Chapter 12

FORECASTING WITH BREAKS

MICHAEL P. CLEMENTS

Department of Economics, University of Warwick

DAVID F. HENDRY

Economics Department, University of Oxford

Contents

Handbook of Economic Forecasting, Volume 1
Edited by Graham Elliott, Clive W.J. Granger and Allan Timmermann

DOI: 10.1016/S1574-0706(05)01012-8

Abstract

A structural break is viewed as a permanent change in the parameter vector of a model. Using taxonomies of all sources of forecast errors for both conditional mean and conditional variance processes, we consider the impacts of breaks and their relevance in forecasting models: (a) where the breaks occur after forecasts are announced; and (b) where they occur in-sample and hence pre-forecasting. The impact on forecasts depends on which features of the models are non-constant. Different models and methods are shown to fare differently in the face of breaks. While structural breaks induce an instability in some parameters of a particular model, the consequences for forecasting are specific to the type of break and form of model. We present a detailed analysis for cointegrated VARs, given the popularity of such models in econometrics.

We also consider the detection of breaks, and how to handle breaks in a forecasting context, including ad hoc forecasting devices and the choice of the estimation period. Finally, we contrast the impact of structural break non-constancies with non-constancies due to non-linearity. The main focus is on macro-economic, rather than finance, data, and on forecast biases, rather than higher moments. Nevertheless, we show the relevance of some of the key results for variance processes. An empirical exercise 'forecasts' UK unemployment after three major historical crises.

Keywords

economic forecasting, structural breaks, break detection, cointegration, non-linear models

JEL classification: C530

1. Introduction

A structural break is a permanent change in the parameter vector of a model. We consider the case where such breaks are exogenous, in the sense that they were determined by events outside the model under study: we also usually assume that such breaks were unanticipated given the historical data up to that point. We do rule out multiple breaks, but because breaks are exogenous, each is treated as permanent. To the extent that breaks are predictable, action can be taken to mitigate the effects we show will otherwise occur. The main exception to this characterization of breaks will be our discussion of non-linear models which attempt to anticipate some shifts.

Using taxonomies of all sources of forecast errors, we consider the impacts of breaks and their relevance in forecasting models:

(a) where the breaks occur after forecasts are announced; and
(b) where they are in-sample and occurred pre-forecasting, focusing on breaks close to the forecast origin.

New generic (model-free) forecast-error taxonomies are developed to highlight what can happen in general. It transpires that it matters greatly what features actually break (e.g., coefficients of stochastic, or of deterministic, variables, or of other aspects of the model, such as error variances). Also, there are major differences in the effects of these different forms of breaks on different forecasting methods, in that some devices are robust, and others non-robust, to various pre-forecasting breaks. Thus, although structural breaks induce an instability in some parameters of a particular model, the consequences for forecasting are specific to the type of break and form of model. This allows us to account for the majority of the findings reported in the major 'forecasting competitions' literature. Later, we consider how to detect, and how to handle, breaks, and the impact of sample size thereon. We will mainly focus on macro-economic data, rather than finance data where typically one has a much larger sample size. Finally, because the most serious consequences of unanticipated breaks are on forecast biases, we mainly consider first moment effects, although we also note the effects of breaks in variance processes.

Our chapter builds on a great deal of previous research into forecasting in the face of structural breaks, and tangentially on related literatures about: forecasting models and methods; forecast evaluation; sources and effects of breaks; their detection; and ultimately on estimation and inference in econometric models. Most of these topics have been thoroughly addressed in previous Handbooks [see Griliches and Intriligator (1983, 1984, 1986), Engle and McFadden (1994), and Heckman and Leamer (2004)], and compendia on forecasting [see, e.g., Armstrong (2001) and Clements and Hendry (2002a)], so to keep the coverage of references within reasonable bounds we assume the reader refers to those sources inter alia.

As an example of a process subject to a structural break, consider the data generating process (DGP) given by the structural change model of, e.g., Andrews (1993):

$$y_t = (\mu_0 + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p}) \\ + \left(\mu_0^* + \alpha_1^* y_{t-1} + \cdots + \alpha_p^* y_{t-p}\right)s_t + \varepsilon_t, \quad (1)$$

where $\varepsilon_t \sim \mathsf{IID}[0, \sigma_\varepsilon^2]$ (that is, Independently, Identically Distributed, mean zero, variance σ_ε^2), and s_t is the indicator variable, $s_t \equiv 1_{(t>\tau)}$ which equals 1 when $t > \tau$ and zero when $t \leqslant \tau$. We focus on breaks in the conditional mean parameters, and usually ignore changes in the variance of the disturbance, as suggested by the form of (1). A constant-parameter pth-order autoregression (AR(p)) for y_t of the form

$$y_t = \mu_{0,1} + \alpha_{1,1} y_{t-1} + \cdots + \alpha_{p,1} y_{t-p} + v_t \quad (2)$$

would experience a structural break because the parameter vector shifts. Let $\boldsymbol{\phi} = (\mu_0\ \alpha_1 \ldots \alpha_p)'$, $\boldsymbol{\phi}^* = (\mu_0^*\ \alpha_1^* \ldots \alpha_p^*)'$ and $\boldsymbol{\phi}_1 = (\mu_{0,1}\ \alpha_{1,1} \ldots \alpha_{p,1})'$. Then the AR($p$) model parameters are $\boldsymbol{\phi}_1 = \boldsymbol{\phi}$ for $t \leqslant \tau$, but $\boldsymbol{\phi}_1 = \boldsymbol{\phi} + \boldsymbol{\phi}^*$ for $t > \tau$ (in Section 5, we briefly review testing for structural change when τ is unknown). If instead, the AR(p) were extended to include terms which interacted the existing regressors with a step dummy D_t defined by $D_t = s_t = 1_{(t>\tau)}$, the extended model (letting $\mathbf{x}_t = (1\ y_{t-1} \ldots y_{t-p})'$)

$$y_t = \boldsymbol{\phi}_{1,d}' \mathbf{x}_t + \boldsymbol{\phi}_{2,d}' \mathbf{x}_t D_t + v_{t,d} \quad (3)$$

exhibits extended parameter constancy – $(\boldsymbol{\phi}_{1,d}'\ \boldsymbol{\phi}_{2,d}') = (\boldsymbol{\phi}'\ \boldsymbol{\phi}^{*\prime})$ for all $t = 1, \ldots, T$, matching the DGP [see, e.g., Hendry (1996)]. Whether a model experiences a structural break is as much a property of the model as of the DGP.

As a description of the process determining $\{y_t\}$, Equation (1) is incomplete, as the cause of the shift in the parameter vector from $\boldsymbol{\phi}$ to $\boldsymbol{\phi} + \boldsymbol{\phi}^*$ is left unexplained. Following Bontemps and Mizon (2003), Equation (1) could be thought of as the 'local' DGP (LDGP) for $\{y_t\}$ – namely, the DGP for $\{y_t\}$ given only the variables being modeled (here, just the history of y_t). The original AR(p) model is mis-specified for the LDGP because of the structural change. A fully-fledged DGP would include the reason for the shift at time τ. Empirically, the forecast performance of any model such as (2) will depend on its relationship to the DGP. By adopting a 'model' such as (1) for the LDGP, we are assuming that the correspondence between the LDGP and DGP is close enough to sustain an empirically relevant analysis of forecasting. Put another way, knowledge of the factors responsible for the parameter instability is not essential in order to study the impact of the resulting structural breaks on the forecast performance of models such as (2).

LDGPs in economics will usually be multivariate and more complicated than (1), so to obtain results of some generality, the next section develops a 'model-free' taxonomy of errors for conditional first-moment forecasts. This highlights the sources of biases in forecasts. The taxonomy is then applied to forecasts from a vector autoregression (VAR). Section 3 presents a forecast-error taxonomy for conditional second-moment forecasts based on standard econometric volatility models. Section 4 derives the properties of forecasts for a cointegrated VAR, where it is assumed that the break occurs at the very end of the in-sample period, and so does not affect the models' parameter estimates.

Alternatively, any in-sample breaks have been detected and modeled. Section 5 considers the detection of in-sample breaks, and Section 6 the selection of the optimal window of data for model estimation as well as model specification more generally in the presence of in-sample breaks. Section 7 looks at a number of ad hoc forecasting methods, and assesses their performance in the face of breaks. When there are breaks, forecasting methods which adapt quickly following the break are most likely to avoid making systematic forecast errors. Section 8 contrasts breaks as permanent changes with non-constancies due to neglected non-linearities, from the perspectives of discriminating between the two, and for forecasting. Section 9 reports an empirical forecasting exercise for UK unemployment after three crises, namely the post-world-war double-decades of 1919–1938 and 1948–1967, and the post oil-crisis double-decade 1975–1994, to examine the forecasts of unemployment that would have been made by various devices: it also reports post-model-selection forecasts over 1992–2001, a decade which witnessed the ejection of the UK from the exchange-rate mechanism at its commencement. Section 10 briefly concludes. Two Appendices A and B, respectively, provide derivations for the taxonomy Equation (10) and for Section 4.3.

2. Forecast-error taxonomies

2.1. General (model-free) forecast-error taxonomy

In this section, a new general forecast-error taxonomy is developed to unify the discussion of the various sources of forecast error, and to highlight the effects of structural breaks on the properties of forecasts. The taxonomy distinguishes between breaks affecting 'deterministic' and 'stochastic' variables, both in-sample and out-of-sample, as well as delineating other possible sources of forecast error, including model misspecification and parameter-estimation uncertainty, which might interact with breaks.

Consider a vector of n stochastic variables $\{\mathbf{x}_t\}$, where the joint density of $\mathbf{x}_t$ at time t is $\mathsf{D}_{\mathbf{x}_t}(\mathbf{x}_t \mid \mathbf{X}_{t-1}^1, \mathbf{q}_t)$, conditional on information $\mathbf{X}_{t-1}^1 = (\mathbf{x}_1, \dots, \mathbf{x}_{t-1})$, where $\mathbf{q}_t$ denotes the relevant deterministic factors (such as intercepts, trends, and indicators). The densities are time dated to make explicit that they may be changing over time. The object of the exercise is to forecast $\mathbf{x}_{T+h}$ over forecast horizons $h = 1, \dots, H$, from a forecast origin at T. A dynamic model $\mathsf{M}_{\mathbf{x}_t}[\mathbf{x}_t \mid \mathbf{X}_{t-1}^{t-s}, \tilde{\mathbf{q}}_t, \boldsymbol{\theta}_t]$, with deterministic terms $\tilde{\mathbf{q}}_t$, lag length s, and implicit stochastic specification defined by its parameters $\boldsymbol{\theta}_t$, is fitted over the sample $t = 1, \dots, T$ to produce a forecast sequence $\{\hat{\mathbf{x}}_{T+h|T}\}$. Parameter estimates are a function of the observables, represented by:

$$\hat{\boldsymbol{\theta}}_{(T)} = \mathbf{f}_T\big(\widetilde{\mathbf{X}}_T^1, \widetilde{\mathbf{Q}}_T^1\big), \tag{4}$$

where $\widetilde{\mathbf{X}}$ denotes the measured data and $\widetilde{\mathbf{Q}}_T^1$ the in-sample set of deterministic terms which need not coincide with $\mathbf{Q}_T^1$. The subscript on $\hat{\boldsymbol{\theta}}_{(T)}$ in (4) represents the influence of sample size on the estimate, whereas that on $\boldsymbol{\theta}_t$ in $\mathsf{M}_{\mathbf{x}_t}[\cdot]$ denotes that the derived parameters of the model may alter over time (perhaps reflected in changed estimates).

Let $\boldsymbol{\theta}_{e,(T)} = \mathsf{E}_T[\hat{\boldsymbol{\theta}}_{(T)}]$ (where that exists). As shown in Clements and Hendry (2002b), it is convenient, and without loss of generality, to map changes in the parameters of deterministic terms into changes in those terms, and we do so throughout.

Since future values of the deterministic terms are 'known', but those of stochastic variables are unknown, the form of the function determining the forecasts will depend on the horizon

$$\hat{\mathbf{x}}_{T+h|T} = \mathbf{g}_h\big(\widetilde{\mathbf{X}}_T^{T-s+1}, \widetilde{\mathbf{Q}}_{T+h}^T, \hat{\boldsymbol{\theta}}_{(T)}\big). \tag{5}$$

In (5), $\widetilde{\mathbf{X}}_T^{T-s+1}$ enters up to the forecast origin, which might be less well measured than earlier data; see, e.g., [Wallis (1993)].[1] The model will generally be a mis-specified representation of the LDGP for any of a large number of reasons, even when designed to be congruent [see Hendry (1995, p. 365)].

The forecast errors of the model are given by $\mathbf{e}_{T+h|T} = \mathbf{x}_{T+h} - \hat{\mathbf{x}}_{T+h|T}$ with expected value

$$\mathsf{E}_{T+h}\big[\mathbf{e}_{T+h|T} \mid \mathbf{X}_T^1, \{\mathbf{Q}^{**}\}_{T+h}^1\big], \tag{6}$$

where we allow that the LDGP deterministic factors (from which the model's deterministic factors $\widetilde{\mathbf{Q}}_{T+h}^T$ are derived) are subject to in-sample shifts as well as forecast period shifts, denoted by $**$ as follows. If we let τ date an in-sample shift ($1 < \tau < T$), the LDGP deterministic factors are denoted by $\{\mathbf{Q}^{**}\}_{T+h}^1 = [\mathbf{Q}_\tau^1, \{\mathbf{Q}^*\}_T^{\tau+1}, \{\mathbf{Q}^{**}\}_{T+h}^{T+1}]$. Thus, the pre-shift in-sample period is $1, \ldots, \tau$, the post-shift in-sample period is $\tau + 1, \ldots, T$, and the forecast period is $T + 1, \ldots, T + h$, where we allow for the possibility of a shift at T. Absences of $**$ and $*$ indicate that forecast and in-sample period shifts did not occur. Thus, $\{\mathbf{Q}^*\}_T^{\tau+1} = \mathbf{Q}_T^{\tau+1}$ implies no in-sample shifts, denoted by $\mathbf{Q}_T^1$, and the absence of shifts both in-sample and during the forecast period gives $\mathbf{Q}_{T+h}^1$. Let $\{\mathbf{Q}^*\}_{T+h}^1 = [\mathbf{Q}_\tau^1, \{\mathbf{Q}^*\}_{T+H}^{\tau+1}]$ refer to an in-sample shift, but no subsequent forecast-period shifts. The deterministic factors $\widetilde{\mathbf{Q}}_T^1$ in the model may also be mis-specified in-sample when the LDGP deterministic factors are given by $\mathbf{Q}_T^1$ ('conventional' mis-specification). Of more interest, perhaps, is the case when the mis-specification is induced by an in-sample shift not being modeled. This notation reflects the important role that shifts in deterministic terms play in forecast failure, defined as a significant deterioration in forecast performance relative to the anticipated outcome, usually based on the historical performance of a model.

We define the forecast error from the LDGP as

$$\boldsymbol{\varepsilon}_{T+h|T} = \mathbf{x}_{T+h} - \mathsf{E}_{T+h}\big[\mathbf{x}_{T+h} \mid \mathbf{X}_T^1, \{\mathbf{Q}^{**}\}_{T+h}^1\big]. \tag{7}$$

By construction, this is the forecast error from using a correctly-specified model of the mean of $\mathsf{D}_{\mathbf{x}_t}(\mathbf{x}_t \mid \mathbf{X}_{t-1}^1, \mathbf{q}_t)$, where any structural change (in, or out, of sample) is known and incorporated, and the model parameters are known (with no estimation error). It

[1] The dependence of $\hat{\boldsymbol{\theta}}_{(T)}$ on the forecast origin is ignored below.

follows that $\mathsf{E}_{T+h}[\boldsymbol{\varepsilon}_{T+h|T} \mid \mathbf{X}_T^1, \{\mathbf{Q}^{**}\}_{T+h}^1] = \mathbf{0}$, so that $\boldsymbol{\varepsilon}_{T+h|T}$ is an innovation against all available information. Practical interest, though, lies in the model forecast error, $\mathbf{e}_{T+h|T} = \mathbf{x}_{T+h} - \hat{\mathbf{x}}_{T+h|T}$. The model forecast error is related to $\boldsymbol{\varepsilon}_{T+h|T}$ as given below, where we also separately delineate the sources of error due to structural change and mis-specification, etc.

$$
\begin{aligned}
\mathbf{e}_{T+h|T} &= \mathbf{x}_{T+h} - \hat{\mathbf{x}}_{T+h|T} \\
&= \left(\mathsf{E}_{T+h}\left[\mathbf{x}_{T+h} \mid \mathbf{X}_T^1, \{\mathbf{Q}^{**}\}_{T+h}^1\right] - \mathsf{E}_{T+h}\left[\mathbf{x}_{T+h} \mid \mathbf{X}_T^1, \{\mathbf{Q}^{*}\}_{T+h}^1\right]\right) && \text{(T1)} \\
&\quad + \left(\mathsf{E}_{T+h}\left[\mathbf{x}_{T+h} \mid \mathbf{X}_T^1, \{\mathbf{Q}^{*}\}_{T+h}^1\right] - \mathsf{E}_{T}\left[\mathbf{x}_{T+h} \mid \mathbf{X}_T^1, \{\mathbf{Q}^{*}\}_{T+h}^1\right]\right) && \text{(T2)} \\
&\quad + \left(\mathsf{E}_{T}\left[\mathbf{x}_{T+h} \mid \mathbf{X}_T^1, \{\mathbf{Q}^{*}\}_{T+h}^1\right] - \mathsf{E}_{T}\left[\mathbf{x}_{T+h} \mid \mathbf{X}_T^1, \widetilde{\mathbf{Q}}_{T+h}^1\right]\right) && \text{(T3)} \\
&\quad + \left(\mathsf{E}_{T}\left[\mathbf{x}_{T+h} \mid \mathbf{X}_T^1, \widetilde{\mathbf{Q}}_{T+h}^1\right] - \mathsf{E}_{T}\left[\mathbf{x}_{T+h} \mid \mathbf{X}_T^{T-s+1}, \widetilde{\mathbf{Q}}_{T+h}^1, \boldsymbol{\theta}_{e,(T)}\right]\right) && \text{(T4)} \\
&\quad + \left(\mathsf{E}_{T}\left[\mathbf{x}_{T+h} \mid \mathbf{X}_T^{T-s+1}, \widetilde{\mathbf{Q}}_{T+h}^1, \boldsymbol{\theta}_{e,(T)}\right]\right. \\
&\qquad \left. - \mathsf{E}_{T}\left[\mathbf{x}_{T+h} \mid \widetilde{\mathbf{X}}_T^{T-s+1}, \widetilde{\mathbf{Q}}_{T+h}^1, \boldsymbol{\theta}_{e,(T)}\right]\right) && \text{(T5)} \\
&\quad + \left(\mathsf{E}_{T}\left[\mathbf{x}_{T+h} \mid \widetilde{\mathbf{X}}_T^{T-s+1}, \widetilde{\mathbf{Q}}_{T+h}^1, \boldsymbol{\theta}_{e,(T)}\right] - \mathbf{g}_h\left(\widetilde{\mathbf{X}}_T^{T-s+1}, \widetilde{\mathbf{Q}}_{T+h}^1, \hat{\boldsymbol{\theta}}_{(T)}\right)\right) && \text{(T6)} \\
&\quad + \boldsymbol{\varepsilon}_{T+h|T}. && \text{(T7)}
\end{aligned}
\tag{8}
$$

The first two error components arise from structural change affecting deterministic (T1) and stochastic (T2) components respectively over the forecast horizon. The third (T3) arises from model mis-specification of the deterministic factors, both induced by failing to model in-sample shifts and ‘conventional’ mis-specification. Next, (T4) arises from mis-specification of the stochastic components, including lag length. (T5) and (T6) denote forecast error components resulting from data measurement errors, especially forecast-origin inaccuracy, and estimation uncertainty, respectively, and the last row (T7) is the LDGP innovation forecast error, which is the smallest achievable in this class.

Then (T1) is zero if $\{\mathbf{Q}^{**}\}_{T+h}^1 = \{\mathbf{Q}^{*}\}_{T+h}^1$, which corresponds to no forecast-period deterministic shifts (conditional on all in-sample shifts being correctly modeled). In general the converse also holds – (T1) being zero entails no deterministic shifts. Thus, a unique inference seems possible as to when (T1) is zero (no deterministic shifts), or non-zero (deterministic shifts).

Next, when $\mathsf{E}_{T+h}[\cdot] = \mathsf{E}_T[\cdot]$, so there are no stochastic breaks over the forecast horizon, entailing that the future distributions coincide with that at the forecast origin, then (T2) is zero. Unlike (T1), the terms in (T2) could be zero despite stochastic breaks, providing such breaks affected only mean-zero terms. Thus, no unique inference is feasible if (T2) is zero, though a non-zero value indicates a change. However, other moments would be affected in the first case.

When all the in-sample deterministic terms, including all shifts in the LDGP, are correctly specified, so $\widetilde{\mathbf{Q}}_{T+h}^1 = \{\mathbf{Q}^{*}\}_{T+h}^1$, then (T3) is zero. Conversely, when (T3) is zero, then $\widetilde{\mathbf{Q}}_{T+h}^1$ must have correctly captured in-sample shifts in deterministic terms, perhaps because there were none. When (T3) is non-zero, the in-sample deterministic factors may be mis-specified because of shifts, but this mistake ought to be detectable. However, (T3) being non-zero may also reflect 'conventional' deterministic mis-specifications. This type of mistake corresponds to omitting relevant deterministic terms, such as an intercept, seasonal dummy, or trend, and while detectable by an appropriately directed test, also has implications for forecasting when not corrected.

For correct stochastic specification, so $\boldsymbol{\theta}_{e,(T)}$ correctly summarizes the effects of $\mathbf{X}_T^1$, then (T4) is zero, but again the converse is false – (T4) can be zero in mis-specified models. A well-known example is approximating a high-order autoregressive LDGP for mean zero data with symmetrically distributed errors, by a first-order autoregression, where forecasts are nevertheless unbiased as discussed below for a VAR.

Next, when the data are accurate (especially important at the forecast origin), so $\widetilde{\mathbf{X}} = \mathbf{X}$, then (T5) is zero, but the converse is not entailed: (T5) can be zero just because the data are mean zero.

Continuing, (T6) concerns the estimation error, and arises when $\hat{\boldsymbol{\theta}}_{(T)}$ does not coincide with $\boldsymbol{\theta}_{e,(T)}$. Biases in estimation could, but need not, induce such an effect to be systematic, as might non-linearities in models or LDGPs. When estimated parameters have zero variances, so $\hat{\mathbf{x}}_{T+h|T} = \mathsf{E}_T[\mathbf{x}_{T+h} \mid \cdot, \boldsymbol{\theta}_{e,(T)}]$, then (T6) is zero, and conversely (except for events of probability zero). Otherwise, its main impacts will be on variance terms.

The final term (T7), $\boldsymbol{\varepsilon}_{T+h|T}$, is unlikely to be zero in any social science, although it will have a zero mean by construction, and be unpredictable from the past of the information in use. As with (T6), the main practical impact is through forecast error variances.

