Exchange Rate Forecasts and Stochastic Trend Breaks

David V. O'Toole

A Thesis submitted for the degree of Doctor of Philosophy

School of Finance and Economics

University of Technology, Sydney

2009

Declaration of Originality

I certify that the work in this thesis has not previously been submitted for another degree nor has it been submitted as part of the requirements for another degree.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Acknowledgements

First I would like to thank Professor Tony Hall as principal supervisor of this thesis. Apart from helping me get funding for this thesis, Tony allowed me the freedom to pursue my interests and provided constructive criticism and advice throughout the research. If Tony expressed doubst about something I had written, then it almost always meant there was more work to be done. I am especially grateful for Tony's guidance at the start of my research - when I told him that I was interested in exchange rate forecasting with cointegrated systems, Tony told me to read Clements & Hendry's two books on forecasting and Pesaran & Timmerman's work on forecasting in the prescence of structural breaks. He also lent me his copy of Halbert White's 'Asymptotic Theory for Econometricians' and pinpointed Peter Phillips' research papers. Looking back now I can see that this early advice has strongly affected my research focus for the better.

I am also grateful to my co-supervisors, Dr Susan Thorpe and Dr Gordon Menzies. I always took heed of Susan's insightful comments about my work, especially after my presentations at the UTS doctoral seminars. Gordon provided valuable advice, early on in the doctorate, to make a clear distinction between my research on forecasting, and the forward discount puzzle. (I followed his suggestion to leave the forward discount puzzle for later!)

I acknowledge funding for my research from the UTS Faculty of Business, without which this thesis would not exist.

UTS also funded my travel to a New Zealand Econometric Study Group conference in Dunedin at which my paper was discussed by P.C.B.Phillips. His suggestions considerably improved what is now Chapter 4.

I presented an early version of Chapter 2 at a UTS doctoral seminar, after which Steve Satchell gave me very valuable feedback. He also pointed out the Dacco & Satchell paper on which Chapter 2 now draws heavily.

I had the opportunity to show some of my work to Rob Engle (who came to Sydney for the Nobel Prize exhibition at UTS). His comments and suggestions resulted in several improvements to Chapter 2 and 3.

It is a pleasure for me to thank my wife Laura and children - Katherine, Olivia and Nicholas (who was born during the doctorate) - for their love and unfailing support. This thesis is dedicated to you. I also would like to thank Mum, Dad, Sandy, Nina, Lara, Greg, Nat, Sian, Jorgie, Iain and Diane.

Contents

Abstract

This thesis examines the forecastability of exchange rates in the presence of trend breaks. In particular, its focus is the predictive power of the interest rate differential for the exchange rate.

Chapter 1 is the Introduction to the thesis. In this Chapter, I briefly review the relevant literature on exchange rate predictability, forecasting in the presence of structural breaks and modelling trends in exchange rate time series.

Models are often evaluated via their out-of-sample forecasts over a single out-of-sample period. However, not all out-of-sample (OOS) periods are of equal difficulty - poor forecast performance of a model over a certain OOS period might actually be evidence in favour of the model if the OOS period was particularly difficult. In Chapter 2, I develop a way to quantify the difficulty of an OOS period affected by trend breaks. This method uses the deficit between the mean square forecast error of the optimal univariate forecast of the trend breaking process and the random walk forecast. This MSFE deficit is what needs to be made up by any extra information in a model in order to beat the random walk. In Chapter 2, I use the degree of difficulty measure in an ex-post analysis of the forecasts of a VEqCM over two separate periods.

Chapter 3 shows that when an out-of-sample period has trend breaks, the forecast error densities generated from a recursive forecasting procedure can have a spurious multimodality at various horizons. This is clearly problematic for any statistics calculated from these densities - in particular, any forecast evaluation statistics or tests. It would also produce misleading value-at-risk calculations. I develop a limit theory explaining why this occurs. In the second half of the Chapter I show how the forecast error density can be disentangled from the trend breaks. This allows an estimate of the extent to which breaks have affected a particular forecast statistic.

In Chapter 4 I use a general trend representation (from the work of P.C.B.Phillips) to model inter-break exchange rate behaviour. I show how this broken trend representation can be used to estimate the trend breaks in an exchange rate series. I show that with the general trend representation, only 'large ' breaks are identified - i.e., small (potentially spurious) breaks can be modelled with an unbroken trend. In an empirical application to the AUD-GBP exchange rate, I find that the estimated breaks can be rationalized using some recent theory on the effect of monetary policy shocks on exchange rate trends.

Chapter 5 concludes the thesis.

Chapter 1

Exchange Rate Prediction and Trend Breaks

1.1 Introduction

Exchange rates are important economic and financial variables. The large currency depreciations that affected South-East Asian countries in the Asian Financial Crisis were unexpected and highly damaging. At the moment there is great interest in whether the US dollar will depreciate and/or lose its status as the international reserve currency. For investors, exchange rate fluctuations can make up a large proportion of a foreign investment return - a risk that is difficult to hedge over long investment horizons, and especially important for investors in places like Australia who have to hold the bulk of their assets in overseas markets in order to fully diversify. Sadly, exchange rates have proven very difficult to forecast. Even in-sample, we don't understand exchange rate fluctuations. As Hodrick (1988) said - 'we do not yet have a model of expected [foreign exchange] returns that fits the data'. [Hodrick's statement was quoted by Engel in his 1996 review of the forward discount puzzle as the state of affairs then, and, to my knowledge, is still true today.] In out-of-sample testing, the situation is at least as bad - Meese & Rogoff (1983) found that macroeconomic variables have no predictive ability for exchange rates. As discussed below, later studies reach similar conclusions.

As a result of Meese & Rogoff (1983), some concluded that macroeconomic variables had no explanatory power for exchange rates (an 'exchange rate disconnect puzzle'). However, the work of Clements & Hendry (1998,

1999) showed that in a world of structural breaks, this conclusion can be a fallacy. They use the example of a scientist rejecting Newton's Laws (a model) after failing to predict the trajectory of a spacecraft that is hit by an asteroid (structural break). My underlying aim in this thesis was to examine the extent to which the exchange rate disconnect puzzle might be an instance of Clements & Hendry's fallacy. As a first step I wanted to determine what impact structural breaks can have on the sort of exchange rate forecasting studies conducted by Meese & Rogoff and others. This work resulted in the first two pieces of research in the thesis.

I focussed on a single exchange rate model - one using the forward premium, and a single type of break - a trend break. The results of the first two chapters are likely to apply to a wider range of models though. The reason for choosing the forward premium is that models using this variable have had some reported success, particularly at long-horizons. The choice of trend break was dictated by the 'long swings' model of Engel & Hamilton (1990) and more recent work reported below. The third research chapter is an attempt to formulate a more general trend break model than the long swings model.

Each Chapter of the thesis begins with a review of the literature relevant to that Chapter, and these are not repeated here. My aim in this introductory Chapter is to delineate my work. In particular, I need to emphasize that, while the thesis uses models that relate the exchange rate return to the forward premium (or interest differential), my work does not impinge on the vast forward discount puzzle or uncovered interest parity literatures. In the next Section I briefly discuss the forward discount puzzle in order to explain how my work differs from it.

1.2 FX Return Predictability and the Meese-Rogoff Puzzle

Foreign exchange return predictability (via the interest differential) has been identified by Cochrane (2005) as one of the 'New facts in finance'. The starting point of the literature is the regression,

$$\Delta s_t = \alpha + \beta(i_t - i_t^\star) + \epsilon_t, \tag{1.1}$$

where s_t is the logarithm of the spot exchange rate (domestic price of foreign currency), and $i_t^\star$ and i_t are the one-period foreign and domestic interest rates. Typically β is found to be negative and significant, and the R^2, while small ($2\% - 3\%$), is non-zero. The negative β suggests the following trading strategy: if foreign bond yields are higher than domestic, buy foreign bonds. Then (1.1) says that the foreign currency will appreciate, further adding to the extra yield difference $i_t^\star - i_t$. Cochrane (2005) provides some reasons why trading on regressions like (1.1) might not work in reality. However, it does imply that foreign exchange returns are predictable on the basis of the interest differential.

There is a close connection between (1.1) and the forward discount puzzle. Under rational expectations, the uncovered interest parity condition (UIP) holds,

$$E_t[\Delta s_t] = i_t - i_t^\star, \tag{1.2}$$

An interpretation of the UIP condition (1.2) is that a foreign bond issuer must offer a yield $i_t^\star$ that equals the rate investors can achieve domestically i_t plus the expected depreciation of the foreign currency (a negative Δs_t is a depreciation of the foreign currency). The condition (1.2) implies that $\alpha = 0$ and $\beta = 1$ in regression (1.1). The negative β has spawned a huge literature known as the 'forward discount puzzle'. Attempts to explain the negative β include time-varying risk premia and systematic expectational errors. Hodrick (1988), Engel (1996) and Sarno & Taylor (2002) survey this vast literature. I emphasize that this thesis does not attempt to explain the negative β.

Clarida & Taylor (1997) provide a theory (see Appendix A of Chapter 2) that suggests that a vector equilibrium correction model with the forward premia of various maturities (or equivalently the interest differentials) should have predictive power for the exchange rate change. Their empirical study found considerable predictive power over a period in the early 1990's. This was one of the few studies to find that exchange rates could be forecasted since Meese & Rogoff (1983). Meese & Rogoff (1983) found that macroeconomic variables (including the interest differential) had no predictive power (relative to the random walk) in out-of-sample testing. The conjecture that Meese & Rogoff had failed to exploit nonlinear relations between exchange rates and fundamentals was refuted by Diebold & Nason (1990). Cheung et al., (2005), found some exchange rate predictability but it varied with exchange rate, model and time period. Despite Clarida & Taylor's (1997) result on the 1990-1993 period, a fair interpretation of the evidence is that exchange rate predictability is, at best, time varying.

1.3 Forecasting, Breaks and Exchange Rate Trends

In this Section I review some of the relevant work on the impact of structural breaks on forecast performance, and also the modelling of exchange rate trends.

1.3.1 Forecasting with Breaks

The two books, Clements & Hendry (1998) and (1999) discuss the impact of structural breaks on forecast performance. They derive a forecast error taxonomy and find that out-of-sample breaks are the main cause of poor forecasts. Hence one focus of their work is finding forecasting devices that are robust to breaks. Pesaran & Timmerman (2004, 2005) focus on the impact of in-sample breaks, the impact of such breaks on parameter estimates, and estimation methods to alleviate such breaks. Dacco & Satchell (1989) looked at forecasting a Markov switching process and found that if the probability of misclassifying the forecast regime is too large (i.e., the probability of a break is too large) then it is not possible to outperform the random walk forecast. I discuss the Dacco & Satchell paper in detail in Chapter 2.

1.3.2 Exchange Rate Trend Modelling

In this Subsection I briefly review some of the approaches to modelling trends in time series - particularly exchange rate trends. A concern of this work is whether stochastic trends or deterministic trends are more appropriate and the implications of these for the persistence of shocks. Recent work by Phillips (2005) has shown that this dichotomy may be artificial, as both models can be seen as alternative representations of the same process. P.C.B.Phillips' approach to modelling trends is reviewed in Chapter 4.

Unit root root testing generally finds that exchange rates are first-difference stationary (Meese & Singleton, 1982). They could, however, be trend-stationary around a suitably complex deterministic trend function - Perron (1989) found that other macroeconomic series thought to be unit root processes, were better modelled as trend stationary around a deterministic function with breaks at the Great Depression and the oil shock. Chapter 4 discusses the application of this idea to exchange rates.

Engel & Hamilton (1990) proposed a Markov switching trend process that models exchange rates well, at least for low (e.g., quarterly) frequencies. (This model is reviewed in Chapter 2.) Typically, the two-state model finds persistent states, one state in which the model has an upward drift, the other downward - Engel & Hamilton call this a 'long swings' process. Cheung & Erlandsson (2005) and Klaassen (2005) are recent examples of work that has found support for the 'long swings' model.

1.4 A Roadmap

As discussed above, this thesis is concerned with the impact of structural breaks on exchange rate forecasting models that use the interest rate differential. Chapter 2 adapts the work of Dacco & Satchell (1999) to show that stochastic trend breaks can cause different time periods to be easier or more difficult to forecast over. Chapter 3 studies the impact of trend breaks on forecast error densities. In this Chapter I develop a limit theory to show that the forecast error densities can be multimodal. In the third research Chapter, Chapter 4 I adapt the trend modelling approach of P.C.B.Phillips to formulate a general broken trend model that I use to model the AUD-GBP exchange rate. Chapter 5 concludes the thesis.

Bibliography

[1] CHEUNG,Y-W., CHINN,M.D. & PASCUAL,A.G. (2005). Empirical exchange rate models of the nineties: Are any fit to survive? *J.Int.Money.&.Finance.***24,** 1150-1175.

[2] CHEUNG,Y-W., & ERLANDSSON,U.G. (2005). Exchange rates and Markov switching dynamics. *J.Bus.Econ.Statistics* **23(3),** 314-320.

[3] CLARIDA,R.H. & TAYLOR,M.P. (1997). The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors. *Rev.Econ. & Stat.* **LXXIX,** 353.

[4] CLARK,T.E. & MCCRACKEN,M.W. (2005) The power of tests of predictive ability in the presence of structural breaks. *J.Econometrics.* **124,** 1-31.

[5] CLEMENTS,M.P. & HENDRY,D.F. (1998). Forecasting Economic Time Series. *C.U.P.*

[6] CLEMENTS,M.P. & HENDRY,D.F. (1999). Forecasting Non-Stationary Economic Time Series. *The MIT Press*

[7] COCHRANE,J.H. (2005). Asset pricing. 2nd ed. *Princeton.*

[8] DACCO,R. & SATCHELL,S. (1999). Why do regime-switching models forecast so badly? *J.Forecasting.* **18,** 1-16.

[9] DIEBOLD,F.X., & NASON,J.A. (1990). Nonparametric exchange rate prediction? *J.Int.Econ.* **28,** 315-332.

[10] ENGEL,C. (1996). The forward discount anomaly and the risk premium: a survey of recent evidence. *J.Empirical Finance.* **3,** 123-192.

[11] ENGEL, C. & HAMILTON,J.D. (1990). Long swings in the dollar: are they in the data and do markets know it? *American Economic Review.* **80,** 689-713.

[12] HODRICK,R. (1988) The empirical evidence on the efficiency of forward and futures foreign exchange markets. *Fundamentals of Pure and Applied Economics.* Harwood Academic. N.Y.

[13] KLAASSEN,F. (2005) Long swings in exchange rates: are they really in the data? *J.Bus.Econ.Statistics* **23(1),** 87-95.

[14] MEESE,R.A. & ROGOFF,K. (1983) Empirical Exchange Rate Models of the Seventies: Do They Fit Out Of Sample? *J.Int.Econ.* **14,** 3-24.

[15] MEESE,R.A. & SINGLETON,K.J. (1982) On unit roots and the empirical modelling of exchange rates. *J.Finance.* **37(4),** 1092-1035.

[16] PERRON,P. (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica.* **57,** 1361-1401.

[17] PESARAN,H. & TIMMERMAN,A. (2004) How costly is it to ignore breaks when forecasting the direction of a time series? *Int.J.Forecasting.* **20,** 411-425.

[18] PESARAN,H. & TIMMERMAN,A. (2005) Small sample properties of forecasts from autoregressive models under structural breaks. *J.Econometrics* **129,** 183-217.

[19] PHILLIPS,P.C.B. (2005b) Challenges of trending time series econometrics. *Mathematics and Computers in Simulation.* **68,** 401-416.

[20] SARNO,L. & TAYLOR,M.P. (2002). The Economics of Exchange Rates. Cambridge University Press.

[21] WEST,K.D. (2005). Forecast Evaluation. *in Handbook of Forecasting, Eds., ELLIOT,G., & TIMMERMAN,A.*

Part I

Research Chapters

Chapter 2

A Degree of Difficulty Measure for Exchange Rate Forecasts

2.1 Introduction

Forecasting models are often judged by their mean square forecast error over a selected out-of-sample period. However, all out-of-sample periods are not equal - some are harder to forecast over than others. The aim of this work is to develop a measure of the 'degree of difficulty' of a given out-of-sample period. This allows us to determine, ex-post, how impressed (or disappointed) we should be by a particular model's performance over that period. A model that forecasts well over an easy period may not be superior to a model that fails over a difficult period. In this Chapter I develop a degree of difficulty measure and apply it to study the time-varying predictive ability of an exchange rate model.

Meese & Rogoff (1983) began a literature known as the 'foreign exchange forecasting puzzle.' Their work showed that none of the models they considered could beat a random walk over their out-of-sample period. This was interpreted as evidence that macroeconomic variables (e.g., interest rate differentials) did not have explanatory power for exchange rates.

Later work by Clarida & Taylor (1997) and Clarida et al., (2003) showed that a vector equilibrium correction model (VEqCM), with the interest differentials at different maturities as predictors, could beat a random walk over their (later) forecast period. More recent work by Cheung et al., (2005) found that exchange rate models (including those using interest differentials) appeared to have time-varying predictive power - they work well in some periods and for some exchange rates, but not for others. This often seems to occur in forecast evaluation studies - a study that finds forecasting power of a certain model over a certain forecasting period is overturned by a later study that uses more recent data. Although, sometimes the reverse happens - a poor forecasting device is found to work well over later data.

One possible cause of this lack of robustness of forecast evaluations is that it is due to structural breaks in the forecast period. Clements & Hendry (1998, 1999) demonstrated that out-of-sample breaks are the main cause of poor out-of-sample forecast performance relative to in-sample fit. However, in a comparison of the predictive ability of two models, out-of-sample breaks will affect the forecasts of both models.

Dacco & Satchell (1999) derived a condition on the stochastic process generating structural breaks that determines when a model is capable of beating the random walk. In this work I have adapted their condition to provide a degree of difficulty measure for a given forecast period. This degree of difficulty measure is applied to exchange rate data which appears to have trend breaks, or 'long swings' - Engel & Hamilton (1990).

The work of Clark & McCracken (2005) has cast doubt on the usefulness of formal forecast evaluation tests in the presence of breaks. Clark & McCracken (2005) demonstrated that when there are structural breaks in the forecasting period, the standard tests for forecast comparison between two models have low power to reject the null hypothesis of equal forecasting power. This poses a problem for an investigator who cannot reject this null hypothesis. Is this because null hypothesis is true, or is it just that the structural breaks in the forecast period have contaminated the test? These unresolved issues are a problem for a forecast evaluation of exchange rates subject to trend breaks. I follow Meese & Rogoff (1983) and Clarida & Taylor (1997) who ignored sampling variability in evaluating their mean square forecast errors (MSFE's), i.e., a forecast having MSFE smaller than that of a random walk is taken as evidence of predictive ability. Clearly it would be good to incorporate the degree of difficulty into a formal testing framework (Diebold & Mariano, 1995).

An important bound on the limits of forecastability was derived by Phillips & Ploberger [Phillips & Ploberger (2003), Phillips (2003)]. This bound is related to the minimal proximity that an estimated model can achieve when the data generating process has different degrees of trending behaviour (Phillips, 1996). In this work I employ the Dacco & Satchell (1999) framework which assumes that the degree of trending is known (i.e., does not need to be estimated). The degree of difficulty measure I derive in that framework is probably a lower bound if the degree of trending had to be estimated. I.e., if a model can't beat a random walk when the trending degree is known, it certainly won't if it had to be estimated.

The Chapter is organized as follows. Section 2 develops the degree of difficulty measure. In Section 3 I study the forecasting ability of a bivariate VEqCM model over two non-overlapping out-of-sample periods. In Section 4 I calculate the degree of difficulty measure for the two periods. Section 5 is the Conclusion.

2.2 Degree of Forecast Difficulty

As discussed in the Introduction, some sample periods are easier to forecast than others. In this Section I develop a way of measuring the difficulty when the out-of-sample data is characterized by trend breaks. This measure is an adaptation of a condition derived by Dacco & Satchell (1999) on the forecastability of series subject to stochastic trend breaks. They showed that such series are unforecastable (on a mean square error basis) if the stochastic break process satisfies a certain condition. I briefly review their framework here.

2.2.1 The Dacco-Satchell Condition

Dacco & Satchell (1999) consider the data generating process (DGP),

$$\Delta y_t = \begin{cases} \mu_1 + \sigma_1 \epsilon_t & \text{if } s_t \text{ is regime 1,} \\ \mu_2 + \sigma_2 \epsilon_t & \text{if } s_t \text{ is regime 2,.} \end{cases} \tag{2.1}$$

where ϵ_t is $i.i.d$ $N(0,1)$, and s_t is the state of a 2-state Markov process.

Dacco & Satchell (1999) consider the forecast rule (where it is assumed that the forecaster knows μ_1 and μ_2.):

$$\Delta \hat{y}_t = \begin{cases} \mu_1 & \text{if } \hat{s}_t = 1, \\ \mu_2 & \text{if } \hat{s}_t = 2. \end{cases} \tag{2.2}$$

where $\hat{s}_t$ is a time$-(t-1)$ forecast of the next state. They define,

$$p_{ij} = P(\hat{s}_{t|t+1} = i, s_{t+1} = j), \tag{2.3}$$

for $i = 1, 2$. I have modified the notation of Dacco & Satchell (1999) with $\hat{s}_{t|t+1}$ to emphasize that this is a forecast using time$-t$ information.

Dacco & Satchell show that in the case where the regime variances are equal, i.e., $\sigma_1 = \sigma_2$, the forecast (2.2) is inferior to the random walk forecast (i.e., has a higher mean square error) if,

$$[p_{12} + p_{21}] \geq \pi(1 - \pi), \tag{2.4}$$

where $\pi = P(s_t = 1)$, the stationary probability of being in state 1 (which is the $(1, 1)-$element of $\lim_{n\to\infty}(p_{ij})^n$). I call (2.4) the Dacco-Satchell (DS) condition. Dacco & Satchell call $[p_{12} + p_{21}]$ the probability of state forecast misclassification.

For unequal variances ($\sigma_1 \neq \sigma_2$) Dacco & Satchell show that the mean square forecast error of the forecast (2.2) is

$$MSE^\star = \sigma_2^2(\frac{p_{11} + p_{12}}{\gamma} + p_{21} + p_{22}) + (\mu_2 - \mu_1)^2(p_{12} + p_{21}), \tag{2.5}$$

where $\gamma = \sigma_2^2/\sigma_1^2$. They also show that the MSE of the random walk forecast is,

$$MSE_{RW} = \pi(1 - \pi)(\mu_1 - \mu_2)^2 + \sigma_2^2(\frac{\pi}{\gamma} + 1 - \pi). \tag{2.6}$$

Hence the more general DS condition is that (2.5) is greater than (2.6).

2.2.2 Degree of Difficulty

The DS condition indicates when the MSFE of the random walk exceeds that of the optimal forecast, conditional on the information set generated by only past values of the variable itself. Formally, when the DS condition fails to hold,

$$D_1 = MSFE^\star[y_t|Y_1^{t-1}] - MSFE_{RW} \geq 0, \tag{2.7}$$

where Y_1^{t-1} is the information set generated by $\{y_{t-1}, y_{t-2}, \ldots, y_1\}$ and $MSFE^\star[y_t|\Omega]$ is the minimum MSFE forecast of y conditional of Ω. We expect that

$$MSFE^\star[y_t|\Omega_2] \leq MSFE^\star[y_t|\Omega_1] \text{ for } \Omega_1 \subseteq \Omega_2, \tag{2.8}$$

i.e., the optimal forecast conditional on a larger information set should have a lower MSFE. In particular, a multivariate model that uses variables other than y_t should improve the MSFE. Note that D_1 - which I call the one-step degree of difficulty, if positive, measures the 'MSFE deficit' facing any model using the extra information in Ω_2 that needs to be made up in order to beat the random walk forecast.

2.2.3 The Sample Dependence of the Probability of Misclassification

In this Subsection I show how the degree of difficulty measure D_1 varies with the forecast period.

The following lemma expresses the probability of state forecast misclassification in terms of the probability of current state misclassification:

Lemma 1. $p(\hat{s}_{t|t+1} = x, s_{t+1} = y) =$

$$\sum_{i=1}^{2}\sum_{j=1}^{2} p(s_{t+1} = y|s_t = j)p(\hat{s}_{t|t+1} = x|\hat{s}_{t|t} = i)p(\hat{s}_{t|t} = i, s_t = j), \tag{2.9}$$

where $\hat{s}_{t|t}$ is the filtered state estimate, i.e., the time-t state inferred from only time-t information (Hamilton, 1994).

