Non-linearities in Macroeconomics: Evaluation of Non-linear Time Series Models

By

Ana Beatriz Camatari Galvão

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Economics

Department of Economics
University of Warwick
August 2001

Contents

List of Figures vi

List of Tables viii

1 Introduction 1
- 1.1 Non-linear Time Series and Macroeconomics . . . 1
- 1.2 Plan of the Thesis . . . 3

2 Univariate Non-linear Time Series Models and the Business Cycles 6
- 2.1 Introduction . . . 6
- 2.2 A Description of the Models . . . 9
  - 2.2.1 TAR Models . . . 10
  - 2.2.2 Markov-Switching Models . . . 13
  - 2.2.3 Endogenous Threshold Models . . . 16
  - 2.2.4 Threshold Moving Average Models . . . 19
  - 2.2.5 Structural Models with Markov-Switching . . . 21
  - 2.2.6 Models for Italian GDP . . . 23
  - 2.2.7 Models for Australian GDP . . . 24
- 2.3 Evaluation Method . . . 26
  - 2.3.1 Definition of Stylised Facts . . . 26
  - 2.3.2 Measures of Fit . . . 31
- 2.4 Results of the Evaluation . . . 34
  - 2.4.1 Evaluation of TAR Models . . . 34
  - 2.4.2 Evaluation of Markov-Switching Models . . . 36
  - 2.4.3 Evaluation of Endogenous Threshold Models . . . 38
  - 2.4.4 Evaluation of Threshold Moving Average Models . . . 39
  - 2.4.5 Evaluation of Structural Models with Markov-Switching . . . 39
  - 2.4.6 Evaluation of Models for Italian GDP . . . 42
  - 2.4.7 Evaluation of Models of Australian GDP . . . 44
  - 2.4.8 Evaluation of Linear Models . . . 47
  - 2.4.9 Discussion of the Results of the Evaluation . . . 49
- 2.5 Conditional Means . . . 52
  - 2.5.1 On the Estimation of Conditional Mean Functions and Surfaces . . . 52
  - 2.5.2 Characteristics of Actual and Simulated Data . . . 54
  - 2.5.3 Analysis of Conditional Means . . . 56
- 2.6 Conclusions . . . 60

3 Non-linear Cointegrated Systems of US Term Structure of Interest Rates 68

3.1 Introduction . . . 68
3.2 Term Structure of Interest Rates and Equilibrium Correction Models . . . 71
3.3 Modelling Threshold Vector Equilibrium Correction models . . . 76
3.3.1 Non-linearity Testing . . . 76
3.3.2 Estimation . . . 81
3.3.3 Model Specification . . . 82
3.4 Evaluating Non-linear Equilibrium Correction Models . . . 83
3.5 Model Estimates . . . 87
3.5.1 Testing and Modelling . . . 87
3.5.2 A Comparison of Specifications . . . 93
3.5.3 Analysis of Estimates . . . 95
3.6 Forecasting Evaluation . . . 98
3.6.1 Forecast Evaluation of the System . . . 98
3.6.2 Forecasting Evaluation of Each Variable . . . 98
3.6.3 Forecasting Evaluation Conditional upon a Regime . . . 103
3.6.4 Forecasting Evaluation and Model Fit . . . 110
3.6.5 Simulation Exercise . . . 111
3.7 Analysis of the Robustness of the Forecasting Results . . . 114
3.7.1 Another Simulation Exercise . . . 116
3.7.2 A New Forecast Competition . . . 119
3.7.3 Comparing with Linear Autoregressive Models . . . 126
3.8 Conclusions . . . 129

4 Non-linearities and Structural Breaks in Predicting US Recessions using the Spread 131

4.1 Introduction . . . 131
4.2 The Spread and Economic Activity . . . 135
4.2.1 Descriptive Statistics and Conditional Means . . . 136
4.2.2 The Simple Rule and the NBER Peaks and Troughs . . . 138
4.3 A linear model . . . 141
4.4 Testing and Modelling Non-linearity . . . 141
4.4.1 Testing and Estimating Threshold and Smooth Transition Models . . 142
4.4.2 Analysis of the Results . . . 147
4.5 Testing and Modelling Structural Breaks . . . 150
4.5.1 Testing and Estimating Parameter Instability in Mean and Variance using Threshold and Smooth Transition Models . . . 151
4.5.2 Analysis of the Results . . . 154
4.6 Testing and Modelling Time-Varying/Structural Break Non-linearity . . . 158
4.6.1 Testing and Estimating Parameter Instability in Threshold Models and in Smooth Transition Models . . . 159
4.6.2 Analysis of the Results . . . 161
4.7 Forecasts of Event Probabilities . . . 166
4.8 Evaluating Models . . . 171
4.8.1 Description of the Models . . . 172
4.8.2 Results of the Evaluation . . . 174
4.9 Conclusions . . . 179
4.10 Appendix: Description of Bivariate Systems of Output Growth and Spread . 182

5 The Economics of Non-linearities in Empirical Models **192**
5.1 Introduction 192
5.2 Asymmetries and the US Business Cycle 193
5.2.1 Business Cycle Asymmetries 194
5.2.2 Non-linear Time Series Models and Business Cycle Asymmetries . . . 199
5.2.3 Business Cycle Theories and the Shape of the Business Cycle 203
5.2.4 Summary 204
5.3 Non-linearities and the Term Structure of Interest Rates 205
5.3.1 Comparing Conditional Means for Model and Data 205
5.3.2 Implications for the Expectations Theory 206
5.3.3 Forecasting the Spread near Business Cycle Turning Points 211
5.3.4 Summary 212
5.4 Predicting Recession in 2001 212
5.4.1 Forecasts for the Out-of-Sample Period 214
5.4.2 Forecasts of the Probability of Events A and B for 2001 214
5.4.3 Comparing with Stock and Watson Recession Index 215
5.4.4 Comparing with the Survey of Professional Forecasters 216
5.4.5 Summary 218
5.5 Conclusions 219

6 Conclusion **221**
6.1 Main findings 222
6.2 Other contributions 224
6.3 Some Open Questions 225

Bibliography **228**

List of Figures

2.1 Cumulative gains (losses) against the last though (peak) for US GDP (dashed lines are NBER peaks and troughs) . . . 30
2.2 Cumulative gains (losses) against the last though (peak) for Italian GDP . . 30
2.3 Cumulative gains (losses) against the last though (peak) for Australian GDP 31
2.4 $E[y_t|y_{t-1}]$ for US GDP growth . . . 56
2.5 $E[y_t|y_{t-1}, y_{t-2}]$ for US GDP growth . . . 58
2.6 $E[y_t|y_{t-1}]$ for US GDP growth and data simulated from SETAR 2, SETAR 4, MRSTAR, MS2, MS DD, MS AR and MS3 . . . 61
2.7 $E[y_t|y_{t-1}]$ for US GDP growth and data simulated from CDR, F&C, asMA, TDMA and AR . . . 62
2.8 $E[y_t|y_{t-1}, y_{t-2}]$ for SETAR2 . . . 63
2.9 $E[y_t|y_{t-1}, y_{t-2}]$ for SETAR4 . . . 63
2.10 $E[y_t|y_{t-1}, y_{t-2}]$ for MRSTAR . . . 63
2.11 $E[y_t|y_{t-1}, y_{t-2}]$ for MS2 . . . 64
2.12 $E[y_t|y_{t-1}, y_{t-2}]$ for MS DD . . . 64
2.13 $E[y_t|y_{t-1}, y_{t-2}]$ for MS AR . . . 64
2.14 $E[y_t|y_{t-1}, y_{t-2}]$ for MS3 . . . 65
2.15 $E[y_t|y_{t-1}, y_{t-2}]$ for CDR . . . 65
2.16 $E[y_t|y_{t-1}, y_{t-2}]$ for F&C . . . 65
2.17 $E[y_t|y_{t-1}, y_{t-2}]$ for asMA . . . 66
2.18 $E[y_t|y_{t-1}, y_{t-2}]$ for TDMA . . . 66
2.19 $E[y_t|y_{t-1}, y_{t-2}]$ for AR . . . 66

3.1 Long-term (10-year T-bond) and short-term (3-month T-bill) interest rates (dashed lines are NBER turning points) . . . 88
3.2 The spread between long- and short-term interest rates (dashed lines are NBER turning points) . . . 88
3.3 Comparing GFESMs for $h = 1, ..., 24$ . . . 99
3.4 Comparing MSFEs for $\Delta r(l)$ for $h = 1, ..., 24$ . . . 99
3.5 Comparing MSFEs for $\Delta r(s)$ for $h = 1, ..., 24$ . . . 100
3.6 Comparing MSFEs for S for $h = 1, ..., 24$ . . . 100
3.7 Comparison of MSFEs for $\Delta r(l)$ conditional on $S_{t-1} > 2.7$ . . . 110
3.8 MSFEs for $\Delta r(s)$ with data generated from the 2R-TVECM, conditional on $S_{t-1} \leq 0$ . . . 112
3.9 MSFEs for S with data generated from the 2R-TVECM, conditional on $S_{t-1} \leq 0$112
3.10 MSFEs for $\Delta r(l)$ with data generated from the 3R-TVECM, conditional on $S_{t-1} > 2.7$ . . . 113

3.11 MSFEs for S with data generated from the 3R-TVECM, conditional on $S_{t-1} > 2.7$ . . . 113
3.12 RMSFEs for S with data generated from the 2R-TVECM, conditional on $S_{t-1} \leq 0$, with VAR as benchmark . . . 117
3.13 RMSFEs for S with data generated from the 3R-TVECM, conditional on $S_{t-1} > 2.7$, with VAR as benchmark . . . 118
3.14 Comparing homoscedastic with heteroscedastic bootstrap forecasts (2R-TVEqCM$_{EG}$ and 3R-TVEqCM$_{EG}$ are in the secondary scale) . . . 119
3.15 Comparison of RMSFEs for $\Delta r(l)$ with VAR as benchmark . . . 120
3.16 Comparison of RMSFEs for $\Delta r(s)$ with VAR as benchmark . . . 121
3.17 Comparison of RMSFEs for S with VAR as benchmark . . . 121
3.18 Comparison of GFESMs with VAR as benchmark . . . 124
3.19 Comparison of RMSFEs for $\Delta r(l)$ with AR as benchmark . . . 126
3.20 Comparison of RMSFEs for $\Delta r(s)$ with AR as benchmark . . . 127
3.21 Comparison of RMSFEs for S with AR as benchmark . . . 127

4.1 The spread between 10-year T-bond and 3-month T-bill and the output growth 136
4.2 Conditional means estimated with smooth splines ($E[y_t|S_{t-p}]$ for $p = 1, ..., 6$ for the sample 1953-1983 in the two upper panels and for $p = 1, ..., 8$ for sample 1984-2000 in the lower panels) . . . 139
4.3 Predicting US recessions with a simple rule . . . 140
4.4 F statistics for all possible break-points for test of structural break in mean . 155
4.5 F statistics for all possible break-points for test of structural break in variance 162
4.6 Predictions of the probability of event A for 1954:3 to 1999:4 by simple rule, VAR, ST, T, 3T, and TVAR . . . 179
4.7 Predictions of the probability of event A for 1954:3 to 1999:4 by SBMV, dST, TVST, PTVST, SBT, and SBTVAR . . . 180
4.8 Predictions of the probability of event B for 1954:3 to 1999:4 by simple rule, VAR, ST, T, 3T, and TVAR . . . 180
4.9 Predictions of the probability of event B for 1954:3 to 1999:4 by SBMV, dST, TVST, PTVST, SBT, and SBTVAR . . . 181

5.1 Gain (loss) against last trough (peak) and output growth for US GDP . . . 195
5.2 The shape of the US business cycle . . . 196
5.3 Comparing shapes of stylised business cycles generated by non-linear models 201
5.4 Conditional mean functions for three-regime TVEqCM and the data . . . 207
5.5 Predicting the spread at the onset of recession . . . 211
5.6 Comparing output growth forecasts for 2000:1 to 2001:1 . . . 215

List of Tables

2.1 Business cycle stylised facts of TAR models 35
2.2 Business cycle stylised facts of Markov-switching models 37
2.3 Business cycle stylised facts of endogenous threshold models 40
2.4 Business cycle stylised facts of threshold moving average models 41
2.5 Business cycle stylised facts of structural models with Markov-switching . . . 43
2.6 Business cycle stylised facts of non-linear univariate models of Italian GDP . 45
2.7 Business cycle stylised factos of non-linear univariate models of Australian GDP 46
2.8 Business cycle stylised facts of AR models 48
2.9 Summary of stylised facts for US business cycles 50
2.10 Model labels and sources 55
2.11 Descriptive statistics of actual and simulated data 55

3.1 Characteristics of the models 90
3.2 Test results 92
3.3 Estimation results 97
3.4 Augmented Diebold and Mariano Forecast accuracy tests for first differences of long-term rates 104
3.5 Encompassing tests for first differences of long-term rates 105
3.6 Augmented Diebold and Mariano forecast acuracy tests for first differences of short-term rates 106
3.7 Encompassing tests for first differences of short-term rates 107
3.8 Augmented Diebold and Mariano forecast accuracy tests for the spread . . . 108
3.9 Encompassing tests for the spread 109

4.1 Descriptive statistics of output growth and the spread 137
4.2 Non-linearity tests 148
4.3 Time-varying/Structural break tests 156
4.4 Time-varying/Structural break tests in non-linear models 164
4.5 Evaluation of probability forecasting (1954:3-1999:4) 175
4.6 Evaluation of probability forecasting (1983:1-1999:4) 176

5.1 Comparison of out-of-sample predictions 215
5.2 Comparing 2001 forecasts 218

To Brother

Acknowledgments

I am grateful to my supervisors Dr. Michael Clements and Prof. Kenneth Wallis for comments, suggestions and support. Specifically, my knowledge on forecasting has been substantially improved by discussions with Mike Clements.

This thesis has benefited from the comments and suggestions of participants in the Conferences and Seminars at which parts of this work were presented: the Econometrics Workshop at University of Warwick, the Young Economist Meeting, the Conference on Business Cycle and Economic Growth at University of Manchester, the Annual Meeting of the Royal Economic Society and the Annual Symposium of the Society for Nonlinear Dynamics and Econometrics. I want particularly to thank Dick Van Dijk for his comments on earlier versions of some parts of this work. Acknowledgments to authors of some computer codes employed in this work are made when a specific problem is referred to in the main body of the thesis.

The development of the research presented in this work was only possible given the financial support provided by the Brazilian government via the "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES".

I am grateful to Kostas Galanakis for his continued support throughout the writing of this thesis.

Declaration

I declare that this thesis is the result of my own work and this includes the parts in which co-authored papers have been derived as indicated. I declare that this thesis has not been submitted for a degree at another university.

Two papers that have been derived based on Chapter 2 of this thesis. The first one is "Univariate Non-linear Time Series Models and the Business Cycle Stylised Facts". This paper was presented at the Econometrics Workshop at University of Warwick (March, 2000), at the Young Economist Meeting (Oxford, March, 2000), at the Conference on Business Cycles and Economic Growth at University of Manchester (July, 2000) and at the Annual Meeting of the Royal Economic Society (St. Andrews, July, 2000). The second one is "Conditional Mean Functions of Non-Linear Models of US Ouput", which is co-authored by my supervisor Michael Clements and is forthcoming in *Empirical Economics.*

Two other papers have been derived based on Chapter 3. The first one is "An Evaluation of Non-Linear Cointegrated Systems of the US Term-Structure of Interest Rates", which is co-authored by Michael Clements. This paper was presented at the Econometrics Workshop at University of Warwick (January, 2001), at the Annual Symposium of the Society of Nonlinear Dynamics and Econometrics (Atlanta, March, 2001) and will be presented at the Annual European Meeting of the Econometric Society (Lausanne, August, 2001). The paper has also been submitted for publication. The second one, which is also co-authored by Michael Clements, is "Non-linear Cointegrated Models of the US Term Structure of Interest Rates". This paper has also been submitted for publication.

The paper "Predictions of the Probability of US Recessions: The Role of Structural Breaks and Non-linearities" has been derived based on Chapter 4. This paper will be presented in a seminar at UFRGS (Porto Alegre, August, 2001).

Summary

This thesis evaluates different specifications of non-linear time series models applied to macroeconomic problems. The evaluations investigate whether linear models are a good representation of the data, and which non-linear specifications are comparatively better in three different applications. In addition, the implications of the evaluation to the understanding of macroeconomic problems and to economic predictions are analysed.

The first evaluation concerns univariate non-linear time series models aimed at reproducing the asymmetries of the business cycles. Using business cycle stylised facts and conditional mean functions and surfaces, the results support the use of non-linear models that can generate a three-phase cycle as the specification that can reproduce all the business cycle features, including the asymmetries in the shape of the cycle.

The second assessment is of models that characterise the non-linearities of the US term structure of interest rates. The forecast evaluation of different specifications of threshold vector equilibrium correction models, which are estimated for long- and short-term interest rates and their spread, shows that the inclusion of non-linearity improves short-horizon forecasts. However, when compared with AR models, the gains from non-linearity only occur when the predictions for the spread are evaluated at long horizons.

The third assessment concerns non-linear bivariate systems that account for the effect of non-linearities and/or structural breaks when the spread is employed as leading indicator. Different specifications are evaluated using their prediction of the probability of two definitions of recessions. Models with non-linearities and structural breaks perform better at predicting the probability of recession than linear models and models with only non-linearity or structural break.

The results of the evaluation of univariate time series models improve the understanding of the connection between these models and business cycle asymmetries. The winner of the forecast competition of bivariate systems of interest rates and their spread indicates that the expectation theory of the term structure of interest only holds for the period in which the spread is negative, even though the spread can predict changes in the long-term rate in a specific state. In addition, the result that structural breaks and non-linearities are important to predict US recessions when the spread is the leading indicator changes the timing of a predicted recession for 2001.

Chapter 1

Introduction

1.1 Non-linear Time Series and Macroeconomics

The development of time-series econometrics applied to macroeconomics has been substantial in the last two decades. The text-books of Hamilton (1994) and Hendry (1995) have good examples of these advances. Most of these developments were made in the linear framework because the linearity assumption is "an important one for proving mathematical and statistical theorems" (Intriligator, 1983). A gradual appreciation of the limitations of the linearity assumption in some settings, such as for characterising business cycles, has promoted the research on non-linear time series models in the last ten years. "The hope in using nonlinear models is that better explanations can be provided of economic events and consequently better forecasts " (Teräsvirta, Tjostheim and Granger, 1994, p. 2921). Therefore, the purposes of the recent developments on non-linear time-series econometrics are structural analysis and forecasting[1], although non-linear time series models have been also applied to policy evaluation in specific macroeconomic problems (Rothman, Van Dijk and Franses, 1999). Non-linear time series models are also employed to improve the knowledge on empirical evidence, which is one of the role of econometrics (Hendry, 1995, chap. 1).

[1] The three purposes of econometrics are "structural analysis, forecasting, and policy evaluation" (Intriligator, 1983, p. 182)

Non-linear time series models have been applied to characterise, for example: (i) business cycle asymmetries (Hamilton, 1989; Beaudry and Koop, 1993; Potter, 1995); (ii) asymmetries in the equilibrium adjustment of labour demand (Pfann and Palm, 1993; Escribano and Pfann, 1998); (iii) the effect of monetary regimes and transaction costs on the ability of the spread to predict interest rate yields (Pfann, Schotman and Tschernig, 1996; Anderson, 1997; Hansen and Seo, 2000); (iv) asymmetries in the effect of monetary policy on output (Weise, 1999); (v) non-linear causality between money and output (Rothman et al., 1999); (vi) asymmetries in the ability of the spread to predict economic activity (Anderson and Vahid, 2000); (vii) non-linearities and structural instability in the Phillips curve (Eliasson, 1999; Hamilton, 2001); (viii) the presence of common non-linear macroeconomic fluctuations among countries (Anderson and Vahid, 1998; Paap and Franses, 1999; Krolzig and Toro, 1999).

Different parametric specifications of non-linear time series models have been proposed (for surveys, see, e.g., Tong, 1990, Granger and Teräsvirta, 1993, Kim and Nelson, 1999c, Franses and Van Dijk, 2000, and Potter, 2000). Threshold autoregressive models (Potter, 1995), smooth transition autoregressive models (Teräsvirta and Anderson, 1992) and Markov-switching models (Hamilton, 1989) are popular specifications employed in univariate applications. Non-linearities have been included in equilibrium correction models using asymmetric adjustment (Escribano and Granger, 1998), thresholds (Balke and Fomby, 1997), smooth transition (Anderson, 1997) and shifts given by a Markov-chain (Krolzig and Toro, 1999). In addition, non-linearities have been allowed in vector autoregressive models in the form of thresholds (Tsay, 1998), smooth transitions (Anderson and Vahid, 1998) and switches given by a Markov-chain (Krolzig, 1997).

Even though non-linear time series models have been applied extensively to macroeconomic data, Granger (2001) concludes that the major weakness of the literature on non-linear macroeconomic empirical models is in the evaluation phase of the study. However, "If a researcher proposes a non-linear time series model, the question will invariably arise: Is the non-linear specification superior to a linear model?"(Hansen, 2000a, p. 47). But this is not the only point that needs evaluation, because

> I would like to suggest that in the future when you are presented with a new piece of theory or an empirical model, you ask these questions: (i) What purpose does it have? What economic decision does it help with? and; (ii) Is there any evidence being presented that allows me to evaluate its quality compared with alternative theories or models? (Granger, 1999, p. 58).

For this reason, an evaluation study should also compare different types of non-linear empirical models, given that they are built for similar purposes to help in taking similar economic decisions. Therefore, this thesis is concerned with the evaluation of different specifications of non-linear time series models applied to macroeconomic problems. The evaluation concerns whether linear models are a good representation of the data and which non-linear specifications are comparatively better. In addition, the implications of the evaluation for the understanding of macroeconomic problems and for forecasting are analysed.

1.2 Plan of the Thesis

With these objectives, the structure of the thesis is as follows. Chapters 2, 3 and 4 present three different evaluations for different types of non-linear time series models applied to business cycle asymmetries, term structure of interest rates and predictions of the probability of recession. The three chapters employ different evaluation methods and assess different non-linear models, considering in turn univariate analysis, equilibrium correction models, and simple vector processes. Some new non-linear specifications, including modelling and testing procedures, are proposed in Chapters 3 and 4 for inclusion in the evaluation. Chapter 5 concerns the economics of the results of the econometric evaluation of empirical models.

Specifically, Chapter 2 evaluates univariate non-linear time series models aimed to reproducing asymmetries of the business cycles. The comparisons consider different specifications of the following non-linear models: Markov-switching models, threshold autoregressive models, endogenous threshold models, threshold moving average models and state-space models with Markov-switching. These models are evaluated according to their ability to reproduce business cycle stylised facts. In addition, a comparative assessment of the relevance of non-linearity is made using a simple autoregressive model. The evaluation of conditional mean functions and surfaces, estimated non-parametrically with data simulated from the models, is employed to observe the dynamic pattern implied by these models. This helps in understanding the type of business cycle asymmetries for which each model is able to account.

Chapter 3 evaluates non-linear equilibrium correction models estimated to characterise non-linearities generated from risk premia, transaction costs and monetary policy regimes in the term structure of interest rates. Different specifications of threshold vector equilibrium correction models based on different testing, estimation and modelling procedures are evaluated and compared with vector equilibrium correction models, vector autoregressive models and univariate autoregressive models. The method of evaluation employs point forecasts generated for a large sample period, including forecast accuracy and forecast encompassing tests, and simulations.

Chapter 4 evaluates non-linear bivariate systems that account for the effect of non-linearities and structural breaks in models that employ the spread to predict economic activity. Non-linearities are incorporated by estimating and testing smooth transition and threshold models. Time-varying and structural break models are tested and estimated to incorporate time changes. Finally, time-varying smooth transition models and structural break threshold models are tested and estimated. These different models are evaluated according to their ability to predict event probabilities, employing three different score rules. The events are defined to characterise recessions. The performance of the models is compared with vector autoregressive models and a simple rule.

Chapter 5 analyses the economic consequences of the results of the evaluation of chapters 2, 3 and 4. First, it shows which types of univariate time series model can reproduce the asymmetric shape of the US business cycle and which types of business cycle theory support the results. Second, the dynamics of the best forecaster for interest rates, given by the evaluation of Chapter 3, is compared with the predictions of the expectations theory of the term structure of interest rates. Third, the predictions of the probability of recession in the US for 2001 from models with non-linearities and structural breaks using the spread as leading indicator compared with other popular forecasters of the recession probabilities.

Chapter 6 summarises the main findings and other contributions of this thesis and indicates directions for future research.

Chapter 2

Univariate Non-linear Time Series Models and the Business Cycles

2.1 Introduction

The asymmetry between the phases of business cycle is extensively investigated in the literature. The tests of steepness and deepness of Sichel (1993) have been applied to macroeconomic series from different countries, showing some support for asymmetries (Holly and Stannett, 1995; Speight, 1997; Speight and McMillan, 1998). Another test for verifying asymmetries – proposed by Verbrugge (1997) – shows evidence of asymmetry for some US macroeconomic series. The asymmetries in business cycles may indicate that the same policy may have different effects depending on the phase of the cycle, and in consequence, the recognition of asymmetries is important for policy making (Boldin, 1999).

Given that business cycles are asymmetric, non-linear models may be necessary to reproduce the data asymmetries, because linear models give symmetric responses to shocks. Non-linear parametric models that are able to account for these asymmetries have been applied to model US output growth, such as Markov-switching models (Hamilton, 1989), threshold models (Potter, 1995), endogenous threshold models (Beaudry and Koop, 1993), threshold moving average models (Elwood, 1998) and structural models with Markov-switching (Kim and Nelson, 1999a)[1]. In addition, non-linear autoregressive models are presented to model the growth of some OECD (Bradley and Jansen, 1997; Peel and Speight, 1998b; Stanca, 1999; Bodman, 1998) and Latin American countries (Greenaway, Leybourne and Sapsford, 1997; Mejía-Reyes, 2000). Non-linearities in the cycles derived from industrial production have also been accounted for using different non-linear specifications (Teräsvirta and Anderson, 1992; Filardo, 1994; Gallegati and Mignaca, 1995; Proietti, 1998). Finally, non-linear models have been employed to model asymmetries and non-linearities in unemployment (Hansen, 1997b; Peel and Speight, 1998a; Skalin and Teräsvirta, 1998; Parker and Rothman, 1998; Koop and Potter, 1999b).

Even though non-linear models can account for characteristics of the data that linear models cannot, the main point is whether employing a linear instead of a non-linear model loses accuracy. The gain or loss of assuming linearity depends on the criterion and the type of non-linear model employed. Non-linear models have their statistical accuracy compared with linear models by means of linear testing against a specific parametric alternative, such as the evaluation of Hansen (1996) on threshold models for US output. In general, linearity tests against different parametric alternatives have been applied to many series, and linearity is rejected for some cases (for surveys of linearity testing, see Granger and Teräsvirta, 1993, and Franses and Van Dijk, 2000).

Another possibility for assessing the gains from assuming non-linearity is to evaluate forecasts. For US output, assuming linearity does not imply any significant forecast accuracy loss (Clements and Krolzig, 1998), probably because of the low degree of non-linearity of the series (Clements, Franses, Smith and Van Dijk, 2000), although evaluating forecasts conditional on the regime (Tiao and Tsay, 1994) and using the forecast density (Clements and Smith, 2000; Lundbergh and Teräsvirta, 2000) results in stronger gains from the inclusion

[1]Literature review on non-linear time series models for US output is presented in section 2.2.

of non-linearity. In the case of unemployment, forecast accuracy gains from non-linearity are stronger for US data (Rothman, 1998; Montgomery, Zarnowitz, Tsay and Tiao, 1998). Non-linear autoregressive models for many macroeconomic series are evaluated by Stock and Watson (1999), who concluded that although some non-linear models have the best forecast accuracy for some series, in general this is not true.

Impulse responses have also been employed to evaluate the degree of asymmetry indicated by non-linear models. They have been calculated for threshold models (Potter, 1995; Koop and Potter, 1999b) and endogenous threshold models (Pesaran and Potter, 1997; Bradley and Jansen, 1997; Jansen and Oh, 1999). In addition, Van Dijk, Franses and Boswijk (2000) show how absorption can be calculated to compare the observed asymmetry in impulse responses. Markov-switching models, threshold models and linear autoregressive models are also evaluated using Bayes factors by Koop and Potter (1999a). They conclude that there is no strong evidence of non-linearity in US GNP. This result is robust to the comparison of non-linear models with models with structural breaks (Koop and Potter, 2000), even though non-linear models are better able to model the behaviour of the data than time-varying models (Koop and Potter, 2001).

Finally, another approach, which is based on the kind of evaluation employed for Real Business Cycle models (King and Plosser, 1994; Harding and Pagan, 2000), is to verify whether non-linear models can account for business cycle stylised facts (Hess and Iwata, 1997b; Harding and Pagan, 2001b). This evaluation method is complementary to the evaluation using impulse responses, forecasting, Bayes factors or linearity tests. By defining a set of stylised facts based on the characteristics of the observed business cycles, the objective is to evaluate the ability of the model to characterise the observed asymmetries. Given that previous evaluations (Hess and Iwata, 1997b; Harding and Pagan, 2001b) support the view that an AR(1) model of US GDP growth is as good as non-linear models in generating business cycle stylised facts, this chapter extends the set of non-linear models analysed, includes new stylised facts and evaluates models for other countries than the US, with the objective of testing the robustness of previous results. In addition, we propose to use conditional mean functions and surfaces, estimated non-parametrically with simulated data, to shed light on the complex dynamics of non-linear autoregressive models, which may be responsible for the ability of the model to account for stylised facts.

Therefore, the main focus of this chapter is whether non-linear univariate models can reproduce the asymmetric stylised facts of the business cycle. In addition, it shows how conditional mean functions and surfaces can be employed to characterise the non-linear dynamics implied by non-linear univariate models.

Section 2.2 reviews the non-linear time series models applied to output growth for US, Italy and Australia, which are the subject of our evaluation. The method of evaluation employing business cycle stylised facts is discussed in section 2.3, and the results of the evaluation are presented in section 2.4. The relevance of conditional mean surfaces to observe the dynamics of non-linear models is analysed in section 2.5, including the presentation of conditional mean functions and surfaces of some models described in section 2.2. Section 2.6 summarises the main results and contributions of this chapter.

2.2 A Description of the Models

Many types of non-linear univariate models, with the objective of capturing the characteristics of business cycles or for forecasting, are published in economics and econometrics journals. The business cycle is studied in the majority of cases by analysing the first-difference of post-war real (seasonally adjusted) GDP and GNP. Non-linear cyclical behaviour is also estimated in unemployment and industrial production series. Most of the literature analyses US data, but business cycle non-linearities have been tested for the majority of OECD countries[2]. The tests indicate that some countries do not present asymmetries in output, e.g. Japan and UK (Goodwin, 1993; Mills, 1995; Bradley and Jansen, 1997). For some countries and series, the literature presents only isolated examples of the application of non-linear univariate time series models. Therefore, we choose to study models estimated for US output and for Australian and Italian GDP because the literature presents different non-linear specifications for similar data sets[3]. Some of these models (5 out of 20) have already been evaluated using stylised facts by Hess and Iwata (1997b) and Harding and Pagan (2001b). However, the method applied below has novel aspects, which may yield new insights.

2.2.1 TAR Models

Threshold autoregressive models are linear conditional on a regime, where the regime is defined by a threshold variable (transition function), a delay and a threshold value. This class of models, which is a generalisation of ARMA time series models, is able to capture global non-linear behaviour (such as limit cycles) because of their piecewise linear construction, and has been applied to many types of time series data (Tong, 1990). Potter (1995) demonstrates that SETAR models can generate asymmetric impulse-response, which depends on the sign and on the magnitude of the shock. Potter (1995) and Peel and Speight (1998b) present SETAR (Self-exciting Threshold Autoregressive) models with two regimes for the US Real GNP. Tiao and Tsay (1994) and Van Dijk and Franses (1999) estimate four-regime TAR models for the output growth.

Potter's (1995) specification evaluated in this work is the one presented by Hess and Iwata (1997b) because the coefficients presented in Table 3 of Potter generate non-stationary

[2]See for example Peel and Speight (1998a) for unemployment, Teräsvirta and Anderson (1992) for industrial production, and Peel and Speight (1998b) and Bradley and Jansen (1997) for output.

[3]Many articles (Hansen, 1997b, Koop and Potter, 1999b, and Montgomery et al., 1998, for example) are published for US unemployment; however, they are not evaluated due to the absence of uniformity of frequency (quarterly or monthly data) and detrending procedures (levels or differences) that creates bias for the evaluation when re-estimation is not done.

data for long simulations[4]. The sample is from 1947:01 to 1992:04 (quarterly data) and the estimated model (Hess and Iwata, 1997b, Appendix) is described as follows:

$$y_t = \begin{cases} -0.007 + 0.302y_{t-1} - 0.600y_{t-2} + 0.028y_{t-5} + \epsilon_{1,t} & \text{if } y_{t-2} \leq 0; \\ 0.004 + 0.326y_{t-1} + 0.195y_{t-2} - 0.060y_{t-5} + \epsilon_{2,t} & \text{if } y_{t-2} > 0; \end{cases} \tag{2.1}$$

$$n_1 = 35;\ \sigma_{\epsilon_1} = 0.0121;\ n_2 = 140;\ \sigma_{\epsilon_2} = 0.0088;$$

where y_t is the first-difference of the ln(GDP), n_i is the number of observations in each regime and $\sigma_{\epsilon_1}, \sigma_{\epsilon_2}$ are the standard deviations of the residuals. Linearity tests have been applied to this model (Hansen, 1996), supporting non-linearity. The key point is the high negative AR2 coefficient in the recession regime (-0.6). This coefficient means that the recoveries are rapid or that the model has an intrinsic stabiliser. Moreover, the coefficient creates asymmetries in the impulse-response of negative shocks. Hess and Iwata (1997b) and Koop and Potter (1999a) evaluate different TAR specifications using, respectively, stylised facts and Bayesian analysis. They find some support for the two-regime SETAR as a good representation of the data.

Peel and Speight (1998b) compare SETAR models for trend-stationary and difference-stationary US GNP. The model with difference-stationary data, based on the 1957:1 to 1991:3 sample, is defined as:

$$y_t = \begin{cases} 0.294 + \epsilon_{1,t} & \text{if } y_{t-2} \leq 0.139;\ \sigma^2_{\epsilon_1} = 1.520; \\ 0.582 + 0.337y_{t-1} + \epsilon_{2,t} & \text{if } y_{t-2} > 0.139;\ \sigma^2_{\epsilon_2} = 0.631; \end{cases} \tag{2.2}$$

where y_t is the first-difference of $100(ln(\text{GNP}))$ and $\sigma^2_{\epsilon_1}, \sigma^2_{\epsilon_2}$ are the variances of the residuals. Compared with the previous two-regime SETAR model, the model of equation 2.2 has a threshold estimated by grid search and no autoregressive term in the first regime. The autoregressive coefficient in the second regime (0.337) is similar to the linear AR specification

[4]Simulated data with infinite moments is also a problem of a Markov-Switching model evaluated by Breunig and Pagan (2001), who argue that the simulation of a non-linear time series model is good way of observing its characteristics.

for similar data set[5].

Tiao and Tsay (1994) demonstrate that the presence of four regimes instead of two improves the performance of the SETAR model for US GNP. The model is based on non-linear tests applied to the residuals of a traditional AR(2) model. The four-regime SETAR (SETAR 4) for 1947:1 to 1991:1 is defined as follow:

$$y_t = \begin{cases} -0.015 - 1.076\, y_{t-1} + \epsilon_{1,t} & \text{if } y_{t-1} \le y_{t-2} \le 0; \quad \sigma_{\epsilon_1} = 0.0062; \\ -0.006 + 0.630 y_{t-1} - 0.756 y_{t-2} + \epsilon_{2,t} & \text{if } y_{t-1} > y_{t-2} \text{ and } y_{t-2} \le 0; \quad \sigma_{\epsilon_2} = 0.0132; \\ 0.006 + 0.438 y_{t-1} + \epsilon_{3,t} & \text{if } y_{t-1} \le y_{t-2} \text{ and } y_{t-2} > 0; \quad \sigma_{\epsilon_3} = 0.0094; \\ 0.004 + 0.443 y_{t-1} + \epsilon_{4,t} & \text{if } y_{t-1} > y_{t-2} > 0; \quad \sigma_{\epsilon_4} = 0.0082; \end{cases} \tag{2.3}$$

where y_t is the first-difference of ln(GNP) and $\sigma_{\epsilon_1}, \sigma_{\epsilon_2}, \sigma_{\epsilon_3}, \sigma_{\epsilon_4}$ are the standard deviations of the residuals. Regime 1 ($y_{t-1} \le y_{t-2} \le 0$) is marked by negative growth two periods ago ($t-2$), worsening in period ($t-1$), and is characterised by an explosive root to bring the economy out of recession. Regime 2 implies negative growth in $t-2$, but improving at $t-1$, with a negative second order autoregressive term similar to Potter's model. Regimes 3 and 4 are similar, and are operative when $t-2$ growth was positive and either slowed in $t-1$ or accelerated. Therefore, recessions in this model are temporary and the growth during expansions does not depend on the economy being slowing down or accelerating. The forecast evaluation of this model indicates that the four-regime SETAR is a better forecaster than an AR(2), conditional on certain states. In general, however, the results are only slightly better than the linear models.

Van Dijk and Franses (1999) also present a four-regime TAR. However, their model has a logistic smooth transition, employing the CDR[6] index and the second-difference of output as transition variables. LM tests for non-linearity are employed to specify the Multiple

[5]The ARIMA (1,1,0) model estimated for comparative purposes (see section 2.4) has autoregressive coefficient of 0.34 for the 1947:1-1999:2 period.

[6]Current Depth of Recession. More examples of application of this variable are presented in section 2.2.3.

Regime Smooth Transition (MRSTAR) model. Their preferred model estimated for 1947:1 to 1995:2, is described as:

$$\begin{aligned}
y_t &= [(0.394 + 0.460y_{t-1} + 0.092y_{t-2})(1 - F(\Delta y_{t-1})) + \\
&\quad (-0.121 + 0.442y_{t-1} + 0.346y_{t-2})F(\Delta y_{t-1})](1 - F(CDR_{t-2})] + \\
&\quad [(0.360 - 0.530y_{t-1} + 0.963y_{t-2})(1 - F(\Delta y_{t-1})) + \\
&\quad (-0.019 + 0.744y_{t-1} - 0.235y_{t-2})F(\Delta y_{t-1})]F(CDR_{t-2}) + \epsilon_t \\
F(\Delta y_{t-1}) &= (1 + \exp[-500(\Delta y_{t-1} - 0.250)/\sigma_{\Delta y_{t-1}}])^{-1} \\
F(CDR_{t-2}) &= (1 + \exp[-500(CDR_{t-2} - 0.064)/\sigma_{CDR_{t-2}}])^{-1} \; \sigma_\epsilon = 0.867;
\end{aligned} \tag{2.4}$$

where y_t is the first-difference of $100(ln(\text{GNP}))$, Δy_t is the second-difference of the $100(\ln(GNP))$, σ_ϵ is the residuals' standard deviation, $F(\Delta y_{t-1})$ and $F(CDR_{t-2})$ are transition functions. The high value of the slope coefficient in the transition functions implies that the switch between regimes is instantaneous, similar to a SETAR model. When the impulse response of this model is evaluated, negative shocks are less persistent than positive ones for shocks with large size.

2.2.2 Markov-Switching Models

While the TAR models were built on developments over traditional ARMA time series models, the Markov-switching (MS) model is based on the combination of models with unobserved variables, a model with switching given by a step function and the evidence of non-linearity by non-parametric tests for US output series (Hamilton, 1989). An appealing characteristic of MS models is that they can be employed as an alternative algorithm to define the turning points of business cycles because the expected duration of contractions and expansions can be predicted from the estimated transition probabilities[7]. The switching between regimes depends on unobserved variables that follow a Markov chain.

[7]The relevance of this type of model to define business cycle turning points is evaluated by Harding and Pagan (2001a).

Hamilton (1989) presents a MS model in which the persistence of shocks is different between the regimes, because they exhibit different expected durations which are calculated based on the transition probabilities. The two-regime MS (MS2) model, which is estimated for 1951:2-1984:4, is described as:

$$\begin{aligned}(y_t - \mu_{s_t}) &= 0.014(y_{t-1} - \mu_{s_{t-1}}) - 0.058(y_{t-2} - \mu_{s_{t-2}}) \\ &\quad - 0.247(y_{t-3} - \mu_{s_{t-3}}) - 0.213(y_{t-4} - \mu_{s_{t-4}}) + \epsilon_t \\ \mu_{s_t} &= -0.3577(1 - S_t) + 1.522 S_t; \\ \Pr[S_t = 1|S_{t-1} = 1] &= 0.9049; \ \Pr[S_t = 0|S_{t-1} = 0] = 0.7550; \quad \sigma_\epsilon = 0.769;\end{aligned} \tag{2.5}$$

where y_t is the first-difference of $100(ln(\text{GNP}))$; $S_t = 1$ defines the expansion regime and $S_t = 0$ defines the contraction regime.

Using the same sample as Hamilton (1989), Durland and McCurdy (1994) relax the hypothesis of constant transition probabilities – "the transitions probabilities are functions of both the inferred current state and also the number of periods that the process has been in that state" (Durland and McCurdy, 1994, p. 279). Duration dependence is found, on the basis of in-sample LR test, for contractions but not for expansions. The Markov-switching model with duration dependence (MS DD) is written as:

$$\begin{aligned}(y_t - \mu_{s_t}) &= -0.017(y_{t-1} - \mu_{s_{t-1}}) - 0.092(y_{t-2} - \mu_{s_{t-2}}) \\ &\quad - 0.255(y_{t-3} - \mu_{s_{t-3}}) - 0.246(y_{t-4} - \mu_{s_{t-4}}) + \epsilon_t \\ \mu_{s_t} &= -0.448(1 - S_t) + 1.594 S_t \\ \Pr[S_t = 1|S_{t-1} = 1, D_{t-1} = d] &= \frac{\exp(4.305 - 0.243d)}{(1 + \exp(4.305 - 0.234d))} \\ \Pr[S_t = 0|S_{t-1} = 0, D_{t-1} = d] &= \frac{\exp(6.516 - 1.348d)}{(1 + \exp(6.516 - 1.348d))}; \quad \sigma_\epsilon = 0.761;\end{aligned} \tag{2.6}$$

where y_t and S_t are defined as before; d is the duration; D_{t-1} is the number of periods the systems has been in the current state (up to some maximum set to 9).

Previous evaluation of the MS2 model shows that it is able to reproduce the business cycle stylised facts as defined by Hess and Iwata (1997b). However, the model does not account for the asymmetric shape of the cycle (Harding and Pagan, 2001b) and is not much better than an AR(4) using Bayes factors (Koop and Potter, 1999a). The two-regime MS specification is not robust when the estimation period is extended, but a three-regime MS is robust (Boldin, 1996). On the other hand, the MS2 is a better forecaster than a SETAR and a three-regime MS, although is not significantly better than an AR(2) (Clements and Krolzig, 1998). The effect of duration dependence in reproducing business cycle stylised facts is that recessions are more likely to happen, generating longer contractions and shorter expansions (Harding and Pagan, 2001b).

McCulloch and Tsay (1994) relax the assumption of Hamilton's model that the dynamic structure is the same in both states. As for the SETAR models, different coefficients are allowed for each state, so that the economy may have different dynamics depending on the regime. The model is estimated using Gibbs sampling. McCulloch and Tsay specify the following Markov-Switching Autoregressive (MS AR) model for 1947:1-1991:1:

$$y_t = \begin{cases} 0.909 + 0.265\, y_{t-1} + 0.029 y_{t-2} - 0.126\, y_{t-3} - 0.110\, y_{t-4} + \epsilon_{1,t} & \text{if } s_t = 1; \\ -0.460 + 0.216 y_{t-1} + 0.628 y_{t-2} - 0.073 y_{t-3} - 0.097 y_{t-4} + \epsilon_{0,t} & \text{if } s_t = 0; \end{cases} \tag{2.7}$$

$$\Pr[s_t = 0|s_{t-1} = 0] = 0.714; \Pr[s_t = 1|s_{t-1} = 1] = 0.882;$$

$$\sigma_{\epsilon_1} = 0.816; \quad \sigma_{\epsilon_0} = 1.017;$$

where y_t is the first-difference of $100(ln(\text{GNP}))$; $s_t = 1$ defines the expansion regime and $s_t = 0$ defines the contraction regime; $\sigma_{\epsilon_0}, \sigma_{\epsilon_1}$ are the standard deviations of the residuals. In the MS AR, the definition of the regimes is different from the MS2 and the MS DD, but the transition probabilities are similar. The coefficients of the second state are not significant but their inclusion improves the overall fit of the model.

The idea that the US business cycles can be better described by three phases – contraction, high-growth recovery and moderate-growth expansion – is argued by Sichel (1994).

Clements and Krolzig (1998) present a three-regime MS (MS3) model with a transition probability matrix that characterises these phases[8]. The main characteristic of this model is its ability to generate longer expansion durations for the period 1959:2-1996:2 when compared to a two-regime model:

$$\begin{aligned} y_t &= 1.44S_{Ht} + 0.87S_{Mt} - 0.06S_{Lt} + \\ &\quad 0.013y_{t-1} - 0.023y_{t-2} - 0.128y_{t-3} - 0.056y_{t-4} + \\ &\quad 0.636\epsilon_h + 0.343\epsilon_M + 0.879\epsilon_L \\ P &= \begin{bmatrix} 0.91 & 0.09 & 0.00 \\ 0.00 & 0.93 & 0.07 \\ 0.13 & 0.02 & 0.85 \end{bmatrix}; \end{aligned} \tag{2.8}$$

where P is the matrix of p_{ij} transition probabilities $(i, j = H, M, L)$; $S_{Ht} = 1$ defines the high growth regime if $s_t = 1$; $S_{Mt} = 1$ defines the moderate growth regime if $s_t = 2$; $S_{Lt} = 1$ defines the low growth regime if $s_t = 3$. Because $p_{32} \simeq 0$, the economy moves directly from recession to high growth, and, given that $p_{13} = 0$, the economy moves from high growth to moderate growth. Moderate growth is the regime that precedes the recession because $p_{21} = 0$. Although the three-regime MS has a good in-sample fit, it does not produce forecasts significantly better than an autoregressive model (Clements and Krolzig, 1998).

2.2.3 Endogenous Threshold Models

The seminal article employing the Current Depth of Recession (CDR) variable – the gap between the current level of output and the economy's historical maximum – is that of Beaudry and Koop (1993). The motivation for the inclusion of this variable is the possibility that the persistence of shocks can be asymmetric, in other words, the persistence of positive shocks is different from that of negative shocks. Pesaran and Potter (1997) show that the

[8]Boldin (1996) also estimates a three-regime MS model. Hess and Iwata (1997b) includes a three-regime MS in their evaluation, but their specifications assume that the variance does not change across regimes.

Beaudry and Koop model can be interpreted as a threshold model with a large number of regimes, defined in a parsimonious way because of the inclusion of the CDR variable.

Beaudry and Koop (1993) do not present the variance of the residuals of their chosen specification (CDR), so the re-estimation done by Hess and Iwata (1997a) is preferred for the representation of their model. The equation for the period of 1949:1 to 1992:4 is written as:

$$y_t = 0.001 + 0.449y_{t-1} + 0.209y_{t-2} + 0.367CDR_{t-1} + \epsilon_t \tag{2.9}$$

$$CDR_t = \max\{X_{t-j}\}_{j\geq 0} - X_t; \qquad \sigma_\epsilon = 0.00904;$$

where y_t is the first-difference of ln(GDP) and X_t is the ln(GDP). Because of the positive coefficient of the CDR variable, the economic growth is greater when CDR is positive than when it is zero. This means that the growth is faster when the output is below the last peak. So that, positive innovations are more persistent than negative ones, because the model has a mechanism of recovery. Hess and Iwata (1997a) criticise Beaudry and Koop (1993) for testing the significance of the CDR variable using a t-test with an assumed Student's t distribution. Moreover, the results for other countries demonstrate that the fast reversion of the negative shocks and the persistence of the positive ones is not a standard result (Bradley and Jansen, 1997). Jansen and Oh (1999) evaluate the forecasting performance of CDR and STAR (Smooth Transition Autoregressive) models for US output and they conclude that the CDR forecasts encompass the linear and the STAR models. However, the CDR fails to reproduce business cycle stylised facts, because it does not generate enough peaks and the produced contractions are not deep enough (Hess and Iwata, 1997b).

Bradley and Jansen (1997) demonstrate that the persistence gap between positive and negative shocks is smaller when the CDR variable is conditional on the economy being in expansion (output growth positive) or contraction. The specified model (CDR_{pos}) for the

US GDP growth from 1950:1 to 1992:4 is:

$$y_t = 0.003 + 0.289y_{t-1} + 0.163y_{t-2} + 0.376CDR2_{t-1} + \epsilon_t \tag{2.10}$$

$$CDR2_t = \begin{cases} \max\{X_{t-j}\}_{j\geq 0} - X_t \text{ if } y_t \geq 0 \\ 0 \qquad\qquad\qquad\qquad \text{if } y_t < 0 \end{cases} ; \quad \sigma_\epsilon = 0.00886;$$

using the same notation as the CDR model. This specification may indicate three regimes in US GDP because the temporary effect of the negative shocks is stronger during recoveries, just after the trough.

The CDR variable represents a floor effect, but the ceiling can also have a dampening effect over business cycles, which motivates the floor and ceiling model proposed by Pesaran and Potter (1997). An Over-Heating (OH) variable is included in the CDR model; as a consequence, three regimes are defined: floor, ceiling and corridor. The model is estimated in second differences due to the floor and ceiling effects, which does not mean that the series is I(2). The asymmetry found by Beaudry and Koop (1993), that positive shocks are more persistent than negative ones, is only true for the floor regime. An inverse effect is identified in the ceiling regime because the ceiling effect means the economy cannot sustain high growth rates. The Floor and Ceiling (F&C) model is described (1954:1-1992:4) as:

$$\Delta_2 X_t = -0.462\Delta_2 X_{t-1} - 0.862CDR_{t-1} - 0.161OH_{t-1} + h_t V_t \tag{2.11}$$

$$F_t = \begin{cases} 1(Y_t < -0.876) & \text{if } F_{t-1} = 0 \\ 1(CDR_{t-1} + Y_t < 0) & \text{if } F_{t-1} = 1 \end{cases}$$

$$CDR_t = \begin{cases} (Y_t + 0.876)F_t & \text{if } F_{t-1} = 0 \\ (CDR_{t-1} + Y_t)F_t & \text{if } F_{t-1} = 1 \end{cases}$$

$$C_t = 1(F_t = 0)1(Y_t > 0.539)1(Y_{t-1} > 0.539)$$

$$OH_t = (OH_{t-1} + Y_t - 0.539)C_t$$

$$\begin{aligned} COR_t &= 1(F_t + C_t = 0) \\ h_{t+1} &= 0.918COR_t + 1.173F_t + 0.685C_t; \end{aligned}$$

where Y_t is the first-difference of $100(ln(\text{GDP}))$; $\Delta_2 X_t$ is the second-difference of $100(ln(\text{GDP}))$; V_t are the residuals; h_{t+1} is the variance of the residuals; F_t , C_t and COR_t are indexes for defining the regimes; OH_t is the over-heating variable; -0.876 and 0.539 are respectively the thresholds for the floor and ceiling regimes. Hess and Iwata (1997b) argue that because when CDR_t is negative pressure starts to build up for output to return to its previous levels the recessions generated by the F&C model are too shallow compared with US business cycle stylised facts. However, the F&C model is more effective for forecasting periods of negative growth than positive ones, but it is not better than the linear model for two-step ahead forecasting (Pesaran and Potter, 1997).

2.2.4 Threshold Moving Average Models

Asymmetric responses to shocks can be represented in asymmetric moving average (asMa) models, which have moving average coefficients dependent on the last period shock being positive or negative (Brännäs and De Gooijer, 1994). The asMA model is a specific case of the Threshold Moving Average (TMA) model, in which the threshold is equal to zero and the delay is equal to one. Asymmetric moving average models are proposed by Brännäs and De Gooijer (1994) with the linearity test developed by Brännäs, De Gooijer and Teräsvirta (1998). Brännäs and De Luna (1998) and Guay and Scaillet (1999) suggest GMM and indirect inference to estimate, respectively, an asMA and a TMA.

TMA models are estimated to observe the asymmetric response to shocks in the US business cycle by Brännäs and De Gooijer (1994) and Elwood (1998). Working independently and using a similar data set (US GNP), the earlier authors reject the null hypothesis of symmetry while the latter author accepts the null hypothesis of symmetry. This disagreement is based on different specifications and estimation methods. We evaluate both models to observe how the dissimilar specifications affect the ability of the model to reproduce business cycle stylised facts. Guay and Scaillet (1999) present a TMA with threshold different from zero for US GNP, however, their model cannot be evaluated because the variance of the residuals is not published.

The asMA proposed by Brännäs and De Gooijer (1994) estimated with data from 1947:1 to 1984:4 is:

$$\begin{aligned} y_t &= 0.007 + \varepsilon_t \\ &+ \begin{cases} 0.69\hat{\varepsilon}_{t-1} + 0.33\hat{\varepsilon}_{t-2} + 0.22\hat{\varepsilon}_{t-3} - 0.11\hat{\varepsilon}_{t-21} + 1.12\hat{\varepsilon}_{t-22} \text{ if } \hat{\varepsilon}_{t-1} \geq 0 \\ 0.61\hat{\varepsilon}_{t-1} + 0.64\hat{\varepsilon}_{t-2} - 0.07\hat{\varepsilon}_{t-3} + 0.48\hat{\varepsilon}_{t-21} - 0.35\hat{\varepsilon}_{t-22} \text{ if } \hat{\varepsilon}_{t-1} < 0 \end{cases} \\ \hat{\sigma}_\varepsilon &= 0.00824; \end{aligned} \tag{2.12}$$

where y_t is the first-difference of ln(GNP). The asymmetries are observed in the size and sign of the estimated parameters and the asMA fits the data better than a MA(3), but only improves forecasts at long horizons.

Elwood (1998) proposes a TDMA (Threshold-Disturbance Moving Average) model, and because the threshold and the delay are the same as in the asMA model, the Elwood's TDMA is an asymmetric moving average model. The model estimated for 1947:1 to 1989:1 is:

$$y_t = \varepsilon_t + \begin{cases} 0.26\hat{\varepsilon}_{t-1} + 0.39\hat{\varepsilon}_{t-2} \text{ if } \hat{\varepsilon}_{t-1} \geq 0 \\ 0.33\hat{\varepsilon}_{t-1} + 0.16\hat{\varepsilon}_{t-2} \text{ if } \hat{\varepsilon}_{t-1} < 0 \end{cases} \tag{2.13}$$

$$\hat{\sigma}_\varepsilon = 0.00989;$$

where y_t is the first-difference mean-difference of ln(GNP). The dissimilarities of size and sign of the coefficients of each regime are much smaller than in the asMA model, implying that the TDMA does not exhibit significant shock asymmetries, as argued by Elwood.

2.2.5 Structural Models with Markov-Switching

Structural models decompose a time series into components, such as cycle, trend, seasonality and irregularity (Kim and Nelson, 1999c). The inclusion of an unobserved component given by a Markov chain is the contribution of Kim and Nelson (1999a) and Luginbuhl and De Vos (1999)[9].

Kim and Nelson (1999a) aim to test the Friedman "plucking" model that predicts that negative shocks are largely transitory while positive shocks are largely permanent. In so doing, they create a structural model with Markov-switching that decomposes US GDP into a trend and a transitory component. The transitory component comprises two types of shocks: an asymmetric discrete shock, which depends on an unobserved variable S_t that follows a first-order Markov-switching process, and a symmetric shock. The trend component is subject to two shocks: shocks to the level and shocks to the growth rate.

The estimation and testing of the model indicate that only the asymmetric discrete shocks influence the transitory component. In this way, the transitory mechanism is only relevant during recessions (when $S_t = 1$) and high growth recoveries, because the state defined by S_t is still equal to one. During normal times of growth ($S_t = 0$), output is explained mainly by the trend and its shocks. An interesting feature of this model is that the transitory mechanism, in the way that it is defined, is a negative factor and has high resemblance to the CDR variable.

The specification evaluated in this work is model 3, without the symmetric shocks to the transitory component. The State-Space model with Markov-switching (SS MS) for

[9]Dynamic factor models, which are a type of structural models, with an unobserved component given by a Markov-chain are proposed by Chauvet (1998) to model business cycles.

the period of 1951:1-1995:3 is given by

$$y_t = \tau_t + c_t \tag{2.14}$$

$$c_t = 1.25c_{t-1} - 0.47c_{t-2} - 0.011S_t$$

$$\tau_t = g_{t-1} + \tau_{t-1} + v_t \qquad v_t \sim N(0, (0.0061^2(1 - S_t) + 0.0098^2 S_t))$$

$$g = g_{t-1} + w_t \qquad w_t \sim N(0, 0.0007^2)$$

$$Pr[S_t = 1/S_{t-1} = 1] = 0.71; \quad Pr[S_t = 0/S_{t-1} = 0] = 0.92;$$

where y_t is $\ln(GDP)$; τ_t is the trend component; c_t is the transitory component; g_t is the growth rate; and $S_t = 1$ defines the recession regime and $S_t = 0$ defines the expansion regime.

The structural model of Luginbuhl and De Vos (1999) has two main dissimilarities with the SS MS (eq. 2.14): the regime switching is included as a component of the structural model in the transition equations (not as shock generator) and the estimation is executed using Bayesian methods. Only five parameters are estimated, compared to the eight parameters of the SS MS. The measurement equation specifies that GDP is decomposed into a trend and a measurement error component. There are three transition equations. Two for the drift components, which are random walks, and one for the trend, which is a random walk plus a linear combination of the drifts. The determination of the linear combination in the equation of the trend depends on a first-order discrete Markov process.

The Unobserved Component Time Series model (UCTSM) is estimated for the period 1947:1-1998:2 as:

$$\begin{aligned}
y_t &= \mu_t + \epsilon_t & \epsilon_t &\sim NID(0, 0.0059) \qquad (2.15)\\
\mu_t &= \mu_{t-1} + (1 - s_{t-1})\alpha_{t-1} + s_t\beta_{t-1} + \eta_t & \eta_t &\sim NID(0, 0.65)\\
\alpha_t &= \alpha_{t-1} + \zeta_t & \zeta_t &\sim NID(0, 0.002)\\
\beta_t &= \beta_{t-1} + \nu & \nu_t &\sim NID(0, 0.0005)\\
P[s_t &= 0/s_{t-1} = 0] = 0.75; P[s_t = 1/s_{t-1} = 1] = 0.91;
\end{aligned}$$

where y_t is ln(GDP); μ_t is the trend component; α_t and β_t are drift components; and $s_t = 1$ defines the expansion regime and $s_t = 0$ defines the contraction regime.

2.2.6 Models for Italian GDP

Stanca (1999) and Bradley and Jansen (1997) extend the application of the MS, TAR and endogenous threshold models to the Italian output. The Italian GDP presents asymmetries similar to those of US data, which were not found for other countries, such as UK (Mills, 1995) and Japan (Bradley and Jansen, 1997; Peel and Speight, 1998b).

The SETAR$_{IT}$ model tested and estimated by Stanca (1999), for the period of 1960:1 to 1995:4, is represented as:

$$y_t = \begin{cases} 0.38 + 0.37\, y_{t-1} + 0.21 y_{t-2} + 0.00 y_{t-3} - 0.24 y_{t-4} + \epsilon_{1,t} & \text{if } y_{t-3} \leq 1.06; \\ 1.09 + 0.23 y_{t-1} + 0.16 y_{t-2} - 0.02 y_{t-3} - 0.21 y_{t-4} + \epsilon_{2,t} & \text{if } y_{t-3} > 1.06; \end{cases} \tag{2.16}$$

$$n_1 = 81;\ \sigma^2_{\epsilon_1} = 0.57;\ n_2 = 70;\ \sigma^2_{\epsilon_2} = 2.06;$$

where y_t is the first-difference of $100(\ln(\text{GDP}_{IT}))$. The sign and the size of the coefficients are different from the ones for the American economy (eqs. 2.1 and 2.2) and the main difference is that the negative second order autoregressive effect in the contraction is not found.

The Markov-switching autoregressive (MS AR$_{IT}$) model[10], such as Hamilton (1990), for the Italian data is written as:

$$y_t = \begin{cases} 0.54 + 0.48\, y_{t-1} + 0.10 y_{t-2} - 0.06 y_{t-3} - 0.18 y_{t-4} + \epsilon_{1,t} & \text{if } s_t = 0; \\ -0.52 - 0.31 y_{t-1} + 0.24 y_{t-2} + 0.48 y_{t-3} + 1.56 y_{t-4} + \epsilon_{2,t} & \text{if } s_t = 1; \end{cases} \tag{2.17}$$

$$\Pr[s_t = 1 | s_{t-1} = 1] = 0.47; \quad \Pr[s_t = 0 | s_{t-1} = 0] = 0.90;$$

$$\sigma^2_{\epsilon_1} = 0.63;\ \sigma^2_{\epsilon_2} = 0.91;$$

[10]The specification defined by Stanca (1999) is estimated with the EM algorithm with autoregressive coefficients and variances changing between regimes. The specification is different from Hamilton (1989) and Goodwin (1993). The last author was not able to obtain meaningful result with the traditional Hamilton specification for Italian GDP.

where y_t is defined as in the SETAR$_{IT}$; and $s_t = 1$ defines high growth regime and $s_t = 0$ for defines the low growth regime.

The inclusion of the CDR variable in an AR(2) model for the Italian GDP significantly improves the fit compared with the linear model (Bradley and Jansen, 1997). First, Bradley and Jansen estimate an endogenous threshold model similar to the Beaudry and Koop (1993) specification of the US data. The results indicate that for large negative shocks, the persistence is smaller than for positive or smaller shocks. Specifically, the observation of the impulse-response indicates a strong recovery in which a –2% shock is converted to +1.4% after 12 quarters. The nature of this asymmetry is similar to that found for the US data. The second specification of the model, employing the CDR variable depending on whether the economy is in contraction (CDR$_{neg}$), is written, for the period of 1960:1 to 1992:1, as:

$$
\begin{aligned}
y_t &= 0.0041 + 0.275 y_{t-1} + 0.152 y_{t-2} + 0.353 CDR1_{t-5} + \epsilon_t \qquad (2.18)\\
CDR1_t &= \begin{cases} \max\{X_{t-j}\}_{j\geq 0} - X_t & \text{if } y_t < 0 \\ 0 & \text{if } y_t \geq 0 \end{cases} \quad ; \quad \sigma = 0.8527;
\end{aligned}
$$

where y_t is the first-difference of $100(ln(\text{GDP}_{IT}))$ and X_t is the $ln(\text{GDP}_{IT})$. In the latter model, the CDR variable is only relevant when the economy is in contraction, in contrast with the CDR$_{pos}$ (eq. 2.10) for US data. The interpretation by the authors of these results is that "negative shocks will be less persistent than positive shocks and this effect occurs due to the positive, offsetting response of output growth to recessions. With five lags, it is solely a peak-to-trough phenomenon" (Bradley and Jansen, 1997, p. 505).

2.2.7 Models for Australian GDP

The evidence of asymmetry in the Australian output is found by Bodman (1998), applying a MS and a MS with duration dependence[11]. CDR, three-regime MS and floor and ceiling specifications were presented by Bodman and Crosby (1998). However, all these spec-

[11]Layton (1994) also estimated a MS model for the Non-farming Australian GDP.

ifications present smaller likelihood values than the two-regime MS model and, consequently, they are rejected by the authors as a reasonable representation of the data.

Bodman (1998) estimates a MS model for Australian GDP, which shows asymmetric behaviour similar to US data (eq. 2.5). The two-regime MS ($MS2_{AU}$) model for period 1960:1 to 1997:3 is:

$$\begin{aligned} (y_t - \mu_{s_t}) &= -0.066(y_{t-1} - \mu_{s_{t-1}}) + 0.105(y_{t-2} - \mu_{s_{t-2}}) \\ &\quad -0.48(y_{t-3} - \mu_{s_{t-3}}) - 0.182(y_{t-4} - \mu_{s_{t-4}}) + \epsilon_t \\ \mu_{s_t} &= -0.437(1 - S_t) + 1.349 S_t \\ \Pr[S_t &= 1|S_{t-1} = 1] = 0.94; \quad \Pr[S_t = 0|S_{t-1} = 0] = 0.73; \quad \sigma_\epsilon = 1.074; \end{aligned} \tag{2.19}$$

where y_t is the first-difference of $100(ln(\text{GDP}_{AU})$; and $S_t = 1$ defines the expansion regime and $S_t = 0$ defines the recession regime. Only the autoregressive coefficient of order four is significant. The transition probabilities are slightly different from the MS2 model (eq. 2.5) implying longer duration of expansions compared to US data.

The inclusion of duration dependence is described in the following specification (MS DD_{AU}):

$$\begin{aligned} (y_t - \mu_{s_t}) &= 0.096(y_{t-1} - \mu_{s_{t-1}}) - 0.071(y_{t-2} - \mu_{s_{t-2}}) \\ &\quad -0.158(y_{t-3} - \mu_{s_{t-3}}) - 0.211(y_{t-4} - \mu_{s_{t-4}}) + \epsilon_t \\ \mu_{s_t} &= -0.489(1 - S_t) + 1.574 S_t \\ \Pr[S_t &= 1|S_{t-1} = 1, D_{t-1} = d] = \frac{\exp(5.028 - 0.324d)}{(1 + \exp(5.028 - 0.324d))} \\ \Pr[S_t &= 0|S_{t-1} = 0, D_{t-1} = d] = \frac{\exp(6.841 - 1.567d)}{(1 + \exp(6.841 - 1.567d))}; \quad \sigma_\epsilon = 0.705; \end{aligned} \tag{2.20}$$

where y_t and S_t are defined as before; d is the duration; D_{t-1} is the number of periods the systems has been in the current state (up to some maximum set to 9).

The results are similar to the ones of the MS DD (eq. 2.6) for US, including duration dependence only during recessions. When the economy has been in recession for more than

4 quarters, there is a probability of 50% that this recession will finish, compared to the probability of 0.01% of moving out of a recession in the next quarter after entering it.

Summarising, we review the 20 univariate non-linear time series models that are evaluated in section 2.4, including the results of previous evaluations of the models.

2.3 Evaluation Method

The non-linear time series models will be evaluated based on their abilities to reproduce business cycle features. This type of evaluation method allows the ability of the models to generate asymmetries found in the business cycle to be observed: asymmetric durations between contractions and expansions and asymmetric shape. The evaluation is complementary to other methods, such as forecasting analysis (e.g., Clements and Krolzig, 1998), impulse responses (e.g., Van Dijk et al., 2000), Bayes factor (e.g., Koop and Potter, 1999a) and conditional means (e.g., Breunig and Pagan, 2001 and section 2.5). Similar methods have also been applied to evaluate dynamic general equilibrium models, mainly Real Business Cycles, by King and Plosser (1994), Simkins (1994) and Harding and Pagan (2000).

2.3.1 Definition of Stylised Facts

The first point for the delineation of the evaluation method is the definition of business cycles. The business cycle is a pattern found in aggregate economic activity in many capitalist economies. The Burns and Mitchell (1946) approach for the determination of such a pattern in a time series starts with the location of turning points and, based on that, the calculation of features, such as durations and amplitudes of the phases defined by those turning points. This work employs an algorithm for the determination of the turning points. In addition, the measure of aggregate economic activity is given by GDP (or GNP) because it is a good measure that was not available in a satisfactory form during the working period of Burns and Mitchell. In this way, the reference cycle definition employed does not consider comovements over many series for the definition of the cycle, which is the case of dynamic factor models (e.g., Stock and Watson, 1989). The use of output for determining the reference cycle was also applied by Simkins (1994), King and Plosser (1994), Hess and Iwata (1997b), Canova (1999) and Harding and Pagan (2001b).

