 115

Introduction

The recent abrupt developments of the financial crisis have made researchers dramatically aware of the importance of having powerful tools for macro models estimation and forecasting. Nonlinear time-series models, and regime-switching models in particular, seem to be a promising alternative and the literature has spent the last ten years in analyzing pros and cons of their usage. This dissertation focuses on the family of Smooth Transition models and attempts to offer a contribution to this debate. The thread linking the three autonomous essays in this thesis is the use of smooth transition modeling and estimation techniques to investigate some still open questions applied macroeconomics, namely 1) How to capture the dynamic adjustment of inflation expectations through time and 2) whether sovereign bond spreads be used to forecast GDP growth. It will be argued all through this dissertation that smooth transition models can provide some interesting answers to these questions.

Smooth Transition (ST) Models (Granger and Teräsvirta, 1993, Van Dijk and Franses 2000) have recently became popular among nonlinear time series models. Indeed, their flexible structure is suitable to model various phenomena: from real exchange rates dynamics (Bereau, Lopez, Villavicencio and Mignon, 2010), to stock returns and house prices movements (Fredj and Yousra, 2004 and Balcilar, Gupta and Shah, 2010) and to the relationship between GDP growth rates with the term spread (Brunetti and Torricelli, 2009). Although these models are unanimously recognized as a powerful tool for capturing some particular characteristics of macro/financial time series, their specification and estimation, and the evaluation of their forecasting performance are anything but simple tasks. This work also aims at filling the gap by shedding some new light on the computational aspects of the estimation of this type of models.

The plan of the thesis is the following. The second essay is mainly of a technical nature and it deals with Smooth Transition models in general, underlining their complex characteristics from the estimation and the asymptotics point of view. The first and third essays instead, are more of an applied type and they show that using smooth transition modeling may improve our understanding of inflation expectations dynamics on the one side, and of the growth-spread relationship on the other side, as these two are phenomena with relevant nonlinear characteristics. Here follows a more detailed description of each of the three essays.

The first essay, *"How Rational are Rational Expectations: New Evidence from Well Known Survey Data"*, is a first attempt to provide an empirical application of smooth transition modeling techniques to expected inflation data focusing on a set of European countries. The so called 'probabilistic approach' is used to derive a quantitative measure of expected inflation from qualitative survey data for France, Italy and the UK; The United States are also included in the dataset through of the Michigan Survey of Consumers expectations series. First, the standard tests to assess the degree of rationality of consumers inflation forecasts are performed. Evidence for cointegration between actual and expected inflation is found in the countries under investigation, but notably the standard Error Correction (ECM) specification does not seem satisfactory as residuals display some unexplained heteroskedasticity. As a consequence a Smooth Transition Error Correction Model (STECM) of the forecast error is specified and estimated via standard Nonlinear Least Squares (NLS). The STECM specification was chosen for two main reasons. Technically, heteroskedasticity in the residuals may signal neglected nonlinearity which a STECM is able to capture. More importantly, the nonlinear ECM model allows us to quantify what we call the *strategic stickiness* in the long-run adjustment process of expectations. With this term we refer to a nonlinear type of weak rationality reminiscent of the inertia in expectations' adjustment that Fehr and Tyran (2001) document in their experimental setting as a by-product of money illusion. The inertia arises from nominal loss aversion in a context of strategic complementarities: people are reluctant to reduce nominal prices after a negative monetary shock because they expect that the others will do the same, actually yielding a higher nominal loss. Our intuition is that a somewhat similar effect might also be in place also in the process of formation of aggregate inflation expectations. Interestingly, our evidence is that consumers expectations do not generally conform to the prescriptions of the rational expectations hypothesis. In particular, we find that the adjustment process towards the long-run Rational Expectations equilibrium is highly nonlinear and it is asymmetric with respect to the size of the past forecast errors. These findings are particularly interesting because they can be interpreted as supporting the hypothesis that consumers are affected by some behavioral bias like for example money illusion. However, it seems that Italian and American consumers' expectations are much more strategically sticky than French and English ones and this is certainly a result that needs to be explained.

Dealing with empirical applications, robustness checks are always in order. Trying to understand whether the STECM estimation results of the first essay were robust enough made us realize that the standard usage of the NLS estimation technique deserved further attention. That is exactly how the second essay was born. More specifically, in *"Some Pitfalls in Smooth Transition Models Estimation: A Monte Carlo Study"*, Monte Carlo simulation is used to assess whether the standard application of Nonlinear Least Squares to Smooth Transition models, yields estimates with desirable asymptotic properties. The paper analyses two notable examples in the ST family, namely Smooth Transition Autoregressive (STAR) models and Smooth Transition Error Correction (STECM) models, and it compares the asymptotic performance of NLS with maximum likelihood based estimation by means different algorithms. It has two main results. First, this standard application of nonlinear least squares with a 'concentrated sum of squares function' is something that needs to be used with caution, as it may yield biased and inconsistent estimates, especially when faced with small samples. Second, the reason beneath the poor performance of both the nonlinear least squares and the maximum likelihood estimators lays in the corresponding loglikelihood function, which displays multiple optima. This issue is thoroughly investigated by means of a global optimization procedure.

Finally, the third essay, *"Forecasting Italian GDP Growth through the Term Spread:*

How Many Stars to the STAR Model?", is devoted to comparing the out-of-sample forecasting performance of smooth transition models vis-a-vis other competing models like linear autoregressive models and time varying parameters models. The ground of the comparison is the prevision of Italian economic activity as a consequence of changes in the term spread, i.e. the difference between long-run and short-run interest rates, in the period from 1997 to 2007. The study aims at being the natural extension of the paper by Brunetti and Torricelli (2009), where a STAR specification was used in-sample, while for sake of simplicity forecasts were obtained by means of a probit model. Indeed, producing one-step forecasts for a STAR model is a manageable task, but multi-step forecasts are much more difficult as the multivariate distribution of the model error will enter the nonlinear data generating process. As a consequence, several numerical methods have been proposed to overcome the problem. As a first step, we robustly estimate Brunetti and Torricelli's logistic STAR model by means of Maximum Likelihood in a reduced time sample (1997-2005). As expected, we see that the use of this estimation technique substantially changes the parameters estimates with respect to the NLS estimates of the original paper. Secondly, n-step ahead forecasts are computed over the period 2005-2007 by using the standard methods proposed by Lundbergh and Teräsvirta (2002), namely *Monte Carlo* and *Bootstrap* Forecasts. Thirdly, by exploiting Granger's (2008) suggestion that 'any non-linear model can be approximated by a time-varying parameter linear model", we forecast the growth-spread relationship by using its equivalent representation as a state-space model with time-varying parameters. The performances of these three forecasting methods are compared by looking at their accuracy in comparison to the one of the linear autoregressive benchmark, and several consecutive forecast horizon are put under investigation. Two are the main results of the paper. For short-term forecast horizons the STAR model is quite effective in predicting GDP growth. However, the farther the forecast horizon, the worst its performance gets in comparison to its equivalent time-varying competitor. In addition, although none of the considered models is able to capture the turbulences of the 2007-2008 financial crisis, using forecasts combination schemes substantially increases the predictions' accuracy.

Chapter 1

How Rational Are Rational Expectations? New Evidence from Well Known Survey Data

Abstract This paper provides further evidence in favor of less than fully rational expectations by making use two instruments, one quite well known, and the other more novel, namely survey data on inflation expectations and Smooth Transition Error Correction Models (STECMs). We use the so called 'probabilistic approach' to derive a quantitative measure of expected inflation from qualitative survey data for France, Italy and the UK. The United States are also included by means of the Michigan Survey of Consumers' expectations series. First, we perform the standard tests to assess the 'degree of rationality' of consumers' inflation forecasts. Afterwards, we specify a STECM of the forecast error, and we quantify the *strategic stickiness* in the long-run adjustment process of expectations stemming from money illusion. Our evidence is that consumers' expectations do not generally conform to the prescriptions of the rational expectations hypothesis. In particular, we find that the adjustment process towards the long-run equilibrium is highly nonlinear and it is asymmetric with respect to the size of the past forecast errors. We interpret these findings as supporting the money illusion hypothesis[1].

[1] I want to thank both the participants of the Irish Society for New Economists meeting (Dublin, September 2010) and those of the second Workshop in Macroeconomics and Financial Economics (Farmagusta, May 2011) for their interesting suggestions on earlier versions of this paper. A special thanks goes to Marco Tucci for his guidance and support throughout my research. I also acknowledge Christofer Carrol, Paul De Grauwe, Pablo Rovira Kaltwasser, Sophie Bereau and the members of the Center of Economic Studies and the University of Leuven for their helpful comments. I am particularly grateful to Christina Gerberding who provided me with her data on experts' inflation expectations, and to Mehmet Balcilar who gently sent me his unreleased R-package for STAR tests. The usual disclaimer applies.

JEL Codes: C 22, D 84, E 31.

Keywords: Nonlinear error correction, inflation expectations, sticky expectations.

1.1 Introduction

Inflation expectations are kept under close watch by many: business consultants, investors, policy makers, and last, but not least, economic researchers. Yet, dealing with expectations is a very complex task since it involves two orders of difficulties. First, expectations are by nature unobservable, hence one needs to find a way to track them as closely as possible. Second, even after a good proxy for expectations is found, one still needs to understand what is the mechanism underlying their formation. More specifically, many efforts of the literature have been concentrated on understanding to which extent do expectations conform to the rational expectations hypothesis (REH) (Muth 1961, and Lucas 1987). On the other side, relatively few have dealt with investigating whether behavioral insights other than the generic notion of 'inattentiveness' play a role in explaining inflation expectations dynamics. This paper aims at filling the gap by using some recent advances in nonlinear time series econometrics.

Recently, the problem of unobservability of expectations has partially been overcome thanks to the availability of direct survey data. These kind of data are very valuable because they yield direct observations of inflation expectations without the need of a priori assumptions on their nature.[2] Nevertheless, the literature is far from having reached a consensus on what mechanism underlies the process of expectations formation and adjustment.[3] In particular, to the best our knowledge there have been no attempts so far to study the long-run adjustment process of inflation expectations without renouncing to the assumption of linearity, implicit in the idea of perfect rationality.[4]

[2] To be more precise, when using survey data one still needs a priori assumptions, but only on the form of the distribution of aggregate inflation expectations. For example, the Carlson and Parkin's (1975) method we also employ to convert qualitative survey data into quantitative ones, assumes a logistic distribution function.

[3] The contributions on the degree of rationality of expectations are several. See for example Berk (1999, 2000), Arnold and Lemmen (2006), Forsells and Kenny (2002), Gerberding (2009) Curto and Milet (2006), and Pjafary and Santoro (2009) among others.

[4] The REH posits that inflation expectations should have three testable characteristics: long-run unbi-

While the literature customarily tests the degree of rationality of expectations within the standard (linear) cointegration framework (Engle and Granger 1987, and Johansen 1991), we use a novel nonlinear cointegration approach enabling us to understand what influences the speed of adjustment of expectations in the long-run, and whether there are significant asymmetries in such adjustment process. More specifically, we use Smooth Transition Error Correction Models (STECMs), a flexible econometric specification which captures the long-run dynamics of variables with a nonlinear-asymmetric adjustment towards the equilibrium.[5] Due to their demanding requirements in terms of large samples availability, STECMs so far have been applied mainly to financial variables like interest rates (Van Dijk and Franses, 2000; henceforth VDF), real exchange rates (Béreau, López Villavicencio and Mignon, 2010), stock returns (Jawadi and Kouba, 2004) and house prices (Balcilar, Gupta and Shah, 2010). Nevertheless, taking into account the intrinsic differences, we are convinced that applying STECMs to inflation expectations can shed some new light on the asymmetries inherent to the long-run adjustment process of expectations, thereby providing useful insights both to policy makers and to researchers.

In this work, we employ the standard 'probabilistic approach' (Carlson and Parking 1975, Berk 1999) to derive a quantitative measure of expected inflation from the European Commission's (EC) Consumer Survey data[6]. Our sample comprehends 298

asedness, 'efficiency' with respect to available information, and mean reversion with respect to the forecast error's long-run 'rational' value. The latter feature was first noted by Bakhshi and Yates (1998), who start from observing that both inflation and inflation expectations generally display a unit root and hence their interpretation of the REH is that in the long run they should cointegrate, possibly with coefficients of the cointegrating vector equal in absolute value. Clearly, such definition involves the notion of a constant (linear) adjustment process.

[5]A STECM model can be viewed as a generalization of the standard linear ECM model proposed by Engle and Granger (1987), allowing for a nonlinear adjustment mechanism. In this type of models the standard constant feedback parameter is replaced by a continuous function, the so called transition function, which is bounded between $(0, 1)$. Generally the transition function is chosen to be either a logistic function, when one tries to capture sign asymmetries or a second order logistic function, when size asymmetries are thought to be more important. For a detailed description STECMs please refer to Anderson (1995), Van Dijk and Franses (2000), and Kapetanios, Shin and Snell (2003).

[6]Even though this method is quite standard in the literature, there are many authors pointing at its drawbacks mainly due to its assumption of a normal distribution of expectations. Indeed, many methods of correction have been proposed (we chose the one of Berk, 1999) but also many alternative methods are

monthly observations (1985-2009) for France, Italy and the UK. For sake of comparability with previous studies, we also include the US in the sample by means of the Michigan Survey of Consumers' expectations series. More specifically, France and Italy are included as inflation targeting countries under the influence of the ECB, while the UK and the US represent our (non inflation targeting) control group. Indeed, many studies point out that inflation targeting might be a key variable to explaining inflation expectations anchoring process to the long-run target[7].

First, we perform the standard tests to assess the 'degree of rationality' of inflation expectations and, like others in this literature, we infer that consumers behave quite differently than what the REH postulates. Afterwards, we use a STECM model of the forecast error to test for what we label *strategic stickiness.* With this term we refer to a nonlinear type of weak rationality reminiscent of the inertia in expectations' adjustment that Fehr and Tyran (2001) document in their experimental setting as a by-product of money illusion[8]. It is the inertia that arises from nominal loss aversion in a context of strategic complementarities: people are reluctant to reduce nominal prices after a negative monetary shock because they expect that the others will do the same, actually yielding a higher nominal loss. Our intuition is that a somewhat similar effect is in place also in the process of formation of aggregate inflation expectations.

From our estimation of the STECM for the forecast error we draw two main results. First, consumers tend to over-estimate inflation both in the short and long-run. Second, *strategic stickiness* does play a role role in shaping the expectations long-run adjustment dynamics, notwithstanding that it is not of the same entity in all countries. Furthermore, big shocks, whatever their sign is, have generally a greater influence in speeding up the adjustment process than small ones.

available. Nevertheless, evaluating which of them performs better is beyond the scope of this paper, and for a more detailed treatment of these issues we suggest to refer to Nardo (2003).

[7]See for example Gürkaynak, Levis and Swanson (2006) and Yigit (2010).

[8]The term money illusion seems to have been coined by Irving Fisher as "the failure to perceive that the dollar, or any other unit of money, expands or shrinks in value" (1928, p.4). Fehr and Tyran (2001) give a somewhat more precise definition, by saying that one is prone to money illusion if i) his/her objective function depends on both nominal and real magnitudes and ii) He/she perceives purely nominal changes affecting his/her opportunity sets. For a thorough treatment of money illusion please see Shafir, Diamond and Trevsky (1997), Fehr and Tyran (2001, 2007, 2008).

It is important to notice that many factors may be responsible for the nonlinear dynamics we find in our data: for example slow information diffusion (Mankiw and Reis 2002, Carrol 2003), Near Rational behavior towards inflation (Akerlof and Yellen 1985, Akerlof, Dickens and Perry 2000, Ball 2000, Maugeri 2010), and in general all the decision heuristics implying less then full adjustment to errors. Indeed, our smooth transition model for the adjustment process can be viewed as a reduced form of structural models of expectations formations accounting for nonlinearities due to a number of less than fully rational decision mechanisms, the most parsimonious of those being money illusion, hence our decision to focus on it throughout the paper.

The rest of this paper is organized as follows. Section 2 gives a general description of our dataset. Section 3 develops the formal procedures we use to assess various theories of expectation formation: first we describe the hypotheses of adaptive expectations and sticky information diffusion, then rational expectations tests both in 'weak' and 'strong' form are discussed, and finally the*strategic stickiness* issue is addressed. Section 5 presents the results of our empirical investigation and section 6 offers some concluding remarks.

1.2 The Data

Increasing availability of direct survey measures of inflation expectations caused a massive interest of the literature in this topic. The pioneering survey study on consumers expectations is the Survey of Consumers devised in the late 40s by George Katona at the University of Michigan. parallely, from 1968 to 1990 the National Bureau of Economic Research, and later the Federal Reserve Bank of Philadelphia, conducted the first survey on the 'professional' views on expectations, i.e. the Survey of Professional Forecasters (SPF). The European Commission has started in 1985 to follow the lead of its foreign rivals, by elaborating surveys on both consumers' and professional forecasters' expectations for the Euro area[9].

[9]Both European surveys are basically designed following the US example. One main difference though, is that while the survey of consumers provides data both at a country level of disaggregation and at the Euro-area level of aggregation, the European SPF is available only for the Euro area as an aggregate.That is the main reason why in order to proxy the experts' expectations we decided not to use the EU-SPF data, but

Our dataset is composed by monthly CPI inflation rates and inflation expectations series both for consumers and for professionals, from January 1985 to October 2009[10]. The sample comprehends three main European countries, France, Italy and the United Kingdom, and the United States. As we already pointed out, we wanted to include two EMU-inflation targeters as opposed to two non targeters because our we also wanted to see whether the monetary policy of the central bank does make a difference in shaping up the adjustment process of expectations[11].

The inflation rate series are taken from, respectively, the Centre for European Economic Research (ZEW), the Italian statistical Office (ISTAT), the English Office for National Statistics, and the US Bureau of Labor Statistics. The series are all unadjusted for seasonality.

The consumers inflation expectations series for France, Italy and the UK are derived by applying the so called 'Probability Approach' (Carlson and Parkin, 1975) to the qualitative data of the European Commission Survey[12]. Following Berk (1999 and 2000), we apply a rescaling of the expectations series by means of 'perceived inflation', as the literature shows that such rescaling dramatically improves the representativeness of the derived expectation measure. Figure 1 displays the series of inflation and consumers expectations over the chosen time sample.

The professionals forecasters' expectations series for Italy, France and the UK are elaborated from the London based firm Consensus Economics. From 1989, this firm asks to renewed experts at the beginning of each month to forecast the development of important macroeconomic variables[13]. The US series is the SPF measure elaborated by the Federal Reserve Bank of Philadelphia. One of the main criticisms made to the use of consumers based measures of expectations is that survey takers might have little incentives to correctly state their perception of future price developments. On the contrary, business experts' opinions should be driven by market forces to track actual inflation as closely as possible. As a matter of fact, a comparison of Figure 3.1

the Consensus Economics data, made available to us by the courtesy of Christina Gerberding.

[10]Actually, the French range of available observations is a little bit shorter than the other, since the inflation rate series starts from 1990.

[11]Needless to say, the choice of the subset of countries is also motivated by data availability considerations.

[12]See the Appendix for more details.

[13]We really thank Christina Gerberding for making these data available to us.

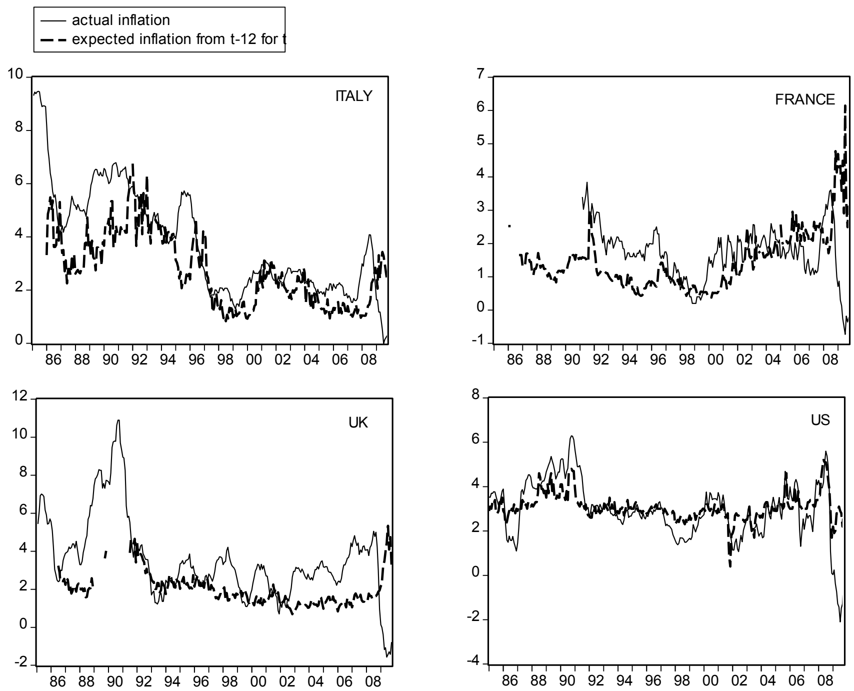


Figure 1.1: Consumers' Forecasts of Inflation

with Figure 1.2 clearly reveals that on average experts have a lower forecast error than consumers.

What is also clear from Figure 3.1, is that consumers were not able to forecast the 2007-2008 financial crisis and the subsequent trough of inflation[14]. Even though there seems to be a strong relationship between actual and expected inflation, consumers have underpredicted and overpredicted inflation much more than experts, at least in the first part of the sample. Moreover, after the switch to the common currency in 1999, European consumers seem to have believed to the low inflation commitment of the European Central Bank (ECB) and anticipated the consequent downward trend in inflation.

To provide a more quantitative evaluation of the forecast performance of both consumers and experts, Table 1.1 provides some standard indicators: the Mean Error (ME), showing the average forecast error over the sample period, the Mean Absolute

[14]Actually, since even the great part of economists were not able to predict the financial crisis, we did not expect consumers to do so. Unfortunately, our series of experts' forecast arrives until 2006 for the majority of countries, hence we cannot give any quantitive judgement of the the experts' forecast performance in 2008.

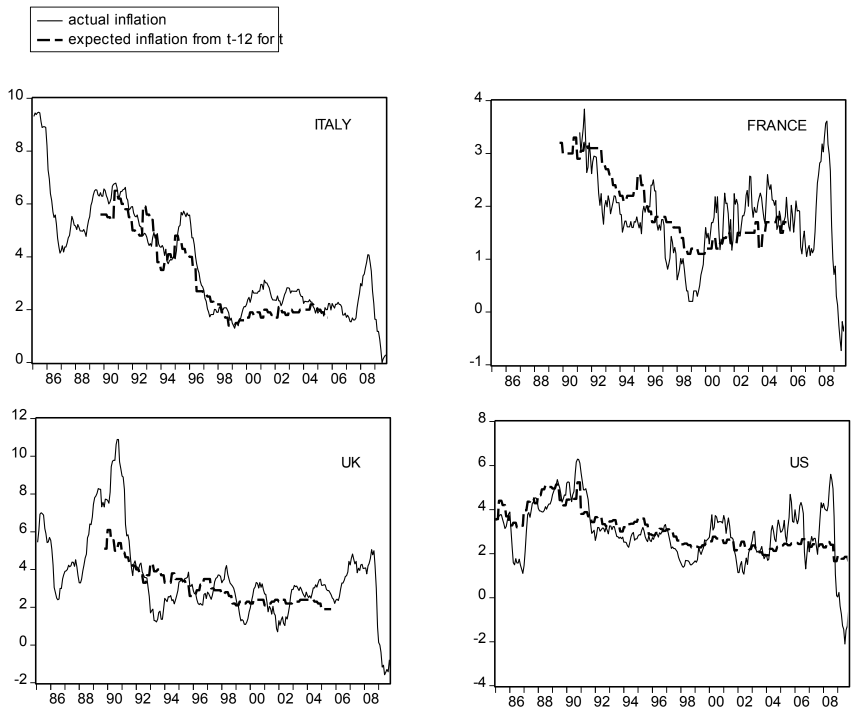


Figure 1.2: Experts' Forecasts of Inflation

	France	Italy	UK	US	France	Italy	UK	US
Full period 1985-2009	Consumers expectations				Experts expectations			
ME	-0.31	-0.99	-1.21	0.11	0.09	-0.37	-0.25	0.12
MAE	0.90	1.13	1.58	0.71	0.48	0.55	1.07	0.76
RMSE	1.11	1.42	2.01	1.00	0.56	0.67	1.52	1.02
Subperiod: 1985-1999	Consumers expectations				Experts expectations			
ME	-0.81	-1.26	-1.25	-0.09	0.41	-0.28	-0.24	0.37
MAE	0.87	1.35	1.45	0.64	0.51	0.54	1.27	0.61
RMSE	1.02	1.66	2.02	0.80	0.59	0.67	1.80	0.77
Subperiod: 1985-1999	Consumers expectations				Experts expectations			
ME	0.15	-0.63	-1.16	0.41	-0.41	-0.55	-0.27	-0.25
MAE	0.93	0.83	1.75	0.81	0.43	0.57	0.72	0.99
RMSE	1.19	1.00	2.00	1.24	0.52	0.66	0.81	1.31

Note: ME= Mean Error, MAE= Mean Absolute Error, RMSE= Root Mean Squared Error

Table 1.1: Forecast performance statistics: Consumers' vs experts' expectations

Error (MAE) which measures how close are predictions to the actual inflation rates, and the Root Mean Squared Error (RMSE) which represents the expected value of the squared error loss, hence it is less sensitive to large forecast errors or outliers. Consistent with our graphical evidence, experts have on average a better forecast performance than consumers, since their MAEs and RMSEs are systematically lower. However, both consumers and experts seem to frequently commit large but counterbalancing errors, as shown by the fact that the ME is always much lower than the MAE. Another interesting finding that emerges from Table 1.1 is that experts seem to have interiorized the credibility strategy of the ECB much more than the public, as shown from the systematically lower MAE and RMSE in the second subsample. On the other hand, consumers do not seem to have a clear idea of this strategy in every country: only Italian consumers have decreased their MAEs and RMSEs in the Euro-era subsample, while French and US citizens have worsened their forecast performance. For English consumers it is not possible to give a precise judgment, since in the second subsample the MAE increases but the RMSE decreases. On the contrary, our US reference point indicates that during the pre Euro-era American consumers had a much clearer picture in their mind of what was happening to inflation than in the following decades.

A final word on comparability of our expectations measures. Our results are broadly in line with the previous findings of the literature, which report a RMSE for European aggregate inflation expectations between 0.47 and 1.29 (Forsells and Kenny, 2002). Again, one could take the study on the US by Lloyd (1999) as a reference: He finds a RMSE for the period 1983-1997 between 1.09 and 1.57, also very close to our estimates.

1.3 Assessing Theories of Rationality

This section briefly describes the different theoretical hypotheses we will test throughout the paper, with special attention to their econometric implementation. The section is organized chronologically, it starts by illustrating the adaptive expectations hypothesis and it complements it with the much newer notion of sticky information diffusion (Mankiw and Reis 2002; Carrol, 2003). Subsequently, the Rational Expectations Hypothesis is thoroughly described in its 'weak' and 'strong' form, although the section gives a prominent role to what we call *strategic stickiness*, that is to say 'weak rationality' with asymmetric adjustment process towards the long-run equilibrium. Section 4 then will conclude the analysis, by dealing with estimation issues and presenting our empirical results.

1.3.1 Adaptive Expectations and Sticky Information Diffusion

The first idea on expectations was that people could revise their predictions according to their past forecast errors. The Adaptive expectations hypothesis was suggested by Irving Fisher in 1930, and then it was formalized by Cagan (1956), Friedman (1957), and Nerlove (1958). The standard way to assess the degree of adaptiveness in consumers expectations is to estimate the following regression

$$\pi_t^e = \theta \pi_{t-1}^e + \xi(\pi_{t-1} - \pi_{t-1}^e) + \epsilon_t \qquad \text{(Adaptiveness)}$$

Here the parameter ξ assumes an important role, since it captures the speed of adjustment of current expectations to the past forecast error. However, as the recent literature on inattentiveness suggests, the speed of this adjustment mechanism depends not only on the subjective 'degree of adaptiveness', but it is also influenced by how fast the information is diffused in the economy and by how costly is obtaining and updating information sets. According to such considerations, equation (Adaptiveness) should be complemented by an equation trying to capture the dynamics of the information diffusion process like the following

$$\pi_t^e = \lambda_1 \pi_{SPF,t}^e + (1 - \lambda_1)\pi_{t-1}^e + \epsilon_t \qquad \text{(Sticky info)}$$

Equation (Sticky info) summarizes the core of Carrol's (2003a) model of 'epidemiological' diffusion of information about inflation, and it posits that households slowly update their information sets from news reports, which are in turn influenced by professional forecasters. In such a context, $\pi_{SPF,t}^e$ is the mean inflation at time t as predicted by experts (i.e. Consensus Economics' forecasts for France, Italy and the UK, and SPF forecasts for the US), and the coefficient λ_1^{-1} is interpreted as the average updating period for households' information sets. Please notice that equation (Sticky info) considers a type of expectations' stickiness which is only due to the intrinsic difficulty to get updated information about inflation from the news. On the other hand, the *strategic stickiness* we put our emphasis on is of a different type, in the sense that it has to do with strategic complementarities among agents' forecasts when they are faced with nominal evaluations.

1.3.2 'Strong' Rationality

The Rational Expectations Hypothesis (REH) (Muth, 1961) in its 'strong version', posits that inflation expectations should have two testable characteristics: long-run unbiasedness, and 'efficiency' with respect to available information. As we will see later, a weaker version of the REH assumes the expectations only display mean reversion with respect to their long-run 'rational' value. The idea of rational expectations is that agents can match on average the predictions of the relevant economic models. This translates into an estimated forecast error which should be centered around zero (unbiasedness property) and should not be correlated with variables included in their information sets at the time predictions were made (orthogonality property). Tests for efficiency in the use of information are extensively undertaken by the current literature, hence in this work we will start our analysis by focusing on the a investigation of unbiasedness property[15].

It is common practice in papers using survey data on expectations to test the strong version of the REH by estimating a series of OLS equations of the following type

$$\pi_t - \pi_t^e = \alpha + \epsilon_t \qquad \text{(Rationality 1)}$$

$$\pi_t = \alpha + \beta\pi_t^e + \epsilon_t \qquad \text{(Rationality 2)}$$

$$\pi_t - \pi_t^e = \alpha + (\beta - 1)\pi_t^e + \epsilon_t \qquad \text{(Rationality 3)}$$

where we indicate with π_t the actual inflation rate for period t, and with π_t^e the expected inflation rate for period t calculated in period $t - 12$. By analyzing the properties of the estimated forecast error of these equations and the accurateness of the parameters estimates, gives an idea of whether the REH is verified on average[16].

[15]See Gerberding (2007) and Forsells and Kenny (2002) among others.

[16]Actually, this very simple approach has attracted some criticisms in the literature. For example, Andolfatto, Hendry and Moran (2008) argue that this type of estimations suffers from small sample bias which

1.3.3 'Weak Rationality'

Many authors claim that the strong version of the REH, the one involving unbiasedness and efficiency, might be 'too strong', given the informational frictions and transaction costs present in reality. What characterizes rational expectations according to many, is that there is mean reversion of expectations towards the correct mean inflation value, that is to say π_t^e and π_t cointegrate in the long-run. The pioneering work on this issue by Bakhshi and Yates (1998) starts exactly by observing that both inflation and inflation expectations are $I(1)$ variables, hence their dynamic interpretation of the REH is that in the long-run they should cointegrate, possibly with coefficients of the cointegrating vector equal in absolute value. This interpretation of the REH yields two main implications: i) no matter how long is the adjustment, time movements of expectations and inflation rates should be linked in the long-run ii) The adjustment from the short-run to the long-run always occurs with the same constant intensity, captured by a linear-constant adjustment function.

In order to be more clear let us assume that π_t^e and π_t are given by

$$\pi_t = \pi_t^e + \varepsilon_{1t} \tag{1.1}$$
$$\pi_t^e = \pi_{t-1}^e + \varepsilon_{2t}$$

Where ε_{it} $i = 1, 2$ has the standard properties. Then the weak version of the REH posits that there is a cointegrating relationship between the two variables of the type

$$\pi_t = \alpha + \beta\pi_t^e + z_t \tag{1.2}$$
$$\text{with } z_t = \rho_1 z_{t-1}, \quad |\rho_1| < 1 \text{ and } \alpha = 0\ ,\ \beta = 1$$

In practice, in order to understand whether (1.2) holds it is customary to perform a standard cointegration analysis on the two series and to test for the appropriate coefficients restrictions.

couples with endogenous learning dynamics, and this explains the induced bias in the forecast error. Furthermore, Symth (2008) sustain that studies testing for 'strong' REH by means of equations similar to (Rationality 1), (Rationality 2) and (Rationality 3) are fatally flawed because they incorrectly assume that expected inflation is measured without error. Here, we are aware of these criticisms and we use this simple analysis only as a starting point for our subsequent nonlinear investigation.

1.3.4 *Strategic Stickiness*: 'Weak Rationality' and Asymmetric Adjustment

Sustaining that people are on average correct in their forecasts of inflation is one thing, sustaining that they perform these correction tasks always in the same way is something different. Indeed, the standard notion of 'weak rationality' (equations 1.1 and 1.2) implicitly assumes that in the long-run there is a constant linear adjustment process linking expectations to the actual mean value of inflation. However, Fehr and Tyran (2001, 2004 and 2008) suggest that expectations of nominal variables often display a sticky and asymmetric adjustment. More specifically, their experiments show that in a context where decisions are confined with nominal magnitudes, people are reluctant to reduce nominal prices after a negative monetary shock because they anticipate that the others will do the same, hence actually magnifying the aggregate nominal inertia. This type of inertia we call*strategic stickiness*, should be much larger after a negative nominal shock than after a positive one and it should also depend on the size of the shock.

Assuming that the data generating processes of π_t^e and π_t still follow (1.1), then *strategic stickiness* implies that there is a cointegrating relationship for the two variables of the type

$$\pi_t = \alpha + \beta\pi_t^e + z_t \tag{1.3}$$
$$\text{with } z_t = F(z_{t-1}) + u_t, \quad z \; stationary, \text{ and } \; u_t \overset{iid}{\sim} (0, \sigma_u^2)$$

Where $F(.)$,the *transition function* is a continuous nonlinear functional form bounded in the $(0, 1)$ interval, capturing asymmetries in the adjustment process stemming from *strategic stickiness*. Notice that here we chose the simple case where the deviation from the long-run equilibrium z_t behaves like a first order stochastic process, but clearly a more general case involves a transition function $F(z_{t-d})$ with an higher lag order $d = \{1, 2, ...\}$.[17]

Our aim is to investigate *strategic stickiness* by specifying a STECM model of

[17] Notice also that the stationarity condition for z_t in this case is more complicated than the standard one, because it depends on the chosen form of the $F(.)$ function.

consumers' forecast error with the general structure:

$$Dy_t = \varphi_1' \mathbf{w}_t + F(z_{t-d}; \gamma, c)\psi_2' \mathbf{w}_t + \varepsilon_t \tag{1.4}$$

where y_t is in our case either π_t or π_t^e, depending on the specification, and x_t is respectively either π_t^e or π_t, the cointegration relationship is indicated by $z_t = \pi_t - \alpha - \beta\pi_t^e$, α and β are the ones estimated during the preliminary cointegration analysis, $\mathbf{w}_t = (1, \widetilde{\mathbf{w}}_t)'$, $\widetilde{\mathbf{w}}_t = (z_{t-1}, Dy_{t-1}, ..., Dy_{t-p+1}; Dx_t, ..., Dx_{t-p+1})'$, for $i = 1, 2$, $m = 2p - 1$. Finally $\varphi_1 = (\varphi_{10}, \varphi_{11}, ..., \varphi_{1m})'$ and $\psi_2 = (\psi_{21}, \psi_{22}, ..., \psi_{2m})'$ are parameters vectors to be estimated.

