NEAR EAST UNIVERSITY

Faculty of Engineering

Department of Computer Engineering

ECONOMIC FORECASTING
(using time series model)

Graduation Project
Com 400

Student :Ömer Faruk Tosun(20000465)

Supervisor :Assoc.Prof.Dr.Rahib Abiyev

Lefkoşa-2005

ACKNOWLEDGEMENT

I am grateful to my thesis advisor Assoc.Prof.Dr.Rahib Abiyev and M.A. Mehmet Göğebakan for helpings.And express my thanks for their guidence, encouragement through out the whole study. It is very much appreciated the time and patience that they have willing devoted when working with me.

Finally, I am grateful to my lecturers, my friends, and my family for giving me chance to complete the study as it is required.

Ömer Faruk Tosun
Lefkoşa - June 2005

ABSTRACT

This project codify extensive recent research on economic forecasting. When a forecasting model coincides with the mechanism generating the data (DGP) in an unchanging world, the theory of economic forecasting is well developed. Forecasts are the conditional expectation, are unbiased, and no other predictor has a smaller mean-square forecast error matrix. Cointegration does not markedly alter that conclusion. Much less is known about forecasting in a non-stationary and evolving world, especially when the model and DGP differ.

The main challenges facing a theory of economic forecasting, however, are to explain the recurrent episodes of systematic mis-forecasting observed historically, and to develop methods which avoid repeating such mistakes in future. To construct an empirically-relevant theory, we allow the model to be mis-specified for a DGP which alters unexpectedly at unknown times. We are able to deduce: what types of changes in economic behaviour are most deleterious for the main types of economic forecasting models; what can be done to improve the performance of such models in the face of structural breaks; and what factors and mistakes do not seem to cause forecast failure.

First, the framework and basic concepts are explained. Most measures of forecast accuracy lack invariance to isomorphic representations of models: invariant measures would help avoid artefacts, but even if forecast accuracy remains ambiguous, forecast failure does not. The model class explored is a vector autoregression (VAR) in integrated-cointegrated variables – a vector equilibriumcorrectionmodel (VEqCM) – subject to structural breaks. VARs in levels and differences are special cases; open models are not considered. The role of causal information in economic forecasting is studied, because non-causal variables may outperform when the model and DGP differ, and the latter suffers structural breaks. This difference from a constant-parameter world helps explain the practical procedures of macro-econometric forecasters.

A taxonomy of forecast errors is delineated for mis-specified, data-based models, facing structural change in the forecast period, from a mis-measured forecast origin. Deterministic factors, especially shifts in equilibrium means, are the main culprit of systematic forecast failure, while other factors influence excess variability. The theory is applied to forecasting in the face of structural breaks, focusing on the differential robustness of differenced VARs and VEqCMs. The distinction between equilibrium correction (the embodiment of cointegration) and error correction (a mechanism for keeping a model on track) is stressed.

The roles of parsimony and collinearity in forecasting highlight the importance of including important, and excluding irrelevant, but changing, variables. Unanticipated deterministic breaks are crucial, as Monte Carlo experiments illustrate. Differencing and intercept corrections can robustify forecasts against such shifts. Empirical examples illustrate the power of the resulting theory.

Another surprise is the difficulty of detecting shifts in parameters other than those concerning deterministic terms. This too is shown, and the worrying implications for 'impulse-response analyses' highlighted.

A linked set of notes addresses the issue of econometric modelling from a general-to-specific (Gets) approach. Disputes about econometric methodology partly reflect a lack of evidence on alternative approaches. We reconsider model selection from a computer-automation perspective, focusing on PcGets. Starting from a general congruent model, standard testing procedures eliminate statistically-insignificant variables, with diagnostic tests checking the validity of reductions, ensuring a congruent final selection. Since jointly selecting and diagnostic testing has eluded theoretical analysis, we study modelling strategies by simulation. Monte Carlo experiments show that PcGets recovers the DGP specification from a general model with size and power close to commencing from the DGP itself. Finally, we also consider the role of selection in forecasting, theory testing, and policy evaluation, and demonstrate the advantages of a Gets approach in all three, with the caveat that forecasting still requires non-standard implementations of estimated models to protect against deterministic shifts.

CONTENTS

INTRODUCTION

To forecast simply requires making a statement about the future. Such statements may be well, or badly, based, accurate or inaccurate on average, precise or imprecise, and model-based or informal: thus, forecasting is potentially a vast subject. The general framework is sketched in §4, and alternative methods of forecasting discussed in §5. We will focus on methods that can be quantitatively evaluated, and hence are model-based, specifically econometric formulations.

Econometric forecasting models usually comprise systems of relationships between variables of interest (such as GNP, inflation, exchange rates etc.), where the relations are estimated from available data, mainly aggregate time-series. The equations in such models have three main components: deterministic terms (like intercepts and trends) that capture the levels and trends, and whose future values are known; observed stochastic variables (like consumers' expenditure, prices, etc.) with unknown future values; and unobserved errors all of whose values (past, present and future) are unknown, though perhaps estimable in the context of a model. Any, or all, of these components, or the relationships between them, could be inappropriately formulated in the model, inaccurately estimated, or could change in unanticipated ways. All nine types of mistake could induce poor forecast performance, either from inaccurate (i.e., biased), or imprecise (i.e., high variance) forecasts. Instead, we find that some mistakes have pernicious effects on forecasts, whereas others are relatively less important in most settings. Moreover, 'correcting' one form of mistake may yield no improvement when others remain. For example, more sophisticated methods for estimating unknown parameters will not help when the problem is an unanticipated trend shift.

Section 3 presents an overview, intentionally lighthearted, of the forecasting enterprise, which nevertheless raises all the main problems and suggests possible solutions. Section 6 discusses sources of forecast failure, and §7 the main concepts needed (unpredictability, forecastability, horizon and moments). Then, §8 develops a forecast-error taxonomy. We set out the assumed form of the data generating process (DGP), and calculate the forecasting model's multi-step forecast errors when the DGP is assumed to change over the forecast period. This is the most mathematical part, but the algebra is fairly straightforward, and is used to direct our search for likely explanations of forecast failure. The expressions for the forecasting model's multi-step forecasts are simply an abstract algebraic representation of the forecasts produced by forecasting models, where the models first have to be specified and estimated. Given the DGP, we can then calculate the forecast errors, and break them down into a number of components. We have already motivated our choice of the form of the DGP, but this should not be taken too literally.

Section 9 considers how to measure forecast accuracy – a surprisingly difficult task – and notes the ambiguity in many measures. Then the role of causal information is examined in §10, again delivering a surprise, this time that irrelevant variables can dominate in forecast accuracy in the general framework we propose. This helps explain many aspects of current forecasting practice, and points towards the efficacy of intercept corrections and differencing. A formal taxonomy is presented in §11

Section 12 distinguishes between error and equilibrium correction. Somewhat paradoxically, models formerly known as 'error-correction models' do not 'error correct' in some states in which models that omit 'error-correction terms' do. This distinction is at the heart of understanding why Box-Jenkins timeseries method can prove hard to beat. Section 13 and 14 explain and illustrate models and methods that can help circumvent forecast failure once the potentially damaging change in economic conditions has occurred. Section 15 considers a number of factors traditionally assigned a role in forecast failure, but which, in the absence of parameter non-constancies, would appear to play only a minor role. Section 16 emphasizes the (non)detectability of breaks in VARs other than deterministic shifts, and §18 illustrates the analysis by the empirical example of UK M1. Finally, §19 considers some of the wider implications beyond the realm of forecasting.

1 ECONOMIC FORECASTING PROBLEM

Economies evolve over time and are subject to intermittent, and sometimes large, unanticipated shifts. Breaks may be precipitated by changes in legislation, sudden switches in economic policy, major discoveries or innovations, or even political turmoil, civil strife and war. Recent examples include the abolition of exchange controls, financial innovation, membership of the European Union, privatization, and the Gulf war. The models used to understand and forecast processes as complicated as national economies are far from perfect representations of behaviour. Moreover, the data series used in model building are often inaccurate, prone to revision, and may be available only after a non-negligible delay. Usually, forecasters are only dimly aware of what changes are afoot, and even when developments can be envisaged, may find it hard to quantify their likely impacts (e.g., the effects of Building Society demutualizations in the UK on consumers' spending in the 1980s). Thus, to understand the properties of economic forecasts requires a theory which allows for: a complicated and changing economy, measured by inaccurate data, using forecasting models which are mis-specified in unknown ways, possibly inconsistently estimated. Surprisingly, it is feasible to develop a theory based on these realistic assumptions, and these lecture notes explain the framework of that theory, highlight its main implications, and demonstrate its empirical relevance.

Such a theory reveals that many of the conclusions which can be established formally for correctlyspecified forecasting models of constant-parameter processes no longer hold. Instead, the theory gives rise to a very different set of predictions about the properties of forecasting tools. We have evaluated these implications both in specific empirical settings and using computer simulations, obtaining a fairly close concordance between theory and evidence. The findings confirm that despite its non-specific assumptions, a theory of forecasting which allows for structural breaks in an economic mechanism for which the econometric model is mis-specified in unknown ways, can provide a useful basis for interpreting, and potentially circumventing, systematic forecast failure in economics.

Our research shows that the treatment of 'equilibrium means' in forecasting models is a crucial factor in explaining forecasting performance. Even in evolving economies, equilibrium means exist which determine values towards which the relevant economy would adjust in the absence of further 'shocks': possible examples include the savings rate, the real rate of interest, the long-run growth rate, and the velocity of circulation. Economic equilibria usually involve combinations of variables, as with all the examples just cited. The key to understanding systematic forecast failure, and its avoidance, turns on four aspects of such equilibrium means. First, their specification and estimation: inadequate representations or inaccurate estimates of equilibrium means can induce poor forecasts. Secondly, the consequences of unanticipated changes in their values are pernicious: the economy then converges to the new equilibrium means, but the forecasting model remains at the old values. Thirdly, successfully modelling movements in equilibrium means can pay handsome dividends, even if only by using corrections and updates to offset changes. Finally, formulating models to minimize the impact of changes in equilibrium means is generally beneficial, even when the cost is a poorer representation of both the economic theory and the data. Various strategies can be adopted to help attenuate the impacts of shifts in equilibrium means, including intercept corrections, over-differencing, co-breaking, and modelling regime switches.

Shifts in equilibrium means inherently involve changes in the levels of some variables, and so entail 'deterministic shifts'. These shifts may occur within the model, or may reflect other changes in the economic mechanism. Unmodelled changes in the intercepts of models are obviously detrimental, but also, for example, non-zero-mean stochastic components may interact with breaks elsewhere in the economy to precipitate forecast failure. Relative to the role played by deterministic shifts, other forms of mis-specification seem to have a less pernicious effect on forecast accuracy. Indeed, the next most important cause of forecast failure, after shifts in deterministic factors over the forecast horizon, are mis-specifications of deterministic terms. For example, omitting a trend in a model when there is one in the data rapidly leads to large forecast errors. And the next source is mis-estimation of deterministic factors: for example, an inaccurately-estimated linear trend can induce serious forecast errors.

Sources of zero-mean forecast errors – such as model mis-specification, parameter-estimation uncertainty, inconsistent estimation, and 'shocks' – all appear much less important determinants of forecast failure, even though they may adversely affect forecast accuracy. Thus, the theory directs attention to the areas that induce forecast failure, and surprisingly suggests that zero-mean mistakes (which include problems such as omitted variables and residual autocorrelation) are of secondary importance. In turn, such results cast doubt on claims that imposing restrictions from general-equilibrium economic theory on forecasting models will improve forecast accuracy. However, some gains do seem to accrue from imposing valid long-run restrictions when the equilibrium means do not shift.

Similarly, the theory reveals that several potential sources of parameter-estimation uncertainty, including high correlations between the explanatory variables in models (usually called collinearity), and a lack of parsimony per se (sometimes called 'overparameterization') are not key culprits, although in conjunction with breaks elsewhere, they may induce serious problems. For example, even when the parameters of a forecasting model remain constant, a break in the correlation structure of the explanatory variables can induce poor forecasts when collinearity is severe (due to variance effects from the least-significant variables). Moreover, the theory indicates how to determine if this last combination is the cause of a forecast mistake: although the ex ante errors are similar to other sources, problems should not be apparent ex post (e.g., collinearity would vanish, and precise coefficient estimates appear), so a clear demarcation from deterministic shifts is feasible in practice, albeit only after the event. An indirect consequence is that little may be gained by inventing 'better' estimation methods, especially if the opportunity cost is less effort devoted to developing more robust forecasting models.

Indeed, in a world plagued by non-constancies, it cannot be demonstrated that effort devoted to model specification and estimation will yield positive returns to forecasting - 'good models, well estimated, and well tested' will not necessarily forecast better than 'poor' ones (in the sense of models which are not well fitting, or fail residual diagnostic tests, etc.). The degrees of congruence or noncongruence of a model with economic theory and data transpire to be neither necessary nor sufficient for forecasting success or failure. However, our forecasting theory clarifies why such a result holds, and why it is not antithetical to developing econometric models for other purposes such as testing theories or conducting economic policy. Indeed, different ways of using models may be required for forecasting as against policy analysis. Moreover, the theory suggests methods by which econometric models can be made more robust to non-constancies: some of these are already in use, but have previously lacked rigorous analyses of their properties.

The impact of 'overfitting' (or 'data mining') on forecast failure seems to have been overemphasized: the results just discussed suggest this should not be a primary cause. Unless sample sizes are very small relative to the number of parameters, parameter-selection effects seem unlikely to downwards bias equation standard errors sufficiently to induce apparent forecast failure. Including irrelevant variables – or excluding important variables – that then change markedly both have adverse effects: the former shifts the forecasts when the data do not; the latter leaves unchanged forecasts when the data alter. Concerns about 'overfitting' address only the former, perhaps at the cost of the latter. In any case, other remedies exist to potential 'overfitting', particularly a more structured approach to empirical modelling based on general to specific principles, which checks that the initial model is a satisfactory specification (i.e., congruent), and the final model is suitably parsimonious, without fitting much better. The role of 'data selection' in all aspects of econometric modelling, testing, forecasting and policy is now susceptible to analysis, and again reveals many new, and often surprising, findings. This is the subject of a separate set of notes on econometric modelling.

Generally, forecast-confidence intervals reflect the 'known uncertainties', namely the quantified variability deriving from model estimation and future shocks, in so far as these resemble the average residuals of the model. In economies with unanticipated intermittent deterministic shifts, such confidence intervals will understate the likely range of outcomes. The problem is that we don't know what we don't know, so it is difficult to account for this source of 'unknown uncertainty'. This issue is distinct from when non-congruent models are used as forecasting devices: care is then required to ensure that their measures of forecast uncertainty accurately characterize the known sources. For example, the usual formulae for forecast-error variances can be wildly incorrect if substantial residual autocorrelation is ignored in estimation and when calculating uncertainty.

Finally, the theory has revealed ways of avoiding systematic forecast failure in economies subject to sudden, unanticipated, large shifts. Most economies have witnessed many such shifts in the last quarter century, and there is no sign that large shocks are abating. When shocks are unanticipated, it would take a magician to conjure ways of avoiding large errors if forecasts are announced before the shocks have occurred: we do not claim prescience. Rather, given an inability to forecast the shock, the theory is relevant to the immediate post-shock forecasts, and clarifies how to avoid a sequence of poor forecasts once a shock has occurred.

1.1 Understanding Economic Forecasts

A forecast is any statement about the future. Such statements may be derived from statistical models or informal methods; be well, or badly, based; accurate or inaccurate; precise or imprecise; and concern short or long horizons: thus, forecasting is potentially a vast subject. We address ten questions, namely:

What is a forecast?
What can be forecast?
How is forecasting done generally?
How is forecasting done by economists?
How can one measure the success or failure of forecasts?
How confident can we be in such forecasts?
How do we analyze the properties of forecasting methods?
What are the main problems?
Do these problems have potential solutions?
What is the future of economic forecasting?

Section 3.1 considers the wide range of expressions in the English language for forecasts and forecasters, and draws an important distinction between forecasting and predicting (anything can be forecast – but not everything can be predicted). Then section 3.2 provides some essential background before section 3.3 describes the main methods of forecasting that have been used in economics. Forecasts may be produced by methods varying from well-tested empirical econometric systems through to those which have no observable basis (such as forecasting the 2002 Derby winner in June 1999), and section 3.4 discusses the potential merits of some of these approaches. Howsoever forecasts are produced, one might expect that their accuracy can be gauged. Unfortunately, there is no unique measure for the accuracy of an economic forecast, as section 3.5 demonstrates; and there is no guarantee that better-based methods will win. Section 3.6 notes some factors that might influence our confidence in forecasts.

Section 3.7 then considers how economists analyze their methods, by contrasting an empirical model's forecasts with those from artificial computer-generated data. To illustrate some of the problems in economic forecasting, §3.8 analyzes UK industrial output since 1700. The main problem transpires to be relatively-sudden, intermittent large shifts in the behaviour of the time series, which we call structural breaks. Historically, growth rates have altered dramatically. We consider the plight of a (mythical, longlived) economist who has been given the task of forecasting UK industrial output over each half-century starting in 1750, and witness how often she would have been badly wrong. Some potential solutions to such forecast failures are suggested in section 3.9 by comparing the outcomes from different methods.

1.1.1 Forecast Terminology

English is a rich language, but it seems to reach one of its peaks of verbosity with synonyms for 'forecasting'. This may be because ante, pre and fore offer an abundance of prefixes. We can construct such interesting sentences as: "Those who can, do; those who can't, forecast" spoke the foresightful oracle of Delphi when she divined the future to foretell the prophecy by a soothsayer whose premonitions included the expectation that one day economists would be able to predict the impact of economic policy....

'Forecast' has an interesting etymology: fore is clear, denoting 'in front' or 'in advance'. The interesting bit is cast – dice, lots, spells (as in to bewitch) and horoscopes are all said to be 'cast'. Together with 'casting a fly', these suggest 'chancing one's luck'. As does 'cast about', and perhaps the older usage of 'casting accounts'. Such connections link the notion to gamblers and perhaps even charlatans. In fact, this is true of many of the other synonyms which abound, including: augury; Cassandra (prophesy without credibility); clairvoyant (seeing things not present to the senses); foreboding; foresee; foreshadow; omen (sign of a future event); precognition (know before the occurrence); presage (indication of a yet to happen); prescience (foreknowledge); portend (warn in advance); scry (to practice crystalgazing); and seer (one who sees into the future); at which point I quit on this almost endless list. As most of these synonyms also have an air of doom about them, we may conclude that forecasting has been a disreputable occupation since time immemorial. While anticipate (look forward to, originally depended on 'ante', with capere – to take, and not related to anti, meaning against); extrapolate (extend current trend); prognosis (to predict the course of a disease); and project (to predict on the basis of past results or present trends) have yet to acquire completely adverse connotations, renaming our activities weather soothsaying, or economic scrying would hardly improve their credibility.

Despite dictionaries sometimes treating forecast and predict as synonyms ('forecast: reckon beforehand, or conjecture about the future,' as against 'predict: forecast, foretell, or prophesy'), common usage suggests somewhat different senses: viz. weather forecast (not weather prediction) whereas, 'it was predictable that the marriage would fail' (but not forecastable). Webster suggests predict implies inference from laws of nature, whereas forecast is more probabilistic. This more nearly matches the way I want to use the terms: whether or not an event is predictable is a property of that event, irrespective of our ability to actually predict it; whereas it is always forecastable, since a forecast is simply a statement. Thus, it makes sense to forecast an unpredictable event – indeed, many may say that has always been true of British weather!

There has long been a market for foreknowledge (e.g., insider trading?), and as economics teaches us that the bigger the market the greater the supply, we corroborate that prediction here. Also, the older a word for the concept of 'making a statement about a future event', or for anyone who does so, the less scientific its connotations: witness prophesy, oracle, seer and soothsayer. Perhaps longevity increases the chances that charlatans will have muscled in on the act. As 'forecasting' does not currently have a great reputation, perhaps we should invent another neologism, such as ante-stating, fore-dicting, or pretelling, (pre-casting having been pre-empted by cement makers, and pre-viewing by the media, whereas pre-vision already has a well-established usage).

Literature has occasionally addressed the topic – as Shakespeare expressed it in Macbeth, (I, iii):

> "If you can look into the seeds of time
> And say which grain will grow and which will not,
> Speak then to me"

This may even be possible with modern technology for the seeds of plants, in so far as the constituents of DNA, the seed's in-built sustenance and so on could be determined in some appropriate scanning device. Shelley was less inscrutable in Ode to the West Wind (1, 57):

> "Scatter as from an unextinguished hearth
> Ashes and sparks, my words among mankind!
> Be through my lips to unawakened earth
> The trumpet of a prophecy! O Wind,
> If Winter comes, can Spring be far behind?"

Here we have a very reliable forecast, at least one not so far refuted on thousands of repetitions. Thus,both of these cases may actually prove successful, unlike much of economic forecasting.

The trouble is that the future is uncertain1 – for two reasons. First, as Maxine Singer expressed the matter in her 'Thoughts of a Nonmillenarian' (Bulletin of the American Academy of Arts and Sciences, 1997, **51**, 2, p39):

> 'Because of the things we don't know (that) we don't know, the future is largely unpredictable. But some developments can be anticipated, or at least imagined, on the basis of existing knowledge.'

Notice her wording: not that the future is unforecastable – clearly it is not, because many statements prognosticating on future possibilities appear annually – merely that it is largely unpredictable. The second reason is the apparent randomness of outcomes within the realms we do understand – call this measurable uncertainty. The first is the basic problem: the second may even make us overly confident about our forecasts.

1.1.2 Some Essential Background

Economies evolve over time and are subject to intermittent, and sometimes large, unanticipated shocks. Economic evolution has its source in scientific discoveries and inventions leading to technical progress which becomes embodied in physical and human capital, whereas breaks may be precipitated by changes in legislation, sudden switches in economic policy, or political turmoil (examples of breaks relevant to the UK include the abolition of exchange controls, the introduction of interest-bearing chequing accounts, and privatization). Thus, data in economics are not stationary, in that measured outcomes have different means and variances at different points in time.

Because their means and variances are changing over time, non-stationary data are exceptionally difficult to model. Consequently, the empirical econometric models used to understand and forecast processes as complicated as national economies are far from perfect representations of behaviour. Moreover, the data series used in model building are often inaccurate and prone to revision. Forecasters may only be dimly aware of what changes are afoot, and even when developments can be envisaged, may find it hard to quantify their likely impacts (e.g., the effects of Building Society demutualizations on consumers' spending).

All these difficulties entail that economic forecasting is fraught with problems, and in practice, forecast failure – a significant deterioration in forecast performance relative to the anticipated outcome – is all too common. Understanding this phenomenon requires a theory of economic forecasting for a complicated and changing economy, measured by inaccurate data, using models which are incorrect in unknown ways. A theory based on these realistic assumptions has been developed recently, and its main implications have demonstrable empirical relevance (see Clements and Hendry, 1998b, 1999b). Unfortunately, many of the conclusions which have been established for correctly-specified forecasting models of stationary processes no longer hold. Fortunately, the new theory suggests ways of circumventing systematic forecast failure in economics.

Poor forecasting is distinct from forecast failure: some variables may be inherently uncertain, so while our forecasts of these are poor absolutely, we are not suddenly confronted by large errors. Indeed, in a social science, forecasts may alter actions, so many events may be inherently unpredictable (viz., changes in equity prices, or perhaps exchange-rate crises): we cannot expect other than poor forecasts of unpredictable events, but we may hope to avoid systematic forecast failure.

Econometric forecasting models are systems of relationships between variables such as GNP, inflation, exchange rates etc. Their equations are then estimated from available data, mainly aggregate time-series. Such models have three main components: deterministic terms introduced to capture averages and steady growth (represented here by intercepts and linear trends, which take the values 1, 1, 1,...; and 1, 2, 3, ... respectively), and whose future values are known; observed stochastic variables with unknown future values (like consumers' expenditure, prices, etc.); and unobserved errors, all of whose values (past, present and future) are unknown (though perhaps estimable in the context of a model). The relationships between any of these three components could be inappropriately formulated, inaccurately estimated, or change in unanticipated ways. Each of the resulting 9 types of mistake could induce poor forecast performance, either from inaccurate (i.e., biased), or imprecise (i.e., high variance) forecasts. Instead, theory suggests that some mistakes have pernicious effects on forecasts, whereas others are relatively less important in most settings. Surprisingly, the key to understanding systematic forecast failure depends on the behaviour of the deterministic terms, even though their future values are known, rather than on the behaviour of variables with unknown future values.

Five aspects of the deterministic terms matter in practice. First, their specification and estimation: inadequate representations or inaccurate estimates of intercepts and trends can induce bad forecasts – knowing the future values of the trend is of little help when it is multiplied by the wrong parameter value (for example, omitting a trend in a model when there is one in the data leads to ever-increasing forecast errors). Secondly, the consequences of unanticipated changes in their values are pernicious: the economy moves, but the forecasting model does not, inducing large forecast errors. Thus, although the future values of the existing deterministic variables are known, there may be different intercepts and trends in the future, and those values are not currently known – see the Singer quote above. Thirdly, deterministic shifts may reflect changes elsewhere in the economy interacting with an incomplete model specification. Next, formulating models to minimize the impact of possible changes in deterministic terms is generally beneficial, even when the cost is a poorer representation by the model of both the economic theory and the data. Finally, successful modelling of changes in deterministic terms pays handsome dividends, even if only by using simple corrections or updates.

Figure 1 illustrates four cases. In the top-left panel, the wrong slope of the trend has been estimated; in the top-right, the intercept has shifted, so the sample mean is wrong in both regimes; in the lowerleft, the data trend has changed but the model has not; and the lower-right panel illustrates that the first-differences of the trends in panel c essentially differ only at the jump point.

Other possible sources of forecast errors – such as mis-specifying the stochastic components or uncertainty due to estimating their parameters – appear less important. Thus, the theory directs attention to areas that may induce forecast failure, and casts serious doubt on competing explanations such as inadequate use of economic theory: it offers no support for claims that imposing restrictions from economic theory will improve forecast accuracy (see e.g., Diebold, 1998). An indirect consequence is that there may be little gain in forecast accuracy by inventing 'better' estimation methods, especially if the opportunity cost is less effort devoted to developing more-robust forecasting models. Also, the new theory suggests that the impact on forecast failure of empirically selecting a model should be small, as should retaining unnecessary estimated parameters (unless sample sizes are very small).

Forecast-confidence intervals seek to measure forecast uncertainty, but only reflect the 'known uncertainties', deriving from model estimation and future shocks which resemble the past, whereas in economics, unanticipated deterministic shifts occur intermittently. Since we don't know what we don't know, it is difficult to account for this 'unknown uncertainty'. Nevertheless, the theory has revealed ways of avoiding systematic forecast failure in economies that are subject to sudden, unanticipated, large shifts. The UK economy has witnessed many such shifts in the last century, and there is no sign that large shocks are abating. When shocks are unanticipated, it would take a magician to conjure ways of avoiding large errors in forecasts announced before such shocks have occurred.

Rather, given an inability to forecast the shock, adaptation is required once a shock has occurred, to avoid a sequence of poor forecasts.

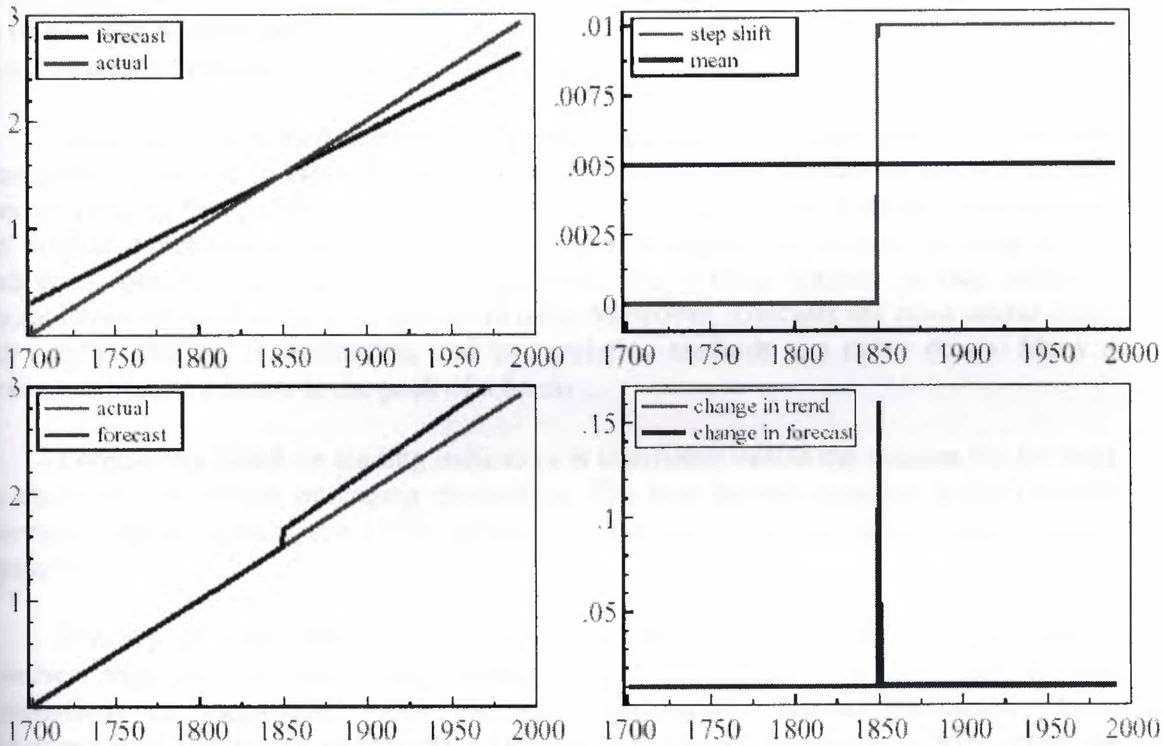


Figure 1 The impacts of incorrect or changing deterministic terms.

1.1.3 Methods of Forecasting

There are many ways of making economic forecasts besides using econometric models. Their success requires that:

(a) there are regularities to be captured;
(b) the regularities are informative about the future;
(c) the proposed method captures those regularities; yet
(d) it excludes non-regularities.

The first two are characteristics of the economic system; the last two of the forecasting method. The history of economic forecasting in the UK suggests that there are some regularities informative about future events, but also major irregularities as well (see e.g., Burns, 1986, Wallis, 1989, Pain and Britton, 1992, and Cook, 1995). However, achieving (c) without suffering from (d) is difficult.

Methods of forecasting include:
(1) guessing, 'rules of thumb' or 'informal models';
(2) extrapolation;
(3) leading indicators;
(4) surveys;
(5) time-series models; and
(6) econometric systems.

Guessing and related methods only rely on luck. While that may be a minimal assumption compared to other methods, guessing is not a generally useful method, even if at every point in time, some 'oracle' manages to forecast accurately. Unfortunately, no-one can predict which oracle will be successful next. Extrapolation is fine so long as the tendencies persist, but that is itself doubtful: the telling feature is that different extrapolators are used at different points in time. Moreover, forecasts are most useful when they predict changes in tendencies, and extrapolative methods can never do so. Many a person has bought a house at the peak of a boom....

Forecasting based on leading indicators is unreliable unless the reasons for the lead are clear, as with orders preceding production. The best known example is the Harvard Barometer, which missed the 1929 collapse. In practice, indicators need to be changed regularly.

Surveys of consumers and businesses can be informative about future events. However, they rely on plans being realized, and if not, usually can offer only ad hoc explanations for departures from outcomes. Historically, time-series models have performed well relative to econometric systems. The theory discussed in x3.2 offers an explanation for that outcome in terms of their relative robustness to deterministic shifts, as illustrated in figure 1, and we will use several simple time-series models below.

Econometric forecasting models were described in x3.2 above. The advantages to economists of formal econometric systems of national economies are to consolidate existing empirical and theoretical knowledge of how economies function, provide a framework for a progressive research strategy, and help explain their own failures, as well as provide forecasts and policy advice. Econometric and time-series models are the primary methods of forecasting in economics.

