DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

An Overview of Forecasting Facing Breaks

Jennifer L. Castle, Michael P. Clements and

David Hendry

Manor Road Building, Oxford OX1 3UQ

An Overview of Forecasting Facing Breaks

Jennifer L. Castle[1], Michael P. Clements[2] and David F. Hendry[1]
[1] Department of Economics, and Institute for New Economic Thinking,
Oxford Martin School, University of Oxford, UK, and
[2] ICMA Centre, University of Reading, UK.[*]

February 1, 2016

Abstract

Economic forecasting may go badly awry when there are structural breaks, such that the relationships between variables that held in the past are a poor basis for making predictions about the future. We review a body of research that seeks to provide viable strategies for economic forecasting when past relationships can no longer be relied upon.

Keywords: Business Cycles; Forecasting; Breaks.
JEL classifications: C51, C22.

1 Overview

The 'Great Recession', 2008–2012 has forcefully reminded us that the Business Cycle is most certainly not dead. As Victor Zarnowitz (2004) expressed it:
'Business cycles, even if less of a threat, are far from conquered and still represent the most serious form of macroeconomic instability'
Thanks to China's growth and automatic stabilizers in 'advanced' economies, the downturn and resulting distress were not on the scale of the Great Depression of the 1930s, but it has still proved difficult to resolve. Moreover, while forecasting remains central to the policy process in most OECD economies, forecast failure—a significant deterioration in forecast performance relative to the anticipated outcome— was a common occurrence over the 'Great Recession'. We consider such a failure for UK GDP below, and contrast that with the ease with which unemployment could be forecasted, suggesting a change in the relationship between these key variables.

From the early days of model-based economic forecasting, the difficulties posed by breaks have been recognized: salient examples include Smith (1929) (*Judging the Forecast for 1929*, published in early 1929), Shoup, Friedman and Mack (1941) ('The times are so different now [i.e., October 1941] from 1935–39 that relations existing then may not exist at all today.'– even prior to the USA entering World War II), and Klein (1947) ('Would the econometrician merely substitute into his equations of peacetime behavior patterns in order to forecast employment in a period during which there will be a war?').

The onsets of Business Cycle downturns are typically not easy to forecast. The 'Great Recession' is but the latest of many historical episodes of forecast failure, revealing problems with the traditional approach to economic forecasting. We have argued that structural breaks are the main culprit: other

putative causes of forecast failure, such as model mis-specification, turn out to be relatively benign in the absence of breaks. Fortunately, there are partial remedies, which fall into two broad groups: 1) automatic devices for robustifying forecasts, as the forecast origin moves forward in time, to avoid systematic failure after future unknown breaks occur, and 2) forecasting as the break unfolds, when there is partial information on the changes that are taking place.

Here we review some of these developments, and illustrate the relative forecasting performances of models during periods when the economy was subject to substantive upheavals, along with the impacts of applying various robustification strategies (remedy 1 above). We do not attempt to forecast using additional information on breaks, although relevant research on this is discussed. The strategies we explore are not always successful and typically come with a cost—inflated forecast-error variances—but have the potential to prevent **systematic** runs of forecast errors of the same sign.

In section 2, we review a traditional approach, explaining what might go wrong when there are breaks. Sections 3 and 4 respectively sketch why model mis-specification by itself rarely causes forecast failure, but why structural breaks do. The discussion of location shifts in section 4 serves to motivate one potential remedy, discussed in section 5, with some alternatives noted. Section 6 briefly considers research on forecasting during a break, using the ever-increasing amounts of data that are available (including social media data) which create the potential for more accurate readings of the state of the economy at the forecast origin. Section 7 presents our two illustrations: forecasts of output growth over the recent recession and of UK unemployment. For the unemployment rate, we take a longer-run view, but also consider a short-sample example and note two earlier episodes where we find that well-specified models in-sample are not the best models out-of-sample. Section 8 concludes.

2   Two theories to economic forecasting

The traditional theory of economic forecasting assumes, at least implicitly, that (see, e.g., Klein, 1971):

1. the forecasting model is a good representation of the process; and

2. the structure of the economy will remain relatively unchanged over the forecast horizon.

Under these assumptions, the natural forecasting strategy, or *operational procedure for forecasting*, is simply to use the best in-sample model, estimated from the best available data. What we have termed forecast failure ought not occur: out-of-sample performance should be broadly similar to how well the model fits the in-sample data. In section 3, we provide a simple illustration to clarify these statements. Unfortunately, as recognized by a number of authors, econometric models are inevitably mis-specified and economies are subject to unanticipated shifts (see, e.g., Stock and Watson, 1996) so episodes of forecast failure have been all too common. As Friedman (2014) recounts, the unpredicted onset of the Great Depression in the USA did great damage to the reputation of economic forecasting by what he calls 'Fortune Tellers'.

To overcome the limitations of traditional forecasting theory, Clements and Hendry (1998, 1999) make two less stringent assumptions that seem more realistic:

1. models are simplified representations, incorrect in many ways; and

2. economies both evolve and occasionally shift abruptly.

In this setting, simply using the best in-sample model may not be the best approach, and as shown by e.g., Clements and Hendry (2006) and Castle, Fawcett and Hendry (2010), forecasting by popular (vector) equilibrium-correction models with well-defined long-run solutions may be especially harmful when there are shifts in equilibrium means. Models which are deliberately mis-specified in-sample,

for example, by omitting equilibrium (or error)-correcting terms, may adapt more rapidly to changed circumstances out-of-sample and produce more accurate forecasts.

It seems fairly intuitive that forecast failure could result from structural breaks that render past relationships between variables a poor guide to the future. It is perhaps less clear that many forms of model mis-specification, including unmodelled changes in parameter values, need not do so. In section 4, we examine the effects of shifts in equilibrium-correction models to motivate one of the robust strategies we subsequently discuss.

3 Model mis-specification and lack of forecast failure

In a stationary world, least-squares estimated equations are consistent for their associated conditional expectations (when second moments exist), so forecasts on average attain their expected accuracy unconditionally (see, e.g., Miller, 1978, and Hendry, 1979). Clements and Hendry (2002, p. 550-2) also illustrate that model mis-specification need not result in forecast failure in the absence of structural breaks.

Providing the data under analysis are and remain stationary, then any model thereof is isomorphic to a mean-zero representation, a well-known result in elementary regression derivations, and a consequence of the famous Frisch and Waugh (1933) theorem. Omitting any subset of variables in any equation will not bias its forecasts because the omission is of zero-mean terms. Let the stationary data generation process (DGP) for the variable, y_t to be forecast be given by:

$$y_t = \beta_0 + \sum_{i=1}^{N} \beta_i z_{i,t-1} + \epsilon_t = \theta_0 + \sum_{i=1}^{N} \beta_i \left(z_{i,t-1} - \kappa_i \right) + \epsilon_t \qquad (1)$$

where $\beta_0, \ldots, \beta_N$ are constant, κ_i is the population mean of z_i, so $\theta_0 = \beta_0 + \sum_{i=1}^{N} \beta_i \kappa_i$ and $\epsilon_t \sim$ $\mathsf{IN}[0, \sigma_\epsilon^2]$, denoting an independent normal random variable with mean $\mathsf{E}[\epsilon_t] = 0$ and variance $\mathsf{V}[\epsilon_t] = \sigma_\epsilon^2$ that is independent of all the $z_{i,t-1}$. The parameters in (1) can be unbiasedly estimated from a sample $t = 1, \ldots, T$. From (1), when the distributions remain the same over a forecast horizon, $h = T + 1, \ldots, T + H$, then forecasting by:

$$\widehat{y}_{T+h|T+h-1} = \widehat{\theta}_0 + \sum_{i=1}^{N} \widehat{\beta}_i \left(z_{i,T+h-1} - \overline{z}_i \right) \qquad (2)$$

where $\overline{z}_i$ is the sample mean of z_i, leads to $\mathsf{E}[y_{T+h} - \widehat{y}_{T+h|T+h-1}] = 0$.

However, the researcher only includes $z_{i,t-1}, i = 1, \ldots, M < N$ in her forecasting model, unaware that $z_{j,t-1}, j = M + 1, \ldots, N$ matter, so forecasts by:

$$\widetilde{y}_{T+h|T+h-1} = \widetilde{\theta}_0 + \sum_{i=1}^{M} \widetilde{\beta}_i \left(z_{i,T+h-1} - \overline{z}_i \right) \qquad (3)$$

Since all the included $(z_{i,T+h-1} - \overline{z}_i)$ still have zero means, and hence $\mathsf{E}[\widetilde{\theta}_0] = \theta_0$, then $\mathsf{E}[y_{T+h} - \widetilde{y}_{T+h|T+h-1}] = 0$. The fit will be inferior to the correctly-specified representation (2), but the resulting model will on average forecast according to its in-sample operating characteristics. Thus, a test for forecast failure based on comparing the in-sample fit of a model like (3) to its out-of-sample forecast performance will not find any indication of forecast failure. In fact, forecasts from mis-specified models may be more or less accurate than those from the estimated DGP depending on the precision with which parameters are estimated, since although they are technically invalid, zero restrictions on coefficients that are close to zero can improve forecast accuracy: see Clements and Hendry (1998, chs.11 & 12).