There are two main reasons why we think this approach is valuable. First, from the econometric point of view it builds on standard cointegration analysis and it amends some of its weaknesses by assessing possible neglected nonlinearities in the ECM adjustment process. Second, from the theoretical point of view, the nonlinear adjustment mechanism is a flexible specification allowing for asymmetric effects of shocks which differ in size and sign: the choice of the transition function can give us precise indications on which type of asymmetry matters more to explain agents' strategic stickiness. Finally, the STECM approach has the further advantage of not having to impose any prior on rationality, and to 'let the data choose' the type of nonlinearity better fitting them by means of the appropriate transition function.

1.4 Results

Our empirical assessment of theories of rationality starts by performing the standard rationality tests that the literature has proposed so far. As we already pointed out, there are already some studies in the literature analyzing the properties of both the EU and the US consumers expectations series[18]. However, most of them employ data up to

[18] For example, Forsells and Kenny (2002) use the EC's consumers data to analyze the properties of expected inflation for the euro area as an aggregate. Arnold and Lemmen (2008) also use the EC's Consumer Survey to assess whether inflation expectations have converged and whether inflation uncertainty has diminished following the introduction of the Euro in Europe. Gerberding (2009) provides an interesting comparison between consumers' and experts' expectations in France, Italy, Germany and UK.

2006, hence it is interesting to see whether the results change with an updated dataset[19]. The section then continues by proposing our strategy to assess 'weak rationality' in the form of *strategic stickiness*. We estimate a STECM model for the countries of interest and we analyze the properties of the estimated transition function so that we can have an indication of what influences the speed of adjustment of expectations in the long-run. The tests for adaptive expectations, sticky information diffusion and rational expectations are all implemented by means of heteroskedasticity corrected OLS, while the STECM estimation for *strategic stickiness* is done by means of nonlinear least squares.

1.4.1 Adaptive Expectations and Sticky Information Diffusion

To which extent so consumers correct their expectations looking at past errors? and how much does the speed of diffusion of news about inflation influences this process? The results of the estimation of both equations (Adaptiveness) and (Sticky info) in Table 1.2 can provide an answer to these questions.

In the adaptive expectations test, the adjustment coefficient to past errors is quite small for France, UK and the US, averaging at 2%; at the same time the average updating time for those countries is estimated to be very different, as people update their information sets respectively once every 4,17 and 10 months. Italy is a special case though, since the adjustment coefficient is very high (14%) but the average updating period is the longest, about 21 months. The estimated θ coefficient in equation (Adaptiveness) is very close to one in all specifications and hence it is of particular interest for two reasons: from the theoretical point of view, there is an high degree of backward looking behavior in expectations formation dynamics; from the econometric point of view, there is an high degree of persistence in inflation expectations, which needs to be handled with the appropriate techniques. As a consequence of that, some of the rationality tests we will apply in the next paragraph handle such persistence with the

[19] Clearly, we are aware that a longer time span comes at the cost of maybe having a structural break and/or one or more outliers in the sample due to the 2008 financial crisis. Nevertheless, since our focus is on the effects of the size of shocks on the adjustment of consumers expectations, we decided keep this long time sample, momentarily leaving the model stability issue in the background.

type of test	Country					
	France					
	θ	ξ	λ_1	$(1/\lambda_1)$	R^2	N
Adaptiveness	0.99	0.03	-	-	0.87	215
	0.00	0.08				
Sticky info	-	-	0.04	25.39	0.85	186
			0.01	0.00		
	Italy					
Adaptiveness	0.94	0.14	-	-	0.88	275
	0.00	0.00				
Sticky info	-	-	0.21	4.69	0.89	182
			0.00	0.00		
	UK					
Adaptiveness	0.98	0.01	-	-	0.86	248
	0.00	0.32				
Sticky info	-	-	0.17	5.88	0.89	263
			0.00	0.00		
	US					
Adaptiveness	0.99	0.02	-	-	0.68	297
	0.00	0.26				
Sticky info	-	-	0.10	9.85	0.70	297
			0.00	0.00		

Notes: small numbers under the estimates are p-values. N is number of observations. Equations are estimated by OLS using covariance matrix corrections suggested by Newey and West (1987).

Table 1.2: Test for adaptive expectations and sticky information diffusion

appropriate techniques.

Rationality

In what follows we assess the REH in its so called 'strong' and 'weak' version. The general way to test for unbiasedness is estimating equation (Rationality 2) and then testing the null $H_0 : (\alpha, \beta) = (0, 1)$. However, since Holden and Peel (1990) showed that the condition $\alpha = 0$ is both necessary and sufficient for unbiasedness, while $(\alpha, \beta) = (0, 1)$ is not necessary, we can simply use equation (Rationality 1) to see whether expectations error are centered around the right value and then test if such value can be conveniently restricted to zero. Equation (Rationality 3) is simply a way to augment equation (Rationality 2) in order to cross-checks the previous results and to see whether all available information is fully exploited. Please notice that all these three equations are expected to have no predictive power under the null of rationality. Table 1.3 gives the results of the three estimation for each country in the sample.

Our estimates of equation (Rationality 1) suggest that in our sample the necessary condition for unbiasedness is never met, the only exception being the US. Furthermore, the sufficient condition is also never satisfied for the all four countries, as indicated by the significant Chi-squared statistics of equations (Rationality 2)[20]. The results of equation (Rationality 3) provide a further confirmation of what we found so far, as the parameters are generally not close to the their theoretical values $(0, 1)$. Our results are in line with the ones of Forsells and Kenny (2002) and Pfajfar and Santoro (2010), and they confirm the poor forecast performance of consumers. Over the full time sample, which probably contains at least one structural break and some outliers due to the current financial crisis, expectations are systematically overestimated (α is always positive, the only exception being the US), as also confirmed by the estimated β which is above 1 in all countries except France.

Clearly that these first tests of 'strong rationality' give such results, does not exclude that other notions of rationality are still in place. A somewhat weaker notion

[20]Notice that the R-squared for the three equations is not always as low as expected, but this has probably to do with our variables being integrated and hence it is probably a spurious result. We will correct for this in the subsequent cointegration analysis.

type of test	Country					
	France					
	α	β	(1-β)	R^2	χ^2 for H_{01} and H_{02}	N
Rationality (1)	0.31	-	-	0.00	4.48	220
	0.04				0.03	
Rationality (2)	1.60	0.08	-	0.01	55.67	220
	0.00	0.60			0.00	
Rationality (3)	1.60	-	1.92	0.45	-	220
	0.00		0.00			
	Italy					
Rationality (1)	0.99	-	-	0.00	58.32	281
	0.00				0.00	
Rationality (2)	0.71	1.11	-	0.66	62.62	281
	0.01	0.00			0.00	
Rationality (3)	0.71	-	0.89	0.02	-	281
	0.45		0.00			
	UK					
Rationality (1)	1.21	-	-	0.00	38.45	263
	0.00				0.00	
Rationality (2)	1.02	1.09	-	0.22	51.25	263
	0.04	0.00			0.00	
Rationality (3)	1.02	-	0.91	0.00	-	263
	0.69		0.42			
	US					
Rationality (1)	-0.11	-	-	0.00	0.72	298
	0.40				0.39	
Rationality (2)	-1.73	1.53	-	0.48	7.38	298
	0.01	0.00			0.03	
Rationality (3)	-1.73	-	0.47	0.10	-	298
	0.01		0.02			

Notes: small numbers under the estimates are p-values. N is number of observations. Chi-squared statistics pertain to the null hypothesis H_{01}: α=0 in equation (1) and H_{02}: (α,β)=(0,1) in equation (2). Equations are estimated by OLS using covariance matrix corrections suggested by Newey and West (1987).

Table 1.3: Test for unbiasedness of consumers' expectations

of rationality might be more appropriate, especially once acknowledged that we are dealing with non stationary variables. Here we follow the approach first introduced by Bakhshi and Yates (1998) and we try to understand if expectations and inflation move together at least in a long-run perspective, i.e. they cointegrate. After performing the standard unit roots test on both variables (not shown), and having confirmed that they are all integrated of order one, we carried on the standard cointegration tests by Johansen (1981). Similarly to Gerberding (2006) we find is that there is a strong evidence for cointegration for France and Italy, while for the UK and US the evidence is a little bit milder. As a consequence we estimated the corresponding bivariate vector ECM of the form

$$D\pi_t = c_0 + c_\pi(\alpha\pi_{t-1} - \beta\pi^e_{t-1}) + \sum_{i=0}^{p} a_i D\pi_{t-i} + \sum_{i=0}^{p} b_i D\pi^e_{t-i} + \varepsilon_{\pi t} \qquad \text{(Inflation)}$$

$$D\pi^e_t = g_0 + g_e(\alpha\pi_{t-1} - \beta\pi^e_{t-1}) + \sum_{i=0}^{p} g_i D\pi^e_{t-i} + \sum_{i=0}^{p} h_i D\pi_{t-i} + \varepsilon_{et} \qquad \text{(Exp. Inflation)}$$

where c_0 and g_0 are constants, c_π and g_e are the ECM adjustment coefficients, and the lag length p is selected in preliminary VAR analysis (not shown). Tables 4 and 5 report the results of the estimation.

	Countries					
	Italy		France			
-lag lenght:	2		2			
-Trace test: H0: at most 1 CE	1.43	[p= 0.27]	5.76		[p=0.49]	
-Rank test: H0: at most 1 CE	1.43	[p= 0.27]	5.76		[p=0.49]	
	α	β	α	β	trend	constant
Cointegrating vector	1.00	-1.41 (-0.08)	1.00	-1.49 (-0.35)	0.02 (0.00)	-3.19
	VECM		VECM			
	inflation	expected inflation	inflation		expected inflation	
ECM adjustment coefficients:	-0.02 (-0.01)	0.10 (-0.02)	-0.04 (-0.02)		0.09 (-0.02)	
R2	0.20	0.21	0.07		0.25	
N	263		205			

Notes: standard errors in parentheses. N is number of observations. Equations are estimated by OLS.

Table 1.4: Test for cointegration between consumers expectations and actual inflation rates: Italy and France

	Countries				
	UK		US		
-lag lenght:	3		2		
-Trace test: H0: at most 1 CE	3.58	[p=0.48]	9.55	[p=0.04]	
-Rank test: H0: at most 1 CE	3.58	[p=0.48]	9.55	[p=0.04]	
	α	constant	α	β	constant
Cointegrating vector	1.00	-3.77	1.00	-3.65	8.31
		(-0.91)		(-0.43)	(-1.34)
	VECM		VECM		
	inflation	expected inflation	inflation	expected inflation	
ECM adjustment coefficients:	-0.05	-0.04	-0.01	0.07	
	(-0.017)	(-0.013)	(-0.01)	(-0.01)	
R2	0.18	0.22	0.19	0.15	
N	229		295		

Notes: standard errors in parentheses. N is number of observations. Equations are estimated by OLS.

Table 1.5: Test for cointegration between consumers expectations and actual inflation rates: UK and US

What we first notice from a general examination of the two tables is that the coefficients of the cointegrating vectors are different in absolute value. This seems to be against the definition of 'weak rationality', but from a broader perspective it also tells us that the existent long-run relationship between π_t^e and π_t involves also a systematic underprediction of inflation for all countries but for the UK, where the β coefficient is instead greater than one. Here we interpret this type long-run relationship as the 'ecologically rational' prediction for inflation, because it can be considered the outcome of one of the most parsimonious heuristic that people have given the available information sets: money illusion[21]. Indeed in a low and stable inflation environment like the EMU, reasoning in nominal terms and underestimating (low) future inflation can be a powerful and efficient rule of thumb to address the complicate issue of forming inflation expectations[22]. Furthermore the fact that the ECM adjustment coefficients are significant and with opposite signs indicates that also a more traditional mechanism is in place , namely the two-way feedback between inflation and expectations. More specifically, having a positive g_e and a negative c_π like in our case, suggests that not only

[21] A decision rule or an heuristic is defined as 'ecologically rational' if it exploits structures of information that are already in the environment, allowing the decision maker to save on information processing and gathering costs. For a broader perspective on this issue see Smith (2002) and Goldstein and Gigerenzer (2002).

[22] See for Akerlof, Dickens and Perry (2000), Lundborg and Sacklen (2006) and Maugeri (2010).

expectations adjust towards their 'ecologically rational value', but also actual inflation adjusts to the level expected by the public, as in the Friedman-Phelps framework.

Strategic Stickiness

The linear cointegration analysis we performed in the previous section leaves some issues unexplored. As we saw, there is some evidence for cointegration between inflation and its expectation, but the cointegrating relationship does not look like the one stemming from a rational behavior due to systematic biases. Furthermore, the VECM estimated residuals are not normally distributed and they display some heteroskedasticity that could arise from neglected nonlinearities. Our guess is that nonlinear asymmetric adjustment stemming from *strategic stickiness* could hide behind these results.

In order to shed more light on these issues, we employ the STECM modelling approach suggested by VDF (2000) and we start by estimating a conditional ECM model for the forecast error, as shown in Table 1.6[23].

As noted earlier, the linear ECM models do not seem to perform badly. Parameters significance is quite satisfactory and the residuals seem to be well behaved, a part from a problem of heteroskedasticity indicated by the high $ARCH(1)$ statistic. However, with the models in CECM form we are able to investigate the issue of neglected nonlinearity by applying the LM test by Luukkonen et al (1988)-VDF (2000) to past forecast errors z_{t-d}[24]. Indeed the results of the test, displayed in Table 1.7, show that the null hypothesis of linearity is rejected for several values of the lag length d of the past forecast error.

Beyond giving evidence of nonlinearities in the adjustment process which could stem from *strategic stickiness*, the test also gives us an indication of which of the past forecast errors is responsible for such nonlinearities, as indicated by the lag order d^* with lowest p-value (underlined in Table 1.7).

An other important indication on the type of *strategic stickiness* characterizing expectations is given from what the data choose to be the appropriate transition func-

[23] The ECM model is conditional in the sense that it isolates either equation (Inflation) or (Exp. Inflation) from the VECM, and it conditions it to an appropriate number of lags of the other endogenous variable.

[24] For technical details about the test, please refer to the appendix where we replicated an entire an example of estimation of a STECM from the paper of VDF (2000).

	France		Italy		UK		US	
dependent variable	Ye		Y		Y		Y	
	coefficients							
	constant	0.00	constant	0.03	constant	0.00	constant	-0.01
		0.89		0.05		0.99		0.72
	Z-1	0.08	Z-1	-0.04	Z-1	-0.03	Z-1	-0.04
		0.00		0.00		0.09		0.01
	Y-2	-0.18	Y-1	0.22	Y-1	0.36	Y-1	0.41
		0.01		0.00		0.00		0.00
	Y-1	-0.35	Y-2	0.22	Y-2	0.11	Y-2	-0.20
		0.03		0.00		0.11		0.00
	Y	-0.10	Y-3	0.18	Y-3	0.14	Ye-1	-0.06
		0.14		0.00		0.03		0.53
			Ye-2	-0.05	Ye-1	-0.04	Ye	0.41
				0.04		0.64		0.00
diagnostics								
	R2	0.25	R2	0.24	R2	0.19	R2	0.30
	DW	2.05	DW	2.01	DW	2.02	DW	1.97
	ARCH(1)	18.68	ARCH(1)	5.63	ARCH(1)	7.12	ARCH(1)	12.85
		0.00		0.02		0.01		0.00

Note: Equations are estimated by Nonlinear Least Squares using covariance matrix suggested by Newey and West (1981). Small numbers below the coefficients are p-values. For notational simplicity, Ye denotes the first difference of expected inflation and Y indicates the first difference of inflation.

Table 1.6: Estimation of the conditional ECM

Country: France							
Test	Null	d=1	d=2	d=3	d=4	d=5	d=6
F Test	H_o'	0.63	0.80	0.03	0.04	0.20	0.00
χ^2 Test	H_o'	0.63	0.79	0.03	0.04	0.20	0.00
Country: Italy							
Test	Null	d=1	d=2	d=3	d=4	d=5	d=6
F Test	H_o'	1.00	0.92	1.00	0.83	0.05	0.64
χ^2 Test	H_o'	1.00	0.91	1.00	0.82	0.05	0.63
Country: UK							
Test	Null	d=1	d=2	d=3	d=4	d=5	d=6
F Test	H_o'	1.00	1.00	0.01	0.00	0.00	0.20
χ^2 Test	H_o'	1.00	1.00	0.01	0.00	0.00	0.19
Country: US							
Test	Null	d=1	d=2	d=3	d=4	d=5	d=6
F Test	H_o'	0.56	0.69	0.63	0.05	0.00	0.00
χ^2 Test	H_o'	0.55	0.68	0.62	0.05	0.00	0.00

Note: p-values for LM-type tests for smooth transition error correction in the forecast error of consumers expectations. The test refers to the Conditional ECM specfication. The null hypothesis is given in the text. Underlined values indicate the lag lenght chosen by the test at the 1% or 5% significance levels.

Table 1.7: LM-type test for smooth transition error correction in consumers' forecast error

tion. In the literature, three types of transition function are generally used. When one suspects that it is *sign asymmetry* that matters more for the adjustment process of the endogenous variable, one should use the logistic transition function. For example, there is evidence that many macroeconomic and financial variables seem to be affected in an asymmetric way by positive and negative shocks[25]. In this case the transition function takes the form

$$F(z_{t-d}) \equiv F(z_{t-d}; \gamma, c) = (1 + \exp\{-\gamma(z_{t-d} - c)\})^{-1} \quad \gamma > 0 \tag{1.5}$$

By substituting (1.5) in (2.7) one obtains the logistic STECM, where positive and negative deviations from the equilibrium relative to the threshold c will give rise to different effects, with z_t being attracted towards 0 with a speed indicated by γ. The higher γ, the faster the transition from the two regimes $(z_{t-d} < c)$ and $(z_{t-d} > c)$, while as γ approaches infinity, the $F(.)$ approaches an indicator function $I[z_{t-d} > c]$.

[25] A popular example is aggregate demand, reacting much more quickly to a negative change in money supply than to a positive one.

Clearly, when γ approaches zero the transition becomes linear as in the standard case.

In some other cases, *size asymmetry* may be more appropriate to describe the dynamics of the variable of interest. For example, large or small misalignments of real effective exchange rates from their 'behavioral equilibrium' values have been shown to have different effects on the adjustment process of the exchange rates itself (Béreau, Villavicencio,and Mignon, 2009). This type of asymmetry can be conveniently modeled through the exponential function

$$F(z_{t-d}) \equiv F(z_{t-d}; \gamma, c) = \left(1 - \exp\left\{-\gamma(z_{t-d} - c)^2\right\}\right) \quad \gamma > 0 \tag{1.6}$$

Here, large (both positive and negative) deviations from the equilibrium gradually change the strength of the adjustment, implying that when $z_{t-d} = c$ the $F(.)$ is zero, while when z_{t-d} either decreases or increases to (minus) infinity, then $F(.)$ approaches one. The problem with the exponential function is that it shrinks to a linear function when γ either approaches zero or infinity. If this is not consistent with the dynamic behavior of the variable of interest, one might use instead the quadratic logistic function

$$F(z_{t-d}) \equiv F(z_{t-d}; \gamma, c) = \left(1 + \exp\left\{-\gamma(z_{t-d} - c_1)(z_{t-d} - c_2)\right\}\right)^{-1} \tag{1.7}$$
$$\gamma > 0 \text{ and } \; c_1 \leq c_2$$

For finite γ, this particular function has a minimum value which is not equal to zero, while for γ going to infinity $F(.)$ is equal to one, both for $z_{t-d} < c_1$ and for $z_{t-d} > c_2$, but it is equal to zero in between. As in the previous case, the transition becomes linear when the speed parameter γ approaches zero.

From the practical point of view, in order to select the appropriate transition variable and transition function for each of our countries, we started from the indications of the nonlinearity test in Table 7, but we also used a 'data specific approach' consisting in fitting various specifications and choosing the best one according to model evaluation criteria. Indeed, this is also Teräsvirta's (1994) suggestion when dealing with nonlinear models, since the available tests might have low power in the presence of possible mispecification errors. For what concerns the choice of the transition function, we also considered the insights from Fehr and Tyran's (2001) experimental evidence, indicating that both the size and signs of the shocks should matter in influencing the degree of *strategic stickiness* of expectations' adjustment. As a consequence, we restricted the possible transition functions to (1.5) and (1.7) , and we chose among the two based on model evaluation criteria.

Our result is that the quadratic logistic function seems to better fit the data in three cases out of four, suggesting that it should be more the size of the past forecast error than the sign determining *strategic stickiness* in the adjustment of consumers' expectations. Once chosen both the transition variable and the transition function, the STECM models were estimated by means of Nonlinear Least Squares as shown in Table 1.8[26].

At a first glance the STECM models seem to perform very well, and certainly better than their linear rivals at least in terms of parameters significance. The estimation of these models clearly involves losing twice as much degrees of freedom compared to the ECMs, but at the same time it results in generally higher R^2 (ranging between 0.27 and 0.56) and not lower Durbin-Watson statistics, a comforting sign. A sign which is a little bit less comforting is that STECM estimation solves the problem of the residuals' heteroskedasticity only in two cases out of four (in Italian and US data). Probably, this is due to the large number of outliers that are still present in the sample and that at this stage we did not attempt to correct. The transition function that the data generally seem to prefer is the quadratic logistic one, with the only exception of the UK which seems to favor the simple logistic. That is an indication that size more than sign asymmetry might be very important in determining the stickiness of expectations, and it is indeed consistent with one particular feature of money illusion: once the size of a nominal shock exceeds a certain (subjective) loss threshold, individuals start to take into considerations the (high) costs of reasoning in nominal terms rather than in real ones[27]. Also notice that the smoothness parameter γ is generally estimated quite

[26]Notice that in our estimation of STECMs we standardize the exponent of the $F(.)$ function by dividing it to the variance of the chosen transition variable. This is an advised choice to render the parameters γ, c_1 and c_2 scale free and it does not influence the other parameters' estimates. See Teräsvirta (1994) for more details.

[27]Akerlof, Dickens and Perry (1985, 2000) name this kind of behavior 'near rational', in the sense that it implies only second order losses. Indeed, money illusion can only be operational in contexts of slow and small nominal price increases: in situations of hyperinflation (e.g. the Nazi Germany during the 30s) people are perfectly aware of their loss or purchasing power, hence money illusion is totally absent.

		France		Italy		UK		US
dependent variable		Ye		Y		Y		Y
		coefficients						
autoregressive parameters	const	-0.01	const	0.02	const	0.01	const	-0.01
		0.74		0.38		0.66		0.38
	Z_{-1}	0.08	Z_{-1}	-0.03	Z_{-1}	-0.05	Z_{-1}	-0.06
		0.00		0.19		0.00		0.00
	Y_{-2}	-0.19	Y_{-1}	0.15	Y_{-1}	0.12	Y_{-1}	0.33
		0.00		0.07		0.04		0.00
	Ye_{-1}	-0.02	Y_{-2}	0.29	Y_{-2}	0.10	Y_{-2}	-0.29
		0.79		0.00		0.11		0.00
	Y	-0.02	Y_{-3}	0.33	Y_{-3}	0.23	Ye_{-1}	-0.06
		0.75		0.00		0.00		0.42
			Ye_{-2}	-0.07	Ye_{-1}	-0.09	Ye	0.37
				0.06		0.22		0.00
transition parameters	const	10.95	const	0.05	const	-0.35	const	0.03
		0.00		0.35		0.51		0.55
	Z_{-1}	7.55	Z_{-1}	-0.03	Z_{-1}	-0.04	Z_{-1}	0.02
		0.00		0.34		0.71		0.64
	Y_{-2}	-1.69	Y_{-1}	0.28	Y_{-1}	0.87	Y_{-1}	0.30
		0.00		0.08		0.00		0.05
	Ye_{-1}	2.53	Y_{-2}	-0.25	Y_{-2}	0.62	Y_{-2}	0.44
		0.00		0.10		0.06		0.02
	Y	4.75	Y_{-3}	-0.29	Y_{-3}	1.00	Ye_{-1}	0.17
		0.00		0.15		0.01		0.64
			Ye_{-2}	0.02	Ye_{-1}	-0.51	Ye	0.40
				0.67		0.11		0.26
Transition function		quadratic logistic		quadratic logistic		logistic		quadratic logistic
Transition variable		Z-4		Z-5		Z-3		Z-6
	γ	2.71	γ	11.74	γ	25.07	γ	369.42
		0.08		0.01		0.01		0.93
	c_1	-2.31	c_1	0.42	c_1	-2.97	c_1	-1.60
		0.00		0.00		0.00		0.00
	c_2	2.34	c_2	-1.85			c_2	0.78
		0.00		0.00				0.00
	R^2	0.56	R^2	0.27	R^2	0.38	R^2	0.39
	DW	1.97	DW	2.03	DW	1.98	DW	1.94
	ARCH(1)	33.16	ARCH(1)	1.69	ARCH(1)	80.03	ARCH(1)	3.52
		0.00		0.19		0.00		0.06

Note: Equations are estimated by Nonlinear Least Squares using covariance matrix suggested by Newey and West (1981). Small numbers below the coefficients are p-values.

Table 1.8: Estimation of the STECM models

imprecisely, while the other threshold parameters have always high significance. This feature of our estimation results should not misinterpreted though. In nonlinear models the standard deviation of the smoothness parameter tends to grow with the size of the parameter itself, and a precise estimate is always difficult to obtain[28]. To have a clearer idea of how the adjustment of expectations is behaved and to understand how *strategic stickiness* affects it, let us examine more closely figure 1.3.

The figure is divided in four panels, each of which graphically illustrates the performance of the STECM for one country. For each of the countries, on the left side we find two panels regarding model performance, both in terms of actual versus fitted values and of residuals' behavior. On the right side instead, we can see how the estimated transition function evolves in time and how it is influenced by the transition variable, with the grey-shaded area showing the location of the estimated thresholds.

For all the four countries it seems that there is still a lot to be done from the model specification point of view. Although the actual and the fitted series correlate very much, the models still fail to capture some of the largest movements in the forecast error, especially at the end of the sample when the recent financial crisis hit. The properties of the estimated transition functions in the upper and lower right panels also deserve some attention.

The quadratic logistic function for both Italy and the US show a very similar pattern, oscillating between zero and one as the observed forecast error either exceeds or stays in the threshold range (c_1, c_2). However, for the Italian case, these oscillations are more frequent in the first part of the sample, while the opposite is true for the US. Parallely, the bottom right panels of both the Italian and the US estimations show that the transmission function becomes linear and equal to zero for values of the past forecast error roughly between respectively $(0, 2)$ and $(-1, 1.5)$. This supports the hypothesis that some kind of behavioral bias resulting in *strategic stickiness* is operational since when forecast errors are within the threshold range people forget to adjust expecta-

[28]As noted by Teräsvirta (1994), when γ is large and at the same time the c parameters are sufficiently close to zero, a negative definite Hessian matrix is difficult to obtain for mere numerical reasons, even when convergence is achieved. That is the reason why joint estimation of the threshold parameters and the other model parameters is generally not advised.

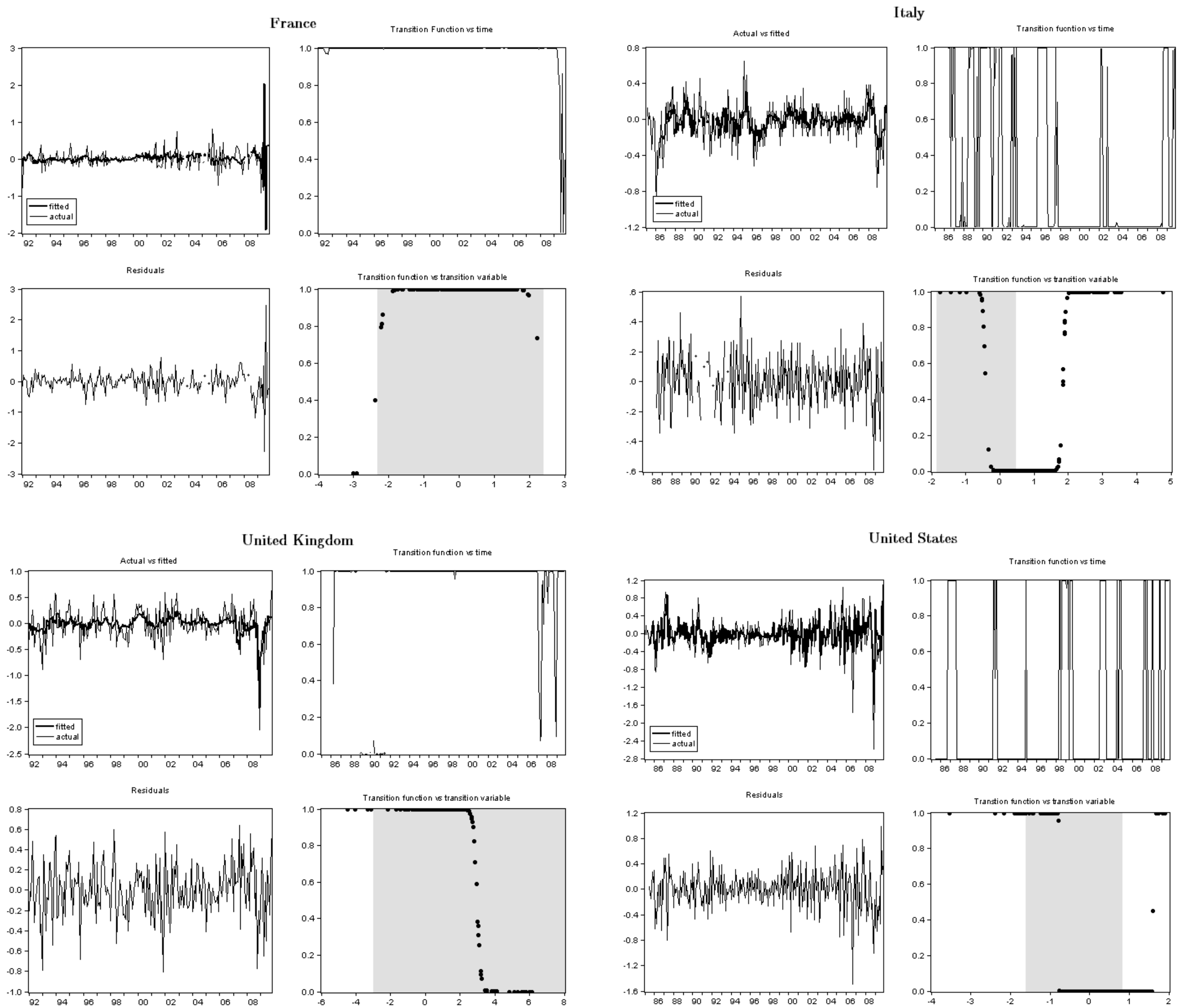


Figure 1.3: Estimated STECM models

tions, while when past errors are either quite small or quite large the adjustment starts to appear. Indeed, within the 'attention thresholds' less than full adjustment might be rational since either the costs are limited or the gains obtained are not very salient.

The case of France is a little bit different. The overall performance of the estimated STECM model seems to beat the one of all the other countries in terms of fit ($R^2 = 0.56$). To confirm that, figure (1.3) shows that the model captures the dynamics of the data for almost all the estimation sample except the financial crisis of 2008, hence the residuals look a little bit better behaved than the other sets of residuals. However, from the bottom left panel we can see that the transition function remains for the majority of the time at its maximum value, with just few observations remaining outside the threshold range. This is consistent with a quite precise estimate of the smoothness parameter $\gamma = -2.71$ (p-value=0.08), the lowest γ we obtain. We interpret these results in terms of a high 'degree of rationality' on behalf of French consumers: for past forecast errors falling outside the quite wide range $(-2, 2)$ we estimated, the adjustment towards the long-run equilibrium is only partial. We interpret this as a scarce evidence *strategic stickiness* on behalf of French consumers.

The case of UK is also peculiar, since it is the logistic function that the data seem to choose. According to Fehr and Tyran's (2001) results, expectations errors should be very sensitive to sign asymmetry, hence the logistic function is the one we were expecting the data would choose more often. However, in our sample this was only the case for English data. [29] From the right side panels we can clearly see that the logistic STECM model fails to capture many movements of the forecast error both in the first half of the sample and after 2008. From the left side panels instead, we notice that the

[29]Nevertheless, we should notice there is one big difference between Fehr and Tyran's experimental setting and our context. Fehr and Tyran were able to implement a fully anticipated monetary shock on experimental subjects and study its effects, while here we can only study the effects of past forecast errors on the aggregated adjustment mechanism of expectations. Indeed, it is possible to think of past forecast errors as incorporating exogenous monetary shocks, and clearly we based our notion of strategic stickiness on such a proxying. However, due to these considerations it is not possible to interpret the sign/size asymmetry favored by our models exactly in the same way as Fehr and Tyran

transition function remains linear and at its maximum as long as the forecast error is below the value of 3: before the threshold is reached, cointegration is almost linear, and English consumer fully adjust their errors to the long-run equilibrium. However, when expectational errors are greater than the threshold *strategic stickiness* kicks in and less than full adjustment occurs. Again, this could be seen as a sign that English consumers behave quite rationally in adjusting their expectations to past forecast errors.

Our general conclusion is that in our data there is some mild evidence of money illusion, or of any other decision heuristic resulting in strategically sticky inflation expectations. Our results however seem to indicate that such behavioral phenomenon is much more pronounced in Italy and in the US, while indeed English and French consumers look more 'rational'. Furthermore, our data seem to suggest that it is the size more than the sign of past forecast errors that matters more in explaining *strategic stickiness*. Given the early stage of our analysis, we want to make it clear though that the good fit of the STECM specification for our data does not exclude the relevance of other types of theoretical models we did not consider to explain a nonlinear adjustment of expectations.

1.5 Concluding Remarks

A model of 'ecological rationality' posits that when agents are confined with complex tasks such as forecasting inflation, they should use the best heuristics methods they have, given the available information sets. Indeed reasoning in nominal terms and ignoring low future inflation might be a powerful rule of thumb in a low and stable inflation environment. This paper has shown that traces of such heuristic behavior can also be found in the aggregate expected inflation time series.

By using standard rationality tests, and novel econometric models like STECMs, we obtain three main results. First, on average European consumers seem to have an upward bias when trying to assess the level of future inflation, being also very much influenced by the speed of diffusion of the available information (stickiness *à la Carrol, 2003*). Secondly, when looking at consumers behavior from a long-run perspective it is possible to notice instead that in equilibrium there is a tendency to underpredict future inflation, especially in periods when inflation is low and stable. Finally, and again from a long-run perspective, we find evidence also for *strategic stickiness*, implied by the fact that small past forecast errors have a much lower influence on the speed of adjustment of expectations than large ones. Size asymmetry seems to play a greater role than sign asymmetry in determining such stickiness. We interpret this findings as a sign that decision heuristics like money illusion are somewhat operating.

Of course one can always question the informative content of expectations series derived from qualitative survey data. Moreover, the use of nonlinear time series techniques implies particular caution because they are sensitive to the choice of the starting parameters and of the optimization algorithm used. In particular, STECMs are admittedly vulnerable to mispecification errors either in small samples, or in samples with multiple outliers. VDF (2000) also show that the availability of high frequency data (i.e. weekly or daily time series) increases the power of the nonlinearity tests and it could be helpful to distinguish 'disguised' nonlinearity from true nonlinearity.

Clearly, all the above considerations can provide fruitful insights for future research in this topic. A particularly promising new line for future research regards the application of panel smooth transition autoregression techniques, like the ones used in Béreau, Lopez, Villavicencio and Mignon (2010) to inflation expectations data. We are aware that our analysis of each expectation series separated one from the other implies a certain loss of potential variability/heterogeneity in the data, hence we hope to amend for this weakness in future work. Finally, it would be very interesting to conduct our investigation of *strategic stickiness* also with disaggregated expectations data, along the lines of what Pfajfar and Santoro (2010) do for the Michigan Consumer Survey data. At any rate, and bearing in mind all these potential improvements of our work, we want to stress our main finding: consumers' inflation expectations do exhibit a nonlinear and asymmetric adjustment to their long-run equilibrium, and this *strategic stickiness* can be traced back to behavioral biases like money illusion.

