



**DEPARTMENT OF ECONOMICS
DISCUSSION PAPER SERIES**

**FORECASTING FROM STRUCTURAL ECONOMETRIC
MODELS**

David F. Hendry and Grayham E. Mizon

Number 597
March 2012

Manor Road Building, Oxford OX1 3UQ

Forecasting from Structural Econometric Models

David F. Hendry

Economics Department and

Institute for New Economic Thinking at the Oxford Martin School, Oxford University, UK,

Grayham E. Mizon*

Southampton University and

Institute for New Economic Thinking at the Oxford Martin School, Oxford University, UK

January 2012

Abstract

Understanding the workings of whole economies is essential for sound policy advice—but not necessarily for accurate forecasts. Structural models play a major role at most central banks and many other governmental agencies, yet almost none forecast the financial crisis and ensuing recession. We focus on the problem of forecast failure that has become prominent during and after that crisis, and illustrate its sources and many surprising implications using a simple model. An application to ‘forecasting’ UK GDP over 2008(1)–2011(2) is consistent with our interpretation.

JEL classifications: C52, C22.

KEYWORDS: Structural Models; Location shifts; Economic Forecasting; *Autometrics*.

1 Introduction

As a committed scholar, Arnold Zellner saw that the most important role of economics was to achieve a fundamental quantitative understanding of the workings of economies in order to be able to provide good policy advice and accurate forecasts.¹ It seems widely agreed that it is essential to find the underlying ‘structure’ to achieve such an understanding. However, there are many different notions of structure and numerous approaches to developing structural models, as well as concerns about the constancy of economic relationships. We provide an illustrative overview of some alternative approaches to forecasting and evaluate their likelihood of success in relation to avoiding forecast failure. The paper builds on the analysis in Hendry and Mizon (2012) both by highlighting the magnitudes of various components in their taxonomy and by an empirical example.

Our paper aims to draw the implications of the developments discussed in section 2 for understanding econometric forecasting of wide-sense non-stationary economic time series. Section 3 discusses forecasting from open economic models, namely ones where there are unmodeled variables, building

*We are grateful to Charles Renfro for inviting us to contribute to this special issue in memory of Arnold Zellner. Financial support from the Open Society Institute and the Oxford Martin School is gratefully acknowledged. We are indebted to Jennifer L. Castle and Neil R. Ericsson for helpful discussions.

¹The first author fondly remembers Arnold’s generous comments on Hendry (1976) that summarized all possible estimators for linear stationary systems. DFH also has never forgotten the kind remarks Arnold made about his use of simple illustrative examples that sought to capture the essence of a problem as in Hendry (1975). Consequently, here we develop a simple forecasting model, which nevertheless has generic implications.

on the taxonomies in Hendry and Mizon (2012). Section 4 provides an extensive scalar autoregressive distributed-lag example, and section 5 seeks to explain the many odd findings that resulted. Section 6 discusses robust forecasting devices which can avoid systematic forecast failure at a cost in increased variance when no breaks occur. Section 7 compares the various approaches to forecast UK GDP across the turbulent and unanticipated collapse over 2008(1)–2011(2). Section 8 concludes.

2 Background

At the start of his career, there were few operational estimators for systems and Zellner devoted considerable effort to developing new ones as in his famous and highly-cited papers Zellner and Theil (1962) and Zellner (1962). The history of the development of the associated software provided by Renfro (2004) shows astonishing progress, driven by antecedent advances in the power and capabilities of computers as well as improved code. Likewise, the conceptual structure has altered, so that time-series aspects have become more salient relative to simultaneity, an issue already emphasized in Zellner and Palm (1974). Vector autoregressive representations (VARs) now play a large role, including as a vehicle for cointegration techniques based on Granger (1981) and Engle and Granger (1987) which have become standard (see Johansen, 1988). Even the basic idea of structure has changed greatly from the early Cowles Commission studies, through Christ (1966), to say Hendry (1995).

Following the period over which Zellner contributed to this literature on structural modeling, from Zellner and Palm (1974), through Zellner (1986), to Zellner and Tobias (1992), the concepts, tools and methodologies advanced greatly. In particular, five surprising results have emerged from our own research building on those papers. Together these provide a framework for explaining why forecast failure can occur even when the best structural model is used and even if the future values of unmodeled variables are known before forecasting.

First, the famous results on rank and order conditions for identification in Koopmans (1950) and Hood and Koopmans (1953) have been modified, as there can exist multiple over-identified representations of a simultaneous system (see Hendry and Mizon, 1993, and Hendry, Lu and Mizon, 2009). It is well known that there are many just-identified representations, but the proof in Koopmans (1950) that for a *given set* of over-identifying restrictions there is a unique structure hid the fact that there may be other different sets of over-identifying restrictions that are also unique and acceptable.

Secondly, while it is crucial that a system be identified, those identified representations can be found by a search algorithm even when the *a priori* identifying restrictions are unknown. The derivation of the test for over-identifying restrictions in Koopmans (1950), based on comparing the likelihood function values of the (so-called) ‘reduced form’ and structure, required that the latter be a special case of the former. The structure is therefore a reduction of the ‘reduced form’, as the structure has fewer parameters when over-identified. Consequently, there exists a reduction path from the ‘reduced form’ to the structure, and that can be located by a computer program (see Hendry and Krolzig, 2005). Indeed, such an algorithm has found multiple over-identified representations all of which satisfy the test for over-identifying restrictions against the common ‘reduced form’. Given the awkwardness of current terminology, it seems preferable to refer to the ‘reduced form’ as the system, and the structural representation as a model of that system (as in Hendry, Neale and Srba, 1988).

This leads to the third change, namely rather than being a pernicious necessity to be avoided at all possible costs, model selection by reduction in a general-to-specific (*Gets*) approach is usually beneficial. A number of studies have demonstrated that unless the prior theory is precisely correct, selection delivers improved models (see Doornik, 2009, Hendry and Doornik, 2009, and Castle, Doornik and Hendry, 2011a). If the theory model is correct, then despite considerable search, selection can still deliver identical estimates for its parameters to those obtained by direct fitting (see Hendry and Johansen, 2011). By

using an extended *Gets* approach, there can even be more candidate variables to be selected from than the number of observations (see Hendry and Krolzig, 2005, and Doornik, 2007). In turn that allows checking for possible location shifts at every observation to be handled during selection using impulse-indicator saturation (IIS: see Hendry, Johansen and Santos, 2008, and Johansen and Nielsen, 2009, for analyses) Castle, Doornik and Hendry (2011b) investigate the performance of IIS in detecting multiple location shifts during selection, including breaks close to the start and end of the sample (as well as correcting for non-normality), and Hendry and Mizon (2011a) provide an empirical application to demand for food in the USA.