1.1.4 On Winning at Forecasting

What determines the winners and losers in a forecasting competition? Many factors undoubtedly play a role, but one aspect can be illustrated by two friends passing time while waiting at a bus-stop. Sue challenges Peter to forecast the behaviour of a student who is standing inside the bus shelter: every 30 seconds they will both write in their diary a forecast for the next 30 seconds as to whether or not the student will have left. Sue has been to my lectures, so always writes what the current state is: when the student is there, she forecasts he will still be there in 30 seconds; and when he has left, she writes that. Thus, in the 5 minutes before the student goes, she is correct 10 times, then wrong once, but thereafter correct for ever. Peter, however, is an economist, so he uses a causal model: students stand at bus stops to get on buses. Thus, if no bus approaches, Peter forecasts the student will stay; but when a bus appears, he forecasts the student will board the bus. Unfortunately, 4 different buses come by, and the student remains stubbornly at the bus stop – then his girl friend appears on her motor bike, the student climbs on and goes away. Peter is wrong 4 times in the five minutes, and if he stuck to his causal model, wrong ever after since the student never got on a bus.

To win a forecasting competition where unanticipated outcomes are feasible, simply forecast the present (perhaps transformed to a stationary distribution). Causal models can go badly wrong in any given instance, and need rapid repair when they do so. However, substitute the phrase 'the volcano will not explode' for 'will remain at the bus stop', and the vacuous nature of Sue's forecast is clear, even if she did win. Thus, economists are right to stick to causal modelling as a basis for forecasting, perhaps mediated by adjustments to offset the unanticipated should it eventuate. We should be pleased with forecast failures – for we learn greatly from them – not ashamed that we lack a full understanding of how economies behave. Thus, I re-iterate an old complaint: when weather forecasters go awry, they get a new super-computer; when economists mis-forecast, we get our budgets cut.

1.1.5 Measuring the Winner

The accuracy and precision of forecasts represent different dimensions: the latter almost always denotes 'with little uncertainty', so that one can say the moon is exactly 5000 miles away and be very precise, but very inaccurate. Conversely, it is accurate to say that the moon lies between 1000 and 1 million miles away, but very imprecise.

To measure accuracy and precision, we usually adopt the notions of 'unbiasedness', so the forecasts are centered on the outcomes, and small variance, so only a narrow range of outcomes is compatible with the forecast statement. Combining bias and variance leads to the mean square forecast error (MSFE) measure that is commonly reported.

Unfortunately, for either multi-period or multi-variable forecasts (which are the norm in economics), no unique measure of a 'winner' is possible in a forecasting competition, even when the metric is agreed. Figure 2 illustrates the problem. The forecast in the top left panel (denoted a) is awful for the levels of the series shown, but is very accurate for the growth rate (top right panel); conversely, forecast b (lower left panel) is fine for the level, but dreadful for the growth (lower right panel). Thus, one must decide on which aspect it is important to be close before a choice is possible. Worse still, MSFE itself is not an obvious criterion: a stockbroker probably does not care how good or bad a model is on MSFE if it is the best for making money!

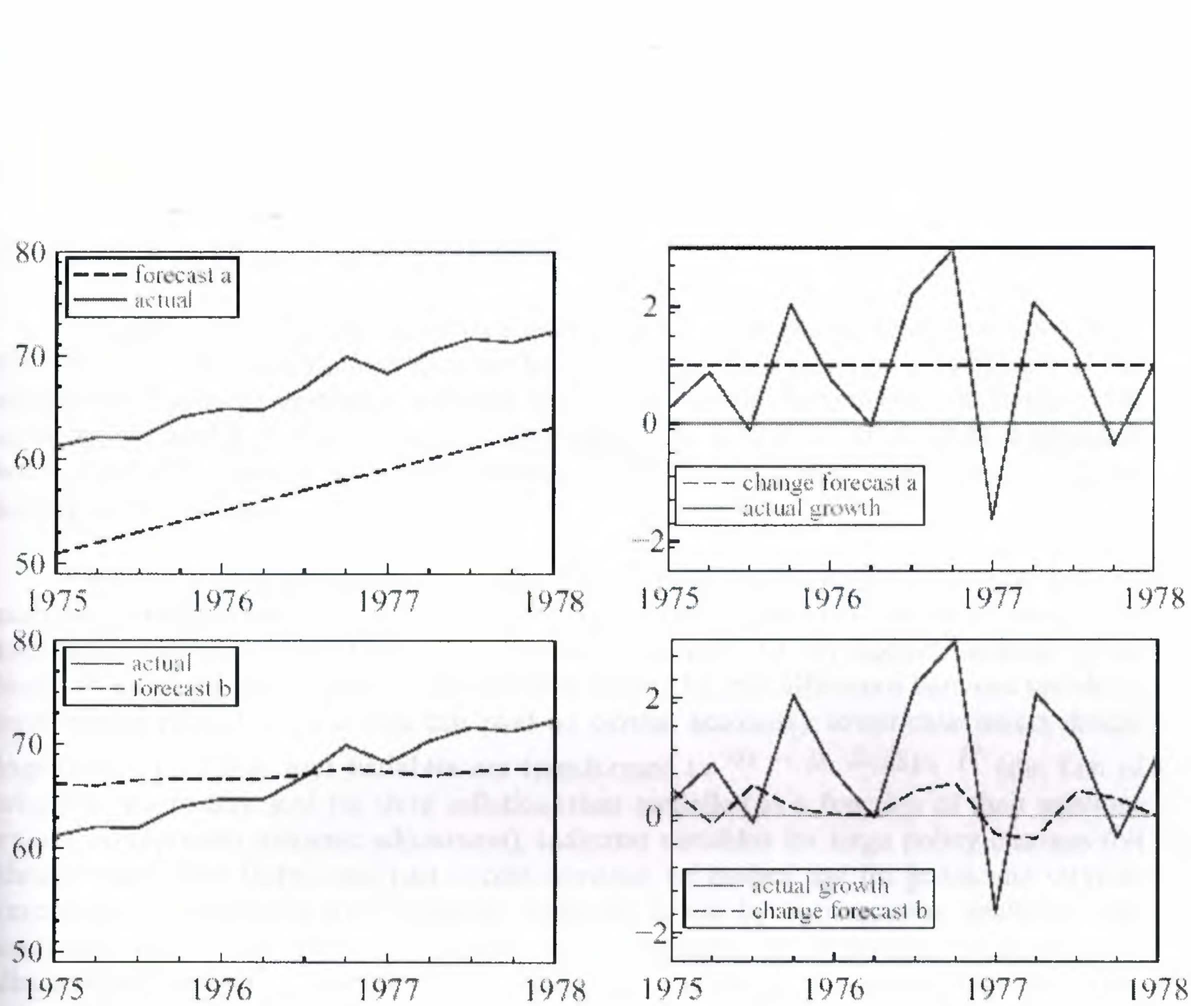


Figure 2 Which wins: forecasts of levels or growth?.

1.1.6 Forecast Confidence Intervals

Forecasts are sometimes presented with estimates of the uncertainty attaching to them, usually in the form of forecast confidence intervals which are expected to cover the likely outcome some percentage of the time (such as 67% or 95%). Naturally, such intervals tend to be wider the longer the forecast horizon. The Bank of England 'rivers of blood and bile' charts show ranges of intervals in ever-lighter shades of red for inflation (green for GNP) as the likelihood falls of the outcome lying outside each bound (see Hatch, 1999). Such estimates are greatly to be welcomed, especially compared to not presenting any measure of uncertainty, merely a forecast number (like 2% inflation) reported as if were exact (surprisingly, that used to be the norm). Figure 3 shows the variation in four economic time series, and figure 4 some forecast confidence intervals.

Since the future is uncertain, outcomes can at best lie within some interval around a forecast. Even when forecast confidence intervals are correctly calculated, outcomes should lie outside that range the converse percentage of the time (e.g., 33% for a 67% interval). But as stressed above, any reported interval is based on 'known uncertainties', and cannot reflect 'what we don't know we don't know': so on average, forecasters will do worse than they anticipate from the conventional calculations (see Ericsson, 1999). By itself, this should not entail a lack of confidence in forecasts, but does serve to emphasize the considerable uncertainty that attaches to possible futures, and the corresponding tentative nature of any claims to foretell what will materialize.

1.1.7 How to Analyze Forecasting Methods

Econometric methods are derived under various assumptions about how economies function, and these assumptions may not be appropriate. To check on the adequacy of our models and methods, simulation methods have proved useful. Implement a facsimile of the econometric model on the computer, and compare the properties of the data it produces with actual outcomes: a serious mis-match would reveal hidden inadequacies. Lets us undertake an example.

First, one must produce an empirical model of the time series to be forecast: here we consider a small monetary system comprising UK narrow money (M1 measure,denoted m), total final expenditure in 1985 prices (demand measure, x), its implicit deflator (price level, p), and the opportunity cost of holding money (R, the difference between the short-term market rate of interest, and that paid on current accounts): lower-case letters denote logs (base e).3 These four variables are transformed to $m-p, x, \Delta p, R$ (the first of which is real money, and the third inflation),then modelled as a function of their previous values (to represent dynamic adjustment), indicator variables for large policy changes (oil shocks and budget shifts), and past excess demands for money and for goods and services (modelled by deviations from long-run relations, found by cointegration analysis). The estimated parameters show the speeds of adjustments in removing excess-demand disequilibria, as well as responses to past changes and major shocks, whereas the properties of the unexplained components (residuals) represent the assumed innovation shocks. The histograms and densities of the four sets of (standardized) residuals from the estimated equations over 1964–1989 are shown in figure 3, together with normal distributions, which provide a reasonable approximation. Also, the top row of figure 4 records the last few periods of fit and the forecasts from that model over the next 12 quarters, together with confidence intervals around the forecasts (we comment on the second row below). The forecasts show steady growth in real money and expenditure, with relatively-constant, low levels of inflation and interest rates. When the model is a good specification, the confidence bands should include the outcomes 95% of the time: since the uncertainty is increasing, the bands are wider the further ahead the forecast. For trending variables like output, they will therefore continue to widen indefinitely, but for stationary series, they will reach an asymptote.

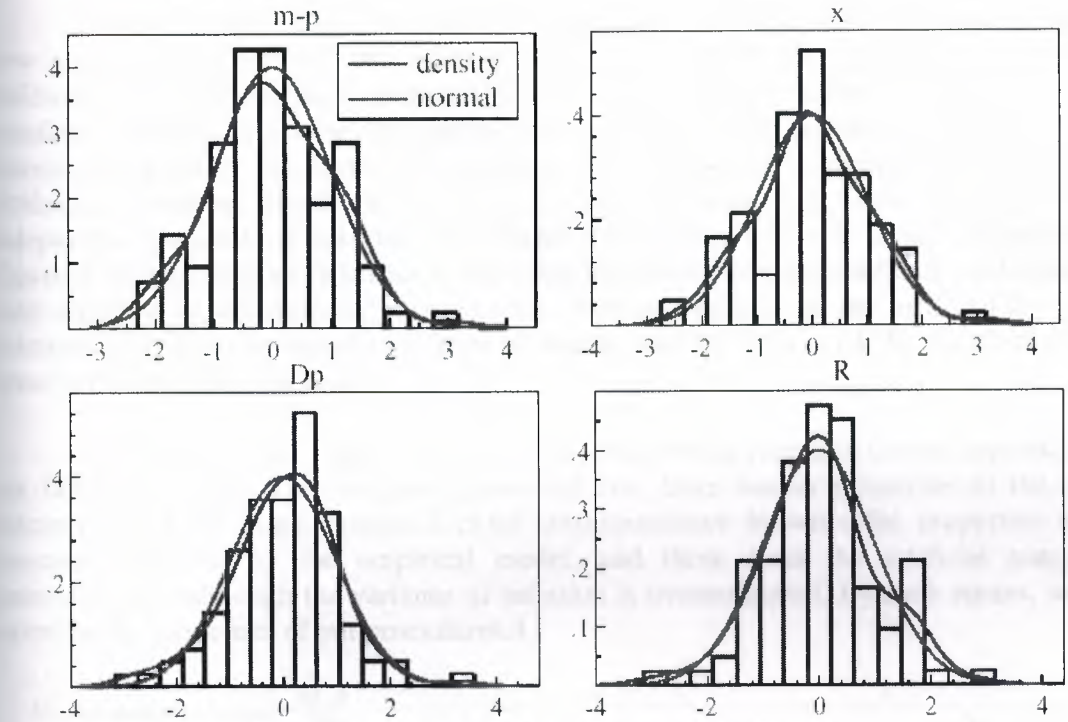


Figure 3 Densities of residuals from the UK M1 model.

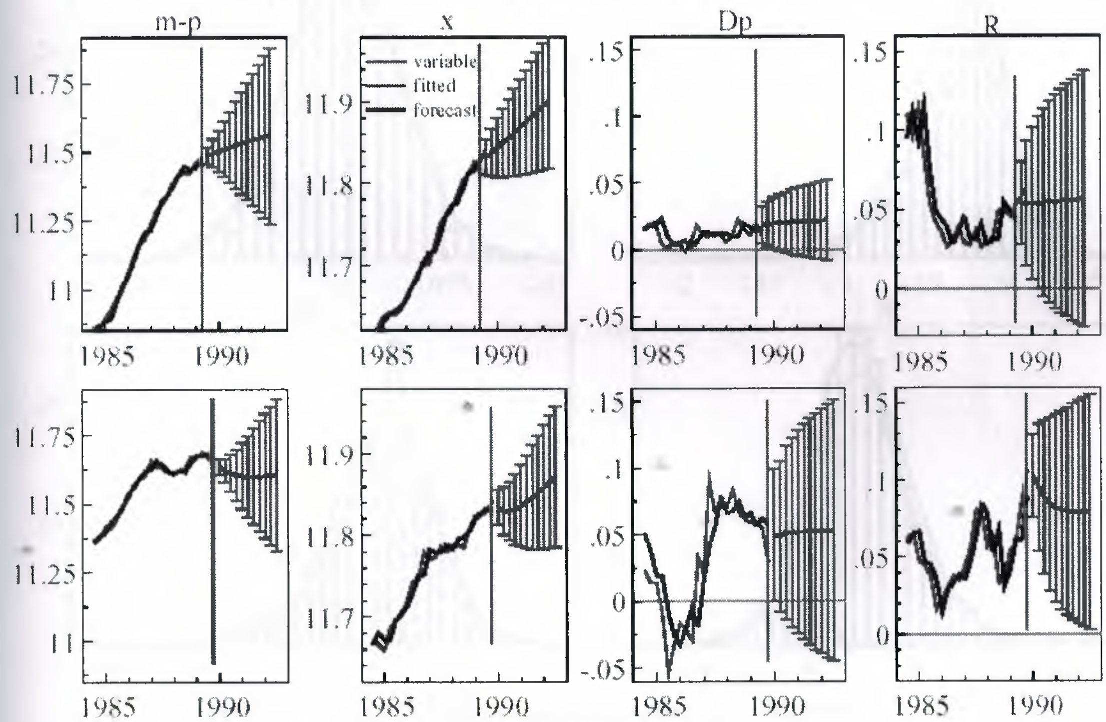


Figure 4 Forecasts of UKM1 and the artificial series.

Given the initial conditions of this system, and the values of all its parameters, we now create a replica of the estimated model on a computer. By replacing the empirical residuals by pseudo-random numbers with the distributions shown in fig. 3, we can simulate artificial data from the model, and re-compute the parameter estimates, tests, forecasts and policy reactions. This exercise can be repeated hundreds of times, thereby producing sampling distributions of the relevant statistics (e.g., how often a test for independent residuals rejects that hypothesis when the errors are in fact independent). Figure 5 records the four densities of the estimated disequilibrium-feedback coefficients in each equation of the system, generated by 1000 replications of the artificial data. The outcome reveals some departures from normality, but the means of the distributions are close to the empirical estimates.

The lower row of figure 4 shows the corresponding forecasts on one replication of our facsimile model. The computer generated data have similar properties to the actual outcomes, and the graph reveals a close correspondence between the properties of the forecasts produced by the empirical model, and those from the artificial computer-generated data, although the variance of inflation is overestimated. By such means, we can ascertain the properties of our procedures.4

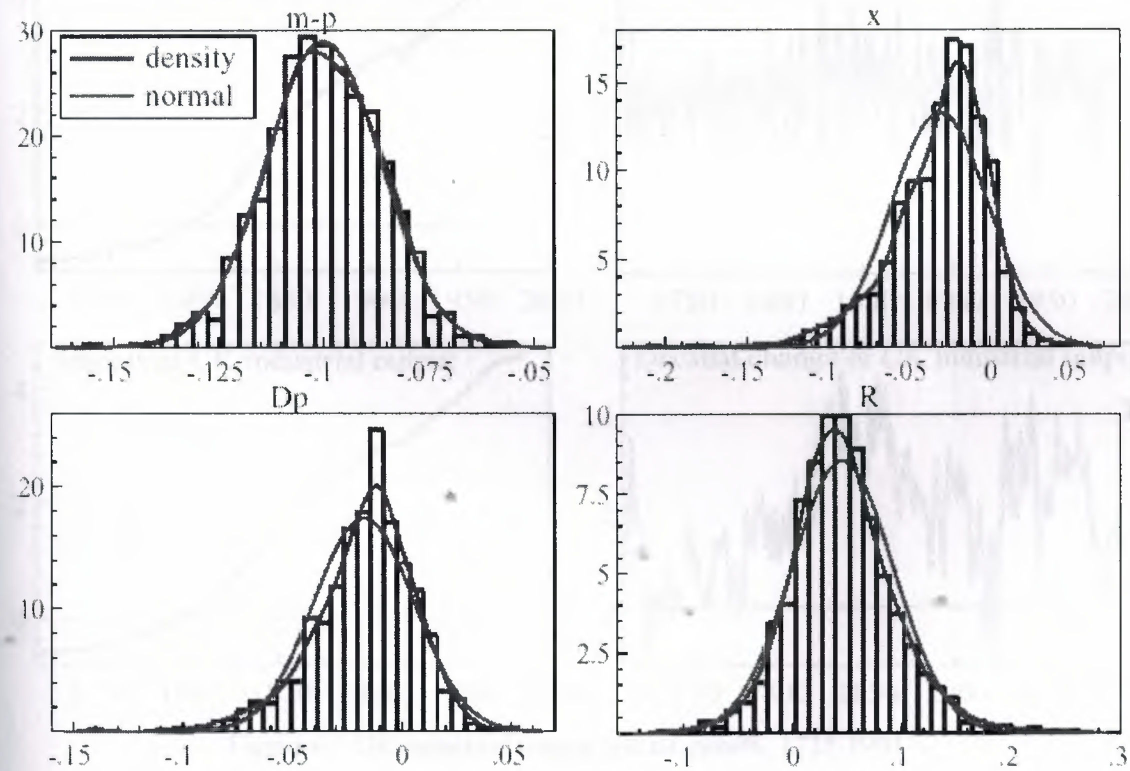


Figure 5 Densities of four estimated coefficients from the artificial data.

1.1.8 Forecasting 300 Years of UK Industrial Output

Figure 6a records the level of UK industrial output on a log scale (denoted y) over 1715–1991.The time series of the log-level is manifestly non-stationary, as the mean has changed greatly over time. Clearly, industrial output has grown dramatically (rather slower if per capita figures are used), but rather unevenly, as highlighted by the changes (first differences) in figure 6b. To smooth the visual appearance, figure 6c reports a decade-long centered moving average, and figure 6d the decadal changes in that moving average, which emphasize the epochal nature of growth. The growth in the series declines at the beginning of the sample, then exhibits a cyclical pattern till around 1775 when there is almost a 'take-off' into sustained growth, with a further substantial rise around 1825 which persists till about 1875 with the onset of the 'great depression' (during which the price level fell for almost 20 years). The crash of 1919–21 is clearly visible, but the 1929–35 depression is not obvious (the UK was much less affected than the USA). Finally, the post-war boom is marked, as is the downturn engineered in 1979–82.

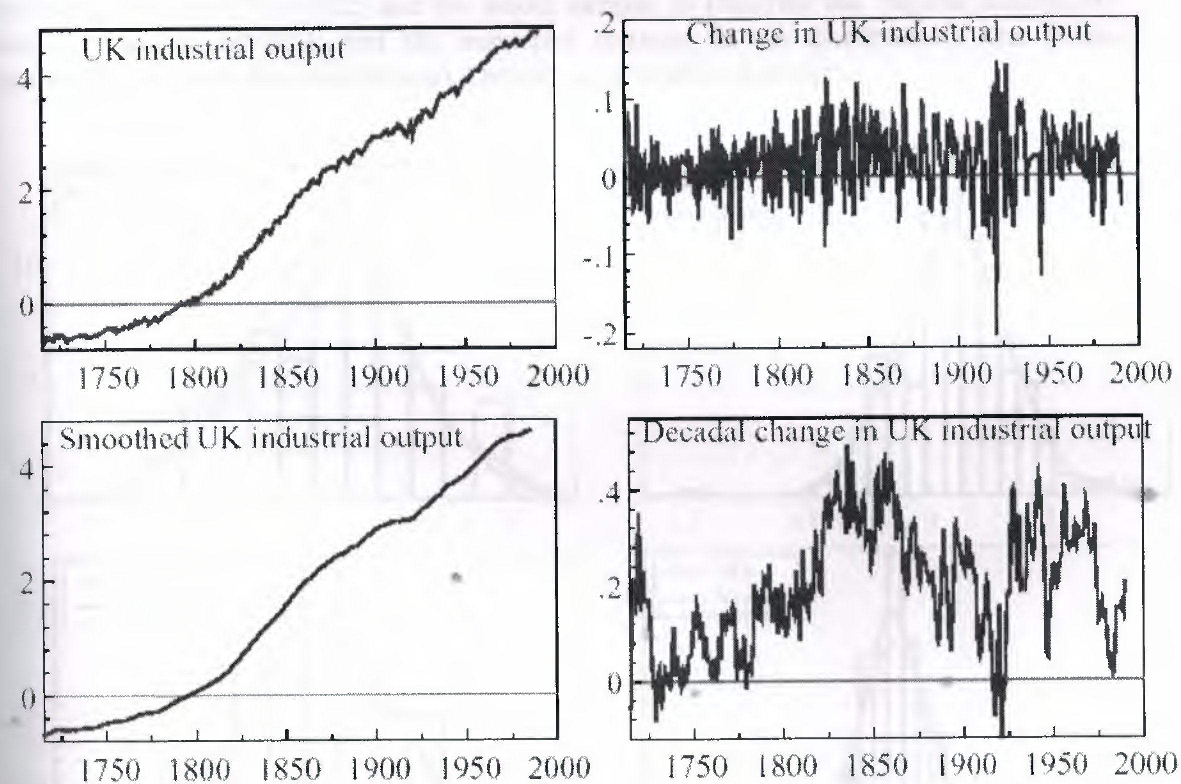


Figure 6 UK industrial output and its growth, 1715-1991.

Table 1 records the means and standard deviations (SD) of $\Delta y_t = y_t - y_{t-1}$ and $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$ over each 50-year sub-period to illustrate the large changes that have occurred in these descriptive statistics.

Table 1 Means and standard deviations (%) of Δy and $\Delta^2 y$.

	1715 -1750	1751 -1800	1801 -1850	1851 -1900	1901 -1950	1951 -1991	**1715 -1991**
				Δy			
Mean	0.86	1.07	2.86	2.77	1.95	1.96	**1.96**
SD	3.58	3.47	5.03	4.09	6.32	3.40	**4.54**
				$\Delta^2 y$			
Mean	0.20	-0.09	0.02	0.00	0.11	-0.23	**0.00**
SD	5.32	5.57	8.01	5.02	9.35	4.29	**6.56**

Across different periods, mean growth rates tripled, and standard deviations almost doubled. Figure 7 shows histograms and densities of the growth rate and its change (acceleration) over 1715–1825 and the whole sample, to illustrate the 'regular uncertainty' noted at the end of x3.1, and the important changes in the distributions (the normal distribution is shown for comparison). Overall, acceleration was zero.

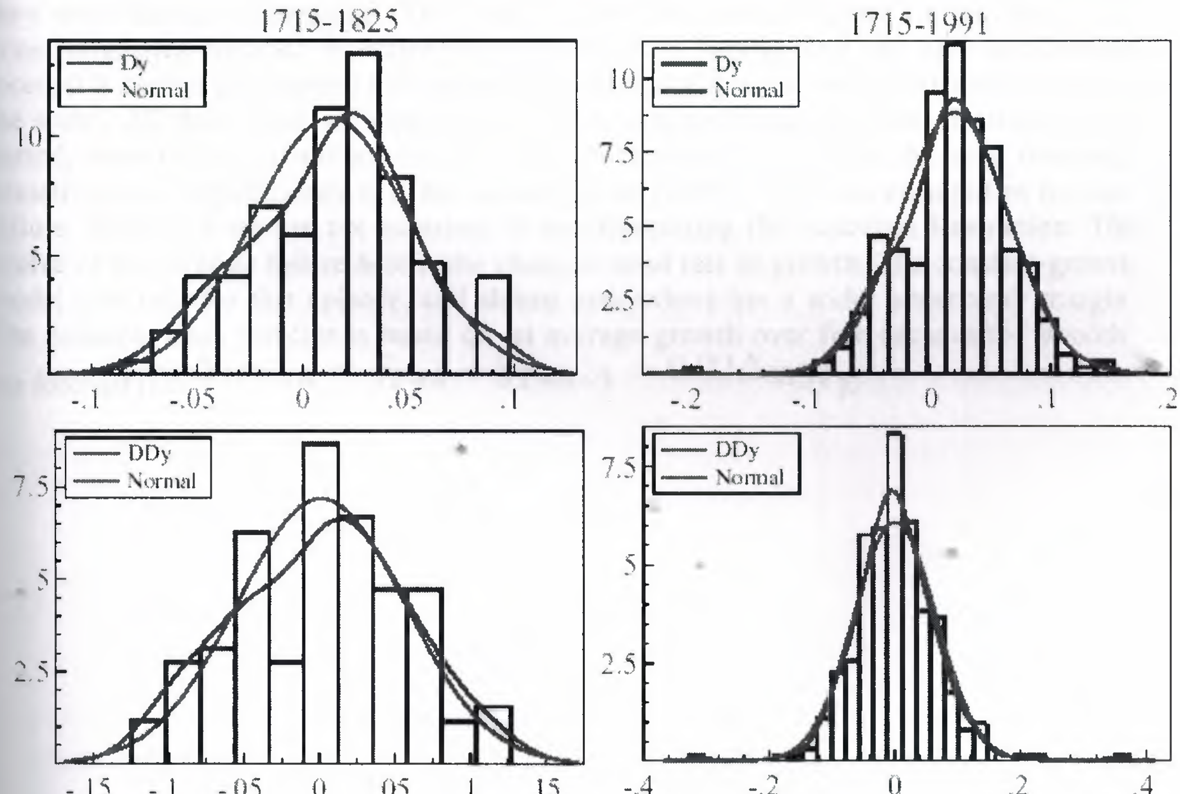


Figure 7 Distributions of growth and acceleration in UK industrial output, 1715-1991.

By itself, the non-stationary level is not necessarily problematic, since we could remove both deterministic and stochastic trends by differencing, and so analyze growth rates. However, there were also great changes in the growth rate, and those would have been harder to forecast: few contemporaneous writers foresaw the consequences of the burgeoning Industrial Revolution till it was well under way, and many of the most vocal focused on its drawbacks in creating 'dark satanic mills', rather than starting a prolonged upswing in general living standards. Nevertheless, we will pretend to 'forecast' industrial output up to 50-years ahead, using models based on the preceding 50-year period: thus, we have forecasts for 1751–1800; 1801–1850; 1851–1900; and 1901–1950; finishing with 1951–1991. Three simple models are used: the first is a linear trend; the second a 'constant change' forecast, and the third is our analogue of 'still standing at the bus-stop', which here corresponds to 'no acceleration'. If the world were non-stochastic, these models would all be identical – but they behave differently in stochastic worlds, due to the incompatible nature of their unexplained components. If the underlying growth rate were constant, all three should deliver unbiased forecasts, differing mainly in precision, but again could differ markedly when growth rates change. However, the third is not reliable beyond the very short-term, so may well perform badly on the long horizons considered here.6

Figure 8 records the three sets of forecasts for successive 50-year horizons, together with forecast confidence intervals which should include the outcomes 95% of the time if they were correctly computed. The trend forecast is a solid line with error bands; the constant-growth forecast is dotted (again with error bands), and the zero-acceleration forecast is shown as a dashed line (without a confidence interval, which otherwise swamps the scale). All three forecasts appear very similar over such long horizons. In almost every period, some of the realizations lie outside the confidence intervals for the trend forecasts, sometimes very significantly as in the second period (1801–1850): this exemplifies forecast failure, albeit that we are not surprised at mis-forecasting the Industrial Revolution. The source of the forecast failure here is the changed trend rate of growth. The constant-growth model also fails for that episode, and almost everywhere has a wider uncertainty margin. The noacceleration forecast is based on an average growth over five decades to 'smooth' the forecast (i.e $\Delta_1 \widehat{y}_{T+h} = \widehat{y}_{T+h} - \widehat{y}_{T+h-1} = 0.02\Delta_{50} y_T$).

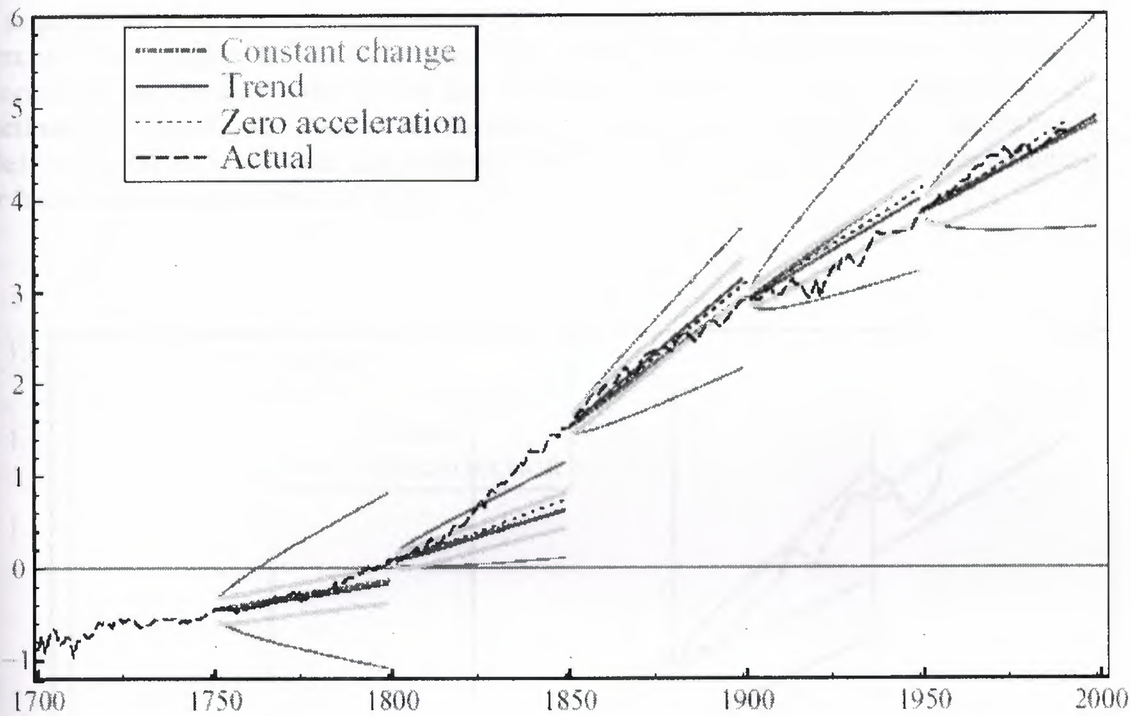


Figure 8 Forecasts of UK industrial output for successive 50-year horizons.

To highlight their distinct underlying behaviour, figure 9 shows the zero-acceleration and constanttrend forecasts over successive 10-year horizons for 1801–1850. The former is dashed, and based on $0.1\Delta_{10}y_T$, and the latter solid, and based on the preceeding 50 observations (we comment on the dotted line below). The outcome illustrates the former's much better performance on shorter horizons: the contrast is stark after 1830. The theory of forecasting correctly predicts which of these forecasting methods will win, assuming the Industrial Revolution induced shifts in the models' deterministic components; and the adaptive method avoids the systematic failure of the constant-trend model.