1.6 Appendix

1.6.1 The European Commission Consumers Survey and the Carlson-Parking's (1975) Method

In the European Commission consumers survey, consumers are asked the following question on future price developments (Question 6): "By comparison with the past 12 months, how do you expect consumer prices will develop in the next 12 months? They will ...

1. increase more rapidly
2. increase at the same rate
3. increase at a slower rate
4. stay about the same
5. fall
6. don't know

The 'Probability Approach' (Carlson-Parking,1975) is based on the idea to interpret the share of responses to each category as estimates of areas under the density function of aggregate inflation expectations, that is to say as probabilities. By specifying a distribution function for these probabilities (generally the logistic or the normal distributions are employed) it is then possible to compute a measure of the mean expected inflation and its standard deviation, together with the two response thresholds δ_t and ε_t. In particular Denoting S_i (for $i = 1, 2, 3, 4, 5$) as the sample proportions opting for each of the five response categories in the survey undertaken in month t, equations (2.1) to (1.11) below give the relevant measures for the derived expectations series.

$$\pi_t^e = -\pi_{t-12}^p \left(\frac{Z_{t-12}^3 + Z_{t-12}^4}{Z_{t-12}^1 + Z_{t-12}^2 - Z_{t-12}^3 - Z_{t-12}^4} \right) \quad (1.8)$$

$$\sigma_t^e = -\pi_{t-12}^p \left(\frac{2}{Z_{t-12}^1 + Z_{t-12}^2 - Z_{t-12}^3 - Z_{t-12}^4} \right) \quad (1.9)$$

$$\delta_t = -\pi_{t-12}^p \left(\frac{Z_{t-12}^1 + Z_{t-12}^2}{Z_{t-12}^1 + Z_{t-12}^2 - Z_{t-12}^3 - Z_{t-12}^4} \right) \quad (1.10)$$

$$\varepsilon_t = -\pi^p_{t-12}\left(\frac{Z^3_{t-12} - Z^4_{t-12}}{Z^1_{t-12} + Z^2_{t-12} + Z^3_{t-12} + Z^4_{t-12}}\right) \tag{1.11}$$

Where π^e_t indicates expected inflation and σ^e_t denotes the standard deviation of the aggregate distribution for inflation expectations, and π^p_{t-12} is the perceived rate of inflation at time $t-12$ used as a scaling factor following Berk (1999). Finally, $N^{-1}[.]$ is the inverse of the assumed probability distribution function which has the following arguments: $Z^1_{t-12} = N^{-1}[1 - S^1_{t-12}]$, $Z^2_{t-12} = N^{-1}[1 - S^1_{t-12} - S^2_{t-12}]$, $Z^3_{t-12} = N^{-1}[1 - S^1_{t-12} - S^2_{t-12} - S^3_{t-12}]$, $Z^4_{t-12} = N^{-1}[S^5_{t-12}]$.

The above expressions for the mean and standard error of expected future inflation express the mean and the uncertainty of expected inflation as a function of the actual and the perceived rate of inflation, which is used as a scaling function. It has been shown by Berk (1999) that using a notion of perceived inflation as a scaling function for the above system significantly improves the accuracy of the derived expectations series. The perceived rate of inflation can be computed by slightly modifying the Carlson Parkin method and applying it to Question 5 of the EC Consumer Survey, related past price developments. The following figure plots the estimated uncertainty (i.e. the standard deviation) for the expectations series we derived using the Carlson Parkin method, together with the inflation rates, both standardized for comparability. As expected, there is a high correlation between the estimated expectations' uncertainty and inflation levels in general.

For a more detailed description of this approach and of the rescaling based on perceived inflation, we suggest to refer to Berk (1999) and Gerberding (2007). For a critical survey of alternative methods to transform qualitative data into quantitative ones see Nardo (2003).

1.6.2 Replicating VDF (2000)

The case for nonlinear adjustment towards the long-run equilibrium is particularly evident for equivalent assets in the context of efficient financial markets (Yadav et al 1994, and Anderson 1995). Indeed, prices of equivalent assets should be such that investors are almost indifferent between holding for example stocks or futures with similar characteristics, or bonds of different maturity from the same issuing company.

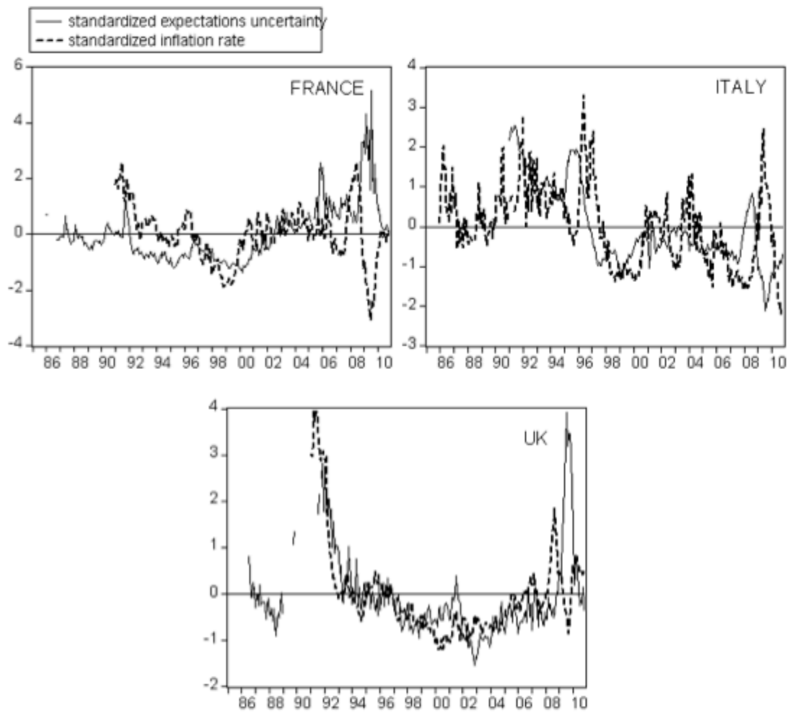


Figure 1.4: Estimated uncertainty for inflation expectations calculated using the Carlson Parkin method

Nevertheless, market frictions (transaction costs, short selling restrictions etc.) which are different for each trader, give rise to asymmetric adjustment of prices deviations from the no arbitrage equilibrium. Coherently with such a framework, VDF (2000) illustrate the properties of STECM models by taking as an example a monthly bivariate interest rate series for the Netherlands composed by one- and twelve-month interbank interest rates (indicated respectively as $R1_t$ and $R12_t$) from January 1981 to December 1985. In this section we are able to replicate their results by using the same STECM estimation strategy we applied in our paper[30].

From simple graphical inspection of Figure 9, one can already see that the two time series display a tendency to move together. Indeed, unit roots tests on the individual series confirm that the two interest rates are individually $I(1)$ (ADF t-statistics are respectively -1.26 and -1.25), while the spread between the two, $S_t = R_{12,t} - R_{1,t}$, can be considered stationary (ADF t-statistic -2.90).

The first step of VDF's procedure to specify a STECM model for S_t consists in

[30]The programs we used in this paragraph and all the other ones we used for model estimation are available upon request.

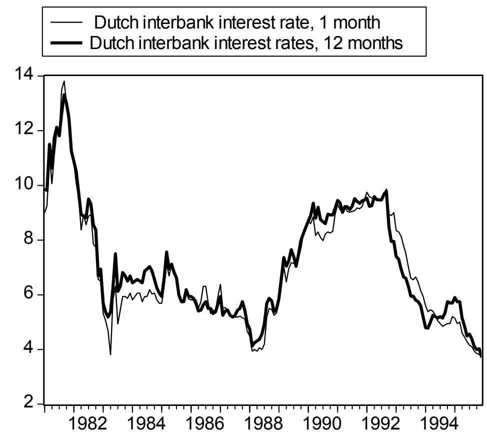


Figure 1.5: Dutch Interbank Interest Rates

fitting a standard linear ECM model. Having had evidence for cointegration through the Johansen trace test, a $VECM(1)$ for the first differences of the two interest rates is fitted with the following result[31]:

$$D(R12)_t = \underset{(0.068)}{0.11} * (R12_{t-1} - \underset{(0.018)}{1.029} * R1_{t-1}) - \underset{(0.148)}{0.206} * D(R12_{t-1}) +$$
$$+ \underset{(0.135)}{0.275} * D(R1_{t-1})$$
$$D(R1)_t = \underset{(0.073)}{0.23} * (R12_{t-1} - \underset{(0.018)}{1.029} * R1_{t-1}) - \underset{(0.159)}{0.096} * D(R12_{t-1}) +$$
$$+ \underset{(0.145)}{0.095} * D(R1_{t-1})$$

The authors then choose to restrict the cointegrating vector to $(1, -1)$, being the standard error of $R1_{t-1}$ equal to 0.018. Moreover, since the error correction coefficient not significant in the equation for the twelve-months interest rate, they estimate a conditional ECM model for $R1_t$ by conditioning it on $R12_t$. The resulting ECM specification is then taken as a basis for the subsequent analysis for nonlinearities in the adjustment process (Here for notational simplicity we renamed $D(R1)_t = X_t$ and $D(R12)_t = Y_t$[32]).

$$X_t = \underset{(0.02)}{-0.02} + \underset{(0.04)}{0.13} * S_{t-1} + \underset{(0.04)}{0.92} * Y_t - \underset{(0.07)}{0.16} * X_{t-1} + \underset{(0.08)}{0.09} * Y_{t-1} \tag{1.12}$$

Equation (1.12) has a problem of residuals non-normality (JB statistic=135.51 and Kurtosis=7.27), which might be caused both by aberrant observations at the begin-

[31] Standard errors are given in parentheses.

[32] As a check of correctness of our procedure, notice that the estimated parameters in equation (1.12) closely replicate the ones of equation (13) in VDF's (1997) paper.

ning of the sample and by neglected nonlinearity. As a consequence of that, the second step of VDF's procedure is testing for possible nonlinearity in the error correction mechanism by means of the Luukkonen et al.'(1988) test, generalized by Teräsvirta (1994), Swanson (1996) and VDF(2000) in the context of smooth transition autoregressive models.

As we already explained, the test is based on a reparametrization of the STECM model by approximating the $F(.)$ function with its third order Taylor approximation, and on computing a LM test for the null hypothesis $H_0' : \varphi_1 = \varphi_2 = \varphi_3 = 0$. Through a similar procedure it is also possible to obtain an estimate of the appropriate lag d^* for the transition variable S_{t-d},and of the most convenient form of the transition function.

Table 1.9 shows that for the choice of the transition variable, lag $d^* = 1$ yields the lowest p-value. Therefore we agree with the choice of VDF to select S_{t-1} as transition variable.

Test	Null	d=1	d=2	d=3	d=4	d=5	d=6
F Test	H_0'	0.01	0.56	0.56	0.12	0.05	0.02
χ^2 Test	H_0'	0.01	0.54	0.154	0.12	0.05	0.01

(*): Author's calculation on data from VDF (1997).

Table 1.9: LM-type test for smooth transition error correction in the CECM from VDF (1997)(*)

The choice of the transition function should be conducted by means of a similar procedure where a sequence of alternative tests is carried out sequentially. However, since there is no guarantee that this sequence will give rise to the right answer, we will use a more practical strategy, as VDF: we estimate the STECM with the three types of transition function and chose the one that best fits the data. After alternative specifications are estimated, and based on different information criteria, we reach the same conclusion of VDF choosing the quadratic logistic transition function.

As suggested by Teräsvirta (1994), we estimate the final STECM model by nonlinear least squares in two steps. First, we use the estimated parameters in equation (1.12) as starting values to estimate the π_2' vector of parameters, imposing sensible starting values for the γ, c_1 and c_2 parameters[33]. Then, we estimate the final STECM model by

[33]Teräsvirta (1994) suggests to initialize the $F(.)$ parameters with $\gamma = 1$ and $c_1 = c_2 = mean(transition\ variable)$.

giving as initial values $\widehat{\pi}_1'$ and $\widehat{\pi}_2'$ obtained in the previous step. Our procedure yields parameters estimates and standard errors which are really close to the ones of VDF:

$$\begin{aligned} X_t = & -\underset{(0.03)}{0.03} + \underset{(0.07)}{0.12} * S_{t-1} + \underset{(0.04)}{0.90} * Y_t - \underset{(0.09)}{0.24} * X_{t-1} + \underset{(0.1)}{0.23} * Y_{t-1} + \\ & + (\underset{(0.12)}{0.32} + \underset{(0.19)}{0.42} * S_{t-1} + \underset{(0.19)}{0.15} * Y_t - \underset{(0.17)}{0.17} * X_{t-1} - \underset{(0.25)}{0.31} * Y_{t-1}) * \\ & * [1 + \exp(\underset{(5.69)}{3.74} / \sigma^2_{S_{t-1}} * (S_{t-1} - \underset{(0.12)}{1.25}) * (S_{t-1} + \underset{(0.08)}{0.40}))^{-1}] \end{aligned} \tag{1.13}$$

The large standard error of the estimated γ is a very common problem in the estimation of smooth transition functions, which makes it generally very difficult the joint estimation of the $F(.)$ parameters. Furthermore, the estimates of γ, c_1 and c_2 and their standard errors are sensitive to rescaling, and to the type of algorithm used. Fortunately, these uncertainties do not affect the other parameters estimates.

Chapter 2

Some Pitfalls in Smooth Transition Models Estimation: A Monte Carlo Study

Abstract Nonlinear Regime Switching models are becoming increasingly popular in recent applied literature, as they allow capturing state-dependent behaviors which would be otherwise impossible to model. However, despite their popularity, the specification and estimation of these type of models is computationally complex and it is far from being a univocally solved issue.
This paper aims at contributing to this debate.[1] In particular, we use Monte Carlo experiments to assess whether employing the standard trick of 'Concentrating the Sum of Squares' (Leybourne, Newbold and Vaugas, 1998) in the application of Nonlinear Least Squares to Smooth Transition models yields estimates with desirable asymptotic properties. Our results confirm that this procedure needs to be used with caution as it may yield biased and inconsistent estimates, especially when faced with small samples.

JEL Codes: C22, C51.

Keywords: Smooth Transition Models, Nonlinear Least Squares, Monte Carlo Methods.

[1] I would like to thank Marco Tucci for his guidance and support throughout my research. A special thanks goes to Giampiero Gallo for suggesting me this title. The usual disclaimer applies.

2.1 Introduction

There seems to be a wide acceptance that relationships between macroeconomic variables may often be nonlinear. Nonetheless, the specification and estimation of nonlinear types of models is nothing but straightforward. Among nonlinear models, Regime Switching (RS) models are becoming increasingly popular in recent applied literature, as they allow capturing state-dependent behaviors which would be otherwise impossible to model. This family of models comprehends Markov Regime Switching models (Hamilton 1989), and Smooth Transition (ST) models, including Smooth Transition Autoregressive models (STAR) (Granger and Teräsvirta, 1993) , and Smooth Transition Error Correction models (STECM) (Van Dijk and Franses 2000 ; Kapetanios Shin and Snell, 2003).

In this work we focus on the estimation of Smooth Transition models, and we put under investigation one of the most routinely employed procedures in this context, namely the 'Concentrating the Sum of Squares' (henceforth CSQ) procedure, first proposed by Leybourne Newbold and Vougas, 1998. We take as a benchmark the first order Smooth Transition Error Correction Model (STECM), nesting in it more simple types of threshold autoregressive models, and we propose a Monte Carlo study on the finite sample properties of the nonlinear least squares estimator when the CSQ procedure is employed[2].

ST models estimation is computationally complex due to at least four order of difficulties. Firstly, ST models estimation, and STECMs estimation in particular, require great data availability and large samples. For this reason, the application of STECMs so far has been confined to problems where large data samples were available, involving for example interest rates (Van Dijk and Franses, 2000), real exchange rates (Bereau, Lopez Villavicencio and Mignon, 2010), stock returns (Fredj and Yousra, 2004), house prices (Balcilar, Gupta and Shah, 2010) and only recently to inflation expectations

[2]The STECM model is a flexible econometric specification extending the basic ECM in order to capture a nonlinear-asymmetric adjustment process towards the equilibrium. In a nutshell, while in the standard ECM *à la* Engle and Granger (1987) the adjustment mechanism is linear (a constant parameter $\alpha < 0$), in the STECM the short-run deviations are linked to the long-run equilibrium by means of a nonlinear function, depending on past values of a so called *transition variable*. Section 2 describes STECMs in more details.

(See previous chapter). Second, as Chan and Theodorakis (2010) point out, there is a lack of knowledge of the general statistical properties for the ST family of models, being the restricted two-regimes case almost the only one thoroughly investigated. Third, when practically estimating ST models, choosing the type of nonlinearity implies choosing also a specific modelling cycle. On this aspect, Granger Teräsvirta and Anderson (1993) distinguish two possible strategies: a *specific-to-general* strategy and a *general-to-specific* one. The former involves first estimating a simple restricted model, and then proceeding to more sophisticated ones only if diagnostics indicate that the simple model is inadequate. The latter implies that one starts to test linearity against the ST model, and if linearity is rejected one specifies sub hypotheses to see which ST model better suits the data. Finally, even when one of the two approaches is chosen, there is a fourth and very relevant order of difficulty, and it is the one on which this paper is centered: while the literature suggests that the best option for estimating ST models is nonlinear least squares (NLS), which in this case should be equivalent to Quasi Maximum likelihood (QMLE), the practical application of NLS is computationally cumbersome. For this reason, many authors suggests to reduce NLS' computational burden by 'Concentrating the sum of squares function'.

Using the CSQ 'trick' to simplify the implementation of NLS is admittedly something that requires caution, nevertheless it has often been advocated as the best solution available to simplify the estimation of ST time series models (Van Dijk Franses 2002; Lundbergh, Tersvirta, Van Dijk 2003; Shittu and Yaya, 2010). CSQ can be summarized as follows: the linear autoregressive part of the model is first estimated by assuming given values of the threshold parameters of the nonlinear part, then the NLS sum of squares function is 'concentrated' with respect to the estimated parameters from first step, and then it is evaluated only with respect to the nonlinear parameters. To the best of our knowledge, no other paper has investigated whether performing the CSQ trick costs too much in terms of loss of asymptotic properties of the estimated ST parameters. This paper aims at filling the gap by means of several Monte Carlo experiments.

Our Monte Carlo design involves two basic scenarios. We use as a reference point the paper by Van and Franses (henceforth VDF) (2000) whose data are available to us, and for which we are able to replicate exactly the parameters estimates. In the first scenario, we use the same parameter values as VDF (2000) [equation (21) page 219] to generate the STECM model, and we compare the performance of NLS-CSQ and Maximum Likelihood under different algorithm, for different values of both the standard deviations of the simulation errors, and the size of the simulated sample. In the second scenario we again propose the same comparison, but we further simplify our model to a STAR(1) model as in Chan McAleer (2002). Both scenarios are used to investigate the asymptotic properties of the NLS estimator under the standard procedure of CSQ, and each experiment involves 10.000 replications and generating series of either $T = 180$ or $T = 90$, there we indicate with T the number of observations. The preliminary results confirm our suspects: using NLS coupled with the CSQ procedure to estimate ST models may yield biased and inconsistent estimates, especially when dealing with small samples. We suspect that the main reason for this lack of robustness has to do with the nonlinear structure of ST models, implying that the objective function to maximize is not well-behaved. These concerns are addressed in the final part of this paper, where we try to obtain a better picture of the loglikelihood function for the simplest case of our experiments, the STAR(1) model, by using a heuristic procedure based on subsequent perturbations of the set of starting values. Finally, these concerns are thoroughly addressed in the final part of the paper, where we employ the global maximization algorithm by Tucci (2002) to obtain a better picture of the loglikelihood function for the simplest case of our experiments, the STAR(1) model.

The rest of this paper is organized as follows. Section 2 gives a general description of the theoretical characteristics of the Smooth Transition family of models. Section 3 starts from a more practical perspective, and it deals with the CSQ procedure applied to the NLS estimator for ST type of models. Section 4 delineates our Monte Carlo design involving two main scenarios, and sections 5 and 6 present the results of our empirical investigation by using mainly graphical and tabular evidence. Finally, in section 7 we show that the loglikelihood function of ST-type of models is not well-behaved and presents several local optima by means of both some heuristic analysis and of the use of Tucci's (2002) global optimization algorithm. Section 8 concludes.

2.2 Smooth Transition Models

Regime-switching behavior can be distinguished according the type of variable affecting the probability to switch. Let us follow the symbology of VDF (2000) by considering the simple two-regimes general model:

$$y_t = (\phi_{0,1} + \phi_{1,1} s_{t-d})(1 - I[s_{t-d} > c]) + (\phi_{0,2} + \phi_{1,2} s_{t-d})(I[s_{t-d} > c]) + \varepsilon_t \qquad (2.1)$$

Where the $\phi_{i,j}$ are constant parameters $i, j = 1, 2$. In this case there are two regimes, $(s_{t-d} \leq c)$ and $(s_{t-d} > c)$ are both determined by the value of the transition variable s_t lagged d times, relative to the threshold value c. In (2.1) the transition between the two regimes is steep, being modeled through the indicator function $I[A]$, which takes value 1 when event A occurs, and 0 otherwise.

Model (2.1) nests the whole family of regime switching models. More precisely, when s_t is an observable transition variable, (2.1) is called a Threshold Autoregressive (TAR) model. When instead the transition is induced by an unobservable Markov process, (2.1) becomes a Markov Switching (MS) model. TAR models can be further complicated to Self-exciting (SE-TAR) models, when $s_{t-d} = y_{t-d}$, and to Smooth Transition autoregressive (S-TAR or STAR) models, when the transition between different regimes is smooth, i.e. $I[.]$ is replaced by a continuous and at least twice-differentiable function $G(y_{t-d}; \gamma, c)$. The possible $G(.)$ functions can be summarized by the n^{th}-order logistic function:

$$\begin{aligned} G(s_{t-d}; \gamma, c_i) &= \left(i + \exp\{-\gamma \prod_{i=1}^{n} (z_{t-d} - c_i) \right)^{-1} \\ c_1 &\leq c_2 \leq ... \leq c_n, \quad \gamma > 0 \end{aligned} \qquad (2.2)$$

which allows for multiple switches between the two extreme regimes each time the function trespasses the thresholds c_i with $i = 1, 2, ...n$. The 'smoothness' or the speed of the transition is determined by the size of the γ parameter. Two popular choices in the literature are associated with $n = 1$, giving rise to the first-order logistic function (or simply the *logistic function*), and $n = 2$ corresponding to the second-order logistic function (or *quadratic logistic function*).

When it is y_{t-1} that induces the (smooth) transition, model (2.1) becomes the popular first order two-regime STAR model (Tersvirta,1994) :

$$y_t = (\phi_{0,1} + \phi_{1,1}y_{t-1})(1 - G(y_{t-1}, \gamma, c)) + (\phi_{0,2} + \phi_{1,2}y_{t-1})G(y_{t-1}, \gamma, c) + \varepsilon_t \qquad (2.3)$$

Extensions to the multi-regime STAR(p) models are obvious, but here we will be confined with the simple first other two-regime case.

Similarly, A Smooth Transition Error Correction (ST-ECM or STECM) model arises when we combine the structure of (4) and the standard ECM model *à la* Engle and Granger (1987). More precisely, assume that y_t and x_t are cointegrated with the following data generating processes:

$$\begin{aligned} y_t &= \beta x_t + z_t \\ Dx_t &= v_t \end{aligned} \qquad (2.4)$$

Where y_t is a non stationary univariate time series, and we consider the simplest case where it is integrated of order one, x_t is also a $I(1)$ variable, D is the first differences operator, and $(z_t, v_t)'$ are the stationary disturbances. In particular, let us suppose that z_t follows an $AR(1)$ process:

$$z_t = \rho z_{t-1} + \varepsilon_t \text{ and } \varepsilon_t \overset{iid}{\sim} (0, \sigma_\varepsilon^2) \qquad (2.5)$$

In this linear case, y_t and x_t are said to be cointegrated with cointegrating vector $(1, -\beta)'$ only if $\rho < 0$. In other words, y_t and x_t are cointegrated if there exists at least one of their linear combinations which is stationary, or $I(0)$. Equation (2.5) represents the standard case where the adjustment process is linear in ρ and it is always of the same strength, regardless the sign or the size of the deviation from the long-run equilibrium.

In the STECM case we speak of nonlinear cointegration because we allow for a nonlinear adjustment process for the error correction mechanism, so z_t follows

$$z_t = G(z_{t-1}; \gamma, c) + \varepsilon_t \text{ and } \varepsilon_t \overset{iid}{\sim} (0, \sigma_\varepsilon^2) \qquad (2.6)$$

Where $G(.)$ is the *transition function* as in (2.2)[3]. This was a simple example where the deviation from the long-run equilibrium z_t is a first order (either linear or nonlinear)

[3]The term nonlinear cointegration may be ambiguous, as in the literature it can indicate both the presence of a nonlinear cointegrating long-run relationship, and a nonlinear short-run adjustment process. To avoid misunderstandings, when we use the term 'nonlinear cointegration' in this article, we are always referring to a nonlinear short-run adjustment process.

stochastic process, but clearly a more general case involves a transition function $G(z_{t-d})$ with an higher lag order $d = \{1, 2, ...\}$. As we already mentioned, the transition function $G(z_{t-d})$ can be conveniently chosen to model different types of nonlinear error correction behaviors, allowing a maximum of $m = n - 1$ regimes. The resulting m-regime STECM(1) model has the following form:

$$\begin{aligned} Dy_t &= (\varphi_{1,1} + \varphi_{1,2} z_{t-1} + \varphi_{1,3} Dy_{t-1} + \varphi_{1,4} Dx_{t-1}) + \\ &\quad + G(z_{t-d}; \gamma, c')(\varphi_{2,1} + \varphi_{2,2} z_{t-1} + \varphi_{2,3} Dy_{t-1} + \varphi_{2,4} Dx_{t-1}) + \varepsilon_t \end{aligned} \tag{2.7}$$

where for sake of generality $c' = (c_1, c_2, ...c_n)$ indicates the vector of threshold parameters. As it is possible to see, a STECM(1) with m regimes is a general specification which nests the m-regime STAR(1) model (in differences) (4) when $\varphi_{i,4} = 0$ for $i = 1, 2$. As a results of its general structure, in the remainder of this paper the STECM(1) model (2.7), with either $m = 1$ or $m = 2$, will be used as the benchmark for our Monte Carlo experiment.

2.3 Concentrating the Sum of Squares Function

To estimate ST models like (4) and (2.7) the following assumptions are needed:

(i) The data generating process is strictly stationary and ergodic.

(ii) The necessary conditions for the existence of moments are satisfied.

(iii) The maximum likelihood estimators of the model parameters are consistent and asymptotically normal.

While on the one hand these are standard assumptions, it must be noted that their verification for nonlinear type of models is nothing but straightforward. In particular, the stationarity conditions for ST models are challenging to be verified as nonlinear time series contain *endogenous dynamics*, i.e. they may fluctuate even in the absence of stochastic shocks[4].

Assuming that assumptions (i)-(iii) hold, the modelling cycle for ST models also requires a specific choice. In this type of literature, it is common to use a general-to-specific modelling cycle. First, one tests whether ST -AR or ST-ECM is more

[4] A complete discussion of the stationary conditions for ST models is beyond the scope of this paper. For a more thorough treatment of this issue please refer to Van Dijk and Franses (2000)

appropriate than the simple linear AR or ECM models (clearly, preliminary cointegration analysis must be carried out when estimating a STECM). An example of linearity test, under assumptions (i)-(iii) for a simple two-regime logistic STAR is given by

$$H_0 : \varphi_{1,1} = \varphi_{2,1}, \quad \varphi_{1,2} = \varphi_{2,2}$$

Notice that the γ and c parameters within the transition function, are not involved in the null hypothesis, since they yield unidentified nuisance parameters. On this behalf, the null hypothesis of linearity can be equivalently reformulated as $H_0' : \gamma = 0$, implying that $G(s_t, 0, c) = 1/2$ and the STAR model becoming

$$y_t = \frac{1}{2}(\varphi_{1,1} + \varphi_{2,1}) + \frac{1}{2}(\varphi_{1,2} + \varphi_{2,2})y_{t-1} + \varepsilon_t$$

which is linear regardless of the validity of H_0. Thus it is important to include parameters in the transition function for purposes of testing. This problem can be avoided by expressing the transition function by its Taylor expansion around $\gamma = 0$, which is a simple but important technique for hypothesis testing in STAR-type models. Once the transition function and the threshold variable have been determined, the parameter of the ST models can be estimated by means of NLS.

Formally, if we define the parameters vector $\varphi = (\varphi_1', \varphi_2', \gamma, c')$ so that both (4) and (2.7) can be summarized as $y_t = F(x_t; \varphi) + \varepsilon_t$. The NLS estimator is given by

$$\widehat{\varphi} = \arg\min_{\varphi} \sum_{t=1}^{T} (y_t - F(x_t; \varphi))^2 = \arg\min_{\varphi} \sum_{t=1}^{T} \varepsilon_t^2. \tag{2.8}$$

where $F(x_t; \varphi)$ is what Chan and Tong (1985) call the *skeleton* of the ST model, that is to say

$$F(x_t; \varphi) \equiv \varphi_1' x_t (1 - G(y_{t-1}, \gamma, c')) + \varphi_2' x_t (G(y_{t-1}, \gamma, c'))$$

Under the additional assumption that the errors ε_t are normally distributed, NLS is equivalent to Maximum Likelihood, otherwise it can interpreted as a Quasi Maximum Likelihood estimator. Wooldridge (1994) and Potscher and Prucha (1997) demonstrated that NLS is asymptotic and normal under appropriate regularity conditions. In theory, problem (2.8) can be solved by means of any nonlinear optimization algorithm, provided that appropriate starting values are chosen[5]. Nevertheless, it is a high

[5] Van Dijk and Franses (2000, p. 90) state that beyond the choice of the starting values, two other issues

dimensionality problem which renders joint estimation of both the linear and nonlinear parameters quite problematic in practice.

For this practical reason, Leybourne, Newbold and Vaugas (1998), first proposed to reduce the computational burden on NLS in ST models estimation by *concentrating one set of parameters out of the objective function* (what we called the CSQ trick). After this seminal paper, this has become the most standard procedure in ST models estimation, to the point that it is even advocated by most econometrics softwares[6]. In a nutshell, CSQ implies that the linear autoregressive part of the model is first estimated by assuming given values of the threshold parameters in $G(y_{t-1}, \gamma, c')$, then the linear autoregressive parameters are concentrated out of the NLS sum of squares function which can be minimized only with respect to the nonlinear ones. Let us describe more in details how this multi-procedure works[7].

Step 1. Since ST models are linear in the autoregressive parameters for given values of γ and c', the deterministic part of the models can conveniently be estimated by OLS. Clearly, the two sets of parameters φ_1 and φ_2 can be estimated separately from each other, hence step1 can be further divided in *step1.a* and *step1.b*.

Indicating with $\mathbf{x}_t(\gamma, c) = \{1, y_{t-1}; G(\gamma, c)\}$ the explicative variables in the STAR(1) case, and with $\mathbf{x}_t(\gamma, c) = \{1, z_{t-1}, Dy_{t-1}, Dx_{t-1}; G(\gamma, c)\}$ those of the STECM(1) case, OLS estimates can be found as

$$\widehat{\varphi}(\gamma, c)' = \{\widehat{\varphi}_{1,1}, \widehat{\varphi}_{1,2}, \widehat{\varphi}_{2,1}, \widehat{\varphi}_{2,2}\}' = \left(\sum_{t=1}^{T} x_t x_t'\right) \sum_{t=1}^{T} x_t y_t' \tag{2.9}$$

Plugging the estimated parameters into the objective function of (2.8) one obtains

deserve particular attention in the optimization procedure. The first one is exactly the one on which is article is focused, namely the CSQ procedure. The other one concerns the significance of the estimated smoothness parameter γ. The issue is that in ST models the standard deviation of the smoothness parameter tends to grow with the size of the parameter itself, hence a precise estimate is always difficult to obtain. More precisely, Teräsvirta(1994) notes that when γ is large and at the same time the c parameters are sufficiently close to zero, a negative definite Hessian matrix is difficult to obtain for mere numerical reasons, even when convergence is achieved. That is the reason behind the apparent low significance of the estimate of γ, which should then be evaluated with diagnostic different from the standard ones.

[6] See for example the Rats letter (2005). Among papers employing the CSQ trick we mention Van Dijk Franses (2002); Lundbergh, Teräsvirta, Van Dijk (2003); Shittu and Yaya, (2010)

[7] For notational simplicity we will describe the procedure referring to model (4), extensions to more complicated models as (2.7) are straightforward.

the sum of squares function concentrated with respect to φ_1 and φ_2 i.e. $SS^C(\gamma, c) = \sum_{t=1}^{T}(y_t - \widehat{\varphi}(\gamma, c)'\mathbf{x}_t(\gamma, c))^2$.

Step 2. The concentrated sum of squares function is then minimized by NLS only with respect to γ and c'. Under standard differentiability assumptions, it can be shown that the minimizer of (2.9) $\widehat{\varphi}^C = \{\widehat{\gamma}(\widehat{\varphi}), \widehat{c}(\widehat{\varphi})\}$ is identical to the vector φ that solves the entire minimization problem (2.8) [8].

Practically speaking, ST models estimation involves first finding appropriate starting values for γ and c' and and keeping them constant for the first two steps. then In *step1.a*, the set of linear parameters $\widehat{\varphi_{1,}}$ is estimated by OLS. In *step1.b*, the $\widehat{\varphi_2}$ set is estimated by applying OLS to the entire STECM using as starting the $\widehat{\varphi_1}$ values obtained from the step. Finally in *step 2*, the γ and c' parameters are found by optimizing the sum of squares function concentrating out $\widehat{\varphi_1}$ and $\widehat{\varphi_2}$.

2.4 Monte Carlo Design

This section takes as a reference the two-regime STECM(1) model (2.7) and it investigates the small samples performance of the NLS estimator when the CSQ trick is employed. We design two main Monte Carlo experiments, for all which we endogenously generate the dependent variable according to the chosen data generating process (DGP). For the exogenous variables we use the dataset of VDF (2000) containing one and twelve months interbank monthly interest rates for the Netherlands (respectively indicated as y_t and x_t). Here follows a brief description of the different experiments.

In Experiment I, both the dependent variable y_t, and the long-run cointegrating relation for the STECM $z_t = (x_t - y_t)$ are endogenously generated according to the

[8] For a more detailed proof see Wooldridge (2002) chapter 12, page 376.

following conditional STECM stationary specification (see VDF, 2000, pp.219)[9]:

$$
\begin{aligned}
Dy_t &= (\varphi_{1,1} + \varphi_{1,2} z_{t-1} + \varphi_{1,3} Dy_{t-1} + \varphi_{1,4} Dx_{t-1} + \varphi_{1,5} Dx_t) + \\
&+ (\varphi_{2,1} + \varphi_{2,2} z_{t-1} + \varphi_{2,3} Dy_{t-1} + \varphi_{2,4} Dx_{t-1} + \varphi_{2,5} Dx_t) * \\
&* \{1 + \exp[-\gamma(z_{t-1} - c_1)(z_{t-1} - c_2)\}^{-1} + \varepsilon_t
\end{aligned} \tag{2.10}
$$

where $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$, and σ_ε is alternatively to set $0.225, 0.131$, and 0.022[10]. The starting values for Dy_t and z_t are set in accordance with the actual values of these variables at the beginning of the sample. The structural parameters are set as in VDF (2000), equation (21), and they can be found in tables 2.1 and 2.2.