Fourthly, Stock and Watson (1996), and Barrell (2001) show empirically that unanticipated location shifts occur intermittently, and Clements and Hendry (2001) relate their occurrence to forecast failure. However, there is a more profound implication of those results: conditional expectations formed at time T from information available at that date for an outcome at $T + 1$ are neither unbiased nor minimum variance when unanticipated location shifts occur (see Hendry and Mizon, 2010: Hendry and Mizon, 2011b, provide an exposition). Because the mean of the distribution has shifted from T to $T + 1$, the previous conditional expectation is not centered on the relevant new mean. Since many theorems about economic forecasting begin with the presumption that conditional expectations provide unbiased and minimum variance predictors, and hence deduce that econometricians should use the conditional expectations of models, a major rethink is required.

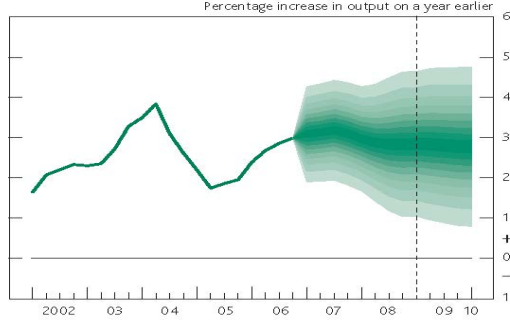
Finally, and a consequence of the previous finding, the quality of forecasts from a model depend on how it is used and what the properties of the forecast period are, as much as on the specification of the model and how it was estimated (see e.g., Clements and Hendry, 1999, and Hendry, 2006). This is the issue on which we focus below.

In part, the increasing realization of unanticipated location shifts as a key problem when modeling and forecasting has necessitated an emphasis on the invariants of the economic process, first discussed by Frisch (1938) and later highlighted by Lucas (1976), inducing a reformulation of the notion of exogeneity (see Engle, Hendry and Richard, 1983, and Engle and Hendry, 1993). Moreover, new tests of that key concept have appeared, including several based on impulse-indicator saturation (see e.g., Hendry and Santos, 2010). Thus, computational power and hugely improved algorithms have built on theoretical understanding to radically alter how econometric models can be constructed and should be used. Such developments building on his pioneering work seem a fitting memorial for Arnold Zellner who emphasised all of those aspects in his many contributions.

3 Economic forecasting

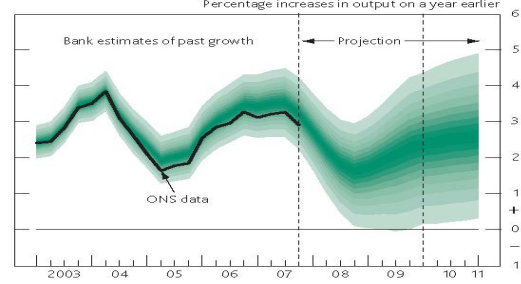
Forecasting is difficult. Chart 1 from the *Bank of England Inflation Report* for February 2007 shows its forecasts at that time for annual changes in UK GDP: ‘steady as she goes’ through the end of 2009. The second Chart updates their forecasts to February 2008 through the end of 2010: a distinct slowdown is now envisaged. However, it is nothing like the 9% fall at annual rates that materialized, which was unanticipated at that stage: Figure 1 records the actual changes in the log of seasonally-adjusted UK GDP at annual rates over 2008(1)–2011(2). As can be seen, there is a sharp drop at the end of 2008. In late 2006 and very early 2007, there was little to suggest the UK’s 16 years of uninterrupted growth was about to end: even if the US sub-prime mortgage problem had been high on the Bank’s radar, it is unlikely anyone foresaw it nearly bringing the whole world economy to its knees.

Chart 1 Current GDP projection based on market interest rate expectations



The fan chart depicts the probability of various outcomes for GDP growth in the future. If economic circumstances identical to today's were to prevail on 100 occasions, the MPC's best collective judgement is that GDP growth over the subsequent three years would lie within the darkest central band on only 10 of those occasions. The fan chart is constructed so that outcomes of GDP growth are also expected to lie within each pair of the lighter green areas on 10 occasions. Consequently, GDP growth is expected to lie somewhere within the entire fan chart on 90 out of 100 occasions. The bands widen as the time horizon is extended, indicating the increasing uncertainty about outcomes. See the box on pages 48–49 of the May 2002 *Inflation Report* for a fuller description of the fan chart and what it represents. The dashed line is drawn at the two-year point.

Chart 1 GDP projection based on market interest rate expectations



The fan chart depicts the probability of various outcomes for GDP growth. To the left of the first vertical dashed line, the distribution reflects the likelihood of revisions to the data over the past; to the right, it reflects uncertainty over the evolution of GDP growth in the future. If economic circumstances identical to today's were to prevail on 100 occasions, the MPC's best collective judgement is that the mature estimate of GDP would lie within the darkest central band on only 10 of those occasions. The fan chart is constructed so that outcomes are also expected to lie within each pair of the lighter green areas on ten occasions. Consequently, GDP growth is expected to lie somewhere within the entire fan on 90 out of 100 occasions. The bands widen as the time horizon is extended, indicating the increasing uncertainty about outcomes. See the box on page 39 of the November 2007 *Inflation Report* for a fuller description of the fan chart and what it represents. The second dashed line is drawn at the two-year point of the projection.

Source: Bank of England

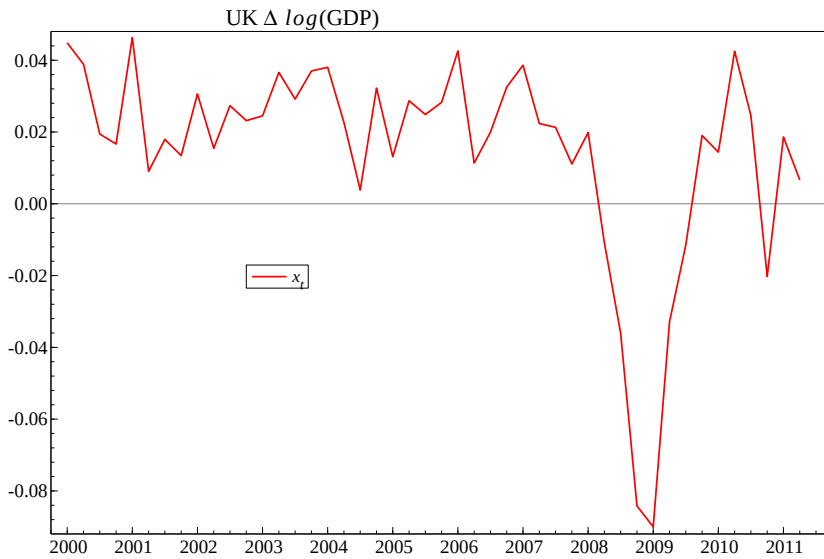


Figure 1: UK GDP year-on-year changes over 2000(1)–2011(2)

Such graphs emphasize that economies are non-stationary and evolving, where even the best economic model differs from the data generation process. Forecast failures and econometric model forecasts being out-performed by so-called ‘naive devices’ date from the early history of econometrics, partly because almost no forecasting models allow for unanticipated location shifts (changes in the previous means of the variables under analysis), although these clearly occur empirically. Note that adding variables which ‘explain’ shifts in-sample will improve forecasts only if their shifts can be forecast in turn.