Thus, model mis-specification *per se* cannot account for forecast failure in stationary processes, because the model's out-of-sample forecast performance will be consistent with what would have been expected based on how well the model fitted the historical data. However, an exception arises to the extent that inconsistently estimated standard errors are used to judge forecast accuracy, or if deterministic terms are mis-specified, violating the mean-zero requirement, as we now discuss. Corsi, Pollock and Prakken (1982) show that residual autocorrelation, perhaps induced by other mis-specifications, leads to excess rejection on parameter-constancy tests as untreated positive residual autocorrelation can downward bias estimated standard errors, which thereby induces excess rejections on constancy tests. In practice, an investigator is likely to add extra lags to remove any residual autocorrelation. This strategy will help make the model congruent (see, e.g., Hendry, 1995) even though it remains mis-specified: by having innovation residuals, excess rejections in parameter constancy tests will not occur. Finally, when $\theta_0 \neq 0$, a failure to include an intercept will lead to biased forecasts from models with either N or M variables.

Conversely, model mis-specification is *necessary* for forecast failure, because otherwise the model coincides with the DGP at all points in time, so never fails. This is consistent with the result in Clements and Hendry (1998) that causal variables will always dominate over non-causal in forecasting when the model coincides with the DGP (or that DGP is stationary), but need not do so when the model is mis-specified for a DGP that is subject to location shifts, the topic we now address. We illustrate the impact of shifts in $\theta_0 \neq 0$ when the forecasting model is the actual DGP (1), then consider a cointegrated system which provides a more realistic representation of the situation confronting forecasting business cycles.

4   Structural breaks and forecast failure

Our intrepid researcher has managed to discover the exact in-sample DGP as in (1):

$$y_t = \theta_0 + \sum_{i=1}^{N} \beta_i \left(z_{t-1} - \kappa_i\right) + \epsilon_t \tag{4}$$

but at the forecast origin T, the DGP shifts to:

$$y_{T+h} = \theta_0^* + \sum_{i=1}^{N} \beta_i \left(z_{i,T+h-1} - \kappa_i\right) + \epsilon_{T+h} \tag{5}$$

Denoting her forecasts from (4) by $\widehat{y}_{T+h|T+h-1}$, then $\mathsf{E}[y_{T+h} - \widehat{y}_{T+h|T+h-1}] = \theta_0^* - \theta_0 \neq 0$, so are biased, and will remain biased until she changes her 'estimate' of θ_0. The change from θ_0 to θ_0^* is a location shift, as the mean value of y_t shifts from the former value to the latter.

Contrast that outcome of failure from one parameter shifting, namely the long-run mean of the dependent variable, to θ_0 staying constant, but every other parameter changing, so all β_i, κ_i, $i = 1, \ldots, N$ shift to β_i^*, κ_i^*, $i = 1, \ldots, N$ where $\mathsf{E}[z_{i,T+h-1}] = \kappa_i^*$. Then $\mathsf{E}[y_{T+h} - \widehat{y}_{T+h|T+h-1}] = \theta_0 - \theta_0 = 0$, so there is no systematic bias, but an increase in the forecast-error variance. Moreover, this result continues to hold even if the forecasting model omits some of the relevant variables. Clearly, not all shifts are equal, but matters get even stranger when we consider a cointegrated system.

We now borrow the model and notation from Castle, Clements and Hendry (2015) to show that vector equilibrium correction models (henceforth, VEqCMs) are not robust when forecasting after breaks, specifically, unanticipated location shifts, and are then liable to systematic forecast failure. Similar analyses and extensions are provided in Clements and Hendry (1999, 2006) *inter alia*.

We consider an n-vector time series $\{\mathbf{x}_t, t = 0, 1, \ldots, T\}$ generated by the cointegrated system:

$$\Delta \mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\alpha} \left(\boldsymbol{\beta}' \mathbf{x}_{t-1} - \boldsymbol{\mu}\right) + \boldsymbol{\phi}'(\mathbf{z}_{t-1} - \boldsymbol{\kappa}) + \boldsymbol{\epsilon}_t \tag{6}$$

where $\epsilon_t \sim \mathsf{IN}_n[\mathbf{0}, \mathbf{\Omega}_\epsilon]$. In addition to lags of the $\Delta\mathbf{x}_t$s that are not explicitly shown for simplicity, $\Delta\mathbf{x}_t$ depends on k explanatory variables denoted $\mathbf{z}_t$, which may include variables other than those in $\Delta\mathbf{x}_{t-i}$, and/or principal components or factors, as discussed in Castle, Clements and Hendry (2013). Either way, the $\mathbf{z}_t$ are assumed to be integrated of order zero, denoted $\mathsf{I}(0)$, so that the form of the model implies that $\mathbf{x}_t$ is $\mathsf{I}(1)$, with r linear combinations $\boldsymbol{\beta}'\mathbf{x}_t$ that cointegrate, so are $\mathsf{I}(0)$, where $\boldsymbol{\beta}$ is n by $r < n$.[1] Hence $\Delta\mathbf{x}_t$ responds to disequilibria between $\mathbf{z}_{t-1}$ and its mean $\mathsf{E}[\mathbf{z}_{t-1}] = \boldsymbol{\kappa}$, so the DGP is equilibrium-correcting in the $\mathbf{z}_t$, as well as to disequilibria between $\boldsymbol{\beta}'\mathbf{x}_{t-1}$ and $\boldsymbol{\mu}$. In this setup, we can analyze the effects of model mis-specification by supposing that the forecasting model omits $\mathbf{z}_{t-1}$, typically because the investigator is unaware of its relevance. In (6), both $\Delta\mathbf{x}_t$ and $\boldsymbol{\beta}'\mathbf{x}_t$ are $\mathsf{I}(0)$, with average growth $\mathsf{E}[\Delta\mathbf{x}_t] = \boldsymbol{\gamma}$ in-sample and equilibrium mean $\mathsf{E}[\boldsymbol{\beta}'\mathbf{x}_t] = \boldsymbol{\mu}$.

Consider now a forecasting model that omits $\mathbf{z}_{t-1}$. The estimated forecasting model becomes:

$$\Delta\widehat{\mathbf{x}}_t = \widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_{t-1} - \widehat{\boldsymbol{\mu}}\right) \tag{7}$$

where $\mathsf{E}[\widehat{\boldsymbol{\gamma}}] = \boldsymbol{\gamma}$ and $\mathsf{E}[\widehat{\boldsymbol{\mu}}] = \boldsymbol{\mu}$ because although the model is mis-specified by omitting $\mathbf{z}_{t-1}$, its effect has a mean of zero. We also suppose that the population value of $\widehat{\boldsymbol{\alpha}}$ is still $\boldsymbol{\alpha}$, despite omitting $\mathbf{z}_{t-1}$, as well as that the population value of $\widehat{\boldsymbol{\beta}}$ remains $\boldsymbol{\beta}$, since cointegrating relationships are little affected by the inclusion or omission of $\mathsf{I}(0)$ variables. Referring back to section 3, providing $\mathsf{E}[\mathbf{z}_{T+h-1}] = \boldsymbol{\kappa}$ over the forecast horizon, this omission will not even bias the forecasts, though it will increase the forecast-error variance.

When there are shifts in the means, so that $\boldsymbol{\gamma}$, $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$ shift to $\boldsymbol{\gamma}^*$, $\boldsymbol{\mu}^*$ and $\boldsymbol{\kappa}^*$ at the forecast origin at time T, the DGP becomes:

$$\Delta\mathbf{x}_{T+1} = \boldsymbol{\gamma}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}^*\right) + (\boldsymbol{\phi}^*)'\left(\mathbf{z}_T - \boldsymbol{\kappa}^*\right) + \boldsymbol{\epsilon}_{T+1} \tag{8}$$

where we have allowed the coefficient vector of the omitted variables to change as well. Then the 1-step ahead forecasts from using (7) to forecast $\Delta\mathbf{x}_{T+1}$ from period T are given by:

$$\Delta\widehat{\mathbf{x}}_{T+1|T} = \widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_T - \widehat{\boldsymbol{\mu}}\right) \tag{9}$$

and the forecasts errors, given by subtracting (9) from (8), have a mean of:

$$(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}) - \boldsymbol{\alpha}\left(\boldsymbol{\mu}^* - \boldsymbol{\mu}\right) - (\boldsymbol{\phi}^*)'\left(\boldsymbol{\kappa}^* - \boldsymbol{\kappa}\right). \tag{10}$$

A closely related issue is the accuracy of data at the forecast origin, as an incorrect value $\widehat{\mathbf{x}}_T$ for $\mathbf{x}_T$ acts like a location shift at the forecast origin. Consequently, nowcasting to produce more accurate values at the forecast origin can be valuable. We do not address real-time forecasting here in order to focus on the impact of shifts. However, in real-time the robustification strategies discussed below will typically be based on data measured with error, subject to subsequent revision as later vintages are released, which might curtail their efficacy in practice, an issue considered for nowcasting in Castle, Fawcett and Hendry (2009) and by Castle *et al.* (2015), who analyse the impact of measurement errors at the forecast origin. Further research in this area is warranted given the relevance of data revisions for macro-data.