$\varphi_{1,1}$	$\varphi_{1,2}$	$\varphi_{1,3}$	$\varphi_{1,4}$	$\varphi_{1,5}$	$\varphi_{2,1}$	$\varphi_{2,2}$	$\varphi_{2,3}$	$\varphi_{2,4}$	$\varphi_{2,5}$
−0.01	0.05	0.81	−0.04	0.07	0.12	0.1	0.41	0.21	−0.21

Table 2.1: Parameter values for the deterministic part

γ	c_1	c_2
7.38	0.42	1.03

Table 2.2: Parameter values for the transition function

In Experiment II, the DGP is simplified to a univariate model, replicating the STAR(1) specification used in Chan and McAleer (2003). More specifically, we leave the parameter values and the erratic structure of the model unchanged, and we generate series according to

$$
\begin{aligned}
Dy_t &= (\varphi_{1,1} + \varphi_{1,3} Dy_{t-1}) + \\
&... + (\varphi_{2,1} + \varphi_{2,3} Dy_{t-1}) * \{1 + \exp[\gamma(Dy_{t-1} - c)\}^{-1} + \varepsilon_t
\end{aligned} \tag{2.11}
$$

[9]To be more precise it is a second -order logistic conditional STECM specification. It is conditional, in the sense that it estimates the model for Dy_t conditioning on Dx_t (The paper by VDF (2000) provides some some reasons for this specification, here we take it for granted since we use the model as a benchmark). For notational simplicity we briefly refer to the model as a STECM(1).

[10] These values for the variance of the simulation error are chosen with the explicit objective to go as close as possible to the estimation results of VDF (2000). In particular, $\sigma_\varepsilon^2 = 0.131$ is exactly the residual sum of squares error in VDF's estimation when they employ weekly data (their lowest), while $\sigma_\varepsilon^2 = 0.225$ is the one resulting from their monthly data estimation. Clearly, $\sigma_\varepsilon^2 = 0.022$ is then chosen to be the more 'optimistic' value for such variance, being 10 times smaller than the one obtained from VDF's monthly estimation.

In this case, the complex STECM model (2.10) is reduced to a simple logistic smooth transition autoregressive model, with the dependent variable in first differences. The construction of this particular scenario is aimed at seeing whether a significant reduction in the number of model parameters (in this case from 13 to 6) and a simplification the type of transition function (from 2^{nd} to 1^{st} order logistic) affects the asymptotic performance of the NLS-CSQ estimator[11].

In each experiment, we apply the following 'treatments':

(A) we estimate the model by NLS estimation employing the CSQ trick (henceforth we will refer to this the NLS-CSQ estimator). As mentioned earlier, this involves a three-step estimation for models (2.10) and (2.11). For further inspection of the performance of NLS-CSQ estimator, we designed three sub-treatments:

A.1) The whole model, either (2.10) or (2.11), is estimated by NLS-CSQ.

A.2) The DGPs are reduced to their purely autoregressive parts, i.e. only the vector of parameters φ_1 is estimated. The Monte Carlo experiment is conducted with these reduced models estimated by simple OLS. We do so to have a reference point for the evaluation of the NLS-CSQ estimator with respect to the more complicated smooth transition models.

A.3) The DGP is reduced by setting $\varphi_{i,3} = \varphi_{i,4} = \varphi_{i,5} = 0$ for $i = 1, 2$ in the case of (2.10), and $\varphi_{1,1} = \varphi_{2,1} = 0$ in the case of (2.11). The Monte Carlo experiment is conducted with these reduced models estimated by NLS-CSQ. We design this sub-treatment in order to be able isolate the effect of each type of parameter of the NLS-CSQ estimator performance.

(B) The parameters of the two models (2.10) and (2.11) are jointly estimated by Maximum Likelihood, and more specifically in treatment *B.1)* we use the Levenberg-

[11] We also performed another experiment which is simply a slight modification of experiment I, where we endogenously generate y_t as in equation (2.10), but we take as an exogenous the cointegrating equation $z_t = (x_t - y_t)$, which is computed with the actual values of the variables. We implemented this scenario because we wanted to make sure that we go as close as possible to the original estimation problem faced by VDF (2000). Indeed, the results from these simulations were not qualitatively different from the ones of experiment I, hence we considered this scenario as a robustness check of the first one and we do not present its results.

Marquardt algorithm[12] and in treatment *B.2)* we use the Berndt, Hall, Hall, and Hausman (BHHH) algorithm[13].

Each experiment is based on 10000 replications of series with either $T = 180$ or $T = 90$ observations[14]. Table (2.3) presents a summary of the different scenarios and of the treatments implemented.

Number of observations	Errors' Variance	Experiment I and II				
T=180	σ_ε=0.225	A.1	A.2	A.3	B.1	B.2
	σ_ε=0.131	A.1	A.2	A.3	B.1	B.2
	σ_ε=0.022	A.1	A.2	A.3	B.1	B.2
T=90	σ_ε=0.225	A.1	A.2	A.3	B.1	B.2
	σ_ε=0.131	A.1	A.2	A.3	B.1	B.2
	σ_ε=0.022	A.1	A.2	A.3	B.1	B.2

Table 2.3: Summary of the Monte Carlo experiments performed

2.5 Experiment I: STECM Estimation

With this experiment we address the main question of the paper: is the NLS-CSQ estimator for STECM models able to yield parameters estimates with desirable asymptotic properties? Here we try to answer this question with respect to our reference model, the two-regime STECM(1) model, and more precisely we take the example of the paper by VDF (2000). VDF (2000) estimate a conditional STECM model using a system of two

[12]The Levenberg-Marquardt algorithm is a local maximization algorithm which interpolates between the Gauss-Newton algorithm and the method of gradient descent. It is thought to be more robust than the Gauss-Newton, since in many cases it finds a solution even if it starts very far off the final minimum. Even if, for well-behaved functions and reasonable starting parameters, it tends to be a bit slower than the Gauss-Newton.

[13]This algorithm follows Newton-Raphson, but replaces the negative of the Hessian by an approximation formed from the sum of the outer product of the gradient vectors for each observation's contribution to the objective function. For least squares and log likelihood functions, this approximation is asymptotically equivalent to the actual Hessian when evaluated at the parameter values which maximize the function. When evaluated away from the maximum, this approximation may be quite poor.

[14]We explicitly chose these time samples because we wanted to analyze the small sample properties of the NLS estimator with the CSQ trick.

monthly interest rate series for the Netherlands composed by one- and twelve-month interbank interest rates (which they indicate as $R1_t$ and $R12_t$) from January 1981 to December 1985. In our symbology these two variables are indicated respectively as y_t and x_t, and again the DGP for Dy_t is given by (2.10). Chan and Mc Aleer (2002) present a somewhat similar experiment with respect to a STAR-GARCH model, but they run the Monte Carlo simulation by fixing the nonlinear parameters in step 1 at their true value. Here instead the CSQ procedure is carried out independently of the true values of γ and c', as in practice the researcher is seldom aware of true transition and threshold values, and She simply initializes them with plausible starting values[15]. In this section, the logistic STECM(1) process defined in (2.10) was simulated with respectively $\sigma_\varepsilon = (0.225, 0.131, 0.022)$ under treatments **(A)**-**(B)** previously described.

The NLS-CSQ estimator, studied in treatment A, confirms our concerns about its asymptotic performance. As we can see from figure 2.1 (upper panels), the bias in percentage terms reaches very high peaks for the two simulation where the series were generated with the two highest variances, while stabilizing to more reasonable values only for the scenario where the series are generated with a variance ten times smaller than VDF's. Moreover, a closer look at the detailed descriptive statistics (reported in table 2.9.5 of the Appendix) shows that there is no hope that the standard deviation of the estimated parameters goes to zero as the number of simulations increases. Reducing the sample size by half clearly rises the maximum bias, but leaves the big picture unchanged. The parameters that seem to create more problems in this type of estimation are the two error correction parameters φ_{12} and φ_{22}, followed by the cumbersome smoothness and threshold parameters γ, c_1 and c_2[16]. These results are compatible with the ones by Chan and Mc Aleer (2002) on logistic STAR(1) models.

In order to further investigate which of the thirteen parameters of the STECM

[15] Teräsvirta(1994) suggests first to standardize the exponent of G by diving it by $\widehat{\sigma}(y)$, making it easier to select a reasonable starting value the so standardized γ. Following this suggestion, throughout all the experiments we decided to set the initial values of the nonlinear parameters as $\gamma = 1$ and $c_1 = c_2 = samplemean(transitionvariable)$ instead of running a grid two/tridimensional grid search. For our dataset, this choice has the advantage of reducing sensibly the computational burden of the Monte Carlo simulations while leaving the results substantially unchanged.

[16] See the appendix for the complete tables reporting descriptive statistics for each parameter of the simulations.

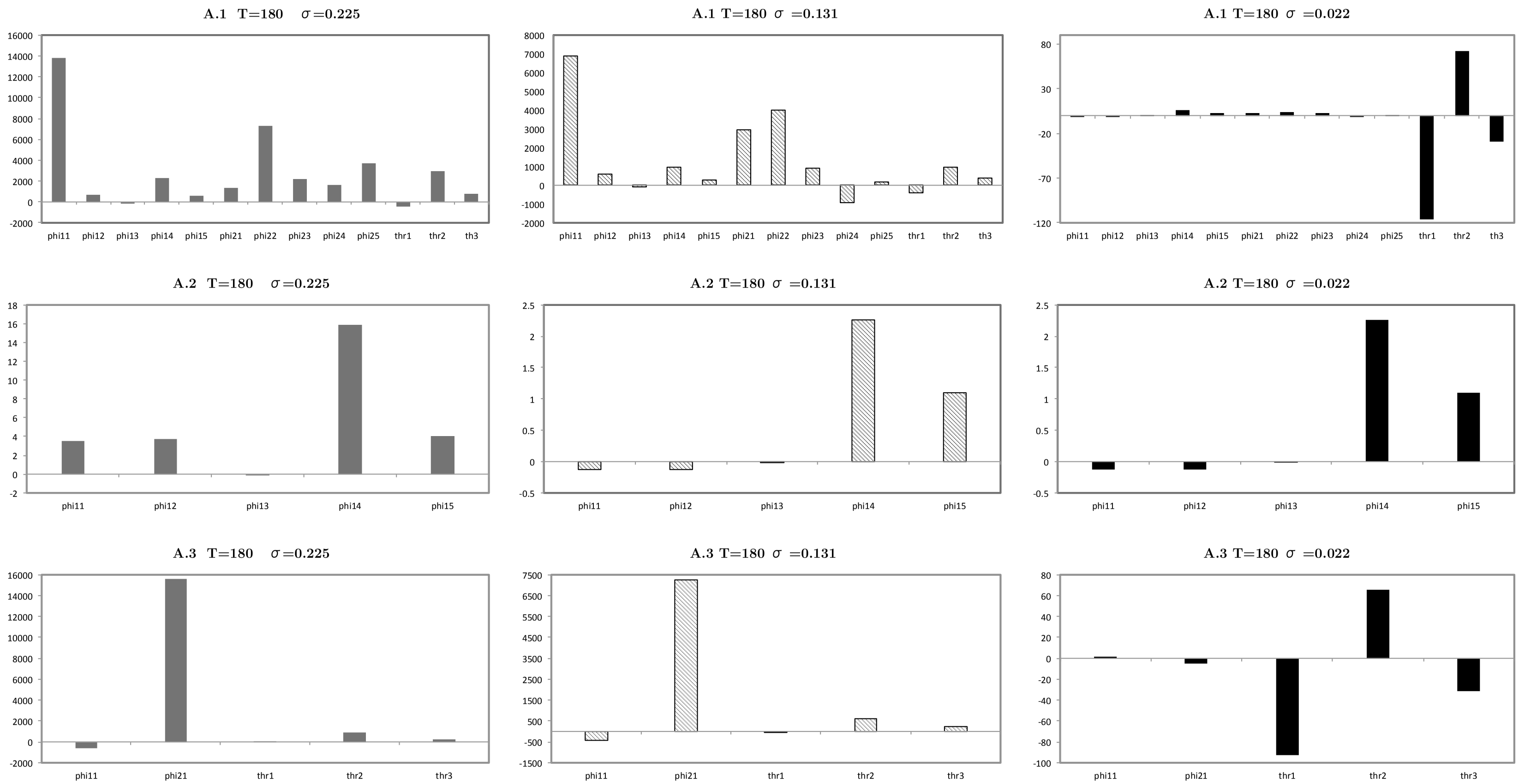


Figure 2.1: Experiment I, Treatments A.1, A.2 and A.3

Note: y-axis=percentage bias; x-axis =parameters of model (2.10), Number of simulations=10000.

create more problems to the NLS-CSQ procedure, treatments A.2 and A.3 were also implemented, as it possible to see from the medium and bottom panels of figure 2.1. Clearly, the minimum percentage bias is obtained when only the first set of autoregressive parameters is estimated via simple OLS in treatment A.2. However, even in this fine-working scenario, the most troublesome parameters remain the error correction ones, φ_{12} and φ_{22}, and the φ_{14} parameter multiplying Dx_t. As a consequence, experiment A.3 seeks to isolate precisely the effects of these two parameters, estimating a reduced form of equation (2.10) after having reduced the DGP. Indeed, From the bottom panels of figure (2.1) we are able to see that the responsibility of the high bias is mainly due to the impact of φ_{12}, φ_{22}, and φ_{14}, overcompensating the influence of the nonlinear parameters. It is worth noticing also that a puzzling result occurs when the sample size is reduced to $T = 90$ in this scenario: we obtain that the higher percentage bias pertains to the simulation with the lowest standard error (rightmost panels), while decreasing when the standard error is gradually raised. This counterintuitive result cannot be clearly explained at this stage and requires further investigation. The direct rival of the NLS-CSQ estimator for Smooth Transition models is the Maximum Likelihood estimator (MLE), which nevertheless is rarely used due to its high computational burden. Furthermore, as already mentioned, joint estimation of the autoregressive and nonlinear parameters in ST models is particularly difficult, and the choice of the starting values and of the optimization algorithm becomes crucial. Chan and Mc Aleer (2002) show that the very celebrated Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is not robust for estimating STAR models. Here, we decided to put under test two of the most widespread optimization algorithms which are usually implemented by commercial econometric softwares, namely the Marquardt-Levemberg (treatment B.1) and the BHHH (treatment B.2) algorithm.

Table 2.9.6 of the appendix clearly shows that in this competition the BHHH is a clear 'looser', resulting is a bias which is even higher than the one of the NLS-CSQ estimator. This is probably due to the fact that given the same set of starting values, chosen to be 'sufficiently close' to the true structural parameters, BHHH is moving in a direction which leads it to stall in a region where the likelihood is flat. Indeed, the algorithm follows Newton-Raphson, but replaces the negative of the Hessian by an approximation formed from the sum of the outer product of the gradient vectors for each observation's contribution to the objective function, and that is why when evaluated away from the maximum the approximation may be quite poor. On the other hand, the Marquardt algorithm modifies the Gauss-Newton algorithm by adding a correction matrix, called 'ridge factor', to the Hessian approximation, which handles numerical problems when the outer product is near singular and may improve the convergence rate. As above, the algorithm updates the parameter values in the direction of the gradient but with a 'diagonal adjustment', and this can explain its improved success rate in our case. Given the above considerations, we will not discuss experiment B.2 any further, and we will focus instead on the comparison between MLE by means of the Marquardt algorithm (B.1) and the three-step NLS-CSQ estimator (A.1).

Figure 2.2 displays the percentage bias obtained as a result ML estimation of model (2.10) with experiment B.1 (dark bars) in comparison with the percentage bias obtained in NLS-CSQ estimation with experiment A.1 (light bars). As we see, the dark bars are always closer to zero than the light bars, indicating that ML can partly correct the asymptotic bias resulting from a three-step procedure like CSQ, at least for a very complex model like a STECM. This result holds for either $T = 180$ or $T = 90$, and for the highest values of the simulation error, namely $\sigma_\varepsilon = 0.225$ and $\sigma_\varepsilon = 0.131$. Then, with the lowest value of the simulation error $\sigma_\varepsilon = 0.022$, we again obtain the counterintuitive result that the estimators' performance is reversed: the lowest (but still relevant) bias is obtained with the NLS-CSQ estimator and ML follows.

In conclusion, in all the treatments but the one with the $T = 90$ and $\sigma_\varepsilon = 0.022$, the bias obtained by ML estimation averages around 10 percent deviation from the actual 'true' value of the parameter and it is always lower than the one obtained from NLS-CSQ. The asymptotic performance of MLE seems to be better than the one of NLS-CSQ, but we suspect it would certainly benefit of some improvement in the choice of the optimization algorithm since the shape of the loglikelihood function is probably not regular. In section 8 we will specifically investigate this issue.

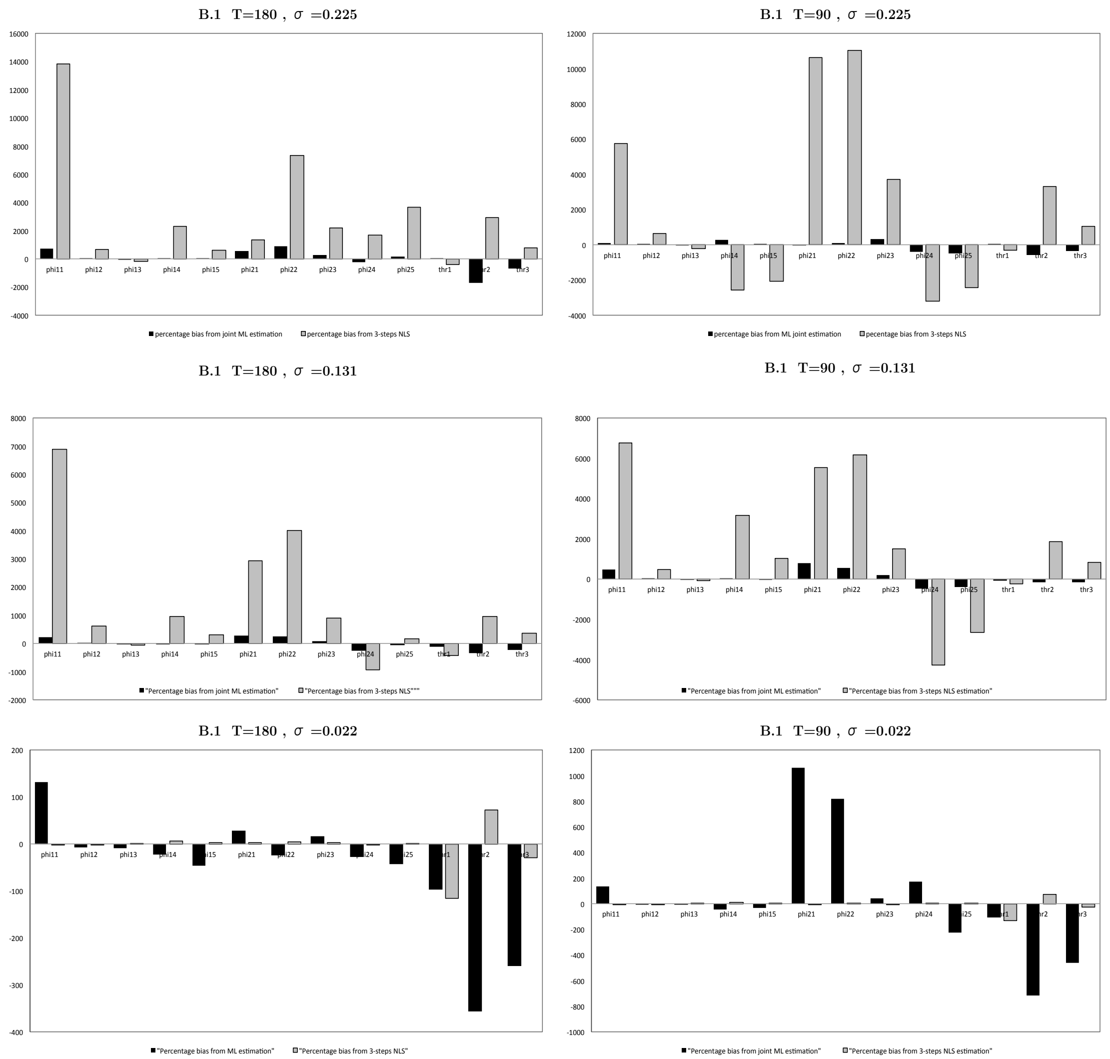


Figure 2.2: Experiment I, Treatment B

Note: y-axis=percentage bias; x-axis =parameters of model (2.10), Number of simulations=10000.

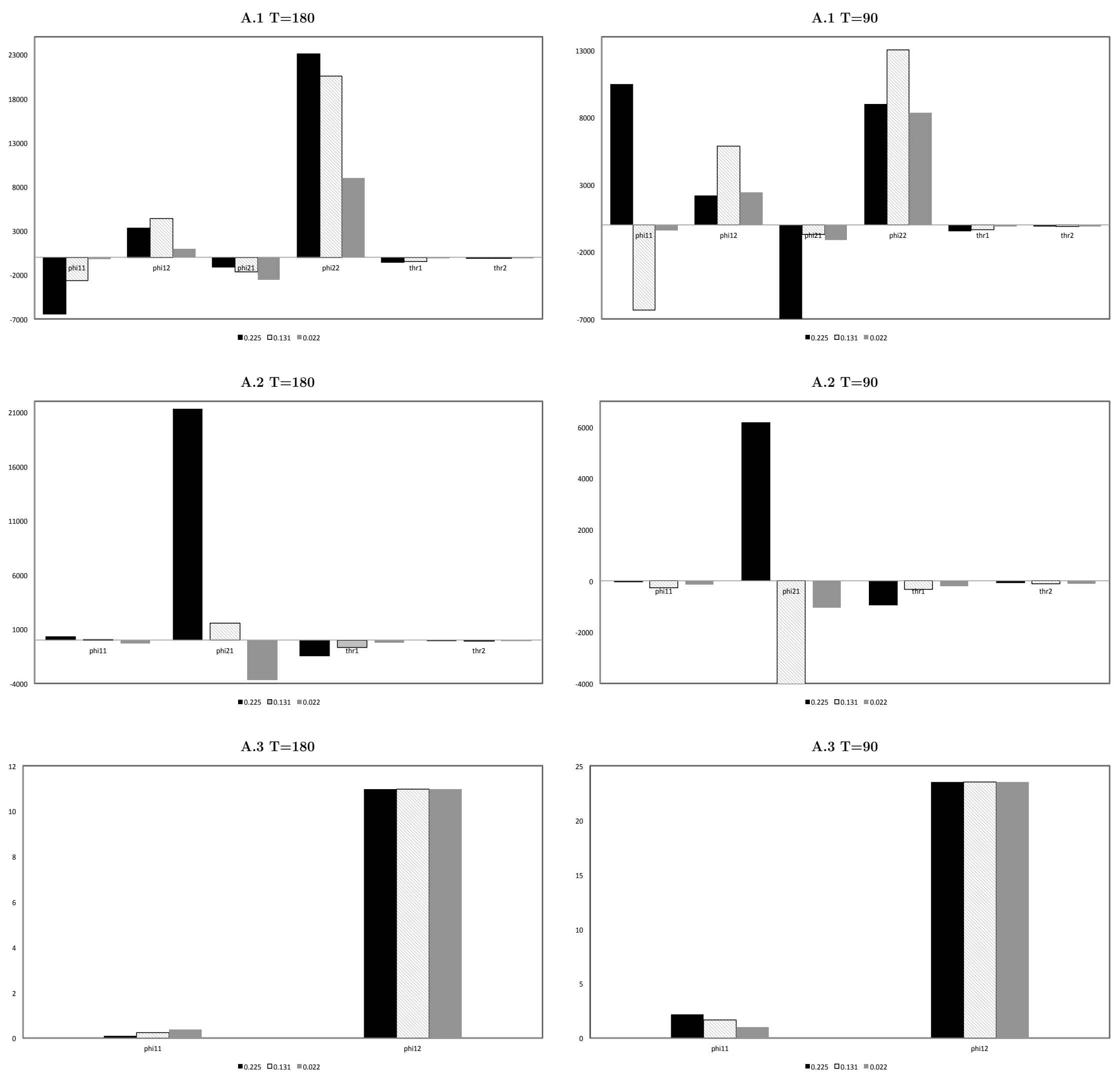


Figure 2.3: Experiment II, treatments A.1, A.2 and A.3

Note: y-axis=percentage bias ; x-axis =parameters of model (2.11), Number of simulations=10000.

2.6 Experiment II: STAR Estimation

This experiment is along the lines of the first one, but here the DGP is simplified to equation (2.11) as in the paper by Chan and Mc Aleer (2002). Again we use the dataset by VDF (2000), and basically we reduce the STECM model to a logistic STAR(1) model in first differences by eliminating the error correction mechanism, the exogenous variable and the threshold parameter c_2. The structural parameters for the experiment are still the ones in tables 2.1 and 2.3.

As figure 2.3 show, the NLS-CSQ estimator still seems to give estimates with less than desirable asymptotic properties. This time our benchmark autoregressive model (treatment A.2, medium panels of figure 2.3) is estimated almost flawlessly, since there is no effect of other endogenous or exogenous variables to take into account. However, the percentage bias obtained in the estimation of the entire STAR model by means of the NLS-CSQ estimator is still non-negligible: the upper panels of figure 2.3 show significant spikes, especially in correspondence of some of the autoregressive parameters. Also in this case, the descriptive statistics shown in table 2.9.7 of the Appendix confirms the diagnosis of inefficiency of the NLS-CSQ estimator by displaying extremely high standard errors for all the parameters estimates. A particular feature of this experiment is that only the scenario with the smallest standard error (grey bars) seems to give a bias close to zero, while the other two systematically give a relevant bias. In particular, in the scenario with $T = 180$ we observe a systematic over prediction of the parameters, in the order of 50 percent for the series generated with the lowest standard error. For the experiment with $T = 90$, the NLS-CSQ procedure still yields the smallest bias when the series is generated with the lowest simulation error (light bars), but there is a relevant over prediction for the series with $\sigma_\varepsilon = 0.131$ (shaded bars) counterbalanced by a relevant under prediction for the series with $\sigma_\varepsilon = 0.225$ (dark bars). Experiment A.3 (figure 2.3, lower panels) confirms that the tendency to yield an excess positive bias may be due to the particular influence of the φ_{21} parameter.

The comparison with the ML estimator may be informative also in this experiment. In figure 2.4 we compare the results of MLE with Marquardt algorithm (dark column) with the ones from the NLS-CSQ estimator (light column), again the bias obtained with BHHH is excluded from the analysis because it incomparably higher than the other two. When the generated sample contains 180 observations (left panels) the performance ranking is as it was in experiment I: MLE by means of Marquardt algorithm yields the lowest bias, followed by the two-step NLS-CSQ estimator and by the MLE with the BHHH algorithm. The same is true for the treatment with $T = 90$ (right panels), with the only difference that when the simulation error is the lowest (rightmost lower panel) surprisingly NLS-CSQ and MLE with BHHH perform equally bad, since the former overestimates the parameters almost to the same extent that the latter underestimates them.

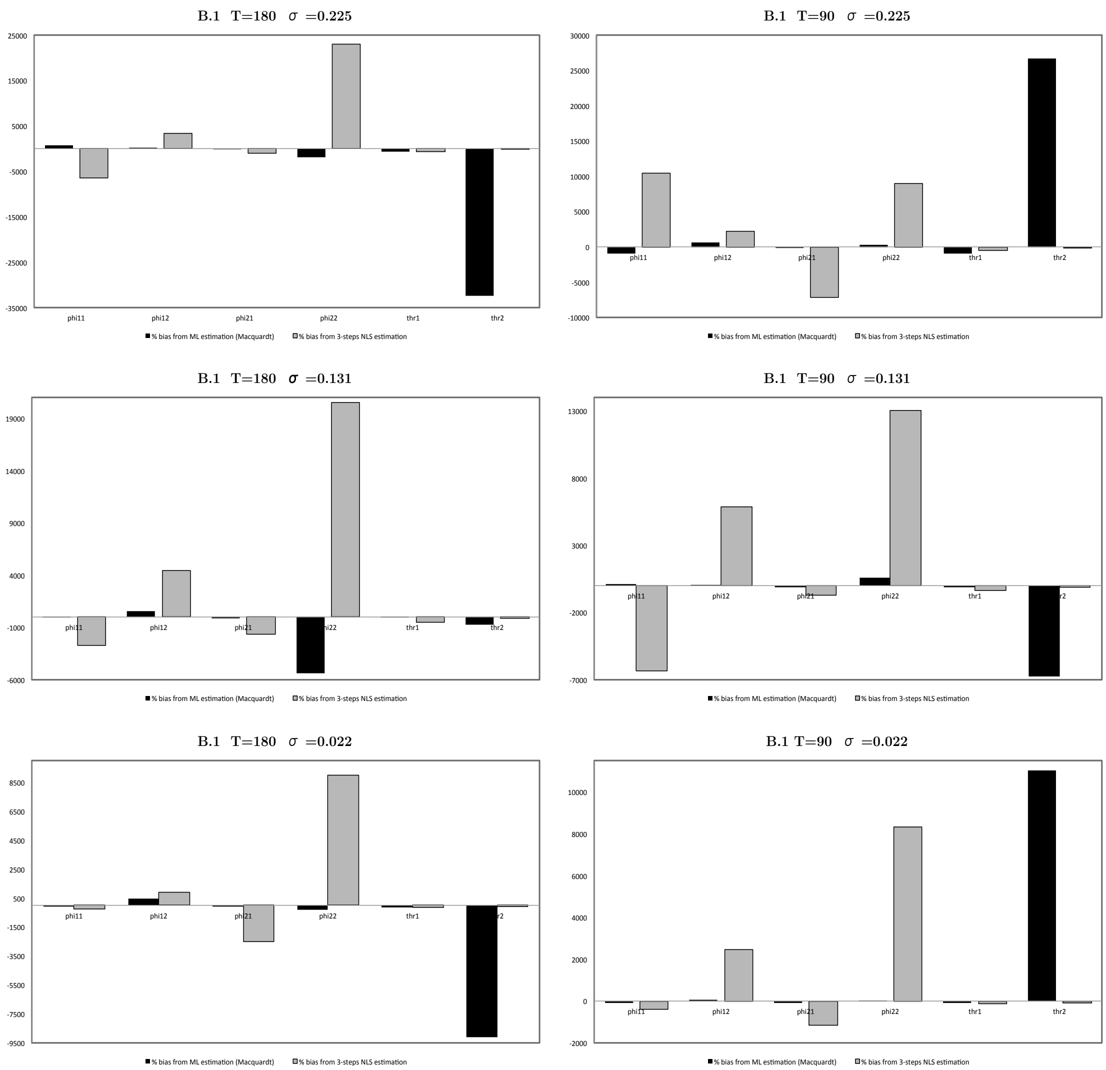


Figure 2.4: Experiment II, treatment B

Note: y-axis=percentage bias ; x-axis =parameters of model (2.11), Number of simulations=10000.

2.7 Investigating of the Shape of the Loglikelihood Function

So far our experiments showed that addressing the problem of ST models estimation both with NLS-CSQ and with MLE in small samples may not yield estimates with desirable asymptotic properties. However, If one has to make a choice, it seems to be then better to use a joint estimation procedure like ML, than a multi-step procedure like NLS-CSQ. Nevertheless, also in this case there is no guarantee that the estimates will be unbiased and consistent, unless the researcher has an idea in advance of the shape of the objective function and consequently of the most appropriate starting values.

Since this latter condition is almost never verified in practice, in this paragraph we start by envisaging a simple heuristic procedure which shows what could happen when the researcher does not have a clue of how 'good' are her starting values [17]. Afterwards, we present a possible way to overcome this problem by using a global optimization algorithm.

Standard ML estimation with a local optimization algorithm but without a clear idea of where the global optimum is located (hence of where the starting values should be) can have dramatic consequences. Let us consider the STAR model (2.11) generated in experiment II, and let us pick the scenario with $T = 180$ and with $\sigma_\varepsilon = 0.131$. This scenario was chosen ad hoc because the frequency of cases where the optimization algorithm found a singularity problem in one of the roots of the likelihood function was low[18]. We perform five subsequent ML estimation by always using the Marquardt-Levemberg algorithm, the one which proved more successful throughout the Monte Carlo experiments. The heuristic procedure consists in using a different set of starting values in each estimation, and to see how the obtained results change in term of

[17]it must be said that the pioneers of ST models (see for example Teräsvirta(1994) and Franses and Van Dijk (2000)) are always very clear on this point, frequently suggesting to perform a grid search to make sure to have appropriate starting values. Nevertheless, it is clear that the more the models gets complicated and the higher the number of parameters, the higher is the probability that the grid search does not give the expected results or that it takes too much time. As a consequence, the 'trick' of standardizing the exponent of the G function is often employed. We also followed this strategy throughout our Monte Carlo experiments.

[18]Indeed the problem of near singularity of the root of the loglikelihood function happens quite frequently, especially in the estimation of complex models like ST. The problem can be handled with the choice of an appropriate algorithm (like Marquardt which has a ridge correction factor) and by modifying appropriately the starting values.

parameters estimates and of likelihood function value. The five sets of starting values are calculated in the following way. The baseline set (*set1*) is appropriately chosen in order to ensure non-singularity. The other four starting values sets are obtained as 2 and -2 times *set 1* (respectively *set 2* and *set 3*), and 4 and -4 times *set 1* (*set 4* and *set 5*). Table 2.4 displays the outcomes of this procedure.

First of all, we notice that in all the five estimations convergence is always achieved and singularity occurs in just one of the cases, indicating that these features of the chosen generated model remain invariant to the choice of the starting points. Since the baseline set of parameters is precisely chosen as to be close to the known 'true parameters' used to generate the series, the first attempt to maximize the likelihood already yields a quite high likelihood, 118.086, even though the estimated parameters are far from the benchmark. However, it is evident that perturbating the starting points implies ending up in different peaks of the likelihood function, jumping from 118.086 to 115.790, then falling to -710.437 and finally stabilizing to 115.791. In each of these estimations we cannot avoid noticing how different from each other the estimates are. Only in a Monte Carlo framework the researcher knows what is the 'true value' of the parameters generating the data. In a routine estimation procedure, She can end up in either one these points of the loglikelihood with almost equal probability, unless She has a truly efficient strategy for selecting starting points.

This simple heuristic procedure is only aimed at giving us a better idea of how the loglikelihood of a simple ST model is shaped. From our example, we can infer that it is not a globally concave function and it displays more than one peak. Given that also Chan and Teodorakis (2010), report a somewhat similar evidence regarding the likelihood of STAR-GARCH models, involving a large flat region and more than one optimum, we suspect that the maximum likelihood estimation problem for ST models could benefit of the use of optimization algorithms different from the 'standard' ones. In the literature, this is not certainly a new discovery. For example, Maringer and Meyer (2008) analyze the problem of STAR models specification and estimation by comparing three optimization heuristics, the Threshold Accepting, the Simulated Annealing and the Differential Evolution. Their results indicate that some improvements may come from the use of these methods for the empirical researcher, but they do not seem to

STARTING VALUES						
	true parameters	set 1	set 2	set 3	set 4	set 5
phi11	-0.01	1	2	-2	4	-4
phi12	-0.04	-0.04	-0.08	0.08	-0.16	0.16
phi21	0.12	0	0	0	0	0
phi22	0.21	0.21	0.42	-0.42	0.84	-0.84
thr1	-7.38	0	0	0	0	0
thr2	0.42	0	0	0	0	0
sigma	0.131	1	2	-2	4	-4
MLE: ESTIMATED VALUES						
	true parameters	lstar1	lstar2	lstar3	lstar4	lstar5
phi11	-0.01	0.009	0.652	-0.276	1.763	-0.173
phi12	-0.04	-0.060	-1.267	0.088	-3.215	-0.723
phi21	0.12	-0.051	-0.775	19.565	-1.917	2.825
phi22	0.21	-0.336	0.644	-0.345	2.529	-3.978
thr1	-7.38	-1192.625	-7.279	-30.756	-6.372	6.021
thr2	0.42	-0.066	-0.267	-107.198	-0.412	-0.481
sigma	0.131	0.125	0.127	-0.234	0.127	-0.127
Log likelihood		118.086	115.790	-710.437	115.791	115.791
number of iterations		226	150	4	462	441
singular covariance proble		no	no	yes	no	no
convercence achieved?		yes	yes	yes	yes	yes

Table 2.4: Heuristic Investigation of the Log-likelihood function of model 2.11

tackle the issue of local optima that we are really concerned about. More recently, Bekiros (2009) proposes a combination of algorithms taken from the machine learning literature (Singular Value Decomposition and Gradient Descent), which should be able to avoid the problem of getting stuck in only one portion of the parameters space, hence reducing the risk of 'over-fitting the model', one of the mayor causes of divergence or entrapment in a local optimum.