Another important point in the definition of the reference cycle is the presence of a previous de-trending process, i.e., whether the classical cycle or the growth cycle is the object of study. Some authors (Goodwin, 1993; Simkins, 1994; Canova, 1999) prefer to analyse growth cycles arguing that they are more frequent and relevant for many OECD economies than classical cycles and that many theoretical and empirical models are applicable only to detrended data. However, Simkins' (1994) calculation of duration of the GNP growth business cycle demonstrates that the durations of expansions and recessions are almost the same, while the NBER results for classical cycles indicate that the durations of contractions are shorter than the ones for expansions. Moreover, different methods for de-trending can generate different reference cycles as demonstrated by Canova (1998), who concludes that there is no best de-trending method because each method can show different features of the data. We follow the NBER and the majority of business cycle analysts in adopting the classical definition of business cycle. The fact that all the models that will be evaluated are based on series in differences is not an obstacle to employing this definition[12].

Because the pattern of the cycle is defined by its turning points, the algorithm for obtaining these points has to be defined. The turning point definitions applied by the NBER, based on the Burns and Mitchell tradition, include *ad hoc* procedures. The best algorithm for reproducing NBER turning points is that of Bry and Boschan (1971). It uses three different ways of smoothing the series before the identification of turning points and

[12]The data simulated from the models are in growth rates, which are composed in log-levels before the application of the algorithm for turning point location, except for the structural models. This does not affect the results because the algorithm employs the differences of the series to locate turning points. In addition, Harding and Pagan (2001a) show that the dating generated by the MS model is for a classical cycle.

censoring procedure. Watson (1994), Simkins (1994), King and Plosser (1994) and Harding and Pagan (2001b) applied extensions and modifications of the algorithm of Bry and Boschan for the definition of turning points. The algorithm includes censoring rules to make sure that the cycles and the phases alternate and have a minimum length. Canova (1999) uses simple rules without any censoring. The problem is that these rules can identify too many cycles when compared to the algorithm of Bry and Boschan because of the lack of censoring (rules for minimum phase length, for example). This lack of robustness is of greater concern when the algorithm is also employed to data simulated from econometric models, even though the lack of censoring does not affect the location of peaks and troughs in the US GDP (Harding and Pagan, 2001a).

The algorithm employed in this work is the Harding and Pagan version of the algorithm of Bry and Boschan for quarterly data (Harding and Pagan, 2001b). The first step is the determination of the potential set of turning points, employing the rule $\{\Delta_2 y_t > 0, \Delta y_t > 0, \Delta y_{t+1} < 0, \Delta_2 y_{t+2} < 0\}$ for classifying a peak at time t (opposite inequalities for a trough). The second step is to ensure that peaks and troughs alternate. The third step requires that the phases are at least 2 quarters long and that the cycles have a length of at least 5 quarters. The absence of smoothing is justified by the fact that quarterly series are already smooth enough[13]. The performance of this algorithm is compared with the FIBCR (Foundation of International Business Cycle Research) turning points by Harding and Pagan (2001b), who show that all turning points are identified and the majority of them match the FIBCR. A good performance is also obtained when compared to the NBER turning points[14].

After determining the cycle turning points, the features of the cycle, which are

[13]The version by King and Plosser (1994) of the Bry and Boschan algorithm with smoothing procedure is applied to quarterly series, however the data evaluated is a monthly decomposition of the quarterly data. In the case of Simkins (1994), the quarterly data is smoothed for the determination of turning points, but this creates problem for identifying one of the troughs.

[14]For post-war data, there are 9 NBER whole cycles. The QBB (Quarterly Bry and Boschan) identifies perfectly five troughs and five peaks. The other four present small differences as one quarter before or after, with exception of the 1949:4 and 1970:1 troughs (two and three quarters before, respectively). See also Harding and Pagan (2001a).

to be considered as stylised facts, have to be chosen. Based on the Burns and Mitchell tradition, comparisons between the reference cycle and the specific cycle using the nine-point graph definition of the cycle are employed by Simkins (1994), King and Plosser (1994), and Balke and Wynne (1995). Another option is the calculation of the average duration and the average amplitude of each phase, as employed by Watson (1994), Hess and Iwata (1997b) and Harding and Pagan (2001b). Neither possibility considers that the differences between cycles may be relevant. The nine-point graph approach, for example, defines the same number of points for expansions and contractions which makes it difficult to characterise the asymmetric duration between the phases[15]. The calculation of the average duration and amplitude allows the characterisation of asymmetry between contractions and expansions, though does not measure dissimilarities between cycles.

Therefore, besides the average amplitude and the average duration of each phase, we employ the cumulative losses (gains) in output from peak (trough) to trough (peak) relative to the previous peak (trough) and the average excess of the cumulative movements over the triangle measure ($0.5(duration * amplitude)$) as business cycle stylised facts (Harding and Pagan, 2001b). The excess is a relevant measure of the shape of the phases over the business cycle. In Figures 2.1, 2.2 and 2.3, the loss (gain) against the last peak (trough) is presented for US, Italy and Australia. The figures show that the growth rate seems to be changing inside a phase. In the case of the US expansion of 1983 to 1990, for example, the shape of the curve is steep (high growth rate) in the beginning of the phase and is smoother after 1984 (lower growth rates). Because most of the expansion phases have a concave shape, the average excess value is positive: 1.1 for US, 1 for Australia and 0.90 for Italy. However, the short contraction phases are better approximated by the triangle, given that the average of the excess of contractions is near zero.

[15] King and Plosser (1994) and Balke and Wynne (1995) present nine-point graphs with the distance between the points given by the average duration of the phase. Simkins (1994) follows the classical approach.

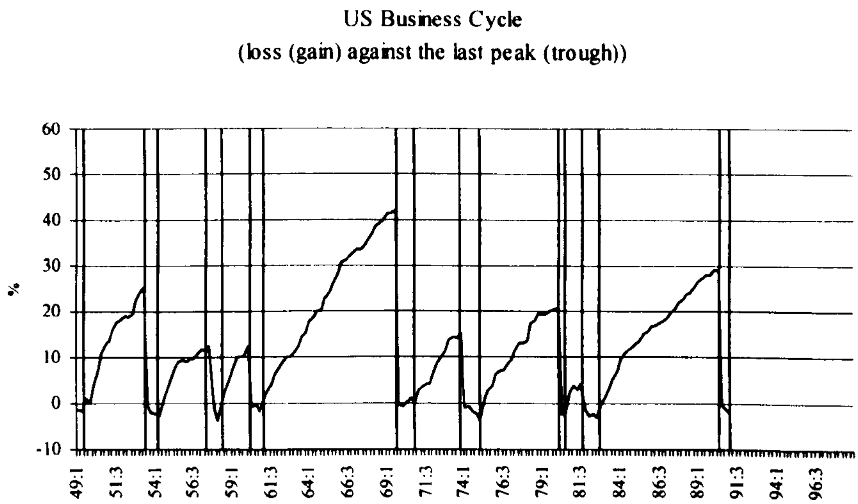

Figure 2.1: Cumulative gains (losses) against the last though (peak) for US GDP (dashed lines are NBER peaks and troughs)

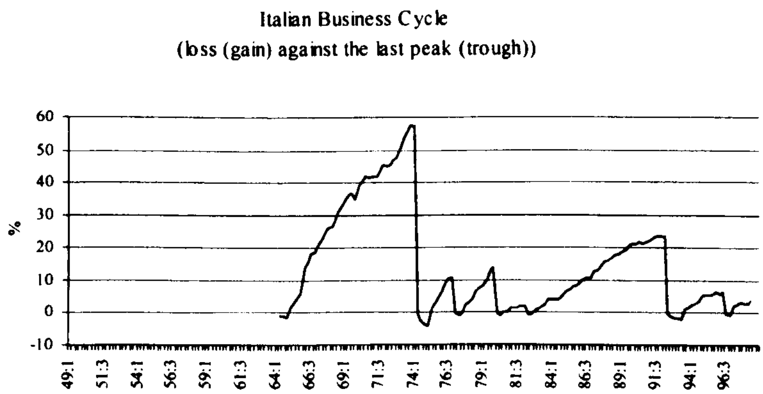

Figure 2.2: Cumulative gains (losses) against the last though (peak) for Italian GDP

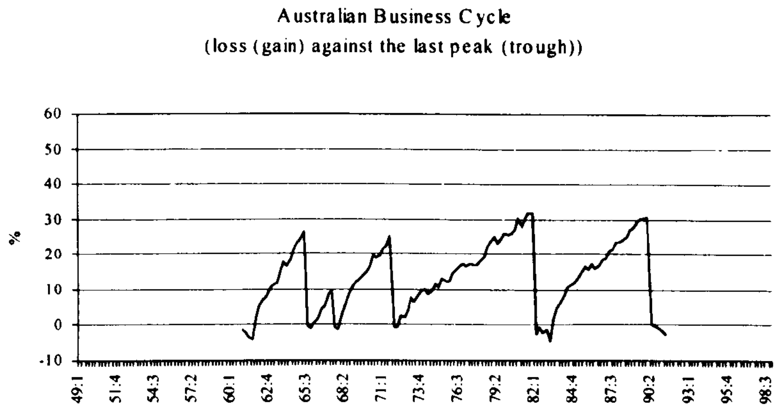

Figure 2.3: Cumulative gains (losses) against the last though (peak) for Australian GDP

An important characteristic of the cycle, observed in Figure 2.1, which is not evaluated employing the average duration and the average amplitude, is the dissimilarities between cycles. Business cycles are not alike, therefore, the lower and the upper quartiles of each stylised fact over the cycles are computed to evaluate the dispersion of the stylised facts. These quartiles are relatively robust to skewness and kurtosis of the distribution of the stylised facts over the cycle.

2.3.2 Measures of Fit

To calculate the ability of a model to reproduce the business cycle features, the model-data is obtained employing a Monte Carlo simulation. 5000 samples of length equal to the sample size in which the model was estimated are generated using pseudo-random numbers drawn from a normal distribution with variance equal to the model residuals[16]. The observed-data stylised facts can be different for each evaluated model because they are computed for the sample size employed to estimate the model[17]. However, there is no need for

[16]We actually generate at least 20 more observations than the sample size to allow the first 20 being discharged. This permits the simulated values to converge to their conditional mean, given that the initial values are zero.

[17]However, the values do not change a lot among samples because the computation of turning points identifies only complete phases. In the case of US, for example, 9 samples have very similar average durations (3 for contractions and 18 for expansions), four samples (using data until 1988) have slightly smaller duration of expansions (16) and two other samples have slightly larger duration of expansions (21) (no change for the

re-estimation and re-specification of the models and the information employed for evaluation is the same as that available for the econometrician at the time of the estimation. The estimated parameters for each model are assumed constant for the simulation procedure[18]. For each Monte Carlo generated sample, the quarterly Bry and Boschan (QBB) algorithm is applied for location of the turning points and, subsequently, for the computation of the stylised facts. The values reported are the mean across samples of the average stylised facts and stylised fact quartiles.

The empirical distribution of the stylised facts generated by the Monte Carlo experiment is employed for the comparison of the observed-data with the model-data. If X is a (nxd) matrix, where n is number of Monte Carlo replications and d is the number of stylised facts, and D is a $(d\mathrm{x}1)$ vector of the observed-data stylised facts, the empirical probability that the model-data stylised facts is bigger than the correspondent observed-data value is

$$p_j = \Pr[X_{ij} > D_j'], \text{ where } j = 1, \dots, d. \tag{2.21}$$

Employing the empirical probabilities, the observed-data stylised fact will be rejected by the empirical distribution of the model-data when p_j is smaller than 0.10 or greater than 0.90. This implies an 80% confidence interval. The confidence interval employed is relatively narrow because models with larger variances of the residuals may create empirical distributions with high dispersions, which implies a higher probability that the observed-data values lie within the interval. The comparison between the confidence intervals, which are defined by the simulated distributions of the stylised facts, and the business cycle stylised facts is similar to the comparison of realisations with forecast intervals (Christoffersen, 1998).

contraction values). In general, sample periods that start before 1951 and finish after 1991 have very similar stylised facts; different data lengths generate only small changes even for the measures of shape of the cycle.

[18]In this point, one could argue that constant parameters mean that we are not accounting for parameter uncertainty that could influence the results. Fair (1993, p. 162) reports Chow argument: "Although coefficient estimates are uncertain, the true coefficients are fixed. In the real world, the reason that economic events are stochastic is because of the stochastic shocks (error terms), not because the coefficients are stochastic". Because the estimated coefficients of the analysed models are not time-varying and the uncertainties over these coefficients are smaller than the ones in the residuals, we think that ignoring parameter uncertainty does not weaken our results.

The decision concerning the degree of interval coverage may influence the conclusions about the accuracy of the model in reproducing the stylised facts. Based on the calculated p_j, we also analyse the robustness of the results for the assumption of, say, 50% coverage.

The disadvantage of applying the empirical distribution for inference is the absence of statistical analysis for the whole of the stylised facts. A Chi-squared test for determined features is employed for standing comparison between groups of observed-data stylised facts and model-data stylised facts. The joint empirical distribution is assumed to be asymptotically normal with means $\boldsymbol{\mu}$ (dx1) and variance-covariance matrix Σ (dxd). The null hypothesis that the means of the model-data stylised fact empirical distributions are equal to the observed-data stylised facts ($\boldsymbol{\mu} = \mathbf{D}$) is tested employing the Q statistic (Simkins, 1994; Hess and Iwata, 1997b)

$$Q = (\mathbf{D} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{D} - \boldsymbol{\mu}) \sim \chi^2(d) \tag{2.22}$$

The main restriction of this statistic is the supposition of an asymptotically normal distribution for the model-data stylised fact empirical distributions; however, Simkins (1994) and Hess and Iwata (1997b) computed the empirical distribution of the statistic and they confirmed that the χ^2 distribution is a good approximation. Because the Q statistic depends on the variance of the simulated data, statistics calculated for unstable models with high kurtosis in the distribution of stylised facts can accept the null hypothesis more times than robust models[19]. This problem is derived from the large residual variances of the estimated model.

Summarising, the evaluation method considers turning points defined by the QBB algorithm, which are employed to calculate the business cycle stylised facts. These stylised facts are computed for the data simulated from the models and for the observed data. They

[19]Note that the variances of each empirical distribution can also be also employed for the calculation of the Monte Carlo error for the estimation of the mean, i.e., $\sqrt{Var(X)/n}$. This error is very small for the majority of the simulations because $n = 5000$. Therefore, for this analysis, we are ignoring the Monte Carlo uncertainty.

are compared using empirical confidence intervals and a test for accuracy in generating the business cycle stylised facts.

2.4 Results of the Evaluation

2.4.1 Evaluation of TAR Models

Table 2.1 presents the results for the TAR models[20]. The two-regime SETAR (eq. 2.1) generates contractions that are longer and deeper than the US business cycle, although the (dur, amp, cum) Q test for both phases supports the model. The evaluation of Hess and Iwata (1997b) has accepted that the SETAR model (Potter, 1995) reproduces the stylised facts, because they apply a different turning point algorithm. On the other hand, the two-regime SETAR model (eq. 2.2) has a better performance. The agreement between model and data is confirmed by the empirical distributions for means and dispersions in both phases, except for the values of the excess of expansion. The good performance of the specification of Peel and Speight (1998b) compared to the one of Potter (1995) supports the argument of Pagan (1997) that the presence of a second order autoregressive coefficient, or the automatic stabiliser, is not necessary to the description of the duration and the amplitude of the cycle found in the US data.

The four-regime SETAR (eq. 2.3) gives a good performance when the Q test statistics are observed. However, the inclusion of more regimes does not make the four-regime SETAR better than the two-regime SETAR (eq. 2.2) as one would expect, even though it is better than the Potter's specification.

On other hand, the MRSTAR model (eq. 2.4) reproduces the majority of stylised facts, but the evaluation of the quartiles indicates that there are too many small simulated

[20]The computations of this chapter make use of GAUSS programing language with codes written by the author, except for the conditional means which are computed using SPLUS 2000. I am grateful to Don Harding for his codes for QBB algorithm and for the simulation of the MS2 and the MS DD employed in Harding and Pagan (2001b).

Table 2.1: Business cycle stylised facts of TAR models

	Phase	US GNP 47:1-91:1	SETAR 2 (eq. 2.1) 47:1-92:4	SETAR 2 (eq. 2.2) 57:1-91:3	SETAR 4 (eq. 2.3) 47:1-91:1	MRSTAR (eq. 2.4) 47:1-95:2
Stylised Facts						
Dur	PT	3.25	5.73* (1.00)	3.91 (0.71)	4.59* (0.92)	3.17 (0.42)
Amp		2.59	4.49* (0.98)	2.38 (0.41)	3.22 (0.80)	1.84 (0.12)
Cum		4.58	16.5* (0.99)	5.98 (0.55)	8.39 (0.82)	4.22 (0.35)
Exc		-0.07	0.22* (0.95)	0.03 (0.79)	0.19* (0.95)	-0.01 (0.80)
Dur	TP	17.5	14.8 (0.20)	19.9 (0.33)	19.4 (0.50)	17.9 (0.41)
Amp		8.79	14.6 (0.13)	19.3 (0.26)	22.0 (0.48)	20.45 (0.40)
Cum		255	192.2 (0.20)	352.9 (0.31)	395.5 (0.54)	371.4 (0.50)
Exc		1.25	0.22* (0.04)	0.20* (0.09)	0.31* (0.08)	-0.01*(0.01)
Quartiles						
Dur	PT	2 4	3.29 7.32* (0.78) (0.92)	2.54 4.79 (0.38) (0.45)	2.90 5.73 (0.67) (0.70)	1.97 3.95 (0.09) (0.27)
Amp		1.75 3.18	2.35 6.03* (0.69) (0.97)	1.38 3.15 (0.26) (0.41)	2.05 4.13 (0.66) (0.78)	0.90* 2.39 (0.02) (0.16)
Cum		3.29 6.58	3.88 20.29* (0.76) (0.93)	1.98 7.81 (0.36) (0.33)	2.76 10.7 (0.25) (0.66)	0.90* 4.90 (0.02) (0.20)
Exc		-0.33 0.23	-0.09 0.50 (0.89) (0.81)	-0.13* 0.19 (0.95) (0.34)	-0.05* 0.40 (0.94)(0.78)	-0.10* 0.10* (0.97) (0.08)
Dur	TP	13 20	7.03* 19.5 (0.05) (0.34)	10.51 26.5 (0.17) (0.25)	9.22 25.79 (0.16) (0.60)	7.11* 24.36 (0.08) (0.55)
Amp		12.63 25.34	5.96* 20.07 (0.06) (0.21)	9.54 26.09 (0.14) (0.30)	9.68 29.7 (0.23) (0.54)	6.64 28.45 (0.12) (0.48)
Cum		103.1 245.4	28.71* 233.3 (0.03) (0.30)	100.1 488.8 (0.19) (0.24)	72.9 500.1 (0.16) (0.60)	39.9* 437.1 (0.01) (0.53)
Exc		0.53 1.92	-0.32* 0.73* (0.01) (0.05)	-0.36* 0.76* (0.07) (0.09)	-0.39* 0.99 (0.05) (0.12)	-0.55* 0.55* (0.01) (0.03)
Q test						
(dur,amp)	PT		3.82 (0.15)	0.80 (0.67)	1.40 (0.62)	2.77 (0.25)
(dur,amp,cum)			5.29 (0.15)	0.83 (0.84)	1.80 (0.62)	5.11 (0.16)
(dur,amp)	TP		2.20 (0.33)	0.56 (0.75)	0.10 (0.95)	0.03 (0.99)
(dur,amp,cum)			3.00 (0.39)	0.94 (0.81)	1.4 (0.50)	0.37 (0.95)
(excess)	PT+TP		5.98 (0.05)	1.63 (0.44)	5.01 (0.08)	6.08 (0.05)
(excess)	TP		3.41 (0.06)	1.27 (0.26)	1.95 (0.16)	5.52 (0.02)
(dur)	PT+TP		4.11 (0.13)	0.35 (0.84)	1.42 (0.50)	0.01 (0.99)
(dur,amp,cum)	PT+TP		8.4 0(0.21)	1.75 (0.94)	1.97 (0.93)	5.46 (0.49)

* observed-data stylised fact is outside a 80% empirical confidence interval.

Note: The GNP (or GDP) stylised facts are calculated using the QBB algorithm for locating turning points to the sample indicated for each model. The first column shows stylised facts for one of the models samples just for helping the understanding of the results. For each replication (5000) of the simulation, the turning points and the stylised facts are calculated. The values reported are the average over these replications. The stylised facts are averages of duration (dur), amplitude (amp), cumulative loss (gain) and excess of the cumulative measure over the triangle approximation (exc). The dispersion (across cycles) of these stylised facts is measured by the lower and the upper quartiles. The amplitude, the cumulative and the excess are in percentage. The values in parenthesis are the proportion in the empirical distribution of values bigger than the observed-data stylised fact. The Q test verify whether the model empirical distribution can generate jointly the indicated stylised facts, with associated p-values in parentheses. PT: Peak-Trough and TP: Trough-Peak.

values compared to the observed-data values. The Q test p-value of 0.99 for the duration of contractions and expansions, which indicates a "perfect" match to the observed-value, is contested by the distribution of the duration of the expansion that generates a large number of cycles with short duration. The disagreement between the statistics may be caused by the lack of robustness of the Q statistic. In addition, this model has average excess equal to zero and generated by a symmetric distribution, which are not characteristic of US business cycles.

Using a confidence interval with smaller coverage (50%), the MRSTAR can account for 5 out of 8 stylised facts, and in the rank follows the two-regime SETAR of Peel and Speight (1998b), and the four-regime SETAR. When quartiles are evaluated with this smaller coverage, however, the SETAR models have better performance than the MRSTAR. In general, the MRSTAR generate stylised facts with higher dispersion than the US data.

2.4.2 Evaluation of Markov-Switching Models

MS models generate data that reproduce most of the business cycle stylised facts, as presented in Table 2.2. The two-state MS model (eq. 2.5) presents statistical accuracy for generating business cycles, except for the shape of the cycle. The inclusion of duration dependence worsens the model performance, generating contractions that are longer than those in the data. These results are also found by Harding and Pagan (2001b). The MS DD is equivalent to the two-state MS model while inaccuracy of generating the shape of the cycle is concerned.

The presence of different autoregressive dynamics between the states (MS AR, eq. 2.7) improves the performance of generating the asymmetric shape of the cycle compared with the MS model. The calculated value of the excess of expansions is now positive; however, the excess of contractions is also positive, which is not a characteristic of the US data.

The three-regime Markov switching (eq. 2.8) accounts for the expansion data better

Table 2.2: Business cycle stylised facts of Markov-switching models

	Phase	US GNP 51:2-84:4	MS2 (eq. 2.5) 51:1-84:4	MS DD (eq. 2.6) 51:1-84:4	MS AR (eq. 2.7) 47:1-91:1	MS3 (eq. 2.8) 59:2-96:2
Stylised Facts						
Dur	PT	3.14	4.08 (0.80)	4.51* (0.98)	4.43 (0.88)	4.02 (0.72)
Amp		2.71	2.80 (0.53)	3.28 (0.85)	4.36 (0.86)	2.48 (0.48)
Cum		4.6	7.11 (0.71)	7.91* (0.95)	14.6 (0.89)	6.21 (0.56)
Exc		-0.05	0.01 (0.66)	0.01 (0.63)	0.20* (0.91)	-0.00 (0.82)
Dur	TP	15.8	18.5 (0.53)	16.4 (0.47)	16.23 (0.32)	27.4 (0.60)
Amp		18.01	25.2 (0.63)	24.4 (0.82)	17.0 (0.22)	24.3 (0.45)
Cum		219.7	427.6 (0.61)	282.2 (0.51)	232.1 (0.29)	638.9 (0.57)
Exc		0.86	-0.01* (0.06)	0.01* (0.05)	0.24* (0.01)	0.74 (0.32)
Quartiles						
Dur	PT	2 4	2.64 4.99 (0.46) (0.51)	3.43 5.50 (0.84) (0.85)	2.55 5.51 (0.44) (0.65)	2.7 4.9 (0.46) (0.47)
Amp		1.71 4.02	1.84 3.59 (0.53) (0.31)	2.45 4.06 (0.86) (0.50)	1.55 5.74 (0.30) (0.80)	1.68 3.11 (0.37) (0.43)
Cum		1.88 7.51	2.69 9.24 (0.59) (0.49)	4.51* 16.54 (0.92) (0.81)	2.17 16.19 (0.17) (0.75)	2.58 8.31 (0.33) (0.38)
Exc		-0.47 0.39	-0.16* 0.17 (0.95) (0.10)	-0.20* 0.20 (0.92) (0.16)	-0.11* 0.36 (0.94) (0.58)	-0.15* 0.13 (0.93) (0.72)
Dur	TP	8 20	9.39 25.15 (0.38) (0.57)	10.72 20.58 (0.70) (0.39)	8.04* 21.36 (0.08) (0.14)	14.3 36.0 (0.31) (0.50)
Amp		12.39 21.45	12.29 34.68 (0.34) (0.73)	15.56 30.79 (0.66) (0.84)	8.63 22.40 (0.12) (0.30)	12.11 32.61 (0.25) (0.49)
Cum		52.06 243.39	121.3 601.5 (0.35) (0.73)	98.5 366.1 (0.67) (0.55)	44.7* 286.8 (0.08) (0.39)	238.4 917.7 (0.32) (0.48)
Exc		0.31 1.82	-0.57* 0.56* (0.06) (0.06)	-0.59* 0.59* (0.07) (0.05)	-0.29* 0.76* (0.02) (0.03)	0.13 1.30 (0.20) (0.25)
Q test						
(dur,amp)	PT		1.00 (0.61)	3.82 (0.15)	1.12 (0.57)	0.65 (0.72)
(dur,amp,cum)			1.00 (0.80)	2.89 (0.27)	1.19 (0.76)	0.65 (0.88)
(dur,amp)	TP		2.82 (0.24)	12.81 (0.00)	1.99 (0.37)	1.23 (0.54)
(dur,amp,cum)			2.82 (0.42)	12.99 (0.01)	2.27 (0.52)	1.46 (0.69)
(excess)	PT+TP		2.07 (0.35)	2.86 (0.24)	7.57 (0.02)	0.60 (0.74)
(excess)	TP		1.81 (0.17)	2.53 (0.11)	6.06 (0.01)	0.08 (0.77)
(dur)	PT+TP		0.72 (0.70)	3.83 (0.15)	1.15 (0.56)	0.51 (0.78)
(dur,amp,cum)	PT+TP		3.85 (0.70)	13.79(0.03)	3.64 (0.73)	2.14 (0.90)

See notes of Table 2.1.

than the two-regime MS, using the Q test p-values. This favours the argument that the inclusion of one more regime improves the reproduction of strong recovery periods, using the definitions of Sichel (1994). The analysis of the confidence interval for the quartiles also indicates that the three-regime MS model is able to account for the observed data stylised facts fairly well[21]. The results of Hess and Iwata (1997b) also imply that the two-regime MS and the three-regime MS reproduce the observed-data stylised facts. However, their three-regime specification, which does not consider switching in the variance and has an autoregressive order equal to one, has a worse performance than the two-regime specification.

Another result shown in Table 2.2 is that only the MS3 model is able to generate the asymmetric shape of the cycles. The division of the expansion phase between high growth and moderate growth may explain the results for the three-regime MS. Because the high growth regime starts the expansion period and then is followed by the moderate growth, the rate of growth within the expansion phase is not the same, as the positive excess of observed-data indicates. However, when a confidence interval of smaller coverage is assumed, the three-regime MS cannot generate the excess of contractions, including the respective quartiles. In addition, when a 50% coverage is employed, the two-regime MS is superior in generating stylised facts and quartiles than the MS with duration dependence and the MS AR.

2.4.3 Evaluation of Endogenous Threshold Models

The CDR model (eq. 2.9) cannot account for the majority of the lower quartiles for both phases, although its performance is better when only durations are evaluated, showing that the model has difficulty in reproducing the amplitude of the cycle (Table 2.3). The model does not explain the asymmetric shape of the cycle, but the generated values are in general positive for the excess of expansions. When CDR is taken as conditional on the economy's growth, the new specification (eq. 2.10) improves the model-data match. However, the model

[21]The empirical densities of expansions are highly skewed. This can explain why the reported means of expansion (Table 2.2) seem too large compared to the values for the observed data.

simulation continues to generate too many small amplitude values for the lower quartiles. The fact that the CDR model generates shallow contractions has also been reported by Hess and Iwata (1997b). In addition, this evaluation shows that this problem also occurs in the expansion phase.

The inclusion of the over-heating variable improves the ability of the simulated model to reproduce the stylised facts. The presence of the three regimes, as in the three-state MS model, is important in reproducing the shape of the expansions. However, the model also generates positive excess values for the contractions, while the corresponding observed-data value is negative.

2.4.4 Evaluation of Threshold Moving Average Models

As observed in Table 2.4, both threshold moving average models have good performance in generating the business cycle stylised facts[22]. Comparing the models, the asMA model is better able to reproduce the contraction quartiles while the TDMA model is better for the expansion quartiles. Thus, some of the contractions generated by the TDMA model are too shallow while some of the expansions of the asMA model have low amplitude. Neither model can generate the shape of the expansions. Even though the asMA is able to generate asymmetric response to shocks, it is not able to generate the asymmetric shape of the cycle.

2.4.5 Evaluation of Structural Models with Markov-Switching

Initial values for some variables are needed for the simulation of the structural models because the models included the trend as part of the data generation process. For the UCTSM (eq. 2.15), the values of the priors of μ_t, α_t and β_t are employed as initial values (μ_0, α_0 and β_0). For the SS MS model (eq. 2.14), the initial values are based on the US GDP for the sample size of the estimation: $\tau_0 = 7.4$ and $g_0 = 0.0081$. The initial value for

[22]The data generated from the TDMA model use the mean growth of the sample (0.0086) as drift, given that the model was estimated with mean-differenced data.

Table 2.3: Business cycle stylised facts of endogenous threshold models

	Phase	US GDP 49:1-92:4	CDR (eq. 2.9) 49:1-92:4	CDR_{pos} (eq. 2.10) 50:1-92:4	F&C (eq. 2.11) 54:1-92:4
Stylised Facts					
Dur	PT	3.13	3.34 (0.63)	3.53 (0.73)	4.33* (0.99)
Amp		2.6	1.81* (0.02)	2.03 (0.16)	2.91 (0.74)
Cum		4.36	3.12* (0.09)	4.37 (0.39)	6.07 (0.86)
Exc		-0.06	0.04 (0.89)	0.03 (0.83)	0.2* (0.96)
Dur	TP	17.75	14.79 (0.19)	17.1 (0.34)	14.3 (0.13)
Amp		20.19	14.31 (0.11)	15.97 (0.21)	14.9 (0.11)
Cum		255.7	187.58 (0.19)	236.94 (0.28)	179.4 (0.13)
Exc		1.06	0.23* (0.04)	0.22* (0.04)	0.5 (0.14)
Quartiles					
Dur	PT	2 4	2.30 3.99 (0.28) (0.24)	2.95 6.31* (0.28) (0.38)	2.90 5.27 (0.69) (0.69)
Amp		2.03 3.01	1.17* 2.35 (0.02) (0.10)	1.68* 4.60 (0.03) (0.27)	1.91 3.81 (0.54) (0.50)
Cum		3.01 6.44	1.43* 4.23* (0.03) (0.09)	2.61* 13.78 (0.04) (0.27)	2.79 8.16 (0.66) (0.51)
Exc		-0.31 0.24	-0.10* 0.16 (0.98) (0.23)	-0.12* 0.25 (0.96) (0.20)	-0.02* 0.39 (0.99) (0.50)
Dur	TP	13 20	7.16* 19.56 (0.05) (0.33)	4.98 11.23 (0.36) (0.15)	7.5 18.7* (0.26) (0.06)
Amp		12.47 25.24	5.90* 19.38 (0.05) (0.19)	3.95 10.51 (0.10) (0.17)	7.45 19.82 (0.10) (0.11)
Cum		101.7 240.1	28.1* 224.6 (0.03) (0.27)	11.2 65.9 (0.20) (0.13)	37.7 228.0* (0.18) (0.06)
Exc		0.63 1.91	-0.33* 0.73* (0.01) (0.05)	-0.17* 0.65* (0.05) (0.05)	-0.10 1.02 (0.12) (0.10)
Q test					
(dur,amp)	PT		7.33 (0.03)	3.79 (0.15)	2.57 (0.28)
(dur,amp,cum)			7.81 (0.05)	4.65 (0.20)	3.23 (0.36)
(dur,amp)	TP		2.70 (0.26)	2.01 (0.37)	1.35 (0.51)
(dur,amp,cum)			3.86 (0.28)	2.29 (0.52)	1.74 (0.63)
(excess)	PT+TP		4.55 (0.10)	3.86 (0.15)	3.8 (0.15)
(excess)	TP		3.21 (0.07)	2.99 (0.08)	0.97 (0.33)
(dur)	PT+TP		0.53 (0.77)	0.40 (0.82)	3.33 (0.19)
(dur,amp,cum)	PT+TP		10.89 (0.09)	6.58 (0.36)	5.15 (0.52)

See notes of Table 2.1.

Table 2.4: Business cycle stylised facts of threshold moving average models

	Phase	US GDP 47:1-89:2	asMA (eq. 2.12) 47:1-84:4	TDMA (eq. 2.13) 47:1-89:2
Stylised Facts				
Dur	PT	3.25	4.10 (0.85)	3.28 (0.47)
Amp		2.59	3.00 (0.66)	1.92 (0.10)
Cum		4.58	8.10 (0.79)	3.91 (0.28)
Exc		-0.07	0.01 (0.77)	-0.01 (0.75)
Dur	TP	15.57	12.45 (0.19)	18.9 (0.62)
Amp		19.06	15.06 (0.19)	20.5 (0.47)
Cum		219.52	164.3 (0.21)	347.2 (0.55)
Exc		1.15	-0.02* (0.01)	0.01* (0.03)
Quartiles				
Dur	PT	2.00 4.00	2.58 5.00 (0.50) (0.56)	2.21 3.87 (0.19) (0.23)
Amp		1.75 3.28	1.60 3.94 (0.35) (0.68)	1.10* 2.57 (0.07) (0.19)
Cum		3.29 6.58	2.20 9.94 (0.16) (0.65)	1.33* 4.91 (0.02) (0.20)
Exc		-0.33 0.23	-0.17 0.18 (0.87) (0.35)	-0.13* 0.12 (0.94) (0.14)
Dur	TP	8.0 20.0	6.06 16.27 (0.16) (0.19)	9.08 25.4 (0.40) (0.58)
Amp		12.4 25.3	6.71 19.94 (0.08) (0.20)	9.21 27.8 (0.20) (0.48)
Cum		52.1 245.4	26.2 191.4 (0.10) (0.21)	64.6 449.9 (0.27) (0.54)
Exc		0.31 1.92	-0.59* 0.55* (0.01) (0.03)	-0.63* 0.64* (0.04) (0.05)
Q test				
(dur,amp)	PT		1.03 (0.60)	2.88 (0.24)
(dur,amp,cum)			1.04 (0.70)	3.59 (0.31)
(dur,amp)	TP		0.50 (0.78)	1.19 (0.55)
(dur,amp,cum)			0.88 (0.83)	1.21 (0.75)
(excess)	PT+TP		6.50 (0.04)	3.88 (0.14)
(excess)	TP		6.18 (0.01)	3.44 (0.06)
(dur)	PT+TP		1.30 (0.52)	0.18 (0.91)
(dur,amp,cum)	PT+TP		1.74 (0.94)	4.76 (0.57)

See notes of Table 2.1.

the transitory component (c_t) is set to zero.

The simulation results are presented in Table 2.5. The SS MS model shows good results: the model generates all the features, including the excess values[23]. Because the trend is a component of the model, some replications generate too strong a trend, which implies that the QBB algorithm cannot find the turning points. The results shown in Table 2.5 are calculated from the replications (85%) where the algorithm found peaks and troughs.

An interesting feature of the SS MS is its ability to explain the asymmetry between contractions and expansions. Two characteristics of the specification of this model may be responsible for this result at least when the confidence interval with large coverage (80%) is employed. First, the presence of negative transitory shocks and permanent positive shocks create an effect similar to the CDR model and the "plucking" model. Second, the presence of shocks in the growth rate, which may create different rates of growth within a phase, which is the argument of Clements and Krolzig (2000b) for the excess value being different from zero.

On the other hand, the simulation results of the UCTSM do not include problems with the identification of peaks and troughs, but generate contractions that are longer and deeper than the US cycles. In addition, the UCTSM produces excess equal to zero for both phases, which is similar to the two-regime MS model.

2.4.6 Evaluation of Models for Italian GDP

The stylised facts of the Italian GDP presented in Table 2.6 bear a resemblance to the ones of the US output, however the duration and the amplitude of expansions are larger. The simulation of the MS AR_{IT} model (eq. 2.17) generates longer and deeper contractions compared with the stylised facts, while the $SETAR_{IT}$ model (eq. 2.16) has a better performance. The $SETAR_{IT}$ (eq. 2.16) has a performance similar to the two-regime

[23]The observation of the empirical densities of the simulated values demonstrates high skewness and kurtosis due to strong outliers, which can bias (upward) the mean across replications reported in Table 2.5.

Table 2.5: Business cycle stylised facts of structural models with Markov-switching

	Phase	US GDP 47:1-98:2	SS MS[a] (eq. 2.14) 51:1-95:3	UCTSM (eq. 2.15) 47:1-98:2
Stylised Facts				
Dur	PT	3.0	3.67 (0.58)	5.53* (0.99)
Amp		2.45	2.53 (0.44)	7.36* (1.00)
Cum		4.11	9.51 (0.52)	30.4* (1.00)
Exc		-0.07	0.03 (0.77)	-0.00 (0.67)
Dur	TP	17.8	27.9 (0.48)	16.9 (0.35)
Amp		20.2	34.3 (0.43)	28.4 (0.79)
Cum		255.7	1154 (0.47)	432.4 (0.63)
Exc		1.06	0.44 (0.15)	0.01* (0.01)
Quartiles				
Dur	PT	2 4	2.52 4.95 (0.34) (0.38)	3.0 7.01 (0.66) (0.89)
Amp		1.8 3.01	1.43 3.79 (0.17) (0.48)	4.02* 9.53* (0.98) (0.99)
Cum		2.10 6.44	2.12 14.5 (0.16) (0.41)	6.62* 36.82* (0.96) (0.99)
Exc		-0.31 0.24	-0.13 0.20 (0.83) (0.30)	-0.27 0.27 (0.60) (0.53)
Dur	TP	13 20	14.35 31.49 (0.40) (0.30)	7.69* 22.7 (0.07) (0.48)
Amp		12.47 25.24	16.76 36.34 (0.29) (0.33)	12.77 38.32 (0.39) (0.79)
Cum		101.7 240.10	463.9 1197 (0.33) (0.30)	69.4 528.2 (0.15) (0.69)
Exc		0.63 1.81	-0.33 0.89 (0.14) (0.14)	-0.54* 0.54* (0.01) (0.02)
Q test				
(dur,amp)	PT		0.41 (0.94)	6.21 (0.05)
(dur,amp,cum)			0.36 (0.83)	8.83 (0.03)
(dur,amp)	TP		0.16 (0.98)	6.54 (0.04)
(dur,amp,cum)			0.16 (0.92)	6.53 (0.09)
(excess)	PT+TP		0.40 (0.82)	6.16 (0.05)
(excess)	TP		0.18 (0.67)	5.95 (0.02)
(dur)	PT+TP		0.35 (0.84)	3.07 (0.22)
(dur,amp,cum)	PT+TP		0.70 (0.99)	14.83 (0.02)

See notes of Table 2.1.

[a] The algorithm identified missing values in 15% of the 5000 replications. The statistics are calculated based only on the valid replications.

SETAR (eq. 2.2) for US GNP. Neither the SETAR_{IT} nor the MS AR_{IT} captures the shape of the expansions because they produce negative excess values for the expansions.

The endogenous threshold model (CDR_{neg}, eq. 2.18) can reproduce all the average features. Again the presence of a phase of strong recovery after recessions, which is represented by the positive coefficient of the CDR variable with 5 lags, conditional on negative growth, helps to create a temporary recession followed by a strong recovery. Consequently, the shape of expansions and contractions can be reproduced. Using a confidence interval with smaller coverage (50%), the CDR_{neg}, however, does not account for the durations and the excesses of both phases. The CDR_{neg} also generates stronger expansions than the data (observing the upper quartiles).

2.4.7 Evaluation of Models of Australian GDP

The stylised facts of the Australian GDP indicate a larger duration of expansions than the US output (see Table 2.7), which are even longer than the Italian ones. Similar to the US output, the inclusion of duration dependence (eq. 2.20) generates longer and deeper contractions, meaning that the model cannot reproduce the contraction stylised facts. In addition, the expansions are not long enough or strong enough.

Although the two-regime MS model (2.19) can generate, with statistical accuracy, the duration, the amplitude and the cumulative effect of both phases, it is not able to reproduce the shape of the cycle and some of the expansion quartiles. Both models reported in Table 2.7 produce values for the average of the excess of the cumulative gain or loss over the triangle approximation equal to zero. Consequently, they do not reproduce the asymmetric shape between contractions and expansions.

The MS2 that can reproduce the duration and amplitude of the expansions in the US, it is not able to reproduce the long phases of prosperity of the Australian business cycle.

Table 2.6: Business cycle stylised facts of non-linear univariate models of Italian GDP

	Phase	Italian GDP 60:1-95:4	SETAR$_{IT}$ (eq. 2.16) 60:1-95:4	MS AR$_{IT}$ (eq. 2.17) 60:1-95:4	CDR$_{neg}$ (eq. 2.18) 60:1-92:1
Stylised Facts					
Dur	PT	3.0	3.7 (0.87)	4.34* (0.93)	3.34 (0.84)
Amp		1.75	2.45 (0.85)	4.03* (0.94)	1.68 (0.42)
Cum		3.47	5.86 (0.81)	13.23* (0.91)	3.33 (0.49)
Exc		-0.10	-0.07 (0.59)	0.09 (0.79)	0.00 (0.78)
Dur	TP	19.80	17.6 (0.26)	16.20 (0.21)	23.12 (0.77)
Amp		21.46	21.01 (0.36)	22.5 (0.43)	23.43 (0.49)
Cum		360.9	310.0 (0.23)	299.9 (0.25)	442.27 (0.43)
Exc		0.94	-0.37* (0.03)	-0.35* (0.06)	0.35 (0.22)
Quartiles					
Dur	PT	2 4	2.51 4.6 (0.38) (0.46)	2.63 5.34 (0.47) (0.60)	2.43 4.02 (0.32) (0.61)
Amp		1.04 1.91	1.35 3.29 (0.64) (0.90)	1.64 4.93* (0.75) (0.96)	1.02 2.18 (0.41) (0.67)
Cum		0.93 5.88	2.02 8.11 (0.78) (0.62)	2.48 14.09 (0.84) (0.74)	1.40 4.53 (0.58) (0.54)
Exc		-0.22 0.02	-0.21 0.08 (0.58) (0.68)	-0.17 0.25 (0.69) (0.87)	-0.10 0.10 (0.81) (0.72)
Dur	TP	7 37	9.35 22.84* (0.54) (0.07)	8.25 21.66* (0.43) (0.06)	13.79 30.78* (0.75) (0.97)
Amp		10.85 23.54	9.98 28.26 (0.31) (0.56)	10 30.3 (0.32) (0.61)	13.69 31.60* (0.49) (0.91)
Cum		39.45 492.43	68.5 392.4 (0.40) (0.18)	60.4 397.5 (0.37) (0.19)	170.3 664.1* (0.62) (0.96)
Exc		0.07 1.19	-1.09* 0.38 (0.05) (0.12)	-1.06* 0.68 (0.08) (0.21)	-0.37 1.03 (0.28) (0.76)
Q test					
(dur,amp)	PT		1.14 (0.57)	1.33 (0.51)	1.48 (0.48)
(dur,amp,cum)			1.35 (0.72)	1.85 (0.60)	1.69 (0.64)
(dur,amp)	TP		0.74 (0.69)	1.42 (0.49)	2.89 (0.24)
(dur,amp,cum)			0.74 (0.86)	1.45 (0.69)	3.12 (0.37)
(excess)	PT+TP		2.96 (0.23)	1.01 (0.60)	0.73 (0.70)
(excess)	TP		2.93 (0.08)	0.77 (0.38)	0.31 (0.48)
(dur)	PT+TP		1.02 (0.60)	1.61 (0.45)	1.27 (0.53)
(dur,amp,cum)	PT+TP		2.00 (0.92)	2.90 (0.82)	5.12 (0.53)

See notes of Table 2.1.

Table 2.7: Business cycle stylised factos of non-linear univariate models of Australian GDP

	Phase	Aust. GDP 60:1-97:3	MS_{AU} (eq. 2.19) 60:1-97:3	MS DD_{AU} (eq. 2.20) 60:1-97:3
Stylised Facts				
Dur	PT	3.3	3.7 (0.64)	4.3* (0.97)
Amp		2.3	3.04 (0.79)	3.12* (0.95)
Cum		4.06	7.4 (0.76)	7.20* (0.97)
Exc		0.11	0.01 (0.25)	0.01 (0.18)
Dur	TP	20.6	23.4 (0.48)	15.7* (0.08)
Amp		24.8	30.8 (0.56)	23.2 (0.31)
Cum		320.8	654.5 (0.61)	232.9 (0.16)
Exc		0.99	0.03* (0.08)	0.01* (0.01)
Quartiles				
Dur	PT	2 5	2.50 4.62 (0.37) (0.25)	3.56 5.19 (0.86) (0.34)
Amp		0.99 4.08	1.81 4.10 (0.81) (0.45)	2.22 3.94 (0.98) (0.41)
Cum		0.80 9.14	2.57 10.24 (0.88) (0.40)	3.93 9.72 (0.99) (0.54)
Exc		-0.13 0.36	-0.19 0.19 (0.45) (0.18)	-0.19 0.19 (0.38) (0.14)
Dur	TP	15 28	11.57 31.47 (0.21) (0.42)	10.9* 19.21* (0.06) (0.06)
Amp		23.16 30.82	15.21 41.47 (0.04) (0.34)	15.99* 28.65 (0.04) (0.34)
Cum		204.8 503.22	184.5 922.4 (0.04) (0.10)	94.2* 296.5 (0.04) (0.10)
Exc		0.49 1.63	-0.63* 0.68* (0.02) (0.04)	-0.60* 0.61* (0.02) (0.04)
Q test				
(dur,amp)	PT		0.64 (0.73)	3.83 (0.15)
(dur,amp,cum)			0.67 (0.88)	3.89 (0.27)
(dur,amp)	TP		0.76 (0.68)	19.2 (0.00)
(dur,amp,cum)			0.91 (0.82)	19.6 (0.00)
(excess)	PT+TP		1.86 (0.40)	5.54 (0.06)
(excess)	TP		1.45 (0.23)	5.17 (0.03)
(dur)	PT+TP		0.22 (0.89)	4.6 (0.10)
(dur,amp,cum)	PT+TP		1.62 (0.95)	19.9 (0.01)

See notes of Table 2.1.

2.4.8 Evaluation of Linear Models

Following Hess and Iwata (1997b) and Pagan (1997), we employ an AR (1) for the first-difference of the $ln(GDP)$ as a good linear representation of the business cycle features. The simulation results for AR (1) models are shown in Table 2.8 for US, Italy and Australia[24]. Other possible linear models that have been proposed are AR(4) and AR(2) models for the growth rate of the output, but the evidence of Pagan (1997) indicate that the AR (1) fares better at replicating business cycle features.

Employing the Q test and the empirical confidence interval for average values, the linear models can generate the duration, the amplitude and the cumulative measure for both phases for all the series. In contrast, when a confidence interval of only 50% coverage is assumed, the models for Australia and Italy do not generate any of the recession stylised facts. This supports the point that non-linearities are more relevant to characterise durations of recessions than of expansions (Tiao and Tsay, 1994). In addition, the linear model generates too many small values for the lower quartile of expansions compared with the US data and for the upper quartile of expansion compared with the Australian data. Moreover, linearity implies that the models cannot generate the asymmetric shape between contractions and expansions because the conditional rate of growth is constant.

Compared with non-linear models for US output, the p-value of the Q test for (dur, amp, cum) for both phases of the AR model is smaller than the p-values of the two-regime SETAR (2), the four-regime SETAR (3), the two-regime MS (5), the MS AR (7), the three-regime MS (8) and SS MS (12). Therefore, the linear model is equivalent or superior to the majority of the non-linear models for the US output when only the average of the stylised facts is employed. For Italy, the SETAR (14) has Q test p-value larger than the linear model. Similarly the MS (17) has a better performance than the linear model for Australia.

[24]The ARIMA (1, 1, 0) for US is $\Delta y_t = 0.0054 + 0.34\Delta y_{t-1} + 0.010\epsilon_t$, for Italy is $\Delta y_t = 0.0056 + 0.33\Delta y_{t-1} + 0.011\epsilon_t$ and for Australia is $\Delta y_t = 0.0099 - 0.07\Delta y_{t-1} + 0.013\epsilon_t$, where Δy_t is the first-difference of the ln(GDP).

Table 2.8: Business cycle stylised facts of AR models

	Phase	US GDP 47:1-98:2	Italian GDP 60:1-98:4	Australian GDP 60:1-97:4
Stylised Facts				
Dur	PT	3.3 (0.66)	3.4 (0.76)	2.8 (0.15)
Amp		1.92 (0.13)	2.23 (0.87)	1.86 (0.18)
Cum		3.88 (0.34)	4.69 (0.72)	2.88 (0.15)
Exc		0.0 (0.83)	0.00 (0.81)	0.00 (0.16)
Dur	TP	18.5 (0.44)	16.6 (0.39)	21.8 (0.45)
Amp		19.5 (0.36)	18.9 (0.52)	24.5 (0.38)
Cum		320.3 (0.45)	274 (0.33)	457.4 (0.48)
Exc		0.02* (0.02)	0.02* (0.06)	0.03* (0.07)
Quartiles				
Dur	PT	2.13 3.84 (0.13) (0.23)	2.21 4.02 (0.19) (0.28)	2.10* 3.17* (0.09) (0.03)
Amp		1.06* 2.50 (0.04) (0.21)	1.26 2.94 (0.90) (0.88)	1.18 2.41* (0.62) (0.03)
Cum		1.22* 4.73 (0.08) (0.19)	1.53 5.9 (0.88) (0.37)	1.30 3.69* (0.76) (0.03)
Exc		-0.12* 0.13 (0.95) (0.13)	-0.14 0.14 (0.76) (0.83)	-0.14 0.14* (0.53) (0.08)
Dur	TP	8.47 24.81 (0.10) (0.59)	8.09 22.24* (0.42) (0.06)	11.7 29.8 (0.21) (0.38)
Amp		8.42 26.6 (0.15) (0.43)	8.63 25.63 (0.90) (0.46)	13.24* 33.24 (0.09) (0.40)
Cum		47.59* 387.3 (0.09) (0.54)	52.46 358.9 (0.70) (0.17)	142.42 673.8 (0.13) (0.33)
Exc		-0.58* 0.62* (0.01) (0.04)	-0.59* 0.63 (0.09) (0.16)	-0.63* 0.70 (0.05) (0.11)
Q test				
(dur,amp)	PT	3.28 (0.19)	1.12 (0.57)	1.12 (0.57)
(dur,amp,cum)		3.73 (0.29)	1.61 (0.66)	1.25 (0.74)
(dur,amp)	TP	0.58 (0.75)	1.07 (0.58)	0.42 (0.81)
(dur,amp,cum)		0.65 (0.88)	1.14 (0.78)	0.74 (0.86)
(excess)	PT+TP	5.12 (0.08)	2.85 (0.24)	2.63 (0.27)
(excess)	TP	4.57 (0.03)	2.32 (0.13)	1.75 (0.19)
(dur)	PT+TP	0.20 (0.91)	0.47 (0.79)	0.90 (0.64)
(dur,amp,cum)	PT+TP	4.53 (0.61)	2.80 (0.83)	2.02 (0.92)

See notes of Table 2.1.

2.4.9 Discussion of the Results of the Evaluation

Comparing the TAR and MS models for the US output indicates that, in general, the two-regime SETAR (eq. 2.2) has a better performance than the two-regime MS (eq. 2.5) mainly because the SETAR model can generate average excess values different from zero. On other hand, the three-state MS (eq. 2.8) has generally better results than the four-regime SETAR (eq. 2.3), because the earlier model can reproduce the asymmetric shape between expansions and contractions. Therefore, the evaluation cannot differentiate between SETAR and MS models, confirming the results of Hess and Iwata (1997b) and Koop and Potter (1999a). The CDR models have in general an inferior performance, mainly for reproducing the amplitude of the US cycles. The inclusion of the over-heating variable improves the performance (floor and ceiling model), but the results are still inferior to the MS and SETAR models. The SS MS of Kim and Nelson (1999a) can reproduce all observed-data features. The structural model (eq. 2.14) with shifts only in the trend (UCTSM) has a poor performance.

For the Italian GDP, the two-regime SETAR (eq. 2.16) is the model with the best performance, even though the model with the CDR variable (eq. 2.18) is able to reproduce the excess values for both phases. For the Australian GDP, the two-regime MS (eq. 2.19) gives better performance. None of the models for Australia is able to reproduce the shape of the cycle and the strong expansions. Therefore, a three-regime MS may be a suggestion for reproducing the Australian stylised facts[25].

Our results confirm that an AR(1) for first differences of output can generate data that reproduces durations and amplitudes of the US business cycle (Pagan, 1997; Hess and Iwata, 1997b; Harding and Pagan, 2001b). Moreover, our analysis extends these results to

[25] Bodman and Crosby (1998) attempt to estimate a three-regime MS for Australian data. However, they include regime shifts only in the mean and not in the variance, this may weaken the ability of the model for generating different growth rates during the expansion phase. In fact, the authors do not find any evidence of a high recovery regime after the trough.

Table 2.9: Summary of stylised facts for US business cycles

	$\frac{\text{duration of contractions}}{\text{duration of cycles}}$	Excess of expansions
Data 47:1–98:2	0.14	1.06
AR(1)	0.15	0.02
SETAR 2 (eq. 2.1)	0.28	0.22
SETAR 2 (eq. 2.2)	0.16	0.20
SETAR 4 (eq. 2.3)	0.19	0.31
MRSTAR (eq. 2.4)	0.15	-0.01
MS2 (eq. 2.5)	0.18	-0.01
MS DD (eq. 2.6)	0.22	0.01
MS AR (eq. 2.7)	0.21	0.24
MS3 (eq. 2.8)	0.13	0.74
CDR (eq. 2.9)	0.18	0.23
CDR_{pos}(eq. 2.10)	0.17	0.22
F&C (eq. 2.11)	0.23	0.50
SS MS (eq. 2.14)	0.12	0.44
UCTSM (eq. 2.15)	0.25	0.01
asMa (eq. 2.12)	0.25	-0.02
TDMA (eq. 2.13)	0.15	0.01

the Italian and the Australian business cycles. However, when smaller confidence intervals are considered, the linear models have problems in generating contraction stylised facts.

Table 2.9 summarises the results for the US output of two types of asymmetries: the duration of the expansions are longer than the contractions and the shape of the expansions is not well approximated by a straight line. As argued by Pagan (1997), the ability of a model to reproduce the first type of asymmetry does not depend on the model being non-linear. In fact, the contractions, on average, only account for 14% of the US cycle, and 15% of the cycle simulated from a AR(1) model[26]. In general, the cycles generated by the non-linear models imply longer contractions: the Potter's SETAR (eq. 2.1) has contractions which are 28% of the cycle size, while the CDR and the MS models generate contractions that are on average 18% of the cycle size. The only exception is the SS MS that generates very short

[26]The duration values employed are based on the average of the stylised facts over simulations and not on the average of the proportion computed for each similation. This is done for easier comparison with the tables shown previously. Because the models are estimated for different samples, the contraction-duration/cycle-duration of the observed data may vary. However, this does not imply large changes in the interpretation of the table, given the small differences of stylised facts as argued in note at page 31.

contractions, probably because this model does not suppose that the trend follows a random walk. Therefore, we just confirm the previous results (Hess and Iwata, 1997b; Harding and Pagan, 2001a) that to reproduce the durations and amplitudes of the US cycle, one would not need a model more complex than an AR(1) for first-differences.

In contrast, non-linearity is important to the generation of the excess of expansions because the positive excess in the data means that the expansion phase may have phases with different average growth rates. The analysis of Pagan (1999) concludes by arguing that the largest positive excess obtained by non-linear models is of .2 from the data generated from a three-regime MS. However, the specification of three-regime MS, which is evaluated by the author, does not allow regime changes in the variance and does not have a strong mechanism for indicating the following phases of cycle: recession, high growth recovery, moderate growth expansion, such as the three-regime MS of Clements and Krolzig (1998). In the present evaluation, four models generate excess values larger than .3: the four-regime SETAR, the three-regime MS, the floor and ceiling and the state-space model with Markov-switching. The first three models have at least three-regimes allowing the separation of the expansion phase between high and moderate growth. The last model has a trend that depends on the growth rate that may change over time by stochastic shocks, implying that the growth rate changes inside the expansion phase. This non-linearity is exploited in the next section, but some preliminary conclusions are indicated: given that this type of asymmetry, first identified by Sichel (1994), is important to describe the business cycles, and not only for US, then non-linear models, specifically models that define two regimes inside the expansion phase, give a better representation of the business cycles than linear models.

Another important result of this section arises from the inclusion of lower and upper quartiles of the business cycle features as stylised facts: linear models cannot generate cycles with the same dispersion as that of the business cycle. When quartiles are considered stylised facts, the non-linear models have comparatively better performance than linear models.

Finally, the goodness-of-fit test based on the Q statistic, which has also been employed by Simkins (1994) and Hess and Iwata (1997b), seems to have very low power. Although the sizes of the p-values have been employed to rank models, the empirical confidence intervals are more informative on the ability of the models to generate business cycle stylised facts.

2.5 Conditional Means

The previous section indicates that most of the non-linear models cannot account for the asymmetric shape of the business cycles, although the fact that the models are non-linear means that they could, in theory, reproduce this asymmetry. These results may be explained by the pattern and the degree of non-linearity of the model. Given the complex dynamics of the non-linear models, we propose to use plots of conditional mean functions and surfaces calculated for data simulated from the model to observe the pattern and degree of non-linearity. Pagan (1999), Harding and Pagan (2001b) and Breunig and Pagan (2001) use plots of conditional mean function functions ($E[y_t|y_{t-1}]$) to evaluate some of the models (CDR, F&C, SETAR, MS2, MS DD) described in section 2.2. In this section, we also calculate conditional mean surfaces, i.e., $E[y_t|y_{t-1}, y_{t-2}]$. The advantage of the surfaces is to allow the observation of the conditional mean given two periods of negative growth, i.e., recessions. Moreover, non-linearity in y_{t-2} is responsible for the recession reversion, implying that negative shocks have temporary effects (e.g., SETAR, four-regime SETAR, CDR, F&C).

2.5.1 On the Estimation of Conditional Mean Functions and Surfaces

The conditional mean functions $E[y_t|y_{t-1}]$ and the conditional mean surfaces $E[y_t| y_{t-1}, y_{t-2}]$ are estimated non-parametrically, corresponding to the follow non-parametric regressions:

$$
\begin{aligned}
y_t &= g_1(y_{t-1}) + \varepsilon_t \\
y_t &= g_2(y_{t-1}, y_{t-2}) + \varepsilon_t.
\end{aligned}
$$

Different non-parametric methods can be employed for estimating g_1 and g_2. Pagan (1999), Harding and Pagan (2001b) and Breunig and Pagan (2001) apply Kernel methods, we employ, instead, local linear regression, which is applied by Fan, Yao and Tong (1996) to observe the characteristics of non-linear dynamic systems. The advantages of employing this type of non-parametric procedure to estimate g_i arise from the robustness of the method compared with Kernels (Hastie and Loader, 1993). Linear regressions are robust to the shape and the closeness of the boundary and also to asymmetric neighborhoods in the interior. Specially for fitting g_2, the multivariate local linear estimator is a simple extension of the single regression, and the properties of the local linear regression estimator (that are similar to parametric regressions) can be easily extended. In addition, the fit of the conditional mean can be conducted under the assumption that the distribution of the residuals is symmetric (it does not need to be Gaussian) with an effective algorithm for correcting outliers (Cleveland, Grosse and Shyu, 1993).

The estimation by local linear regressions is equivalent to weighted least squares with the weights depending on a kernel function and a bandwidth (Simonoff, 1996; Pagan and Ullah, 1999). Let x be a vector of measurement of p predictors, the linear regression estimator can be written as:

$$
\hat{g}(x) = b(x)'(B'W(x)B)^{-1}B'W(x)y,
$$

where $b(x)$ is a expansion of x on basis of polynomials; B is the matrix of evaluations of b at the sample $x_i s$ and $W(x)$ is the diagonal weight matrix: $W_i(x) = K(x - x_i)$ (Hastie and Loader, 1993). We employ local linear regression, so when $p = 1$, $b(x)$ is an expansion over a constant and y_{t-1}, and when $p = 2$, the monomials are the constant, y_{t-1} and y_{t-2}. A tricube kernel is employed to calculate the weight matrix and the smoothness depends on the size of the neighborhood αT employed in the calculation, given the sample size T. To reduce the spurious effect of outliers, we employed the *loess* procedure that gives less weight to observations with relatively large residuals, which is available in S-PLUS 2000 (MathSoft, Inc) (Cleveland et al., 1993; Simonoff, 1996, chap. 5).

2.5.2 Characteristics of Actual and Simulated Data

Given that the models evaluated in this work for Australia and Italy very much resemble the models estimated for the US data, concerning their abilities to reproduce business cycle stylised facts, we analyse conditional means only from the non-linear models estimated for US data. In addition, the structural models: SS MS and UCTSM (see section 2.4.5) are not evaluated because they generate non-stationary data, given that the trend is included as a component of the model. An alternative would be the application of a detrending method to the data simulated from these models, but, given that the choice of detrending method could change our results (Canova, 1998), we decide not to do so[27]. Table 2.10 presents a summary with the name, the source and a description of the models, given that section 2.2 presented a detailed description of them.

15,000 observations are simulated from each model and the descriptive statistics of these series are presented in Table 2.11. The mean growth of the US GDP is 3.3% per year[28]. However the AR model, which can reproduce the durations of the cycle, produces data with a mean of only 2%. The inclusion of non-linearity increases this value to 4.1% in the case of the MS2 and the MS DD models. The models have similar standard deviations, except for the MS3 which has a standard deviation of 0.8 compared with 1.1/1.2 from the other models. The smaller variance can be explained by the relatively smaller residual standard

[27]We include in this evaluation only the most popular two-regime SETAR specification: Potter's (1995) model. For the same reason, we do not include the CDR_{pos}.

[28]The descriptive statistics presented in Table 2.11 are based on quarterly growth rates. The values presented in the text are annualised rates.

Table 2.10: Model labels and sources

Label	Source	Description
SETAR 2	Potter (1995)	two-regime SETAR model
SETAR 4	Tiao and Tsay (1994)	four-regime SETAR model
MRSTAR	Van Dijk and Franses (1999)	four-regime STAR with transition variables given by $\Delta_2 y_t$ and CDR_{t-2}
MS2	Hamilton (1989)	two-regime MS-in-mean, homoscedastic model
MS DD	Durland and McCurdy (1994)	two-regime MS-in-mean with duration dependence
MS AR	McCulloch and Tsay (1994)	two-regime MS model, with slopes and mean changing
MS3	Clements and Krolzig (1998)	three-regime MS-in-intercept, heteroscedastic model
CDR	Beaudry and Koop (1993)	current depth of recession model
F&C	Pesaran and Potter (1997)	floor and ceiling model
asMA	Brännäs and De Gooijer (1994)	threshold moving average model with threshold equal to zero
TDMA	Elwood (1998)	threshold moving average model with threshold equal to zero

Table 2.11: Descriptive statistics of actual and simulated data

	mean	Std. Dev.	min	max	%<-2	%>3
SETAR 2	0.50	1.2	-4.9	4.7	3	1.2
SETAR 4	0.80	1.1	-4.2	6.5	3	2.0
MRSTAR	0.88	1.0	-3.3	4.9	0.24	1.7
MS2	1.0	1.2	-3.1	5.0	0.65	2.5
MS DD	1.0	1.2	-3.7	5.0	0.94	3.1
MS AR	0.63	1.2	-6.0	4.7	3.2	1.1
MS3	0.70	0.8	-3.7	4.9	0.94	3.1
CDR	0.79	1.0	-3.3	4.8	0.30	2.0
F&C	0.64	1.1	-4.6	6.1	1.1	1.4
asMA	0.71	1.3	-3.8	6.1	1.4	4.4
TDMA	0.83	1.1	-3.2	4.9	0.4	2.1
AR	0.51	1.1	-4.0	4.7	0.98	0.91
Data	0.81	1.1	-2.9	4.2		

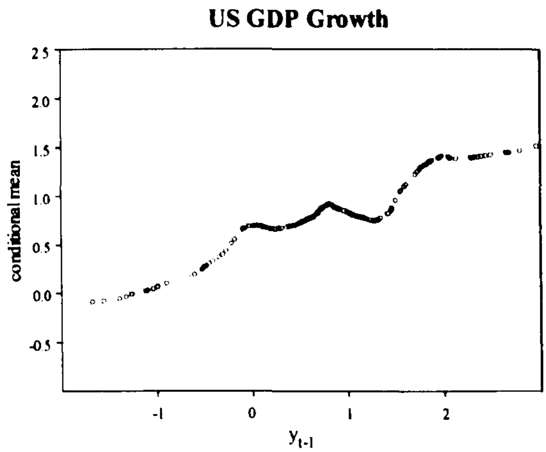

Figure 2.4: $E[y_t|y_{t-1}]$ for US GDP growth

deviations of the MS3 model compared with other models, showing a reasonable acceptable fit with the data after 1959[29] (see eq. 2.8). 95% of the observations generated by the models are in the interval $(-2, 3)$[30], so we define this interval for the computation of the conditional mean surface.

2.5.3 Analysis of Conditional Means

In this subsection, we analyse the conditional mean functions and conditional mean surfaces for the models described in the last part using loess. We assume $\alpha = 0.3$, which is an adequate degree of smoothness because non-linearities can be identified without strong variations (higher values of α over-smooth the functions and surfaces, while lower values do not deliver a straight line (or a plane surface) for the conditional mean of the linear model).

Conditional mean functions

Figure 2.4 is a scatter plot of $\hat{y}$, estimated by loess with a constant and y_{t-1} as predictors, against y_{t-1}. The figure confirms the evidence that recessions are followed by

[29]This overfitting could also explain why this model does not improve forecasting compared to an AR(4) (Clements and Krolzig, 1998).

[30]Actually, this interval has 94.2% of the values simulated from the asMa model.

periods of aceleration of the growth rates given that the shape of the curve is convex when $y_{t-1} \in (-0.5, 0)$ and gets flatter between $(0, 1.5)$ (Sichel, 1994). Indeed the US is expected to grow at rates of 2.9% percent a year when $y_{t-1} = 0$ and 3.6% when the last period growth rate is around 3.3%[31]. Therefore, the moderate growth period is probably associated with a ceiling effect (Goodwin and Sweeney, 1993; Pesaran and Potter, 1997). The presence of these two regimes inside the expansion phase is the main explanation for the positive value of the excess of the cumulative gains relative to the last trough over the triangle approximation (Clements and Krolzig, 2000b; Harding and Pagan, 2001a). To observe how the models fit this interesting data non-linearity, we plot the conditional functions of data and models in Figures 2.6 and 2.7 for the interval $y_{t-1} \in (-0.5, 1)$.

We can divide the pattern exhibited in the conditional functions into three types: (i) models with almost no change in the inclination of the conditional mean function; (ii) models with changes that are opposite to those of the data: smooth inclination until $y_{t-1} = 0.2$ and steeper after that; and (iii) models with conditional functions with non-linearity similar to the data: steep line until y_{t-1} gets positive and flatter after that. The majority of the models, headed by the AR(1) model, are in the first category: SETAR 2, MRSTAR, MS2, MS AR, MS DD, TDMA. The last two models exhibit some non-linearity, which it is not strong enough to define a pattern, however, the MS DD generates deeper recessions compared with the fast reversions of the TDMA. Models that strongly generate asymmetric responses to shocks are in the second category: the CDR and the asMA. The fact that negative shocks have only temporary effects in the latter models does not mean that there is a strong recovery after the recession, probably because they do not generate deep recessions. Finally, three models are in the last category: the SETAR 4, the MS3 and the F&C. Not by coincidence, these models are the same ones which are able to generate excess of expansions bigger than .3, as discussed in the last section.