The theory of economic forecasting has been well developed along lines first proposed in Haavelmo (1944) when the econometric model coincides with a stationary DGP. Consider an $n \times 1$ vector $\mathbf{x}_t \sim D_{\mathbf{x}_t}(\mathbf{x}_t | \mathbf{X}_{t-1}, \psi)$ for $\psi \in \Psi \subseteq \mathcal{R}^k$, where $D_{\mathbf{x}_t}(\cdot)$ denotes data the density of $\mathbf{x}_t$, conditional on $\mathbf{X}_{t-1} = (\dots \mathbf{x}_1 \dots \mathbf{x}_{t-1})$. A forecast $\tilde{\mathbf{x}}_{T+h|T} = \mathbf{f}_h(\mathbf{X}_T)$ is desired at T for $\mathbf{x}_{T+h}$ in time $T + h$ using all the available information. The main issue is how to select $\mathbf{f}_h(\cdot)$, and the well-known answer is the

conditional expectation, $\widehat{\mathbf{x}}_{T+h|T} = \mathbb{E}[\mathbf{x}_{T+h} | \mathbf{X}_T]$ which is unbiased, as $\mathbb{E}[(\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}) | \mathbf{X}_T] = \mathbf{0}$. Also $\widehat{\mathbf{x}}_{T+h|T}$ has smallest mean-square forecast-error matrix of unbiased predictors:

$$\mathbb{M}[\widehat{\mathbf{x}}_{T+h|T} | \mathbf{X}_T] = \mathbb{E}[(\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T})(\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T})' | \mathbf{X}_T].$$

However, the DGP is not known, so there are six main requirements for learning about $D_{\mathbf{X}_T^1}(\cdot)$, the joint density of all the available data, and its parametrization ψ , involving:

- (i) the specification of the set of relevant variables $\{\mathbf{x}_t\}$;
- (ii) the measurement of those $\mathbf{x}_t$;
- (iii) the formulation of $D_{\mathbf{X}_T^1}(\cdot)$;
- (iv) modeling the relationships between the variables;
- (v) estimating ψ ; and
- (vi) the properties of $D_{\mathbf{X}_T^1}(\cdot)$ determine the ‘intrinsic’ uncertainty; all of which introduce in-sample uncertainties. Then, the forecast uncertainty is determined by:
- (vii) the unknown properties of $D_{\mathbf{X}_{T+H}^{T+1}}(\cdot)$;
- (viii) growing as H increases;
- (ix) especially for integrated data; and
- (x) is further increased by any changes in $D_{\mathbf{X}_{T+H}^{T+1}}(\cdot)$ or in ψ .

We structure our analysis of these forecasting issues around a simple example.

4 Scalar autoregressive-distributed lag example

Consider an autoregressive-distributed lag DGP with a known exogenous $\{z_t\}$:

$$x_t = \mu + \rho x_{t-1} + \gamma z_t + \epsilon_t \text{ where } \epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2] \quad (1)$$

which is stationary in sample with all parameters $(\rho, \gamma, \sigma_\epsilon^2)$ constant and $|\rho| < 1$. When ρ, γ are known and x_T and z_{T+1} are observed without error, the optimal forecast for x_{T+1} is:

$$\widehat{x}_{T+1|T} = \mu + \rho x_T + \gamma z_{T+1} \quad (2)$$

producing an unbiased forecast:

$$\mathbb{E}[(x_{T+1} - \widehat{x}_{T+1|T}) | x_T, z_{T+1}] = \mathbb{E}[\mu - \mu + (\rho - \rho)x_T + (\gamma - \gamma)z_{T+1} + \epsilon_{T+1}] = 0$$

which is zero with the smallest possible conditional and unconditional variance determined by $D_{\mathbf{X}_T^1}(\cdot)$:

$$\mathbb{V}[(x_{T+1} - \widehat{x}_{T+1|T}) | x_T, z_{T+1}] = \mathbb{V}[\epsilon_{T+1}] = \sigma_\epsilon^2.$$

Here, $D_{\mathbf{X}_T^1}(\cdot)$ implies $D_{\mathbf{X}_{T+1}^{T+1}}(\cdot)$ and $D_{\mathbf{X}_{T+1}^{T+1}}(\cdot) = \text{IN}[\mu + \rho x_T + \gamma z_{T+1}, \sigma_\epsilon^2]$. The focus of much of the literature on forecasting has been on the additional consequences of estimating the parameters of structural models like (1), often assumed to coincide with the DGP.

However, matching the earlier list of 10 requirements are ten potential problems:

- (ia) the specification may be incomplete if (e.g.) $\mathbf{x}_t$ is not a scalar;
- (iia) the measurements might be incorrect if (e.g.) $\widetilde{\mathbf{x}}_t$ was observed, not $\mathbf{x}_t$;
- (iiia) the formulation might be inadequate if (e.g.) an intercept was needed or z_t was omitted;
- (iva) modeling might have gone wrong if (e.g.) x_{t-2} was incorrectly selected;
- (va) estimating ρ adds a bias, $(\rho - \mathbb{E}[\widehat{\rho}])x_T$, and a variance $\mathbb{V}[\widehat{\rho}]x_T^2$ (as does estimating γ if z_t is included);
- (via) the ‘intrinsic’ uncertainty from $D_{\mathbf{X}_T^1}(\cdot)$ is ever present;

- (viii) the forecast analysis above assumed $\epsilon_{T+1} \sim \text{IN}[0, \sigma_\epsilon^2]$ but $V[\epsilon_{T+1}]$ could differ;
- (viii) a multi-step forecast error $\sum_{h=1}^H \rho^{h-1} \epsilon_{T+h}$ would arise when $H > 1$, like that in the first Chart above, leading to $V[\sum_{h=1}^H \rho^{h-1} \epsilon_{T+h}] = \frac{1-\rho^{2H}}{1-\rho^2} \sigma_\epsilon^2$;
- (ix) so if $\rho = 1$, the forecast variance would be trending as $H \sigma_\epsilon^2$; and
- (x) if μ , ρ or γ changed, forecast failure could occur.