When (9) is still used to forecast the outcomes from (8) 1-step ahead even after several periods have elapsed:

$$\Delta\widehat{\mathbf{x}}_{T+h|T+h-1} = \widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-1} - \widehat{\boldsymbol{\mu}}\right) \tag{11}$$

then the resulting forecast error $\widehat{\boldsymbol{\epsilon}}_{T+h|T+h-1} = \Delta\mathbf{x}_{T+h} - \Delta\widehat{\mathbf{x}}_{T+h|T+h-1}$ remains biased as:

$$\mathsf{E}\left[\widehat{\boldsymbol{\epsilon}}_{T+h|T+h-1}\right] = (\boldsymbol{\gamma}^* - \boldsymbol{\gamma}) - \boldsymbol{\alpha}\left(\boldsymbol{\mu}^* - \boldsymbol{\mu}\right) - (\boldsymbol{\phi}^*)'\left(\boldsymbol{\kappa}^* - \boldsymbol{\kappa}\right) \tag{12}$$

---

[1]See, for example, Johansen (1988), or Banerjee, Dolado, Galbraith and Hendry (1993) for textbook treatments of integration and cointegration; and Hendry and Juselius (2000, 2001) for surveys.

Even assuming $\mathsf{E}[\mathbf{z}_{T+h-1}] = \boldsymbol{\kappa}^* = \boldsymbol{\kappa}$, the first two components in (12) will continue to cause systematic forecast failure. The problem is that the model lacks adaptability, a difficulty for all members of the equilibrium-correction class including regressions, vector autoregressions (VARs) as well as cointegrated systems: the equilibrium correction always corrects back to the *old* equilibrium, determined by $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$, **irrespective of how much the new equilibrium has shifted**.

Importantly, forecast failure does **not** require a model to be mis-specified in-sample, seen by forecasting from the pre-break DGP. The resulting forecasts are given by:

$$\Delta\widehat{\mathbf{x}}_{T+h|T+h-1} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+h-1} - \boldsymbol{\mu}\right) + \boldsymbol{\phi}'(\mathbf{z}_{T+h-1} - \boldsymbol{\kappa}) \tag{13}$$

so even when $\boldsymbol{\kappa}^* = \boldsymbol{\kappa}$, has a forecast error of:

$$\mathsf{E}\left[\widehat{\boldsymbol{\epsilon}}_{T+h|T+h-1}\right] = (\boldsymbol{\gamma}^* - \boldsymbol{\gamma}) - \boldsymbol{\alpha}\left(\boldsymbol{\mu}^* - \boldsymbol{\mu}\right) \tag{14}$$

which has the same form as (10) when $\boldsymbol{\kappa}^* = \boldsymbol{\kappa}$, creating persistent forecast errors of the same sign as the forecast origin moves through time despite using the in-sample DGP.

These examples treat the parameters as 'variation free' in that each can be shifted separately from any of the others. That is unlikely in practice, as e.g., a fall in the equilibrium mean is liable to alter the growth rate. Hence, although the terms in (14) could in principal cancel, that does not seem likely. Similarly, changes in $\boldsymbol{\kappa}$ are likely to alter $\boldsymbol{\mu}$. Finally, although we have implemented the impacts of shifts as instantaneous, they are more than likely to take time to complete in dynamic systems, so that, e.g., in the early stages after a shift from $\boldsymbol{\kappa}$ to $\boldsymbol{\kappa}^*$, one would anticipate that $\mathsf{E}[\mathbf{z}_{T+1}] \neq \boldsymbol{\kappa}^*$. Thus, location shifts will usually not produce a neat step in observable data, but smoother responses of the shape often seen in time series.

5   Robust forecasting devices

Robustification against the adverse effects of a 'break' on forecasts is a form of 'insurance', in that there is a cost, but the strategies confer benefits in a bad state of nature. The cost typically manifests in a higher forecast-error variance, and the benefits are largely unbiased forecasts subsequent to the occurrence of breaks. Changes in the probability of bad states and the costs of insuring affect the efficacy of robustification strategies, which therefore depends on factors such as the frequencies and magnitudes of location shifts, the underlying predictability of the series, amongst other things. We consider a number of robustification strategies for forecasting after location shifts, as well as improved adaptability to breaks, and averaging across forecasting devices or information sources. Robustification includes: intercept correction, differencing, forecast-error correction mechanisms, and pooling:

Intercept correction (IC) is a widely-adopted strategy, which can offset location shifts. Unfortunately, ICs can also exacerbate a break, as with stochastic regime-switching processes where a correction is in-built, and may require pre-testing for inclusion each period in every equation, yet the forms, timings, and durations of shifts are unknown. Although a work-horse of forecasters, ICs can be inadequate: many past failures occurred despite their use.

Differencing robustifies forecasts after location shifts, partly by adding a unit root, and partly by reducing deterministic polynomials by one order. In the literature, strategies have typically focused on differencing prior to model specification and estimation, as in Box and Jenkins (1970) and differenced vector autoregressions. We consider the impact of differencing *after* estimation, so parts of models (such as their equilibrium-correction feedbacks) are transformed to differences in forecast mode, which increases forecast-error variances, but may reduce bias in the face of location shifts.

Forecast-error correction mechanisms (FErCMs): a classic FErCM is the exponentially-weighted moving-average model, which does well in forecasting competitions (see, e.g., Makridakis and Hibon, 2000) and

while primarily designed to correct recent past measurement errors, is also relatively robust to location shifts.

Forecast pooling has generated a vast literature—see Clemen (1989) for an early bibliography. Hendry and Clements (2004) show that when unanticipated location shifts have different effects on differently mis-specified models, the pooled model may give more accurate forecasts.

Information pooling is an alternative to pooling forecasts. Current approaches include diffusion indices and factor models: see Stock and Watson (1998), Forni, Hallin, Lippi and Reichlin (2000) and Castle *et al.* (2013).

As an example of a robustification strategy, we consider differencing, using the model and notation established in section 4. Suppose instead of forecasting with the estimated model, and calculating forecasts using (11), we take the first difference of the estimated model, yielding for 1-step forecasts:

$$\widetilde{\Delta \mathbf{x}}_{T+h|T+h-1} = \Delta \mathbf{x}_{T+h-1} + \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'\Delta \mathbf{x}_{T+h-1}. \tag{15}$$

An immediately apparent feature of (15) is the absence of the parameters shifted above; less obvious at a glance is that (15) is double differenced, in that $\Delta^2 \mathbf{x}_{T+h}$ is being forecast by the difference of the equilibrium-correction term. Castle *et al.* (2015) provide a number of possible interpretations of this differencing procedure, and of why the resulting forecasts might be more accurate. One suggestion they offer is to re-write the expression in (15) as:

$$\widetilde{\Delta \mathbf{x}}_{T+h|T+h-1} = \Delta \mathbf{x}_{T+h-1} + \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-1} - \widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-2}\right) = \widetilde{\boldsymbol{\gamma}} + \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-1} - \widetilde{\boldsymbol{\mu}}\right) \tag{16}$$

and then regard $\Delta \mathbf{x}_{T+h-1}$ as a highly adaptive estimator $\widetilde{\boldsymbol{\gamma}}$ of the current growth rate, and the previous value of the cointegrating combination, $\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-2} = \widetilde{\boldsymbol{\mu}}$ as an estimator of $\boldsymbol{\mu}$. We use $\widetilde{\boldsymbol{\gamma}}$ and $\widetilde{\boldsymbol{\mu}}$ as shorthand for these estimates, although they both depend on T and h. In this interpretation, both $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$ are replaced by instantaneous estimators that are unbiased both before and some time after the population parameters have shifted, since for $h > 2$, $\mathsf{E}[\boldsymbol{\beta}'\mathbf{x}_{T+h-2}] \simeq \boldsymbol{\mu}^*$ and $\mathsf{E}[\Delta \mathbf{x}_{T+h-1}] \simeq \boldsymbol{\gamma}^*$.

This reinterpretation of the approach originally due to Hendry (2006) suggests a *class* of forecasting devices given by:

$$\overline{\Delta \mathbf{x}}_{T+h|T+h-1} = \frac{1}{r}\sum_{i=1}^{r}\Delta \mathbf{x}_{T+h-i} + \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-1} - \frac{1}{m}\sum_{i=1}^{m}\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-1-i}\right) \tag{17}$$

where the instantaneous estimates $\widetilde{\boldsymbol{\gamma}}$ and $\widetilde{\boldsymbol{\mu}}$ are replaced by local averages when $r > 1$ and $m > 1$. The performance of these differenced models is considered in section 7. Castle *et al.* (2015) analyze the performance of these forecasts under a number of scenarios, including when the DGP is unchanged, so that the original model's forecasts are optimal, when there are measurement errors, and crucially, when there are location shifts, as in section 4. For the latter, they find that these strategies reduce root mean-square forecast errors (RMSFEs) when the parameter shift is 'large' relative to the variance of the disturbance term in the original model. Further, when r and m span the available sample, the resulting estimators are close to $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\mu}}$.