Here, depart from this type of literature by abandoning the devices typical of local optimizations algorithms and embracing a global optimization approach: we performed a ML estimation by means of a global algorithm which also gives us a picture of the loglikelihood function as a by-product of the estimation. Weise (2009) synthetically defines global optimization algorithms as the ones that "employ measures that prevent convergence to local optima and increase the probability of finding a global optimum." [19]. We used Tucci's (2002) global optimization algorithm EZGRAD, since it is particularly easy to use and it adapts very well even to very complicate objective functions like ours. The algorithm combines the gradient procedure first implemented by Tucci (1998) and an 'accordion type' of boundary mechanism which renders the algorithm more robust. In a nutshell, the algorithm works as follows: the main building block of the program starts a standard gradient procedure from different initial points. When the gradient optimum obtained starting from two different values of the control variable is the same, it is concluded that the function under consideration does not present non-convexities. Otherwise the costs associated with the local optima are compared and the lowest is chosen as the global solution of the optimization problem. The main difference with respect to the majority of the other algorithms is that it starts with very few points scattered all over the search space, and it checks if the function satisfies the 'regularity conditions'. If not it increases exponentially the number of points evaluated, hence it is fast with well-behaved functions and slow with more complex ones.

The implementation of MLE by means of EZGRAD with around 22,000 replications yields some interesting results which we can summarize in three mayor points.[20]

[19] For a thorough treatment of global optimization algorithms please refer to the above mentioned book.

[20] More exactly the number of replications is 229378. It is important to notice also that the algorithm works with a minimization program, hence the loglikelihood had to be specified with a negative sign. For this reason,

First of all, without giving the algorithm any indications about starting values other than their domain, it is able to reach a quite good local maximum within the likelihood function, namely the value of 114.2229. Indeed, this is one of the advantages of the global optimization algorithms, they do not depend on the researcher's guess on starting values. As we saw previously, Marquardt reaches a 'better' maximum, 118.086, only when provided with starting values close to the true parameters, but this result is very sensitive even to their small perturbations.

Secondly, there is a high number local maxima yielding a value of the likelihood quite close to 114.2229. For example, a likelihood value around 112 is obtained in 105 replications, while 113 is obtained in 39 cases out of 22,000. To give a more evident representation of this feature of the loglikelihood, we selected six other local minima in the neighborhood of 112 (a value which is actually quite close to 114.2229) and we depicted the values of estimated parameters in each of these points in a bar graph. Figure 2.6 speaks quite eloquently: from the different heights of the columns we see that the oscillation range for each of the parameters can be very wide and yet almost the same loglikelihood value is obtained. This feature is particularly pronounced for the parameter φ_{22} (dark grey column in the top panel) and for the transition parameter γ (lightly shaded column in the bottom panel).

Finally, EZGRAD gives us an indication of the approximative shape of the loglikelihood function, by saving its value for each iteration of the search. Figure 2.5 shows only a portion of this loglikelihood, and in particular the one where the best local minimum of 114.229 is found. This piece of information is particularly valuable since it further confirms our thoughts about the irregular shape of this function. It is possible to see from the figure that there are many minimum peaks, and many of them are not far from the values of 114.229. With an effort of imagination, we can infer how such loglikelihood would look on a multi-dimensional graph: it would probably display wide flat regions or 'ridges', implying that standard local algorithms may frequently fail due to near singularity problems.

all the figures pertaining to the application of EZGRAD will display the negative of the loglikelihood value on the vertical axis, and when we refer to the values of the loglikelihood function the minus sign is always implicit.

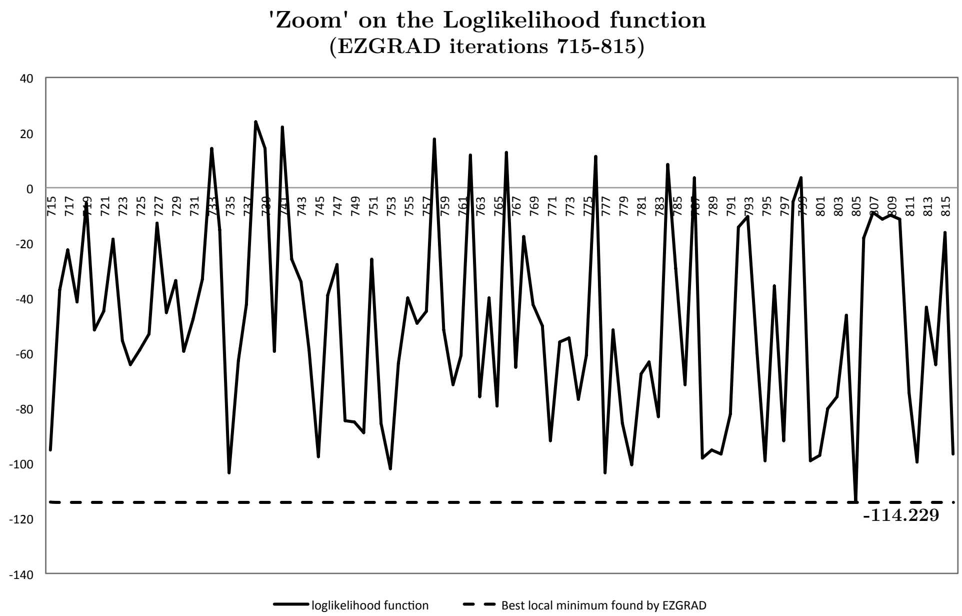


Figure 2.5: Empirical loglikelihood function estimated by EZGRAD for model (2.11)
Note: Only iterations from 1 to 900 are shown.

Vertical axis: negative of loglikelihood value. Horizontal axis: iterations of the minimization procedure.

2.8 Concluding Remarks

The paper examined the structural and asymptotic properties of the Nonlinear Least Squares estimator of Smooth transition models when the usually standard procedure of 'Concentrating the sum of squares function' is employed. We presented numerical simulation results regarding the finite sample properties of the NLS-CSQ estimator for the STECM(1) and STAR(1) models, and we compared these results with the ones obtained by joint Maximum Likelihood estimation.

Experiment I focused on a 13-parameters STECM(1) model, whose size and complexity generally provides the rationale for using the CSQ procedure. Nevertheless, we showed that the gains of using the NLS-CSQ procedure in terms of reduced computational burden and lower estimation time may be more than offset by the bias and lack of robustness of the resulting estimates. As a further check of this result, Experiment II proposed the same simulation scenario on a much simpler 5-parameters STAR(1) model. Even in this case, the MLE seems to outperform the NLS-CSQ estimator, which proves to be particularly flawed in small samples. It is worth mentioning that

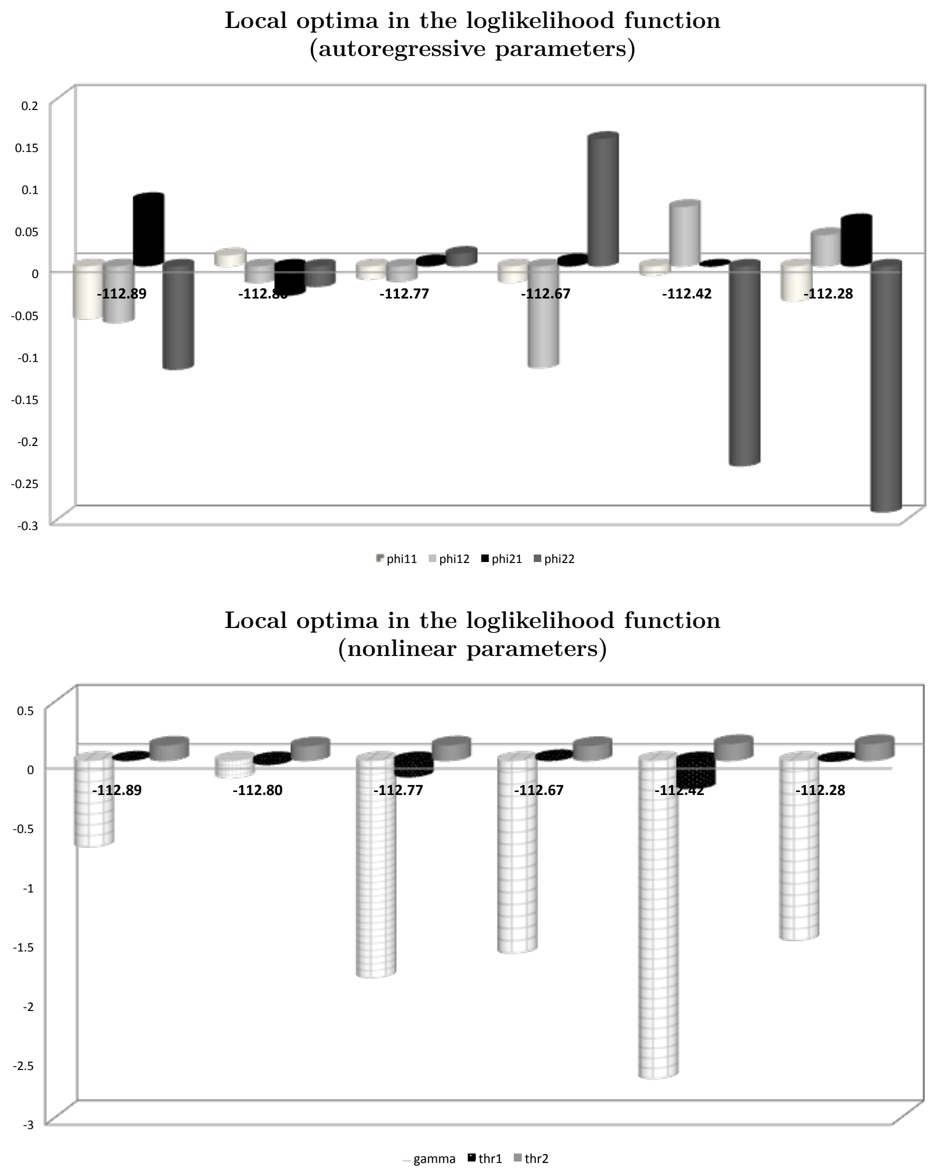


Figure 2.6: Different points on the empirical loglikelihood function of model (2.11) giving rise to local optima in the neighborhood of 112

Note: Vertical axis: negative of loglikelihood value.

Horizontal axis: autoregressive parameters $\varphi_{11}, \varphi_{12}, \varphi_{21}, \varphi_{22}$ (upper panel) and $\gamma, c_1(thr1), c_2(thr2)$ (lower panel).

our measure of a 'small sample' here is a $T = 180$ observations span, a very common (and even desirable) situation in applied macroeconomics problem. A simple heuristic analysis on the loglikelihood function of a logistic STAR model confirmed our suspects that estimation difficulty is most likely caused by a non-regular likelihood function with more than one maximum, and this is in line with Chan and Teodorakis' (2010) results. This result was strengthened when we employed a global optimization algorithm to maximize the loglikelihood function.

In sum, our findings can be summarized with these three major points:

(1) For the estimation of ST models it is better to avoid the NLS-CSQ estimator, unless *a)* the model is particularly simple and/or *b)* the available sample of observations is very large, ensuring a very small residual standard error in the estimation.

(2) When none of the conditions above are met, it is safer to perform a joint estimation of the ST model parameters by means of MLE.

(3) Particular caution must be used in the choice of the numerical algorithm beneath MLE. Given that there is no guarantee that a grid-search will result in starting values which are sufficiently close to the optimum, the use of a global optimization algorithm is strongly advised.

2.9 Appendix: Detailed Descriptive Statistics

T=180, Experiment I, Treatment A1

σ	0.0220					0.131					0.225				
	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis
phi11	-0.010	-0.010	0.003	0.020	3.086	-0.698	-0.013	23.442	-11.377	1096.861	-1.394	-0.013	44.054	-13.962	664.160
phi12	0.050	0.050	0.004	-0.035	2.965	0.360	0.053	12.376	-0.529	822.659	0.373	0.054	23.750	-6.423	638.653
phi13	0.810	0.810	0.004	-0.049	2.936	0.201	0.793	9.850	-27.251	1269.379	-0.609	0.769	19.909	-8.448	412.603
phi14	-0.042	-0.042	0.037	-0.030	2.927	-0.424	-0.057	49.451	-0.782	553.187	-0.969	-0.055	80.475	-4.671	339.745
phi15	0.072	0.072	0.032	0.036	2.912	0.284	0.085	44.181	-5.315	706.809	0.511	0.080	75.917	-1.151	341.401
phi21	0.123	0.117	0.043	0.477	2.579	3.652	0.065	110.084	9.581	659.374	1.710	0.047	261.322	9.852	792.509
phi22	0.104	0.097	0.047	0.661	3.269	4.109	0.032	78.052	24.248	880.231	7.413	0.018	170.051	20.624	880.044
phi23	0.420	0.399	0.115	0.213	1.756	4.098	0.248	59.643	26.546	1457.174	9.374	0.276	128.004	10.613	448.725
phi24	0.209	0.197	0.144	0.467	3.573	-1.775	0.132	246.800	17.005	1220.025	3.729	0.113	552.938	0.355	836.911
phi25	-0.211	-0.198	0.126	-0.538	3.569	-0.556	-0.145	211.015	-28.558	1717.554	-7.906	-0.125	512.126	-11.827	910.509
thr1	1.192	1.074	2.258	92.709	9019.422	24.160	0.547	711.012	-17.644	1592.495	24.249	0.157	766.266	70.361	6328.702
thr2	0.723	0.721	0.319	0.079	1.769	4.377	0.465	200.601	18.036	1269.648	12.728	0.289	251.921	6.037	183.307
th3	0.729	0.723	0.320	-0.044	2.763	4.786	0.439	184.341	1.119	410.473	9.154	0.301	275.791	-8.378	502.399

T=90, Experiment I, Treatment A1

σ	0.0220					0.131					0.225				
	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis
phi11	-0.010	-0.010	0.004	0.060	3.140	-0.686	-0.016	27.876	-9.828	482.069	-0.583	-0.016	48.675	9.024	938.127
phi12	0.050	0.050	0.005	-0.038	3.137	0.283	0.057	14.419	1.325	439.243	0.362	0.055	25.144	-6.905	487.793
phi13	0.810	0.810	0.005	-0.059	2.934	0.194	0.787	11.154	-6.990	427.213	-1.003	0.759	21.508	-14.775	363.023
phi14	-0.044	-0.044	0.045	-0.022	2.992	-1.307	-0.062	59.701	-2.490	419.790	0.989	-0.057	100.740	4.927	439.516
phi15	0.074	0.074	0.040	0.034	3.022	0.783	0.090	54.697	0.320	460.864	-1.393	0.081	87.589	-4.206	255.485
phi21	0.119	0.108	0.053	0.724	3.192	6.785	0.061	127.698	5.660	361.246	12.850	0.047	358.640	19.903	902.083
phi22	0.100	0.091	0.059	0.737	3.848	6.264	0.025	92.260	16.974	571.624	11.112	0.026	234.895	29.008	1479.608
phi23	0.409	0.373	0.134	0.287	1.628	6.564	0.241	63.470	16.888	435.935	15.575	0.306	165.993	18.730	523.288
phi24	0.211	0.192	0.232	0.518	4.908	-8.735	0.111	247.226	-7.150	475.352	-6.514	0.081	575.954	-12.187	755.472
phi25	-0.213	-0.192	0.199	-0.623	4.929	5.333	-0.119	232.850	1.011	563.604	4.941	-0.099	476.968	4.898	550.004
thr1	2.410	1.621	8.562	64.559	5078.956	11.505	0.483	1044.256	-76.818	7104.768	14.646	0.207	467.085	-4.122	974.636
thr2	0.725	0.726	0.344	0.000	2.661	8.187	0.444	207.392	8.614	224.483	14.336	0.265	248.252	9.057	237.654
th3	0.724	0.725	0.347	-0.214	4.826	9.765	0.417	206.299	5.655	193.372	11.857	0.269	242.849	3.801	173.745

Table 2.9.5: Detailed descriptive statistics for Experiment I

T=180, Experiment I, Treatment B1

σ	0.0220					0.131					0.225				
	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis
phi11	-0.023	-0.009	0.471	-57.014	4759.859	-0.031	-0.008	0.503	-0.599	92.705	-0.080	-0.007	1.220	-0.171	99.970
phi12	0.046	0.047	0.224	-35.921	2773.491	0.059	0.036	0.345	-4.332	351.734	0.086	0.051	1.120	-2.653	751.864
phi13	0.730	0.808	0.624	-22.004	1305.996	0.585	0.446	0.623	5.280	275.889	0.704	0.769	2.251	27.669	1427.298
phi14	-0.031	-0.015	0.813	-57.918	4723.947	-0.029	-0.002	0.800	1.918	173.127	-0.058	-0.047	2.212	12.363	558.142
phi15	0.037	0.046	0.566	-30.417	2188.584	0.043	0.025	0.727	-3.989	207.414	0.072	0.072	2.334	-16.608	737.957
phi21	0.154	0.075	1.433	35.013	1605.574	0.450	-0.006	23.997	9.005	1047.211	0.814	0.025	59.446	42.343	3189.652
phi22	0.076	0.064	0.412	10.790	465.896	0.342	0.047	15.888	13.549	1947.713	0.986	0.034	38.495	44.281	3168.670
phi23	0.475	0.480	1.208	20.097	1334.416	0.767	0.830	8.552	-22.261	2858.355	1.518	0.245	40.940	52.768	4427.470
phi24	0.151	0.087	1.173	19.714	1256.649	-0.324	0.042	19.871	-22.523	1318.634	-0.299	0.080	36.119	-6.438	428.221
phi25	-0.118	-0.079	1.021	10.475	688.800	-0.098	-0.013	15.753	-8.158	1230.148	-0.531	-0.086	34.573	-15.404	726.329
thr1	-0.143	0.124	31.524	-64.832	4553.221	0.911	-0.002	45.799	10.927	936.243	-7.896	-0.116	80.263	-2.439	145.527
thr2	-1.077	0.452	129.657	-72.749	7064.495	-1.066	0.000	46.402	1.973	215.228	-6.694	-0.287	86.839	-3.844	143.697
th3	-1.640	0.466	129.488	-79.380	7051.640	-1.455	0.000	47.065	-12.284	479.212	-5.850	-0.288	90.994	5.816	605.767

T=90, Experiment I, Treatment B1

σ	0.0220					0.131					0.225				
	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis
phi11	-0.023	-0.011	0.239	-11.598	422.884	-0.056	-0.009	0.670	-2.586	91.040	-0.020	-0.004	3.420	26.311	1990.274
phi12	0.047	0.050	0.145	-3.974	377.100	0.062	0.038	0.411	-3.337	126.271	0.077	0.040	4.060	33.054	2220.485
phi13	0.729	0.809	0.374	2.562	205.849	0.617	0.693	0.664	-0.062	127.958	0.586	0.745	9.246	-43.472	3501.808
phi14	-0.022	-0.014	0.473	7.776	426.176	-0.057	-0.015	1.216	17.223	1044.316	-0.145	-0.044	14.020	-72.046	6540.303
phi15	0.047	0.044	0.445	3.963	639.867	0.064	0.035	1.228	-38.021	2806.730	0.098	0.067	4.827	4.374	349.448
phi21	1.389	0.089	51.125	46.642	3453.518	1.054	0.003	23.791	2.021	580.076	0.063	-0.001	47.989	-16.295	723.000
phi22	0.920	0.073	53.100	87.513	8282.770	0.651	0.048	14.898	10.071	542.233	0.178	0.048	34.785	-32.662	2242.074
phi23	0.574	0.268	10.142	12.441	2632.661	1.141	0.603	10.167	14.323	520.629	1.644	0.287	24.716	17.861	762.856
phi24	0.575	-0.006	59.429	57.468	3702.599	-0.772	0.030	19.162	-15.263	626.956	-0.629	0.035	35.734	-2.325	586.934
phi25	0.262	0.030	62.564	2.108	4150.277	0.608	0.002	18.835	32.233	2007.219	0.842	-0.038	38.442	29.642	1518.455
thr1	0.350	-0.040	47.559	80.501	7668.367	-0.400	-0.003	42.020	34.114	2904.754	-9.944	-0.345	64.000	-19.365	701.123
thr2	-2.598	0.107	74.914	21.601	1001.860	-0.316	0.000	40.772	5.814	254.332	-1.968	-0.224	58.484	6.160	366.101
th3	-3.714	0.084	84.377	6.116	1759.650	-0.437	0.000	40.344	-1.414	180.981	-2.602	-0.227	71.754	15.822	830.321

Table 2.9.6: Detailed statistics for experiment I, treatment B.1

T=180, Experiment II, Treatment A1

σ	0.0220					0.131					0.225				
	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis
phi11	0.014	-0.003	0.826	-37.267	2460.610	0.258	0.000	24.521	33.939	3380.455	0.630	0.001	48.131	34.763	2256.082
phi12	-0.405	-0.032	13.782	33.933	3186.886	-1.808	-0.041	71.711	-53.169	3127.788	-1.376	-0.030	82.834	-32.354	2086.953
phi21	-2.915	-0.010	71.579	-4.304	138.112	-1.877	-0.022	477.461	-0.783	230.968	-1.202	-0.022	922.780	-1.213	274.650
phi22	19.116	-0.086	418.620	10.188	330.933	43.310	-0.087	1095.536	19.420	658.131	48.569	-0.074	1443.133	15.158	501.788
thr1	3.138	0.081	59.029	0.469	124.929	30.454	2.321	413.555	1.863	171.971	37.467	2.202	785.556	-4.064	595.911
thr2	-0.006	-0.007	0.299	-45.613	4122.381	0.006	-0.031	1.392	43.243	2582.635	-0.073	-0.045	3.992	-67.834	5229.403

T=90, Experiment I, Treatment A1

σ	0.0220					0.131					0.225				
	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis
0.029	-0.002	1.300	59.308	4834.470	0.627	-0.001	52.121	97.352	9634.850	-1.058	0.000	51.142	-43.857	2058.219	
-1.018	-0.028	38.311	-83.400	7697.987	-2.380	-0.037	167.231	-97.946	9716.369	-0.911	-0.028	100.641	-19.512	2396.061	
-1.257	-0.011	58.097	-0.823	178.167	-0.741	-0.016	453.758	-1.208	261.404	-8.413	-0.013	957.631	-25.809	1545.615	
17.714	-0.117	402.636	8.817	533.717	27.528	-0.106	1096.020	13.600	654.092	19.069	-0.089	1567.257	26.195	2196.032	
0.820	0.175	53.375	-10.685	327.657	17.261	2.474	399.564	-8.184	1200.891	26.980	3.548	492.355	6.342	213.605	
-0.017	-0.005	0.738	-74.296	6516.672	0.018	-0.029	2.925	48.653	3920.260	-0.062	-0.042	4.044	-9.690	1340.791	

Table 2.9.7: Detailed statistics for experiment II, treatment A.1

T=180, Experiment II, Treatment B1

σ	0.0220					0.131					0.225				
	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis
phi11	-0.001	0.000	0.105	-58.250	3984.809	-0.002	-0.003	4.427	-18.116	2991.348	-0.080	-0.003	5.835	-10.970	1380.015
phi12	-0.235	-0.040	13.471	-90.821	8606.756	-0.266	-0.118	16.392	-41.174	3177.614	-0.112	-0.070	59.830	41.510	5357.822
phi21	-0.001	0.000	0.105	-58.250	3984.809	-0.002	-0.003	4.427	-18.116	2991.348	-0.080	-0.003	5.835	-10.970	1380.015
phi22	-0.432	0.210	45.692	-47.337	3047.263	-11.054	0.047	671.020	-60.925	4045.645	-3.611	0.022	519.117	-64.790	5462.730
thr1	1.210	0.000	152.089	61.660	4517.873	-6.615	-0.008	626.481	-43.587	2803.593	42.408	0.123	3110.201	62.028	4334.365
thr2	-37.796	0.000	3779.012	-99.981	9997.462	-2.753	0.000	676.047	-21.718	3365.851	-135.781	0.001	14069.113	-99.723	9963.480

T=90, Experiment I, Treatment B1

σ	0.0220					0.131					0.225				
	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis	mean	median	standard deviation	skewness	kurtosis
phi11	-0.001	0.000	0.071	-3.926	788.268	-0.019	-0.003	8.533	32.718	4363.312	0.081	-0.002	3.787	30.952	1546.813
phi12	-0.063	-0.040	3.281	16.750	1225.626	-0.074	-0.135	29.594	39.056	3448.831	-0.281	-0.083	19.981	-45.824	2992.314
phi21	-0.001	0.000	0.071	-3.926	788.268	-0.019	-0.003	8.533	32.718	4363.312	0.081	-0.002	3.787	30.952	1546.813
phi22	0.221	0.210	10.555	-29.662	1420.124	1.458	0.127	213.787	23.306	3230.580	0.848	-0.018	264.842	29.505	2514.355
thr1	-0.048	0.000	59.288	14.990	1791.632	3.581	0.054	3352.846	55.828	5844.949	67.017	0.051	6497.508	98.791	9839.607
thr2	46.669	0.000	4548.734	99.936	9991.478	-27.841	0.000	1795.364	-69.079	5179.962	112.529	0.000	11310.896	98.756	9833.042

Table 2.9.8: Detailed statistics for experiment II, treatment B.1

Chapter 3

Forecasting Italian GDP Growth through the Term Spread: How Many Stars to the STAR Model?

Abstract

This paper analyzes the link between the Italian GDP growth rate and the interest rate spread by means of a nonlinear approach.[1]

First, reconnect to the work of Brunetti and Torricelli (2009) by robustly estimating a Smooth Transition Autoregressive (STAR) model of the Italian growth-spread relationship in the years 1997-2005. Then, we focus on the predictive performance of the STAR model by proposing a sort of 'horse race' between different forecasting models at consecutive forecast horizons. Simulation based forecasts of the STAR model are evaluated against forecasts from an equivalent state-space system with time-varying parameters, and against a benchmark autoregressive model. Moreover, these competing models are also pooled together with a variety of weighting schemes and their relative performance is evaluated. We find that for short-term forecast horizons the STAR model is quite effective in predicting GDP growth. However, the farther the forecast horizon, the worst its performance gets in comparison to its equivalent time-varying competitor. In addition, although none of the considered models is able to

[1] I would like to thank Marianna Brunetti for the precious suggestions and for providing me her dataset, and a special thank goes to Marco Tucci for his guidance and support throughout my research. All remaining errors are mine.

capture the turbulences of the 2007 financial crisis, we show that using forecast combination schemes substantially increases the predictions' accuracy.

JEL Codes: C51, C53, E37.

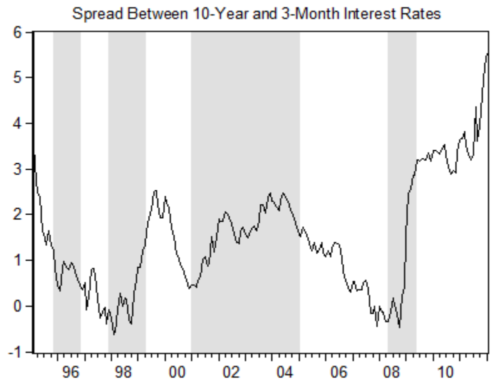
Keywords: Smooth Transition Models, Term Spread, Forecast Combination.

3.1 Introduction

Imagine we had a theoretical model being able to explain how much GDP growth changes in reaction of changes in the slope of the yield curve, i.e. the spread between long and short term interest rates. Using such model for making predictions in advance would not only have significantly helped Greece or Italy during the past stormy months, but it would have provided also a guideline for the actions of governments and central banks. Unfortunately, a unanimous theoretical explanation of the growth-spread relationship does not exist, and as Benati and Goodhart (2008) put it, this still remains a "stylized fact in search of a theory". It may be informative to look at figure 3.1, displaying the Italian interest rate spread and the Italian recession periods according to the OECD chronology. Indeed, most recessions are preceded by a decline in the slope of the yield curve, and often also by an inversion of the yield curve with short-term bond yields exceeding those on long-term maturities.

The association between the Italian term spread and the business cycle can be measured more precisely by looking at the correlations between the term spread and the year-over-year percentage change in real gross domestic product. The right panel of figure 3.1 presents the contemporaneous correlation between the two variables for the period 1995-2010, as well as correlations at respectively six lags and leads of the term spread relative to GDP growth. As anticipated by the graphical analysis, the correlation between GDP growth and lagged term spreads are generally positive and significant, while the correlations with the leads of the term spread are uniformly negative and significant. This can be interpreted as an indication that the more steeply sloped is the yield curve, the higher the rate of future GDP growth. Moreover, a higher GDP growth rate in one quarter is associated with a less steeply sloped yield curve in subsequent quarters.

Figure 3.1: Italian Term Spread and Recessions



Correlation of Italian GDP Growth and Lagged and Future Term Spreads

lagged term spread			future term spread		
t	-0.216		t	-0.216	
t-1	0.045		t+1	-0.3826	**
t-2	0.2899	**	t+2	-0.4051	**
t-3	0.4298	**	t+3	-0.3441	**
t-4	0.3976	**	t+4	-0.2651	**
t-5	0.3208	**	t+5	-0.1837	
t-6	0.2066		t+6	-0.0966	

Note: (Left panel) The term spread is calculated as the difference between the yields on 10-year and 3-month government bonds, here it is measured as quarterly averages of monthly observations. The shaded areas denote recessions according to the OECD chronology.
(Right panel) The figures refer to quarterly data from 1995Q1 to 2010Q12; the two stars indicate that correlations are significant at the 5 percent level. Real GDP growth has been calculated as the percentage change from the previous year on the real GDP series [source: OECD].

From the theoretical viewpoint, the Rational Expectation Hypothesis provides one of the possible rationales for using the term spread as an indicator of market expectations about future economic conditions. In this framework, long-term interest rates are averages of appropriate expected future short-term interest rates, the link being summarized by the so called yield curve. A change in the term spread corresponds to a change of the slope of the yield curve, hence when the market anticipates a recession, a reduction in expected future short-term interest rates is anticipated by a flattening yield curve. Other possible theories that have been proposed to explain this relationship span from a purely monetary explanation, to a more theoretical one relating to with consumers' expenditure smoothing behavior.

On the empirical side, the literature finds that the spread predicts output growth more accurately in some countries and some periods than in others and in particular, that the term spread's power to forecast output has diminished since the mid-1980s. Several recent studies also find evidence of significant nonlinearities, such as threshold effects, and smooth regime changes within the relationship. Among those, Brunetti and Torricelli (2009) focus on the Italian case and find evidence for significant regime switches.[2] There are two main reasons why a linear representation of the link between these two

[2] See Wheelock and Wohar (2009) for a comprehensive survey.

variables cannot be considered as satisfactory: the relationship is asymmetric, namely it differs depending on past values of the spread being positive or negative, and it also displays regime-switching behavior, as the informational content of the spread changes with the regime in place[3]. As a consequence, Smooth Transition (ST) Models have often been used for the purpose of estimating this relationship and formulating forecasts at different horizons.[4]

Unfortunately, one of the weaknesses of this approach is that ST models are neither easy to specify, nor to estimate and produce forecasts with. As a consequence, many authors in this literature have estimated smooth transition models but then they come up with alternative specifications more practically useful for predictive purposes.[5] Interestingly, these problems underlined by the literature could depend on computational issues related the technique used to estimate and forecast with ST models, more than on theoretical issues. Our research so far has revealed that the use of some standard econometric procedures for nonlinear estimation problems with a high degree of complexity, may result in poor estimates characterized by an even poorer asymptotic performance[6].

In this work, the starting point is Brunetti and Torricelli's (2009) logistic STAR specification for the growth-spread relationship, but we improve on that work in a twofold way. On the one hand, we use an updated version of their dataset from 1997 to 2011: the monthly dataset focuses on Italy and it is made of three and ten months interest rates on Italian government bonds and by the industrial production index. Furthermore, we re-estimate their model by means of Maximum Likelihood, while they used Nonlinear Least Squares. Indeed, even if these two methods should be equivalent in

[3]On this topic see for example Galbraith and Tcakz (2000), Venetis et al (2003), and more recently Brunetti and Torricelli (2009).

[4]For sake of clarity, let us reproduce the structure of the popular first order two-regime STAR model (Teräsvirta, 1994), which is: $y_t = (\phi_{0,1} + \phi_{1,1}y_{t-1})(1 - G(y_{t-1}, \gamma, c)) + (\phi_{0,2} + \phi_{1,2}y_{t-1})G(y_{t-1}, \gamma, c) + \varepsilon_t$. $G(y_{t-d}; \gamma, c)$ is a continuous and at least twice-differentiable function, and the 'smoothness' or the speed of the transition is determined by the size of the γ parameter, which is strictly positive. Pioneering studies on ST models are the ones by Granger and Teräsvirta(1993) and Van Dijk and Franses (2000)

[5]For example Brunetti and Torricelli (2009) are forced to use a binary probit model to predict recession probabilities.

[6]As we showed in chapter two

principle, our previous results point out that for the estimation of ST models it is better to avoid the (almost univocally used) Nonlinear Least Squares estimator and instead, it is safer to perform a joint estimation of the parameters by means of Maximum Likelihood. As a result, we obtain different parameters estimates.[7] On the other hand, we focus on the predictive properties of the estimated growth-spread relationship by running a sort of 'horse race' between competing forecasting models. This choice is motivated by the fact that, although producing one-step forecasts for a STAR model is a manageable task, multi-step forecasts are much more difficult because the multivariate distribution of the model error will enter the nonlinear data generating process. Exactly for this reason, several numerical methods have been proposed to produce multi-step forecasts and in this work we will focus on three main macro-classes. First, we use the re-estimated Brunetti and Torricelli's LSTAR model and we compute forecasts by using the methods proposed by Lundbergh and Teräsvirta (2002), namely *Monte Carlo* and *Bootstrap* forecasts; the comparative analysis will be carried on over different forecasts horizons (respectively 1, 3, 6, 12, 24 months ahead). Afterwards, we exploit Granger's (2008) suggestion that "any non-linear model can be approximated by a time-varying parameter linear model", and produce forecasts with an equivalent representation of the growth-spread STAR relationship as a state-space model with time-varying parameters. This state-space model can be robustly estimated via Kalman Filter, once maximum likelihood estimates of the structural parameters are obtained through prediction error decomposition (Harvey, 1981). The final set of competing forecasts are generated using different forecast combination methods and weighting schemes. Indeed, there is a growing body of theoretical and empirical papers arguing in favor of combined forecasts, especially when data complexity is such that there is no single model that dominates the others in terms of predictive performance (Timmerman, 2006). Furthermore, forecast pooling may be a good remedy to the so

[7] The problem with maximum likelihood estimation however is that it requires good starting values to achieve the optimum, and to this end a grid-search is usually employed. However, performing a grid search on the whole set of parameters may be computationally intensive while carrying it on a subset of parameters does not guarantee to end-up close to the global optimum. Indeed, in chapter 2 we showed that choosing starting values not sufficiently close to the optimum in Smooth Transition model estimation may have serious consequences since the loglikelihood of this type of models displays multiple optima.

called 'Winner's curse', involving that a model which displays a very good in-sample fit, typically produces very poor forecasts (Hansen, 2009).[8] Finally, the competing models' out-of-sample performance is evaluated relatively to a benchmark autoregressive model. The comparison is carried on by complementing a graphical analysis with the evaluation of various forecasts metrics, including 'relative efficiency' measures, root mean squared forecasts errors (RMSFE), and Diebold-Mariano statistics (Diebold and Mariano, 1995). The final objective of this research project is to give a qualified answer to the important question: how well does the term spread predict the growth rate of Italian economic activity? The issue is clearly of major importance for both researcher and policy-makers. With the combination of the robust estimation approach we are using and of our composite forecasting technique, we are able to shed some new light on this important question.