Proof. Basically Lemma 1 expresses $P(\hat{s}_{t|t+1} = x, s_{t+1} = y)$ in terms of the four possible time$-t$ starting points. First, note that
$P(\hat{s}_{t|t+1} = x, s_{t+1} = y)$

$$\begin{aligned} &= \sum_{i=1}^{2}\sum_{j=1}^{2} p(\hat{s}_{t|t+1} = x, s_{t+1} = y, \hat{s}_{t|t} = i, s_t = j) \\ &= \sum_{i=1}^{2}\sum_{j=1}^{2} p(\hat{s}_{t|t+1} = x, s_{t+1} = y|\hat{s}_{t|t} = i, s_t = j)p(\hat{s}_{t|t} = i, s_t = j). \end{aligned} \tag{2.10}$$

Now, the following identity,

$$p(X, Y|Z) = p(X|Y, Z)p(Y|Z), \tag{2.11}$$

follows from

$$\begin{aligned} p(X, Y|Z)p(Z) = p(X, Y, Z) &= p(X|Y, Z)p(Y, Z) \\ &= p(X|Y, Z)p(Y|Z)p(Z). \end{aligned} \tag{2.12}$$

Setting,

$$Y = \hat{s}_{t|t+1},\ X = s_{t+1},\ Z = (\hat{s}_{t|t}, s_t), \tag{2.13}$$

in (2.12) gives,

$p(\hat{s}_{t|t+1} = x, s_{t+1} = y|\hat{s}_{t|t} = i, s_t = j)$

$$\begin{aligned} &= p(s_{t+1} = y|\hat{s}_{t|t+1} = x, \hat{s}_{t|t} = i, s_t = j)p(\hat{s}_{t|t+1} = x|\hat{s}_{t|t} = i, s_t = j), \\ &= p(s_{t+1} = y|s_t = j)p(\hat{s}_{t|t+1} = x|\hat{s}_{t|t} = i). \qquad (2.14) \end{aligned}$$

The final step follows because s_{t+1} is independent of $(\hat{s}_{t|t}, \hat{s}_{t|t+1})$ conditional on s_t, and $\hat{s}_{t|t+1}$ is independent of s_t conditional on $\hat{s}_{t|t}$. Substitute (2.14) back into (2.10) and the result follows. □

The next step in deriving the sample-dependent degree of difficulty measure is collected in a second lemma:

Lemma 2. $p(\hat{s}_{t|t} = i, s_t = j) =$

$$\lim_{T\to\infty} \frac{1}{T} \sum_{\tau=t-T}^{t} p(\hat{s}_{\tau|\tau} = i|\Delta y_\tau) \left(\frac{p(\Delta y_\tau|s_\tau = j)}{p(\Delta y_\tau)}\right) p(s_\tau = j). \qquad (2.15)$$

Proof.

$$\begin{aligned} p(\hat{s}_{t|t} = i, s_t = j) &= \int_{\Delta y} p(\hat{s}_{t|t} = i, s_t = j, \Delta y_t) d\Delta y_t \\ &= \int_{\Delta y} p(\hat{s}_{t|t} = i, s_t = j|\Delta y_t) p(\Delta y_t) d\Delta y_t \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_{\tau=t-T}^{t} p(\hat{s}_{\tau|\tau} = i, s_\tau = j|\Delta y_\tau), \qquad (2.16) \end{aligned}$$

where $\{\Delta y_\tau\}_{\tau=1}^{\infty}$ is an infinite sample from $p(\Delta y_t)$, the stationary distribution of Δy_t under the model (2.1).

Now substituting $p(Y|Z) = p(Z|Y)p(Y)p(Z)^{-1}$ into (2.11) gives,

$$p(X, Y|Z) = \frac{p(X|Y, Z)p(Z|Y)p(Y)}{p(Z)}. \qquad (2.17)$$

Using (2.17), I can re-write (2.16) as
$p(\hat{s}_{t|t} = i, s_t = j) =$

$$\lim_{T\to\infty} \frac{1}{T} \sum_{\tau=t-T}^{t} p(\hat{s}_{\tau|\tau} = i|s_\tau = j, \Delta y_\tau) \left(\frac{p(\Delta y_\tau|s_\tau = j)}{p(\Delta y_\tau)}\right) p(s_\tau = j). \qquad (2.18)$$

The result follows from the observation that the filtered state estimate is independent of the actual state conditional on the observed data, i.e.,

$$p(\hat{s}_{\tau|\tau} = i | s_\tau = j, \Delta y_\tau) = p(\hat{s}_{\tau|\tau} = i | \Delta y_\tau). \tag{2.19}$$

□

Note that the state estimation algorithm only uses the observed Δy_τ to form an optimal estimate - the actual state s_τ is unobserved. Note also that $p(\hat{s}_{\tau|\tau} = i|\Delta y_\tau)$ is the filtered state probability estimated on data up to time$-\tau$. The time series of these probabilities can be obtained as the output of any routine that estimates the Markov switching model (2.1)- e.g., the Gauss program written by James Hamilton to model exchange rate trends (Engel & Hamilton, 1990).

The two Lemmas imply the following Proposition:

Proposition 1. $p(\hat{s}_{t|t+1} = x, s_{t+1} = y) =$

$$\lim_{T\to\infty} \frac{1}{T} \sum_{\tau=t-T}^{t} \sum_{i=1}^{2} \sum_{j=1}^{2} p(s_{t+1} = y|s_t = j) p(\hat{s}_{t|t+1} = x|\hat{s}_{t|t} = i)$$
$$p(\hat{s}_{\tau|\tau} = i|\Delta y_\tau) \left(\frac{p(\Delta y_\tau | s_\tau = j)}{p(\Delta y_\tau)} \right) p(s_\tau = j) \tag{2.20}$$

Proof. Apply lemmas 1 and 2 in succession. □

2.2.4 The k-Step Ahead Forecast Difficulty

Proposition 1 also yields an expression for $p(\hat{s}_{t|t+k} = x, s_{t+1} = y)$. Denoting the transition matrix by Π, and using the fact that $p(s_{t+k} = i|s_t = j)$ is the (i, j)'th entry of Π^k [see e.g., Stirzaker (1994)], the k-step version of (2.20) is,

$p(\hat{s}_{t|t+k} = x, s_{t+k} = y)$

$$= \lim_{T\to\infty} \frac{1}{T} \sum_{\tau=t-T}^{t} \sum_{i=1}^{2} \sum_{j=1}^{2} (\Pi^k)_{yj} (\Pi^k)_{xi} p(\hat{s}_{\tau|\tau} = i|\Delta y_\tau) \left(\frac{p(\Delta y_\tau | s_\tau = j)}{p(\Delta y_\tau)} \right) p(s_\tau = j) \tag{2.21}$$

This is sufficient to calculate the equal variance Dacco-Satchell condition (2.4) and hence the equivalent k-step ahead degree of difficulty D_k. In order to calculate the general D_k (i.e., with $\sigma_1 \neq \sigma_2$) the k-step means and variances need to be estimated from the model

$$\Delta_k y_t = \begin{cases} \mu_1^k + \sigma_1^k \epsilon_t^k & \text{if } s_t \text{ is regime 1,} \\ \mu_2^k + \sigma_2^k \epsilon_t^k & \text{if } s_t \text{ is regime 2,.} \end{cases} \tag{2.22}$$

where $\Delta_k y_t = y_t - y_{t-k}$. This model can be estimated by sampling the data at frequency k and estimate the original (one-step) Markov switching model (2.1).

Note that there is an inconsistency in assuming that the number of regimes does not change as the sampling rate is lowered. In particular, a 2-state model for weekly-sampled data implies a 3-state process for fortnightly data, viz., {up-up, up-down, down-up, down-down} where the two middle states are the same. If one assumes that the weekly model is correctly specified, then a 2-state model for fortnightly data will be misspecified. Going the other way, it is not obvious what a correctly specified 2-state fortnightly model implies for weekly data. Given that the original 2-state model (Engel & Hamilton, 1990) was fitted to quarterly data, and is still regarded as a reasonable model, it is also not apparent that the number of states should be increased as sampling frequency decreases. One way to proceed would be to conduct specification testing at each sampling frequency to find the 'correct' number of states (Hamilton, 1996). Of course, such testing may still lead to the inconsistency mentioned above. It is straightforward to generalize the expressions for the degree of difficulty to more than two states, as noted in Dacco & Satchell (1999). Once the number of states is determined at each frequency, the relevant degree of difficulty can then be calculated.

2.2.5 The Degree of Difficulty of a Forecast Period

Proposition 1 shows that as the sample size increases, the p_{xy} and hence the Dacco-Satchell condition (2.4) will approach their limiting values. In this sense, the Dacco-Satchell condition can be interpreted as an asymptotic result as the forecast evaluation period becomes large. Alternatively, assuming that y_t is an ergodic process, the Dacco-Satchell condition indicates the ensemble average outcome of a large number of forecast evaluations over different out-of-sample periods.

In small samples, however, the left hand side of Dacco-Satchell condition, $p_{12} + p_{21}$ will depend on the realized sample. And, in particular, the degree of difficulty measure D_1, which is calculated from $p_{12} + p_{21}$ will differ from its limiting value, and reflect the degree of difficulty of forecasting that particular sample.

2.2.6 Estimating The Degree of Difficulty

In order to calculate p_{xy} (and hence D_1) from an observed sample $\{\Delta y_1, \ldots, \Delta y_T\}$ several quantities in (2.20) need to be calculated. These can all be computed from the standard output from a program that estimates a Markov switching model. As mentioned above, $p(\hat{s}_{t|t} = i|\Delta y_t)$, the filtered state probability, is generated automatically by the Gauss program of James Hamilton. Similarly, the transition probability matrix $\pi_{ij} = p(s_{t+1} = i|s_t = j)$ is a direct output from the program. I assume that the estimation period is large enough that the state forecast function $\phi_{ij} = P(\hat{s}_{t|t+1} = i|\hat{s}_{t|t} = j)$ is equal to the transition matrix, i.e., $\phi_{ij} = \pi_{ij}$. The conditional and unconditional probabilities of an observation $p(\Delta y_t|s_t = j)$ and $p(\Delta y_t)$ are straightforward to calculate once the parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ and the transition matrix are estimated (see Hamilton, 1994). Finally $p(s_t = j)$ is the (j, j)'th entry of $\lim_{N\to\infty} \pi_{ij}^N$.

The code to do all these calculations was written in the Ox programming language (Doornik, 2002) with an OxGauss interface to a Gauss program (written by James Hamilton) that estimates a two-state Markov switching model.

2.3 Forecasting the Australian Dollar

Meese and Rogoff (1983) showed that the random walk model is a stiff benchmark for any forecasting model of currencies and this has given rise to a 'foreign exchange forecasting puzzle.' However, some models do seem to have predictive ability over certain sub-periods [Clarida & Taylor (1997), Cheung et al., (2005)]. In this Section I formulate a similar model to Clarida & Taylor (1997) and conduct forecasts over two non-overlapping out-of-sample periods. Clements & Krolzig (1998) advocate the use of two out-of-sample periods to monitor the robustness of a forecast evaluation.

2.3.1 Data

Our data is weekly-sampled exchange rate data on spot and 3 month forward Australian dollar exchange rates from 1997-1 (week 1 of 1997) until 2003-23. All exchange rates are expressed as the US dollar price of an Australian dollar (so, actually, the data is the USD/AUD exchange rate). All variables were converted to logarithms to avoid the problems of the Seigel paradox (see Sarno & Taylor, 2002). Unit root tests in Appendix B show that both the spot and 3 month forward exchange rates appear to be I(1) processes.

2.3.2 Formulating a VEqCM

Given that both the spot and forward rates are I(1) processes, a VAR in first differences could be formulated. However, such a model will be misspecified if there is cointegration between the series [i.e., if a linear combination of the two is I(0)]. If the spot and forward rates are cointegrated then a vector equilibrium correction model (VEqCM) should produce superior forecasts to the first differenced VAR. In this Subsection I formulate a vector equilibrium-correction model of the spot and 3−month forward exchange rates. I denote the vector of variables as,

$$\mathbf{x_t} = [s_t, f_t^3]', \tag{2.23}$$

where s_t is the time series of spot rates and f_t^3 the 3 month forward exchange rate.

By following a general-to-specific modelling approach (starting with third-order VAR), I arrived (via an F test on retained regressors) at a second-order VAR model for the levels $\mathbf{x_t}$. This was then written in the isomorphic VEqCM representation,

$$\Delta\mathbf{x_t} = \pi\mathbf{x_{t-1}} + \beta\Delta\mathbf{x_{t-1}} + \epsilon_\mathbf{t}. \tag{2.24}$$

As we have shown, $\mathbf{x_t}$ is an I(1)-process while $\Delta\mathbf{x_t}$ is I(0). Hence, to avoid an unbalanced regression, π should not be of full rank.

2.3.3 Determining the Cointegration Rank

Under the null hypothesis that the long-run matrix π has rank r the (Johansen) Trace statistic should be very close to zero because the smallest $p-r$ eigenvalues should be close to zero. The Trace statistic has a nonstandard distribution under the null hypothesis that the rank of the long-run matrix is r. The empirical results of the Trace test on our forward exchange rate data is presented in Table 2.1. The test rejects the hypothesis that the cointegrating rank is equal to 0 but cannot reject the hypothesis that the rank is less than or equal to 1. The Max Eigenvalue test (not reported here) supported this conclusion. Hence we take the cointegrating rank to be 1.

H_0 : rank <=	Trace test	pvalue
0	80.99	[0.000] **
1	1.8119	[0.178]

Table 2.1: Testing for Cointegrating Rank:

2.3.4 Estimating the VEqCM

The fact that the long-run matrix π is not of full rank implies that it can be decomposed as,

$$\pi = \alpha\beta', \tag{2.25}$$

where α and β have full column rank.

The theory on forward rate systems in Appendix A says that the cointegrating vector should be the forward premium, $f_t^3 - s_t$. This implies that $\beta = [-1, 1]'$. Actual estimates of the cointegrating vector typically are very close to $[-1, 1]'$, see Clarida & Taylor (1997). Given the theory of Appendix A and empirical confirmation, I assume that the cointegration vector is known to be $[-1, 1]'$. This is the framework of the empirical application in Horvath & Watson (1995) who demonstrate that VEqCM-based tests of cointegration have substantially increased power when the cointegrating vector is imposed rather than estimated. By imposing the known cointegrating vector on the system, the VEqCM can be estimated as an unrestricted (bivariate) VAR in first differences with the premium as a regressor in each equation (Cochrane, 2005). The result from estimating the bivariate VEqCM on data up to 2000-1 is presented in Table 2.2. The model was estimated using the PcFiml class in the Ox programming language (Doornik, 2002).

Note that embedded in this bivariate VEqCM is a spot-forward rate regression with the typical (but difficult to explain) negative coefficient in front of the forward premium. (The forward rate unbiasedness hypothesis says that the coefficient should be +1.) The model predicts that if, say, the forward premium is positive, i.e., $s < f$, then $\Delta s < 0$, i.e., the exchange rate will depreciate (i.e., the AUD will depreciate against the USD, because I am expressing the exchange rate as the AUD price of a US dollar). This is a consistent finding in single-equation forward-spot regressions and appears to imply that the premium will widen even more - a disequilibrium, not equilibrium-correcting dynamics. However, the negative coefficient on the forward premium in the equation for change in the forward rate is of greater magnitude, i.e., the forward rate will depreciate even more than the spot rate, causing the premium to narrow - an implication that is only apparent from the full bivariate system. I re-emphasize that exchange rate dynamics in which the exchange rate appreciates on a forward premium are well known in the literature (Hodrick, 1988; Gourinchas & Tornell, 2004; Eichenbaum & Evans, 1995).

The model predicts that the spot exchange rate will trend (either appreciating or depreciating) until the forward premium returns to zero. Trending behaviour is a double edged sword in a forecasting model - when the model is right it will do very well, but when it gets the direction wrong, it will be very wrong - as Phillips (2003) points out in the context of using nonstationary regressors in forecasting: trends can 'embody a powerful signal' but 'can also be powerfully wrong in prediction'.

2.3.5 Recursive Forecasts

The forecast evaluation study of Meese and Rogoff (1983) (and Clarida and Taylor, 1997) employed a recursive forecasting methodology. I follow that method to generate an ensemble of forecasts from $1-$ up to $40-$steps ahead.

Explanatory	Δs_t		$\Delta f_{3,t}$	
Variable	Coeff	S.E.	Coeff	S.E.
Δs_{t-1}	-0.28	0.17	0.062	0.159
$\Delta f_{3,t-1}$	0.26	0.17	-0.072	0.160
$(f_3 - s)_{t-1}$	-0.29	0.15	-0.537	0.141
constant	-0.001	0.009	-0.002	0.0008

Table 2.2: Vector Error Correction Model - USD/AUD:

This is done as follows: a single forecast (for each horizon) is produced by estimating a model on data up to $2000-1$ (week 1 of 2000), and generating forecasts for $2000-2$ (the forecast origin) up to $2000-41$. I then push the forecast origin ahead by a single period and generate forecasts for $2000-3$ to $2000-42$. This procedure is repeated 30 times to give an ensemble of forecasts. This ensemble allows the us to calculate forecast statistics for, in this case, any horizon from $h=1$ to $h=40$. The procedure to do this was written in the Ox econometric programming language (Doornik, 2002).

2.3.6 Time-Varying Predictive Ability

One aim of this paper is to determine the extent to which time-varying predictive ability of exchange rate models can be explained by differing degrees of difficulty of the out-of-sample periods. Hence I conduct two forecast evaluations over two non-overlapping out-of-sample periods. The first period (which I call the $2000-$period) has forecast origins from $2000-2$ to $2000-31$. Each forecast is a dynamic $40-$step ahead forecast. For the second period (the $2002-$period) I also produce 30 forecasts with origins $2002-2$ to $2002-31$. Each of these forecasts is also a dynamic $40-$step ahead forecast. Figure 2.1 shows these two out-of-sample periods and their associated forecast origins.

Figures 2.2 and 2.3 show samples of the VEqCM forecast changes for the two periods. Figure 2.2 clearly shows that the VEqCM has forecasted the trend correctly. Figure 2.3 shows the opposite - illustrating Phillips (2003) statement about the possibility of trends being powerfully wrong in prediction.

In order to be consistent with Meese & Rogoff (1983) and Clarida & Taylor (1997), I disregard sampling variability in comparing the mean square forecast errors (MSFE's) of the VEqCM and random walk models over the two periods - i.e., I regard model A to have superior predictive ability to model B if A has a lower MSFE. This is also consistent with the degree of difficulty measure which simply indicates whether a model is expected to have a higher or lower MSFE - not whether the difference in MSFE's is significant - as in Diebold & Mariano (1995) for instance. Note that Clark & McCracken (2005) showed that forecast evaluation statistics appear to have quite low power in the presence of structural breaks - an issue likely to affect a formal forecast evaluation in the presence of the Markov switching trend breaks here.

The MSFE statistics for the VEqCM and random walk change forecasts

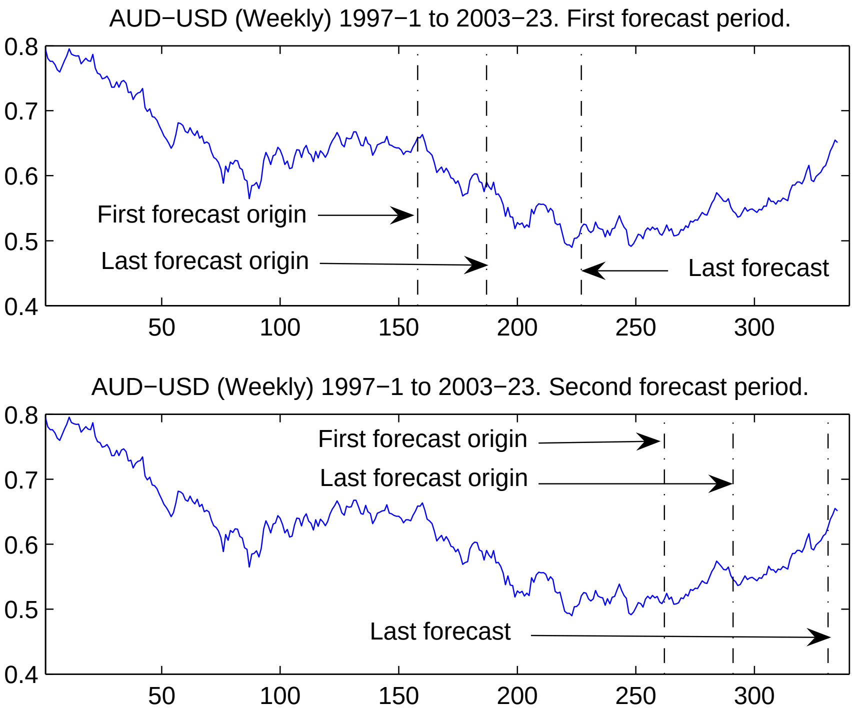

Figure 2.1: The AUD-USD exchange rate. Panel (a) shows the first forecast period. panel (b) shows the second forecast period.

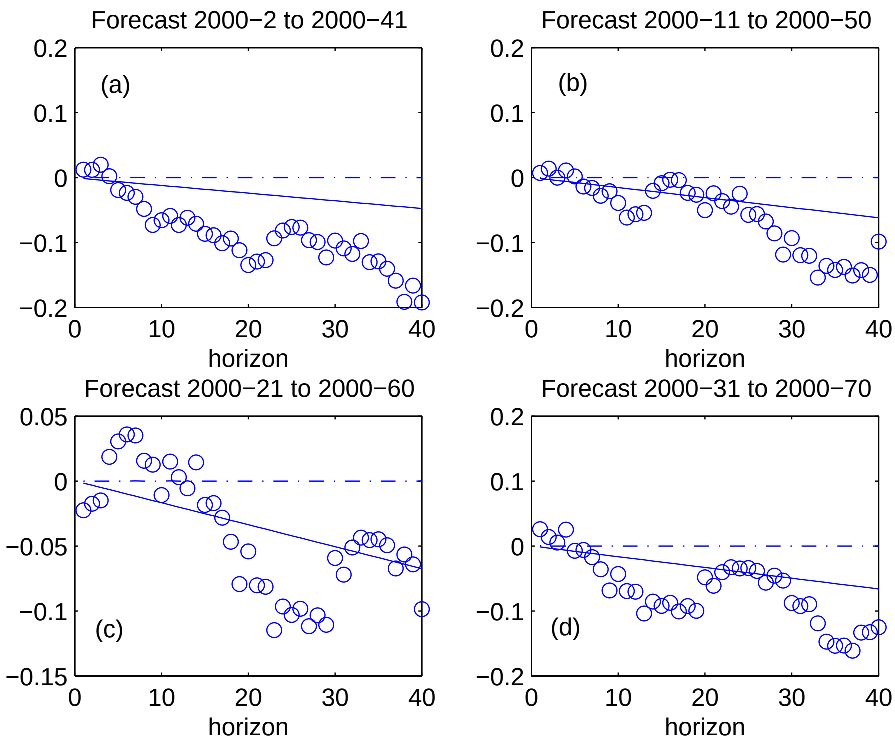

Figure 2.2: A selection of dynamic 40-step forecast changes for the USD-AUD exchange rate over the 2000 period. Panels (a)-(d) show forecast origins 2000-2, 2000-11, 2000-21, 2000-31. The random walk forecast is denoted by the broken line. The VEqCM forecast is the unbroken line. The actual change in the exchange rate is denoted by a circle.

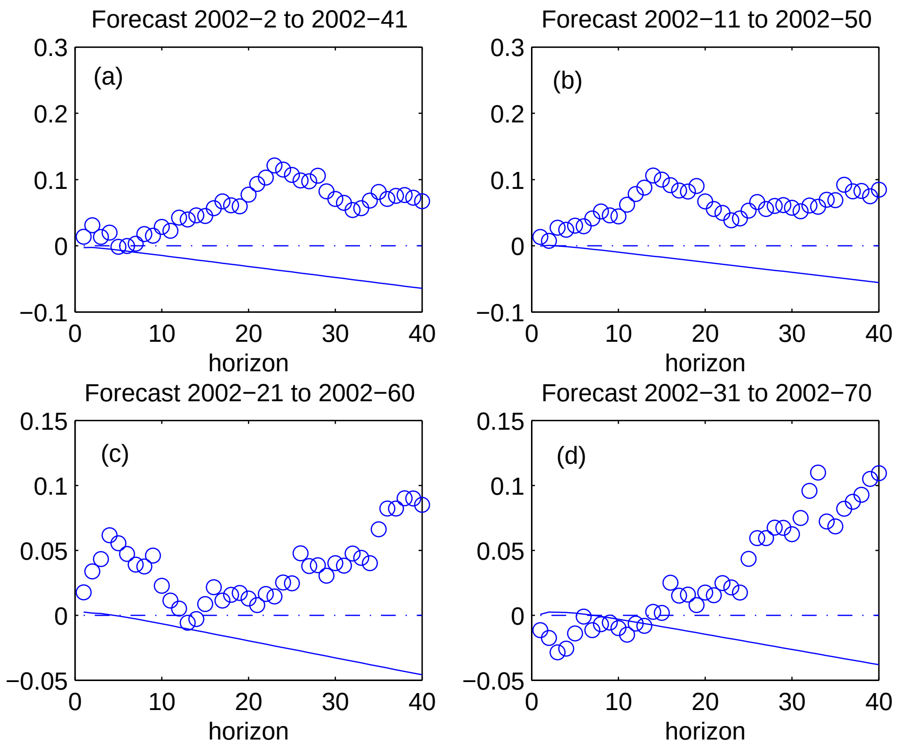

Figure 2.3: A selection of dynamic 40-step forecast changes for the USD-AUD exchange rate over the 2002 period. Panels (a)-(d) show forecast origins 2002-2, 2002-11, 2002-21, 2002-31. The random walk forecast is denoted by the broken line. The VEqCM forecast is the unbroken line. The actual change in the exchange rate is denoted by a circle.

over the 2000 and 2002 periods are presented in Figure 2.4. The MSFE's for the 2000 period [panel (a)] clearly point to the VEqCM being the superior model, at least beyond the 15−step horizon. The MSFE's for the 2002 period [panel (b)] tell the opposite story. In this period, the divergence between the models also occurs at around the 15−step horizon but here it is the random walk with the smaller MSFE. Appendix C shows a similar conclusion when forecast error densities are examined.