Interactions between all of these difficulties could compound the problem for forecasters. Fortunately, as shown in Clements and Hendry (1999), most of the problems do not lead to the kind of mis-forecasting seen in the charts above, as we now illustrate. Unfortunately, as we also illustrate, even after determining the source of such forecast failures, many serious difficulties remain for the structural modeler.

To simplify the first round of analysis, we set $\mu = 0$ in the DGP (1) where this is known. That enables us to contrast forecasts like (2) knowing (ρ, γ) with estimating those parameters from a sample of accurate data over $t = 1, \dots, T$:

$$\tilde{x}_{T+1|T} = \hat{\rho}x_T + \hat{\gamma}z_{T+1} \quad (3)$$

Next, we mis-specify the model by omitting z_t , then also change ρ . The graphs below are based on a single random draw of $T = 50$ observations from (1) with $z_t \sim \text{IN}[0, 1]$ when $\rho = 0.8$, $\gamma = 1$, and $\sigma_\epsilon^2 = 1$. A later theoretical analysis will confirm the results are general despite the specificity of the illustration, but the graphs highlight the key issues.

Figure 2 panel a, records the forecasts both when ρ and γ are known and constant, and when they are estimated. $\hat{x}_{T+h|T+h-1}$ from (2) is shown with error bars of $\pm 2\hat{\sigma}$, whereas forecasts when estimating (ρ, γ) are shown as $\tilde{x}_{T+h|T+h-1}$ with forecast interval bands. As can be seen, the two sets of forecasts are almost identical, with only a small increase in uncertainty from estimation. Thus, estimation *per se* does not seem to be a major problem, and is well known to be of probabilistic order $O_p(T^{-1})$.

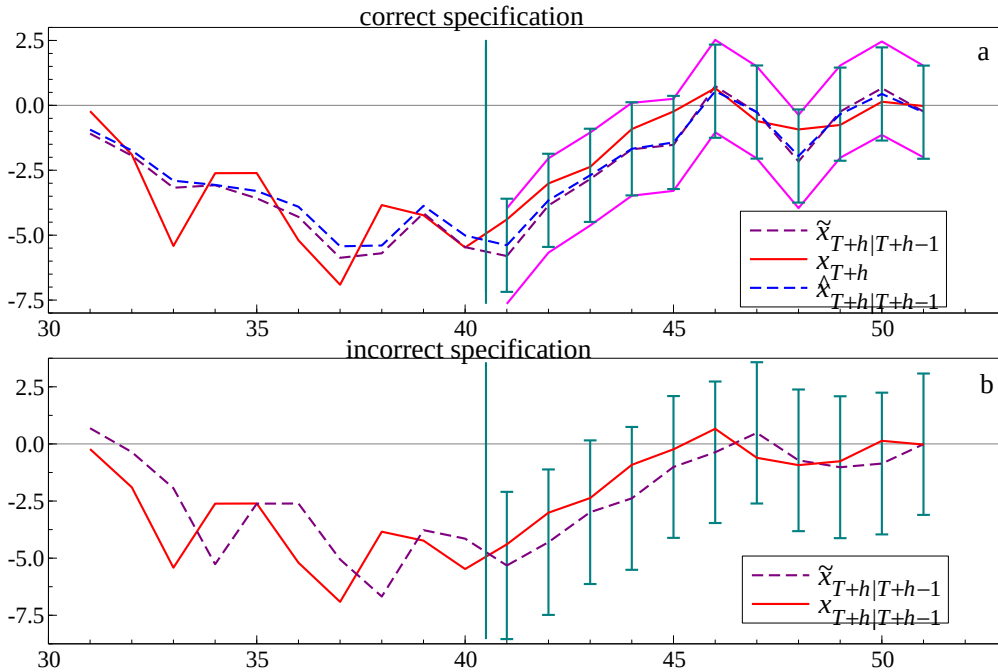


Figure 2: Forecasts under correct and incorrect specification

Next, we consider the impact on forecast accuracy and precision of incorrect specification, here corresponding to inadvertently omitting z_t , shown in Figure 2 panel b. When z_t is omitted both in

estimation and forecasting the resulting forecasts are:

$$\tilde{x}_{T+1|T} = \tilde{\rho}x_T \quad (4)$$

These are clearly poorer than under correct specification, but well within the *ex ante* forecast intervals shown. Thus, forecast failure is not just a mis-specification problem either.

Finally in this group, at $T = 40$ we shift ρ to $\rho^* = 0.4$, then back to $\rho = 0.8$ at $T = 46$ till $T = 50$, so the DGP reverts to its previous state, mimicking a short, sharp regime switch in the dynamics, denoted:

$$x_{T+h} = \rho^*x_{T+h-1} + \gamma z_{T+h} + \epsilon_{T+h} \quad (5)$$

over $h = 1, \dots, 5$, before again becoming:

$$x_{T+h} = \rho x_{T+h-1} + \gamma z_{T+h} + \epsilon_{T+h} \quad (6)$$

over $h = 6, \dots, 10$. To allow for possible interactions with mis-specification, we first return to forecasting from (3), then also use (4). Figure 3 panel c reports the former: there is only a slight impact from halving ρ , then almost none from doubling it again, so forecast failure is not just a problem of changing parameters.