The remainder of this paper goes as follows. Section 2 gives a brief literature review on the term spread as a possible predictor of future economic activity, while section 3 outlines the methodology we have used for both the in-sample model estimation and the out-of-sample model evaluation. In particular, the 'horse race' forecasting exercise is thoroughly described in the final part of the section. Section 4 is devoted to illustrate the motivations that led us to choose exactly the 1997-2005 period for our estimation, and section 5 finally describes the results of our estimation and goes deeply into the details of the forecasting exercise. Finally, section 5 concludes.

3.2 Background

So far the literature has offered three main explanations for the leading indicator property of the term spread, one based on Rational Expectations theory, one connected to Monetary Policy, and a third one relating to the Intertemporal Consumption theory.

According to the former, the combination of the Fisher relation between nominal interest rates and expected inflation and the Rational Expectations Hypothesis (REH) implies that the risk premium can be seen as constant over time. Given these premises,

[8] Among the various contributions on forecast combination it is worth mentioning at least Stock and Watson, 2004, Hendry and Clements, 2002, and Marcellino, 2004.

the term spread should contain information about future economic activity and inflation. More precisely, when the market anticipates a recession, a reduction in expected future short-term interest rates is reflected by a flattening yield curve and the opposite happens for a boom. Some notable examples of paper making use of this approach are the ones by Crespo et al. (2005) and Haug and Siklos (2006).

The Monetary Policy explanation instead relies on the fact that a policy tightening is likely to cause the yield curve to flatten or possibly invert. This is due to the fact that both long-term and short-term interest rates are expected to rise, but the latter will rise more than the former if the effect of monetary policy is expected to be temporary. A more rigorous description of such a mechanism is provided by the models with price rigidities by Estrella (2005), and Estrella and Trubin (2006). Interestingly, these models also suggest that the term spread might forecast output growth better, the more responsive the monetary authority is to deviations of output growth from potential. The predictive ability of the spread should then decrease the higher the 'degree of conservativeness' of the central bank.

Finally, the intertemporal consumption approach (Harvey, 1988 and Hu, 1993) directly links changes in the slope of the yield curve to changes in real activity as a consequence of the consumption smoothing behavior of individuals. For example, when consumers expect a recession they might want to sell short-term financial instruments and purchase long-term bonds to avoid abrupt changes in their consumption pattern. This results in a flatter real term structure of interest rates. However, it is nominal term structure that matters when we want to make predictions: only if inflation is slightly persistent nominal shocks will be reflected in short-term expected inflation, causing changes in the yield curve. If inflation persistence is near-unity instead, expected inflation will change by an equal amount for all time horizons and the yield curve will not be affected.

From the empirical viewpoint, there is an incredible number of papers which have found the spread useful in predicting real economic activity. The earlier contributions have concentrated on how accurately the term spread forecasts output growth by using linear models of the type

$$\Delta y_t = \alpha + \beta S_t + \phi(L)\Delta y_{t-1} + \epsilon_t \tag{3.1}$$

Where Δy_t is the growth rate of an indicator of real economic activity (for example GDP or industrial production), S_t indicates the term spread, $\phi(L)$ is a lag polynomial and ϵ_t is an error term. Regressions of the type of equation (3.1) have been estimated extensively on US data, showing relevant predictive ability until the mid-80s, but less so for more recent time spans[9]. There are several other studies investigating the same issue for countries like Canada, Japan, Denmark, UK and Germany, while to our knowledge only Brunetti and Torricelli (2009) have specifically focused on Italian data[10].

The diminished performance of the term spread forecasts of output growth in recent years seems not to be questionable. For this reason, concerns have been raised over the fact that the predictive performance of the term spread may show substantial regime-switches, and that linear regressions based on the yield spread may suffer from parameter instability. Nonlinear models have already been used widely to study the relationship between growth and the slope of the yield curve.

Venetis, Paya, and Peel (2003) use nonlinear smooth transition models that can accommodate regime-type nonlinear behavior and time-varying parameters to examine the predictive power and stability of the term spreadoutput growth relationship. Using data for the United States, United Kingdom, and Canada, they find that the relationship between output growth and the term spread is stronger when past values of the term spread do not exceed a positive threshold value. Duarte, Venetis, and Paya (2005) use both linear regression and nonlinear models to examine the predictive accuracy of the term spread output growth relationship among euro-area countries. The authors find that linear indicator and nonlinear threshold indicator models predict output growth well at four-quarter horizons and that the term spread is a useful indicator of future output growth and recessions in the euro area. Korenok et al. (2009) find relevant asymmetries in the growth-spread relationship at the sectoral level for the US by using a Markov switching model. Audrino and Medeiros (2011) use a Tree Smooth Transition model for US short-term interest rates to discover that leading indicators for inflation and real activity are the most relevant predictors in characterizing their

[9] See for example Laurent (1989), Harvey (1988), and Estrella and Hardouvelis (1991)

[10] Other papers in this literature with a 'non US focus' are, for example, Stock and Watson (2003), Venetis, Paya, and Peel (2003), Nakaota (2005), and Benati and Goodhart (2008)

regime switching structure. Notably, Giacomini and Rossi (2006) find that a sort of 'breakdown of predictability' seems to have occurred in the data since the mid-80s, and it may be argued that this is due to the increased stability of output and other macroeconomic variables. It remains to be seen then how incorporating data for the recession that began in 2007 affects the performance of forecasting models that use the term spread to predict economic activity and whether the additional information sheds light on alternative explanations for the forecasting relationship. This paper has the ambition also to be a first attempt in addressing these issues.

3.3 Methodology

As noted by Brunetti and Torricelli (2009), the relationship between future economic activity and the slope of the yield curve may well be asymmetric- it should change depending on past values of the spread being positive or negative- and it may also display regime-switching behavior- the informational content of the spread should be different in booms or recessions. These two main characteristics suggest that a non-linear approach to modelling the growth-spread relationship may be appropriate. We reconnect to Brunetti and Torricelli's work and we also estimate a nonlinear model for the growth-spread relationship, encompassing regime switches and possible structural breaks. However, since our focus is on the effect of increased volatility in output growth and interest rates on the predictive ability of this relationship, our dataset also incorporates recent data for the turbulent years 2008-2011, which are used to evaluate the the out-of-sample performance of the different models. Hereafter we briefly describe the estimation methodology we have used and the set-up for the forecasting exercise.

3.3.1 Estimating the STAR model

Following previous contributions in this literature we will be using the following two-regime STAR specification:

$$\begin{aligned}\Delta y_t^k =& (\alpha_1 + \phi_1'(L)\Delta y_{t-1}^k + \beta_1'(L)S_{t-1})+ \\ &+(\alpha_2 + \phi_2'(L)\Delta y_{t-1}^k + \beta_2'(L)S_{t-1}) \cdot G(S_{t-d};\gamma,c') + \epsilon_t\end{aligned} \tag{3.2}$$

Where symbols have the usual meaning and the lag polynomial for S_t contains only significant lags for the term spread selected among $d = (1, 3, 6, 12, 18)$. The transition function $G(S_{t-d}; \gamma, c')$ is a continuous and at least twice-differentiable function, normally belonging to the exponential or logistic family. More specifically, in our case tests suggest to use the logistic function, that is to say:

$$G(S_{t-d}; \gamma, c) = (1 + \exp\{-\gamma(S_{t-d} - c)\})^{-1} \quad \gamma > 0$$

The 'smoothness' or the speed of the transition is determined by the size of the γ parameter, which is strictly positive, while the regimes are determined by the transition variable S_{t-d} being smaller or greater than the threshold c.

The modelling cycle for STAR model is a general-to-specific one, involving first a linearity test, and then if linearity is rejected at various conventional significance levels, the choice of the appropriate transition function and transition variable. Hence, first one tests whether ST-AR is more appropriate than the simple linear AR by means of the standard Luukkonen et al. (1988) linearity test. The test is based on a re-parametrization of the STAR model (3.2) by approximating the $G(.)$ function with its third order Taylor approximation around $\gamma = 0$, that is to say:

$$\Delta y_t^k = \alpha_{00} + \sum_i (\beta_{0i} S_{t-d} + \beta_{1i} S_{t-i} S_{t-d} + \beta_{2i} S_{t-i} S_{t-d}^2 + \beta_{3i} S_{t-i} S_{t-d}^3) + \epsilon_t \qquad (3.3)$$

Afterwards, a LM test for the null hypothesis $H_0 : \beta_{1i} = \beta_{2i} = \beta_{3,i} = 0$ can be computed. Through a similar procedure it is also possible to obtain an estimate of the appropriate lag d for the transition variable S_{t-d}, and of the most convenient form of the transition function.

The parameters of the STAR model are normally estimated by a nonlinear optimization routine, like Nonlinear Least Squares (NLS). However, the literature is often silent on three main orders of difficulties that arise:

1. STAR models have a complex structure resulting in a conditional loglikelihood function with multiple optima, and NLS may not always reach the global optimum. To this end, the choice of appropriate starting-values for the optimization routine becomes crucial.[11]

[11] See for example Chan and McAleer (2002),

2. The significance of the nonlinear parameters' estimates is not easily interpretable, and this is particularly true for the significance of the smoothness parameter γ. This is due to the fact that the standard deviation of the smoothness parameter tends to grow with the size of the parameter itself.[12]

3. STAR models present also forecasting difficulties related to the fact that *(a)* while for linear models it is possible to find a recursive relationship between forecasts at subsequent horizons, it is not possible for nonlinear models, and *(b)* the danger of over-fitting the model is always incumbent[13].

We argue that researchers have not put enough attention to these three aspects. The novelty in this work that complements the one of Brunetti and Torricelli is that we take extra care in addressing issues (1) and (3).
In order to avoid getting stuck in one of the local optima, we use Maximum Likelihood estimation and not NLS. In order to find 'good' starting values a grid search is employed as follows: for each value of γ and c the residual sum of squares is computed and the values that correspond to the minimum of that sum are taken as starting values. It should also be noted that in order to make γ scale-free, it is divided by the sample standard deviation of the chosen transition variable. Finally, the unknown parameters are estimated by using the Marquardt algorithm to maximize the conditional maximum likelihood function. A general-to-specific approach is adopted to select the significant spreads: all lagged spreads (1, 3, 6, 12, 18 months) are initially included, then the nonsignificant ones are sequentially eliminated and the nonlinear models re-estimated until parsimonious final specifications are found.[14]

[12]More precisely, Teräsvirta (1994) notes that when smoothness parameter is large and at the same time the threshold parameters are sufficiently close to zero, a negative definite Hessian matrix is difficult to obtain for mere numerical reasons, even when convergence is achieved. That is the reason behind the apparent low significance of the estimate of the smoothness parameter, which should then be evaluated with diagnostics different from the standard ones.

[13]On these two aspect see respectively Teräsvirta (2006) and Bekiros (2009).

[14]As the initial estimates for γ are always very high, indicating that only a few observations are actually near the threshold c, they are replaced with a ceiling value of 100 and the models are re-estimated. This procedure is in line with Venetis et al. (2003) and could in principle lead to inconsistent estimates; however, provided that γ is sufficiently large, the bias should be practically negligible.

3.3.2 The Forecasting Exercise

Forecasting with Smooth Transition models is still an open topic in the literature. As mentioned earlier, producing one-step ahead predictions from a STAR model is something manageable, but multi-step forecasts are much more difficult as the multivariate distribution of the model error will enter the nonlinear data generating process. It is for this reason that Brunetti and Torricelli did not produce forecasts with the STAR model they estimated, and used a simpler logit model to investigate whether the term-spread can also help predicting recession probabilities.

Here, we address the issue by putting up a sort of 'horse race' between the STAR and other forecasting models: the objective of the experiment is to assess their ability to forecast the Italian GDP growth rate from March 2005 on, for different forecast horizons.

Clearly, the first of the competitors is the estimated growth-spread LSTAR model. We produce h-step ahead forecasts from model (3.2) by means of numerical simulation as in Lundbergh and Teräsvirta (2002). To illustrate the method, let us take for simplicity the two regime logistic STAR model:

$$y_t = (\phi_{0,1} + \phi_{1,1} y_t)(1 - G(y_t, \gamma, c) + (\phi_{0,2} + \phi_{1,2} y_t) G(y_t, \gamma, c) \tag{3.4}$$

The one step ahead forecast can be naively computed by ignoring the error term as

$$y^F_{t+1|t} = (\phi_{0,1} + \phi_{1,1} y_t)(1 - G(y_t, \gamma, c) + (\phi_{0,2} + \phi_{1,2} y_t) G(y_t, \gamma, c) \tag{3.5}$$

Afterwards, $y^F_{t+1|t}$ can be employed to generate h-step ahead predictions using numerical simulation to approximate the multidimensional integral involved in the recursions. For example, the two-step ahead prediction is computed as

$$\begin{aligned} y^F_{t+2|t} \cong& (\phi_{0,1} + \phi_{1,1} y^F_{t+1|t}) + \\ &+ \frac{1}{M} \sum_{m=1}^{M} [(\phi_{0,2} + \phi_{1,2}(y^F_{t+1|t} + \epsilon^m_{t+1})] \cdot G[(y^F_{t+1|t} + \epsilon^m_{t+1}), \gamma, c)] \end{aligned} \tag{3.6}$$

where each of the m values of ϵ^m_{t+1} is either drawn independently from the error distribution of (3.2) -*Monte Carlo Forecasts*- or it is drawn independently from the set of residuals from the estimated model without replacement -*Bootstrap Forecasts*. Extensions to the STAR case containing also exogenous variables, like in our case, for model (3.2), are straightforward but notationally complex. It is important to notice that often nonlinear forecasts produce forecasts that are deemed unrealistic simply for computational reasons. As a consequence, we also implement a so called ‘insanity filter’, substituting those forecasts which produce insanely high forecast errors (with respect to the observed changes in the actual time series) with the last observed actual value.[15]

The second type of forecasts we want to evaluate are produced from an equivalent state-space representation of (3.2) with time varying parameters. In this respect, we exploit Granger’s (2008) intuition that “any non-linear model can be approximated by a time-varying parameter linear model” by providing the following state-space representation of (3.2):

$$\begin{aligned} Dy_t^k &= a(L)Dy_t^k + \boldsymbol{s}_t'\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t \\ \boldsymbol{\beta}_t &= \boldsymbol{T}'\boldsymbol{\beta}_{t-1} + u_t \end{aligned} \tag{3.7}$$

Where, $a(L)$ is a lag polynomial, $\boldsymbol{\beta}$ is a vector of stochastically time-varying parameters, $\boldsymbol{T}$ is a conformable diagonal matrix of coefficients to be estimated and $u_t \sim N(0, \sigma_u^2)$, where σ_u^2 is the error variance from the estimated model (3.2). System (3.7) is estimated via Kalman Filter, once maximum likelihood estimates of parameters in $\boldsymbol{A}$ and $\boldsymbol{T}$ are obtained through prediction error decomposition (Harvey, 1981). Afterwards, forecasts can be easily produced again via Kalman Filter.[16]

Both Simulation Based and State-space system based forecasts are evaluated with respect to the forecasts produced through the benchmark autoregressive model

$$\begin{aligned} \Delta y_t^k =& \alpha_1 + \phi_1(L)\Delta y_{t-1}^k + \beta_{1,1}S_{t-1} + \beta_{1,3}S_{t-3} + \\ &+ \beta_{1,6}S_{t-6} + \beta_{1,12}S_{t-12} + \beta_{1,18}S_{t-18} + \epsilon_t \end{aligned} \tag{3.8}$$

As we already mentioned, model (3.8) has a poor in-sample performance but it is still useful as a reference point for the out-of-sample forecasting exercise.

[15] Swanson and White (1995, 1997a,b) are those who call the procedure ‘insanity filter’. Afterwards, the method has then been used widely. Some notable examples are Stock and Watson (1998) and Teräsvirta, van Dijk and Medeiros (2005) and more recently Koch and Teräsvirta (2011) . In this work, the implementation of the filter goes as follows: if the h-step ahead predicted change exceeds the maximum h-step change observed during the estimation period, it is replaced by the last observed change.

[16] This method is particularly interesting, since STAR models like (3.4) can always be rewritten as a linear model with stochastically time varying coefficients $\{\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2 \cdot G(y_{t-d}, \gamma, c)\}$ as noted by Teräsvirta et al. (2005).

As a final step, we use *Thick modelling* to produce an alternative series of forecasts and compare it with the previously obtained ones. *Thick modelling* postulates that using the nonlinear model within a combination of forecasts from other (possibly nonlinear) models may result in a better forecasting performance. The minimization of some kind of loss function gives the researcher the possibility to obtain appropriate weights through which combining the forecasts[17]. The reasons for the success of simple combination schemes are still poorly understood by the literature, however it is often the case in empirical applications that pooled forecasts obtained by appropriate combination of several models outperform single model based forecasts. On top of that, Hansen (2007) suggests that forecast combination schemes are one of the possible solution to the so called 'Winner's curse' problem, involving a strong connection between the in-sample and the out-of-sample fit of a model and hence a low out-of-sample performance of the best in sample fitting model. Moreover, Hendry and Clements (2004) show that forecast pooling can prove successful when forecasting models are differentially misspecified, and especially when time series are subject to structural shifts: these are characteristics that are both likely to occur in our dataset. Combined forecasts are produced according to the following simple scheme

$$\Delta y_{t+h}^{k,F} = \sum_i \omega_i \Delta y_{t+h}^{k,F_i} \tag{3.9}$$

where ω_i are appropriately chosen weights, with i indicating that they refer respectively to models (3.8), (3.2) and (3.7). While the weighting scheme could in principle be very sophisticated, involving that the weights are inversely correlated with the forecasting performance of each model, for our dataset the simplest schemes are the ones that proved more successful. As a consequence, the ω_i weights are either estimated in a simple OLS fashion, or the are set to be time varying and to follow a first order autoregressive process. Hence, hereafter we will indicate with $\Delta y_{t+h}^{k,CLIN}$ and $\Delta y_{t+h}^{k,CTVP}$ the forecasts obtained by pooling the results of the other models, with respectively

[17] The seminal paper by Granger and Jeon (2004) may be useful to give a synthetic definition of this approach: "Thick modelling consists of using many alternative specifications of similar quality, using each to produce the output required for the purpose of the modelling exercise, such as a set of forecasts, policy scenarios, elasticity estimates, or tests of some hypothesis, and then combine or synthesize the results." (Granger and Jeon, 2004, p.325)

linear and time varying weights.

The forecasting exercise involves using different forecast metrics to evaluate the relative performance of each model. The basic measure we will be using is the Root Mean Squared Forecast Error (RMSFE):

$$RMSFE = \sqrt{\frac{1}{H} \cdot \sum_{h=1}^{H} (\Delta \widehat{y}_{t+h}^{k} - \Delta y_{t+h}^{k})}$$

where H indicates the forecast horizon, and hence also the number of steps ahead the prediction is carried for, $\Delta \widehat{y}_{t+h}^{k}$ is the forecast obtained for time $t+h$ with each of the analyzed models, while Δy_{t+h}^{k} is the actual value of observation $t+h$. Since the RMSFE *per sé* cannot be used when comparing the performance of more than one model, we will also employ the measure of Relative Efficiency (RE) proposed by Mincer and Zarnowitz (1969), given by the ratio between the RMSFE of the model i under examination and the one of the respective AR benchmark.

$$RE = \frac{RMSFE^{i}}{RMSFE^{AR}}$$

Whenever the relative efficiency measure is greater or equal than the value of 1, it is considered as a sign that the forecast accuracy of the benchmark model exceeds the one of the other model. In addition to that, we will employ also the commonly used Diebold-Mariano test (Diebold and Mariano, 1995) test in order to assess whether the differences emerging in the forecast performance of the different models are significant from the statistical point of view. In particular, the test evaluates the significance of the null hypothesis of equal predictive accuracy of the selected model and its benchmark, based on the statistic $DM = \frac{\overline{D}}{\sqrt{\sigma_{D}^{2}}}$, where D is the large sample mean of the loss differential between the chosen forecast and the benchmark, approximately distributed with mean $\overline{D}$ and variance σ_{D}^{2}. The simplicity of this tests lies in the fact that in large samples DM is distributed as a standard normal and hence t-statistic critical values can be used to evaluate the test.

3.4 The Dataset and the Chosen Time Sample

Our whole dataset is composed by monthly data from January 1993 to December 2011 containing the logarithm of the Italian industrial production index (y), and the interest rates on government bonds with respectively ten year (i_t^{10yr}) and three months maturity (i_t^{3mts}). All the series have been retrieved from Datastream and they are adjusted for seasonality. A standard measure of the term spread is then obtained as $S_t = i_t^{10yr} - i_t^{3mts}$. Just like Brunetti and Torricelli (2009), we construct the annualized growth rate of real economic activity over the next k months as $\Delta y_t^k = (1200/k)(y_{t+k} - y_t)$. The specifications we focus on are respectively with k equal $3, 6, 12$ or 24 months ahead. To emphasize our focus on forecasting we split the sample in two: data from January 1993 to February 2005 are used to produce coefficient estimates, and data from March 2005 to December 2011 are kept for the out-of-sample forecasting exercise. The choice of the estimation sample is motivated by the fact that, first, there is evidence for structural instability in the data around 1992, the period of the Italian Lira devaluation, and this was already documented by Brunetti and Torricelli (2009). Furthermore, we will show that there might have been a structural shift also in late 2006 until 2008, due to the financial crisis.

Indeed, having to deal with possible nonlinearities in time series data, implies that one needs to be particularly cautious to avoid spurious nonlinear relationships. For example, Yoon (2011) shows that nonlinear estimates of STAR type of models would yield significant and plausible estimation results when they are applied to the data series following models with a structural break. Moreover, not only structural instability may be disguised as spurious nonlinearity, but also other forms of non stationarity like unit roots. To avoid these potential dangers, here we perform a battery of unit root and structural change tests.

Table 3.4.1 (upper panel) shows that the Augmented Dickey-Fuller test, the Phillips-Perron test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) all indicate that our annualized growth indicators do not display evidence of unit root, while the interest rates and the spread series do. The evidence is a little bit more mixed for the industrial production series, for which the KPSS test only indicates the possibility of first order non stationarity. The Johansen test for cointegration between the two interest rate series (lower panel) further confirms that indeed there may exist a long run cointegrating relationship among the two, reason that justifies using the spread as a stationary series in our specification. However, the literature acknowledges that sometimes unit

Table 3.4.1: Stationarity and Cointegration Tests

test type	Dy_3	Dy_6	Dy_12	Dy_24	Ind. prod.	Spread	i_3mts	i_10yr
ADF test	0.0007	0.0012	0.0205	0.0205	0.0378	**0.1934**	**0.1453**	**0.1474**
KPSS test (KPSS critical value 5%= 0.463)	0.2783	0.2927	0.3412	0.3412	**0.5732**	**1.5061**	**1.1753**	**1.1221**
PP test	0.0007	0.0012	0.0205	0.0205	0.0378	**0.1934**	**0.1453**	**0.1474**

	Cointegration Test between interest rates			
	Trace Test		Maximum Eigenvalue Test	
number of cointegrating vectors	0	p=0.0106	0	p=0.0063
	1*	p=0.8586	1*	p=0.8586

Note: The upper part of the table includes p-values for the ADF and PP tests, while for the KPSS test the test-statistic and the critical values are reported. We signal in bold each time the tests indicate the possible presence of a unit root. The lower part of the table reports the p-values for Johansen cointegration test for the long run and short run interest rates series. We indicate with an asterisk the number of cointegrating vector that the test indicates at the 5 % confidence level.

root tests may have low power, especially in case of structural instability. The time span considered in this work is admittedly at risk for that, since it comprises at least one exceptional event: the global financial crisis which has started in 2007. We further investigate this issue by performing various structural stability tests.

We start by a linear relationship of the following type being estimated for each forecast horizon k:

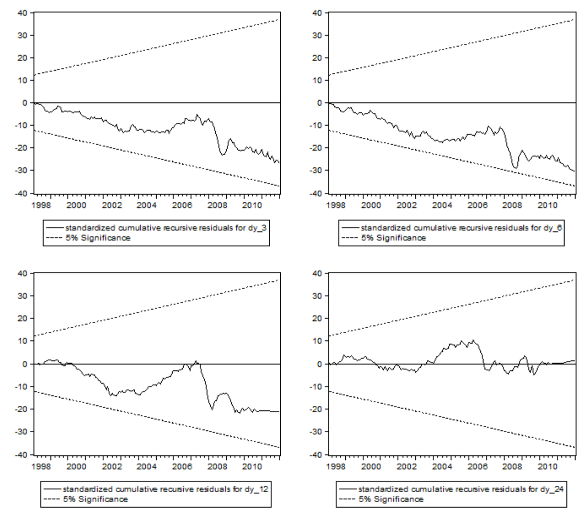
$$\begin{aligned} \Delta y_t^k = &\alpha_1 + \phi_1(L)\Delta y_{t-1}^k + \beta_{1,1}S_{t-1} + \beta_{1,3}S_{t-3} + \\ &+ \beta_{1,6}S_{t-6} + \beta_{1,12}S_{t-12} + \beta_{1,18}S_{t-18} + \epsilon_t \end{aligned} \tag{3.10}$$

This is actually the kind of relationship that has customarily been estimated in many of the studies we previously cited: there is a basic autoregressive component, involving in our case up to 12 lags of the dependent variable, and various lags of the term spread as an exogenous component. Here we include only five relevant lagged spreads, and more precisely we include S_{t-i} with $i = 1, 3, 6, 12, 18$ months. Equation (3.10) is estimated recursively and the residuals of the equation are then tested for structural instability by means of the CUSUM test (Brown, Durbin, and Evans, 1975), both in the version involving standardized recursive residuals and in the one on squared standardized residuals.

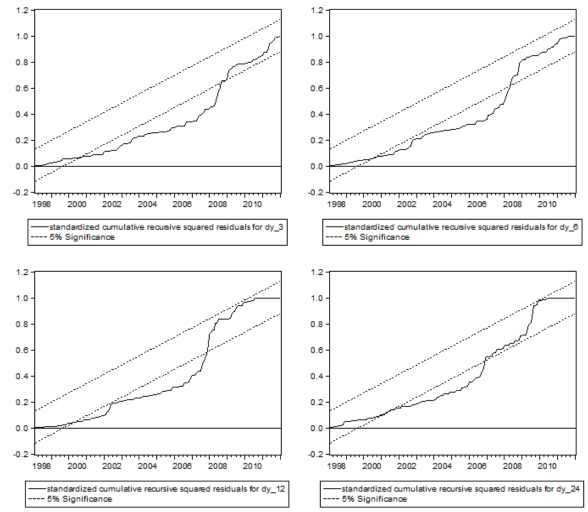
Figure (3.2) shows the results of these tests. In the basic CUSUM test (figure (3.2), panel (A)), the cumulative sum of recursive residuals is plotted together with the 5 percent critical lines. As we can see, in this test the cumulative sum never goes outside the area between the two critical lines, hence suggesting no parameter instability. Nevertheless, signs of variance instability are evident from figure (3.2) panel (B). The CUSUM test based on squared residuals exactly aims at capturing potential instability in the process generating the variance of residuals. Indeed, in figure (3.2) panel (B) the standardized cumulative squared residuals exceed the 5 % critical values in many parts of the sample. In particular, for the 3, 6 and 12 months ahead growth rates equations the variance instability seems to characterize the whole time span from 2000 to 2008, while for the one year ahead equation, it seems to end in late 2006. This is no surprise, since appropriate tests already outlined the presence of heteroskedasticity in the estimated residuals of (3.10).

A somewhat more precise indication of parameter instability can be given by performing a sequence of Chow Forecast tests on the n-step ahead forecast errors. This is particularly informative since it does not require the specification of a precise date for a breakpoint to occur, as in the standard Chow test, while it computes all n step ahead forecast errors, starting with the smallest possible sample size for estimating the forecasting equation and then adding one observation at a time. The plot from this test shows the recursive residuals at the top, and significant probabilities, based on the F-statistic, in the lower portion of the graph (left vertical axis). The lower part of each panel displays the probability values for those sample points where the hypothesis of parameter constancy would be rejected at respectively the 5, 10, or 15 % levels. The points with p-values less than 0.05 correspond to those points where the recursive residuals go outside the two standard error bounds. Figure (3.2) panel (C) really clarifies the issue: for all the forecast horizons the test indicates that between 2007 and 2008 a structural break is very likely. All the above considerations translate in a precise indication for our forecasting exercise: the 2007-2008 period is very volatile and it is not wise to include it in the estimation sample, at least at this stage.
The choice of the estimation window is crucial when dealing with possible parameter instability. On the one hand, it may be safer to include the most recent data

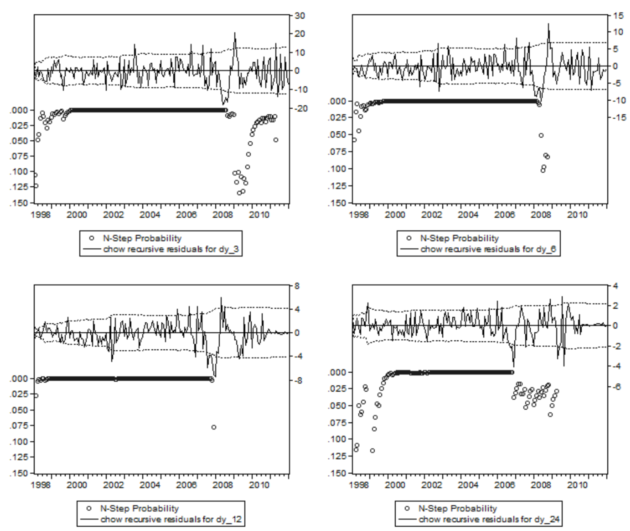
Figure 3.2: Structural Instability Tests on Model (3.10)



(A)



(B)



(C)

Note: Panel (A): CUSUM Tests on the Recursive Standardized Residuals

Panel (B): CUSUM Tests on the Recursive Standardized Squared Residuals

Panel (C): Recursive Chow Breakpoint Tests.

The right axis is for recursive residuals, and the the left axis is for probability values related to the hypothesis of parameter constancy being rejected at the 5, 10, or 15 percent levels.

Table 3.4.2: Linearity Tests for Each Forecast Horizon

K=3					K=6				
d=1	d=3	d=6	d=12	d=18	d=1	d=3	d=6	d=12	d=18
0.9204	0.7207	<u>0.2898</u>	0.4990	0.7751	0.2569	0.2887	<u>0.1268</u>	0.9243	0.3835
K=12					**K=24**				
d=1	d=3	d=6	d=12	d=18	d=1	d=3	d=6	d=12	d=18
0.1801	<u>0.0979</u>	0.7334	0.5055	0.5493	0.5661	<u>0.1787</u>	0.7345	0.2176	0.5243

Note: The table contains p-values for the Luukkonen et al. (1988) linearity test for model (3.10). For each forecast horizon k we underline the lag d chosen by the test, yielding the lowest p-value.

available after the last structural break. On the other hand, this needs not to be the most rewarding choice, as Pesaran and Timmerman (2007) show that there can be a trade-off between reducing the forecast error bias and increasing the forecast error variance. Furthermore, theoretical and simulation results confirm that the forecasting performance can typically be improved if some pre-break information is also included. As a consequence, we chose to use data from the beginning of the sample until 2005 to obtain parameters estimates, and we leave the remaining part of the sample for the forecasting exercise, being aware that the estimated model may perform quite poorly during the financial crisis period.

In sum, the results obtained from the estimation of a linear model like (3.10) are unsatisfactory for all the k specifications. In the majority of cases, the estimated coefficients do not carry the expected signs and all the lagged term spread result not significant. Furthermore, estimated residuals do not show the standard normality and homoskedasticty properties they should possess[18]. Following Brunetti and Torricelli (2009), we argue that a nonlinear specification could solve at least some of these problems and we proceed with a nonlinear approach. To this purpose, we implement the Luukkonen et al. (1988) linearity test for all the forecast horizons, as reported in table 3.4.2. Overall, the null hypothesis of linearity of model (3.10) is uniformly rejected by the test. While the table only reports p-values connected to the choice of the appropriate lag length d, the choice of the transition function indicated by the test (not

[18]Results for the OLS estimation are not reported but they are available on request.

Table 3.5.1: Estimated Coefficients of model (3.2) for $k = 3, 6, 12, 24$

K=3			K=6			K=12			K=24		
d=6			d=6			d=3			d=3		
coefficient	estimate	p-value	coefficient	estimate	p-value	coefficient	estimate	p-value	coefficient	estimate	p-value
α_1	0.3942	0.3285	α_1	0.5177	0.0320	α_1	-0.1348	0.7986	α_1	2.0887	0.0000
						$\beta_{1,1}$	2.7934	0.0000	$\beta_{1,1}$	-0.6832	0.0000
$\beta_{1,12}$	-0.6181	0.0895	$\beta_{1,12}$	-0.8553	0.0038	$\beta_{1,12}$	-3.6473	0.0000	$\beta_{1,12}$	-0.9297	0.0000
						$\beta_{1,18}$	1.6517	0.0048	$\beta_{1,18}$	-0.5319	0.0000
α_2	-5.8335	0.0043	α_2	-2.4770	0.0428	α_2	0.1936	0.7641	α_2	-3.4992	0.0000
			$\beta_{2,3}$	3.5355	0.0000	$\beta_{2,1}$	-3.8277	0.0000	$\beta_{2,1}$	0.9435	0.0000
			$\beta_{2,6}$	-3.1429	0.0000						
$\beta_{2,12}$	5.7594	0.0000	$\beta_{2,12}$	2.6406	0.0000	$\beta_{2,12}$	3.8640	0.0000	$\beta_{2,12}$	1.6104	0.0000
$\beta_{2,18}$	-3.1682	0.0206	$\beta_{2,18}$	-1.4225	0.0015	$\beta_{2,18}$	-2.0162	0.0012	$\beta_{2,18}$	0.5120	0.0076
c	1.7086	0.0000	c	1.3436	0.0000	c	0.4755	0.0000	c	1.3930	0.0000
AIC	11.2210		AIC	8.9537		AIC	7.4614		AIC	4.1030	
BIC	11.8845		BIC	9.5377		BIC	7.9126		BIC	4.6338	
logl	-519.2164		logl	-412.2550		logl	-344.8782		logl	-178.9936	
JB	4.9971	0.0822	JB	0.5460	0.7611	JB	0.4463	0.8000	JB	4.0096	0.1347
ARCH(1)	0.4199	0.5170	ARCH(1)	0.8550	0.3551	ARCH(1)	0.0913	0.7626	ARCH(1)	0.0572	0.8110

Note: The model has been estimated by means of maximum likelihood and a Marquardt optimization algorithm, in the time span from January 1997 to February 2005. For reasons of space the table does not report the estimated coefficients of the AR component, although overall significant.

reported) is in line with Brunetti and Torricelli's results, and it always indicates the logistic one as the most appropriate one. We can now proceed with the nonlinear estimation approach.

3.5 Results

Model (3.2) is estimated by means of maximum likelihood, and a general-to-specific approach is used to select the relevant lags of the spread variable. Although in Brunetti and Torricelli only some of parameters estimates are reported, it is already possible to see from table (3.5.1) that our results are quite different from theirs.