Hence, the conclusion about whether or not the VEqCM is a superior model or not is dependent on the period we choose to forecast. As mentioned before, often in the forecasting literature, studies that find forecasting ability of a model are overturned by later studies using new data. The analysis here shows that a study on the AUD/USD conducted in 2000 − 2001 would have been overturned by a 2002 − 2003 study.

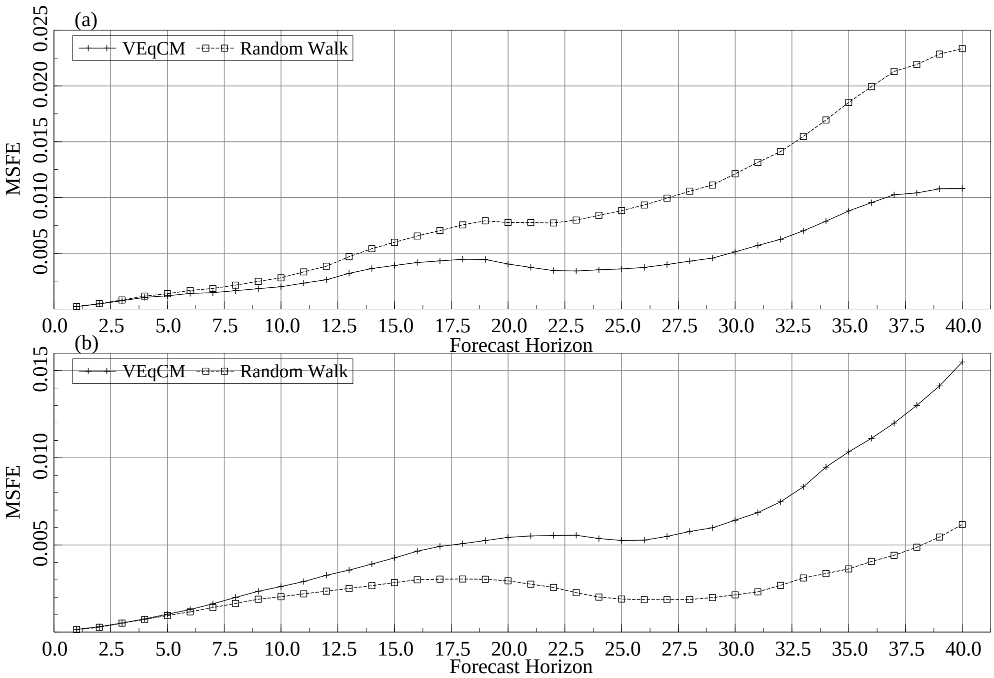

Figure 2.4: Mean square forecast errors from VEqCM and Random Walk forecast changes over different periods. All forecasts are 40−step dynamic forecasts. (a) Forecasts with forecast origins from 2000 − 2 to 2000 − 31. (b) Forecasts with forecast origins from 2002 − 2 to 2002 − 31.

2.4 Ex-Post Degree of Difficulty Analysis

In this Section I use the degree of difficulty measure to see whether the time-varying predictive ability result of the previous Section can be explained by the second out-of-sample period being harder to forecast over than the first. An alternative is that the two periods are of similar difficulty, and the time-varying predictability is an intrinsic feature of the VEqCM, or, by implication, the interest rate differential between Australia and the US (by covered interest parity, the interest differential is equal to the forward premium plus or minus transaction costs).

An assumption of this analysis is that the data generating process for the spot exchange rate has a stochastic trend break component that is reasonably well modelled by a two-state Markov switching process. Apart from the original work of Engel & Hamilton (1990), Klaassen (2005) and Cheung & Erlandsson (2005) are recent studies that have also found support for the Markov switching model.

2.4.1 Degree of Difficulty

The Markov switching model (2.1) was estimated on the USD-AUD data sampled at frequencies from $1, 1/2, 1/3, \ldots, 1/40$ to match the forecast horizons from the previous Section. For the weekly and fortnightly data I found it necessary to run a low-pass filter over the data (modes higher than the 60'th were omitted). Without this filtering, it appears that the low frequency trend is dominated by the noise. The estimates for selected horizons are reported in Table 2.3. Note that the mean and variance increase as the frequency $(1/k)$ decreases. It is important to note that, due to the sampling, the sample size gets quite small for the longer horizons making the accuracy of those estimates suspect.

The filtered regime probabilities for the two forecast periods, 2000 and 2002 are shown in Figures 2.5 and 2.6. The impression from looking at the regime plots is that the first period appears to be affected by more breaks than the second - the first period has 5 breaks to the second period's 2. This suggests that there is a greater chance of misclassifying the states in period 1, and hence misclassifying the forecast state, as outlined in Section 2. This is surprising given that the VEqCM forecasted better over the first period.

Figure 2.7 plots the degrees of difficulty for k-step ahead forecasts for the two periods. (Note that these degrees of difficulty have been divided by

Parameter	$k = 1$	$k = 2$	$k = 4$	$k = 13$	$k = 26$
μ_1	0.595	1.056	1.261	4.29	6.323
μ_2	-0.642	-1.300	-3.066	-2.55	-6.219
σ_1^2	0.277	1.130	5.628	15.86	51.52
σ_2^2	0.279	1.094	3.529	24.25	15.69
p_{11}	0.889	0.815	0.749	0.909	0.647
p_{22}	0.886	0.793	0.478	0.943	0.762

Table 2.3: Markov switching long swings model estimates for AUD-USD 1997-1 to 2004-30. The columns show the estimates for data sampled at weekly, fortnightly, monthly, quarterly and half-yearly frequencies.

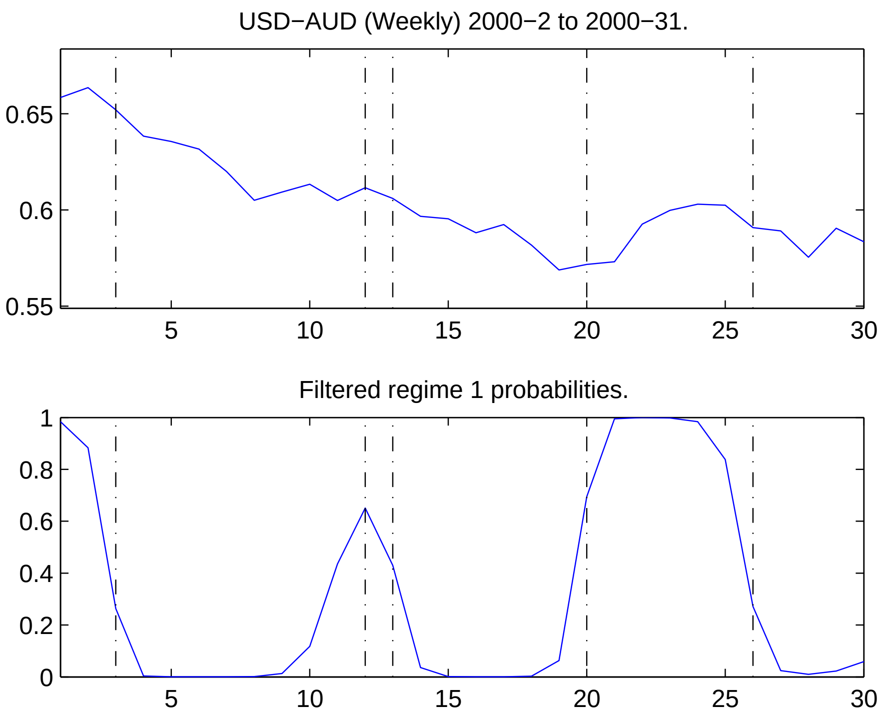

Figure 2.5: The AUD-USD exchange rate 2000-2 to 2000-31 with filtered regime probabilities.

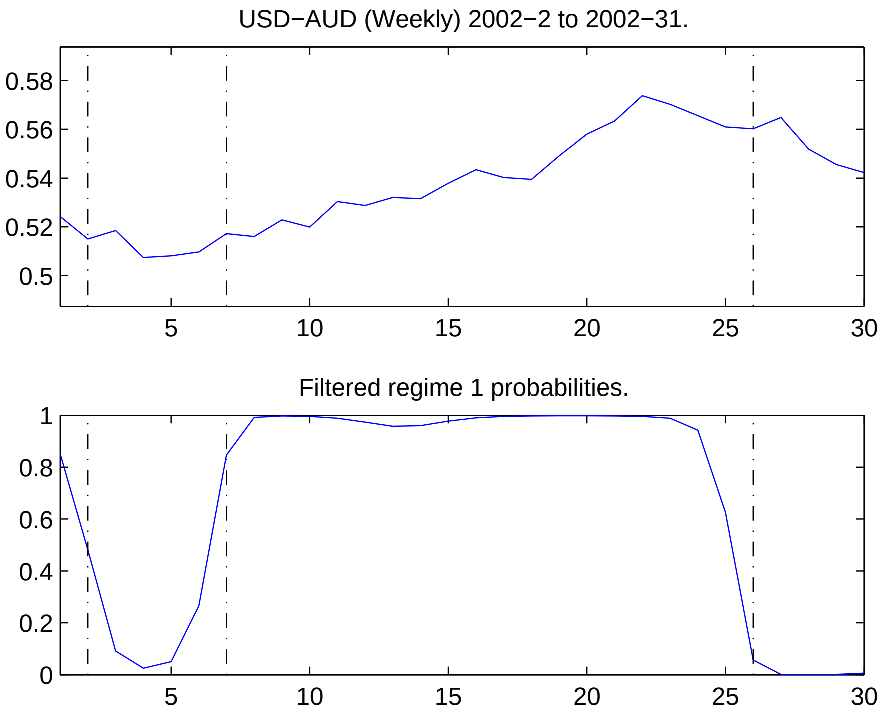

Figure 2.6: The AUD-USD exchange rate 2002-2 to 2002-31 with filtered regime probabilities.

10000 - this is because the VEqCM was estimated on Δs_t while the Markov switching model used $100\Delta s_t$.) Figure 2.7 accords with the visual impression of Figures 2.5 and 2.6 - the second period is the easier to forecast from. Note, however, that even though the VEqCM did not beat the random walk (Figure 2.4), it does appear to have reduced the MSFE deficit as measured by the degree of difficulty. For instance, at the 40-step ahead horizon, the degree of difficulty measure in Figure 2.7 shows that there is an MSFE deficit of about 0.012; at the same horizon, Figure 2.4b shows that the VEqCM has a MSFE of about 0.015 compared to the random walk forecast MSFE of 0.006, a difference of 0.009.

The VEqCM, seems to have done very well by beating the random walk over the more difficult 2000 period. So, although the difference in performance cannot be attributed to differing degrees of difficulty of the two periods, the VEqCM has improved on the optimal univariate forecast in both periods. In other words, there does appear to be predictive information in the interest differential that the VEqCM is exploiting. However, the results here suggest that this predictive ability is intrinsically time-varying - it is not driven solely by forecast periods of differing degrees of difficulty.

2.5 Conclusion

This Chapter has developed a way of measuring the degree of difficulty of forecasting a time series subject to Markov switching regime changes. The measure is the difference between the optimal univariate forecast and the random walk forecast over a period - it can be thought of as the mean square error deficit that any extra information has to make up, in order to beat the random walk forecast over that period.

This measure was used in an ex-post analysis of VEqCM forecasts of the Australian dollar exchange rate. Interestingly, the degree of difficulty measure found that the period in which the VEqCM failed to beat the random walk was actually the easier of the two periods. Despite this, the degree of difficulty measure showed that the VEqCM did manage to reduce the MSFE deficit over both periods, indicating that the extra information in the VEqCM had predictive content. However, it appears that the predictive power is truly time-varying.

Further work in this area could extend the degree of difficulty measure to more complex break processes and interbreak behaviour. It would also be

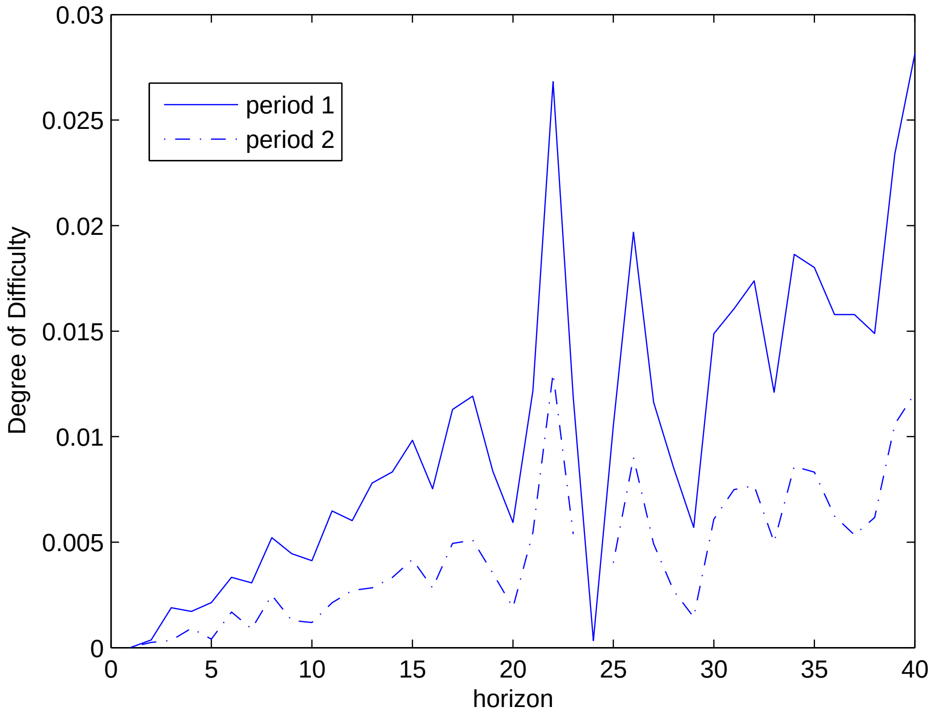

Figure 2.7: The degree of difficulty estimates for the two out-of-sample periods for forecast horizons 1 to 40. Period 1 (the unbroken line) is the 2000 period. Period 2 (the broken line) is the 2002 period. The degree of difficulty has been divided by 10000 to convert to the same units as the MSFE's from the VEqCM.

advantageous to incorporate the empirical bound of Phillips (1996), which takes into account the maximal proximity an estimated model can be from a trending DGP.

2.A Appendix: Forward Rate Systems

In this Section I present the theory that suggests that a vector equilibrium correction model (VEqCM), with forward premia as cointegrating vectors, is a promising candidate for foreign exchange forecasting. This theory is not new, but follows the framework of Clarida & Taylor, (1997) which drew upon Hall et al., (1992). The framework demonstrates that if spot and forward exchange rates are unit-root processes, and if deviations from the risk-neutral efficient markets hypothesis (RNEMH) are stationary, then the forward premia will be stationary, the system will be cointegrated with a common stochastic trend and the forward premia will form a basis for the cointegrating space.

First, assume that spot exchange rate s_t is a unit-root process. This is one of the strongly supported empirical facts from the exchange rate literature and, as Appendix B shows, is also supported by our data. Hence, I can write,

$$s_t = z_t + \nu_t, \tag{2.26}$$

where ν_t is a stationary process and z_t is a unit-root process of the form,

$$z_t = \gamma + z_{t-1} + e_t. \tag{2.27}$$

Following Clarida & Taylor (1997) I define the time-t deviation from the risk-neutral efficient markets hypothesis at horizon $h(j)$ as,

$$\phi_{h(j),t} = f_{h(j),t} - E[s_{t+h(j)}|\Omega_t]. \tag{2.28}$$

Combining Equations (2.26) and (2.28) gives,

$$f_{h(j),t} = h(j)\gamma + z_t + E[v_{t+h(j)}|\Omega_t] + \phi_{h(j),t}. \tag{2.29}$$

Now subtracting (2.26) from both sides of (2.29) gives an expression for the forward premia,

$$f_{h(j),t} - s_t = h(j)\gamma + E[v_{t+h(j)} - v_t|\Omega_t] + \phi_{h(j),t}. \tag{2.30}$$

Equation (2.30) makes it clear that as long as $\phi_{h(j),t}$ the deviation from the RNEMH is stationary, then the forward premia will be stationary too.

There are several important conclusions to be made here. First, note that Equation (2.30) shows that there exist linear combinations of the spot and forward rates that are stationary, despite the individual series being nonstationary. This implies that the system consisting of the spot and j forward rates is cointegrated. Second, it should be clear that there are j such linear transformations of the form (2.30), one for each forward premium. It is also easy to see that the vectors representing these j linear transformations are linearly independent. Hence the cointegrating rank of the system is j. Finally, the Granger representation theorem states that if a system of I(1) series are cointegrated then there exists a vector equilibrium correction model (VEqCM) representation of the system. This implies Granger causality in at least one direction. However, in the presence of breaks, Granger causality does not necessarily imply that one variable will improve out-of-sample forecasting ability (Clements & Hendry, 1999). In fact, Clements & Hendry, (1998, 1999) show that non-causal forecasting devices can out-forecast causal models. This result is relevant when I compare the forecasting ability of a random walk and a VEqCM in the presence of trend breaks.

2.B Unit Root Testing

In this Appendix I determine whether the spot and forward rates exhibit I(1) nonstationarity. A time series x_t is integrated of order one, i.e., I(1), if it is nonstationary but its first difference series $\Delta x_t = x_t - x_{t-1}$ is stationary. Hence I test the levels of spot and forward exchange rates for nonstationarity and then test the first differences.

Variable	Spot	3 Month
Level	-1.64	-1.666
First Difference	-5.748	-5.736

Critical values used in ADF test: 5% = -3.434, 1% = -4.008.

Table 2.4: Augmented Dickey-Fuller Statistics:

Table 2.4 contains the Augmented Dickey-Fuller (ADF) Statistics for the levels and first differences of the spot rate and the forward rate. The ADF statistics for the levels of the spot and forward rates are around -1.7. Hence we cannot reject the null hypothesis that each series is nonstationary.[1] By contrast, the second row in Table 2.4 shows that the first differences of both series appears to be stationary. More precisely, the ADF statistic values of around -5.7 all easily reject the null hypothesis that the series are nonstationary (the 1% critical value is -4.008). Hence, both series appear to be I(1) processes, i.e., their levels are nonstationary but their first differences are stationary. This is in agreement with the general conclusion of the literature on foreign exchange that spot and forward exchange rates are well described by I(1) processes (see, e.g., Sarno & Taylor, 2003).

2.C Appendix: Forecast Error Densities

This appendix presents the forecast error densities from the VEqCM over the two out-of-sample periods. Recall, our conclusion from examining the mean square forecast errors is that the VEqCM beats the random walk in one forecasting period but fails to in a subsequent period. This conclusion from examining mean square forecast errors, which only use the first two moments of the forecast error distribution, could conceivably change on examination of the forecast error densities. In this Appendix I show that it doesn't appear to.

2.C.1 Unsmoothed Forecast Error Densities

Figure 2.8 shows the forecast errors from the recursive forecasts over the same two forecasting periods as in the previous Section. The point of interest is the forecast error density at horizon h. We can determine the width of these densities (an indication of forecast error variance) from vertical cross-sections of Figure 2.8. While we could do this for all horizons, we focus on the horizon-40 forecast errors. We discuss the intermediate horizons in the next Subsection.

For the VEqCM in the 2000 forecasting period [panel (a) of Figure 2.8] the horizon-40 forecast errors have a range of $[0.0028, 0.195]$, while the random walk for the same period and horizon-40 has a range of $[0.0675, 0.239]$.

[1] It is possible that a longer timespan of data, even if sampled at a lower frequency (e.g., monthly) may give a more powerful unit root test that could reject this null (see Pierse & Snell, 1995 and references therein).

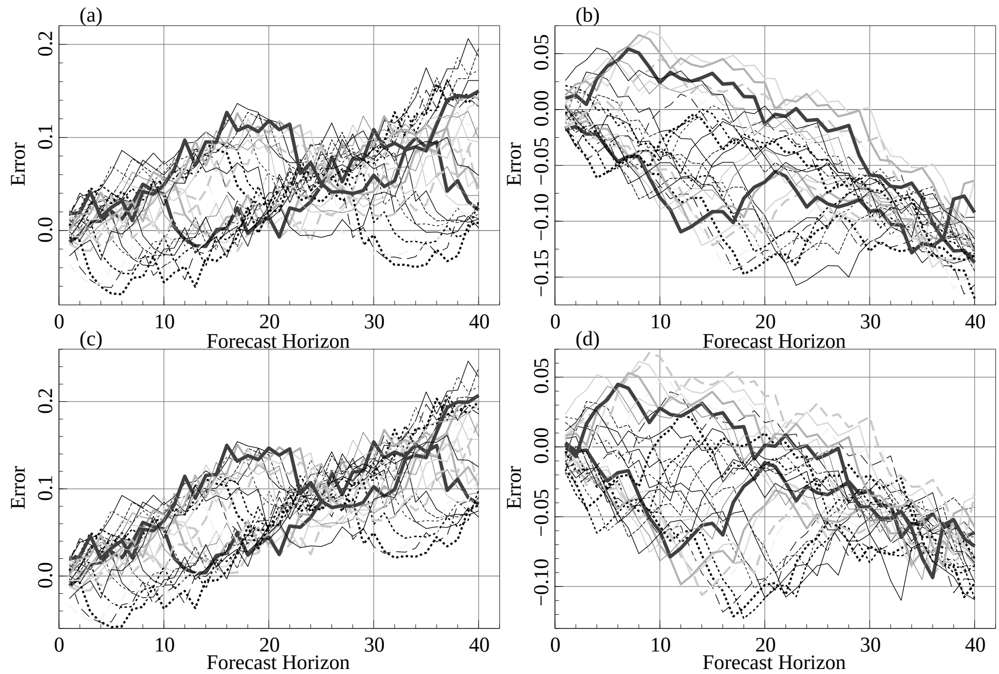

Figure 2.8: Recursive forecasts from VEqCM and Random Walk models over different periods overlaid on the same graph. All forecasts are 40−step dynamic forecasts. (a) VEqCM forecasts with forecast origins from $2000-2$ to $2000-31$. (b) VEqCM forecasts with forecast origins from $2002-2$ to $2002-31$. (c) Random walk forecasts with forecast origins from $2000-2$ to $2000-31$. (d) Random walk forecasts with forecast origins from $2002-2$ to $2002-31$.

Clearly the VEqCM is preferred for the 2000 period. However, for the 2002 forecasting period, the horizon-40 ranges are $[-0.169, -0.063]$ and $[-0.109, -0.033]$ for the VEqCM and random walk respectively. Hence for the 2002 period the random walk is preferred. This reversal is the same as when we considered mean square forecast errors.

2.C.2 Smoothed Forecast Error Densities

Another way to obtain the forecast error densities is to use kernel density smoothing. This allows us to see the forecast error densities more clearly, although we must keep in mind that these are estimated densities. Figure 2.9 is an overlay of the forecast error densities of the VEqCM and random walk models at selected horizons for the 2000 forecasting period. There is not much to choose between the two models until around the 17$-$step horizon when the peak of the VEqCM error distribution stays closer to zero.

Figure 2.10 is the same as Figure 2.9 but for the 2002 forecasting period. Like the 2000 period forecasts, both models are fairly close before horizon 17 but for horizons greater than this the random walk is more highly peaked around zero. These observations favour the random walk model for the 2002 period.

The bimodality apparent at some horizons in Figures 2.9 and 2.10 leads us to question the value of basing forecast evaluation on statistics such as the mean square forecast error. This is an issue that we discuss in the next Chapter.

Our analysis of the forecast error densities (smoothed or not) leads us to draw the same conclusion as just looking at mean square forecast errors. There is a sample dependence in our forecast evaluations that does not allow us to determine whether the VEqCM is a superior forecasting device when benchmarked against the random walk model.