The taxonomy in (8) includes elements for the seven main sources of forecast error, partitioning these by whether or not the corresponding expectation is zero. However, several salient features stand out. First, the key distinction between whether the expectations in question are zero or non-zero. In the former case, forecasts will not be systematically biased, and the main impact of any changes or mis-specifications is on higher moments, especially forecast error variances. Conversely, if a non-zero mean error results from any source, systematic forecast errors will ensue. Secondly, and a consequence of the previous remark, some breaks will be easily detected because at whatever point in time they happened, 'in-sample forecasts' immediately after a change will be poor. Equally, others may be hard to detect because they have no impact on the mean forecast errors. Thirdly, the impacts of any transformations of a model on its forecast errors depend on which mistakes have occurred. For example, it is often argued that differencing doubles the forecast-error variance: this is certainly true of $\boldsymbol{\varepsilon}_{T+h|T}$, but is not true in general for $\mathbf{e}_{T+h|T}$. Indeed, it is possible in some circumstances to reduce the forecast-error variance by differencing; see, e.g., Hendry (2005). Finally, the taxonomy applies to any model form, but to clarify some of its implications, we turn to its application to the forecast errors from a VAR.

2.2. VAR model forecast-error taxonomy

We illustrate with a first-order VAR, and for convenience assume the absence of in-sample breaks so that the VAR is initially correctly specified. We also assume that the $n \times 1$ vector of variables $\mathbf{y}_t$ is an $\mathsf{I}(0)$ transformation of the original variables $\mathbf{x}_t$: Section 4.1 considers systems of cointegrated $\mathsf{I}(1)$ variables. Thus,

$$\mathbf{y}_t = \boldsymbol{\phi} + \boldsymbol{\Pi}\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t,$$

with $\boldsymbol{\epsilon}_t \sim \mathsf{IN}_n[\mathbf{0}, \boldsymbol{\Omega}_\epsilon]$, for an in-sample period $t = 1, \ldots, T$. The unconditional mean of $\mathbf{y}_t$ is $\mathsf{E}[\mathbf{y}_t] = (\mathbf{I}_n - \boldsymbol{\Pi})^{-1}\boldsymbol{\phi} \equiv \boldsymbol{\varphi}$, and hence the VAR(1) can be written as

$$\mathbf{y}_t - \boldsymbol{\varphi} = \boldsymbol{\Pi}(\mathbf{y}_{t-1} - \boldsymbol{\varphi}) + \boldsymbol{\epsilon}_t.$$

The h-step ahead forecasts conditional upon period T are given by, for $h = 1, \ldots, H$,

$$\hat{\mathbf{y}}_{T+h} - \hat{\boldsymbol{\varphi}} = \widehat{\boldsymbol{\Pi}}(\hat{\mathbf{y}}_{T+h-1} - \hat{\boldsymbol{\varphi}}) = \widehat{\boldsymbol{\Pi}}^h(\hat{\mathbf{y}}_T - \hat{\boldsymbol{\varphi}}), \tag{9}$$

where $\hat{\boldsymbol{\varphi}} = (\mathbf{I}_n - \widehat{\boldsymbol{\Pi}})^{-1}\hat{\boldsymbol{\phi}}$, and '^'s denote estimators for parameters, and forecasts for random variables. After the forecasts have been made at time T, $(\boldsymbol{\phi}, \boldsymbol{\Pi})$ change to $(\boldsymbol{\phi}^*, \boldsymbol{\Pi}^*)$, where $\boldsymbol{\Pi}^*$ still has all its eigenvalues less than unity in absolute value, so the process remains $\mathsf{I}(0)$. But from $T + 1$ onwards, the data are generated by

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

so both the slope and the intercept may alter. The forecast-error taxonomy for $\hat{\boldsymbol{\epsilon}}_{T+h|T} = \mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h|T}$ is then given by

$$\begin{aligned}\hat{\boldsymbol{\epsilon}}_{T+h|T} \simeq\, & \left(\mathbf{I}_n - \left(\boldsymbol{\Pi}^*\right)^h\right)\left(\boldsymbol{\varphi}^* - \boldsymbol{\varphi}\right) && \text{(ia) equilibrium-mean change} \\ & + \left(\left(\boldsymbol{\Pi}^*\right)^h - \boldsymbol{\Pi}^h\right)(\mathbf{y}_T - \boldsymbol{\varphi}) && \text{(ib) slope change} \\ & + \left(\mathbf{I}_n - \boldsymbol{\Pi}_p^h\right)(\boldsymbol{\varphi} - \boldsymbol{\varphi}_p) && \text{(iia) equilibrium-mean mis-specification} \\ & + \left(\boldsymbol{\Pi}^h - \boldsymbol{\Pi}_p^h\right)(\mathbf{y}_T - \boldsymbol{\varphi}) && \text{(iib) slope mis-specification} \\ & + \left(\boldsymbol{\Pi}_p^h + \mathbf{C}_h\right)(\mathbf{y}_T - \hat{\mathbf{y}}_T) && \text{(iii) forecast-origin uncertainty} \\ & - \left(\mathbf{I}_n - \boldsymbol{\Pi}_p^h\right)(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_p) && \text{(iva) equilibrium-mean estimation} \\ & - \mathbf{F}_h\left(\widehat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}_p\right)^\nu && \text{(ivb) slope estimation} \\ & + \textstyle\sum_{i=0}^{h-1}\left(\boldsymbol{\Pi}^*\right)^i\boldsymbol{\epsilon}_{T+h-i} && \text{(v) error accumulation.}\end{aligned} \tag{10}$$

The matrices $\mathbf{C}_h$ and $\mathbf{F}_h$ are complicated functions of the whole-sample data, the method of estimation, and the forecast-horizon, defined in (A.1) and (A.2) below – see, e.g., Calzolari (1981). $(\cdot)^\nu$ denotes column vectoring, and the subscript p denotes a plim (expected values could be used where these exist). Details of the derivations are given in Clements and Hendry (1999, Chapter 2.9) and are noted for convenience in Appendix A.

This taxonomy conflates some of the distinctions in the general formulation above (e.g., mis-specification of deterministic terms other than intercepts) and distinguishes others (equilibrium-mean and slope estimation effects). Thus, the model mis-specification terms (iia) and (iib) may result from unmodeled in-sample structural change, as in the general taxonomy, but may also arise from the omission of relevant variables, or the imposition of invalid restrictions.

In (10), terms involving $\mathbf{y}_T - \boldsymbol{\varphi}$ have zero expectations even under changed parameters (e.g., (ib) and (iib)). Moreover, for symmetrically-distributed shocks, biases in $\widehat{\mathbf{\Pi}}$ for $\mathbf{\Pi}$ will not induce biased forecasts [see, e.g., Malinvaud (1970), Fuller and Hasza (1980), Hoque, Magnus and Pesaran (1988), and Clements and Hendry (1998) for related results]. The $\boldsymbol{\epsilon}_{T+h}$ have zero means by construction. Consequently, the primary sources of systematic forecast failure are (ia), (iia), (iii), and (iva). However, on ex post evaluation, (iii) will be removed, and in congruent models with freely-estimated intercepts and correctly modeled in-sample breaks, (iia) and (iva) will be zero on average. That leaves changes to the 'equilibrium mean' $\boldsymbol{\varphi}$ (not necessarily the intercept $\boldsymbol{\phi}$ in a model, as seen in (10)), as the primary source of systematic forecast error; see Hendry (2000) for a detailed analysis.

3. Breaks in variance

3.1. Conditional variance processes

The autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982), and its generalizations, are commonly used to model time-varying conditional processes; see, inter alia, Engle and Bollerslev (1987), Bollerslev, Chou and Kroner (1992), and Shephard (1996); and Bera and Higgins (1993) and Baillie and Bollerslev (1992) on forecasting. The forecast-error taxonomy construct can be applied to variance processes. We show that ARCH and GARCH models can in general be solved for long-run variances, so like VARs, are a member of the equilibrium-correction class. Issues to do with the constancy of the long-run variance are then discussed.

The simplest ARCH(1) model for the conditional variance of u_t is $u_t = \eta_t \sigma_t$, where η_t is a standard normal random variable and

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2, \tag{11}$$

where $\omega, \alpha > 0$. Letting $\sigma_t^2 = u_t^2 - v_t$, substituting in (11) gives

$$u_t^2 = \omega + \alpha u_{t-1}^2 + v_t. \tag{12}$$

From $v_t = u_t^2 - \sigma_t^2 = \sigma_t^2(\eta_t^2 - 1)$, $\mathsf{E}[v_t \mid \mathbf{Y}_{t-1}] = \sigma_t^2 \mathsf{E}[(\eta_t^2 - 1) \mid \mathbf{Y}_{t-1}] = 0$, so that the disturbance term $\{v_t\}$ in the AR(1) model (12) is uncorrelated with the regressor, as required. From the AR(1) representation, the condition for covariance stationarity of $\{u_t^2\}$ is $|\alpha| < 1$, whence

$$\mathsf{E}\left[u_t^2\right] = \omega + \alpha \mathsf{E}\left[u_{t-1}^2\right],$$

and so the unconditional variance is

$$\sigma^2 \equiv \mathsf{E}\left[u_t^2\right] = \frac{\omega}{1-\alpha}.$$

Substituting for ω in (11) gives the equilibrium-correction form

$$\sigma_t^2 - \sigma^2 = \alpha\left(u_{t-1}^2 - \sigma^2\right).$$

More generally, for an ARCH(p), $p > 1$,

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_p u_{t-p}^2 \tag{13}$$

provided the roots of $(1 - \alpha_1 z - \alpha_2 z^2 + \cdots + \alpha_p z^p) = 0$ lie outside the unit circle, we can write

$$\sigma_t^2 - \sigma^2 = \alpha_1\left(u_{t-1}^2 - \sigma^2\right) + \alpha_2\left(u_{t-2}^2 - \sigma^2\right) + \cdots + \alpha_p\left(u_{t-p}^2 - \sigma^2\right), \tag{14}$$

where

$$\sigma^2 \equiv \mathsf{E}\left[u_t^2\right] = \frac{\omega}{1 - \alpha_1 - \cdots - \alpha_p}.$$

The generalized ARCH [GARCH; see, e.g., Bollerslev (1986)] process

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{15}$$

also has a long-run solution. The GARCH(1, 1) implies an ARMA(1, 1) for $\{u_t^2\}$. Letting $\sigma_t^2 = u_t^2 - v_t$, substitution into (15) gives

$$u_t^2 = \omega + (\alpha + \beta) u_{t-1}^2 + v_t - \beta v_{t-1}. \tag{16}$$

The process is stationary provided $\alpha + \beta < 1$. When that condition holds

$$\sigma^2 \equiv \mathsf{E}\left[u_t^2\right] = \frac{\omega}{1 - (\alpha + \beta)},$$

and combining the equations for σ_t^2 and σ^2 for the GARCH(1, 1) delivers

$$\sigma_t^2 - \sigma^2 = \alpha\left(u_{t-1}^2 - \sigma^2\right) + \beta\left(\sigma_{t-1}^2 - \sigma^2\right). \tag{17}$$

Thus, the conditional variance responds to the previous period's disequilibria between the conditional variance and the long-run variance and between the squared disturbance and the long-run variance, exhibiting equilibrium-correction type behavior.

3.2. GARCH model forecast-error taxonomy

As it is an equilibrium-correction model, the GARCH(1, 1) is not robust to shifts in σ^2, but may be resilient to shifts in ω, α and β which leave σ^2 unaltered. As an alternative to (17), express the process as

$$\sigma_t^2 = \sigma^2 + \alpha\left(u_{t-1}^2 - \sigma_{t-1}^2\right) + (\alpha + \beta)\left(\sigma_{t-1}^2 - \sigma^2\right). \tag{18}$$

In either (17) or (18), α and β multiply zero-mean terms provided σ^2 is unchanged by any shifts in these parameters. The forecast of next period's volatility based on (18) is given by

$$\hat{\sigma}_{T+1|T}^2 = \hat{\sigma}^2 + \hat{\alpha}\left(\hat{u}_T^2 - \hat{\sigma}_T^2\right) + \left(\hat{\alpha} + \hat{\beta}\right)\left(\hat{\sigma}_T^2 - \hat{\sigma}^2\right) \tag{19}$$

recognizing that $\{\alpha, \beta, \sigma^2\}$ will be replaced by in-sample estimates. The '$\hat{\ }$' on u_T denotes this term is the residual from modeling the conditional mean. When there is little dependence in the mean of the series, such as when $\{u_t\}$ is a financial returns series sampled at a high-frequency, u_T is the observed data series and replaces $\hat{u}_T$ (barring data measurement errors).

Then (19) confronts every problem noted above for forecasts of means: potential breaks in σ^2, α, β, mis-specification of the variance evolution (perhaps an incorrect functional form), estimation uncertainty, etc. The 1-step ahead forecast-error taxonomy takes the following form after a shift in ω, α, β to $\omega^*, \alpha^*, \beta^*$ at T to:

$$\sigma_{T+1}^2 = \sigma^{2*} + \alpha^*\left(u_T^2 - \sigma_T^2\right) + \left(\alpha^* + \beta^*\right)\left(\sigma_T^2 - \sigma^{2*}\right),$$

so that letting the subscript p denote the plim:

$$\begin{aligned}
\sigma_{T+1}^2 - \hat{\sigma}_{T+1|T}^2 \\
= \left(1 - \left(\alpha^* + \beta^*\right)\right)\left(\sigma^{2*} - \sigma^2\right) &\quad \text{[1] long-run mean shift} \\
+ \left(1 - \left(\hat{\alpha} + \hat{\beta}\right)\right)\left(\sigma^2 - \sigma_p^2\right) &\quad \text{[2] long-run mean inconsistency} \\
+ \left(1 - \left(\hat{\alpha} + \hat{\beta}\right)\right)\left(\sigma_p^2 - \hat{\sigma}^2\right) &\quad \text{[3] long-run mean variability} \\
+ \left(\alpha^* - \alpha\right)\left(u_T^2 - \sigma_T^2\right) &\quad \text{[4] } \alpha \text{ shift} \\
+ \left(\alpha - \alpha_p\right)\left(u_T^2 - \sigma_T^2\right) &\quad \text{[5] } \alpha \text{ inconsistency} \\
+ \left(\alpha_p - \hat{\alpha}\right)\left(u_T^2 - \sigma_T^2\right) &\quad \text{[6] } \alpha \text{ variability} \\
+ \hat{\alpha}\left(u_T^2 - \mathsf{E}_T\left[\hat{u}_T^2\right]\right) &\quad \text{[7] impact inconsistency} \\
+ \hat{\alpha}\left(\mathsf{E}_T\left[\hat{u}_T^2\right] - \hat{u}_T^2\right) &\quad \text{[8] impact variability} \\
+ \left[\left(\alpha^* + \beta^*\right) - (\alpha + \beta)\right]\left(\sigma_T^2 - \sigma^2\right) &\quad \text{[9] variance shift} \\
+ \left[(\alpha + \beta) - \left(\alpha_p + \beta_p\right)\right]\left(\sigma_T^2 - \sigma^2\right) &\quad \text{[10] variance inconsistency} \\
+ \left[\left(\alpha_p + \beta_p\right) - \left(\hat{\alpha} + \hat{\beta}\right)\right]\left(\sigma_T^2 - \sigma^2\right) &\quad \text{[11] variance variability} \\
+ \hat{\beta}\left(\sigma_T^2 - \mathsf{E}_T\left[\hat{\sigma}_T^2\right]\right) &\quad \text{[12] } \sigma_T^2 \text{ inconsistency} \\
+ \hat{\beta}\left(\mathsf{E}_T\left[\hat{\sigma}_T^2\right] - \hat{\sigma}_T^2\right) &\quad \text{[13] } \sigma_T^2 \text{ variability.}
\end{aligned} \tag{20}$$

The first term is zero only if no shift occurs in the long-run variance and the second only if a consistent in-sample estimate is obtained. However, the next four terms are zero on average, although the seventh possibly is not. This pattern then repeats, since the next block of four terms again is zero on average, with the penultimate term possibly non-zero, and the last zero on average. As with the earlier forecast error taxonomy, shifts in the mean seem pernicious, whereas those in the other parameters are much less serious contributors to forecast failure in variances. Indeed, even assuming a correct in-sample specification, so terms [2], [5], [7], [10], [12] all vanish, the main error components remain.

4. Forecasting when there are breaks

4.1. Cointegrated vector autoregressions

The general forecast-error taxonomy in Section 2.1 suggests that structural breaks in non-zero mean components are the primary cause of forecast biases. In this section, we examine the impact of breaks in VAR models of cointegrated $\mathsf{I}(1)$ variables, and also analyze models in first differences, because models of this type are commonplace in macroeconomic forecasting. The properties of forecasts made before and after the structural change has occurred are analyzed, where it is assumed that the break occurs close to the forecast origin. As a consequence, the comparisons are made holding the models' parameters constant. The effects of in-sample breaks are identified in the forecast-error taxonomies, and are analyzed in Section 6, where the choice of data window for model estimation is considered. Forecasting in cointegrated VARs (in the absence of breaks) is discussed by Engle and Yoo (1987), Clements and Hendry (1995), Lin and Tsay (1996), and Christoffersen and Diebold (1998), while Clements and Hendry (1996) (on which this section is based) allow for breaks.

The VAR is a closed system so that all non-deterministic variables are forecast within the system. The vector of all n variables is denoted by $\mathbf{x}_t$ and the VAR is assumed to be first-order for convenience:

$$\mathbf{x}_t = \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 t + \boldsymbol{\Upsilon}\mathbf{x}_{t-1} + \boldsymbol{\nu}_t, \tag{21}$$

where $\boldsymbol{\nu}_t \sim \mathsf{IN}_n[\mathbf{0}, \boldsymbol{\Omega}]$, and $\boldsymbol{\tau}_0$ and $\boldsymbol{\tau}_1$ are the vectors of intercepts and coefficients on the time trend, respectively. The system is assumed to be integrated, and to satisfy $r < n$ cointegration relations such that [see, for example, Johansen (1988)]

$$\boldsymbol{\Upsilon} = \mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}',$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times r$ matrices of rank r. Then (21) can be reparametrized as a vector equilibrium-correction model (VECM)

$$\Delta\mathbf{x}_t = \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 t + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\nu}_t. \tag{22}$$

Assuming that $n > r > 0$, the vector $\mathbf{x}_t$ consists of $\mathsf{I}(1)$ variables of which r linear combinations are $\mathsf{I}(0)$. The deterministic components of the stochastic variables $\mathbf{x}_t$ depend on $\boldsymbol{\alpha}$, $\boldsymbol{\tau}_0$ and $\boldsymbol{\tau}_1$. Following Johansen (1994), we can decompose $\boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 t$ as

$$\boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 t = \boldsymbol{\alpha}_\perp \boldsymbol{\zeta}_0 - \boldsymbol{\alpha}\boldsymbol{\lambda}_0 - \boldsymbol{\alpha}\boldsymbol{\lambda}_1 t + \boldsymbol{\alpha}_\perp \boldsymbol{\zeta}_1 t, \tag{23}$$

where $\boldsymbol{\lambda}_i = -(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\tau}_i$ and $\boldsymbol{\zeta}_i = (\boldsymbol{\alpha}'_\perp \boldsymbol{\alpha}_\perp)^{-1}\boldsymbol{\alpha}'_\perp \boldsymbol{\tau}_i$ with $\boldsymbol{\alpha}'\boldsymbol{\alpha}_\perp = \mathbf{0}$, so that $\boldsymbol{\alpha}\boldsymbol{\lambda}_i$ and $\boldsymbol{\alpha}_\perp \boldsymbol{\zeta}_i$ are orthogonal by construction. The condition that $\boldsymbol{\alpha}_\perp \boldsymbol{\zeta}_1 = \mathbf{0}$ rules out quadratic trends in the levels of the variables, and we obtain

$$\Delta \mathbf{x}_t = \boldsymbol{\alpha}_\perp \boldsymbol{\zeta}_0 + \boldsymbol{\alpha}\big(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\lambda}_0 - \boldsymbol{\lambda}_1 t\big) + \boldsymbol{\nu}_t. \tag{24}$$

It is sometimes more convenient to parameterize the deterministic terms so that the system growth rate $\boldsymbol{\gamma} = \mathsf{E}[\Delta \mathbf{x}_t]$ is explicit, so in the following we will adopt

$$\Delta \mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\alpha}\big(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1 t\big) + \boldsymbol{\nu}_t, \tag{25}$$

where one can show that $\boldsymbol{\gamma} = \boldsymbol{\alpha}_\perp \boldsymbol{\zeta}_0 + \boldsymbol{\alpha}\boldsymbol{\psi}$, $\boldsymbol{\mu}_0 = \boldsymbol{\psi} + \boldsymbol{\lambda}_0$ and $\boldsymbol{\mu}_1 = \boldsymbol{\lambda}_1$ with $\boldsymbol{\psi} = (\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}(\boldsymbol{\lambda}_1 - \boldsymbol{\beta}'\boldsymbol{\alpha}_\perp \boldsymbol{\zeta}_0)$ and $\boldsymbol{\beta}'\boldsymbol{\gamma} = \boldsymbol{\mu}_1$.

Finally, a VAR in differences (DVAR) may be used, which within sample is misspecified relative to the VECM unless $r = 0$. The simplest is

$$\Delta \mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\eta}_t, \tag{26}$$

so when $\boldsymbol{\alpha} = \mathbf{0}$, the VECM and DVAR coincide. In practice, lagged $\Delta \mathbf{x}_t$ may be used to approximate the omitted cointegrating vectors.