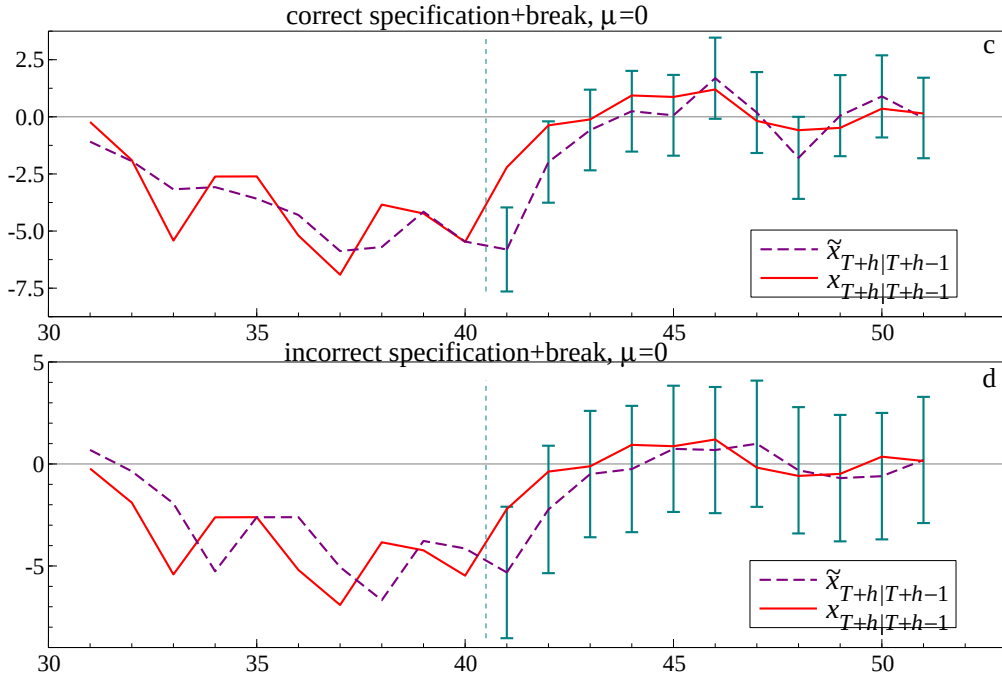


Figure 3: Correct and incorrect specifications with changed ρ

Next, we use (4), so all of estimation uncertainty, mis-specification, and non-constancy occur, yet again there is little noticeable impact from halving then doubling ρ as seen in figure 3 panel d. Whatever underlies outcomes as in the GDP Figure 1, those three mistakes do not explain it. Nor could reasonable measurement errors, nor a shift in γ , irrespective of the inclusion or omission of z_t (see Hendry and Mizon, 2012), nor using the wrong lag value, such as x_{t-2} . Of course, none of these additional mistakes will help, but they are not the source of the problem of forecast failure.

Now reconsider the same shifts in ρ but where $\mu = 10$ in:

$$x_t = \mu + \rho x_{t-1} + \gamma z_t + \epsilon_t \quad (7)$$

Irrespective of the correct specification of the model in-sample—whether (3) or (4) is used—there is a catastrophic impact from halving ρ , here shown in figure 4 panel e for forecasts from (3). The data are not unlike those for UK GDP above, first dropping then returning. Here, the first 5 forecasts are badly biased, and each is **above** the previous outcome despite the plunging values of x . Yet forecast failure vanishes on doubling ρ again. Manifestly, the value of μ , often treated as a nuisance parameter, is actually fundamental. So is the problem just the non-zero value of the intercept—or an interaction with model mis-specification?

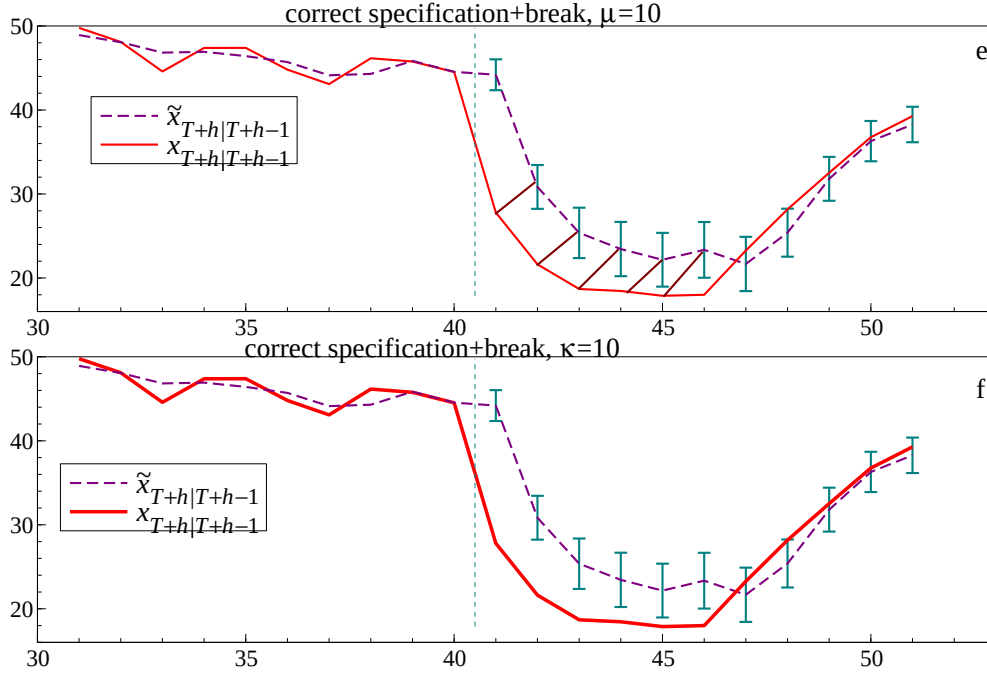


Figure 4: Correct specifications with non-zero θ and changed ρ

To clarify, again for the model which was correctly specified in-sample, consider forecasts from (3) for the same breaks in ρ as in (5) and (6), but setting $\mu = 0$, where that is known to the investigator. Instead, we let $E[z_t] = \kappa = 10$, noting that the model is correctly specified in-sample, there is a zero intercept, and the forecasts use known future z_{T+h} . Despite all those advantageous features, forecast failure is again manifest in figure 4 panel f: it is almost identical to the outcome in panel e.

Even that is far from the whole story as can be seen as follows. When the model is incorrectly specified by omitting z_t as in (4), forecasting after the same breaks in ρ as in (5) and (6), but now with both $\mu = 10$, and $\kappa = 10$ where the former is also shifted hugely to $\mu^* = 50$ at $T = 41$ then back to $\mu = 10$ at $T = 46$ so:

$$x_{T+h} = \mu^* + \rho^* x_{T+h-1} + \gamma z_{T+h} + \epsilon_{T+h} \quad (8)$$

Although almost everything seems to be wrong, there is no forecast failure using $\hat{x}_{T+h|T+h-1} = \hat{\mu} + \hat{\rho} x_{T+h-1}$ over the ten forecasts in figure 5.