Finally, differencing the model as in (15) is equivalent to a particular form of intercept correction, specifically, adding the estimated model residual at the forecast origin to the forecast. To see this, write (15) as:

$$\widetilde{\Delta \mathbf{x}}_{T+h|T+h-1} = \Delta \mathbf{x}_{T+h-1} - \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-2} - (\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\mu}}) + \left[\widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-1} + (\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\mu}})\right]$$

$$= \widehat{\boldsymbol{\epsilon}}_{T+h-1|T+h-2} + \Delta \widehat{\mathbf{x}}_{T+h|T+h-1} \tag{18}$$

where we have added and subtracted $\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\mu}}$ on the right-hand side. Intercept corrections of this form were considered as a possible remedy by Clements and Hendry (1996), but have a long history as 'add factors', or a means of putting the forecasts 'back on track' in macro-modelling.

6 Partial information to help to forecast a break

In some instances, it may be possible to forecast a break, but that requires (i) the break to be predictable; (ii) there is information relevant to that predictability; (iii) the information is available at the forecast origin; (iv) the forecasting model already embodies that source of information; (v) there is an operational method for selecting an appropriate model; and (vi) the resulting forecasts are usefully accurate. Castle, Fawcett and Hendry (2011) consider the conditions under which this will be possible, and in so doing distinguish between two information sets: one for 'normal forces' and one for 'break drivers'. The break drivers need not be conventional economic data, but could encompass legislative changes, acts of terrorism, war or natural disasters, or other events. High-frequency information such as *Google Trends* and prediction markets may also be useful for determining shifts. Break drivers could be modelled as a non-linear ogive, so would need to feature within non-linear, dynamic models with multiple breaks, leading to the possibility that there may be more candidate variables, N, than observations, T. Automatic model selection algorithms, available in software such as *Autometrics*, allow for $N > T$, and so facilitate such an approach (see Doornik, 2009a, 2009b).

When, as often seems likely, accurate forecasting of breaks is not possible, it may be possible to model breaks during their progress, perhaps by threshold models as in Teräsvirta, Tjøstheim and Granger (2011). However, Castle, Fawcett and Hendry (2011) find that modelling the progress of a break requires theoretical assumptions about the shape of the break function and restrictions on the number of its parameters to be estimated, so after a location shift, is not much better than using robust devices.

Indicator saturation using flexible break functions provides a new possibility when similar breaks have occurred previously. Pretis, Schneider, Smerdon and Hendry (2015) apply this approach in dendrochronology to estimate the impacts of volcanic eruptions on temperature. Modelling breaks at or near the forecast origin can also improve nowcasting: see *inter alia*, Castle and Hendry (2010), Marcellino and Schumacher (2010), and the survey article in Bánbura, Giannone and Reichlin (2011). Breaks at the forecast origin with just one data point are observationally equivalent to measurement errors, which may be revised later. This is an important problem for nowcasting, as 'corrections' for breaks or measurement errors work in opposite directions, so incorrect attribution can exacerbate a nowcast error, although higher frequency data and the behaviour of revisions can help distinguish the source after a few periods.

7 Empirical illustration

We provide two illustrations of forecasting. The first considers UK output growth, denoted Δy_t during the great recession, and the second extends the example of forecasting the UK unemployment rate, $U_{r,t}$, in Clements and Hendry (2006). That the second half of the last century and first decade of this century were punctuated by relatively sudden shifts is evident from Figure 1, which plots the annual changes in quarterly UK real GDP over this period, as well as the dates of estimated shifts in the growth rate. Many of the shifts found correspond to major unanticipated policy changes, so even the growth rate is not a stationary process, although its time series 'looks like' erratic cyclical behaviour.

The 'well-specified' models we consider include univariate (AR) and vector autoregressions (VARs), and cointegrated systems, VEqCMs, selected by *Autometrics* as parsimonious, congruent reductions of a more general unrestricted model (GUM), allowing for in-sample outliers and location shifts. Outliers and shifts are detected by impulse-indicator saturation (IIS: see Hendry, Johansen and Santos, 2008, Johansen and Nielsen, 2009, and Castle, Doornik and Hendry, 2012) and step-indicator saturation respectively (SIS: see Castle, Doornik, Hendry and Pretis, 2015). Thus, if $\left\{1_{\{j=t\}}, t = 1, \ldots, T\right\}$ denotes the complete set of T impulse indicators and $\left\{1_{\{j \le t\}}, t = 2, \ldots, T\right\}$ the corresponding set of $T - 1$ step indicators, then $\left\{1_{\{j=t\}}, t = 1, \ldots, T\right\}$ and $\left\{1_{\{j \le t\}}, t = 2, \ldots, T\right\}$ are included in the initial set of candidate variables in the GUM to search over. Exemplars of the robust device class include random walks
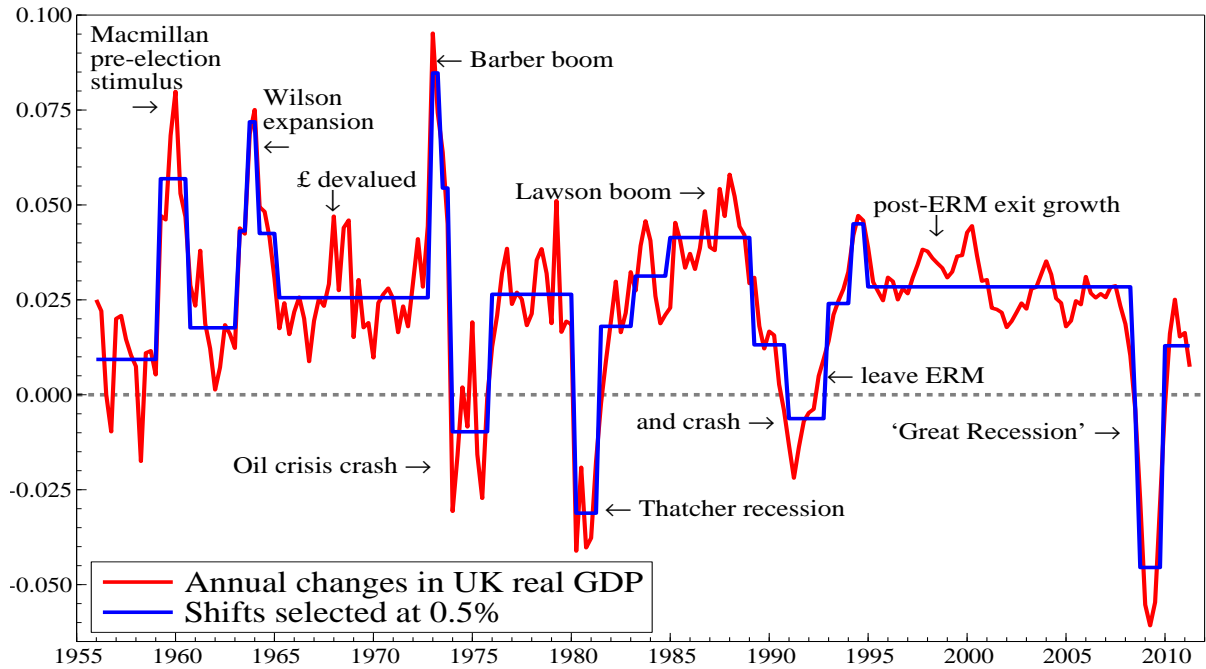
Figure 1: Year-on-year changes in quarterly UK real GDP, $\Delta_4 y_t$, with steps selected by SIS at 0.5%.

(RWs), double differenced devices like (15) (DDDs), and smoothed variants thereof as in (17) above (SDDs), differenced VARs (DVARs) and linear combinations of all of these (Pooled), defined in Table 1.

Forecasts for variables denoted $x_{T+j}, j = 1, \ldots, H$ are evaluated using 1-step biases and RMSFEs:

$$\text{Bias} = \frac{1}{H} \sum_{j=1}^{H} \left(x_{T+j} - \widehat{x}_{T+j|T+j-1} \right)$$

$$\text{RMSFE} = \sqrt{\frac{1}{H} \sum_{j=1}^{H} \left(x_{T+j} - \widehat{x}_{T+j|T+j-1} \right)^2}$$

where x_t is $\Delta_4 y_t$, Δy_t or $U_{r,t}$.[2] RMSFE is valid for 1-step ahead forecasts, but Clements and Hendry (1993a, 1993b) propose using the generalised forecast-error second-moment criterion (GFESM) which accounts for covariances between forecasts at different horizons: Hendry and Martinez (2015) provide an extension.

7.1 Output Growth

We apply a range of forecasting models to UK GDP growth over the 'Great Recession'. We use the preliminary release of GDP data available at the end of the forecast horizon, so although the forecasts are not based on real-time data, they replicate the information available in 2011Q2, so *ex post* revisions only available after the forecast horizon are not used. We focus on a single vintage rather than real-time data to directly analyse the impact of structural breaks. Although in practice policy-makers must work with real-time data, using the time-series dimension of real-time vintages includes any changes

---

[2]Detailed results are available on request.

in methodology, concepts or classifications, inducing apparent structural breaks which are an artifact of changing the measurement system.[3] By analysing a single vintage we abstract from these additional important effects. For the multivariate forecasts, we extend the dataset to include a measure of inflation and interest rates.[4] Table 1 summarises the forecasting models examined.