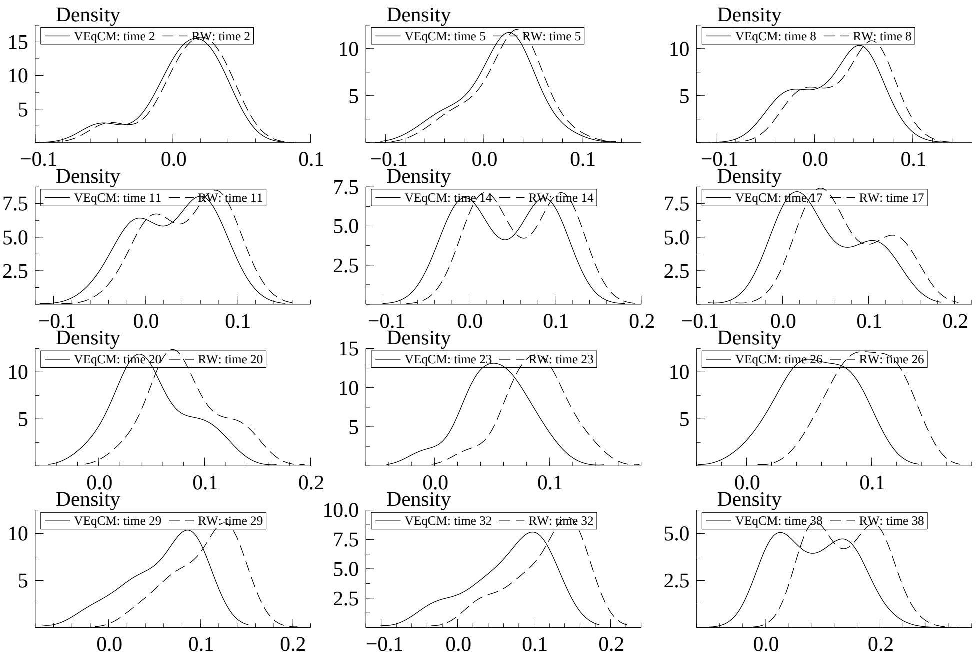

Figure 2.9: Smoothed forecast error densities from VEqCM and Random Walk models over different periods. All forecasts are 40−step dynamic forecasts with forecast origins from $2000 - 2$ to $2000 - 31$.

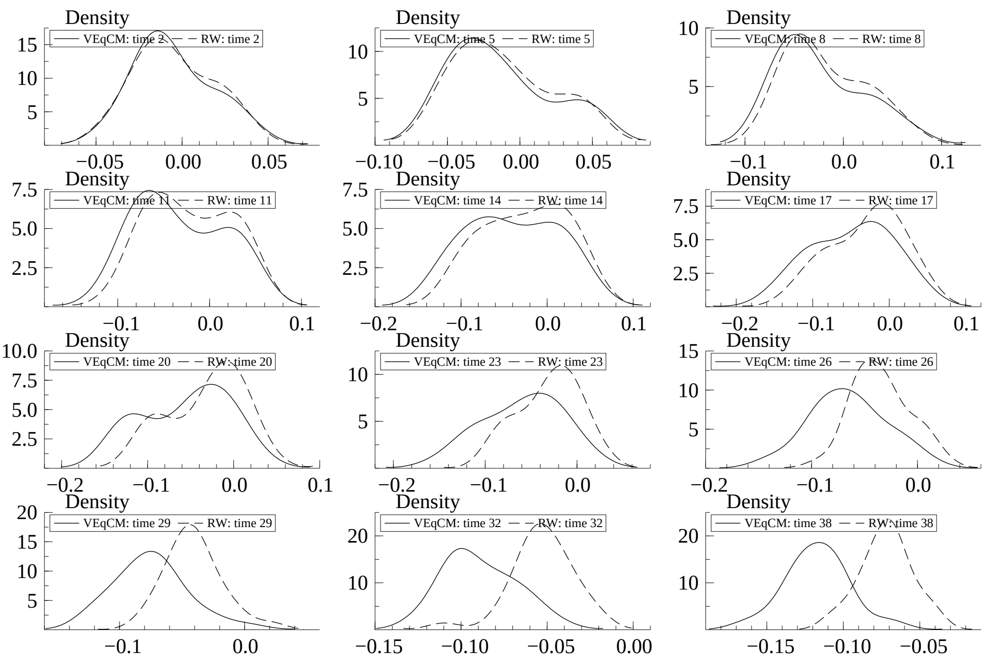

Figure 2.10: Smoothed forecast error densities from VEqCM and Random Walk models over different periods. All forecasts are 40−step dynamic forecasts with forecast origins from $2002-2$ to $2002-31$.

Bibliography

[1] CHEUNG,Y-W., CHINN,M.D. & PASCUAL,A.G. (2005). Empirical exchange rate models of the nineties: Are any fit to survive? *J.Int.Money.&.Finance.***24,** 1150-1175.

[2] CHEUNG,Y-W., & ERLANDSSON,U.G. (2005). Exchange rates and Markov switching dynamics. *J.Bus.Econ.Statistics* **23(3),** 314-320.

[3] CLARIDA,R.H. & TAYLOR,M.P. (1997). The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors. *Rev.Econ. & Stat.* **LXXIX,** 353.

[4] CLARIDA,R.H., SARNO,L., TAYLOR,M.P. & VALENTE,G. (2003). The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: a Step Beyond. *J.Int.Econ.* **60,** 61-83.

[5] CLARK,T.E. & MCCRACKEN,M.W. (2005). The power of tests of predictive ability in the presence of structural breaks. *J.Econometrics.* **124,** 1-31.

[6] CLEMENTS,M.P. & HENDRY,D.F. (1998). Forecasting Economic Time Series. *C.U.P.*

[7] CLEMENTS,M.P. & HENDRY,D.F. (1999). Forecasting Non-Stationary Economic Time Series. *The MIT Press*

[8] CLEMENTS,M.P. & 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.

[9] COCHRANE,J. (2005). Time Series for Macroeconomics and Finance. *University of Chicago, mimeo.*

[10] DACCO,R. & SATCHELL,S. (1999). Why do regime-switching models forecast so badly? *J.Forecasting.* **18,** 1-16.

[11] DIEBOLD,F.X. & MARIANO,R.S. (1995). Comparing predictive accuracy. *J.Bus.Econ.Statistics.* **13(3),** 253-263.

[12] DOORNIK.J.A. (2002). Object-Oriented Matrix Programming Using Ox, 3rd ed. *London: Timberlake Consultants Press and Oxford: www.nuff.ox.ac.uk/Users/Doornik.*

[13] EICHENBAUM,M. & EVANS,C.L. (1995). Some empirical evidence on the effects of shocks to monetary policy on exchange rates. *The Quarterly Journal of Economics.* November, 1995, 975-1009.

[14] ENGEL, C. & HAMILTON,J.D. (1990). Long swings in the dollar: are they in the data and do markets know it? *American Economic Review.* **80,** 689-713.

[15] GOURINCHAS, P-O. & TORNELL,A. (2004). Exchange rate puzzles and distorted beliefs. *J.Int.Econ* **64,** 303-333.

[16] HALL,A.D., ANDERSON,H. & GRANGER,C.W.J. (1992). A Cointegration Analysis of Treasury Bill Yields. *Rev.Econ. & Stat.* **74.**

[17] HAMILTON,J.D. (1994). Time Series Analysis. *Princeton, NJ: Princeton University Press.*

[18] HAMILTON,J.D. (1996). Specification testing in Markov-switching time series models. *J.Econometrics.* **70,** 127-157.

[19] HODRICK,R. (1988) The empirical evidence on the efficiency of forward and futures foreign exchange markets. *Fundamentals of Pure and Applied Economics.* Harwood Academic. N.Y.

[20] HORVATH,M.T.K. & WATSON,M.W. (1995) Testing for Cointegration When Some of the Cointegrating Vectors are Prespecified. *Econometric Theory* **11(5),** 952-984.

[21] KLAASSEN,F. (2005) Long swings in exchange rates: are they really in the data? *J.Bus.Econ.Statistics* **23(1),** 87-95.

[22] MEESE,R. & ROGOFF,K. (1983) Empirical Exchange Rate Models of the Seventies: Do They Fit Out Of Sample? *J.Int.Econ.* **14,** 3-24.

[23] PIERSE,R.G. & SNELL,A.J. (1995). Temporal aggregation and the power of tests for a unit root. *J.Econometrics.* **65,** 333-345.

[24] PLOBERGER,W., PHILLIPS,P.C.B. (2003). Empirical limits for time series econometric models. *Econometrica.***71(2),** 627-673.

[25] PHILLIPS,P.C.B. (2003). Laws and limits of econometrics. *Economic Journal.***113,** C26-C52.

[26] PHILLIPS,P.C.B. (1996). Econometric model determination. *Econometrica.***64(4),** 763-812.

[27] SARNO,L. & TAYLOR,M.P. (2002). The Economics of Exchange Rates. Cambridge University Press.

[28] STIRZAKER,D. (1994). Elementary Probability. Cambridge University Press.

Chapter 3

Long Horizon Exchange Rate Forecasts When There Are Trend Breaks

3.1 Introduction

Forecasting models are typically evaluated by holding back some data, called the out-of-sample (OOS) period, and using the model to generate an ensemble of forecast errors over this OOS period. This error ensemble is then used for testing various hypotheses. In this Chapter I show:

1. If the variable being forecasted has trend breaks, then so will the long-horizon forecast errors of any predictor,

2. This induces multimodality in the forecast error density of an error ensemble generated by a recursive forecasting procedure. This multimodality is spurious in the sense that it disappears asymptotically.

3. If the break-generating mechanism is assumed to be known, then one can obtain the the asymptotic forecast error distribution via Monte Carlo methods.

Formally, suppose that we have k $h-$step ahead recursive forecasts

$$\{f_{T|T+h}, f_{T+1|T+h+1}, \ldots, f_{T+k|T+h+k}\} \tag{3.1}$$

of a variable y_t over the OOS period $\{y_{T+h}, y_{T+h+1}, \ldots, y_{T+h+k}\}$. The ensemble of forecast errors generated from this recursive procedure is $\{e_{t|t+h}\}_{t=T}^{T+k}$ where $e_{t|t+h} = f_{t|t+h} - y_{t+h}$. Out-of-sample structural breaks are the main cause of forecast failure - a model with good in-sample fit generating poor out-of-sample forecasts [Clements & Hendry (1998, 1999)]. Now, an OOS break in the region $[T, T+h]$ will affect the distribution of the first forecast error $e_{T|T+h}$. However, the final member of the forecast error ensemble $e_{T+k|T+h+k}$ will probably be unaffected by the break (but may be affected by a later break). These issues make it unlikely that the members of the error ensemble are drawings from a common distribution. In fact, I show later that the forecast error density at any horizon is a mixture of densities. This is a problem for forecast evaluation tests using such ensembles, and implies that forecast error confidence intervals calculated from these ensembles can be seriously wrong. It also suggests a new procedure for producing the 'fan charts' published by central banks.

Forecast error distributions (and confidence intervals) should incorporate the potential impact of out-of-sample breaks. The theory I develop suggests a procedure for generating forecast error distributions that incorporate this break uncertainty. This procedure does require taking a stand on the break generating process. There is evidence that exchange rates are subject to Markov switching trend breaks (Engel & Hamilton, 1990). In the empirical Section I apply the procedure to the forecast errors of forward rate models of the AUD-USD exchange rate subject to Markov trend breaks.

The focus of this work is on the impact of out-of-sample breaks on the forecast error ensemble. There is already a considerable literature on the impact of in-sample breaks on forecasts, e.g., (Pesaran & Timmerman, 2005). The issues of model uncertainty, which is closely related to that of structural breaks also features in the literature on monetary rules. Also, this work does not contribute to the problem of how to forecast in the prescence of breaks, e.g., intercept corrections and differencing [Clements & Hendry (1998, 1999)], Bayesian model averaging (Pesaran et al., 2006) and sample selection methods (Pesaran & Timmerman, 2007) to name a few.

3.2 Forecasts of Exchange Rates with Long Swings

In this Section I present the results of a simulation study of a bivariate model of the exchange rate and the forward premium, in which the exchange rate is subject to Markov switching trend breaks. The simulations indicate two things: first, the forecast errors inherit the common regime-switching trend. Second, regime switching forecast errors, when combined with a recursive forecasting procedure, produce forecast error densities that can be bimodal - a problem when interpreting forecast statistics calculated from these densities. These findings are explained theoretically in the following Section.

3.2.1 Trend Breaks and Exchange Rate Forecast Errors

In this Subsection I show that if the exchange rate follows a broken trend process then so will the forecast errors for any predictor.

Following Engel & Hamilton (1990), suppose the exchange rate y_t follows a random walk with a time-varying drift. This can be modelled as,

$$\Delta y_{t+1} = \mu_{s_t} + \sigma\epsilon_{t+1}. \tag{3.2}$$

where ϵ_t are standard normal i.i.d., for all t. The drift μ_{s_t} varies due to the state variable s_t which takes values in a finite state-space $\{1, \ldots, k\}$. A trend break (in a given period) is a change in the value of s_t. Engel & Hamilton (1990) assumed that $k = 2$ and that s_t followed a Markov process (they also allowed the volatility to be state dependent, via $\sigma = \sigma_{s_t}$). The results of this Section, and the next, do not depend on the dynamics of s_t. The time$-t$, $h-$step ahead, forecast error from a predictive model $f_{t|t+h}$ is then equal to,

$$f_{t|t+h} - y_{t+h} = (f_{t|t+h} - y_t) - \sum_{i=1}^{h} \Delta y_{t+i}. \tag{3.3}$$

Equation (3.3) shows that the $h-$step ahead forecast error inherits the broken stochastic trend from (3.2). In particular, the distribution of this error over an out-of-sample period depends on the realized states (or trend breaks) over that period, i.e.,

$$\{s_{t+1}, s_{t+2}, \ldots s_{t+h}\}. \tag{3.4}$$

3.2.2 Recursive Forecasts and Trend Breaks

The recursive forecast procedure for generating an ensemble of forecasts is as follows. Essentially an initial forecast origin, T say, is selected and an $h-$step forecast $f_{T|T+h}$ produced. Then the forecast origin is incremented by one step to $T+1$ and a second $h-$step forecast $f_{T+1|T+h+1}$ is generated. This procedure is repeated k times to generate an ensemble of $h-$step ahead forecasts $\{f_{t|t+h}\}_{t=T}^{T+k}$. Forecast statistics are then calculated from this ensemble. If the DGP is given by (3.2) and hence, any forecast error by (3.3) then the forecast error ensemble inherits some unintended properties.

Suppose, by way of illustration, that we wish to forecast the exchange rate (at a weekly frequency) 6 weeks ahead starting in the first week of the year 2000. Also suppose that there are structural breaks in the weeks $2000-3$ and $2000-8$, i.e., that the state variable s_t changes value on those dates. Table 3.1 is a schematic that shows the effect of these structural breaks on the recursive forecasting method. Note that as the forecast origin moves forward, the point at which a break occurs comes earlier in the forecast. Eventually, the break point moves into the estimation period. This introduces a spurious threshold effect where some forecasts are directly affected by the break (but at different stages of the forecast) while in others, the impact of the break is via the estimated parameters of the model. In the case of an expanding estimation window, the effect of this extra data point on parameter estimates may be minimal.

Another spurious threshold effect introduced is that there are a different number of breaks in different forecasts. So, for instance, the forecast with forecast origin $2000-3$ is the only forecast with two breaks in the forecasting period. On the other hand, the forecast with origin $2000-9$ has no breaks in the forecast period. Looking vertically down table 3.1 we can see another strange threshold effect. As the forecast horizon crosses from 3 to 4 we see that the number of breaks in each horizon's ensemble drops from 2 to 1.

One would hope that these aberrant forecasts would be outweighed in an averaging procedure by the other breaks. However, these recursive forecasts do not generally produce large ensemble sizes. For example, if you have weekly data and you hold back a year of data for forecast evaluation, your ensemble will only have 52 forecasts.

Estimation			Origin	Horizon				
e_{K-2}	e_{K-1}	e_K	h_1	h_2	h_3	h_4	h_5	h_6
1999-50	1999-51	1999-52	2000-1	2000-2	**2000-3**	2000-4	2000-5	2000-6
1999-51	1999-52	2000-1	2000-2	**2000-3**	2000-4	2000-5	2000-6	2000-7
1999-52	2000-1	2000-2	**2000-3**	2000-4	2000-5	2000-6	2000-7	**2000-8**
2000-1	2000-2	**2000-3**	2000-4	2000-5	2000-6	2000-7	**2000-8**	2000-9
2000-2	**2000-3**	2000-4	2000-5	2000-6	2000-7	**2000-8**	2000-9	2000-10
2000-3	2000-4	2000-5	2000-6	2000-7	**2000-8**	2000-9	2000-10	2000-11
2000-4	2000-5	2000-6	2000-7	**2000-8**	2000-9	2000-10	2000-11	2000-12
2000-5	2000-6	2000-7	**2000-8**	2000-9	2000-10	2000-11	2000-12	2000-13
2000-6	2000-7	**2000-8**	2000-9	2000-10	2000-11	2000-12	2000-13	2000-14

Table 3.1: Schematic of recursive forecasts. The schematic shows how breaks impact recursive forecasts with forecast origins moving forward by one step each forecast. The forecast origin is $2000 - 1$. Break dates at $2000 - 3$ and $2000 - 8$ are represented by boldface dates. It can be seen that as the forecast origin moves forward, breaks impact earlier periods of the forecast and eventually move into the in-sample period.

3.2.3 Bimodal Forecast Error Densities

Here I show the results of a small simulation of a bivariate system that demonstrates some of the problems discussed in the previous Subsection. The system I consider is

$$\Delta y_t = \mu_{s_t} + \beta \pi_t^m + \sigma \epsilon_{1,t}, \tag{3.5}$$
$$\pi_t^m = \rho \pi_{t-1}^m + \epsilon_{2,t}, \tag{3.6}$$

where $\pi_t = \phi_t^{t+m} - y_t$ is the $m-$period forward premium. Consider using the forward exchange rate as the predictor of the future spot exchange rate. (The results of this Section are similar for a range of potential predictive models - the forward rate forecast was chosen as one of the simplest for illustrative purposes. In Section 3.4 I will examine forecast errors from different models over real data and demonstrate similar qualitative results.) The period t forward forecast error at horizon h is

$$e_{t|t+h} = \phi_t^{t+m} - y_{t+h}. \tag{3.7}$$

Figure 3.1 shows 30 forward forecast errors for horizons 1 to 50 from (3.5). The values used in the simulation were $\rho = 0.8$, $\beta = 1$, $\sigma = 0.5$. The state variable follows a two-state Markov process with $p_{11} = p_{22} = 0.9$ and $\mu_1 = 0.2$, $\mu_2 = -0.2$. Close inspection of Figure 3.1 shows the effect discussed where the later the origin of a forecast, the earlier the break occurs. It can also be seen that some forecasts are affected by breaks that the others are not. This seems to induce a bimodality to the forecast error distributions at certain horizons. This is more clearly seen in Figure 3.2 which plots the smoothed densities for selected horizons.

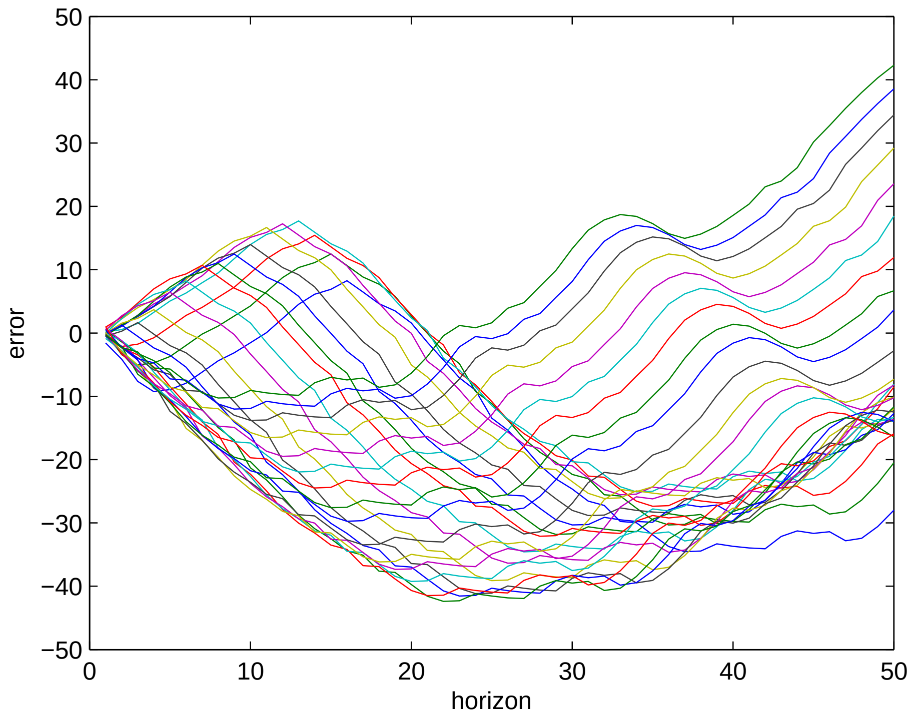

Figure 3.1: Forecasts from the simulated VEqCM. Each forecast is a 50−step dynamic forecast.

These smoothed densities were produced with the whole ensemble of 30 forecasts. The bimodality occurs at several horizons, most markedly at $h = 5$ and $h = 40$. While a genuine bimodality in a forecast error distribution would be of interest, that which emerges spuriously here is an artefact of the recursive forecasting procedure.

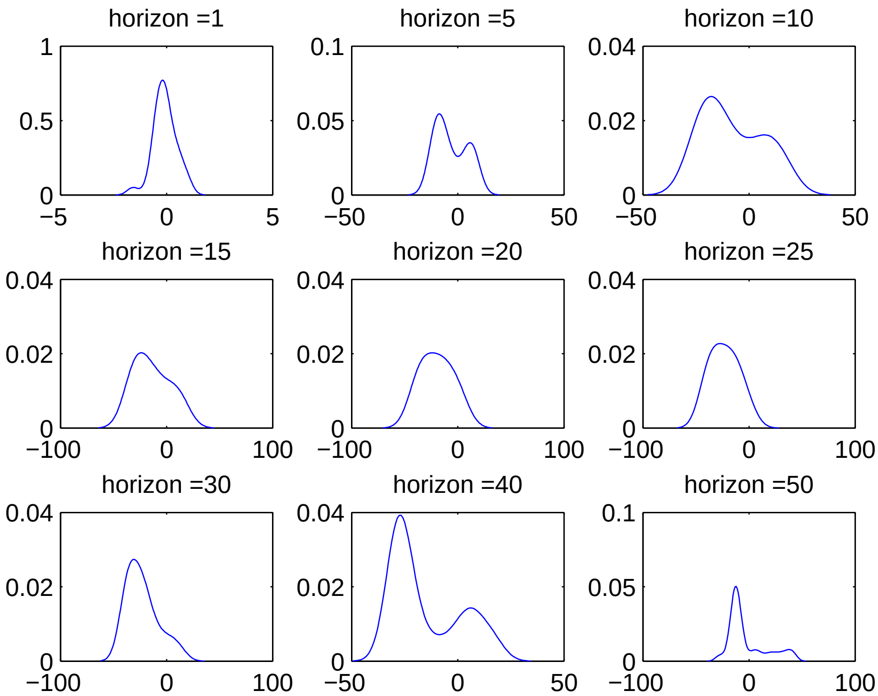

Figure 3.2: Smoothed density forecasts extracted from the recursive forecasts from the VEqCM with 50−step dynamic forecasts. Each density is for the forecast errors at a particular horizon.

3.3 A Theory of Forecast Error Ensembles Under Trend Breaks

In this Section I develop a theory that goes some way towards explaining the simulation results of the previous Section. The theory draws on the functional central limit theory for regime-switching trends developed by Cavaliere (2003), reviewed in the next Subsection (which does not contain original work).

3.3.1 Asymptotics for Markov-Switching Trends

Cavaliere (2003) defines a Markov-switching trend as follows:

Definition 1 (Markov-switching trend). *Let μ_t be an irreducible, aperiodic markov chain with state space $(\mu^1, \ldots, \mu^K)'$ (with $\mu^i \neq \mu^j$ for at least one i, j with $i \neq j$). A Markov-switching trend is a stochastic process $\{\tau_t\}_0^\infty$ that satisfies $\tau_t = \tau_{t-1} + \mu_t$ and $\tau_0 = 0$.*

The stochastic segmented trend model of Engel & Hamilton (1990) is an example of what Cavaliere calls an $I(1)$ process with Markov-switching trend.

Definition 2 ($I(1)$ process with Markov-switching trend). *An $I(1)$ process with markov-switching trend is a stochastic process $\{X_t\}_0^\infty$ satisfying the decomposition $X_t = \tau_t + Z_t$ where $\{\tau_t\}$ is a Markov-switching trend and $\{Z_t\}$ is $I(1)$, $\{\tau_t\}$ and $\{Z_t\}$ being stochastically independent at all leads and lags.*

The proof of Theorem 3 in Cavaliere (2003) derives a functional central limit theorem (FCLT) for an $I(1)$ process with Markov-switching trend X_t. In particular,

$$\begin{aligned}
X_t &= \tau_t + Z_t && (3.8)\\
&= (\tau_0 + \sum_{i=1}^{t} \mu_i) + (X_0 + \sum_{i=1}^{t} u_i) && (3.9)\\
&= (\tau_0 + \bar{\mu}t + \sum_{i=1}^{t}(\mu_i - \bar{\mu})) + (X_0 + \sum_{i=1}^{t} u_i) && (3.10)\\
&= X_0 + \tau_0 + \bar{\mu}t + \sum_{i=1}^{t} v_i && (3.11)
\end{aligned}$$

where $v_i = \mu_t - \bar{\mu} + u_t$. Cavaliere's FCLT for v_t is collected in the following Lemma [the proof of which is on p.212 of Cavaliere (2003)].