5 Understanding these forecast errors

The unique explanation for this rather strange set of outcomes is in fact simple: location shifts, or alternatively expressed in this context, shifts in the long-run mean. When they occur, forecast failure results

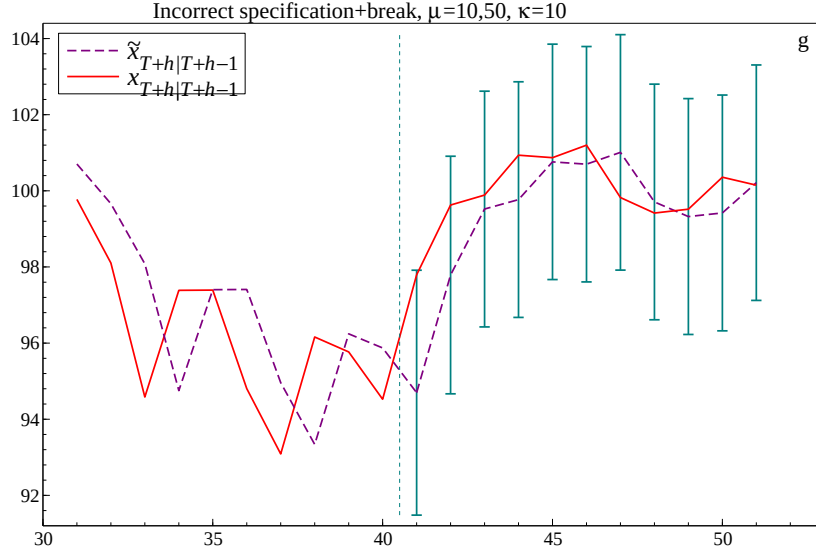


Figure 5: Incorrect specification with μ and ρ changed twice

irrespective of the goodness or otherwise of the in-sample model; when they do not occur, there is no obvious failure irrespective of how many DGP parameters shift and by much. Changes in $E[x_t]$ are the culprit.

Let $\lambda = \gamma/(1 - \rho)$ and rewrite the DGP in (1) as:

$$\Delta x_t = (\rho - 1)(x_{t-1} - \theta - \lambda(z_t - \kappa)) + \epsilon_t \quad (9)$$

and note that in general from (1) or (9):

$$E[x_t] = \frac{(\mu + \gamma\kappa)}{(1 - \rho)} = \theta \neq 0.$$

In the first three cases, and the last, $E[x_{T+h}] = 0$ before and after the shifts in ρ . In the second set of cases, $E[x_{T+h}]$ shifts from $\theta = 50$ to $\theta^* = 17$ in both cases *e* and *f*, but $\theta = \theta^*$ in *g*. All models in this class are equilibrium correction so fail systematically when $E[x_{T+h}]$ changes from θ to θ^* , as forecasts converge back to θ , irrespective of the new parameter values in the DGP. The class of equilibrium-correction models (EqCMS) is huge: all regressions; dynamic systems; VARs; DSGEs; ARCH; GARCH; and some other volatility models. Location shifts are a pervasive and pernicious problem affecting all EqCMS. To establish that claim formally, we next explain the taxonomy of all possible sources of forecast errors based on Hendry and Mizon (2012).

The DGP is as above, but written as:

$$x_t = \theta + \rho(x_{t-1} - \theta) + \gamma(z_t - \kappa) + \epsilon_t \quad (10)$$

where $\epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2]$, $E[x_t] = \theta$ and $E[z_t] = \kappa$ with $\gamma \neq 0$, but z_t is omitted from the forecasting model:

$$x_t = \mu + \rho x_{t-1} + v_t$$

Here we only consider 1-step ahead forecasts facing a single break, which occurs immediately after time T , with the post-break DGP from $h = 1, \dots$:

$$x_{T+h} = \theta^* + \rho^*(x_{T+h-1} - \theta^*) + \gamma^*(z_{T+h} - \kappa^*) + \epsilon_{T+h} \quad (11)$$

The estimated, mis-specified, forecasting model is denoted:

$$\hat{x}_{T+1|T} = \hat{\theta} + \hat{\rho} (\hat{x}_T - \hat{\theta}) \quad (12)$$

estimated over $t = 1, \dots, T$, with parameter estimates $(\hat{\theta}, \hat{\rho})$. Let $\mathbf{E}[\hat{\theta}] = \theta_e$ and $\mathbf{E}[\hat{\rho}] = \rho_e$, where $_e$ denotes the expected value (when that exists). Forecasting takes place from an estimated $\hat{x}_T$ at the forecast origin and yields the forecast error $\hat{\epsilon}_{T+1|T} = x_{T+1} - \hat{x}_{T+1|T}$:

$$\hat{\epsilon}_{T+1|T} = \theta^* - \hat{\theta} + \rho^* (x_T - \theta^*) - \hat{\rho} (\hat{x}_T - \hat{\theta}) + \gamma^* (z_{T+1} - \kappa^*) + \epsilon_{T+1} \quad (13)$$

All the main sources of forecast error occur using (12) when (11) is the DGP:

stochastic breaks: (ρ, γ) changing to (ρ^*, γ^*) ;

deterministic breaks: (θ, κ) shifting to (θ^*, κ^*) ;

omitted variables: z_t excluded over both estimation and forecast periods;

biased (and inconsistent) parameter estimates: $\rho_e \neq \rho$, $\theta_e \neq \theta$ in general;

estimation uncertainty: $\mathbf{V}[\hat{\rho}, \hat{\theta}] \neq \mathbf{0}$;

forecast-origin uncertainty: $\hat{x}_T$;

innovation errors: ϵ_{T+h} .

The taxonomy of sources of forecast errors reveals all the possible effects, although we have omitted some small order interaction terms and estimation covariances of $\mathbf{O}_p(T^{-1})$ for simplicity. The calculations involved expand every term so that each final component corresponds to a single effect (e.g.):

$$\theta^* - \hat{\theta} = (\theta^* - \theta) + (\theta - \theta_e) + (\theta_e - \hat{\theta}) \quad (14)$$

so it is decomposed into shift, bias and estimation effects. Most previous taxonomies have focused on closed models, but reach similar conclusions to those discussed below.

Taxonomy for 1-step ahead forecast errors

$\hat{\epsilon}_{T+1 T} \simeq$	Element	Expectation	Variance	
$(1 - \rho^*) (\theta^* - \theta)$	(ia)	$(1 - \rho^*) (\theta^* - \theta)$	0	
$+ (\rho^* - \rho) (x_T - \theta)$	(ib)	0	$(\rho^* - \rho)^2 \mathbf{V}[x_T]$	
$+ (1 - \rho) (\theta - \theta_e)$	(iia)	$(1 - \rho) (\theta - \theta_e)$	0	
$+ (\rho - \rho_e) (x_T - \theta)$	(iib)	0	$(\rho - \rho_e)^2 \mathbf{V}[x_T]$	
$-\rho (\hat{x}_T - x_T)$	(iii)	$-\rho (\mathbf{E}[\hat{x}_T] - x_T)$	$\rho^2 \mathbf{V}[\hat{x}_T - x_T]$	
$-(1 - \rho) (\hat{\theta} - \theta_e)$	(iva)	0	$\mathbf{O}_p(T^{-1})$	
$-(\hat{\rho} - \rho_e) (x_T - \theta)$	(ivb)	$\simeq 0$	$\mathbf{O}_p(T^{-1})$	
$+\gamma^* (z_{T+1} - \kappa^*)$	(v)	0	$(\gamma^*)^2 \mathbf{V}[z_{T+1}]$	
$+\epsilon_{T+1}$	(vi)	0	σ_ϵ^2	(15)