AR	GUM:$\Delta_4 y_t = \beta_0 + \sum_{i=1}^{8} \beta_i \Delta_4 y_{t-i} + \epsilon_t$.
	Lags selected using *Autometrics* at 5%.
AR_IS	GUM:$\Delta_4 y_t = \beta_0 + \sum_{i=1}^{8} \beta_i \Delta_4 y_{t-i} + \sum_{j=1}^{T} \delta_j 1_{\{j=t\}} + \sum_{j=2}^{T} \delta_{s,j} 1_{\{j \leq t\}} + \epsilon_t$.
	Selection of lags and indicators using *Autometrics* at 0.1%.
AR_IC; AR_IS_IC	$\Delta_4 \widetilde{y}_{\iota,T+h\|T+h-1} = \Delta_4 \widehat{y}_{T+h\|T+h-1} + \widehat{\epsilon}_{T+h-1}$. Forecasts & residuals from AR or AR_IS.
RW	$\Delta_4 \widehat{y}_{T+h\|T+h-1} = \Delta_4 y_{T+h-1}$.
DDD	$\Delta_4 \widehat{y}_{T+h\|T+h-1} = \Delta_4 y_{T+h-1} + \Delta\Delta_4 y_{T+h-1}$.
SDD (r=m=4)	$\Delta_4 \widehat{y}_{T+h\|T+h-1} = \frac{1}{4} \sum_{i=1}^{4} \Delta_4 y_{T+h-i} + \widehat{\rho} \left(\Delta_4 y_{T+h-1} - \frac{1}{4} \sum_{i=1}^{4} \Delta_4 y_{T+h-1-i} \right)$.
	$\widehat{\rho} = 0.766$ is estimate of persistence parameter from an AR(1) model.
SDD (r=m=2)	$\Delta_4 \widehat{y}_{T+h\|T+h-1} = \frac{1}{2} \sum_{i=1}^{2} \Delta_4 y_{T+h-i} + \widehat{\rho} \left(\Delta_4 y_{T+h-1} - \frac{1}{2} \sum_{i=1}^{2} \Delta_4 y_{T+h-1-i} \right)$.
VAR	$\mathbf{X}_t = \boldsymbol{\gamma} + \sum_{i=1}^{4} \boldsymbol{\Gamma}_i \mathbf{X}_{t-i} + \boldsymbol{\eta}_t$ where $\mathbf{X}_t = (\Delta_4 y_t, \Delta_4 p_{t,r,t})'$.
VAR_IS	$\mathbf{X}_t = \boldsymbol{\gamma} + \sum_{i=1}^{4} \boldsymbol{\Gamma}_i \mathbf{X}_{t-i} + \sum_{j=1}^{T} \boldsymbol{\delta}_j 1_{\{j=t\}} + \sum_{j=2}^{T} \boldsymbol{\delta}_{s,j} 1_{\{j \leq t\}} + \boldsymbol{\eta}_t$
	Selection of lags and indicators across all equations using *Autometrics* at 0.1%.
DVAR, ._IS	$\widehat{\mathbf{X}}_{T+h} = \mathbf{X}_{T+h-1} + \sum_{i=1}^{4} \widehat{\boldsymbol{\Gamma}}_i \Delta\mathbf{X}_{T+h-1}$.
	$\widehat{\boldsymbol{\Gamma}}_i, \forall i$, is taken from estimation of VAR or VAR_IS.
Pooled	$\Delta_4 \widetilde{y}_{T+h\|T+h-1} = \frac{1}{12} \sum_{j=1}^{12} \Delta_4 \widehat{y}_{j,T+h\|T+h-1}$.
	$\Delta_4 \widehat{y}_{j,T+h\|T+h-1}$ for $j = 1, \ldots, 12$, denotes 1-step ahead forecasts for above models.

Table 1: Forecasting models and devices for GDP growth.

The in-sample estimation period is 1958Q1 to 2007Q4 with 14 forecast period observations from 2008Q1 to 2011Q2. 1-step ahead forecasts are computed for annual and quarterly GDP growth (for quarterly GDP growth replace $\Delta_4 y_t$ with Δy_t in table 1). Note that differenced impulse and step indicators have no effect on the forecasts. Table 2 records the bias and RMSFE, with smallest bias or RMSFE in bold and second smallest in italics: as shown in Clements and Hendry (1993a), the ranking varies with the transformation of Δ_4 or Δ. Figure 2 records the forecasts for a range of forecasting models and devices for annual GDP growth, and figure 3 records the squared forecast errors from these models: both figures are plotted on the same vertical axis scale for comparison.

The non-robust models, including AR, AR_IS, VAR and VAR_IS, are designed to capture the in-sample data characteristics and therefore have the smallest in-sample standard errors of the models considered. However, they are not robust to location shifts, an example of which is clearly evident in the data in 2008. The figures demonstrate the large forecast errors made using these models over the 'Great Recession'. In contrast, intercept correction or differencing results in no significant forecast failure over this deep recession. Alternative robust devices also avoid the significant forecast failure evident in the 'model-based' forecasts, at the cost of a large forecast error variance seen by the wider $\pm 2\widehat{\sigma}$ bands.

---

[3]Examples of methodological changes to GDP measurement over the sample period include the European System of Accounts 1995, introduced in 1998, results from the Annual Business Inquiry and improvements to the Business Register from which the ONS conducts its surveys in 2001, and annual chain-linking along with improvements in some price deflators in 2003.

[4]The GDP data are given by the July 2011 vintage release of ABMI, seasonally adjusted real GDP at market prices. The measure of inflation used is the GDP deflator, vintage July 2011, ONS code: YBHA/ABMI. The data are taken from the Bank of England real-time database, see http://www.bankofengland.co.uk/statistics/pages/gdpdatabase/default.aspx. The interest rate is the Bank of England base rate available from www.bankofengland.co.uk. For the multivariate models, the GDP deflator data is only available to 2011Q1 for the given vintage, so we compute forecasts for the 13 out-of-sample observations rather than 14 observations for all the other forecast devices.
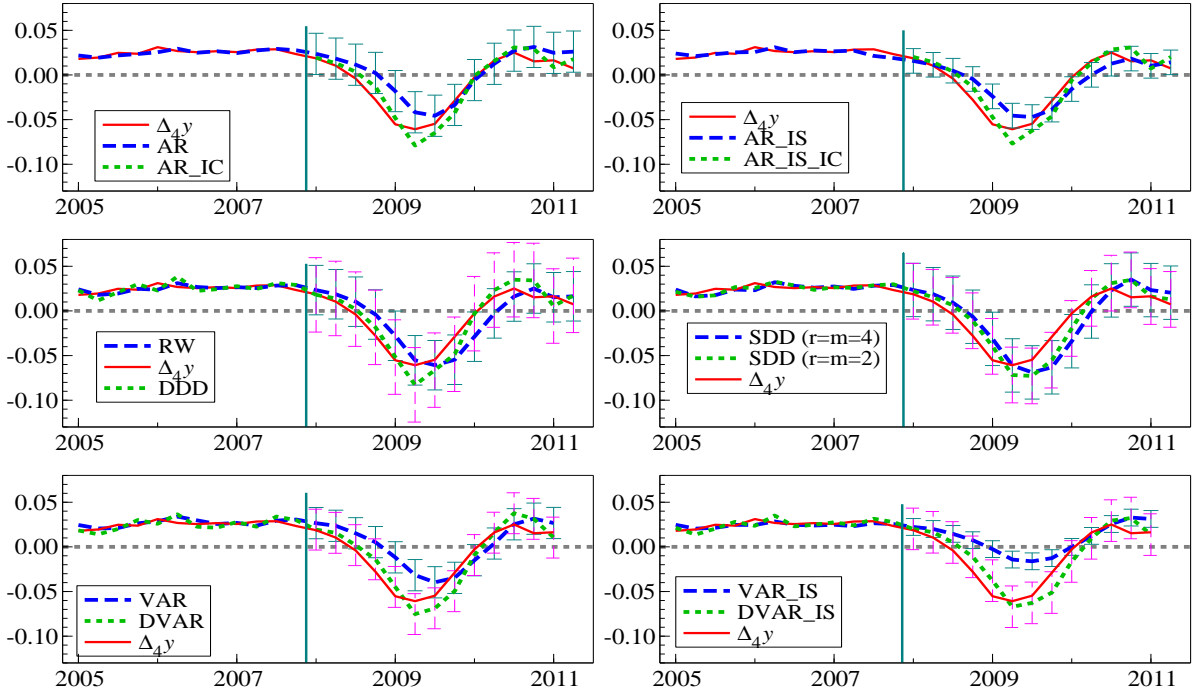
Figure 2: Forecasts for annual GDP growth.

The best performing models over the great recession are devices that robustify the 'congruent' in-sample models, using their parameter estimates. For annual GDP growth, the AR models with intercept correction are preferred on a RMSFE criterion, though the DVAR is preferred for quarterly output growth. Although the VAR produces some of the worst forecasts, using the same parameter estimates in a robust DVAR delivers some of the best forecasts. This highlights the divergence between selecting models based on in-sample congruency criteria, and using the resulting models for forecasting.