Lemma 3 (Cavaliere (2003)). *Let v_t be defined as above. If v_t is strong-mixing with size $-1/(1-2p)$ for $p > 2$, $\sup_t E|v_t|^p < \infty$ and $0 < \lambda_v^2 < \infty$ where $\lambda_v^2 := lim_{T\to\infty} E[T^{-1}(\sum v_t)^2]$, then*

$$T^{-1/2} \sum_{t=1}^{[sT]} v_t \Rightarrow \lambda_v B(s). \tag{3.12}$$

3.3.2 Regime-Switching Trends with Fixed Out-of-Sample Breaks

In this Subsection I present the functional central limit theory for the trend break process when the location of the out-of-sample breaks is fixed. This idea of fixing the location of breaks is common in the break dating literature and has been called snapshot asymptotics (Phillips, 1998).

Corollary 1. *Under the conditions of Lemma 3, if X_t is an $I(1)$ process with Markov-switching trend with $X_0 = 0$, and we fix the location of the breaks, i.e., the state space $(\mu_1, \ldots, \mu_T)$ is fixed, then,*

$$T^{-1/2}(X_{[sT]} - \bar{\mu}_1^{[sT]}[sT]) \Rightarrow \lambda_v B(s), \tag{3.13}$$

where $\bar{\mu}_1^{[sT]} = (1/[sT]) \sum_{i=1}^{[sT]} \mu_i$.

Proof. I first consider the case where the state space is not fixed. From (3.11) $X_t - X_0 - \tau_0 - \bar{\mu}t = \sum_i^t v_i$. Using Lemma 3 and the fact that $X_0 = \tau_0 = 0$ by assumption, this implies that

$$T^{-1/2}(X_{[sT]} - \bar{\mu}[sT]) \Rightarrow \lambda_v B(s) \tag{3.14}$$

Now, if the state space $(\mu_1, \ldots, \mu_{[sT]})$ is fixed then $\bar{\mu}$ will not approach its asymptotic mean, but will simply be equal to $\bar{\mu}_1^{[sT]} = (1/[sT]) \sum_{i=1}^{[sT]} \mu_i$. □

3.3.3 Recursive Forecasts of Regime-Switching Trends

In this Subsection I show that the forecast error densities have a mixed normal density. The significance of this is that mixtures of normals can be multimodal - leading to the simulation results presented in the previous Section.

Proposition 2. *If Y_t is an $I(1)$ process with Markov switching trend, and $\hat{Y}_{t|t+h}$ is an h-step ahead forecast of Y_t then the recursively-generated forecast error ensemble $e_{t|t+h} = \hat{Y}_{t|t+h} - Y_{t+h}$ for $t = T_0 + 1, \ldots, T_0 + K$, ($T_0$ is the forecast origin) follows a mixed normal distribution.*

Proof. If we define,

$$\lambda_j^2 = lim_{H\to\infty} var(H^{-1/2} \sum_{i=T_0+j}^{T_0+j+H} v_i), \tag{3.15}$$

then

$$H^{-1/2}(Y_{T_0+[sH]} - \bar{\mu}_{T_0}^{T_0+[sH]}[sH]) \Rightarrow \lambda_j B_j(s). \tag{3.16}$$

where,

$$\bar{\mu}_{T_0}^{T_0+[sH]} = \frac{1}{[sH]} \sum_{i=T_0+1}^{T_0+[sH]} \mu_i. \tag{3.17}$$

Now, $H^{-1/2}e_{T_0|T_0+[sH]} = H^{-1/2}(\hat{Y}_{T_0|T_0+[sH]} - Y_{T_0+[sH]})$, which implies that the $[sH]-$step ahead forecast error, $e_{T_0|T_0+[sH]} = \hat{Y}_{T_0|T_0+[sH]} - Y_{T_0+[sH]}$ follows the following invariance principle,

$$H^{-1/2}e_{T_0+j|T_0+j+[sH]} \Rightarrow \hat{Y}_{T_0+j|T_0+j+[sH]} - \bar{\mu}_{T_0+j}^{T_0+j+[sH]}[sH] - \lambda_j B_j(s). \tag{3.18}$$

Now, $B_j(1) \sim N(0,1)$ and $\lambda_j = \lambda$ for all j. Hence, if $Z \sim N(0,1)$,

$$H^{-1/2}e_{T_0+j|T_0+j+[sH]} \sim \hat{Y}_{T_0+j|T_0+j+[sH]} - (\lambda Z + \bar{\mu}_{T_0+j}^{T_0+j+[sH]}[sH]). \tag{3.19}$$

This expression is the distribution of the forecast error, *conditional* on the realized out-of-sample regime shifts. This shows that the conditional distribution of each forecast error $\{e_{T_0+j|T_0+j+[sH]}\}$, for $j = 1, \ldots, K$ is a scaled and mean-shifted normal distribution. The mean shift is determined the realized regime shifts in the forecast period. Note that there are likely to be a multiplicity of different break configurations that produce the same mean shift, and those mean shifts with larger multiplicities (entropies) should occur more often. Note also, that in the case where the state space and forecast horizon are both finite, there are a finite number of possible mean shifts $\{\bar{\mu}_k\}_{k=1}^K$ that can occur. We can form the unconditional distribution of the forecast errors by weighting each conditional distribution by the probability of the realized mean shift, yielding:

$$H^{-1/2} e_{T_0+j|T_0+j+[sH]} \sim \hat{Y}_{T_0+j|T_0+j+[sH]} - \sum_{k=1}^{K} p(\bar{\mu}_k)(\lambda Z + \bar{\mu}_k[sH]), \quad (3.20)$$

which demonstrates that the unconditional distribution of the forecast errors is a mixture of normals. □

3.4 Empirical Application: The AUD-USD Exchange Rate

In this Section I present the forecasts of the vector equilibrium correction model discussed in Chapter 2. Figure 3.3 shows nine forecasts from the VEqCM. Close inspection of Figure 3.3 shows the effect discussed where the later the origin of a forecast, the earlier the break occurs. It can also be seen that some forecasts are affected by breaks that the others are not. This seems to induce a bimodality in the forecast error distributions at certain horizons, $h = 15$ and $h = 40$ for instance. While a genuine bimodality in a forecast error distribution would be of interest, that which emerges spuriously here is an artefact of the recursive forecasting procedure.

In order to get an idea of the forecast error densities produced by the recursive method, Figure 3.4 plots the smoothed densities for selected horizons. These smoothed densities were produced with the whole ensemble of 30 forecasts. The bimodality occurs at several horizons, most markedly at $h = 14$ and $h = 38$. As mentioned before, this bimodality is simply an unwanted effect of the recursive forecasting procedure.

3.5 Estimating Break Distortion

This Section deals with the practical problem of deciding, ex-post, to what degree a particular forecast error ensemble has been distorted by out-of-sample breaks. The approach I take here involves proposing a stochastic process for the mechanism generating the breaks. At the end of this Section I deal with the potential criticism that the break-generating process is almost certainly misspecified.

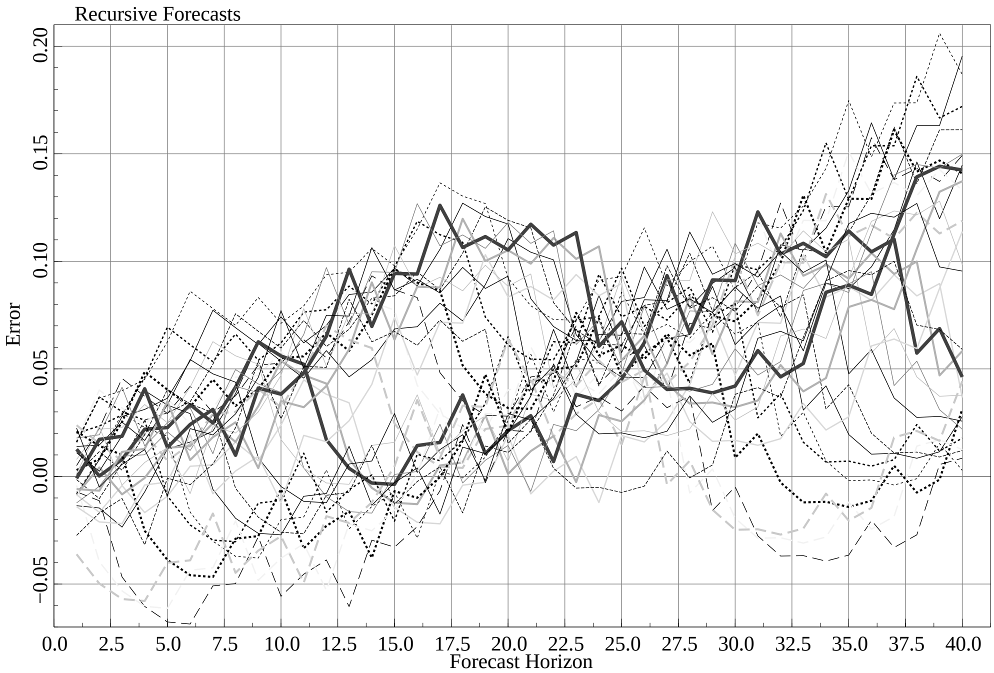

Figure 3.3: Recursive forecasts from the VEqCM. Each forecast is a $40-$step dynamic forecast. The forecasts are for forecast origins from $2000-7$ to $2000-15$.

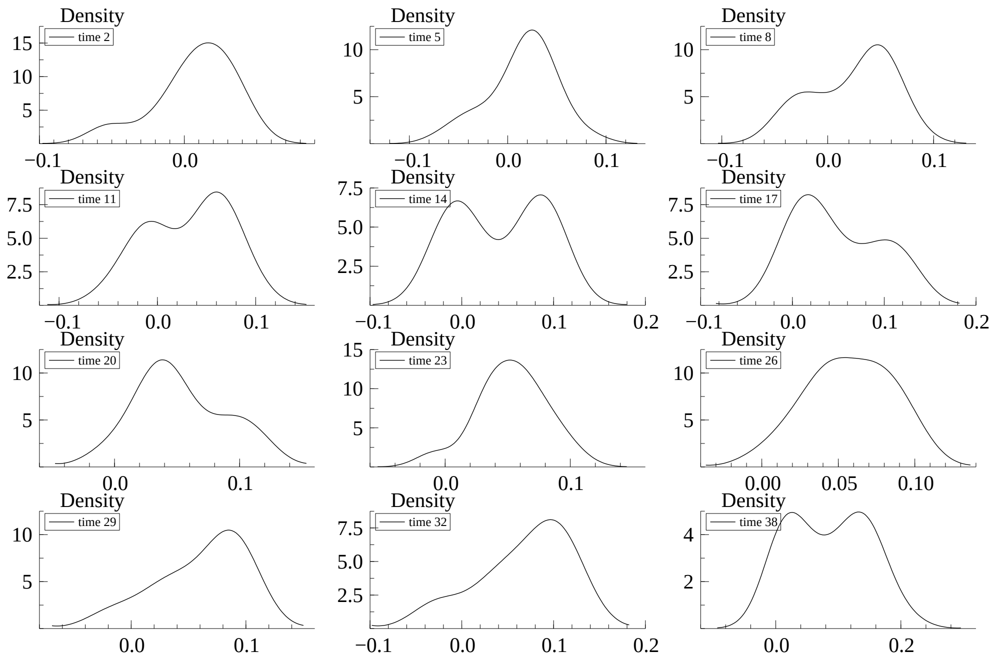

Figure 3.4: Smoothed density forecasts extracted from the recursive forecasts from the VEqCM with 40−step dynamic forecasts and forecast origins from $2000-2$ to $2000-31$. Each density is for the forecast errors at a particular horizon.

3.5.1 Break Configurations

I introduce the term 'break configuration' to denote the position of breaks in the forecast period. Suppose that we have an 20−step-ahead forecast of a trend-break process starting at the forecast origin T_0+1 and that the Markov chain driving the trend regimes has only two states, and up state $(\mu_t > 0)$ and down state $\{\mu_t < 0\}$, i.e., $s_t \in \{u, d\}$. One possible break configuration of this forecast error process is that the process starts in the 'up' regime and that there are breaks in periods 7 and 15. Following Chib (1998), the break configuration can be re-parametrized in terms of the realized unobserved states - the breaks are the dates at which this state changes value. Defining $C_{T_0+k,h}$ to be the break configuration from T_0+k to $T_0+k+h-1$, the example break configuration (breaks in periods 7 and 15) can be written as,

$$C_{T_0+1,20} = \{\mu_{T_0+t}\}_{t=1}^{20} = \{uuuuuu|dddddddd|uuuuuu\}. \tag{3.21}$$

3.5.2 Averaging Over Break Configurations

The joint distribution of the forecast model error at horizon h and the break configuration can be factorized,

$$f(e_{T_0|T_0+h}, C_{T_0+1,h}) = g(e_h|C_{T_0+1,h})p(C_{T_0+1,h}). \tag{3.22}$$

This tells us that the forecast error density is a mixture of distributions where the weighting factor is the probability of a given break configuration. Expression (3.22) also makes explicit the fact that a single forecast is a realization from a conditional density $g(e_h|C_B = C^{\star}_{T_0+1,h})$. So, a forecast statistic $\bar{Q}$ generated by K $H-$horizon forecasts, is an average

$$\bar{Q}^K = \frac{1}{K}\sum_{k=1}^{K} Q(e_{T_0+k|T_0+k+h}) \tag{3.23}$$

where Q is a function of the forecast errors - for instance, taking $Q(e) = e^2$ leads to (3.23) being the mean square forecast error. Note that in (3.23) each term in the sum is equally weighted, i.e., we are implicitly assuming that each break configuration $C_{T_0+k,h}$, $k = 1, \dots, K$ is equally likely. The question is, how representative is the ensemble of forecast errors generated by the recursive forecasting procedure. The quantity (3.23) is an approximation to the integral,

$$\begin{aligned}\bar{Q}^{\infty} &= \int_C Q(e_{T_0|T_0+h}) f(e_{T_0|T_0+h}, C_{T_0+1,h}) dC \quad (3.24)\\ &= \int_C Q(e_{T_0|T_0+h}) g(e_h|C_{T_0+1,h}) p(C_{T_0+1,h}) dC \quad (3.25)\\ &\simeq \frac{1}{N}\sum_{i=1}^{N} Q(e_{T_0|T_0+h}) g(e_{T_0|T_0+h}|C^i_{T_0+1,h}) p(C^i_{T_0+1,h}) \quad (3.26)\end{aligned}$$

The (lack of) representativeness of the recursive forecast statistic (3.23) can then be measured by the distance,

$$d(\bar{Q}^K, \bar{Q}^{\infty}) = |\bar{Q}^K - \bar{Q}^{\infty}|. \quad (3.27)$$

We could evaluate (3.26) by Monte Carlo methods if we knew the process generating the breaks - this would allow us to draw random variates from $g(e_{T_0|T_0+h}|C^i_{T_0+1,h})$ and weight each $Q(e_{T_0|T_0+h})$ by the correct probabilities $p(C^i_{T_0+1,h})$. Even without knowing the break-generating process, we can propose a simple stochastic process for the breaks (e.g., the Markov-switching model) and see how large (3.27) is. Even with a misspecified break-generating process, a large value of (3.27) would be an indication that breaks had distorted the recursive forecasts.

3.6 Modelling Forecast Error Dynamics

In order evaluate (3.26) we need to have a model of the forecast error dynamics. I propose to directly model the forecast error process as a Markov regime switching trend process.

3.6.1 Untangling the Breaks

We want to model (i) the breaks in forecast error processes and, (ii) the inter-break dynamics of forecast errors. Suppose we have an ensemble of forecasts with different forecast origins as Table 3.1. As can be seen from Table 3.1, we would lose information about the breaks if we averaged naïvely over different dynamic forecasts. We propose a contemporaneous averaging where we vertically align the dates of each forecast and calculate an average for

Origin	2000-1	2000-2	**2000-3**	2000-4	2000-5	2000-6	2000-7	**2000-8**
2000-1	2000-1	2000-2	**2000-3**	2000-4	2000-5	2000-6		
2000-2		2000-2	**2000-3**	2000-4	2000-5	2000-6	2000-7	
2000-3			**2000-3**	2000-4	2000-5	2000-6	2000-7	**2000-8**

Table 3.2: Schematic of contemporaneously averaging forecasts. Each forecast is a 6−step forecast. The forecast origin is incremented one step for successive forecasts. Break dates are represented by boldface dates.

each date. This is necessary because when there are breaks the forecast error processes are not stationary. Table 3.2 is a schematic of this contemporaneous averaging procedure. The forecasts are first differenced before averaged giving us a contemporaneously averaged forecast change.

3.6.2 A Segmented Trend Model for Forecast Errors

The model we employ for forecast errors from both the VEqCM and random walk model is a Markov switching segmented trend model along the lines of Engel & Hamilton (1990).

$$\Delta E_h \sim \begin{cases} N(\mu_1, \sigma_1^2) & \text{if } h \text{ is regime 1,} \\ N(\mu_2, \sigma_2^2) & \text{if } h \text{ is regime 2,} \end{cases} \tag{3.28}$$

Before the model was fitted, a low-pass filter was applied to the forecast error series. Without such filtering the regime switching model picks up the high-frequency movements in the error processes which is not desirable. The filtering was done by applying a Fourier transform to the data, setting the Fourier modes beyond a cut-off of 100 to zero, and then applying the inverse Fourier transform. The results were robust to different cut-offs, so long as the very high frequency components (modes beyond 250) were excluded.

Our model choice was strongly influenced by a desire for simplicity. It is certainly possible that more complex models may better model the forecast error processes of the VEqCM and random walk. The procedure for integrating over breaks presented in this work does not rely on a Markov switching model being used for forecast errors.

3.6.3 Results

The model (3.28) is estimated by the EM algorithm. Ox code was written to interface with the Gauss program written by James Hamilton (Engel & Hamilton, 1990). This program was run in OxGauss.

Table 3.3 contains the estimates of the segmented trend model for the forecast errors from the VEqCM and random walk model. Figure 3.5 plots the smoothed regime probabilities for the two forecast error processes. The close similarity of the inferred regimes is to be expected because both forecast error processes share a common stochastic regime switching trend.

Regime Estimates	VEqCM Coeff	RW Coeff
μ_1	0.578	0.678
μ_2	-0.154	-1.654
σ_1	0.320	0.307
σ_2	0.652	1.511
p_{11}	0.944	0.944
p_{22}	0.842	0.847

Table 3.3: Markov Switching Model for Forecast Errors

While the inferred regimes are almost identical, the dynamics of the two error processes are different. In state 1 (the 'up' state), the random walk errors tend to increase more than the VEqCM errors, i.e., $\mu_1^{RW} > \mu_1^{VEqCM}$. Similarly, in the 'down' state, $\mu_2^{RW} < \mu_2^{VEqCM}$. The random walk errors have a considerably higher volatility in state 2. It is possible to calculate the 'steady state' mean square forecast errors for both error processes (I thank Steve Satchell for this observation). Recall that the mean square error can be decomposed into the error variance and the square of the bias. The unconditional bias and variance can be written

$$bias = p_1\mu_1 + p_2\mu_2, \quad (3.29)$$
$$variance = p_1\sigma_1^2 + p_2\sigma_2^2 \quad (3.30)$$

where $p_i = \lim_{n\to\infty}(P^n)_{ii}$ is the steady state probability and P is the estimated transition matrix. Approximating these limits by the 10000'th power gives the follwing estimates:

$$p_1^{VEqCM} = 0.7383, \qquad p_1^{RW} = 0.7321, \quad (3.31)$$
$$p_2^{VEqCM} = 0.2617, \qquad p_2^{RW} = 0.2679. \quad (3.32)$$

Inserting these into (3.29) and (3.30) gives an unconditional bias of 0.386 and standard deviation of 0.407 for the VEqCM and 0.053 and 0.630 for the random walk. This implies a steady state $MSFE_{VEqCM} = 0.315$ and $MSFE_{RW} = 0.400$. Interestingly, the random walk forecast, despite being unbiased (as we would expect if exchange rates are martingales) has the higher MSFE.

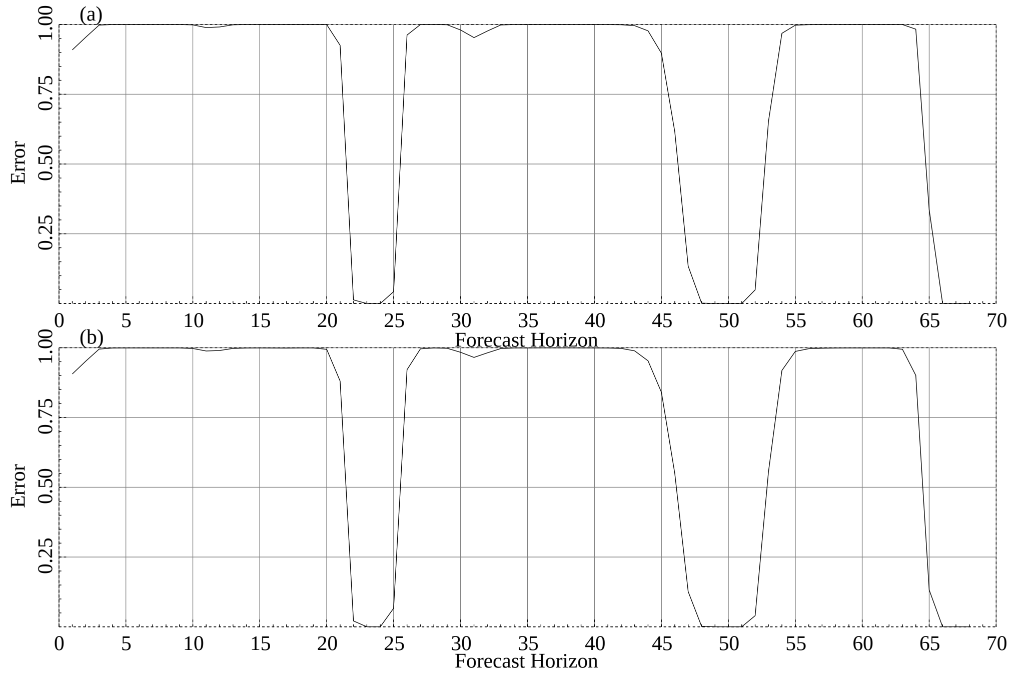

Figure 3.5: Smoothed regime probabilities for the forecast errors from the VEqCM and random walk models for the period $2000 - 3$ to $2001 - 17$. (a) VEqCM, (b) Random walk.

3.7 Monte Carlo Sampling

In this Section I use the model of forecast errors introduced in the previous Section to generate the forecast error ensembles needed to evaluate (3.26). I also discuss the implications for the 'foreign exchange forecasting puzzle,' i.e., can the VEqCM beat the random walk model of exchange rates?

3.7.1 Sampling Break Configurations

We need to draw sample break configurations with the correct probabilities. The assumption of Markov breaks means that we can use the transition probability matrix to do this. The estimated transfer probability matrix can be written as,

$$\mathbf{P} = \begin{pmatrix} p(1,1) & p(1,2) \\ p(2,1) & p(2,2) \end{pmatrix} \tag{3.33}$$

We want to determine the probability of any break configuration C_B, given the transfer matrix (3.33). A break configuration tells us the sequence of regimes during the forecast period. We can determine the probability of observing the regimes of neighbouring data points from (3.33). The probability of (3.33) generating a break configuration C_B is,

$$p(C_B) = p(C_B^1) \prod_{i=2}^{N} p(C_B^{i-1}, C_B^i), \tag{3.34}$$

where N is the number of steps in the forecasting period and C_B^i is the regime at step i. The unconditional probability of being in regime j is $p(j)$. We can generate samples from $p(C^i_{T_0+1,h}))$ by starting from an initial regime and using the transition matrix (3.33) to generate the subsequent break configuration.

3.7.2 The Conditional Density of Forecast Errors

Using Corollary 1 the conditional density of the $h-$horizon forecast error is

$$g(e_{T_0|T_0+h}|C^i_{T_0+1,h}) \sim N(h^{1/2}\bar{\mu}_1^h, \lambda_v^2). \tag{3.35}$$

In order to use this expression we would need a consistent estimate of the long-run variance λ_v^2. Given that the simulations here are only for horizons up to 40 steps ahead, and the the aim of the simulations is to approximate the break distortion, I use the simpler approach of simply drawing a random variate according to (3.33) for each state of the simulated break configuration.

3.7.3 Simulations

Following the terminology of Berliner (2001) the procedure outlined in the previous Subsection is a Monte Carlo ensemble forecast. Berliner (2001) deals with numerical weather forecasting where the ensembles are generated by using random initial conditions for each numerical simulation of the weather model. Our application here, by contrast, is an ensemble over random break configurations.

Figure (3.6) shows an ensemble of forecasts generated in this way. The whole Monte Carlo experiment comprised 5000 replications - clearly we can only show a limited number in Figure (3.6). It is difficult to distinguish the VEqCM and random walk forecasts in Figure (3.6) but the VEqCM forecast range is narrower random walk at the 40−step horizon.