[31] Note that the Figure 2.4 is calculated using quarterly growth rates.

US GDP Growth

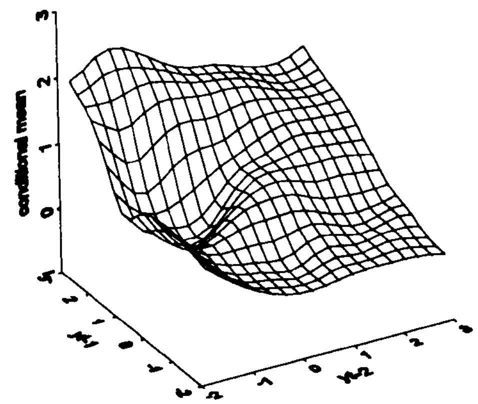

Figure 2.5: $E[y_t|y_{t-1}, y_{t-2}]$ for US GDP growth

Therefore, the analysis of the conditional functions is useful for verifying that not all non-linear models fit data non-linearities well: only models with at least three-regimes can describe the fact that a high growth recovery phase happens after the recession.

Conditional mean surfaces

Figure 2.5 shows $\hat{y}$, estimated by loess using a constant, y_{t-1} and y_{t-2} as predictors, against y_{t-1} and y_{t-2}. The plot of the surface allows observation of the reversion when both y_{t-1} and y_{t-2} are negative, given that the possibility of being a boundary effect is excluded because we employed *loess* that is robust for this type of problem. Another interesting feature is that the surface inclination depends on y_{t-2} being positive or negative, given that y_{t-1} is positive: the surface is steeper when y_{t-2} is negative. Both characteristics confirm that recessions are short lived and are followed by a high growth recovery.

The conditional mean surfaces for the models are shown in Figures 2.8, 2.9, 2.10, 2.11, 2.12, 2.13, 2.14, 2.15, 2.16, 2.17, 2.18, and 2.19. The reversion effect observed in the data surface when both y_{t-1} and y_{t-2} are negative is only verified in the surface of the F&C

(fig. 2.16) and on a smaller scale in the surfaces of the SETAR 4 (fig. 2.9) and the CDR (fig. 2.15). The fact that the impulse responses of these models indicate that negative shocks are temporary may be the cause of these patterns. The majority of the models have a flat surface in the part that corresponds to negative values of y_{t-1} and y_{t-2}: SETAR 2, MS2, MS DD, MS3, asMA, TDMA. Finally, surfaces that do not change the inclination given negative values are: the MRSTAR (fig. 2.10), the MS AR (fig. 2.13) and the AR (fig. 2.19). Actually the surfaces of the MRSTAR and MS AR models do not show non-linearity. Compared with the results obtained in section 2.4, it seems hard to understand why the SETAR 4 and the F&C generate longer recessions when they are the models that generate stronger recession reversion. The reason may be that the recession needs to be long enough for the dynamics of the models to switch to the regime that implies recession reversion. Before entering into the recession reversion regime, the models suggest high negative growth for at least two quarters.

The characteristic of the data is that the surface inclination depends on y_{t-2} being positive or negative, given that y_{t-1} is positive. This characteristic is stronger in SETAR models, but Markov-Switching models seem to account for a ceiling effect when values of y_{t-1} and y_{t-2} are highly positive. However, this ceiling effect looks symmetrical with the floor effect which happens during recessions (i.e., $y_{t-1} < 0, y_{t-2} < 0$) for the MS2 and the MS DD. In the case of the MS3, there is a regime of moderate growth characterised by a pattern in the middle of the surface, but the inclination of the surfaces of the MS2 and the MS DD are constant outside these "floor" and "ceiling" regions.

It is not immediately clear that models with three regimes are a better representation of the data non-linearities when observing the surfaces. However, it is possible to identify interesting non-linear patterns in the dynamics of the models. Actually, for some of the non-linear models, the important result is how weak are the non-linearities generated by these models, compared with the non-linearity observed in the US data. Therefore, the reason why some non-linear models, such as the MRSTAR and the TDMA, can generate the durations and amplitudes of the cycles is not because they capture better the data non-linearity, but because they generate data which are similar to data from an AR(1).

2.6 Conclusions

Non-linear time series models can reproduce business cycles stylised facts if they have one of two characteristics: (i) they generate data that are similar to data from an AR(1), implying constant conditional mean; or (ii) they have a mechanism that creates two different regimes inside the expansion phase.

Non-linear models with the first characteristic generate the duration and the amplitude of the business cycles, but they have difficulty in reproducing the dispersion and the shape of the cycle. In fact, if a model is needed to account only for asymmetric durations between the business cycle phases, an AR(1) model with drift for the first-difference of the $\ln(GDP)$ is suitable. This is not only true for the US business cycles, but for the Italian and Australian business cycle as well.

Non-linear models with the second characteristic generate the asymmetric shape of the business cycle, which means that they generate data that do not have a constant conditional mean, as an AR(1). The analysis of conditional mean functions and surfaces shows that the positive excess of expansions are caused by the fact that expansions are followed by high growth recoveries that later turn into a moderate growth expansion, as a result of a 'ceiling' effect.

Two models (out of 16 for US output) are able to account for these two characteristics: the three-regime MS model (Clements and Krolzig, 1998) and the state-space model with Markov-switching (Kim and Nelson, 1999a). Therefore, one must be careful when estimating non-linear models for characterising business cycles: the fact of being non-linear does not mean that the model generates the asymmetries present in the observed cycle (more

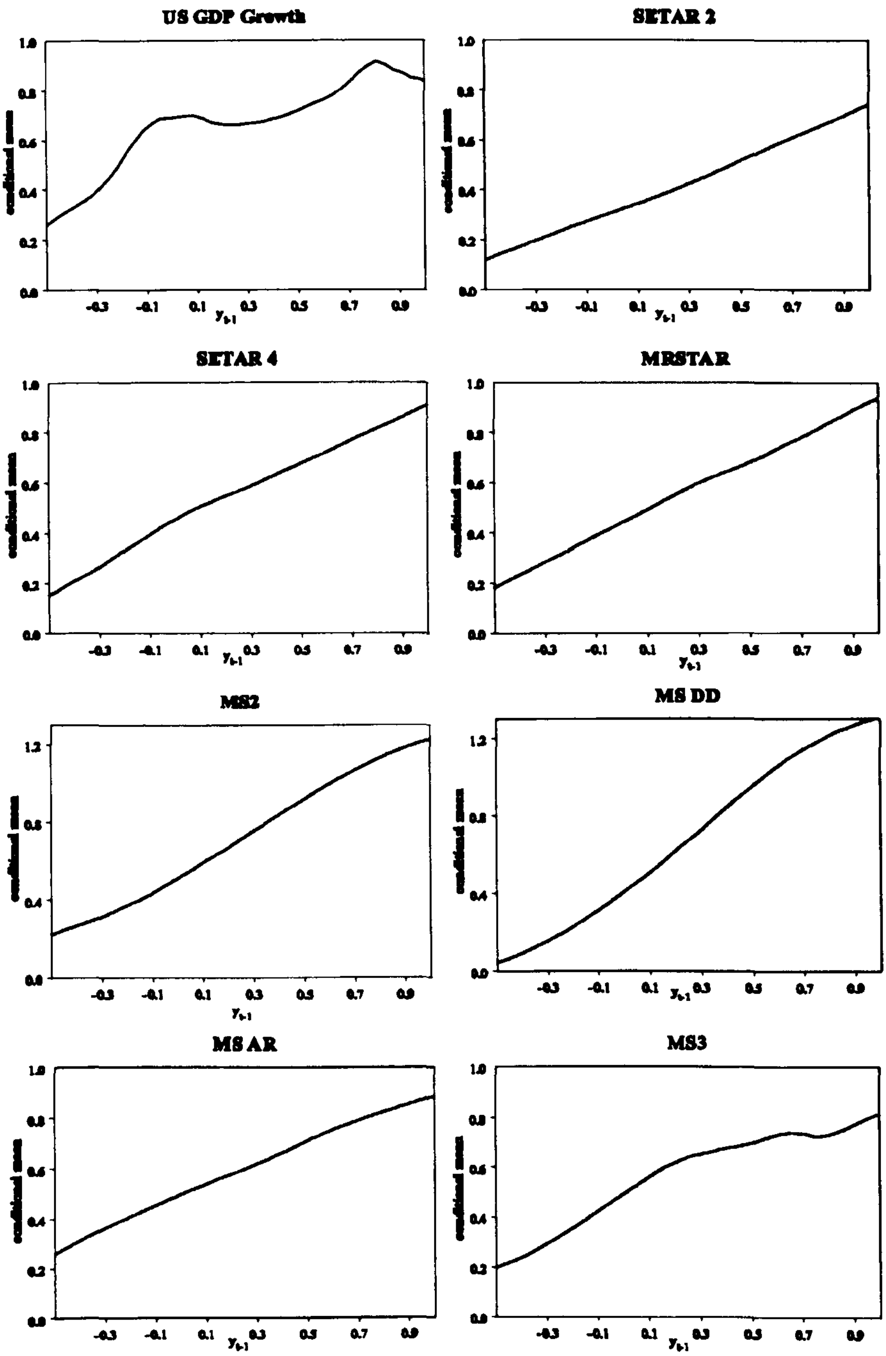

Figure 2.6: $E[y_t|y_{t-1}]$ for US GDP growth and data simulated from SETAR 2, SETAR 4, MRSTAR, MS2, MS DD, MS AR and MS3

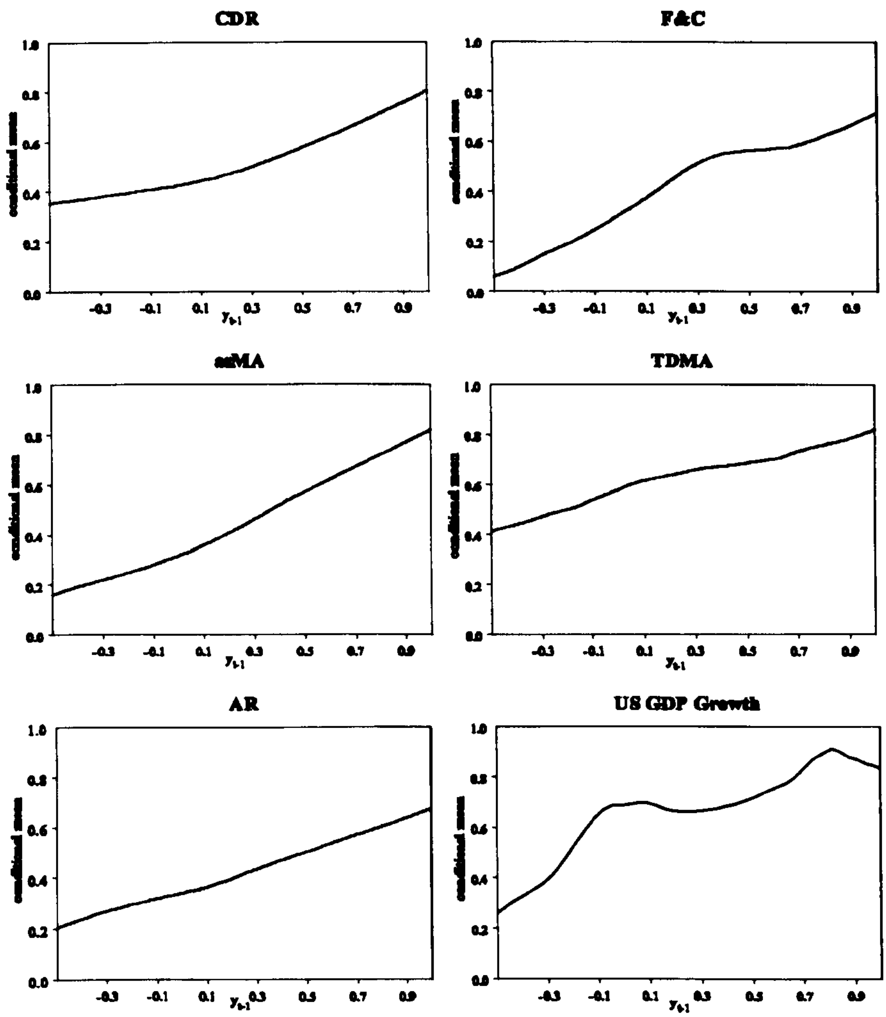

Figure 2.7: $E[y_t|y_{t-1}]$ for US GDP growth and data simulated from CDR, F&C, asMA, TDMA and AR

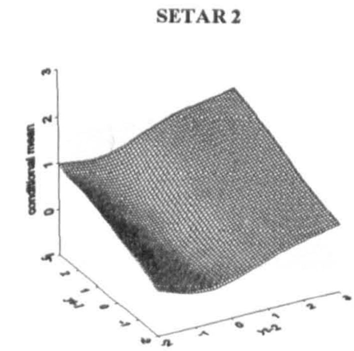

Figure 2.8: $E[y_t|y_{t-1}, y_{t-2}]$ for SETAR2

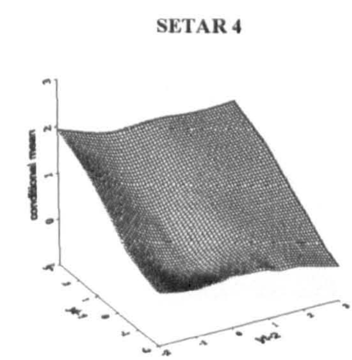

Figure 2.9: $E[y_t|y_{t-1}, y_{t-2}]$ for SETAR4

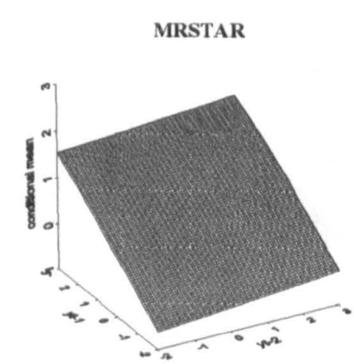

Figure 2.10: $E[y_t|y_{t-1}, y_{t-2}]$ for MRSTAR

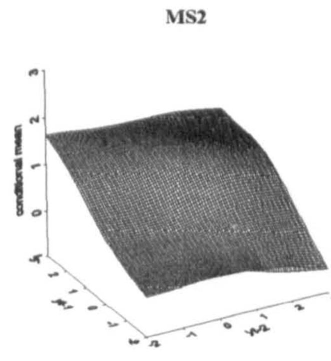

Figure 2.11: $E[y_t|y_{t-1}, y_{t-2}]$ for MS2

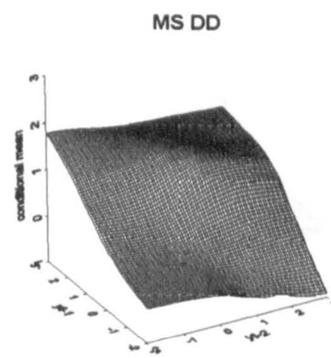

Figure 2.12: $E[y_t|y_{t-1}, y_{t-2}]$ for MS DD

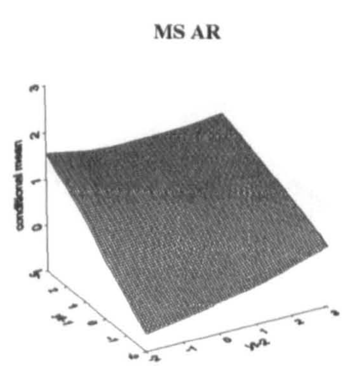

Figure 2.13: $E[y_t|y_{t-1}, y_{t-2}]$ for MS AR

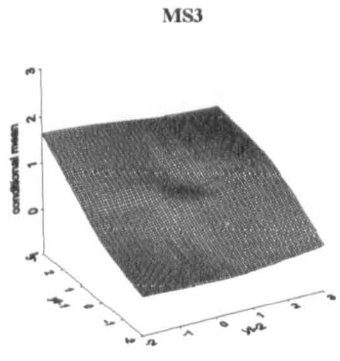

Figure 2.14: $E[y_t|y_{t-1}, y_{t-2}]$ for MS3

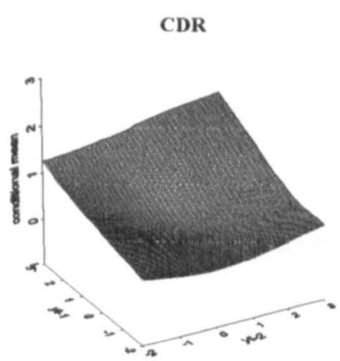

Figure 2.15: $E[y_t|y_{t-1}, y_{t-2}]$ for CDR

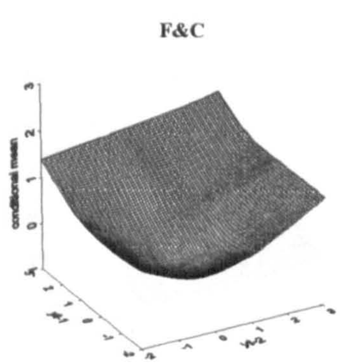

Figure 2.16: $E[y_t|y_{t-1}, y_{t-2}]$ for F&C

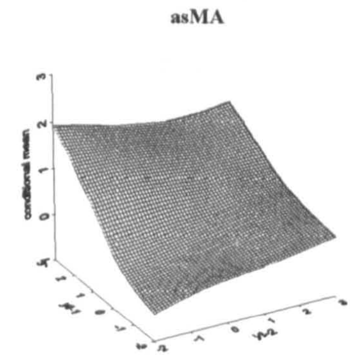

Figure 2.17: $E[y_t|y_{t-1}, y_{t-2}]$ for asMA

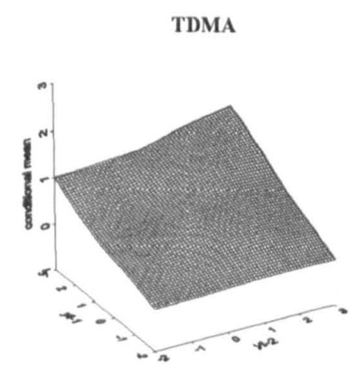

Figure 2.18: $E[y_t|y_{t-1}, y_{t-2}]$ for TDMA

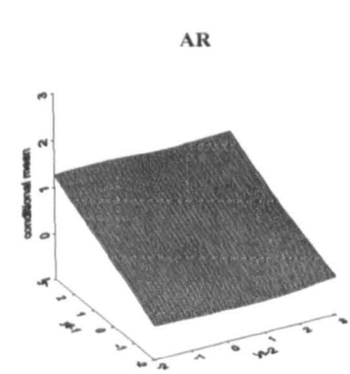

Figure 2.19: $E[y_t|y_{t-1}, y_{t-2}]$ for AR

details in Chapter 5). The analysis of the conditional mean functions and surfaces is a good instrument to observe which types of non-linearity, if any, a non-linear model is capturing.

Chapter 3

Non-linear Cointegrated Systems of US Term Structure of Interest Rates

3.1 Introduction

Equilibrium correction models are popular representations of the dynamics of macroeconomics data because they can describe the long-run relationship between the variables in conjunction with short-run adjustments. The specification of the long-run equilibrium is based on tests for unit roots and cointegration. The forecast performance of cointegration systems depends on which outcome of the system is evaluated: levels, differences or cointegrating combinations (Clements and Hendry, 1995, Christoffersen and Diebold, 1998 and references therein).

However, linear equilibrium correction models indicate that the equilibrium adjustment does not depend on the size and on the sign of the deviation. In the presence of transaction costs, arbitrage opportunities only occur when the difference between the prices in each market is large enough to mean net gains to investors, implying that disequilibria are only corrected when they are larger than a threshold. In markets in which transaction costs may be significant, non-linear equilibrium correction models are the indicated specification. Non-linear error correction models have been estimated for the relationship between spot and future prices (Dwyer, Locke and Yu, 1996; Martens, Kofman and Vorst, 1998; Tsay, 1998) and for the relationship between different interest rate maturities (Anderson, 1997; Enders and Granger, 1998; Van Dijk and Franses, 2000).

Similarly, linear equilibrium correction models cannot account for the effect of economic policies on the adjustment to equilibrium. Roberds, Runkle and Whiteman (1996) argue that the spread, the cointegrating relationship between the long-term and short-term interest rates, helps to forecast the short-term interest rates only during certain monetary policy regimes. Rudebusch (1995) shows that whether the spread is relevant to forecasting short-term interest rates depends on the perception of the market of the likelihood of the maintenance of the current FED funds rate target. This phenomenon has been characterised in the literature with regime switching models (Gray, 1996; Pfann et al., 1996; Ang and Bekaert, 1998).

Non-linear and asymmetric equilibrium models can be also applied when positive and negative equilibrium deviations imply different adjustments. For example, asymmetric costs of hiring and firing mean that labour demand adjusts asymmetrically to long-run equilibrium (Escribano and Pfann, 1998).

The effect of non-linearities on economic forecasting has been evaluated in the case of univariate models (Tiao and Tsay, 1994; Clements and Krolzig, 1998; Rothman, 1998; Montgomery et al., 1998; Stock and Watson, 1999; Lundbergh and Teräsvirta, 2000; Clements et al., 2000). The results of these evaluations typically conclude that non-linear models do not forecast better than linear ones in the case of US GDP, but improve the forecasting of the US unemployment rate. Regarding interest rates, the survey of Fauvel, Paquet and Zimmermann

(1999) is not optimistic about the applicability of non-linear models, such as neural network and regime switching models, to model and forecast interest rates in a univariate setting, although the evaluation of Swanson and White (1995) suggests that neural network models are good forecasters of interest rates.

The forecasting performance of different specifications of asymmetric error correction has been compared by Escribano and Granger (1998), who found some weak evidence that the inclusion of asymmetries in the equilibrium correction adjustment improves forecasting performance. The problem is that models that fit best in-sample do not necessarily forecast better because the models may overfit the data (Ramsey, 1996; Escribano and Granger, 1998; Stock and Watson, 1999).

However, the effect of the inclusion of non-linearity in cointegrated systems has not been evaluated. This chapter fills this gap, evaluating the forecasting performance of non-linear vector equilibrium correction models compared to linear vector equilibrium correction models. In other words, this work evaluates whether the linear restriction reduces the forecasting performance compared to more flexible non-linear equilibrium correction models. Specifically, threshold vector equilibrium correction models (TVEqCM) are evaluated. To check the robustness of the results to the assumption of cointegration, we also compare TVEqCMs with threshold VARs and autoregressive models.

There is a growing literature on the presence of non-linear cointegration between long-term and short-term interest rates. Tests for threshold or non-linear cointegration have been developed (Balke and Fomby, 1997; Enders and Granger, 1998; Enders and Siklos, 2001; Corradi, Swanson and White, 2000; Hansen and Seo, 2000) and non-linear equilibrium correction models for the short-term interest rate have been estimated and tested (Anderson, 1997; Tsay, 1998; Van Dijk and Franses, 2000). We apply different non-linear specifications of bivariate systems to short- and long-term US interest rates. The models are derived by employing testing and specification methods found in the literature.

A literature review of the application of equilibrium correction models to the term structure of interest rates is presented in section 3.2. The different specification procedures and non-linearity tests for threshold equilibrium correction models are presented in section 3.3. Forecasting performance is evaluated using the measures and tests described in section 3.4. The testing and estimates of the models are analysed in section 3.5 and the results of the forecast evaluation are presented in section 3.6. The robustness of the results of the forecasting evaluation is analysed in section 3.7. Discussion of the main results are presented in the conclusions in the last section.

3.2 Term Structure of Interest Rates and Equilibrium Correction Models

The dynamics of the interest rates depend on their term structure. The expectations theory implies that the yield to maturity k is the weighted average of the expected one period yield plus a risk premium:

$$r_t(k) = \frac{1}{k}\left[\sum_{j=1}^{k} E_t r_{t+j-1}(1)\right] + L_t(k) \tag{3.1}$$

$r_t(k)$ is the yield at maturity k; E_t is the expected value at time t; L_t is the risk premium. The arbitrage between bond markets with different maturities guarantees that equation 3.1 holds. Investors observing profitable price differences between markets of similar assets will buy or sell bonds, causing price differences to disappear. However, the presence of the risk premium implies that the yield curve will be upward sloping during most of the time.

Assuming that the yields are integrated of order one ($I(1)$), the possibility of cointegration of the yield spread can be observed re-writing equation 3.1 as:

$$r_t(k) - r_t(1) = \frac{1}{k}\left[\sum_{i=1}^{k-1}\sum_{j=1}^{i} E_t \Delta r_{t+j}(1)\right] + L_t(k) \tag{3.2}$$

If the RHS of equation 3.2 is stationary, this equation implies that each yield $r_t(k)$ is cointegrated with $r_t(1)$ and that the spreads are stationary, and consequently, the spreads between any two yields should be cointegrating (Hall, Anderson and Granger, 1992). If the spreads are cointegrated, a Vector Equilibrium Correction Model (VEqCM) for two yields, say short-term (s) and long-term (l), can be written as:

$$\Delta \mathbf{r}_t = c(L)\Delta \mathbf{r}_{t-1} - \boldsymbol{\alpha}(S_{t-1} - \mu) + \boldsymbol{\epsilon}_t \tag{3.3}$$

where $\mathbf{r}_t$ is the $[r_t(l), r_t(s)]'$ vector; Δ is the first difference operator; $c(L)$ is a matrix of coefficients in the lag operator; S_{t-1} is the spread $[(r_t(l) - r_t(s))]$; μ is the equilibrium spread; $\boldsymbol{\alpha}$ is the adjustment vector of the long-run attractor; $\boldsymbol{\epsilon}_t$ is the disturbances vector. Yields of different maturities can only move in dissimilar directions in the short-run; in the long-run they will move together. The possibility of μ being different from zero is due to the risk premium. In addition, equation 3.3 implies that $[1, -1]'$ is the cointegration vector for $[r_t(l), r_t(s)]'$.

The empirical evidence of cointegration between yields of different maturities and the importance of the equilibrium correction representation for improving the short-run forecast of interest rates are being contested in the literature (Pagan, Hall and Martin, 1996). The usefulness of the spread for forecasting the short-term rates depends on the maturities chosen (values of s and l) (Rudebusch, 1995) and on the monetary policy (Roberds et al., 1996). In the case that a policy of stabilising interest rates has market credibility, the short-term interest rates expected by the market are equal to the current short-term rates (Mankiw and Miron, 1986). This makes the short-term rates a random walk and the spread equal to the term premium, consequently the fluctuations of the spread do not help to predict the short-term interest rate. Therefore, if the Fed targets the stabilisation of the short-term interest rate, then $\boldsymbol{\alpha}$ in equation 3.3 is only statistically different from zero when the risk premium changes or when the policy is not credible. For other policies, such as the targeting of the monetary base, as adopted in 1979-1982 in the US, the short-term interest rate suffers strong fluctuations, which are temporary, generating periods of high volatility. In the latter case, the spread may help to forecast the short-term rate (Sola and Driffill, 1994; Rudebusch, 1995; Roberds et al., 1996; Gray, 1996). Another reason why the spread may not help to predict interest rates is the presence of a time-varying risk premium that may create bias in the estimation of equation 3.3 (Tzavalis and Wickens, 1998). In addition, the expectations theory of the term structure of the interest rates ignores transaction cost effects. When these costs are contemplated, the adjustment to the equilibrium deviation occurs only when the arbitrage gains are larger than the transaction costs (Anderson, 1997).

These considerations - the diversity of monetary policy, the risk premium and the presence of transaction costs - can be accommodated in a non-linear equilibrium correction model, where the speed of adjustment depends on the regime. This means that the adjustment to the equilibrium, represented by $\boldsymbol{\alpha}$ in equation 3.3, may depend on the size and sign of the disequilibrium. For example, Enders and Granger (1998) propose a Momentum Threshold Autoregressive model (M-TAR) to characterise dissimilar speeds of adjustment given the sign of the disequilibrium. When $\Delta S_{t-1} \geq (\mu + \pi) = c$, the speed of adjustment to the equilibrium is $\boldsymbol{\alpha}$; when $\Delta S_{t-1} < c$, the speed of adjustment is $\boldsymbol{\beta}$. The Momentum Threshold Vector Equilibrium Correction Model (M-TVEqCM) can be written as:

$$\Delta \mathbf{r}_t = c(L)\Delta \mathbf{r}_{t-1} - \boldsymbol{\alpha} S_{t-1} G(\Delta S_{t-1}) - \boldsymbol{\beta} S_{t-1}(1 - G(\Delta S_{t-1})) + \epsilon_t, \tag{3.4}$$

where $G(\Delta S_{t-1})$ is a Heaviside indicator function that is equal to 1 when $\Delta S_{t-1} \geq c$ and is equal to 0 otherwise, characterising two regimes in the equilibrium correction mechanism. This model was extended by Enders and Siklos (2001) to include w_{t-1}, the estimated long-run relationship between $r(l)$ and $r(s)$, instead of S_{t-1}. Likewise, Hansen and Seo (2000) present a two-regime Threshold Vector Equilibrium Correction model (TVEqCM) for interest rates, using a grid search for the estimation of the threshold and the cointegration vector. The main objective of the authors, however, is to test for threshold cointegration, using the *sup* of an LM type test. The hypothesis of threshold cointegration is supported by the spread of some maturities.

The speed of adjustment may also depend on the size of the disequilibrium. When transaction costs create a band of inaction, Threshold Equilibrium Correction Models (TEqCM) can be a good representation of the data. Anderson (1997) supposes that inside a band $c_1 = \mu + \tau_l < S_{t-1} < \mu + \tau_u = c_2$, the equilibrium adjustment does not occurs, and the adjustment occurs at different speeds outside this band. The author then proposes a three-regime TEqCM:

$$\Delta r_t(s) = \begin{cases} c(L)\Delta \mathbf{r}_{t-1} - \alpha S_{t-1} + \varepsilon_{1t} & \text{if } S_{t-1} < c_1 \\ d(L)\Delta \mathbf{r}_{t-1} + \varepsilon_{2t} & \text{if } c_1 < S_{t-1} < c_2 \\ a(L)\Delta \mathbf{r}_{t-1} - \beta S_{t-1} + \varepsilon_{3t} & \text{if } S_{t-1} > c_2, \end{cases} \tag{3.5}$$

where the endogenous variable is the short-term yield, and the long-term interest rate may enter as an explanatory variable in the vector $\Delta \mathbf{r}_{t-1}$. Each regime has different short-term dynamics and adjustment to the long-run equilibrium, given that $\alpha \neq \beta$ and the variances of the disturbances are regime-dependent. Likewise, Tsay (1998) estimates a three-regime TVEqCM, which includes short- and long-term yields as endogenous variables. However, although Tsay employs the spread as a transition variable, the model does not include an equilibrium correction term.

Because different investors might have different transaction costs, the effect of transaction costs might be smooth in the aggregate (Anderson, 1997). In this case, a Smooth Transition Equilibrium Correction model (STEqCM) is indicated as a representation of the relationship between the spread and the yields (Anderson, 1997; Van Dijk and Franses, 2000). A STEqCM for the short-term yields can be written as:

$$\Delta r_t(s) = c(L)\Delta \mathbf{r}_{t-1} - \alpha S_{t-1} + [d(L)\Delta \mathbf{r}_{t-1} - \beta S_{t-1}]G(S_{t-1}; \gamma, \delta, \mu) + \varepsilon_t, \tag{3.6}$$

where $G(S_{t-1};\gamma,\delta,\mu)$ is a transition function, which depends on the transition variable S_{t-1} and the parameters γ, δ and μ. Anderson applies a transition function that allows adjustment to be asymmetric around the long-run equilibrium μ:

$$G(S_{t-1}) = 1 - \exp\left\{-\gamma(S_{t-1}-\mu)^2\left[0.5 + \frac{1}{1+\exp(-\delta(S_{t-1}-\mu))}\right]\right\}; \tag{3.7}$$

and assumes that $\alpha = 0$. Thus, the adjustment to the long-run equilibrium is weaker when deviations from μ are small, because many traders will be inside the inaction band; larger deviations produce stronger adjustments to equilibrium. The specification of Van Dijk and Franses (2000) employs a quadratic logistic function:

$$G(S_{t-1}) = \frac{1}{1+\exp\{-\gamma(S_{t-1}-c_1)(S_{t-1}-c_2)\}}. \tag{3.8}$$

The latter STEqM (eq. 3.6 with transition function defined by eq. 3.8) nests the TEqCM (equation 3.5) when $\gamma \to \infty$, given that the dynamic coefficients are the same in the upper and the lower regimes.

Moreover, Corradi et al. (2000) found some evidence of non-linearity in the cointegration between long- and short-term interest rates. The authors, instead of estimating a non-linear vector equilibrium correction specification, propose tests of linear cointegration against the non-linear cointegration, based on an exponential function.

Summarising, equilibrium correction models can be derived from the propositions of the expectations theory of the term structure of interest rates, implying that changes in the short- and long-term interest rates can be predicted by changes in the spread, defined as the equilibrium. However, the present literature evidences that the adjustment to the long-run equilibrium may be non-linear, depending the size and the sign of the disequilibrium.

3.3 Modelling Threshold Vector Equilibrium Correction models

The non-linear models presented in section 3.2 are estimated for different data sets. The number of regimes and the definition of the transition variable are, in the majority of the models, not tested. Because a reasonable forecasting competition should be based on well specified models, over the same data set, this section discusses the testing, the modelling and the estimation procedures for non-linear equilibrium correction models.

3.3.1 Non-linearity Testing

Threshold vector equilibrium correction models are extensions of univariate Threshold Autoregressive (TAR) models, and inherit the non-standard aspects of testing for non-linearity that arise from the presence of nuisance parameters under the null hypothesis when likelihood-based approaches are used (see, e.g., Hansen (1996)).

The benchmark model is a linear model VEqCM

$$\Delta\mathbf{r}_t = \mathbf{c} + \sum_{j=1}^{p} \Phi_j \Delta\mathbf{r}_{t-j} + \boldsymbol{\alpha} S_{t-1} + \epsilon_t, \tag{3.9}$$

$\Delta\mathbf{r}_t$ is the vector $[\Delta r_t(l), \Delta r_t(s)]'$; $\boldsymbol{\alpha}$ is (2X1), given that the cointegration rank is 1; S_t is $[r_t(l) - r_t(s)]$; Φ_j for $j = 1, ..., p$ are (2X2) matrices, ϵ_t is vector of disturbances $[u_{1t}, u_{2t}]'$. The autoregressive order is set to minimise an information criteria.

A possible non-linear alternative hypothesis is that interest rates follow a TVEqCM model:

$$\Delta\mathbf{r}_t = \mathbf{c}^{(i)} + \sum_{j=1}^{p} \Phi_j^{(i)} \Delta\mathbf{r}_{t-j} + \boldsymbol{\alpha}^{(i)} S_{t-1} + \epsilon_t^{(i)} \text{ if } r^{(i-1)} < z_{t-d} \leq r^{(i)}, \tag{3.10}$$

where $i = 1, 2$ in the case of a two-regime model (with $r^{(0)} = -\infty, r^{(2)} = +\infty$) and $i = 1, 2, 3$ in a three-regime model (with $r^{(0)} = -\infty, r^{(3)} = +\infty$); z_{t-d} is the transition variable with delay d; $r^{(i)}$ are the thresholds; $\alpha^{(i)}$ is a vector (2X1), so the spread is an explanatory variable for both equations in the system.

Tsay (1998)

The non-linearity test proposed by Tsay (1998) is the vector extension of the Tsay (1989) test for non-linearity based on an 'arranged regression'. The problem of testing for a threshold becomes that of testing for a change-point. Unlike likelihood-based approaches, which rely on simulated critical values, the test has an asymptotic chi-squared distribution. The arranged regression orders the observations according to the size of z_{t-d}, and assumes that the threshold variable (z_t) and the delay, d, as well as the autoregressive order, p, are known. The model is then estimated by recursive LS, and the predictive residuals are obtained. Under the null that the model is linear, these residuals are uncorrelated with the explanatory variables in the arranged regression. The test is constructed by regressing the (standardized) predictive residuals on the explanatory variables, and testing for the significance of the latter (Tsay, 1998, p. 1189-91). The simulation results presented by Tsay indicate that the test has good power when d is well specified and when a large part of the sample period can be designed as the starting value for the recursive least square estimation. Using Tsay's method, the non-linearity test chooses the best delay and the best transition variable to use for the estimation of a threshold model.

Balke and Fomby (1997)

The test proposed by Balke and Fomby (1997) is a two-step procedure for testing threshold cointegration. The first step is to test for cointegration using a standard method, such as OLS and an ADF test, as in Engle and Granger (1987) or the ML procedure of Johansen (1988). The analysis of cointegration tests when the equilibrium correction is non-linear has been done by Balke and Fomby (1997) and Van Dijk and Franses (2000). The second step tests for non-linearity in the cointegrating combination, using a test of linearity against SETAR structure, such as Hansen (1996; 2000a). In this work, Hansen's (2000a) approach, which allows testing using two-regime and three-regime models under the alternative, is employed for testing non-linearity in the second step. The F-test for non-linearity has p-values calculated by a bootstrap procedure that take into account heteroscedasticity in the residuals in the linear model.

Enders and Granger (1998) and Enders and Siklos (2001)

Testing for cointegration (as in of step 1 of Balke and Fomby above) may have low power when the process is I(0), but exhibits non-linear mean reversion (Balke and Fomby, 1997). Enders and Granger (1998) propose a unit root test with an asymmetric adjustment under the alternative hypothesis, where the process is either a TAR (Threshold Autoregressive) or an M-TAR (Momentum-TAR). First the spread is regressed against a constant, then the residuals $\hat{S}$ are used to estimate the following regression:

$$\Delta \hat{S}_t = I_t \rho_1 \hat{S}_{t-1} + (1 - I_t)\rho_2 \hat{S}_{t-1} + u_t, \tag{3.11}$$

where the Heaviside function $I_t = 1$ for $\hat{S}_{t-1} \geq 0$ and $I_t = 0$, otherwise, for the TAR alternative; and $I_t = 1$ for $\Delta\hat{S}_{t-1} \geq 0$ and $I_t = 0$, otherwise, for the M-TAR alternative. Enders and Granger obtain by simulation critical values for the unit root null that $\rho_1 = \rho_2 = 0$ against both these alternatives, and for various modifications of the above set up. Conditional on rejecting the null and finding $\rho_1 < 0$ and $\rho_2 < 0$, tests that $\rho_1 = \rho_2$ have standard distributions. Enders and Siklos (2001) generalise these ideas to tests for cointegration, that is, when the variable is an estimated residual from an Engle-Granger (1987) static regression of one integrated variable on another (or several). In the case of the term structure, the residual is $w_t = r_t(l) - \hat{\theta} r_t(s)$, where $\hat{\theta}$ is estimated by OLS. The null hypothesis is now interpreted as a test that $r(l)$ and $r(s)$ are not cointegrated, and the alternative is of 'asymmetric cointegration', whereby the series can be shown to be related by a non-linear equilibrium-correction model. Enders and Siklos (2001) obtain, by simulation, critical values of two test statistics of the null that $\rho_1 = \rho_2 = 0$ (against both TAR and MTAR alternatives). These are an F-test of $\rho_1 = \rho_2 = 0$ and a 't − max' statistic, which is the larger of the two individual t-tests of $\rho_1 = 0$ and $\rho_2 = 0$. A Monte Carlo evaluation of these tests suggests that the F-test is to be preferred, and has reasonable power when the process is an M-TAR, but otherwise is dominated by the ADF test. We report both tests. Further complications arise when the threshold is not known (we implicitly assume that it is zero for w_t). An incorrect assumption concerning the threshold reduces the power of the test as argued by Berben and Van Dijk (1999).

Generalisation of Hansen (2000a)

Instead of testing S_t (or an estimated cointegrating relationship between $r\,(l)$ and $r\,(s)$) for non-linearity as outlined above, threshold effects can be tested for by comparing the linear system (3.9) against the non-linear alternative (3.10). This requires a multivariate extension of Hansen (2000a). The two- and three-regime TVEqCM can be written as

$$\begin{aligned} \Delta \mathbf{r}_t &= \left[\mathbf{c}^{(1)} + \sum_{j=1}^{p} \Phi_j^{(1)} \Delta \mathbf{r}_{t-j} + \boldsymbol{\alpha}^{(1)} S_{t-1}\right] I_{1t}(r) + \\ & \quad \left[\mathbf{c}^{(2)} + \sum_{j=1}^{p} \Phi_j^{(2)} \Delta \mathbf{r}_{t-j} + \boldsymbol{\alpha}^{(2)} S_{t-1}\right] I_{2t}(r) + \epsilon_{2t} \end{aligned} \tag{3.12}$$

$$\begin{aligned} \Delta \mathbf{r}_t &= \left[\mathbf{c}^{(1)} + \sum_{j=1}^{p} \Phi_j^{(1)} \Delta \mathbf{r}_{t-j} + \boldsymbol{\alpha}^{(1)} S_{t-1}\right] G_{1t}(r_1, r_2) + \\ & \quad \left[\mathbf{c}^{(2)} + \sum_{j=1}^{p} \Phi_j^{(2)} \Delta \mathbf{r}_{t-j} + \boldsymbol{\alpha}^{(2)} S_{t-1}\right] G_{2t}(r_1, r_2) + \\ & \quad \left[\mathbf{c}^{(3)} + \sum_{j=1}^{p} \Phi_j^{(3)} \Delta \mathbf{r}_{t-j} + \boldsymbol{\alpha}^{(3)} S_{t-1}\right] G_{3t}(r_1, r_2) + \epsilon_{3t}, \end{aligned} \tag{3.13}$$

where $I_{1t} = I(S_{t-1} \leq r), I_{2t} = I(S_{t-1} > r), G_{1t} = I(S_{t-1} \leq r_1), G_{2t} = I(r_1 < S_{t-1} \leq r_2)$ and $G_{3t} = I(S_{t-1} > r_2)$, where $I(\cdot)$ is an indicator function. We denote the estimated covariance matrices of ϵ_{2t} and ϵ_{3t} by $\hat{\Omega}_{2r}$ and $\hat{\Omega}_{3r}$, and let $\hat{\Omega}_{1r}$ be the covariance matrix of the VEqCM[1]. An LR test for linearity against the two-regime specification is:

$$LR_{12} = (T)(\ln(\det(\hat{\Omega}_{1r})) - \ln(\det(\hat{\Omega}_{2r})), \tag{3.14}$$

where T is the number of observations effectively employed in the estimation.

The asymptotic distribution is an extension for multivariate models of the one derived by Hansen (1996), as argued in Hansen (2000a). The bootstrap can be used to obtain finite sample approximation. Because the bootstrap efficiency depends on the hypothesis that the residuals are independent, corrections for heteroscedasticity were proposed by Hansen (2000a). The bootstrap distribution is calculated from data generated by the linear model by re-sampling its residuals. The residuals are corrected for heteroscedasticity before re-sampling, using a regression of the squared residuals on the squared regressors, as described in Hansen (2000a).

Hansen and Seo (2000)

Rather than firstly estimating the cointegrating relationship, Hansen and Seo (2000) suggest a single step approach that jointly estimates the cointegrating vector and the threshold. The non-linear model is:

$$\begin{aligned} \Delta \mathbf{r}_t &= \left[\mathbf{c}^{(1)} + \sum_{j=1}^{p} \Phi_j^{(1)} \Delta \mathbf{r}_{t-j} + \boldsymbol{\alpha}^{(1)} w_{t-1}(\theta)\right] d_{1t}(\theta, r) \\ &+ \left[\mathbf{c}^{(2)} + \sum_{j=1}^{p} \Phi_j^{(2)} \Delta \mathbf{r}_{t-j} + \boldsymbol{\alpha}^{(2)} w_{t-1}(\theta)\right] d_{2t}(\theta, r) + \epsilon_t \end{aligned} \tag{3.15}$$

where $w_t(\theta) = r_t(l) - \theta r_t(s)$ is the equilibrium correction term, $d_{1t}(\theta, r) = I(w_{t-1}(\theta) \leq r)$ and $d_{2t}(\theta, r) = I(w_{t-1}(\theta) > r)$. Their method involves estimating equation 3.15 at each point in a suitable grid of values defined over both θ and r, and choosing the pair $\left(\hat{\theta}, \hat{r}\right)$ that minimizes $\log\left|\hat{\Omega}(\theta, r)\right|$ (where $\hat{\Omega}$ is the estimated residual covariance matrix). In practice, because of

[1]The models are estimated on the same sample and for the same autoregressive order p.

the limitations of the estimation by grid search, the delay is given, and the TVEqCM is restricted to having two regimes, as above.

The authors also proposed an 'LM-like' non-linearity test. An LM test is calculated of the linear model against equation 3.15 with $\theta = \hat{\theta}$, the ML value in the linear VEqCM, and r taking on each of a pre-assigned set of values in the interval (r_L, r_U) (such that a minimum number of observations occur in each regime, say 10%). The test statistic is the supremum:

$$SupLM = \sup_{r_L \leq r \leq r_U} LM(\hat{\theta}, r). \tag{3.16}$$

Hansen and Seo derive the asymptotic distribution of this statistic, and propose a bootstrap procedure to obtain the finite sample approximation.

3.3.2 Estimation

Tsay (1998) considers estimation of the TVEqCM (3.10) by conditional multivariate least squares assuming the number of regimes, the autoregressive order p, and the threshold variable z_t are known. Equation 3.10 is estimated for all permissible combinations of the delay d and the thresholds r_1 and r_2 (for the 3-regime model), subject to a minimum number of observations in each regime, $r_2 > r_1$ and d being a (typically low) integer. The estimates of the thresholds and d are those values for which the residual sum of squares is minimized. The asymptotic results and properties of the estimators are discussed by Tsay. Instead of minimising the sum of the squares of the residuals, Tsay suggests employing the AIC, given that when the autoregressive order and the number of regimes are fixed, AIC is asymptotically equivalent to selecting the model with the smallest generalised residual variance, but not to selecting the model with the smallest sum of square of the residuals.

To reduce the computational burden of estimating the three-regime model, which is important when bootstrapping is employed to calculate the finite sample distribution of the non-linearity test, Hansen (2000a) proposes a one-step-at-a-time algorithm, that is a sequential procedure. The threshold value estimated for a two-regime model is employed as one of the thresholds of the three-regime model, and a grid search for the second threshold is then conducted, with the same delay as in the two-regime model. The objective function of the grid search is $\ln\left|\hat{\Omega}(r_1, r_2)\right|$, where $\hat{\Omega}$ is the estimated variance-covariance matrix of the residuals, given the threshold values r_1 and r_2 and assuming constant variance across regimes. This procedure is iterated at least once, to refine the estimation of the threshold values. Bai (1997) proved the consistency of this sequential approach for models of multiple structural breaks.

In addition, the cointegration vector and the threshold can be jointly estimated as described in section 3.3.1.

3.3.3 Model Specification

The first point to specify a TVEqCM is the definition of the number of regimes. When the number of regimes s is unknown, Tsay (1998) suggests making a selection based on the AIC, defined by:

$$AIC = \sum_{i=1}^{s} \left[T_i \ln\left(\left|\hat{\Omega}_i\right|\right) + 2k(kp+1)\right], \tag{3.17}$$

where p is the autoregressive order, $k = 2$ in the bivariate system, $\hat{\Omega}_i$ is the estimated residual covariance matrix of regime i, and T_i is the number of observations in regime i. Thus the AIC is calculated for each combination of the threshold values, the delay and for $s = 2, 3$, given p determined by the order of a VAR.

In addition, tests with two-regime model under the null and three-regime under the alternative can be employed to specify the number of regimes when the Balke and Fomby (1997) approach is followed. This is an F-test, similar to the non-linearity test, employing bootstrap p-values. Likewise, while testing the two-regime TVEqCM (eq. 3.12) against the three-regime TVEqCM (eq. 3.13), an LR test can be applied:

$$LR_{23} = (T)(\ln(\det(\hat{\Omega}_{2r})) - \ln(\det(\hat{\Omega}_{3r})) \tag{3.18}$$

For bootstrapping the distribution of the LR_{23} statistic, the data has to be simulated from the system 3.12. As suggested by Hansen (2000a), we allow for regime heteroscedasticity in bootstrapping the residuals.

Alternatively, the number of regimes can be defined by theoretical considerations, like the ones employed by Anderson (1997) to define a three-regime model. As discussed in section 3.2, theoretical arguments over the term structure can be employed to define two-regime models (adjustment depend upon descending and ascending yield curves) and three-regime models (transaction costs).

Specific tests for non-linearity may also indicate the type of non-linear model required, such as the test of Enders and Granger (1998) for whether there is asymmetric adjustment to the spread. Testing for a unit root against M-TAR and TAR alternatives may indicate which of these two models should be used.

3.4 Evaluating Non-linear Equilibrium Correction Models

Clements and Hendry (1995) and Christoffersen and Diebold (1998) find that there is a gain in forecast accuracy at longer horizons when the ability to predict the cointegrating combination is evaluated (here, the spread). Christoffersen and Diebold suggest the MSFE_{tri}, which is the sum of the MSFEs for the cointegrating relation and for the first difference of one of the variables in the bivariate system, as an adequate measure to observe the forecasting effects of cointegration. However, the MSFE_{tri} is not invariant to scale. Specifically in this work, the means and standard deviations of, say, $\Delta r(s)$ are a fraction of the means and standard deviations of S, as a result the MSFE of S dominates the MSFE, meaning biased results. In addition, Clements and Hendry (1993) show why forecast evaluation using the standard (root) mean squared forecast error criterion, (R)MSFE, may depend on the transformation of the variables adopted (e.g., the levels of the original variables, their first differences or growth rates, or a mixture of first differences and stationary combinations).

Because forecasts of both rates and the spread are of interest in this work, we consider these separately for the most part. We look at forecasts of the differences of the interest rates as these should be stationary. The MSFE for the variable x at horizon h is denoted by $\text{MSFE}_{x,h}$, and is calculated by averaging the squares of the h-step ahead forecast errors over a number of forecast origins, T. Given the described problem on summing MSFEs, we report an overall measure of systems performance referred to as the GFESM (general forecast-error second-moment matrix) by Clements and Hendry (1993), because this is invariant to whether we evaluate the models in terms of their ability to forecast the changes of the rates, or the short rate and the spread, etc. Given the h-step ahead vector of forecast errors for predicting the variables in the bivariate system $\mathbf{e}_{T+h}$, the measure employed is the determinant of the GFESM at power $1/h$:

$$GFESM_h = \left|E[\mathbf{EE}']\right|^{(1/h)}; \text{ where } \mathbf{E}' = [\mathbf{e}'_{T+1}, ..., \mathbf{e}'_{T+h}]. \tag{3.19}$$

We test whether the MSFEs (for variable x in predicting at h step-ahead) of the various models are significantly different from each other using the test of equal forecast accuracy of Diebold and Mariano (1995), with the small-sample corrections suggested by Harvey, Leybourne and Newbold (1997). That is, to test the null of equal accuracy at h steps-ahead we calculate:

$$d_t = e_{i,t}^2 - e_{j,t}^2$$

$$DM = [\hat{V}(\bar{d})]^{-1/2}\bar{d}$$

$$ADM = \left[\frac{n + 1 - 2h + n^{-1}h(h-1)}{n}\right]^{1/2} DM, \tag{3.20}$$

where $t = 1, ..., n$ indexes the n h-step ahead forecasts that are available, and i and j index the two models. So $e_{i,t}$ is the error in forecasting the value at t, made at $t-h$, using model i, $\bar{d}$ is the sample mean of the loss differential d_t, where loss is symmetric and defined in terms of the squares of the residuals. The estimated variance of the sample mean is denoted by $\hat{V}(\bar{d})$, and depends on the sample autocovariances of d_t. Under the null of equal forecast accuracy, ADM is asymptotically standard normal. Thus values of the statistic in the left tail suggest model i is more accurate, and values in the right tail that j is more accurate. Harvey et al. (1997) suggest comparing the Augmented Diebold and Mariano (ADM) test to the t-distribution to reduce size distortions, and Clark (1999) confirms that these modifications improve the small-sample performance of the test.

We can test whether model i forecast encompasses model j, that is, whether once we have model i, there is no useful additional information contained in the forecasts of model j, by modifying d_t to:

$$d_t = [(e_{i,t} - e_{j,t})]e_{i,t}. \tag{3.21}$$

Harvey, Leybourne and Newbold (1998) show that this test is equivalent to the forecast encompassing test of Chong and Hendry (1986). The condition that a model forecast encompasses another is more stringent than that the model is more accurate on the ADM test. One model may be more accurate than another on ADM but, nevertheless, the dominated model contains useful information not incorporated in the superior model.

West (1996) and West and McCracken (1998) draw attention to the impact of parameter uncertainty on the size of the tests of equal forecast accuracy and of encompassing. Specifically for the case of cointegrated systems, Corradi, Swanson and Olivetti (1999) show that the Diebold and Mariano forecast accuracy test can be applied when the loss function is quadratic. In the case of evaluation of nested models, such as in this work, the asymptotic distributions of the forecast accuracy tests are non-standard when parameter uncertainty is considered (Clark and McCraken, 2000). In general, the effects of parameter uncertainty on the forecast accuracy tests can be minimised by using an adequate proportion of in-sample to out-of-sample observations. In our evaluation, using rolling forecasts, the smallest estimation period contains 360 monthly observations (without starting point correction). The forecasts are rolling for 100 (n) periods (each time the model is re-estimated) until the 460^{th} observation. For each n, 24 (h) step-ahead ex-ante forecasts are generated. Therefore, we expect that parameter uncertainty will have secondary importance in this work.

An alternative way to rank competitor forecasters is to employ simulation. An advantage is that the forecast evaluation is based on a situation in which the 'non-linearity' that occurred in the past occurs in the forecast period. Moreover, one can vary the sample size to investigate the effects of parameter estimation uncertainty. A disadvantage of this approach is that it requires that the model being used to generate data is a good representation of the process, which brings into question the relevance for 'real world' comparisons. Clements and Krolzig (1998), Clements and Smith (1999a) and Lundbergh and Teräsvirta (2000) applied this type of analysis to compare non-linear models against linear ones. The idea is to simulate data from one of the models in the competition and to use this data to estimate the models and to calculate the forecast errors, including MSFEs, or some other measure, for each h. After repeating this some thousand times, the mean of the MSFEs of each replication is calculated and compared among models.

The predictions from the non-linear models are generated by employing the bootstrap[2]. In this way, we do not need to make any assumption about the distribution of the residuals, only that the residuals are independent. Some preliminary results using the Monte Carlo simulation to generate forecasts (supposing the residuals are bivariate normal) indicated that bootstrapping has a slightly better performance for the same number of repli-

[2]For considerations about the different methods to forecast non-linear models, see Granger and Teräsvirta (1993)

cations (500); although the values converge when the number of replications gets larger.

3.5 Model Estimates

Most analyses of interest rate maturities employ either the Fama (Hall et al., 1992; Corradi et al., 2000) or the McCulloch and Kwon (1993) data set (Pagan et al., 1996; Hansen and Seo, 2000). The main disadvantage of these data sets is that the available sample ends at the beginning of the 90s. The Fred website (www.stls.frb.org/fred/data/irates.html) provides data for different maturities from 1960 until the present. Comparing the same data period (1960-1991) for the three-month treasury bill and the 10-year treasury rate to the same maturities of the McCulloch and Kwon (1993) data (economics.sbs.ohio-state.edu/mccull.html), we conclude that the series are virtually the same. For this analysis, we employ monthly data of the 3-month treasury bill at secondary market and the 10-year treasury constant maturity for the period 1960:1 to 2000:4. Anderson (1997) and Hansen and Seo (2000) employ three and six month maturities, but we think that the chosen maturities are more representative of the short- and long-term interest rates. The data is presented in Figure 3.1. The spread is defined as $r(120) - r(3)$ and it is presented in Figure 3.2. The estimation period is 1960:1 to 1989:12, the remaining sample is employed for forecast evaluation. The 'in-sample' analysis in this section presents results for the beginning and the end of the forecast sample, i.e., for the samples of 1960:1-1989:12 and 1960:1-1998:4.

3.5.1 Testing and Modelling

Applying the procedures for specification of TVEqCMs outlined in section 3, we estimate seven different TVEqCMs (3 with 2 regimes and 4 with 3 regimes). The main characteristics of these models are described in Table 3.1[3]. All the models have the autoregressive

[3]The estimation and forecasting evaluation is performed with codes written by the author using GAUSS. The 2R-TVEqCM$_u$ is tested using a modification of Bruce Hansen's code, employed in Hansen (2000a). The 2R-TVEqCM$_{joint}$ is tested and estimated using a modification of Hansen's and Seo's code, employed in Hansen and Seo (2000). Both codes were obtained at www.ssc.wisc.edu/~bhansen.

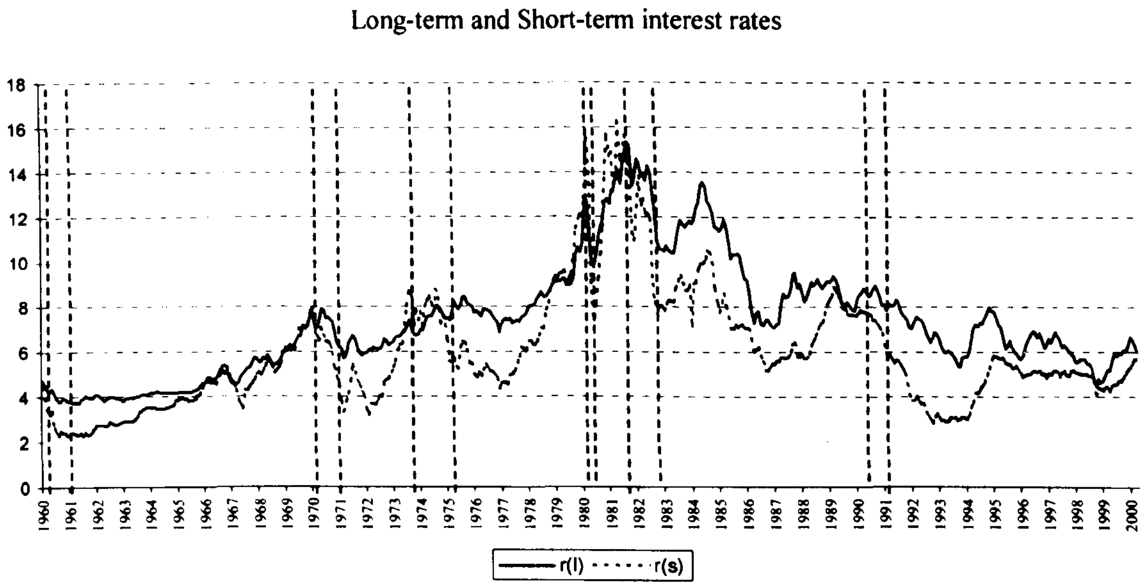

Figure 3.1: Long-term (10-year T-bond) and short-term (3-month T-bill) interest rates (dashed lines are NBER turning points)

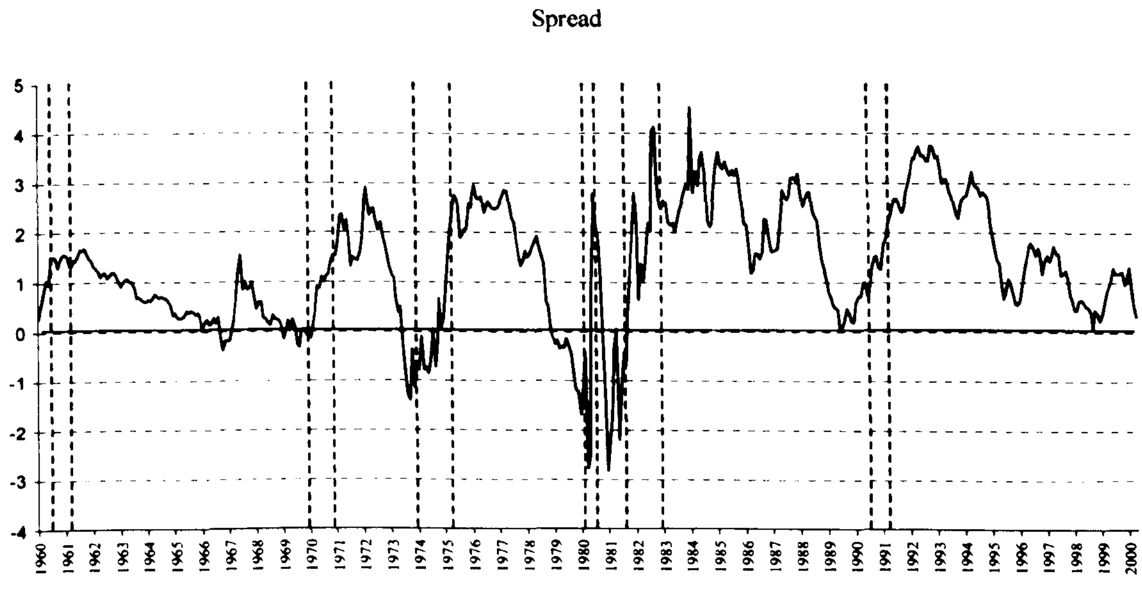

Figure 3.2: The spread between long- and short-term interest rates (dashed lines are NBER turning points)

order (p) defined by the linear model estimation. The delay is set to 1, because although non-linearity is rejected for $d = 1, 2, 3, 4$ (employing $p = 2$) (see Table 3.2), the rejection is strongest for $d = 1$. Moreover, previous studies (Anderson, 1997; Hansen and Seo, 2000) set the delay equal to 1. The decision to set $p = 2$ results from the calculation of information criteria for a VEqCM. We employ the Schwarz Information Criteria (SIC) to define p, because it is well known that the AIC (Akaike) may over-parameterize the model, over-fitting the data, and harming the forecast performance[4]. In section 3.6, these considerations are discussed further. Therefore, each TVEqCM is estimated given $p = 2$, $d = 1$ and with a minimum number of observations in each regime of 10%.

For all the models, except the MTVEqCM and the 2R-TVEqCM$_{joint}$, the spread ($S_t = r(l) - r(s)$) is imposed as the equilibrium correction term. This assumption is based on the unit root tests and on cointegration tests, presented in Table 3.2. There is some ambiguity over the spread: on the basis of an ADF test it is $I(1)$, the Phillips-Perron tests indicates that it is $I(0)$ (panel 2), and the Johansen systems-based trace test for cointegration finds that $r(l)$ and $r(s)$ are cointegrated, but the restriction that the interest rates have equal and opposite sign (defining the spread) is rejected. Following previous works, and taking into account the low power of unit root tests, we use the spread as the cointegrating relationship in the models, except for the MTVEqCM and the 2R-TVEqCM$_{joint}$ where the long-run relationship estimated.

For the MTVEqCM and the 2R-TVEqCM$_{joint}$, the cointegration vector is defined as $w_t = r_t(l) - \theta r_t(s)$. In the case of the MTVEqCM, θ is estimated as in the Engle and Granger (1987) procedure, and then asymmetric adjustment is tested in w_t using the Enders and Siklos (2001) test, which is shown in panel 4 and 5 of Table 3.2. For both non-linear specifications, M-TAR and TAR, the test statistics reject the null hypothesis of no cointegration. The F-

[4] A preliminary forecasting evaluation with $p = 7$, which is the autoregressive order supported by AIC, concluded that all the models have worse forecast performance compared with models with $p = 2$.

Table 3.1: Characteristics of the models

Models	Specification	Non-linearity Testing	Estimation of Thresholds[a]
VEqCM			
2R-TVEqCMh	two-regime (from theoretical indication) regime-dependent variance		grid search on multivariate model to minimise AIC (eq. 3.17)
MTVEqCM	two-regime (from non-linearity test) short-run coef. and variance are constant across regimes	Enders and Granger (1998)	grid search on univariate model of w_t to minimise SSE
2R-TVEqCMjoint	two-regime (from non-linearity test) constant variance	Hansen and Seo (2000)	grid search on multivariate model to minimise $\ln(\det(\Omega(\theta, r))$
3R-TVEqCMi	three-regime (using AIC) regime-dependent variance	Tsay (1998)	grid search on multivariate model to minimise AIC (eq. 3.17) with $r_1 \in [-0.5, 0.95]$ and $r_2 \in [1.5, 3.5]$
3R-TVEqCMu	three-regime (using F-test in a TAR for spread) regime-dependent variance	Balke and Fomby (1997) with Hansen (2000) non-linearity test	one-step-at-a-time grid search in univariate model of the spread to minimise the SSE
3R-TVEqCMf	three-regime (from theoretical indication) regime-dependent variance		grid search on multivariate model to minimise AIC (eq. 3.17)
3R-TVEqCM	three-regime (using LR test of eq. 3.18) constant variance	LR test (eq. 3.14)	one-step-at-a-time grid search in a multivariate model to minimise $\ln(\det(\Omega(r_1, r_2))$

Note: Thresholds are estimated conditional on at least 10% of the observations in each regime, except for the 3R-TVRqCMi. [a] Conditional on thresholds, transition variable, delay, autoregressive order and number of regimes, the models are estimated by conditional multivariate least squares.

test for $\rho_1 = \rho_2$ does not reject the null when an M-TAR is specified. However, this is not a strong evidence against M-TAR asymmetries because the F-test assumes that the threshold is equal to zero, while the estimated value is 0.39, reducing the power of the test as discussed by Berben and Van Dijk (1999). We choose the M-TAR specification as the basis for the TVEqCM to be in line with Enders and Granger (1998) and Enders and Siklos (2001) on similar data sets.

For the 2R-TVEqCM$_{joint}$, a grid for θ is defined using the asymptotic normal interval of the estimate of θ of the linear model, and θ is then estimated by grid search jointly with the threshold value. The 2R-TVEqCM$_{joint}$ is tested against the linear specification using the *supLM* test of Hansen and Seo (2000). Table 3.2, panel 6 suggests that the linear VEqCM is clearly rejected, supporting threshold cointegration.

The TVEqCM is the alternative of two other testing procedures. The first employs the approach of Balke and Fomby (1997). Non-linearity is tested based on the estimation of a SETAR model for the spread with $p = 1$ and $d = 1$. The non-linearity F-test is calculated when an AR is under the null and a SETAR is under the alternative, and also when the null is a two-regime SETAR and the alternative is a three-regime SETAR (Hansen, 2000a). Statistics and p-values of these tests are presented in panels 8 and 9 in Table 3.2, given that p-values are calculated by bootstrap, taking into account heteroscedasticity of the residuals under the null. Non-linearity is rejected, but it is not possible to reject the two-regime SETAR in favour of the three-regime SETAR. However, we estimate a TVEqCM with the thresholds defined by a three-regime SETAR to check the implications for forecasting.

The second method is the generalisation of the non-linearity test of Hansen (2000a) to the multivariate framework, as explained in section 3. The p-values of the LR tests, presented in panels 10 and 11 in Table 3.2, are calculated using the heteroscedasticity corrected bootstrap. The test rejects the hypothesis of linearity, and also the hypothesis of two-regimes, contradicting the results for the univariate case.

Table 3.2: Test results

	Tests	1960:1-1989:12	1960:1-1998:4
1	Augmented Dickey-Fuller Unit Roots	$r(s)$, -1.699 $r(l), -1.572$ $\Delta r(s), -8.651^*$ $\Delta r(l), -7.870^*$ S, -2.497	$r(s)$, -1.890 $r(l)$, -1.736 $\Delta r(s), -9.633^*$ $\Delta r(l), -7.109^*$ S -2.795
2	Phillips-Perron Unit Root	S, -3.716*	S -3.899^*
3	Johansen cointegration (trace)	$r(l) - \Pi r(s)$, 16.88*	$r(l) - \Pi r(s)$ 18.6*
4	Enders and Granger Asymmetric Cointegration (TAR)	Φ, 9.84* t-max, -3.96* $\rho_1 = \rho_2$,2.84[0.09]	Φ, 10.91* t-max, -4.26* $\rho_1 = \rho_2$,4.42*[0.04]
5	Enders and Granger Asymmetric Cointegration (M-TAR)	Φ, 8.39* t-max, -3.20* $\rho_1 = \rho_2$,0.11[0.74]	Φ, 8.64* t-max, -3.07* $\rho_1 = \rho_2$,0.01[0.91]
6	Hansen and Seo Threshold cointegration	25.183* [0.00]	34.527* [0.00]
7	Tsay System non-linearity	$d = 1$, 116.98*[0.00] $d = 2$, 68.83*[0.00] $d = 3$, 105.23*[0.00] $d = 4$, 87.39*[0.00]	$d = 1$, 110.54*[0.00] $d = 2$, 63.94*[0.00] $d = 3$, 130.13*[0.00] $d = 4$, 195.43*[0.00]
8	Threshold Cointegration linear X three-regime	45.17* [0.002]	52.43* [0.00]
9	Threshold Cointegration Two-regime X three-regime	11.09 [0.17]	10.08 [0.14]
10	System non-linearity (eq. 3.14) linear X three-regime	149.71* [0.01]	182.88* [0.01]
11	System non-linearity (eq. 3.18) two-regime X three-regime	62.24* [0.02]	71.29* [0.03]

* null hypothesis rejected at least at 5% significance level.