We now consider the implications of the forecast-error taxonomy, commencing from the foot of the table:

(vi): the innovation error has $\mathbf{E}[\epsilon_{T+1}] = 0$ and $\mathbf{V}[\epsilon_{T+1}] = \sigma_\epsilon^2$ so there is no forecast bias, and an $\mathbf{O}_p(1)$ variance, which is irreducible if $\{\epsilon_t\}$ is an innovation;

(v): the omitted variable leads to $\mathbf{E}[\gamma^* (z_{T+1} - \kappa^*)] = 0$ and $\mathbf{V}[\gamma^* (z_{T+1} - \kappa^*)] = \sigma_z^2$, so again induces no bias despite its omission and the change in its parameter values, but adds an $\mathbf{O}_p(1)$ variance, reducible by including $\{z_t\}$ in the model, offset by the parameter change and an estimation variance of $\mathbf{O}_p(T^{-1})$;

(ivb): slope estimation has $\mathbf{E}[(\hat{\rho} - \rho_e) (x_T - \theta)] \simeq 0$ as $\mathbf{E}[\hat{\rho} - \rho_e] = 0$ by definition, and also $\mathbf{E}[x_T - \theta] = 0$,

but adds an estimation variance of $O_p(T^{-1})$;

(iva): equilibrium-mean estimation also has $E[(1 - \rho)(\hat{\theta} - \theta_e)] = 0$, with an estimation variance of $O_p(T^{-1})$;

(iii): forecast-origin uncertainty, $E[\rho(\hat{x}_T - x_T)]$, is zero only if the forecast origin is unbiasedly estimated, and will have a variance $O_p(1)$ unless known for certain;

(iib) slope mis-specification again has $E[(\rho - \rho_e)(x_T - \theta)] = 0$, and adds an $O_p(1)$ variance unconditionally;

(iia) equilibrium-mean mis-specification, where $\theta \neq \theta_e$, is possible if some in-sample location shifts were not modeled, but can be eliminated by a congruent model-selection procedure that removes breaks;

(ib) the slope change, $E[(\rho^* - \rho)(x_T - \theta)] = 0$ as $E[x_T - \theta] = 0$ irrespective of $\rho^* \neq \rho$;

(ia) that leaves just the equilibrium-mean change—which is the fundamental problem: $\theta^* \neq \theta$ induces systematic forecast failure.

Once in-sample breaks have been removed, even from good forecast origin estimates:

$$E[\hat{\epsilon}_{T+1|T}] \simeq (1 - \rho^*)(\theta^* - \theta) \quad (16)$$

and that bias persists at $\hat{\epsilon}_{T+2|T+1}$ etc., so long as (12) is used, even though no further breaks ensue. Keeping μ constant, but shifting ρ to ρ^* , induces a shift in θ to θ^* . The power of that insight is exemplified by:

(a) change *both* μ and ρ by large magnitudes but with $\theta = \theta^*$ generates an outcome that is isomorphic to $\mu = \mu^* = 0$, so no break is detected as seen above; and

(b) when $\mu = \mu^* = 0$, then $\kappa \neq \kappa^*$ induces forecast failure by shifting θ even when z_{t-1} is correctly included and is known over the forecast horizon;

(c) as figure 6 shows, it is feasible to essentially replicate the same break by changing μ , γ and ρ in many combinations, so no economists, economic agents, could tell what had shifted the outcome till long afterwards.

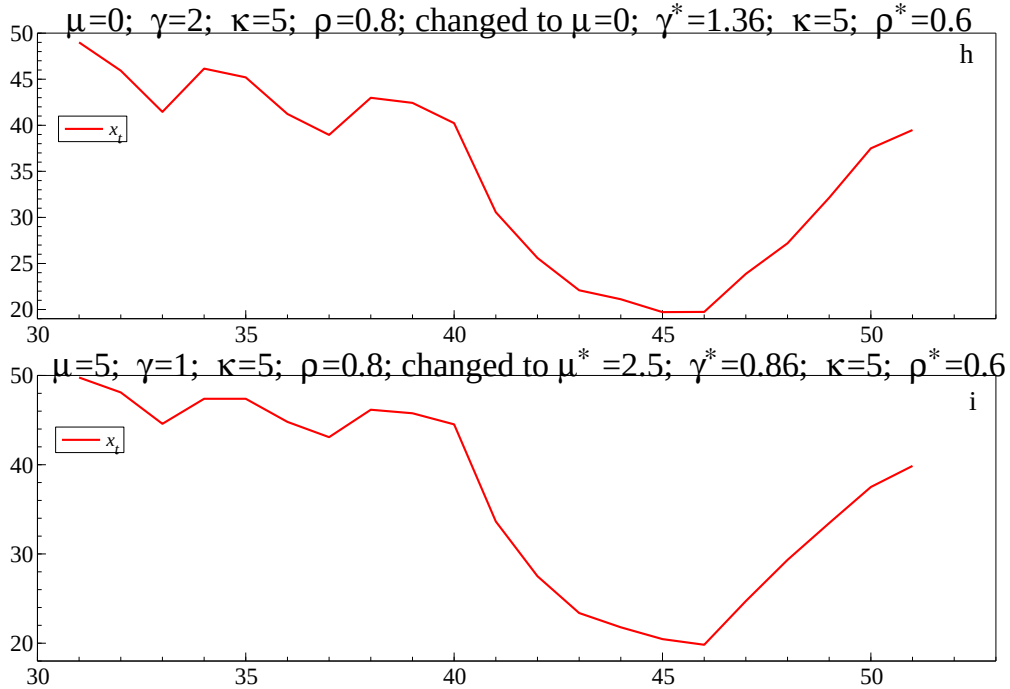


Figure 6: Many parameters shift

Such results apply to all equilibrium-correction models—they fail systematically when $E[x]$ changes as the models' forecasts converge to θ irrespective of value of θ^* . However, when ρ^* is changed back to $\rho = 0.8$, the parameter values revert to those of the original DGP, so the old equilibrium is restored, and forecasts rapidly converge back to $E[x]$. This suggests the original model also ‘recovers’ when the DGP reverts.

6 Robust forecasting devices

Robust forecasting devices may forecast better in such shifting processes as measured by RMSFE than any structural model. Difference the mis-specified estimated model (12), so that $\Delta \tilde{x}_{T+h|T+h-1} = \hat{\rho} \Delta x_{T+h-1}$ or:

$$\tilde{x}_{T+h|T+h-1} = x_{T+h-1} + \hat{\rho} \Delta x_{T+h-1} \quad (17)$$

Despite using the ‘wrong’ $\hat{\rho}$ for the first 5 forecasts, being incorrectly differenced, and omitting the relevant variable, nevertheless Figure 7 demonstrates that the robust forecasting device (17) avoids most of the last nine forecast errors for the DGP compared to (say) Panel c. The overall RMSFE is 6.6 for the in-sample model versus 5.5 for the robust forecasts here; but 3.8 versus 2.0 over the last nine forecasts, so is nearly halved, avoiding systematic failure.