7.2 Unemployment rate

As a comparison to the GDP growth forecasts—which highlighted the need for robust forecasting devices when location shifts are present—we now consider forecasts of the UK unemployment rate over roughly the same period. We use an annual dataset of the unemployment rate over 1865–2014, updated over 1870–1913 from archival records by Boyer and Hatton (2002), and develop 'model-based' forecasts of the unemployment rate based on Clements and Hendry (2006), who show that the unemployment rate and the real interest rate minus the real growth rate, which we denote $R_{r,t}$, are cointegrated, or co-break. Therefore, our dataset consists of the annual unemployment rate and $R_{r,t} = R_{l,t} - \Delta p_t - \Delta y_t$, where $R_{l,t}$ is the long-term bond interest rate, Δp_t is annual inflation measured by the implicit deflator of GDP, and Δy_t is annual real GDP growth. This 'structural' model is based on steady-state growth theory, such that the unemployment rate rises when the real interest rate exceeds the real growth rate, and vice versa. Hendry (2001) explains the basis for this relationship, which has held relatively constantly for 150 years with a long-run equilibrium unemployment rate of 5% having a 1–1 response to $R_{r,t}$, with a t-value of 10, established well before the Great Recession.

Figure 4 compares short-sample models of UK unemployment: a) using $R_{r,t}$; b) using Δy_t. As can be seen, the forecasts of the former are far better. A fall in GNP of 6% would usually lead to a large

	$\widehat{\Delta_4 y}_{T+h\|T+h-1}$		$\widehat{\Delta y}_{T+h\|T+h-1}$	
	Bias	RMSFE	Bias	RMSFE
AR	-1.128	1.646	-1.524	3.175
AR_IS	-0.263	1.382	-1.828	3.450
AR_IC	-0.103	**0.980**	-0.662	*2.195*
AR_IS_IC	-0.080	*1.033*	-0.668	2.232
RW	-0.114	1.614	-0.650	2.249
DDD	*-0.026*	1.159	*-0.645*	2.512
SDD (r=m=4)	-0.091	1.802	-0.681	2.303
SDD (r=m=2)	-0.055	1.347	**-0.642**	2.202
VAR	-1.214	2.035	-1.496	3.308
VAR_IS	-1.947	2.624	-2.128	3.715
DVAR	**-0.024**	1.144	-0.754	**2.168**
DVAR_IS	-0.098	1.191	-0.769	2.400
Pooled	-0.435	1.164	-1.003	2.570

Table 2: 1-step ahead Bias and RMSFE for a range of forecasting devices for $\Delta_4 y_{T+h\|T+h-1}$ and $\Delta y_{T+h\|T+h-1}$, ($\times 100$). Bold denotes smallest Bias or RMSFE, and italics denote the second smallest.
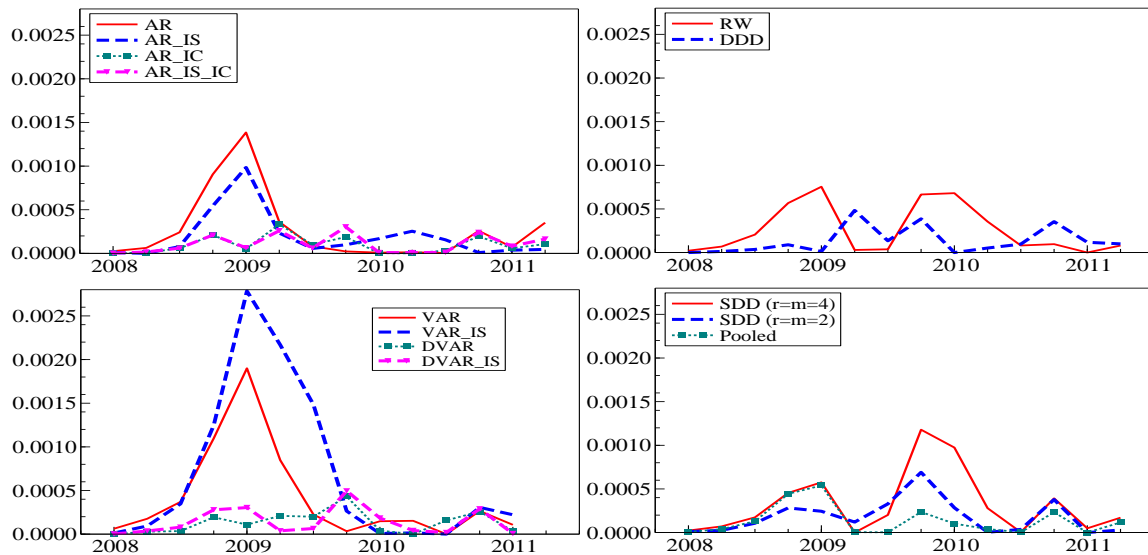

Figure 3: Squared forecast error comparisons on the same scales for annual GDP growth, plotting $(\Delta_4 y_{T+h} - \widehat{\Delta_4 y}_{T+h\|T+h-1})^2$ for the models and devices considered.
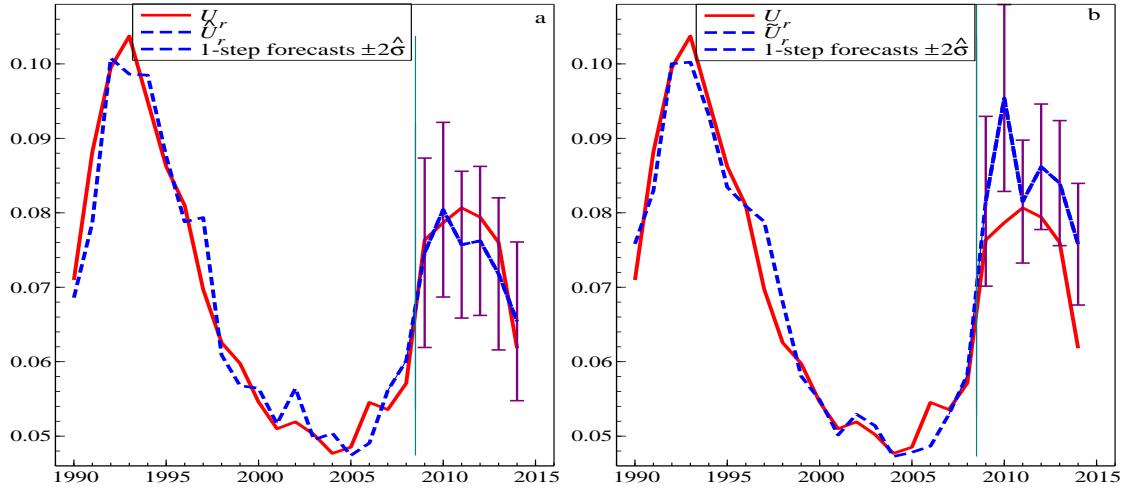
Figure 4: 1-step forecasts from short-sample models of UK unemployment estimated over 1990–2008: a) using $R_{r,t}$; b) using Δy_t.

rise in unemployment in the UK, which did not happen, but is what panel b) illustrates. $R_{r,t}$ reflects the dramatic reduction in $R_{l,t}$ brought about by the Bank of England lowering its discount rate to 0.5% and undertaking massive quantitative easing to lower longer-term interest rates, even though inflation remained above target for some time.

The models are estimated over 1865–2008, with 6 observations retained for forecasting from 2009–2014. As there are few forecast observations, a comparison was also made with two other periods in which structural breaks had or were occurring, namely 1944–1949 and 1975–1980. Table 3 lists the range of models and devices considered. For 1-step ahead forecasts, the differenced VAR/VEqCM and intercept-corrected models' forecasts are identical.

Figures 5 and 6 record the forecasts and squared forecast errors respectively for the models and devices considered. The key result is that there is no forecast failure for unemployment over this 'Great Recession' period. Despite the difficulty of forecasting GDP growth, there is little difficulty in forecasting unemployment regardless of the model or device used. 'Model-based' forecasts with IIS and SIS for in-sample breaks and outliers forecast the best, both on Bias and RMSFE. Robustification through differencing or intercept correction yields no benefit in the situation where there is no evident location shift in the data. This is most clearly observed in figure 6, where in 2010 the differenced/intercept-corrected devices for both the VAR and EqCM are worse than the model-based forecasts.

These results are sample specific: for alternative periods where the unemployment rate exhibits a location shift, then the robust devices beat the 'model-based' forecasts. For the period 1944–1949, the RW and DDD are the preferred devices based on minimizing Bias and RMSFE, whereas for 1975–1980, VAR_IS_IC and the DDD are preferred. Thus, best in-sample fit is no guarantee of forecast success when shifts can occur.