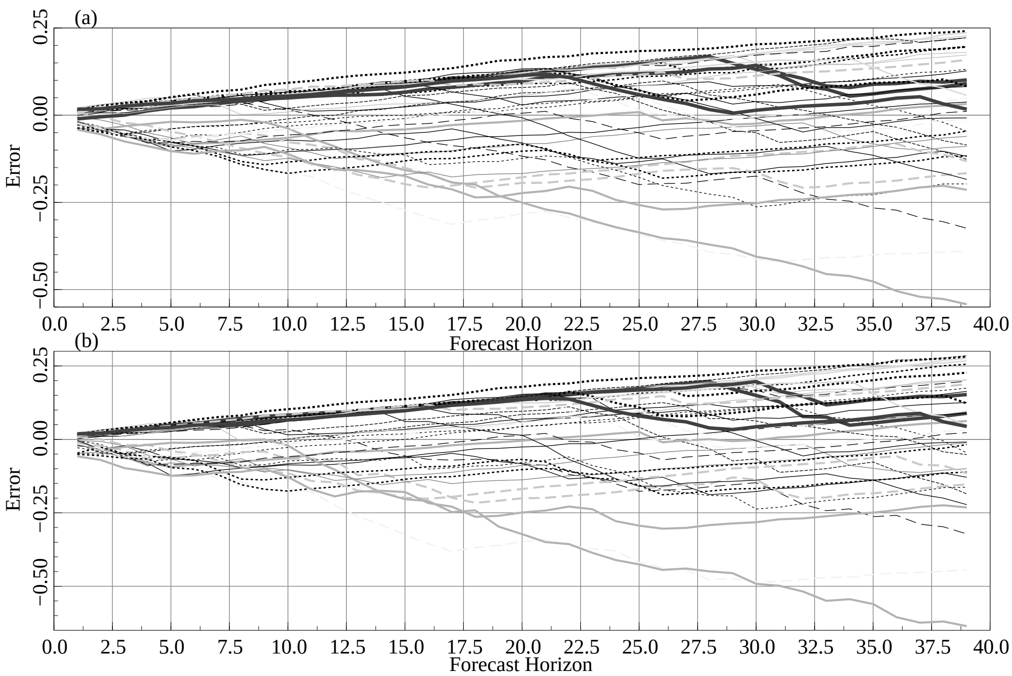

Figure 3.6: Monte Carlo ensemble forecasts for the forecast errors from VEqCM and random walk models. These simulated forecast errors are for a 30−step horizon. (a) VEqCM simulated forecast errors. (b) Random walk simulated forecast errors.

A better way to visualize the forecast ensemble is to estimate the smoothed forecast error densities at each horizon. A selection of these smoothed densities are presented in Figure (3.7). The quantiles of these densities could be calculated yielding confidence intervals that incorporate the possibility of out-of-sample Markov breaks. Alternatively the highest density regions could be calculated (Hyndman, 1995, 1996).

The densities in Figure (3.7) show that generally the probability mass of the VEqCM forecast errors are closer to zero than those of the random walk model. In the next Subsection we calculate statistics from these distributions that supports this conclusion.

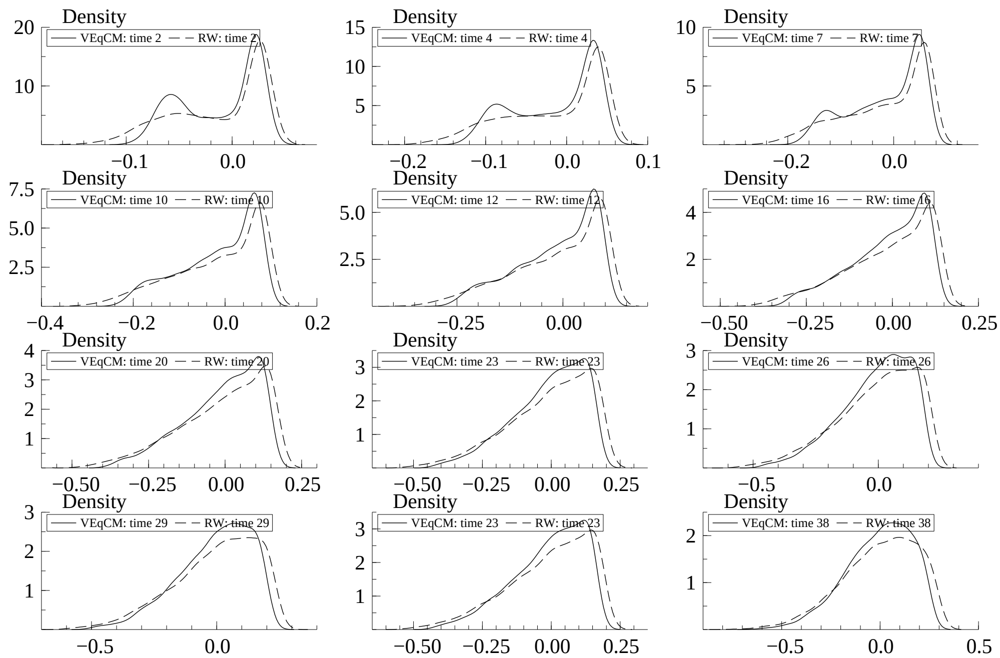

Figure 3.7: Smoothed density forecasts extracted from the simulated forecast errors for the VEqCM and random walk models.

3.7.4 Break-Averaged Forecast Statistics

Now that we have 'break-averaged' forecast error distributions for each horizon, we can calculate the moments of these distributions and other forecast statistics that have also been averaged over different break configurations.

Figure (3.8) shows a selection of break-averaged forecast statistics for the VEqCM and random walk models. The implications of these statistics for the foreign exchange forecasting puzzle are discussed in the next Subsection.

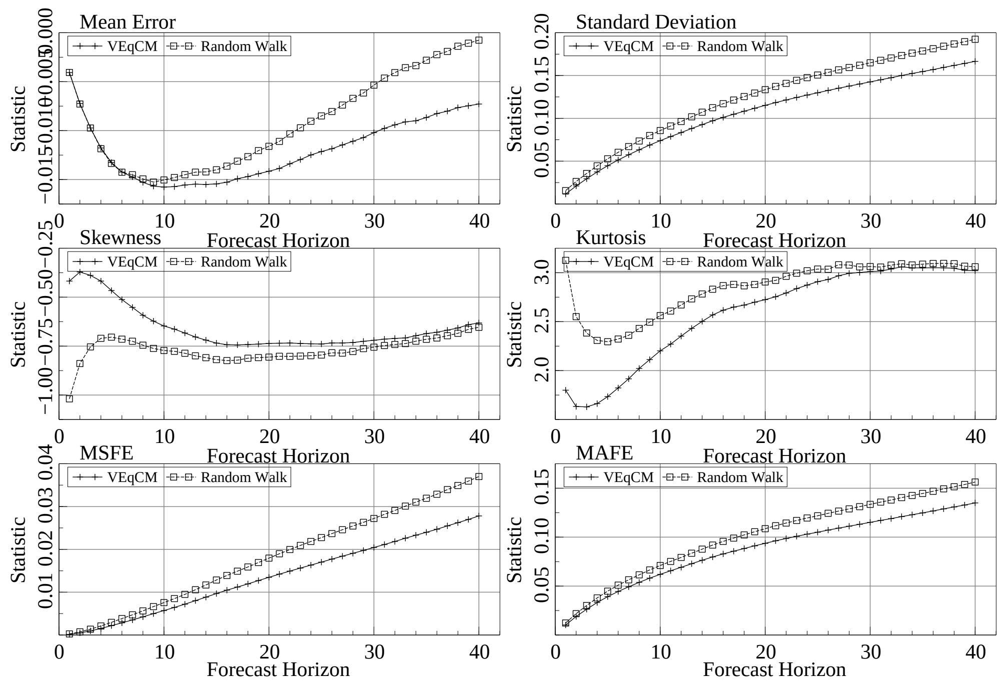

Figure 3.8: Moments of the forecast error distributions along with MSFE and MAFE for VEqCM and random walk models.

This idea of 'break-averaged' foreast statistics is a way to make explicit the dependence of the results of forecast evaluation study on the out-of-sample break configuration in that study. The Markov switching model for forecast errors formulated here may not be suitable for all studies - different models may be required in other circumstances. However, the framework is independent of the forecast error model.

3.8 The Information Content of Forward Premia and Break Distortion

The predictive ability of the VEqCM can be interpreted as an indicator of the information content in forward exchange rates. Clarida & Taylor (1997) provided a theory that shows that forward premia can be biased predictors of the future exchange rate change while still having predictive content. Their empirical work showed that a VEqCM with the forward premia as cointegrating vectors did beat a random walk over their out-of-sample period.

In this Section I use the break distortion measure (3.27) to study the impact of breaks on the predictive performance of the VEqCM, which is a smaller (bivariate) version of the Clarida & Taylor 5-equation system. Figure (3.9) shows that the VEqCM has a lower mean square forecast error (MSFE) over the 2000 period for the AUD-USD exchange rate. By comparing this with the simulated MSFE, we can get an idea of the break distortion over the 2000 period.

Figure (3.10) shows the simulated mean square forecast errors for the VEqCM and random walk models. This clearly shows the VEqCM to still have predictive ability over a wide range of (simulated) break configurations. This is not to say that the VEqCM will beat the random walk over all possible break configurations, simply that, if both models are compared over a large enough variety of break configurations, the VEqCM should be superior. The other forecast statistics [Figure (3.8)]: mean absolute forecast error, standard deviation, and kurtosis support also support this conclusion.

Figure (3.11) shows the break distortion for the MSFE's of the VEqCM and random walk models - this is (3.27) where $Q(e) = e^2$. There is clearly some impact of the trend breaks on both the VEqCM and random walk forecasts. Forecasting models are often compared via the difference of their MSFE's. Figure(3.12) shows the difference of the two curves in Figure (3.11).

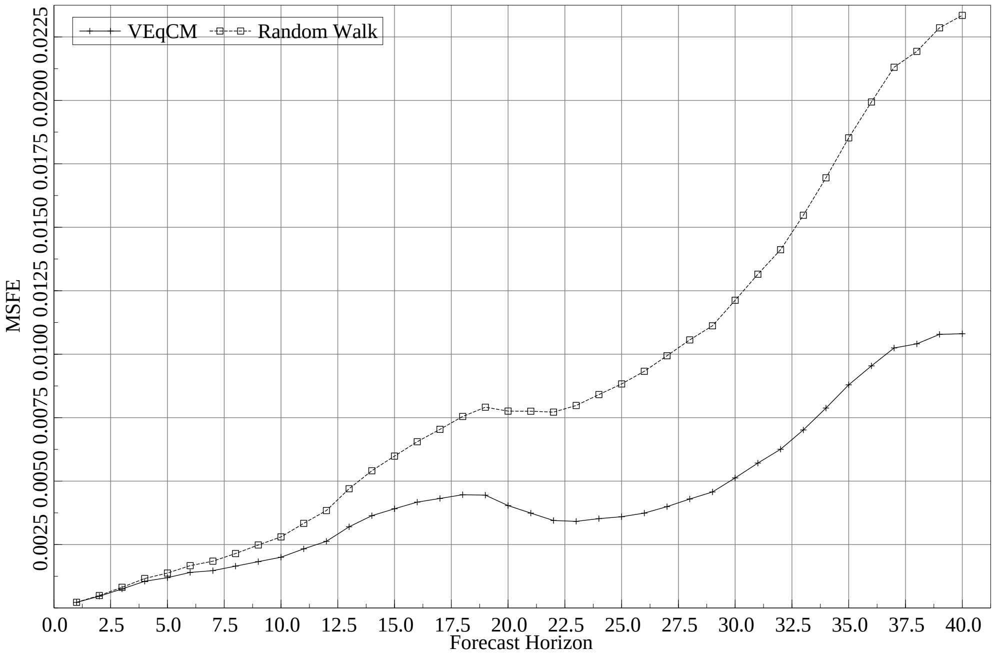

Figure 3.9: Mean square forecast errors for the VEqCM and random walk models with forecast origins 2000-2 to 2000-31.

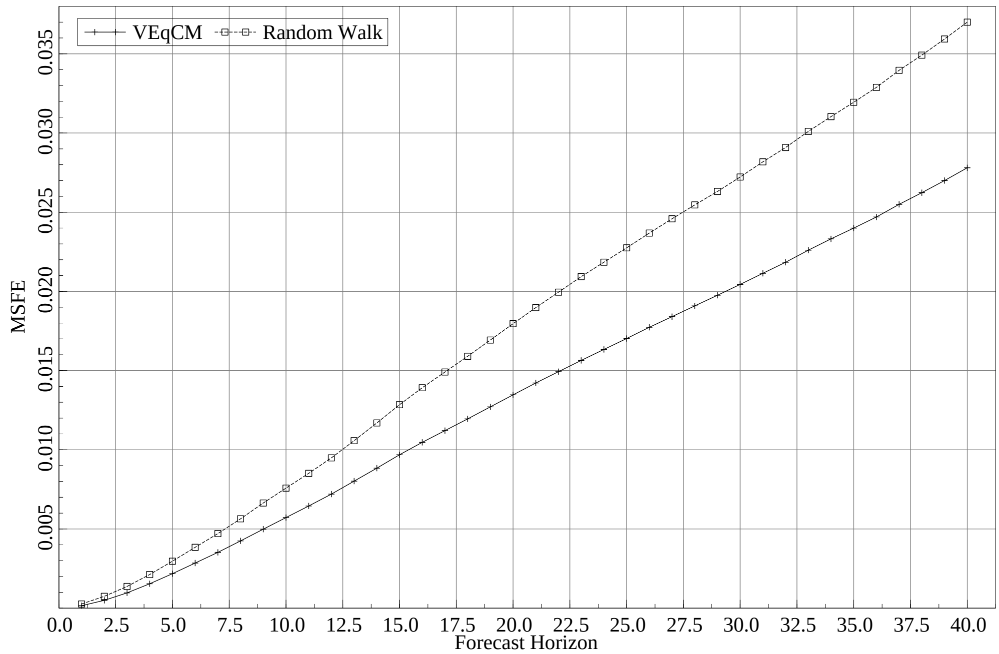

Figure 3.10: Simulated Mean square forecast errors for the VEqCM and random walk models.

Interestingly the impact of breaks on the MSFE difference is greatest at the longer horizons. It is at these long horizons that foreign exchange return predictability is often found. Figure (3.12) casts some doubt about the extent to which this is driven (or retarded) by trend breaks. Interestingly, the distortion begins a decline at around the 25-step horizon, but continues to increase for the VEqCM. This may be a cause for concern for the robustness of long-horizon return predictability findings.

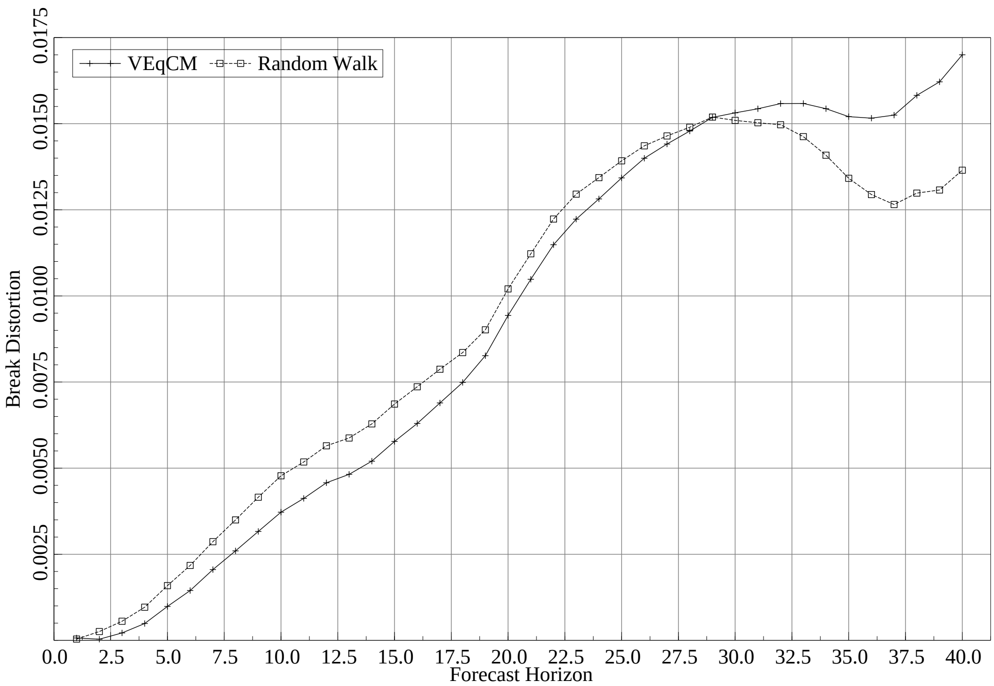

Figure 3.11: Estimated Break Distortion of Mean Square Forecast Errors for the VEqCM and random walk models.

A reasonable conclusion from both Figures (3.12) and (3.11) is that breaks have affected the forecast comparison in the 2000 period. We cannot say at this point whether this is sufficient to cast doubt on the forecasts over 2000.

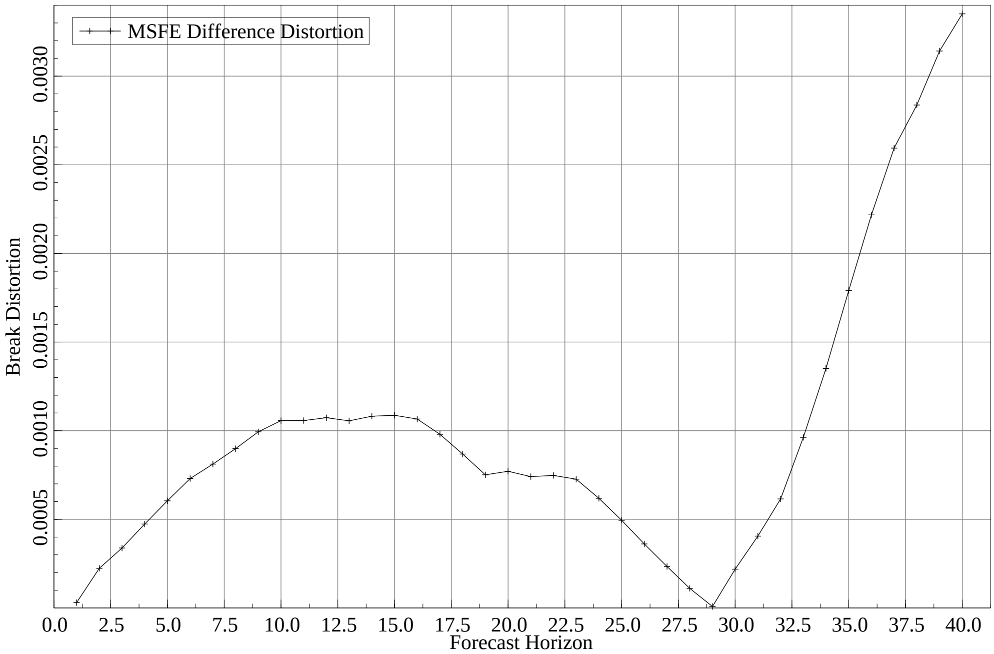

Figure 3.12: Break Distortion of Difference of theMean Square Forecast Errors for the VEqCM and random walk models.

3.9 Conclusion

This work has provided a theory and simulation evidence showing how stochastic trend breaks in the out-of-sample period can distort the forecast error ensembles produced by recursive forecasting procedures. This was also shown to be an issue in an empirical study using the forward premium to predict the echange rate change.

Having demonstrated the problem of breaks I presented a way of gauging whether trend breaks have distorted a given forecast evaluation. The method of averaging forecast statistics over a wide range of break configurations allows the investigator to get an idea of the sample-dependence in aforecast evaluation. Further work is needed to determine how large the break distortion needs to be before one should regard the forecasting study as suspect.

The idea of modelling forecast error dynamics under breaks is not dependent on the specific forecast error model that we formulated, i.e., a Markov switching model. It is possible that more detailed models, including those that specifically model structural breaks in the forecasting model, might provide superior results. Further work is needed to see how robust our method is to misspecification of the trend break generating process.

Bibliography

[1] BANERJEE,A., DOLADO,J., GALBRAITH,J.W. & HENDRY,D.F. (1993). Cointegration, Error-Correction, and the Econometric Analysis of Non-Stationary Data. Oxford University Press.

[2] BERLINER,L.M. (2001). Monte carlo ensemble forecasting. *Statistics and Computing.* **11,** 269-275.

[3] CAVALIERE,G. (2003). Asymptotics for unit root tests under Markov regime-switching. *Econometrics Journal.* **6,** 193-216.

[4] CHIB,S. (1998) Estimation and comparison of multiple change-point models. *J.Econometrics.* **86,** 221-241.

[5] CHONG,Y.Y. & HENDRY,D.F. (1986). Econometric evaluation of linear macroeconomic models. *Review of Economic Studies.* **53,** 671-690.

[6] CLARIDA,R.H. & TAYLOR,M.P. (1997). The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors. *Rev.Econ. & Stat.* **LXXIX,** 353.

[7] CLARIDA,R.H., SARNO,L., TAYLOR,M.P. & VALENTE,G. (2003). The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: a Step Beyond. *J.Int.Econ.* **60,** 61-83.

[8] CLARK,T.E. & MCCRACKEN,M.W. (2005) The power of tests of predictive ability in the presence of structural breaks. *J.Econometrics.* **124,** 1-31.

[9] CLEMENTS,M.P. & HENDRY,D.F. (1998). Forecasting Economic Time Series. *C.U.P.*

[10] CLEMENTS,M.P. & HENDRY,D.F. (1999). Forecasting Non-Stationary Economic Time Series. *The MIT Press*

[11] ENGEL, C. & HAMILTON,J.D. (1990). Long swings in the dollar: are they in the data and do markets know it? *American Economic Review.* **80,** 689-713.

[12] HYNDMAN (1995). Highest-density forecast regions for nonlinear and nonnormal time series. *J.Forecasting* **14,** 431-41.

[13] HYNDMAN (1996). Computing and graphing highest-density regions. *Am.Statistician* **50,** 120-6.

[14] KOOP,G. (1992). Model Uncertainty. *Rev.Econ. & Stat.* **74.**

[15] MEESE,R. & ROGOFF,K. (1983). Empirical Exchange Rate Models of the Seventies: Do They Fit Out Of Sample? *J.Int.Econ.* **14,** 3-24.

[16] PESARAN,M.H., PETTENUZZO,D., & TIMMERMAN,A. (2006). Forecasting Time Series Subject To Multiple Structural Breaks. *Rev.Econ.Studies,* **73,** 1057-1084.

[17] PESARAN,M.H., & TIMMERMAN,A. (2005). Small sample properties of forecasts from autoregressive under structural breaks. *J.Econometrics,* **129,** 183-217.

[18] PESARAN,M.H., & TIMMERMAN,A. (2007). Selection of estimation window in the prescence of breaks. *J.Econometrics,* **137,** 134-161.

[19] PHILLIPS,P.C.B. (1998). New tools for understanding spurious regression. *Econometrica,* **66(6),** 1299-1325.

[20] SARNO,L. & TAYLOR,M.P. (2002). The Economics of Exchange Rates. Cambridge University Press.

[21] STOCK,J.H. & WATSON,M.W. (1996). Evidence on structural instability in macroeconomic time series relations. *J.Bus.Econ.Stats.* **14(1),** 11-30.

Chapter 4

Broken Trend Modelling of Exchange Rates

4.1 Introduction

Perron (1989) showed that a number of series that were thought to be unit root processes [see Nelson & Plosser (1982)] were actually better modelled as stationary processes around a deterministic broken trend. He showed that the different regimes of trending were driven by infrequent but significant shocks - the 1929 Crash and the 1973 oil shock. Cochrane (1991) pointed out that any realization of a unit root process will, after testing, be found to be stationary around an 'overly creative' broken trend formulation. Indeed Phillips (2005) criticized the Garcia & Perron (1996) analysis of mean shifts in the real rate of interest for allowing too many breaks. Perron's (1989) result was striking because his broken trend formulation was very simple (a one time change in the slope of a linear trend) and the break dates (1929 and 1973) were easily rationalizable. However, what if a broken trend analysis is not so convincing? For instance, suppose that in order to get a stationary de-trended series, the broken trend requires a large number of inexplicable break dates or overly complex interbreak behaviour. When should we start to suspect that the investigator has been 'overly creative'?

In this chapter I develop a broken trend representation for a realization of a time series that is general enough to allow varying degrees of complexity. The broken trend representation is an extension of the work of Phillips (2001, 2005b) and Phillips & Ploberger (2003) on representations of Brownian motion in terms of deterministic Fourier modes with random coefficients. With this representation, moderate shifts in the sample realization of the time series can be modelled with the Fourier modes - i.e., they are consistent with a realization from a random walk - and do not require extra breaks. I select the broken trend representation using information criteria. This penalizes overly complex (creative) broken trend models, i.e., those with too many breaks or Fourier modes.

For the empirical application of the method I use the AUD-GBP exchange rate. Unit root tests of exchange rates (not augmented with broken trends) typically find that the unit root hypothesis cannot be rejected [Meese & Singleton (1982); Sarno & Taylor (2002)]. Furthermore, Meese & Rogoff (1983) demonstrated that the random walk model could not be beaten in forecasting of exchange rates out-of-sample. This phenomenon has been dubbed the 'Exchange Rate Disconnect Puzzle' - exchange rate movements appear to be uncorrelated with any fundamental economic variables. If exchange rates are really stationary about a broken trend processes then this has an implication for the Meese-Rogoff puzzle - if exchange rate movements might be explained at all, it might be by infrequent 'trend shocks' similarly to Perron (1989). A first step in explaining exchange rate movements via infrequent trend shocks is accurate inference about when these trends occur. I use the Bai & Perron (1998, 2003) break dating technique to do this.