1.1.9 Some Potential Solutions

The above example is deliberately contrived to illustrate several of the potential solutions to forecasting when a data process is non-stationary. First, differencing removes deterministic terms (so second differencing removes a linear trend), and reduces step shifts to 'blips'. Nothing can prevent a failure if there is an unanticipated break, but once the break is past, some forecasting methods (in differences) are much more robust than others (in levels). Secondly, updating estimates helps adapt to changing data: linear-trend forecasts based on the previous 10 data points only are much more accurate here, despite the resulting decrease in the precision of the estimated coefficients. Thirdly, when the first of a sequence of forecasts is in error, often the remainder suffer similarly. Consequently, an 'intercept shift' equal to the last observed error can considerably improve forecast performance, as shown by the dotted line in figure 9: every forecast is much improved, sometimes sufficiently to 'win'. To succeed in forecasting competitions, econometric models will have to mimic the adaptability of the best forecasting devices, while retaining their foundations in economic analysis.

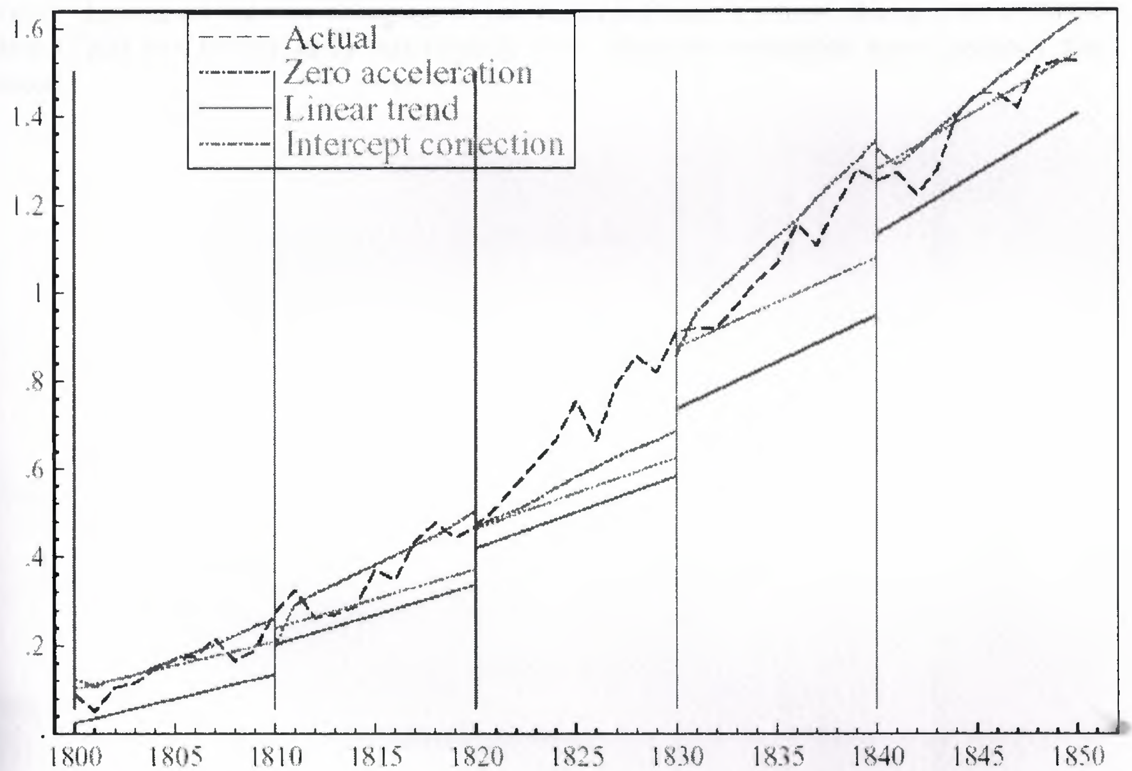


Figure 9 Zero-acceleration and constant-trend forecasts over 10-year horizons, 1801–1850.

1.2 Technical Note

To summarize the effect of shifts, let y_t denote a vector of non-integrated (I(0)) time series with prebreak unconditional expectations and variances denoted by $E[y_t]$ and $V[y_t]$ respectively. Let the corresponding entities based on treating the model as the data generation process be denoted Em [y_t] and Vm [y_t]: these are the means and variances of the outputs from the model. Then forecast failure, and conversely the detectability of breaks, depends strongly on the difference $\mathsf{E}[\mathbf{y}_t] - \mathsf{E}_\mathsf{m}[\mathbf{y}_t]$ so long as $\mathsf{V}_\mathsf{m}[\mathbf{y}_t]$ does not alter markedly. Consequently, parameter changes in the DGP that leave $\mathsf{E}[\mathbf{y}_t] \simeq \mathsf{E}_\mathsf{m}[\mathbf{y}_t]$ suffer from a detectability problem unless they generate very large variance increases. Since I(1) vector autoregressions (VARs) can be reparameterized by differencing and cointegration transformations as I(0) vector equilibrium-correction models (VEqCMs) where all variables are expressed as deviations around their (perhaps pre-break) means, the same logic applies: only shifts in those means induce departures that are readily detectable. This strongly guides the formulation for 'detectors' of, and 'solutions' to, systematic forecast failure in a world of structural breaks. Moreover, 'impulse-response analyses' depend on the non-changing of the very coefficients whose changes are difficult to detect, and can be seriously mis-leading even when no detectable non-constancy has occurred.

2 A FRAMEWORK FOR ECONOMIC FORECASTING

For an econometric theory of forecasting to deliver relevant conclusions about empirical forecasting, it must be based on assumptions that adequately capture the appropriate aspects of the real world to be forecast. We distinguish six facets: [A] the nature of the DGP; [B] the knowledge level about that DGP; [C] the dimensionality of the system under investigation; [D] the form of the analysis; [E] the forecast horizon; and [F] the linearity or otherwise of the system. Then we have:

[A] Nature of the DGP

[i] stationary;
[ii] cointegrated stationary;
[iii] evolutionary, non-stationary.

[B] Knowledge level

[i] known DGP, known θ;
[ii] known DGP, unknown θ;
[iii] unknown DGP, unknown θ.

[C] Dimensionality of the system

[i] scalar process;
[ii] closed vector process;
[iii] open vector process.

[D] Form of analysis

[i] asymptotic analysis;
[ii] finite sample results, perhaps simulation based.

[E] Forecast horizon

[i] 1-step;
[ii] multi-step.

[F] Linearity of the system

[i] linear;
[ii] non-linear.

An exhaustive analysis under this taxonomy would generate 216 cases! Many of these are not directly relevant: we focus on [A](iii)+[B](iii)+[C](ii)+[D](i)+[E](ii)+[F](i), using estimated econometric systems.

2.1 Alternative Methods of Forecasting

There are many ways of making economic forecasts besides using econometric models. Their success requires that (a) there are regularities to be captured; (b) the regularities are informative about the future (c) the method captures those regularities; and (d) excludes non-regularities. The first two are characteristics of the economic system; the last two of the forecasting method.

The history of economic forecasting in the UK suggests that there are some regularities informative about future events, but also major irregularities as well (see e.g., Burns, 1986, Wallis, 1989, Pain and Britton, 1992, and Cook, 1995). The dynamic integrated systems with intermittent structural breaks that are formalized below seem consistent with such evidence. However, achieving (c) without suffering from (d) is difficult, and motivates the conceptual structure proposed below, as well as the emphasis on issues such as parsimony and collinearity, and the re-examination of the role of causal information when forecasting models are mis-specified.

Methods of forecasting include guessing; 'rules of thumb' or 'informal models'; naive extrapolation; leading indicators; surveys; time-series models; and econometric systems. Scalar versions of time-series models include Kalman (1960) or Box and Jenkins (1976). Autoregressive integrated moving average models (ARIMAs) are a dominant class of time-series models as the Wold decomposition theorem (Wold, 1938) states that any purely indeterministic stationary time series can be expressed as an infinite moving average (MA); see Cox and Miller (1965) , p.286–8, for a lucid discussion. The multivariate successor to Box–Jenkins is the vector autoregressive representation, see Doan, Litterman and Sims (1984). In the USA this approach has claimed some successes.

Formal econometric systems of national economies consolidate existing empirical and theoretical knowledge of how economies function, provide a framework for a progressive research strategy, and help explain their own failures as well as provide forecasts. Economic forecasting based on econometric and multivariate time-series models will be our primary methods.

2.2 Sources Of Forecast Failure

The possible sources of mistakes that can induce multi-step forecast errors from econometric models of possibly cointegrated I(1) processes can be delineated in a formal taxonomy. This highlights which sources induce forecast-error biases, and which have variance effects. The framework comprises:

(1) a forecasting model formulated in accordance with some theoretical notions,
(2) selected by some empirical criteria,
(3) but mis-specified (to an unknown extent) for the DGP,
(4) with parameters estimated (possibly inconsistently),
(5) from (probably inaccurate) observations,
(6) which are generated by an integrated-cointegrated process,
(7) subject to intermittent structural breaks.

Such assumptions more closely mimic the empirical setting than those often underlying investigations of economic forecasting, and we explored this framework in detail in Clements and Hendry (1998b, 1999b). The resulting forecast-error taxonomy includes a source for the effects of each of 2.–7., partitioned (where appropriate) for deterministic and stochastic influences: see §8.

Our analysis utilizes the concepts of a DGP and a model thereof, and attributes the major problems of forecasting to structural breaks in the model relative to the DGP. Between the actual DGP and the empirical forecasting model, there lies a 'local DGP of the variables being modelled', denoted the LDGP: see Bontemps and Mizon (1996). Using a VEqCM as the DGP in a two-tier system (say) entails that the VEqCM is the LDGP in the three-tier stratification. Changes in growth rates or equilibrium means in the VEqCM could be viewed as resulting from a failure to model the forces operative at the level of the DGP. The correspondence between the LDGP and DGP is assumed to be close enough to sustain an analysis of forecasting, checked by what happens in practice (via empirical illustrations, where the outcomes depend on the actual, but unknown, mapping between the forecasting model and the economy). We first formalize the relevant concepts before describing the DGP and the forecasting models, then record the taxonomy of forecast errors, focusing on the biases and variances of the various practical models.

2.3 Concepts

2.3.1 Unpredictability

ν_t is an unpredictable process with respect to It-1 if:

$$\mathsf{D}_{\nu_t}(\nu_t \mid I_{t-1}) = \mathsf{D}_{\nu_t}(\nu_t) \qquad (1)$$

so the conditional and unconditional distributions coincide. Unpredictability is invariant under nonsingular contemporaneous transforms: if ν_t is unpredictable, so is $\mathbf{B}\nu_t$ where $|\mathbf{B}| \neq 0$.The definition is equivalent to the statistical independence of ν_t from I_{t-1}: it does not connote 'wild', and knowing $\mathsf{D}_{\nu_t}(\nu_t)$ may be highly informative relative to not knowing it.

However, unpredictability is not invariant under intertemporal transforms since if **ut** = ν_t + **Af(It-1)**:

$$\mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t \mid I_{t-1}) \neq \mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t)$$

when $\mathbf{A} \neq \mathbf{0}$. Unpredictability is relative to the information set used; e.g., it can happen that for $J_{t-1} \subset I_{t-1}$:

$$\mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t \mid J_{t-1}) = \mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t) \text{ yet } \mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t \mid I_{t-1}) \neq \mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t).$$

However, $J_{t-1} \subset I_{t-1}$ does not preclude predictability. Unpredictability may also be relative to the time period, in that we could have:

$$\mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t \mid I_{t-1}) = \mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t) \text{ for } t = 1, \ldots, T \quad (2)$$

yet:

$$\mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t \mid I_{t-1}) \neq \mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t) \text{ for } t = T+1, \ldots, T+H, \quad (3)$$

or vice versa. Finally, unpredictability may be relative to the horizon considered in that:

$$\mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t \mid I_{t-2}) = \mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t) \text{ yet } \mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t \mid I_{t-1}) \neq \mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t).$$

But the converse, that :

$$\mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t \mid I_{t-1}) = \mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t) \text{ yet } \mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t \mid I_{t-2}) \neq \mathsf{D}_{\mathsf{u}_t}(\mathbf{u}_t)$$

is not possible as $I_{t-2} \subseteq I_{t-1}$ by definition.

Sequential factorization of the joint density of $\mathbf{X}_T^1$ yields the prediction representation:

$$\mathsf{D}_{\mathsf{X}}(\mathbf{X}_T^1 \mid I_0, \cdot) = \prod_{t=1}^{T} \mathsf{D}_{\mathsf{x}_t}(\mathbf{x}_t \mid I_{t-1}, \cdot). \quad (4)$$

Predictability therefore requires combinations with I_{t-1}: the 'causes' must already be in train. These need not be direct causes, and could be very indirect: e.g., a variable's own lags may 'capture' actual past causes. Thus, when the relevant I_{t-1} is known, structure is not necessary for forecasting, even under changed conditions. Unfortunately, that I_{t-1} is known is most unlikely in economics, with important implications for understanding why 'ad hoc' methods can work well, as seen below.

Finally, explains the 'paradox' that (e.g.) the change in the log of real equity prices is unpredictable, but the level is predictable $x_t = \Delta x_t + x_{t-1}$.

2.3.2 Moments

Tend to focus on first and second moments if these exist: ν_t is unpredictable in mean if:

$$\mathsf{E}[\nu_t \mid I_{t-1}] = \mathsf{E}[\nu_t] \ \forall t.$$

Similarly for variance, unpredictable if:

$$\mathsf{V}[\nu_t \mid I_{t-1}] = \mathsf{V}[\nu_t] \ \forall t.$$

Converse of e.g. ARCH, GARCH and stochastic volatility. Unpredictable in mean is not invariant under non-linear transforms (e.g.):

$$\mathsf{E}[\nu_t \mid I_{t-1}] = \mathsf{E}[\nu_t] \quad \text{but} \quad \mathsf{E}[\nu_t\nu_t' \mid I_{t-1}] \neq \mathsf{E}[\nu_t\nu_t'],$$

but is minimum MSFE.

2.3.3 Horizon

If weakly stationary, the horizon H is such that:

$$\mathsf{V}[\nu_{T+H} \mid I_T] > \alpha \mathsf{V}[x_{T+H}].$$

Here, α may be 0.95, 0.99 etc.
If non-stationary (integrated of order one: I(1)) and inherently positive, use:

$$\sqrt{\mathsf{V}[\nu_{T+H} \mid I_T]} > \kappa \overline{x}.$$

Here, κ may be 0.25, 0.5 etc. If in logs, do not need to scale by sample mean.

2.3.4 Forecastability

A forecasting rule is any systematic operational procedure for making statements about future events. Events are forecastable relative to a loss measure if the rule produces a lower expected loss than the historical mean. Predictability is necessary but not sufficient for forecastability. Also need (a)–(d) above, which are sufficient, but not necessary. Thus, past is more explicable than future is forecastable (cf. stock-market commentators). Intertemporal transforms affect predictability, so no unique measure of forecast accuracy exists. This adds to the difficulty of theoretical analysis. New unpredictable components can enter in each period, so for integrated processes, $\mathsf{V}[x_{T+h}|I_T]$ $\mathsf{V}[x_{T+h}|I_T]$ is non-decreasing in h. Otherwise, can increase or decrease over horizons. Cannot prove that need 'genuinely' relevant information to forecast. Can show that 'irrelevant' or non-causal variables can be 'best available' forecasting devices in absence of omniscience.

2.3.5 Implications

These concepts have a number of important implications applicable to any statistical forecasting method. First, predictability is necessary but not sufficient for forecastability. From (1), since the conditional mean of an unpredictable process is its unconditional mean, predictability is necessary for forecastability. However, it is not sufficient: the relevant information set may be unknown in practice. There is a potential ambiguity in the phrase 'information set' in the contexts of predictability and forecasting: I_{t-1} denotes the conditioning set generated by the relevant events, whereas forecastability also requires knowledge of how I_{t-1} enters the conditional density in (1). For example, $\mathbf{v}_{t-1}$ may matter, but in an awkward non-linear way that eludes empirical modelling.

Secondly, translating 'regularity' as a systematic relation between the entity to be forecast and the available information, then conditions (a)–(d) above are sufficient for forecastability. They may not be necessary in principle (e.g., inspired guessing; precognition etc.), but for statistical forecasting, they seem close to necessary as can be seen by considering the removal of any one of them (e.g., if no regularities exist to be captured).

Thirdly, if the occurrence of large ex ante unpredictable shocks (such as earthquakes, or oil crises), induces their inclusion in later information sets (moving from (2) to (3) above), the past will be more explicable than the future is forecastable. Consequently, when the 'true' I_{t-1} is unknown, to prevent the baseline innovation error variance being an underestimate, forecast-accuracy evaluation may require 'unconditioning' from within-sample rare events that have been modelled post hoc. Conversely, forecast-period events determine the outcome of forecast evaluation tests.

Fourthly, from (4), intertemporal transforms affect predictability, so no unique measure of predictability, and hence of forecast accuracy, exists. Linear dynamic econometric systems are invariant under linear transforms in that they retain the same error process, and transformed estimates of the original are usually the direct estimates of the transformed system: such transforms are used regularly in empirical research. But by definition, the predictability of the transformed variables is altered by any transforms that are intertemporal (e.g., switching from yt on y_t on y_{t-1} to Δy_t on y_{t-1}).7 This precludes unique generic rankings of methods, adding to the difficulty of theoretical analysis and practical appraisal.

Next, since new unpredictable components can enter in each period, forecast error variances could increase or decrease over increasing horizons from any given T, as a consequence of (2) versus (3). For integrated processes, $\mathsf{V}[x_{T+h}|I_T]$ is non-decreasing in h when the innovation distribution is homoscedastic. Otherwise, when the initial forecast period T increases with real time, forecast uncertainty will be non-decreasing in h unless the innovation variance is ever-decreasing (since h-steps ahead from T becomes h − 1 from T + 1).

Finally, and the focus of §10, when the 'true' I_{t-1} is unknown one cannot prove that 'genuinely' relevant information must always dominate non-causal variables in forecasting. Rather, one can show in examples that the latter can be the 'best available' forecasting devices on some measures in the absence of omniscience (i.e., when the model is not the DGP). First, however, we need to explain the class of processes and models under analysis, and consider how forecast accuracy will be measured.

3 THE DGP AND MODELS

We need to be clear about what we mean by forecast failure. This is defined as significant misforecasting relative to the previous record (in-sample, or earlier forecasts), whereas poor forecasting is judged relative to some standard, either absolute (perhaps because of a policy requirement for accuracy), or relative to a rival model. Notice that forecasts may be poor simply because the series is inherently volatile, but this is not the same as forecast failure, the phenomenon we are primarily interested in explaining.

A further useful distinction is between ex-ante forecast failure and ex-post predictive failure. The ex ante notion relates to incorrect statements about as yet unobserved events, and could be due to many causes, including data errors or false assumptions about non-modelled variables which are corrected later, so the relevant model remains constant when updated. Thus ex-ante forecast failure is primarily a function of forecast-period events. Ex-post predictive failure is rejection on a valid parameter-constancy test against the observed outcomes, and occurs when a model is non-constant on the whole available information set, and is a well-established notion.

In Clements and Hendry (1999a) tables 2.1 and 2.2, we enumerate the possible sources of forecast error in systems of equations, and suggest that in practice, unmodelled shifts in deterministic factors may play an important role in forecast failure. A simpler taxonomy is outlined below. In section 8.1 the economy is represented by a cointegrated-integrated vector autoregression, subject to intermittent structural breaks, which can thus be written as a vector equilibrium9 correction system (VEqCM). The EqCM form is used for many of the equations in extant large-scale economic models that are routinely used for forecasting. The taxonomy allows for structural change in the forecast period, the model and DGP to differ over the sample period, the parameters of the model to be estimated from the data, and the forecasts to commence from incorrect initial conditions. Thus, all potential sources are included. This treatment also suggests why non-congruent models (e.g., vector autoregressions in differences that eschew cointegration or equilibria) need not fail when forecasting.

3.1 The Data Generation Process

For exposition, the data-generation process (DGP) is defined over the period t=1.......T by a firstorder vector autoregressive process (VAR) in the n variables $\mathbf{x}_t$

$$\mathbf{x}_t = \boldsymbol{\tau} + \boldsymbol{\Upsilon}\mathbf{x}_{t-1} + \boldsymbol{\nu}_t \text{ where } \boldsymbol{\nu}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \boldsymbol{\Omega}_\nu\right],$$

denoting an independent normal error with expectation $\mathsf{E}[\nu_t] = \mathbf{0}$ and variance matrix $\mathsf{V}[\nu_t] = \Omega_\nu$. The DGP is integrated of order unity (I(1)), and satisfies r < n cointegration relations such that:

$$\Upsilon = \mathbf{I}_n + \alpha\beta' \quad (6)$$

where α and β are $n \times r$ matrices of rank r. Then (5) can be reparameterized as the vector equilibriumcorrection model (VEqCM):

$$\Delta \mathbf{x}_t = \tau + \alpha\beta' \mathbf{x}_{t-1} + \nu_t \quad (7)$$

where $\Delta \mathbf{x}_t$ and $\beta' \mathbf{x}_t$ are I(0). Let:

$$\tau = \gamma - \alpha\mu \quad (8)$$

where μ is $r \times 1$ and $\beta'\gamma = \mathbf{0}$ so in deviations about means:

$$(\Delta \mathbf{x}_t - \gamma) = \alpha\left(\beta' \mathbf{x}_{t-1} - \mu\right) + \nu_t \quad (9)$$

where the system grows at the unconditional rate $\mathsf{E}[\Delta \mathbf{x}_t] = \gamma$ with long-run solution $\mathsf{E}\left[\beta' \mathbf{x}_t\right] = \mu$.

3.2 I(0) Representation

The notation $\mathbf{x}_t$ denotes I(1) vector, so we use $\mathbf{y}_t' = (\mathbf{x}_t'\beta : \Delta \mathbf{x}_t'\beta_\perp)$ (taking the r cointegrating vectors and (n – r) linear combinations of the $\Delta \mathbf{x}_t$) for the n-dimensional I(0) representation:

$$\mathbf{y}_t = \phi + \Pi \mathbf{y}_{t-1} + \epsilon_t \quad (10)$$

where:

$$\phi = \begin{pmatrix} -\beta'\alpha\mu \\ \beta_\perp'(\gamma - \alpha\mu) \end{pmatrix} \text{ and } \Pi = \begin{pmatrix} \Lambda & 0 \\ \beta_\perp'\alpha & 0 \end{pmatrix} \quad (11)$$

as:

$$\beta' x_t = -\beta' \alpha \mu + \Lambda \beta' x_{t-1} + \beta' \nu_t$$

where $\Lambda = I_r + \beta' \alpha$ denotes the dynamic matrix of the cointegration vectors, with all its eigenvalues inside the unit circle. Thus:

$$y_t = \begin{pmatrix} \beta' x_t \\ \beta'_{\perp} \Delta x_t \end{pmatrix} = \begin{pmatrix} -\beta' \alpha \mu \\ \beta'_{\perp} (\gamma - \alpha \mu) \end{pmatrix} + \begin{pmatrix} \Lambda \\ \beta'_{\perp} \alpha \end{pmatrix} \beta' x_{t-1} + \begin{pmatrix} \beta' \\ \beta'_{\perp} \end{pmatrix} \nu_t \quad (12)$$

While it is clearly restrictive to exclude any dynamics from Δx_{t-1}, the resulting algebra is much simpler, and we doubt if the analysis is seriously misled by doing so. For stationary processes, the restrictions in (11) can be ignored subject to the eigenvalues of Π remaining inside the unit circle. As this formulation is obtained by pre-multiplying (5) by the non-singular matrix $(\beta : \beta_{\perp})'$, it is isomorphic to the original.

3.3 The Model Class

The form of the model coincides with (5) as a linear representation of **xt**, but is potentially mis-specified:

$$x_t = \tau_p + \Upsilon_p x_{t-1} + u_t \quad (13)$$

where the parameter estimates $(\hat{\tau} : \hat{\Upsilon} : \hat{\Omega}_{\nu})$ are possibly inconsistent, with $\tau_p \neq \tau$ and $\Upsilon_p \neq \Upsilon$.Empirical econometric models like (13) are not numerically calibrated theoretical models, but have error processes which are derived, and so are not autonomous: see Gilbert (1986), Hendry (1995a), andSpanos (1986) inter alia. The theory of reduction explains the origin and status of such empirical models in terms of the implied information reductions relative to the process that generated the data.

Three specific models of interest are:

$$\Delta x_t = \gamma + (\alpha \beta' x_{t-1} - \mu) + \nu_t \quad (14)$$

$$\Delta \mathbf{x}_t = \gamma + \xi_t \quad (15)$$

and:

$$\Delta^2 \mathbf{x}_t = \mathbf{u}_t \quad (16)$$

(14) is correctly-specified in-sample, but will not be out-of-sample if the DGP alters over the forecast period. (14) is the dominant class of forecasting model, so its behaviour in the face of forecast-period parameter non-constancy needs to be understood if we are to cast light on actual experience. The second model (15) is a VAR in the differences of the variables (denoted DV), but is correctly specified when $\alpha = \mathbf{0}$ in (9), in which case $\xi_t = \nu_t$. Thus, its mis-specification in-sample is owing to omitting the cointegrating vectors, not differencing the data. The third model (denoted DDV) does difference the variables in (15), and is based on the assumption that economic variables do not accelerate or decelerate continually, namely $\mathsf{E}[\Delta^2 \mathbf{x}_t] = \mathbf{0}$, leading to forecasts of 'same change'.

3.4 Measuring Forecast Accuracy

As Clements and Hendry (1993) show, measures of forecast accuracy are often based on the forecast error second moment matrix:

$$\mathbf{V}_h \equiv \mathsf{E}\left[\mathbf{e}_{T+h}\mathbf{e}'_{T+h}\right] = \mathsf{V}\left[\mathbf{e}_{T+h}\right] + \mathsf{E}\left[\mathbf{e}_{T+h}\right]\mathsf{E}\left[\mathbf{e}'_{T+h}\right] \quad (17)$$

where $\mathbf{e}_{T+h}$ is a vector of h-step ahead forecast errors. This is the MSFE matrix, equal to the forecast error (co)variance matrix when forecasts are unbiased. Comparisons based on (17) may yield inconsistent rankings between forecasting models or methods for multi-step ahead forecasts.

Analyses could begin with the specification of a loss function, from which the optimal predictor can be derived, but a well-defined mapping between forecast errors and their costs, is not typically the case in macro-economics. However, the problem with measures based on (17) is that they lack invariance to non-singular, scale-preserving linear transformations, although the model class is invariant.

Clements and Hendry (1993) show analytically that for multi-step forecasts the trace of $\mathbf{V}_h$ is a measure which lacks invariance. Similarly, neither is the determinant of $\mathbf{V}_h$; and in fact, taking the matrix as a whole is insufficient to ensure invariance: $\mathbf{d}'\mathbf{V}_h\mathbf{d}$ is the smallest for every non-zero vector **d** is the MSFE matrix criterion.

Denote a linear forecasting system by the succinct notation:

$$\Gamma \mathbf{s}_t = \mathbf{u}_t \quad (18)$$

where $\mathbf{u}_t \sim \mathsf{ID}_{n+k}(\mathbf{0}, \boldsymbol{\Omega})$ i.e. independently distributed, zero mean with covariance matrix . $\boldsymbol{\Omega}$, $\mathbf{s}_t' = (\mathbf{x}_t' : \mathbf{z}_t')$, $\mathbf{x}_t$ are the n variables to be forecast and $\mathbf{z}_t$ are k available predetermined variables (perhaps just $\mathbf{x}_{t-1}$) and $\Gamma = (\mathbf{I} : -\mathbf{B})$ say. The parameters are $(\mathbf{B} : \boldsymbol{\Omega})$, where $\boldsymbol{\Omega}$ is symmetric, positive semi-definite. Then the likelihood and generalized variance of the system in (18) are invariant under scale-preserving, non-singular transformations of the form:

$$\mathbf{M}\Gamma\mathbf{P}^{-1}\mathbf{P}\mathbf{s}_t = \mathbf{M}\mathbf{u}_t$$

or

$$\Gamma^*\mathbf{s}_t^* = \mathbf{u}_t^* \quad \text{so} \quad \mathbf{u}_t^* \sim \mathsf{ID}_{n+k}\left[\mathbf{0}, \mathbf{M}\boldsymbol{\Omega}\mathbf{M}'\right] \quad (19)$$

In (19), $\mathbf{s}_t^* = \mathbf{P}\mathbf{s}_t$, $\mathbf{M}$ and $\mathbf{P}$ are respectively $n \times n$ and $(k+n) \times (k+n)$ non-singular matrices where abs $(|\mathbf{M}|) = 1$ and **P** is upper block-triangular, for example:

$$\mathbf{P} = \begin{bmatrix} \mathbf{I}_n & \mathbf{P}_{12} \\ \mathbf{0} & \mathbf{P}_{22} \end{bmatrix} \quad \text{so that} \quad \mathbf{P}^{-1} = \begin{bmatrix} \mathbf{I}_n & -\mathbf{P}_{12}\mathbf{P}_{22}^{-1} \\ \mathbf{0} & \mathbf{P}_{22}^{-1} \end{bmatrix}$$

Since we need to be able to calculate $\mathbf{P}^{-1}$, a restriction on **P** is that $|\mathbf{P}_{22}| \neq 0$. Then

$$\Gamma^* \equiv \mathbf{M}\Gamma\mathbf{P}^{-1} = \mathbf{M}\left(\mathbf{I} : -(\mathbf{P}_{12} + \mathbf{B})\mathbf{P}_{22}^{-1}\right) = \mathbf{M}(\mathbf{I} : -\mathbf{B}^*) \quad (20)$$

The systems (18) and (19) are isomorphic. Forecasts and forecast confidence intervals made in the original system and transformed after the event to $\mathbf{x}_t^*$ or made initially from the transformed system will be identical, and this remains true when the parameters are estimated by maximum likelihood.

If we let $\mathbf{V}_h$ denote the MSFE matrix for $\mathbf{x}_t$, and $\mathbf{V}_h^*$ the MSFE matrix for $\mathbf{x}_t$ for any other method, then for transformations involving **M** only:

(i) the trace measure for x_t is not in general equivalent to that for x_t^*:

$$tr\,(\mathbf{M}\mathbf{V}_h\mathbf{M}') \neq tr\,(\mathbf{V}_h)$$

(ii) the determinant of the matrix is invariant:

$$|\mathbf{M}\mathbf{V}_h\mathbf{M}'| = |\mathbf{V}_h| \text{ when } |\mathbf{M}| = 1$$

(iii) the Granger–Newbold matrix measure is invariant:

$\mathbf{d}^{*\prime}\mathbf{V}_h\mathbf{d}^* < \mathbf{d}^{*\prime}\mathbf{V}_h^*\mathbf{d} \quad \forall \mathbf{d}^* \neq \mathbf{0}$ implies $\mathbf{d}'\mathbf{M}\mathbf{V}_h\mathbf{M}'\mathbf{d} < \mathbf{d}'\mathbf{M}\mathbf{V}_h^*\mathbf{M}'\mathbf{d}$ $\forall \mathbf{d} \neq \mathbf{0}$, and for all scale-preserving **M**.

For transformations using **P** both the determinant and the MSFE matrix criteria are not invariant for multi-step forecasts. For h > 1 the invariance property requires taking account of covariance terms between different step-ahead forecast errors, leading to a generalized MSFE matrix:

$$\mathbf{\Phi}_h = \mathsf{E}\left[\mathbf{E}_h\mathbf{E}_h'\right]$$

where $\mathbf{E}_h$ is obtained by stacking the forecast step errors up to and including the h-step ahead errors

$$\mathbf{E}_h' = \left[\mathbf{e}_{T+1}', \mathbf{e}_{T+2}', \ldots, \mathbf{e}_{T+h-1}', \mathbf{e}_{T+h}'\right]$$

This is the GFESM.