4.2. VECM forecast errors

We now consider dynamic forecasts and their errors under structural change, abstracting from the other sources of error identified in the taxonomy, such as parameter-estimation error. A number of authors have looked at the effects of parameter estimation on forecast-error moments [including, inter alia, Schmidt (1974, 1977), Calzolari (1981, 1987), Bianchi and Calzolari (1982), and Lütkepohl (1991)]. The j-step ahead forecasts for the levels of the process given by $\hat{\mathbf{x}}_{T+j|T} = \mathsf{E}_T[\mathbf{x}_{T+j} \mid \mathbf{x}_T]$ for $j = 1, \dots, H$ are

$$\hat{\mathbf{x}}_{T+j|T} = \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1(T+j) + \boldsymbol{\Upsilon}\hat{\mathbf{x}}_{T+j-1|T} = \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^i \boldsymbol{\tau}(i) + \boldsymbol{\Upsilon}^j \mathbf{x}_T, \tag{27}$$

where we let $\boldsymbol{\tau}_0 + \boldsymbol{\tau}_1(T+j-i) = \boldsymbol{\tau}(i)$ for notational convenience, with forecast errors $\hat{\boldsymbol{\nu}}_{T+j|T} = \mathbf{x}_{T+j} - \hat{\mathbf{x}}_{T+j|T}$. Consider a one-off change of $(\boldsymbol{\tau}_0 : \boldsymbol{\tau}_1 : \boldsymbol{\Upsilon})$ to $(\boldsymbol{\tau}_0^* : \boldsymbol{\tau}_1^* : \boldsymbol{\Upsilon}^*)$ which occurs either at period T (before the forecast is made) or at period $T+1$ (after the forecast is made), but with the variance, autocorrelation, and distribution of the disturbance term remaining unaltered. Then the data generated by the process for the next H periods is given by

$$\mathbf{x}_{T+j} = \boldsymbol{\tau}_0^* + \boldsymbol{\tau}_1^*(T+j) + \boldsymbol{\Upsilon}^*\mathbf{x}_{T+j-1} + \boldsymbol{\nu}_{T+j}$$

$$= \sum_{i=0}^{j-1} (\boldsymbol{\Upsilon}^*)^i \boldsymbol{\tau}^*(i) + \sum_{i=0}^{j-1} (\boldsymbol{\Upsilon}^*)^i \boldsymbol{\nu}_{T+j-i} + (\boldsymbol{\Upsilon}^*)^j \mathbf{x}_T. \tag{28}$$

Thus, the j-step ahead forecast error can be written as

$$\hat{\boldsymbol{\nu}}_{T+j|T} = \left(\sum_{i=0}^{j-1} (\boldsymbol{\Upsilon}^*)^i \boldsymbol{\tau}^*(i) - \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^i \boldsymbol{\tau}(i) \right) + \sum_{i=0}^{j-1} (\boldsymbol{\Upsilon}^*)^i \boldsymbol{\nu}_{T+j-i} + \left((\boldsymbol{\Upsilon}^*)^j - \boldsymbol{\Upsilon}^j \right) \mathbf{x}_T. \tag{29}$$

The expectation of the j-step forecast error conditional on $\mathbf{x}_T$ is

$$\mathsf{E}\left[\hat{\boldsymbol{\nu}}_{T+j|T} \mid \mathbf{x}_T \right] = \left(\sum_{i=0}^{j-1} (\boldsymbol{\Upsilon}^*)^i \boldsymbol{\tau}^*(i) - \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^i \boldsymbol{\tau}(i) \right) + \left((\boldsymbol{\Upsilon}^*)^j - \boldsymbol{\Upsilon}^j \right) \mathbf{x}_T \tag{30}$$

so that the conditional forecast error variance is

$$\mathsf{V}\left[\hat{\boldsymbol{\nu}}_{T+j|T} \mid \mathbf{x}_T \right] = \sum_{i=0}^{j-1} (\boldsymbol{\Upsilon}^*)^i \boldsymbol{\Omega} (\boldsymbol{\Upsilon}^*)^{i\prime}.$$

We now consider a number of special cases where only the deterministic components change. With the assumption that $\boldsymbol{\Upsilon}^* = \boldsymbol{\Upsilon}$, we obtain

$$\begin{aligned} \mathsf{E}[\hat{\boldsymbol{\nu}}_{T+j|T}] &= \mathsf{E}[\hat{\boldsymbol{\nu}}_{T+j|T} \mid \mathbf{x}_T] \\ &= \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^i \left(\left[\boldsymbol{\tau}_0^* + \boldsymbol{\tau}_1^*(T+j-i) \right] - \left[\boldsymbol{\tau}_0 + \boldsymbol{\tau}_1(T+j-i) \right] \right) \\ &= \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^i \left[(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}) + \boldsymbol{\alpha}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0^*) + \boldsymbol{\alpha}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_1^*)(T+j-i) \right], \end{aligned} \tag{31}$$

so that the conditional and unconditional biases are the same. The bias is increasing in j due to the shift in $\boldsymbol{\gamma}$ (the first term in square brackets) whereas the impacts of the shifts in $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ eventually level off because

$$\lim_{i \to \infty} \boldsymbol{\Upsilon}^i = \mathbf{I}_n - \boldsymbol{\alpha} (\boldsymbol{\beta}' \boldsymbol{\alpha})^{-1} \boldsymbol{\beta}' \equiv \mathbf{K},$$

and $\mathbf{K}\boldsymbol{\alpha} = \mathbf{0}$. When the linear trend is absent and the constant term can be restricted to the cointegrating space (i.e., $\boldsymbol{\tau}_1 = \mathbf{0}$ and $\boldsymbol{\zeta}_0 = \mathbf{0}$, which implies $\boldsymbol{\lambda}_1 = \mathbf{0}$ and therefore $\boldsymbol{\mu}_1 = \boldsymbol{\gamma} = \mathbf{0}$), then only the second term appears, and the bias is $O(1)$ in j. The formulation in (31) assumes that $\boldsymbol{\Upsilon}$, and therefore the cointegrating space, remains unaltered. Moreover, the coefficient on the linear trend alters but still lies in the cointegrating space. Otherwise, after the structural break, $\mathbf{x}_t$ would be propelled by quadratic trends.

4.3. DVAR forecast errors

Consider the forecasts from a simplified DVAR. Forecasts from the DVAR for $\Delta\mathbf{x}_t$ are defined by setting $\Delta\mathbf{x}_{T+j}$ equal to the population growth rate $\boldsymbol{\gamma}$,

$$\Delta\tilde{\mathbf{x}}_{T+j} = \boldsymbol{\gamma} \tag{32}$$

so that j-step ahead forecasts of the level of the process are obtained by integrating (32) from the initial condition $\mathbf{x}_T$,

$$\tilde{\mathbf{x}}_{T+j} = \tilde{\mathbf{x}}_{T+j-1} + \boldsymbol{\gamma} = \mathbf{x}_T + j\boldsymbol{\gamma} \quad \text{for } j = 1, \ldots, H. \tag{33}$$

When $\boldsymbol{\Upsilon}$ is unchanged over the forecast period, the expected value of the conditional j-step ahead forecast error $\tilde{\boldsymbol{\nu}}_{T+j|T}$ is

$$\mathsf{E}[\tilde{\boldsymbol{\nu}}_{T+j|T} \mid \mathbf{x}_T] = \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^i \left[\boldsymbol{\tau}_0^* + \boldsymbol{\tau}_1^*(T + j - i)\right] - j\boldsymbol{\gamma} + \left(\boldsymbol{\Upsilon}^j - \mathbf{I}_n\right)\mathbf{x}_T. \tag{34}$$

By averaging over $\mathbf{x}_T$ we obtain the unconditional bias $\mathsf{E}[\tilde{\boldsymbol{\nu}}_{T+j}]$.

Appendix B records the algebra for the derivation of (35):

$$\mathsf{E}[\tilde{\boldsymbol{\nu}}_{T+j|T}] = j\left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}\right) + \mathbf{A}_j\boldsymbol{\alpha}\left[\left(\boldsymbol{\mu}_0^a - \boldsymbol{\mu}_0^*\right) - \boldsymbol{\beta}'\left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}^a\right)(T + 1)\right]. \tag{35}$$

In the same notation, the VECM results from (31) are

$$\mathsf{E}[\hat{\boldsymbol{\nu}}_{T+j|T}] = j\left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}\right) + \mathbf{A}_j\boldsymbol{\alpha}\left[\left(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0^*\right) - \boldsymbol{\beta}'\left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}\right)(T + 1)\right]. \tag{36}$$

Thus, (36) and (35) coincide when $\boldsymbol{\mu}_0^a = \boldsymbol{\mu}_0$, and $\boldsymbol{\gamma}^a = \boldsymbol{\gamma}$ as will occur if either there is no structural change, or the change occurs after the start of the forecast period.

4.4. Forecast biases under location shifts

We now consider a number of interesting special cases of (35) and (36) which highlight the behavior of the DVAR and VECM under shifts in the deterministic terms. Viewing $(\boldsymbol{\tau}_0, \boldsymbol{\tau}_1)$ as the primary parameters, we can map changes in these parameters to changes in $(\boldsymbol{\gamma}, \boldsymbol{\mu}_0, \boldsymbol{\mu}_1)$ via the orthogonal decomposition into $(\boldsymbol{\zeta}_0, \boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1)$. The interdependencies can be summarized as $\boldsymbol{\gamma}(\boldsymbol{\zeta}_0, \boldsymbol{\lambda}_1)$, $\boldsymbol{\mu}_0(\boldsymbol{\zeta}_0, \boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1)$, $\boldsymbol{\mu}_1(\boldsymbol{\lambda}_1)$.

Case I: $\boldsymbol{\tau}_0^* = \boldsymbol{\tau}_0$, $\boldsymbol{\tau}_1^* = \boldsymbol{\tau}_1$. In the absence of structural change, $\boldsymbol{\mu}_0^a = \boldsymbol{\mu}_0$ and $\boldsymbol{\gamma}^a = \boldsymbol{\gamma}$ and so

$$\mathsf{E}[\hat{\boldsymbol{\nu}}_{T+j|T}] = \mathsf{E}[\tilde{\boldsymbol{\nu}}_{T+j|T}] = \mathbf{0} \tag{37}$$

as is evident from (35) and (36). The omission of the stationary $\mathsf{I}(0)$ linear combinations does not render the DVAR forecasts biased.

Case II: $\boldsymbol{\tau}_0^* \neq \boldsymbol{\tau}_0$, $\boldsymbol{\tau}_1^* = \boldsymbol{\tau}_1$, but $\boldsymbol{\zeta}_0^* = \boldsymbol{\zeta}_0$. Then $\boldsymbol{\mu}_0^* \neq \boldsymbol{\mu}_0$ but $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}$:

$$\mathsf{E}[\hat{\boldsymbol{\nu}}_{T+j|T}] = \mathbf{A}_j\boldsymbol{\alpha}\left(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0^*\right), \tag{38}$$

$$\mathsf{E}[\tilde{\boldsymbol{\nu}}_{T+j|T}] = \mathbf{A}_j \boldsymbol{\alpha}\left(\boldsymbol{\mu}_0^a - \boldsymbol{\mu}_0^*\right). \tag{39}$$

The biases are equal if $\boldsymbol{\mu}_0^a = \boldsymbol{\mu}_0$; i.e., the break is after the forecast origin. However, $\mathsf{E}[\tilde{\boldsymbol{\nu}}_{T+j}] = \mathbf{0}$ when $\boldsymbol{\mu}_0^a = \boldsymbol{\mu}_0^*$, and hence the DVAR is unbiased when the break occurs prior to the commencement of forecasting. In this example the component of the constant term orthogonal to $\boldsymbol{\alpha}$ ($\boldsymbol{\zeta}_0$) is unchanged, so that the growth rate is unaffected.

Case III: $\boldsymbol{\tau}_0^* \neq \boldsymbol{\tau}_0$, $\boldsymbol{\tau}_1^* = \boldsymbol{\tau}_1$ (as in Case II), but now $\boldsymbol{\lambda}_0^* = \boldsymbol{\lambda}_0$ which implies $\boldsymbol{\zeta}_0^* \neq \boldsymbol{\zeta}_0$ and therefore $\boldsymbol{\mu}_0^* \neq \boldsymbol{\mu}_0$ and $\boldsymbol{\gamma}^* \neq \boldsymbol{\gamma}$. However, $\boldsymbol{\beta}'\boldsymbol{\gamma}^* = \boldsymbol{\beta}'\boldsymbol{\gamma}$ holds (because $\boldsymbol{\tau}_1^* = \boldsymbol{\tau}_1$) so that

$$\mathsf{E}[\hat{\boldsymbol{\nu}}_{T+j|T}] = j\left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}\right) + \mathbf{A}_j \boldsymbol{\alpha}\left(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0^*\right), \tag{40}$$

$$\mathsf{E}[\tilde{\boldsymbol{\nu}}_{T+j|T}] = j\left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}\right) + \mathbf{A}_j \boldsymbol{\alpha}\left(\boldsymbol{\mu}_0^a - \boldsymbol{\mu}_0^*\right). \tag{41}$$

Consequently, the errors coincide when $\boldsymbol{\mu}_0^a = \boldsymbol{\mu}_0$, but differ when $\boldsymbol{\mu}_0^a = \boldsymbol{\mu}_0^*$.

Case IV: $\boldsymbol{\tau}_0^* = \boldsymbol{\tau}_0$, $\boldsymbol{\tau}_1^* \neq \boldsymbol{\tau}_1$. All of $\boldsymbol{\mu}_0$, $\boldsymbol{\mu}_1$ and $\boldsymbol{\gamma}$ change. If $\boldsymbol{\beta}'\boldsymbol{\gamma}^* \neq \boldsymbol{\beta}'\boldsymbol{\gamma}$ then we have (35) and (36), and otherwise the biases of Case III.

4.5. Forecast biases when there are changes in the autoregressive parameters

By way of contrast, changes in autoregressive parameters that do not induce changes in means are relatively benign for forecasts of first moments. Consider the VECM forecast errors given by (29) when $\mathsf{E}[\mathbf{x}_t] = \mathbf{0}$ for all t, so that $\boldsymbol{\tau}_0 = \boldsymbol{\tau}_0^* = \boldsymbol{\tau}_1 = \boldsymbol{\tau}_1^* = \mathbf{0}$ in (21):

$$\hat{\boldsymbol{\nu}}_{T+j|T} = \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^{*i} \boldsymbol{\nu}_{T+j-i} + \left(\boldsymbol{\Upsilon}^{*j} - \boldsymbol{\Upsilon}^j\right)\mathbf{x}_T. \tag{42}$$

The forecasts are unconditionally unbiased, $\mathsf{E}[\hat{\boldsymbol{\nu}}_{T+j|T}] = \mathbf{0}$, and the effect of the break is manifest in higher forecast error variances

$$\mathsf{V}[\hat{\boldsymbol{\nu}}_{T+j|T} \mid \mathbf{x}_T] = \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^{*i} \boldsymbol{\Omega} \boldsymbol{\Upsilon}^{*i\prime} + \left(\boldsymbol{\Upsilon}^{*j} - \boldsymbol{\Upsilon}^j\right)\mathbf{x}_T \mathbf{x}_T' \left(\boldsymbol{\Upsilon}^{*j} - \boldsymbol{\Upsilon}^j\right)'.$$

The DVAR model forecasts are also unconditionally unbiased, from

$$\tilde{\boldsymbol{\nu}}_{T+j|T} = \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^{*i} \boldsymbol{\nu}_{T+j-i} + \left(\boldsymbol{\Upsilon}^{*j} - \mathbf{I}_n\right)\mathbf{x}_T,$$

since $\mathsf{E}[\tilde{\boldsymbol{\nu}}_{T+j|T}] = \mathbf{0}$ provided $\mathsf{E}[\mathbf{x}_T] = \mathbf{0}$.

When $\mathsf{E}[\mathbf{x}_T] \neq \mathbf{0}$, but is the same before and after the break (as when changes in the autoregressive parameters are offset by changes in intercepts) both models' forecast errors are unconditionally unbiased.

4.6. Univariate models

The results for $n = 1$ follow immediately as a special case of (21):

$$x_t = \tau_0 + \tau_1 t + \Upsilon x_{t-1} + \nu_t. \tag{43}$$

The forecasts from (43) and the 'unit-root' model $x_t = x_{t-1} + \gamma + \upsilon_t$ are unconditionally unbiased when Υ shifts provided $\mathsf{E}[x_t] = 0$ (requiring $\tau_0 = \tau_1 = 0$). When $\tau_1 = 0$, the unit-root model forecasts remain unbiased when τ_0 shifts provided the shift occurs prior to forecasting, demonstrating the greater adaptability of the unit-root model. As in the multivariate setting, the break is assumed not to affect the model parameters (so that γ is taken to equal its population value of zero).

5. Detection of breaks

5.1. Tests for structural change

In this section, we briefly review testing for structural change or non-constancy in the parameters of time-series regressions. There is a large literature on testing for structural change. See, for example, Stock (1994) for a review. Two useful distinctions can be drawn: whether the putative break point is known, and whether the change in the parameters is governed by a stochastic process. Section 8 considers tests against the alternative of non-linearity.

For a known break date, the traditional method of testing for a one-time change in the model's parameters is the Chow (1960) test. That is, in the model

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \varepsilon_t \tag{44}$$

when the alternative is a one-off change:

$$\mathsf{H}_1(\pi)\colon \boldsymbol{\alpha} = \begin{cases} \boldsymbol{\alpha}_1(\pi) & \text{for } t = 1, 2, \ldots, \pi T, \\ \boldsymbol{\alpha}_2(\pi) & \text{for } t = \pi T + 1, \ldots, T, \end{cases}$$

where $\boldsymbol{\alpha}' = (\alpha_1\ \alpha_2 \ldots \alpha_p)$, $\pi \in (0, 1)$, a test of parameter constancy can be implemented as an LM, Wald or LR test, all of which are asymptotically equivalent. For example, the Wald test has the form

$$\mathsf{F}_T(\pi) = \frac{RSS_{1,T} - (RSS_{1,\pi T} + RSS_{\pi T+1,T})}{(RSS_{1,\pi T} + RSS_{\pi T+1,T})/(T - 2p)},$$

where $RSS_{1,T}$ is the 'restricted' residual sum of squares from estimating the model on all the observations, $RSS_{1,\pi T}$ is the residual sum of squares from estimating the model on observations 1 to πT, etc. These tests also apply when the model is not purely autoregressive but contains other explanatory variables, although for $\mathsf{F}_T(\pi)$ to be asymptotically chi-squared all the variables need to be $\mathsf{I}(0)$ in general.

When the break is not assumed known a priori, the testing procedure cannot take the break date π as given. The testing procedure is then non-standard, because π is identified under the alternative hypothesis but not under the null [Davies (1977, 1987)]. Quandt (1960) suggested taking the maximal $\mathsf{F}_T(\pi)$ over a range of values of $\pi \in \Pi$, for Π a pre-specified subset of (0, 1). Andrews (1993) extended this approach to non-linear models, and Andrews and Ploberger (1994) considered the 'average' and 'exponential' test statistics. The asymptotic distributions are tabulated by Andrews (1993), and depend on p and Π. Diebold and Chen (1996) consider bootstrap approximations to the finite-sample distributions.

Andrews (1993) shows that the sup tests have power against a broader range of alternatives than $\mathsf{H}_1(\pi)$, but will not have high power against 'structural change' caused by the omission of a stationary variable. For example, suppose the DGP is a stationary AR(2):

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t$$

and the null is $\phi_{1,t} = \phi_{1,0}$ for all t in the model $y_t = \phi_{1,t} y_{t-1} + \varepsilon_t$, versus H_1^*: $\phi_{1,t}$ varies with t. The omission of the second lag can be viewed as causing structural change in the model each period, but this will not be detectable as the model is stationary under the alternative for all $t = 1, \ldots, T$. Stochastic forms of model mis-specification of this sort were shown in Section 2.1 not to cause forecast bias.

In addition, Bai and Perron (1998) consider testing for multiple structural breaks, and Bai, Lumsdaine and Stock (1998) consider testing and estimating break dates when the breaks are common to a number of time series. Hendry, Johansen and Santos (2004) propose testing for this form of non-constancy by adding a complete set of impulse indicators to a model using a two-step process, and establish the null distribution in a location-scale IID distribution.

Tests for structural change can also be based on recursive coefficient estimates and recursive residuals. The CUSUM test of Brown, Durbin and Evans (1975) is based on the cumulation of the sequence of 1-step forecast errors obtained by recursively estimating the model. As shown by Krämer, Ploberger and Alt (1988) and discussed by Stock (1994), the CUSUM test only has local asymptotic power against breaks in non-zero mean regressors. Therefore, CUSUM test rejections are likely to signal more specific forms of change than the sup tests. Unlike sup tests, CUSUM tests will not have good local asymptotic power against $\mathsf{H}_1(\pi)$ when (44) does not contain an intercept (so that y_t is zero-mean).

As well as testing for 'non-stochastic' structural change, one can test for randomly time-varying coefficients. Nyblom (1989) tests against the alternative that the coefficients follow a random walk, and Breusch and Pagan (1979) against the alternative that the coefficients are random draws from a distribution with a constant mean and finite variance.

From a forecasting perspective, in-sample tests of parameter instability may be used in a number of ways. The finding of instability may guide the selection of the window of data to be used for model estimation, or lead to the use of rolling windows of observations to allow for gradual change, or to the adoption of more flexible models, as discussed in Sections 6 and 7.

As argued by Chu, Stinchcombe and White (1996), the 'one shot' tests discussed so far may not be ideal in a real-time forecasting context as new data accrue. The tests are designed to detect breaks on a given historical sample of a fixed size. Repeated application of the tests as new data becomes available, or repeated application retrospectively moving through the historical period, will result in the asymptotic size of the sequence of tests approaching one if the null rejection frequency is held constant. Chu, Stinchcombe and White (1996, p. 1047) illustrate with reference to the Ploberger, Krämer and Kontrus (1989) retrospective fluctuation test. In the simplest case that $\{Y_t\}$ is an independent sequence, the null of 'stability in mean' is H_0: $\mathsf{E}[Y_t] = 0$, $t = 1, 2, \ldots$ versus H_1: $\mathsf{E}[Y_t] \neq 0$ for some t. For a given n,

$$FL_n = \max_{k<n} \sigma_0^{-1} \sqrt{n}(k/n) \left| \frac{1}{k} \sum_{t=1}^{k} y_t \right|$$

is compared to a critical value c determined from the hitting probability of a Brownian motion. But if FL_n is implemented sequentially for $n+1, n+2, \ldots$ then the probability of a type 1 error is one asymptotically. Similarly if a Chow test is repeatedly calculated every time new observations become available.

Chu, Stinchcombe and White (1996) suggest monitoring procedures for CUSUM and parameter fluctuation tests where the critical values are specified as boundary functions such that they are crossed with the prescribed probability under H_0. The CUSUM implementation is as follows. Define

$$\widetilde{Q}_n^m = \hat{\sigma}^{-1} \sum_{i=m}^{m+n} \omega_i,$$

where m is the end of the historical period, so that monitoring starts at $m+1$, and $n \geqslant 1$. The ω_i are the recursive residuals, $\omega_i = \hat{\varepsilon}_i / \sqrt{\upsilon_i}$, where $\hat{\varepsilon}_i = y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_{i-1}$, and

$$\upsilon_i = 1 + \mathbf{x}_i' \left(\sum_{j=1}^{i-1} \mathbf{x}_j \mathbf{x}_j' \right)^{-1} \mathbf{x}_i,$$

with

$$\hat{\boldsymbol{\beta}}_i = \left(\sum_{j=1}^{i} \mathbf{x}_j \mathbf{x}_j' \right)^{-1} \left(\sum_{j=1}^{i} \mathbf{x}_j y_j \right),$$

for the model

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + \varepsilon_t,$$

where $\mathbf{x}_t$ is $k \times 1$, say, and $\mathbf{X}_j = (\mathbf{x}_1 \ldots \mathbf{x}_j)$, etc. $\hat{\sigma}^2$ is a consistent estimator of $\mathsf{E}[\varepsilon_t^2] = \sigma^2$. The boundary is given by

$$\sqrt{n+m-k}\sqrt{c + \ln\left(\frac{n+m-k}{m-k}\right)}$$

(where c depends on the size of the test). Hence, beginning with $n = 1$, $|\widetilde{Q}_n^m|$ is compared to the boundary, and so on for $n = 2, n = 3$, etc. until $|\widetilde{Q}_n^m|$ crosses the boundary, signalling a rejection of the null hypothesis H_0: $\boldsymbol{\beta}_t = \boldsymbol{\beta}$ for $t = n+1, n+2, \ldots$. As for the one-shot tests, rejection of the null may lead to an attempt to revise the model or the adoption of a more 'adaptable' model.