VAR	$\mathbf{X}_t = \boldsymbol{\gamma} + \sum_{i=1}^{4} \boldsymbol{\Gamma}_i \mathbf{X}_{t-i} + \boldsymbol{\eta}_t$ where $\mathbf{X}_t = (U_{r,t}, R_{r,t})'$.	
	Lags selected using *Autometrics* at 5%.	
VAR_IS	$\mathbf{X}_t = \boldsymbol{\gamma} + \sum_{i=1}^{4} \boldsymbol{\Gamma}_i \mathbf{X}_{t-i} + \sum_{j=1}^{T} \boldsymbol{\delta}_j \mathbf{1}_{\{j=t\}} + \sum_{j=2}^{T} \boldsymbol{\delta}_{s,j} \mathbf{1}_{\{j \leq t\}} + \boldsymbol{\eta}_t, \mathbf{X}_t = (U_{r,t}, R_{r,t})'$	
	Selection of lags and indicators across all equations using *Autometrics* at 0.1%.	
EqCM	$\Delta U_{r,t} = \beta_0 + \beta_1 ECM_{t-1} + \sum_{i=1}^{4} \lambda_{Ur,i} \Delta U_{r,t-i} + \sum_{i=1}^{4} \lambda_{Rr,i} \Delta R_{r,t-i} + \epsilon_t$	
	$ECM_t = U_{r,t} - \mu_0 - \mu_1 R_{r,t}$; with parameters obtained from solving for long-run of ADL(4).	
	ECM_{t-1} forced. Lags selected using *Autometrics* at 5%.	
EqCM_IS	$\Delta U_{r,t} = \beta_0 + \beta_1 ECM_{t-1} + \sum_{i=1}^{4} \lambda_{Ur,i} \Delta U_{r,t-i} + \sum_{i=1}^{4} \lambda_{Rr,i} \Delta R_{r,t-i} + \sum_{j=1}^{T} \delta_j \mathbf{1}_{\{j=t\}}$	
	$+ \sum_{j=2}^{T} \delta_{s,j} \mathbf{1}_{\{j \leq t\}} + \epsilon_t$	
	$ECM_t = U_{r,t} - \mu_0 - \mu_1 R_{r,t}$; from long-run of ADL(4) with IIS and SIS.	
	ECM_{t-1} forced. Selection of lags and indicators using *Autometrics* at 0.1%.	
....IC	$\widehat{Ur}_{\iota,T+h|T+h-1} = \widehat{Ur}_{T+h|T+h-1} + \widehat{\epsilon}_{T+h-1}$.	
	The forecast and residual is taken from VAR, VAR_IS, EqCM or EqCM_IS.	
RW	$\widehat{U}_{r,T+h|T+h-1} = Ur_{T+h-1}$.	
DDD	$\widehat{U}_{r,T+h|T+h-1} = Ur_{T+h-1} + \Delta U_{r,T+h-1}$.	
SDD (r=m=4)	$\widehat{U}_{r,T+h|T+h-1} = \frac{1}{4} \sum_{i=1}^{4} U_{r,T+h-i} + \widehat{\rho} \left(Ur_{T+h-1} - \frac{1}{4} \sum_{i=1}^{4} U_{r,T+h-1-i} \right)$.	
	$\widehat{\rho} = 0.906$ is estimate of persistence parameter from an AR(1) model.	
SDD (r=m=2)	$\widehat{U}_{r,T+h|T+h-1} = \frac{1}{2} \sum_{i=1}^{2} U_{r,T+h-i} + \widehat{\rho} \left(Ur_{T+h-1} - \frac{1}{2} \sum_{i=1}^{2} U_{r,T+h-1-i} \right)$.	
	$\widehat{\rho} = 0.906$ is estimate of persistence parameter from an AR(1) model.	
DVAR, ._IS	$\widehat{\mathbf{X}}_{T+h} = \mathbf{X}_{T+h-1} + \sum_{i=1}^{4} \widehat{\boldsymbol{\Gamma}}_i \Delta \mathbf{X}_{T+h-1}$, where $\mathbf{X}_t = (U_{r,t}, R_{r,t})'$	
	$\widehat{\boldsymbol{\Gamma}}_i, \forall i$, is taken from estimation of VAR or VAR_IS after *Autometrics* selection.	
DEqCM, ._IS	$\widehat{\Delta U}_{r,T+h} = \Delta U_{r,T+h-1} + \widehat{\beta}_1 \Delta ECM_{T+h-1} + \sum_{i=1}^{4} \widehat{\lambda}_{Ur,i} \Delta^2 U_{r,T+h-i} + \sum_{i=1}^{4} \widehat{\lambda}_{Rr,i} \Delta^2 R_{r,T+h-i}$	
	Estimated parameters imposed from EqCM or EqCM_IS after *Autometrics* selection.	
Pooled	$\widetilde{U}_{r,T+h|T+h-1} = \frac{1}{16} \sum_{j=1}^{12} \widehat{U}_{r,j,T+h|T+h-1}$.	
	$\widehat{U}_{r,j,T+h|T+h-1}$ for $j = 1, \ldots, 16$, denotes 1-step ahead forecasts for the models listed above.	

Table 3: Forecasting models and devices for the unemployment rate.

	Bias	RMSFE
VAR	0.197	0.901
VAR_IS	-0.086	**0.790**
EqCM	0.139	1.078
EqCM_IS	**-0.007**	*0.932*
VAR_IC; DVAR	-0.242	1.101
VAR_IS_IC; DVAR_IS	-0.185	1.129
EqCM_IC; DEqCM	-0.333	1.366
EqCM_IS_IC; DEqCM_IS	-0.296	1.292
RW	*0.078*	0.995
DDD	-0.296	1.053
SDD(r=m=4)	-0.280	1.031
SDD(r=m=2)	-0.261	1.013
Pooled	-0.157	0.996

Table 4: 1-step ahead Bias and RMSFE over 2009–2014 for a range of forecasting devices for $U_{r,T+h|T+h-1}$, $(\times 100)$. Bold denotes smallest Bias or RMSFE and italics the second smallest.
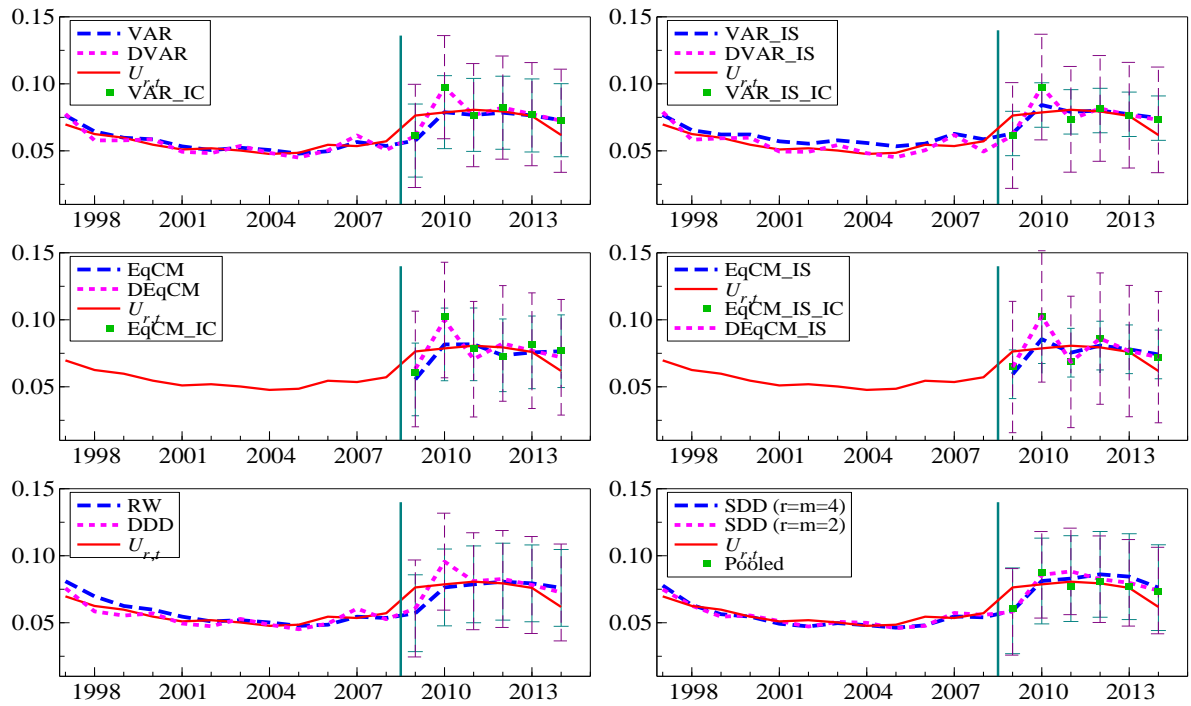
Figure 5: 1-step forecasts for annual UK unemployment rate from the long sample with $\pm 2\widehat{\sigma}$.
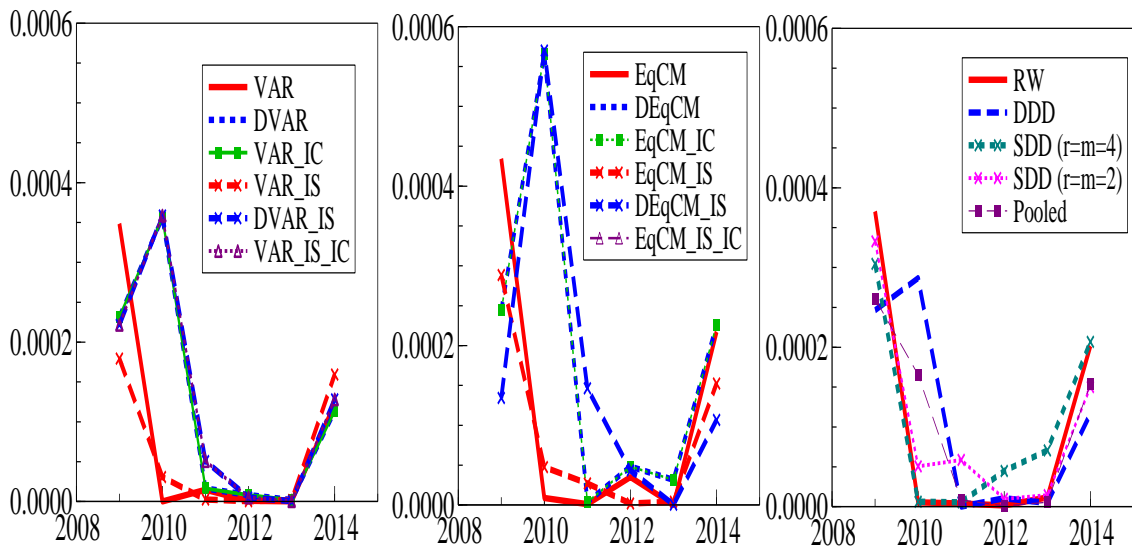

Figure 6: Squared forecast error comparisons for annual unemployment-rate models. The figures plot $(U_{r,T+h} - \widehat{U}_{r,T+h|T+h-1})^2$ for the models and devices considered.