The nonlinear specification yields a very good overall fit, with very significant coefficients which also display the expected sign. Moreover, the estimated residuals do not display the heteroskedasticity and non normality features they showed previously in the linear specifications, as it is possible to see from the Jarque Bera and ARCH

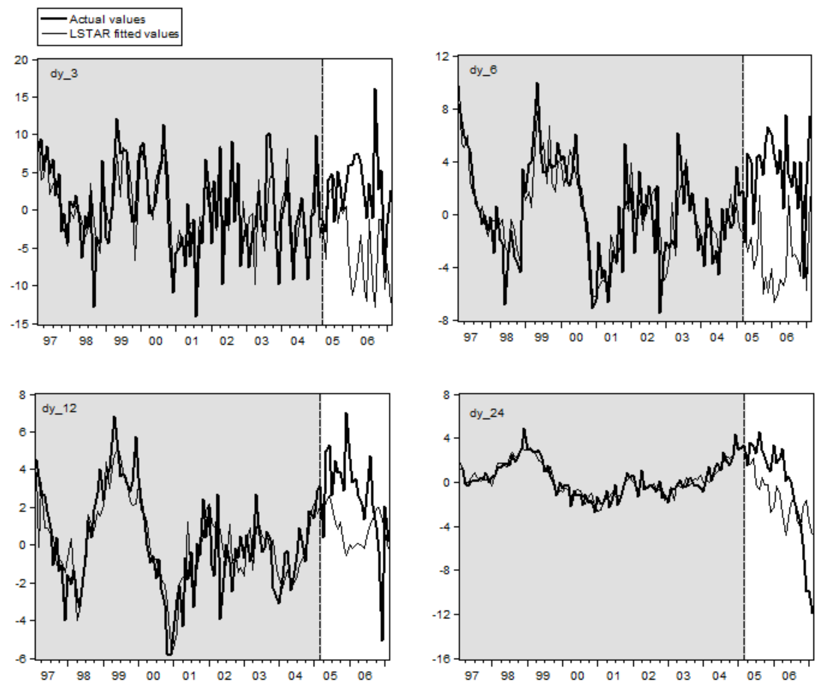
statistics at the bottom of the table. Across all forecast horizon, lagged term spreads always result very significant, although in each specification the chosen lags differ. In particular, for the GDP growth rate 3 and 6 months ahead, the term spread of 12 and 18 months before seem to have the property of leading indicator. For longer forecast horizons, 12 and 24 months ahead, also the term spread lagged one month seems to play a role. Notably, the β_1 coefficients in the first regime have always opposite signs than the β_2, indicating that our presumption about the effect of the spread being reversed across regimes was grounded.
Table (3.5.1) is the starting point for the forecasting exercise. We made use of a general-to-specific approach allowing us to select the most parsimonious specification of the LSTAR model for each forecast horizon. At a first stage, we will examine the in-sample and out-of sample performance of the LSTAR model only, thereby complementing the work of Brunetti and Torricelli. As a second step, we will devote our attention to the proper out-of-sample forecast 'horse race', and compare the forecast accuracy of models (3.2), (3.7) and (3.10). The comparison will be carried on over two evaluation periods across specifications: the first one comprehends only 4 months, spanning from March 2005 to June 2006, while the second one is longer, involving 24 months, from March 2005 to March 2007. In each scenario of the exercise, the forecasts obtained from the three models under examination are paralleled with those obtained by pooling the forecasts according to the 'Thick Modelling approach'.

3.5.1 Forecasting with the STAR model

Figure (3.3) provides a synthetic view of the in-sample and out-of-sample performance of model (3.2) across specifications[19]. As we can see, for most specifications the in-sample fit of the LSTAR model is quite satisfactory, as in the shaded area of the graphs the fitted values proceed very close to the actual ones. However, there are some marked differences across specifications when looking at the predicted values from march 2005 to July 2007. Clearly, the forecasts depart quite substantially from the actual values,

[19]Here we show some pictures relating to the forecast performance of Monte Carlo Forecasts from the LSTAR model, the standard method in the literature. The next section takes into consideration also the Bootstrap forecasts and the other type of forecasting models

Figure 3.3: LSTAR Model Performance: Actual values versus Fitted

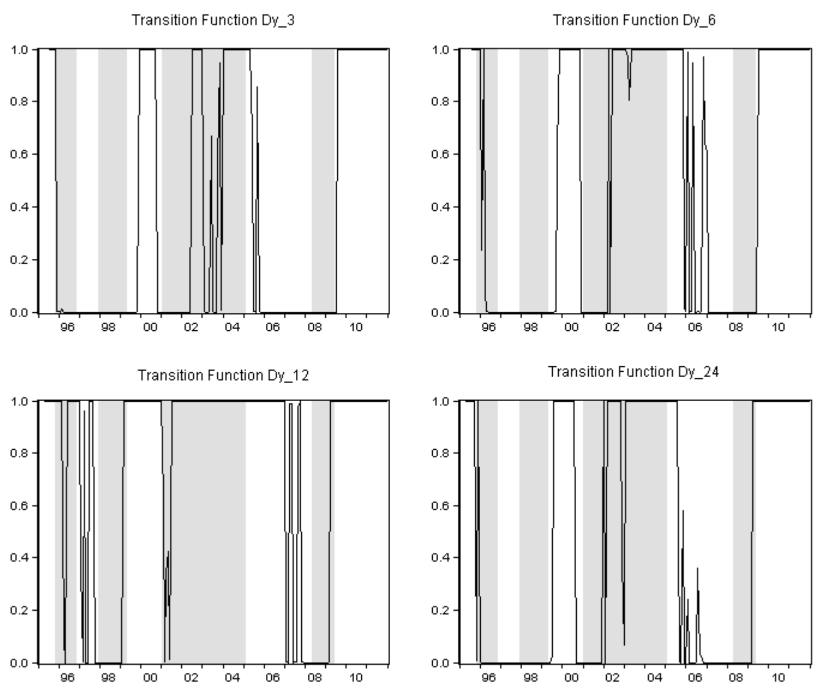


Note: Thick lines indicate the actual values of respectively, series Δy_t^3, Δy_t^6, Δy_t^{12}, Δy_t^{24}. Thin lines indicated fitted values of model (3.2) across specification: in the first part of the sample (1997-2005) -shaded area- the models are fitted with parameter estimates from table (3.5.1); in the second part (2005-2007), fitted values are calculated as Monte Carlo Forecasts from the estimated LSTAR models.

especially in the specifications for Δy_t^6 and Δy_t^{12}. Moreover, the Monte Carlo forecasts from model (3.2) not only depart from the actual values in a substantial manner, but also they seem not to be able to predict the signs of the GDP growth rates, as they move in exactly the opposite direction of the actual series. This effect is very pronounced across all specifications, with a negative correlation between actual and forecast values being around -0.8.

It is also interesting to look at the fitted transition function from the estimated LSTAR models, as this could give us an intuition on how the spread dynamics evolve across the business cycle. Figure (3.4) shows the estimated transition functions for each specification along with the OECD business cycle chronology, and recessions are indicated as grey areas. The entire time span can be divided in three main sub-periods: the first one from 1997 to 2000, the second capturing the years until late 2006, and the third one characterized by the turbulences of the financial crisis, from 2007 to 2011.

Figure 3.4: LSTAR Model Performance: Transition Functions vs Time



Note: Transition functions are calculated with estimated parameters from table (3.5.1). Shaded areas are recessions according to the OECD chronology.

In the first sub-period the two short-term specifications (upper panels) and the one for 24 months ahead (lower right panel) display a similar pattern, with the transition function remaining at zero when the recessions kick in, and peaking towards 1 when recessions are about to end. Only the one year ahead growth rate specification (lower left panel) seems to behave differently, as the transition function oscillates substantially between zero and one also during the recession. Also during the pre-crisis and crisis periods, all the specifications except the one with Δy_t^{12}, show a similar pattern. The transition functions oscillate during the first 2000-2004 recession and fall to zero in 2006, anticipating the next recession. The switching mechanism triggered by the lagged term spread we have identified seems a good candidate to predict business cycle dynamics: will this be enough for the LSTAR model to be the winner of our 'horse race'?

3.5.2 Forecasting with multiple models: a 'Horse Race'

Giving an intuitive albeit detailed description of the results of our forecasts 'horse race' is not an easy task but in this section we try our best. For each of the four specifications, we implemented four main scenarios where nonlinear simulation based forecasts are compared with their linear and time varying rivals. More specifically the scenarios differ among each other because of the following four characteristics that may vary:

1. *Treatment 1*: Forecasts are generated as Monte Carlo Forecasts, where the distribution of the errors is the same as the one of model (3.2).

2. *Treatment 2*: Forecasts are generated as Bootstrap Forecasts, by drawing errors independently and without replacement from the set of residuals from the estimated LSTAR model.

3. *Naive start*: The one-step ahead of the forecasts is calculated in a *naive* manner, as suggested by Lundbergh and Teräsvirta (2002). This implies ignoring the influence of the first error term and setting $y^F_{t+1|t} = (\phi_{0,1} + \phi_{1,1}y_t)(1 - G(y_t, \gamma, c)) + (\phi_{0,2} + \phi_{1,2}y_t)G(y_t, \gamma, c)$.

4. *Simulated start*: The one-step ahead of the forecasts is calculated by ignoring the structure of the forecasting model and simply by setting $y^F_{t+1|t} = y_t + \epsilon_t$ where ϵ_t

is drawn from the same assumed distribution which is be used to compute the h-step ahead forecasts.

Moreover, as we already explained, for each of these four scenarios not only the ‘pure’ forecasts from models (3.2), (3.7) and (3.10) will be compared, but also combined forecasts from these models will be examined. In particular two additional treatments will be added:

1. *Linearly combined (CL) forecasts*: Forecasts from models (3.2), (3.7) and (3.10) are pooled according to equation (3.9). The ω_i weights are obtained via OLS estimation of $\Delta y^k_{t+1} = \omega_1 \Delta y^{k,AR}_{t+1} + \omega_2 \Delta y^{k,STAR}_{t+1} + \omega_3 \Delta y^{k,TVP}_{t+1}$ with the restriction that $\sum_i \omega_i = 1$.

2. *Time-varying weights combined (CTVP) forecasts*: Forecasts from models (3.2), (3.7) and (3.10) are pooled according to equation (3.9). The ω_i weights are obtained by applying the Kalman Filter to the state space system

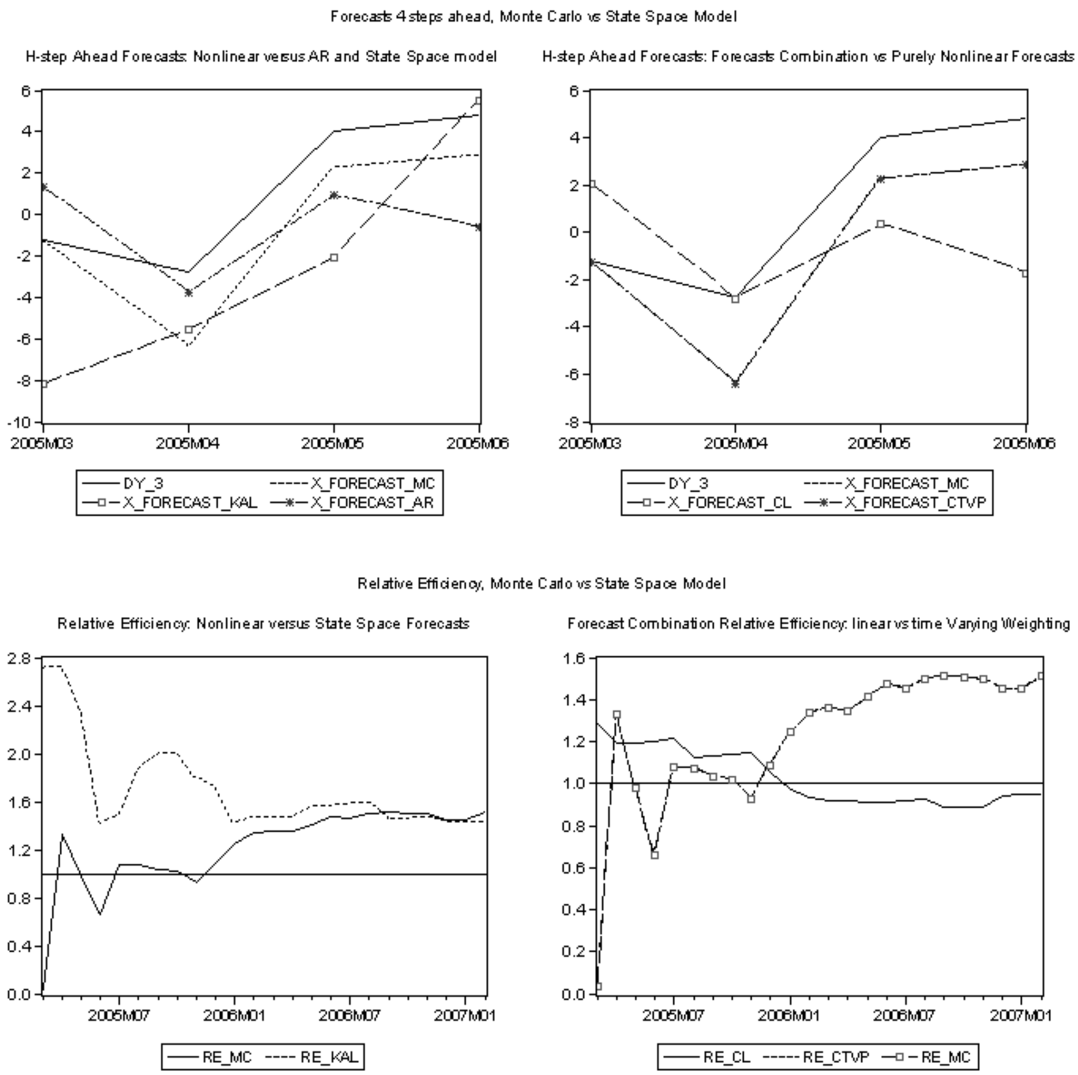
$$\Delta y^k_{t+1} = \omega_{1i,t} \Delta y^{k,AR}_{t+1} + \omega_{2i,t} \Delta y^{k,STAR}_{t+1} + \omega_{3i,t} \Delta y^{k,TVP}_{t+1}$$

$$\omega_{i,t} = \omega_{i,t-1} + u_t; \text{ i=AR, STAR, TVP}$$

In order to avoid the disorientation of the reader, we will only present results from selected treatments of the forecasting ‘horse race’ by making use of graphical evidence. Detailed tables with the results of each forecasting exercise for all the specifications can be found in the appendix.

Let us first examine the results with the Δy^3_t specification, visualized in figure (3.5) and complemented by the detailed table (3.7.1) in the appendix. The Monte Carlo and the Bootstrap forecasts scenarios look very similar, both with or without the *naive start* option, hence we selected the picture regarding the Monte Carlo forecasts treatment as most representative. The relative efficiency of the forecasts from the LSTAR model is below one only at very short-term forecast horizons (1 and 3 months ahead), while it increases from 6 steps ahead on. The lower left panel of figure (3.5) gives us also an idea of the comparative performance on model (3.2) and its equivalent state-space specification (3.7) over a longer forecast horizon, namely 2 years ahead. In this case, when

Figure 3.5: Δy_t^3 Specification: Four-Step-Ahead Forecasts

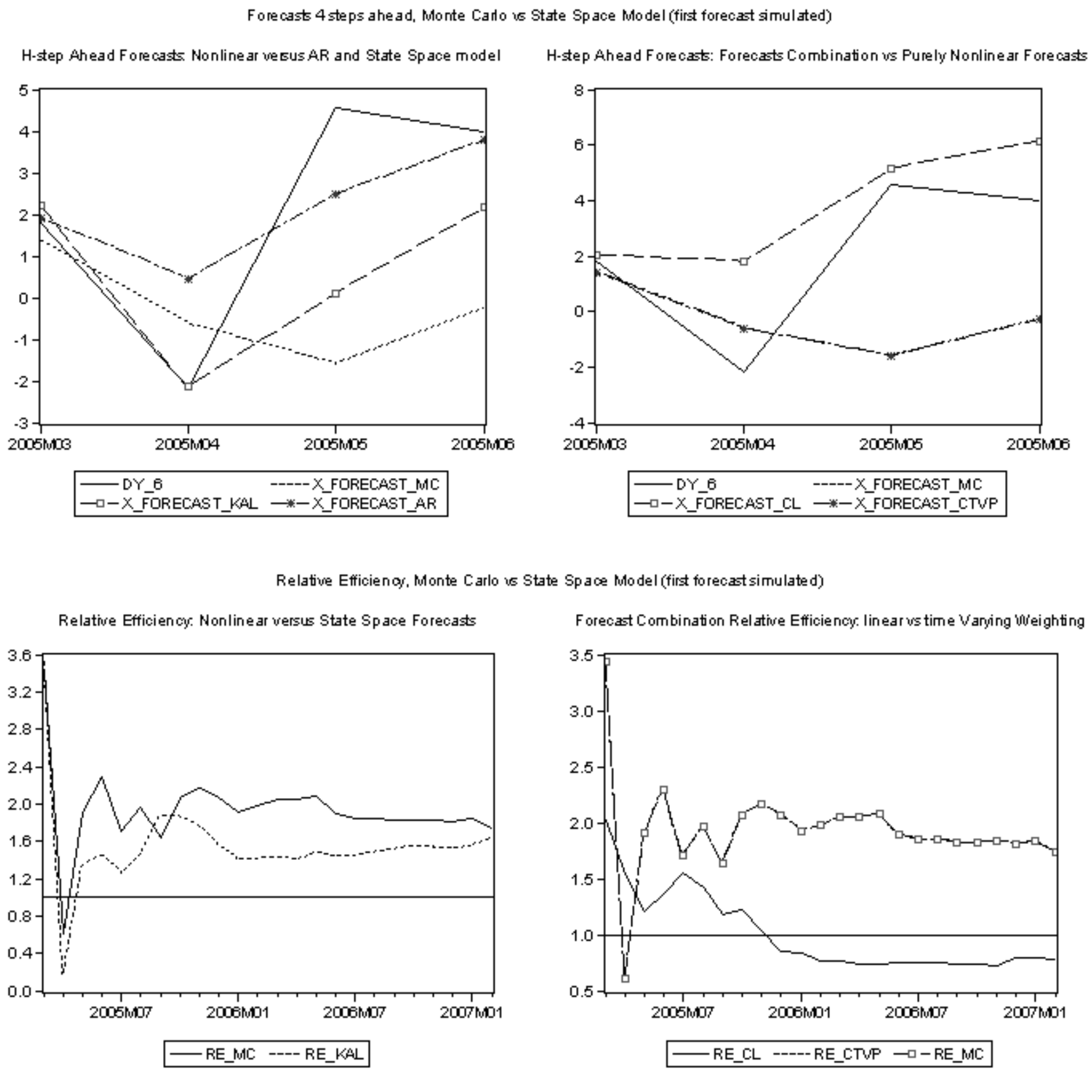


Note: (Upper left panel) Four-Step-Ahead Monte Carlo Forecasts (dashed line) are compared with Forecasts from the AR model (–⋆– line) and from the Time Varying Parameter model (–□– line). Actual values of Δy_t^3 are indicated with a continuos line.
(Upper right panel) our-Step-Ahead Monte Carlo Forecasts (dashed line) are compared with Combined forecast from the whole set of models examined in the left panel with either a linear weighting scheme (–□– line) or a time-varying weighting scheme (–⋆– line).
(Upper lower panels) Relative Efficiency measures for the forecasts evaluated, where the benchmark model is the AR (3.8) model. Whenever the relative efficiency measure is greater or equal than one it is an indication that the benchmark model yields a lower forecast error than the one under examination. Relative efficiency for the pure Monte Carlo forecasts (continuous line) is compared with the one from the time varying model (dashed line, left panel), with the one from combined forecasts with a linear weighting scheme (–□– line, right panel) and with a time varying weighting scheme (dashed line, right panel).

evaluating the predictions of the two models until 12 months ahead the LSTAR model is definitely better. However, the performance of the two becomes almost equal and much worse than the AR specification when looking at farther away horizons. Interestingly, it seems that using forecast combination schemes improves the performance of the forecasts: as it shows from the right side panels of the figure, forecasts from model (3.9) with time varying weights beats the Monte Carlo forecasts at short-term horizons, while using a linear weighting improves the predictive performance over longer time spans. Remarkably, for most forecasts the Diebold-Mariano test also indicate that the differences among the performance of all of these models are statistically significant.
The results with specification Δy_t^6 are particularly interesting because they suggest that the choice on how to compute the one-step ahead forecast may not be innocuous. Indeed, table (3.7.2) indicates that in the *simulated start* case the Diebold-Mariano test always rejects the null of equal predictive accuracy between each couple of models, while this is not the case for the *naive start* scenario. For this reason, we selected exactly the Monte Carlo forecast with *simulated start* experiment to be displayed in figure (3.6). In this experiment the Monte Carlo and Bootstrap forecasts from the LSTAR model perform almost equally bad at all the horizons. Instead, the time varying equivalent of the LSTAR model shows its potential by achieving a lower relative efficiency than the standard LSTAR at each forecast horizon, although both of them being are beat by the AR model in this scenario. By looking at the lower let panel of figure (3.6) however, we see that this small advantage in terms of relative efficiency of model (3.7) vanishes completely when forecasting from 12 months ahed on: at longer horizons it seems that the LSTAR model beats its time varying rival. Nevertheless, the pure nonlinear forecasts are outperformed by the combined ones also this time: the right lower panel of the figure clearly shows that the linearly combined forecasts yield a much lower relative efficiency than the Monte Carlo forecast and the time varying weighted forecasts, and this is true for forecast horizons from 3 to 24 months ahead.

The specifications with Δy_t^{12} and Δy_t^{24} are a little bit more cumbersome to interpret since in these cases the Diebold-Mariano test often fails to reject the null of equal predictive accuracy. We report these test statistics in tables (3.7.3) and (3.7.4) but we

Figure 3.6: Δy_t^6 Specification: Four-Step-Ahead Forecasts

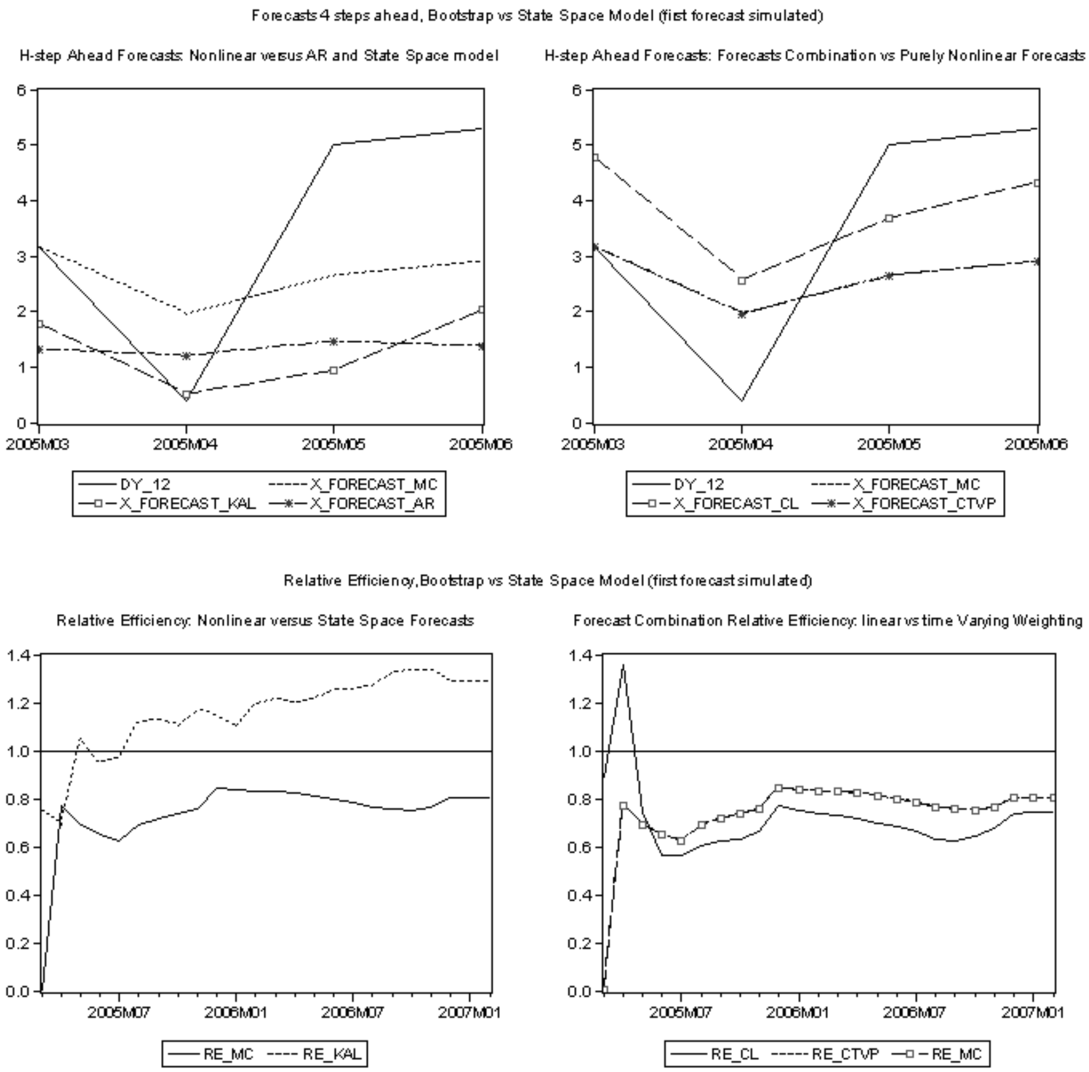


Note: (Upper left panel) Four-Step-Ahead Monte Carlo Forecasts with the first forecast generated with the *naive start* option (dashed line) are compared with Forecasts from the AR model (–⋆– line) and from the Time Varying Parameter model (–□– line). Actual values of Δy_t^6 are indicated with a continuos line.

(Upper right panel) four-Step-Ahead Monte Carlo Forecasts (dashed line) are compared with Combined forecast from the whole set of models examined in the left panel with either a linear weighting scheme (–□– line) or a time-varying weighting scheme (–⋆– line).

(Upper lower panels) Relative Efficiency measures for the forecasts evaluated, where the benchmark model is the AR (3.8) model. Whenever the relative efficiency measure is greater or equal than one it is an indication that the benchmark model yields a lower forecast error than the one under examination. Relative efficiency for the pure Monte Carlo forecasts-option .4- (continuous line) is compared with the one from the time varying model (dashed line, left panel), with the one from combined forecasts with a linear weighting scheme (–□– line, right panel) and with a time varying weighting scheme (dashed line, right panel).

Figure 3.7: Δy_t^{12} Specification: Four-Step-Ahead Forecasts

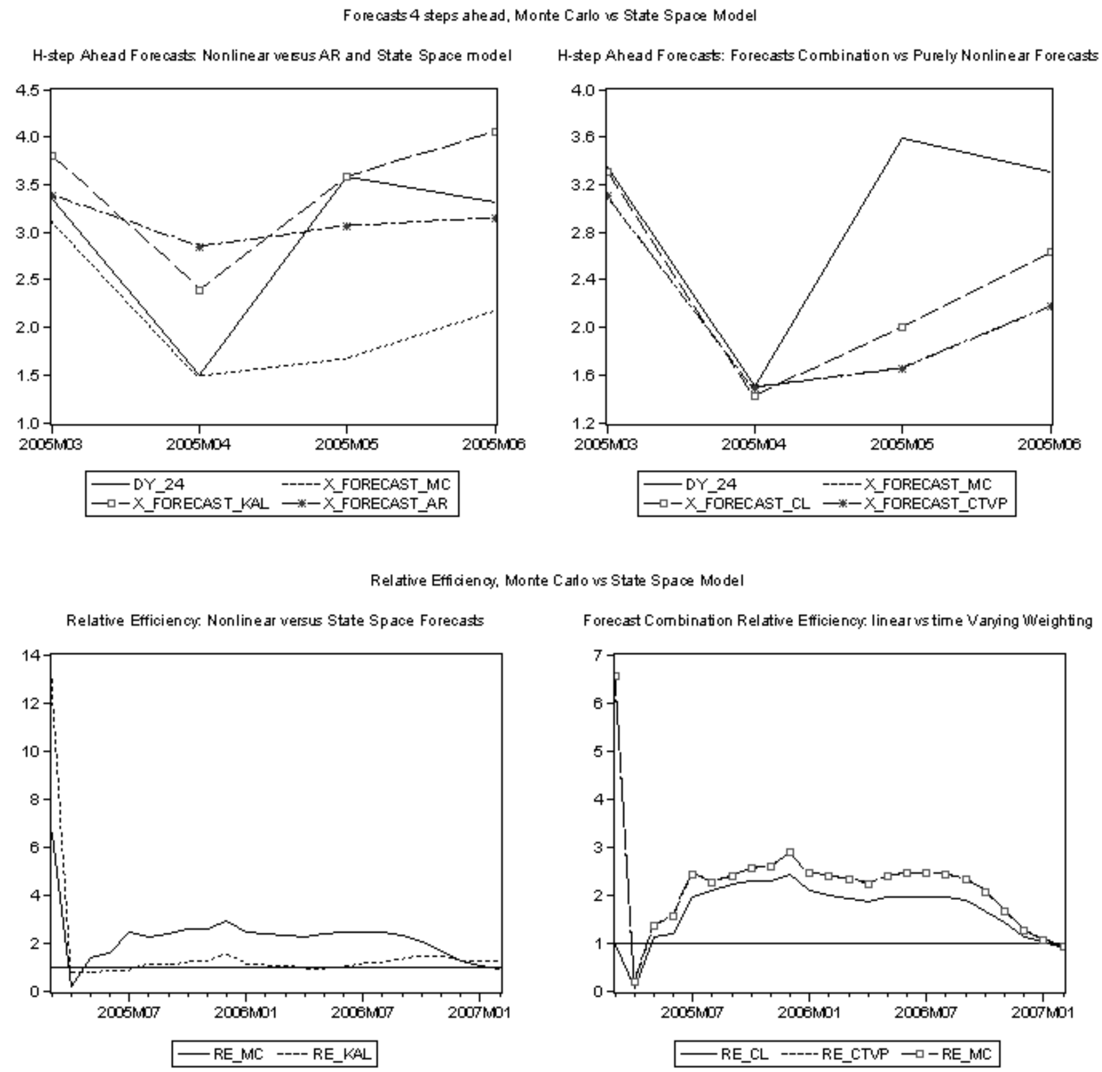


Note: (Upper left panel) Four-Step-Ahead Bootstrap Forecasts with the first forecast generated with option .4 (dashed line) are compared with Forecasts from the AR model (–⋆– line) and from the Time Varying Parameter model (–□– line). Actual values of Δy_t^6 are indicated with a continuos line.

(Upper right panel) four-Step-Ahead Bootstrap Forecasts (dashed line) are compared with Combined forecast from the whole set of models examined in the left panel with either a linear weighting scheme (–□– line) or a time-varying weighting scheme (–⋆– line).

(Upper lower panels) Relative Efficiency measures for the forecasts evaluated, where the benchmark model is the AR (3.8) model. Whenever the relative efficiency measure is greater or equal than one it is an indication that the benchmark model yields a lower forecast error than the one under examination. Relative efficiency for the pure Bootstrap forecasts-option .4- (continuous line) is compared with the one from the time varying model (dashed line, left panel), with the one from combined forecasts with a linear weighting scheme (–□– line, right panel) and with a time varying weighting scheme (dashed line, right panel).

Figure 3.8: Δy_t^{24} Specification: Four-Step-Ahead Forecasts



Note: (Upper left panel) Four-Step-Ahead Monte Carlo Forecasts (dashed line) are compared with Forecasts from the AR model (–⋆– line) and from the Time Varying Parameter model (–□– line). Actual values of Δy_t^{24} are indicated with a continuos line.
(Upper right panel) our-Step-Ahead Monte Carlo Forecasts (dashed line) are compared with Combined forecast from the whole set of models examined in the left panel with either a linear weighting scheme (–□– line) or a time-varying weighting scheme (–⋆– line).
(Upper lower panels) Relative Efficiency measures for the forecasts evaluated, where the benchmark model is the AR (3.8) model. Whenever the relative efficiency measure is greater or equal than one it is an indication that the benchmark model yields a lower forecast error than the one under examination. Relative efficiency for the pure Monte Carlo forecasts (continuous line) is compared with the one from the time varying model (dashed line, left panel), with the one from combined forecasts with a linear weighting scheme (–□– line, right panel) and with a time varying weighting scheme (dashed line, right panel).

also remain a bit skeptical on their value in detecting differences in forecast accuracy. A number of recent papers have documented problems with procedures that test whether differences in out-of- sample forecast error are statistically significant[20]. On top of that, for these two specifications the insanity filter is much more operational than in the other two cases, suggesting that maybe the out-of-sample performance of the LSTAR model would benefit of some improvements also in the model specification. We leave these aspects aside for the moment, and we try to see what insights can we gain from these two final parts of the forecasting exercise. When we try to forecast the GDP growth rate during the next 12 months, figure (3.7), the spread-growth LSTAR model we have estimated does not seem to be very helpful. Its relative efficiency, with Bootstrap forecasts and the*simulated start* method, is below one only for 3-steps ahead predictions. For all the other horizons, the equivalent time-varying state space model (3.7) seems to be a little bit better, while the simple AR model still seems to be most successful predictor. When looking at farther horizons however, the pure nonlinear model turns again to outperform its rivals. From the combined forecasts side, the linear weighting scheme seems more convincing than all the competitors, although this time the pure LSTAR and the linearly combined forecast follow a very similar pattern. Forecasting GDP growth rate one year in advance looks even tougher. Figure (3.8) and the complementary table (3.7.4) in the appendix clearly show that the one with Δy_t^{24} as the dependent variable is the least satisfactory of our specifications. This time the LSTAR model yields a relative efficiency lower than one only at the 24 step ahead prediction, being surpassed by model (3.7) in all the other cases. However, the linear model still seems to yield lower forecast errors than these two. Among the pooled forecasts, the linearly weighted one seem to outperform all the rivals, especially when longer time spans are considered (lower right panel of the the figure).

3.6 Concluding Remarks

This paper investigated the impact of the Italian interest rates spread on the growth rate of real economic activity. The purpose of our empirical analysis is to compare the

[20]see for example Clark (1996)

forecasting power of a nonlinear LSTAR model, widely used in the previous literature, with the one other types of models like the autoregressive one, the time varying parameters model, and the pooled forecasts model.

We find that the lagged term-spread always provides information about the future path of economic activity that is not captured by a purely autoregressive model. Moreover, modelling the impact of the spread on economic activity in a nonlinear fashion seems to improve both the in-sample and out-of-sample fit of the model. We take as a benchmark the work of Brunetti and Torricelli (2009), where a LSTAR growth-spread model is estimated but not used for forecasting, and we produce Monte Carlo and Bootstrap forecasts with our LSTAR specification up to 24 months ahead. Hence, the study differs from the previous ones on the Italian case both for choice to use the LSTAR model for predictive purposes, which is quite demanding, and for the fact that its out-of-sample performance is closely evaluated in comparison with other rival models within a 'horse race' experiment.

Our results suggest that for the Italian case it is worth using a smooth transition specification also for forecasting purposes, at least for two main order of reasons. First, once we took care that the model is robustly estimated, using the LSTAR model solves (at least partially) the heteroskedasticity problem often plaguing the residuals of linear models customarily used in this literature. Secondly, the LSTAR specification gives us the possibility of obtaining a precise indication of the regime-switches characterizing the growth-spread relationship, thereby producing also a robust leading indicator for business cycle oscillations.

These strengths of the LSTAR model, however, are not sufficient to make it the winner of the forecasting competition. In this respect, it is not really possible to award a 'one size fits it all' winner. Certainly, for one to three months forecasts, the nonlinear model seems to be more reliable than the pure autoregressive model and the equivalent time varying model. However, when the forecast horizon is expanded to 6-24 months ahead, the performance of the LSTAR model is seriously challenged by the one of its equivalent state-space representation. This is particularly interesting since in some cases it could be quite handy to estimate a time-varying model instead of a STAR: it allows the researcher to exploit the properties of the Kalman filter and to possibly reduce the bias-variance trade-off of the model. Finally, our results strongly suggest that using forecast combination schemes substantially improves on the out-of-sample performance, and this is line with other studies in the forecasting literature. In particular, for our dataset it seems that the linearly combined forecast beat the other competitors both at short and long-run horizons.