For example, j_hj is unaffected by transforming the data by **M**(where $|\mathbf{M}|$ = 1). Denote the vector of stacked forecast errors from the transformed model as $\tilde{\mathbf{E}}_h'$, so that

$$\tilde{\mathbf{E}}_h' = \left[\mathbf{e}_{T+1}'\mathbf{M}', \mathbf{e}_{T+2}'\mathbf{M}', \ldots, \mathbf{e}_{T+h-1}'\mathbf{M}', \mathbf{e}_{T+h}'\mathbf{M}'\right]$$

or $\tilde{\mathbf{E}}_h = (\mathbf{I} \otimes \mathbf{M})\,\mathbf{E}_h$. Thus

$$\left|\tilde{\mathbf{\Phi}}_h\right| = \left|\mathsf{E}\left[\tilde{\mathbf{E}}_h\tilde{\mathbf{E}}_h'\right]\right| = \left|\mathsf{E}\left[(\mathbf{I}\otimes\mathbf{M})\,\mathbf{E}_h\mathbf{E}_h'(\mathbf{I}\otimes\mathbf{M}')\right]\right| = \left|\mathsf{E}\left[\mathbf{E}_h\mathbf{E}_h'\right]\right| \times |\mathbf{I}\otimes\mathbf{M}|^2 = \left|\mathsf{E}\left[\mathbf{E}_h\mathbf{E}_h'\right]\right|$$

since $|\mathbf{I}\otimes\mathbf{M}| = |\mathbf{I}\otimes\mathbf{M}'| = 1$

Transforming by **P** leaves the error process $\{\mathbf{e}_{T+i}\}$, and therefore $\mathbf{E}_h$, unaffected, demonstrating invariance to **P** transforms.

Generalizing the Granger-Newbold matrix criterion to apply to the pooled or stacked forecast error second moment matrix Φ_h, then the model denoted by $\sim$ dominates that denoted by ^ if:

$$\widehat{\Phi}_h - \widetilde{\Phi}_h \succ 0$$

that is, if the difference between the two estimates of the GFESM matrix is positive definite. Dominance on this measure is sufficient but not necessary for dominance on the determinant of the GFESM matrix:

$$\widehat{\Phi}_h - \widetilde{\Phi}_h \succ 0 \Rightarrow \left|\widehat{\Phi}_h\right| > \left|\widetilde{\Phi}_h\right|$$

since $\widehat{\Phi}_h$ and $\widetilde{\Phi}_h$ are positive definite. GFESM is close to predictive likelihood: see Bjørnstad (1990).

4 CAUSAL INFORMATION IN ECONOMIC FORECASTING

We now consider the role of causal information in economic forecasting first when the model coincides with the mechanism, then when it does not; the mechanism is allowed to be non-constant over time. In the first case, causal information is always useful, and produces better forecasts than non-causal. Adding further variables produces no improvement. Even when the model is mis-specified, causallyrelevant information generally improves forecasts providing the mechanism generates stationary data. Such a result cannot be shown for a mis-specified model of a non-constant mechanism, and non-causal additional variables potentially can be more useful than causally-relevant ones so long as the model remains mis-specified.

To demonstrate these claims, we assume all parameters are known: estimation uncertainty would reinforce the main conclusion. While sufficiently poor estimates would weaken any conclusions from the first case, our concern is to establish that causally-relevant variables cannot be relied upon to produce the 'best' forecasts when the model is mis-specified, and parameter uncertainty would strengthen this finding.

4.1 Model Coincides with the Mechanism

Consider the DGP in (5) for the n I(1) variables $\mathbf{x}_t$. Here, (5) is both the model and the DGP, although it could be written in a lower-dimensional parameter space in terms of I(0) transformations of the original variables as in (14) above. The notation is simplest when the mechanism is constant, so we prove the result for 1-step forecasts in that setting first.

The in-sample conditional expectation of $\mathbf{x}_{T+1}$ given $\mathbf{x}_T$ is:

$$\mathsf{E}\left[\mathbf{x}_{T+1} \mid \mathbf{x}_T\right] = \boldsymbol{\tau} + \boldsymbol{\Upsilon}\mathbf{x}_T$$

and this delivers the (matrix) minimum MSFE. Under the present assumptions, the resulting forecast error is a homoscedastic innovation against all further information:

$$\mathsf{E}\left[\boldsymbol{\nu}_{T+1} \mid \mathbf{x}_T\right] = \mathbf{0} \text{ and } \mathsf{V}\left[\boldsymbol{\nu}_{T+1} \mid \mathbf{x}_T\right] = \boldsymbol{\Omega} \quad (21)$$

Consequently, adding any further variables $\mathbf{z}_{t-1}$ to (5) will not improve the forecast accuracy of mean or variance.

Conversely, replacing any $x_{i,t-1}$ by any or all elements from $\mathbf{z}_{t-1}$ will lead to inefficient forecasts unless there is perfect correlation between $x_{i,t}$ and $\mathbf{z}_t$. Denote the resulting regressor vector by $\overline{\mathbf{x}}_{t-1}$, then, forecasting from:

$$\mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\Gamma}\overline{\mathbf{x}}_{t-1} + \mathbf{e}_t$$

where $\mathsf{E}[\mathbf{e}_t|\overline{\mathbf{x}}_{t-1}] = \mathbf{0}$ using:

$$\widetilde{\mathbf{x}}_{T+1} = \boldsymbol{\gamma} + \boldsymbol{\Gamma}\overline{\mathbf{x}}_T$$

the forecast error is:

$$\mathbf{e}_{T+1} = \mathbf{x}_{T+1} - \widetilde{\mathbf{x}}_{T+1} = (\boldsymbol{\tau} - \boldsymbol{\gamma}) + \boldsymbol{\Upsilon}\mathbf{x}_T - \boldsymbol{\Gamma}\overline{\mathbf{x}}_T + \boldsymbol{\nu}_{T+1}$$

Let $\mathbf{x}_t = \boldsymbol{\zeta} + \boldsymbol{\Psi}\overline{\mathbf{x}}_t + \mathbf{w}_t$ (say) with $\mathsf{E}[\mathbf{w}_t|\overline{\mathbf{x}}_t] = \mathbf{0}$ and $\mathsf{V}[\mathbf{w}_t|\overline{\mathbf{x}}_t] = \boldsymbol{\Phi}$, so:

$$\mathbf{e}_{T+1} = (\boldsymbol{\tau} - \boldsymbol{\gamma} + \boldsymbol{\Upsilon}\boldsymbol{\zeta}) + (\boldsymbol{\Upsilon}\boldsymbol{\Psi} - \boldsymbol{\Gamma})\overline{\mathbf{x}}_T + \boldsymbol{\Upsilon}\mathbf{w}_T + \boldsymbol{\nu}_{T+1}$$

with mean:

$$\mathsf{E}[\mathbf{e}_{T+1} \mid \overline{\mathbf{x}}_T] = (\boldsymbol{\tau} - \boldsymbol{\gamma} + \boldsymbol{\Upsilon}\boldsymbol{\zeta}) + (\boldsymbol{\Upsilon}\boldsymbol{\Psi} - \boldsymbol{\Gamma})\overline{\mathbf{x}}_T = \mathbf{0} \quad (22)$$

so that $\boldsymbol{\gamma} = \boldsymbol{\tau} + \boldsymbol{\Upsilon}\boldsymbol{\zeta}$ and $\boldsymbol{\Upsilon}\boldsymbol{\Psi} = \boldsymbol{\Gamma}$; and variance:

$$\mathsf{V}[\mathbf{e}_{T+1} \mid \overline{\mathbf{x}}_T] = \boldsymbol{\Omega} + \boldsymbol{\Upsilon}\boldsymbol{\Phi}\boldsymbol{\Upsilon}' \quad (23)$$

Thus, the forecasts are conditionally unbiased (22), but inefficient (23).

Next, in a non-constant DGP, the taxonomy shows that the main non-constancies of interest concern direct or indirect changes in the deterministic components of (5). Either $\boldsymbol{\tau}$ can change, or if $\boldsymbol{\Upsilon}$ changes, the unconditional means of the I(0) components alter. We only consider the former. Let $\boldsymbol{\tau}$ change to $\boldsymbol{\tau}^*$, so the DGP in the forecast period becomes:

$$\mathbf{x}_{T+1} = \boldsymbol{\tau}^* + \boldsymbol{\Upsilon}\mathbf{x}_T + \boldsymbol{\nu}_T \quad (24)$$

Since the model also switches to (24) by being the mechanism, the forecast errors have the same properties as in (21), and the previous result is unchanged. Its converse, that (24) will dominate incorrect models, is more tedious to show, but follows from a generalization of the argument in (22) and (23).

Such powerful results are not surprising; but the assumption that the model coincides with the mechanism is extremely strong and not empirically relevant.

4.2 Model Does Not Coincide with the Mechanism

First, we show that if the process is stationary, predictive failure is unconditionally unlikely, irrespective of how badly the model is specified (see Hendry, 1979), but that causal information dominates noncausal. Even so, non-causal might help, if it acts as a proxy for the omitted causal variables.. Then we provide an example where causal information does not help once structural breaks are introduced.

Reparameterize the system as in (14):

$$\Delta \mathbf{x}_t = \gamma + \alpha \left(\beta' \mathbf{x}_{t-1} - \mu \right) + \nu_t \quad (25)$$

There are many ways in which a model could be mis-specified for the mechanism in (25), but we only consider omission of the I(0) cointegrating components. Denote the model by:

$$\Delta \mathbf{x}_t = \delta + \rho \left(\beta_1' \mathbf{x}_{t-1} - \mu_1 \right) + \eta_t \quad (26)$$

where β_1' is (perhaps a linear transform of) a subset of the r cointegrating vectors in (25), and μ_1 is the unconditional expectation of $\beta_1' \mathbf{x}_t$. Then, as $\mathsf{E}[\beta_1' \mathbf{x}_{t-1}] = \mu_1$, $\delta = \gamma$, and hence for known parameters in (26) and forecast $\widehat{\Delta \mathbf{x}}_{T+1} = \gamma + \rho(\beta_1' \mathbf{x}_T - \mu_1)$

$$\mathsf{E}\left[\widehat{\Delta \mathbf{x}}_{T+1}\right] = \gamma$$

So forecasts are unconditionally unbiased, though inefficient. Adding any omitted I(0) linear combinations of $\mathbf{x}_{t-1}$ will improve forecasts, as will adding any $\Delta \mathbf{x}_{t-1}$ which proxy for omitted $\beta_2' \mathbf{x}_{t-1}$.

Thus, the notion of basing forecasting on 'causal models' still has substance, perhaps qualified by the need to estimate parameters from small samples of badly-measured data. However, once the model is not the mechanism and the mechanism is non-constant, the dominance of causal information over non-causal cannot be shown. We consider a counter example where non-causal information dominates causal on at least one forecast criterion, unless omniscience is assumed. The result may help explainsome of the apparent success of the approach in Box and Jenkins (1976).

4.3 Example

Consider the following I(1) DGP:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \Pi_{1,2} y_{2,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \quad (27)$$

which holds till time τ, then changes to:

$$\begin{pmatrix} y_{1,\tau+i} \\ y_{2,\tau+i} \end{pmatrix} = \begin{pmatrix} \Pi^*_{1,2} y_{2,\tau+i-1} \\ y_{2,\tau+i-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,\tau+i} \\ \epsilon_{2,\tau+i} \end{pmatrix} \quad (28)$$

Only the first block in (27) is modelled, with forecasts generated by the correct insample system:

$$\widehat{y}_{1,\tau+i} = \Pi_{1,2} y_{2,\tau+i-1} \quad (29)$$

so that after the break, the forecast error is:

$$\mathsf{E}\left[y_{1,\tau+i} - \widehat{y}_{1,\tau+i} \mid y_{2,\tau+i-1}\right] = \left(\Pi^*_{1,2} - \Pi_{1,2}\right) y_{2,\tau+i-1}$$

Since y2;t is a vector random walk, it will almost always be non-zero, so the forecast error could be large as well as persistent.

By way of contrast, consider the following non-causal forecasting rule:

$$\widetilde{y}_{1,\tau+i} = y_{1,\tau+i-1}$$

This is a purely extrapolative mechanism, but 'error corrects' to the previous level. Thus, using the fact that the unconditional growth, $\mathsf{E}[\Delta y_{2,\tau+i-1}] = 0$;

$$\begin{aligned} \mathsf{E}\left[y_{1,\tau+i} - \widetilde{y}_{1,\tau+i} \mid y_{2,\tau+i-1}\right] &= \mathsf{E}\left[\Delta y_{1,\tau+i} \mid y_{2,\tau+i-1}\right] \\ &= \begin{cases} \left(\Pi^*_{1,2} - \Pi_{1,2}\right) y_{2,\tau-1} & \text{for } i = 1 \\ 0 & \text{for } i > 1. \end{cases} \end{aligned}$$

By construction, the lagged y_1 is non-causal, yet forecasts based on it are less biased after $\tau + 1$. For possible changes in parameter values, such shifts could dominate any variance losses.

5 THE FORMAL FORECAST ERRORS TAXONOMY

Given the information set $\{\mathbf{y}_t\}$, and the knowledge that the system is linear with one lag, using estimated parameters ('^'s on parameters denote estimates, and on random variables, forecasts), the h-step ahead forecasts at forecast-origin T for $h = 1, \ldots, H$ are:

$$\hat{\mathbf{y}}_{T+h} - \hat{\boldsymbol{\varphi}} = \hat{\mathbf{\Pi}} \left(\hat{\mathbf{y}}_{T+h-1} - \hat{\boldsymbol{\varphi}} \right) = \hat{\mathbf{\Pi}}^h \left(\hat{\mathbf{y}}_T - \hat{\boldsymbol{\varphi}} \right) \quad (30)$$

where $\hat{\boldsymbol{\varphi}} = (\mathbf{I}_n - \hat{\mathbf{\Pi}})^{-1}\hat{\boldsymbol{\phi}}$. In this simple system, the h-step ahead forecast equals the estimated equilibrium mean, plus the deviation therefrom at the estimated forecast origin, scaled by the h^{th} power of the dynamic matrix. Although the forecast origin $\hat{\mathbf{y}}_T$ is uncertain, we assume $\mathsf{E}[\hat{\mathbf{y}}_T] = \boldsymbol{\varphi}$, so on average $\hat{\mathbf{y}}_T$ is unbiased (otherwise, an additional term arises in the taxonomy from that bias). By using (30), we do not assume that the forecaster knows the DGP, since the taxonomy allows for the possibility that all the parameter estimates are inconsistent, reflecting any extent of model mis-specification. For example, $\hat{\mathbf{\Pi}}$ might be restricted to zero (dynamic mis-specification); the wrong variables used in the various equations, or the intercepts suppressed despite $\boldsymbol{\phi} \neq \mathbf{0}$. A subscript p on a parameter denotes the plim$_{T\to\infty}$ (under constant parameters) of the corresponding estimate.

Because the system is dynamic, the impacts of breaks differ with the time-lapse since the break. Thus, after a structural break, the system becomes non-stationary in that its first and second moments are not constant. Consequently, every moment has to be calculated explicitly, depending on the timing of the break. We consider the case when a single permanent break occurs at the forecast announcement: unknown to the forecaster, at time T, the parameters $(\boldsymbol{\phi} : \mathbf{\Pi})$ change to $(\boldsymbol{\phi}^* : \mathbf{\Pi}^*)$ where $\mathbf{\Pi}^*$ still has all its eigenvalues inside the unit circle. Thus, from T + 1 onwards, the data are generated by:

$$\mathbf{y}_{T+h} = \boldsymbol{\phi}^* + \mathbf{\Pi}^* \mathbf{y}_{T+h-1} + \boldsymbol{\epsilon}_{T+h}, \quad h = 1, \ldots \quad (31)$$

Letting $\boldsymbol{\phi}^* = (\mathbf{I}_n - \mathbf{\Pi}^*) \boldsymbol{\varphi}^*$:

$$\begin{aligned} \mathbf{y}_{T+h} - \boldsymbol{\varphi}^* &= \mathbf{\Pi}^* \left(\mathbf{y}_{T+h-1} - \boldsymbol{\varphi}^* \right) + \boldsymbol{\epsilon}_{T+h} \\ &= \left(\mathbf{\Pi}^* \right)^h \left(\mathbf{y}_T - \boldsymbol{\varphi}^* \right) + \sum_{i=0}^{h-1} \left(\mathbf{\Pi}^* \right)^i \boldsymbol{\epsilon}_{T+h-i} \end{aligned} \quad (32)$$

The future outcomes, as a deviation from the new equilibrium, are the appropriate power of the new dynamic matrix, times the deviation at the forecast origin, but measured from the new equilibrium, plus the accumulated 'discounted' future errors.

From (30) and (32), the h-step ahead forecast errors $\hat{\epsilon}_{T+h} = \mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h}$ are:

$$\hat{\epsilon}_{T+h} = \varphi^* - \hat{\varphi} + (\mathbf{\Pi}^*)^h (\mathbf{y}_T - \varphi^*) - \hat{\mathbf{\Pi}}^h (\hat{\mathbf{y}}_T - \hat{\varphi}) + \sum_{i=0}^{h-1} (\mathbf{\Pi}^*)^i \epsilon_{T+h-i} \tag{33}$$

(33) can be rearranged in various ways and further refined. The following brings out some of the key terms without unduly complicating matters. We abstract from forecast origin uncertainty $(\delta_y = \hat{\mathbf{y}}_T - \mathbf{y}_T = \mathbf{0})$ and parameter estimation uncertainty, that is $\delta_\varphi = \hat{\varphi} - \varphi_p$ and $\delta_\Pi = \hat{\mathbf{\Pi}} - \mathbf{\Pi}_p$, the deviations of sample estimates from population parameters, are assumed to be zero. δ_φ and δ_Π have only variance effects, assuming finite-sample estimation biases are zero, and then are only of order $O\left(T^{-1}\right)$.

In the present formulation, the rows (ib) and (iib) alone induce biases, whereas the Slope change and Slope mis-specification terms (rows (ia) and (iia)) have unconditional expectations of zero, since $\mathsf{E}(\mathbf{y}_T - \varphi) = \mathbf{0}$, and only affect forecast-error variances. Thus systematic forecast-error biases result when:

$$\mathsf{E}\left[\hat{\epsilon}_{T+h}\right] = \left(\mathbf{I}_n - (\mathbf{\Pi}^*)^h\right)(\varphi^* - \varphi) + \left(\mathbf{I}_n - \mathbf{\Pi}_p^h\right)(\varphi - \varphi_p) \tag{34}$$

is non-zero.

First consider the term involving $(\varphi - \varphi_p)$ in (34). Almost all estimation methods ensure that residuals have zero means in-sample, so provided ' has remained constant in-sample, this term is zero by construction. However, if ' has previously altered, and that earlier shift has not been modelled, then φ_p will be a weighted average of the in-sample values, and hence will not equal the end-of-sample value φ. One advantage of developing models that are congruent in-sample, even when the objective is forecasting, is to minimize such effects. When $\varphi = \varphi_p$, forecasts will be biased only to the extent that the long-run mean shifts from the in-sample population value.

Table 2 Simplified forecast-error taxonomy.

$$
\begin{aligned}
\epsilon_{T+h} \simeq\ & \left((\mathbf{\Pi}^{*})^{h}-\mathbf{\Pi}^{h}\right)(\mathbf{y}_{T}-\varphi) && (ia)\ \text{Slope change}\\
&+\left(\mathbf{I}_{n}-(\mathbf{\Pi}^{*})^{h}\right)(\varphi^{*}-\varphi) && (ib)\ \text{Equilibrium-mean change}\\
&+\left(\mathbf{\Pi}^{h}-\mathbf{\Pi}_{p}^{h}\right)(\mathbf{y}_{T}-\varphi) && (iia)\ \text{Slope mis-specification}\\
&+\left(\mathbf{I}_{n}-\mathbf{\Pi}_{p}^{h}\right)(\varphi-\varphi_{p}) && (iib)\ \text{Equilibrium-mean mis-specification}\\
&+\sum_{i=0}^{h-1}(\mathbf{\Pi}^{*})^{i}\epsilon_{T+h-i} && (iii)\ \text{Error accumulation.}
\end{aligned}
$$

then φ_p will be a weighted average of the in-sample values, and hence will not equal the end-of-sample value φ. One advantage of developing models that are congruent in-sample, even when the objective is forecasting, is to minimize such effects. When $\varphi = \varphi_p$, forecasts will be biased only to the extent that the long-run mean shifts from the in-sample population value.

Next, consider the case when $\varphi^* \neq \varphi$. A systematic bias results: since $(\mathbf{\Pi}^*)^h \to \mathbf{0}$ as $h \to \infty$, this is increasing in h, and eventually rises to the full effect of $(\varphi^* - \varphi)$. Consequently, a sequence of same-signed, increasing magnitude, forecast errors should result from a deterministic shift (here, in the equilibrium mean). Moreover, such effects do not die out as the horizon increases, but converge to the full impact of the shift.

By way of contrast, changes in the dynamics, and dynamic parameter mis-specifications, are both multiplied by mean-zero terms, so vanish on average: indeed, they would have no effect whatever on the forecast errors if the forecast origin equalled the equilibrium mean. Conversely, the larger the disequilibrium at the time of a shift in the dynamics, the larger the resulting impact.

Consider now variance effects, which contribute to the MSFE. The variances of the cumulated errors (assumed independent over time):

$$\mathsf{V}[\epsilon_{T+h}] = \sum_{i=0}^{h-1} (\mathbf{\Pi}^*)^i \,\mathbf{\Omega}_\epsilon\, (\mathbf{\Pi}^*)^{i\prime} \tag{35}$$

are likely to dominate in practice – parameter estimation effects are of order T^{-1} and typically are small in comparison.

5.1 Forecast-Error Biases and Variances in 3 Models

The unconditional biases of the one and two step ahead forecasts from three contending models (14), (15) and (16) are reported in tables 3 (pre-break forecasting) and 4 (post-break forecasting). The symbol $\nabla_\alpha = \alpha^* - \alpha$ etc., and $\mathbf{B} = \mathbf{I}_r - \Lambda = \beta'\alpha$ where $\Lambda = \mathbf{I}_r + \beta'\alpha$.

The various methods perform about equally badly when forecasting before a break that occurs over the forecast horizon, whereas the biases are much smaller for DDV if the break has happened before the forecast is made, albeit that this is not taken into account in the forecast. Since DV and DDV are 'noncausal' forecasting devices, we see that non-causal information can dominate causal when forecasting in the face of deterministic shifts.

Table 5 shows the excess in the 1-step variances over Ω, for various values of α, the parameter change that most affects the variances, where $\mathbf{A}^* = \alpha^*\Lambda - \alpha$ and $\mathbf{H}^* = \mathbf{I}_n - \alpha^*\beta'$.

Table 3 1 and 2-steps pre break.

	VEqCM	DV	DDV
	1-step forecasts		
$\mu \to \mu^*$	$\alpha\nabla_\mu$	$\alpha\nabla_\mu$	$\alpha\nabla_\mu$
$\gamma \to \gamma^*$	∇_γ	∇_γ	∇_γ
$\alpha \to \alpha^*$	0	0	0
	2-step forecasts		
$\mu \to \mu^*$	$\alpha\mathbf{B}\nabla_\mu$	$\alpha\mathbf{B}\nabla_\mu$	$\alpha\mathbf{B}\nabla_\mu$
$\gamma \to \gamma^*$	$2\nabla_\gamma$	$2\nabla_\gamma$	$2\nabla_\gamma$
$\alpha \to \alpha^*$	0	0	0

Table 4 1 and 2-steps post break.

	VEqCM	DV	DDV
1-step forecasts			
$\mu \to \mu^*$	$\alpha\nabla_\mu$	$\alpha\Lambda\nabla_\mu$	$\alpha\beta'\alpha\nabla_\mu$
$\gamma \to \gamma^*$	∇_γ	∇_γ	$\mathbf{0}$
$\alpha \to \alpha^*$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$
2-step forecasts			
$\mu \to \mu^*$	$\alpha\mathbf{B}\nabla_\mu$	$\alpha\Lambda\mathbf{B}\nabla_\mu$	$\alpha\left(\mathbf{I}_r + \mathbf{B}\right)\beta'\alpha\nabla_\mu$
$\gamma \to \gamma^*$	$2\nabla_\gamma$	$2\nabla_\gamma$	$\mathbf{0}$
$\alpha \to \alpha^*$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$

The unstarred matrices replace any starred parameter with its unstarred value: thus, $\mathbf{A} = \alpha\Lambda - \alpha = \alpha\left(\beta'\alpha\right)$. The pattern is clear, and the only exception is the DDV when $\alpha \neq \alpha^*$, although the fact that $\mathbf{V}^* = \Lambda^*\mathbf{V}\Lambda^{*\prime} + \beta'\Omega\beta$ allows considerable rearrangement without altering the substance.

Next, table 6 records the excess in the 2-step variances over Ω, where $\mathbf{C}^* = \alpha^*\left(\mathbf{I}_r + \Lambda^*\right)$, $\mathbf{D}^* = \mathbf{C}^* - \mathbf{C}$, $\mathbf{F}^* = \mathbf{C}^*\Lambda - 2\alpha$. As before, $\mathbf{C} = \alpha\left(\mathbf{I}_r + \Lambda\right)$, so $\mathbf{D} = \mathbf{0}$, and $\mathbf{F} = \alpha(\beta'\alpha)\left(2\mathbf{I}_r + \Lambda\right)$ and $\mathbf{G} = \alpha\beta'(2\mathbf{I}_n + \alpha\beta') - 2\mathbf{I}_n$; where $\mathbf{F}^{**} = \mathbf{C}^*\Lambda^* - 2\alpha^*$.

The pattern is similar to the 1-step outcomes, although the values are larger, and the formulae more complicated: the rapid increase in the DDV variance is especially noticeable.

5.2 Discussion

When forecasting before a break, all three models are susceptible to forecast failure, and there is little to choose between them, although the VEqCMhas the smallest variance component when it is correctly specified and no break occurs. When forecasting after a break, the DDV has the greatest robustness to a deterministic shift, but the largest and most rapidly-increasing forecast-error variances in general. The DV lies between, depending on what deterministic terms change.

Nevertheless, the longer the multi-step evaluation horizon, the less well the DDV, and probably the DV will perform, partly from their variance terms, and partly because most breaks will be afterforecasting, a case in which these models offer no gains. Conversely, the shorter the horizon, for a sequence of horizons, the more likely some breaks will precede forecasting, and consequently, DDV and DV may outperform the VEqCM, even when it is correctly-specified in-sample .

Table 5 1-step unconditional variances.

$T \to T+1$	VEqCM	DV	DDV
$\alpha = 0$	0	0	$\boldsymbol{\Omega}$
$\alpha = \alpha^*$	0	$\alpha \mathbf{V} \alpha'$	$\mathbf{H}\boldsymbol{\Omega}\mathbf{H}' + \mathbf{A}\mathbf{V}\mathbf{A}'$
$\alpha \neq \alpha^*$	$\nabla_\alpha \mathbf{V} \nabla'_\alpha$	$\alpha^* \mathbf{V} \alpha^{*\prime}$	$\mathbf{H}^*\boldsymbol{\Omega}\mathbf{H}^{*\prime} + \mathbf{A}^*\mathbf{V}\mathbf{A}^{*\prime}$
$T+1 \to T+2$	VEqCM	DV	DDV
$\alpha = 0$	0	0	$\boldsymbol{\Omega}$
$\alpha = \alpha^*$	0	$\alpha \mathbf{V} \alpha'$	$\mathbf{H}\boldsymbol{\Omega}\mathbf{H}' + \mathbf{A}\mathbf{V}\mathbf{A}'$
$\alpha \neq \alpha^*$	$\nabla_\alpha \mathbf{V}^* \nabla'_\alpha$	$\alpha^* \mathbf{V}^* \alpha^{*\prime}$	$\mathbf{H}^*\boldsymbol{\Omega}\mathbf{H}^{*\prime} + \alpha^*\beta'\alpha^*\mathbf{V}\alpha^{*\prime}\beta\alpha^{*\prime}$

Here and in the text $\alpha = 0$ implicitly implies $\alpha^* = 0$.

Table 6 2-step unconditional variances.

$T \to T+2$	VEqCM	DV	DDV
$\alpha = 0$	$\boldsymbol{\Omega}$	$\boldsymbol{\Omega}$	$5\boldsymbol{\Omega}$
$\alpha = \alpha^*$	$\boldsymbol{\Upsilon}\boldsymbol{\Omega}\boldsymbol{\Upsilon}'$	$\boldsymbol{\Upsilon}\boldsymbol{\Omega}\boldsymbol{\Upsilon}' + \mathbf{C}\mathbf{V}\mathbf{C}'$	$\boldsymbol{\Upsilon}\boldsymbol{\Omega}\boldsymbol{\Upsilon}' + \mathbf{G}\boldsymbol{\Omega}\mathbf{G}' + \mathbf{F}\mathbf{V}\mathbf{F}'$
$\alpha \neq \alpha^*$	$\boldsymbol{\Upsilon}^*\boldsymbol{\Omega}\boldsymbol{\Upsilon}^{*\prime} + \mathbf{D}^*\mathbf{V}\mathbf{D}^{*\prime}$	$\boldsymbol{\Upsilon}^*\boldsymbol{\Omega}\boldsymbol{\Upsilon}^{*\prime} + \mathbf{C}^*\mathbf{V}\mathbf{C}^{*\prime}$	$\boldsymbol{\Upsilon}^*\boldsymbol{\Omega}\boldsymbol{\Upsilon}^{*\prime} + \mathbf{G}^*\boldsymbol{\Omega}\mathbf{G}^{*\prime} + \mathbf{F}^*\mathbf{V}\mathbf{F}^{*\prime}$
$T+1 \to T+3$	VEqCM	DV	DDV
$\alpha = 0$	$\boldsymbol{\Omega}$	$\boldsymbol{\Omega}$	$5\boldsymbol{\Omega}$
$\alpha = \alpha^*$	$\boldsymbol{\Upsilon}\boldsymbol{\Omega}\boldsymbol{\Upsilon}'$	$\boldsymbol{\Upsilon}\boldsymbol{\Omega}\boldsymbol{\Upsilon}' + \mathbf{C}\mathbf{V}\mathbf{C}'$	$\boldsymbol{\Upsilon}\boldsymbol{\Omega}\boldsymbol{\Upsilon}' + \mathbf{G}\boldsymbol{\Omega}\mathbf{G}' + \mathbf{F}\mathbf{V}\mathbf{F}'$
$\alpha \neq \alpha^*$	$\boldsymbol{\Upsilon}^*\boldsymbol{\Omega}\boldsymbol{\Upsilon}^{*\prime} + \mathbf{D}^*\mathbf{V}^*\mathbf{D}^{*\prime}$	$\boldsymbol{\Upsilon}^*\boldsymbol{\Omega}\boldsymbol{\Upsilon}^{*\prime} + \mathbf{C}^*\mathbf{V}^*\mathbf{C}^{*\prime}$	$\boldsymbol{\Upsilon}^*\boldsymbol{\Omega}\boldsymbol{\Upsilon}^{*\prime} + \mathbf{G}^*\boldsymbol{\Omega}\mathbf{G}^{*\prime} + \mathbf{F}^{**}\mathbf{V}\mathbf{F}^{**\prime}$

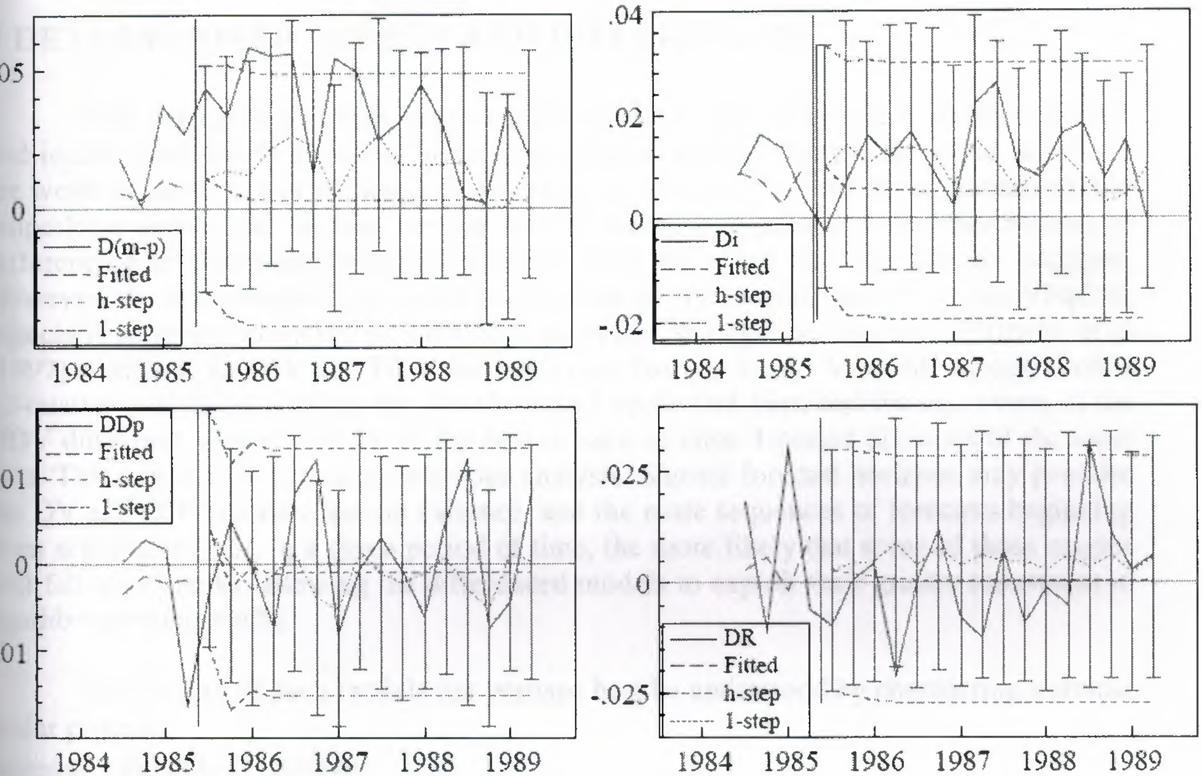
This behaviour is precisely what was observed by Eitrheim, Husebø and Nymoen (1997) in their study of the forecasting performance of the Norges Bank model. Over the longest (12 quarter) evaluation horizon, the Bank's model performed well, followed by a DV modelled to be congruent: the equivalent of the DDV did worst. But over a sequence of three 4-period divisions of the same evaluation data, the DDV did best more often than any other method. The empirical illustration confirms similar results for a small monetary model of the UK.