Notes: (1) and (3) are based on p=7; (2) based on truncation lag equal to 5; (4) and (5) are based on positive and negative deviations of the estimated cointegrating combination and on p=1; (6),(8),(9),(10) and (11) are based on heteroscedasticity-corrected statistic and on 500 bootstrap samples; (7) is also computed with heteroscedasticity-consistent statistic. For all the non-linearity tests, we set d = 1 and p = 2 , except for (7) which considers d=1,...,4.

Some models are specified (2R-TVEqCM$_h$ and 3R-TVEqCM$_f$) without any prior non-linearity testing, allowing grid search to choose the model that best fits the data by selecting the thresholds to minimise the AIC, conditional on at least 10% of the sample in each regime. In some cases, the search is over a grid interval given by the researcher, based on the descriptive statistics of values of the threshold variable. In the case of the 3R-TVEqCM$_i$, the intervals of possible threshold values are defined as $r_1 \in [-0.5, 0.95]$ and $r_2 \in [1.5, 3.5]$. The grid is performed with 40 points in the interval for r_1 and 30 points in the interval for r_2, and the choice between two-regime or three-regime TVEqCM is carried out using the AIC values. The AIC for the three-regime model is better than a two-regime specification, which can also be observed in Table 3.3.

The variance-covariance matrix of the system depends on the regime in the case of the 2R-TVEqCM$_h$, the 3R-TVEqCM$_i$, the 3R-TVEqCM$_u$ and the 3R-TVEqCM$_f$. The procedure to estimate thresholds may also differ following the supposition on the variance of residuals. In general, when the variance depends on the regime, the grid search minimises the AIC criterion (eq. 3.17), which is based on the sum of $\ln(|\hat{\Omega}^i|)$ of each regime. When the variance is constant across regimes, the grid search minimises $\ln(|\hat{\Omega}|)$ with $\hat{\Omega}$ being defined by the residuals of the full sample. One-step-at-a-time is an algorithm to reduce the computational burden of grid search when the criterion is $\ln(|\hat{\Omega}|)$ (section 3.3). Conditional on the threshold value, however, all the models are estimated by conditional multivariate least squares (Tsay, 1998).

3.5.2 A Comparison of Specifications

At the risk of being repetitive, this section describes formally the specification of the estimated models with the objective of clarifying similarities and differences. The summary of the model characteristics is shown in Table 3.1.

The TVEqCMs with two regimes can be nested in the followed equation:

$$\begin{bmatrix} \Delta r_t(l) \\ \\ \Delta r_t(s) \end{bmatrix} = \begin{bmatrix} (\Phi_1 \Delta r_{t-1} + \alpha_1 w_{t-1}) I(w_{t-1}) + (\Phi_2 \Delta r_{t-1} + \alpha_2 w_{t-1})(1 - I(w_{t-1})) \\ \\ (\Phi_3 \Delta r_{t-1} + \alpha_3 w_{t-1}) I(w_{t-1}) + (\Phi_4 \Delta r_{t-1} + \alpha_4 w_{t-1})(1 - I(w_{t-1})) \end{bmatrix} + \begin{bmatrix} u_{1t} \\ \\ u_{2t} \end{bmatrix},$$

where $\Delta r_{t-1} = (1, \Delta r_{t-1}(l), \Delta r_{t-2}(l), \Delta r_{t-1}(s), \Delta r_{t-2}(s))'$, $I(w_{t-1}) = 1$ if $w_{t-1} \leq r$ and $I(w_{t-1}) = 0$ if $w_{t-1} > r$ and $w_{t-1} = r_{t-1}(l) - \theta r_{t-1}(s)$. The disturbances are contemporaneously correlated ($cov(u_{1t}, u_{2t}) \neq 0$). Given that $\epsilon_t = (u_{1t}, u_{2t})'$, the variance-covariance matrix of the residuals is $\Omega = (\epsilon'\epsilon)/T$. The regime-dependent variance-covariance Ω^i can be calculated for each regime i with ϵ_t^i taken given the values of the transition function $I(w_{t-1})$ that defines the regimes.

The 2R-TVEqCM$_h$ has the the spread as cointegration vector ($\theta = 1$) and the variance is regime-dependent. MTVEqCM has θ estimated in a first step, $I(w_{t-1}) = I(\Delta w_{t-1})$ (a momentum TAR), $\Phi_2 = \Phi_4 = 0$ and the variance of the residuals are the same across regimes. The 2R-TVEqCM$_{joint}$ has θ jointly estimated with r and the same residual variances across regimes.

All the three-regime models have $w_{t-1} = r_{t-1}(l) - r_{t-1}(s) = S_{t-1}$, so they can be nested in the following equations:

$$\Delta r_t(l) = \begin{cases} (\Phi_1 \Delta r_{t-1} + \alpha_1 S_{t-1}) I_1(S_{t-1}) + \\ (\Phi_2 \Delta r_{t-1} + \alpha_2 S_{t-1})(1 - I_1(S_{t-1})) I_2(S_{t-1}) + \\ (\Phi_3 \Delta r_{t-1} + \alpha_3 S_{t-1})(1 - I_2(S_{t-1})) + u_{1t} \end{cases}$$

$$\Delta r_t(s) = \begin{cases} (\Phi_4 \Delta r_{t-1} + \alpha_4 S_{t-1}) I_1(S_{t-1}) + \\ (\Phi_5 \Delta r_{t-1} + \alpha_5 S_{t-1})(1 - I_1(S_{t-1})) I_2(S_{t-1}) + \\ (\Phi_6 \Delta r_{t-1} + \alpha_6 S_{t-1})(1 - I_2(S_{t-1})) + u_{2t} \end{cases}$$

where $r_1 < r_2$, $I_1(S_{t-1}) = 1$ if $S_{t-1} \leq r_1$, $I_2(S_{t-1}) = 1$ if $S_{t-1} \leq r_2$ and $cov(u_{1t}, u_{2t}) \neq 0$. The 3R-TVEqCM$_i$ has regime-dependent variance, and the only difference with the 3R-TVEqCM$_f$ is that its thresholds are estimated using an interval defined by the researcher and not by the assumption of at least 10% of the observations in each regime. The thresholds of the 2R-TVEqCM$_i$ and the 3R-TVEqCM$_f$ are estimated by grid search to minimise the AIC (eq. 3.17). 3R-TVEqCM$_u$ has also the same characteristics of the 3R-TVEqCM$_i$ and the 3R-TVEqCM$_f$ but the thresholds (r_1 and r_2) are estimated in a preliminary step based on the estimation of a threshold autoregressive model for the spread. Finally, the thresholds of the 3R-TVEqCM are estimated one-step-at-a-time to minimise $\ln(\left|\hat{\Omega}(r_1, r_2)\right|$, given that $\hat{\Omega} = (\epsilon' \epsilon)/T$ for each allowable combination of r_1 and r_2 and that the information of the regime-dependent variances Ω^i is not employed for the estimation of the thresholds.

3.5.3 Analysis of Estimates

The results of estimating the models for both 1960:1-1989:12 (the initial estimation period) and 1960:1-1998:4 (the sample including the forecast period) are presented in Table 3.3. The short-run dynamics are summarised by the long-run dynamic growth multipliers[5] to save space.

A number of interesting points arise. The estimates of the lower regime threshold are virtually the same for all the models and both periods, at around zero, so the lower regime is typically characterized by $r(s) > r(l)$. The exceptions are the MTVEqCM and the 2R-TVEqCM$_{joint}$. The threshold variable for the MTVEqCM is the change in the spread, and the threshold is estimated as a 0.39 point increase. The threshold value of the 2R-TVECM$_{joint}$ is difficult to interpret, because the threshold variable, at $r(l) - 1.54r(s)$ and $r(l) - 1.4r(s)$ for the two samples, is quite different from the spread. For the three-regime models the value of the threshold between the middle and upper regimes is around 2 3/4 for the 3R-TVEqCM$_u$ and the 3R-TVEqCM for both periods, but for the 3R-TVEqCM$_i$ and

[5]The long-run multiplier of the effect of $\Delta r(s)$ on $\Delta r(l)$ is regime specific:

$$\left(\Phi^{(i)}_{1,(1,2)} + \Phi^{(i)}_{2,(1,2)}\right) / (1 - \Phi^{(i)}_{1,(1,1)} - \Phi^{(i)}_{2,(1,1)})$$

where in $\Phi^{(i)}_{j,(s,l)}$ the superscript i refers to the regime, j to the lag, and (s,l) to the s,l^{th} element of that coefficient matrix. The multiplier of $\Delta r(l)$ on $\Delta r(s)$ is similarly defined.

the 3R-TVEqCM$_f$ is close to 1 1/2 in the first period and 2 for the whole period.

The adjustment to equilibrium in the linear model is small in absolute value (but statistically significant) for both equations. However, allowing for a threshold effect, the coefficient of the spread is much larger at around 0.6 for the $\Delta r(s)$ equation in the lower regime ($S_{t-1} < 0.6$). Thus, *ceteris paribus*, increasing spreads are necessary to re-adjust to the equilibrium, as a result of decreasing short-term interest rates $\Delta r_{t+1}(s) < 0$. Although the short-term interest rate falls when the spread is negative, there is little evidence that $r(l)$ responds.

In the upper regime (spreads in excess of 2.7 in the case of the 3R-TVEqCM$_u$ and the 3R-TVEqCM), $\Delta r(l)$ tends to respond to reduce the discrepancy. In the case of the 3R-TVEqCM, there is a numerically large but not statistically significant coefficient on the spread in the short-term interest rate equation, suggesting short rates will increase. The evidence that upper-regime spreads imply lower long-term interest rates is stronger for the 3R-TVEqCM$_u$ and the 3R-TVEqCM, than for the 3R-TVEqCM$_i$ and the 3R-TVEqCM$_f$, because the upper regime threshold is higher. In all the three-regime models, the coefficient on the spread is close to zero and statistically insignificant in the middle regime, approximately $0 < S_{t-1} < 2\ 3/4$.

The best-fitting model on AIC is the 3R-TVEqCM$_f$, because this model has the greater flexibility in the way the regimes are defined (in the 3R-TVEqCM$_u$, for example, the regimes are based on a univariate SETAR) and the variances of the residuals are allowed to depend on the regime (restricted to be the same in the 3R-TVEqCM). The three-regime TVEqCMs generally fit better than the two-regime models, even discounting the penalty for the inclusion of more parameters.

Table 3.3: Estimation results

M	T	Eq	Lower regime			Middle regime			Upper regime				
			Mult. Δr(s) Δr(l)	S_{t-1}, w_{t-1} α_L	T_L	Mult. Δr(s) Δr(l)	S_{t-1}, w_{t-1} α_M	T_M	Mult. Δr(s) Δr(l)	S_{t-1}, w_{t-1} α_U	T_U	r_1,r_2 (θ)	AIC SIC
VEqCM	A	Δr(l)	0.009	-0.025									-4.02
		Δr(s)	0.440	0.058*									-3.88
	B	Δr(l)	-0.007	-0.026*									-4.32
		Δr(s)	0.411	0.046*									-4.21
2R-TVEqCM h	A	Δr(l)	-0.132	0.037	61				0.077	-0.022	295	0.06	-4.27
		Δr(s)	1.217	0.608*					0.077	0.036			-4.01
	B	Δr(l)	-0.132	0.037	61				0.059	-0.020	395	0.06	-4.63
		Δr(s)	1.217	0.608*					0.122	0.027			-4.41
M-TVEqCM	A	Δr(l)	0.010	-0.011						-0.037		0.39	-4.05
		Δr(s)	0.440	0.044						0.127*		(1.14)	-3.90
	B	Δr(l)	-0.003	-0.009						-0.038		0.39	-4.36
		Δr(s)	0.418	0.033*						0.112*		(1.16)	-4.23
2R-TVEqCM joint	A	Δr(l)	0.142	0.006	52				-0.006	-0.010	303	-4.35	-4.27
		Δr(s)	1.883	0.185*					0.329	0.004		(1.54)	-4.00
	B	Δr(l)	0.199	0.008	83				-0.070	-0.014	372	-2.5	-4.56
		Δr(s)	1.609	0.203*					0.282	0.001		(1.40)	-4.34
3R-TVEqCM i	A	Δr(l)	-0.118	0.039	58	0.281	-0.039	171	-0.139	-0.139*	127	0.05,	-4.56
		Δr(s)	1.244	0.595*		-0.232	-0.048		0.195	0.087		1.65	-4.17
	B	Δr(l)	-0.118	0.039	58	0.198	-0.033	245	-0.090	-0.121*	153	0.05,	-4.82
		Δr(s)	1.244	0.595*		0.247	-0.026		0.171	0.102		2.0	-4.49
3R-TVEqCM u	A	Δr(l)	-0.132	0.037	61	0.099	-0.006	254	-0.071	-0.407*	41	0.06,	-4.48
		Δr(s)	1.217	0.608*		-0.085	0.004		0.217	0.187		2.74	-4.09
	B	Δr(l)	-0.132	0.037	61	0.023	-0.002	344	0.171	-0.240	51	0.06,	-4.74
		Δr(s)	1.217	0.608*		0.105	0.015		0.083	0.338		2.86	-4.41
3R-TVEqCM f	A	Δr(l)	-0.132	0.037	61	0.287	-0.032	168	-0.139	-0.139	127	0.07,	-4.57
		Δr(s)	1.217	0.608*		-0.873	-0.021		0.195	0.087		1.66	-4.18
	B	Δr(l)	-0.132	0.037	61	0.222	-0.020	240	-0.099	-0.112*	155	0.06,	-4.84
		Δr(s)	1.217	0.608*		0.052	0.005		0.169	0.101		1.92	-4.51
3R-TVEqCM	A	Δr(l)	-0.132	0.037	61	0.100	-0.021	244	-0.074	-0.426*	51	0.06,	-4.30
		Δr(s)	1.217	0.608*		-0.087	-0.019		0.183	0.016		2.62	-3.91
	B	Δr(l)	-0.132	0.037	61	0.075	-0.017	321	-0.031	-0.309*	74	0.06,	-4.61
		Δr(s)	1.217	0.608*		-0.013	0.003		0.286	0.042		2.71	-4.29

* statistically significant at 5%.

Notes: The models are summarised in Table 3.1. The estimates are based on two samples T: A for 60-89 and B for 60-98:4. The long-run multipliers (Mult.) are calculated for each equation treating Δr(s) or Δr(l) as exogenous. The α_i for i = L, M, U (lower, middle and upper regime) are the coefficients of the spread variable (S_{t-1}) or the cointegrated relationship ($w_{t-1} = r_{t-1}(l) - \theta r_{t-1}(s)$). T_i refers to the number of observations in each regime.

3.6 Forecasting Evaluation

The forecasts are generated for the period 1990:1 until 1998:4. Each time the forecast origin is moved forward by one, a new observation is included in the model and the model is re-estimated, given the delay, the autoregressive value and the number of regimes. Therefore, non-linearity tests on the sample up to 1989 are supposed to hold for the whole sample, which is a sensible assumption given that the result of the tests presented in Table 3.2 does not change with the sample size. For each origin, one to twenty-four step-ahead forecasts are generated using the bootstrap, and the forecast errors are calculated.

3.6.1 Forecast Evaluation of the System

Because of the invariance of the trace of MSFE matrix to scale and data transformation, we employ the determinant of the GFESM to evaluate the system (eq. 3.19), as described in section 3.4. The ratio of the GFESM for $h = 1, ..., 24$ is presented in Figure 3.3. Values larger than 1 mean that the model has smaller GFESM than the VEqCM. Non-linearity can improve forecasts by as much as 70% at 1 step-ahead and 5% at 24 step-ahead. The 2R-TVEqCM$_{joint}$ has the best forecast performance at all horizons, followed by the 2R-TVEqCM$_h$, the 3R-TVEqCM and the 3R-TVEqCM$_u$. Thus, using this measure, the two-regime models seem to have better forecast performance.

3.6.2 Forecasting Evaluation of Each Variable

Figures 3.4, 3.5 and 3.6 present the ratio of the MSFEs for each series, $\Delta r(l)$, $\Delta r(s)$, S for forecast horizons $h = 1, ..., 24$. The linear model MSFE forms the numerator of the ratio, so values in excess of 1 indicate an improved performance relative to the linear model. The inclusion of non-linearity has almost no influence on the forecasting of growth of the long-term interest rates. In addition, the 3R-TVEqCM$_i$ and the 3R-TVEqCM$_f$ have forecasts that are 7% worse than the linear ones for the first three steps-ahead.

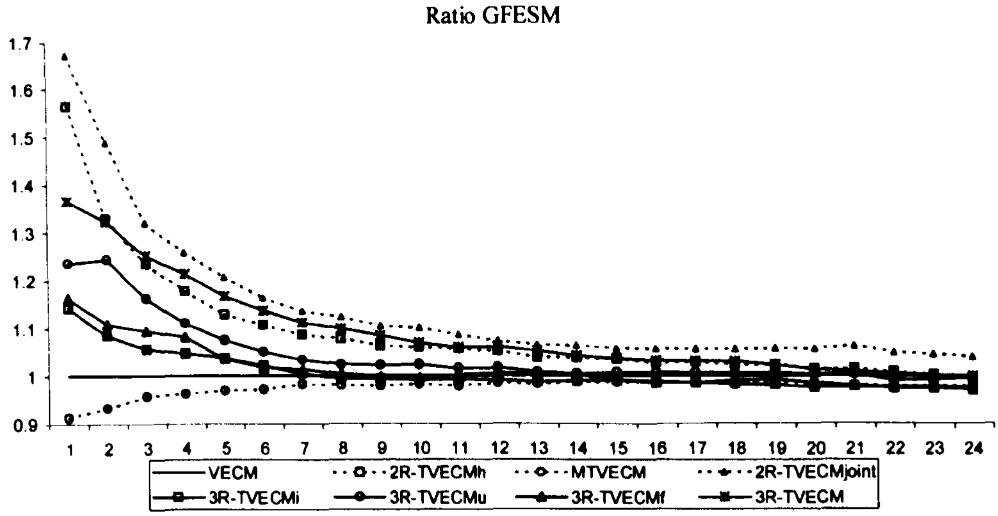

Figure 3.3: Comparing GFESMs for $h = 1, ..., 24$

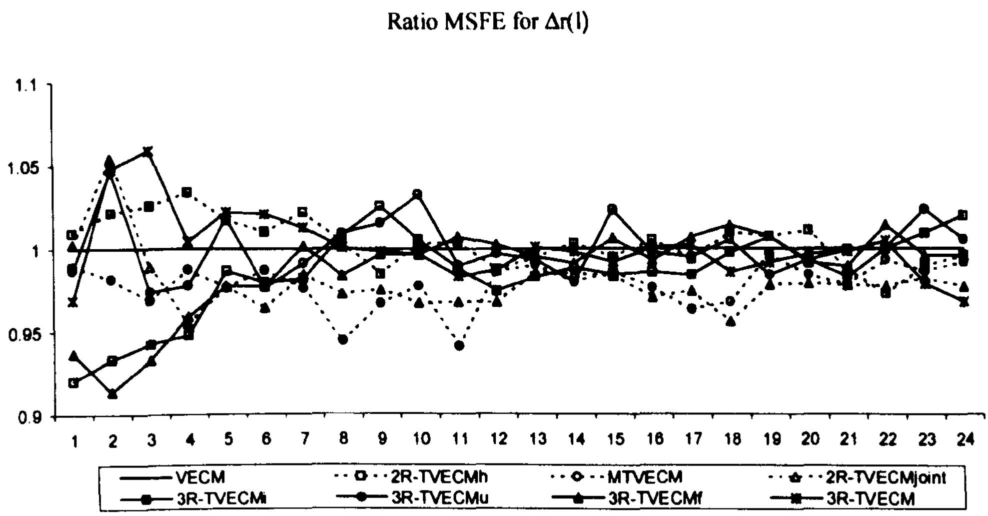

Figure 3.4: Comparing MSFEs for $\Delta r(l)$ for $h = 1, ..., 24$

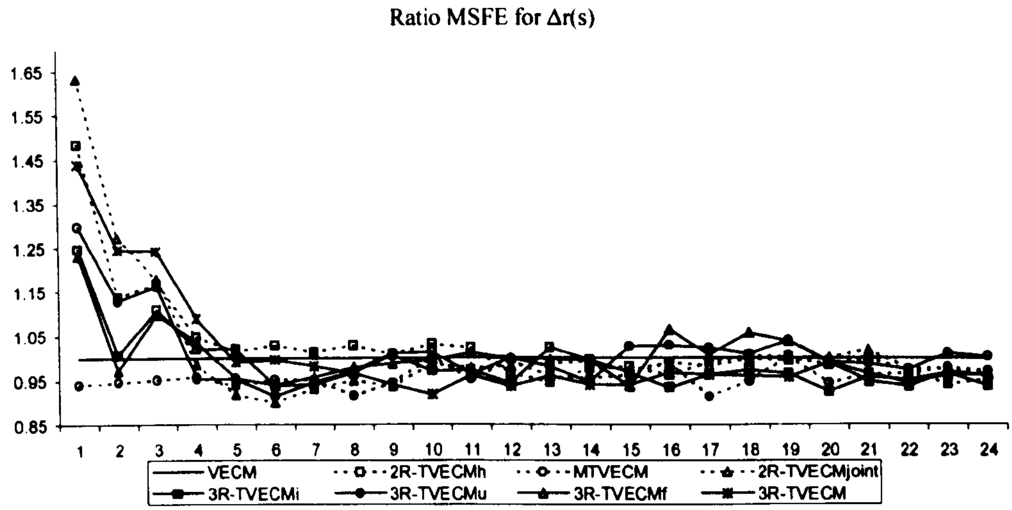

Figure 3.5: Comparing MSFEs for $\Delta r(s)$ for $h = 1, ..., 24$

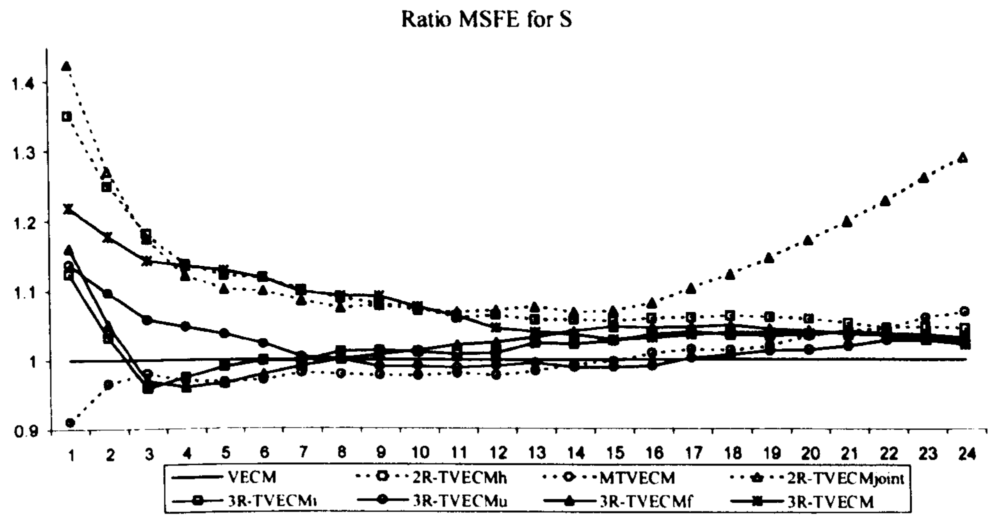

Figure 3.6: Comparing MSFEs for S for $h = 1, ..., 24$

Tables 3.4 and 3.5 present the p-values of forecast accuracy and encompassing tests for predicting the growth of the long-term interest rate. The rank of the ADM test is based on the number of times that a model has better forecast accuracy compared to other models (using 10% significance level). The rank of the encompassing test classifies the models depending on how many times a model is not encompassed by other models (at 10%). Thus, at one step-ahead, the linear model is a good forecaster compared with the non-linear models. For two steps-ahead, the 2R-TVEqCM$_h$, the 2R-TVEqCM$_{joint}$, the 3R-TVEqCM$_u$ and the 3R-TVEqCM give better performance than the other models. The 2R-TVECM$_h$ has better accuracy than most of the models at $h = 4$. At eight steps-ahead, the MTVEqCM and the 2R-TVEqCM$_{joint}$ have inferior performance. The encompassing test of Table 3.5 has better discriminating power, and at one step-ahead, the 3R-TVEqCM$_i$ produces poor forecasts, while the 2R-TVEqCM$_h$ and the 2R-TVEqCM$_{joint}$ have the best performance. At $h = 2$, the 2R-TVEqCM$_{joint}$ and the 3R-TVEqCM$_u$ have more information. Similar to the accuracy test, at $h = 8$, the encompassing test indicates that models that do not have the spread as the cointegration vector are poor forecasters. Summarising, there are some significant gains in the forecasting of non-linearities at short horizons, but the linear model is a good forecaster at $h = 1, 8$.

The improvement in the accuracy of forecasts of $\Delta r\,(s)$ from allowing for non-linearity is much more marked. The 2R-TVEqCM$_h$, the 2R-TVEqCM$_{joint}$ and the 3R-TVEqCM have MSFEs around 50% lower at a horizon of one, and the 3R-TVEqCM is 10% more accurate at four steps-ahead. This supports the finding of non-linearity in the short-term interest rate by Pfann et al. (1996), and also the single equation equilibrium correction models for the short-term rate (Anderson, 1997; Van Dijk and Franses, 2000). The ADM tests for these forecast errors are presented in Table 3.6. At one step-ahead, the MTVEqCM has the worst forecasts, and the 2R-TVEqCM$_{joint}$, the 3R-TVEqCM and the 2R-TVEqCM$_h$ are good forecasters. At two steps-ahead, the results are similar to one step-ahead, except that the 3R-TVEqCM$_i$ and the 2R-TVEqCM$_f$ generate forecasts as poor as the linear model. At four steps-ahead, the 2R-TVEqCM$_h$ and the 3R-TVEqCM still forecast significantly better than the other models. It is hard to discriminate among models at eight steps-ahead, thus the linear model is a relatively good forecaster. Overall, the 2R-TVEqCM$_h$, the 2R-TVEqCM$_{joint}$ and the 3R-TVEqCM appear to be best. The encompassing test, presented in Table 3.7, confirms the forecast accuracy test: the MTVEqCM, the 3R-TVEqCM$_i$ and the 3R-TVEqCM$_f$ are the worse forecasters compared to the other non-linear models, and the 2R-TVEqCM$_h$, the 2R-TVEqCM$_{joint}$ and the 3R-TVEqCM encompass the linear model at one and two steps-ahead. When $h = 8$, the linear model is a better forecaster than the non-linear models, but it is equivalent to the 2R-TVEqCM$_h$. Summarising, the presence of non-linearity improves the forecasts of $\Delta r(s)$ at short horizons, when the 2R-TVEqCM$_h$, the 2R-TVEqCM$_{joint}$ and the 3R-TVEqCM are the chosen specification.

The MSFE ratios for the spread indicate gains between 15% and 35% with the inclusion of non-linearity at one step-ahead, and Figure 3.6 suggests some improvement even at long horizons, mainly for the MTVEqCM and the 2R-TVEqCM$_{joint}$, which have $\theta \neq 1$. The p-values of forecast accuracy and the encompassing tests for the spread are presented in Tables 3.8 and 3.9. The VEqCM and the MTVEqCM have inferior forecast accuracy and are encompassed by the other models for one step-ahead forecasts. The 2R-TVEqCM$_h$, the 2R-TVEqCM$_{joint}$ and the 3R-TVEqCM are the best forecasts at $h = 1, 2$, observing the encompassing test of Table 3.9. Additionally, both tests suggest that 3R-TVEqCM$_i$, the 3R-TVEqCM$_u$ and the 3R-TVEqCM$_f$ have equivalently poor forecasts for $h = 1, 2, 4$. At four steps-ahead, the 2R-TVEqCM$_h$ and the 3R-TVEqCM are the best forecasters. At $h = 8$, both tests indicate that the 2R-TVEqCM$_h$ and the 3R-TVEqCM generate better forecasts than other non-linear models and also the linear model. Therefore, the tests confirm that the inclusion of non-linearity improves the forecasting at longer horizons (until $h = 8$). However, we cannot evaluate longer horizons ($h = 24$), because of the bias in calculating the autocorrelation function when h is large.

The 2R-TVEqCM$_h$ forecasts well, despite the fact that negative spreads do not occur over the forecast period, and that it is during these times that it differs from the linear model and incorporates a significant 'levels effect'. But notice that the linear model and the 2R-TVEqCM$_h$ suggest quite different dynamic responses (as summarized in the dynamic multipliers reported in Table 3.3). So the failure of the linear model to appropriately model behaviour when the spread is negative affects the dynamic responses of the linear model, so that its performance is inferior even at more commonly observed values of the spread.

3.6.3 Forecasting Evaluation Conditional upon a Regime

The literature suggests that non-linear models often record increased gains over linear comparators in some states of nature, but not others (see, for example, Tiao and Tsay, 1994, Clements and Smith, 1997; 1999a). Comparing models on MSFE over the whole forecast period, as we have done, is likely to under-estimate the gains that arise conditional on being in a specific regime.

The form of the estimated non-linear models suggests that responses to the equilibrium correction term differ markedly from those implied by the VEqCM at low (i.e., negative) and high (in excess of 2 3/4) values of the spread, so that we would like to consider forecasts of $r\,(s)$ and S when the value of the spread at the forecast origin is negative, and forecasts of $r\,(l)$ when the value of the spread exceeds 2.7. For the majority of the data points the spread is between these two extremes, so the VEqCM, whose parameter estimates are effectively an average of the regime-specific values, is characterized by very modest mean reversion, and its forecasts are not too dissimilar to those of the non-linear models. However, the spread is never negative over the forecast period. Instead, results of a forecast evaluation conditioned on $S_{t-1} < 0$ are reported based on the simulated data in section 3.6.5.

Table 3.4: Augmented Diebold and Mariano Forecast accuracy tests for first differences of long-term rates

$h = 1$

model j	model i: 1	2	3	4	5	6	7	8	Rank
1	-								1
2	0.354	-							1
3	0.732	0.741	-						1
4	0.477	0.581	0.375	-					2
5	0.927	0.936	0.910	0.895	-				2
6	0.610	0.693	0.518	0.623	0.123	-			2
7	0.873	0.890	0.839	0.832	0.192	0.795	-		2
8	0.667	0.723	0.605	0.683	0.244	0.654	0.335	-	2

$h = 2$

model j	1	2	3	4	5	6	7	8	Rank
1	-								2
2	0.126	-							1
3	0.884	0.922	-						2
4	0.138	0.215	0.099	-					1
5	0.826	0.881	0.750	0.894	-				2
6	0.167	0.268	0.108	0.553	0.061	-			1
7	0.860	0.903	0.800	0.911	0.853	0.954	-		2
8	0.291	0.372	0.232	0.530	0.093	0.491	0.064	-	1

$h = 4$

model j	1	2	3	4	5	6	7	8	Rank
1	-								2
2	0.001	-							1
3	0.738	0.965	-						3
4	0.933	0.997	0.804	-					3
5	0.922	0.987	0.842	0.562	-				3
6	0.749	0.939	0.582	0.261	0.225	-			3
7	0.812	0.944	0.729	0.462	0.242	0.650	-		3
8	0.433	0.876	0.315	0.074	0.048	0.183	0.156	-	2

$h = 8$

model j	1	2	3	4	5	6	7	8	Rank
1	-								2
2	0.379	-							2
3	0.999	0.999	-						3
4	0.868	0.842	0.229	-					3
5	0.386	0.446	0.010	0.119	-				1
6	0.214	0.407	0.005	0.039	0.487	-			1
7	0.733	0.752	0.016	0.385	0.905	0.755	-		2
8	0.480	0.558	0.009	0.182	0.636	0.632	0.219	-	2

Models: (1)VEqCM; (2) 2R-TVEqCMh; (3) MTVEqCM; (4) 2R-TVEqCMjoint; (5) 3R-TVEqCMi; (6) 3R-TVEqCMu; (7) 3R-TVEqCMf; (8) 3R-TVECM.
Note: p-values for the forecast accuracy test (3.20). Values underlined mean that H0 that model i is as accurate as model j is rejected at 10%. Rank classifies model j by the number of times that has better forecast accuracy compared to the other models.

Table 3.5: Encompassing tests for first differences of long-term rates

$h = 1$

		model i								
		1	2	3	4	5	6	7	8	Rank
m	1	-	0.410	0.151	0.168	0.002	0.065	0.004	0.025	2
o	2	0.168	-	0.099	0.144	0.001	0.066	0.003	0.030	1
d	3	0.573	0.474	-	0.202	0.002	0.088	0.007	0.042	2
e	4	0.134	0.252	0.072	-	0.001	0.049	0.002	0.016	1
l	5	0.435	0.472	0.392	0.234	-	0.345	0.596	0.149	4
j	6	0.126	0.255	0.104	0.117	0.005	-	0.009	0.096	3
	7	0.301	0.341	0.258	0.123	0.090	0.210	-	0.084	3
	8	0.096	0.160	0.091	0.105	0.009	0.243	0.012	-	2

$h = 2$

		1	2	3	4	5	6	7	8	Rank
m	1	-	0.757	0.066	0.608	0.030	0.540	0.027	0.338	3
o	2	0.058	-	0.022	0.548	0.016	0.419	0.015	0.272	2
d	3	0.812	0.783	-	0.653	0.046	0.610	0.042	0.392	3
e	4	0.033	0.075	0.019	-	0.009	0.133	0.009	0.090	1
l	5	0.430	0.506	0.306	0.442	-	0.642	0.073	0.604	4
j	6	0.043	0.076	0.019	0.189	0.005	-	0.004	0.213	1
	7	0.502	0.560	0.386	0.467	0.736	0.663	-	0.641	5
	8	0.076	0.114	0.048	0.150	0.011	0.232	0.005	-	2

$h = 4$

		1	2	3	4	5	6	7	8	Rank
m	1	-	0.997	0.174	0.045	0.011	0.174	0.050	0.419	2
o	2	<0.001	-	0.012	0.001	0.001	0.041	0.010	0.062	1
d	3	0.631	0.912	-	0.132	0.034	0.259	0.076	0.490	3
e	4	0.902	0.993	0.715	-	0.256	0.573	0.307	0.873	5
l	5	0.738	0.911	0.603	0.369	-	0.461	0.560	0.841	5
j	6	0.638	0.903	0.382	0.134	0.051	-	0.126	0.692	4
	7	0.563	0.785	0.500	0.249	0.104	0.349	-	0.632	5
	8	0.294	0.773	0.149	0.042	0.012	0.104	0.054	-	2

$h = 8$

		1	2	3	4	5	6	7	8	Rank
m	1	-	0.504	0.001	0.097	0.508	0.660	0.206	0.393	1
o	2	0.273	-	<0.001	0.100	0.413	0.439	0.169	0.302	1
d	3	0.497	0.487	-	0.688	0.503	0.484	0.249	0.421	1
e	4	0.858	0.822	0.159	-	0.824	0.920	0.584	0.802	2
l	5	0.287	0.308	0.005	0.072	-	0.353	0.051	0.216	1
j	6	0.122	0.270	0.002	0.013	0.339	-	0.164	0.241	1
	7	0.659	0.647	0.003	0.312	0.830	0.653	-	0.673	2
	8	0.351	0.404	<0.001	0.143	0.462	0.479	0.133	-	2

Models: (1)VEqCM; (2) 2R-TVEqCMh; (3) MTVEqCM; (4) 2R-TVEqCMjoint; (5) 3R-TVEqCMi; (6) 3R-TVEqCMu; (7) 3R-TVEqCMf; (8) 3R-TVECM.

Note: p-value for forecasting encompassing test (3.21). Values underlined mean that the H0 that model i encompass model j is rejected at 10%. Rank classifies model j by how many times it is not encompassed by other models, compared with the other models.

Table 3.6: Augmented Diebold and Mariano forecast acuracy tests for first differences of short-term rates

$h = 1$

		model i								
		1	2	3	4	5	6	7	8	Rank
m	1	-								4
o	2	<0.001	-							2
d	3	0.960	>0.999	-						5
e	4	<0.001	0.053	<0.001	-					1
l	5	0.002	0.998	0.001	0.999	-				3
j	6	0.018	0.942	0.011	0.992	0.339	-			3
	7	0.009	0.998	0.004	0.999	0.705	0.725	-		3
	8	0.003	0.651	0.002	0.944	0.043	0.076	0.030	-	2

$h = 2$

		1	2	3	4	5	6	7	8	Rank
m	1	-								3
o	2	0.012	-							2
d	3	0.442	0.991	-						4
e	4	0.007	0.056	0.007	-					1
l	5	0.477	0.907	0.295	0.993	-				4
j	6	0.072	0.551	0.050	0.914	0.113	-			2
	7	0.628	0.973	0.400	0.999	0.807	0.956	-		4
	8	0.008	0.114	0.007	0.569	0.018	0.150	0.003	-	2

$h = 4$

		1	2	3	4	5	6	7	8	Rank
m	1	-								2
o	2	0.002	-							1
d	3	0.841	0.944	-						2
e	4	0.573	0.770	0.390	-					2
l	5	0.369	0.700	0.242	0.289	-				2
j	6	0.880	0.987	0.524	0.631	0.815	-			2
	7	0.226	0.587	0.127	0.208	0.349	0.129	-		2
	8	0.013	0.223	0.019	0.183	0.194	0.002	0.232	-	1

$h = 8$

		1	2	3	4	5	6	7	8	Rank
m	1	-								1
o	2	0.200	-							1
d	3	0.922	0.912	-						2
e	4	0.858	0.894	0.343	-					2
l	5	0.771	0.820	0.108	0.373	-				2
j	6	0.688	0.788	0.172	0.399	0.478	-			2
	7	0.631	0.726	0.016	0.357	0.302	0.443	-		1
	8	0.692	0.755	0.023	0.432	0.512	0.546	0.620	-	1

Note: See notes of Table 3.4.

Table 3.7: Encompassing tests for first differences of short-term rates

$h = 1$

		model i								
		1	2	3	4	5	6	7	8	Rank
m	1	-	0.993	0.013	0.955	0.693	0.169	0.421	0.487	6
o	2	<0.001	-	<0.001	0.297	<0.001	<0.001	<0.001	0.019	2
d	3	0.909	0.983	-	0.963	0.744	0.170	0.478	0.464	7
e	4	<0.001	<0.001	<0.001	-	<0.001	<0.001	<0.001	0.007	1
l	5	<0.001	0.877	<0.041	0.761	-	0.057	0.080	0.456	3
j	6	<0.001	0.171	<0.001	0.210	0.004	-	0.002	0.396	4
	7	<0.001	0.914	<0.001	0.682	0.360	0.069	-	0.505	3
	8	<0.001	0.062	<0.001	0.074	0.002	0.011	<0.001	-	1

$h = 2$

		1	2	3	4	5	6	7	8	Rank
m	1	-	0.938	0.011	0.955	0.108	0.686	0.044	0.813	4
o	2	0.002	-	0.001	0.267	0.006	0.142	0.001	0.449	2
d	3	0.795	0.926	-	0.963	0.206	0.744	0.112	0.831	5
e	4	<0.001	0.017	0.001	-	<0.001	0.002	<0.001	0.007	1
l	5	0.092	0.442	0.060	0.761	-	0.352	0.076	0.737	3
j	6	0.017	0.223	0.015	0.210	0.005	-	0.001	0.402	2
	7	0.105	0.531	0.055	0.682	0.600	0.436	-	0.858	4
	8	<0.001	0.012	<0.001	0.074	0.001	0.033	0.004	-	1

$h = 4$

		1	2	3	4	5	6	7	8	Rank
m	1	-	0.949	0.500	0.155	0.349	0.059	0.437	0.945	4
o	2	<0.001	-	0.126	0.066	0.061	0.004	0.128	0.546	2
d	3	0.640	0.751	-	0.292	0.488	0.206	0.595	0.901	5
e	4	0.255	0.409	0.149	-	0.396	0.133	0.372	0.567	5
l	5	0.161	0.323	0.084	0.123	-	0.071	0.302	0.601	4
j	6	0.761	0.953	0.217	0.311	0.559	-	0.647	0.990	5
	7	0.068	0.244	0.020	0.049	0.132	0.045	-	0.557	3
	8	0.003	0.093	0.003	0.065	0.080	<0.001	0.102	-	1

$h = 8$

		1	2	3	4	5	6	7	8	Rank
m	1	-	0.663	0.039	0.053	0.096	0.226	0.240	0.232	1
o	2	0.103	-	0.057	0.034	0.093	0.131	0.177	0.154	1
d	3	0.855	0.816	-	0.460	0.561	0.617	0.874	0.815	4
e	4	0.695	0.712	0.186	-	0.273	0.400	0.447	0.417	4
l	5	0.562	0.678	0.018	0.151	-	0.298	0.271	0.273	3
j	6	0.585	0.675	0.070	0.219	0.264	-	0.371	0.209	3
	7	0.478	0.602	0.001	0.221	0.047	0.257	-	0.196	2
	8	0.599	0.634	0.001	0.297	0.275	0.276	0.387	-	3

Note: See notes of table 3.5.

Table 3.8: Augmented Diebold and Mariano forecast accuracy tests for the spread

$h = 1$

		model i								
		1	2	3	4	5	6	7	8	Rank
m	1	-								4
o	2	≤0.001	-							1
d	3	≥0.999	≥0.999	-						5
e	4	≤0.001	0.102	≤0.001	-					1
l	5	0.054	0.996	0.004	0.999	-				3
j	6	0.090	0.979	0.016	0.993	0.436	-			3
	7	0.024	0.986	0.002	0.997	0.072	0.393	-		2
	8	0.017	0.906	0.002	0.963	0.105	0.105	0.236	-	2

$h = 2$

		1	2	3	4	5	6	7	8	Rank
m	1	-								4
o	2	≤0.001	-							2
d	3	0.925	0.999	-						5
e	4	0.007	0.378	0.001	-					1
l	5	0.367	0.966	0.257	0.978	-				3
j	6	0.191	0.898	0.135	0.908	0.211	-			3
	7	0.304	0.940	0.210	0.949	0.224	0.702	-		2
	8	0.086	0.714	0.067	0.738	0.067	0.091	0.116	-	2

$h = 4$

		1	2	3	4	5	6	7	8	Rank
m	1	-								3
o	2	0.010	-							2
d	3	0.804	0.978	-						3
e	4	0.186	0.561	0.131	-					3
l	5	0.579	0.861	0.485	0.800	-				3
j	6	0.362	0.723	0.281	0.633	0.180	-			3
	7	0.628	0.881	0.531	0.815	0.728	0.862	-		3
	8	0.124	0.512	0.093	0.467	0.078	0.118	0.061	-	1

$h = 8$

		1	2	3	4	5	6	7	8	Rank
m	1	-								2
o	2	0.046	-							1
d	3	0.649	0.867	-						2
e	4	0.354	0.530	0.323	-					2
l	5	0.462	0.754	0.404	0.630	-				2
j	6	0.497	0.748	0.419	0.623	0.565	-			2
	7	0.497	0.783	0.422	0.637	0.622	0.499	-		2
	8	0.125	0.479	0.124	0.465	0.125	0.120	0.069	-	1

Note: See notes of Table 3.4.

Table 3.9: Encompassing tests for the spread

$h = 1$

		model i								
		1	2	3	4	5	6	7	8	Rank
m	1	-	>0.999	0.001	0.996	0.347	0.032	0.428	0.065	5
o	2	≤0.001	-	≤0.001	0.656	≤0.001	≤0.001	≤0.001	≤0.001	2
d	3	<0.999	>0.999	-	0.999	0.659	0.086	0.707	0.148	6
e	4	≤0.001	0.019	≤0.001	-	≤0.001	≤0.001	≤0.001	≤0.001	2
l	5	0.001	0.858	≤0.001	0.917	-	0.020	0.795	0.098	4
j	6	≤0.001	0.134	≤0.001	0.267	0.012	-	0.030	0.416	4
	7	≤0.001	0.733	≤0.001	0.831	0.028	0.010	-	0.047	3
	8	≤0.001	0.022	≤0.001	0.075	≤0.001	0.001	0.003	-	1

$h = 2$

		1	2	3	4	5	6	7	8	Rank
m	1	-	0.994	0.081	0.957	0.167	0.102	0.178	0.242	5
o	2	≤0.001	-	≤0.001	0.303	0.006	0.007	0.014	0.026	1
d	3	0.862	0.989	-	0.978	0.241	0.149	0.254	0.289	6
e	4	≤0.001	0.123	≤0.001	-	0.003	0.009	0.007	0.020	1
l	5	0.063	0.807	0.032	0.771	-	0.293	0.585	0.599	4
j	6	0.005	0.259	0.003	0.213	0.018	-	0.043	0.695	2
	7	0.045	0.745	0.025	0.664	0.106	0.242	-	0.511	3
	8	0.002	0.125	0.002	0.121	0.009	0.021	0.022	-	1

$h = 4$

		1	2	3	4	5	6	7	8	Rank
m	1	-	0.978	0.125	0.574	0.202	0.338	0.178	0.614	4
o	2	0.004	-	0.007	0.221	0.055	0.128	0.051	0.215	2
d	3	0.710	0.943	-	0.673	0.252	0.390	0.231	0.639	4
e	4	0.068	0.310	0.051	-	0.084	0.172	0.076	0.237	2
l	5	0.336	0.667	0.241	0.490	-	0.622	0.123	0.804	4
j	6	0.143	0.434	0.094	0.237	0.077	-	0.060	0.764	3
	7	0.399	0.725	0.295	0.521	0.542	0.708	-	0.858	4
	8	0.030	0.188	0.017	0.109	0.035	0.053	0.031	-	1

$h = 8$

		1	2	3	4	5	6	7	8	Rank
m	1	-	0.936	0.269	0.426	0.410	0.372	0.383	0.797	3
o	2	0.033	-	0.084	0.275	0.153	0.188	0.147	0.391	2
d	3	0.560	0.794	-	0.487	0.452	0.418	0.441	0.779	3
e	4	0.195	0.331	0.200	-	0.211	0.247	0.209	0.329	3
l	5	0.335	0.614	0.275	0.363	-	0.319	0.249	0.772	3
j	6	0.363	0.648	0.289	0.381	0.445	-	0.341	0.830	3
	7	0.375	0.675	0.298	0.375	0.486	0.316	-	0.881	3
	8	0.074	0.329	0.065	0.226	0.076	0.084	0.045	-	1

Note: See notes of Table 3.5.

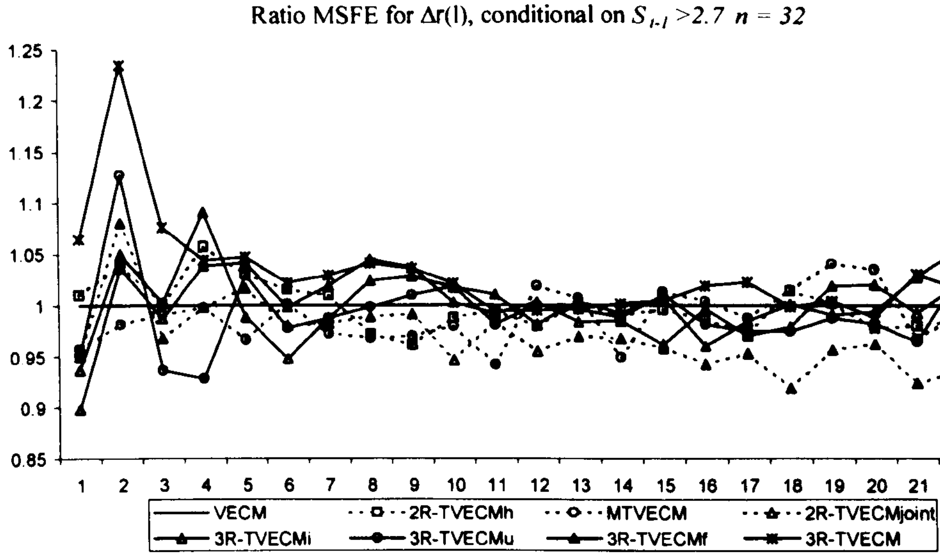

Figure 3.7: Comparison of MSFEs for $\Delta r(l)$ conditional on $S_{t-1} > 2.7$

The evaluation of forecasts for $\Delta r(l)$ when the spread is larger than 2.7 is presented in Figure 3.7. In this case, the three-regime TVEqCMs, with emphasis on the 3R-TVEqCM$_u$ and the 3R-TVEqCM, estimate a negative and significant effect of the spread on the $\Delta r(l)$. Comparing Figure 3.7 with Figure 3.4, it is possible to observe that the 3R-TVEqCM$_u$ and the 3R-TVEqCM give a better performance. The 3R-TVEqCM improves the forecast by around 23% for the second step and in around 10% for the other horizons until four. Therefore, the inclusion of a third regime improves the forecast of $\Delta r(l)$, conditional on a large spread.

3.6.4 Forecasting Evaluation and Model Fit

Table 3.3 presents values of AIC and SIC for the first and the last sample employed in the forecasting evaluation to see the extent to which model fit and forecast performance are correlated. On the basis of model fit, as measured by SIC, only MTVEqCM is worse than the linear. However, the best forecasting models – the 2R-TVEqCM$_h$ and the 3R-TVEqCM – have SIC values in the middle range while 3R-TVEqCM$_f$ has the smallest SIC. The finding that models that fit better do not necessarily forecast better is in line with a growing body of evidence, e.g., Ramsey (1996), Escribano and Granger (1998) and Clements and Krolzig (1998), who discuss this from a number of different perspectives.

3.6.5 Simulation Exercise

One of the problems of this forecast evaluation is that, as can be observed in Figure 3.2, the value of the spread is never negative in the out-of-sample period. Thus we resort to Monte Carlo simulation to examine the relative performances of the models when negative spreads do occur in the forecast period. This exercise is meaningful because there is no reason to suppose that the event of negative spread would not happen again in the future, implying that the fact that it did not occur in the forecasting period may affect the results of the forecasting evaluation[6].

Based on the Monte Carlo evaluation proposed by Clements and Krolzig (1998), we generate data from the best two- and three-regime models, the 2R-TVEqCM$_h$ and the 3R-TVEqCM, using pseudo-random numbers transformed to have the appropriate covariance matrix, and the estimated (full-sample) values of the models parameters. The number of observations is similar to that in the empirical exercise, and the last 24 observations are kept back for forecasting. The VEqCM, the 2R-TVEqCM$_h$ and the 3R-TVEqCM are then estimated, and used to generate forecasts. Our results are based on Monte Carlo estimates of the MSFEs from 2000 replications, where on each of the replications, Monte Carlo is used to generate a single sequence of multi-period forecasts.

When the 2R-TVEqCM$_h$ is the DGP, we calculate MSFEs for forecasts with $S_{t-1} \leq 0$ in the origin. This is the regime that did not occur in the out-of-sample period, but represents 10% of the simulations. Conditional on this regime, MSFEs for $\Delta r(s)$ and S are presented in Figures 3.8 and 3.9. Non-linearity gains are of 40% at one step-ahead for $\Delta r(s)$

[6] *Postscript*: In fact, the spread is negative in the last three months of 2000.

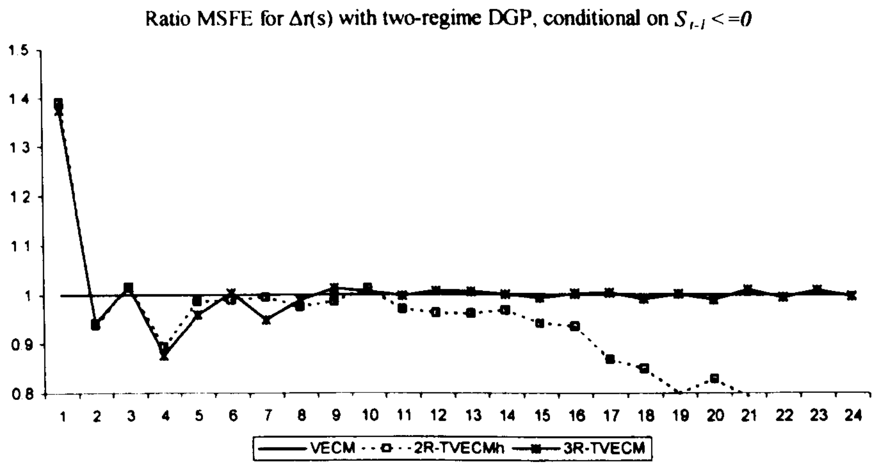

Figure 3.8: MSFEs for $\Delta r(s)$ with data generated from the 2R-TVECM, conditional on $S_{t-1} \leq 0$

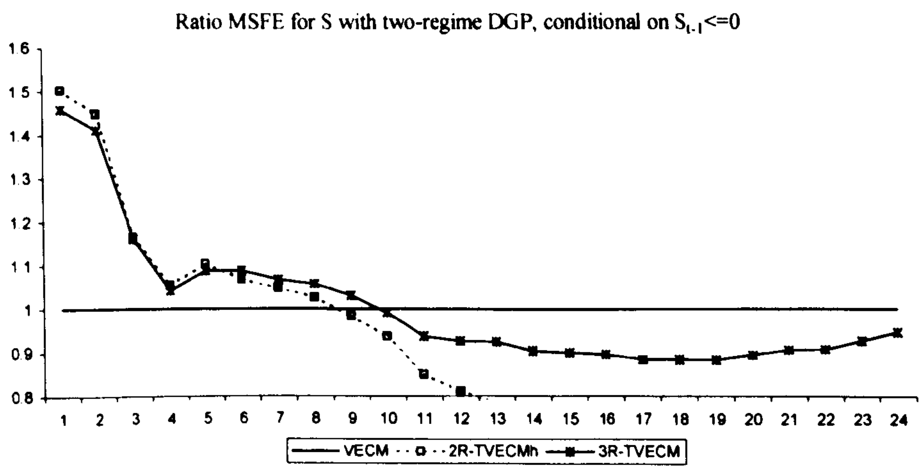

Figure 3.9: MSFEs for S with data generated from the 2R-TVECM, conditional on $S_{t-1} \leq 0$

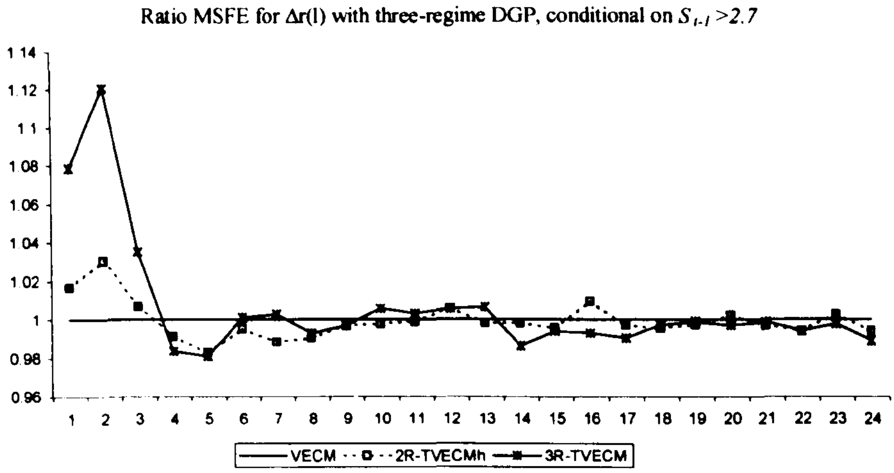

Figure 3.10: MSFEs for $\Delta r(l)$ with data generated from the 3R-TVECM, conditional on $S_{t-1} > 2.7$

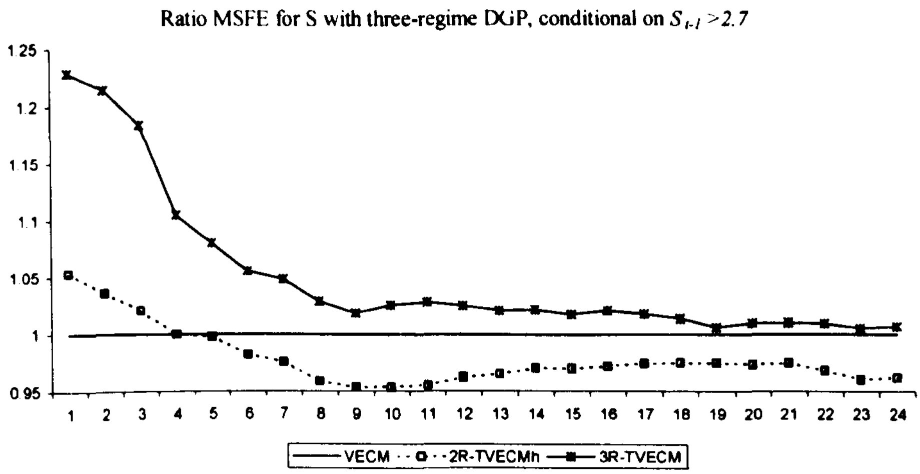

Figure 3.11: MSFEs for S with data generated from the 3R-TVECM, conditional on $S_{t-1} > 2.7$

but are not sustained at longer horizons. In contrast, the gains for predicting the spread are larger and until $h = 8$.

For the case in which the 3R-TVEqCM is the DGP, we consider the forecasts when $S_{t-1} > 2.7$, representing 20% of the simulations. The gains of a third regime (Figures 3.10 and 3.11) are larger compared to Figures 3.8 and 3.9, although Figures 3.8 and 3.9 are not a good benchmark because the 2R-TVEqCM$_h$ and the 3R-TVEqCM have the same dynamics in the lower regime. The gains of 25% in MSFESs for the forecast spread (Figure 3.11) highlight the relevance of allowing for a third regime, at least in this simulation study.

Summarising, non-linear models improve the short-horizon forecasts of $\Delta r(s)$ and of $\Delta r(l)$, conditional on the regime. Non-linearities reduce MSFEs in predicting the spread at longer horizons ($h = 8$). In addition, only conditional on $S_{t-1} > 2.7$, forecasts from three-regime models are significantly better than two-regimes.

3.7 Analysis of the Robustness of the Forecasting Results

In the previous section, we evaluated models estimated employing testing and modelling procedures suggested by the literature. The results indicate that TVEqCMs have better forecasting performance at short horizons when compared with a VEqCM. Because the non-linearity in the TVEqCMs affects both long-run and short-run dynamics, the improvement in accuracy may be because non-linearity improves the response to the long-run disequilibrium or as a result of non-linearity in short-run dynamics. The setup of the evaluation of the last section does not allow us to observe these differences. Another feature that should be analysed is how much does the value of the cointegration vector affect long-horizon forecasts. This is connected with the fact that the models with the same cointegration vector have similar long-horizon behaviour for the prediction of the spread. For example, the MTVEqCM and the 2R-TVEqCM$_h$ do not have MSFE similar to the models that assume

$\theta = 1$ at long horizons ($h = 24$).

In addition, the fact that some models account for regime-dependent variance and some others do not may affect the forecasting performance. The assumption of regime-dependence may not affect the estimation of the parameters but it affects forecasts that rely on bootstrapping. When the thresholds are estimated using the minimisation of the residual sum of squares, it does not matter if the model is written as

$$\mathbf{z}_t = (\boldsymbol{\beta}^{(1)}\mathbf{x}_{t-1} + \boldsymbol{\alpha}^{(1)}w_{t-1})I(w_{t-1}) + (\boldsymbol{\beta}^{(2)}\mathbf{x}_{t-1} + \boldsymbol{\alpha}^{(2)}w_{t-1})(1 - I(w_{t-1})) + \boldsymbol{\epsilon}_t \tag{3.22}$$

or as

$$\mathbf{z}_t = (\boldsymbol{\beta}^{(1)}\mathbf{x}_{t-1} + \boldsymbol{\alpha}^{(1)}w_{t-1} + \boldsymbol{\epsilon}_t^{(1)})I(w_{t-1}) + (\boldsymbol{\beta}^{(2)}\mathbf{x}_{t-1} + \boldsymbol{\alpha}^{(2)}w_{t-1} + \boldsymbol{\epsilon}_t^{(2)})(1 - I(w_{t-1})), \tag{3.23}$$

the estimates of $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ are the same because the models have the same residual sum of squares. The latter equation shows how the vector $\boldsymbol{\epsilon}_t$ can be decomposed into $\mathbf{v}_t = (\boldsymbol{\epsilon}_t^{(1)}I(w_{t-1}), \boldsymbol{\epsilon}_t^{(2)}(1 - I(w_{t-1})))'$. This decomposition does not affect the sum of the square of the residuals of the model ($\sum_{i=1}^{T}\boldsymbol{\epsilon}_i^2 = \sum_{i=1}^{T}\mathbf{v}_i^2$). However, when it is supposed that $\boldsymbol{\epsilon}_t^{(1)} \sim N(0, \Omega^{(1)})$ and $\boldsymbol{\epsilon}_t^{(2)} \sim N(0, \Omega^{(2)})$ (regime dependent-variance), the variance-covariance matrix of the coefficients ($\Omega^i \otimes (X'X)^{-1}$) and the forecasts may change. The forecasts of the non-linear model are affected because, depending on w_{T+j-1}, the residuals will be drawn from different distributions and this affects $\mathbf{z}_{T+j}$, which is used to calculate w_{T+j}. Variances depending on the regime may change the probability of a forecast being generated from each regime.

The thresholds estimated for the 3R-TVEqCM$_f$ and the 3R-TVEqCM, which have, respectively, regime-dependent variance and variance constant across regimes, are different. This results from the fact that the thresholds are not chosen to minimise the residual sum of squares ($tr[\boldsymbol{\epsilon}'\boldsymbol{\epsilon}]$) but to minimise the determinant of the variance-covariance matrix of the residuals $\Omega = ((\boldsymbol{\epsilon}'\boldsymbol{\epsilon})/T)$. This is so because the disturbances are supposed to be contemporaneously correlated, thus the determinant is employed instead of the trace. Because the sum of $\det(\Omega^{(i)})$ over regimes is not the same as $\det(\Omega)$ (effect of covariances), the models have different estimated thresholds. One could argue that this difference may result from the application of the one-step-at-a-time algorithm to the grid search for thresholds in the 3R-TVEqCM. However, comparing the complete grid search, splitting the sample among regimes, with the one-step-at-a-time method, when the residual sum of squares is the criterion, the thresholds (0.06, 2.06) are the same (as one would expect given the consistency proved by Bai, 1997).

Therefore, this section analyses the robustness of the results of the last section concerning three characteristics: (a) cointegration vector (no cointegration, using the spread, or using the first step of the Engle and Granger (1987) procedure); (b) non-linearity effect (only in the short-run dynamics (TVAR), or in both short- and long-run dynamics in a cointegration system); and (c) assumptions about the variance of residuals (forecasting using regime-dependent bootstrap residuals (heteroscedastic) and without regime-dependent bootstrap residuals (homoscedastic)). The analysis of how much the forecasting performance is from non-linearities in the short- or in the long-run dynamics is conducted using a simulation exercise similar to the one of section 3.6.5. The three characteristics are jointly observed using a different group of competitors in a new forecasting competition in section 3.7.2. The comparison of the RMSFEs of the models of section 3.7.2 with linear autoregressive models is analysed in section 3.7.3.

3.7.1 Another Simulation Exercise

The objective of this simulation exercise is to compare the forecasting performance of TVEqCMs (non-linearity in the short-run and in long-run dynamics of the system) and TVARs (no equilibrium correction term, but non-linearity in the dynamics). This comparison is conducted conditional on $S_{t-1} \leq 0$ and $S_{t-1} > 2.7$. As argued in section 3.6.5, negative

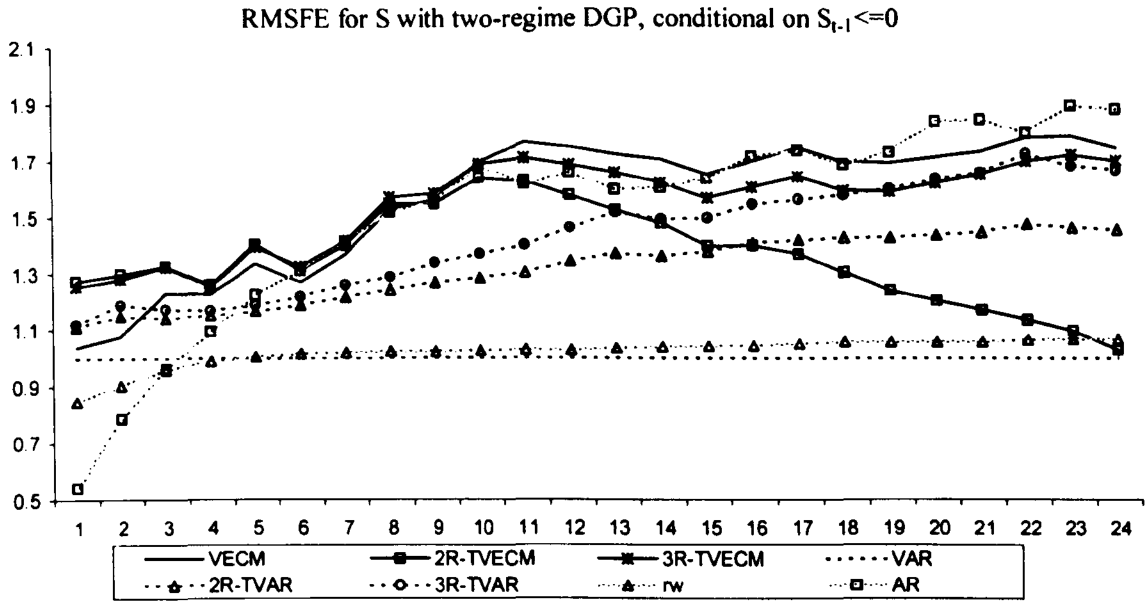

Figure 3.12: RMSFEs for S with data generated from the 2R-TVECM, conditional on $S_{t-1} \leq 0$, with VAR as benchmark

spreads do not occur in the forecasting period, thus we employ a simulation exercise. As before, data is simulated from the 2R-TVEqCM$_h$ and the 3R-TVEqCM. Based on each vector of simulated data (with size as the observed data), we estimate a VEqCM, a VAR, a 2R-TVEqCM$_h$, a 3R-TVEqCM, a two-regime TVAR and a three-regime TVAR. The VAR is employed as the benchmark as in previous analysis on the forecasting of cointegration systems (Clements and Hendry, 1995). Because only when the predictions for the cointegration vector are analysed is the effect of cointegration observed (Clements and Hendry, 1995; Christoffersen and Diebold, 1998), only then can the forecasts for the spread be evaluated.