5.2. *Testing for level shifts in ARMA models*

In addition to the tests for structural change in regression models, the literature on the detection of outliers and level shifts in ARMA models [following on from Box and Jenkins (1976)] is relevant from a forecasting perspective; see, inter alia, Tsay (1986, 1988), Chen and Tiao (1990), Chen and Liu (1993), Balke (1993), Junttila (2001), and Sánchez and Peña (2003). In this tradition, ARMA models are viewed as being composed of a 'regular component' and possibly a component which represents anomalous exogenous shifts. The latter can be either outliers or permanent shifts in the level of the process. The focus of the literature is on the problems caused by outliers and level shifts on the identification and estimation of the ARMA model, viz., the regular component of the model. The correct identification of level shifts will have an important bearing on forecast performance. Methods of identifying the type and estimating the timing of the exogenous shifts are aimed at 'correcting' the time series prior to estimating the ARMA model, and often follow an iterative procedure. That is, the exogenous shifts are determined conditional on a given ARMA model, the data are then corrected and the ARMA model re-estimated, etc.; see Tsay (1988) [Balke (1993) provides a refinement] and Chen and Liu (1993) for an approach that jointly estimates the ARMA model and exogenous shifts.

Given an ARMA model

$$y_t = f(t) + \left[\theta(L)/\phi(L)\right]\varepsilon_t,$$

where $\varepsilon_t \sim \mathsf{IN}[0, \sigma_\varepsilon^2]$, $\theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q$, $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$, then $[\theta(L)/\phi(L)]\varepsilon_t$ is the regular component. For a single exogenous shift, let

$$f(t) = \omega_0 \left[\frac{\omega(L)}{\delta(L)}\right]\xi_t^{(d)},$$

where $\xi_t^{(d)} = 1$ when $t = d$ and $\xi_t^{(d)} = 0$ when $t \neq d$. The lag polynomials $\omega(L)$ and $\delta(L)$ define the type of exogenous event. $\omega(L)/\delta(L) = 1$ corresponds to an additive outlier (AO), whereby y_d is ω_0 higher than would be the case were the exogenous component absent. When $\omega(L)/\delta(L) = \theta(L)/\phi(L)$, we have an innovation outlier (IO).

The model can be written as

$$y_t = \frac{\theta(L)}{\phi(L)}\bigl(\varepsilon_t + \omega_0 \xi_t^{(d)}\bigr),$$

corresponding to the period d innovation being drawn from a Gaussian distribution with mean ω_0. Of particular interest from a forecasting perspective is when $\omega(L)/\delta(L) = (1-L)^{-1}$, which represents a permanent level shift (LS):

$$\begin{aligned} y_t &= \bigl[\theta(L)\big/\phi(L)\bigr]\varepsilon_t, && t < d,\\ y_t - \omega_0 &= \bigl[\theta(L)\big/\phi(L)\bigr]\varepsilon_t, && t \geqslant d. \end{aligned}$$

Letting $\pi(L) = \phi(L)/\theta(L)$, we obtain the following residual series for the three specifications of $f(t)$:

$$\begin{aligned} &\text{IO:} && e_t = \pi(L)y_t = \omega_0 \xi_t^{(d)} + \varepsilon_t,\\ &\text{AO:} && e_t = \pi(L)y_t = \omega_0 \pi(L)\xi_t^{(d)} + \varepsilon_t,\\ &\text{LS:} && e_t = \pi(L)y_t = \omega_0 \pi(L)(1-L)^{-1}\xi_t^{(d)} + \varepsilon_t. \end{aligned}$$

Hence the least-squares estimate of an IO at $t = d$ can be obtained by regressing e_t on $\xi_t^{(d)}$: this yields $\hat{\omega}_{0,\mathrm{IO}} = e_t$. Similarly, the least-squares estimate of an AO at $t = d$ can be obtained by regressing e_t on a variable that is zero for $t < d$, 1 for $t = d$, $-\pi_k$ for $t = d + k$, $k > 1$, to give $\hat{\omega}_{0,\mathrm{AO}}$. Similarly for LS.

The standardized statistics:

$$\begin{aligned} &\text{IOs:} && \tau_{\mathrm{IO}}(d) = \hat{\omega}_{0,\mathrm{IO}}(d)/\hat{\sigma}_\varepsilon,\\ &\text{AOs:} && \tau_{\mathrm{AO}}(d) = \bigl(\hat{\omega}_{0,\mathrm{AO}}(d)/\hat{\sigma}_\varepsilon\bigr)\sqrt{\textstyle\sum_{t=d}^{T}\bigl(\pi(L)\xi_t^{(d)}\bigr)^2},\\ &\text{LSs:} && \tau_{\mathrm{LS}}(d) = \bigl(\hat{\omega}_{0,\mathrm{LS}}(d)/\hat{\sigma}_\varepsilon\bigr)\sqrt{\textstyle\sum_{t=d}^{T}\bigl(\pi(L)(1-L)^{-1}\xi_t^{(d)}\bigr)^2} \end{aligned}$$

are discussed by Chan and Wei (1988) and Tsay (1988). They have approximately normal distributions. Given that d is unknown, as is the type of the shift, the suggestion is to take:

$$\tau_{\max} = \max\{\tau_{\mathrm{IO},\max}, \tau_{\mathrm{AO},\max}, \tau_{\mathrm{LS},\max}\},$$

where $\tau_{j,\max} = \max_{1 \leqslant d \leqslant T}\{\tau_j(d)\}$, and compare this to a pre-specified critical value. Exceedance implies an exogenous shift has occurred.

As $\phi(L)$ and $\theta(L)$ are unknown, these tests require a pre-estimate of the ARMA model. Balke (1993) notes that when level shifts are present, the initial ARMA model will be mis-specified, and that this may lead to level shifts being identified as IOs, as well as reducing the power of the tests of LS.

Suppose $\phi(L) = 1 - \phi L$ and $\theta(L) = 1$, so that we have an AR(1), then in the presence of an unmodeled level shift of size μ at time d, the estimate of ϕ is inconsistent:

$$\operatorname*{plim}_{T\to\infty} \hat{\phi} = \phi + \left[\frac{(1-\phi)\mu^2(T-d)d/T^2}{\sigma_\varepsilon^2/(1-\phi^2) + \mu^2(T-d)d/T^2}\right]; \tag{45}$$

see, e.g., Rappoport and Reichlin (1989), Reichlin (1989), Chen and Tiao (1990), Perron (1990), and Hendry and Neale (1991). Neglected structural breaks will give the appearance of unit roots. Balke (1993) shows that the expected value of the $\tau_{\text{LS}}(d)$ statistic will be substantially reduced for many combinations of values of the underlying parameters, leading to a reduction in power.

The consequences for forecast performance are less clear-cut. The failure to detect structural breaks in the mean of the series will be mitigated to some extent by the induced 'random-walk-like' property of the estimated ARMA model. An empirical study by Junttila (2001) finds that intervention dummies do not result in the expected gains in terms of forecast performance when applied to a model of Finnish inflation.

With this background, we turn to detecting the breaks themselves when these occur in-sample.

6. Model estimation and specification

6.1. Determination of estimation sample for a fixed specification

We assume that the break date is known, and consider the choice of the estimation sample. In practice the break date will need to be estimated, and this will often be given as a by-product of testing for a break at an unknown date, using one of the procedures reviewed in Section 5. The remaining model parameters are estimated, and forecasts generated, conditional on the estimated break point(s); see, e.g., Bai and Perron (1998).[2] Consequently, the properties of the forecast errors will depend on the pre-test for the break date. In the absence of formal frequentist analyses of this problem, we act as if the break date were known.[3]

Suppose the DGP is given by

$$y_{t+1} = 1_{(t\leqslant\tau)}\boldsymbol{\beta}_1'\mathbf{x}_t + (1 - 1_{(t\leqslant\tau)})\boldsymbol{\beta}_2'\mathbf{x}_t + u_{t+1} \tag{46}$$

so that the pre-break observations are $t = 1, \ldots, \tau$, and the post-break $t = \tau+1, \ldots, T$. There is a one-off change in all the slope parameters and the disturbance variance, from σ_1^2 to σ_2^2.

[2] In the context of assessing the predictability of stock market returns, Pesaran and Timmermann (2002a) choose an estimation window by determining the time of the most recent break using reversed ordered CUSUM tests. The authors also determine the latest break using the method in Bai and Perron (1998).

[3] Pastor and Stambaugh (2001) adopt a Bayesian approach that incorporates uncertainty about the locations of the breaks, so their analysis does not treat estimates of breakpoints as true values and condition upon them.

First, we suppose that the explanatory variables are strongly exogenous. Pesaran and Timmermann (2002b) consider the choice of m, the first observation for the model estimation period, where $m = \tau + 1$ corresponds to only using post-break observations. Let $\mathbf{X}_{m,T}$ be the $(T - m + 1) \times k$ matrix of observations on the k explanatory variables for the periods m to T (inclusive), $\mathbf{Q}_{m,T} = \mathbf{X}'_{m,T}\mathbf{X}_{m,T}$, and $\mathbf{Y}_{m,T}$ and $\mathbf{u}_{m,T}$ contain the latest $T - m + 1$ observations on y and u, respectively. The OLS estimator of $\boldsymbol{\beta}$ in

$$\mathbf{Y}_{m,T} = \mathbf{X}_{m,T}\boldsymbol{\beta}(m) + \mathbf{v}_{m,T}$$

is given by

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_T(m) &= \mathbf{Q}_{m,T}^{-1}\mathbf{X}'_{m,T}\mathbf{Y}_{m,T} \\
&= \mathbf{Q}_{m,T}^{-1}\left(\mathbf{X}'_{m,\tau} : \mathbf{X}'_{\tau+1,T}\right)\begin{pmatrix}\mathbf{Y}_{m,\tau} \\ \mathbf{Y}_{\tau+1,T}\end{pmatrix} \\
&= \mathbf{Q}_{m,T}^{-1}\mathbf{Q}_{m,\tau}\boldsymbol{\beta}_1 + \mathbf{Q}_{m,T}^{-1}\mathbf{Q}_{\tau+1,T}\boldsymbol{\beta}_2 + \mathbf{Q}_{m,T}^{-1}\mathbf{X}'_{m,T}\mathbf{u}_{m,T},
\end{aligned}$$

where, e.g., $\mathbf{Q}_{m,\tau}$ is the second moment matrix formed from $\mathbf{X}_{m,\tau}$, etc. Thus $\hat{\boldsymbol{\beta}}_T(m)$ is a weighted average of the pre and post-break parameter vectors. The forecast error is

$$\begin{aligned}
e_{T+1} &= y_{T+1} - \hat{\boldsymbol{\beta}}_T(m)'\mathbf{x}_T \\
&= u_{T+1} + (\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1)'\mathbf{Q}_{m,\tau}\mathbf{Q}_{m,T}^{-1}\mathbf{x}_T - \mathbf{u}'_{m,T}\mathbf{X}_{m,T}\mathbf{Q}_{m,T}^{-1}\mathbf{x}_T, \qquad (47)
\end{aligned}$$

where the second term is the bias that results from using pre-break observations, which depends on the size of the shift $\boldsymbol{\delta}_{\boldsymbol{\beta}} = (\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1)$, amongst other things. The conditional MSFE is

$$\begin{aligned}
\mathsf{E}\left[e_{T+1}^2 \mid \mathcal{I}_T\right] &= \sigma_2^2 + \left(\boldsymbol{\delta}'_{\boldsymbol{\beta}}\mathbf{Q}_{m,\tau}\mathbf{Q}_{m,T}^{-1}\mathbf{x}_T\right)^2 \\
&\quad + \mathbf{x}'_T\mathbf{Q}_{m,T}^{-1}\mathbf{X}'_{m,T}\mathbf{D}_{m,T}\mathbf{X}_{m,T}\mathbf{Q}_{m,T}^{-1}\mathbf{x}_T, \qquad (48)
\end{aligned}$$

where $\mathbf{D}_{m,T} = \mathsf{E}[\mathbf{u}_{m,T}\mathbf{u}'_{m,T}]$, a diagonal matrix with σ_1^2 in the first $\tau - m + 1$ elements, and σ_2^2 in the remainder. When $\sigma_2^2 = \sigma_1^2 = \sigma^2$ (say), $\mathbf{D}_{m,T}$ is proportional to the identity matrix, and the conditional MSFE simplifies to

$$\mathsf{E}\left[e_{T+1}^2 \mid \mathcal{I}_T\right] = \sigma^2 + \left(\boldsymbol{\delta}'_{\boldsymbol{\beta}}\mathbf{Q}_{m,\tau}\mathbf{Q}_{m,T}^{-1}\mathbf{x}_T\right)^2 + \sigma^2\mathbf{x}'_T\mathbf{Q}_{m,T}^{-1}\mathbf{x}_T.$$

Using only post-break observations corresponds to setting $m = \tau + 1$. Since $\mathbf{Q}_{m,\tau} = \mathbf{0}$ when $m > \tau$, from (48) we obtain

$$\mathsf{E}\left[e_{T+1}^2 \mid \mathcal{I}_T\right] = \sigma_2^2 + \sigma_2^2\left(\mathbf{x}'_T\mathbf{Q}_{\tau+1,T}^{-1}\mathbf{x}_T\right)$$

since $\mathbf{D}_{\tau+1,T} = \sigma_2^2\mathbf{I}_{T-\tau}$.

Pesaran and Timmermann (2002b) consider $k = 1$ so that

$$e_{T+1} = u_{T+1} + (\beta_2 - \beta_1)\theta_m x_T - v_m x_T, \qquad (49)$$

where

$$\theta_m = \frac{Q_{m,\tau}}{Q_{m,T}} = \frac{\sum_{t=m}^{\tau} x_{t-1}^2}{\sum_{t=m}^{T} x_{t-1}^2} \quad \text{and} \quad v_m = \mathbf{u}'_{m,T}\mathbf{X}_{m,T} Q_{m,T}^{-1} = \frac{\sum_{t=m}^{T} u_t x_{t-1}}{\sum_{t=m}^{T} x_{t-1}^2}.$$

Then the conditional MSFE has a more readily interpretable form:

$$\mathsf{E}\big[e_{T+1}^2 \mid \mathcal{I}_T\big] = \sigma_2^2 + \sigma_2^2 x_T^2 \left(\sigma_2^2 \delta_\beta^2 \theta_m^2 + \frac{\psi\theta_m + 1}{\sum_{t=m}^{T} x_{t-1}^2}\right),$$

where $\psi = (\sigma_1^2 - \sigma_2^2)/\sigma_2^2$. So decreasing m (including more pre-break observations) increases θ_m and therefore the squared bias (via $\sigma_2^2\delta_\beta^2\theta_m^2$) but the overall effect on the MSFE is unclear.

Including some pre-break observations is more likely to lower the MSFE the smaller the break, $|\delta_\beta|$; when the variability increases after the break period, $\sigma_2^2 > \sigma_1^2$, and the fewer the number of post-break observations (the shorter the distance $T - \tau$). Given that it is optimal to set $m < \tau + 1$, the optimal window size m^* is chose to satisfy

$$m^* = \underset{m=1,\ldots,\tau+1}{\operatorname{argmin}} \big\{\mathsf{E}\big[e_{T+1}^2 \mid \mathcal{I}_T\big]\big\}.$$

Unconditionally (i.e., on average across all values of x_t) the forecasts are unbiased for all m when $\mathsf{E}[x_t] = 0$. From (49):

$$\mathsf{E}[e_{T+1} \mid \mathcal{I}_T] = (\beta_2 - \beta_1)\theta_m x_T - v_m x_T \tag{50}$$

so that

$$\mathsf{E}[e_{T+1}] = \mathsf{E}\big(\mathsf{E}[e_{T+1} \mid \mathcal{I}_T]\big) = (\beta_2 - \beta_1)\theta_m \mathsf{E}[x_T] - v_m \mathsf{E}[x_T] = 0. \tag{51}$$

The unconditional MSFE is given by

$$\mathsf{E}\big[e_{T+1}^2\big] = \sigma^2 + \omega^2(\beta_2 - \beta_1)^2 \frac{\nu_1(\nu_1 + 2)}{\nu(\nu + 2)} + \frac{\sigma^2}{\nu - 2}$$

for conditional mean breaks ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) with zero-mean regressors, and where $\mathsf{E}[x_t^2] = \omega^2$ and $\nu_1 = \tau - m + 1$, $\nu = T - m + 1$.

The assumption that $\mathbf{x}_t$ is distributed independently of all the disturbances $\{u_t,\ t = 1, \ldots, T\}$ does not hold for autoregressive models. The forecast error remains unconditionally unbiased when the regressors are zero-mean, as is evident with $\mathsf{E}[x_t] = 0$ in the case of $k = 1$ depicted in Equation (51), and consistent with the forecast-error taxonomy in Section 2.1. Pesaran and Timmermann (2003) show that including pre-break observations is more likely to improve forecasting performance than in the case of fixed regressors because of the finite small-sample biases in the estimates of the parameters of autoregressive models. They conclude that employing an expanding window of data may often be as good as employing a rolling window when there are breaks. Including pre-break observations is more likely to reduce MSFEs when the degree of persistence of the AR process declines after the break, and when the mean of the process is unaffected. A reduction in the degree of persistence may favor the use of pre-break observations by offsetting the small-sample bias. The small-sample bias of the AR parameter in the AR(1) model is negative:

$$\mathsf{E}\big[\hat{\beta}_1\big] - \beta_1 = \frac{-(1+3\beta_1)}{T} + O\big(T^{-3/2}\big)$$

so that the estimate of β_1 based on post-break observations is on average below the true value. The inclusion of pre-break observations will induce a positive bias (relative to the true post-break value, β_2). When the regressors are fixed, finite-sample biases are absent and the inclusion of pre-break observations will cause bias, other things being equal. Also see Chong (2001).

6.2. *Updating*

Rather than assuming that the break has occurred some time in the past, suppose that the change happens close to the time that the forecasts are made, and may be of a continuous nature. In these circumstances, parameter estimates held fixed for a sequence of forecast origins will gradually depart from the underlying LDGP approximation. A moving window seeks to offset that difficulty by excluding distant observations, whereas updating seeks to 'chase' the changing parameters: more flexibly, 'updating' could allow for re-selecting the model specification as well as re-estimating its parameters. Alternatively, the model's parameters may be allowed to 'drift'. An assumption sometimes made in the empirical macro literature is that VAR parameters evolve as driftless random walks (with zero-mean, constant-variance Gaussian innovations) subject to constraints that rule out the parameters drifting into non-stationary regions [see Cogley and Sargent (2001, 2005) for recent examples]. In modeling the equity premium, Pastor and Stambaugh (2001) allow for parameter change by specifying a process that alternates between 'stable' and 'transition' regimes. In their Bayesian approach, the timing of the break points that define the regimes is uncertain, but the use of prior beliefs based on economics (e.g., the relationship between the equity premium and volatility, and with price changes) allows the current equity premium to be estimated. The next section notes some other approaches where older observations are down weighted, or when only the last few data points play a role in the forecast (as with double-differenced devices).

Here we note that there is evidence of the benefits of jointly re-selecting the model specification and re-estimating its resulting parameters in Phillips (1994, 1995, 1996), Schiff and Phillips (2000), and Swanson and White (1997), for example. However, Stock and Watson (1996) find that the forecasting gains from time-varying coefficient models appear to be rather modest. In a constant parameter world, estimation efficiency dictates that all available information should be incorporated, so updating as new data accrue is natural. Moreover, following a location shift, re-selection could allow an additional unit root to be estimated to eliminate the break, and thereby reduce systematic forecast failure, as noted at the end of Section 5.2; also see Osborn (2002, pp. 420–421) for a related discussion in a seasonal context.

7. Ad hoc forecasting devices

When there are structural breaks, forecasting methods which adapt quickly following the break are most likely to avoid making systematic forecast errors in sequential real-time forecasting. Using the tests for structural change discussed in Section 5, Stock and Watson (1996) find evidence of widespread instability in the postwar US univariate and bivariate macroeconomic relations that they study. A number of authors have noted that empirical-accuracy studies of univariate time-series forecasting models and methods often favor ad hoc forecasting devices over properly specified statistical models [in this context, often the ARMA models of Box and Jenkins (1976)].[4] One explanation is the failure of the assumption of parameter constancy, and the greater adaptivity of the forecasting devices. Various types of exponential smoothing (ES), such as damped trend ES [see Gardner and McKenzie (1985)], tend to be competitive with ARMA models, although it can be shown that ES only corresponds to the optimal forecasting device for a specific ARMA model, namely the ARIMA(0, 1, 1) [see, for example, Harvey (1992, Chapter 2)]. In this section, we consider a number of ad hoc forecasting methods and assess their performance when there are breaks. The roles of parameter estimation updating, rolling windows and time-varying parameter models have been considered in Sections 6.1 and 6.2.

7.1. Exponential smoothing

We discuss exponential smoothing for variance processes, but the points made are equally relevant for forecasting conditional means. The ARMA(1, 1) equation for u_t^2 for the GARCH(1, 1) indicates that the forecast function will be closely related to exponential smoothing. Equation (17) has the interpretation that the conditional variance will exceed the long-run (or unconditional) variance if last period's squared returns exceed the long-run variance and/or if last period's conditional variance exceeds the unconditional. Some straightforward algebra shows that the long-horizon forecasts approach σ^2. Writing (17) for σ_{T+j}^2, we have

$$\begin{aligned}\sigma_{T+j}^2 - \sigma^2 &= \alpha\left(u_{T+j-1}^2 - \sigma^2\right) + \beta\left(\sigma_{T+j-1}^2 - \sigma^2\right)\\ &= \alpha\left(\sigma_{T+j-1}^2 \boldsymbol{\nu}_{T+j-1}^2 - \sigma^2\right) + \beta\left(\sigma_{T+j-1}^2 - \sigma^2\right).\end{aligned}$$

Taking conditional expectations

$$\begin{aligned}\sigma_{T+j|T}^2 - \sigma^2 &= \alpha\left(\mathsf{E}\left[\sigma_{T+j-1}^2 \boldsymbol{\nu}_{T+j-1}^2 \mid \mathbf{Y}_T\right] - \sigma^2\right) + \beta\left(\mathsf{E}\left[\sigma_{T+j-1}^2 \mid \mathbf{Y}_T\right] - \sigma^2\right)\\ &= (\alpha+\beta)\left(\mathsf{E}\left[\sigma_{T+j-1}^2 \mid \mathbf{Y}_T\right] - \sigma^2\right)\end{aligned}$$

[4] One of the earliest studies was Newbold and Granger (1974). Fildes and Makridakis (1995) and Fildes and Ord (2002) report on the subsequent 'M-competitions', Makridakis and Hibon (2000) present the latest 'M-competition', and a number of commentaries appear in *International Journal of Forecasting* 17.

using

$$\mathsf{E}\big[\sigma_{T+j-1}^2 \boldsymbol{\nu}_{T+j-1}^2 \mid \mathbf{Y}_T\big] = \mathsf{E}\big[\sigma_{T+j-1}^2 \mid \mathbf{Y}_T\big]\mathsf{E}\big[\boldsymbol{\nu}_{T+j-1}^2 \mid \mathbf{Y}_T\big] = \mathsf{E}\big[\sigma_{T+j-1}^2 \mid \mathbf{Y}_T\big],$$

for $j > 2$. By backward substitution ($j > 0$),

$$\begin{aligned}\sigma_{T+j|T}^2 - \sigma^2 &= (\alpha+\beta)^{j-1}\big(\sigma_{T+1}^2 - \sigma^2\big) \\ &= (\alpha+\beta)^{j-1}\big[\alpha\big(u_T^2 - \sigma^2\big) + \beta\big(\sigma_T^2 - \sigma^2\big)\big] \end{aligned} \tag{52}$$

(given $\mathsf{E}[\sigma_{T+1}^2 \mid \mathbf{Y}_T] = \sigma_{T+1}^2$). Therefore $\sigma_{T+j|T}^2 \to \sigma^2$ as $j \to \infty$.