6 EQUILIBRIUM CORRECTION AND ERROR CORRECTION

The preceding analysis suggests that equilibrium mean shifts may be an important cause of sustained forecast failure. If forecasts are made prior to the shift having occurred, then any forecasting model or device that did not anticipate the change is likely to go badly wrong. As time progresses, the forecaster who habitually makes forecasts each month (say) will eventually forecast from a 'post-shift' origin. We show below that the forecaster who uses a VEqCM will continue to make biased forecasts, while the forecasts produced by the user of the DV model will eventually 'error-correct' to the changed state of affairs, albeit that these forecasts may be less precise.

The problem with EqCMs is that they force variables back in to relations that reflect the previous equilibria —if the equilibrium means have altered to new values, then EqCM models will correct to inappropriate values. Because the new, changed levels are viewed by the estimated model as disequilibria, forecasts will continually be driven off course. UK M1 provides one potential example of equilibriummean shifts following the introduction in 1984 of interest-bearing retail sight deposits: these sharply lowered the opportunity costs of holding M1, shifting the long-run equilibrium mean, which — when not modelled appropriately — induced substantial forecast errors: see fig. 27.

The forecast errors depicted there are from a 4-variable system of money, income, inflation and the interest rate. The model omits the own rate of interest on M1 following the 1984 legislative change. Despite in-sample congruency, well-determined estimates, and theoretically-supported cointegration in an equation for UK M1 that had remained constant for almost a decade, the forecasts are for systematic falls in real M1 during the most rapid rise that has occurred historically. Almost all the forecasthorizon data lie outside the ex ante 1-step 95% prediction intervals for UK M1. Such an outcome is far from 'error correction', prompting the renaming of cointegration combinations to equilibrium correction. By way of comparison, figure 10 shows the combined 1-step and multi-step forecasts for a DV model. By eliminating the equilibrium-correction terms, the DV suffers from residual autocorrelation ($\mathsf{F}_{\mathsf{ar}}(80, 231) = 1.51^{**}$), and its confidence intervals calculated by the usual formulae are incorrect, probably overstating the actual uncertainty. Nevertheless, the absence of bias in the forecasts compared to those from the VEqCM is striking.



Consequently, VEqCMs will be reliable in forecasting only if they contain all the variables needed to track changed states of nature—here the VEqCMfails because it omits the change in the interest rate variable. However, the graph of the DV model forecasts of UK M1 suggest such models may be more robust to equilibrium mean shifts, and this is in fact a general result, as we discuss in the next section.

7 DETERMINISTIC SHIFTS AND DIFFERENCING

Such forecasting models neglect the long-run relations that tie variables together, and indeed need have no 'causal' basis at all. But within the framework within which we are working, as set out in section 2, there can be no presumption that causal models should outperform non-causal. In fact, we are able to establish analytically the effectiveness of differencing in producing unbiased forecasts when the mean shift has already occurred. However, the forecast-error variances of DV and DDVs exceed those from the VEqCM, especially at longer horizons, so on MSFE the VEqCM might be favoured. Eitrheim et al (1997) compared the DV and DDV models to the Norges–Bank 'VEqCM' model. Over a 12-quarter evaluation horizon, the Bank's model performed best, and the equivalent of the DDV did worst, whereas the DDV did best in each of three 4-period divisions of the same data. This outcome is consistent with our analysis. Shorter forecast horizons may penalise the DV and DDV models less on variance, and the more sequences of forecasts beginning from separate origins, in a given period of time, the more likely that some of those origins will fall after breaks, allowing the differenced models to exploit their greater robustness to equilibrium mean shifts.

This facility of such models can perhaps best be understood by considering a simple scalar process.
Suppose a variable y_t follows:

$$y_t = \mu + \delta 1_{\{t>T_1\}} + \epsilon_t \qquad (36)$$

where $\epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2]$. In (36), $1_{\{t>T_1\}}$ is an indicator with the value zero till time $T_1 < T$, after which it is unity. This allows the intercept to take on two values: μ when $t \leq T_1$, and $\mu + \delta$ when $t > T_1$. The forecasting model which predicts the pre-break mean of the process is the equivalent of the VEqCM. The 'DV' model eliminates the mean, and is here the driftless random walk $\Delta y_t = u_t$, (u_t hypothesised to be $\mathsf{IN}[0, \sigma_u^2]$) with h-step ahead forecasts from an origin T denoted by $\widetilde{y}_{T+h} = y_T$ for all h.

Consider the error in a 1-step forecast made at the time of the break using the random walk predictor:

$$y_{T_1+1} - \widetilde{y}_{T_1+1} = \epsilon_{T_1+1} - \epsilon_{T_1} + \delta$$

which has a bias of δ and a forecast variance of $2\sigma_\epsilon^2$. Using the pre-break mean the bias is the same but the variance is halved. But now move the forecast origin forward by one period, and repeat the exercise.

The random walk model now corrects to the changed mean of the process:

$$y_{T_1+2} - \widetilde{y}_{T_1+2} = \epsilon_{T_1+2} - \epsilon_{T_1+1}$$

so the forecasts are now unbiased, and the variance is still $2\sigma_{\epsilon}^2$. The 'correct' model pre-break continues to have the same variance and bias as before, so that for large enough δ, eliminating the mean using the random walk predictor delivers the smaller MSFE. The contrast between the pre-break mean and the random walk predictors is analogous to that between the VEqCM and the DV models, when the VEqCM is correctly-specified but for the equilibrium mean shift that occurs prior to forecasting. Intercept corrections can have similar effects to differencing in the face of mean shifts, as the next section now illustrates.

8 DETERMINISTIC SHIFTS AND INTERCEPT CORRECTIONS

Published macroeconomic forecasts are rarely purely model-based, and adjustments are often made to arrive at a final forecast. These adjustments can be rationalised in a variety of ways, but here we focus on their role in offsetting regime shifts. The example below is a slight generalisation of that in the previous section, whereby the DGP is autoregressive:

$$y_t = \mu + \delta 1_{\{t \geq T_1\}} + \rho y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{37}$$

$|\rho| < 1$, but the mean shift is again assumed unknown to the investigator. Now, for $T > T_1$, $\mathsf{E}[y_{T+1}] = (\mu + \delta)/(1 - \rho)$ and:

$$\mathsf{E}\left[y_{T+1} \mid y_T\right] = \mu + \delta + \rho y_T$$

The model is mis-specified in that the investigator uses:

$$\widehat{y}_{T+1} = \widehat{\mu} = T^{-1} \sum_{t=1}^{T} y_t$$

to forecast. Since:

$$\mathsf{E}\left[\widehat{\mu}\right] = T^{-1}\left(\sum_{t=1}^{T_1} \mathsf{E}\left[y_t\right] + \sum_{t=T_1+1}^{T} \mathsf{E}\left[y_t\right]\right) = \frac{\mu + \delta}{1 - \rho} - \kappa \frac{\delta}{1 - \rho} \tag{38}$$

where $\kappa = T^{-1} T_1$, then unconditionally the bias is given by the second term in (38).

At T, there was a residual of $\widehat{u}_T = y_T - \widehat{\mu}$, so to set the model 'back on track' (i.e., fit the last observation perfectly), the intercept correction buT is often added to the forecast to yield:

$$\widehat{y}_{i,T+1} = \widehat{\mu} + \widehat{u}_T = y_T \tag{39}$$

Thus, the IC forecast changes the forecasting model to a random walk, thereby losing all the information about the equilibrium mean. However, one consequence is that:

$$\mathsf{E}\left[\widehat{y}_{\iota,T+1}\right] = \mathsf{E}\left[y_T\right] = \frac{\mu+\delta}{1-\rho}$$

which is unconditionally unbiased, despite the mis-specification of the model for the DGP. Further:

$$y_{T+1} - \widehat{y}_{\iota,T+1} = \mu + \delta + (\rho - 1)\, y_T + \epsilon_{T+1} = (\rho - 1)\left(y_T - \frac{\mu+\delta}{1-\rho}\right) + \epsilon_{T+1}$$

so that the unconditional MSFE is:

$$\mathsf{E}\left[\left(y_{T+1} - \widehat{y}_{\iota,T+1}\right)^2\right] = \sigma_\epsilon^2 + (1-\rho)^2\,\mathsf{V}\left[y_T\right] = \frac{2\sigma_\epsilon^2}{1+\rho} \quad (40)$$

as against the minimum obtainable (for known parameters) of σ_ϵ^2. Clearly, such ICs as $\widehat{u}_T$ have excellent properties in this setting.

Intercept corrections can be shown to work similarly in more complex settings, and in particular, for VEqCM models of VEqCM DGPs in the face of shifts in both equilibrium means and underlying growth rates. Forecast-error bias reductions are generally bought at the cost of less precise forecasts, and their efficacy depends on the size of the deterministic shift relative to the horizon to be forecast. Figure 31 illustrates for the UK money demand example. The form of correction employed there makes the same adjustment to all forecast origins based on an average of the two errors at the beginning of the period. The correction is only applied to the money equation, where it shifts upward the forecasts of $\Delta(m-p)$,and partially corrects the under-prediction.

9 LESS IMPORTANT SOURCES OF FORECAST FAILURE

Other factors besides unmodelled parameter shifts are signalled by the forecast-error taxonomy as potential sources of forecast failure. These include: model mis-specification; parameter estimation uncertainty, possibly induced by collinearity, a lack of parsimony and model selection; and forecast origin mis-measurement. While these may exacerbate the effects of non-constancies in deterministic factors, of themselves they seem unlikely to constitute primary causes of forecast failure. To draw on an analogy from Kuhn (1962), all these aspects may matter in 'normal forecasting', and contribute to a worse forecast performance, but shifts in deterministic terms are dramatically more important during 'forecasting d´eb^acles'.

9.1 Model Mis-Specification

Model mis-specification *per se* cannot account for forecast failure. In the absence of changed economic conditions (required for the '*per se*' part of the statement), a model's out-of-sample forecast performance will be as expected based on its in-sample fit to the data. If model mis-specification results in serially autocorrelated errors,tests of parameter constancy which ignore this may be misleading but this is a secondary consideration.

9.2 Estimation Uncertainty

Estimation uncertainty seems unlikely to be a source of forecast failure by itself, since the in-sample and out-of-sample fits will be similar in the absence of any changes in the underlying process. First, we illustrate by an empirical example the impact on computed prediction intervals of adding parameter variances to the variances arising from the innovation errors. The four dimensional system is again the UK money demand system. We consider multi-step forecasts over 1985(3)–1986(2) from an estimation sample of 1978(3)–1985(2) for the levels of the variables. There are two models, a relatively unrestricted VAR, and a VEqCM, the latter developed in Hendry and Doornik (1994) to offer a parsimonious, congruent representation of the data. Figures 11 and 12 show the resulting forecasts with the bars based on the innovation errors only, and bands showing the overall 95% prediction intervals once parameter estimation variances are included.There are distinct differences between the bands and bars for the VAR (which has 12 parameters in every equation for T = 28), but virtually none in the VEqCM. The models in which we wish to explain forecast failure in practise more closely resemble the VEqCM than the VAR, so parameter estimation uncertainty seems unlikely to be the answer. Moreover, increasing the estimation sample by extending it back to 1964(3) noticeably reduces the bands for the VAR.

9.3 Collinearity

'Collinearity' can nott of itself account for forecast failure, although interacting with a break in an exogenous variable it might. Consider the case of forecasting UK house prices over 1972–75. As the first column of fig. 13 confirms, very poor 1-step forecasts result when the model is estimated up to 1972 and then used to predict. The forecast-period residuals are large and all of the same sign, indicating dramatic forecast failure when viewed alongside the in-sample residuals.Such an outcome is consistent with an equilibrium shift in the equation, or with the model being mis-specified in terms of omitting a relevant variable that changes significantly during the 1972–75 period.However,the 'forecast failure'in this instance appears to have been induced by the many changes in financial markets following the Competition and Credit Control regulations introduced in late 1971,which altered collinearity between the many regressors in the econometric model(taken from Hendry(1984):his equation(17),but replacing the term $(\Delta p_{h,t-1})^3$ by $\Delta p_{h,t-1}$)

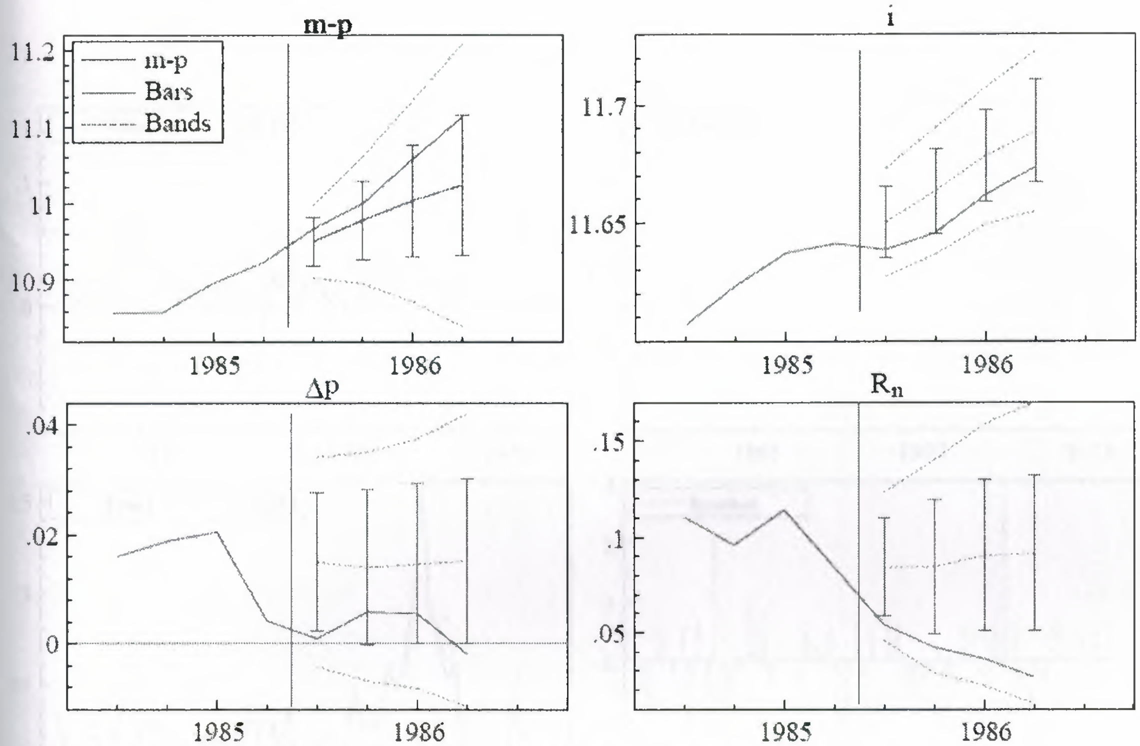


Figure 11 Forecasts and 95% prediction bars and bands for the monetary VAR.

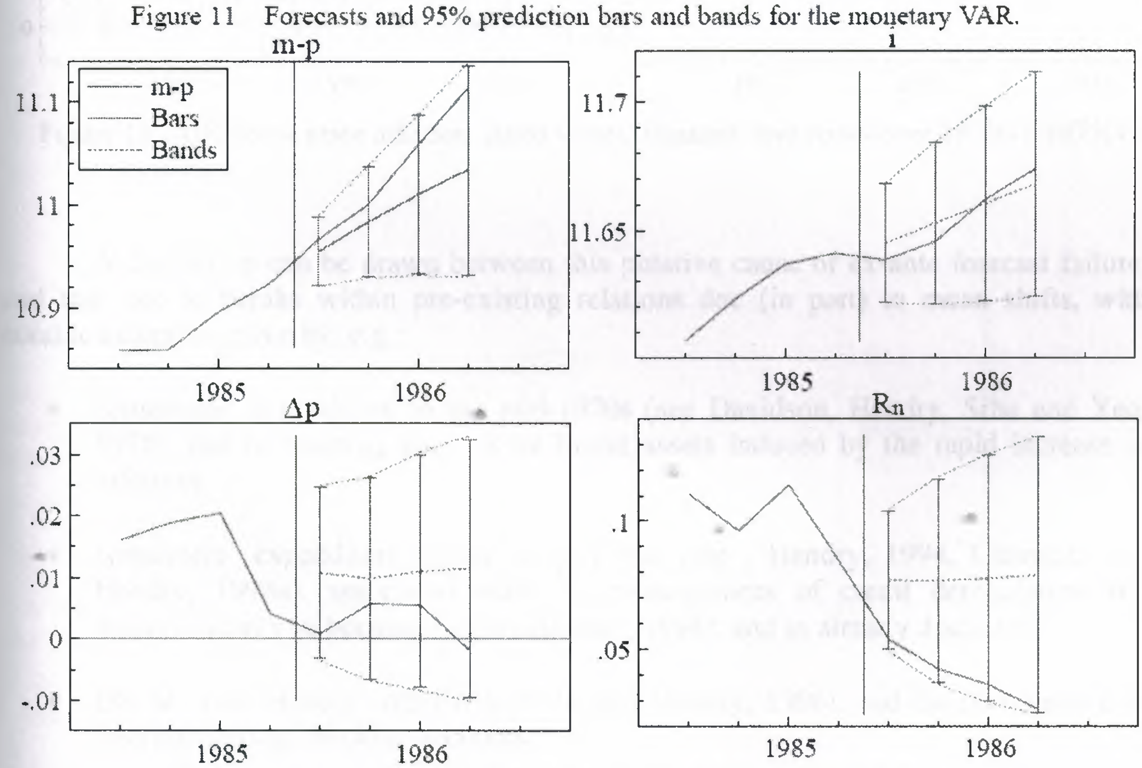


Figure 12 Forecasts and 95% prediction bars and bands for the monetary model.

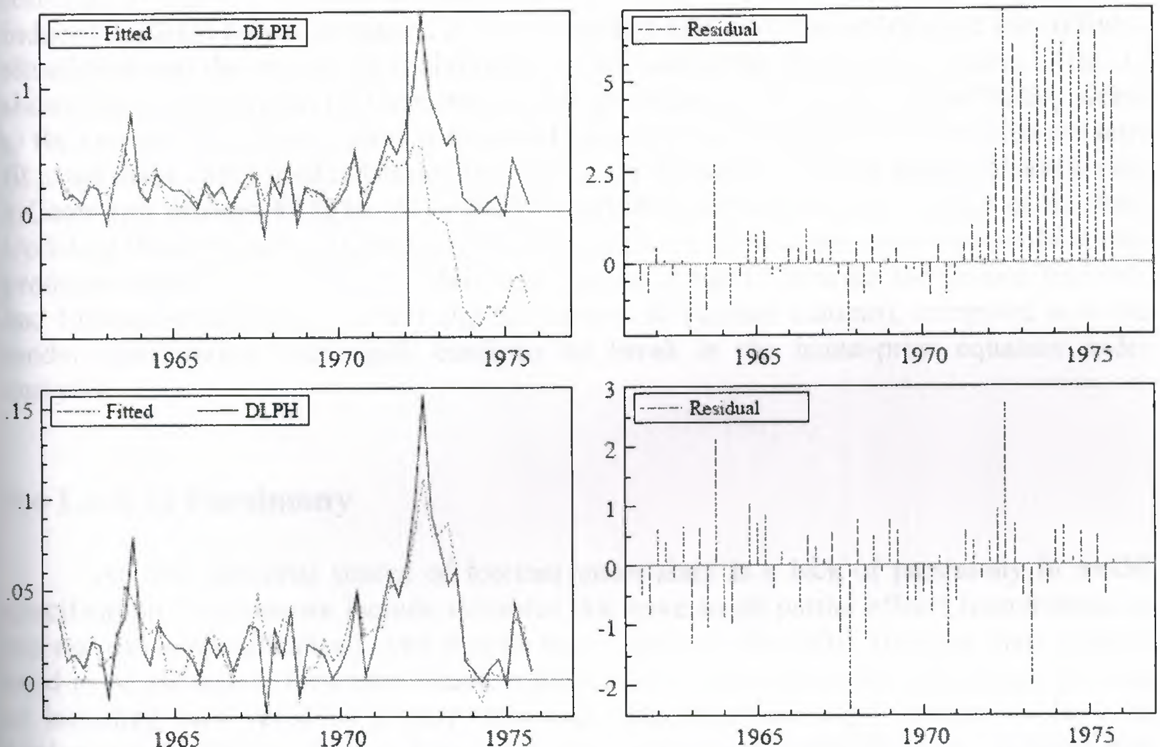


Figure 13 UK house price inflation: fitted values, forecasts and errors over 1972(1)–1975(4).

A distinction can be drawn between this putative cause of ex-ante forecast failure, and that due to breaks within pre-existing relations due (in part) to mean shifts, with notable examples given by, e.g.:

- consumers' expenditure in the mid-1970s (see Davidson, Hendry, Srba and Yeo, 1978), due to omitting the loss on liquid assets induced by the rapid increase in inflation;

- consumers' expenditure in the early 1990s (see , Hendry, 1994, Clements and Hendry, 1998a), associated with the consequences of credit deregulation and negative equity in housing (seeMuellbauer, 1994); and as already discussed,

- UK M1 (see Hendry andMizon, 1993, and Hendry, 1996), and the introduction of interest-bearing checking accounts.

How can one differentiate those outcomes (which deliver similar graphs for forecast residuals to fig. 13) from the present case where we believe a break in the regressor set induced apparent failure in what is in fact a constant relation? The hallmark of the collinear situation is that the ex-post fit is similar to the in-sample fit. The second column in fig. 13 shows the ex-post results on the same scaling: the equation does not fit significantly worse to the ex-post data. Panel c plots both ex-ante and ex-post residuals to confirm that the later fit is not at the expense of a deterioration earlier on. Formally, over the last 16 observations, a Chow test delivers $FCh(16, 28) = 1.75$, $p = 0{:}094$, despite the large change in the data. Updating the consumers' expenditure, or M1 equations, without the required modifications, produces significantly worse fits. Here, the contrast in fig. 13 between the ex-ante forecasts and forecast errors (first column) and the ex-post fit (second column), computed with the model specification unchanged, confirms no break in the house-price equation under analysis.

9.4 Lack of Parsimony

Another potential source of forecast uncertainty is a lack of parsimony in model specification. Suppose we include variables that have small partial effects (conditional on the remaining specification) even though they appear in the DGP. Because their impacts need to be estimated, their elimination could improve forecast accuracy. But then the cost of including such variables is only somewhat inaccurate forecasts – we do not have an explanation for forecast failure. Forecast failure could result if variables were included that changed substantially in the forecast period, again pointing to the key role of parameter non-constancies.

9.5 Overfitting

'Overfitting' is close to 'lack of parsimony', and following Todd, is fitting 'not only the most salient features of the historical data, which are often the stable, enduring relationships' but also 'features which often reflect merely accidental or random relationships that will not recur' Todd (1990, p.217). The latter is called sample dependence in Hendry (1995b).

It is useful to distinguish two cases: simply having a 'generously parameterized model', and using the in-sample data evidence to select the variables to be retained. The first is a transient problem in a progressive research strategy, in that an extended data sample will reveal the accidental nature of the irrelevant effects by their becoming less significant. Moreover, although the resulting forecasts may be inaccurate, as argued in §15.4, systematic forecast failure will only occur if the data properties of the incorrectly included variables change during the forecast period.

The second, commencing with an 'over-parameterized representation', and then using general-tosimple modelling, need not lead to overfitting: simplification could either attenuate or exacerbate sample dependence. The former could occur if genuinely irrelevant factors were eliminated, whereas the latter could happen if the influences of accidental aspects were captured more 'significantly' by being retained in a smaller parameterization (also a transient problem). The Monte Carlo results in Hoover and Perez (1999) suggest that general-to-simple procedures, extended as they suggest, often deliver a final equation that is close to the one which generated their data, supporting their efficacy in achieving the former.

10 THE DETECTABILITY OF BREAKS IN VARs

To highlight the different extents to which various possible breaks affect forecast failure, we report a Monte Carlo simulation of a bivariate cointegrated I(1) VAR subject to shifts: the experiment is implemented in Ox by PcNaive for Windows (see Doornik, 1999, and Doornik and Hendry, 1998). The data are generated by:

$$\begin{aligned} \Delta x_{1,t} &= \gamma_1 + \alpha_1 (x_{1,t-1} - x_{2,t-1} - \mu_1) + \epsilon_{1,t} \\ \Delta x_{2,t} &= \gamma_2 + \alpha_2 (x_{1,t-1} - x_{2,t-1} - \mu_1) + \epsilon_{2,t} \end{aligned} \tag{41}$$

where $\epsilon_{i,t} \sim \mathsf{IN}[0, \sigma_{ii}]$, with $\mathsf{E}[\epsilon_{1,t}\epsilon_{2,s}] = 0 \ \forall t$; s, so in (41), **K** is diagonal with elements $\sqrt{\sigma_{ii}}$. We consider 4 types of experiment:

(A) a constant DGP (to establish test size);
(B) breaks in the coefficients of the feedbacks (α_1 and α_2);
(C) breaks in the long-run mean (μ_1);
(D) breaks in the growth rates (γ_1 and γ_2).

Four different full-sample sizes are considered: T = 24, 60, 100, and 200 (denoted a, b, c, d on graphs), the last of which is relatively large for macro-economic models. Breaks occur at t = 0.5T, and revert to the original parameter values at t = 0.75T to mimic a second break. The design seeks to show the effects of more information at larger T with fixed relative break points: as will be seen, 'undetectable' breaks remain hard to find even at the largest sample size considered here. Since breaks are involved, these experiments could not have been conducted as one recursive experiment from T = 10 to T = 200, except when studying null rejection frequencies. However, for graphical presentation, the individual graph lines between a, b, c, d are only distinguished by a different symbol where that clarifies an important feature, even when the recorded outcomes overlap.

The unrestricted VAR with intercept and one lag is estimated and tested. When breaks occur, modelling cointegrated processes is difficult, and as a VAR is usually the first step, constancy tests should be implemented at that stage (i.e., prior to checking cointegration): detectability may increase if a VeqCM is imposed, but doing so is unlikely to affect the rankings across the outcomes in our experiments A–D. Throughout, 500 replications were used, and rejection frequencies at both 0.05 and 0.01 nominal test sizes were recorded (so have standard errors of about 0.01 and 0.004 respectively).

The experimental formulation is invariant to what induces changes in $\gamma_i - \alpha_i\mu_1$, but large growthrate changes seem unlikely for real economic variables. It may seem surprising that the 'absolute' size of μ_1 can matter, since even after log transforms, the measurement units affect μ_1: for example, a switch from a price index normalized at unity to one normalized at 100 radically alters μ_1 in (say) a log-linear money-demand model without affecting either α or σ. Nevertheless, changes in μ_1 (relative to error standard deviations) also need to be judged absolutely, not as percentages: thus, using ∇ to denote parameter changes, $\nabla\mu_1/\sigma$ matters per se, and this cannot depend on the measurement system, only on agents' behaviour. When μ_1/σ is large (small), a given effect will be a small (large) percent, but will have the same detectability for a given α. For example, for both broad and narrow money demand in the UK after financial innovation in the mid 1980s (see Ericsson, Hendry and Prestwich, 1998, and Hendry and Ericsson, 1991), $\nabla\mu_1/\sigma \simeq$ 25–30, in models that excluded appropriate modifiers. The rise in the savings rate in the mid 1970s was of roughly the same absolute magnitude (see e.g., Hendry, 1994). For 'standard' values of α, (around 0.1–0.2) these numbers translate into 'permanent' equilibrium shifts of 2.5σ to 6σ. Such considerations determined the values of the parameters in the experimental design.

Two baseline sets of dynamics are considered $\alpha_1 = -0.1$, and $\alpha_2 = 0$ (so $x_{2,t}$ is both weakly and strongly exogenous for the parameters of the first equation: see Engle, Hendry and Richard, 1983); and α_1 = −0:1, and α_1 = 0:1 (so $x_{2,t}$ is neither weakly nor strongly exogenous for the parameters of the first equation). For α_1, the change is −0.05. We investigate μ_1 = 1, changed by an amount of +0.3 for its break (so $\nabla\alpha_1\mu_1 = 3\sigma_{11}$). Also, using $\beta' = (1 : -1)$ enforces γ_1 = γ_2 which was set to 0.01 (roughly 4% p.a. for quarterly data): the change considered is to double both of these to 0.02 in (D), which would constitute a very dramatic increase in long-term growth. Thus, two ratios of $\gamma_i/\sqrt{\sigma_{ii}}$ are examined, namely unity and 2 (see Hylleberg and Mizon, 1989), inducing the derived values of $\sqrt{\sigma_{ii}}$ = 0.01 throughout (roughly 1% for the 1-step ahead forecast standard error under constant parameters). Notice that γ and μ correspond to elements of φ rather than ϕ. In total, there are 8 baseline experiments, and 4 changes to α, γ , and μ making 32 experiments in total. These are reported graphically, each panel showing all four sample size outcomes for both p values. The critical values for the constancy tests are those for a known break point, which delivers the highest possible power for the test used. The graphs serve to illustrate the outcomes visually, showing that rejection frequencies are everywhere low in some cases, confirming that the highest power is immediately before the first break, whereas the second break is often less detectable when the first has not been modelled, and sometimes showing that the tests are actually biased after the second break.

10.1 Test Size

The relation of the actual to the nominal size for vector constancy tests has not been much investigated, so the experiments in (A) check their size in an I(1), cointegrated setting, with and without feedback in the second equation. As fig. 14 reveals, the results are reasonably reassuring when the EqCM enters both relations: with 500 replications, the approximate 95% confidence intervals are (0.03, 0.07) and (0.002, 0.018) for 5% and 1% nominal, and these are shown on the graphs as dotted and dashed lines respectively, revealing that few null rejection frequencies lie outside those bounds once the sample size exceeds 60. At the smallest sample sizes, there is some over-rejection, though never above 9% for the 5% nominal or 3% for the 1% nominal. When the EqCM enters the first relation only, there is a systematic, but small, excess rejection: around 6% instead of 5%, and 1.5% instead of 1%. However, these outcomes are not sufficiently discrepant to markedly affect the outcomes of the 'power' comparisons below.