Figures 3.12 and 3.13 present the RMSFE for $h = 1, ..., 24$, using a VAR(2) as benchmark. Values greater than 1 mean that the RMSFE of the model is smaller than the RMSFE of a VAR. Conditional on being in the regime given by negative spreads in the last period, all the models are better than the VAR at one step ahead. The performance of non-linear models improves comparatively with increasing horizons. The effect of cointegration is an improvement of 5% at one step-ahead and of 70% at 24 steps-ahead, and of threshold

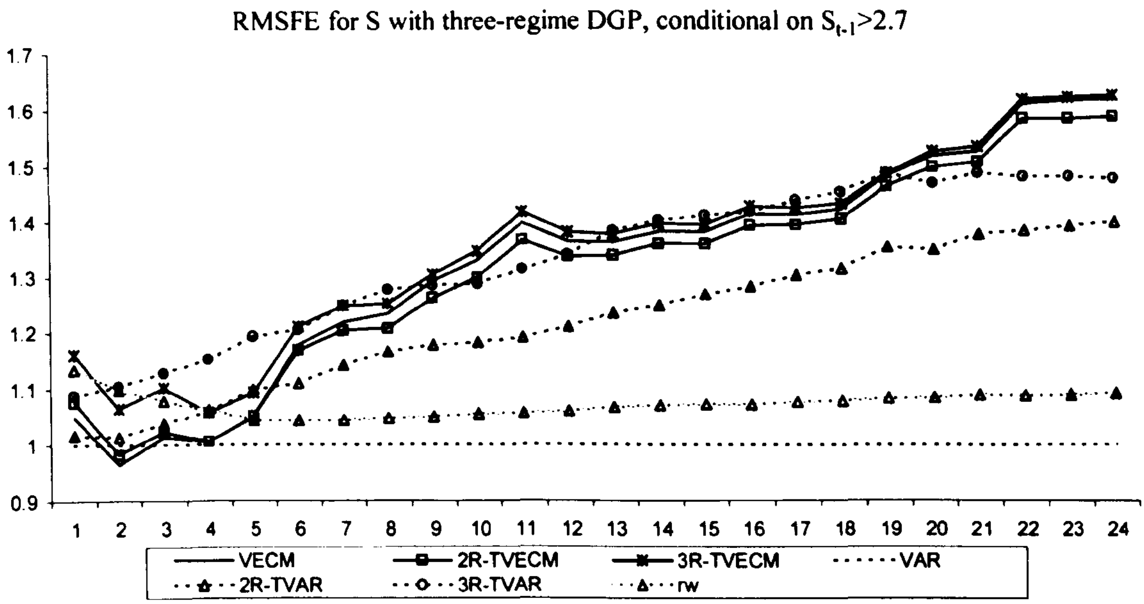

Figure 3.13: RMSFEs for S with data generated from the 3R-TVECM, conditional on $S_{t-1} > 2.7$, with VAR as benchmark

cointegration is 25% at one step-ahead and also 70% at 24 steps-ahead. Non-linearity, in the short-run, improves the forecast more than the inclusion of the equilibrium correction mechanism (VEqCM compared to the 2R-TVAR and the 3R-TVAR). Three-regime models have better long horizon performances than two-regime models[7]. At one-step ahead, the gain of threshold non-linearity is of 20% and that of the threshold equilibrium correction term is of more than 15 percentage points. Therefore, 3R-TVEqCM, for example, has a good performance because the thresholds induce non-linear effects in the short-run and the long-run dynamics of the cointegration system.

However, in Figure 3.13, the effects of non-linearity (even being conditional on a specific state of nature) and cointegration are smaller. At one step-ahead, the 3R-TVEqCM is only 15% better able than the VAR to predict the spread, probably because the data are simulated from the 3R-TVEqCM. The inclusion of cointegration only improves forecasts significantly after 5 steps, when models with equilibrium correction have similar MSFEs.

[7]After the evaluation of model with heteroscedastic and homoscedastic errors in the next part, it will be clear that this difference may be because the two-regime model forecasts are generated using heteroscedastic bootstrap while the three-regime forecasts are computed using homoscedastic bootstrap.

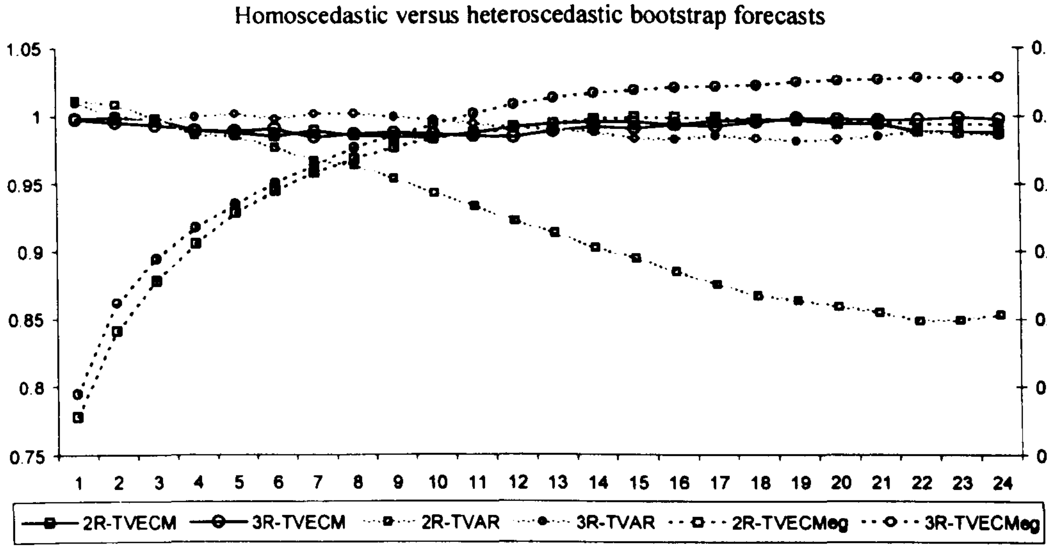

Figure 3.14: Comparing homoscedastic with heteroscedastic bootstrap forecasts (2R-TVEqCM$_{EG}$ and 3R-TVEqCM$_{EG}$ are in the secondary scale)

The difference between TVARs and TVEqCMs is small for three-regime models: 10 percent points at one-step-ahead and 15 points at 24 steps-ahead. This difference is slightly larger for two-regime models.

Therefore, non-linearity implies gains in forecasting performance not only because the adjustment to the equilibrium is non-linear but also because the coefficients of the short-run dynamics are calculated conditional on the regime defined by the threshold.

3.7.2 A New Forecast Competition

When the differences of testing and estimation procedures employed to specify the models of section 3.6 are disregarded, the differences among specifications depend on the number of regimes (2 or 3), the cointegration vector (w_{t-1} or S_{t-1}) and the assumption about the variance of residuals (regime-dependent or constant across regimes). However, it is not possible to observe how much of the forecasting performance of the model depends on the cointegration vector and how much on the variance being regime-dependent.

To obtain a more precise inference about the forecasting results, we estimated three

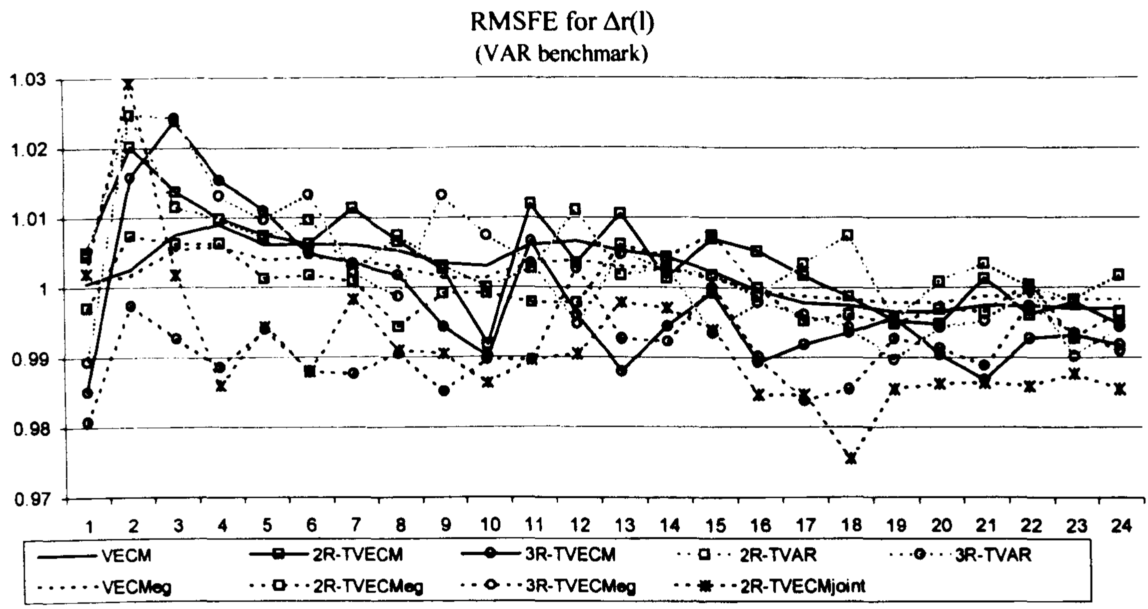

Figure 3.15: Comparison of RMSFEs for $\Delta r(l)$ with VAR as benchmark

different linear models: a VAR(2), a VEqCM with the spread as cointegrating vector ($\theta = 1$) and a VEqCM with θ estimated by OLS - the first step of the Engle and Granger procedure (VEqCM_{EG})[8]. The inclusion of non-linearity in the VAR implies the estimation of a two-regime (2R-TVAR) and a three-regime TVAR (3R-TVAR), with S_{t-1} as the transition variable. The inclusion of non-linearity in the VEqCM affects both long- and short-run model dynamics and a 2R-TVEqCM and a 3R-TVEqCM are estimated. Likewise TVEqCMs are estimated when $\theta \neq 1$: 2R-TVEqCM_{EG} and 3R-TVEqCM_{EG}. In addition, we also estimate a model with the cointegration and the threshold value estimated jointly (Hansen and Seo, 2000), called, as before, 2R-TVEqCM_{joint}. Compared with the evaluation of the previous section, the 2R-TVEqCM is equivalent to the 2R-TVEqCM_h, except for assuming variance constant across regimes for the computation of the forecasts[9]. The 2R-TVEqCM_{joint} and the 3R-TVEqCM have the same specification as before.

[8]The cointegration vector is $w_t = r(l) - \theta r(s)$. Models using $w_t = r(l) - \theta r(s) - \mu$ have significantly worse performance in short-horizons forecasts and have similar performance at long-horizons.

[9]The assumption of constant variance over regimes arises from the results of the comparison between forecasts generated asumming homoscedasticity and heteroscedasticity, presented in Figure 3.14.

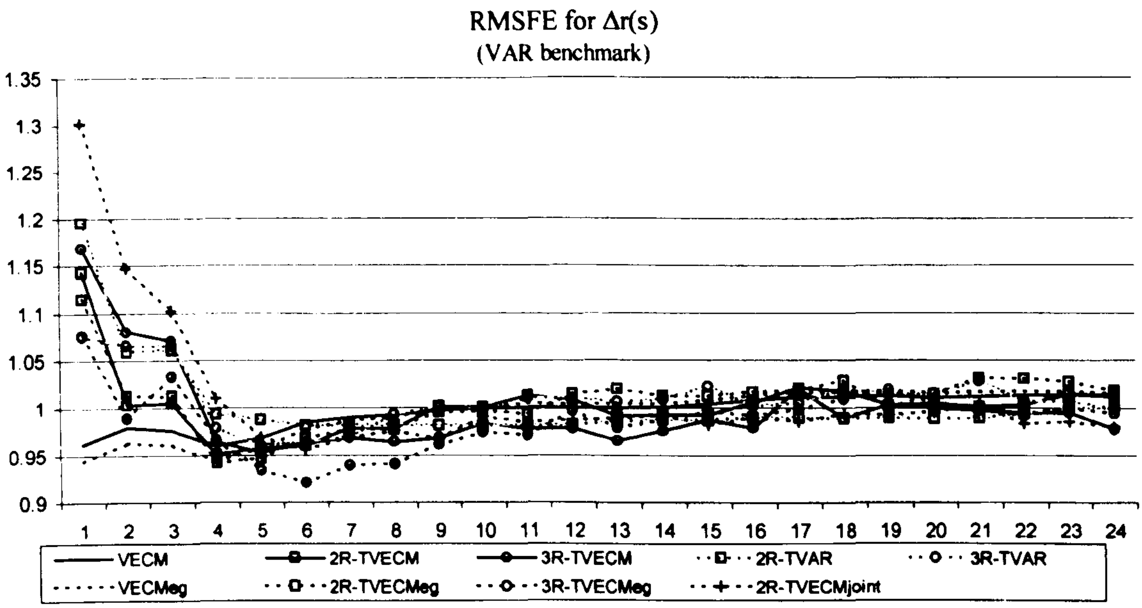

Figure 3.16: Comparison of RMSFEs for $\Delta r(s)$ with VAR as benchmark

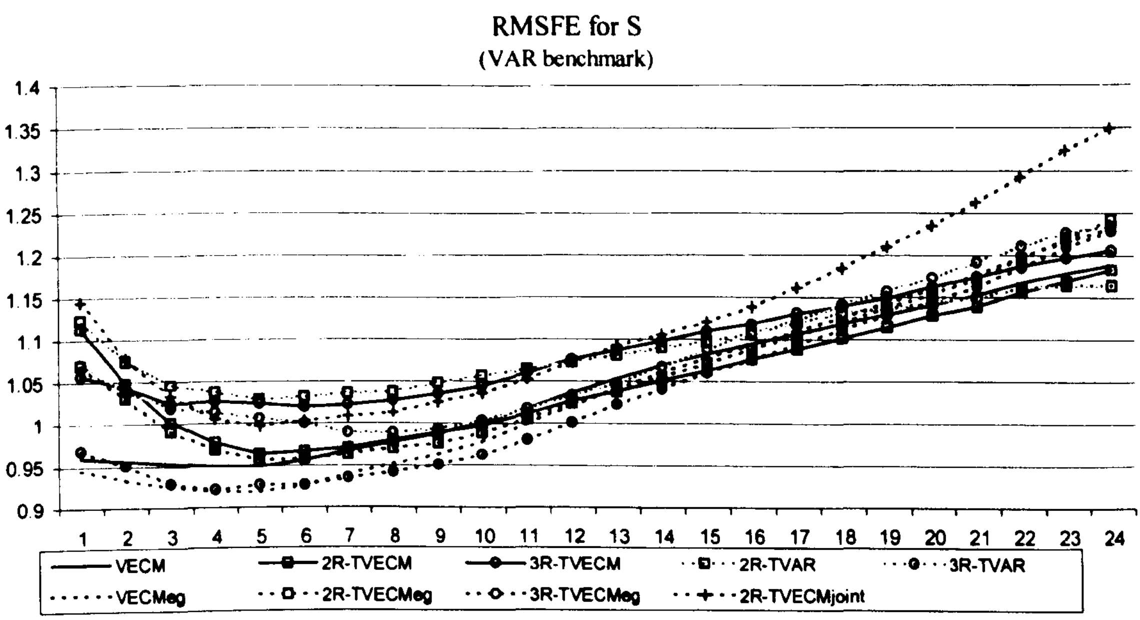

Figure 3.17: Comparison of RMSFEs for S with VAR as benchmark

The models are estimated for the same sample period used in the last section (1960-1989) and the same out-of-sample period is considered (1990:1-1998:4). For each observation in the out-of-sample period, the models are re-estimated (including the thresholds) and forecasts from 1 to 24 steps-ahead are generated. The threshold of two regime models are estimated to maximise $\ln(\det(\hat{\Omega}))$ using the form of equation 3.22, except for the 2R-TVEqCM$_{joint}$ that follows the procedure outlined in section 3.3 based on Hansen and Seo (2000). The thresholds of three-regime models are also calculated to minimise $\ln(\det(\hat{\Omega}))$ using the three regimes equivalent of 3.22, but the one-step-at-a-time algorithm is employed to reduce the computational burden of the grid search. The forecasts of the threshold models are generated using the bootstrap procedure with 1000 replications.

Two bootstrap procedures are employed: a homoscedastic and a heteroscedastic one. The first type of bootstrap draws values from only one vector of residuals $\hat{\epsilon}$ and the second one draws from $\hat{\epsilon}^{(i)}$ ($i = 1, 2$ for two regime models and $i = 1, 2, 3$ for three regime models), where the vector $\hat{\epsilon}^{(i)}$ that is actually employed to calculate the forecast at $T + j$ depends on the value of the transition variable in the last period (w_{T+j-1}) and the threshold values. Figure 3.14 shows the effect of the type of forecasting procedure on the RMSFE. Values smaller than one mean that the RMSFE (calculated over 100 data points for each step-ahead) of threshold models with forecasts computed using the homoscedastic bootstrap are smaller than the heteroscedastic bootstrap. Note that the values for the 2R-TVEqCM$_{EG}$ and the 3R-TVEqCM$_{EG}$ are plotted using the secondary scale. For most of the threshold models, differences between RMSFEs from forecasting procedures are very small, and the homoscedastic version is much better for two models (2R-TVEqCM$_{EG}$ and 3R-TVEqCM$_{EG}$). Therefore, the variances across regimes do not seem to be significantly different in importance for point forecasting, although this difference should be larger when interval or density forecasts are evaluated. In the analysis that follows, the forecasting of the threshold models are calculated using the homoscedastic bootstrap.

Figures 3.15 and 3.16 present the RMSFEs of a VAR model to predict $\Delta r(l)$ and $\Delta r(s)$ compared with the other 9 models employed in the forecast competition. As in Figure 3.4, Figure 3.15 shows that the inclusion of non-linearity and/or an equilibrium correction term does not improve the forecasting of $\Delta r(l)$[10].The largest improvement is generated by the 2R-TVEqCM$_{joint}$, the 2R-TVAR, the 3R-TVAR and 2R-TVEqCM at two steps-ahead. However, they fare only 2% better than the VAR. On other hand, the prediction of $\Delta r(s)$ is improved when cointegration and non-linearity are included. At long horizons, the models are equivalent because it is not possible to observe cointegration effects in long-horizons when the RMSFEs of the variables in first difference are analysed (Clements and Hendry, 1995; Christoffersen and Diebold, 1998). At one-step-ahead the 2R-TVEqCM$_{joint}$ fares 20% better than the VAR, which is followed by the 2R-TVAR, the 3R-TVECM, the 2R-TVECM, the 2R-TVECM$_{EG}$ and the 3R-TVAR, all with RMSFEs at least 7% smaller than the VAR. In general, it seems that two-regime models are better than three-regimes. The RMSFE is larger when non-linearity in the adjustment to the equilibrium is included, which can be verified by comparing the results for the 2R-TVAR and the 2R-TVEqCM with the 2R-TVECM$_{EG}$. However, given the RMSFE for the 2R-TVEqCM$_{joint}$, this may be result of a poorly estimated cointegration vector. On other hand, the 3R-TVEqCM is better than the 3R-TVAR in 10 percentage points at one-step-ahead and it is equivalent to the 3R-TVAR for larger h.

Therefore, this new forecasting competition confirms the results of the previous one that the 2R-TVEqCM$_{joint}$ and the 3R-TVEqCM are good forecasters of $\Delta r(s)$ at short-horizons, although the performance of the 2R-TVEqCM (similar to the 2R-TVEqCM$_h$) shows that what matters is the non-linearity of the short-run coefficients (2R-TVAR) and not of the adjustment to the long-run equilibrium. The results also confirm that the restriction

[10]Note that in Figure 3.15, we are employing the RMSFE as measure of forecast accuracy, then the compared figures are based in different scales that do not affect the rank order, but affects the value of the ratio.

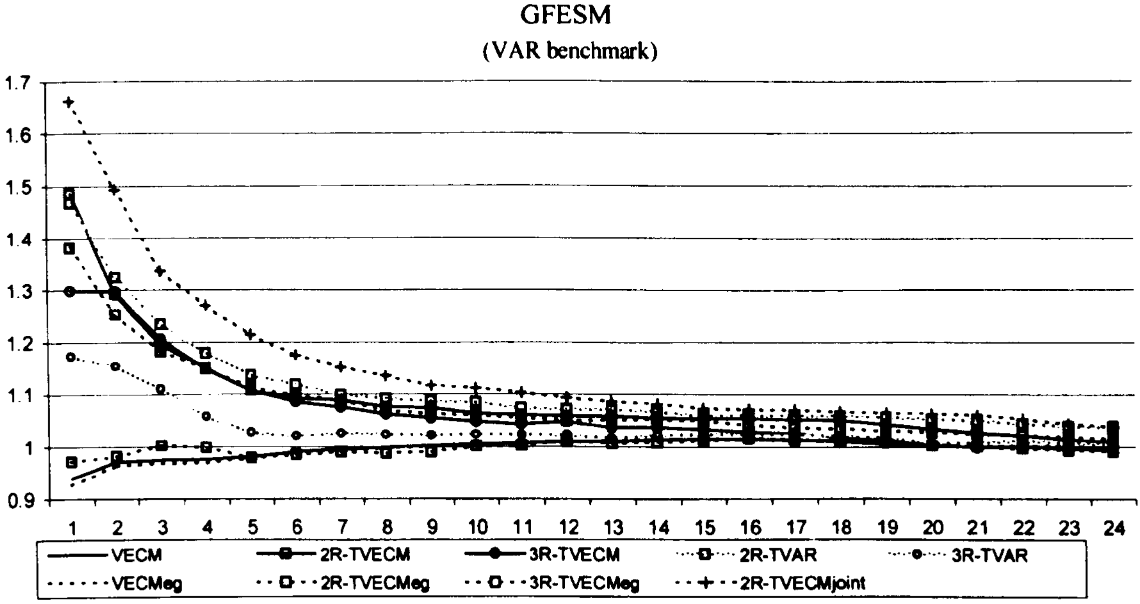

Figure 3.18: Comparison of GFESMs with VAR as benchmark

that $\theta = 1$ does not affect the forecast of $\Delta r(s)$ (the 2R-TVEqCM$_h$ compared with the MTVEqCM) and that non-linearities do not improve the forecasts of $\Delta r(l)$.

The results of the RMSFE in predicting the spread are presented in Figure 3.17. At short-horizons, the gains of cointegration (VEqCM) are negative, but threshold cointegration reduces RMSFE in 15% at one step-ahead. At long horizons, the cointegration systems fare at least 15% better than the VAR while this value in the simulation exercise, conditional on the regime, goes as far as 70%. This effect of cointegration - comparing 2R-TVEqCM$_{joint}$ with 2R-TVAR at $h = 24$ - when the forecasts for the cointegrating relation are evaluated has been reported previously (Clements and Hendry, 1995; Christoffersen and Diebold, 1998).

The analysis of the rank of RMSFE at 24 steps-ahead sheds light on the results of the MTVEqCM and the 2R-TVEqCM$_{joint}$ at long horizons. The performance in predicting the spread at long horizons depends on the estimated cointegration vector. An isolated winner is the 2R-TVEqCM$_{joint}$ that estimates jointly the threshold and the cointegration vector and is able to better characterised the data when long-horizon forecasts are need

(2 years ahead). The rank follows with models with the cointegration vector estimated using the Engle and Granger (1987) procedure and models with spread as cointegration vector. The difference of RMSFEs between VEqCM and VEqCM$_{EG}$ is not large. Therefore, the imposition of $w_t = r(l) - r(s)$ as the cointegration vector does not imply reduction in forecasting performance when the restriction is relaxed but θ is estimated by OLS in a first step. This restriction only implies losses when compared with a procedure that jointly estimates θ and r, implying that the procedure to model TVEqCMs proposed by Hansen and Seo (2000), is better than the procedure suggested by Balke and Fomby (1997) (at least in this specific data set).

Another robustness check is the calculation of GFESM (eq. 3.19) for each one of the models for each step ahead. The determinant of the GFESM is calculated for the forecast errors to predict $[\Delta r(s), S]'$ and the measure is invariant to include $\Delta r(l)$ instead of $\Delta r(s)$. The ratio of GFESM of the VAR over each model GFESM at h steps-ahead is plotted in Figure 3.18. The inclusion of cointegration (VEqCM and VEqCM$_{EG}$) does not decrease the GFESM at short-horizons; and because of the weight that the GFESM gives to forecast errors from small h, the GFESMs of these models have only a small increase at longer horizons. The 2R-TVEqCM$_{joint}$ has the smallest GFESM for all horizons, followed by the models 2R-TVEqCM, 2R-TVAR, 2R-TVEqCM$_{EG}$ and 3R-TVEqCM. Consequently, Figure 3.18 supports the results of the previous section, confirming the good forecast performance of the 2R-TVEqCM$_{joint}$, the 2R-TVEqCM$_h$ and the 3R-TVEqCM. An important information obtained from Figure 3.18 is that the 2R-TVEqCM$_h$ is a good forecaster because it has short-run coefficients changing and not because of the changing long-run adjustment over regimes. In contrast, non-linearities in both adjustment to the equilibrium and short-run dynamics are responsible for the performance of the 3R-TVEqCM (which fares better than 3R-TVAR).

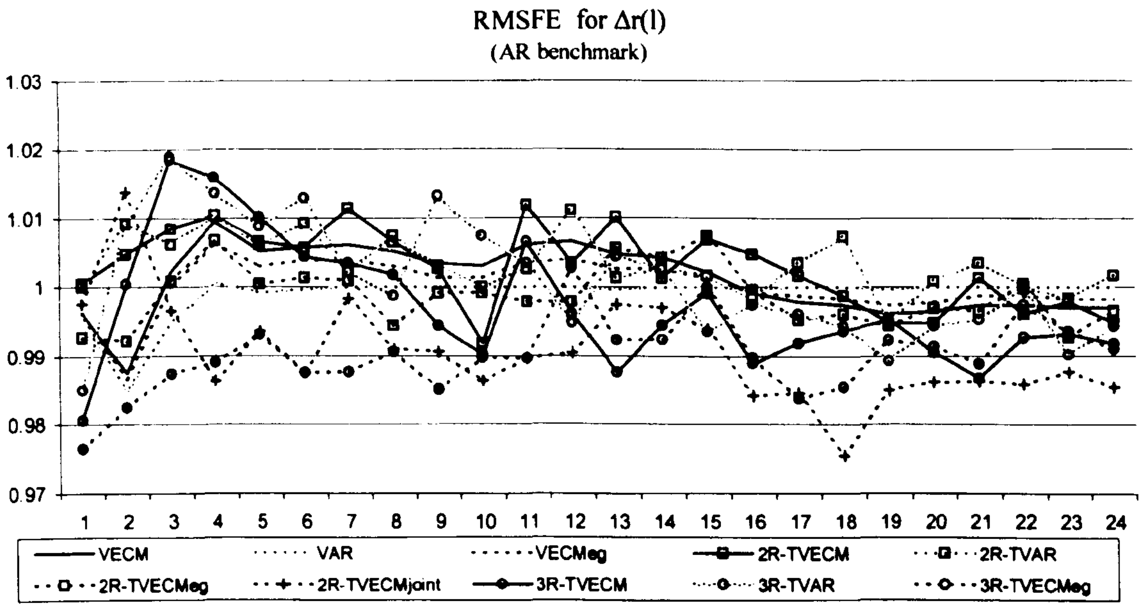

Figure 3.19: Comparison of RMSFEs for $\Delta r(l)$ with AR as benchmark

3.7.3 Comparing with Linear Autoregressive Models

Some practitioners may ask whether it is worth estimating VEqCMs to predict interest rates instead of using a simple AR model. Given the autoregressive order of the VAR, we estimate AR(2) models for $\Delta r(s)$ and $\Delta r(l)$. In addition, an AR(1) is estimated for the spread. We use these autoregressions to generate 24 step-ahead forecasts, using the same forecast procedure outlined in the last section. The RMSFEs of these AR models are employed as benchmark in Figures 3.19, 3.20 and 3.21. For the spread, we also estimate a random walk, because in the middle regime, the spread may not be stationary.

We do not expect to find any gains by allowing for cointegration in predicting $\Delta r(l)$ and $\Delta r(s)$ at long horizons compared to AR models, given the results of Christoffersen and Diebold (1998). AR models are the best forecasters for both rates at short horizons (Figures 3.19 and 3.20). For the spread, non-linearity gains are observed at short and long horizons (Figure 3.21). The gains are of 5% for most of the horizons, using the TVEqCM$_{joint}$, the 2R-TVAR and the 3R-TVECM. However, a random walk is equivalent to these models when

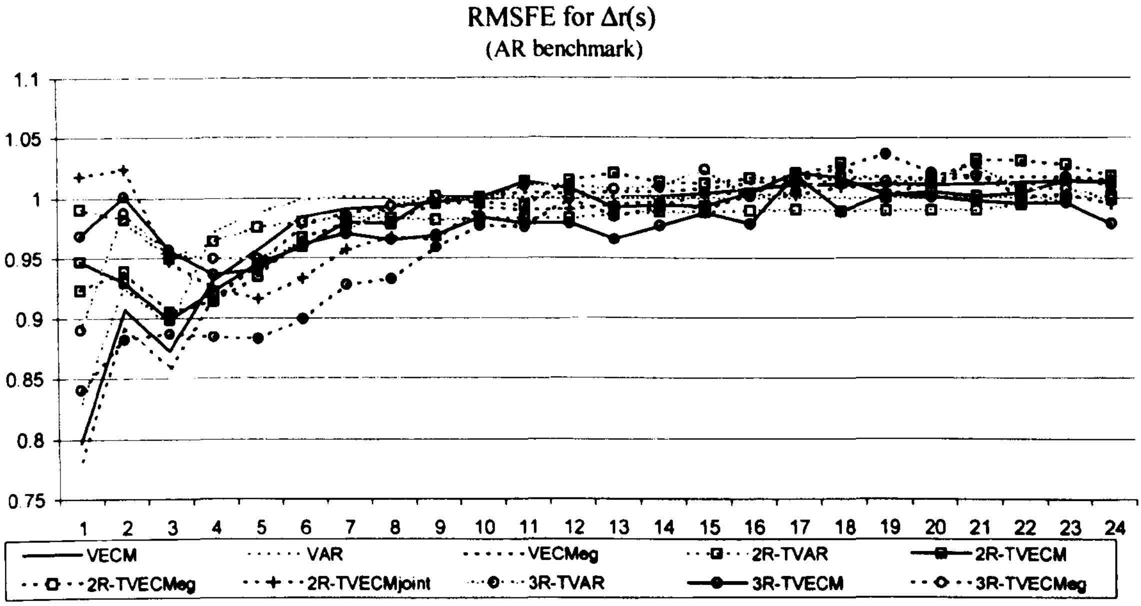

Figure 3.20: Comparison of RMSFEs for $\Delta r(s)$ with AR as benchmark

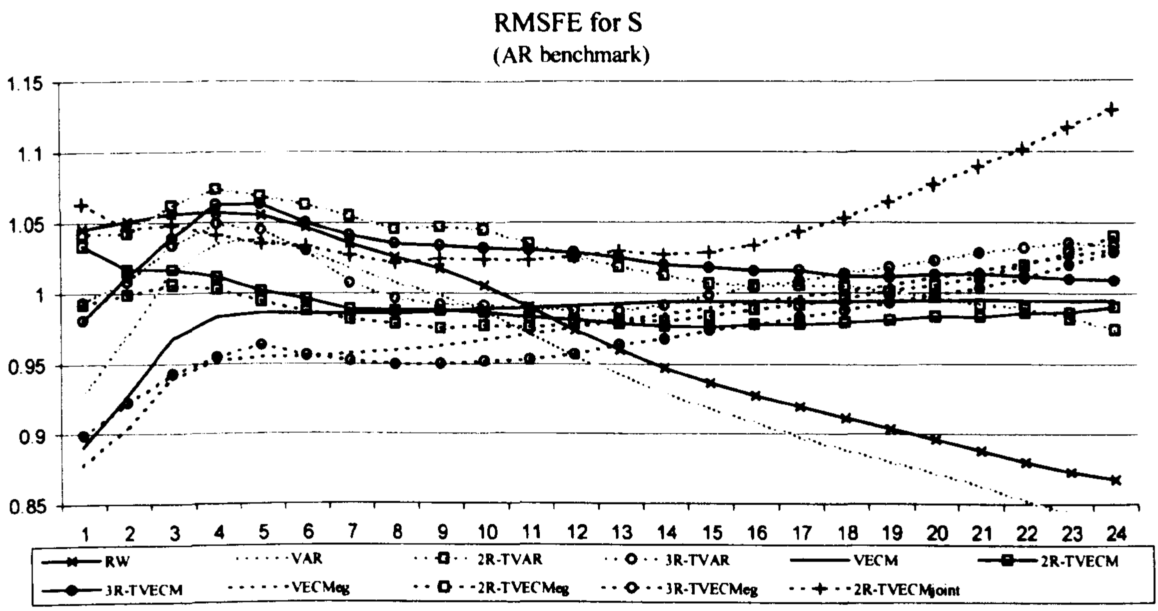

Figure 3.21: Comparison of RMSFEs for S with AR as benchmark

$h = 1,...,9$. Therefore, a simple random walk is a good forecaster of the spread at short horizons. At long horizons, the gains of more complex models are stronger. Observe that the VAR has a bad performance at long-horizons, which can cause bias in the results of Figure 3.17.

Using the simulation exercise described in section 3.7.1, the forecast performance for a random walk and an AR(1), which are estimated with simulated data, are presented in Figures 3.12 and 3.13. Although at long horizons, the RMSFE of the AR(1) is similar to the other models, at short horizons, the forecasts from the AR(1) are at least 50% worse and from the random walk are 15% worse (Figure 3.12). However, the random walk seems to be a good forecaster at short-horizons, when the three-regime is the DGP, conditional on $S_{t-1} > 2.7$ (Figure 3.13).

At short horizons, the models that are equivalent to the random walk are the 2R-TVAR, the 3R-TVEqCM and the 2R-TVEqCM$_{joint}$ (Figure 3.21). One of the characteristics of these models is that the regime with largest number of observations (with higher probability of occurrence) is the one in which the coefficients of the adjustment to the long-run equilibrium are zero in both equations. This can be interpreted as an absence of cointegration between $r(l)$ and $r(s)$, which may imply that the spread ($r(l)$-$r(s)$) follows a random walk given that the levels of both interest rates are I(1). This argument can explain the results of Figure 3.21 compared with the results of the simulation (Figures 3.12 and 3.11) that calculates the RMSFE conditional on the regimes in which the adjustment coefficient to the equilibrium is statistically significant.

In conclusion, the results of this section support, in general, the forecasting evaluation of the previous section. In addition, the results of this section also offer some explanation of those obtained in the previous section, in particular:

(A) The 2R-TVEqCM$_{joint}$ has good performance because it has the best procedure to estimate threshold cointegration.

(B) Regime-dependent variance does not improve the point forecasts of the estimated TVEqCM (3R-TVEqCM$_i$, 3R-TVEqCM$_u$ and 3R-TVEqCM$_f$ compared with 3R-TVEqCM).

(C) In some cases non-linearities in the short-run dynamics are more important than in the adjustment to the equilibrium.

(D) The estimation of the cointegration vector ($\theta \neq 1$) improves the spread predictions when the VAR is the benchmark.

Furthermore, a new result is also obtained: non-linear models do not forecast significantly better than simple autoregressive models, and a random walk is a good forecaster of the spread at short-horizons. However, non-linearity and cointegration generate forecast gains at long horizons when the spread is forecasted. When only linear models are considered, allowing for cointegration also results in gains in long horizons (Clements and Hendry, 1995). The inclusion of non-linearity improves the performance compared with models that only allow for cointegration.

3.8 Conclusions

This chapter applies testing and specifications techniques commonly employed to threshold autoregressions and regressions to systems with the objective of evaluating the effect of non-linearities in cointegrated systems. In doing so, we demonstrate how methods to test threshold autoregressive models can be applied to test threshold equilibrium correction models. Threshold vector equilibrium correction models are tested and estimated to model the US term structure of interest rates, showing that the usefulness of the spread to forecast short- and long-term interest rates depend on the regime.

Non-linearity improves short horizons forecasts of VEqCM, even though one of the regimes do not occur in the forecasting period. In addition, when compared with VAR predictions, non-linearity (jointly with cointegration) also improves long horizons forecasts of the spread (cointegrating relation). In contrast, TVEqCMs, in general, do not forecast better than simple AR of $\Delta r(l)$ and $\Delta r(s)$, although some specifications have (R)MSFEs for $\Delta r(l)$ and $\Delta r(s)$ similar to the AR at short horizons. When predictions for the spread are assessed, however, the gains from modelling non-linearity and cointegration are observed compared to an AR at long horizons. The results of the forecasting evaluation also contribute to the literature, showing that the presence of cointegration improves forecasts of the cointegrating relationship at long horizons not only comparing VECMs with VARs, but also comparing TVECMs with TVARs.

The forecasting evaluation can be employed to compare the efficacy of the methods proposed in the literature to model threshold equilibrium correction models. The joint estimation of the threshold and the cointegration vector (Hansen and Seo, 2000) generates a TVEqCM with better forecast accuracy than estimating the TVEqCM either using the thresholds computed from a threshold autoregressive model for the cointegrating relationship (Balke and Fomby, 1997; Enders and Siklos, 2001) or by grid search assuming that the cointegration vector is known.

Chapter 4

Non-linearities and Structural Breaks in Predicting US Recessions using the Spread

4.1 Introduction

The spread between long and short-term interest rates, which represents the term structure of interest rate, is a good predictor of economic activity. The forecasting performance depends on the measure of economic activity (Bonser-Neal and Morley, 1997), however, specifically, the spread is a good predictor of the growth rate of US GDP (Estrella and Hardouvelis, 1991; Hamilton and Kim, 2000, and the surveys of Berk, 1998 and Stock and Watson, 2001). The interest rates employed to calculate the spread also affect its predictive ability. The spread between the 3-month T-bill and the 10-year T-bond (Hamilton and Kim, 2000) and the spread between the 1-year T-bond and the 10-year T-bond (Lahiri and Wang, 1996) are good predictors.

The information contained in the spread variable goes beyond monetary policy, because even if other indicators of monetary policy are included as predictors in a regression to explain output growth, the spread keeps its predictive power (see Anderson and Vahid, 2000, who use non-linear models). This is also true for the inclusion of lagged oil price changes and lagged output (Hamilton and Kim, 2000). The spread variable reflects future expected short rates and changes in the risk premium. The expectations hypothesis of the term structure and the temporary influence of monetary policy suggest a positive correlation between the spread and future economic growth. This is so because a tight-temporary monetary policy produces negative spreads and reduces output growth, given that short-term interest rates rise relatively more than the long-term ones, following the expectation hypothesis, whereas positive spreads are associated with "easy" monetary policy and high economic growth. This positive correlation can also be explained by market expectations of future growth that may be reflected in the long-term rate. Finally, cyclical factors in the risk premium that change the volatility of interest rates may be responsible for the relationship. Hamilton and Kim (2000), however, could not find empirical evidence to support the last hypothesis. The spread also contains information on future inflation as surveyed by Berk (1998) and Stock and Watson (2001).

The spread is also a good predictor of the probability of recession in the US. Estrella and Mishkin (1998) use probit models to show that the spread can predict US recessions four quarters in advance, and Bernard and Gerlach (1998) obtain similar results for Germany and Canada. A drawback of the probit approach is that it is necessary to have data on the occurrence of recession before the estimation of the model, and this is only obtained with some time lag. This is taken into account by Lahiri and Wang (1996), who propose a Markov-Switching model as a filter to get information on the probability of recession from the spread. This model gives good performance in predicting recessions in the US and also in Germany (Ivanova, Lahiri and Seitz, 2000).

The ability of the spread to predict recessions means that policymakers can use the spread as a variable to help to frame monetary policy. However, this may change the predictive power of the spread. Based on the Lucas (1976) critique, Estrella and Hardouvelis (1991, p. 575) argue that "the estimated historical correlations are not necessarily policy invariant", so the invariance of the predictive power of the spread is an important question. Haubrich and Dombrosky (1996), Dotsey (1998) and Stock and Watson (2001) report that the predictive power of the spread between long- and short-term interest rates has decreased after 1985. Stock and Watson show that a model with the spread as predictor of output growth does not have significantly smaller MSFE compared with an autoregressive model for the period 1985-1999. The failure of the Stock and Watson (1989) indicator index to predict the 1990-91 recession has been attributed to the fact that the index gave too large a weight to the spread (Dotsey, 1998). However, employing Markov-switching models to obtain the probability of recession, Lahiri and Wang (1996) show that the spread predicted the last recession. Likewise, Dueker (1997) and Estrella and Mishkin (1998), using probit models, demonstrate that the spread is still better than other leading indicators at predicting recessions in the US. The test results of Estrella, Rodrigues and Schich (2000) help to understand these contradictory results. After testing models to predict output growth and models to predict recession using spread as the leading indicator, the authors found that unstable parameters are a characteristic of models used to predict output growth but not recessions.

Lahiri and Wang (1996, p. 306) argue that the "optimal forecasting horizon for predicting troughs [using the spread] is significantly shorter than the horizons for predicting peaks". Galbraith and Tkacz (2000) employing non-linearity tests for threshold effects, also found significant evidence of an asymmetric relationship between the spread and output growth. The predictive power of the spread is almost zero when the spread is greater than 2% at annual rates. This is also the result given by the simple non-linear model proposed by Dotsey (1998). Anderson and Vahid (2000) built non-linear autoregressive leading indicators employing the spread based on smooth transition regressions. They found that non-linearities improve the accuracy of predicting the probability of recession of the US economy.

Even if the spread is still useful to predict recessions, the literature indicates that its predictive power has been decreasing. If the spread depends on monetary policy and on the market perception of this policy, which is mainly reflected in the long-term rate, changes in policy may affect the reasons why the spread is high or low. Moreover, as Berk (1998) argues, the same value of the spread can be caused by different types of policies. Another problem is that when the arbitrage between the long and short maturity markets is not instantaneous, the relationship between the spread and output growth is not clear (Berk, 1998). For example, as described by Hardouvelis (1994), the expectations hypothesis of the term structure does not hold because the long-term interest rates over-react to the tightening of monetary policy, implying high spread values. High spreads mean that the predictive power of the spread is lower as indicated by the non-linear modelling of Dotsey (1998) and Galbraith and Tkacz (2000).

A possible break in the relationship between the spread and output growth may be connected with the lower volatility of output growth after 1984 (Kim and Nelson, 1999b; McConnell and Perez-Quiros, 2000). Lower variability in output is the result of decreasing variability in the production of durables goods, probably because of an improvement in inventory management. Kim and Nelson and McConnell and Perez-Quiros suggest, respectively, a Bayesian and a Classical Markov-Switching model to represent US real GDP growth. The models have two different unobserved processes driving changes in the mean and in the variance. Kim and Nelson show that the difference between the conditional means of expansion and contraction regimes is smaller after 1984 than before. In addition, McConnell and Perez-Quiros argue that the Markov-Switching model only predicts the 1990-91 recession if a structural break in the variance in 1984 is taken into account.

In contrast, the variability of the long-term interest rate has increased (Watson, 1999) because of the persistence of monetary policy which smooths short-term interest rates, reducing their variability. Structural break tests indicate parameter instability in autoregressive models of long-term interest rates and of short-term interest rates (Sensier and Van Dijk, 2001). However, because the variability of long-term interest rate has increased while the short-term interest rate volatility decreased, the volatility of the spread may not have changed.

This chapter tests whether the relationship between output growth and the spread has changed in section 4.5. It also investigates the possibility of non-linearities in the relationship (section 4.4), given the asymmetries in the business cycle described in Chapter 2 and the non-linear behaviour of the spread analysed in Chapter 3. Furthermore, it tests the possibility of changing non-linearity, i.e., parameter instability in non-linear models in section 4.6. The models indicated by the tests are evaluated by their abilities to predict the probabilities of recession in section 4.8. That is, we evaluate how non-linearities and time-varying coefficients can improve the prediction of event (recessions) probabilities, given the evaluation method described in section 4.7, using bivariate models with the spread and output growth as endogenous variables.

Before proceeding to the examination of test procedures, we analyse the characteristics of the data. The next section presents descriptive statistics of the spread and output growth, a non-parametric estimation of this relationship and a simple way of extracting information from the spread about the probability of recession. A simple linear model, used as benchmark, is then presented in section 4.3.

4.2 The Spread and Economic Activity

The data employed in this work were taken from the Fred database on February 2001 (http:// www.stls. frb.org /fred/index.html). Output growth is the first-difference of the log of real GDP at chained 1996 prices (*100). The spread is the difference between the

10-year T-bond and 3-month T-bill. Averaging is employed to transform monthly interest rate data to quarterly, and the data period is from 1953:2 to 2000:4. The relationship between output growth and spread is mainly positive as observed in Figure 4.1, except for the middle of the 60's that has decreasing spreads in a period of high growth. Low values of the spread precede recessions, except for 1966, therefore, the spread is signaling a recession in 2001.

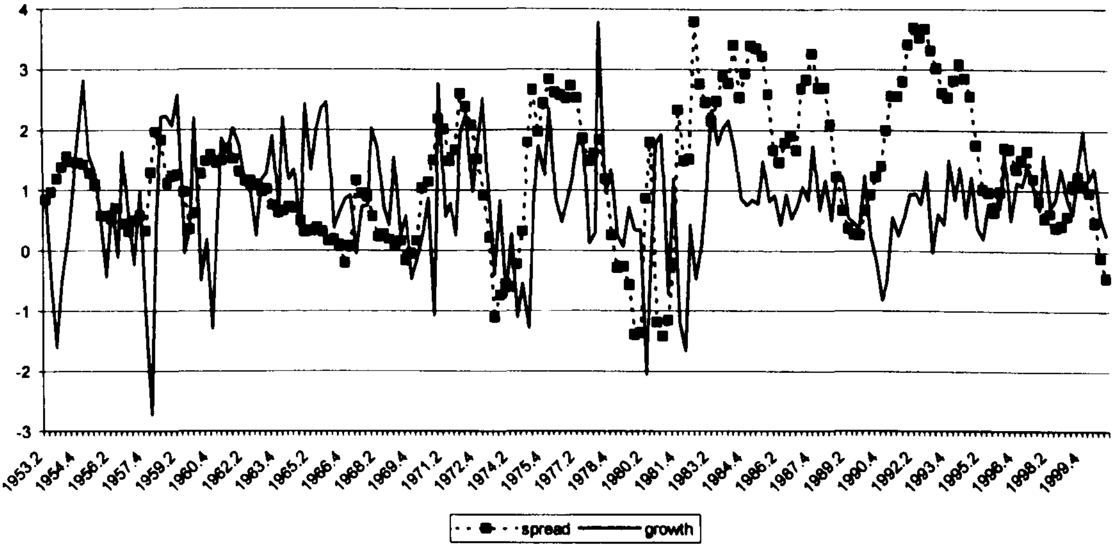

Figure 4.1: The spread between 10-year T-bond and 3-month T-bill and the output growth

4.2.1 Descriptive Statistics and Conditional Means

Table 4.1 presents descriptive statistics of these data - means, variances and lagged cross-correlations - for different samples to observe possible structural changes. The sample split in 1984 is also employed by Stock and Watson (2001) and follows the reported structural break in the output growth variance (Kim and Nelson, 1999b; McConnell and Perez-Quiros, 2000). The variance of output growth for the period 1984:1-2000:4 is half that of the previous period, while the mean of the spread in the latter period is twice that in 1953:2-1969:4. The changes in the characteristics of these two variables may change the predictive ability of the spread with respect to output growth, which is indicated by the correlations. The highest

Table 4.1: Descriptive statistics of output growth and the spread

	μ_y	μ_S	σ_y	σ_S
1953:2-1969:4	0.883	0.816	1.092	0.533
1970:1-1983:4	0.700	1.203	1.188	1.380
1953:2-1983:4	0.800	**0.992**	**1.135**	1.024
1984:1-2000:4	0.845	**1.840**	**0.535**	1.093

Correlations (y_t, S_{t-j})

	(S_{t-1})	(S_{t-2})	(S_{t-3})	(S_{t-4})	(S_{t-5})	(S_{t-6})	(S_{t-7})	(S_{t-8})
1953:2-1969:4	0.228	0.294	**0.350**	0.244	0.024	-0.071	-0.054	-0.133
1970:1-1983:4	0.452	**0.581**	0.507	0.467	0.373	0.191	0.111	-0.004
1953:2-1983:4	0.338	**0.442**	0.414	0.365	0.251	0.109	0.062	-0.034
1984:1-2000:4	0.113	0.128	0.078	0.081	0.112	**0.149**	-0.001	-0.154

μ_i is the mean of variable i for the specified period, given that i = y (output growth), S (spread). σ_i is the standard deviation of variable i. The lower panel exhibits correlations between y_t and S_{t-j} for j = 1,...,8.

correlation between the spread and output growth is 0.44 at lag 2 for the period 1953:2 -1983:4 and 0.15 for the period 1984:1-2000:4 at lag 6. Therefore, the predictive power and the optimal predictive horizons seem to have changed.

Non-parametric estimation of output growth conditional on the spread is able to show possible non-linearities in the spread-economic activity relationship. These conditional means $(E[y_t|S_{t-p}])$[1] are estimated for the spread from lag 1 to 6 for the sample from 1953:2 to 1983:4 and from lag 1 to 8 for the sample 1984:1 to 2000:4 (Figure 4.2), given that the spread seems to predict at longer horizons in the second sample. The relationship between the spread and output growth is strongly positive, linear and the spread is able to predict negative growth for lags 2, 3, and 4 in the first sample. Some non-linearities in this relationship are shown when the lag is 5 (the spread does not predict increasing growth when S_{t-5} is in the interval $[1, 2]$). In the second sample, the predictive power of the spread seems much smaller: the spread does not even predict negative growth as the estimated curve is flat. Positive relationships occur when the spread is smaller than 1/2 for lags higher than 4.

[1]Smoothing splines are employed to estimate conditional means (see, e.g., Simonoff (1996, chap. 5.6)), using SPLUS2000. They are local cubic functions with the smoothing parameter determined by the number of degrees of freedom of the regression, which is set to 6.

An interesting case is the plot for $E[y_t|S_{t-7}]$ of the second sample. The correlation is near zero, but this masks some important non-linearity in the relationship. Increasing spreads predict increasing growth when $S_{t-7} < 1/2$ and when $S_{t-7} > 2\ 1/2$, however increasing spreads predict decreasing growth when S_{t-7} is in the interval $[1/2, 2\ 1/2]$ (Figure 4.2). Note that these are predictions made almost two years ahead. This may explain why the spread's predictive ability for economic activity after 1984 has been questioned (Haubrich and Dombrosky, 1996; Dotsey, 1998), but the predictive ability for recessions seems strong (Estrella and Mishkin, 1998). However, these are only preliminary results, without relevant statistical testing. Testing for non-linearity and structural breaks are the objectives of this chapter.

4.2.2 The Simple Rule and the NBER Peaks and Troughs

Given the findings of Estrella and Mishkin (1998) for the US and Bernard and Gerlach (1998) for Germany and Canada, we estimated a probit model to predict recessions using the spread as the explanatory variable. With S_{t-3} as the explanatory variable for the whole sample, we found that the coefficient was almost -1. Based on this, we suggest a simple rule to calculate probability of recession in the US using the spread:

$$Pr[recession_t] = 1 - \Phi(S_{t-d}) \tag{4.1}$$

where Φ is the standard normal cumulative distribution. This rule does not need any information about previous turning points, only the actual value of the spread and the optimal predictive horizon d. The analysis of the descriptive statistics suggest the use of $d = 3$. This rule is presented in Figure 4.3 against the NBER turning points. We also use a mixed rule that uses $d = 6$ for the observations after 1983. The simple rule gives a good performance of predicting recessions if we suppose that the recession happens when $Pr[recession] > 0.35$[2].

[2]This alternative truncation implies that the simple rule needs scaling to predict recessions when the recession event is defined as probability of recession greater than 0.50. The probit estimation does not suggest

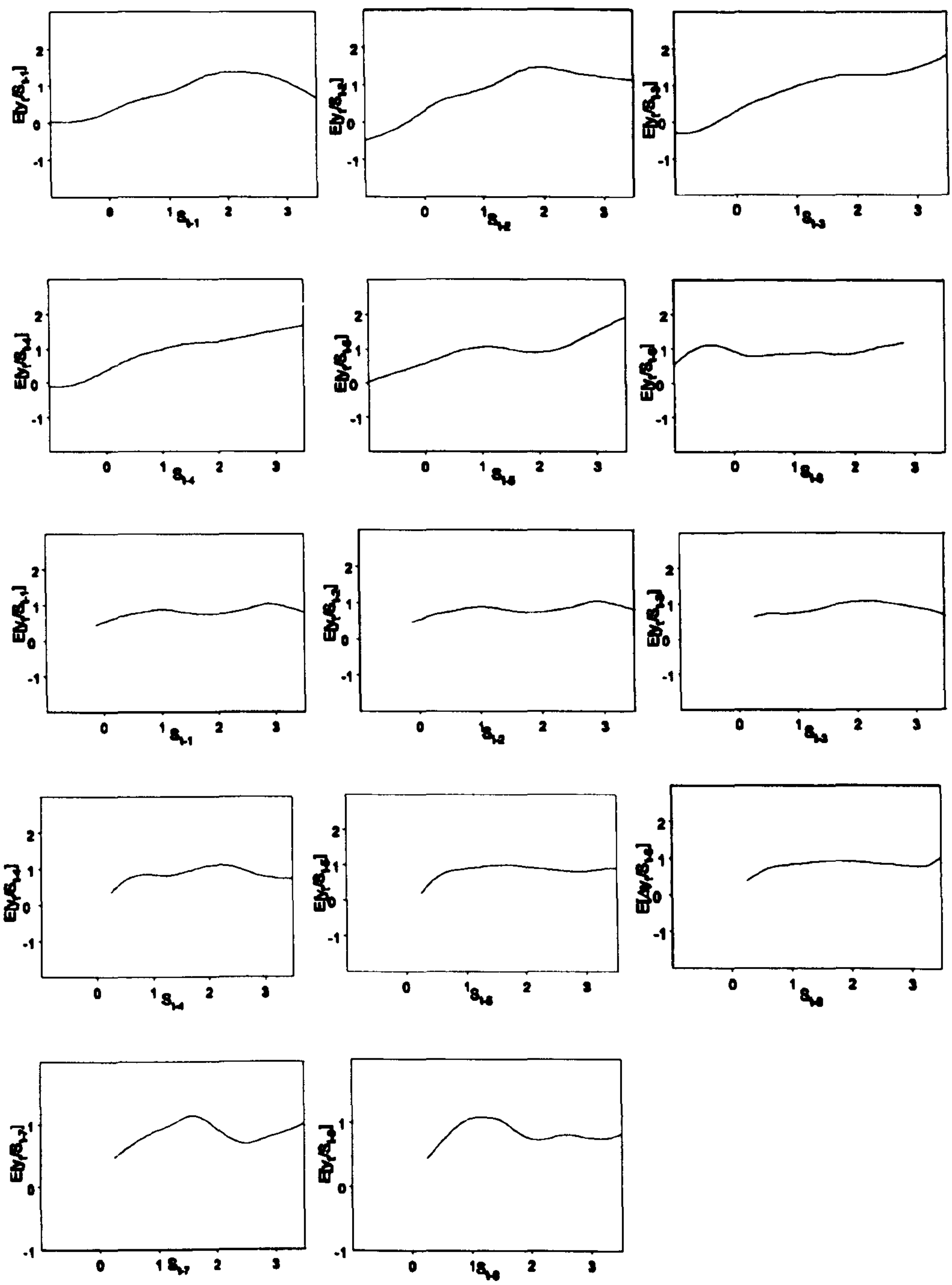

Figure 4.2: Conditional means estimated with smooth splines ($E[y_t|S_{t-p}]$ for $p = 1, ..., 6$ for the sample 1953-1983 in the two upper panels and for $p = 1, ..., 8$ for sample 1984-2000 in the lower panels)

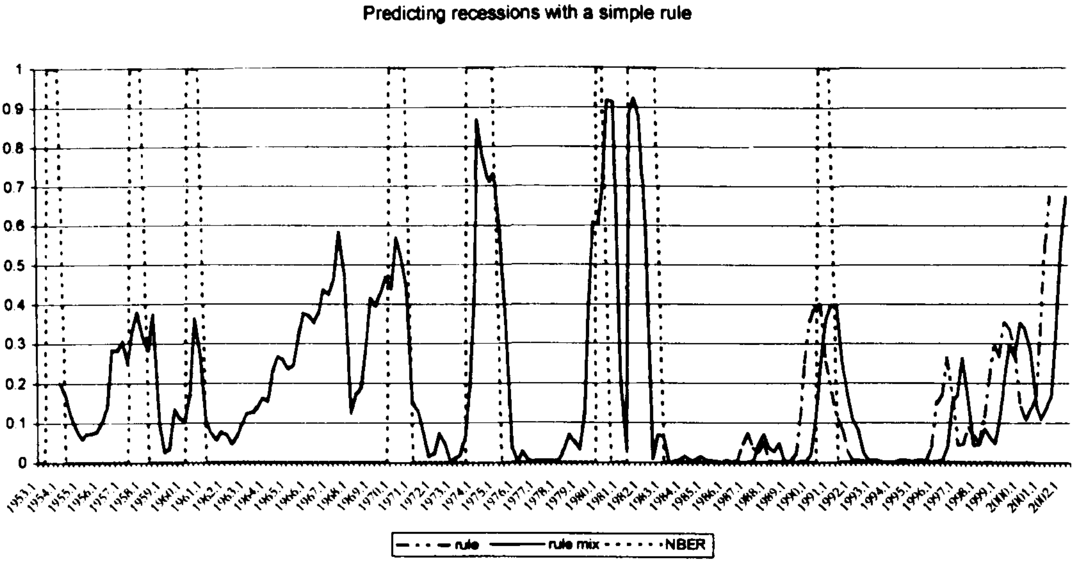

Figure 4.3: Predicting US recessions with a simple rule

Using this alternative cut-off point, there is only one false alarm in the 60's when the government was reducing long-term interest rates to increase economic growth. The fit for the recessions in the 50's is not very impressive, but the spread can predict the 1990-91 recession. The change in d after 1983 does improve the forecasting performance for the last recession. Accepting $d = 6$, a recession is predicted for 2002:1, using $d = 3$ the predicted recession is in 2001:2.

This preliminary analysis of the data suggests a structural break in the mean of the spread and in the variance of output growth in 1984. The analysis of the plots of conditional mean shows that the predictive of the spread in the more recent sample is reduced if non-linearities are not taken into account. Statistical tests for these hypotheses are conducted in the next sections. Furthermore, a simple rule, without any sophisticated estimation techniques, extracts information from the spread to predict NBER recessions correctly.

any scaling (constant equal to zero). The simple rule is evaluated with other models in section 4.8 when the optimal predictive horizon is assumed to be equal to 5 and a recession occurs when $Pr[recession] > 0.50$.

4.3 A linear model

The possibility of non-linearities and structural breaks in the relationship between the spread and output growth is tested in a linear model. We define a bivariate VAR(3) using information criteria and LR tests, based on a maximum autoregressive order of 6. The F tests on the autoregressive lags could not reject the null hypothesis that the third lag of output growth is insignificant. Therefore, our benchmark model has the same explanatory variables for both equations, but a maximum lag of 2 for output growth (y_t) and a maximum lag of 3 for the spread (S_t). Equation 4.A3[3] presents the estimated coefficients for the period 1953:1 to 1999:4, given that the observations for 2000 are left for posterior forecasting analysis. The long-run multiplier of the spread on the output growth is 0.3, supporting the positive relationship between spread and output and the relevance of this variable to predict growth. The sum of the autoregressive coefficients of the spread in the spread equation (0.9) shows that shocks in the spread are very persistent.

4.4 Testing and Modelling Non-linearity

Non-linearities in the relationship between the spread and economic activity have been reported by Dotsey (1998), Anderson and Vahid (2000) and Galbraith and Tkacz (2000). They can be explained by asymmetries in the reaction to monetary policy and asymmetries in business cycles. The predictive power of the spread decreases when its past value is large, suggesting that the spread has better predictive power for contractions than for high growth phases.

[3] All the estimated models of this section and also sections 4.4, 4.5, and 4.6 are described in an Appendix to this chapter.

4.4.1 Testing and Estimating Threshold and Smooth Transition Models

Non-linearities can be tested employing a specific non-linear alternative to the null of non-linearity. Anderson and Vahid (2000) choose the smooth transition regression models and Galbraith and Tkacz (2000) choose threshold regressions. In this work we employ both alternatives. The aim is to estimate a non-linear model using output growth and the spread as endogenous variables. Calling output growth variable y_t, the spread S_t, the transition variables s_{t-d}, and $x_{t-1} = (1, y_{t-1,\ldots,} y_{t-p_1}, S_{t-1}, .., S_{t-p_2})'$, a non-linear model can be written as:

$$\begin{bmatrix} y_t \\ S_t \end{bmatrix} = \begin{bmatrix} \beta_1 x_{t-1}(1 - F_1(s_{t-d_1})) + \beta_2 x_{t-1}(F_1(s_{t-d_1})) \\ \beta_3 x_{t-1}(1 - F_2(s_{t-d_2})) + \beta_4 x_{t-1}(F_2(s_{t-d_2})) \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \tag{4.2}$$

If the non-linear functions $F_1(s_{t-d})$ and $F_2(s_{t-d})$ are equal for each equation of the model, we suppose that $cov(u_{1t}, u_{2t}) \neq 0$. Because the explanatory variables of each equation are the same (x_{t-1}), the model is a generalisation of a VAR model. However, to predict output growth using the spread, and specifically recessions (section 4.8), we can relax the assumption that $F_1(s_{t-d_1}) = F_2(s_{t-d_2})$, given that $cov(u_{1t}, u_{2t}) = 0$. Under these conditions, we can estimate each equation of the model separately. In addition, the assumption that x_{t-1} is the same for each equation and regime can be relaxed. Anderson and Vahid (2000) called the latter type of model a non-linear autoregressive leading indicator model. Another alternative is to consider common non-linearity in the model; that is, there is a linear combination of the variables of $(y_t, S_t)'$ such that the the conditional expectation is linear on x_{t-1}. This implies that following restrictions are imposed in equation 4.2: $F_1(s_{t-d_1}) = F_2(s_{t-d_2})$ and $\beta_2 = \beta_4$. Tests for common non-linearity are proposed by Anderson and Vahid (1998) and applied to a non-linear autoregressive leading indicator model, using the spread as the leading indicator in Anderson and Vahid (2000). The authors found that common non-linearity is rejected by the data and that a model with common non-linearity has significant inferior performance in forecasting recessions compared to a model without this restriction.

Given the strong restrictions imposed in a non-linear model by common non-linearity and the disappointing results of Anderson and Vahid (2000), this work does not attempt to estimate a non-linear model with common non-linearity. We employ two approaches for estimating non-linear models such as equation 4.2: each equation is estimated separately (where x_{t-1} may be different for each regime, $cov(u_{1t}, u_{2t}) = 0$, $F_1(s_{t-d_1}) \neq F_2(s_{t-d_2})$) and both equations are jointly estimated (where x_{t-1} is the same for all regimes, $cov(u_{1t}, u_{2t}) \neq 0$, $F_1(s_{t-d_1}) = F_2(s_{t-d_2})$).

Let z_{it} be y_t when $i = 1$ and S_t when $i = 2$, then each equation of the non-linear model (eq. 4.2) can be written as a smooth transition regression:

$$z_{it} = \beta_1 x_{t-1}(1 - G(s_{t-d})) + \beta_2 x_{t-1}(G(s_{t-d})) + u_{it},$$

$t = 1, ..., T$, where $x_{t-1} = (1, y_{t-1}, y_{t-2}, S_{t-1}, S_{t-2}, S_{t-3})'$; s_{t-d} is the transition variable and $G(s_{t-d})$ is a smooth transition function that can be logistic: $G(s_{t-d}) = 1/(1+\exp\{-\gamma(s_{t-d} - c)\})$ or exponential: $G(s_{t-d}) = 1 - \exp\{-\gamma(s_{t-d} - c)^2\}$. Smooth transition regressions are estimated by conditional non-linear least squares, using the information that, given the parameters in the transition function, the minimisation is a linear problem (for more details, see Van Dijk, Teräsvirta and Franses, 2001). Using initial parameters indicated by a grid search, the sum of the squares of the residuals is minimised using the Newton-Raphson algorithm, and the c parameter is constrained to be inside the interval of the minimum and maximum of the transition variable (trimmed by 5%)[4]. The smooth parameter is standardised by the standard deviation of the transition variable, which helps with the estimation of the parameters, as suggested by Teräsvirta (1998). Significance tests (t-tests) are employed to determine the autoregressive orders of each regression equation, using the specification of the linear model (eq. 4.A3) as a benchmark. Therefore smooth transition models may have different x_{t-1} in each regime.

[4]The models and tests of this chapter are computed using Gauss programming language with code written by the author. The code to estimate Smooth Transition models is based partially on Dick van Dijk's code employed in Franses and Van Dijk (2000) and Van Dijk, Strikholm and Teräsvirta (2001).

A threshold regression with two regimes can be written as:

$$z_{it} = \beta_1 x_{t-1}(1 - I(s_{t-d})) + \beta_2 x_{t-1}(I(s_{t-d})) + u_{it},$$

where $I(s_{t-d}) = 0$ when $s_{t-d} \leq r$ and $I(s_{t-d}) = 1$ when $s_{t-d} > r$. This regression can be easily extended to allow three regimes (including one more step function with a different threshold). Conditional on the threshold r, variances of residuals are computed for each regime, that is, $\sigma^2_{u_i} = \sigma^2_L(1 - I(s_{t-d})) + \sigma^2_U I(s_{t-d})$. These regime-dependent variances are employed to calculate the variance-covariance matrix of the coefficients. The threshold of two-regime threshold regressions is estimated by grid search over all possible threshold values $r \in [r_L, r_U]$. The upper and the lower values of this interval are calculated as follows. The values ($t = 1, ..., T$) of the transition variable are sorted and a proportion π of the observations is trimmed in each end with $0 < \pi < 1$. The thresholds of models with three regimes are estimated one-step-at-a-time, using the method suggested by Hansen (2000a), as discussed in Chapter 3. The delay is jointly estimated with the thresholds (by grid search) given that $d_L = 1$ and $d_U = d_{\max}$. Conditional on the threshold value, the delay and the autoregressive order, the threshold models are estimated by OLS. The threshold regression can be easily written as a model of VAR type (eq. 4.2), assuming that $cov(u_{1t}, u_{2t}) \neq 0$ and that each equation has the same $I(s_{t-d})$. In this case, conditional on the threshold value, the model is estimated by multivariate least squares that is equivalent to OLS estimation for each equation. The thresholds of threshold VAR models are estimated using the same procedures employed for the regressions.

For both types of parametric non-linearity, the testing procedure faces the problem of nuisance parameters under the null hypothesis. This problem is solved when the smooth transition models are the alternative, by employing Taylor approximations of the transition function, allowing the utilisation of standard test distributions (Granger and Teräsvirta, 1993; Teräsvirta, 1998; Van Dijk, Teräsvirta and Franses, 2001). For threshold models, Hansen

(1996) developed asymptotic theory and a numerical method to calculate the asymptotic distribution, although bootstrapping is shown to have a better performance in finite samples (Hansen, 2000a).

Employing the Teräsvirta (1998) method, we use the following auxiliary regression to test non-linearity:

$$z_{it} = \beta_0 x_{t-1} + \beta_1 \tilde{x}_{t-1} s_{t-d} + \beta_2 \tilde{x}_{t-1} s_{t-d}^2 + \beta_3 \tilde{x}_{t-1} s_{t-d}^3 + \tilde{u}_{it} \tag{4.3}$$

where $\tilde{x}_{t-1} = (y_{t-1}, y_{t-2}, S_{t-1}, S_{t-2}, S_{t-3})'$ and s_{t-d} is taken to be one of the variables in the $\tilde{x}_{t-1}$. Three tests are realised for each transition variable: ST_1, ST_2, ST_3. ST_1 tests if $\beta_1 = 0$, supposing that β_2 and β_3 are zero in the auxiliary regression (4.3). ST_2 tests if $\beta_2 = 0$, supposing that β_3 is zero in the auxiliary regression, and, finally, ST_3 tests if $\beta_3 = 0$ in 4.3. We employ the F version of the test because of its better small sample properties[5]. These statistics can be used as model selection criteria to decide between the logistic or the exponential transition functions. If the smallest p-value (that implies rejection of linearity) is achieved with the ST_2 test, a model with exponential transition function is suggested, but if the stronger rejection is from the ST_1 or ST_3 test, a logistic transition function is indicated. These non-linearity tests have been employed to model non-linear autoregressive leading indicators in Anderson and Vahid (2000) and they have good power even in small samples as reported by Teräsvirta (1998). We can also generalise the test using a VAR as null hypothesis and writing equation 4.3 with $\mathbf{x}_t = (y_t, S_t)'$, $\epsilon_t = (u_{1t}, u_{2t})$ and $\epsilon_t \sim N(0, \Omega)$. The F statistic is calculated using $trace[\epsilon'\epsilon]$ as sum square of the residuals of the models[6]. Van Dijk (1999) applies this type of testing to a bivariate VAR of spot and future prices. In addition, he shows, using Monte Carlo evaluation, that the test can be wrongly sized

[5]Given the number of restrictions r, the sample size T and the number of parameters estimated under the alternative hypothesis k, the F statistic is $\frac{(S_0 - S_1)/r}{S_1/(T-k)} \sim F(r, T-k)$, where S_0 are the residual sum square under the null hypothesis and S_1 is the residual sum square under the alternative hypothesis.