Contrast the EWMA formula for forecasting $T+1$ based on $\mathbf{Y}_T$:

$$\begin{aligned}\tilde{\sigma}_{T+1|T}^2 &= \frac{1}{\sum_{s=0}^{\infty}\lambda^s}\big(u_T^2 + \lambda u_{T-1}^2 + \lambda^2 u_{T-2}^2 + \cdots\big) \\ &= (1-\lambda)\sum_{s=0}^{\infty}\lambda^s u_{T-s}^2, \end{aligned} \tag{53}$$

where $\lambda \in (0, 1)$, so the largest weight is given to the most recent squared return, $(1-\lambda)$, and thereafter the weights decline exponentially. Rearranging gives

$$\tilde{\sigma}_{T+1|T}^2 = u_T^2 + \lambda\big(\tilde{\sigma}_{T|T-1}^2 - u_T^2\big).$$

The forecast is equal to the squared return plus/minus the difference between the estimate of the current-period variance and the squared return. Exponential smoothing corresponds to a restricted GARCH(1, 1) model with $\omega = 0$ and $\alpha + \beta = (1-\lambda) + \lambda = 1$. From a forecasting perspective, these restrictions give rise to an ARIMA(0, 1, 1) for u_t^2 (see (16)). As an integrated process, the latest volatility estimate is extrapolated, and there is no mean-reversion. Thus the exponential smoother will be more robust than the GARCH(1, 1) model's forecasts to breaks in σ^2 when λ is close to zero: there is no tendency for a sequence of 1-step forecasts to move toward a long-run variance. When σ^2 is constant (i.e., when there are no breaks in the long-run level of volatility) and the conditional variance follows an 'equilibrium' GARCH process, this will be undesirable, but in the presence of shifts in σ^2 may avoid the systematic forecast errors from a GARCH model correcting to an inappropriate equilibrium.

Empirically, the estimated value of $\alpha + \beta$ in (15) is often found to be close to 1, and estimates of ω close to 0. $\alpha + \beta = 1$ gives rise to the Integrated GARCH (IGARCH) model. The IGARCH model may arise through the neglect of structural breaks in GARCH models, paralleling the impact of shifts in autoregressive models of means, as summarized in (45). For a number of daily stock return series, Lamoureux and Lastrapes (1990) test standard GARCH models against GARCH models which allow for structural change through the introduction of a number of dummy variables, although Maddala and Li (1996) question the validity of their bootstrap tests.

7.2. Intercept corrections

The widespread use of some macro-econometric forecasting practices, such as intercept corrections (or residual adjustments), can be justified by structural change. Published forecasts based on large-scale macro-econometric models often include adjustments for the influence of anticipated events that are not explicitly incorporated in the specification of the model. But in addition, as long ago as Marris (1954), the 'mechanistic' adherence to models in the generation of forecasts when the economic system changes was questioned. The importance of adjusting purely model-based forecasts has been recognized by a number of authors [see, inter alia, Theil (1961, p. 57), Klein (1971), Klein, Howrey and MacCarthy (1974), and the sequence of reviews by the UK ESRC Macroeconomic Modelling Bureau in Wallis et al. (1984, 1985, 1986, 1987), Turner (1990), and Wallis and Whitley (1991)]. Improvements in forecast performance after intercept correction (IC) have been documented by Wallis et al. (1986, Table 4.8, 1987, Figures 4.3 and 4.4) and Wallis and Whitley (1991), inter alia.

To illustrate the effects of IC on the properties of forecasts, consider the simplest adjustment to the VECM forecasts in Section 4.2, whereby the period T residual $\hat{\boldsymbol{\nu}}_T = \mathbf{x}_T - \hat{\mathbf{x}}_T = (\boldsymbol{\tau}_0^* - \boldsymbol{\tau}_0) + (\boldsymbol{\tau}_1^* - \boldsymbol{\tau}_1)T + \boldsymbol{\nu}_T$ is used to adjust subsequent forecasts. Thus, the adjusted forecasts are given by

$$\dot{\mathbf{x}}_{T+h} = \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1(T+h) + \boldsymbol{\Upsilon}\dot{\mathbf{x}}_{T+h-1} + \hat{\boldsymbol{\nu}}_T, \tag{54}$$

where $\dot{\mathbf{x}}_T = \mathbf{x}_T$, so that

$$\dot{\mathbf{x}}_{T+h} = \hat{\mathbf{x}}_{T+h} + \sum_{i=0}^{h-1} \boldsymbol{\Upsilon}^i \hat{\boldsymbol{\nu}}_T = \hat{\mathbf{x}}_{T+h} + \mathbf{A}_h \hat{\boldsymbol{\nu}}_T. \tag{55}$$

Letting $\hat{\boldsymbol{\nu}}_{T+h}$ denote the h-step ahead forecast error of the unadjusted forecast, $\hat{\boldsymbol{\nu}}_{T+h} = \mathbf{x}_{T+h} - \hat{\mathbf{x}}_{T+h}$, the conditional (and unconditional) expectation of the adjusted-forecast error is

$$\mathsf{E}[\dot{\boldsymbol{\nu}}_{T+h} \mid \mathbf{x}_T] = \mathsf{E}[\hat{\boldsymbol{\nu}}_{T+h} - \mathbf{A}_h \hat{\boldsymbol{\nu}}_T] = [h\mathbf{A}_h - \mathbf{D}_h]\big(\boldsymbol{\tau}_1^* - \boldsymbol{\tau}_1\big), \tag{56}$$

where we have used

$$\mathsf{E}[\hat{\boldsymbol{\nu}}_T] = \big(\boldsymbol{\tau}_0^* - \boldsymbol{\tau}_0\big) + \big(\boldsymbol{\tau}_1^* - \boldsymbol{\tau}_1\big)T.$$

The adjustment strategy yields unbiased forecasts when $\boldsymbol{\tau}_1^* = \boldsymbol{\tau}_1$ irrespective of any shift in $\boldsymbol{\tau}_0$. Even if the process remains unchanged there is no penalty in terms of bias from intercept correcting. The cost of intercept correcting is in terms of increased uncertainty. The forecast error variance for the type of IC discussed here is

$$\mathsf{V}[\dot{\boldsymbol{\nu}}_{T+h}] = 2\mathsf{V}[\hat{\boldsymbol{\nu}}_{T+h}] + \sum_{j=0}^{h-1}\sum_{i=0}^{h-1} \boldsymbol{\Upsilon}^j \boldsymbol{\Omega} \boldsymbol{\Upsilon}^{i\prime}, \quad j \neq i, \tag{57}$$

which is more than double the conditional expectation forecast error variance, $\mathsf{V}[\hat{\boldsymbol{\nu}}_{T+h} \mid \mathbf{x}_T]$. Clearly, there is a bias-variance trade-off: bias can be reduced at the cost of an inflated forecast-error variance. Notice also that the second term in (57) is of the order of h^2, so that this trade-off should be more favorable to intercept correcting at short horizons. Furthermore, basing ICs on averages of recent errors (rather than the period T error alone) may provide more accurate estimates of the break and reduce the inflation of the forecast-error variance. For a sufficiently large change in $\boldsymbol{\tau}_0$, the adjusted forecasts will be more accurate than those of unadjusted forecasts on squared-error loss measures. Detailed analyses of ICs can be found in Clements and Hendry (1996, 1998, Chapter 8, 1999, Chapter 6).

7.3. Differencing

Section 4.3 considered the forecast performance of a DVAR relative to a VECM when there were location shifts in the underlying process. Those two models are related by the DVAR omitting the disequilibrium feedback of the VECM, rather than by a differencing operator transforming the model used to forecast [see, e.g., Davidson et al. (1978)]. For shifts in the equilibrium mean at the end of the estimation sample, the DVAR could outperform the VECM. Nevertheless, both models were susceptible to shifts in the growth rate. Thus, a natural development is to consider differencing once more, to obtain a DDVAR and a DVECM, neither of which includes any deterministic terms when linear deterministic trends are the highest needed to characterize data.

The detailed algebra is presented in Hendry (2005), who shows that the simplest double-differenced forecasting device, namely:

$$\Delta^2 \mathbf{x}_{T+1|T} = \mathbf{0} \tag{58}$$

can outperform in a range of circumstances, especially if the VECM omits important explanatory variables and experiences location shifts. Indeed, the forecast-error variance of (58) need not be doubled by differencing, and could even be less than that of the VECM, so (58) would outperform in both mean and variance. In that setting, the DVECM will also do well, as (in the simplest case again) it augments (58) by $\boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{T-1}$ which transpires to be the most important observable component missing in (58), provided the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ do not change. For example, consider (25) when $\boldsymbol{\mu}_1 = \mathbf{0}$, then differencing all the terms in the VECM but retaining their parameter estimates unaltered delivers

$$\Delta^2 \mathbf{x}_t = \Delta\boldsymbol{\gamma} + \boldsymbol{\alpha}\Delta\left(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}_0\right) + \boldsymbol{\xi}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{t-1} + \boldsymbol{\xi}_t. \tag{59}$$

Then (59) has no deterministic terms, so does not equilibrium correct, thereby reducing the risks attached to forecasting after breaks. Although it will produce noisy forecasts, smoothed variants are easily formulated. When there are no locations shifts, the 'insurance' of differencing must worsen forecast accuracy and precision, but if location shifts occur, differencing will pay.

7.4. Pooling

Forecast pooling is a venerable ad hoc method of improving forecasts; see, inter alia, Bates and Granger (1969), Newbold and Granger (1974), Granger (1989), and Clements and Galvão (2005); Diebold and Lopez (1996) and Newbold and Harvey (2002) provide surveys, and Clemen (1989) an annotated bibliography. Combining individual forecasts of the same event has often been found to deliver a smaller MSFE than any of the individual forecasts. Simple rules for combining forecasts, such as averages, tend to work as well as more elaborate rules based on past forecasting performance; see Stock and Watson (1999) and Fildes and Ord (2002). Hendry and Clements (2004) suggest that such an outcome may sometimes result from location shifts in the DGP differentially affecting different models at different times. After each break, some previously well-performing model does badly, certainly much worse than the combined forecast, so eventually the combined forecast dominates on MSFE, even though at each point in time, it was never the best.

An improved approach might be obtained by trying to predict which device is most likely to forecast best at the relevant horizon, but the unpredictable nature of many breaks makes its success unlikely – unless the breaks themselves can be forecast. In particular, during quiescent periods, the DDV will do poorly, yet will prove a robust predictor when a sudden change eventuates. Indeed, encompassing tests across models would reveal the DDV to be dominated over ‘normal’ periods, so it cannot be established that dominated models should be excluded from the pooling combination.

Extensions to combining density and interval forecasts have been proposed by, e.g., Granger, White and Kamstra (1989), Taylor and Bunn (1998), Wallis (2005), and Hall and Mitchell (2005), inter alia.

8. Non-linear models

In previous sections, we have considered structural breaks in parametric linear dynamic models. The break is viewed as a permanent change in the value of the parameter vector. Non-linear models are characterized by dynamic properties that vary between two or more regimes, or states, in a way that is endogenously determined by the model. For example, non-linear models have been used extensively in empirical macroeconomics to capture differences in dynamic behavior between the expansion and contraction phases of the business cycle, and have also been applied to financial time series [see, inter alia, Albert and Chib (1993), Diebold, Lee and Weinbach (1994), Goodwin (1993), Hamilton (1994), Kähler and Marnet (1994), Kim (1994), Krolzig and Lütkepohl (1995), Krolzig (1997), Lam (1990), McCulloch and Tsay (1994), Phillips (1991), Potter (1995), and Tiao and Tsay (1994), as well as the collections edited by Barnett et al. (2000), and Hamilton and Raj (2002)]. Treating a number of episodes of parameter instability in a time series as non-random events representing permanent changes in the model will have different implications for characterizing and understanding the behavior of the time series, as well as for forecasting, compared to treating the time series as being governed by a non-linear model. Forecasts from non-linear models will depend on the phase of the business cycle and will incorporate the possibility of a switch in regime during the period being forecast, while forecasts from structural break models imply no such changes during the future.[5]

Given the possibility of parameter instability due to non-linearities, the tests of parameter instability in linear dynamic models (reviewed in Section 5) will be misleading if non-linearities cause rejections. Similarly, tests of non-linearities against the null of a linear model may be driven by structural instabilities. Carrasco (2002) addresses these issues, and we outline some of her main findings in Section 8.1. Noting the difficulties of comparing non-linear and structural break models directly using classical techniques, Koop and Potter (2000) advocate a Bayesian approach.

In Section 8.2, we compare forecasts from a non-linear model with those from a structural break model.

8.1. Testing for non-linearity and structural change

The structural change (SC) and two non-linear regime-switching models can be cast in a common framework as

$$\begin{aligned} y_t = {} & (\mu_0 + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p}) \\ & + \left(\mu_0^* + \alpha_1^* y_{t-1} + \cdots + \alpha_p^* y_{t-p}\right) s_t + \varepsilon_t, \end{aligned} \tag{60}$$

where ε_t is $\mathsf{IID}[0, \sigma^2]$ and s_t is the indicator variable. When $s_t = 1$ ($t \geqslant \tau$), we have an SC model in which potentially all the mean parameters undergo a one-off change at some exogenous date, τ. The first non-linear model is the Markov-switching model (MS). In the MS model, s_t is an unobservable and exogenously determined Markov chain. In the 2-regime case, s_t takes the values of 1 and 0, defined by the transition probabilities

$$p_{ij} = \Pr(s_{t+1} = j \mid s_t = i), \quad \sum_{j=0}^{1} p_{ij} = 1, \ \forall i, j \in \{0, 1\}. \tag{61}$$

The assumption of fixed transition probabilities p_{ij} can be relaxed [see, e.g., Diebold, Rudebusch and Sichel (1993), Diebold, Lee and Weinbach (1994), Filardo (1994), Lahiri and Wang (1994), and Durland and McCurdy (1994)] and the model can be generalized to allow more than two states [e.g., Clements and Krolzig (1998, 2003)].

The second non-linear model is a self-exciting threshold autoregressive model [SETAR; see, e.g., Tong (1983, 1995)] for which $s_t = 1_{(y_{t-d} \leqslant r)}$, where d is a posi-

[5] Pesaran, Pettenuzzo and Timmermann (2004) use a Bayesian approach to allow for structural breaks over the forecast period when a variable has been subject to a number of distinct regimes in the past. Longer horizon forecasts tend to be generated from parameters drawn from the 'meta distribution' rather than those that characterize the latest regime.

tive integer. That is, the regime depends on the value of the process d periods earlier relative to a threshold r.

In Section 5, we noted that testing for a structural break is complicated by the structural break date τ being unknown – the timing of the change is a nuisance parameter which is unidentified under the null that $[\mu_0^* \; \alpha_1^* \; \ldots \; \alpha_p^*]' = \mathbf{0}$. For both the MS and SETAR models, there are also nuisance parameters which are unidentified under the null of linearity. For the MS model, these are the transition probabilities $\{p_{ij}\}$, and for the SETAR model, the value of the threshold, r. Testing procedures for non-linear models against the null of linearity have been developed by Chan (1990, 1991), Hansen (1992, 1996a), Garcia (1998), and Hansen (1996b).

The main findings of Carrasco (2002) can be summarized as:

(a) Tests of SC will have no power when the process is stationary, as in the case of the MS and SETAR models [see Andrews (1993)] – this is demonstrated for the 'sup' tests.

(b) Tests of SETAR non-linearity will have asymptotic power of one when the process is SC or MS (or SETAR), but only power against local alternatives which are $T^{1/4}$, rather than the usual $T^{1/2}$.

Thus, tests of SC will not be useful in detecting parameter instability due to non-linearity, whilst testing for SETAR non-linearity might be viewed as a portmanteau pre-test of instability. Tests of SETAR non-linearity will not be able to detect small changes.

8.2. Non-linear model forecasts

Of the two non-linear models, only the MS model minimum MSFE predictor can be derived analytically, and we focus on forecasting with this model.[6] To make matters concrete, consider the original Hamilton (1989) model of the US business cycle. This posits a fourth-order ($p = 4$) autoregression for the quarterly percentage change in US real GNP $\{y_t\}$ from 1953 to 1984:

$$y_t - \mu(s_t) = \alpha_1\big(y_{t-1} - \mu(s_{t-1})\big) + \cdots + \alpha_4\big(y_{t-4} - \mu(s_{t-4})\big) + u_t, \tag{62}$$

where $\varepsilon_t \sim \mathsf{IN}[0, \sigma_\varepsilon^2]$ and

$$\mu(s_t) = \begin{cases} \mu_1 > 0 & \text{if } s_t = 1 \text{ ('expansion' or 'boom')}, \\ \mu_0 < 0 & \text{if } s_t = 0 \text{ ('contraction' or 'recession')}. \end{cases} \tag{63}$$

[6] Exact analytical solutions are not available for multi-period forecasts from SETAR models. Exact numerical solutions require sequences of numerical integrations [see, e.g., Tong (1995, §4.2.4 and §6.2)] based on the Chapman–Kolmogorov relation. As an alternative, one might use Monte Carlo or bootstrapping [e.g., Tiao and Tsay (1994) and Clements and Smith (1999)], particularly for high-order autoregressions, or the normal forecast-error method (NFE) suggested by Al-Qassam and Lane (1989) for the exponential-autoregressive model, and adapted by De Gooijer and De Bruin (1997) to forecasting with SETAR models. See also Chapter 8 by Teräsvirta in this Handbook.

Relative to (60), $[\alpha_1^* \ \ldots \ \alpha_p^*] = \mathbf{0}$ so that the autoregressive dynamics are constant across regimes, and when $p = 0$ (no autoregressive dynamics) $\mu_0 + \mu_0^*$ in (60) is equal to μ_1. The model (62) has a switching mean rather than intercept, so that for $p > 0$ the correspondence between the two sets of 'deterministic' terms is more complicated. Maximum likelihood estimation of the model is by the EM algorithm [see Hamilton (1990)].[7]

To obtain the minimum MSFE h-step predictor, we take the conditional expectation of y_{T+h} given $\mathbf{Y}_T = \{y_T, y_{T-1}, \ldots\}$. Letting $\hat{y}_{T+j|T} = \mathsf{E}[y_{T+j} \mid \mathbf{Y}_T]$ gives rise to the recursion

$$\hat{y}_{T+h|T} = \hat{\mu}_{T+h|T} + \sum_{k=1}^{4} \alpha_k (\hat{y}_{T+h-k|T} - \hat{\mu}_{T+h-k|T}) \tag{64}$$

with $\hat{y}_{T+h|T} = y_{T+h}$ for $h \leqslant 0$ and where the predicted mean is given by

$$\hat{\mu}_{T+h|T} = \sum_{j=1}^{2} \mu_j \Pr(s_{T+h} = j \mid \mathbf{Y}_T). \tag{65}$$

The predicted regime probabilities

$$\Pr(s_{T+h} = j \mid \mathbf{Y}_T) = \sum_{i=0}^{1} \Pr(s_{T+h} = j \mid s_T = i) \Pr(s_T = i \mid \mathbf{Y}_T)$$

only depend on the transition probabilities $\Pr(s_{T+h} = j \mid s_{T+h-1} = i) = p_{ij}$, $i, j = 0, 1$, and the filtered regime probabilities $\Pr(s_T = i \mid \mathbf{Y}_T)$ [see, e.g., Hamilton (1989, 1990, 1993, 1994) for details].

The optimal predictor of the MS-AR model is linear in the last p observations and the last regime inference. The optimal forecasting rule becomes linear in the limit when $\Pr(s_t \mid s_{t-1}) = \Pr(s_t)$ for $s_t, s_{t-1} = 0, 1$, since then $\Pr(s_{T+h} = j \mid \mathbf{Y}_T) = \Pr(s_t = j)$ and from (65), $\hat{\mu}_{T+h} = \mu_y$, the unconditional mean of y_t. Then

$$\hat{y}_{T+h|T} = \mu_y + \sum_{k=1}^{4} \alpha_k (\hat{y}_{T+h-k|T} - \mu_y), \tag{66}$$

so to a first approximation, apart from differences arising from parameter estimation, forecasts will match those from linear autoregressive models.

Further insight can be obtained by writing the MS process $y_t - \mu(s_t)$ as the sum of two independent processes:

$$y_t - \mu_y = \mu_t + z_t,$$

[7] The EM algorithm of Dempster, Laird and Rubin (1977) is used because the observable time series depends on the s_t, which are unobservable stochastic variables.

such that $\mathsf{E}[\mu_t] = \mathsf{E}[z_t] = 0$. Assuming $p = 1$ for convenience, z_t is

$$z_t = \alpha z_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right],$$

a linear autoregression with Gaussian disturbances. μ_t represents the contribution of the Markov chain:

$$\mu_t = (\mu_2 - \mu_1)\zeta_t,$$

where $\zeta_t = 1 - \Pr(s_t = 0)$ if $s_t = 0$ and $-\Pr(s_t = 0)$ otherwise. $\Pr(s_t = 0) = p_{10}/(p_{10} + p_{01})$ is the unconditional probability of regime 0. Using the autoregressive representation of a Markov chain:

$$\zeta_t = (p_{11} + p_{00} - 1)\zeta_{t-1} + v_t,$$

then predictions of the hidden Markov chain are given by

$$\hat{\zeta}_{T+h|T} = (p_{11} + p_{00} - 1)^h \hat{\zeta}_{T|T},$$

where $\hat{\zeta}_{T|T} = \mathsf{E}[\zeta_T \mid \mathbf{Y}_T] = \Pr(s_T = 0 \mid \mathbf{Y}_T) - \Pr(s_T = 0)$ is the filtered probability $\Pr(s_T = 0 \mid \mathbf{Y}_T)$ of being in regime 0 corrected for the unconditional probability. Thus $\hat{y}_{T+h|T} - \mu_y$ can be written as

$$\begin{aligned}
\hat{y}_{T+h|T} - \mu_y &= \hat{\mu}_{T+h|T} + \hat{z}_{T+h|T} \\
&= (\mu_0 - \mu_1)(p_{00} + p_{11} - 1)^h \hat{\zeta}_{T|T} \\
&\quad + \alpha^h\left[y_T - \mu_y - (\mu_0 - \mu_1)\hat{\zeta}_{T|T}\right] \\
&= \alpha^h(y_T - \mu_y) + (\mu_0 - \mu_1)\left[(p_{00} + p_{11} - 1)^h - \alpha^h\right]\hat{\zeta}_{T|T}. \qquad (67)
\end{aligned}$$

This expression shows how the difference between the MS model forecasts and forecasts from a linear model depends on a number of characteristics such as the persistence of $\{s_t\}$. Specifically, the first term is the optimal prediction rule for a linear model. The contribution of the Markov regime-switching structure is given by the term multiplied by $\hat{\zeta}_{T|T}$, where $\hat{\zeta}_{T|T}$ contains the information about the most recent regime at the time the forecast is made. Thus, the contribution of the non-linear part of (67) to the overall forecast depends on both the magnitude of the regime shifts, $|\mu_0 - \mu_1|$, and on the persistence of regime shifts $p_{00} + p_{11} - 1$ relative to the persistence of the Gaussian process, given by α.