One of the first broken trend models was the 'long swings' exchange rate model of Engel & Hamilton (1990). In this model the exchange rate change is the sum of a stochastic trend and a linear trend with a slope that changes according to a two-state Markov process. The breaks in the smoothed state sequence from the Engel & Hamilton (1990) 'long swings' model is another way of dating 'trend shocks' and I compare the two dating procedures in the chapter.

The chapter is organized as follows. Section 2 introduces a model called a broken trend process and explains how it can be estimated. Section 3 applies the model to the AUD-GBP exchange rate. Model selection criteria are used to select a model which is then compared to the interest rate differential. A Markov switching model is estimated for the same data and its break dating capacity is compared. In Section 5 I discuss the interpretatibility of the breaks. Section 5 concludes the chapter with some suggestions for further work.

4.2 Broken Trend Processes

In this Section I define a broken trend process, how it can be estimated using the method of Bai & Perron (1998, 2003), and it's implications for broken trend modelling.

4.2.1 The Karhunen-Loève Representation of a Stochastic Trend

Here I briefly review the Karhunen-Loève representation of a stochastic trend following Phillips (2005). Phiilips (2005) starts with a random walk $\Delta X_t = u_t$. The partial sum process is

$$X_{[n\cdot]} = \sum_{k=0}^{[n\cdot]} u_t. \tag{4.1}$$

If u_t is a sequence of i.i.d random scalars with finite variance σ^2, then the partial sum process obeys a functional central limit theorem, i.e.,

$$n^{-1/2}\sigma^{-1}X_{[n\cdot]} \rightarrow_d B(\cdot), \tag{4.2}$$

where $B(\cdot)$ is a Brownian motion.

A Brownian motion has a Karhunen-Loève (almost sure) representation:

$$B(r) = \sqrt{2}\sum_{k=1}^{\infty} \frac{\sin[(k-1/2)\pi r]}{(k-1/2)\pi}\xi_k, \tag{4.3}$$

where ξ_k are i.i.d $N(0, \sigma)$ (Papoulis, 1991).

Given a series of observations $\{X_t : t = 1, \ldots, n\}$ Phillips (2001) showed that the realizations of the ξ_k in the K-L representation (4.3) can be recovered by estimating a regression of the form,

$$n^{-1/2}(X_t - \mu) = \sum_{k=1}^{K} \hat{a}_k \phi_k(\frac{t}{n}) + \frac{\hat{u}_{tK}}{\sqrt{n}}. \tag{4.4}$$

where $\phi_k(r) = \sqrt{2}\sin[(k-1/2)\pi r]$. (The $\sqrt{n}$ scaling transforms X_t into an element of $D[0, 1]$, the space of cadlag functions on $[0, 1]$.) Phillips (2001) showed that $\hat{a}_k = \xi_k + o_{a.s.}(1)$.

Note that (4.3) is a representation of a stochastic process in terms of deterministic functions (trends) with random coefficients. A realization of the random walk can be generated from realizations of the stochastic ξ_k. To summarize: the partial sum process of a random walk converges to a Brownian motion, and this has a representation as a random series. The sum in (4.3) can be done over only a finite subset of Fourier modes, e.g., just the first K - in which case it could approximate almost any deterministic trend function. While full covergence to a stochastic trend (Brownian path) requires an infinite number of Fourier modes in (4.3), a large finite number of modes would be expected to be a reasonable approximation. It appears that there may be a smooth transition from deterministic to stochastic trends as K goes to infinity, and that there is a K_c that marks the crossover between the two. As pointed out in Phillips (1998) in the context of unit root testing against a trend stationary alternative, the either/or choice between deterministic and stochastic trends is ill-posed - both are just alternative representations of the same process.

4.2.2 A Broken Trend Representation

Here I develop a general representation of a broken trend by allowing the K-L representation to vary across sub-periods. Taking break fractions $\{l_i\}_1^{N-1} = \{l_1, l_2, \ldots, l_{N-1}\}$, we can allow the K Karhunen-Loève coefficients of X_t and μ to shift across sub-intervals:

$$n^{-1/2}X_t = \begin{cases} n_1^{-1/2}\mu_1 + \sum_{k=1}^{K_1} \hat{a}_{1k}\phi_k(\frac{t}{n}) + \frac{\hat{u}_{tK}}{\sqrt{n}} & \text{if} \quad r \in [0, l_1], \\ n_2^{-1/2}\mu_2 + \sum_{k=1}^{K_2} \hat{a}_{2k}\phi_k(\frac{t}{n}) + \frac{\hat{u}_{tK}}{\sqrt{n}} & \text{if} \quad r \in (l_1, l_2], \\ & \vdots \\ n_N^{-1/2}\mu_N + \sum_{k=1}^{K_N} \hat{a}_{Nk}\phi_k(\frac{t}{n}) + \frac{\hat{u}_{tK}}{\sqrt{n}} & \text{if} \quad r \in (l_{N-1}, 1]. \end{cases} \tag{4.5}$$

In (4.5) n_i is the number of observations in sub-interval i.

This is a very general representation that allows breaks in the level (μ) and the trend, i.e., the Fourier modes. Allowing different numbers of Fourier modes in each sub-interval (i.e., different K_i), also allows some regions to be modelled with trends of different degrees of stochasticity. In this paper, however, I restrict $K_i = K_j$ for all i and j so that the representation can be estimated using the Bai-Perron procedure.

4.2.3 Segmented Brownian Motion

This Subsection considers the $K \to \infty$ limit of the broken trend representation (4.5). Note that (4.5) is constructed by piecing together N different K-L representations, each defined on $[0,1]$. Using (4.3) and (4.4) each of these K-L representations converges to a (different) realization of Brownian motion on $[0,1]$. For instance, under the condition that $\mu_1 = O_p(n_1^{1/2})$, the first term in (4.5), converges to $x_1 + B_1(r)$ for $r \in [0,1]$, as $K_1 \to \infty$ and $n_1 \to \infty$, where $x_1 = \lim_{n_1 \to \infty} \frac{\mu_1}{\sqrt{n_1}}$.

Repeating this for each sub-interval, i.e., taking the limits $K_i \to \infty$ and $n_i \to \infty$, with $\mu_i = O_p(n_i^{1/2})$ and $\lim_{n_i \to \infty} \frac{\mu_i}{\sqrt{n_i}} = x_i$, yields a process I call segmented Brownian motion and denote by $B^\star(r; \{l_i\}_1^N)$:

$$B^\star(r; \{l_i\}_1^N) = \begin{cases} x_1 + B_1(r) & \text{if} \quad r \in [0, l_1], \\ x_2 + B_2(r) & \text{if} \quad r \in (l_1, l_2], \\ & \vdots \\ x_N + B_N(r) & \text{if} \quad r \in (l_{N-1}, 1]. \end{cases} \tag{4.6}$$

A sample path of this process is shown in Figure 4.1 where realizations from four Brownian motions are pieced together to form the segmented Brownian process.

Note that the Brownian motions only differ in their means $x_1, \ldots, x_N$, i.e., $B_1(r) = B_2(r) = \ldots = B_N(r)$, implying that (in the $K \to \infty$ limit) the only breaks between regimes are level shifts. Hence, the segmented Brownian motion is a form of jump-diffusion process (with an unspecified jump intensity). Note that this is not the case for $K < \infty$ where trend elements do vary between regimes.

Clearly, the process (4.6), which is the $K \to \infty$ limit of (4.5) exhibits martingale behaviour in between breaks. It is possible that (4.5) could be a martingale even for $K < \infty$. Immediately after a break, it may not be possible for agents (using a time-t information set) to determine whether the post-break regime is stochastic or deterministic, i.e., it may take some time to estimate K and the parameters for the new regime. Also, agents' inference about whether a break has occurred will depend on the power of their break detection procedure.

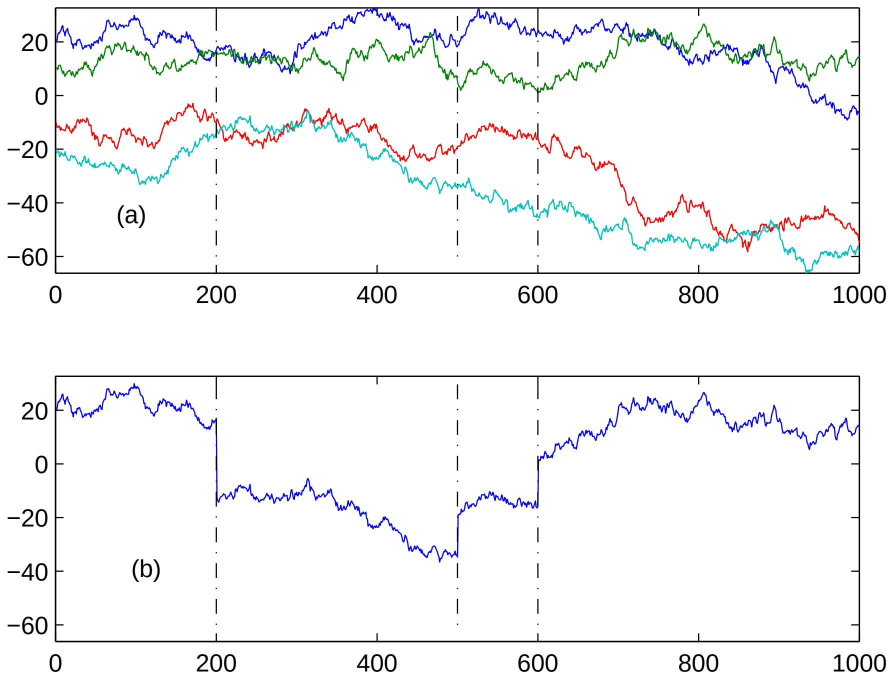

Figure 4.1: A schematic of segmented Brownian motion. Panel (a) shows sample paths of four random walks. Panel (b) Shows how segmented Brownian motion is constructed by selecting different sample paths in different segments.

4.2.4 Estimating the Broken Trend Representation

To estimate the Broken Trend Representation, we need to do this regression (4.4) on each subinterval of the partition generated by the (unknown) break fractions $\{l_i\}_1^N$. The Bai & Perron (1998, 2003) procedure is a way to estimate regression (4.4) with multiple unknown break dates.

The Bai & Perron break dating procedure relies on the following condition: If w_t is the time-t vector of regressors, and $\{T_1^0, \ldots T_m^0\}$ are the true break points, then,

$$\frac{1}{l_i} \sum_{t=T_i^0+1}^{T_i^0+[l_i \nu]} w_t w_t' \rightarrow_p Q_i(\nu), \tag{4.7}$$

which is non-random positive definite, uniformly in $\nu \epsilon [0, 1]$.

If this holds, then the estimates of the break fractions are consistent. This condition rules out trending regressors but does allow regressors of the form $\{\phi(t/T)^p\}_{p=0}^P$. Fortunately, the K-L basis functions are of this form, so we can use the Bai & Perron break date estimation procedure.

4.2.5 Eliminating Nonessential Breaks

As mentioned above, Phillips (2005) criticized the Garcia & Perron (1996) analysis for allowing too many breaks. Essentially, by only allowing the conditional mean to shift between regimes, any large moves in the realized interest rate were fitted with an extra break. The advantage of allowing all Fourier coefficients to vary between regimes (as is done here) is that interbreak moves that are consistent with a realization of a Brownian motion aren't modelled with unnecessary mean shifts. Only those moves that are inconsistent with Brownian motion will be modelled with a break, i.e., only the large moves. This should result in fewer, but significant estimated break dates, avoiding the danger that the broken trend formulation is 'overly creative'. In addition, we would hope that the dates of these large breaks might be easier to rationalize.

Recalling that for a Brownian motion $B(r)$,

$$B(r + \delta) - B(r) \sim N(0, \delta), \tag{4.8}$$

only increments of the (scaled and normalized) series of $O(\sqrt{\delta})$ would be regarded as breaks. When considering the whole series $\{X_t : t = 1, \ldots, n\}$

with unconditional variance σ_X^2, this condition is that an increment must be $O(\sigma_X \sqrt{n})$ to be regarded as a break.

4.3 Empirical Application: The AUD-GBP 1997 - 2004

In this Section we use the theory in Section 3 to identify the trend breaks in the AUD-GBP exchange rate.

4.3.1 Data

Daily data from 1/1/1997 until 16/7/2004 was downloaded from Datastream. This data was then sampled at a weekly frequency by selecting the observation for the Wednesday of each week. This produced a time series of 394 observations from 1997-1 to 2004-30 (using a year-week format).

4.3.2 Model Selection

For any time series it is possible to get a fit as close as desired by either increasing the number of K-L modes N_K or the number of allowed structural breaks N_B (with the sample size being the upper limit). These are two extremes that are not likely to be informative. If we denote the class of broken trend models by $BT(N_K, N_B)$ then we want to select the model within this class that provides the best in-sample fit while being parsimonious. This suggests that we use information criteria to select a model. Secondly, since one of the aims of this paper is break dating in exchange rate data, we would also like the model to find breaks that make sense when compared with economic variables. In this paper we compare the inferred breaks with the interest rate differential between Australia and the U.K.

The information criteria we employ are the Bayes Information Criterion (BIC) and the modified BIC of Liu, Wu, and Zidek (1997) which we denote by LWZ. In our regressions, the BIC always selected the model with the maximum number of allowed breaks. By contrast, the LWZ criterion selected more parsimonious models. The LWZ information criteria for all models are listed in Table 4.1. The model with smallest LWZ is $BT(4, 2)$.

The models selected by the LWZ criterion for each number of modes are fitted in Figures 4.2 to 4.5. Figure 4.2 is a model that only allows shifts in the

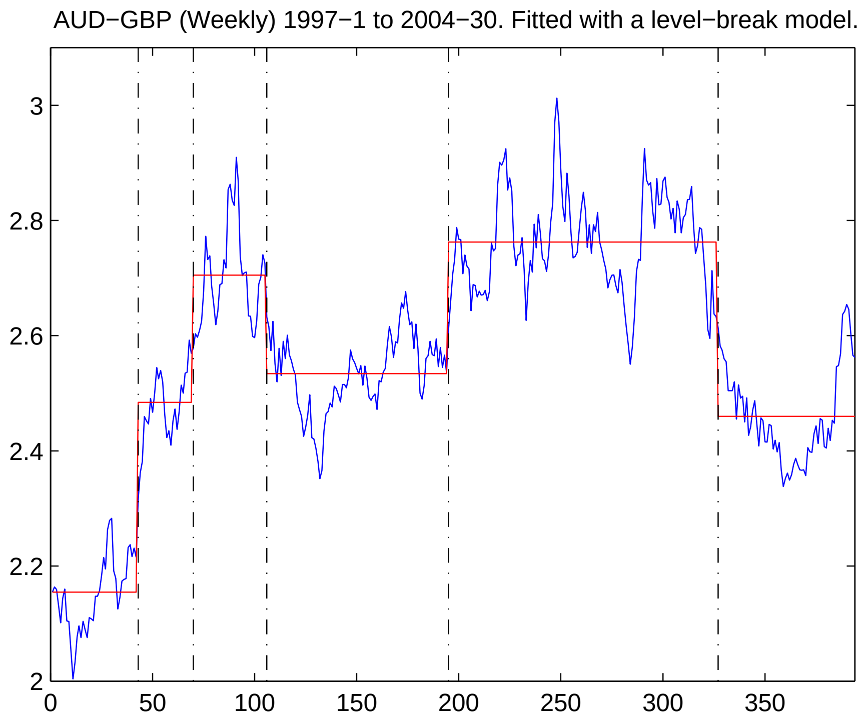

Figure 4.2: AUD GBP exchange rate fitted with 1 K-L mode and 5 breaks.

mean. This is similar to the Garcia & Perron (1996) model for US interest rates. The estimated break dates for this model are $\{43, 70, 106, 195, 327\}$. Looking at Figure 4.2 the model does seem to pick up shifts in the conditional mean of the series. However, it appears that in certain places these shifts in conditional mean are reached via trends. This is particularly the case up to observation 100.

This suggests that a model that allows shifts in the mean and trend might be better suited for the AUD-GBP series. The fit of the model selected out of this class, i.e, $BT(2, 8)$ is shown in Figure 4.3. The estimated break dates for this model are $\{42, 86, 133, 218, 245, 289, 320, 371\}$. Of concern with this model is the proliferation of breaks needed to fit the series. In the next Section I will discuss whether the breaks identified by this model can be rationalized/interpreted.

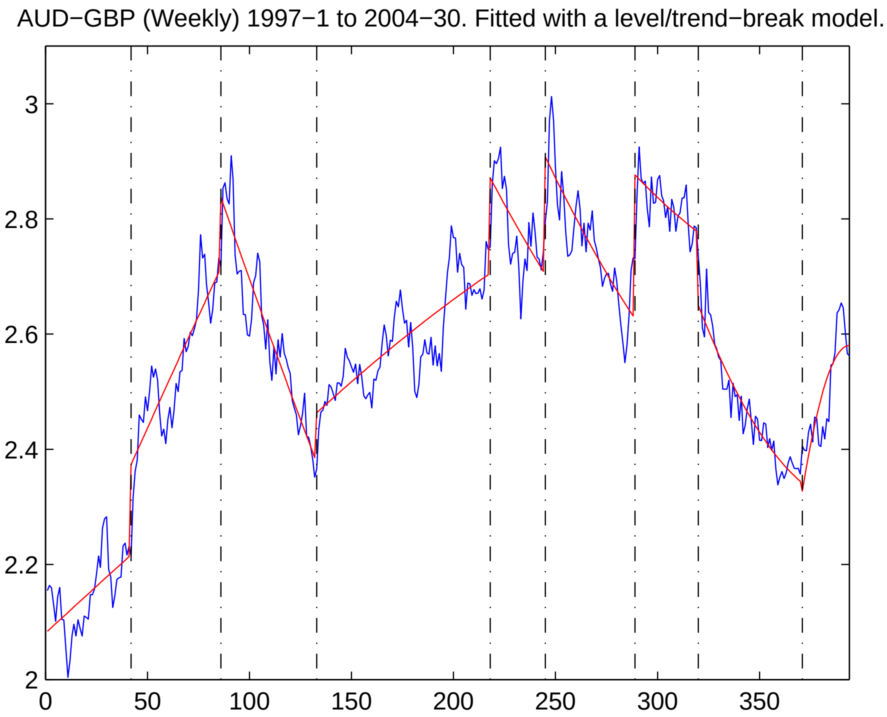

Figure 4.3: AUD GBP exchange rate fitted with 2 K-L modes and 8 breaks.

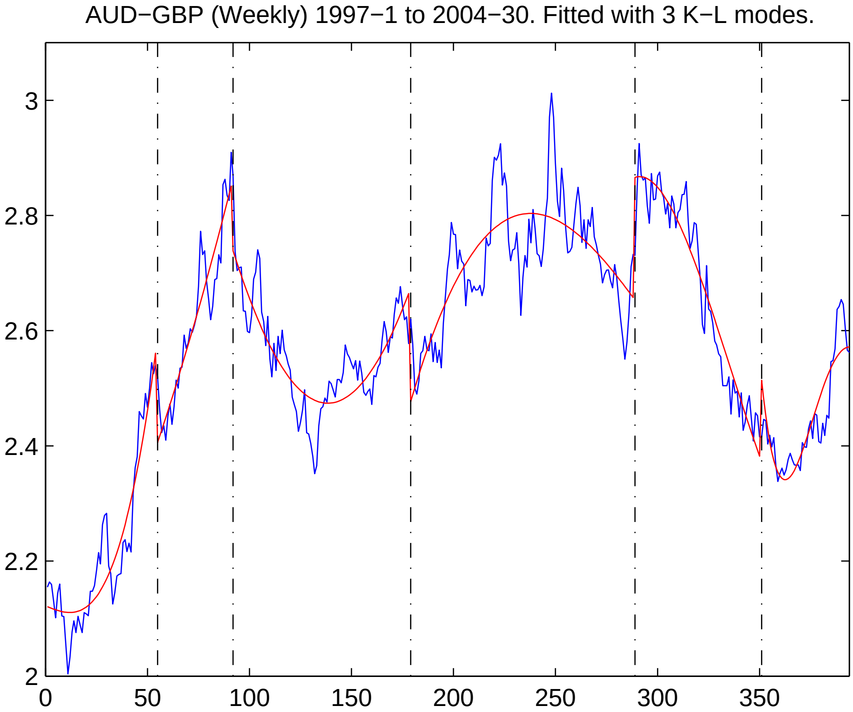

Figure 4.4: AUD GBP exchange rate fitted with 3 K-L modes and 5 breaks.

Figure 4.4 shows the fit of the $BT(3,5)$ model. The estimated breaks in this occur at $\{55, 92, 179, 289, 351\}$. While the fit of the model is good, the extra Fourier mode allows (apparent) trend shifts, e.g., around observation 130, to be modelled without breaks. This applies more strongly to the model $BT(4,2)$ in Figure 4.5 where only two breaks are needed to fit the data better than all the other models (according to the LWZ criterion). Notice that the two breaks are really jumps - sharp mean shifts that cannot be well fit by the K-L modes. These jumps are $\{86, 282\}$. In the next Section I discuss the interpretability of these two break dates.

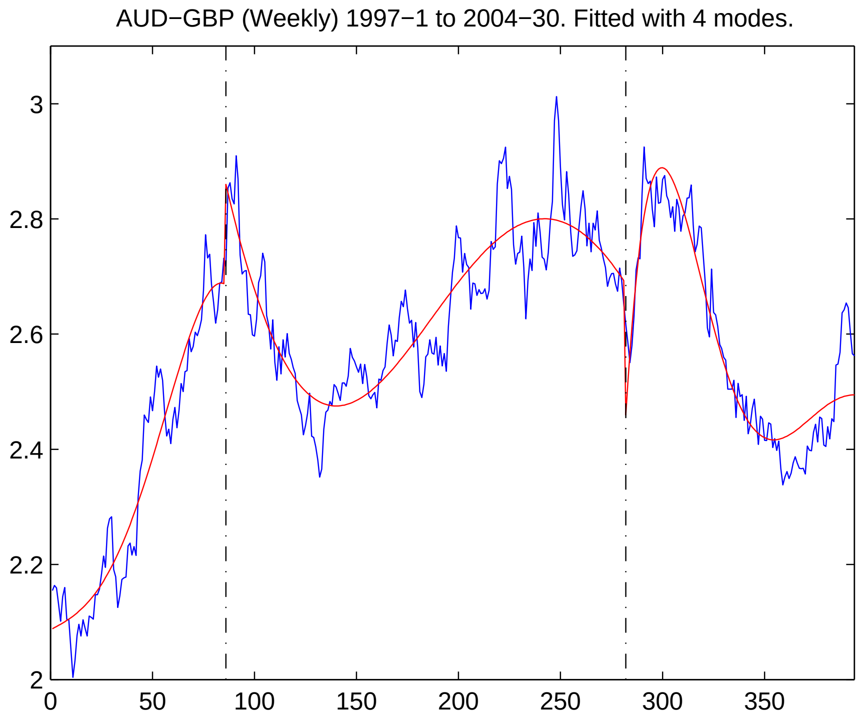

Figure 4.5: AUD GBP exchange rate fitted with 4 K-L modes and 2 breaks.

4.3.3 A Markov Switching Long Swings Model

The long swings model of Engel & Hamilton (1990) is a popular model for exchange rate trends. Because the model is also used for dating regime shifts in exchange rates, it is important to compare it to our regime dating approach.

Engel & Hamilton (1990) modelled the change in exchange rates as a drifting random walk, but where the drift (and variance of the innovation) depends on an unobserved 2-state Markov chain. They modelled the percentage change in the exchange rate model as,

$$100\frac{\Delta E_t}{E_t} \sim \begin{cases} N(\mu_1, \sigma_1^2) & \text{if } s_t \text{ is regime 1,} \\ N(\mu_2, \sigma_2^2) & \text{if } s_t \text{ is regime 2,} \end{cases} \tag{4.9}$$

where s_t is the state (1 or 2) of the Markov chain in period t. The transition probabilities for the chain to go from state i to j are denoted by p_{ij}. Persistent trends (or long swings) are reflected by p_{11} and p_{22} having values near 1.

There is an issue with the frequency of the data to be fit by a long swings model. When the data are sampled at daily, weekly or monthly frequencies, we do not find persistent trends. For instance at a weekly frequency, we find $p_{11} = 0.767$ and $p_{22} = 0.388$. We follow Engel & Hamilton (1990) and sample our data at a quarterly frequency. This results in a time series of 30 observations.

The model was estimated by an implementation of the EM algorithm written in Gauss by James Hamilton (Engel & Hamilton, 1990). This program was run in OxGauss. The estimates are reported in Table 4.2. These estimates are consistent with the hypothesis that exchange rates have persistent regimes (high p_{11} and p_{22}), and that these regimes comprise up ($\mu_1 > 0$) and down ($\mu_2 < 0$) trends.