8 Conclusions

We reviewed a range of issues confronting empirical modellers and forecasters facing wide-sense non-stationary processes subject to both stochastic trends and location shifts. A key finding is that traditional models, such as the VAR or VEqCM need not be the best models out-of-sample. The stark contrast in forecast performance of these models for GDP growth and the unemployment rate highlight the importance of out-of-sample location shifts for forecast rankings. The GDP growth results are in line with the forecasting theory descibed above: robust devices are able to adapt more rapidly to location shifts. The severity of the recession was not captured well by the 'model-based' forecasts, but systematic mistakes were avoided by using intercept corrections and differencing. This does not render the in-sample models useless. The parameter estimates from these models were essential in forming the robust devices. Hence, the 'model-based forecasting theory' is still used, but in a transformed way to make the forecasts more robust after location shifts. The devices that used such information tended to forecast better than the *ad hoc* robust devices such as the RW and DDD that do not use such information. The Pooled forecast never came first or second here.

It is remarkable that the unemployment rate can be forecast with ease over the 'Great Recession', a period that experienced substantial structural change.[5] The 2% fall in real GDP at the end of the 'Lawson boom' in 1991–2 led to unemployment rising from 7% to 10%, whereas the **6%** fall in real GDP in 2008–9 was followed by unemployment rising from 5.5% to just 8%, an incredibly small response for the largest output fall in the post-war period. Nevertheless, the results for unemployment also support our forecasting theory, which established that forecast failure is mainly due to structural breaks, hence the absence of location shifts in the data over the forecast horizon entails that the 'model-based' forecasts should do well. What may have occurred is a significant extent of co-breaking in the determining variables of the unemployment rate, so that the location shifts that are clearly evident in the real growth rate, a key variable in $R_{r,t}$, are cancelled by shifts in the real interest rate. This suggests 'Quantitative Easing' (QE) could have been useful. Factors such as more flexible labour markets, more part-time work and shorter contract hours, an increase in underemployment, people moving out of the labour force during the recession, more self-employment, and more flexible pay enabling firms to weather weaker demand without lay-offs, all helped limit the rise in unemployment. Our simplistic 'traditional' models, the VAR and EqCM, do not model any of these interactions, and yet they forecast well, so the plethora of effects on the unconditional mean of the data must have approximately cancelled. The various robustification methods give markedly different results, so being able to predict which is likely to do well is one of the challenges for the future. And as business cycles have not ceased, their modelling and forecasting remains a key activity to avoid excessive social costs from the volatility of economic activity.

References

Bánbura, M., Giannone, D., and Reichlin, L. (2011). Nowcasting. In Clements, and Hendry (2011), Ch. 7.

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

Bårdsen, G., Kolsrud, D., and Nymoen, R. (2012). Forecast robustness in macroeconometric models. Working paper, Norwegian University of Science and Technology, Trondheim.

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

---

[5]Similar findings of some variables easy to forecast and others hard are reported by Bårdsen, Kolsrud and Nymoen (2012).

Boyer, G. R., and Hatton, T. J. (2002). New estimates of British unemployment, 1870–1913. *Journal of Economic History*, **62**, 643–675.

Castle, J. L., Clements, M. P., and Hendry, D. F. (2013). Forecasting by factors, by variables, by both or neither?. *Journal of Econometrics*, **177**, 305–319.

Castle, J. L., Clements, M. P., and Hendry, D. F. (2015). Robust approaches to forecasting. *International Journal of Forecasting*, **31**, 99–112.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2012). Model selection when there are multiple breaks. *Journal of Econometrics*, **169**, 239–246.

Castle, J. L., Doornik, J. A., Hendry, D. F., and Pretis, F. (2015). Detecting location shifts during model selection by step-indicator saturation. *Econometrics*, **3(2)**, 240–264.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2009). Nowcasting is not just contemporaneous forecasting. *National Institute Economic Review*, **210**, 71–89.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2010). Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics*, **158**, 25–36.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2011). Forecasting breaks and during breaks. In Clements, and Hendry (2011), pp. 315–353.

Castle, J. L., and Hendry, D. F. (2010). Nowcasting from disaggregates in the face of location shifts. *Journal of Forecasting*, **29**, 200–214.

Castle, J. L., and Shephard, N. (eds.)(2009). *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, **5**, 559–583.

Clements, M. P., and Hendry, D. F. (1993a). On the limitations of comparing mean squared forecast errors. *Journal of Forecasting*, **12**, 617–637. (With Discussion).

Clements, M. P., and Hendry, D. F. (1993b). On the limitations of comparing mean squared forecast errors: A reply. *Journal of Forecasting*, **12**, 669–676.

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

Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

Clements, M. P., and Hendry, D. F. (2002). Explaining forecast failure in macroeconomics. In Clements, M. P., and Hendry, D. F. (eds.), *A Companion to Economic Forecasting*, pp. 539–571. Oxford: Blackwells.

Clements, M. P., and Hendry, D. F. (2006). Forecasting with breaks. In Elliott, G., Granger, C. W. J., and Timmermann, A. (eds.), *Handbook of Econometrics on Forecasting*, pp. 605–657. Amsterdam: Elsevier.

Clements, M. P., and Hendry, D. F. (eds.)(2011). *Oxford Handbook of Economic Forecasting*. Oxford: Oxford University Press.

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

Doornik, J. A. (2009a). Autometrics. In Castle, and Shephard (2009), pp. 88–121.

Doornik, J. A. (2009b). Econometric model selection with more variables than observations. Working paper, Economics Department, University of Oxford.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized factor model: Identification and estimation. *Review of Economics and Statistics*, **82**, 540–554.

Friedman, W. A. (2014). *Fortune Tellers: The Story of America's First Economic Forecasters*. Princeton: Princeton University Press.

Frisch, R., and Waugh, F. V. (1933). Partial time regression as compared with individual trends. *Econometrica*, **1**, 221–223.

Hendry, D. F. (1979). The behaviour of inconsistent instrumental variables estimators in dynamic systems with autocorrelated errors. *Journal of Econometrics*, **9**, 295–314.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (2001). Modelling UK inflation, 1875–1991. *Journal of Applied Econometrics*, **16**, 255–275.

Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, **135**, 399–426.

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

Hendry, D. F., Johansen, S., and Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, **33**, 317–335. Erratum, 337–339.

Hendry, D. F., and Juselius, K. (2000). Explaining cointegration analysis: Part I. *Energy Journal*, **21**, 1–42.

Hendry, D. F., and Juselius, K. (2001). Explaining cointegration analysis: Part II. *Energy Journal*, **22**, 75–120.

Hendry, D. F., and Martinez, A. B. (2015). Evaluating multi-step system forecasts with relatively few forecast-error observations. Unpublished paper, Economics Department, Oxford University.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254.

Johansen, S., and Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, and Shephard (2009), pp. 1–36.

Klein, L. R. (1947). The use of econometric models as a guide to economic policy. *Econometrica*, **15**, 111–151.

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

Makridakis, S., and Hibon, M. (2000). The M3 Competition: results, conclusions and implications. *International Journal of Forecasting*, **16**, 451–476.

Marcellino, M., and Schumacher, C. (2010). Factor-MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, **72**, 518–550.

Miller, P. J. (1978). Forecasting with econometric methods: A comment. *Journal of Business*, **51**, 579–586.

Pretis, F., Schneider, L., Smerdon, J. E., and Hendry, D. F. (2015). Detecting volcanic impacts in temperature reconstructions by designed break-indicator saturation. Working paper, Economics Department, Oxford University.

Shoup, C., Friedman, M., and Mack, R. P. (1941). Amount of Taxes Needed in June 1942 To Avert Inflation: A Preliminary Report Submitted to a Joint Committee of the Carnegie Corporation and

the Institute of Public Administration. Mimeo, Institute of Public Administration, Washington, D.C.

Smith, B. B. (1929). Judging the forecast for 1929. *Journal of the American Statistical Association*, **24**, 94–98.

Stock, J. H., and 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., and Watson, M. W. (1998). Diffusion indices. Working paper, 6702, NBER, Washington.

Teräsvirta, T., Tjøstheim, D., and Granger, C. W. J. (2011). *Modelling Nonlinear Economic Time Series*. Oxford: Oxford University Press.

Zarnowitz, V. (2004). An important subject in need of much new research. *Journal of Business Cycle Measurement and Analysis*, **1**, 1–7.