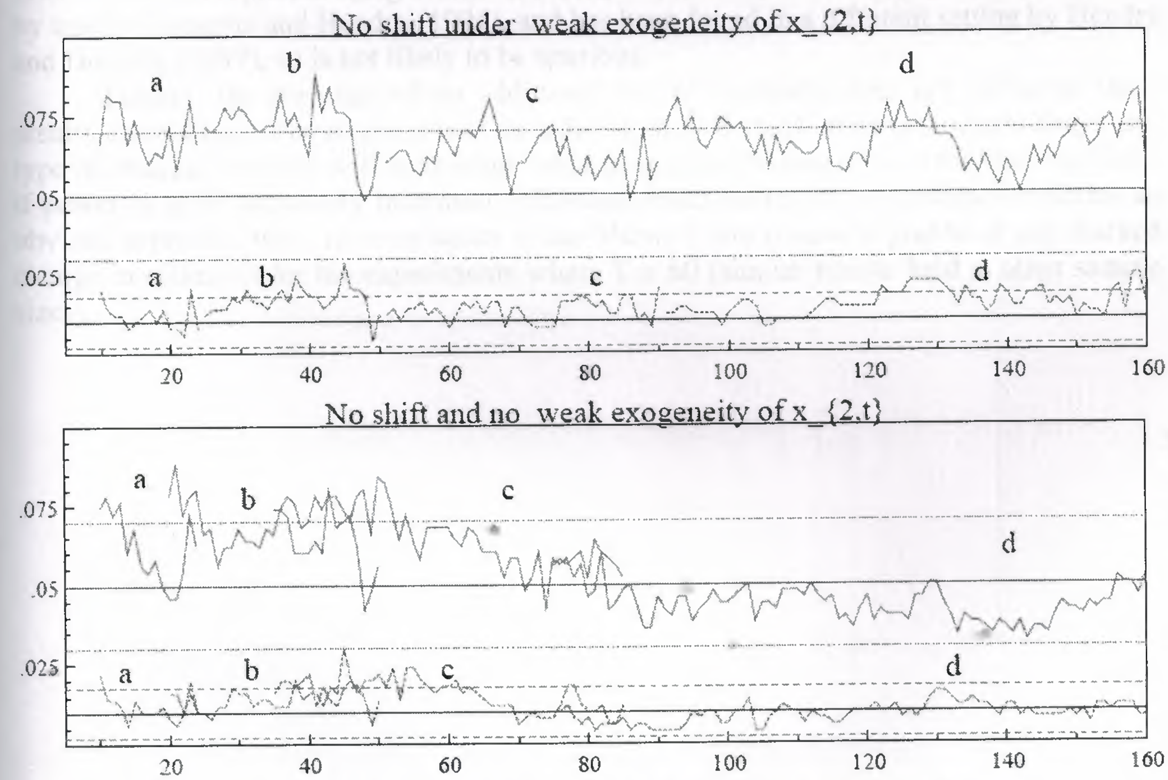


Figure 14 Constancy-test null rejection frequencies in a cointegrated process.

10.2 Dynamic Shift

Experiments in (B) demonstrate that a change in the strength of reaction to a zero-mean disequilibrium is not readily detectable. This is despite the fact that the intercept also shifts in the VAR representation:

$$\begin{aligned} \mathbf{x}_t &= \gamma - \alpha\mu + \alpha\beta'\mathbf{x}_{t-1} + \epsilon_t && \text{pre break} \\ \mathbf{x}_s &= \gamma - \alpha^*\mu + \alpha^*\beta'\mathbf{x}_{s-1} + \epsilon_s && \text{post break.} \end{aligned}$$

One might have anticipated detectability from $\nabla\alpha\mu \neq 0$, particularly since that change numerically exceeds the equivalent jump in (C) which we show below is easily detected. Nevertheless, despite the induced intercept shift, changes in the dynamics alone are not easily detectable. Figure 15 records the outcomes: the powers are so low, do not increase with sample size, and indeed barely reflect any breaks, that one might question whether the Monte Carlo was correctly implemented: be assured it was, but anyway, this effect is easy to replicate using PcNaive. Moreover, it was predicted by the analysis above, by that in Clements and Hendry (1994), and has been found in a different setting by Hendry and Doornik (1997), so is not likely to be spurious.

Further, the presence of an additional EqCM feedback does not influence these results, even though one might expect an induced shift. Perhaps other tests could detect this type of change, but they will need some other principle of construction than that used here if power is to be noticeably increased. Although direct testing of the parameters seems an obvious approach, there is no evidence in the Monte Carlo recursive graphs of any marked change in estimates for the experiments where T = 60 (similar results held at other sample sizes).

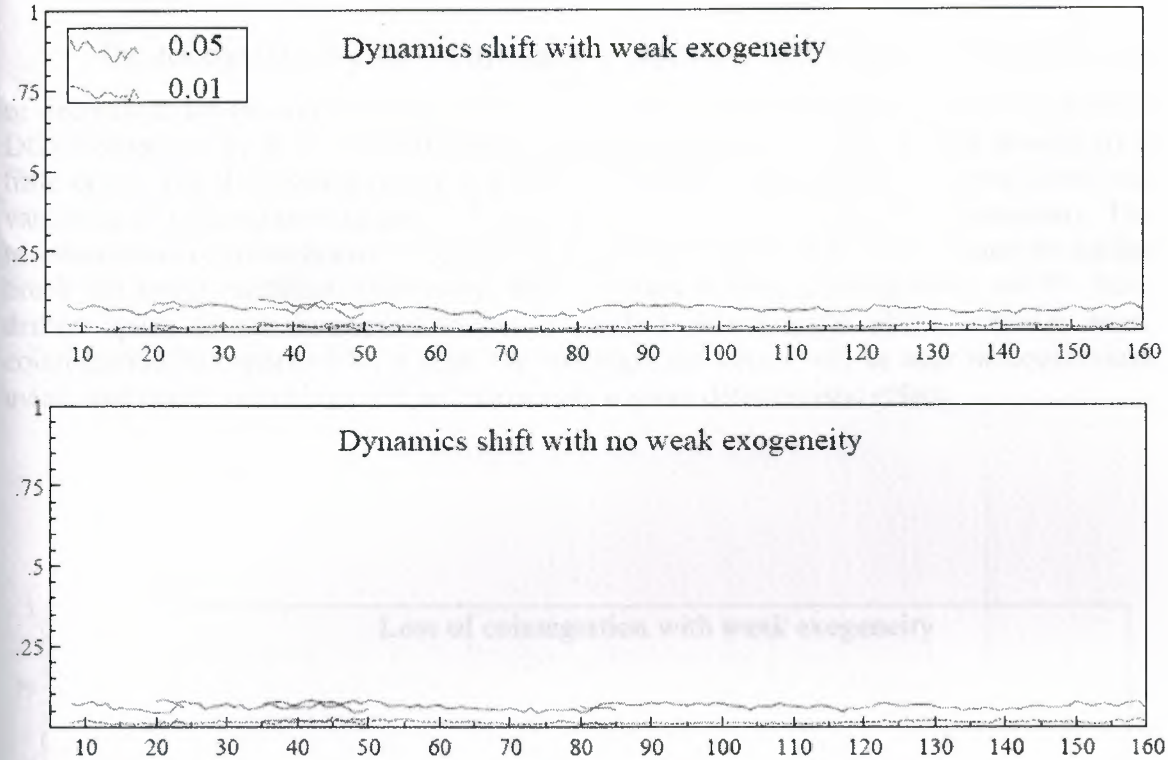


Figure 15 Constancy-test rejection frequencies for changes in α.

The remarkable feature of this set of experiments is that most of the parameters in the system have been changed, namely from:

$$\mathbf{\Gamma} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}; \ \boldsymbol{\tau} = \begin{pmatrix} 0.11 \\ -0.09 \end{pmatrix} \tag{42}$$

to:

$$\mathbf{\Gamma}^* = \begin{pmatrix} 0.85 & 0.15 \\ 0.1 & 0.9 \end{pmatrix}; \ \boldsymbol{\tau}^* = \begin{pmatrix} 0.16 \\ -0.09 \end{pmatrix} \tag{43}$$

yet the data hardly alter. Moreover, increasing the size of the shift in α_1, and indeed making both αs shift, does not improve the detectability: for example, using $\alpha_1^* = -0.2$, and $\alpha_2^* = -1.5$ causes little perceptible increase in the rejection frequency, or movement in the recursive estimates. This remains true even when a VEqCM is used, with known cointegrating vector.

The detectability of a shift in dynamics is dependent on whether the _s are increased or decreased: for example, setting $\alpha_1^* = \alpha_2^* = 0$, so that cointegration vanishes and the DGP becomes a VAR in first differences, delivers the graphs in fig. 16. The powers are a little better: but the highest power is still less than 30% even though in some sense, the vanishing of cointegration might be deemed a major structural change to an economy. The re-instatement of cointegration is somewhat more detectable than its loss, despite the earlier break not being modelled: intuitively, after a period without cointegration, the xs have drifted apart, so the re-imposition has a marked deterministic effect, whereas when cointegration has operated for a time, the cointegration vector will be near its equilibrium mean, and hence switching it off will have only a small deterministic effect.

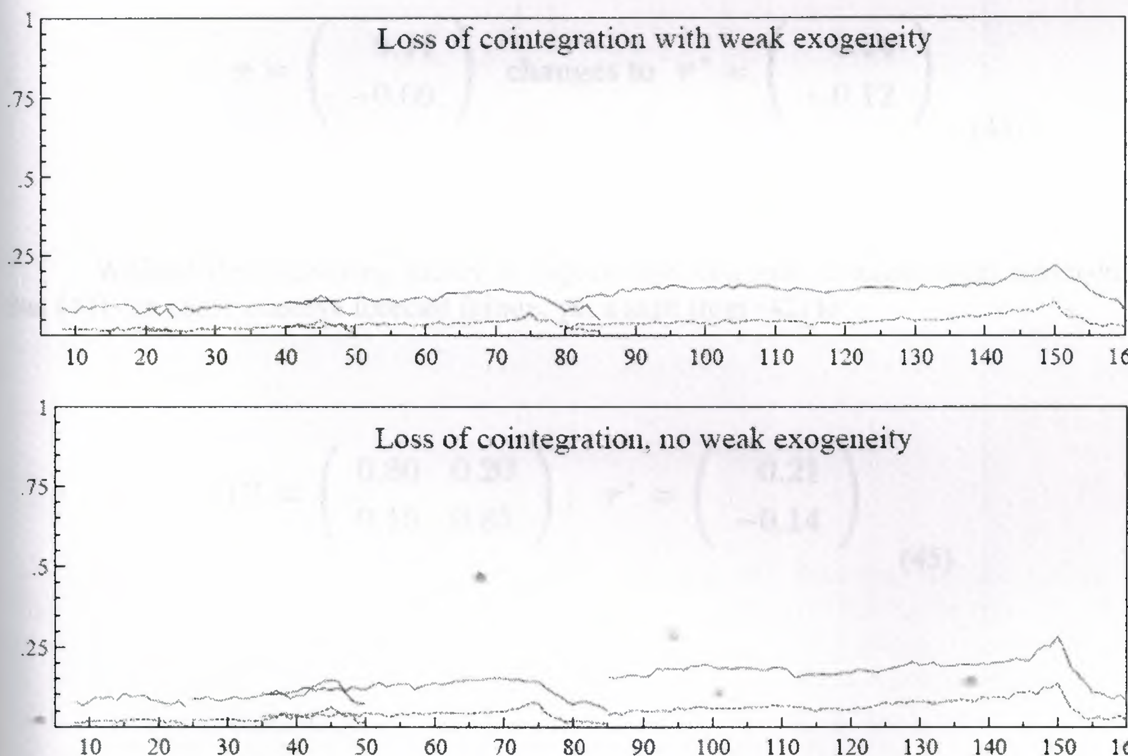


Figure 16 Constancy-test rejection frequencies for loss of cointegration.

10.3 Equilibrium-Mean Shift

Experiments in (C) show the contrasting ease of detecting equilibrium-mean breaks. Figure 17 confirms the anticipated outcomes for the mean shifts: the break-point test rejection frequencies were close to their nominal size of 5% under the null, but the break in the equilibrium mean is easily detected, even at quite small sample sizes, especially when the EqCM enters both relations, which also serves to sharpen the location of the break. Because the relative positions of the breaks are held fixed, the power only increases slowly at p = 0.05 as T grows, but has a more marked effect for p = 0.01: but larger samples do not ensure higher powers for detecting breaks.

To emphasize the detectability issue, note that the corresponding VAR here has the same $\boldsymbol{\Gamma}$ as (42) throughout, and:

$$\boldsymbol{\tau} = \begin{pmatrix} 0.11 \\ -0.09 \end{pmatrix} \text{ changes to } \boldsymbol{\tau}^* = \begin{pmatrix} 0.14 \\ -0.12 \end{pmatrix} \tag{44}$$

Without the underlying theory to explain this outcome, it might seem astonishing that (44) can cause massive forecast failure, yet a shift from (42) to:

$$\boldsymbol{\Gamma}^* = \begin{pmatrix} 0.80 & 0.20 \\ 0.15 & 0.85 \end{pmatrix}; \quad \boldsymbol{\tau}^* = \begin{pmatrix} 0.21 \\ -0.14 \end{pmatrix} \tag{45}$$

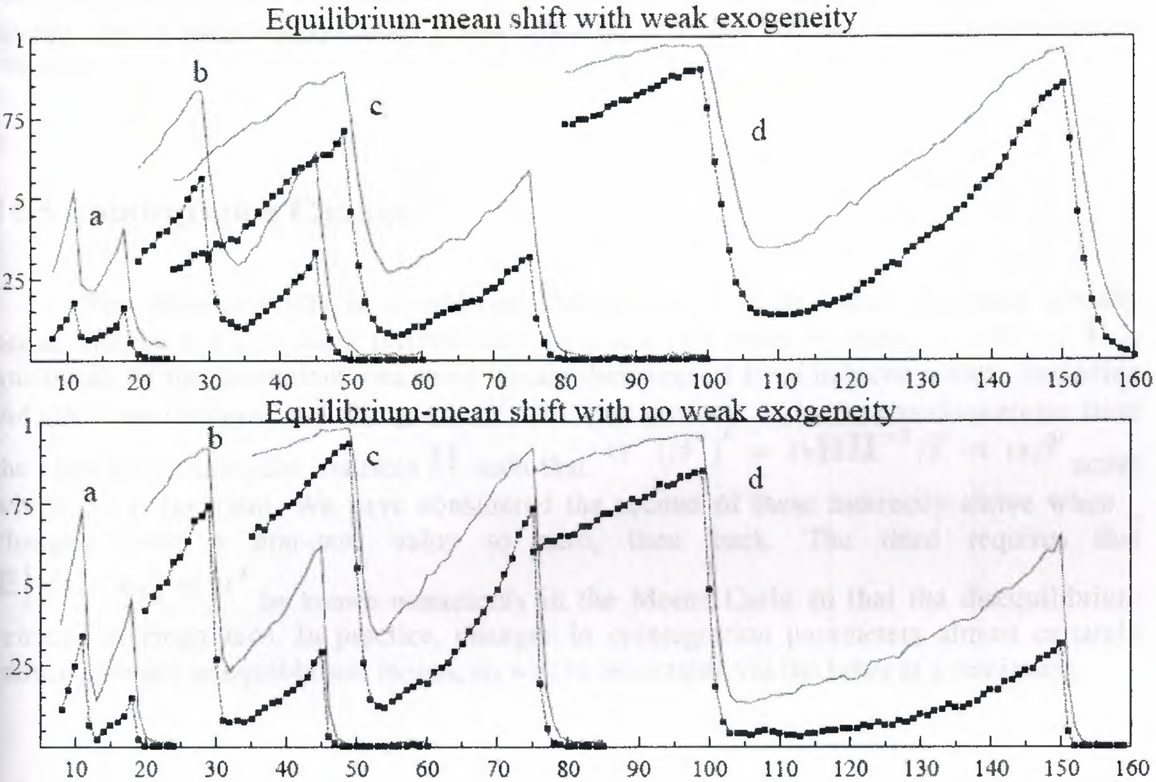


Figure 17 Constancy-test rejection frequencies for changes in μ.

is almost completely undetectable.

10.4 Growth-Rate Shift

Finally, experiments in (D) examine the ease of detecting the corresponding doubling of the growth rate. This is a large change in real growth, but that is equivalent to a fraction of the change in (C), and at about 5% rejection, is again almost undetectable on these tests at small T, but does become increasingly easy to spot as the sample size grows. This is sensible, since the data exhibit a broken trend, but the model does not, so larger samples with the same relative break points induce effects of a larger magnitude. Thus, the type of structural break to be detected affects whether larger samples will help.

The model formulation also matters: if a VEqCMis used, the first growth break shows upmuchmore strongly. The increased initial detectability is because the VEqCM 'isolates' the shift in γ, whereas the VAR 'bundles' it with $\gamma-\alpha\mu$, where the second component can be relatively much larger, camouflaging the shift. Moreover, VAR estimates of $\gamma-\alpha\mu$ generally have very large standard errors because of the I(1) representation, whereas estimates of are usually quite precise. Even so, a large growth-rate change has a surprisingly small effect even at T = 200 on the recursively-estimated intercept.

10.5 Cointegration Changes

The main difficulty in considering changes in β is to ensure that these actually occur, and are not just linear recombinations, potentially offset by changes to α so Γ is unaltered. At the same time, one must isolate their impact from induced effects, including induced rank changes, and changes to μ. The first problem is due to transformations from the class of non-singular matrices $\mathbf{H}$ such that $\alpha^* (\beta^*)' = \alpha\mathbf{H}\mathbf{H}^{-1}\beta' = \alpha\beta'$ under which Γ is invariant. We have considered the second of these indirectly above when _ changed from a non-zero value to zero, then back. The third requires that $\mathsf{E}\left[(\beta^*)' \mathbf{x}_t\right] = \mu^*$ be known numerically in the Monte Carlo so that the disequilibrium remains at mean zero. In practice, changes in cointegration parameters almost certainly induce changes in equilibrium means, so will be detectable via the latter at a minimum.

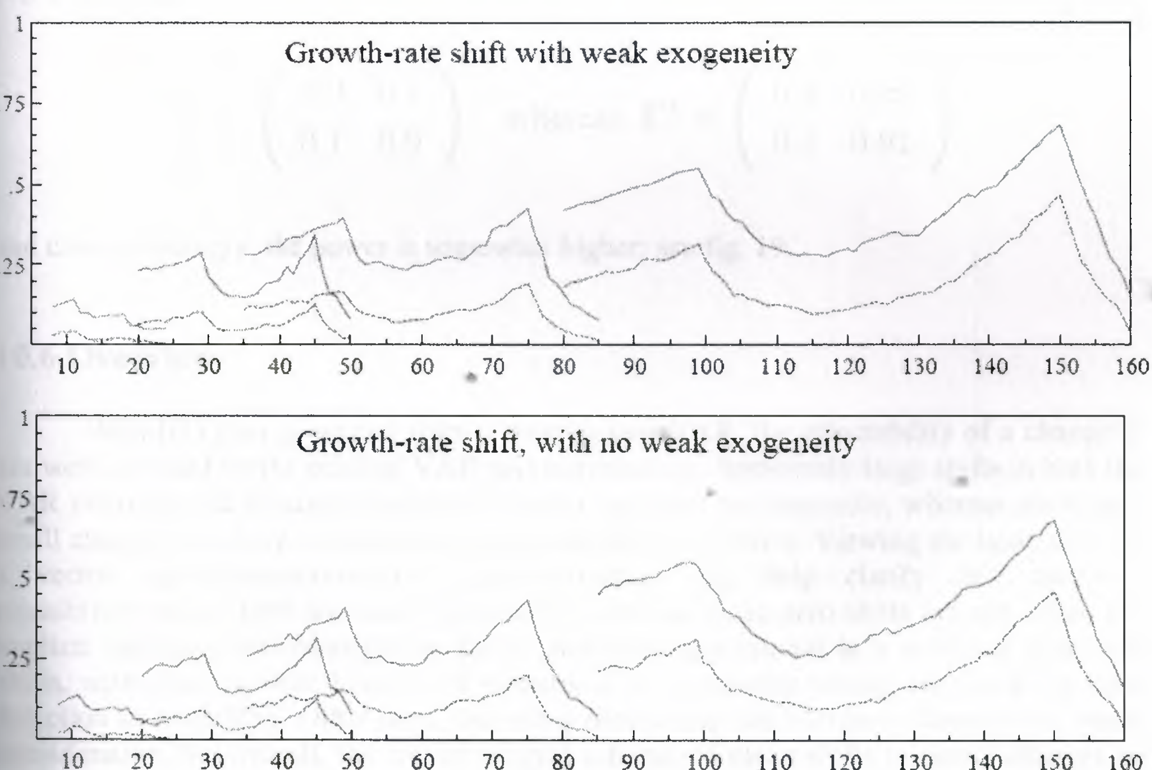


Figure 18 Constancy-test rejection frequencies for changes in γ.

The experiment to illustrate this case therefore set $\mu = \mu^* = 0$, with the other design parameters as before, and changed β' from (1 : −1) to (1 : −0.9), altering $\tau = \gamma$ to ensure $\beta'\gamma = 0$ both before and after the shift, and commencing the simulation from $\mathbf{y}'_0 = (0 : 0)$, but discarding the first 20 generated observations. Since the process drifts, power should rise quickly as the sample size grows, both because of increased evidence, and the increased data values. Moreover, the re-instatement should be more detectable than the initial change, matching when cointegration is first lost then re-appears. Conversely, the induced change to $\mathbf{\Gamma}$ is not very large when $\alpha' = (-0.1 : 0)$:

$$\mathbf{\Gamma} = \begin{pmatrix} 0.9 & 0.1 \\ 0.0 & 1.0 \end{pmatrix}; \text{ whereas } \mathbf{\Gamma}^* = \begin{pmatrix} 0.9 & 0.09 \\ 0.0 & 1.0 \end{pmatrix}$$

with appropriate changes to τ. In one sense, the very small shift in $\mathbf{\Gamma}$ is remarkably detectable, and reveals how important a role is played by shifts in off-diagonal elements in a VAR: we note with interest that the so-called 'Minnesota' prior shrinks these terms towards zero (see Doan et al., 1984). Only a slightly bigger change results when α' = (−0.1 : 0.1), as:

$$\mathbf{\Gamma} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}; \text{ whereas } \mathbf{\Gamma}^* = \begin{pmatrix} 0.9 & 0.09 \\ 0.1 & 0.91 \end{pmatrix}$$

and correspondingly, the power is somewhat higher: see fig. 19.

10.6 Overview

With I(1) data generated from a cointegrated VAR, the detectability of a change is not well reflected by the original VAR parameterization. Apparently-large shifts in both the VAR intercept and dynamic coefficient matrix need not be detectable, whereas seemingly small changes can have a substantial, and easily detected, effect. Viewing the issue through a vector equilibrium-correction parameterization can help clarify this outcome: equilibrium-mean shifts are readily detectable, whereas mean-zero shifts are not. Thus, the implicit variation-free assumptions about parameters are crucial in a world of structural shifts, with consequential benefits of robustness in forecasting versus drawbacks of non-detection in modelling. Other tests, including monitoring and variance-change tests, merit consideration, but overall, the results suggest a focus on mean shifts to detect changes of concern to economic forecasting.

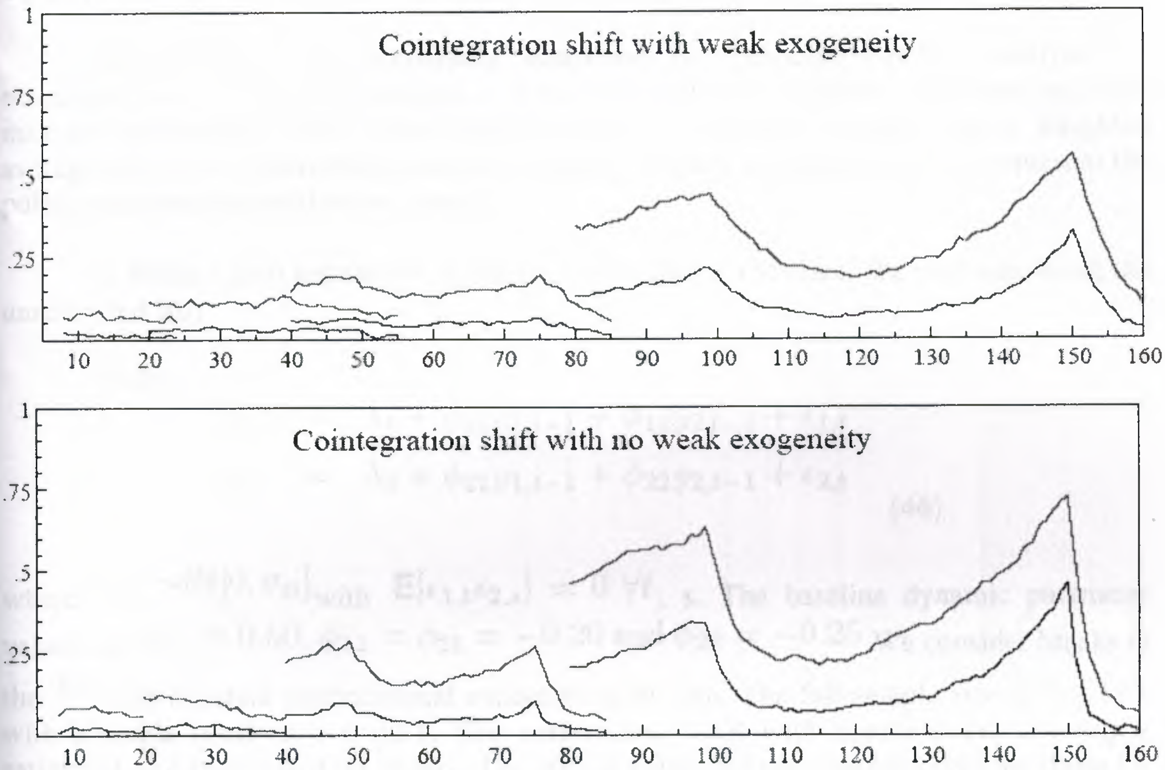


Figure 19 Rejection frequencies for a change in the cointegration parameter.

11 IMPULSE-RESPONSE ANALYSES

Such findings are potentially disastrous for 'impulse-response' analyses of economic policy. Since the changes in VAR intercepts and dynamic coefficient matrices may not be detected even when tested for, but the recorded estimates are a weighted average across the different regimes, the resulting impulse responses need not represent the policy outcomes that will in fact occur.

A Monte Carlo simulation, similar to those above, illustrates the problem, using the unrestricted I(0)

VAR:

$$\begin{aligned} y_{1,t} &= \lambda_1 + \phi_{11} y_{1,t-1} + \phi_{12} y_{2,t-1} + \epsilon_{1,t} \\ y_{2,t} &= \lambda_2 + \phi_{21} y_{1,t-1} + \phi_{22} y_{2,t-1} + \epsilon_{2,t} \end{aligned} \quad (46)$$

where $\epsilon_{i,t} \sim \mathsf{IN}[0, \sigma_{ii}]$ with $\mathsf{E}[\epsilon_{1,t}\epsilon_{2,s}] = 0 \; \forall t$, s. The baseline dynamic parameter values are $\phi_{11} = 0.50$, $\phi_{12} = \phi_{21} = -0.20$ and $\phi_{22} = -0.25$ We consider breaks in the ϕ_{ij} with constant unconditional expectations of zero. The full-sample size is T = 120, with a single break at t = 0.5T. The unrestricted VAR with intercept and one lag is estimated, and then tested for breaks. The critical values for theconstancy tests are those for a known break point, which delivers the highest possible power for the test used.

The rejection frequencies are reported graphically for both p values: 1000 replications are used, and rejection frequencies at both 0.05 and 0.01 nominal test sizes are recorded (standard errors about 0.007 and 0.003 respectively). The graphs serve to illustrate the outcomes visually, showing that rejection frequencies are everywhere low in most cases, but confirming that the highest power is immediately before the break.

11.1 Test Size

As fig. 20 reveals, the null rejection frequencies in the I(0) baseline data are reassuring: with 1000 replications, the approximate 95% confidence intervals are (0.036, 0.064) and (0.004, 0.016) for 5% and 1% nominal, and these are shown on the graphs as dotted and dashed lines respectively. The actual test sizes are close to their nominal levels.

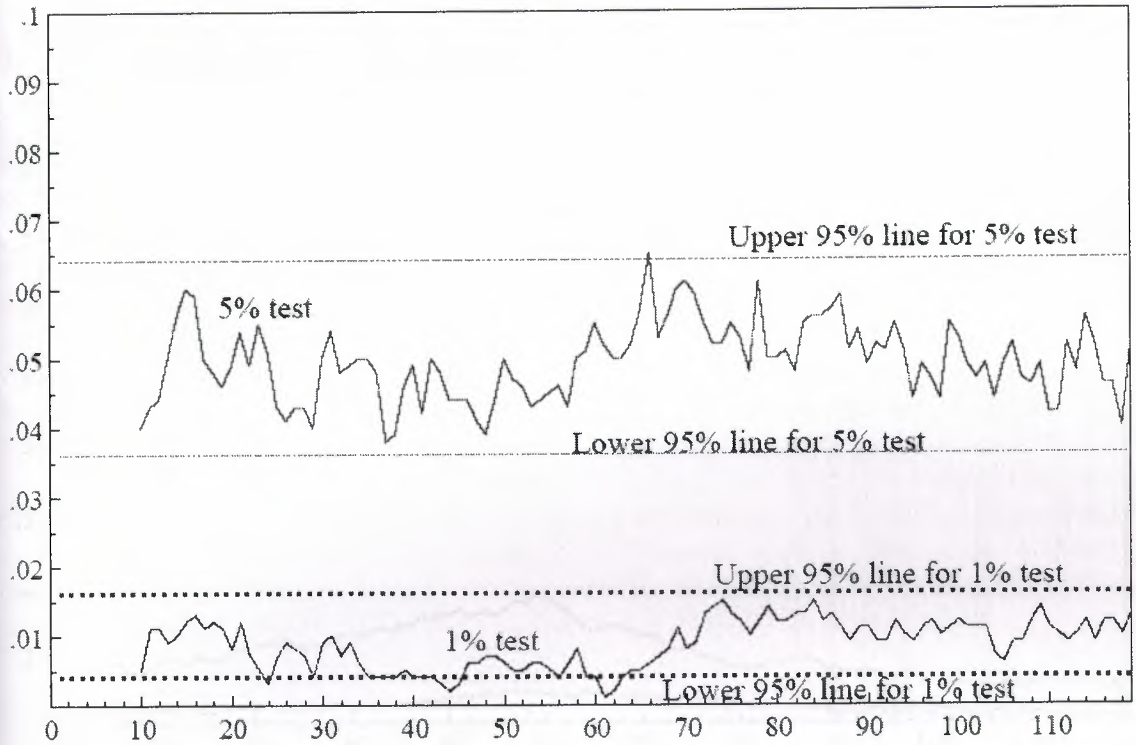


Figure 20 Constancy-test rejection frequencies for the I(0) null.

11.2 I(0) dynamic Shift

The detectability of a shift in dynamics is low when the DGP is an I(0) VAR with $\lambda = \mathbf{0}$. We consider a large parameter shift, from:

$$\Phi = \begin{pmatrix} 0.50 & -0.20 \\ -0.20 & -0.25 \end{pmatrix} \text{ to } \Phi^* = \begin{pmatrix} 0.50 & 0.20 \\ 0.20 & 0.25 \end{pmatrix} \tag{47}$$

The first element is left constant to highlight the changes in the other impulses. This break delivers the graph in fig. 21. The highest power is less than 25%, even though the change constitutes a major structural break for the model economy.