[6]Specifically, $\frac{trace[\epsilon_0'\epsilon_0] - trace[\epsilon_1'\epsilon_1]/r}{trace[\epsilon_0'\epsilon_0]/(2T-k)} \sim F(r, 2T-k)$, where k is the sum of parameters estimated in both equations and ϵ_0 is the residual matrix under the null hypothesis and ϵ_1 is the residual matrix under the alternative.

and lack power when the sample size is smaller than 200, therefore the test results based on the VAR should be analyzed with care. These non-linearity tests can be calculated using the heteroscedasticity robust procedure outlined in Granger and Teräsvirta (1993, p. 69-70). However, because the heteroscedasticity in the null hypothesis can be the result of non-linearity that is not taken into account in the conditional mean, this correction removes most of the power of the test to detect non-linearity (Lundbergh and Teräsvirta, 1998). This reduction of power is considerably larger than the one associated with the statistics proposed by Hansen (1996), who detected only a mild reduction. Therefore, the linearity test using the auxiliary regression (eq. 4.3) presented is not robust to an unknown form of heteroscedasticity.

The Hansen (1996) test for linearity against the alternative hypothesis of a threshold model does not have a conventional asymptotic distribution. F statistics are calculated for comparison between the linear model and the threshold model for each possible value of the threshold. The values of the transition variable, given the delay, determine the range of possible threshold values. We take the spread as the transition variable and denote $d_U = d_{\max}$ and $d_L = 1$, and r_U and r_L are defined by $\pi = .10$. The *supF* statistic is the maximum F statistic obtained in a search over possible thresholds and delays:

$$\sup F = \max_{\substack{r_L \leq r \leq r_U \\ d_L \leq d \leq d_U}} \left(\frac{S_L - S(r,d)}{S(r,d)/(T-k)} \right), \tag{4.4}$$

where k is the number of parameters estimated under the alternative hypothesis; S_L is the sum of the squares of the residuals of the linear model (eq. 4.A3) and $S(r,d)$ is the sum of the squares of the residuals of the threshold model given a combination of r and d. The p-value of this test is calculated using the asymptotic results of Hansen (1996). Hansen (2000a) argues that a bootstrap is a better approximation for finite samples, but his bootstrap approach implies being able to simulate values of y_t and S_t, which is not possible when we test each equation separately. Another alternative is to apply the fixed regressor bootstrap employed in the context of parameter instability testing (Hansen, 2000b). However, the gains of using fixed regressor bootstrap to obtain p-values arise from its robustness to breaks in the regressors, as a consequence of this type of bootstrap not being a finite sample approximation. In addition to the asymptotic p-values, asymptotic heteroscedasticity corrected p-values are calculated (Hansen, 1996).

The threshold non-linearity test is also conducted in the VAR, applying a LR test. Given the estimated variance-covariance matrices ($\hat{\Omega}_j$) of the residuals $(u_{1t}, u_{2t})'$, the supLR for testing non-linearity with a two-regime threshold model under the alternative is

$$\sup LR_{12} = \max_{\substack{r_L \leq r \leq r_U \\ d_L \leq d \leq d_U}} (T(\det(\hat{\Omega}_1) - \det(\hat{\Omega}_2))), \tag{4.5}$$

where $j = 1$ is the index for the linear model, $j = 2$ is the index for a two-regime threshold model. A test of a two-regime against a three-regime threshold model is also performed using this same type of test (LR_{23}). The p-value of the LR test is calculated using heteroscedasticity corrected bootstrap[7]. These tests are also described in Chapter 3, where they are employed to test TVEqCMs.

4.4.2 Analysis of the Results

The statistics and p-values of the non-linearity tests presented in Table 4.2 show that the logistic function is the appropriate smooth transition function for both equations. When the non-linearity is tested in each regression, S_{t-2} is chosen as the transition variable for the output equation and S_{t-3} is the best alternative for the spread equation. When non-linearity is tested jointly for both equations, S_{t-1} is selected as the transition variable. The tests using the threshold regressions as alternative hypothesis indicate S_{t-1} as the transition variable for both equations, while the heteroscedastic corrected bootstrap suggests S_{t-2} for the output equation. The results of the $supLR_{12}$ test confirm the T_1 tests that S_{t-1} is the

[7]For the LR_{12} test, the residuals of the linear model are standardised by the fitted values of a regression of the squared of the errors on x_{t-1}^2. For the LR_{23} test, the bootstrap assume that the errors of the two-regime model are regime-dependent.

Table 4.2: Non-linearity tests

Test	trans. var./ equation	y_{t-1}	y_{t-2}	S_{t-1}	S_{t-2}	S_{t-3}	S_{t-4}
ST_1	output	0.650 (0.662)	1.456 (0.204)	3.476 (0.005)	3.637 (0.004)	2.069 (0.072)	
	spread	1.801 (0.115)	6.046 (0.000)	4.886 (0.001)	2.148 (0.062)	6.595 (0.000)	
	both	0.948 (0.488)	2.579 (0.005)	3.840 (0.001)	3.230 (0.001)	3.172 (0.001)	
ST_2	output	0.259 (0.935)	2.816 (0.018)	1.912 (0.095)	1.185 (0.317)	0.446 (0.816)	
	spread	2.035 (0.076)	4.009 (0.002)	2.253 (0.051)	6.209 (0.000)	6.192 (0.000)	
	both	0.703 (0.722)	3.099 (0.001)	1.999 (0.033)	2.489 (0.008)	1.677 (0.082)	
ST_3	output	0.418 (0.836)	0.870 (0.503)	0.293 (0.917)	0.243 (0.943)	1.314 (0.260)	
	spread	2.409 (0.039)	5.636 (0.000)	2.863 (0.017)	2.144 (0.063)	6.583 (0.000)	
	both	0.894 (0.540)	1.890 (0.046)	0.917 (0.518)	0.694 (0.730)	2.329 (0.011)	
T_1	output	8.61 (0.840) [0.836]	21.73 (0.022) [0.086]	22.28 (0.021) [0.086]	22.13 (0.024) [0.078]	15.36 (0.232) [0.352]	
	spread	33.99 (0.000) [0.176]	41.32 (0.000) [0.158]	46.53 (0.000) [0.042]	28.63 (0.000) [0.271]	35.87 (0.000) [0.269]	
$\sup LR_{12}$	both			46.085 [0.009]			61.348 [0.002]
$\sup LR_{23}$	both			24.827 [0.343]			29.217 [0.220]

Note: ST_1, ST_2, ST_3 are non-linearity tests with smooth transition regression alternative; T_1 is a supF test for threshold non-linearity with p-value given by homoscedastic ( ) and heteroscedastic [ ] asymptotic dist.; $\sup LR_{12}$ is a test for threshold non-linearity (hetero. corrected) with maximum d = 3 and d = 5, the alternative is a 2 regime model; $\sup LR_{23}$ tests a two-regime against a three-regime threshold model; number of bootstraps = 1000.

best transition variable when the maximum delay is assumed to be 3 (the implicit assumption of the previous test), but when the maximum delay is equal to 5 the test supports S_{t-4} as the transition variable. The alternative hypothesis of a three-regime threshold model is rejected by the data, using the results of the $\sup LR_{23}$ test.

The results of the non-linearity tests suggest a smooth transition model (ST) (eqs. 4.A4 and 4.A5) and a threshold (T) model (eqs. 4.A6 and 4.A7) when the equations are estimated separately for the spread and output growth. In addition, with the same non-linearity function for both equations, the tests suggest a threshold system with S_{t-4} as transition variable (TVAR, eq. 4.A8). Given the results of Anderson and Vahid (2000), based on models with common smooth transition dynamics and the fact that the test for the spread equation strongly suggest S_{t-3} as transition variable while the system version suggests S_{t-1}, we decide not estimating a smooth transition model with the same transition function in each equation.

The long-run multiplier of the spread on output is larger for the first regime than for the second in the three models (ST, T and TVAR). In the case of the ST model (eq. 4.A4), for example, the long-run effect of changes in the spread is 0.62 in the first regime and zero in the second regime, compared with 0.3 for the linear model (eq. 4.A3). This confirms the results of Galbraith and Tkacz (2000) that the predictive power of the spread decreases at larger values: in their case the break is at 2, in our case is at 1.7.

Given the evidence of Chapter 3 that a three-regime model can describe some relevant characteristics of the spread dynamics, we consider a three-regime model. The three-regime alternative is not supported when the $supLR_{23}$ is employed, but this could be because either the output equation strongly rejects the additional regime or the heteroscedasticity correction reduces the power of the test to detect non-linearities (see, e.g., Hansen, 1996, when using asymptotic distribution). A test for remaining non-linearity in the spread equation of the ST model (eq. 4.A5) rejects the null hypothesis, so two additive transition functions, as suggested by Van Dijk and Franses (1999), can help to model this non-linearity. Anderson and Vahid (2000) propose predicting recessions with smooth transition models similar to the ST model (eqs. 4.A4 and 4.A5). They also found that a logistic function with S_{t-3} as transition variable is inadequate to represent the non-linear behaviour of the spread. Equations 4.A14 and 4.A10 are regressions for the spread with three regimes. The fit of the smooth transition regression (dST) is better than the threshold regression (3T) and these equations give different transition variables as a result of the non-linearity testing reported in Table 4.2. Tests on the residuals of equation 4.A14 indicate that there is still some remaining non-linearity. Because the non-linearity testing also has power against heteroscedasticity, a model that allows some form of heteroscedasticity in the residuals is a possible solution to this problem. The three-regime threshold model (3T) is able to account for some heteroscedasticity because it has regime-dependent variances of the residuals.

Confirming previous results, we found strong evidence of non-linearity in the relationship between spread and output. The tests indicate threshold (T and TVAR) and smooth transition models (ST).

4.5 Testing and Modelling Structural Breaks

Instability in the regression of output growth on the spread has been reported by Haubrich and Dombrosky (1996), Estrella et al. (2000) and Stock and Watson (2001). However, this instability is not found in probit models to predict recession using the spread (Estrella et al., 2000). The parameter instability when tested on a regression between output growth and the spread may be influenced by changes in the variance of the output series, such as reported by Kim and Nelson (1999b) and McConnell and Perez-Quiros (2000). The output growth may be better represented by structural break models and time-varying models than non-linear models (Koop and Potter, 2000; Koop and Potter, 2001). In this section, we estimate time-varying and structural break models for the relationship between output growth and spread. These two type of models can be nested in:

$$\begin{bmatrix} y_t \\ S_t \end{bmatrix} = \begin{bmatrix} \beta_1 x_{t-1}(1-F_1(t)) + \beta_2 x_{t-1}(F_1(t)) \\ \beta_3 x_{t-1}(1-F_2(t)) + \beta_4 x_{t-1}(F_2(t)) \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \tag{4.6}$$

In the case of structural break models, $F_i(t)$ is a step function, which means that a discrete break occurs at a point in time, switching the parameters from β_1 or β_3 to β_2 or β_4. In the time-varying models, this switch is smooth, in consequence, the structural break is continuous and $F_i(t)$ is a logistic function. The tests and modelling described in this section allow $F_1(t)$ to be different from $F_2(t)$, which means that the break points are different in each equation. We also impose the restriction that $F_1(t) = F_2(t)$ in some cases. We consider in this section only complete structural change, which means that all the coefficients in β_1 and β_3 change over time. Some attempts to estimate partial structural change models (not presented) show that the complete structural change models are statistically preferred to partial change models.

Alternatively, we test whether there is any structural break in the variance, given that the mean equation is linear. The model in this case employs a step function $I_i(t)$ and the variance equation is written as:

$$\begin{bmatrix} \sigma^2_{u_{1t}} \\ \sigma^2_{u_{2t}} \end{bmatrix} = \begin{bmatrix} \sigma^2_A(1-I_1(t)) + \sigma^2_B(I_1(t)) \\ \sigma^2_C(1-I_2(t)) + \sigma^2_D(I_2(t)) \end{bmatrix}.$$

4.5.1 Testing and Estimating Parameter Instability in Mean and Variance using Threshold and Smooth Transition Models

For testing and estimating a structural break model (eq. 4.6), we suppose initially that $F_1(t) \neq F_2(t)$ and that the equations of the model are estimated separately. Given the linear regression $z_{it} = \beta x_{t-1} + u_{it}$ for $t = 1, ..., T$, the structural change in this regression arises in the coefficient β (Hansen, 2000b). A structural break regression can be written as:

$$z_{it} = \beta_1 x_{t-1} I(t) + \beta_2 x_{t-1}(1 - I(t)) + u_{it},$$

where $I_t = 1$ if $t \leq \tau$ and $I_t = 0$ if $t > \tau$; and $\beta_2 = \beta_1 + \theta$. The null hypothesis for the test for structural change in the conditional mean is that $\theta = 0$. Given τ, one may calculate a different residual variance for each regime. The break point τ can be estimated using grid search over the values of $\tau \in [\tau_L, \tau_u]$, where the limits of this interval are calculated by trimming t by π. Likewise, a model with structural break in the variance can be written as:

$$z_{it} = \beta x_{t-1} + u_{it} \tag{4.7}$$

$$\sigma^2_{u_{it}} = \sigma^2_B I(t) + \sigma^2_A (1 - I(t)),$$

where $\sigma^2_A = \sigma^2_B + \theta_v$ and the null hypothesis of no structural break in the variance is that $\theta_v = 0$.

Testing structural breaks in the mean and the variance for unknown break points implies the presence of a nuisance parameter under the null hypothesis, as in the non-linearity testing. Hansen (2001) presents an up-to-date literature review on the advances in this area. The standard test for one structural break is the *supLR*, proposed by Andrews (1993). The idea is to calculate the LR statistic for all possible structural breaks, given that $\tau \in [\tau_L, \tau_U]$, where $\tau_L = \pi T$ and $\tau_L = (1 - \pi)T$[8]. Then the maximum LR value over $[\tau_L, \tau_U]$ is the test statistic. The test is equivalent to the $\sup F$ described in eq. 4.4 with the difference that x_{t-1} is ordered chronologically rather than ranked by the threshold value.

The p-values can be calculated numerically by the Hansen (1997a) procedure, given the number of restrictions (6 given the defined x_{t-1}) and the trimming factor π. Hansen (2000b) shows that a fixed regressor bootstrap works better than the asymptotic values in small samples. The residuals can be corrected for heteroscedasticity before they are

[8] Note that it may be necessary to adjust the values of πT and $(1 - \pi)T$ to the next integer to get feasible numbers for τ.

bootstrapped to generate values of the endogenous variable. In this work, we use the approach of Hansen (2000b) for testing structural breaks in regression coefficients. The *supF* is a test for one unknown structural break using an F statistic (as eq. 4.4) and p-value numerically calculated from its asymptotic distribution. Explicitly, the $\sup F$ is:

$$\sup F = \max_{\tau_L \leq \tau \leq \tau_U} \left(\frac{S_L - S(\tau)}{S(\tau)/(T-k)} \right). \tag{4.8}$$

P-values by fixed regressor bootstrap are also calculated using homoscedastic and heteroscedastic bootstrap (Hansen, 2000b)[9].

In the case of testing for structural breaks jointly in both equations ($F_1(t) = F_2(t)$), we use two approaches. The first one is to calculate the multivariate version of the $\sup F$ test (eq. 4.8). A similar approach is proposed by Bai, Lumsdaine and Stock (1998), who apply a multivariate version of Andrews (1993). For this test, we employ the asymptotic p-value, given $\pi = 0.10$ and 12 restrictions (Andrews, 1993). The second approach is to apply the $\sup LR_{12}$ (eq. 4.5), employed to test non-linearity in section 4.4, using a time trend as the transition variable. Using this same approach, a test for a model with two structural breaks against one structural break is also applied ($\sup LR_{23}$), which is similar to the test for multiple breaks proposed by Bai (1999). The p-values, as in the previous section, are calculated by bootstrap with a correction for heteroscedasticity.

A similar procedure can be employed to test and to estimate a model with a structural break only in the variance equation such as equation 4.7. Instead of seeking a break in a regression of, say, output growth against lagged values of the spread and output growth, the break has to be located in regressions of the square of the residuals of each equation of the linear model (eq. 4.A3) against a constant. Therefore, the $\sup F$ (eq. 4.8) is employed to test for a structural break in the variance: as above, p-values are calculated using the

[9]In the homoscedastic bootstrap, values are drawn from a normal distribution and regressed against the regressors under the null hypothesis and the regressors under the alternative hypothesis. The residual variances are calculated for both regressions and the F statistic is calculated. In the heteroscedastic bootstrap, the values are drawn from a normal distribution multiplied by the errors of the model under the alternative hypothesis.

asymptotic distribution and homoscedastic and heteroscedastic fixed regressor bootstraps.

Lin and Teräsvirta (1994) suggest testing for smooth structural breaks, instead of a discrete jump: a test for continuous structural change. The tests employ logistic transition functions with the time trend as the transition variable. Therefore, a time varying regression can be written as:

$$z_{it} = \beta_1 x_{t-1}(1 - G(t)) + \beta_2 x_{t-1}(G(t)) + u_{it},$$

where $G(t) = 1/(1 + \exp\{-\gamma(t - c)\})$. Using Teräsvirta's (1998) approach for modelling smooth transition models, we can apply the same test setup as employed to test smooth transition non-linearity in the last section but with the time trend as the only possible transition variable ($s_{t-d} = t$) with eq. 4.3 as auxiliary regression. Table 4.3 presents these tests for continuous structural change in the mean, called ST_1, ST_2, and ST_3.

4.5.2 Analysis of the Results

The tests for structural breaks in the mean and in the variance and also for time-varying parameters are presented in Table 4.3[10]. The tests for parameter instability strongly reject the null hypothesis of no changes in the output-growth equation and, to a lesser extent, in the spread equation. The tests with the alternative hypothesis of continuous change cannot reject the null hypothesis, and the heteroscedasticity corrected version of the test for discrete changes also cannot reject the null hypothesis when the spread regression is analysed. The multivariate version of the $\sup F$ rejects the null hypothesis, implying that there is a evidence of a common break point. As argued by Bai et al. (1998), multivariate models have more information about a possible structural break in the relationship between two endogenous variables. However, the break-points of each equation are different when estimated separately: 1959:2 and 1981:1. To understand how in this case the multivariate

[10] The codes for the $supF$ and the $supF_{var}$ tests are based on Hansen's code employed in the papers Hansen (2000b) and Hansen (2001).

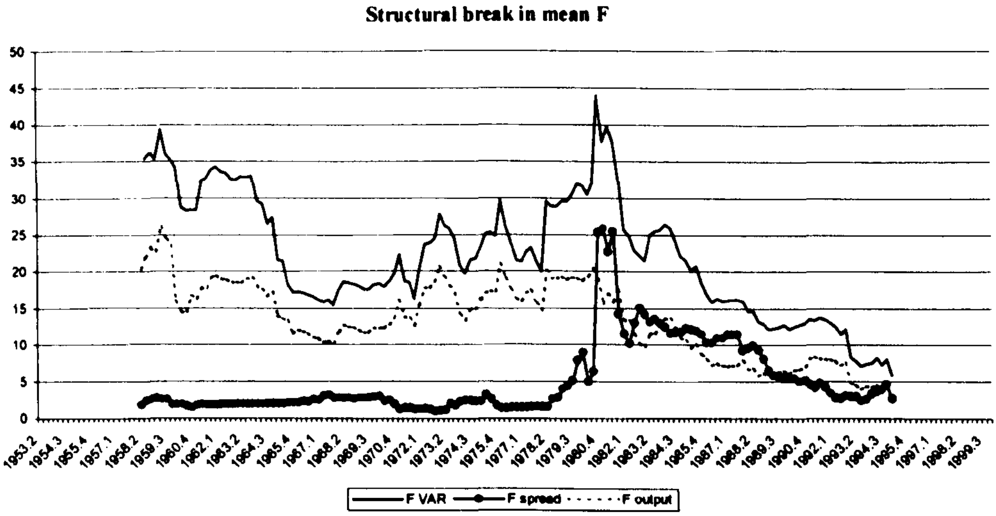

Figure 4.4: F statistics for all possible break-points for test of structural break in mean

sup F reject the null hypothesis, we plot the F statistics for each separate equation and for both equations in Figure 4.4. The figure also helps to understand how precisely the break is estimated. The plot of the F statistic of the spread equation has a more evident inverted V shape when compared with the F of the output equation, meaning that there is less uncertainty in the location of the break-point in the spread equation. The break identified by the multivariate F is essentially the same as the one of the spread equation. Observe that the break can be spurious as a consequence of being identified in a period in which interest rates are extremely volatile, that is, the structural break in the conditional mean can be created by a change in the conditional variance. In fact, the $\sup LR_{12}$ has its p-value calculated using the heteroscedastic bootstrap and does not find evidence of a structural break. Therefore, the tests suggests that there is a structural break in the output growth equation and maybe in the spread equation.

The tests for a structural break in the variance strongly support the existence of a break in the variance of the output growth equation. The same test when it is assumed that there is a structural break in the mean does not change this result (not shown). The

Table 4.3: Time-varying/Structural break tests

Test/equation	output	spread	both
ST_1	3.124 (0.006)	0.570 (0.754)	2.403 (0.005)
ST_2	1.092 (0.369)	0.478 (0.824)	0.916 (0.531)
ST_3	2.240 (0.042)	0.401 (0.877)	1.688 (0.068)
$\sup F$	26.121 (0.006) [0.011] {0.024}	25.880 (0.007) [0.011] {0.150}	43.991 (0.001)
Break Dates	1959:2	1981:1	1980:4
$\text{sup}LR_{12}$			28.327 (0.153)
Break Dates			1980:4
$\sup LR_{23}$			11.820 (0.952)
Break Dates			1971:2, 1980:4
$\sup F_{var}$	23.468 (0.001) [0.001] {0.005}	12.403 (0.007) [0.013] {0.158}	
Break Dates	1983:2	1966:1	

Note: ST_1, ST_2, ST_3 are tests for continuous structural change with smooth transition alternative; the supF tests for changes in the mean with p-values given by asymp. ( ), homo. boots. [ ] and hetero. boots. { }; supLR_{12}, supLR_{23} are tests for structural break in systems with hetero. boots. p-value; supF_{var} is equivalent to supF to test structural break in the variance; number of bootstraps = 1000.

break point estimated for the variance (1983:2) of the output equation is similar to the value (1984:1) of univariate models of output growth (Kim and Nelson, 1999b; McConnell and Perez-Quiros, 2000). In the case of the spread, the heteroscedastic fixed regressor bootstrap suggests again that the 1966:1 break in the variance is not statistically significant. This could be the result of a poor estimated break point when changes are detected in the conditional mean and in the conditional variance. Therefore, we follow Hansen's (2001) suggestion of searching jointly for a structural break in mean and in variance. The structural break in mean and in variance model (SBMV) is defined as

$$z_{it} = \beta_1 x_{t-1} I(t) + \beta_2 x_{t-1}(1 - I(t)) + u_{it}$$

$$\sigma^2_{u_{it}} = \sigma^2_B I(t) + \sigma^2_A (1 - I(t))$$

Conditional on the break point τ, this regression is estimated by maximum likelihood. The break point is estimated by grid search, using $\pi = 0.20$. The point that gives the maximum value of the maximum likelihood for the spread is at 1981:1 and for output at 1980:4. Although each equation was estimated separately, the break point is essentially the same, occurring in the period when a new monetary policy regime created strong interest rate volatility (Watson, 1999). Therefore, we suggest a structural break model with changes in the mean and the variance equations as good data representation (SBMV, eqs. 4.A11 and 4.A12).

Given the failure to reject the null hypothesis of time-varying parameters in the spread equation, another model suggested by the test results of Table 4.3 is a smooth time-varying model for output. When a model with one smooth transition function is estimated using the trend as transition variable, the parameter constancy test on the residuals of this model rejects the null hypothesis. This suggests a model with two transition functions to account for the possibility of two structural breaks (equation 4.A13).

Comparing the long-run multiplier of the spread on output growth of the SBMV model (eq. 4.A11) with the T model (eq. 4.A6), we can observe that the models are suggesting different ways of breaking the data. The multiplier is 0.87 for the first regime in the threshold model and 0.66 in SBMV; and it is -0.14 in the second regime of the T model and 0.21 for the SBMV model.

Therefore, the data supports parameter instability in the output growth equation and, to a lesser extent, in the spread equation. Although threshold models can account for parameter instability (Koop and Potter, 2000), the analysis of the dynamic multipliers indicates structural breaks may coexist with non-linearities.

4.6 Testing and Modelling Time-Varying/Structural Break Non-linearity

The tests applied in section 4.5 are based on the linear model (eq. 4.A3). In this section, we apply tests for continuous or discrete structural breaks, using some of the non-linear models of section 4.4 as the null hypothesis. Although non-linear models can capture some characteristics of structural break models (Clements and Smith, 1999b; Koop and Potter, 2000; Koop and Potter, 2001), it may be the case that the break also implies changes in the non-linear parameters. Time-varying non-linear models have been proposed by Lundbergh, Teräsvirta and Van Dijk (2000) and these have been applied to capture changes in seasonality in industrial production by Van Dijk, Strikholm and Teräsvirta (2001).

The models estimated and tested in this section to predict output growth using the spread can be nested in:

$$\begin{aligned} y_t &= x_{t-1}\beta_1[(1-F_1(s_{t-d_1}))(1-F_1(t))] + x_{t-1}\beta_2[F_1(s_{t-d_1})(1-F_1(t))] \\ &\quad + x_{t-1}\beta_3[(1-F_2(s_{t-d_2}))F_1(t)] + x_{t-1}\beta_4[F_2(s_{t-d_2})F_1(t)] + u_{1t} \\ S_t &= x_{t-1}\beta_5[(1-F_3(s_{t-d_3}))(1-F_2(t))] + x_{t-1}\beta_6[F_3(s_{t-d_3})(1-F_2(t))] \\ &\quad + x_{t-1}\beta_7[(1-F_4(s_{t-d_4}))F_2(t)] + x_{t-1}\beta_8[F_4(s_{t-d_4})F_2(t)] + u_{2t}. \end{aligned}$$

As previously, $F_i(s_{t-d_i})$ and $F_i(t)$ can be logistic functions or step functions; x_{t-1} may not be same for all regimes when logistic functions are employed and $cov(u_{1t}, u_{2t}) = 0$. The equations can be estimated separately or jointly, the latter assumes that the $F_i(s_{t-d_i})$ and $F_i(t)$ are the same for both equations and that $cov(u_{1t}, u_{2t}) \neq 0$.

4.6.1 Testing and Estimating Parameter Instability in Threshold Models and in Smooth Transition Models

In the case of threshold regressions, the possibility of parameter instability is tested using the method of Hansen (2000b). The main supposition is that the non-linearity is the same in both sub-samples, because, given the threshold value, the model can be treated as a linear model with dummy variables. Specifically, the model under the null is:

$$z_{it} = \beta_1 x_{t-1}(1 - I(s_{t-d})) + \beta_2 x_{t-1}(I(s_{t-d})) + u_{it},$$

where $I(s_{t-d}) = 0$ when $s_{t-d} \leq r$ and $I(s_{t-d}) = 1$ when $s_{t-d} > r$. Consequently, the model under the alternative hypothesis is:

$$\begin{aligned} z_{it} &= [\beta_1 x_{t-1}(1 - I(s_{t-d})) + \beta_2 x_{t-1}(I(s_{t-d}))](1 - I(t)) \\ &\quad + [\beta_3 x_{t-1}(1 - I(s_{t-d})) + \beta_4 x_{t-1}(I(s_{t-d}))](I(t)) + \tilde{u}_{it}, \end{aligned} \tag{4.9}$$

where $I(t) = 0$ when $t \leq \tau$ and $I(t) = 1$ when $t > \tau$. Grid search is employed to estimate τ, given that s_{t-d} and r in $I(s_{t-d})$ are known. In the case of the spread equation and of the multivariate model, we obtain $S(\hat{r}, \hat{d}, \tau)$ for each $\tau \in [\tau_L, \tau_U]$ of equation 4.9 using the following regression:

$$z_{it} = \theta_1 x_{t-1} + \theta_2 x_{t-1}[I(s_{t-d})] + \theta_3 x_{t-1}[I(t)] + \theta_4 x_{t-1}[I(t)I(s_{t-d})] + \tilde{u}_{it}{}^{11}.$$

The trimming proportion π is of 10% for regressions and 20% for the multivariate model. These values are employed to calculate a $\sup F$ statistics such as:

$$\sup F = \max_{\tau_L \leq \tau \leq \tau_U} \left(\frac{S(\hat{r}, \hat{d}) - S(\hat{r}, \hat{d}, \tau)}{S(\hat{r}, \hat{d}, \tau)/(T-k)} \right), \tag{4.10}$$

where $k = 24$, giving x_{t-1} with 6 variables. The p-values of the *supF* tests are calculated using the asymptotic distribution and fixed regressor bootstrap (homoscedastic and heteroscedastic) for the output equation and only asymptotic values are calculated for the spread equation and the system.

Using the residuals of the T model (eqs. 4.A6 and 4.A7), we also test for a structural break in the variance, with the $supF_{var}$ statistics described in section 4.5.

The threshold regressions with a structural break in the mean (4.9) can be nested in the time-varying smooth transition autoregressions proposed by Lundbergh et al. (2000), who suggest two methods to specify time-varying smooth transition autoregressions. These methods are extended to regressions in this work. The first one is to test for parameter constancy (Teräsvirta, 1998) in the smooth transition regressions presented in section 4.4 and remaining non-linearity in the time-varying models of section 4.5[12]. The former is computed for the ST model (eqs. 4.A4 and 4.A5), and the latter for the time-varying model with only one smooth transition function[13]. The second method is to test time-varying non-linearity with a linear model under the null hypothesis and, after the rejection of the null hypothesis,

[12]Testing for remaining non-linearity in structural break models implies threshold models being estimated for each sub-sample indicated by the structural break. One of the sub-samples in the case of the output growth equation, for example, has only 22 observations which is a small enough sample to create strong size and power distortions in the test. Therefore, the testing employed verifies structural breaks in threshold models but not non-linearity in structural break models.

[13]Given the number of parameters in the null hypothesis, the sample size and the fact that the test of remaining non-linearity involves the estimation of an auxiliary regression with twice the parameters of the null hypothesis, we do not test for remaining non-linearity in equation 4.A13.

test non-linearity and time variation separately to examine the need for a time-varying non-linear model, instead of a smooth transition model or a time-varying model.

The parameter constancy tests follow the same pattern as the ST_1, ST_2 and ST_3 presented in the previous sections, here called PC_1, PC_2 and PC_3. The test for a time-varying smooth transition model, similar to the one proposed by Lundbergh et al. (2000) is based on the following auxiliary regression:

$$z_{it} = \beta_0 x_{t-1} + \beta_1 \tilde{x}_{t-1} s_{t-d} + \beta_2 \tilde{x}_{t-1} t + \beta_3 \tilde{x}_{t-1} s_{t-d} t + \tilde{u}_{it}, \tag{4.11}$$

where t is a time trend. The test for time-varying smooth transition (TVST) uses an F-test for the null hypothesis $\beta_1 = \beta_2 = \beta_3 = 0$. Given the rejection of the null hypothesis, we employ the $TVSTR_{NL}$ test with H0: $\beta_1 = \beta_3 = 0$ and the $TVSTR_{TV}$ test with $H0 : \beta_2 = \beta_3 = 0$, using equation 4.11 as auxiliary regression. Only if both $TVSTR_{NL}$ and $TVSTR_{TV}$ reject H0 is a time-varying smooth transition model supported. We apply this testing procedure for all variables in $\tilde{x}_{t-1}$, however the test has smaller p-values if the transition variables selected in section 4.4 are employed, thus the reported statistics in Table 4.4 have $s_{t-d} = S_{t-2}$ for output; $s_{t-d} = S_{t-3}$ for the spread and $s_{t-d} = S_{t-1}$ for the system.

4.6.2 Analysis of the Results

The results of tests for discrete or continuous structural breaks in non-linear models are presented in Table 4.4. We do not report the tests for remaining non-linearity in the time-varying regression with one transition function, but they confirm that there is no remaining non-linearity. The $supF$ for structural break in the mean and the $\sup F_{var}$ for structural break in the variance suggest that the inclusion of non-linearity does not affect the existence of instability in the equations. Indeed the break dates are the same for each equation and also for the system. Figure 4.5 shows how the presence of non-linearity improves the sharpness of the estimation of the breakpoint (compare F VAR with F TVAR).

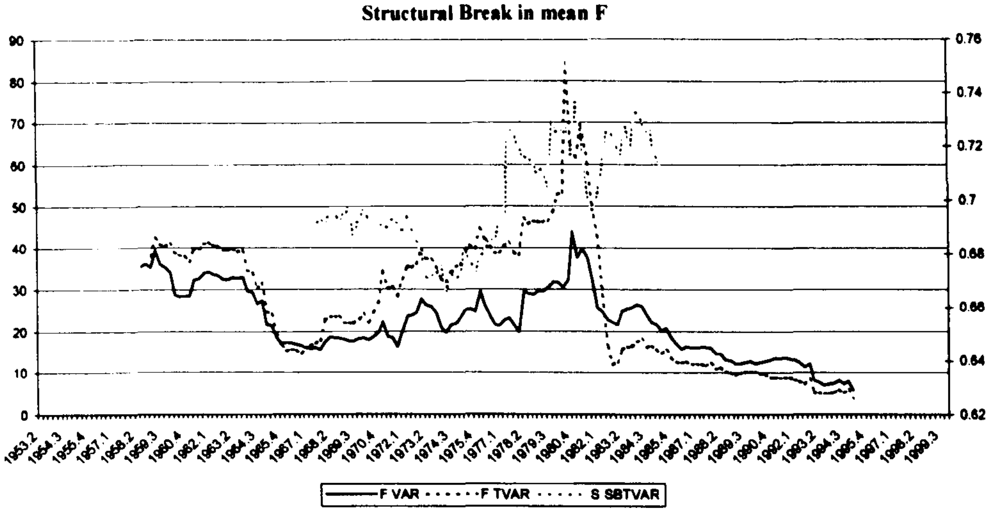

Figure 4.5: F statistics for all possible break-points for test of structural break in variance

However, the test is calculated supposing that the transition function is the same for both sub-samples. We relax this assumption using a grid search to estimate the following model:

$$\begin{aligned} z_{it} &= [\beta_1 x_{t-1}(1 - I_1(s_{t-d_1})) + \beta_2 x_{t-1}(I_1(s_{t-d_1}))](1 - I(t)) \\ &\quad + [\beta_3 x_{t-1}(1 - I_2(s_{t-d_2})) + \beta_4 x_{t-1}(I_2(s_{t-d_2}))](I(t)) + u_{it}, \end{aligned}$$

where $I_1(s_{t-d_1}) = 0$ when $s_{t-d_1} \leq r_1$ and $I_1(s_{t-d_1}) = 1$ when $s_{t-d_1} > r_1$; and $I_2(s_{t-d_2}) = 0$ when $s_{t-d_2} \leq r_2$ and $I_2(s_{t-d_2}) = 1$ when $s_{t-d_2} > r_2$. Conditional on τ (the structural break point), one threshold model is estimated in each sub-sample. The squares of the residuals in each sub-sample are summed and $\hat{\tau}$ is the value that minimises this criterion. Specifically,

$$\hat{\tau}, \hat{r}_1, \hat{d}_1, \hat{r}_2, \hat{d}_2 = \min_{\tau_L \leq t \leq \tau_U} \Big(\min_{\substack{r_{1L} \leq r_1 \leq r_{1U} \\ d_{1L} \leq d_1 \leq d_{1U}}} S_1(\tau, r_1, d_1) + \min_{\substack{r_{2L} \leq r_2 \leq r_{2U} \\ d_{2L} \leq d_2 \leq d_{2U}}} S_2(\tau, r_2, d_2) \Big), \tag{4.12}$$

where S_1 is the residual sum of squares of the threshold model estimated using a grid search over the delay and the threshold value for the first sub-sample and S_2 is the same for the second sub-sample. The proportion of trimming over t is 0.30 and over r is 0.12 for regressions and 0.18 for the system, with $d_U = 5$[14]. The residual sum of squares from the estimation of a

[14]The trimming proportions are larger than the previous models because a reasonable number of observa-

structural break threshold VAR over $\tau \in [\tau_L, \tau_U]$, standardized by $2T-k$, is also presented in Figure 4.5 using the secondary scale. Observe that allowing the transition function $I_i(s_{t-d})$ to change in each sub-sample implies less precise estimation of the break point. However, the estimated break point (1981:1) of the SBTVAR is virtually the same as the break point found for the TVAR (1980:3).

We suggest a structural break threshold model (SBT) (eqs. 4.A17 and 4.A18) and a structural break threshold VAR (SBTVAR) (eq. 4.A19) as good data representations. Given the evidence of a structural break in the variance even when a threshold regression describes the mean equation, we analyse the ability of regime-dependent variances of structural break threshold models to describe changing variances over time in the output growth equation. In the case of the SBT model (eqs. 4.A17 and 4.A18), for example, the standard deviations of the residuals of the first sub-sample are twice those of the second sub-sample and the break point is 1983:4. As a result, the SBT model accounts for a structural break in the variance similar to that described in the univariate models of Kim and Nelson (1999b) and McConnell and Perez-Quiros (2000).

Both general-to-specific and specific-to-general approaches for specifying time-varying smooth transition models are represented in the test statistics of Table 4.4, and both suggest a time-varying smooth transition model for the output equation. In the case of the spread equation, the parameter constancy tests do not indicate the need for the inclusion of a new transition equation but the TVSTR tests suggest that time-varying smooth transition models may be a good idea. The estimation of a time-varying smooth transition model for the spread is a hard task because the algorithm does not converge (given maximum interactions of 1000) when S_{t-3} is employed as the transition variable and the autoregressive lags are given by the linear model. Moreover, small changes in the autoregressive orders create large differences in the estimated parameters in the transition function. This may be a result of

tions is necessary to estimate a threshold regression for each sub-sample.

Table 4.4: Time-varying/Structural break tests in non-linear models

Test/equation	output	spread	both
PC_1	1.681 (0.106)	0.469 (0.920)	
PC_2	1.372 (0.162)	0.724 (0.810)	
PC_3	2.052 (0.010)	0.876 (0.644)	
$TVSTR$	2.480 (0.003)	2.938 (0.000)	0.942 (0.559)
$TVSTR_{NL}$	3.477 (0.000)	3.661 (0.000)	1.511 (0.084)
$TVSTR_{TV}$	3.210 (0.001)	3.228 (0.001)	1.308 (0.182)
$\sup F$	33.402 (0.025) [0.043] {0.038}	45.869 (0.000)	84.440 (0.000)
Break Dates	1960:1	1980:3	1980:3
$\sup F_{var}$	19.194 (0.000) [0.020] {0.051}	13.513 (0.004) [0.066] {0.331}	
Break Dates	1984:1	1966:3	

Note: PC_1, PC_2 and PC_3 are parameter constancy tests; TVSTR is a test for time-varying smooth transition non-linearity; $TVSTR_{NL}$ and $TVSTR_{TV}$ give the indication if a TVST model is necessary; supF and $\sup F_{var}$ are explained in the notes of Table 4.3.

the fact that when the smoothing parameter γ is large, the model is close to a threshold model, making the estimation of γ a hard task (Teräsvirta, 1998). With the small sample available, it is unlikely that the number of observations of the transition variable S_{t-3} in the small neighborhood of c is sufficient to estimate γ with precision. Using lag 1 for both variables and S_{t-1} as the transition variable, we estimate equation 4.A16, but this should be interpreted with caution because the in-sample fit is worse than the linear model. Similar difficulties were found in estimating a time-varying smooth transition model for output, but the algorithm converges easily when the specification of eq. 4.A15 is employed.

Because of its large smoothing parameter, the time-varying smooth transition model (TVST) (eqs. 4.A15 and 4.A16) can be seen as a threshold model. The long-run multiplier of the spread with respect to output in equation 4.A15 is larger in the first sub-sample than in the second (1.47 and 0.81 compared to 0.03 and 0.0), supporting the idea that the impact of changes in the spread on output growth has been reduced. The dynamics of the SBTVAR can explain the result of Stock and Watson (2001), that the spread does not help to predict output growth in the period 1984-1999. The assumption of linearity does not allow to observe that the spread only helps to predict output growth when $S_{t-5} \leq 1$ (eq. 4.A19). Therefore, for 77% of observations of the period 1981-1999, the inference that the spread lost its power to predict output growth is true, but not for the 23% of the observations in which the inequality $S_{t-5} \leq 1$ is valid. The dynamics of the structural break threshold model also explain why there is no instability in predicting recessions using the spread but the output growth-spread relationship is unstable (Estrella et al., 2000): the spread is a good predictor when it is small or negative, implying small or negative growth (recession), but not when it is high, meaning that large spreads do not predict strong output growth. The latter dynamic behaviour is stronger in the more recent period (after 1981).

In summary, we have evidence of changing non-linear behaviour in the relationship between output growth and the spread. This can be represented using a time-varying smooth transition model for output or a multivariate structural break threshold model.

4.7 Forecasts of Event Probabilities

The main concern of the forecasting literature has been point forecasts and, to a lesser extent, interval forecasts. However, recent papers have presented tools to evaluate density forecasts (see Tay and Wallis (2000) for a survey). The main advantage of density forecasts, when compared with point forecasts, is to account for the uncertainty of a prediction that is important when the forecast is employed in decision-making (Tay and Wallis, 2000). In addition, the evaluation of density forecasts is a better way of discriminating between competitors (Clements and Smith, 2000). However, a popular evaluation method based on the probability integral transform (Diebold, Gunther and Tay, 1998) is not useful for macroeconomic data because it has low power of discrimination between linear and non-linear models in sample sizes commonly employed (Clements et al., 2000).

A special case of the density forecast is the probability forecast of future events. Instead of being concerned with all feasible values of the variable being predicted, the event probability forecast refers to the probability of a specific event. Event probabilities may be employed to generate event forecasts. Events may be chosen to meet the information demands of policy-makers and business planners about an economic scenario. Fair (1993), for example, motivates his paper about event probability forecasting by the fact that in 1989 and 1990 policy makers and business planners wanted to know whether there would be a recession. Another example is the analysis of Garratt, Lee, Pesaran and Shin (2000) on the probability of UK inflation being inside the Bank of England's target range.

Predicting the probabilities of US recessions, which are relatively rare events with large potential financial consequences for individuals and companies, has been the focus of attention of econometric researchers. Earlier non-parametric works include models to extract information about the probability of recession from composite leading indicators (Neftci, 1982; Diebold and Rudebusch, 1989). Hamilton (1989) introduces Markov-switching modelling that began a large literature on predicting recessions using filter probabilities from the unobserved component of Markov-switching models (e.g., Lahiri and Wang, 1996 and Ivanova et al., 2000). Another popular model for predicting recessions is the probit model (e.g., Estrella and Mishkin, 1998 and Estrella and Mishkin, 1997), that is, a non-linear regression on leading indicators of a binary variable 0 (expansions) or 1 (recessions); the NBER chronology of expansions and recession have been commonly used in such models. In addition, a non-linear filter to extract information on the probability of recession, using a group of 5 leading indicators based on models for prediction of earthquakes, has been proposed by Keilis-Borok, Stock, Soloviev and Mikhalev (2000). Koskinen and Öller (2001) present a non-linear filter calibrated to extract information from past GDP changes and the composite leading indicator for predicting the probability of business cycle turning points.

Filardo (1999), after evaluating five models to predict recessions, concludes that recession prediction is more accurate when different models are predicting the event. This is the motivation for Camacho and Perez-Quiros (2000) to propose a linear combination rule using a Markov-switching model and a non-parametric model to forecast recessions.

The models suggested by this literature consider recessions as unobserved components or a binary variable and naturally lead to recession probabilities. However, the concept that predicting a recession is nothing more than predicting a specific event implies that event probabilities are derived from density forecasts. Stock and Watson (1989) show how to obtain the probabilities of recession from a dynamic factor model of leading indicators using stochastic simulation. Stochastic simulation has been used to extract the probability of an event from macroeconometric models (Fair, 1993; Garratt et al., 2000) and non-linear models (Anderson and Vahid, 2000). Specifically in the case of non-linear models, the evaluation of probabilities or density forecasts does not add any computational burden over conditional expectations because non-linear models required these to be generated by stochastic simulation (Monte Carlo or Bootstrap, see, e.g., Granger and Teräsvirta, 1993). The analysis of the ability of a model to predict an event can be used for both out-of-sample (Garratt et al., 2000) or in-sample (Fair, 1993; Anderson and Vahid, 2000) evaluation.

In this work, the models estimated in the last sections will be evaluated according to their ability to predict the probability of two events:

A. "At least two consecutive quarters of negative growth in real GDP over the next five quarters";

B. "At least two quarters of negative growth in real GDP over the next five quarters" (Anderson and Vahid, 2000, p. 6).

The first event is a popular simple rule to define recessions without employing NBER turning points. It does not need to coincide with the NBER recessions for three reasons: the rule is based on a single series at a quarterly frequency, no censoring is applied, and the event is forward-looking. Because event A leads NBER recessions, models that predict event A with good performance can be employed to calculate leading recession probabilities indexes for the US economy, such as the experimental leading recession index (XRI) of Stock and Watson (1989). Comparing Figure 4.3, which includes NBER turning points, with Figure 4.6, which presents Event A turning points, the forward-looking characteristic of the rule compared with the NBER turning points is evident. The rule to obtain event A fails to account for the recession in 1960, but the rule employed in event B can capture the 1960 recession. The evident drawback of the rule that calculates event B is that it does not differentiate the two recessions at the beginning of 1980's. Event B happens from 1979:3 to 1982:1 (Figure 4.9). Therefore, these characteristics support the application of both events to evaluate the models.

The procedure to extract the probabilities of event A and B from the models is the same as the one described by Anderson and Vahid (2000). Define $\mathbf{x}_t$ as the vector of endogenous variables $(y_t,\ S_t)'$, $\mathbf{X}^{t-1} = \{\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, ..\mathbf{x}_1\}$ as the history of $\mathbf{x}_t$ and $\mathbf{x}_t = f(\mathbf{X}^{t-1}; \boldsymbol{\beta}) + \boldsymbol{\epsilon}_t$ as the forecasting model where $\boldsymbol{\beta}$ is the matrix of parameters and $\boldsymbol{\epsilon}_t$ are iid with $Var(\boldsymbol{\epsilon}_t) = \Omega$. $\boldsymbol{\epsilon}_t$ is assumed to have a multivariate normal distribution. Given $\hat{\boldsymbol{\beta}}$ and $\hat{\Omega}$, the trial sequence of forecasts for $\{\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{x}_{t+2}, \mathbf{x}_{t+3}, \mathbf{x}_{t+4}\}$ conditional on $\mathbf{X}^{t-1}$ is built as follows. A random vector $\boldsymbol{\epsilon}_t$ is drawn from the distribution $\boldsymbol{\epsilon} \sim N(0, \hat{\Omega})$ and it is used to calculate $\hat{\mathbf{x}}_t$, given $\mathbf{X}^{t-1}$ and $\hat{\boldsymbol{\beta}}$. $\hat{\mathbf{x}}_t$ is added to "history" to form $\hat{\mathbf{X}}^t$. Then a new draw $(\boldsymbol{\epsilon}_{t+1})$ is made from $N(0, \hat{\Omega})$ and it is employed to calculate $\hat{\mathbf{x}}_{t+1}$, given $\hat{\mathbf{X}}^t$ and $\hat{\boldsymbol{\beta}}$, and is used to form $\hat{\mathbf{X}}^{t+1}$. This procedure is continued until the sequence of forecasts is complete $\{\hat{\mathbf{x}}_t, \hat{\mathbf{x}}_{t+1}, \hat{\mathbf{x}}_{t+2}, \hat{\mathbf{x}}_{t+3}, \hat{\mathbf{x}}_{t+4}\}$. This sequence of forecasts can be called S_1, and the same trial is repeated to obtain a set of 2000 forecast sequences. The probability of event A (B) is the proportion of these 2000 sequences in which the event A (B) occurs (P_t). Given the effective sample size T (1954:1-1999:4 for the majority of the models), a series of event probabilities P_t for $t = 1, ..., T$ can be obtained.

In the case of threshold models, the forecasting model can be also written as $\mathbf{x}_t^j = f^j(\mathbf{X}^{t-1}; \boldsymbol{\beta}^j) + \boldsymbol{\epsilon}_t^j$, where $j = 1, 2$ for models with two regimes, $j = 1, 2, 3$ for models with three regimes and $j = 1, 2, 3, 4$ for structural threshold models. Therefore, $Var(\boldsymbol{\epsilon}_t^j)$ depends on the regime (defined by the threshold and the transition variable), so for each regime with different number of observations n^j $(T = \sum_{i=1}^{s} n_i)$, there is a different Ω^j and $\boldsymbol{\epsilon}_t^j$ is supposed to be multivariate normal with variance $\hat{\Omega}^j$. In this framework, for each step to obtain the forecast sequences $(h = 0, ..., 4)$ for, say, a two-regime threshold model, either vector $\boldsymbol{\epsilon}_{t+h}^1$ is drawn from $\boldsymbol{\epsilon}^1 \sim N(0, \hat{\Omega}^1)$ or vector $\boldsymbol{\epsilon}_{t+h}^2$ is drawn from $\boldsymbol{\epsilon}^2 \sim N(0, \hat{\Omega}^2)$ depending on $\hat{s}_{T+h-1-d} < r$ or $\hat{s}_{T+h-1-d} > r$. The vector $\boldsymbol{\epsilon}_{t+h}^j$ is employed to compute $\hat{\mathbf{x}}_{t+h}$ that includes the transition variable that defines the regimes $\hat{s}_{T+h-1-d}$.

The accuracy of the predictions are evaluated using the quadratic probability score (QPS) and the log probability score (LPS) (Diebold and Rudebusch, 1989). The first one ranges from 0 to 2, with 0 being perfect accuracy. The second one ranges from 0 to ∞. LPS and QPS imply different loss functions with large mistakes more heavily penalised under LPS. Let P_t be the prediction probability of the event A or B by the model for the next five periods starting at t and R_t is binary variable that is equal to 1 if the event occurs in the actual data and equal to 0 otherwise, then the Briers score (QPS) and the logarithm score (LPS) are written as:

$$QPS = \frac{1}{T}\sum_{t=1}^{T} 2(P_t - R_t)^2 \text{ and} \tag{4.13}$$

$$LPS = -\frac{1}{T}\sum_{t=1}^{T}[(1 - R_t)\ln(1 - P_t) + R_t \ln(P_t)]. \tag{4.14}$$

Pesaran and Skouras (2001) criticise the application of the QPS to analyse probability prediction accuracy in the context of a two-state-two-action decision problem because the measure cannot be derived from a decision based problem. Instead they suggest the use of the Kuipers score. Suppose for example that one wants to analyse how well a model can predict event A, a future recession. The simulation procedure described generates predicting probabilities of recession (event A). One can define two states as two different indications given by the model: the economy will be in recession or the economy will be in expansion. Suppose that the recession is imminent when the predicted probability is larger than 1/2. So one can calculate event forecasts (E_t): $E_t = 1$ when $P_t > 1/2$ and $E_t = 0$ when $P_t \leq 1/2$. Comparing these events forecasts with the actual outcomes (R_t), the following contingency matrix can be written:

		Actual Outcomes	
		recession ($R_t = 1$)	expansion ($R_t = 0$)
forecasts	recession ($E_t = 1$)	Hits	False Alarms
	expansion ($E_t = 0$)	Misses	Correct rejections

The Kuipers score is defined as the difference between the proportion of recessions that were correctly forecasted ($H = \frac{hits}{(hits+misses)}$) and the proportion of expansions that were incorrectly forecasted ($FA = \frac{false\ alarms}{(false\ alarms+correct\ rejections)}$):

$$KS = H - FA. \tag{4.15}$$

Granger and Pesaran (2000) show how to derive a relationship between the Kuipers score and the market-timing test of Pesaran and Timmermann (1992). The null hypothesis is that the forecast of the recession has no economic value against the alternative hypothesis that has economic value. Given the Kuipers Score (KS) and defining $\hat{\pi}_r$ the estimated probability that the realisations are recessions (event A) and $\hat{\pi}_f$ the estimated probability that the recession is forecasted, the PT (Pesaran and Timmermann) can be written as:

$$PT = \frac{\sqrt{T}(KS)}{\left[\frac{\hat{\pi}_r(1-\hat{\pi}_r)}{\hat{\pi}_f(1-\hat{\pi}_f)}\right]^{1/2}} \tag{4.16}$$

and this statistic is asymptotically N(0,1). In this work, we employ three types of scoring rules: the QPS (eq. 4.13), the LPS (4.14) and the KS (4.15). Previous works on the evaluation of models using event probabilities (Fair, 1993, Camacho and Perez-Quiros, 2000 and Anderson and Vahid, 2000) applied mainly the QPS that has a loss function similar to mean square forecast errors.

4.8 Evaluating Models

The objective of this section is to evaluate how non-linearities and structural breaks can improve the prediction of recessions when the spread is the leading indicator. We evaluate the models presented in sections 4.4, 4.5 and 4.6. Predicting recessions means predicting event A, which is the probability of at least two consecutive periods of negative growth in a five quarters prediction, and, to lesser extent, predicting event B, which is the probability of at least two periods of negative output growth in a five quarters prediction. It is important to observe that using in-sample forecasts, we are evaluating the ability of the spread to lead downturns, using the model as a filter. However, the model is not a filter in the sense employed by Camacho and Perez-Quiros (2000), because the probabilities of the events are obtained by the simulation of forecast sequences and not by the extraction of the values of the transition function or the unobserved component. The Briers score (eq. 4.13), the LPS (eq.

4.14) and the KS (eq. 4.15) are employed to compare the forecast probabilities by the model and the realisations. The test of Pesaran and Timmermann (1992) (PT) was calculated but it is not reported because the null hypothesis is strongly rejected by all the models when the full in-sample period is evaluated, meaning that all the models produce forecasts with economic value.

4.8.1 Description of the Models

The equations of the models evaluated are presented in an Appendix to this chapter. We include a model that predicts constant probability, which is equal to the mean occurrence of events A and B, for the whole sample (eq. 4.17). In addition, the simple rule discussed in section 4.2 is also analysed, however some suppositions have to be made. Because the events are based on the 5-step ahead forecasts and information available until $t-1$, we suppose that the optimal lead for prediction using the spread and the rule is 5 ($\Pr(recesion_t) = 1 - \Phi(S_{t-5})$). The same equation is employed to predict events A and B because it is not possible to discriminate these two events given that the forecasts of the simple rule are not generated by simulation. Therefore, $P_t = 1 - \Phi(S_{t-1})$, where P_t is the probability of event A or B at time t.

We also evaluate the linear model (VAR), four non-linear models (ST, T, 3T and TVAR), one time varying model (only for the output equation) (dST), one model with structural break in the mean and in the variance (SBMV), two models with time-varying non-linearity (TVST and PTVST) and two models with structural break and non-linearity (SBT and SBTVAR). The following general equation will be used to nest the models to be evaluated:

$$
\begin{aligned}
y_t &= x_{t-1}\beta_1[(1-F_1(s_{t-d_1}))(1-F_1(t))] + x_{t-1}\beta_2[F_1(s_{t-d_1})(1-F_1(t))] \\
&\quad + x_{t-1}\beta_3[(1-F_2(s_{t-d_2}))F_1(t)] + x_{t-1}\beta_4[F_2(s_{t-d_2})F_1(t)] + u_{1t} \\
S_t &= x_{t-1}\beta_5[(1-F_3(s_{t-d_3}))(1-F_2(t))] + x_{t-1}\beta_6[F_3(s_{t-d_3})(1-F_2(t))] \\
&\quad + x_{t-1}\beta_7[(1-F_4(s_{t-d_4}))F_2(t)] + x_{t-1}\beta_8[F_4(s_{t-d_4})F_2(t)] + u_{2t}
\end{aligned}
$$

In the smooth transition framework, $F_i(.)$ is a logistic function equal to $G_i(.)$. In the analysis that follows, we suppose that x_{t-1} is the same for all regimes, but in the case of smooth transition models, this is not always true given that the equations are separately estimated ($cov(u_{1t}, u_{2t}) = 0$) and t tests are employed to define the model dynamics. The ST model[15] (4) has a different smooth transition function for each equation, i.e., $F_1(t) = F_2(t) = F_2(s_{t-d_2}) = F_4(s_{t-d_4}) = 0$, $F_1(s_{t-d_1}) = G_1(S_{t-2})$ and $F_3(s_{t-d_3}) = G_2(S_{t-3})$. The dST model (9) has two smooth transitions in each equation using the trend as transition variable in the output equation and S_{t-3} as the transition variable in the spread equation, i.e., $\beta_3 = \beta_6 = 0$, $F_1(s_{t-d_1}) = F_2(s_{t-d_2}) = G_1(t)$, $F_1(t) = G_2(t)$, $G_1(t)G_2(t) \approx G_2(t)$, $F_3(s_{t-d_3}) = F_4(s_{t-d_4}) = G_3(S_{t-3})$, $F_2(t) = G_4(S_{t-3})$ and $G_3(S_{t-3})G_4(S_{t-3}) \approx G_4(S_{t-3})$ (Van Dijk and Franses, 1999, p. 317). The TVST (10) model has two smooth transitions for each equation, with transition variables given by the trend and s_{t-d}, i.e, $F_1(s_{t-d_1}) = F_2(s_{t-d_2}) = G_1(S_{t-2}), F_3(s_{t-d_3}) = F_4(s_{t-d_4}) = G_3(S_{t-2}), F_1(t) = G_2(t)$ and $F_2(t) = G_4(t)$. The PTVST model has the same output equation as the TVST models and the same spread equation as the ST model. The latter model is proposed because the tests of sections 4.5 and 4.6 and the estimated result of TVST model indicate that the time-varying smooth transition model is not a good representation of the spread equation.

The threshold models assume that $F_i(.)$ is a step function equal to $I_i(.)$. The T model (5) has two regimes and one different transition variable, delay and threshold for

[15]The numbers in parentheses designate the models (not equations) described in the Appendix.

each equation, i.e., $F_1(t) = F_2(t) = F_2(s_{t-d_2}) = F_4(s_{t-d_4}) = 0$, $F_1(s_{t-d_1}) = I_1(S_{t-1})$ and $F_3(s_{t-d_3}) = I_2(S_{t-1})$. The 3T model (6) is a threshold model with three regimes, i.e., given that $r_1 < r_2$ and that $r_3 < r_4$, $F_2(s_{t-d_2}) = F_4(s_{t-d_4}) = 1, F_1(s_{t-d_1}) = I_1(S_{t-1})$, $F_1(t) = I_2(S_{t-1})$, $F_3(s_{t-d_3}) = I_3(S_{t-1})$, $F_2(t) = I_4(S_{t-1})$. The TVAR (7) model imposes the restriction of same transition variable, delay and threshold for each equation but allows contemporaneous correlation between the residuals of both equations, i.e., $cov(u_{1t}, u_{2t}) \neq 0$, $F_1(t) = F_2(t) = F_2(s_{t-d_2}) = F_4(s_{t-d_4}) = 0$ and $F_1(s_{t-d_1}) = F_3(s_{t-d_3}) = I_1(S_{t-4})$. The SBMV model (8) has a structural break in the mean and in the variance equations, i.e, $F_1(s_{t-d_1}) = F_2(s_{t-d_2}) = F_3(s_{t-d_3}) = F_4(s_{t-d_4}) = 1$ and $F_1(t) = I_1(t)$ and $F_2(t) = I_2(t)$ and the equation of the variance follows a similar pattern (eqs. 4.A11 and ??). The SBT model (11) has in each equation different breaks and delays, and transition variables and thresholds are estimated for each sub-sample given by the break, i.e., $F_1(s_{t-d_1}) = I_1(S_{t-1}), F_2(s_{t-d_2}) = I_2(S_{t-5}), F_1(t) = I_1(t), F_3(s_{t-d_3}) = I_3(S_{t-4}), F_4(s_{t-d_4}) = I_4(S_{t-1})$ and $F_2(t) = I_2(t)$. The SBTVAR model (11) imposes the restriction that the non-linearity and the break are the same for each equation, allowing contemporaneous correlation between the residuals of both equations, which means that $cov(u_{1t}, u_{2t}) \neq 0$, $F_1(t) = F_2(t) = I(t), F_1(s_{t-d_1}) = F_3(s_{t-d_3}) = I_1(S_{t-3})$ and $F_2(s_{t-d_2}) = F_4(s_{t-d_4}) = I_2(S_{t-5})$.

4.8.2 Results of the Evaluation

The measures of probability forecasting performance are presented in Table 4.5. The ST model, which is very similar to the non-linear autoregressive leading indicator model proposed by Anderson and Vahid (2000), has the smallest QPS in predicting event A. When the ST was compared with a linear model and a smooth transition model with common non-linearity in both equations, Anderson and Vahid (2000) found that ST gives the smallest QPS and also LPS. In Table 4.5, however, the simple rule and the structural break threshold models (SBT and SBTVAR) present better performance using the LPS measure, which gives

Table 4.5: Evaluation of probability forecasting (1954:3-1999:4)

	Event A			Event B		
	QPS	LPS	KS	QPS	LPS	KS
(1) constant	0.245	0.410	0.000	0.367	0.553	0.000
(2) simple rule	0.137	0.234	0.559	0.252	0.393	0.356
(3) VAR	0.178	0.310	0.078	0.268	0.426	0.318
(4) ST	0.135	0.253	0.462	0.248	0.400	0.417
(5) T	0.152	0.274	0.430	0.240	0.386	0.333
(6) 3T	0.156	0.277	0.276	0.243	0.393	0.420
(7) TVAR	0.154	0.281	0.513	0.258	0.412	0.434
(8)SBMV	0.184	0.317	0.352	0.272	0.430	0.504
(9) dST	0.184	0.306	0.311	0.232	0.381	0.516
(10) TVST	0.237	0.390	0.467	0.291	0.452	0.758
(4/10) PTVST	0.221	0.368	0.865	0.248	0.398	0.662
(11) SBT	0.149	0.243	0.419	0.230	0.347	0.530
(12) SBTVAR	0.146	0.245	0.405	0.239	0.361	0.434

Note: The scores are calculated based on the probabilities of event A or B for each time in the period indicated. For each time, the event probabilities are calculated using 5-step-ahead forecasts generated assuming that the coefficients are known and equal to the one estimated for the full sample, but with data available only until t-1. For LPS, see eq. 4.14; for QPS, see eq. 4.13; for KS, see eq. 4.15.

more weight to large deviations, providing some initial evidence that non-linearity does not capture the effects of the structural break. The SBT also has the best performance at predicting event B using QPS and LPS. The inclusion of an additional transition function in the spread equation while using a double time-varying model for output (dST) improves the performance in predicting event B compared to the smooth transition model (ST).

The time-varying smooth transition models have both QPS and LPS for event A larger than the linear model, however when the KS is evaluated, the VAR scores 8% for event A and the partial time-varying smooth transition model (PTVST) scores 87% - the highest KS among the models. For both events, the smooth transition models have better KS than the equivalent threshold models. The model that includes a time-varying transition function only in the output equation (PTVST) fares better than the model with time-varying transition function in both equations(TVST), which was expected, given the bad fit of the time-varying threshold model for the spread equation (eq. 4.A16). Observing Figure 4.7,

Table 4.6: Evaluation of probability forecasting (1983:1-1999:4)

	Event A			Event B		
	QPS	LPS	KS	QPS	LPS	KS
(1) constant	0.145	0.285	0.000	0.191	0.358	0.000
(2) simple rule	0.109	0.178	0.000	0.109	0.178	0.000
(3) VAR	0.117	0.225	0.000	0.135	0.262	0.000
(4) ST	0.125	0.236	0.000	0.130	0.256	0.000
(5) T	0.118	0.222	0.000	0.123	0.242	0.000
(6) 3T	0.109	0.207	0.000	0.121	0.244	0.000
(7) TVAR	0.125	0.236	0.000	0.135	0.265	0.000
(8) SBMV	0.123	0.236	0.000	0.157	0.297	0.034
(9) dST	0.123	0.238	0.000	0.160	0.300	0.016
(10) TVST	0.161	0.311	0.000	0.190	0.354	0.200
(4/10) PTVST	0.163	0.319	0.000	0.163	0.319	0.000
(11) SBT	0.112	0.151	0.494	0.111	0.156	0.494
(12) SBTVAR	0.094	0.144	0.016	0.094	0.152	0.016

Note: See notes of Table 4.5.

TVST fails to predict the 1980 recession most probably because the spread equation has a break in 1980 forcing the model to enter into a regime with bad fit.

Figures 4.6 and 4.7 show that the smooth transition models did not predict the occurrence of event A in 1990. In contrast, the structural break threshold model and the structural break threshold VAR predicted this event[16]. The time-varying smooth transition models have smooth transitions with large enough parameters to be equivalent to a step function, but they suppose that the smooth transition function is the same for both subsamples. In principle, we could relax this supposition by including one more transition function. However, given the sample size and the results of threshold models, it is reasonable to suppose that γ (smoothing parameter) is large, implying that it is hard to get a precise estimated of γ because there is not enough observations in the neighborhood of c (Teräsvirta, 1998). On the other hand, allowing for different delays and thresholds in the threshold model is not hard given that conditional on the break, we can estimate separate models for each

[16]Anderson and Vahid (2000) argue that this recession being atypical, it was hard to predict it. Fintzen and Stekler (1999) demonstrate it was possible to predict the 1990-91 recession with the available data some months before the peak. Using only the information available in the spread, the recession can be considered atypical because only when a break in the beginning of the 80's is taken into account, this recession is predictable. Whether this break could be detected with data available until 1989 is beyond the scope of this work, but an interesting topic for future research.

sub-sample, using grid search to obtain thresholds (eq. 4.12) .

Another important characteristic of the structural break threshold models that may have helped to predict the event A is that they allow for regime-dependent variance. McConnell and Perez-Quiros (2000) show that a univariate Markov-switching model of output growth accounts for the 1990-91 recession only if there is a shift in the variance in 1984. The plots for the threshold models exhibit the predicted probabilities from the simulation procedure with regime-dependent variances and without. The presence of this type of heteroscedasticity improves the scores for all threshold models (not reported) and in the case of the models with structural break improves significantly the probability fit after 1983.

However, employing a SBT may not be done without a cost: the model generates a strong false alarm at the end of 1999. This occurs because the spread during 1998 was smaller than 0.8 that is the threshold for the recession regime of the second sub-sample, which has a large negative constant and large impact of the spread on output growth. This effect is milder in the SBTVAR because this model has coefficients with smaller size in similar regimes although the threshold, 1, is larger. This false alarm is derived from a monetary tightening by the Fed to control inflation at the beginning of 1998 that was followed by a financial crisis generated by the Russian default in August, 1998. As described by Marshall (2001), the crisis was characterised by rapid increases in uncertainty, which implies decreasing values of the spread. Including this information in the SBT model implies that the predictions are based on the recession regime because that is the type of regularity found previously in the historical data and incorporated into the model. However, the Fed policy action of cutting interest rates, switching the policy from controlling inflation to fighting against an imminent financial crisis, was enough to calm the markets after a couple of months and to lead the US economy to the longest period of expansion in history (Marshall, 2001). Thus, the recession predicted for 1999 by model SBT did not materialise, because of the credible Fed action.

Another interesting result is that only the double smooth transition model predicts event A in 1958. This may be connected with the fact that the output equation in this model has two transition functions with trend as transition variables, allowing for two structural breaks, the first one being in 1958. However the model does not perform well in predicting subsequent events, even though it is the only smooth transition model that has event probabilities greater than 0.25 during the 1990 event.

When QPS and LPS are computed only for the 1983-1999 period (Table 4.6), the threshold model with three regimes (3T) gives better performance than the equivalent model with two regimes (T). This difference is not suggested by Table 4.5. The possibility of one more break was rejected by the $\sup LR_{23}$ test presented in Table 4.2. Therefore, the presence of a third regime is more significant in the more recent sample. The long-run multiplier of the spread on output growth is -0.2 in the third regime, which confirms the intervals with negative correlation in the sample 1984-2000 found in analysis of the conditional mean plots in section 4.2.

The evaluation of the scores to predict event B does not show many differences compared with event A, although all the models present larger values of QPS and LPS, meaning that all the models have worse forecast accuracy for this event. Some differences are observed when KS is calculated for the 1983-1999 period. The KS of the PTVST model is significantly different from zero for event B but not for event A (significance evaluated using the PT test, eq. 4.16). Table 4.6 confirms the evaluation of the figures: only the SBT model has KS significantly different from zero in predicting event A.