8.3. *Empirical evidence*

There are a large number of studies comparing the forecast performance of linear and non-linear models. There is little evidence for the superiority of non-linear models across the board. For example, Stock and Watson (1999) compare smooth-transition models [see, e.g., Teräsvirta (1994)], neural nets [e.g., White (1992)], and linear autoregressive models for 215 US macro time series, and find mixed evidence – the non-linear models sometimes record small gains at short horizons, but at longer horizons the linear models are preferred. Swanson and White (1997) forecast nine US macro series using a variety of fixed-specification linear and non-linear models, as well as flexible specifications of these which allow the specification to vary as the in-sample period is extended. They find little improvement from allowing for non-linearity within the flexible-specification approach.

Other studies focus on a few series, of which US output growth is one of the most popular. For example, Potter (1995) and Tiao and Tsay (1994) find that the forecast performance of the SETAR model relative to a linear model is markedly improved when the comparison is made in terms of how well the models forecast when the economy is in recession. The reason is easily understood. Since a majority of the sample data points (approximately 78%) fall in the upper regime, the linear AR(2) model will be largely determined by these points, and will closely match the upper-regime SETAR model. Thus the forecast performance of the two models will be broadly similar when the economy is in the expansionary phase of the business cycle. However, to the extent that the data points in the lower regime are characterized by a different process, there will be gains to the SETAR model during the contractionary phase.

Clements and Krolzig (1998) use (67) to explain why MS models of post war US output growth [such as those of Hamilton (1989)] do not forecast markedly more accurately than linear autoregressions. Namely, they find that $p_{00} + p_{11} - 1 = 0.65$ in their study, and that the largest root of the AR polynomial is 0.64. Because $p_{00} + p_{11} - 1 \simeq \alpha$ in (67), the conditional expectation collapses to a linear prediction rule.

9. Forecasting UK unemployment after three crises

The times at which causal-model based forecasts are most valuable are when considerable change occurs. Unfortunately, that is precisely when causal models are most likely to suffer forecast failure, and robust forecasting devices to outperform, at least relatively. We are not suggesting that prior to any major change, some methods are better at anticipating such shifts, nor that anyone could forecast the unpredictable: what we are concerned with is that even some time after a shift, many model types, in particular members of the equilibrium-correction class, will systematically mis-forecast.

To highlight this property, we consider three salient periods, namely the post-world-war double-decades of 1919–1938 and 1948–1967, and the post oil-crisis double-decade 1975–1994, to examine forecasts of the UK unemployment rate (denoted $U_{r,t}$). Figure 1 records the historical time-series of $U_{r,t}$ from 1875 to 2001 within which our three episodes lie. The data are discussed in detail in Hendry (2001), and the 'structural' equation for unemployment is taken from that paper.

The dramatically different epochs pre World War I (panel a), inter war (b), post World War II (c), and post the oil crisis (d) are obvious visually as each panel unfolds. In (b) there is an upward mean shift in 1920–1940. Panel (c) shows a downward mean shift and lower variance for 1940–1980. In the last panel there is an upward mean shift and higher

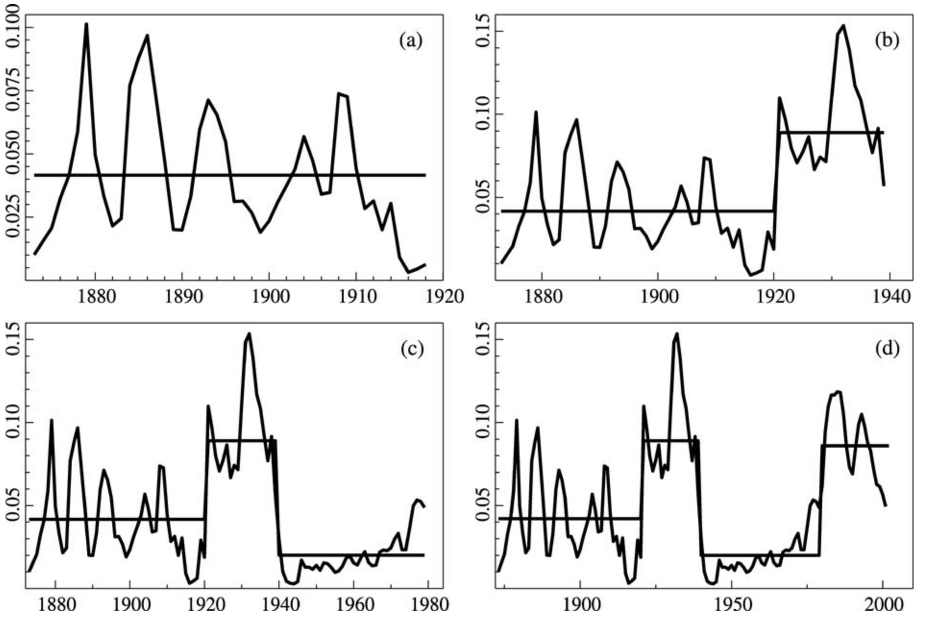

Figure 1. Shifts in unemployment.

variance from 1980 onwards. The unemployment rate time series seems distinctly non-stationary from shifts in both mean and variance at different times, but equally does not seem to have a unit root, albeit there is considerable persistence. Figure 2a records the changes in the unemployment rate.

The difficulty in forecasting after the three breaks is only partly because the preceding empirical evidence offers little guidance as to the subsequent behavior of the time series at each episode, since some 'naive' methods do not have great problems after breaks. Rather, it is the lack of adaptability of a forecasting device which seems to be the culprit.

The model derives the disequilibrium unemployment rate (denoted U_t^d) as a positive function of the difference between $U_{r,t}$ and the real interest rate $(R_{l,t} - \Delta p_t)$ minus the real growth rate (Δy_t). Then $U_{r,t}$ and $(R_{l,t} - \Delta p_t - \Delta y_t) = R_t^r$ are 'cointegrated' [using the PcGive test, $t_c = -3.9^{**}$; see Banerjee and Hendry (1992) and Ericsson and MacKinnon (2002)], or more probably, co-breaking [see Clements and Hendry (1999) and Hendry and Massmann (2006)]. Figure 2b plots the time series of R_t^r. The derived excess-demand for labor measure, U_t^d, is the long-run solution from an AD(2, 1) of $U_{r,t}$ on R_t^r with $\hat{\sigma} = 0.012$, namely,

$$U_t^d = U_{r,t} - \underset{(0.01)}{0.05} - \underset{(0.22)}{0.82} R_t^r, \qquad T = 1875\text{–}2001. \tag{68}$$

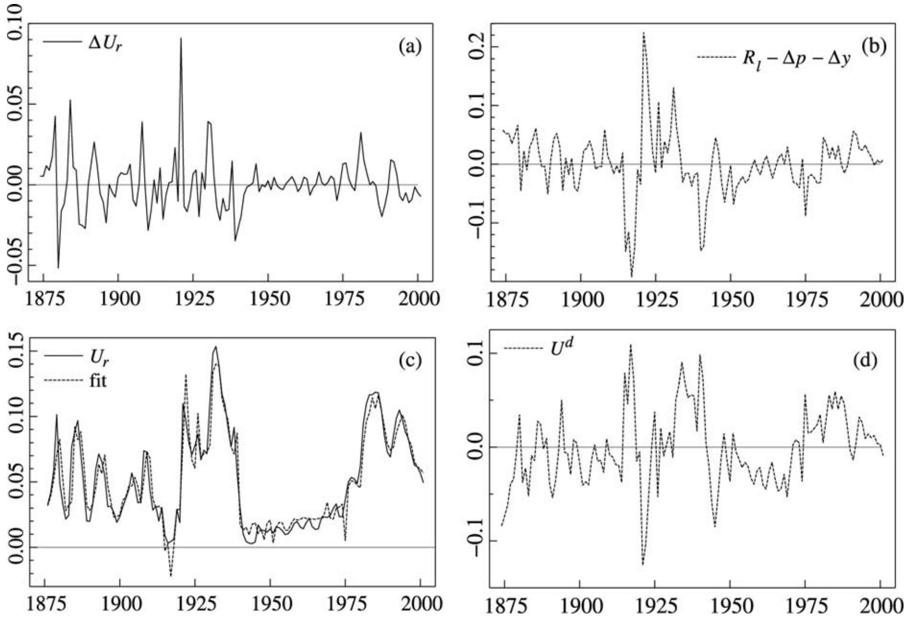

Figure 2. Unemployment with fitted values, $(R_{l,t} - \Delta p_t - \Delta y_t)$, and excess demand for labor.

The derived mean equilibrium unemployment is slightly above the historical sample average of 4.8%. U_t^d is recorded in Figure 2d.

Technically, given (68), a forecasting model for $U_{r,t}$ becomes a four-dimensional system for $U_{r,t}$, $R_{l,t}$, Δp_t, and Δy_t, but these in turn depend on other variables, rapidly leading to a large system. Instead, since the primary interest is illustrating forecasts from the equation for unemployment, we have chosen just to model $U_{r,t}$ and R_t^r as a bivariate VAR, with the restrictions implied by that formulation. That system was converted to an equilibrium-correction model (VECM) with the long-run solution given by (68) and $R^r = 0$. The full-sample FIML estimates from PcGive [see Hendry and Doornik (2001)] till 1991 were

$$\begin{aligned}
&\Delta U_{r,t} = \underset{(0.07)}{0.24}\Delta R_t^r - \underset{(0.037)}{0.14}U_{t-1}^d + \underset{(0.078)}{0.16}\Delta U_{r,t-1},\\
&\Delta R_t^r = -\underset{(0.077)}{0.43}R_{t-1}^r,\\
&\hat{\sigma}_{U_r} = 1.27\%, \quad \hat{\sigma}_{R^r} = 4.65\%, \quad T = 1875\text{–}1991,\\
&\chi^2_{\text{nd}}(4) = 76.2^{**}, \quad \mathsf{F}_{\text{ar}}(8, 218) = 0.81, \quad \mathsf{F}_{\text{het}}(27, 298) = 1.17.
\end{aligned} \tag{69}$$

In (69), $\hat{\sigma}$ denotes the residual standard deviation, and coefficient standard errors are shown in parentheses. The diagnostic tests are of the form $\mathsf{F}_j(k, T - l)$ which denotes an approximate F-test against the alternative hypothesis j for: second-order vector serial correlation [F_{ar}, see Guilkey (1974)] vector heteroskedasticity [$\mathsf{F}_{\mathrm{het}}$, see White (1980)]; and a chi-squared test for joint normality [$\chi^2_{\mathrm{nd}}(4)$, see Doornik and Hansen (1994)]. $*$ and $**$ denote significance at the 5% and 1% levels, respectively. All coefficients are significant with sensible signs and magnitudes, and the first equation is close to the OLS estimated model used in Hendry (2001). The likelihood ratio test of over-identifying restrictions of the VECM against the initial VAR yielded $\chi^2_{\mathrm{Id}}(8) = 2.09$. Figure 2c records the fitted values from the dynamic model in (69).

9.1. Forecasting 1992–2001

We begin with genuine ex ante forecasts. Since the model was selected from the sample $T = 1875$–1991, there are 10 new annual observations available since publication that can be used for forecast evaluation. This decade is picked purely because it is the last; there was in fact one major event, albeit not quite on the scale of the other three episodes to be considered, namely the ejection of the UK from the exchange rate mechanism (ERM) in the autumn of 1992, just at the forecast origin. Nevertheless, by historical standards the period transpired to be benign, and almost any method would have avoided forecast failure over this sample, including those considered here. In fact, the 1-step forecast test over 10 periods for (69), denoted $\mathsf{F}_{\mathrm{Chow}}$ [see Chow (1960)], delivered $\mathsf{F}_{\mathrm{Chow}}(20, 114) = 0.15$, consistent with parameter constancy over the post-selection decade. Figure 3 shows the graphical output for 1-step and 10-step forecasts

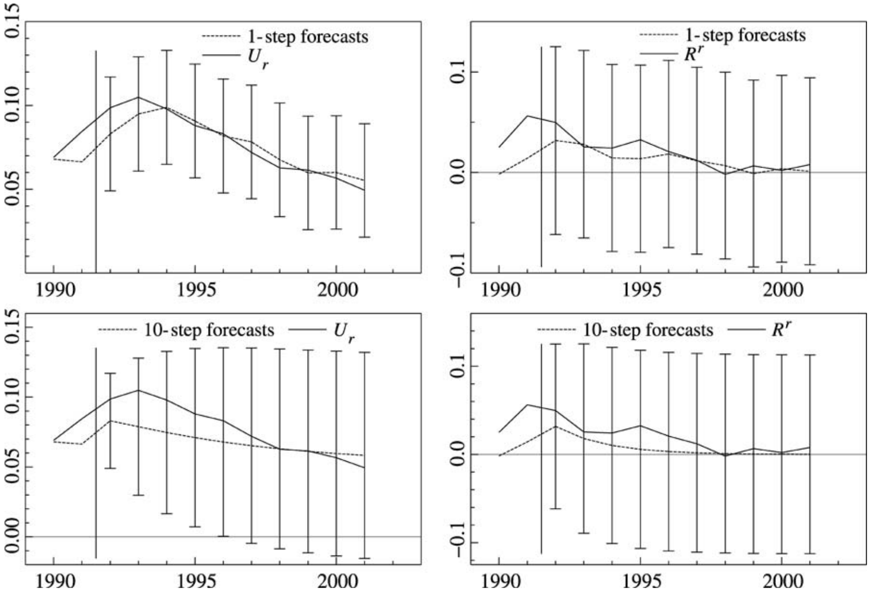

Figure 3. VECM 1-step and 10-step forecasts of $U_{r,t}$ and R^r_t, 1992–2001.

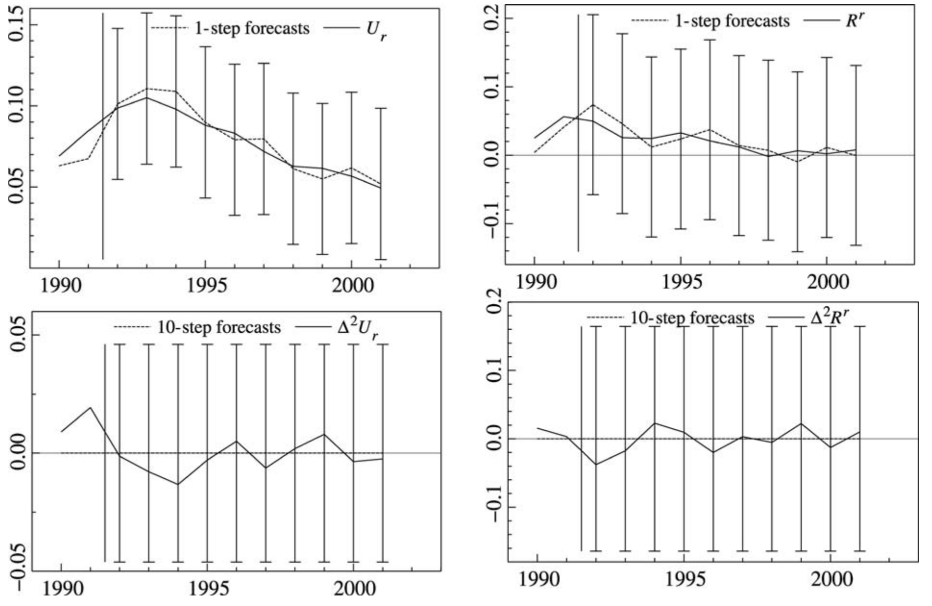

Figure 4. DVECM 1-step forecasts of $U_{r,t}$, R_t^r, and 10-step forecasts of $\Delta^2 U_{r,t}$, $\Delta^2 R_t^r$, 1992–2001.

of $U_{r,t}$ and R_t^r over 1992–2001. As can be seen, all the outcomes lie well inside the interval forecasts (shown as $\pm 2\hat{\sigma}_f$) for both sets of forecasts. Notice the equilibrium-correction behavior manifest in the 10-step forecasts, as U_r converges to 0.05 and R^r to 0: such must occur, independently of the outcomes for $U_{r,t}$ and R_t^r.

On all these criteria, the outcome is successful on the out-of-selection-sample evaluation. While far from definitive, as shown in Clements and Hendry (2005), these results demonstrate that the model merits its more intensive scrutiny over the three salient historical episodes.

By way of comparison, we also record the corresponding forecasts from the differenced models discussed in Section 7.3. First, we consider the VECM (denoted DVECM) which maintains the parameter estimates, but differences all the variables [see Hendry (2005)]. Figure 4 shows the graphical output for 1-step forecasts of $U_{r,t}$ and R_t^r and the 10-step forecasts of $\Delta^2 U_{r,t}$ and $\Delta^2 R_t^r$ over 1992–2001 (throughout, the interval forecasts for multi-step forecasts from mis-specified models are not adjusted for the – unknown – mis-specification). In fact, there was little discernible difference between the forecasts produced by the DVECM and those from a double-difference VAR [DDVAR, see Clements and Hendry (1999) and Section 7.3].

The 1-step forecasts are close to those from the VECM, but the entailed multi-step levels forecasts from the DVECM are poor, as the rise in unemployment prior to the forecast origin turns to a fall throughout the remainder of the period, but the forecasts continue to rise: there is no free lunch when insuring against forecast failure.

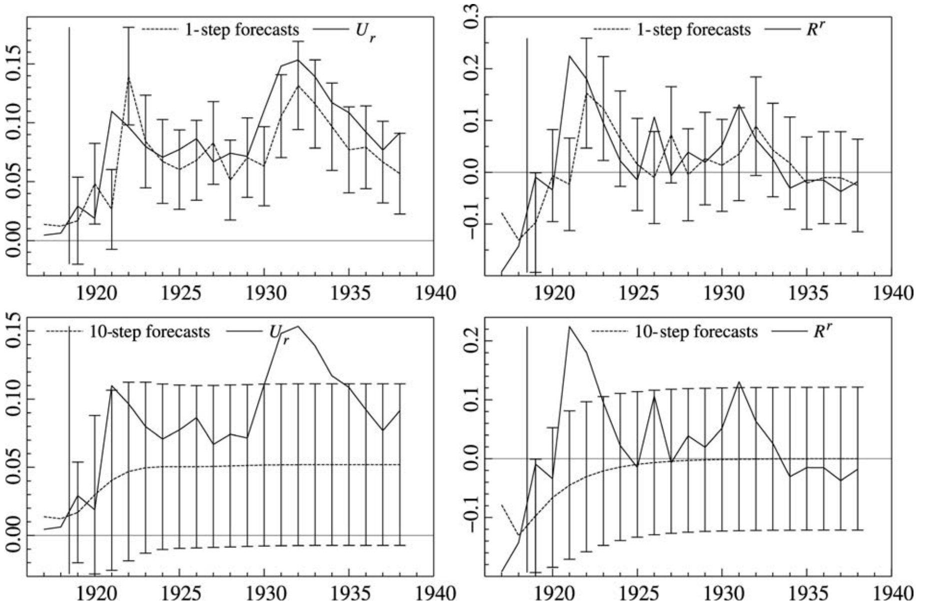

Figure 5. VECM 1-step and 10-step forecasts of $U_{r,t}$ and R_t^r, 1919–1938.

9.2. Forecasting 1919–1938

Over this sample, $\mathsf{F}_{\text{Chow}}(40, 41) = 2.81^{**}$, strongly rejecting the model re-estimated, but not re-selected, up to 1918. The graphs in Figure 5 confirm the forecast failure, for both 1-step and 10-step forecasts of $U_{r,t}$ and R_t^r. As well as missing the post-World-War I dramatic rise in unemployment, there is systematic under-forecasting throughout the Great Depression period, consistent with failing to forecast the substantial increase in R_t^r on both occasions. Nevertheless, the results are far from catastrophic in the face of such a large, systematic, and historically unprecedented, rise in unemployment.

Again using our comparator of the DVECM, Figure 6 shows the 1-step forecasts, with a longer historical sample to highlight the substantial forecast-period change (the entailed multi-step levels' forecasts are poor). Despite the noticeable level shift in $U_{r,t}$, the differenced model forecasts are only a little better initially, overshooting badly after the initial rise, but perform well over the Great Depression, which is forecasting long after the earlier break. $\mathsf{F}_{\text{Chow}}(40, 42) = 2.12^{**}$ is slightly smaller overall despite the initial 'bounce'.

9.3. Forecasting 1948–1967

The model copes well with the post-World-War II low level of unemployment, with $\mathsf{F}_{\text{Chow}}(40, 70) = 0.16$, with the outcomes shown in Figure 7. However, there is sys-

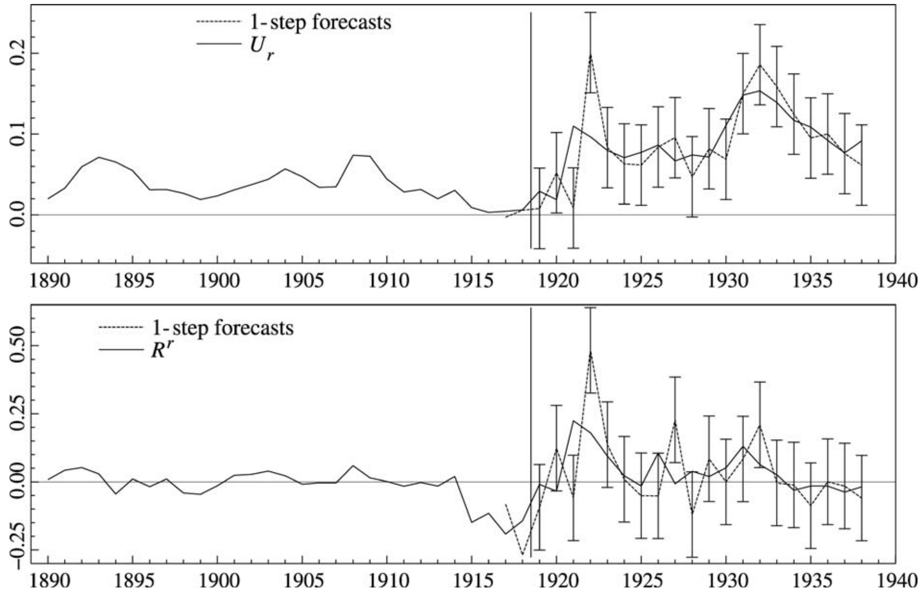

Figure 6. DVECM 1-step forecasts of $U_{r,t}$ and R_t^r, 1919–1938.