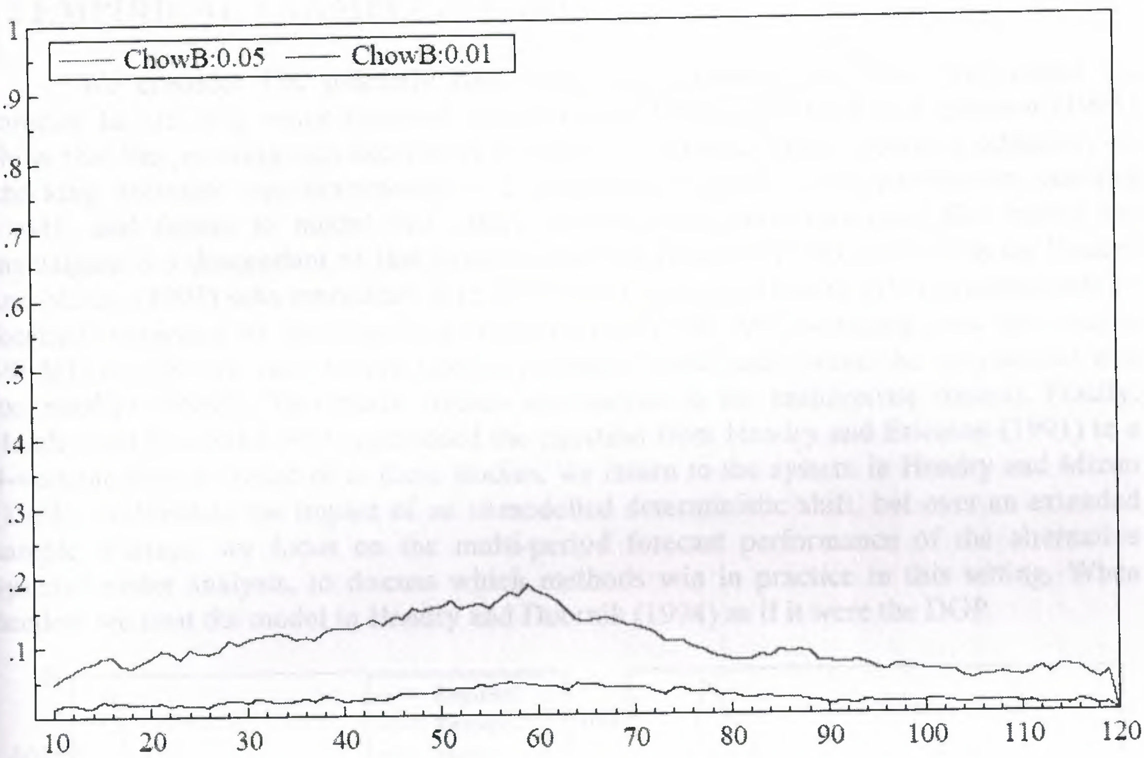


Figure 21 Constancy-test rejection frequencies for the I(0) structural break.

11.3 Impulse responses

Finally, we record the impulse responses from the pre- and post- break models, and the model fitted across the regime shifts in fig. 22. The contrast is marked: despite the near undetectability of the break, the signs of most of the impulses have altered, and those obtained from the fitted model sometimes reflect one regime, and sometimes the other. Overall, mis-leading policy advice would follow.

12 EMPIRICAL EXAMPLE: UK M1

We consider UK quarterly data (seasonally adjusted), to study the impact on forecast failure of a major financial innovation in 1984(2). Hendry and Ericsson (1991) show that the resulting introduction of non-zero own interest rates (learning adjusted) on checking accounts was tantamount to a deterministic shift in the equilibrium demand forM1, and failure to model that effect induced very poor forecasts. The model we investigate is a descendant of that first proposed in Hendry (1979), and builds on Hendry and Mizon (1993) who embedded it in a 4-variable system. Hendry (1996) considered the forecast behaviour of the single-equation model of UK M1 estimated over the sample 1963(3) to 1983(2), and showed that its forecasts failed badly when the data period was extended to 1989(2). This study extends his analysis to the multivariate context. Finally, Hendry and Doornik (1994) embedded the equation from Hendry and Ericsson (1991) in a 4-variable system. Relative to these studies, we return to the system in Hendry and Mizon (1993), to illustrate the impact of an unmodelled deterministic shift, but over an extended sample. Further, we focus on the multi-period forecast performance of the alternative systems under analysis, to discuss which methods win in practice in this setting. When needed, we treat the model in Hendry and Doornik (1994) as if it were the DGP.

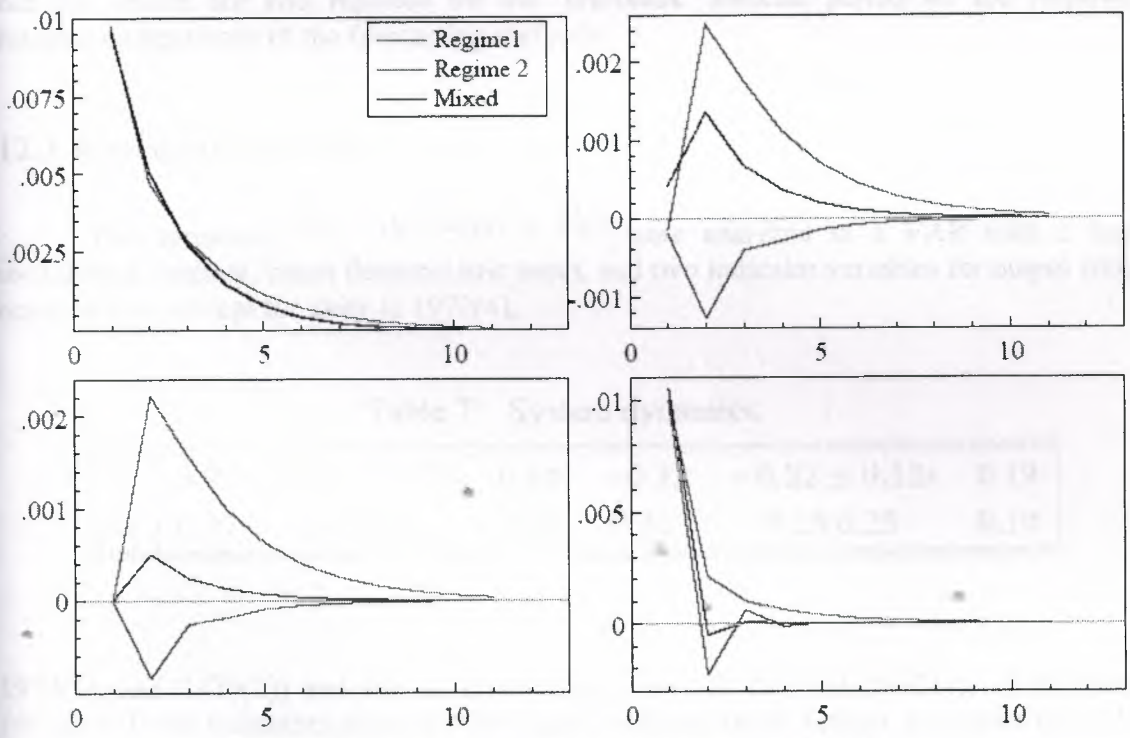


Figure 22 Impulse response comparisons in an I(0) VAR.

Let M denote nominal M1, I total final expenditure, P its deflator, and R the interest rate on threemonth Local-Authority bills: lower case denotes logs, and $\Delta = (1 - L)$ is the first difference, when L is the lag operator. We consider forecasting over two distinct historical periods. For the first, the sample period is 1964(3)–1985(2), after initial values for lags, with the remaining observations for 1985(3)– 1989(2) retained for out-of-sample forecasting. The four variables $(m - p, \Delta p, i, R)$ appear to be I(1), so we begin by developing a dynamic system, undertake a cointegration analysis, then simplify to a model in I(0) space. As expected, the system's multi-step forecast performance is very poor. Adding a step-shift dummy to allow a separate intercept (autonomous growth) over the forecast period rescues the forecasts, similar to those from the 'correct' model: this is a form of intercept correction (see Hendry and Clements, 1994). We also develop 'time-series' models which do not fail on forecasting, as the test period commences after the structural break. Thus this choice of forecast period illustrates the efficacy of intercept corrections and 'time-series' models when a major break has occurred prior to forecasting. The second exercise selects the immediately preceding 16 observations as the forecast period, i.e., 1981(3)–1985(2), to assess the costs of these strategies over a period when the dynamic system remains a reasonably good approximation to the DGP. The specifications of the models for the second exercise are carried over directly from those estimated on the longer sample – given its insample constancy, re-specifying the model would probably make little difference. Below we emphasize the 'post-break' forecast period, where the hitherto well-specified simultaneous equation model exhibits spectacular forecast failure, but full results are also reported for the 'pre-break' forecast period for the empirical forecast comparisons of the forecasting methods.

12.1 A 4-equation VAR

The variables $((m-p_t), \Delta p_t, i_t, R_t)$ were analyzed in a VAR with 2 lags, including a constant, linear deterministic trend, and two indicator variables for output (dout equal to zero, except for unity in 1972(4),

Table 7 System dynamics.

λ	0.97	$0.86 \pm 0.17\iota$	0.64	-0.32	$-0.22 \pm 0.12\iota$	0.19
$\lvert\lambda\rvert$	0.97	0.87,0.87	0.64	0.32	0.25,0.25	0.19

1973(1), and 1979(2)) and the oil crises (doil, unity in 1973(3), 1973(4), 1974(2) and 1979(3)). These indicators adjust for the largest residuals in the system; the issues raised by how the dummies enter the system are discussed in Doornik et al. (1998). All computations and graphics were produced by GiveWin and PcFiml (see Doornik and Hendry, 1996, 1999). The money-demand equation had residual serial correlation, but otherwise the outcomes are consistent with a congruent system.

The eigenvalues of the long-run matrix are -0.41, -0.05, and $-0.11 \pm 0.19\iota$ (using ι to denote $\sqrt{-1}$ to avoid confusion with income, i), so the rank seems non-zero, and is likely to be one or perhaps two. The eigenvalues of the companion matrix (denoted λ) are shown in table 7. Only one root is very close to unity, two have moduli near 0.9, and the remainder are small.

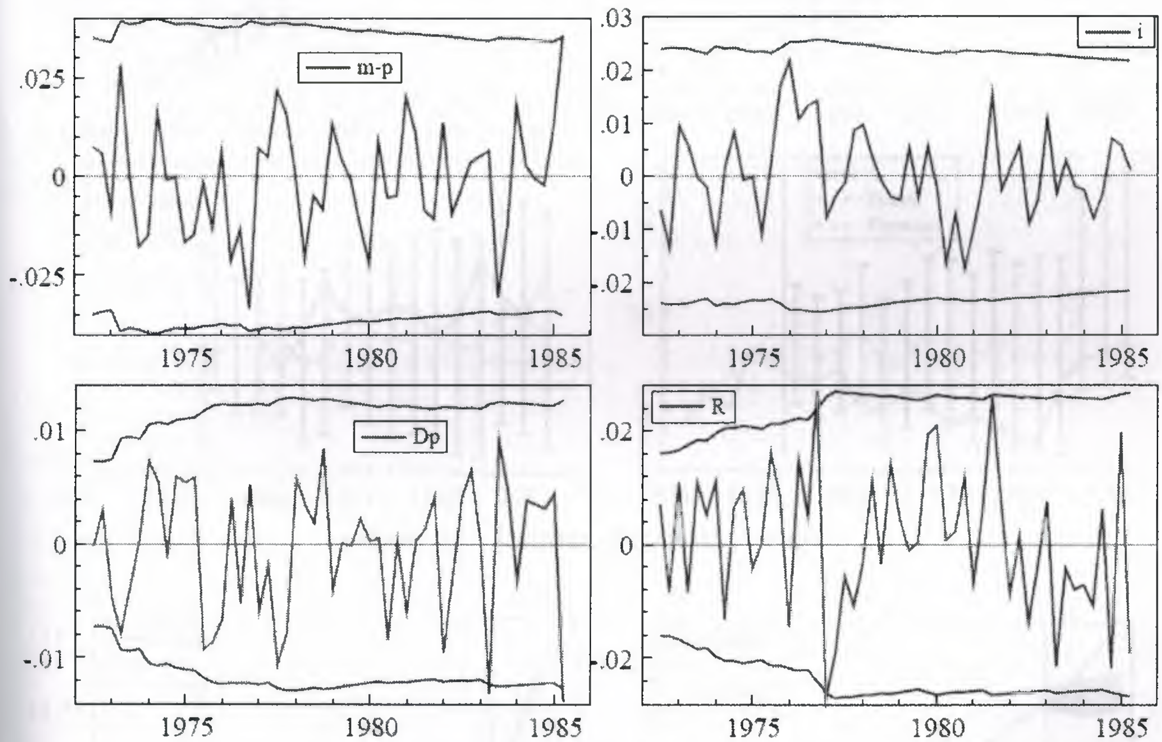


Figure 23 System recursive 1-step residuals.

All first lags were significant, but the second lags and the trend were insignificant (on F(4; 72), at 5% or less). Figure 23 shows the in-sample recursively-computed system 1-step residuals with 95% confidence bands: the equations for Δp and R are somewhat non-constant, although the system breakpoint Chow (1960) test did not exceed the 1% critical value within sample. Figure 24 reports the 1-step ahead out-of-sample forecasts with approximate 95% confidence intervals: there is some evidence of mis-forecasting in the money and interest-rate equations, but overall, the performance is respectable, consistent with the constancy-test outcome of F(64; 72) = 0.87. Finally, fig. 25 records the fitted and actual values for each variable together with the 16-steps ahead forecasts and their approximate 95% confidence intervals: the excellent fit but awful multi-step forecast performance of this unrestricted system is manifest. This outcome is the combination of the post-forecasting deterministic shift due to financial innovation, interacting with the I(1) formulation and the over-parameterization. We address these last two issues in turn.

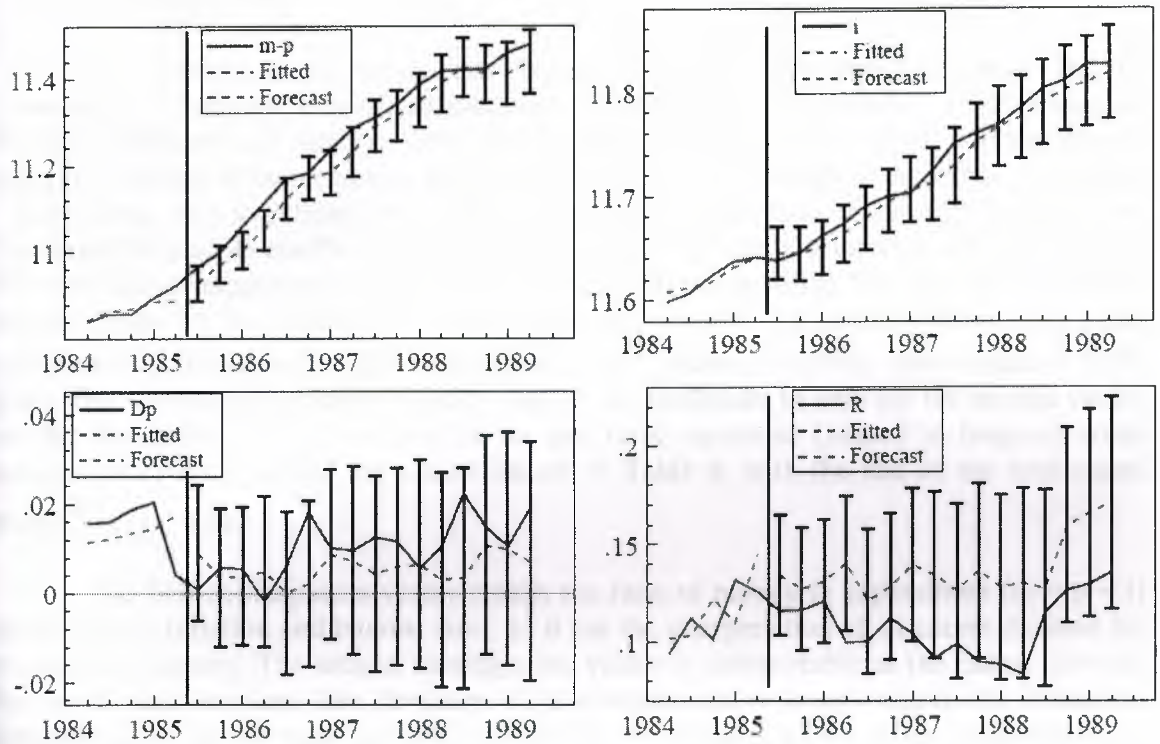


Figure 24 System1-step ahead forecasts.

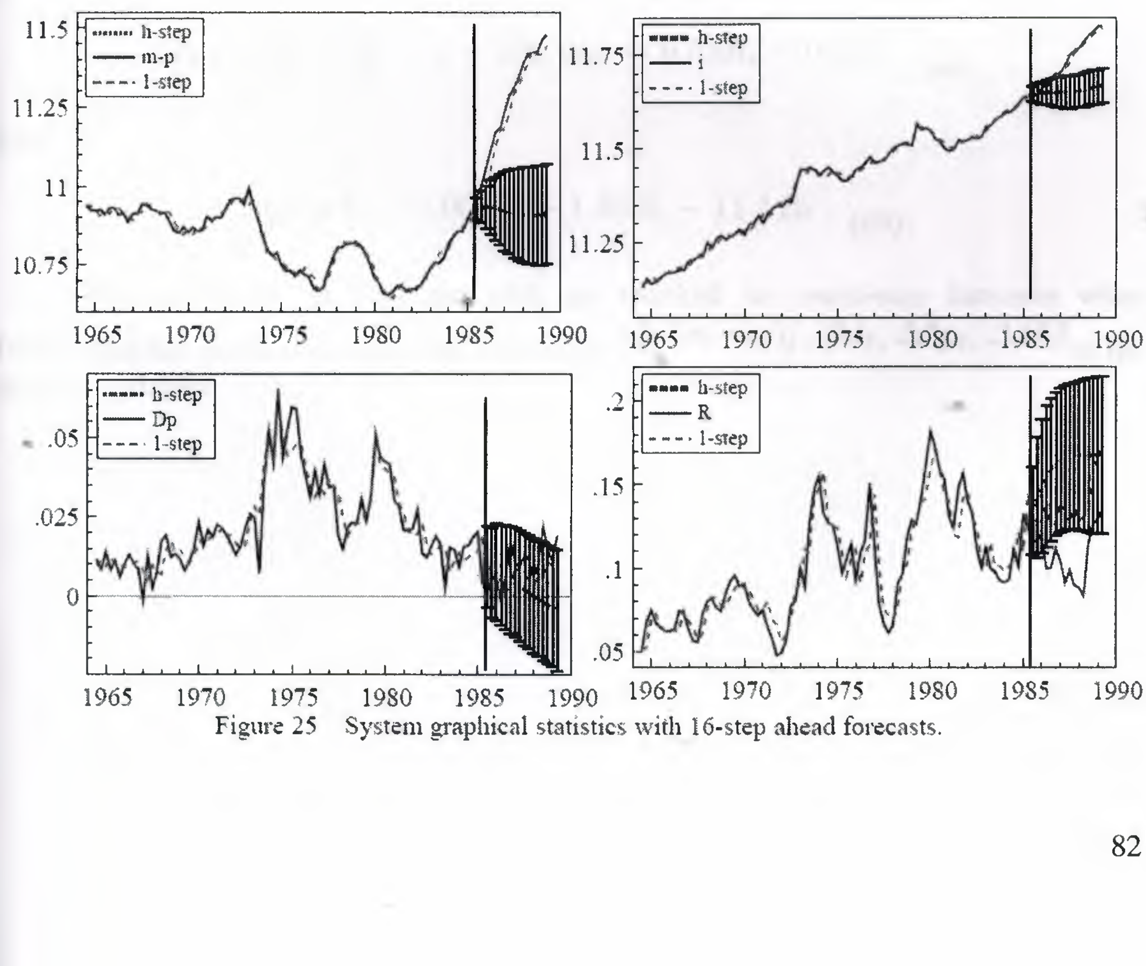


Figure 25 System graphical statistics with 16-step ahead forecasts.

12.2 Cointegration

The cointegration analysis restricted the trend to the cointegration space, but the constant and dummies entered unrestrictedly (see Doornik and Hendry, 1999, Banerjee, Dolado, Galbraith and Hendry, 1993, and Johansen, 1995). The null of no cointegration is strongly rejected at conventional I(1) critical values, and although a second cointegrating vector is not very significant, we retain it following Hendry and Mizon (1993), given the interpretability of its coefficients after restrictions. To uniquely determine and interpret the two possible cointegration vectors, we removed the trend from the first, and m−p from the second. Then we restricted the income coefficient to −1 in the first vector, and the trend coefficient in the second to the mean value of Δi (namely, 0:0062, approximately 2.5% p.a.), also eliminating inflation. Finally, we set the feedbacks to zero for the second vector on the first equation, and the first on the last three equations (related to long-run weak exogeneity) which yielded the results shown in Table 8, with the test of the restrictions being $\chi^2(7) = 9.58$.

The first cointegration vector relates the ratio of money to expenditure (m − p − i) negatively to inflation and interest rates, so it has the interpretation of an excess demand for transactions money. The second cointegration vector is interpretable as the excess demand for goods and services (the deviation of expenditure from trend, negatively related to interest rates), and its main influence is onto the i equation, so we retain these two long-run relations. Thus, the two, zero-mean, I (0) linear combinations defining the equilibrium-correction mechanisms (EqCMs) are:

$$c_{1,t} = m_t - p_t - i_t + 6.67\Delta p_t + 6.53R_t - 0.223 \qquad (48)$$

and:

$$c_{2,t} = i_t - 0.0062t + 1.22R_t - 11.125 \qquad (49)$$

The definitions in (48) and (49) are required for multi-step forecasts when formulating the model in terms of the differences $(\Delta(m-p)_t, \Delta i_t, \Delta^2 p_t, \Delta R_t)$ of the original variables.

Table 8 Restricted Cointegration analysis.

$$\left[\begin{array}{ccc} \widehat{\alpha} & 1 & 2 \\ m-p & \underset{(0.014)}{-0.098} & \underset{(-)}{0} \\ i & \underset{(-)}{0} & \underset{(0.030)}{-0.124} \\ \Delta p & \underset{(-)}{0} & \underset{(0.016)}{-0.019} \\ R & \underset{(-)}{0} & \underset{(0.035)}{-0.007} \end{array}\right], \quad \left[\begin{array}{cccccc} \widehat{\beta}' & m-p & i & \Delta p & R & t \\ 1 & \underset{(-)}{1} & \underset{(-)}{-1} & \underset{(1.61)}{6.67} & \underset{(0.67)}{6.53} & \underset{(-)}{0} \\ 2 & \underset{(-)}{0} & \underset{(-)}{1} & \underset{(-)}{0} & \underset{(0.31)}{1.22} & \underset{(-)}{-0.0062} \end{array}\right]$$

12.3 The I(0) System

Going from the second-order VAR in the levels of the variables $((m-p_t), \Delta p_t, i_t, R_t)$ to a simultaneousequations model involves a number of steps (see, e.g., Hendry and Mizon, 1993, Clements and Mizon, 1991), any of which might potentially affect forecast performance. We first considered the impact of imposing cointegration. The initial system in the levels of the variables is given an equivalent representation in terms of differences, cointegrating combinations, and (two) lagged level terms. The 1) level terms are then deleted from all four equations to assess the impact of imposing unit roots and cointegration, as studied by Clements and Hendry (1995) via Monte Carlo, and empirically in a simplified monetary model, as distinct from parsimony per se, so the insignificant I(0) terms are retained. Imposing cointegration made little difference to forecasting performance relative to the unrestricted VAR.

Table 9 FIML model estimates.

$$\Delta(m-p)_t = -\underset{(0.06)}{0.16}\,\Delta(m-p-i)_{t-1} - \underset{(0.16)}{0.68}\left(\Delta^2 p_t + \Delta R_t\right) - \underset{(0.009)}{0.098}\,c_{1,t-1}$$

$$\Delta i_t = \underset{(0.006)}{0.050}\,Dout - \underset{(0.022)}{0.11}\,c_{2,t-1} + \underset{(0.07)}{0.20}\,\Delta i_{t-1} + \underset{(-)}{0.0062}$$

$$\Delta^2 p_t = -\underset{(0.08)}{0.34}\,\Delta^2 p_{t-1} + \underset{(0.004)}{0.028}\,Doil - \underset{(0.0007)}{0.0015} - \underset{(0.015)}{0.029}\,c_{2,t-1}$$

$$\Delta R_t = \underset{(0.10)}{0.14}\,\Delta R_{t-1} + \underset{(0.007)}{0.014}\,Doil + \underset{(0.06)}{0.14}\,\Delta(m-p)_{t-1}$$

12.4 A Simultaneous-Equations Model

A model of the I(0) system was developed by sequential simplification, based on earlier findings, and delivered the estimates shown in Table 9, augmented by the definitions in (48) and (49). This resulted in only 13 estimated I(0) parameters plus the three I(1) from Table 8 (so should avoid any overparameterization problems that may have affected the initial system), but was an acceptable reduction as the likelihood-ratio test of all the restrictions yielded $\chi^2_{\text{or}}(23) = 14.1\ (\text{p} > 0.92)$, which does not reject.

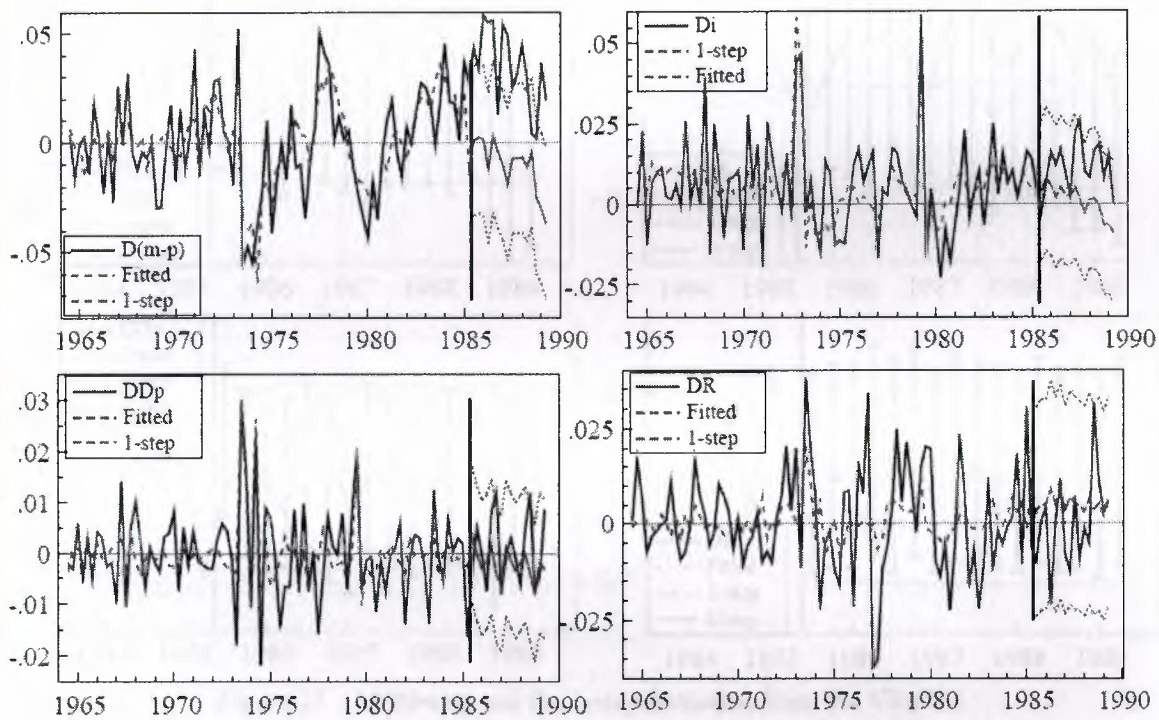


Figure 26 Fitted and actual values with 1-step forecasts, from the VEqCM.

The model is a valid reduction of the initial system, but $\text{F}_{\text{Ch}}(64, 80) = 3.99^{**}$,so parameter constancy out-of-sample is strongly rejected, and the 1-step forecast performance is poor relative to the in-sample fit, as fig. 26 shows, especially for $\Delta(m - p)$.

12.5 Multi-Step Forecasts

The model's multi-step forecast performance is not as poor as that of the system, suggesting some benefit from the I(0) reduction and parsimony, perhaps because the explanatory variables in Table 9 have small mean values. Figure 27 shows both the multi-step and the 1-step forecasts for comparison, only over the forecast horizon. For the last three variables, the forecast bands for the multi-step forecasts hardly increase as the horizon increases, consistent with their nearly non-dynamic nature, so the forecasts quickly become the unconditional means of their respective growth rates. Further, the bands are not much larger than the corresponding 1-step bars. The bands increase at first for m− p, where the EqCM plays a key role, but again the forecasts converge to the mean growth, although now the 1-step bars are distinctly narrower. Nevertheless, despite the structural break revealed by the parameter constancy test, multi-step predictive failure is not nearly so manifest.

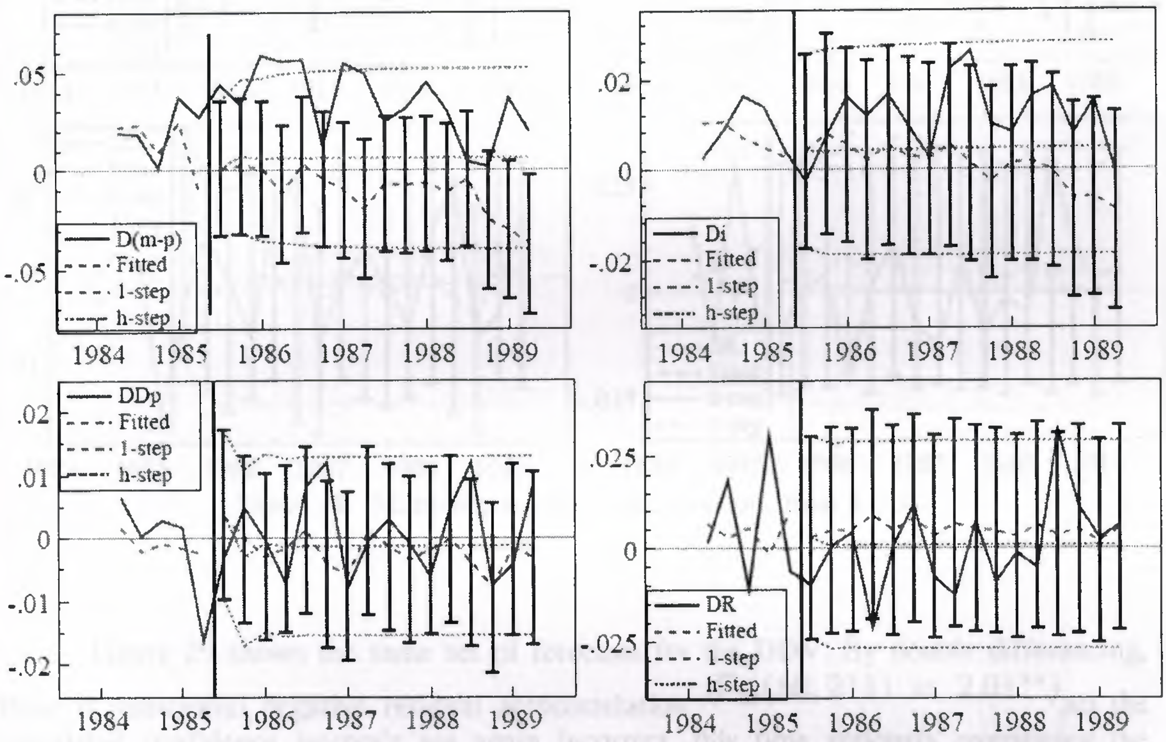


Figure 27 Multi-step and the 1-step forecasts, from the VEqCM.

12.6 Forecast Comparisons

The theory above predicts that the DV and DDV should be less susceptible to a deterministic structural break in the equilibrium mean than the VEqCM, but have larger forecast standard errors. The former corresponds to dropping the EqCMs from the VEqCM, replacing the $\Delta(m-p)$equation by its 'reduced form', and eliminating insignificant variables in the resulting model. Figure 28 shows the combined 1- step and multi-step forecasts. By eliminating the equilibrium-correction terms, the DV suffers from residual autocorrelation ($(\mathsf{F}_{ar}(80,231) = 1.51^{**})$, so its confidence intervals calculated by the usual formulae are incorrect, probably understating the actual uncertainty. Nevertheless, the absence of bias in the forecasts conforms to the earlier theory, when forecasting after a break.