The flexibility of threshold models compared to smooth transition models is an important advantage when forecast probabilities are evaluated. Allowing for different non-linear behaviours for the periods before and after the break, the structural break threshold models have scored better after 1983. The fact that the power of the spread to predict the events is asymmetric over the cycle does not override the characteristic that the standard deviations of the shocks after the break (around 1981) are half that of the previous period.

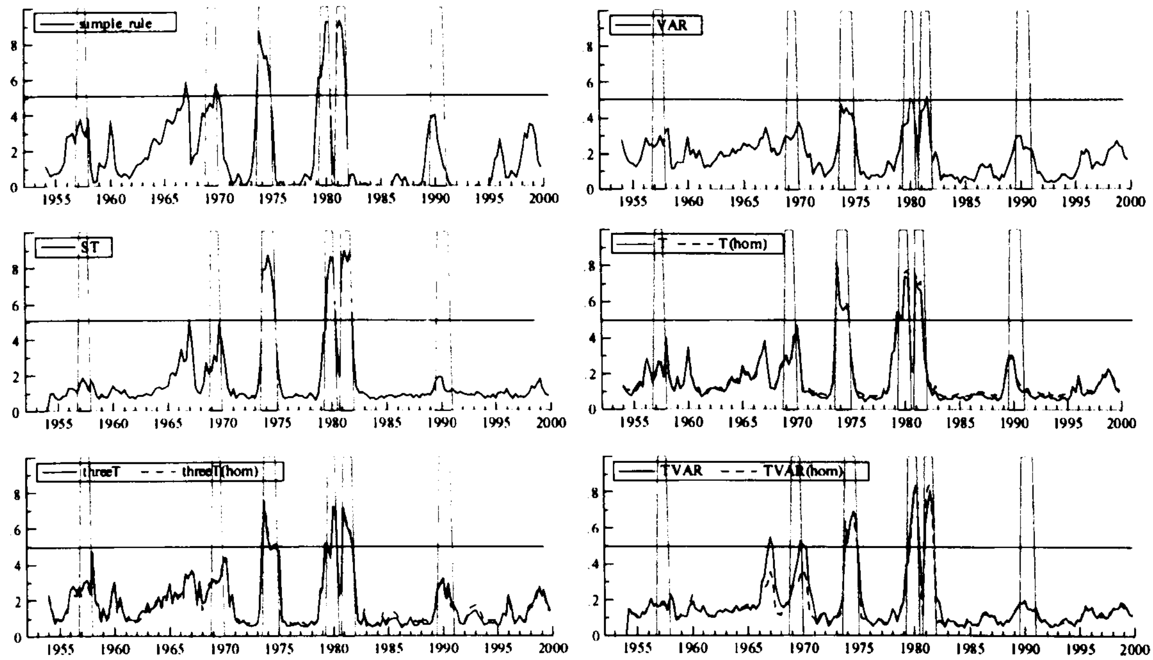

Figure 4.6: Predictions of the probability of event A for 1954:3 to 1999:4 by simple rule, VAR, ST, T, 3T, and TVAR

Therefore, the structural break threshold models are able to account for three characteristics of the data presented in the literature: the output-growth relation with the spread is unstable (Haubrich and Dombrosky, 1996; Dotsey, 1998; Stock and Watson, 2001); it is weaker when the value of the spread is large (Dotsey, 1998; Galbraith and Tkacz, 2000); the variability of output growth after 1984 is half that of the previous period (McConnell and Perez-Quiros, 2000).

4.9 Conclusions

The main contribution of this chapter is to evaluate the effect of non-linearities and structural breaks on the ability of bivariate models of output growth and the spread in predicting event probabilities, specifically the probability of recessions. The results suggest that non-linearities are necessary to predict recessions using the spread as leading indicator. This result was also obtained by Anderson and Vahid (2000), using smooth transition

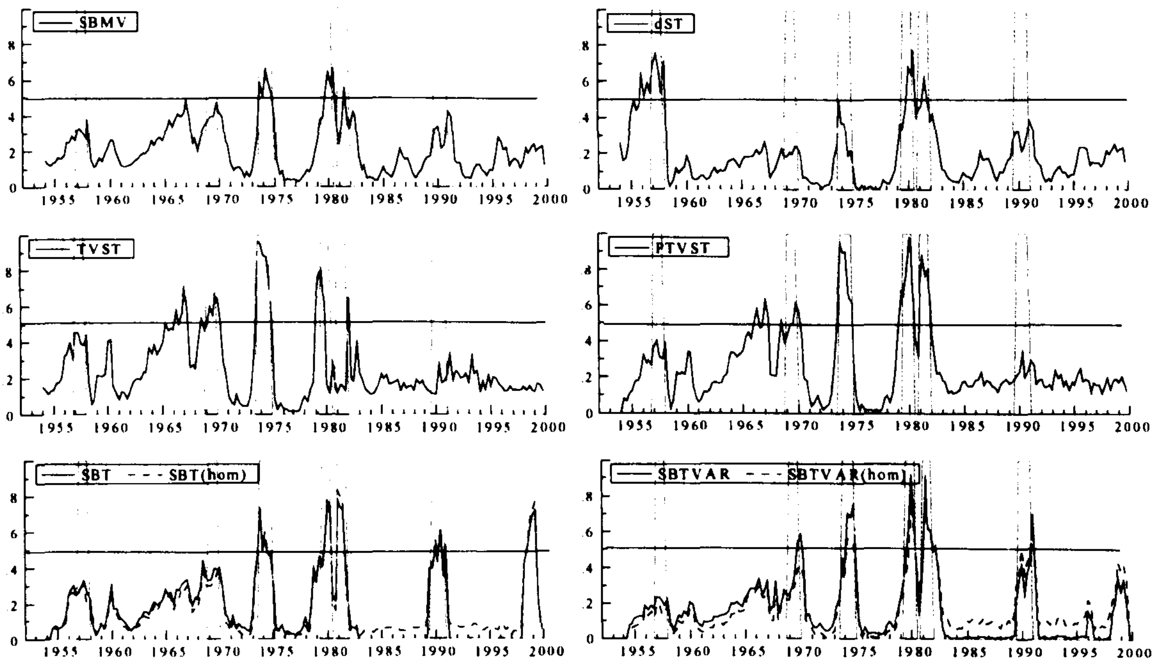

Figure 4.7: Predictions of the probability of event A for 1954:3 to 1999:4 by SBMV, dST, TVST, PTVST, SBT, and SBTVAR

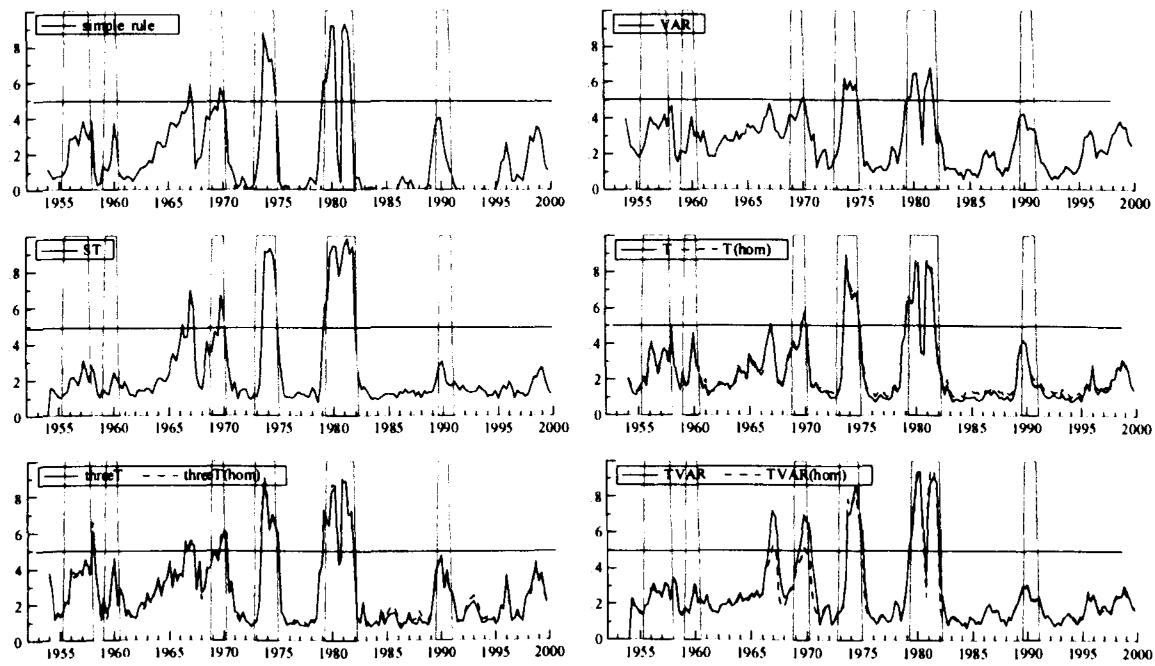

Figure 4.8: Predictions of the probability of event B for 1954:3 to 1999:4 by simple rule, VAR, ST, T, 3T, and TVAR

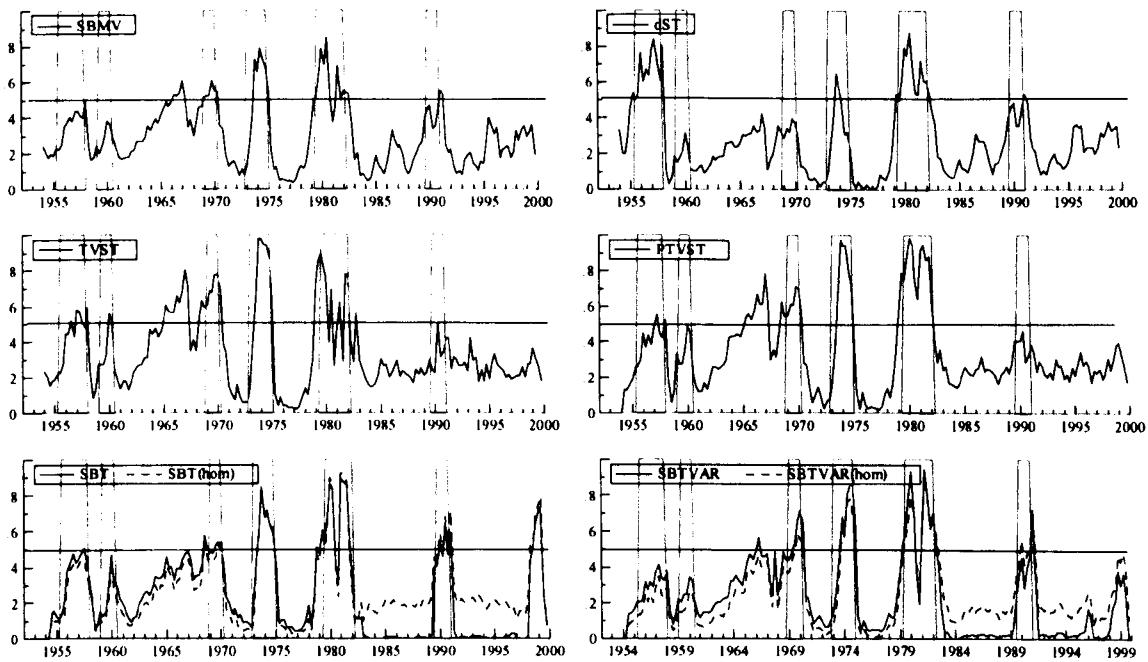

Figure 4.9: Predictions of the probability of event B for 1954:3 to 1999:4 by SBMV, dST, TVST, PTVST, SBT, and SBTVAR

models, while our evaluation also includes threshold models. However, non-linear models are equivalent in terms of probability forecast accuracy to a simple rule based on probit models.

The presence of structural break and non-linearity is needed for the prediction of the 1990-91 recession. Koop and Potter (2000; 2001) suggest that non-linearity tests may indicate threshold models when the data is generated from a structural break model or a time-varying model. The results of the tests of section 4.6 and the evaluation of 4.8 suggest that non-linearity and structural breaks (time-varying parameters) are *both* necessary for a model to be able to capture the dynamic spread-output growth relationship in the period 1954-1999. We show how a structural break threshold model may account for two reported characteristics of the output growth-spread relationship: it is non-linear (Galbraith and Tkacz, 2000; Anderson and Vahid, 2000) and it is unstable over time (Haubrich and Dombrosky, 1996; Stock and Watson, 2001). The structural break threshold model can also account for a further characteristic of the output growth series: its variability has decreased after 1984 (McConnell and Perez-Quiros, 2000). Similar to the results in the Chapter 3, models with three regimes seem to be a better representation of the data for the observations after 1984.

Applying different models to extract information about output growth from the spread, we show that the spread is a reasonable leading indicator. The Kuipers score is around 50% for some models which means (when the model does not generate a false alarm, which is the case for most of the models) that employing the spread as a leading indicator one could predict 50% of the recessions (event A) happening between 1954 and 1999. Observe that these values are based on forecasts with only past information on the endogenous variables but with parameters estimated for the full sample. The implications of assuming models with non-linearities and structural breaks to extract information from the spread to predict a potential recession in 2001 are analysed in Chapter 5.

4.10 Appendix: Description of Bivariate Systems of Output Growth and Spread

These models are tested and discussed in sections 4.4, 4.5 and 4.6 and evaluated in section 4.8. The effective sample employed to estimated the following models is from 1954:1 to 1999:4, except for models ST, T, SBT and SBTVAR that have starting date, respectively, of 1954:2, 1954:2, 1954:3 and 1954:3. y_t is the output growth $(100 * (\ln RGDP_t - \ln RGDP_{t-1}))$ and S_t is the spread between a 10-year T-bond and a 3-month T-Bill. SSE is the residual sum of squares and σ_{u_i} is the standard deviation of u_{it}. PA is the probability of event A and PB is the probability of event B calculated for the observed data (the events are defined in section 4.7).

1. *Constant* (predicts constant probability of events A and B):

$$\begin{aligned} CA_t &= \frac{1}{T}\sum_{i=1}^{T} PA_i \\ CB_t &= \frac{1}{T}\sum_{i=1}^{T} PB_i \end{aligned} \tag{4.A1}$$

2. *Simple rule* (gives the same probability for events A and B):

$$SR_t = 1 - \Phi(S_{t-1}), \tag{4.A2}$$

where $\Phi()$ is the cdf of a standard normal distribution.

3. *Vector Autoregressive model* (VAR):

$$y_t = \underset{[0.13]}{0.305} + \underset{[0.08]}{0.245} y_{t-1} + \underset{[0.08]}{0.081} y_{t-2} - \underset{[0.12]}{0.013} S_{t-1} + \underset{[0.15]}{0.322} S_{t-2} - \underset{[0.13]}{0.108} S_{t-3} \tag{4.A3}$$

$$SSE_1 = 137.23, \sigma_{u_1} = 0.88$$

$$S_t = \underset{[0.09]}{0.280} - \underset{[0.04]}{0.076} y_{t-1} - \underset{[0.06]}{0.103} y_{t-2} + \underset{[0.09]}{1.034} S_{t-1} - \underset{[0.13]}{0.351} S_{t-2} + \underset{[0.13]}{0.215} S_{t-3}$$

$$SSE_2 = 49.65, \sigma_{u_2} = 0.53$$

4 . *Smooth transition model* (ST):

$$\begin{aligned} y_t = (&-\underset{[0.38]}{0.122} - \underset{[0.18]}{0.234} y_{t-1} - \underset{[0.41]}{0.247} S_{t-1} + \underset{[0.41]}{0.726} S_{t-2} \\ &-\underset{[0.17]}{0.363} S_{t-3} + \underset{[0.16]}{0.806} S_{t-4})(1 - G(S_{t-2})) \\ &+(\underset{[0.12]}{0.758} + \underset{[0.09]}{0.260} y_{t-1}(G(S_{t-2})) \end{aligned} \tag{4.A4}$$

$$G(S_{t-2}) = (1/(1 + \exp(-9.88(S_{t-2} - 0.08)/\sigma_{S_{t-2}}))) \quad SSE_1 = 119.145 \quad \sigma_{u_1} = 0.825$$

$$\begin{aligned} S_t = (&\underset{[0.96]}{1.218} - \underset{[0.15]}{0.632} y_{t-1} - \underset{[0.11]}{0.396} y_{t-2} + \underset{[0.10]}{1.580} S_{t-1} \\ &-\underset{[0.23]}{3.056} S_{t-2} + \underset{[0.80]}{2.377} S_{t-3})(1 - G(S_{t-3})) \\ &+(\underset{[0.07]}{0.232} - \underset{[0.03]}{0.082} y_{t-1} - \underset{[0.05]}{0.088} y_{t-2} + \underset{[0.08]}{1.071} S_{t-1} - \underset{[0.08]}{0.147} S_{t-2}) G(S_{t-3}) \end{aligned} \tag{4.A5}$$

$$G(S_{t-3}) = (1/(1 + \exp(-94.37(S_{t-3} + 0.563)/\sigma_{S_{t-3}}))) \quad SSE_2 = 30.79 \quad \sigma_{u_2} = 0.423$$

5. *Threshold model* (T):

$$y_t = (-\underset{[0.18]}{0.045} + \underset{[0.14]}{0.255}y_{t-1} + \underset{[0.15]}{0.409}y_{t-2} - \underset{[0.27]}{0.187}S_{t-1} \tag{4.A6}$$

$$+\underset{[0.31]}{1.001}S_{t-2} - \underset{[0.26]}{0.523}S_{t-3})(I_1(S_{t-1}))$$

$$+(\underset{[0.19]}{0.851} + \underset{[0.08]}{0.278}y_{t-1} - \underset{[0.07]}{0.033}y_{t-2} - \underset{[0.15]}{0.121}S_{t-1} +$$

$$\underset{[0.19]}{0.024}S_{t-2} - \underset{[0.13]}{0.085}S_{t-3})(1 - I_1(S_{t-1}))$$

$$I_1(S_{t-1}) = 1(S_{t-1} \leq 0.667)$$

$$\sigma_{u_1} = 0.916I_1(S_{t-1}) + 0.807(1 - I_1(S_{t-1})) \quad SSE_1 = \; 121.98$$

$$S_t = (-\underset{[0.07]}{0.411} - \underset{[0.04]}{0.025}y_{t-1} - \underset{[0.05]}{0.211}y_{t-2} + \underset{[0.09]}{1.233}S_{t-1} \tag{4.A7}$$

$$-\underset{[0.11]}{0.494}S_{t-2} + \underset{[0.08]}{0.103}\, S_{t-3})(I_1(S_{t-1}))$$

$$+(-\underset{[0.42]}{0.522} + \underset{[0.09]}{0.022}y_{t-1} + \underset{[0.08]}{0.048}y_{t-2} + \underset{[0.17]}{0.961}S_{t-1}$$

$$-\underset{[0.17]}{0.072}S_{t-2} + \underset{[0.12]}{0.233}S_{t-3})(1 - I_1(S_{t-1}))$$

$$I_1(S_{t-1}) = 1(S_{t-1} \leq 1.79)$$

$$\sigma_{u_2} = 0.445I_1(S_{t-1}) + 0.360(1 - I_1(S_{t-1})) \quad \text{SSE} = 39.63$$

6. *Threshold Vector Autoregressive model* (TVAR):

$$y_t = (\underset{[0.23]}{-0.016} - \underset{[0.25]}{0.006}y_{t-1} + \underset{[0.26]}{0.295}y_{t-2} + \underset{[0.24]}{0.032}S_{t-1} \tag{4.A8}$$

$$+\underset{[0.34]}{0.829}S_{t-2} - \underset{[0.33]}{0.004}S_{t-3})I_1(S_{t-4})$$

$$+(\underset{[0.14]}{0.515} - \underset{[0.07]}{0.253}y_{t-1} + \underset{[0.07]}{0.040}y_{t-2} + \underset{[0.16]}{0.076}S_{t-1}$$

$$+\underset{[0.21]}{0.127}S_{t-2} - \underset{[0.14]}{0.095}S_{t-3})(1 - I_1(S_{t-4}))$$

$$\sigma_{u_1} = 1.313I_1(S_{t-4}) + 0.841(1 - I_1(S_{t-4})) \qquad \mathrm{SSE}_1 = 125.54$$

$$S_t = (\underset{[0.19]}{0.798} - \underset{[0.20]}{0.149}y_{t-1} - \underset{[0.21]}{0.424}y_{t-2} + \underset{[0.20]}{0.882}S_{t-1}$$

$$-\underset{[0.28]}{1.027}S_{t-2} + \underset{[0.27]}{1.021}S_{t-3})I_1(S_{t-4})$$

$$+(\underset{[0.07]}{0.162} - \underset{[0.04]}{0.081}y_{t-1} - \underset{[0.04]}{0.049}y_{t-2} + \underset{[0.07]}{1.098}S_{t-1}$$

$$-\underset{[0.07]}{0.186}S_{t-2} + \underset{[0.06]}{0.023}S_{t-3})(1 - I_1(S_{t-4}))$$

$$I_1(S_{t-4}) = 1(S_{t-4} \leq 0.1933)$$

$$\sigma_{u_2} = 1.076I_1(S_{t-4}) + 0.391(1 - I_1(S_{t-4})) \qquad \mathrm{SSE}_2 = 37.33$$

$$\sigma_{u_1}\sigma_{u_2} = -0.637(S_{t-4}) + 0.024(1 - I_1(S_{t-4}))$$

7. *Three-regime Threshold model (3T):*

$$y_t = (\underset{[0.19]}{0.171} - \underset{[0.15]}{0.188}y_{t-1} + \underset{[0.16]}{0.449}y_{t-2} - \underset{[0.24]}{0.216}S_{t-1} \tag{4.A9}$$

$$+\underset{[0.22]}{1.104}S_{t-2} - \underset{[0.22]}{0.277}S_{t-3})I_1(S_{t-1})$$

$$+(\underset{[0.27]}{-0.030} + \underset{[0.12]}{0.407}y_{t-1} + \underset{[0.11]}{0.110}y_{t-2} + \underset{[0.35]}{0.658}S_{t-1}$$

$$-\underset{[0.45]}{0.837}S_{t-2} + \underset{[0.30]}{0.583}S_{t-3})I_2(S_{t-1})$$

$$+(\underset{[0.34]}{1.484} + \underset{[0.10]}{0.219}y_{t-1} - \underset{[0.10]}{0.023}y_{t-2} - \underset{[0.18]}{0.336}S_{t-1}$$

$$+\underset{[0.20]}{0.125}S_{t-2} - \underset{[0.14]}{0.023}S_{t-3})(1 - I_1(S_{t-1}) - I_2(S_{t-1}))$$

$$\sigma_{u_1} = 0.649 I_1(S_{t-1}) + 0.902 I_2(S_{t-1}) + 0.782(1 - I_1(S_{t-1}) - I_2(S_{t-1}) \quad SSE_1 = 113.50$$

$$I_1(S_{t-1}) = 1(S_{t-1} \leq 0.183) \quad I_2 = 1(0.183 < S_{t-1} \leq 1.473)$$

$$\begin{aligned} S_t = (&\underset{[0.19]}{0.167} - \underset{[0.15]}{0.196} y_{t-1} - \underset{[0.14]}{0.143} y_{t-2} + \underset{[0.23]}{0.938} S_{t-1} \\ &- \underset{[0.22]}{0.493} S_{t-2} - \underset{[0.22]}{0.328} S_{t-3}) I_1(S_{t-1}) \\ &+ (\underset{[0.09]}{0.299} - \underset{[0.04]}{0.059} y_{t-1} - \underset{[0.04]}{0.181} y_{t-2} + \underset{[0.11]}{1.281} S_{t-1} \\ &- \underset{[0.12]}{0.503} S_{t-2} + \underset{[0.80]}{0.175} S_{t-3}) I_2(S_{t-1}) \\ &+ (-\underset{[0.42]}{0.522} + \underset{[0.09]}{0.022}\, y_{t-1} + \underset{[0.08]}{0.050} y_{t-2} + \underset{[0.17]}{0.961} S_{t-1} \\ &- \underset{[0.17]}{0.073} S_{t-2} + \underset{[0.12]}{0.233} S_{t-3})(1 - I_1(S_{t-1}) - I_2(S_{t-1})) \end{aligned} \tag{4.A10}$$

$$\sigma_{u_2} = 0.645 I_1(S_{t-1}) + 0.364 I_2(S_{t-1}) + 0.554(1 - I_1(S_{t-1}) - I_2(S_{t-1}) \quad SSE_2 = 36.56$$

$$I_1(S_{t-1}) = 1(S_{t-1} \leq 0.203) \quad I_2 = 1(0.203 < S_{t-1} \leq 1.79)$$

8. *Structural Break in the Mean and the Variance:*

$$\begin{aligned} y_t = (&\underset{[0.16]}{0.186} + \underset{[0.11]}{0.197} y_{t-1} + \underset{[0.09]}{0.043} y_{t-2} + \underset{[0.20]}{0.497} S_{t-1} \\ &- \underset{[0.34]}{0.206} S_{t-2} + \underset{[0.25]}{0.211} S_{t-3}) I(t) \\ &+ (\underset{[0.18]}{0.192} + \underset{[0.12]}{0.290} y_{t-1} + \underset{[0.12]}{0.247} y_{t-2} - \underset{[0.10]}{0.305} S_{t-1} \\ &+ \underset{[0.14]}{0.593} S_{t-2} - \underset{[0.14]}{0.189} S_{t-3})(1 - I(t)) \end{aligned} \tag{4.A11}$$

$$I(t) = 1(t \leq 1980:4) \quad \sigma_{u_1} = \underset{[0.13]}{0.950} I(t) + \underset{[0.05]}{0.574}(1 - I(t)) \quad SSE_1 = 122.42$$

$$\begin{aligned} S_t \quad = \quad (&\underset{[0.10]}{0.195} - \underset{[0.04]}{0.096} y_{t-1} - \underset{[0.04]}{0.075} y_{t-2} + \underset{[0.09]}{1.097} S_{t-1} \\ &- \underset{[0.24]}{0.622} S_{t-2} + \underset{[0.24]}{0.449} S_{t-3}) I(t) \\ &+ (\underset{[0.17]}{0.553} - \underset{[0.12]}{0.179} y_{t-1} - \underset{[0.15]}{0.314} y_{t-2} + \underset{[0.94]}{0.934} S_{t-1} \\ &- \underset{[0.17]}{0.075} S_{t-2} - \underset{[0.12]}{0.078} S_{t-3})(1 - I(t)) \end{aligned} \tag{4.A12}$$

$$I(t) = 1(t \leq 1981:1) \qquad \sigma_{u_2} = \underset{[0.06]}{0.473} I(t) + \underset{[0.04]}{0.501}(1 - I(t)) \quad SSE_2 = \; 43.16$$

9. *Double Smooth Transition model* (dST):

$$y_t = (\underset{[0.60]}{-1.136} + \underset{[0.20]}{0.686} y_{t-1} - \underset{[0.63]}{0.139} y_{t-2} \tag{4.A13}$$

$$+\underset{[0.97]}{2.741} S_{t-1} - \underset{[0.74]}{2.999} S_{t-2} + \underset{[0.97]}{1.843} S_{t-3})$$

$$+(\underset{[0.55]}{1.487} - \underset{[0.25]}{0.616} y_{t-1} + \underset{[0.23]}{0.217} y_{t-2}$$

$$-\underset{[0.65]}{2.519} S_{t-1} + \underset{[1]}{3.269} S_{t-2} - \underset{[0.75]}{1.916} S_{t-3}) G_1(t)$$

$$+(\underset{[0.26]}{-0.160} + \underset{[0.17]}{0.180} y_{t-1} + \underset{[0.16]}{0.156} y_{t-2}$$

$$-\underset{[0.21]}{0.524} S_{t-1} + \underset{[0.25]}{0.226} S_{t-2}) G_2(t)$$

$$G_1(t) = (1/(1 + \exp(-500(t - 22.07)/\sigma_{(t)})))$$

$$G_2(t) = (1/(1 + \exp(-500(t - 108.5)/\sigma_{(t)})))$$

$$SSE_1 = \; 108.51 \quad \sigma_{u_1} = 0.806$$

$$S_t = (\underset{[1.04]}{1.074} + \underset{[0.13]}{0.654} y_{t-1} - \underset{[0.14]}{0.376} y_{t-2} \tag{4.A14}$$

$$+\underset{[0.11]}{1.573} S_{t-1} - \underset{[0.25]}{3.046} S_{t-2} + \underset{[0.92]}{2.248} S_{t-3})$$

$$+(\underset{[1.04]}{-0.763} - \underset{[0.13]}{0.732} y_{t-1} - \underset{[0.14]}{0.304} y_{t-2}$$

$$-\underset{[0.14]}{0.572} S_{t-1} + \underset{[0.27]}{2.910} S_{t-2} - \underset{[0.92]}{2.380} S_{t-3}) G_1(S_{t-3})$$

$$+(\underset{[0.33]}{0.659} - \underset{[0.10]}{0.067} y_{t-1} - \underset{[0.16]}{0.110} y_{t-2}$$

$$+\underset{[0.16]}{0.280} S_{t-1} - \underset{[0.17]}{0.264} S_{t-2}) G_2(S_{t-3})$$

$$G_1(S_{t-3}) = (1/(1 + \exp(-22.5(S_{t-3} + 0.550)/\sigma_{S_{t-3}})))$$

$$G_2(S_{t-3}) = (1/(1 + \exp(-500(S_{t-3} - 1.651)/\sigma_{S_{t-3}})))$$

$$SSE_2 = \; 26.997 \quad \sigma_{u_2} = 0.402$$

10. *Time-Varying Smooth Transition Model* (TVST):

$$\begin{aligned} y_t = (&\underset{[0.17]}{0.243}y_{t-1} + \underset{[0.27]}{0.437}S_{t-1} + \underset{[0.40]}{0.555}S_{t-2} \\ &+\underset{[0.37]}{0.122}S_{t-3})(1 - G_1(S_{t-1}))(1 - G_2(t)) \\ &+(\underset{[0.12]}{0.246}y_{t-1} + \underset{[0.32]}{0.993}S_{t-1} - \underset{[0.45]}{0.938}S_{t-2} \\ &+\underset{[0.26]}{0.559}S_{t-3})(G_1(S_{t-1}))(1 - G_2(t)) \\ &+(\underset{[0.30]}{1.205} - \underset{[0.18]}{0.689}y_{t-1} - \underset{[0.19]}{1.339}S_{t-1} \\ &+\underset{[0.30]}{1.694}S_{t-2})(1 - G_1(S_{t-1}))G_2(t) \\ &+(\underset{[0.20]}{0.475} + \underset{[0.10]}{0.486}y_{t-1} - \underset{[0.09]}{0.167}S_{t-1} + \underset{[0.11]}{0.349}S_{t-2} \\ &-\underset{[0.12]}{0.182}S_{t-3})G_1(S_{t-1})G_2(t) \end{aligned} \tag{4.A15}$$

$$G_1(S_{t-1}) = (1/(1 + \exp(-500(S_{t-2} - 0.42)/\sigma_{S_{t-2}})))$$

$$G_2(t) = (1/(1 + \exp(-500(t - 108.6)/\sigma_{(t)})))$$

$$SSE_1 = 113.89 \quad \sigma_{u_1} = 0.826$$

$$\begin{aligned} S_t = (&\underset{[0.07]}{0.130} - \underset{[0.03]}{0.136}y_{t-1} + \underset{[0.06]}{0.987}S_{t-1})(1 - G_1(S_{t-1}))(1 - G_2(t)) \\ &+(\underset{[0.52]}{0.673} - \underset{[0.06]}{0.247}y_{t-1} + \underset{[0.21]}{0.789}S_{t-1})(G_1(S_{t-1}))(1 - G_2(t)) \\ &+(\underset{[0.16]}{0.669} - \underset{[0.07]}{0.248}y_{t-1} + \underset{[0.17]}{0.693}S_{t-1})(1 - G_1(S_{t-1}))G_2(t) \\ &+(-\underset{[0.20]}{0.405} + \underset{[0.15]}{0.249}y_{t-1} + \underset{[0.16]}{1.003}S_{t-1})G_1(S_{t-1})G_2(t) \end{aligned} \tag{4.A16}$$

$$G_1(S_{t-1}) = (1/(1 + \exp(-500(S_{t-1} - 1.613)/\sigma_{S_{t-1}})))$$

$$G_2(t) = (1/(1 + \exp(-500(t - 105.4)/\sigma_{(t)})))$$

$$SSE_2 = 46.79 \quad \sigma_{u_2} = 0.522$$

11. *Structural Break Threshold model* (SBT):

$$y_t = (\underset{[0.15]}{0.088} + \underset{[0.11]}{0.148}y_{t-1} + \underset{[0.11]}{0.218}y_{t-2} + \underset{[0.21]}{0.113}S_{t-1} \tag{4.A17}$$

$$+\underset{[0.26]}{0.452}S_{t-2} - \underset{[0.21]}{0.076}S_{t-3})(1 - I_1(S_{t-1}))I(t)$$

$$+(\underset{[0.68]}{2.266} + \underset{[0.14]}{0.057}y_{t-1} - \underset{[0.13]}{0.332}y_{t-2} - \underset{[0.30]}{0.662}S_{t-1}$$

$$-\underset{[0.29]}{0.003}S_{t-2} + \underset{[0.21]}{0.507}S_{t-3})I_1(S_{t-1})I(t)$$

$$+(-\underset{[0.90]}{2.530} + \underset{[0.24]}{1.187}y_{t-1} + \underset{[0.43]}{0.565}y_{t-2} + \underset{[1.03]}{2.684}S_{t-1}$$

$$-\underset{[1.21]}{1.307}S_{t-2} + \underset{[0.86]}{0.481}S_{t-3})(1 - I_2(S_{t-5}))(1 - I(t))$$

$$+(\underset{[0.21]}{0.918} - \underset{[0.14]}{0.135}y_{t-1} + \underset{[0.13]}{0.225}y_{t-2} + \underset{[0.15]}{0.036}S_{t-1}$$

$$+\underset{[0.23]}{0.047}S_{t-2} - \underset{[0.14]}{0.137}S_{t-3})I_2(S_{t-5})(1 - I(t))$$

$$\sigma_{u_1} = 0.98(1 - I_1(S_{t-1}))I(t) + 0.84I_1(S_{t-1})I(t)$$

$$+0.61(1 - I_2(S_{t-5}))(1 - I(t)) + 0.40I_2(S_{t-5})(1 - I(t))$$

$$I_1(S_{t-1}) = 1(S_{t-1} > 1.58) \quad I_2(S_{t-5}) = 1(S_{t-5} > 0.817) \quad I(t) = 1(t \le 1983:4)$$

$$SSE_1 = 107.67$$

$$S_t = (\underset{[0.29]}{0.527} - \underset{[0.16]}{0.333}y_{t-1} + \underset{[0.28]}{0.219}y_{t-2} + \underset{[0.34]}{0.335}S_{t-1} \tag{4.A18}$$

$$+\underset{[0.50]}{0.047}S_{t-2} - \underset{[0.46]}{0.156}S_{t-3})(1 - I_1(S_{t-4}))I(t)$$

$$+(\underset{[0.07]}{0.057} + \underset{[0.03]}{0.002}y_{t-1} - \underset{[0.03]}{0.083}y_{t-2} + \underset{[0.11]}{1.260}S_{t-1}$$

$$-\underset{[0.15]}{0.819}S_{t-2} + \underset{[0.11]}{0.516}S_{t-3})I_1(S_{t-4})I(t)$$

$$+(\underset{[0.09]}{0.556} - \underset{[0.09]}{0.076}y_{t-1} - \underset{[0.08]}{0.379}y_{t-2} + \underset{[0.11]}{1.254}S_{t-1}$$

$$-\underset{[0.14]}{0.506}S_{t-2} + \underset{[0.11]}{0.142}S_{t-3})(1 - I_2(S_{t-1}))(1 - I(t))$$

$$+(-\underset{[0.44]}{0.599} + \underset{[0.09]}{0.033}y_{t-1} + \underset{[0.09]}{0.091}y_{t-2} + \underset{[0.17]}{0.968}S_{t-1}$$

$$-\underset{[0.17]}{0.062}S_{t-2} + \underset{[0.12]}{0.222}S_{t-3})I_2(S_{t-1})(1 - I(t))$$

$$\sigma_{u_2} = 0.62(1 - I_1(S_{t-4}))I(t) + 0.19 I_1(S_{t-4})I(t)$$

$$+0.52(1 - I_2(S_{t-1}))(1 - I(t)) + 0.56 I_2(S_{t-1})(1 - I(t))$$

$$I_1(S_{t-4}) = 1(S_{t-4} > 0.313) \quad I_2(S_{t-1}) = 1(S_{t-1} > 1.79) \quad I(t) = 1(t \leq 1968:3)$$

$$SSE_2 = 34.40$$

12. *Structural Break Threshold Vector Autogressive model* (SBTVAR):

$$y_t = (\underset{[0.18]}{-0.191} - \underset{[0.19]}{0.121} y_{t-1} + \underset{[0.19]}{0.841} y_{t-2} + \underset{[0.23]}{0.123} S_{t-1} \tag{4.A19}$$

$$+\underset{[0.37]}{1.567} S_{t-2} - \underset{[0.34]}{1.216} S_{t-3})(1 - I_1(S_{t-3}))(1 - I(t))$$

$$+(\underset{[0.22]}{0.501} + \underset{[0.11]}{0.144} y_{t-1} - \underset{[0.11]}{0.037} y_{t-2} + \underset{[0.22]}{0.180} S_{t-1}$$

$$+\underset{[0.31]}{0.159} S_{t-2} + \underset{[0.24]}{0.044} S_{t-3}) I_1(S_{t-3})(1 - I(t))$$

$$(\underset{[0.38]}{-0.641} + \underset{[0.25]}{0.791} y_{t-1} + \underset{[0.20]}{0.211} y_{t-2} - \underset{[0.22]}{0.092} S_{t-1}$$

$$+\underset{[0.28]}{1.063} S_{t-2} - \underset{[0.28]}{0.386} S_{t-3})(1 - I_2(S_{t-5}))I(t)$$

$$+(\underset{[0.21]}{0.678} + \underset{[0.14]}{0.204} y_{t-1} + \underset{[0.13]}{0.146} y_{t-2} - \underset{[0.16]}{0.054} S_{t-1}$$

$$+\underset{[0.20]}{0.158} S_{t-2} - \underset{[0.13]}{0.119} S_{t-3}) I_2(S_{t-5}) I(t)$$

$$\sigma_{u_1} = 0.61(1 - I_1(S_{t-3}))I(t) + 1.02 I_1(S_{t-3})I(t)$$

$$+0.63(1 - I_2(S_{t-5}))(1 - I(t)) + 0.50 I_2(S_{t-5})(1 - I(t))$$

$$S_t = (\underset{[0.25]}{0.416} + \underset{[0.26]}{0.340} y_{t-1} - \underset{[0.26]}{0.361} y_{t-2} + \underset{[0.31]}{1.604} S_{t-1}$$

$$\underset{[0.50]}{-2.436} S_{t-2} + \underset{[0.47]}{1.470} S_{t-3})(1 - I_1(S_{t-3}))(1 - I(t))$$

$$+(\underset{[0.07]}{0.126} - \underset{[0.04]}{0.097} y_{t-1} - \underset{[0.04]}{0.070} y_{t-2} + \underset{[0.07]}{1.071} S_{t-1}$$

$$\underset{[0.10]}{-0.248} S_{t-2} + \underset{[0.08]}{0.147} S_{t-3}) I_1(S_{t-3})(1 - I(t))$$

$$(\underset{[0.18]}{1.216} - \underset{[0.12]}{0.252} y_{t-1} - \underset{[0.09]}{0.377} y_{t-2} + \underset{[0.11]}{0.386} S_{t-1}$$

$$\underset{[0.14]}{-0.405} S_{t-2} - \underset{[0.14]}{0.676} S_{t-3})(1 - I_2(S_{t-5})) I_t$$

$$+(\underset{[0.18]}{0.295} - \underset{[0.12]}{0.058}y_{t-1} - \underset{[0.10]}{0.014}y_{t-2} + \underset{[0.13]}{1.224}S_{t-1}$$

$$-\underset{[0.17]}{0.202}S_{t-2} - \underset{[0.11]}{0.143}S_{t-3})I_2(S_{t-5})I(t)$$

$$\sigma_{u_2} = 0.83(1 - I_1(S_{t-3}))I(t) + 0.33I_1(S_{t-3})I(t)$$

$$+0.31(1 - I_2(S_{t-5}))(1 - I(t)) + 0.42I_2(S_{t-5})(1 - I(t))$$

$$I_1(S_{t-3}) = 1(S_{t-3} > 0.1566) \qquad I_2(S_{t-5}) = 1(S_{t-5} > 0.98) \quad I(t) = 1(t < 1981:1)$$

$$\sigma_{u_1}\sigma_{u_2} = -0.24(1 - I_1(S_{t-3}))I(t) - 0.02I_1(S_{t-3})I(t)$$

$$-0.04(1 - I_2(S_{t-5}))(1 - I(t)) + 0.07I_2(S_{t-5})(1 - I(t))$$

$$SSE_1 = 110.31 \quad SSE_2 = 28.21$$

Chapter 5

The Economics of Non-linearities in Empirical Models

5.1 Introduction

Macroeconomists can employ the results of the econometric evaluation of previous chapters to suggest 'stylised facts' that a good theoretical model should take into account. The success of Markov-switching models in characterising business cycle turning points in many macroeconomic series, for example, is the motivation for business cycle theorists to develop models that allow endogenous switches between equilibria (Azariadis and Smith, 1998; Eudey and Perli, 1999). In this chapter, we use the 'stylised facts' generated by the non-linear time series model applied to the term structure of interest rates to understand whether the expectations theory holds. Furthermore, the best ranked models can be employed to provide models for generating forecasts used in taking policy decisions. Policy makers usually employ various types of models to generate forecasts, although most of the short-horizon forecasts that determine decisions of central bankers are judgemental (Pagan and Robertson, 2001). Their main concern is on forecasts of output and inflation, which support monetary policy decisions. In this context, improved forecasts of output, and also of the probability of recession, are useful information for policy making. Therefore, in this chapter an analysis of the previous results is given according to their economic implications for development and testing of economic theory and for generating better forecasts for policy makers.

The evaluation of the previous chapters indicates non-linearities in the shape of the US business cycle, in the ability of the spread to predict short- and long-term interest rates and in the relationship between output-growth and the spread. Section 5.2 analyses how the concepts of business cycle asymmetries presented in the literature are associated with the shape of the cycle employed in Chapter 2 to observe which types of asymmetries are generated by non-linear time series models and the reason for these. In addition, we describe business cycle theories that support the type of asymmetries found in the US business cycle. Section 5.3 analyses the implications of the dynamics of the best forecaster from Chapter 3 for the theory of the term structure of interest rates. It also shows how non-linearity is important for forecasting the spread just before a turning point. Section 5.4 evaluates how the inclusion of non-linearity and structural breaks in a model to predict recessions, using the spread, affects the forecast of output growth and the probability of a recession for 2001, compared with two other popular recession forecasters.

5.2 Asymmetries and the US Business Cycle

Non-linear univariate models can generate the asymmetric shape of the US cycle only if they account for a period of high-growth recovery after the trough, followed by a period of moderate growth. This conclusion was the main contribution of the discussion of Chapter 2. Conditional mean functions and surfaces are computed to verify the ability of the dynamics of non-linear models to reproduce the asymmetric shape of the cycle. In this section, the relationship between the measure of shape employed and other types of business cycle asymmetries that have been tested in the literature, such as steepness, deepness and sharpness is analysed. Given the average excess as a measure of business cycle shape, we evaluate which type of asymmetries the non-linear time series models of Chapter 2 generate. Finally, the results of the evaluation are employed to indicate which type of business cycle theories account for the asymmetric shape of the business cycle.

5.2.1 Business Cycle Asymmetries

As discussed in section 2.3, the business cycle analysis in this work is built on the classical definition of business cycles, meaning that turning points are located in the log-level of real GDP. Econometric models are generally estimated with detrended data, assuming a stochastic trend. Since the information on the trend can be recovered, the classical business cycle definition can be employed to evaluate econometric models. Moreover, the turning points extracted with Markov-switching models are usually compared with NBER turning points, not with the growth cycle (see, e.g., Hamilton (1989) and Durland and McCurdy (1994)). Another definition of the business cycle employs a filtering procedure to extract the cycle from a time series. In this work, because we employ the classical definition, the business cycle is generated by an interaction between trend and fluctuations.

Turning points located using the QBB algorithm described in section 2.3 are employed to calculate the gains (losses) in terms of the increase (decrease) in ln(GDP) at quarter t compared with the last trough (peak), which are presented in Figure 5.1. For example, at 1965:1, this value is 25% per cent, meaning that the economy has grown 25% since the expansion phase started. If the growth rate of the expansion (contraction) phase is constant, the area under the gain (loss) curve is equal to the area of a right-angled triangle with height given by the maximum amplitude in the phase and with base given by the duration of the phase. However, this does not seem to be so in the case of the US business cycle, as indicated by Figure 5.1. Thus, a good measure of the shape of the cycle inside each phase is to

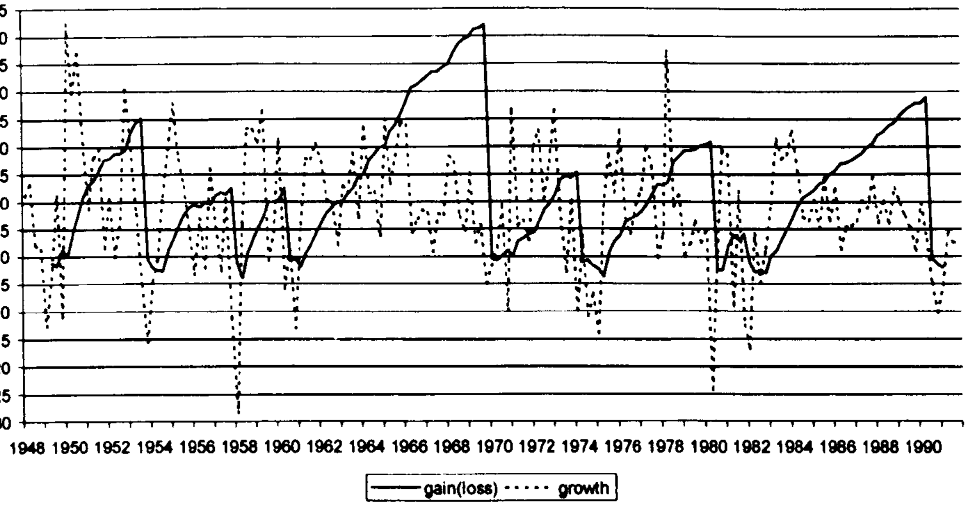

Figure 5.1: Gain (loss) against last trough (peak) and output growth for US GDP

calculate the difference between the sum of the gains (losses) for each t over the expansion (contraction) phase and the area of the triangle created by the duration and the amplitude. This is the excess measure defined by Harding and Pagan (2001b), and employed as a stylised fact in Chapter 2.

A stylised illustration of the shape of the US business cycle employing the mean of the excess values over cycles to define the shape is described in Figure 5.2[1]. The turning points together with the two parallel lines with constant growth rate (floor and the ceiling) define the amplitudes and the durations of the business cycles. Given these features, the straight line represents the triangle representation of the shape of the cycle. The curve represents the loss (gains) against last peak (though), which defines the shape of the cycle. Because the excess is highly positive during expansions, the shape is concave. This means that higher growth rates are observed just after the trough and moderate growth rates (following the growth rate of the ceiling) occur during the largest part of the expansions. The contractions have an excess equal to zero or slightly negative, which means that the shape of the cycle in

[1]Figure 5.2 is an illustration: the duration and the amplitudes are not in perfect scale with the average duration and amplitude of the US cycle. However, the main characteristic is present: contractions are shorter than expansions and have smaller amplitude.

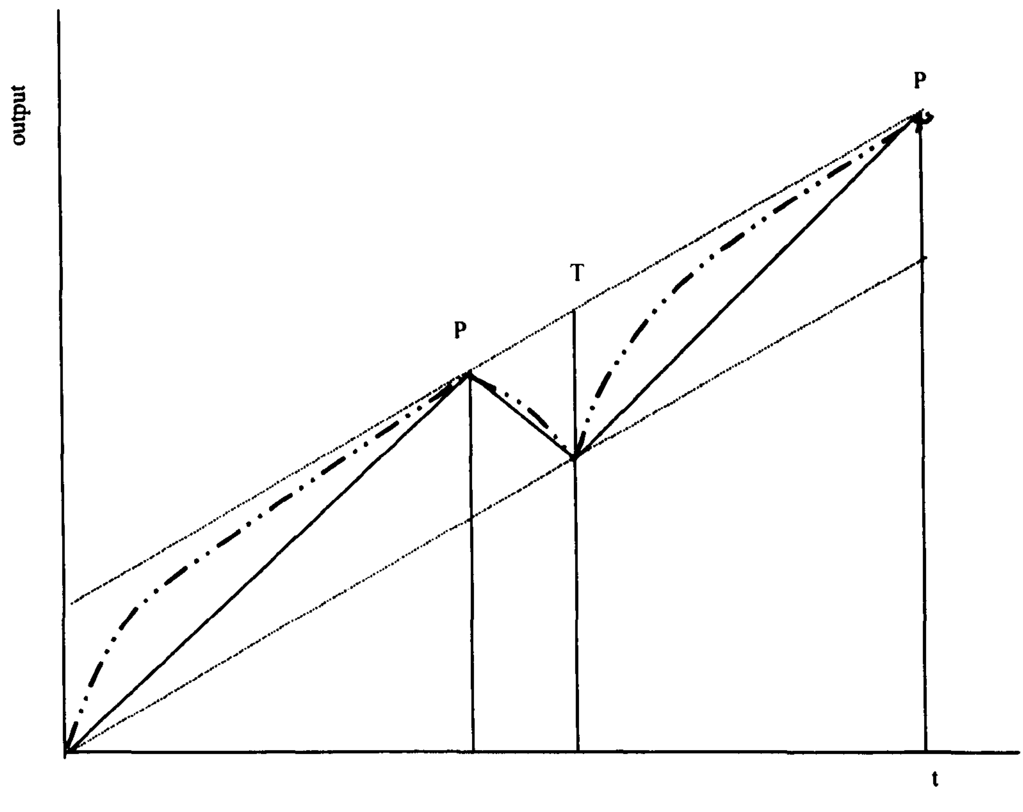

Figure 5.2: The shape of the US business cycle

this phase is linear, or slightly convex. A similar business cycle shape is presented by King and Plosser (1994) and Balke and Wynne (1995), which is extracted from the log-level of real output after dividing the cycle in eight sub-phases. In fact, Figure 5.2 is remarkably similar to the shape of the cycle described in figure 1 of Balke and Wynne (1995, p. 654). In contrast, the triangle approximation has high resemblance with the "stylised two-phase cycle" of Sichel (1994, p. 275). Therefore, a straight line is a good approximation of the shape of contractions because the growth rate is almost constant in this phase, but it is not a good approximation of the expansions because the growth rates are larger at the beginning than at the end of the phase.

Sichel (1993) defines two concepts of asymmetries: steepness, which happens when contractions are steeper than expansions (or vice-versa), and deepness, which occurs when troughs are deeper than peaks are tall. Using both Hodrick-Prescott and Beveridge-Nelson filters to obtain the cycle, Sichel does not find evidence of steepness and only weak evidence of deepness in the US GNP. DeLong and Summers (1986, p. 167) also find no evidence that "economic downturns are brief and severe relative to a trend, whereas upturns are longer and more gradual". The latter authors conclude that the evidence of asymmetry is not strong enough to make linear models a poor approximation of business cycles. These analyses, based on the concept that the business cycle is the result of a band-pass filter, have at least two problems: the type of filtering method may imply different 'cycles' (Canova, 1998), and it is not necessary to have a cycle component to generate data similar to the US business cycle (Pagan, 1997).

For example, Figure 5.1 also presents the growth rates of US GDP (multiplied by 10), which can be seen as the cycle component after a stochastic trend is removed. It is hard to identify steepness or deepness in this series simply by inspecting the graph, although symmetry tests have power to identify these asymmetries if they are presented in the data (Clements and Krolzig, 2000a). The fact that the evidence of steepness and deepness in

US output is not strong (DeLong and Summers, 1986; Sichel, 1993; Verbrugge, 1997) while the excess measure clearly shows that expansions are not mirror reflections of contractions, depends on the definition of the business cycle employed. We identify asymmetries in the shape of the cycle determined by turning points in the log-level of GDP whereas steepness and deepness are tested using the cycle extracted from de-trending (Clements and Krolzig, 2000a) or filtering (Sichel, 1993). In fact, Balke and Wynne (1995) found asymmetries in the shape of the classical cycle, but not asymmetries in the shape of the growth cycle.

McQueen and Thorley (1993) describe a turning point asymmetry – the sharpness – in which peaks are sharp and troughs are round. Sharpness means that there is a large difference between the growth rates before and after the turning points. One way of testing this asymmetry is based on a three-state Markov-chain because a neutral state is needed to capture gradual changes or flat growth rates. The results of McQueen and Thorley indicate strong evidence of this type of asymmetry in industrial production. Additionally, Clements and Krolzig (2000a) present evidence of sharpness in the GDP using parametric tests. This evidence that troughs are sharp and peaks are round supports the concept that recessions are followed by high-growth recoveries and that peaks occur when the economy is experiencing moderate growth. Therefore, the shape of the business cycle illustrated in Figure 5.2, based on the excess measure, confirms sharpness around the trough.

Round peaks and sharp troughs are also present in the Balke and Wynne (1995, p. 643) definition of the shape of the cycle as "the pattern of variation in growth rates of the key aggregates over the course of expansions and recessions". The authors test symmetry by comparing the growth rates of the 8 sub-phases of the cycle given by the turning points. The first sub-phase is defined as the time period of the initial trough and the fifth phase is the time period of the subsequent peak. "The second, third and fourth phases break the expansion into three time intervals of equal length, while the sixth, seventh and eighth phases break the subsequent recession into three time intervals of equal length" (Balke and Wynne, 1995, p. 643). The average growth rates of each sub-phase is calculated by regressing the growth rate of output against dummy variables for each one of the sub-phases. The results of the symmetry tests indicate that the cycle is concave during expansions and linear during contractions, supporting the sharpness asymmetry.

Summarising, the stylised shape of the business cycle, determined by the value of excess of the cumulative loss (gain) over the triangle approximation, exhibits sharpness. Steepness and deepness are not characteristics of the shape of the classical cycle because these types of asymmetries depend on the business cycle being defined as deviations around the trend. Therefore, given the definition of cycle employed, we cannot use the measure of shape of the cycle to assess steepness and deepness in the data generated by the models.

5.2.2 Non-linear Time Series Models and Business Cycle Asymmetries

The business cycles generated by non-linear time series models arise from the interaction of the impulse given by the stochastic part of the models and the propagation given by the dynamics of the models[2]. The shocks, in most of the models, are drawn from a single symmetric normal distribution, but, in some other cases, they are selected from normal distributions with different variances conditional on the regime. The propagation mechanism includes the short-run non-linear dynamics and a stochastic trend.

Figure 5.3 presents a stylised illustration of the shape of the cycle generated from different non-linear time series models. The shape of the cycle is characterised by the average excess of the cumulative gains (losses) over the triangle approximation for contractions and expansions computed with data simulated from the models, which are described in the tables of Chapter 2. To draw the figure, we assume that the models are able to generate the amplitudes and durations, so that the differences between their abilities to generate business

[2]Tong (1990) shows that piecewise linear approximations of non-linear dynamics systems can generate endogenous fluctuations, such as limit cycles. However, the parameters estimated for the threshold models do not imply endogenous fluctuations - the presence of the stochastic shocks is necessary to create cycles.

cycle stylised facts arise only from the shape of the cycle. In addition, being a stylised figure, the amplitude of contractions is too high compared with the amplitude of expansions, but this is done in order to make the shape of cycle easier to see.

The two-regime MS (MS2) has an average excess equal to zero, which means that the shape of the cycle is well approximated by a straight line. When it is assumed that the MS2 can correctly account for the turning points, each phase has a constant growth rate equal to μ_{exp} or μ_{con}, given that the autoregressive coefficients are equal to zero. This means that y_t does not depend on y_{t-1}, where y is the first-difference of output. During expansions, the presence of the dynamics implies that the inclination may change but in such a way that $E[y_t|y_{t-1}]$ is symmetric around μ_{exp}. Giving that one of the characteristics of the US business cycles is that higher growth rates occur at the beginning of the expansions, it is likely that large positive deviations from the μ_{exp} create, via the small negative autoregressive coefficients (eq. 2.5), $E[y_t|y_{t-1}] < \mu_{\text{exp}}$, which implies a slightly convex shape. Because the average excess is equal to zero, the curve should be concave later in the expansion so that the area under the curve is equal to the area of the triangle formed by the duration and the amplitude. During contractions, the deviations from the mean are smaller and $E[y_t|y_{t-1}]$ is more likely to be constant and equal to μ_{con}. A stylised illustration of the shape of the MS2 is labeled as two-phase symmetrical in Figure 5.3. A similar shape is also generated by the Markov-switching model with duration dependence (MS DD) and the unobserved component structural time series model (UCSTM) that share the dynamic characteristics of the MS2.

The plots of the conditional mean for the multiple-regime STAR model (MRSTAR) (Figures 2.6 and 2.10) indicate that conditional mean function and surfaces have constant inclination for both phases. This can be explained by the analysis of the impulse-response plots of the MRSTAR presented by Van Dijk and Franses (1999). For small shocks, the responses are virtually the same for each one of the regimes. Because the simulation employs shocks that are at maximum equal to $|\sigma|$, the dynamic responses of each regime are very sim-

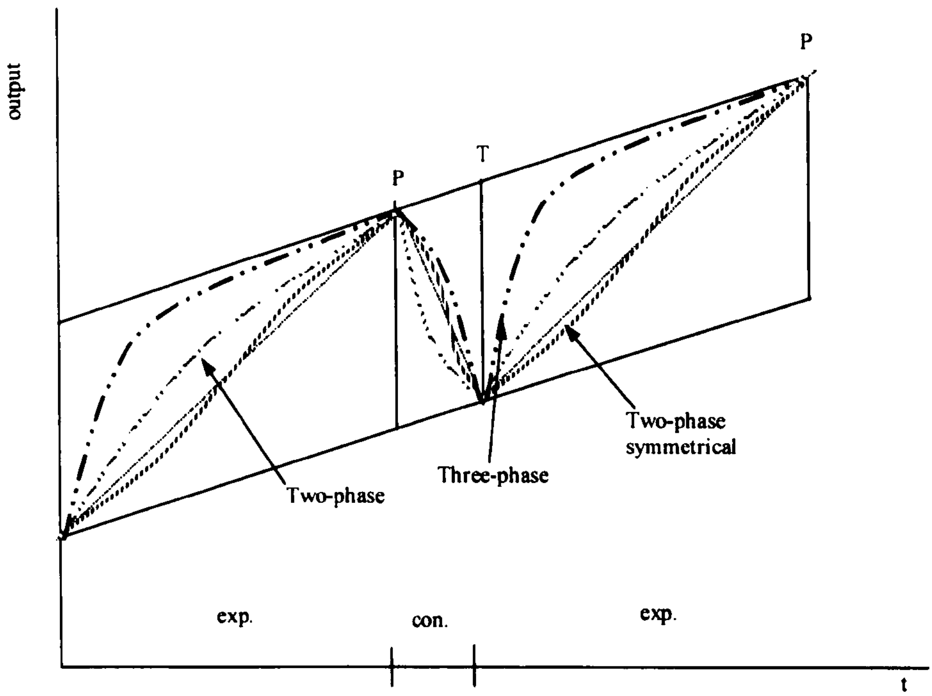

Figure 5.3: Comparing shapes of stylised business cycles generated by non-linear models

ilar. This creates a conditional mean function with a constant inclination. As a consequence, the shape of the cycle generated by this model can be well approximated by a straight line because the growth rate does not change inside the phases.

The shape of the cycle generated by the threshold moving average models are also well approximated by the triangle because the drift is not conditional on the regime and the asymmetric shocks are not strong enough to generate significant asymmetries. Finally, linear models do not generate asymmetries in the propagation mechanism, so even though they can generate durations and amplitudes, the shape of the cycle generated by these models is linear.

The values of the excess from the two-regime SETAR (SETAR2) indicate that the shape of the expansions is concave (excess = 0.22), but not as concave as the observed business cycle (excess = 1.2). Moreover, the model gives a concave shape for the contractions (excess = 0.22), implying that the peaks are not as round as the business cycle and the troughs are not very sharp. This is probably caused by the reversion mechanism presented in the contraction regime of this model: the growth rates are highly negative at the beginning of contraction and they get smaller as the duration of the contraction increases. These dynamics do not generate short contractions because, given the coefficient on the AR1 in the recession regime (0.3), some time is necessary for the recession to become deep enough for the reversion to start. Assuming that the model generates the correct duration and amplitude (which is not always the case), the stylised shape of the cycle generated by the SETAR2 can be classified as two-phase in Figure 5.3.

During contractions, a concave shape is also generated by the four-regime SETAR (SETAR4), the two-regime Markov-switching with slopes and mean changing (MS AR), and the Floor and Ceiling (F&C). These models generate concave contractions because of a fast negative growth rate at the beginning of recessions, although they generally account for the other contraction stylised facts.

Therefore, to get the shape of the business cycle, it is not necessary to have a mechanism that creates smaller negative growth rates to reverse the recession, but rather a high recovery phase in the expansion phase. This point can be illustrated by contrasting the performance of CDR model and the model with the CDR variable conditional on the growth rate being positive (CDR_{pos}). The CDR_{pos} has a better performance in generating stylised facts because the depth of recession only matters in the creation of a strong recovery and has no influence on the smoothness of the shape of the recession phase. As a consequence, the floor is important in the creation of temporary recessions and strong recoveries, but the recovery does not depend on the duration of the contraction but on the deepness of the contraction[3].

[3]This also explain why Markov-switching models with duration-dependence models are so bad in generating stylised facts of the contractions (see, sections 2.4.2 and 2.4.7).

Finally, as discussed in section 2.4, the three-regime MS (MS3) (cont. exc = 0; exp. exc = 0.7) can account for the shape of the US business cycle because it represents a three-phase cycle[4]. Figure 5.3 presents the shape of the cycle generated by a model that characterises a three-phase cycle, such as the MS3 and the State-Space model with Markov-Switching (SS MS). The latter model generates the asymmetric shape of the cycle, using an interaction between a permanent component – which depends on a stochastic trend, the growth rate and asymmetric shocks – and a temporary component – which is activated during the recession regime. The F&C can also generate the shape of the expansions, but it does not generate the shape of contractions. Thus, the shape of the cycle of the F&C is given by the three-phase curve during expansions and by the two-phase during contractions.

Furthermore, the Real Business Cycle (RBC) models cannot generate the shape of the business cycle described in Figure 5.2 (Balke and Wynne, 1995; Harding and Pagan, 2000). Like the RBC models, non-linear models estimated for US output that generate a two-phase cycle cannot generate the asymmetric shape of the US business cycle. Whilst they produce cycles with correct amplitude and durations, so does a random walk with drift (Pagan, 1997).

5.2.3 Business Cycle Theories and the Shape of the Business Cycle

Only non-linear time series models that generate the three-phase cycle are able to reproduce the asymmetric shape of the US business cycle. The floor and ceiling model of Hicks (1950) shows how an unstable equilibrium can be converted into non-linear cycles when a floor and a ceiling are imposed. It characterises recessions as deviations from the feasible ceiling that are followed by strong recoveries when the economy hits the floor, indicating

[4]Actually, Sichel (1994, p. 271) argues that "the three-phase Markov model is not especially informative about the particular pattern of three phases". The problem is, as argued by Clements and Krolzig (2000a) when analysing the three-regime MS of Hess and Iwata (1997b), that the specification on which this argument is based does not have regime-changing variances and the regime probabilities do not imply the three-phase business cycle.

a three-phase cycle. This type of cycle can also be characterised by temporary deviations from a permanent component, such as that suggested by the 'plucking' model of Friedman (1969, p. 264). The evaluation of the SS MS confirms that temporary recessions given a permanent trend are able to characterise the shape of the US business cycle. In addition, the characterisation that recessions are gaps beneath the potential output (DeLong and Summers, 1988) is well supported by our results.

In contrast, the idea that business cycles are generated by switches between two equilibria (Azariadis and Smith, 1998) is not supported by our results. The reason is that the recession should be characterised as a temporary phenomenon and not as a new equilibrium (a point of view that is also supported by Pagan, 1997).

5.2.4 Summary

The main contribution of this section is to show that there is no simple relationship between non-linearities in time series models and the asymmetric shape of the US business cycles, using the classical concept of the cycle. Models that generate two-phase cycles are not adequate to reproduce the shape of the cycle. In addition, tests for asymmetries based on a two-phase business cycle are not very useful for inferring asymmetries in the shape of the classical cycle. The US business cycle has three phases, and this also seems to be the case in Australia and Italy (observing Figures 2.2 and 2.3). Any kind of theoretical or empirical model that attempts to characterise the business cycle needs to have a mechanism that delivers the three phases[5,6]. Linear or two-phase non-linear models can generate durations and amplitudes, but when it is necessary to predict how long it takes the economy to come

[5]This is true even when we remember that the business cycles are not alike. The quartiles of the excess of expansions of the US business cycle are 0.6 and 1.81, implying that the concave shape is robust.

[6]Kim and Murray (2001) argue that the 1990-91 recession is atypical: the transitory mechanism did not have an important effect during the recession, which was not followed by a high-growth recovery. The filtered and smoothed probabilities of the three-regime MS presented in Clements and Krolzig (1998) indicate a weak but existent high growth regime after the 1990-91 recession. Generally, the problem of models with unobserved components given by a Markov chain is that the transition probabilities can only predict recession for 1990-91 when a structural break in the variance is included in 1984 (McConnell and Perez-Quiros, 2000).

back to the last peak after the recession, these models will always over-predict this timing.

5.3 Non-linearities and the Term Structure of Interest Rates

The results of Chapter 3 support three different specifications of threshold vector equilibrium correction models (2R-TVEqCM$_{joint}$, the 2R-TVEqCM$_h$ and the 3R-TVEqCM[7]) as being reasonable to forecast short- and long-term interest rates, and their spread. In this section, we analyse the economic implications of the characteristics of one of the winners of the forecast competition of Chapter 3. Specifically, we investigate the 3R-TVEqCM, which is a three-regime TVEqCM. The 2R-TVEqCM$_{joint}$ is hard to analyse because the cointegrating relation $[1, -1.4]'$ is different from that implied by the expectations theory $[1, 1]'$, even though this may be used as evidence that expectations theory does not hold. In contrast, 3R-TVEqCM has a forecasting performance similar to the 2R-TVEqCM$_{joint}$, except for predicting the spread at long horizons, and it assumes that the spread is the cointegrating relation. The 2R-TVEqCM$_h$ is not a good model to characterise the relationship between the spread and rates because, when the restriction that the spread is not a predictor in the interest rate equations is imposed, the forecasting performance improves (2R-TVAR).

Before studying the characteristics of the three-regime TVEqCM, we observe whether this model can represent the non-linearities found in the data using non-parametric conditional means.

5.3.1 Comparing Conditional Means for Model and Data

One way of assessing the match between the models and the observed data is by calculating the non-parametric estimates of the conditional mean of the change in each rate against the lagged spread, $E[\Delta r_t(.) \mid S_{t-1}]$. A similar approach was employed in Chapter 2 to observe the dynamics generated by different types of non-linear univariate models.

[7]The characteristics of these models are presented in Table 3.1.

In addition, Breunig and Pagan (2001) advocate an approach of this sort to see whether empirical Markov-switching models are a good description of the data. The complexity of the three-regime threshold models suggests the employment of non-parametric conditional means to assess the responses of each rate to the lagged spread. The conditional means are estimated using smoothing splines (see, Simonoff (1996, chap. 5.6)). The degree of smoothness is controlled by the number of degrees of freedom of the regression, which is set to 6.

The TVEqCM is simulated to obtain $\Delta r_t(.)$ and S_t, with the parameters estimated for the sample 1960-1998:4. As discussed in section 3.7 in the case of generating forecasts, the simulation may depend on whether the variance-covariance of the residuals is allowed to change among regimes or not. We use Monte Carlo to simulate 5000 observations from the three-regime TVEqCM; and two types of simulation procedures are employed: with the variance-covariance of the full sample residuals (HOM) and with different variance-covariance using the residuals of each regime (HET).