We have left at least one important topic for future research: how did the financial crisis impact on the forecasting performance of the growth-spread relationship? we admittedly postpone this question since we have found evidence of a structural break exactly during the financial crisis and including this data in our sample would have endangered our results. A fruitful avenue in this direction could be averaging forecasts over different estimation windows (Pesaran and Timmerman,2007)[21].

In sum, we may conclude that the STAR approach offers an interesting alternative in modelling and forecasting the relationship between the term spread and economic growth in Italy, since this phenomenon presents some evident regime-switching characteristics. Nonetheless, more work is certainly needed to understand whether more parsimonious alternatives to the STAR model, like time-varying models, or simply different approaches to nonlinearity, like Markov-switching or neural networks models, can add something to our understanding of the predictive ability of the yield spread for future economic growth.

3.7 Appendix: Detailed Tables for the Forecasting Exercise

[21] In particular, Schrimpf and Wang (2010) show that in the presence of multiple structural breaks using this method may help to reduce the forecast errors bias, and sometimes also the forecast error variance.

Table 3.7.1: Out-of-sample Predictive power of Δy_t^3

Naive start

Naive start	Monte Carlo Forecasts				Forecasts Combination			
	RMSFE		Relative Efficiency		RMSFE		Relative Efficiency	
h	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP
1	0.0854	6.9513	2.7187	0.0334	3.2769	0.0854	1.2816	0.0334
3	2.3439	5.5829	2.3408	0.9828	2.8381	2.3439	1.1900	0.9828
6	4.1231	7.1780	1.8693	1.0737	4.3159	4.1231	1.1239	1.0737
12	8.4926	9.3078	1.4693	1.3406	5.9070	8.4926	0.9325	1.3406
24	10.5764	10.0008	1.4286	1.5108	6.6533	10.5764	0.9504	1.5108

Bootsrap Forecasts				Forecasts Combination			
RMSFE		Relative Efficiency		RMSFE		Relative Efficiency	
STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP
0.0854	6.9513	2.7187	0.0334	3.2739	0.0854	1.2804	0.0334
2.3262	5.5829	2.3408	0.9753	2.8350	2.3262	1.1887	0.9753
4.0950	7.1780	1.8693	1.0664	4.3216	4.0950	1.1254	1.0664
8.4888	9.3078	1.4693	1.3400	5.9032	8.4888	0.9319	1.3400
10.5719	10.0008	1.4286	1.5102	6.6482	10.5719	0.9497	1.5102

Diebold Mariano Test

		TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR
DM test type 1	-2.5762	-3.8872	0.5927	-2.5762
DM test type 2	-2.3882	-3.9217	-0.2389	-2.3882

Diebold Mariano Test

	STAR vs AR	TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR
DM test type 1	-2.5649	-3.8872	0.5961	-2.5649
DM test type 2	-2.3734	-3.9217	-0.2415	-2.3734

Simulated start

Simulated start	Monte Carlo Forecasts				Forecasts Combination			
	RMSFE		Relative Efficiency		RMSFE		Relative Efficiency	
h	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP
1	0.0251	6.9513	2.7187	0.0098	3.2609	0.0251	1.2753	0.0098
3	2.3308	5.5829	2.3408	0.9773	2.8331	2.3308	1.1879	0.9773
6	4.1200	7.1780	1.8693	1.0729	4.3114	4.1200	1.1227	1.0729
12	8.4843	9.3078	1.4693	1.3393	5.9086	8.4843	0.9327	1.3393
24	10.5760	10.0008	1.4286	1.5107	6.6537	10.5760	0.9505	1.5107

Bootstrap Forecasts				Forecasts Combination			
RMSFE		Relative Efficiency		RMSFE		Relative Efficiency	
STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP
3.7043	6.9513	2.7187	1.4488	4.3720	3.7043	1.7099	1.4488
3.7682	5.5829	2.3408	1.5799	3.2180	3.7682	1.3492	1.5799
4.6488	7.1780	1.8693	1.2106	4.6143	4.6488	1.2016	1.2106
9.1171	9.3078	1.4693	1.4392	5.8278	9.1171	0.9200	1.4392
10.6679	10.0008	1.4286	1.5239	6.6312	10.6679	0.9472	1.5239

Diebold Mariano Test

		TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR
DM test type 1	-2.5730	-3.8872	0.5919	-2.5730
DM test type 2	-2.3787	-3.9217	-0.2371	-2.3787

Diebold Mariano Test

	STAR vs AR	TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR
DM test type 1	-2.8081	-3.8872	0.6416	-2.8081
DM test type 2	-2.8939	-3.9217	-0.3539	-2.8939

Note: For each specification, the the table reports RMSFEs and a measure of Relative Efficiency, i.e. the RMSFE ratio with respect to the benchmark AR model. The forecast horizons considered are respectively 1, 3, 6, 12 and 24 months ahead, starting from March 2005. For the Relative Efficiency measure a number less than one indicates that the model under evaluation in the numerator has superior forecasting ability than the benchmark. The Diebold-Mariano test statistics for the null hypothesis of equal predictive ability of each model compared with the AR is also reported.

Table 3.7.2: Out-of-sample Predictive power of Δy_t^6

Naive start	Monte Carlo Forecasts				Forecasts Combination				Bootsrap Forecasts				Forecasts Combination			
	RMSFE		Relative Efficiency		RMSFE		Relative Efficiency		RMSFE		Relative Efficiency		RMSFE		Relative Efficiency	
h	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP
1	3.2201	0.4333	3.5165	26.1358	1.1534	3.2201	9.3613	26.1358	3.2201	0.4333	3.5165	26.1358	1.1534	3.2201	9.3613	26.1358
3	4.6780	2.5981	1.3479	2.4271	2.6150	4.6780	1.3567	2.4271	4.6780	2.5981	1.3479	2.4271	2.6150	4.6780	1.3567	2.4271
6	5.2437	3.0768	1.4692	2.5040	3.0684	5.2437	1.4652	2.5040	5.2437	3.0768	1.4692	2.5040	3.0684	5.2437	1.4652	2.5040
12	7.8987	5.5131	1.4213	2.0364	3.1806	7.8987	0.8200	2.0364	7.8987	5.5131	1.4213	2.0364	3.1806	7.8987	0.8200	2.0364
24	7.6870	7.1457	1.6394	1.7636	3.5437	7.6870	0.8130	1.7636	7.6870	7.1457	1.6394	1.7636	3.5437	7.6870	0.8130	1.7636

Diebold Mariano Test

	STAR vs AR	TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR
DM test type 1	-4.1894	-3.3214	1.8690	-4.1894
DM test type 2	-5.5756	-4.5566	2.1269	-5.5756

Diebold Mariano Test

	STAR vs AR	TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR
DM test type 1	-4.1894	-3.3214	1.8690	-4.1894
DM test type 2	-5.5756	-4.5566	2.1269	-5.5756

Simulated start	Monte Carlo Forecasts				Forecasts Combination				Bootsrap Forecasts				Forecasts Combination			
	RMSFE		Relative Efficiency		RMSFE		Relative Efficiency		RMSFE		Relative Efficiency		RMSFE		Relative Efficiency	
h	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP
1	0.4231	0.4333	3.5165	3.4342	0.2486	0.4231	2.0176	3.4342	0.1681	0.4333	3.5165	1.3646	0.0075	0.1681	0.0612	1.3646
3	3.6722	2.5981	1.3479	1.9052	2.3299	3.6722	1.2088	1.9052	3.5514	2.5981	1.3479	1.8425	2.2668	3.5514	1.1760	1.8425
6	4.1042	3.0768	1.4692	1.9598	2.9635	4.1042	1.4151	1.9598	3.9066	3.0768	1.4692	1.8655	2.9239	3.9066	1.3962	1.8655
12	7.6655	5.5131	1.4213	1.9763	2.9820	7.6655	0.7688	1.9763	7.6393	5.5131	1.4213	1.9695	2.9344	7.6393	0.7565	1.9695
24	7.5589	7.1457	1.6394	1.7342	3.4241	7.5589	0.7856	1.7342	7.5459	7.1457	1.6394	1.7312	3.3989	7.5459	0.7798	1.7312

Diebold Mariano Test

	STAR vs AR	TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR
DM test type 1	-3.4675	-3.3214	2.0406	-3.4675
DM test type 2	-4.5308	-4.5566	2.4130	-4.5308

Diebold Mariano Test

	STAR vs AR	TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR
DM test type 1	-3.3597	-3.3214	2.0793	-3.3597
DM test type 2	-4.3583	-4.5566	2.5070	-4.3583

Note: For each specification, the the table reports RMSFEs and a measure of Relative Efficiency, i.e. the RMSFE ratio with respect to the benchmark AR model. The forecast horizons considered are respectively 1, 3, 6, 12 and 24 months ahead, starting from March 2005. For the Relative Efficiency measure a number less than one indicates that the model under evaluation in the numerator has superior forecasting ability than the benchmark. The Diebold-Mariano test statistics for the null hypothesis of equal predictive ability of each model compared with the AR is also reported.

Table 3.7.3: Out-of-sample Predictive power of Δy_t^{12}

Naive start	Monte Carlo Forecasts				Forecasts Combination				Bootsrap Forecasts				Forecasts Combination			
	RMSFE		Relative Efficiency		RMSFE		Relative Efficiency		RMSFE		Relative Efficiency		RMSFE		Relative Efficiency	
h	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP
1	0.2351	0.4638	12.9471	6.5623	0.0341	0.2351	0.9529	6.5623	0.2351	0.4638	12.9471	6.5623	0.0331	0.2351	0.9247	6.5623
3	1.1243	0.5738	0.6907	1.3534	0.9162	1.1243	1.1029	1.3534	1.1261	0.5738	0.6907	1.3555	0.9167	1.1261	1.1034	1.3555
6	2.6364	1.2576	1.0754	2.2545	2.4459	2.6364	2.0916	2.2545	2.6375	1.2576	1.0754	2.2554	2.4454	2.6375	2.0911	2.2554
12	3.0110	1.3229	1.0426	2.3730	2.5253	3.0110	1.9903	2.3730	3.0131	1.3229	1.0426	2.3747	2.5241	3.0131	1.9893	2.3747
24	3.9054	4.9692	1.1803	0.9277	3.7868	3.9054	0.8995	0.9277	3.9071	4.9692	1.1803	0.9281	3.7871	3.9071	0.8996	0.9281
	Diebold Mariano Test								Diebold Mariano Test							
		STAR vs AR	TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR					STAR vs AR	TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR			
	DM test type 1	0.3039	-1.7056	0.5497	0.3039				DM test type 1	0.3016	-1.7056	0.5491	0.3016			
	DM test type 2	-0.8492	-1.7968	-0.5689	-0.8492				DM test type 2	-0.8503	-1.7968	-0.5703	-0.8503			
Simulated start	Monte Carlo Forecasts				Forecasts Combination				Bootstrap Forecasts				Forecasts Combination			
	RMSFE		Relative Efficiency		RMSFE		Relative Efficiency		RMSFE		Relative Efficiency		RMSFE		Relative Efficiency	
h	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP
1	0.2602	0.4638	12.9471	7.2631	0.0389	0.2602	1.0868	7.2631	0.1395	0.4638	12.9471	3.8934	0.2688	0.1395	7.5030	3.8934
3	0.8523	0.5738	0.6907	1.0259	0.7399	0.8523	0.8907	1.0259	0.8596	0.5738	0.6907	1.0348	0.8108	0.8596	0.9760	1.0348
6	1.7477	1.2576	1.0754	1.4945	2.3462	1.7477	2.0063	1.4945	1.7299	1.2576	1.0754	1.4793	2.3147	1.7299	1.9794	1.4793
12	2.0121	1.3229	1.0426	1.5858	2.0592	2.0121	1.6229	1.5858	1.9943	1.3229	1.0426	1.5717	2.0359	1.9943	1.6046	1.5717
24	3.3521	4.9692	1.1803	0.7962	2.9434	3.3521	0.6992	0.7962	3.3459	4.9692	1.1803	0.7948	2.9384	3.3459	0.6980	0.7948
	Diebold Mariano Test								Diebold Mariano Test							
		STAR vs AR	TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR					STAR vs AR		TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR		
	DM test type 1	0.7323	-1.7056	0.8065	0.7323				DM test type 1	0.7339		-1.7056	0.8087	0.7339		
	DM test type 2	-0.2667	-1.7968	0.2997	-0.2667				DM test type 2	-0.2508		-1.7968	0.2792	-0.2508		

Note: For each specification, the the table reports RMSFEs and a measure of Relative Efficiency, i.e. the RMSFE ratio with respect to the benchmark AR model. The forecast horizons considered are respectively 1, 3, 6, 12 and 24 months ahead, starting from March 2005. For the Relative Efficiency measure a number less than one indicates that the model under evaluation in the numerator has superior forecasting ability than the benchmark. The Diebold-Mariano test statistics for the null hypothesis of equal predictive ability of each model compared with the AR is also reported.

Table 3.7.4: Out-of-sample Predictive power of Δy_t^{24}

Naive start	Monte Carlo Forecasts				Forecasts Combination				Bootsrap Forecasts				Forecasts Combination			
	RMSFE		Relative Efficiency		RMSFE		Relative Efficiency		RMSFE		Relative Efficiency		RMSFE		Relative Efficiency	
h	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP
1	0.2351	0.4638	12.9471	6.5623	0.0341	0.2351	0.9529	6.5623	0.2351	0.4638	12.9471	6.5623	0.0331	0.2351	0.9247	6.5623
3	1.1243	0.5738	0.6907	1.3534	0.9162	1.1243	1.1029	1.3534	1.1261	0.5738	0.6907	1.3555	0.9167	1.1261	1.1034	1.3555
6	2.6364	1.2576	1.0754	2.2545	2.4459	2.6364	2.0916	2.2545	2.6375	1.2576	1.0754	2.2554	2.4454	2.6375	2.0911	2.2554
12	3.0110	1.3229	1.0426	2.3730	2.5253	3.0110	1.9903	2.3730	3.0131	1.3229	1.0426	2.3747	2.5241	3.0131	1.9893	2.3747
24	3.9054	4.9692	1.1803	0.9277	3.7868	3.9054	0.8995	0.9277	3.9071	4.9692	1.1803	0.9281	3.7871	3.9071	0.8996	0.9281
Diebold Mariano Test									Diebold Mariano Test							
		STAR vs AR	TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR					STAR vs AR	TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR			
	DM test type 1	0.3039	-1.7056	0.5497	0.3039				DM test type 1	0.3016	-1.7056	0.5491	0.3016			
	DM test type 2	-0.8492	-1.7968	-0.5689	-0.8492				DM test type 2	-0.8503	-1.7968	-0.5703	-0.8503			
Simulated start	Monte Carlo Forecasts				Forecasts Combination				Bootstrap Forecasts				Forecasts Combination			
	RMSFE		Relative Efficiency		RMSFE		Relative Efficiency		RMSFE		Relative Efficiency		RMSFE		Relative Efficiency	
h	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP	STAR	TVP	TVP	STAR	CL	CTVP	CL	CTVP
1	0.2602	0.4638	12.9471	7.2631	0.0389	0.2602	1.0868	7.2631	0.1395	0.4638	12.9471	3.8934	0.2688	0.1395	7.5030	3.8934
3	0.8523	0.5738	0.6907	1.0259	0.7399	0.8523	0.8907	1.0259	0.8596	0.5738	0.6907	1.0348	0.8108	0.8596	0.9760	1.0348
6	1.7477	1.2576	1.0754	1.4945	2.3462	1.7477	2.0063	1.4945	1.7299	1.2576	1.0754	1.4793	2.3147	1.7299	1.9794	1.4793
12	2.0121	1.3229	1.0426	1.5858	2.0592	2.0121	1.6229	1.5858	1.9943	1.3229	1.0426	1.5717	2.0359	1.9943	1.6046	1.5717
24	3.3521	4.9692	1.1803	0.7962	2.9434	3.3521	0.6992	0.7962	3.3459	4.9692	1.1803	0.7948	2.9384	3.3459	0.6980	0.7948
Diebold Mariano Test									Diebold Mariano Test							
		STAR vs AR	TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR					STAR vs AR	TVP vs AR	Pooled Forecasts (linear) vs AR	Pooled Forecasts (time varying) vs AR			
	DM test type 1	0.7323	-1.7056	0.8065	0.7323				DM test type 1	0.7339	-1.7056	0.8087	0.7339			
	DM test type 2	-0.2667	-1.7968	0.2997	-0.2667				DM test type 2	-0.2508	-1.7968	0.2792	-0.2508			

Note: For each specification, the the table reports RMSFEs and a measure of Relative Efficiency, i.e. the RMSFE ratio with respect to the benchmark AR model. The forecast horizons considered are respectively 1, 3, 6, 12 and 24 months ahead, starting from March 2005. For the Relative Efficiency measure a number less than one indicates that the model under evaluation in the numerator has superior forecasting ability than the benchmark. The Diebold-Mariano test statistics for the null hypothesis of equal predictive ability of each model compared with the AR is also reported.

Bibliography

[1] G.A. Akerlof, W.T. Dickens, and G.L. Perry. Near-rational wage and price setting and the long-run phillips curve. *Brookings Papers on Economic Activity*, 2000(1):1–44, 2000.

[2] G.A. Akerlof and J.L. Yellen. Can small deviations from rationality make significant differences to economic equilibria? *The American Economic Review*, 75(4):708–720, 1985.

[3] H.M. Anderson. Transaction costs and nonlinear adjustment towards equilibrium in the us treasury bill market. *Oxford Bulletin of Economics and Statistics*, 59(4):465–484, 1997.

[4] D. Andolfatto, S. Hendry, and K. Moran. Are inflation expectations rational? *Journal of Monetary Economics*, 55(2):406–422, 2008.

[5] I.J.M. Arnold and J.J.G. Lemmen. Inflation expectations and inflation uncertainty in the eurozone: evidence from survey data. *Review of World Economics*, 144(2):325–346, 2008.

[6] F. Audrino and M.C. Medeiros. Modeling and forecasting short-term interest rates: The benefits of smooth regimes, macroeconomic variables, and bagging. *Journal of Applied Econometrics*, 26(6):999–1022, 2011.

[7] H. Bakhshi and A. Yates. Are uk inflation expectations rational? Technical report, Bank of England working papers, 1998.

[8] M. Balcilar, R. Gupta, and Z.B. Shah. An in-sample and out-of-sample empirical investigation of the nonlinearity in house prices of south africa. *Economic Modelling*, 28(3):891–899, 2011.

[9] L. Ball. Near-rationality and inflation in two monetary regimes. Technical report, National Bureau of Economic Research, 2000.

[10] R. Batchelor and A.B. Orr. Inflation expectations revisited. *Economica*, 55(219):317–31, 1988.

[11] S.D. Bekiros. A robust algorithm for parameter estimation in smooth transition autoregressive models. *Economics Letters*, 103(1):36–38, 2009.

[12] L. Benati and C. Goodhart. Investigating time-variation in the marginal predictive power of the yield spread. *Journal of Economic Dynamics and Control*, 32(4):1236–1272, 2008.

[13] S. Bereau, A. L. Villavicencio, and V. Mignon. Nonlinear adjustment of the real exchange rate towards its equilibrium value: A panel smooth transition error correction modelling. *Economic Modelling*, 27(1):404 – 416, 2010.

[14] J.M. Berk. Measuring inflation expectations: a survey data approach. *Applied Economics*, 31(11):1467–1480, 1999.

[15] RL Brown, J. Durbin, and JM Evans. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37(2):149–192, 1975.

[16] M. Brunetti and C. Torricelli. Economic activity and recession probabilities: information content and predictive power of the term spread in italy. *Applied Economics*, 41(18):2309–2322, 2009.

[17] M.F. Bryan and G. Venkatu. *The demographics of inflation opinion surveys.* Number Oct 15. Federal Reserve Bank of Cleveland, 2001.

[18] P. Cagan. The monetary dynamics of hyper-inflation. In M. Friedman, editor, *Studies in the Quantity Theory of Money,*. University of Chicago Press, Chicago, 1956.

[19] E. Cannon and G.P. Cipriani. Euro-illusion: A natural experiment. *Journal of Money Credit and Banking*, 38(5):1391, 2006.

[20] J.A. Carlson and M. Parkin. Inflation expectations. *Economica*, 42(166):123–138, 1975.

[21] F. Chan and M. McAleer. Maximum likelihood estimation of star and star-garch models: Theory and monte carlo evidence. *Journal of Applied Econometrics*, 17(5):509 – 534, 2002.

[22] F. Chan and B. Theoharakis. Estimating m-regimes star-garch model using qmle with parameter transformation. *Mathematics and Computers in Simulation*, 81(7):1385–1396, 2011.

[23] K. S. Chan and H. Tong. On the use of the deterministic lyapunov function for the ergodicity of stochastic difference equations. *Advances in applied probability*, 17(3):666 – 678, 1985.

[24] C.R. Chen, P.P. Lung, and F.A. Wang. Stock market mispricing: Inflation illusion or resale option? Technical report, University of Dayton Working Paper, 2006.

[25] T.E. Clark. *Finite-sample properties of tests for forecast equivalence*. Federal reserve bank of Kansas city, 1996.

[26] R. Cooper and A. John. Coordinating coordination failures in keynesian models. *The Quarterly Journal of Economics*, 103(3):441, 1988.

[27] V. Corradi, N.R. Swanson, H. White, and Park) Pennsylvania State University (U. Testing for stationarity-ergodicity and for comovements between nonlinear discrete time markov processes. *Journal of Econometrics*, 96(1):39–74, 2000.

[28] J. Crespo Cuaresma, E. Gnan, and D. Ritzberger-Grünwald. The term structure as a predictor of real activity and inflation in the euro area: a reassessment. *Review of World Economics*, 141(2):318–342, 2005.

[29] R. Curtin. Inflation expectations: theoretical models and empirical tests. In *University of Michigan, paper presented at a conference of National Bank of Poland, Warsaw in*, 2006.

[30] R.B. Davies. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64:247–254, 1977.

[31] F.X. Diebold and R.S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–265, 1995.

[32] J. Doepke, J. Dovern, U. Fritsche, and J. Slacalek. *European inflation expectations dynamics*. Deutsche Bundesbank, 2005.

[33] A. Duarte, I.A. Venetis, and I. Paya. Predicting real growth and the probability of recession in the euro area using the yield spread. *International Journal of Forecasting*, 21(2):261–277, 2005.

[34] S.D. Dutt and D. Ghosh. Inflation expectations: are they rational? an empirical examination. *Applied Economics Letters*, 7(2):103–106, 2000.

[35] R.F. Engle and C.W.J. Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: Journal of the Econometric Society*, 55:251–276, 1987.

[36] A. Escribano and S. Mira. Nonlinear cointegration and nonlinear error-correction models. Technical report, Working Papers Series in Statistics and Econometrics, 96-54, Universidad Carlos III de Madrid, 1996.

[37] EstimaRATS. Transition autoregression models. The RATS Letter 18, No.2 pag.3, July 2005.

[38] A. Estrella. Why does the yield curve predict output and inflation? *The Economic Journal*, 115(505):722–744, 2005.

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

[40] A. Estrella and M.R. Trubin. The yield curve as a leading indicator: some practical issues. *Current Issues in Economics and Finance*, 12(5)(Jul):pp. 1–7, 2006.

[41] E. Fehr and J.R. Tyran. Does money illusion matter? *American Economic Review*, 91(5):1239–1262, 2001.

[42] E. Fehr and J.R. Tyran. Limited rationality and strategic interaction: the impact of the strategic environment on nominal inertia. *Econometrica*, 76(2):353–394, 2008.

[43] I. Fisher. *The Rate of Interest: Its nature, determination and relation to economic phenomena*. The Macmillan Company, 1907.

[44] I. Fisher. *The money illusion*. Kessinger Pub, 2006.

[45] M. Forsells and G. Kenny. The rationality of consumer's inflation expectations: Survey-based evidence for the euro area. Technical report, ECB Working Paper, the European Central Bank, 2002.

[46] P. H. Franses and D. Van Dijk. *Nonlinear time series models in empirical finance*. Cambridge Univ Press, 2000.

[47] J. Fredj and K. Yousra. Threshold cointegration between stock returns: An application of stecm models. Econometrics 0412001, EconWPA, December 2004.

[48] M. Friedman. *A theory of the consumption function*. Princeton University Press, 1957.

[49] John W. Galbraith and Greg Tkacz. 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, 2000.

[50] C. Gerberding. The information content of survey data on expected price developments for monetary policy. Technical report, Discussion Paper Series 1: Economic Studies, 2001.

[51] C. Gerberding. Household versus expert forecasts of inflation: New evidence from european survey data. In P. J. N. Sinclair, editor, *Inflation Expectations*. Routledge, 2009.

[52] R. Giacomini and B. Rossi. How stable is the forecasting performance of the yield curve for output growth? *Oxford Bulletin of Economics and Statistics*, 68:783–795, 2006.

[53] D.G. Goldstein and G. Gigerenzer. Models of ecological rationality: The recognition heuristic. *Psychological review*, 109(1):75, 2002.

[54] C. W. Granger, T. Teräsvirta, and H.M. Anderson. Modeling nonlinearity over the business cycle. In *Business Cycles, Indicators and Forecasting*, NBER Chapters, pages 311–326. National Bureau of Economic Research, Inc, December 1993.

[55] C.W.J. Granger. Non-linear models: Where do we go next-time varying parameter models? *Studies in Nonlinear Dynamics & Econometrics*, 12(3):1, 2008.

[56] C.W.J. Granger and Y. Jeon. Thick modeling. *Economic Modelling*, 21(2):323–343, 2004.

[57] Refet S. Grkaynak, Andrew T. Levin, and Eric T. Swanson. Does inflation targeting anchor long-run inflation expectations? evidence from long-term bond yields in the u.s., u.k., and sweden. Working Paper Series 2006-09, Federal Reserve Bank of San Francisco, 2006.

[58] J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, :357 – 384, 1989.

[59] B.E. Hansen. Least squares model averaging. *Econometrica*, 75(4):1175–1189, 2007.

[60] P.R. Hansen. In sample fit and out of sample fit: Their joint distribution and its implications for model selection. Technical report, working-paper, 2009.

[61] A.C. Harvey. *Time series models*. Oxford: Philip Allan, 1981.

[62] C.R. Harvey. The real term structure and consumption growth. *Journal of Financial Economics*, 22(2):305–333, 1988.

[63] A.A. Haug and P.L. Siklos. The behavior of short-term interest rates: international evidence of non-linear adjustment. *Studies in Nonlinear Dynamics & Econometrics*, 10(4):6, 2006.

[64] D.F. Hendry and M.P. Clements. Pooling of forecasts. *The Econometrics Journal*, 7(1):1–31, 2004.

[65] K. Holden and D.A. Peel. On testing for unbiasedness and efficiency of forecasts. *The Manchester School*, 58(2):120–127, 1990.

[66] P. Howitt. Money illusion. In *Money: the New Palgrave*, page 244. WW Norton & Company, 1989.

[67] Z. Hu. The yield curve and real activity. Technical report, Staff Papers-International Monetary Fund, 1993.

[68] S. Johansen. Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2-3):231–254, 1988.

[69] S. Johansen. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580, 1991.

[70] G. Kapetanios, Y. Shin, and A. Snell. Testing for a unit root in the nonlinear star framework. *Journal of Econometrics*, 112(2):359 – 379, 2003.

[71] A.B. Kock and T. Teräsvirta. Forecasting performance of three automated modelling techniques during the economic crisis 20072009. Technical report, CREATES Research Papers, School of Economics and Management, University of Aarhus, 2011.

[72] O. Korenok, B. Mizrach, and S. Radchenko. A note on demand and supply factors in manufacturing output asymmetries. *Macroeconomic Dynamics*, 13(02):263–277, 2009.

[73] R.D. Laurent. Testing the spread. *Federal Reserve Bank of Chicago Economic Perspectives*, 13(4):22–34, 1989.

[74] S. Leybourne, P. Newbold, and D. Vougas. Unit roots and smooth transitions. *Journal of Time Series Analysis*, 19(1):83 – 97, 1998.

[75] R.E. Lucas and Yrjö Jahnsson Foundation. *Models of business cycles*. Blackwell Oxford, 1987.

[76] S. Lundbergh and T. Teräsvirta. *A Companion to Economic Forecasting*, chapter Forecasting with Smooth Transition Auto regressive Models, page 485. Wiley-Blackwell, 2002.

[77] S. Lundbergh and T. Teräsvirta. *Forecasting with smooth transition autoregressive models*. Wiley Online Library, 2002.

[78] S. Lundbergh, T. Tersvirta, and D. Van Dijk. Time-varying smooth transition autoregressive models. *Journal of Business and Economic Statistics*, 21(1):104 – 121, 2003.

[79] P. Lundborg and H. Sacklen. Low-inflation targeting and long-run unemployment. *The Scandinavian Journal of Economics*, 108(3):397–418, 2006.

[80] R. Luukkonen, P. Saikkonen, and T. Teräsvirta. Testing linearity against smooth transition autoregressive models. *Biometrika*, 75(3):491–499, 1988.

[81] M. Marcellino. Forecasting emu macroeconomic variables. *International Journal of Forecasting*, 20(2):359–372, 2004.

[82] D.G. Maringer and M. Meyer. Smooth transition autoregressive models-new approaches to the model selection problem. *Studies in Nonlinear Dynamics & Econometrics*, 12(1):5, 2008.

[83] N. Maugeri. Macroeconomic implications of near rational behavior: an application to the italian phillips curve. Technical report, Department of Economics, University of Siena Working Paper Series, 2010.

[84] David G. McMillan. Bubbles in the dividends price ratio? evidence from an asymmetric exponential smooth-transition model. *Journal of Banking and Finance*, 31(3):787 – 804, 2007.

[85] J.A. Mincer and V. Zarnowitz. *Economic Forecasts and Expectations*, chapter The Evaluation of Economic Forecasts, pages 1–46. New York: National Bureau of Economic Research, 1969.

[86] J.F. Muth. Rational expectations and the theory of price movements. *Econometrica: Journal of the Econometric Society*, 29:315–335, 1961.

[87] H. Nakaota. The term structure of interest rates in japan: the predictability of economic activity. *Japan and the World Economy*, 17(3):311–326, 2005.

[88] M. Nardo. The quantification of qualitative survey data: a critical assessment. *Journal of economic surveys*, 17(5):645–668, 2003.

[89] M. Nerlove. *The dynamics of supply: Estimation of farmers' response to price*. AMS Press, 1982.

[90] M. Nerlove. Expectations, plans, and realizations in theory and practice. *Econometrica*, 51(5):1251–79, 1983.

[91] H. Nielsen. Essays on expectations, 2003.

[92] B. Nobay, I. Paya, and D.A. Peel. Inflation dynamics in the us: global but not local mean reversion. *Journal of Money, Credit and Banking*, 42(1):135–150, 2010.

[93] M.H. Pesaran and A. Pick. Forecast combination across estimation windows. *Journal of Business and Economic Statistics*, 29(2):307–318, 2011.

[94] M.H. Pesaran and A. Timmermann. Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161, 2007.

[95] D. Pfajfar and E. Santoro. Heterogeneity, learning and information stickiness in inflation expectations. *Journal of Economic Behavior & Organization*, 75(3):426–444, 2010.

[96] B.M Potscher and I.R. Prucha. *Dynamic nonlinear econometric models: Asymptotic theory*. Springer Verlag, 1997.

[97] P. Raghubir and J. Srivastava. Effect of face value on product valuation in foreign currencies. *Journal of Consumer Research*, 29(3):335–347, 2002.

[98] A. Schrimpf and Q. Wang. A reappraisal of the leading indicator properties of the yield curve under structural instability. *International Journal of Forecasting*, 26(4):836–857, 2010.

[99] O. I. Shittu and O. O. S. Yaya. On fractionally integrated logistic smooth transitions in time series. *parameters*, 1:2, 2010.

[100] V.L. Smith. Constructivist and ecological rationality in economics. *The American Economic Review*, 93(3):465–508, 2003.

[101] J.H. Stock and M.W. Watson. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. Technical report, National Bureau of Economic Research, 1998.

[102] J.H. Stock and M.W. Watson. Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41(3):788–829, 2003.

[103] J.H. Stock and M.W. Watson. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430, 2004.

[104] N.R. Swanson and H. White. A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business & Economic Statistics*, 13(3):265–75, 1995.

[105] N.R. Swanson and H. White. Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, 13(4):439–461, 1997a.

[106] N.R. Swanson and H. White. A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics*, 79(4):540–550, 1997b 1997b.

[107] M.P. Taylor, D.A. Peel, and L. Sarno. Nonlinear mean-reversion in real exchange rates: Toward a solution to the purchasing power parity puzzles. *International Economic Review*, 42(4):1015–1042, 2001.

[108] T. Teräsvirta. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89(425):208–218, 1994.

[109] T. Teräsvirta. Forecasting economic variables with nonlinear models. *Handbook of Economic Forecasting*, 1:413–457, 2006.

[110] T. Teräsvirta, D. van Dijk, and M.C. Medeiros. Smooth transition autoregressions, neural networks, and linear models in forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting*, 21:755–774, 2005.

[111] A. Timmermann. Forecast combinations. *Handbook of economic forecasting*, 1:135–196, 2006.

[112] M.P. Tucci. The nonconvexities problem in adaptive control models: A simple computational solution. *Computational Economics*, 12(3):203–222, 1998.

[113] M.P. Tucci. A note on global optimization in adaptive control, econometrics and macroeconomics. *Journal of Economic Dynamics and Control*, 26(9-10):1739–1764, 2002.

[114] A. Tversky and D. Kahneman. Rational choice and the framing of decisions. *Economic Theory*, 251:S278, 1986.

[115] T. Tyszka and K. Przybyszewski. Cognitive and emotional factors affecting currency perception. *Journal of economic psychology*, 27(4):518–530, 2006.

[116] D. Van Dijk and P.H. Franses. Nonlinear error-correction models for interest rates in the Netherlands. In William A. Barnett, editor, *Nonlinear econometric modeling in time series: proceedings of the Eleventh International Symposium in Economic Theory*. Cambridge University Press, 2000.

[117] D. Van Dijk, T. Teräsvirta, and P. Franses. Smooth transition autoregressive models. *Econometric Reviews*, 21:1–47, 2002.

[118] Ioannis A. Venetis, Ivan Paya, and David A. Peel. Re-examination of the predictability of economic activity using the yield spread: a nonlinear approach. *International Review of Economics & Finance*, 12(2):187 – 206, 2003.

[119] P.P. Wakker, V. Köbberling, and C. Schwieren. Prospect-theorys diminishing sensitivity versus economics intrinsic utility of money: how the introduction of the euro can be used to disentangle the two empirically. *Theory and Decision*, 63(3):205–231, 2007.

[120] T. Weise. *Global Optimization Algorithms-Theory and Application.* Self-Published, 2009.

[121] K.D. West and W.K. Newey. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

[122] D.C. Wheelock and M.E. Wohar. Can the term spread predict output growth and recessions? a survey of the literature. *Federal Reserve Bank of St. Louis Review*, 91(5 Part 1):419–40, 2009.

[123] J. M. Wooldridge. Estimation and inference for dependent processes. *Handbook of econometrics*, 4:2639 – 2738, 1994.

[124] J. M. Wooldridge. *Econometric analysis of cross section and panel data.* The MIT press, 2002.

[125] P.K. Yadav, P.F. Pope, and K. Paudyal. Threshold autoregressive modeling in finance: The price differences of equivalent assets. *Mathematical Finance*, 4(2):205–221, 1994.

[126] T.M. Yigit. Inflation targeting: An indirect approach to assess the direct impact. *Journal of International Money and Finance*, 29(7):1357–1368, 2010.

[127] G. Yoon. Structural breaks and smooth transition autoregressive processes: an application to the us stock value ratios. *Applied Economics*, 43(18):2313–2320, 2011.