Because of the necessity of sampling the data at a low frequency it is not possible to make a direct comparison between the fit of the long swings model and the broken trend models - essentially we have different dependent variables. However, we can approximate the information criteria on a weekly basis by scaling the quarterly error variance. In particular, if we accept the $\sqrt{t}$ scaling of volatility then $\sigma_{13}^2 = 13\sigma_1^2$. The error variance (not conditioned on the regime) is,

$$\sigma^2 = p_1\sigma_1^2 + p_2\sigma_2^2.$$

For our data $p_1 = 12/30$ and $p_2 = 18/30$ so $\sigma^2 = 1.446$. Hence the BIC is $\ln \sigma^2 + k \ln T/T = 1.049$. This is much larger than the BIC for all the broken

Number of Breaks	Model 1 mode	2 modes	3 modes	4 modes
0	-3.175	-3.489	-3.918	-3.946
1	-3.811	-4.080	-4.461	-4.902
2	-4.052	-4.729	-4.814	-5.233^{**}
3	-4.572	-4.827	-4.984	-5.190
4	-4.584	-4.926	-5.149	-5.146
5	-4.779^{*}	-5.091	-5.157^{*}	-5.110
6	-4.767	-5.139	-5.148	-5.090
7	-4.756	-5.164	-5.156	-5.101
8	-4.753	-5.199^{*}	-5.138	-5.088
9	-4.749	-5.176	-5.134	-5.046

Table 4.1: LWZ information criteria for models in the class $BT(N_K, N_B)$. A star indicates the model selected by LWZ conditional on the number of modes. A double star indicates the unconditional model selected. Note 1: The 1 mode model is a broken level model. Note 2: This model allows breaks in both level and trend.

Variable	Estimate
μ_1	3.957
	(1.686)
μ_2	-3.210
	(1.744)
σ_1^2	16.266
	(9.686)
σ_2^2	20.502
	(9.166)
p_{11}	0.778
	(0.176)
p_{22}	0.872
	(0.138)

Table 4.2: Markov switching long swings model estimates for AUD-GBP 1997-1 to 2004-30. Standard errors are in brackets.

trend models considered here and provides strong evidence against using the long swings model for break dating.

The Markov switching model allows us to determine the inferred state sequence, i.e., the sequence of states of the Markov chain most likely to have generated the data. The dated breaks are the changepoints in this state sequence. Figure 4.6 shows the smoothed state sequence, i.e., the sequence inferred from all of the data. The four break dates identified $(92, 131, 248, 365)$ are comparable with four of the breaks identified by the level-and-trend break model, $(86, 133, 245, 371)$. Looking at Figure 4.6, these appear to be the major trend breaks in the AUD-GBP over the sample period. However, the long swings model misses the level shift around observation 42 that all the broken trend models pick up. This is not surprising given that the long swings model does not attempt to model level shifts.

The break dating performance of the Markov switching model has been criticized in the context of business cycle dating (Harding & Pagan, 2002). In reply, Hamilton (2002) agreed that the simple two-state model was inadequate for dating, but has been superceded my more sophisticated models. The same could be said here in relation to trend break dating. Also, the Markov switching model does explicitly model the break dynamics. An extension of the broken trend model here would be to allow for stochastic break dynamics (e.g., with Poisson or Markov switching dynamics).

4.4 Explaining Trends with Monetary Policy Shocks

Phillips (2003) observed that economic theory rarely provides guidance on the form, or causes, of trending behaviour of a time series -'no-one understands trends, but everyone sees them in the data'. In searching for explanatory variables for the trend breaks identified in the previous Section, the interest rate differential, as a prime variable in theories of exchange rates, is an obvious first choice. The uncovered interest parity theory (UIP) or, equivalently, the forward discount hypothesis (FDH) predicts that an exchange rate should depreciate after a positive interest rate shock. The consensus is that UIP fails to hold empirically, i.e., a positive shock to the interest differential between, say, Australia and the UK, tends to lead to an appreciation of the Australian dollar, rather than the depreciation expected under UIP

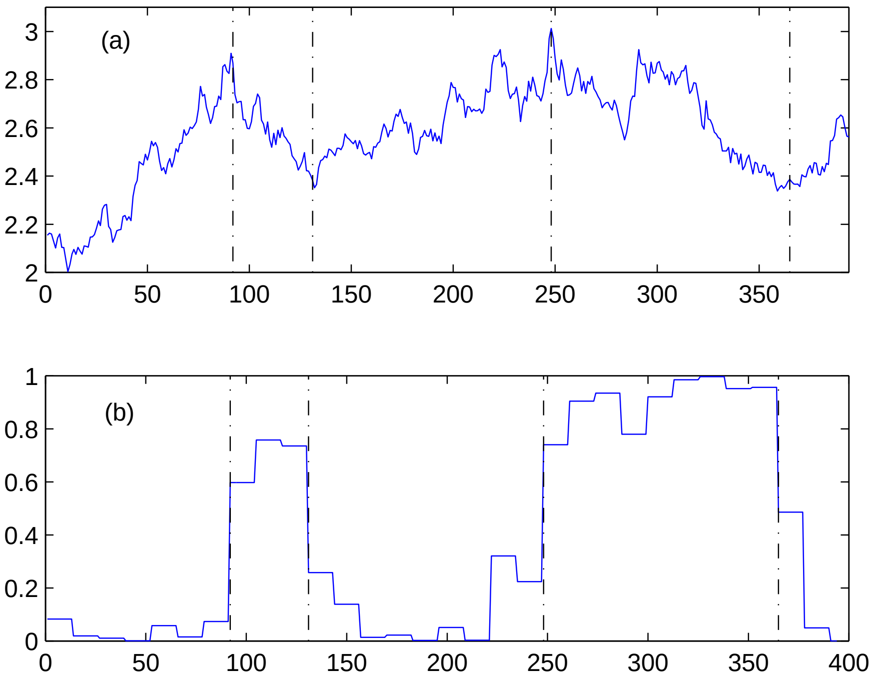

Figure 4.6: (a) The AUD GBP exchange rate and (b) the inferred state sequence from the 2-state Markov switching model.

[See Hodrick (1988), Engel (1996), Sarno & Taylor (2002)].

4.4.1 Monetary Shocks and Exchange Rate Trends

Eichenbaum & Evans (1995) found that contractionary shocks to US monetary policy led to persistent appreciations, i.e., trends, in the USD. They identified a 'delayed overshooting' effect whereby unanticipated contractionary monetary shocks cause the exchange rate to gradually appreciate before depreciating some months later. This empirical finding contradicts the instantaneous overshooting finding of Dornbusch (1976), i.e., an immediate appreciation followed by a depreciation. A dissenting voice to the existence of delayed overshooting is Faust & Rogers (2003) who find that it is sensitive to the identifying assumptions made in Eichenbaum & Evans (1995). Faust & Rogers (2003) find that the peak of overshooting can be either delayed or immediate, depending on the identification scheme used. Both immediate and delayed overshooting involve a gradual domestic depreciation (i.e., an upward trend) after the peak response of the exchange rate to the monetary shock.

Gourinchas & Tornell (2004) explained delayed overshooting via agents' misperceptions about the persistence of interest rate shocks. In particular, they showed that trending behaviour can occur when agents systematically underestimate the persistence of interest rate shocks. It will be useful for our analysis to briefly outline the main results of Gourinchas & Tornell (2004). They start by assuming that agents falsely believe that the interest differential x_t has the following dynamics:

$$x_t = z_t + \nu_t, \tag{4.10}$$

$$z_t = \lambda z_{t-1} + \epsilon_t. \tag{4.11}$$

In reality, $\sigma_\nu^2 = 0$, implying that the interest differential follows an AR(1) process with autoregressive coefficient λ. The variance ratio of the persistent to temporary shocks, $\eta = \sigma_\nu^2/\sigma_\epsilon^2$ is interpreted as the relative importance of transitory to permanent shocks. They differentiate between agents' subjective forecasts, e.g., $E_t^S[x_{t+1}]$ and the rational expectations forecast $E_t[x_{t+1}]$. Gourinchas & Tornel assume that 'subjective' UIP holds, i.e.,

$$E_t^S[e_{t+1}] - e_t = x_t, \tag{4.12}$$

and solve (4.12) forward for the exchange rate level yielding,

$$e_t = e_t^r + \frac{1}{1-\lambda}(E_t[x_{t+1}] - E_t^S[x_{t+1}]), \tag{4.13}$$

where,

$$e_t^r = \bar{e}_t - \frac{x_t}{1-\lambda}, \tag{4.14}$$

is the rational expectations exchange rate (that assumes $\eta = 0$). The term $\bar{e}_t$ is the conditional long-run equilibrium exchange rate [see Gourinchas & Tornell (2004) for a justification of this]. Under rational expectations (4.14) implies that a positive shock to the interest differential will result in an immediate appreciation of the currency (a decrease in e_t). The term $E_t[x_{t+1}] - E_t^S[x_{t+1}]$ in (4.13) is positive implying that the actual appreciation is less than under rational expectations.

I use the rational expectations solution (4.14) to do a heuristic analysis of the exchange rate trends we found in the last Section. Given that (4.14) holds in every period, we can subtract the $t+j$ period condition from (4.14) to yield,

$$\Delta_{t+j}e_t = -(1+\lambda)^{-1}\Delta_{t+j}x_t, \tag{4.15}$$

where $\Delta_{t+j}e_t = e_{t+j} - e_t$ and I assumed $\bar{e}_t = \bar{e}_{t+j}$. This clearly shows the negative relation between the interest differential shock and the exchange rate change. The condition applies to unexpected interest differential shocks, but it is plausible that the predictability of the realized shock decreases rapidly with the horizon j. Note that a similar relation holds when agents misperceive the persistance of shocks (i.e., using (4.13) but the speed of change is less). I propose to use (4.15), conditioned on the the estimated breaks, to rationalize the exchange rate breaks from the previous Section.

4.4.2 Rationalizing the Trend Breaks

In this Subsection I use interest rate data to examine the extent to which monetary policy shocks can 'explain' the trends estimated by the $BT(4,2)$ model. As Phillips (2005b) emphasizes, the search for a formalism for explaining trends is a current research frontier - apart from the I(1)-I(0) dichotomy of cointegration, we don't have a theory for relating time series of different fractional orders of integration. Given this, I take some licence in heuristically relating the interest differential time series with the AUD-GBP

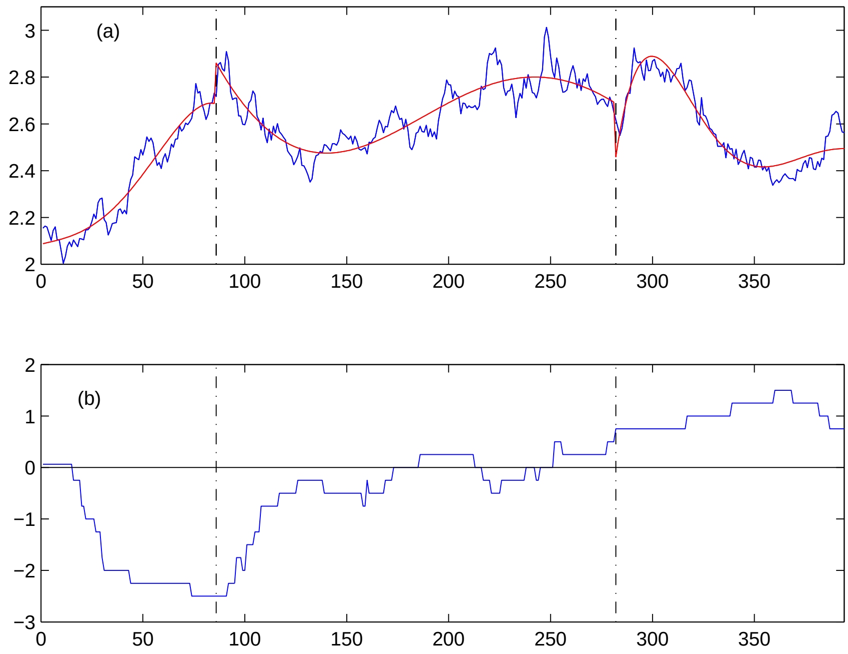

Figure 4.7: Panel (a): AUD GBP exchange rate fitted with 4 K-L modes and 2 breaks. Panel (b): The cash rate differential between Australia and the UK. (The regimes in (b) are just those from (a).)

exchange rate. The data series we use here for monetary shocks is the difference between the Reserve Bank of Australia's target for the overnight cash rate and the Bank of England's repo rate. This is plotted in panel (b) of Figure (4.7). Panel (a) of Figure (4.7) plots the break dates selected by the $BT(4,2)$ model chosen from the LWZ criterion.

A qualitative analysis of Figure (4.7) shows that, conditional on the $BT(4,2)$ regimes, the negative relationship (4.15) between interest differential shocks and exchange rate trends appears to hold. Table 4.3 shows the changes in the AUD-GBP exchange rate and the interest differential for the three break fractions identified by the $BT(4,2)$ model. Also included in the table are the same calculations assuming a break at 371. This break was identified by the BT models with both 3 and 2 modes, but, as it occurs at

Subperiod	Δs_t	Δx_t
$0 \rightarrow 86$	+0.232	-2.56%
$86 \rightarrow 282$	-0.038	+3.25%
$282 \rightarrow 394$	-0.021	0.00%
$282 \rightarrow 371$	-0.084	+0.50%
$371 \rightarrow 394$	+0.063	-0.50%

Table 4.3: Changes in the log exchange rate and the interest differential over different break fractions.

the end of the sample, may not have been large enough to be identified as a break by the 4-mode model.

The slope of the OLS line fitting the first three points in Table 4.3 is -4.46. Using (4.15) this corresponds to the autoregressive coefficient of the interest differential $\lambda = 0.78$. With the possible break at time 371 we get a slope of -4.69 and $\lambda = 0.79$. The actual interest differential has an estimated coefficient of 0.997 with a 95% confidence interval of $(0.988, 1.01)$. While this disagrees with (4.15) we note that series with infrequent mean shifts can masquerade as unit root processes. The appearance of panel (b) of Figure (4.7) suggests this might be a possibility. Table 4.3 certainly is in qualitative agreement with (4.15). The fact that the breaks identified by the $BT(4, 2)$ model can be rationalized is an indicator that it the estimated break dates are significant. By contrast, Panel (b) of Figure (4.8) shows the breaks of the $BT(2, 6)$ model along with the interest differential. This shows a series of hard to interpret breaks in the region $(133, 250)$ not to mention the break at 320. It appears that the $BT(2, 6)$ is not doing much more than curve fitting the AUD-GBP with spurious breaks (Phillips, 2005a).

4.5 Conclusion

This chapter proposed a new broken trend model for economic and financial time series. The method combines the general trend representation of Phillips (2005) with the popular Bai Perron break dating method. As noted by (Cochrane, 1991) any series can be made stationary around a sufficiently complex deterministic trend break function. The inclusion of extra modes precludes the need for unrealistic or overly numerous breaks, avoiding

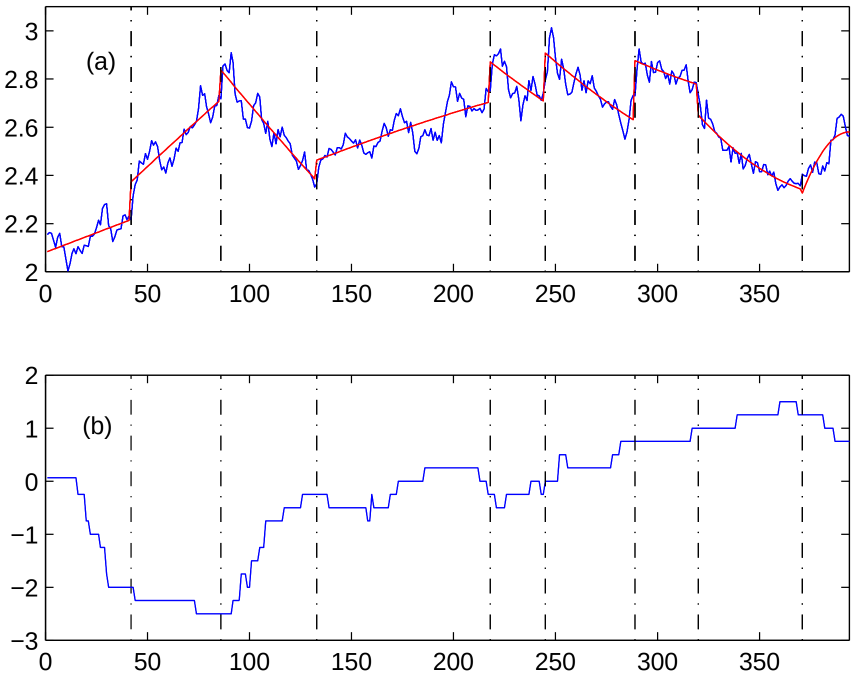

Figure 4.8: Panel (a): AUD GBP exchange rate fitted with 2 K-L modes and 8 breaks. Panel (b): The cash rate differential between Australia and the UK. (The regimes (b) are those estimated in (a).)

the curve fitting criticism when trend breaks are estimated (Phillips 2005a). The method does not require an either/or choice on whether the trends are modelled as stochastic or deterministic processes - the transition from deterministic to stochastic occurs smoothly as the number of Fourier modes selected by the information criterion increases. I showed that as the number of modes goes to infinity, the broken trend model converges to a realization of a jump-diffusion process.

The broken trend process was fitted to an AUD-GBP exchange rate series and modelled the breaks well. On an information criterion basis the method was shown to be superior to that of using the smoothed state sequence from a Markov switching long swings model. The break dates specified by the best-fitting model (on an information criterion basis) were rationalizable using the cash rate differential between Australia and the UK.

Further work might jointly model the breaks in the interest differential and the exchange rate in an attempt to 'explain' the trends in the exchange rate. Additional further work might look at model selection criteria within the broken trend process framework and different information criteria, e.g., the PIC (Ploberger & Phillips, 2003).

Bibliography

[1] BAI,J. & PERRON,P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66,** 47-78.

[2] BAI,J. & PERRON,P. (2003). Computation and analysis of multiple structural change models. *J.Appl.Econometrics* **18,** 1-22.

[3] COCHRANE,J. (1991). Pitfalls and Opportunities: What macroeconomists should know about unit roots. Comment. *NBER Macroeconomics Annual 1991 Vol.6.* 201-210.

[4] DORNBUSCH,R. (1976). Expectations and exchange rate dynamics. *J.Political Economy* **84,** 1161-1176.

[5] EICHENBAUM,M. & EVANS,C.L. (1995). Some empirical evidence on the effects of shocks to monetary policy on exchange rates. *The Quarterly Journal of Economics.* November, 1995, 975-1009.

[6] ENGEL,C. (1996). The forward discount anomaly and the risk premium: a survey of recent evidence. *J.Empirical Finance.* **3,** 123-192.

[7] ENGEL,C. & HAMILTON,J.D. (1990). Long swings in the dollar: are they in the data and do markets know it? *American Economic Review.* **80,** 689-713.

[8] FAUST,J. & ROGERS,J.H. (2003) Monetary policy's role in exchange rate behavior. *J.Monetary.Econ.* **50,** 1403-1424.

[9] GARCIA, R. & PERRON,P. (1996). An analysis of the real rate of interest under regime shifts. *Review of Economics & Statistics.* 111-125.

[10] GOURINCHAS, P-O. & TORNELL,A. (2004). Exchange rate puzzles and distorted beliefs. *J.Int.Econ* **64,** 303-333.

[11] HAMILTON,J. (2002). Comment on: A comparison of two business cycle dating methods. *J.Economic Dynamics & Control.* **27,** 1691-1693.

[12] HARDING,D. & PAGAN,A. (2002). A comparison of two business cycle dating methods. *J.Economic Dynamics & Control.* **27,** 1681-1690.

[13] HODRICK,R. (1988) The empirical evidence on the efficiency of forward and futures foreign exchange markets. *Fundamentals of Pure and Applied Economics.* Harwood Academic. N.Y.

[14] LIU, J., WU, S., & ZIDEK,J.V. (1997). On segmented multivariate regressions. *Statistica Sinica.* **7,** 497-525.

[15] MEESE,R. & ROGOFF,K. (1983). Empirical Exchange Rate Models of the Seventies: Do They Fit Out Of Sample? *J.Int.Econ.* **14,** 3-24.

[16] MEESE,R. & SINGLETON,K. (1982). On unit roots and the empirical modelling of exchange rates. *J.Finance.* **37(4),** 1029-1035.

[17] NELSON,C. & PLOSSER,C. (1982). Trends and random walks in macroeconomic time series. *J.Monetary Economics.* **10,** 139-162.

[18] PERRON,P. (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica.* **57,** 1361-1401.

[19] PAPOULIS,A. (1991). Probability, Random Variables, and Stochastic Processes. (Third Edition.) McGraw-Hill.

[20] PHILLIPS,P.C.B. (1998) New tools for understanding spurious regressions. *Econometrica.* **66(6),** 1299-1325.

[21] PHILLIPS,P.C.B. (2001) New unit root asymptotics in the prescence of deterministic trends. *J.Econometrics.* **11,** 323-353.

[22] PHILLIPS,P.C.B. (2003) Laws and limits of econometrics. *The Economic Journal* **13,** C26-C52.

[23] PHILLIPS,P.C.B. (2005a) Econometric Analysis of Fisher's Equation. *American Journal of Economics & Sociology* **64(1),** 125-168.

[24] PHILLIPS,P.C.B. (2005b) Challenges of trending time series econometrics. *Mathematics and Computers in Simulation.* **68,** 401-416.

[25] PLOBERGER,W. & PHILLIPS,P.C.B. (2003) Empirical limits for time series econometric models. *Econometrica. 71,* 627-673.

[26] SARNO,L. & TAYLOR,M.P. (2002). The Economics of Exchange Rates. Cambridge University Press.

Chapter 5

Overview

This thesis examined the impact of trend breaks on exchange rate forecasts. The thesis was motivated by the question of whether trend breaks in the exchange rate can explain, at least in part, the poor performance of exchange rate forecasting models.

Chapter 2 developed a degree of difficulty measure that applies to series subject to Markov structural breaks - as exchange rates appear to be [Engel & Hamilton (1990)]. This Chapter showed that different out-of-sample periods for such Markov switching processes can differ in their forecast difficulty. This is a potential factor in explaining the observed time-varying predictive ability of some exchange rate models. The degree of difficulty is the difference between the mean square forecast errors of the optimal univariate forecast and the random walk forecast - a quantity first derived in Dacco & Satchell (1999). Chapter 2 showed how this degree of difficulty measure can be calculated for any forecast horizon and applied it to forecasts of the AUD-USD exchange rate.

Chapter 3 developed a theory that explains how multimodal forecast error densities can be generated when forecasting a series subject to trend breaks, drawing on the work of Cavaliere (2003). This multimodality is a problem for forecast statistics calcluated from this error density. Similarly, it will be misleading if used for value-at-risk (or similar) calculations. Chapter 3 also showed how the impact of the trend breaks on a particular forecast statistic can be gauged - essentially, the sample forecast statistic can be compared with the asymptotic value calculated by simulation.

Chapter 4 was motivated by comments in Phillips (2005) about the 'impoverished class of trend mechanisms' available to model time series. In

Chapters 2 and 3 I employed the Markov switching 'long swings' model [Engel & Hamilton (1990)] in which inter-break exchange rate behaviour is modelled as a linear trend. Chapter 4 uses the Karhunen-Loève representation from Phillips (2005) to allow for more general inter-break trends. Essentially, the general trend formulation can model reasonably complex behaviour without breaks, while simpler trend formulations require the insertion of breaks between two or more trends. These breaks are likely to be spurious, in the sense that they are compensating for an impoverished trend formulation. Chapter 4 shows that only large breaks survive as the order of the broken trend model is increased - only those of $O(\sqrt{T})$ for sample size T. There is a good chance that these large breaks are explicable. Chapter 4 shows that the breaks identified for the AUD-GBP exchange rate do appear to be rationalizable in terms of monetary policy shocks.

As discussed in Chapter 1, one of the motivations for this work was to see if trend breaks can explain, in part, the exchange rate disconnect puzzle that emerged from Meese & Rogoff (1983). The degree of difficulty measure can be used in an ex-post analysis of a forecast - so, a seemingly poor forecast may not be so bad when allowance is made for the degree of difficulty. The approach of Chapter 3 can also be used ex-post to determine the impact of trend breaks on a particular forecast. These tools can't provide a conclusive answer to the question of whether the exchange rate disconnect puzzle is caused by trend breaks (other factors are probably present). However, they make a start towards quantifying the impact of breaks on exchange rate forecasts.

Bibliography

[1] CAVALIERE,G. (2003). Asymptotics for unit root tests under Markov regime-switching. *Econometrics Journal.* **6,** 193-216.

[2] DACCO,R. & SATCHELL,S. (1999). Why do regime-switching models forecast so badly? *J.Forecasting.* **18,** 1-16.

[3] ENGEL, C. & HAMILTON,J.D. (1990). Long swings in the dollar: are they in the data and do markets know it? *American Economic Review.* **80,** 689-713.

[4] MEESE,R. & ROGOFF,K. (1983). Empirical Exchange Rate Models of the Seventies: Do They Fit Out Of Sample? *J.Int.Econ.* **14,** 3-24.

[5] PHILLIPS,P.C.B. (2001). Trending time series and macroeconomic activity: Some present and future challenges. *J.Econometrics.***100,** 21-27.

[6] PHILLIPS,P.C.B. (2005) Challenges of trending time series econometrics. *Mathematics and Computers in Simulation.* **68,** 401-416.