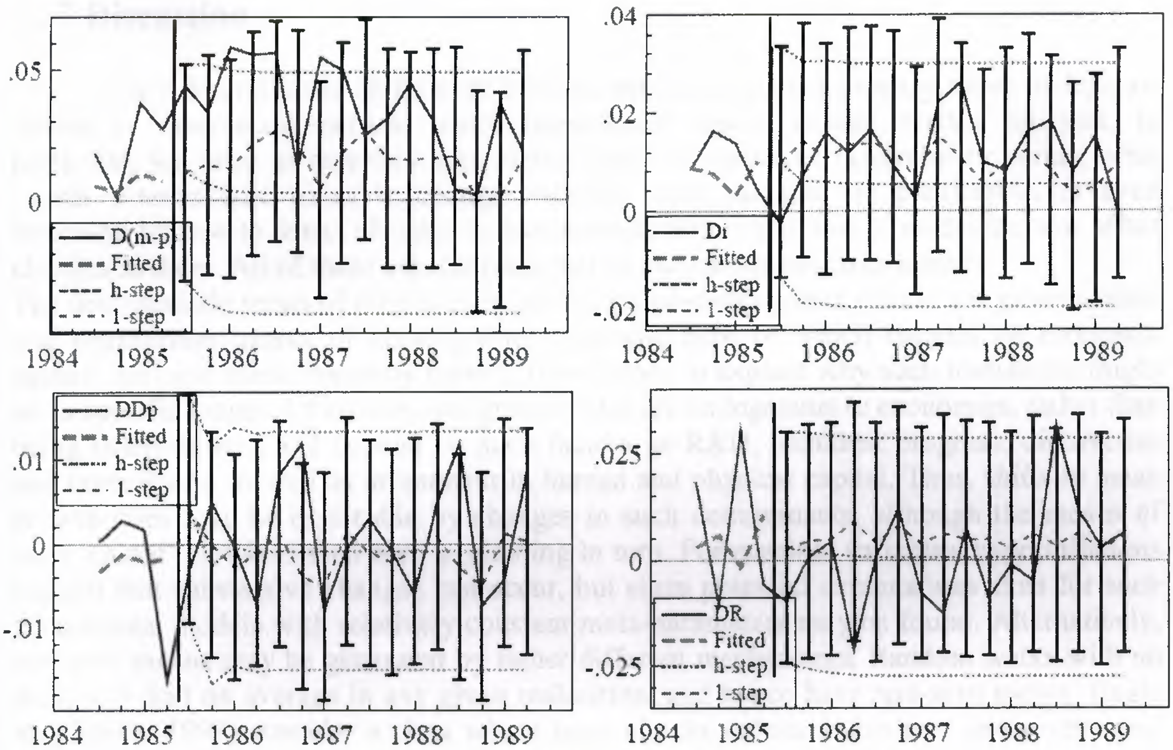


Figure 28 Multi-step and the 1-step forecasts, from the DV.

Figure 29 shows the same set of forecasts for the DDV. By double differencing, there is substantial negative residual autocorrelation $(\mathsf{F}_{\mathrm{ar}}(80, 211) = 2.01^{**})$,so the calculated confidence intervals are again incorrect, this time seriously overstating the uncertainty. Nevertheless, the bias performance seems good visually.

Next, fig. 30 compares all three multi-step forecasts, in the space of $(\Delta(m-p), \Delta i, \Delta R, \Delta^2 p)$. The actual multi-step forecasts are very similar for all three forecasting devices, namely zero $(\Delta R, \Delta^2 p)$ or the unconditional growth rate $(\Delta(m-p), \Delta i)$.In this representation, the DDV has easily the largest confidence intervals, and they increase rapidly in the horizon, matching the theoretical calculations (although they are upwards biased by the negative residual serial correlation). Between the VEqCM and the DV, the VEqCM has the wider intervals for money demand where the EqCM is strongest, but they are closely similar for the other three variables.

12.7 Discussion

Since deterministic shifts over the forecast horizon are a primary cause of forecast failure in macroeconometrics, many inter-related issues require further analysis. In particular, we need to ascertain what determines the values of deterministic terms; what causes deterministic terms to change; whether such changes are predictable, or even forecastable; how to detect changes in deterministic terms; and how to model them or offset changes in them. All of these aspects form part of our current research agenda.
The deterministic terms of relevance to an integrated-cointegrated system are growth rates, and equilibrium means of cointegrating relations, both of which depend on economic agents' decision rules. Presently there is little theory to explain why such transforms might lie in specific ranges. Of course, real growth rates are endogenous to economies, rather than being deterministic, and depend on such factors as R&D, technical progress, discoveries and innovations, as well as investment in human and physical capital. Thus, shifts in mean growth rates may be explicable by changes in such determinants, although the means of such 'causal' variables then need explaining in turn. For nominal variables, hyperinflations suggest that substantive changes can occur, but since potential explanations exist for such phenomena, models with relatively constant meta-parameters may be found. Alternatively, non-zero means may be generated by rather different mechanisms. Random walks with no drift, still drift on average in any given realization, and hence have non-zero means. Engle and Smith (1998) consider a class where large shocks induce a unit root in an otherwise stable process, so have a permanent effect akin to the type of deterministic shift discussed above. Given the rarity of big shocks, any realization will again have a non-zero mean with a high probability.

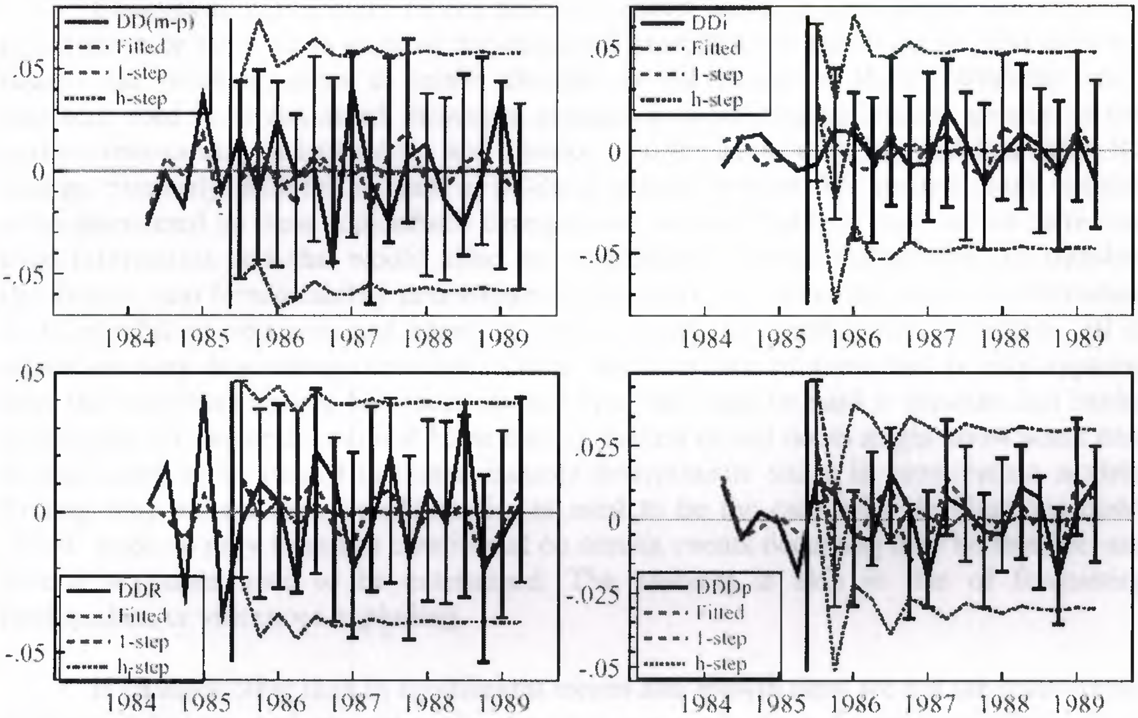


Figure 29 Multi-step and the 1-step forecasts, from the DDV.

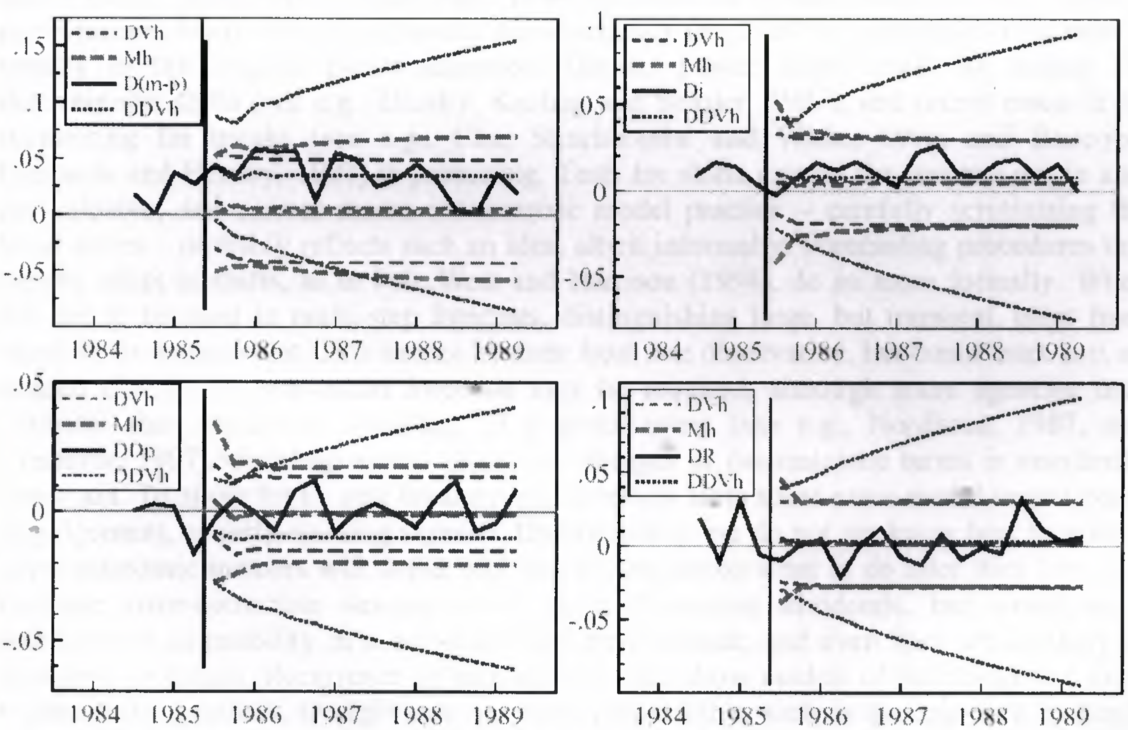


Figure 30 Multi-step forecasts in differences from the three forecasting devices.

Changes to equilibrium means probably reflect shifts in unmodelled determinants, as seems to be the case in most of the empirical examples we know. Again, that does not resolve the problem, since it entails changes in the means of those variables, which therefore need to be explained. However, changes to legislation can matter greatly, as can various innovations in technology and finance, and these are only 'unmodelled' after the change. Similarly, with the impacts of political turmoil or wars. We suspect much remains to be discovered on these topics.Such changes may be predictable in part, in that there may exist information sets that would affect the conditional forecast distribution. To translate that feature into forecastability necessitates knowing ex ante what the relevant information is, being able to obtain it, and knowing how it enters the conditional distribution, all of which are very demanding conditions. Often, the relevance of some fact is only apparent after the associated change has occurred, and even then may be hard to measure and harder to forecast: for example, political horse trading behind closed doors might allow some item of legislation to be passed that then induces deterministic shifts in econometric models. Timing may always prove problematic, as used to be the case with devaluations under 'fixed' pegs, so only forecasts conditional on certain events occurring may be feasible, and several scenarios need to be entertained. The problem is akin to that of forecasting earthquakes, or volcanoes exploding.

If changes other than in equilibrium means and growth rates are not the main source of forecast failure, and are not easily detected in-sample, then directed tests focused on deterministic shifts may provide more powerful tests for breaks likely to harm forecast performance. Many tests of parameter non-constancy check all the parameters of a model, usually in the original parameterization. Greater power might result by testing for deterministic shifts (see e.g., Hendry, Krolzig and Sensier, 1997), and recent research on monitoring for breaks (see e.g., Chu, Stinchcombe and White, 1996, and Banerjee, Clements and Hendry, 1997) is promising. Tests for shifts around the forecast origin also are valuable, and current macro-econometric model practice – carefully scrutinizing the latest errors – probably reflects such an idea, albeit informally. Forecasting procedures that rapidly adapt to shifts, as in Pole,West and Harrison (1994), do so more formally. When ICs are to be used in multi-step forecasts, distinguishing large, but transient, blips from mean shifts is important. This cannot be done from one observation, but could from two, so sudden changes in announced forecasts may be required, although some agencies may 'smooth' their successive forecasts of a given event (see e.g., Nordhaus, 1987, and Clements, 1997).Modeling and/or offsetting changes in deterministic terms is manifestly important. To allow for ex ante breaks needs foresight from some extra-model source (such as judgement, or early-warning signals). Unfortunately, we do not yet know how to predict when economic meteors will strike, and can but advise on what to do after they have hit. Genuine error-correction devices could repay handsome dividends, but would need considerable adaptability in a non-stationary environment; and even then are unlikely to anticipate problems. Recurrence of such shifts would allow models of their behavior, as in regime-shift equations, though there are other possibilities, such as the approach in Engle and Smith (1998). Improved ICs may be possible: Clements and Hendry (1998d) and §18.9 discuss some ideas.

12.8 Modelling Shifts

Regime-shifting models, such as Markov-switching autoregressions, as in Hamilton (1989, 1993), or self-exciting threshold autoregressive models, as in Tong (1978), seek to model changes by including stochastic and deterministic shifts in their probability structure. By separately modelling expansionary and contractionary phases of business cycles, say, an implicit assumption is that the shifts are regular enough to be modelled. In practice, the forecasting superiority of such approaches is controversial: conditional on being in a particular regime, these models may yield gains (see, e.g., Tiao and Tsay, 1994, Clements and Smith, 1999), but unconditionally there is often little improvement over linear models on criteria such as MSFE (see, e.g., Pesaran and Potter, 1997, Clements and Krolzig, 1998). However, they may be favoured on qualitative measures of forecast performance, or by approaches that evaluate the whole forecast density (Clements and Smith, 2000). Nevertheless, given the prominence of deterministic shifts as an explanation for forecast failure, efforts to model such shifts may yield significant rewards.

12.9 Intercept Corrections

When the source of a model's mis-specification is known, it is usually corrected, but in many settings, mis-specifications are unknown, so are difficult to correct. One widely-used tool is intercept correction (denoted IC), which sets the model 'back on track' to start from the actual forecast origin x_T . Hendry and Clements (1994) develop a general theory of intercept corrections, and Clements and Hendry (1996) show that such corrections can robustify forecasts against breaks that have happened, but only at the cost of an often substantial increase in forecast-error variance. The form of correction envisaged in that analysis is such that the correction alters as the forecast origin moves through the sample – the correction is always based on the error(s) made at, or immediately prior to, the origin. However, those forms of correction require a steadily expanding information set, and to treat the intercept-correcting strategy on a par with the other forecasting models, in this section we consider a simpler correction. This form of intercept correction can be implemented by adding an indicator variable equal to unity from the last sample observation onwards, so that the same correction is applied at all forecast origins. Figure 27a shows why such an IC will work here: immediately prior to forecasting, the model is under- fitting by a substantial amount, and 'shifting' the forecast origin to the data will offset much of the later mis-forecasting. To reduce the forecast-error variance, the IC can be set to unity for the last few sample observations: here we chose two (namely 1985(1), (2)). Further, to highlight the effects, we only entered the IC in the first equation (where it was significant at the 5% level: it was insignificant if added to the remaining equations). Figure 31 shows the impact in the I(0) representation, and fig. 32 in the I(1). The IC shifts upward the sequence of forecasts of $\Delta(m-p)$,but still underestimates the resulting growth, and hence the level. Equally, the improvement in the UK's rate of output growth from 0.62% per quarter over the estimation sample to 1.17% over the forecast period leads to substantial under-prediction of the final level for the i equation with c_2. Thus, these outcomes are all in line with the theory.

13 METHODOLOGICAL IMPLICATIONS

Our research has implications for econometric research outside the forecasting arena. Most importantly, in a world of deterministic shifts forecast performance is not a good guide to model choice, unless the sole objective is short-term forecasting. Thus, there are no grounds for selecting the best forecasting

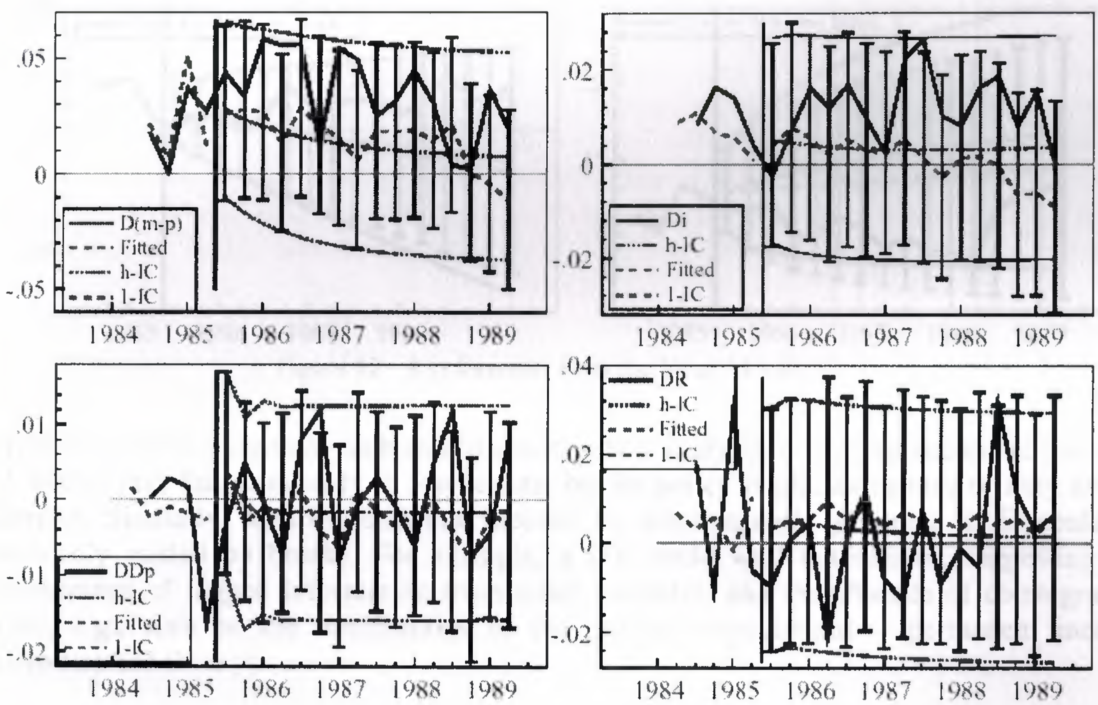


Figure 31 I(0) forecasts from the VEqCM with IC.

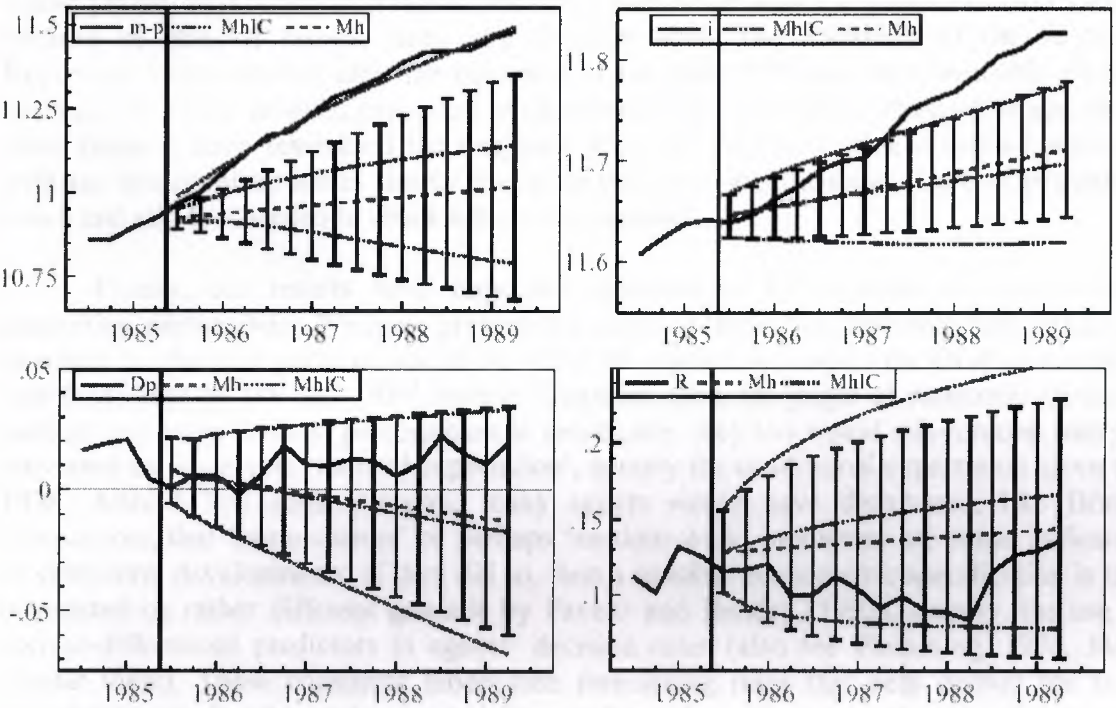


Figure 32 I(1) forecasts from the VEqCM with IC.

model for other purposes, such as economic policy analysis, or testing economic theories. A model may fail grievously in forecasting, but its policy implications may or may not be correct. Similarly, tests of economic theories by whole-sample goodness of fit could be seriously misled by breaks. For example, a DV could well outperform, suggesting the irrelevance of lagged information from other variables, and the absence of cointegration (see, e.g., tests of the implications of the rational expectations – permanent income consumption theory).

Further, if forecast failure is primarily due to forecast-period deterministic shifts, then there are no possible within-sample tests of later failure. The UK M1 example illustrates this point. Whether the model breaks down after the introduction of interest-bearing checking accounts depends on how the model is updated over this period – specifically, whether the interest rate variable is modified for this change, not on the model's fit to the data over the original period. Equally, we have shown that non-congruent models may not fail, and congruent fail, so conventional diagnostic tests do not suffice either as indicators of potential failure. Consequently, the methodology by which a model is developed empirically may have little to do with its later forecasting success or failure, in stark contrast to the claims in Hess, Jones and Porter (1997).

Conversely, as a usable knowledge base, and for fostering empirical understanding, theory-related, congruent, encompassing models remain undominated. Forecast failure does not, though it might, entail an invalid theoretical model; it does reveal forecast data that are different from the in-sample observations, and hence an incomplete empirical model for the whole period. It is a non sequitur to reject the theory on which a model is based simply because of forecast failure: thus, any decision about the invalidity of the so-called Keynesian macro-models after the oil-crises of the mid-1970s was not justifiable on that basis alone. Those models, expanded to incorporate the effects that changed, might in all other respects have resembled the originals. Equally, they may have remained rejected, with the deterministic shifts simply revealing their mis-specifications. Careful evaluation was – and still is – needed to check which case applied.

Finally, our results have important implications for theories of expectations-generating mechanisms. We have presented a range of theoretical and empirical situations in which the forecast performance of the VEqCM – which represents the DGP in-sample – was dominated by DV and DDV models. Consider, then, the plight of economic agents in such an economy: without precognition or prescience, they too would mis-forecast badly if they used the in-sample 'rational expectation', namely the conditional expectation given the DGP. After a few such mistakes, many agents would have discovered, like British Chancellors, that 'same change' or perhaps 'random-walk' predictors are better indicators of short-term developments. If they did so, then a sensible econometric specification is that postulated on rather different grounds by Favero and Hendry (1992), namely, the use of second-differenced predictors in agents' decision rules (also see Flemming, 1976, for a similar view). These constitute model-free forecasting rules that help deliver the least biased forecasts feasible under deterministic shifts, and are immune to the type of argument advanced by Lucas (1976).

Thus, the study of forecasting has wide-reaching methodological implications for economic and econometric research.

CONCLUSION

Despite the relatively weak assumptions that the economy under analysis is non-stationary and subject to unanticipated structural breaks, that the model may differ from the mechanism in unknown ways, and that it requires both selection and estimation from available data, a useful theory of economic forecasting can be developed. The resulting implications can differ considerably from those obtained when the model is assumed to coincide with a constant mechanism. For example, causal information cannot be shown to uniformly dominate non-causal in such a setting, so that the preferred model on forecast accuracy criteria may omit relationships between variables that are operative in the economy, and important for policy. Also, intercept corrections have no theoretical justification in stationary worlds with correctly-specified empirical models, but in a world subject to structural breaks of unknown form, size, and timing, serve to 'robustify' forecasts against deterministic shifts. The efficacy of intercept corrections confirms that the best forecasting model is not necessarily the best policy model.

The taxonomy of potential sources of forecast errors clarifies the roles of model mis-specification, sampling variability, error accumulation, forecast origin mis-measurement, intercept shifts, and slopeparameter changes. All sources reduce forecastability, but forecast failure seems primarily attributable to deterministic shifts. The consequences of many forms of structural break can be derived analytically, and this reveals that different models may be differentially susceptible to structural breaks, as occurs from over-differencing. Intercept-corrections which exploit this susceptibility to eliminate the impacts of breaks could be useful.

Allen and Fildes (2000) overview the empirical evidence on 'what works' in practice for econometric forecasting. They suggest that the evidence is consonant with the claim that simple models, which are admissible reductions of VARs with relatively generous lag specifications, estimated by least squares, and tested for constant parameters will do best on average. They also note the following as unresolved issues:

(1) the role of causal variables, particularly when such variables must be forecast by auxiliary models;
(2) whether congruent models outperform non-congruent, and hence:
(3) the value-added of mis-specification testing in selecting forecasting models; and
(4) whether cointegration restrictions improve forecasts.

We have shown that causal variables cannot be proved to dominate non-causal; that congruent models need not outperform non-congruent, so rigorous mis-specification testing need not help for selecting forecasting models; and that equilibrium-mean shifts induce forecast failure, so cointegration will improve forecasting only if the implicit means remain constant. All of these converses occur under equilibrium-mean shifts that induce unanticipated departures of the model's unconditional mean $\mathsf{E}_m[y_{T+h}]$ from that of the data $\mathsf{E}[y_{T+h}]$.It follows that forecasting success is no better an index for model selection than forecast failure is for model rejection. Thus, any focus on 'out-of-sample' forecast performance (perhaps because of fears over 'data-mining') would appear to be unsustainable (see, e.g., Newbold, 1993, p.658), as would the belief that a greater reliance on economic theory will help forecasting (see, e.g., Diebold, 1998), because that does not tackle the root problem.

If the economy were reducible by transformations to a stationary stochastic process, where the resulting unconditional moments were constant over time, unanticipated departures of $\mathsf{E}_m[\mathbf{y}_{T+h}]$ from $\mathsf{E}[\mathbf{y}_{T+h}]$ would not occur, so well-tested, causally-relevant, congruent models which embodied valid theory restrictions would both fit best, and by encompassing, also dominate in forecasting on average. The prevalence historically of unanticipated deterministic shifts suggests that such transformations do not exist. Nevertheless, the case for continuing to use econometric systems probably depends on their competing reasonably successfully in the forecasting arena. Cointegration, co-breaking, and modelselection procedures as good as PcGets, with rigorous testing should help, but none of these ensures immunity to forecast failure from new breaks. Thus, for forecasting there is a powerful case for adopting robust approaches: a key development must be error-correction methods that do not eliminate policy-relevant sources of information (such as cointegration). An approach which incorporates causal information in a congruent econometric system for policy, but operates with robustified forecasts, merits consideration.

REFERENCES

1.Allen, P. G., and Fildes, R. A. (2000). Econometric forecasting strategies and techniques. Forthcoming, J. S. Armstrong, (ed.) Principles of Forecasting, Kluwer Academic Press, USA.

2.Banerjee, A., Clements, M. P., and Hendry, D. F. (1997). Monitoring for structural change in forecasting. Mimeo, Oxford Institute of Economics and Statistics, Oxford.

3.Banerjee, A., Dolado, J. J., Galbraith, J.W., and Hendry, D. F. (1993). Co-integration, Error Correction and the Econometric Analysis of Non-Stationary Data. Oxford: Oxford University Press.

4.Bjørnstad, J. F. (1990). Predictive likelihood: A review. Statistical Science,242-265.

5.Bontemps, C., and Mizon, G. E. (1996). Congruence and encompassing. Economics department, mimeo, European University Institute.

6.Box, G. E. P., and Jenkins, G.M. (1976). Time Series Analysis, Forecasting and Control. San Francisco: Holden-Day. First published, 1970.

7.Burns, T. (1986). The interpretation and use of economic predictions. In Proceedings of the Royal Society, No. A407, pp. 103–125.

8.Central Statistical Office, U. K. (1993). Economic Trends Annual Supplement. London: H.M.S.O.

9.Chong, Y. Y., and Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. Review of Economic Studies, 671–690. Reprinted in Granger, C.W. J. (ed.) (1990), Modelling Economic Series. Oxford: Clarendon Press.

10. Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. Econometrica, 591–605.

11. Chu, C.-S., Stinchcombe, M., and White, H. (1996). Monitoring structural change. Econometrica, 1045–1065.

12. Clements, M. P. (1997). Evaluating the rationality of fixed-event forecasts. Journal of Forecasting, 225–239.

13. Clements, M. P., and Hendry, D. F. (1993). On the limitations of comparing mean squared forecast errors. Journal of Forecasting, 617–637. With discussion.

14. Clements, M. P., and Hendry, D. F. (1994). Towards a theory of economic forecasting. In Hargreaves, C. (ed.), Non-stationary Time-series Analysis and Cointegration, pp. 9–52. Oxford: Oxford University Press.

15. Clements, M. P., and Hendry, D. F. (1995). Forecasting in cointegrated systems. Journal of Applied Econometrics, 127–146.

16. Clements, M. P., and Hendry, D. F. (1996). Intercept corrections and structural change. Journal of Applied Econometrics, 475–494.

17. Clements, M. P., and Hendry, D. F. (1998a). Forecasting economic processes. International Journal of Forecasting, 111–131.

18. Clements, M. P., and Hendry, D. F. (1998b). Forecasting Economic Time Series. Cambridge: Cambridge University Press. The Marshall Lectures on Economic Forecasting.

19. Clements, M. P., and Hendry, D. F. (1998c). On winning forecasting competitions in economics. Spanish Economic Review, 123–160.

20. Clements, M. P., and Hendry, D. F. (1998d). Using time-series models to correct econometric model forecasts. mimeo, Institute of Economics and Statistics, University of Oxford.

21. Clements, M. P., and Hendry, D. F. (1999a). Forecasting Non-Stationary Economic Time Series. Cam67 bridge, Mass.: MIT Press. The Zeuthen Lectures on Economic Forecasting.

22. Clements, M. P., and Hendry, D. F. (1999b). Forecasting Non-stationary Economic Time Series: The Zeuthen Lectures on Economic Forecasting. Cambridge, Mass.: MIT Press.

23. Clements, M. P., and Krolzig, H.-M. (1998). A comparison of the forecast performance of Markovswitching and threshold autoregressive models of US GNP. Econometrics Journal, C47–75.

24. Clements, M. P., and Mizon, G. E. (1991). Empirical analysis of macroeconomic time series: VAR and structural models. European Economic Review, 887–932.

25. Clements, M. P., and Smith, J. (1999). A Monte Carlo study of the forecasting performance of empirical SETAR models. Journal of Applied Econometrics, 124–141.

26. Clements, M. P., and Smith, J. (2000). Evaluating the forecast densities of linear and non-linear models: Applications to output growth and unemployment. Journal of Forecasting. Forthcoming.

27. Cook, S. (1995). Treasury economic forecasting. mimeo, Institute of Economics and Statistics, University of Oxford.

28. Corsi, P., Pollock, R. E., and Prakken, J. C. (1982). The Chow test in the presence of serially correlated errors. In Chow, G. C., and Corsi, P. (eds.), Evaluating the Reliability ofMacro-Economic Models. New York: John Wiley.

29. Cox, D. R., and Miller, H. D. (1965). The Theory of Stochastic Processes: Chapman and Hall.

30. Crafts, N. F. R., and Harley, C. K. (1992). Output growth and the British Industrial Revolution: A restatement of the Crafts–Harley view. Economic History Review, 703–730.

31. Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, J. S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. Economic Journal, 661–692. Reprinted in Hendry, D. F. (1993), Econometrics: Alchemy or Science? Oxford: Blackwell Publishers.

32. Diebold, F. X. (1998). The past, present and future of macroeconomic forecasting. The Journal of Economic Perspectives, 175–192.

33. Doan, T., Litterman, R., and Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions. Econometric Reviews, 1–100.