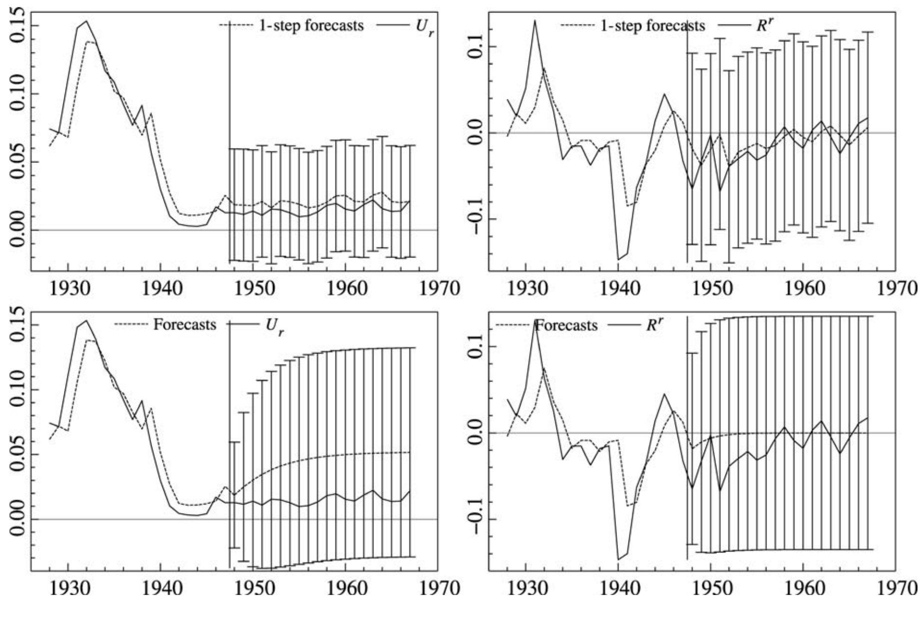

Figure 7. VECM 1-step and 10-step forecasts of $U_{r,t}$ and R_t^r, 1948–1967.

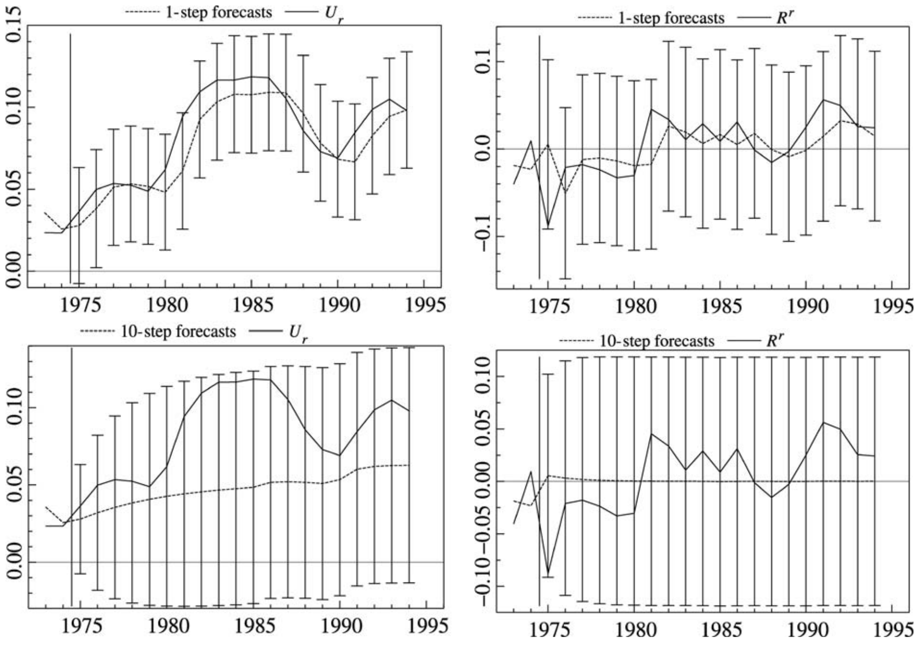

Figure 8. VECM 1-step and 10-step forecasts of $U_{r,t}$ and R_t^r, 1975–1994.

tematic over-forecasting of the level of unemployment, unsurprisingly given its exceptionally low level. The graph here emphasizes the equilibrium-correction behavior of U_r converging to 0.05 even though the outcome is now centered around 1.5%. The DVECM delivers $\mathsf{F}_{\text{Chow}}(40, 71) = 0.12$ so is closely similar. The forecasts are also little different, although the forecast intervals are somewhat wider.

9.4. *Forecasting 1975–1994*

Finally, after the first oil crisis, we find $\mathsf{F}_{\text{Chow}}(40, 97) = 0.61$, so surprisingly no forecast failure results, although the outcomes are poor as Figure 8 shows for both 1-step and 10-step forecasts of $U_{r,t}$ and R_t^r. There is systematic under-forecasting of the level of unemployment, but the trend is correctly discerned as upwards. Over this period, $\mathsf{F}_{\text{Chow}}(40, 98) = 0.53$ for the DVECM, so again there is little impact from removing the equilibrium-correction term.

9.5. *Overview*

Despite the manifest non-stationarity of the UK unemployment rate over the last century and a quarter, with location and variance shifts evident in the historical data, the empirical forecasting models considered here only suffered forecast failure occasionally, although they were often systematically adrift, under- or over-forecasting. The differenced VECM did not perform much better even when the VECM failed. A possible explanation may be the absence of deterministic components from the VECM in (69) other than that embedded in the long-run for unemployment. Since $\hat{\sigma}_{U_r} = 1.27\%$, a 95% forecast interval spans just over 5% points of unemployment so larger shifts are needed to reject the model.

It is difficult to imagine how well real-time forecasting might have performed historically: the large rise in unemployment during 1919–1920 seems to have been unanticipated at the time, and induced real hardship, leading to considerable social unrest. Conversely, while the Beveridge Report (Social Insurance and Allied Services, HMSO, 1942, followed by his Full Employment in a Free Society and The Economics of Full Employment, both in 1944) essentially mandated UK Governments to keep a low level of unemployment using Keynesian policies, nevertheless the outturn of 1.5% on average over 1946–1966 was unprecedented. And the Thatcher reforms of 1979 led to an unexpectedly large upturn in unemployment, commensurate with inter-war levels. Since the historical period delivered many unanticipated 'structural breaks', across many very different policy regimes (from the Gold Standard, floating, Bretton Woods currency pegs, back to a 'dirty' floating – just to note exchange-rate regimes), overall, the forecasting performance of the unemployment model considered here is really quite creditable.

10. Concluding remarks

Structural breaks in the form of unforeseen location shifts are likely to lead to systematic forecast biases. Other factors matter, as shown in the various taxonomies of forecast errors above, but breaks play a dominant role. The vast majority of forecasting models in regular use are members of the equilibrium-correction class, including VARs, VECMs, and DSGEs, as well as many popular models of conditional variance processes. Other types of models might be more robust to breaks. We have also noted issues to do with the choice of estimation sample, and the updating of the models' parameter estimates and of the model specification, as possible ways of mitigating the effects of some types of breaks. Some ad hoc forecasting devices exhibit greater adaptability than standard models, which may account for their successes in empirical forecasting competitions. Finally, we have contrasted non-constancies due to breaks with those due to non-linearities.

Appendix A: Taxonomy derivations for Equation (10)

We let $\boldsymbol{\delta}_{\varphi} = \hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_p$, where $\boldsymbol{\varphi}_p = (\mathbf{I}_n - \boldsymbol{\Pi}_p)^{-1}\boldsymbol{\phi}_p$, $\boldsymbol{\delta}_{\Pi} = \widehat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}_p$, and $\hat{\mathbf{y}}_T - \mathbf{y}_T = \boldsymbol{\delta}_y$.

First, we use the approximation:

$$\widehat{\boldsymbol{\Pi}}^h = (\boldsymbol{\Pi}_p + \boldsymbol{\delta}_{\Pi})^h \simeq \boldsymbol{\Pi}_p^h + \sum_{i=0}^{h-1} \boldsymbol{\Pi}_p^i \boldsymbol{\delta}_{\Pi} \boldsymbol{\Pi}_p^{h-i-1} \triangleq \boldsymbol{\Pi}_p^h + \mathbf{C}_h. \tag{A.1}$$

Let $(\cdot)^\nu$ denote a vectorizing operator which stacks the columns of an $m \times n$ matrix $\mathbf{A}$ in an $mn \times 1$ vector $\mathbf{a}$, after which $(\mathbf{a})^\nu = \mathbf{a}$. Also, let $\otimes$ be the associated Kronecker product, so that when $\mathbf{B}$ is $p \times q$, then $\mathbf{A} \otimes \mathbf{B}$ is an $mp \times nq$ matrix of the form $\{b_{ij}\mathbf{A}\}$. Consequently, when $\mathbf{ABC}$ is defined,

$$(\mathbf{ABC})^\nu = (\mathbf{A} \otimes \mathbf{C}')\mathbf{B}^\nu.$$

Using these, from (A.1),

$$\begin{aligned}
\mathbf{C}_h(\mathbf{y}_T - \boldsymbol{\varphi}_p) &= \big(\mathbf{C}_h(\mathbf{y}_T - \boldsymbol{\varphi}_p)\big)^\nu \\
&= \left(\sum_{i=0}^{h-1} \boldsymbol{\Pi}_p^i \otimes (\mathbf{y}_T - \boldsymbol{\varphi}_p)' \boldsymbol{\Pi}_p^{h-i-1\,\prime}\right)\boldsymbol{\delta}_\Pi^\nu \\
&\mathrel{\hat{=}} \mathbf{F}_h \boldsymbol{\delta}_\Pi^\nu.
\end{aligned} \tag{A.2}$$

To highlight components due to different effects (parameter change, estimation inconsistency, and estimation uncertainty), we decompose the term $(\boldsymbol{\Pi}^*)^h(\mathbf{y}_T - \boldsymbol{\varphi}^*)$ into

$$(\boldsymbol{\Pi}^*)^h(\mathbf{y}_T - \boldsymbol{\varphi}^*) = (\boldsymbol{\Pi}^*)^h(\mathbf{y}_T - \boldsymbol{\varphi}) + (\boldsymbol{\Pi}^*)^h(\boldsymbol{\varphi} - \boldsymbol{\varphi}^*),$$

whereas $\widehat{\boldsymbol{\Pi}}^h(\hat{\mathbf{y}}_T - \hat{\boldsymbol{\varphi}})$ equals

$$\begin{aligned}
&(\boldsymbol{\Pi}_p^h + \mathbf{C}_h)\boldsymbol{\delta}_y - (\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_p) + (\mathbf{y}_T - \boldsymbol{\varphi}_p) \\
&\quad = (\boldsymbol{\Pi}_p^h + \mathbf{C}_h)\boldsymbol{\delta}_y - (\boldsymbol{\Pi}_p^h + \mathbf{C}_h)\boldsymbol{\delta}_\varphi + (\boldsymbol{\Pi}_p^h + \mathbf{C}_h)(\mathbf{y}_T - \boldsymbol{\varphi}_p) \\
&\quad \mathrel{\hat{=}} (\boldsymbol{\Pi}_p^h + \mathbf{C}_h)\boldsymbol{\delta}_y - (\boldsymbol{\Pi}_p^h + \mathbf{C}_h)\boldsymbol{\delta}_\varphi + \mathbf{F}_h\boldsymbol{\delta}_\Pi^\nu + \boldsymbol{\Pi}_p^h(\mathbf{y}_T - \boldsymbol{\varphi}) - \boldsymbol{\Pi}_p^h(\boldsymbol{\varphi}_p - \boldsymbol{\varphi}).
\end{aligned}$$

Thus, $(\boldsymbol{\Pi}^*)^h(\mathbf{y}_T - \boldsymbol{\varphi}^*) - \widehat{\boldsymbol{\Pi}}^h(\hat{\mathbf{y}}_T - \hat{\boldsymbol{\varphi}})$ yields

$$\begin{aligned}
&\big((\boldsymbol{\Pi}^*)^h - \boldsymbol{\Pi}_p^h\big)(\mathbf{y}_T - \boldsymbol{\varphi}) - \mathbf{F}_h\boldsymbol{\delta}_\Pi^\nu - (\boldsymbol{\Pi}_p^h + \mathbf{C}_h)\boldsymbol{\delta}_y \\
&\quad - (\boldsymbol{\Pi}^*)^h(\boldsymbol{\varphi}^* - \boldsymbol{\varphi}) + \boldsymbol{\Pi}_p^h(\boldsymbol{\varphi}_p - \boldsymbol{\varphi}) + (\boldsymbol{\Pi}_p^h + \mathbf{C}_h)\boldsymbol{\delta}_\varphi.
\end{aligned} \tag{A.3}$$

The interaction $\mathbf{C}_h\boldsymbol{\delta}_\varphi$ is like a 'covariance', but is omitted from the table. Hence (A.3) becomes

$$\begin{aligned}
&\big((\boldsymbol{\Pi}^*)^h - \boldsymbol{\Pi}^h\big)(\mathbf{y}_T - \boldsymbol{\varphi}) + (\boldsymbol{\Pi}^h - \boldsymbol{\Pi}_p^h)(\mathbf{y}_T - \boldsymbol{\varphi}) \\
&\quad - (\boldsymbol{\Pi}^*)^h(\boldsymbol{\varphi}^* - \boldsymbol{\varphi}) + \boldsymbol{\Pi}_p^h(\boldsymbol{\varphi}_p - \boldsymbol{\varphi}) \\
&\quad - (\boldsymbol{\Pi}_p^h + \mathbf{C}_h)\boldsymbol{\delta}_y - \mathbf{F}_h\boldsymbol{\delta}_\Pi^\nu + \boldsymbol{\Pi}_p^h\boldsymbol{\delta}_\varphi.
\end{aligned}$$

The first and third rows have expectations of zero, so the second row collects the 'non-central' terms.

Finally, for the term $\boldsymbol{\varphi}^* - \hat{\boldsymbol{\varphi}}$ we have (on the same principle):

$$(\boldsymbol{\varphi}^* - \boldsymbol{\varphi}) + (\boldsymbol{\varphi} - \boldsymbol{\varphi}_p) - \boldsymbol{\delta}_\varphi.$$

Appendix B: Derivations for Section 4.3

Since $\boldsymbol{\Upsilon} = \mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'$, for $j > 0$,

$$\begin{aligned}\boldsymbol{\Upsilon}^j &= (\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}')^j = \boldsymbol{\Upsilon}^{j-1}(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}') = \boldsymbol{\Upsilon}^{j-1} + \boldsymbol{\Upsilon}^{j-1}\boldsymbol{\alpha}\boldsymbol{\beta}' = \cdots \\ &= \mathbf{I}_n + \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^i \boldsymbol{\alpha}\boldsymbol{\beta}', \end{aligned} \tag{B.1}$$

so

$$(\boldsymbol{\Upsilon}^j - \mathbf{I}_n) = \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^i \boldsymbol{\alpha}\boldsymbol{\beta}' = \mathbf{A}_j \boldsymbol{\alpha}\boldsymbol{\beta}' \tag{B.2}$$

defines $\mathbf{A}_j = \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^i$. Thus,

$$\mathsf{E}[(\boldsymbol{\Upsilon}^j - \mathbf{I}_n)\mathbf{w}_T] = \mathbf{A}_j \boldsymbol{\alpha} \mathsf{E}[\boldsymbol{\beta}'\mathbf{x}_T] = \mathbf{A}_j \boldsymbol{\alpha} \boldsymbol{f}_T, \tag{B.3}$$

where $\mathbf{f}_T = \mathsf{E}[\boldsymbol{\beta}'\mathbf{x}_T] = \boldsymbol{\mu}_0^a + \boldsymbol{\beta}'\boldsymbol{\gamma}^a(T+1)$, say, where the values of $\boldsymbol{\mu}_0^a = \boldsymbol{\mu}_0$ and $\boldsymbol{\gamma}^a = \boldsymbol{\gamma}$ if the change occurs after period T, and $\boldsymbol{\mu}_0^a = \boldsymbol{\mu}_0^*$ and $\boldsymbol{\gamma}^a = \boldsymbol{\gamma}^*$ if the change occurs before period T.

Substituting from (B.3) into (34):

$$\mathsf{E}[\tilde{\boldsymbol{\nu}}_{T+j}] = \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^i [\boldsymbol{\gamma}^* - \boldsymbol{\alpha}\boldsymbol{\mu}_0^* - \boldsymbol{\alpha}\boldsymbol{\mu}_1^*(T+j-i)] - j\boldsymbol{\gamma} + \mathbf{A}_j \boldsymbol{\alpha} \boldsymbol{f}_T. \tag{B.4}$$

From (B.1), as $\boldsymbol{\Upsilon}^i = \mathbf{I}_n + \mathbf{A}_i \boldsymbol{\alpha}\boldsymbol{\beta}'$,

$$\mathbf{A}_j = \sum_{k=0}^{j-1} \boldsymbol{\Upsilon}^k = \sum_{k=0}^{j-1} (\mathbf{I}_n + \mathbf{A}_k \boldsymbol{\alpha}\boldsymbol{\beta}') = j\mathbf{I}_n + \left(\sum_{k=0}^{j-1} \mathbf{A}_k\right)\boldsymbol{\alpha}\boldsymbol{\beta}' = j\mathbf{I}_n + \mathbf{B}_j \boldsymbol{\alpha}\boldsymbol{\beta}'. \tag{B.5}$$

Thus from (B.4), since $\boldsymbol{\beta}'\boldsymbol{\gamma} = \boldsymbol{\mu}_1$ and $\boldsymbol{\beta}'\boldsymbol{\gamma}^* = \boldsymbol{\mu}_1^*$,

$$\begin{aligned}\mathsf{E}[\tilde{\boldsymbol{\nu}}_{T+j}] &= \mathbf{A}_j \boldsymbol{\gamma}^* - \mathbf{A}_j \boldsymbol{\alpha}\boldsymbol{\mu}_0^* - \mathbf{A}_j \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\gamma}^*(T+j) + \sum_{i=1}^{j-1} i \boldsymbol{\Upsilon}^i \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\gamma}^* - j\boldsymbol{\gamma} \\ &\quad + \mathbf{A}_j \boldsymbol{\alpha} \boldsymbol{f}_T \\ &= j(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}) + \mathbf{A}_j \boldsymbol{\alpha} \boldsymbol{f}_T - \boldsymbol{\mu}_0^* - \boldsymbol{\beta}'\boldsymbol{\gamma}^* T \\ &\quad + \left(\sum_{i=1}^{j-1} i \boldsymbol{\Upsilon}^i - j\mathbf{A}_j + \mathbf{B}_j\right)\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\gamma}^* \\ &= j(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}) + \mathbf{A}_j \boldsymbol{\alpha}(\boldsymbol{\mu}_0^a - \boldsymbol{\mu}_0^* - \boldsymbol{\beta}'[\boldsymbol{\gamma}^* - \boldsymbol{\gamma}^a](T+1)) \\ &\quad + \mathbf{C}_j \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\gamma}^*, \end{aligned} \tag{B.6}$$

where $\mathbf{C}_j = (\mathbf{D}_j + \mathbf{B}_j - (j-1)\mathbf{A}_j)$ when $\mathbf{D}_j = \sum_{i=1}^{j-1} i\boldsymbol{\Upsilon}^i$. However, $\mathbf{C}_j\boldsymbol{\alpha}\boldsymbol{\beta}' = \mathbf{0}$ as follows. Since $\boldsymbol{\Upsilon}^j = \mathbf{I}_n + \mathbf{A}_j\boldsymbol{\alpha}\boldsymbol{\beta}'$ from (B.2), then

$$j\mathbf{A}_j\boldsymbol{\alpha}\boldsymbol{\beta}' = j\boldsymbol{\Upsilon}^j - j\mathbf{I}_n,$$

and so eliminating $j\mathbf{I}_n$ using (B.5):

$$(\mathbf{B}_j - j\mathbf{A}_j)\boldsymbol{\alpha}\boldsymbol{\beta}' = \mathbf{A}_j - j\boldsymbol{\Upsilon}^j.$$

Also,

$$\mathbf{D}_j = \sum_{i=1}^{j} i\boldsymbol{\Upsilon}^i - j\boldsymbol{\Upsilon}^j = \sum_{i=1}^{j} \boldsymbol{\Upsilon}^i - j\boldsymbol{\Upsilon}^j + \left(\sum_{i=1}^{j-1} i\boldsymbol{\Upsilon}^i\right)\boldsymbol{\Upsilon} = \mathbf{A}_j\boldsymbol{\Upsilon} - j\boldsymbol{\Upsilon}^j + \mathbf{D}_j\boldsymbol{\Upsilon}.$$

Since $\boldsymbol{\Upsilon} = \mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'$,

$$\mathbf{D}_j\boldsymbol{\alpha}\boldsymbol{\beta}' = j\boldsymbol{\Upsilon}^j - \mathbf{A}_j - \mathbf{A}_j\boldsymbol{\alpha}\boldsymbol{\beta}'.$$

Combining these results,

$$\begin{aligned}\mathbf{C}_j\boldsymbol{\alpha}\boldsymbol{\beta}' &= \big(\mathbf{D}_j + \mathbf{B}_j - (j-1)\mathbf{A}_j\big)\boldsymbol{\alpha}\boldsymbol{\beta}' \\ &= j\boldsymbol{\Upsilon}^j - \mathbf{A}_j - \mathbf{A}_j\boldsymbol{\alpha}\boldsymbol{\beta}' + \mathbf{A}_j - j\boldsymbol{\Upsilon}^j + \mathbf{A}_j\boldsymbol{\alpha}\boldsymbol{\beta}' = \mathbf{0}.\end{aligned} \tag{B.7}$$

References

Al-Qassam, M.S., Lane, J.A. (1989). “Forecasting exponential autoregressive models of order 1”. Journal of Time Series Analysis 10, 95–113.

Albert, J., Chib, S. (1993). “Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts”. Journal of Business and Economic Statistics 11, 1–16.

Andrews, D.W.K. (1993). “Tests for parameter instability and structural change with unknown change point”. Econometrica 61, 821–856.

Andrews, D.W.K., Ploberger, W. (1994). “Optimal tests when a nuisance parameter is present only under the alternative”. Econometrica 62, 1383–1414.

Armstrong, J.S. (Ed.) (2001). Principles of Forecasting. Kluwer Academic, Boston.

Bai, J., Lumsdaine, R.L., Stock, J.H. (1998). “Testing for and dating common breaks in multivariate time series”. Review of Economics and Statistics 63, 395–432.

Bai, J., Perron, P. (1998). “Estimating and testing linear models with multiple structural changes”. Econometrica 66, 47–78.

Baillie, R.T., Bollerslev, T. (1992). “Prediction in dynamic models with time-dependent conditional variances”. Journal of Econometrics 52, 91–113.

Balke, N.S. (1993). “Detecting level shifts in time series”. Journal of Business and Economic Statistics 11, 81–92.

Banerjee, A., Hendry, D.F. (1992). “Testing integration and cointegration: An overview”. Oxford Bulletin of Economics and Statistics 54, 225–255.

Barnett, W.A., Hendry, D.F., Hylleberg, S., et al. (Eds.) (2000). Nonlinear Econometric Modeling in Time Series Analysis. Cambridge University Press, Cambridge.

Bates, J.M., Granger, C.W.J. (1969). “The combination of forecasts”. Operations Research Quarterly 20, 451–468. Reprinted in: Mills, T.C. (Ed.) (1999). Economic Forecasting. Edward Elgar.

Bera, A.K., Higgins, M.L. (1993). "ARCH models: Properties, estimation and testing". Journal of Economic Surveys 7, 305–366.