The estimates obtained for the model and the data are plotted in Figure 5.4. The plots underline the finding that S_{t-1} is positively related to $\Delta r_t\,(s)$ for $S_{t-1} < 0$ (the relationship is accentuated for the homoscedastic model). The long rate reacts negatively to the spread when the latter exceeds 2 1/2, and positively when $S_{t-1} < 0$, but in this case the response is more muted than that of $\Delta r\,(s)$. These features are also apparent in the conditional mean functions estimated for the US data, indicating that the models are able to capture the salient features of of the dynamic linkages between $r\,(l)$ and $r\,(s)$ over the period 1960 - 1998.

5.3.2 Implications for the Expectations Theory

Given that the 3R-TVEqCM captures data non-linearities as described in the last subsection, we discuss some of the key features of this model in the light of the literature on

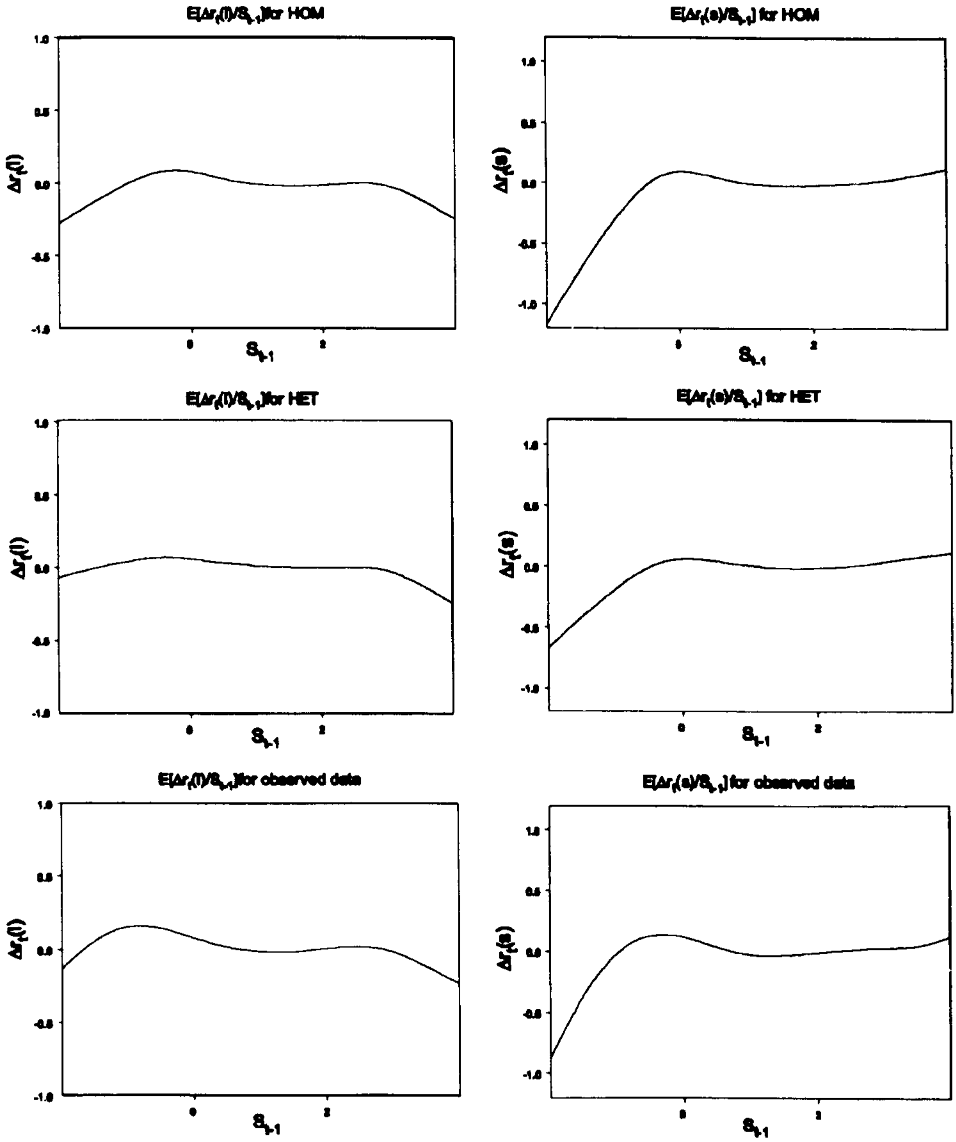

Figure 5.4: Conditional mean functions for three-regime TVEqCM and the data

the term structure of interest rates in this subsection. From Figure 5.4, it is apparent that the 3R-TVEqCM is characterized by significant feedback of negative spreads to lower short rates, and to a lesser extent, of negative spreads on to lower long rates. Negative spreads are the results of a reverse yield curve:

> Suppose, for example, investors believe that the prevailing level of bond yields is unusually high relative to historical precedent and that lower rates in the future are more probable than higher ones. (...) If investors act in accordance with these expectations, they will tend to bid up the prices (force down the yields) of long-term bonds and sell off short-term securities, causing their price to fall (yields to rise) (Malkiel, 1970, p. 7).

Although this argument does not examine the effect of the risk premium as a result of the maturity, it is a good explanation for why inverse yield curves may occur when the risk premium is small. From Figure 3.2, it is apparent that the spread flattens and becomes negative at the end of 1966, during most of 1969, from June 1973 to October 1974, from November 1978 to April 1980 and from November 1980 to August 1981. After each one of these periods, save 1966[8], the US economy went into recession, as defined by the NBER. One could then argue that negative spreads lead recession periods (Estrella and Hardouvelis, 1991; Hamilton and Kim, 2000). One of the reasons for this relationship is that the spread captures the effect of monetary policy. When a shock hits the economy, the expectation of inflation may change in the direction of higher future inflation rates, increasing the long-term interest rates and in consequence the spread. As a response to this shock, the government may actively increase the short-term interest rate to reduce inflation, which may contribute to the economy moving into a recession. The fast rising short-term interest rate causes the effect described by Malkiel (1970), and the spread is temporarily negative. On the other hand, during the recession period, when the policy is reversed, short-term interest rates will fall and the spread will become positive again.

[8]In 1966 the FED tried to implement 'Operation Twist', whereby action on the market was aimed at increasing short-term interest rates, to reduce the balance-of-payment deficit, and at the same time lowering long rates to promote economic growth (Malkiel, 1970). This might explain why the yield curve went negative without a recession ensuing.

When the dynamics of the growth of the interest rate can be described by a regime with negative spread, the coefficient of the spread is around 0.6 for the short-term rate equation. This means that the negative spread implies a pressure to reduce short-term interest rates while there is no effect in the long-term. Therefore, the spread helps to forecast short-term interest rates, as predicted by the expectations theory. This characteristic has been reported in the literature by Pfann et al. (1996) and can be associated with periods of predictability of the interest rate as defined by Mankiw and Miron (1986). On the other hand, in the middle regime, the spread does not help to forecast short-term and long-term interest rates, in contradiction with the expectations theory. A reason is that short-term interest rates behave as a random walk in periods of low volatility (Gray, 1996), which is one of the characteristics of the middle regime period. The random walk arises from the persistent smoothing of the short-term interest rate by the monetary authority, which implies that the market expectation of the short-term interest rate is the same as its current value (Mankiw and Miron, 1986).

After 1984 the occurrences of negative spreads are replaced instead by episodes when long rates are unusually high relative to short rates. In terms of the 3R-TVEqCM, the lower regime no longer occurs, and the third regime becomes more prevalent. This may be a result of changes in the monetary policy in the direction of more aggressively smoothed short-term interest rates (Clarida, Gali and Gertler, 2000), which follow the high volatility and negative spreads of 1979-1982, when the FED targeted monetary aggregates rather than smoothing the short-term rate. This new monetary regime increases the credibility of the monetary authority by the pursuance of persistent and well defined policies. Watson (1999) argues that a feature of the second half of the 1980's and 1990's is that even small increases in the persistence of the short-term interest rate, especially at already high levels of persistence, would translate into substantially higher volatility in the long term rate. Therefore, small changes in short rates can engineer large changes in long rates, as noted by Campbell (1995) in his discussion of the impact of monetary policy on the bond market in the Spring of 1994.

However, whatever the source(s) of the large positive spreads, our models do not exhibit the positive correlation between long rates and the spread predicted by the theory:

> When long rates are unusually high relative to short rates, long rates do not decline to restore the usual yield curve, as one might suppose. Instead long rates tend to rise; the yield spread falls only because short rates rise even faster (Campbell, 1995, p. 137-8).

In our model, long rates decrease in response to the long-rate being too high. But as the conditional mean functions in Figure 5.4 show, the data, as well as the models, are at odds with this prediction, as is the evidence of Hardouvelis (1994) and Campbell (1995). One strand of argument is that over-reaction of the public to monetary authority policies, aimed at preventing the economy overheating, increases the variability of long rates, so that in subsequent periods the disequilibria between the expected values of the short-term rate and the actual long-term rate leads to decreasing long rates, establishing a negative correlation between the spread and long rates[9]. The responses to over (under)-reactions of the long-term interest rate to changes in monetary policy are characterised in our models by allowing for non-linearities in a cointegrated system. The upper regime of 3R-TVEqCM captures, for example, the periods from May 1987 to October 1987 and from January 1994 to November 1994, when long rates over-reacted to tightening monetary policy. In addition, when long-term interest rates under-reacted to the easing of policy after the 1982 and 1991 recessions, and after the Stock Market crash of October 1987, high spreads and predictable subsequent declines in long-term interest rates occur.

Summarising, a non-linear model that presents different dynamics among different regimes is a better representation compared to linear models because the possibility that the dynamics depends on the level of the spread can be accommodated, which is an important

[9] Of course, if policy is credible, the long-rate should reflect the expected future low inflation environment brought about by higher short rates. A time-varying risk premium could be another possible explanation for the negative correlation between the spread and long-term interest rate. Hardouvelis (1994) finds little support for this, whereas Tzavalis and Wickens (1998) take the opposite position.

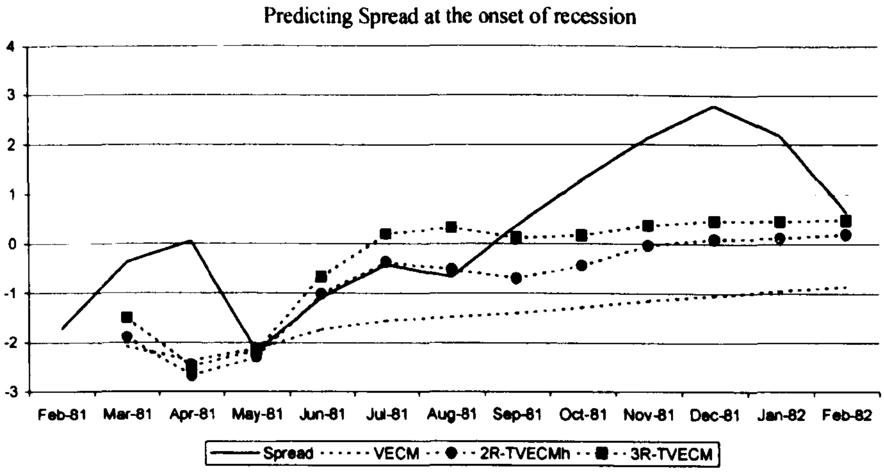

Figure 5.5: Predicting the spread at the onset of recession

characteristic of the three-month Treasury bill and also of the 10-year Treasury bond. The spread does not help to forecast the short-term interest rate when the spread is positive, and does not forecast the long-term interest rate when it is smaller than 2 3/4. For these regimes, the interest rates have relative low volatility and are highly persistent. When the spread is negative, the expectations theory holds and the spread forecasts increasing short-term rates. When the spread is larger than 2 3/4, it forecasts decreasing long-term interest rates which is not in agreement with the expectations theory.

5.3.3 Forecasting the Spread near Business Cycle Turning Points

As discussed in the previous subsection, the literature reports that the spread is a good predictor of recessions because it is negative in the months before the peak (inverse yield curve) and it is positive during the recessions (Estrella and Hardouvelis, 1991; Estrella and Mishkin, 1998; Hamilton and Kim, 2000). In this section, we compare forecasts for the spread generated for linear and non-linear models near the July 1981 peak defined by the NBER.

Figure 5.5 presents the forecasts from the VEqCM, the 2R-TVEqCM$_h$ and the 3R-

TVEqCM with origin four months before the peak of July 1981. The predictions are 12 steps-ahead forecasts using only the information available until February 1981. As can be seen from Figure 5.5, the spread is negative at the forecast origin, and after eight steps-ahead, positive. This characterises the switch between expansions and contractions in the value of the spread. The predictions of the VEqCM are similar to the non-linear models for $h = 1, 2, 3$, but at longer horizons, the linear model predicts negative spreads for all the periods, while the non-linear models predict near positive values.

Therefore, Figure 5.5 may be interpreted as an illustrative example of how non-linearities matter in predicting the spread at the onset of recessions.

5.3.4 Summary

The predictions of the expectation theory of the term structure of the interest rate are analysed in a three-regime TVEqCM, which is one of the winners of the forecasting evaluation of Chapter 3 and fits well the dynamics between interest rates and the spread. The theoretical indication that the spread helps to predict changes in interest rates is accepted by the data when the yield curve is inverse: the negative spread predicts decreasing short-term interest rates. When the yield curve is upward, the spread does not help to predict both interest rates, except if the positive inclination is large, meaning spreads greater than 2 3/4. In the latter case, the large positive spread predicts decreasing long-term interest rates, on the contrary to the expectations theory predictions that the long-term interest rate must increase under these conditions.

5.4 Predicting Recession in 2001

Last March, The Economist wrote that "the dismal scientists have a dismal record in predicting recessions"(*Don't Mention the R-word*, 2001). These criticisms show the poor reputation of economic forecasters in predicting the turning points of the US economy. On the other hand, early turning point predictions are necessary for effective utilisation of economic policy to stabilise fluctuations. Examining the results of Chapter 4, one can observe that models using the spread as a leading indicator do not give good predictions for every single recession. Only two models (out of 11) predicted the 1990-91 recession. These models have an important characteristic, namely, a structural break in the beginning of 80's that affect the non-linear dynamics and the conditional variance. The reduction of the variance of output growth is explained by the reduction of the variance of inventories, created by better inventory management (McConnell and Perez-Quiros, 2000). Changes in the spread may be created by changes in the monetary regimes (Watson, 1999).

The best forecasters in the assessment of Chapter 4 can be employed to answer the contemporary question in the financial media whether the US is going to have a recession in 2001[10]. Negative predictions started to be made after strong reductions in stock prices and in the index of consumer confidence in the last quarter of 2000. Using the rule that negative spreads predict recessions in the near future, the last two quarters of 2000 presented negative spreads, which were translated to high probability of recession by the simple rule, as discussed in section 2.2. The objective of this section, therefore, is to evaluate how the models that performed best in-sample in the last section perform out-of-sample, and specifically to evaluate the probabilities of recession for 2001. We also compare the results of the models of this work with the Stock and Watson Leading recession index and the Survey of Professional Forecasters.

For this exercise, we choose the models that better calibrate the data probabilities for events A and B in the period 1983:1999, because of the strong evidence of structural break in the output growth-spread relationship in this work and in the literature (Stock and Watson, 2001). We evaluate two non-linear models (smooth transition (ST) and three-regime

[10] This section was written in April 2001, using the information available at that time.

threshold (3T)), one time varying smooth transition model (PTVST) and two structural break threshold models (SBT and SBTVAR). The threshold model with three regimes is analysed instead of the two-regime version because the three-regime model is a better representation of the data for the more recent sample, although they give similar performance when the whole sample period is considered. These models are described in the appendix of Chapter 4.

5.4.1 Forecasts for the Out-of-Sample Period

The point forecasts for the out-of-sample period 2000:1-2000:4 are plotted in Figure 5.6 and the mean squared forecast errors are presented in Table 5.1. The 3T model gives the smallest MSFE and the SBT gives the largest; the MSFE of the latter being four times larger than the former. Observe in Figure 5.6 that the variability of point forecasts for the PTVST and the SBT is larger than for the ST and the 3T. This result may be an indication that the SBT and also the SBTVAR are overfitting the data which may lead to poor out-of-sample performance (Ramsey, 1996), even compared with other non-linear models that may also be overfitting[11]. However, this characteristic improves the in-sample ability to predict events A and B. Therefore, less weight should be given to point forecasts from the SBT and SBTVAR compared to event probability forecasts.

5.4.2 Forecasts of the Probability of Events A and B for 2001

The probability that two quarters of negative growth will occur in the next five quarters (event A) is 22% using the PTVST model and 81% using the SBTVAR. The different probabilities among models depend on how the negative spreads of the last quarters of 2000 affect the non-linearity of the model. The SBT model has the strongest negative growth

[11] The SBT model, for example, is the result of the estimation of 24 parameters for each equation plus two thresholds and a break data. On other hand, the ST model has 9 parameters plus the smoothing parameter and threshold for the output growth equation, and 11 parameters plus smoothing parameter and threshold for the spread equation.

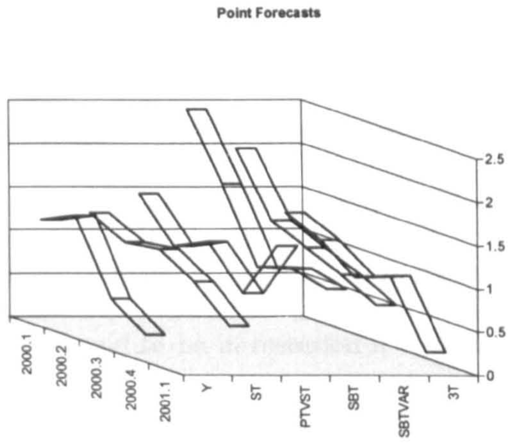

Figure 5.6: Comparing output growth forecasts for 2000:1 to 2001:1

Table 5.1: Comparison of out-of-sample predictions

	ST	3T	PTVST	SBT	SBTVAR	XRI
$MSFE$ (2000:1–2000:4)	0.204	0.153	0.226	0.630	0.381	
$\Pr(A_{2001:1})$	0.569	0.490	0.222	0.550	0.810	
$\Pr(B_{2001:1})$	0.752	0.676	0.567	0.561	0.853	
$\Pr(y_{2001:1} < 0$ and $y_{2001:2} < 0)$	0.112	0.160	0.005	0.000	0.072	0.07

Notes: ST, 3T, PTVST, SBT and SBTVAR are described in the Appendix; XRI is experimental leading recession index of Stock and Watson, 1989.

reaction which generates a false recessive alarm in 1999. The average probability in predicting event A is 61%, which means that is quite likely that a recession will happen by 2002:1. The probability of event B has a higher mean: 68%, which means that models are predicting a slowdown.

5.4.3 Comparing with Stock and Watson Recession Index

Table 5.1 also shows the probabilities of negative growth in the first two quarters of 2001. These probabilities are calculated to compare with the (monthly) Stock and Watson

XRI index. The Stock and Watson experimental leading recession index (XRI) [12] extracts the probability of recession using eight components (after the 1997 revision) in a dynamic factor model (Stock and Watson, 1989, 1993). Using different series such as interest rates and manufacturers' unfilled orders, the authors try to explore comovements between economic variables and to detect recession when a downturn is signalled by different sectors of the economy. A monthly period is said to be in recession if that month is either in a sequence of six consecutive declines of the composite index below some boundary or in a sequence of nine declines below the boundary with no more than one increase during the middle seven months (Stock and Watson, 1989, p. 357). This definition of recession is employed to identify recessions in the observed data and also to calculate the leading recession index. To compare with the probabilities extracted from the models presented in this work, we suppose that this recession definition is equivalent to obtaining two negative output growth predictions in a two-quarter horizon. Given that the last calculation of the experimental leading index was still above the boundary of the recession period, the possibility that the second definition of recession is relevant can be neglected.

The experimental leading recession index calculated with information until 2000:12 is presented in Table 5.1. The average probability that a recession would happen in the first two quarters of 2001 calculated for the models of this work is 7.8%, similar to the Stock and Watson XRI (7%). Because the effect of the negative spread is more delayed in the SBT, this model is still predicting growth for this period. The models without structural break present higher probability (around 13%).

5.4.4 Comparing with the Survey of Professional Forecasters

The Survey of Professional Forecasters (SPF) collects information on forecasts of economic series, such as output, unemployment, interest rates and inflation, made by pri-

[12] Definition of the index and historical values are at http://ksghome.harvard.edu/~.JStock.Academic.Ksg /xri/INDEX.HTM.

vate sector forecasters (34 of them), organised by the Federal Reserve Bank of Philadelphia (Croushore, 1993). The information is collected every three months and the results are published on http://www.phil.frb.org/econ/spf. To compare the predictions of our models with the Survey of Professional Forecasts, we present in Table 5.2 the point forecasts for all the quarters of 2001 and the first quarter of 2002. The forecasts in our models were calculated with 5000 replications by Monte Carlo simulation. As discussed before, overparameterisation may influence the results of more flexible models. The SPF predicts a small positive growth of 0.8% at annual rate for the first quarter of 2001 and higher positive growth for the following quarters. Our models predict higher growth rates for the first quarter and negative rates for later periods. The 3T model supports the hypothesis of a soft landing of the US economy, while the ST model predicts a short mild recession for the second and third quarters of the year. The PTVST model does not present any meaningful pattern. The SBTVAR model predicts a long recession and the SBT model predicts a deep recession. Given this disagreement in the point forecasts, the main conclusion that can be drawn from Table 5.2 is that our models differ from the optimist predictions of the SPF and they suggest either very low growth rates or negative growth rates for the last three-quarters of the year. However it is hard to find support from these numbers that a recession, as defined by the NBER, is (will be) happening in 2001.

The SPF also publishes the mean (over forecasters) of the estimated probability of negative output growth for each one of the next 5 quarters (see also Tay and Wallis, 2000). The mean of the risk of negative growth for the period 2001:1 to 2002:1, published by the SPF in February, 2001, using information until January, 2001, is in the lower panel of Table 5.2. By way of comparison, we compute the probability of negative output growth for the 2001:1-2002:1 period from our models, applying parameters estimated for the 1954-1999 sample and data until 2000:4. Again compared with the SPF prediction, the models are more optimistic for the first quarter of 2001 and more pessimistic for the last two quarters of 2001, when the

Table 5.2: Comparing 2001 forecasts

	Point Forecasts for Real GDP at annual rate %					
	SPF	ST	3T	PTVST	SBT	SBTVAR
2001:1	0.8	2.0	0.8	5.8	3.8	3.1
2001:2	2.2	-0.2	0.4	-0.5	3.4	-1.5
2001:3	3.3	-0.4	0.2	3.6	0.8	-4.3
2001:4	3.7	0.0	0.8	1.2	-2.9	-2.4
2002:1	3.1	0.5	1.6	1.0	-8.0	-2.1
	Risk of negative growth %					
2001:1	37	27	38	4	1	7
2001:2	32	52	47	55	1.5	75
2001:3	23	54	49	25	47	86
2001:4	18	50	43	41	56	73
2002:1	13	44	35	39	59	72

Notes: ST, 3T, PTVST, SBT and SBTVAR are described in the Appendix; SPF is the results from the Survey of Professional Forecasters, published in February, 2001

probability of negative growth is on average 52% for both quarters[13]. However, only after the publication of GDP data for the fourth quarter of 2001 can a proper evaluation of the predictions of Table 5.2 be made.

5.4.5 Summary

For the out-of-sample period (2000:1-2001:4), the evaluation of event probabilities is not useful because the events did not occur in this period, but the MSFEs for point forecasts are calculated. Large MSFEs for structural break threshold models suggest that these models, although they are good event probability forecasters, are overfitting the data, generating poor point forecasts. The evaluation of the models in predicting growth and the probability of a recession in 2001 shows that the models are capturing different characteristics of the data because they indicate different directions using the same leading indicator. The predictions of some models proposed in this work for 2001 agree with the Stock and Watson (1989) experimental leading recession index but they are too pessimistic when compared with the

[13]Diebold, Tay and Wallis (1999) evaluate the SPF density forecasts of inflation. They conclude that the forecasts are not optimal. However, whether the predictions for the risk of negative growth are sub-optimal has not been evaluated.

Survey of Professional Forecasters predictions. In general, the models predict a mild recession or at least one quarter of negative growth for the third and the fourth quarters of 2001.

Postscript: The preliminary results of the Real GDP growth, published by the US Department of Commerce in July 27, 2001, for the first two quarters of 2001 are 1.3% and 0.7% at annual rate. This confirms the predictions of the probability of recession by the SBT, the 3T and also the SPF, given that these forecasters do not predict a probability of recession greater than 50% for 2001:2. However, these are still preliminary results and only when the data for the whole year is available, a stringent evaluation of these forecasts can be done.

Postscript 2: The GDP growth rates published by the US Department of Commerce in November 30, 2001, for the first, second and third quarter of 2001 are 1.3%, 0.3% and -1.1%. Observing Table 5.2, the ST and the SBTVAR predicted negative growth for 2001:3. Therefore, this preliminary evaluation supports the non-linear autoregressive leading indicator of Anderson and Vahid (2000) and, in lesser extent, the structural break threshold model proposed in this thesis as the models that anticipated the current US recession.

5.5 Conclusions

The evaluation of univariate non-linear time series models is useful for understanding which type of non-linear dynamic is necessary to reproduce the US business cycle asymmetries, indicating that the business cycle is better characterised by three phases. This result contrasts with the idea that the business cycle can be viewed as endogenous switches between two equilibria (Eudey and Perli, 1999) and supports the idea that business cycles are the result of transitory deviations from potential output (the ceiling).

The assessment of the dynamics of the three-regime threshold equilibrium model applied to the US term structure of interest rates can help to clarify the results of the empirical tests of the expectations theory of the term structure. It also shows that the dynamics for the short- and the long-term rates depend on the regime, which can be characterised, for example, by monetary regimes (Fuhrer, 1996).

The presence of a structural break in non-linear bivariate systems that employ the spread as the leading indicator affects the prediction of the probability of recessions and negative growth for 2001. The structural break threshold model predicts negative growth only after 2001:4 while the smooth transition model predicts this event for 2001:2. The models with non-linearity and structural breaks indicate a deep recession for 2001:4-2002:1, while the Survey of Professional Forecasters forecasts 3% growth. Observe that these predictions are made without taking into account the action of the Federal Reserve of reducing interest rates, which could reverse these negative predictions or create a soft landing, such as the one that occurred after the Russian financial crisis in 1998/99.

Chapter 6

Conclusion

In general, this thesis furthers our understanding of the comparative properties of different non-linear time series models applied to macroeconomics, which can help to motivate the development of economic theory. The assessment performed in this work demonstrates that (i) the simple presence of a non-linearity does not signify that non-linear time series models can generate the asymmetries of the classical business cycles; (ii) non-linearities can improve forecasts of multivariate linear models at short horizons; (iii) non-linearities and structural break are important to predict US recession probabilities using the spread as leading indicator. Therefore, this thesis indicates that non-linearities are relevant to the dynamics of macroeconomic data in some specific cases and it is important to delineate the circumstances in which these gains can be achieved. Our econometric results support (i) the three-phase characterisation of the US classical business cycle, (ii) the ability of the spread to predict long- and short-term interest rates in some monetary regimes, and (iii) the ability of the spread to predict recessions, even after 1984.

6.1 Main findings

The assessment of the ability of non-linear time series models to reproduce business cycle stylised facts supports previous findings (Hess and Iwata, 1997b; Harding and Pagan, 2001b) that a first-order autoregressive model of the first-difference of output can reproduce the durations and amplitudes of the US business cycle. This simple linear model is also capable of reproducing amplitudes and durations of the Italian and Australian business cycle. However, non-linearities are needed to reproduce the shape of the cycle. Non-linear models only reproduce the shape of the US business cycle when they are able to characterise a three-phase cycle: recessions are followed by high growth recoveries that eventually give way to a moderate growth phase. To characterise a three-phase cycle, a model does not need to reflect progressively smaller negative growth rates inside the contraction phase, but strong recoveries just after the trough. Examples of these type of models are the three-regime Markov-switching model (Clements and Krolzig, 1998) and the state-space model with Markov-switching (Kim and Nelson, 1999a). The former model has transition probabilities defined in such a way that the switches between regimes follow the pattern of a three-phase cycle, including changes in the variance. The latter model considers recessions as a temporary event given by a combination of asymmetric shocks in the trend and a negative temporary component that only works during recessions and just after the trough. Our analysis of the ability of non-linear models to reproduce the shape of the cycle concludes that the two-phase cycle – contractions followed by expansions – which is the implicit assumption of non-linear models with two regimes, is not a good representation of the US business cycle. This does not mean that these models cannot capture turning points or the durations of the cycles correctly but that they cannot capture the asymmetric shape of the classical cycle.

The inclusion of non-linearity in vector equilibrium correction models improves short-horizon forecasts of the first-differences of US short- and long-term interest rates and of their spread. These forecast gains are generated in part by the non-linear short-run dynamics of the model. Non-linearities also result in gains when threshold vector autoregressive models are compared with vector autoregressive models. However, when AR models are employed as the benchmark, neither cointegration, non-linearities, nor the ability of the spread to predict interest rates generate any forecast improvement for the rates. Even though when forecasts for the spread are evaluated, one can observe strong forecast accuracy gains at long horizons from the threshold vector equilibrium correction model that jointly estimates the cointegration vector and the threshold. In addition, the comparison of threshold VARs with threshold VEqCMs indicates that the modelling of cointegration improves forecasts of the cointegrating relation, in support of existing results based on linear models (Clements and Hendry, 1995). The result that AR models are equivalent or better forecasters than VEqCMs when the first-difference of interest rates are evaluated (Christoffersen and Diebold, 1998) is extended to threshold VEqCMs compared to AR models.

Our analysis of the non-linearities implied by the dynamics of the threshold vector equilibrium correction model, which has a good forecast performance and fits well the conditional mean function describing non-linearities in the first-difference of interest rates conditional on the spread, finds that the expectations theory of the term structure of interest rates only holds when the spread is negative. In this case, negative spreads forecast increases in short-term interest rates. When the last-period spread is positive but smaller than $2\,\frac{1}{2}$, the spread does not predict future changes in interest rates, contrary to the expectations theory. For values of the past period spread larger than $2\,\frac{1}{2}$, the spread helps to predict long-term interest rates, but in the opposite direction compared to the theory's predictions which are that long-term interest rates will increase.

Tests for non-linearities and structural instability find evidence of a structural break and non-linearities in the ability of the spread to predict US output growth. The results confirm that non-linearity is necessary to predict the probability of recession using bivariate systems of output growth and the spread (Anderson and Vahid, 2000). A new finding is that only when a break in the transition functions is assumed at the beginning of the 80s is the 1990/91 recession predictable using non-linear bivariate systems. The probabilities of two types of recessional event are extracted from the models, employing stochastic simulation. Using a score measure that takes into account both successful predictions and false alarms, threshold models with three regimes, time-varying smooth transition models and structural break threshold models all perform well in the more recent sample period.

The analysis of the dynamics of the structural break threshold model shows that it supports previous findings that the relation between output growth and the spread is non-linear (Galbraith and Tkacz, 2000; Anderson and Vahid, 2000) and unstable (Stock and Watson, 2001), and that there is a structural break in the variance of output growth (Kim and Nelson, 1999b; McConnell and Perez-Quiros, 2000). The probability of negative growth in the second quarter of 2001 employing the models with structural break and non-linearity is 1.5% while a three-regime threshold model gives a figure of 45%. In contrast, the results are similar when this probability is evaluated for 2001:4: 56% and 43%, respectively. Therefore, whether or nor we allow for structural break affects the timing and the strength of the prediction of recession in a five horizon forecast using data up to December 2000.

6.2 Other contributions

With the aim of assessing different non-linear specifications applied to macroeconomic time series, this thesis makes two contributions for testing and modelling non-linear time series models: an LR test for non-linearities in vector equilibrium correction models, and a new model with a structural break that affects the thresholds and the delay of a threshold model. The LR test is employed to test non-linearity in the vector equilibrium correction model of short- and long-term interest rates. It is also employed to verify whether a two- or a three-regime specification is adequate. In addition, the LR test is employed to test threshold non-linearity and structural breaks in VARs in Chapter 4. The structural break threshold model is applied successfully to describe non-linearity, instability and variance changes in the ability of the spread to predict economic activity. In addition, testing, modelling and estimation procedures applied to threshold vector equilibrium models are extensions to the literature on thresholds in univariate and regression models.

This thesis also presents contributions to the evaluation methods employed. The inclusion of the quartiles as stylised facts is a further way of capturing the important characteristic that business cycles are not alike. The application of conditional mean surfaces estimated non-parametrically with simulated data is another contribution to the evaluation of non-linear time series models. Although Pagan (1999) and Breunig and Pagan (2001) have employed conditional means to evaluate non-linear time series models, the method employed in Chapter 2 innovates in two directions: it uses conditional mean surfaces to observe the dynamics created conditional upon the information of the last two periods, which is important in observing recoveries from recessions, and it employs *loess* as a more robust non-parametric estimation of conditional mean functions and effective estimator for conditional mean surfaces. Finally, the utilisation of the Kuipers score to evaluate event probabilities is not very common in the literature, although Granger and Pesaran (2000) explicitly advocate this accuracy measure. Compared to the other two popular score measures – QPS and LPS – the Kuipers score seems more effective in evaluating economically meaningful events: failure to predict recessions, and false alarms.

6.3 Some Open Questions

The availability of different technologies to test 'threshold cointegration' (reviewed in Chapter 3) indicates that Monte Carlo evaluation could help to understand the comparative properties of the tests. The empirical applications show that tests that assume unknown thresholds are of better assistance in specifying threshold equilibrium correction models, but it is not clear whether it is best to test threshold non-linearity in the cointegration relation or equilibrium correction. The heavy computational burden of these tests may be a problem for this type of evaluation. In addition, the effect of heteroscedasticity corrections to define statistics for tests of non-linearity is not clear. In Chapter 4, the assumption of heteroscedasticity changes the results of F-tests with smooth transition models and threshold models under the alternative hypothesis. Of course, the problem arises from the fact that heteroscedasticity may be generated from a non-linear relation being specified as a linear one.

The conclusions in favour of a three-phase cycle as a good representation of the US business cycle are based on data up to the last turning point in 1991. The reduction in the volatility of output growth may imply that recoveries are not as strong as before, given that the reduction of the variance may be caused by the lack of strong recoveries. In fact, Sichel (1993) argues that the three–phase business cycle is caused by inventory cycles, whereas McConnell and Perez-Quiros (2000) find that it is inventory investment that is responsible for the reduction in the variance of output growth. Researchers will have problems in analysing changes in US business cycle phases in the absence of new turning points.

Our analysis of the structure of the regimes defined by the three-regime threshold vector equilibrium correction model indicates that the occurrence of regimes may also have changed. The possibility of changes in the non-linear dynamics between interest rates of different maturities and the spread can be exploited using extensions of the structural break threshold models employed in Chapter 4.

In addition, models with thresholds in the conditional mean and in the variance, including conditional heteroscedasticity (e.g., Li and Li, 1996), could be employed to improve the estimation of the system of interest rates and to specify the spread equation in the bivariate system to predict output growth. This may solve the problem of remaining non-linearity in the latter case by including autoregressive heteroscedasticity in the residuals.

A final possibility is to apply structural break threshold models to predict recessions using different leading indicators, such as some of the financial variables indicated by Stock and Watson (2001). Spreads of different maturities could also be employed as leading indicators. With a Kuipers score of around 50% for the models proposed in this work, there is still space for new research to improve the economists' ability for predicting recessions.

Bibliography

Anderson, H. M. (1997). Transaction costs and non-linear adjustment towards equilibrium in the US Treasury bill market, *Oxford Bulletin of Economics and Statistics* **59**: 465–84.

Anderson, H. M. and Vahid, F. (1998). Testing multiple equation systems for common nonlinear components, *Journal of Econometrics* **84**: 1–36.

Anderson, H. M. and Vahid, F. (2000). Predicting the probability of a recession with nonlinear autoregressive leading indicator models, *Monash University, Department of Econometrics and Business Statistics Working Paper* **3/2000**.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point, *Econometrica* **61**: 821–56.

Ang, A. and Bekaert, G. (1998). Regime switches in interest rates, *National Bureau of Economic Research, Working Paper* **6508**.

Azariadis, C. and Smith, B. (1998). Financial intermediation and regime switching in business cycles, *American Economic Review* **88**: 516–36.

Bai, J. (1997). Estimating multiple breaks one at a time, *Econometric Theory* **13**: 315–52.

Bai, J. (1999). Likelihood ratio tests for multiple structural break, *Journal of Econometrics* **91**: 299–323.

Bai, J., Lumsdaine, R. L. and Stock, J. H. (1998). Testing for and dating common breaks in multivariate time series, *Review of Economic Studies* **65**: 395–432.

Balke, N. S. and Fomby, T. B. (1997). Threshold cointegration, *International Economic Review* **38**: 627–45.

Balke, N. S. and Wynne, M. A. (1995). Recessions and recoveries in Real Business Cycle models, *Economic Inquiry* **33**: 640–63.

Beaudry, P. and Koop, G. (1993). Do recessions permanently change output?, *Journal of Monetary Economics* **31**: 149–63.

Berben, R.-P. and Van Dijk, D. (1999). Unit root tests and asymmetric adjustment: A reassessment, *Econometric Institute, Erasmus Rotterdam, Research Report* **1999-02/A**.

Berk, J. M. (1998). The information content of the yield curve for monetary policy: A survey, *De Economist* **146**: 303–20.

Bernard, H. and Gerlach, S. (1998). Does the term structure predict recessions? The international evidence, *CEPR, Discussion Papers* **1892**.

Bodman, P. M. (1998). Asymmetry and duration dependence in Australian GDP and unemployment, *The Economic Record* **74**: 399–411.

Bodman, P. M. and Crosby, M. (1998). The Australian business cycle: Joe Paoloka or dead cat bounce?, *University of Melbourne, Department of Economics, Research Paper* **669**.

Boldin, M. D. (1996). A check on the robustness of Hamilton's Markov switching model approach to the economic analysis of the business cycle, *Studies in Nonlinear Dynamics and Econometrics* **1**: 35–46.

Boldin, M. D. (1999). Should policy makers worry about asymmetries in the business cycle?, *Studies in Nonlinear Dynamics and Econometrics* **3**: 203–20.

Bonser-Neal, C. and Morley, T. R. (1997). Does the yield spread predict real economic activity? A multicountry analysis, *Federal Reserve Bank of Kansas City, Economic Review* **82**: 38–53.

Bradley, M. D. and Jansen, D. W. (1997). Nonlinear business cycle dynamics: Cross-country evidence on the persistence of aggregate shocks, *Economic Inquiry* **35**: 495–509.

Brännäs, K. and De Gooijer, J. G. (1994). Autoregressive-asymmetric moving average models for business cycle data, *Journal of Forecasting* **13**: 529–44.

Brännäs, K., De Gooijer, J. G. and Teräsvirta, T. (1998). Testing linearity against nonlinear moving average models, *Communications in Statistics - Theory and Methods* **27**: 2025–35.

Brännäs, K. and De Luna, X. (1998). Generalized method of moment and indirect estimation of the ARasMA model, *Computational Statistics* **13**: 485–94.

Breunig, R. and Pagan, A. R. (2001). Some simple methods for assessing Markov switching models, *Australian National University (mimeo)* .

Bry, G. and Boschan, C. (1971). *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, NBER, New York.

Burns, A. and Mitchell, W. (1946). *Measuring Business Cycles*, NBER, New York.

Camacho, M. and Perez-Quiros, G. (2000). That is what the US leading indicators lead, *European Central Bank, Working Paper* **27**.

Campbell, J. Y. (1995). Some lessons from the yield curve, *Journal of Economic Perspectives* **9**: 129–52.

Canova, F. (1998). Detrending and business cycle facts, *Journal of Monetary Economics* **41**: 475–512.

Canova, F. (1999). Does detrending matter for the determination of the reference cycle and the selection of turning points?, *Economic Journal* **109**: 126–50.

Chauvet, M. (1998). An econometric characterization of business cycle dynamics with factor structure and regime switching, *International Economic Review* **39**: 969–96.

Chong, Y. Y. and Hendry, D. F. (1986). Econometric evaluation in linear macro-economic models, *Review of Economic Studies* **53**: 671–90.

Christoffersen, P. F. (1998). Evaluating interval forecasts, *International Economic Review* **39**: 841–62.

Christoffersen, P. F. and Diebold, F. X. (1998). Cointegration and long-horizon forecasting, *Journal of Business and Economic Statistics* **16**: 450–58.

Clarida, R., Gali, J. and Gertler, M. (2000). Monetary policy rules and macroeconomic stability: Evidence and some theory, *Quaterly Journal of Economics* **115**: 147–80.

Clark, T. E. (1999). Finite-sample properties of tests for equal forecast accuracy, *Journal of Forecasting* **18**: 489–504.

Clark, T. E. and McCraken, M. W. (2000). Tests of equal forecast accuracy and encompassing for nested models, *Federal Reserve of Kansas City Working Paper (revised version)* **99-11**.

Clements, M. P., Franses, P. H., Smith, J. and Van Dijk, D. (2000). On SETAR non-linearity and forecasting, *Econometric Institute, Erasmus Rotterdam, Research Report* **1999-14/A**.

Clements, M. P. and Hendry, D. F. (1993). On the limitations of comparing mean squared forecast errors, *Journal of Forecasting* **12**: 617–37.

Clements, M. P. and Hendry, D. F. (1995). Forecasting in cointegrated systems, *Journal of Applied Econometrics* **10**: 127–46.

Clements, M. P. and Krolzig, H.-M. (1998). A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP, *Econometrics Journal* **1**: C47–C75.

Clements, M. P. and Krolzig, H.-M. (2000a). Business cycle asymmetries: Characterisation and testing based on Markov-switching autoregressions, *University of Warwick (mimeo)*

.

Clements, M. P. and Krolzig, H.-M. (2000b). Modelling business cycle features using switching regime models, *University of Warwick (mimeo)* .

Clements, M. P. and Smith, J. (1997). The performance of alternative forecasting methods for SETAR models, *International Journal of Forecasting* **13**: 463–75.

Clements, M. P. and Smith, J. (1999a). A Monte Carlo study of the forecasting performance of empirical SETAR models, *Journal of Applied Econometrics* **14**: 123–41.

Clements, M. P. and Smith, J. (1999b). Testing self-exciting threshold autoregressive models against structural change models, *University of Warwick (mimeo)* .

Clements, M. P. and Smith, J. (2000). Evaluating the forecast densities of linear and non-linear models: Applications to output growth and unemployment, *Journal of Forecasting* **19**: 255–76.

Cleveland, W. S., Grosse, E. and Shyu, W. M. (1993). Local regression models, *in* J. M. Chambers and T. J. Hastie (eds), *Statistical Models in S*, Chapman and Hall, London, pp. 309–76.

Corradi, V., Swanson, N. R. and Olivetti, C. (1999). Predictive ability with cointegrated variables, *Texas AM University (mimeo)* .

Corradi, V., Swanson, N. R. and White, H. (2000). Testing for stationary-ergodicity and for comovements between nonlinear discrete time Markov processes, *Journal of Econometrics* **96**: 39–73.

Croushore, D. (1993). Introducing: The Survey of Professional Forecasters, *Federal Reserve Bank of Philadelphia, Business Review* **nov./dec.**: 3–13.

DeLong, J. B. and Summers, L. H. (1986). Are business cycles symmetrical?, *in* R. J. Gordon (ed.), *The American Business Cycle: Continuity and Change*, NBER, Chicago, pp. 166–79.

DeLong, J. B. and Summers, L. H. (1988). How does macroeconomic policy affect output?, *Brooking Papers on Economic Activity* **2**: 433–94.

Diebold, F. X., Gunther, T. A. and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management, *International Economic Review* **39**: 863–83.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy, *Journal of Business and Economic Statistics* **13**: 253–65.

Diebold, F. X. and Rudebusch, G. D. (1989). Scoring the leading indicators, *Journal of Business* **62**: 369–91.

Diebold, F. X., Tay, A. S. and Wallis, K. F. (1999). Evaluating density forecasts of inflation: The survey of professional forecasters, *in* R. F. Engle and H. White (eds), *Cointegration, Causality and Forecasting: A Fetschrift in Honour of Clive W. J. Granger*, Oxford University Press, Oxford, pp. 76–90.

Don't Mention the R-Word: Are Economic Forecasters Wishful Thinkers or Wimps? (2001). *The Economist, printed edition* **March 1**.

Dotsey, M. (1998). The predictive content of the interest rate term spread for future economic growth, *Federal Reserve of Richmond, Economic Quartely* **84**: 31–51.

Dueker, M. J. (1997). Strengthening the case for the yield curve as a predictor of US recessions, *Federal Reserve Bank of St. Louis Review* **mar./apr.**: 41–51.

Durland, J. M. and McCurdy, T. H. (1994). Duration-dependent transitions in a Markov model of US GNP growth, *Journal of Business and Economic Statistics* **12**: 279–88.

Dwyer, G. P., Locke, P. and Yu, W. (1996). Index arbitrage and nonlinear dynamics between the SP500 futures and cash, *Review of Financial Studies* **9**: 301–32.

Eliasson, A. (1999). Is the short-run Phillips curve nonlinear? Empirical evidence for Australia, Sweden and the United States, *SSE/EFI Working Papers in Economics and Finance* **330**.

Elwood, S. K. (1998). Is the persistence of shocks to output asymmetric?, *Journal of Monetary Economics* **41**: 411–26.

Enders, W. and Granger, C. W. J. (1998). Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates, *Journal of Business and Economics Statistics* **16**: 304–11.

Enders, W. and Siklos, P. L. (2001). Cointegration and threshold adjustment, *Journal of Business and Economic Statistics* **19**: 166–176.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing, *Econometrica* **55**: 251–76.

Escribano, A. and Granger, C. W. J. (1998). Investigating the relationship between gold and silver prices, *Journal of Forecasting* **17**: 81–107.

Escribano, A. and Pfann, G. (1998). Non-linear error correction, asymmetric adjustment and cointegration, *Economic Modelling* **15**: 197–216.

Estrella, A. and Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity, *Journal of Finance* **46**: 555–76.

Estrella, A. and Mishkin, F. S. (1997). The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank, *European Economic Review* **41**: 1375–401.

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

Estrella, A., Rodrigues, A. P. and Schich, S. (2000). How stable is the predictive power of the yield curve? Evidence from Germany and the United States, *Federal Reserve of New York, Staff Report* **113**.

Eudey, G. and Perli, R. (1999). Regime-switching in expectations over the business cycle, *Federal Reserve Bank of Philadelphia, Working Paper* **99-17**.

Fair, R. C. (1993). Estimating event probabilities from macroeconometric models using stochastic simulation, *in* J. H. Stock and M. W. Watson (eds), *Business Cycles, Indicators, and Forecasting*, NBER, University of Chicago Press, Chicago, pp. 157–78.

Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems, *Biometrika* **83**: 189–206.

Fauvel, Y., Paquet, A. and Zimmermann, C. (1999). A survey on interest rate forecasting, *Center for Research on Economic Fluctuations and Employment, Université du Québec à Montréal, Cahier de Recherche* **87**.

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics, *Journal of Business and Economic Statistics* **12**: 299–308.

Filardo, A. J. (1999). How reliable are recession prediction models?, *Federal Reserve Bank of Kansas City, Economic Review* **second quarter**: 35–55.

Fintzen, D. and Stekler, H. O. (1999). Why did forecasters fail to predict the 1990 recession?, *International Journal of Forecasting* **15**: 309–23.

Franses, P. H. and Van Dijk, D. (2000). *Non-Linear Time Series Models in Empirical Finance*, Cambridge University Press, Cambridge.

Friedman, M. (1969). *The Optimum Quantity of Money and Other Essays*, Macmillan, London, chapter The Monetary Studies of the National Bureau, pp. 261–84.

Fuhrer, J. C. (1996). Monetary policy shifts and long-term interest rates, *Quarterly Journal of Economics* **111**: 1183–1209.

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

Gallegati, M. and Mignaca, D. (1995). Nonlinearities in business cycle: SETAR models and G7 industrial production data, *Applied Economics Letters* **2**: 422–27.

Garratt, A., Lee, K., Pesaran, M. H. and Shin, Y. (2000). Forecast uncertainties in macroeconometric modelling: An application to the UK economy, *Trinity College, Cambridge (mimeo)* .

Goodwin, T. H. (1993). Business-cycle analysis with a Markov-switching models, *Journal of Business and Economic Statistics* **11**: 331–39.

Goodwin, T. H. and Sweeney, R. J. (1993). International evidence on Friedman's theory of the business cycle, *Economic Inquiry* **31**: 178–93.

Granger, C. W. J. (1999). *Empirical Modeling in Economics: Specification and Evaluation*, Cambridge University Press, Cambridge.

Granger, C. W. J. (2001). An overview of nonlinear macroeconometric empirical models, *Macroeconomic Dynamics, forthcoming* .

Granger, C. W. J. and Pesaran, M. H. (2000). Economic and statistical measures of forecast accuracy, *Journal of Forecasting* **19**: 537–60.

Granger, C. W. J. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics* **42**: 27–62.

Greenaway, D., Leybourne, S. and Sapsford, D. (1997). Modeling growth (and liberalization) using smooth transitions analysis, *Economic Inquiry* **35**: 798–814.

Guay, A. and Scaillet, O. (1999). Indirect inference, nuisance parameter and threshold moving average, *Centre for Research on Economic Fluctuations and Employment, Université du Québec à Montréal, Cahier de Recherche* **95**.

Hall, A. D., Anderson, H. M. and Granger, C. W. J. (1992). A cointegration analysis of Treasury bill yields, *Review of Economics and Statistics* **74**: 116–26.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* **57**: 357–84.

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

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton, NJ.

Hamilton, J. D. (2001). A parametric approach to flexible nonlinear inference, *Econometrica* **69**: 537–73.

Hamilton, J. D. and Kim, D. H. (2000). A re-examination of the predictability of economic activity using the yield spread, *NBER Working Papers* **7954**.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* **64**: 413–30.

Hansen, B. E. (1997a). Approximate asymptotic p-values for structural-change tests, *Journal of Business and Economic Statistics* **15**: 60–67.

Hansen, B. E. (1997b). Inference in TAR models, *Studies in Nonlinear Dynamics and Econometrics* **2**: 1–14.

Hansen, B. E. (2000a). Testing for linearity, *in* D. A. R. George, L. Oxley and S. Potter (eds), *Surveys in Economic Dynamics*, Blackwell, Oxford, pp. 47–72.

Hansen, B. E. (2000b). Testing for structural change in conditional models, *Journal of Econometrics* **97**: 93–115.

Hansen, B. E. (2001). The new econometrics of structural change: Understanding and dating changes in US productivity, *University of Wisconsin (mimeo)* .

Hansen, B. E. and Seo, B. (2000). Testing for threshold cointegration, *University of Wisconsin (mimeo)* .

Harding, D. and Pagan, A. R. (2000). Knowing the cycle, *in* R. Backhouse and A. Salanti (eds), *Macroeconomics and the Real World*, Oxford University Press, Oxford, pp. 23–41.

Harding, D. and Pagan, A. R. (2001a). A comparison of two business cycle dating methods, *University of Melbourne (mimeo)* .

Harding, D. and Pagan, A. R. (2001b). Dissecting the cycle: A methodological investigation, *Journal of Monetary Economics, forthcoming* .

Hardouvelis, G. A. (1994). The term structure spread and future changes in long and short rates in the G7 countries, *Journal of Monetary Economics* **33**: 255–83.

Harvey, D. I., Leybourne, S. J. and Newbold, P. (1997). Testing the equality of prediction mean squared errors, *International Journal of Forecasting* **13**: 281–91.

Harvey, D. I., Leybourne, S. J. and Newbold, P. (1998). Tests for forecast encompassing, *Journal of Business and Economic Statistics* **16**: 254–59.

Hastie, T. and Loader, C. (1993). Local regression: Automatic kernel carpentry, *Statistical Science* **8**: 120–43.

Haubrich, J. G. and Dombrosky, A. M. (1996). Predicting real growth using the yield curve, *Federal Reserve Bank of Cleveland, Economic Review* **32**: 26–34.

Hendry, D. F. (1995). *Dynamic Econometrics*, Advanced Texts in Econometrics, Oxford University Press, Oxford.

Hess, G. D. and Iwata, S. (1997a). Asymmetric persitence in GDP? A deeper look at depth, *Journal of Monetary Economics* **40**: 535–54.

Hess, G. D. and Iwata, S. (1997b). Measuring and comparing business-cycle features, *Journal of Business and Economic Statistics* **15**: 432–44.

Hicks, J. R. (1950). *A Contribution to the Theory of the Trade Cycle*, Clarendon Press, Oxford.

Holly, S. and Stannett, M. (1995). Are there asymmetries in UK consumption? A time series analysis, *Applied Economics* **27**: 767–72.

Intriligator, M. D. (1983). Economic and econometric models, *in* Z. Griliches and M. D. Intriligator (eds), *Handbook of Econometrics, Vol. 1*, Elsevier, Amsterdam, pp. 181–222.

Ivanova, D., Lahiri, K. and Seitz, F. (2000). Interest rate spreads as predictors of German inflation and business cycles, *International Journal of Forecasting* **16**: 39–58.

Jansen, D. W. and Oh, W. (1999). Modeling nonlinearity of business cycles: Choosing between the CDR and the STAR models, *Review of Economics and Statistics* **81**: 344–49.

Johansen, S. (1988). Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* **12**: 231–54.

Keilis-Borok, V., Stock, J. H., Soloviev, A. and Mikhalev, P. (2000). Pre-recession pattern of six economic indicators in the USA, *Journal of Forecasting* **19**: 65–80.

Kim, C.-J. and Murray, C. J. (2001). Permanent and transitory components of recessions, *Empirical Economics, forthcoming* .

Kim, C.-J. and Nelson, C. R. (1999a). Friedman's plucking model of business fluctuations: Tests and estimates of permanent and transitory components, *Journal of Money, Credit and Banking* **31**: 317–34.

Kim, C.-J. and Nelson, C. R. (1999b). Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle, *Review of Economics and Statistics* **81**: 608–16.

Kim, C.-J. and Nelson, C. R. (1999c). *State-Space Model with Regime Switching*, MIT Press, Cambridge.

King, R. and Plosser, C. (1994). Real business cycles and the test of the Adelmans, *Journal of Monetary Economics* **33**: 405–38.

Koop, G. and Potter, S. M. (1999a). Bayes factors and nonlinearity: Evidence from economic time series, *Journal of Econometrics* **88**: 251–81.

Koop, G. and Potter, S. M. (1999b). Dynamic asymmetries in US unemployment, *Journal of Business and Economic Statistics* **17**: 298–312.

Koop, G. and Potter, S. M. (2000). Nonlinearity, structural breaks, or outliers in economic time series?, *in* W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjostheim and A. Würtz (eds), *Nonlinear Econometric Modeling in Time Series*, Cambridge University Press, Cambridge, pp. 61–78.

Koop, G. and Potter, S. M. (2001). Are apparent findings of nonlinearity due to structural instability in economic time series?, *Econometrics Journal* **4**: 37–55.

Koskinen, L. and Öller, L.-E. (2001). A classifying procedure for signaling turning points, *SSE/SFI Working Paper Series in Economics and Finance* **427**.

Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*, Springer-Verlag, Berlin.

Krolzig, H.-M. and Toro, J. (1999). A new approach to the analysis of shocks and the cycle in a model of output and employment, *European Institute, EUI Working Papers in Economics* **99-30**.

Lahiri, K. and Wang, J. G. (1996). Interest rate spreads as predictors of business cycles, *in*

G. S. Maddala and C. R. Rao (eds), *Handbook of Statistics, Vol. 14*, Elsevier, Amsterdam, pp. 297–315.

Layton, A. P. (1994). Further on the nature of the Australian business cycle, *The Economic Record* **70**: 12–18.

Li, C. W. and Li, W. K. (1996). On a double-threshold autoregressive heteroscedastic time series model, *Journal of Applied Econometrics* **11**: 253–74.

Lin, C.-F. J. and Teräsvirta, T. (1994). Testing the constancy of regression parameters against continuous structural change, *Journal of Econometrics* **62**: 211–28.

Lucas, R. E. (1976). Econometric policy evaluations: A critique, *in* K. Brunner and A. H. Meltzer (eds), *The Phillips Curve and Labor Markets*, North Holland, Amsterdam, pp. 19–46.

Luginbuhl, R. and De Vos, A. (1999). Bayesian analysis of an unobserved-component time series model of GDP with Markov-switching and time-varying growths, *Journal of Business and Economic Statistics* **17**: 456–65.

Lundbergh, S. and Teräsvirta, T. (1998). Modelling economic high-frequency time series with STAR-GARCH models, *SSE/EFI Working Paper Series in Economics and Finance* **291**.

Lundbergh, S. and Teräsvirta, T. (2000). Forecasting with smooth transition autoregressive models, *SSE/EFI Working Papers in Economics and Finance* **390**.

Lundbergh, S., Teräsvirta, T. and Van Dijk, D. (2000). Time-varying smooth transition autoregressive models, *SSE/EFI Working Papers in Economics and Finance* **376**.

Malkiel, B. G. (1970). *The Term Structure of Interest Rates: Theory, Empirical Evidence and Applications*, General Learning Corporation, Morristown, NJ.

Mankiw, N. G. and Miron, J. A. (1986). The changing behavior of the term structure of interest rates, *Quaterly Journal of Economics* **101**: 211–228.

Marshall, D. A. (2001). The crisis of 1998 and the role of the Central Bank, *Federal Reserve Bank of Chicago, Economic Perspectives* **First Quarter**: 2–23.

Martens, M., Kofman, P. and Vorst, T. C. F. (1998). A threshold error-correction model for intraday futures and index returns, *Journal of Applied Econometrics* **13**: 245–63.

McConnell, M. M. and Perez-Quiros, G. (2000). Output fluctuations in the United States: What has changed since early 1980s?, *American Economic Review* **90**: 1464–76.

McCulloch, J. H. and Kwon, H. C. (1993). US term structure data, 1947-1991, *Ohio State University, Working Paper.* **93-6**.

McCulloch, R. E. and Tsay, R. S. (1994). Statistical analysis of economic time series via Markov switching models, *Journal of Time Series Analysis* **15**: 523–39.

McQueen, G. and Thorley, S. (1993). Asymmetric business cycle turning points, *Journal of Monetary Economics* **31**: 341–62.

Mejía-Reyes, P. (2000). Asymmetries and common cycles in Latin America: Evidence from Markov-switching models, *Economia Mexicana* **9**: 189–225.

Mills, T. C. (1995). Are there asymmetries or nonlinearities in UK output?, *Applied Economics* **27**: 1211–17.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S. and Tiao, G. C. (1998). Forecasting the US unemployment rate, *Journal of American Statistical Association* **93**: 478–93.

Neftci, S. N. (1982). Optimal prediction of cyclical downturns., *Journal of Economic Dynamics and Control* **4**: 225–41.

Paap, R. and Franses, P. H. (1999). Do the US and Canada have a common nonlinear cycle in unemployment?, *Econometric Institute, Erasmus Rotterdam, Research Report* **1999-07/A**.

Pagan, A. R. (1997). Policy, theory, and the cycle, *Oxford Review of Economic Policy* **13**: 19–33.

Pagan, A. R. (1999). The getting of macroeconomic wisdom, *Australian National University (mimeo)* .

Pagan, A. R., Hall, A. D. and Martin, V. (1996). Modeling the term structure, *in* G. S. Maddala and C. R. Rao (eds), *Handbook of Statistics 14*, North-Holland, Amsterdam, pp. 91–118.

Pagan, A. R. and Robertson, J. (2001). Forecasting for policy, *in* M. P. Clements and D. Hendry (eds), *A Companion to Economic Forecasting, Forthcoming*, Blackwell.

Pagan, A. R. and Ullah, A. (1999). *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

Parker, R. E. and Rothman, P. (1998). The current depth of recession and unemployment rate forecasts, *Studies in Nonlinear Dynamics and Econometrics* **2**: 151–58.

Peel, D. and Speight, A. E. H. (1998a). The nonlinear time series properties of unemployment rates: Some further evidence, *Applied Economics* **30**: 287–94.

Peel, D. and Speight, A. E. H. (1998b). Threshold nonlinearities in output: Some international evidence, *Applied Economics* **30**: 323–33.

Pesaran, M. H. and Potter, S. M. (1997). A floor and ceiling model of US output, *Journal of Economic Dynamics and Control* **21**: 661–95.

Pesaran, M. H. and Skouras, S. (2001). Decision-based methods for forecast evaluation, *in* M. P. Clements and D. F. Hendry (eds), *A Companion to Economic Forecasting, Forthcoming*, Blackwell.

Pesaran, M. H. and Timmermann, A. G. (1992). A simple nonparametric test of predictive performance, *Journal of Business and Economic Statistics* **10**: 461–65.

Pfann, G. A., Schotman, P. C. and Tschernig, R. (1996). Nonlinear interest rate dynamics and implications for the term structure, *Journal of Econometrics* **74**: 149–76.

Pfann, G. and Palm, F. (1993). Asymmetric adjustment costs in non-linear labour demand models for the Netherlands and UK manufacturing sectors, *Review of Economic Studies* **60**: 397–412.

Potter, S. M. (1995). A nonlinear approach to US GNP, *Journal of Applied Econometrics* **10**: 109–25.

Potter, S. M. (2000). Nonlinear time series modelling: An introduction, *in* D. A. R. George, L. Oxley and S. Potter (eds), *Surveys in Economic Dynamics*, Blackwell, Oxford, pp. 1–24.

Proietti, T. (1998). Characterizing asymmetries in business cycles using smooth-transition structural time-series models, *Studies in Nonlinear Dynamics and Econometrics* **3**: 141–56.

Ramsey, J. B. (1996). If nonlinear models cannot forecast, what use are they?, *Studies in Nonlinear Dynamics and Econometrics* **1**: 65–86.

Roberds, W., Runkle, D. and Whiteman, C. H. (1996). A daily view of yield spreads and short-term interest rate movements, *Journal of Money, Credit and Banking* **28**: 34–53.

Rothman, P. (1998). Forecasting asymmetric unemployment rates, *Review of Economics and Statistics* **80**: 164–168.

Rothman, P., Van Dijk, D. and Franses, P. H. (1999). A multivariate STAR analysis of the relationship between money and output, *Econometric Institute, Erasmus Rotterdam, Research Report* **1999-45/A**.

Rudebusch, G. D. (1995). Federal Reserve interest rate targeting, rational expectations, and the term structure, *Journal of Monetary Economics* **35**: 245–74.

Sensier, M. and Van Dijk, D. (2001). Short-term volatility versus long-term growth: Evidence in US macroeconomic time series, *Centre for Growth and Business Cycle Research, University of Manchester, Discussion Paper* **8**.

Sichel, D. E. (1993). Business cycle asymmetry: A deeper look, *Economic Inquiry* **31**: 224–236.

Sichel, D. E. (1994). Inventories and the three phases of the business cycle, *Journal of Business and Economic Statistics* **12**: 270–77.

Simkins, S. P. (1994). Do Real Business Cycle models really exhibit business cycle behaviour?, *Journal of Monetary Economics* **33**: 381–404.

Simonoff, J. S. (1996). *Smoothing Methods in Statistics*, Springer-Verlag, New York.

Skalin, J. and Teräsvirta, T. (1998). Modelling asymmetries and moving equilibria in unemployment rates, *SSE/EFI Working Paper Series in Economics and Finance,* **262**.

Sola, M. and Driffill, J. (1994). Testing the term structure of interest rates using a stationary vector autoregression with regime switching, *Journal of Economic Dynamics and Control* **18**: 601–28.

Speight, A. E. H. (1997). Testing for business cycle asymmetry: Some international evidence, *Applied Economics Letters* **4**: 603–6.

Speight, A. E. H. and McMillan, D. G. (1998). Testing for asymmetries in UK macroeconomic time series, *Scottish Journal of Political Economy* **45**: 158–170.

Stanca, L. (1999). Asymmetries and nonlinearities in Italian macroeconomic fluctuations, *Applied Economics* **31**: 483–91.

Stock, J. H. and Watson, M. W. (1989). New indexes of coincident and leading economic indicators, *NBER Macroeconomics Annual* **4**: 351–94.

Stock, J. H. and Watson, M. W. (1993). A procedure for predicting recessions with leading indicators: Econometric issues and recent experience, *in* J. H. Stock and M. W. Watson (eds), *Business Cycles, Indicators, and Forecasting*, NBER, University of Chicago Press, Chicago, pp. 95–156.

Stock, J. H. and Watson, M. W. (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, *in* R. F. Engle and H. White (eds), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive Granger*, Oxford University Press, Oxford, pp. 1–44.

Stock, J. H. and Watson, M. W. (2001). Forecasting output and inflation: The role of asset prices, *NBER Working Paper* **8180**.

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

Tay, A. S. and Wallis, K. F. (2000). Density forecasting: A survey, *Journal of Forecasting* **19**: 235–54.

Teräsvirta, T. (1998). Modeling economic relationships with smooth transition regressions, *in* A. Ullah and E. David (eds), *Handbook of Applied Economic Statistics*, Marcel Dekker, New York, pp. 507–52.

Teräsvirta, T. and Anderson, H. M. (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models, *Journal of Applied Econometrics* **7**: S119–S136.

Teräsvirta, T., Tjostheim, D. and Granger, C. W. J. (1994). Aspects of modelling nonlinear time series models, *in* R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics, Vol. 4*, Elsevier, Amsterdam, pp. 2917–57.

Tiao, G. C. and Tsay, R. S. (1994). Some advances in non-linear and adaptive modelling in time series, *Journal of Forecasting* **13**: 109–31.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*, Oxford University Press, Oxford.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes, *Journal of the American Statistical Association* **84**: 231–40.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models, *Journal of American Statistical Association* **93**: 1188–1202.

Tzavalis, E. and Wickens, M. (1998). A re-examination of the rational expectations hypothesis of the term structure: Reconciling the evidence from long-run and short-run tests, *International Journal of Finance and Economics* **3**: 229–39.

Van Dijk, D. (1999). *Smooth Transition Models: Extensions and Outlier Robust Inference*, PhD thesis, Tinbergen Institute.

Van Dijk, D., Franses, P. and Boswijk, P. (2000). Asymmetric and common absorption of shocks in nonlinear autoregressive models, *Econometric Institute, Erasmus Rotterdam, Research Report* **2000-01/A**.

Van Dijk, D. and Franses, P. H. (1999). Modeling multiple regimes in the business cycle, *Macroeconomic Dynamics* **3**: 311–40.

Van Dijk, D. and Franses, P. H. (2000). Nonlinear error-correction models for interest rates in the Netherlands, *in* W. A. Barnett, D. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjostheim and A. Wurtz (eds), *Nonlinear Econometric Modelling in Time Series: Proceedings of the Eleventh International Symposium in Economic Theory*, Cambridge University Press, Cambridge, pp. 203–27.

Van Dijk, D., Strikholm, B. and Teräsvirta, T. (2001). The effects of institutional and technological change and business cycle fluctuations on seasonal patterns in quarterly industrial production series, *SSE/EFI Working Paper Series in Economics and Finance* **429**.

Van Dijk, D., Teräsvirta, T. and Franses, P. H. (2001). Smooth transition autoregressive models - A survey of recent developments, *Econometric Reviews, forthcoming* .

Verbrugge, R. (1997). Investigating cyclical asymmetries, *Studies in Nonlinear Dynamics and Econometrics* **2**: 15–22.

Watson, M. W. (1994). Business-cycle durations and postwar stabilization of the US economy, *American Economic Review* **84**: 24–46.

Watson, M. W. (1999). Explaining the increased variability in long-term interest rates, *Federal Reserve Bank of Richmond, Economic Quaterly* **85**: 71–96.

Weise, C. L. (1999). The asymmetric effects of monetary policy: A nonlinear vector autoregression approach, *Journal of Money, Credit and Banking* **31**: 85–108.

West, K. D. (1996). Asymptotic inference about predictive ability, *Econometrica* **64**: 1067–1084.

West, K. D. and McCracken, M. W. (1998). Regression-based tests of predictive ability, *International Economic Review* **39**: 817–840.