Bianchi, C., Calzolari, G. (1982). "Evaluating forecast uncertainty due to errors in estimated coefficients: Empirical comparison of alternative methods". In: Chow, G.C., Corsi, P. (Eds.), Evaluating the Reliability of Macro-Economic Models. Wiley, New York. Chapter 13.

Bollerslev, T. (1986). "Generalised autoregressive conditional heteroskedasticity". Journal of Econometrics 51, 307–327.

Bollerslev, T., Chou, R.S., Kroner, K.F. (1992). "ARCH modelling in finance – A review of the theory and empirical evidence". Journal of Econometrics 52, 5–59.

Bontemps, C., Mizon, G.E. (2003). "Congruence and encompassing". In: Stigum, B.P. (Ed.), Econometrics and the Philosophy of Economics. Princeton University Press, Princeton, pp. 354–378.

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

Breusch, T.S., Pagan, A.R. (1979). "A simple test for heteroscedasticity and random coefficient variation". Econometrica 47, 1287–1294.

Brown, R.L., Durbin, J., Evans, J.M. (1975). "Techniques for testing the constancy of regression relationships over time (with discussion)". Journal of the Royal Statistical Society B 37, 149–192.

Calzolari, G. (1981). "A note on the variance of ex post forecasts in econometric models". Econometrica 49, 1593–1596.

Calzolari, G. (1987). "Forecast variance in dynamic simulation of simultaneous equations models". Econometrica 55, 1473–1476.

Carrasco, M. (2002). "Misspecified structural change, threshold, and Markov switching models". Journal of Econometrics 109, 239–273.

Chan, K.S. (1990). "Testing for threshold autoregression". The Annals of Statistics 18, 1886–1894.

Chan, K.S. (1991). "Percentage points of likelihood ratio tests for threshold autoregression". Journal of the Royal Statistical Society, Series B 53, 691–696.

Chan, N.H., Wei, C.Z. (1988). "Limiting distributions of least squares estimates of unstable autoregressive processes". Annals of Statistics 16, 367–401.

Chen, C., Liu, L.-M. (1993). "Joint estimation of model parameters and outlier effects in time series". Journal of the American Statistical Association 88, 284–297.

Chen, C., Tiao, G.C. (1990). "Random level-shift time series models, ARIMA approximations and level-shift detection". Journal of Business and Economic Statistics 8, 83–97.

Chong, T. (2001). "Structural change in AR(1) models". Econometric Theory 17, 87–155.

Chow, G.C. (1960). "Tests of equality between sets of coefficients in two linear regressions". Econometrica 28, 591–605.

Christoffersen, P.F., Diebold, F.X. (1998). "Cointegration and long-horizon forecasting". Journal of Business and Economic Statistics 16, 450–458.

Chu, C.S., Stinchcombe, M., White, H. (1996). "Monitoring structural change". Econometrica 64, 1045–1065.

Clemen, R.T. (1989). "Combining forecasts: A review and annotated bibliography". International Journal of Forecasting 5, 559–583. Reprinted in: Mills, T.C. (Ed.) (1999). Economic Forecasting. Edward Elgar.

Clements, M.P., Galvão, A.B. (2005). "Combining predictors and combining information in modelling: Forecasting US recession probabilities and output growth". In: Milas, C., Rothman, P., van Dijk, D. (Eds.), Nonlinear Time Series Analysis of Business Cycles. Contributions to Economic Analysis Series. Elsevier. In press.

Clements, M.P., Hendry, D.F. (1995). "Forecasting in cointegrated systems". Journal of Applied Econometrics 10, 127–146. Reprinted in: Mills, T.C. (Ed.) (1999). Economic Forecasting. Edward Elgar.

Clements, M.P., Hendry, D.F. (1996). "Intercept corrections and structural change". Journal of Applied Econometrics 11, 475–494.

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

Clements, M.P., Hendry, D.F. (1999). Forecasting Non-Stationary Economic Time Series. MIT Press, Cambridge, MA.

Clements, M.P., Hendry, D.F. (Eds.) (2002a). A Companion to Economic Forecasting. Blackwells, Oxford.
Clements, M.P., Hendry, D.F. (2002b). "Explaining forecast failure in macroeconomics". In: Clements and Hendry (2002a), pp. 539–571.
Clements, M.P., Hendry, D.F. (2005). "Evaluating a model by forecast performance". Oxford Bulletin of Economics and Statistics 67, 931–956.
Clements, M.P., Krolzig, H.-M. (1998). "A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP". Econometrics Journal 1, 47–75.
Clements, M.P., Krolzig, H.-M. (2003). "Business cycle asymmetries: Characterisation and testing based on Markov-switching autoregressions". Journal of Business and Economic Statistics 21, 196–211.
Clements, M.P., Smith, J. (1999). "A Monte Carlo study of the forecasting performance of empirical SETAR models". Journal of Applied Econometrics 14, 124–141.
Cogley, T., Sargent, T.J. (2001). "Evolving post World War II inflation dynamics". NBER Macroeconomics Annual 16, 331–373.
Cogley, T., Sargent, T.J. (2005). "Drifts and volatilities: Monetary policies and outcomes in the post World War II US". Review of Economic Dynamics 8, 262–302.
Davidson, J.E.H., Hendry, D.F., Srba, F., Yeo, J.S. (1978). "Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom". Economic Journal 88, 661–692. Reprinted in: Hendry, D.F. (1993). Econometrics: Alchemy or Science? Blackwell Publishers, Oxford and Oxford University Press, 2000.
Davies, R.B. (1977). "Hypothesis testing when a nuisance parameter is present only under the alternative". Biometrika 64, 247–254.
Davies, R.B. (1987). "Hypothesis testing when a nuisance parameter is present only under the alternative". Biometrika 74, 33–43.
De Gooijer, J.G., De Bruin, P. (1997). "On SETAR forecasting". Statistics and Probability Letters 37, 7–14.
Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). "Maximum likelihood estimation from incomplete data via the EM algorithm". Journal of the Royal Statistical Society, Series B 39, 1–38.
Diebold, F.X., Chen, C. (1996). "Testing structural stability with endogenous breakpoint: A size comparison of analytic and bootstrap procedures". Journal of Econometrics 70, 221–241.
Diebold, F.X., Lee, J.H., Weinbach, G.C. (1994). "Regime switching with time-varying transition probabilities". In: Hargreaves, C. (Ed.), Non-Stationary Time-Series Analyses and Cointegration. Oxford University Press, Oxford, pp. 283–302.
Diebold, F.X., Lopez, J.A. (1996). "Forecast evaluation and combination". In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics, vol. 14. North-Holland, Amsterdam, pp. 241–268.
Diebold, F.X., Rudebusch, G.D., Sichel, D.E. (1993). "Further evidence on business cycle duration dependence". In: Stock, J., Watson, M. (Eds.), Business Cycles Indicators, and Forecasting. University of Chicago Press and NBER, Chicago, pp. 255–280.
Doornik, J.A., Hansen, H. (1994). "A practical test for univariate and multivariate normality". Discussion Paper, Nuffield College.
Durland, J.M., McCurdy, T.H. (1994). "Duration dependent transitions in a Markov model of U.S. GNP growth". Journal of Business and Economic Statistics 12, 279–288.
Engle, R.F. (1982). "Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation". Econometrica 50, 987–1007.
Engle, R.F., Bollerslev, T. (1987). "Modelling the persistence of conditional variances". Econometric Reviews 5, 1–50.
Engle, R.F., McFadden, D. (Eds.) (1994). Handbook of Econometrics, vol. 4. Elsevier Science, North-Holland, Amsterdam.
Engle, R.F., Yoo, B.S. (1987). "Forecasting and testing in co-integrated systems". Journal of Econometrics 35, 143–159.
Ericsson, N.R., MacKinnon, J.G. (2002). "Distributions of error correction tests for cointegration". Econometrics Journal 5, 285–318.
Filardo, A.J. (1994). "Business cycle phases and their transitional dynamics". Journal of Business and Economic Statistics 12, 299–308.

Fildes, R.A., Makridakis, S. (1995). "The impact of empirical accuracy studies on time series analysis and forecasting". International Statistical Review 63, 289–308.
Fildes, R.A., Ord, K. (2002). "Forecasting competitions – Their role in improving forecasting practice and research". In: Clements and Hendry (2002a), pp. 322–253.
Fuller, W.A., Hasza, D.P. (1980). "Predictors for the first-order autoregressive process". Journal of Econometrics 13, 139–157.
Garcia, R. (1998). "Asymptotic null distribution of the likelihood ratio test in Markov switching models". International Economic Review 39, 763–788.
Gardner, E.S., McKenzie, E. (1985). "Forecasting trends in time series". Management Science 31, 1237–1246.
Goodwin, T.H. (1993). "Business-cycle analysis with a Markov-switching model". Journal of Business and Economic Statistics 11, 331–339.
Granger, C.W.J. (1989). "Combining forecasts – Twenty years later". Journal of Forecasting 8, 167–173.
Granger, C.W.J., White, H., Kamstra, M. (1989). "Interval forecasting: An analysis based upon ARCH-quantile estimators". Journal of Econometrics 40, 87–96.
Griliches, Z., Intriligator, M.D. (Eds.) (1983). Handbook of Econometrics, vol. 1. North-Holland, Amsterdam.
Griliches, Z., Intriligator, M.D. (Eds.) (1984). Handbook of Econometrics, vol. 2. North-Holland, Amsterdam.
Griliches, Z., Intriligator, M.D. (Eds.) (1986). Handbook of Econometrics, vol. 3. North-Holland, Amsterdam.
Guilkey, D.K. (1974). "Alternative tests for a first order vector autoregressive error specification". Journal of Econometrics 2, 95–104.
Hall, S., Mitchell, J. (2005). "Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESRC fan charts of inflation". Oxford Bulletin of Economics and Statistics 67, 995–1033.
Hamilton, J.D. (1989). "A new approach to the economic analysis of nonstationary time series and the business cycle". Econometrica 57, 357–384.
Hamilton, J.D. (1990). "Analysis of time series subject to changes in regime". Journal of Econometrics 45, 39–70.
Hamilton, J.D. (1993). "Estimation, inference, and forecasting of time series subject to changes in regime". In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), Handbook of Statistics, vol. 11. North-Holland, Amsterdam.
Hamilton, J.D. (1994). Time Series Analysis. Princeton University Press, Princeton.
Hamilton, J.D., Raj, B. (Eds.) (2002). Advances in Markov-Switching Models. Applications in Business Cycle Research and Finance. Physica-Verlag, New York.
Hansen, B.E. (1992). "The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP". Journal of Applied Econometrics 7, S61–S82.
Hansen, B.E. (1996a). "Erratum: The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP". Journal of Applied Econometrics 11, 195–198.
Hansen, B.E. (1996b). "Inference when a nuisance parameter is not identified under the null hypothesis". Econometrica 64, 413–430.
Harvey, A.C. (1992). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.
Heckman, J.J., Leamer, E.E. (Eds.) (2004). Handbook of Econometrics, vol. 5. Elsevier Science, North-Holland, Amsterdam.
Hendry, D.F. (1995). Dynamic Econometrics. Oxford University Press, Oxford.
Hendry, D.F. (1996). "On the constancy of time-series econometric equations". Economic and Social Review 27, 401–422.
Hendry, D.F. (2000). "On detectable and non-detectable structural change". Structural Change and Economic Dynamics 11, 45–65. Reprinted in: Hagemann, H., Landesman, M., Scazzieri, R. (Eds.) (2002). The Economics of Structural Change. Edward Elgar, Cheltenham.
Hendry, D.F. (2001). "Modelling UK inflation, 1875–1991". Journal of Applied Econometrics 16, 255–275.
Hendry, D.F. (2005). "Robustifying forecasts from equilibrium-correction models". Special Issue in Honor of Clive Granger, Journal of Econometrics. In press.

Hendry, D.F., Clements, M.P. (2004). "Pooling of forecasts". The Econometrics Journal 7, 1–31.

Hendry, D.F., Doornik, J.A. (2001). Empirical Econometric Modelling Using PcGive 10, vol. I. Timberlake Consultants Press, London.

Hendry, D.F., Johansen, S., Santos, C. (2004). "Selecting a regression saturated by indicators". Unpublished Paper, Economics Department, University of Oxford.

Hendry, D.F., Massmann, M. (2006). "Co-breaking: Recent advances and a synopsis of the literature". Journal of Business and Economic Statistics. In press.

Hendry, D.F., Neale, A.J. (1991). "A Monte Carlo study of the effects of structural breaks on tests for unit roots". In: Hackl, P., Westlund, A.H. (Eds.), Economic Structural Change, Analysis and Forecasting. Springer-Verlag, Berlin, pp. 95–119.

Hoque, A., Magnus, J.R., Pesaran, B. (1988). "The exact multi-period mean-square forecast error for the first-order autoregressive model". Journal of Econometrics 39, 327–346.

Johansen, S. (1988). "Statistical analysis of cointegration vectors". Journal of Economic Dynamics and Control 12, 231–254. Reprinted in: Engle, R.F., Granger, C.W.J. (Eds.) (1991). Long-Run Economic Relationships. Oxford University Press, Oxford, pp. 131–152.

Johansen, S. (1994). "The role of the constant and linear terms in cointegration analysis of nonstationary variables". Econometric Reviews 13, 205–229.

Junttila, J. (2001). "Structural breaks, ARIMA model and Finnish inflation forecasts". International Journal of Forecasting 17, 207–230.

Kähler, J., Marnet, V. (1994). "Markov-switching models for exchange rate dynamics and the pricing of foreign-currency options". In: Kähler, J., Kugler, P. (Eds.), Econometric Analysis of Financial Markets. Physica Verlag, Heidelberg.

Kim, C.J. (1994). "Dynamic linear models with Markov-switching". Journal of Econometrics 60, 1–22.

Klein, L.R. (1971). An Essay on the Theory of Economic Prediction. Markham Publishing Company, Chicago.

Klein, L.R., Howrey, E.P., MacCarthy, M.D. (1974). "Notes on testing the predictive performance of econometric models". International Economic Review 15, 366–383.

Koop, G., Potter, S.M. (2000). "Nonlinearity, structural breaks, or outliers in economic time series". In Barnett et al. (2000), pp. 61–78.

Krämer, W., Ploberger, W., Alt, R. (1988). "Testing for structural change in dynamic models". Econometrica 56, 1355–1369.

Krolzig, H.-M. (1997). Markov Switching Vector Autoregressions: Modelling, Statistical Inference and Application to Business Cycle Analysis. Lecture Notes in Economics and Mathematical Systems, vol. 454. Springer-Verlag, Berlin.

Krolzig, H.-M., Lütkepohl, H. (1995). "Konjunkturanalyse mit Markov-regimewechselmodellen". In: Oppenländer, K.H. (Ed.), Konjunkturindikatoren. Fakten, Analysen, Verwendung. München Wien, Oldenbourg, pp. 177–196.

Lahiri, K., Wang, J.G. (1994). "Predicting cyclical turning points with leading index in a Markov switching model". Journal of Forecasting 13, 245–263.

Lam, P.-S. (1990). "The Hamilton model with a general autoregressive component. Estimation and comparison with other models of economic time series". Journal of Monetary Economics 26, 409–432.

Lamoureux, C.G., Lastrapes, W.D. (1990). "Persistence in variance, structural change, and the GARCH model". Journal of Business and Economic Statistics 8, 225–234.

Lin, J.-L., Tsay, R.S. (1996). "Co-integration constraint and forecasting: An empirical examination". Journal of Applied Econometrics 11, 519–538.

Lütkepohl, H. (1991). Introduction to Multiple Time Series Analysis. Springer-Verlag, New York.

Maddala, G.S., Li, H. (1996). "Bootstrap based tests in financial models". In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics, vol. 14. North-Holland, Amsterdam, pp. 463–488.

Makridakis, S., Hibon, M. (2000). "The M3-competition: Results, conclusions and implications". International Journal of Forecasting 16, 451–476.

Malinvaud, E. (1970). Statistical Methods of Econometrics, second ed. North-Holland, Amsterdam.

Marris, R.L. (1954). "The position of economics and economists in the Government Machine: A comparative critique of the United Kingdom and the Netherlands". Economic Journal 64, 759–783.
McCulloch, R.E., Tsay, R.S. (1994). "Bayesian analysis of autoregressive time series via the Gibbs sampler". Journal of Time Series Analysis 15, 235–250.
Newbold, P., Granger, C.W.J. (1974). "Experience with forecasting univariate time series and the combination of forecasts". Journal of the Royal Statistical Society A 137, 131–146.
Newbold, P., Harvey, D.I. (2002). "Forecasting combination and encompassing". In: Clements, M.P., Hendry, D.F. (Eds.), A Companion to Economic Forecasting. Blackwells, Oxford, pp. 268–283.
Nyblom, J. (1989). "Testing for the constancy of parameters over time". Journal of the American Statistical Association 84, 223–230.
Osborn, D. (2002). "Unit root versus deterministic representations of seasonality for forecasting". In: Clements, M.P., Hendry, D.F. (Eds.), A Companion to Economic Forecasting. Blackwells, Oxford, pp. 409–431.
Pastor, L., Stambaugh, R.F. (2001). "The equity premium and structural breaks". Journal of Finance 56, 1207–1239.
Perron, P. (1990). "Testing for a unit root in a time series with a changing mean". Journal of Business and Economic Statistics 8, 153–162.
Pesaran, M.H., Pettenuzzo, D., Timmermann, A. (2004). "Forecasting time series subject to multiple structural breaks". Mimeo, University of Cambridge and UCSD.
Pesaran, M.H., Timmermann, A. (2002a). "Market timing and return prediction under model instability". Journal of Empirical Finance 9, 495–510.
Pesaran, M.H., Timmermann, A. (2002b). "Model instability and choice of observation window". Mimeo, University of Cambridge.
Pesaran, M.H., Timmermann, A. (2003). "Small sample properties of forecasts from autoregressive models under structural breaks". Journal of Econometrics. In press.
Phillips, K. (1991). "A two-country model of stochastic output with changes in regime". Journal of International Economics 31, 121–142.
Phillips, P.C.B. (1994). "Bayes models and forecasts of Australian macroeconomic time series". In: Hargreaves, C. (Ed.), Non-Stationary Time-Series Analyses and Cointegration. Oxford University Press, Oxford.
Phillips, P.C.B. (1995). "Automated forecasts of Asia-Pacific economic activity". Asia-Pacific Economic Review 1, 92–102.
Phillips, P.C.B. (1996). "Econometric model determination". Econometrica 64, 763–812.
Ploberger, W., Krämer, W., Kontrus, K. (1989). "A new test for structural stability in the linear regression model". Journal of Econometrics 40, 307–318.
Potter, S. (1995). "A nonlinear approach to US GNP". Journal of Applied Econometrics 10, 109–125.
Quandt, R.E. (1960). "Tests of the hypothesis that a linear regression system obeys two separate regimes". Journal of the American Statistical Association 55, 324–330.
Rappoport, P., Reichlin, L. (1989). "Segmented trends and non-stationary time series". Economic Journal 99, 168–177.
Reichlin, L. (1989). "Structural change and unit root econometrics". Economics Letters 31, 231–233.
Sánchez, M.J., Peña, D. (2003). "The identification of multiple outliers in ARIMA models". Communications in Statistics: Theory and Methods 32, 1265–1287.
Schiff, A.F., Phillips, P.C.B. (2000). "Forecasting New Zealand's real GDP". New Zealand Economic Papers 34, 159–182.
Schmidt, P. (1974). "The asymptotic distribution of forecasts in the dynamic simulation of an econometric model". Econometrica 42, 303–309.
Schmidt, P. (1977). "Some small sample evidence on the distribution of dynamic simulation forecasts". Econometrica 45, 97–105.
Shephard, N. (1996). "Statistical aspects of ARCH and stochastic volatility". In: Cox, D.R., Hinkley, D.V., Barndorff-Nielsen, O.E. (Eds.), Time Series Models: In Econometrics, Finance and other Fields. Chapman and Hall, London, pp. 1–67.

Stock, J.H. (1994). "Unit roots, structural breaks and trends". In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics. North-Holland, Amsterdam, pp. 2739–2841.
Stock, J.H., Watson, M.W. (1996). "Evidence on structural instability in macroeconomic time series relations". Journal of Business and Economic Statistics 14, 11–30.
Stock, J.H., Watson, M.W. (1999). "A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series". In: Engle, R.F., White, H. (Eds.), Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive Granger. Oxford University Press, Oxford, pp. 1–44.
Swanson, N.R., White, H. (1997). "Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models". International Journal of Forecasting 13, 439–462.
Taylor, J.W., Bunn, D.W. (1998). "Combining forecast quantiles using quantile regression: Investigating the derived weights, estimator bias and imposing constraints". Journal of Applied Statistics 25, 193–206.
Teräsvirta, T. (1994). "Specification, estimation and evaluation of smooth transition autoregressive models". Journal of the American Statistical Association 89, 208–218.
Theil, H. (1961). Economic Forecasts and Policy, second ed. North-Holland, Amsterdam.
Tiao, G.C., Tsay, R.S. (1994). "Some advances in non-linear and adaptive modelling in time-series". Journal of Forecasting 13, 109–131.
Tong, H. (1983). Threshold Models in Non-Linear Time Series Analysis. Springer-Verlag, New York.
Tong, H. (1995). Non-Linear Time Series. A Dynamical System Approach. Clarendon Press, Oxford. First published 1990.
Tsay, R.S. (1986). "Time-series model specification in the presence of outliers". Journal of the American Statistical Association 81, 132–141.
Tsay, R.S. (1988). "Outliers, level shifts and variance changes in time series". Journal of Forecasting 7, 1–20.
Turner, D.S. (1990). "The role of judgement in macroeconomic forecasting". Journal of Forecasting 9, 315–345.
Wallis, K.F. (1993). "Comparing macroeconometric models: A review article". Economica 60, 225–237.
Wallis, K.F. (2005). "Combining density and interval forecasts: A modest proposal". Oxford Bulletin of Economics and Statistics 67, 983–994.
Wallis, K.F., Whitley, J.D. (1991). "Sources of error in forecasts and expectations: UK economic models 1984–88". Journal of Forecasting 10, 231–253.
Wallis, K.F., Andrews, M.J., Bell, D.N.F., Fisher, P.G., Whitley, J.D. (1984). Models of the UK Economy, A Review by the ESRC Macroeconomic Modelling Bureau. Oxford University Press, Oxford.
Wallis, K.F., Andrews, M.J., Bell, D.N.F., Fisher, P.G., Whitley, J.D. (1985). Models of the UK Economy, A Second Review by the ESRC Macroeconomic Modelling Bureau. Oxford University Press, Oxford.
Wallis, K.F., Andrews, M.J., Fisher, P.G., Longbottom, J., Whitley, J.D. (1986). Models of the UK Economy: A Third Review by the ESRC Macroeconomic Modelling Bureau. Oxford University Press, Oxford.
Wallis, K.F., Fisher, P.G., Longbottom, J., Turner, D.S., Whitley, J.D. (1987). Models of the UK Economy: A Fourth Review by the ESRC Macroeconomic Modelling Bureau. Oxford University Press, Oxford.
White, H. (1980). "A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity". Econometrica 48, 817–838.
White, H. (1992). Artificial Neural Networks: Approximation and Learning Theory. Oxford University Press, Oxford.