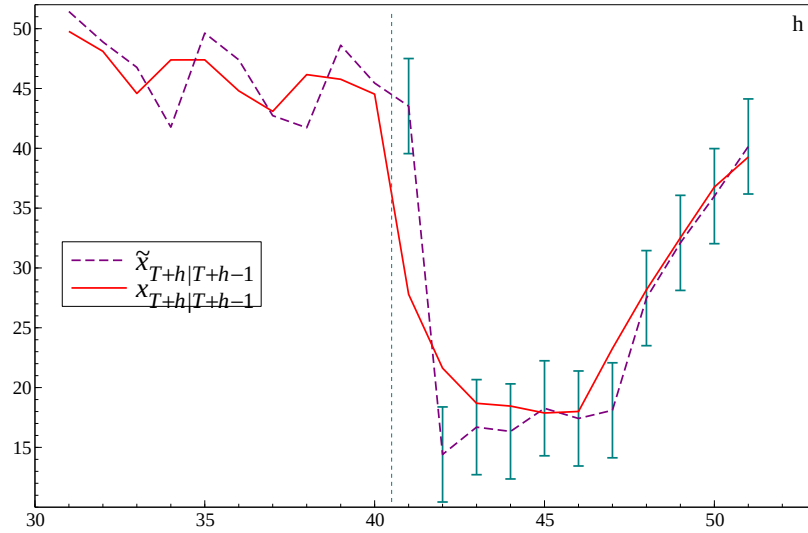


Figure 7: Differenced wrong model with changed ρ

A taxonomy of sources of forecast error for (17) clarifies why. Let $\tilde{\epsilon}_{T+h|T+h-1} = x_{T+h} - \tilde{x}_{T+h|T+h-1}$. We will use accurate $\hat{x}_T$, with $\hat{\rho} = \rho$ for simplicity but neither is crucial. Then from (17):

$$\tilde{\epsilon}_{T+1|T} = (1 - \rho^*) (\theta^* - \theta) - (1 - \rho^*) (x_T - \theta) + \gamma^* (z_{T+1} - \kappa^*) - \rho \Delta x_T + \epsilon_{T+1}$$

so taking expectations, using $E[x_T] = \theta$, for $h \geq 1$:

$$E[x_{T+h}] \simeq \theta^* + (\rho^*)^h (\theta - \theta^*) \quad (18)$$

so:

$$E[\Delta x_{T+h}] \simeq (1 - \rho^*) (\rho^*)^{h-1} (\theta^* - \theta) \quad (19)$$

hence:

$$E[\tilde{\epsilon}_{T+1|T}] \simeq (1 - \rho^*) (\theta^* - \theta) \quad (20)$$

which is equal to the in-sample DGP forecast bias.

However, at $T + 2$ the differenced-device taxonomy becomes:

$$\tilde{\epsilon}_{T+2|T+1} = -(1 - \rho^*) (x_{T+1} - \theta^*) + \gamma^* (z_{T+2} - \kappa^*) - \rho \Delta x_{T+1} + \epsilon_{T+2} \quad (21)$$

so from (18) and (19):

$$E[\tilde{\epsilon}_{T+2|T+1}] \simeq (1 - \rho^*) (\rho^* - \rho) (\theta^* - \theta) \quad (22)$$

Unless ρ^* has the opposite sign to ρ , there is a valuable offset from the $-\rho \Delta x_{T+1}$ component, helping explain the earlier forecast outcomes even though $\rho^* \neq \rho$.

Finally, at $T + 3$:

$$E[\tilde{\epsilon}_{T+3|T+2}] = \rho^* (\rho^* - \rho) (1 - \rho^*) (\theta^* - \theta)$$

which is close to zero above, with $\rho^* (1 - \rho^*) < 0.25$. Consequently, once $h - 1 > 2$ -periods after the break, using:

$$\tilde{x}_{T+h|T+h-1} = x_{T+h-1} + \hat{\rho} \Delta x_{T+h-1} \quad (23)$$

then:

$$\tilde{\epsilon}_{T+h|T+h-1} = (1 - \hat{\rho}) \Delta x_{T+h-1} = (1 - \hat{\rho}) (\rho^* - 1) (x_{T+h-2} - \theta^* - \lambda^* (z_{T+h-1} - \kappa^*)) + (1 - \hat{\rho}) \epsilon_{T+h-1} \quad (24)$$

as from equation (9):

$$\Delta x_{T+h-1} = (\rho^* - 1) (x_{T+h-2} - \theta^* - \lambda^* (z_{T+h-1} - \kappa^*)) + \epsilon_{T+h-1} \quad (25)$$

which therefore:

- a) correctly reflects the new equilibrium through $(x_{T+h-2} - \theta^* - \lambda^* (z_{T+h-1} - \kappa^*))$;
- b) includes the effect from z_{T+h-1} even though that variable was omitted from the forecasting device;
- c) adjusts at the speed $(\rho^* - 1)$;
- d) uses the well-determined in-sample estimate $\hat{\rho}$, albeit that has shifted;
- e) where $(1 - \hat{\rho})$ in (24) acts like a ‘damped trend’.

These taxonomy findings match previous graphs, and are little affected by estimating ρ . For the ‘all parameters change’ DGP, the robust device avoids almost all but the first forecast error in figure 8 panel (I), despite nearly all the parameters shifting, in stark contrast to the in-sample DGP forecasts in panel (II), which show massive forecast failure for the first six forecast errors.

The robust device avoids systematic forecast failure after a location shift, at an insurance cost when no shifts occur. Returning to the first correct specification and compared to Figure 2 from (2), the RMSFE of (17) is 1.0 against the estimated DGP forecasts of 0.87.

7 Forecasting UK GDP across 2008(1)–2011(2)

The autoregressive model selected by *Autometrics* over 1989(2)–2007(4) allowing for breaks was:

$$\hat{y}_t = \underset{(0.083)}{0.537} y_{t-1} + \underset{(0.003)}{0.012} - \underset{(0.013)}{0.071} 1_{1990(3)} \quad (26)$$

$$\hat{\sigma} = 0.013 \quad \chi^2(2) = 2.27 \quad F_{ar}(5, 6) = 1.60$$

$$R^2 = 0.49 \quad F_{het}(2, 71) = 0.05 \quad F_{reset}(2, 70) = 1.43$$

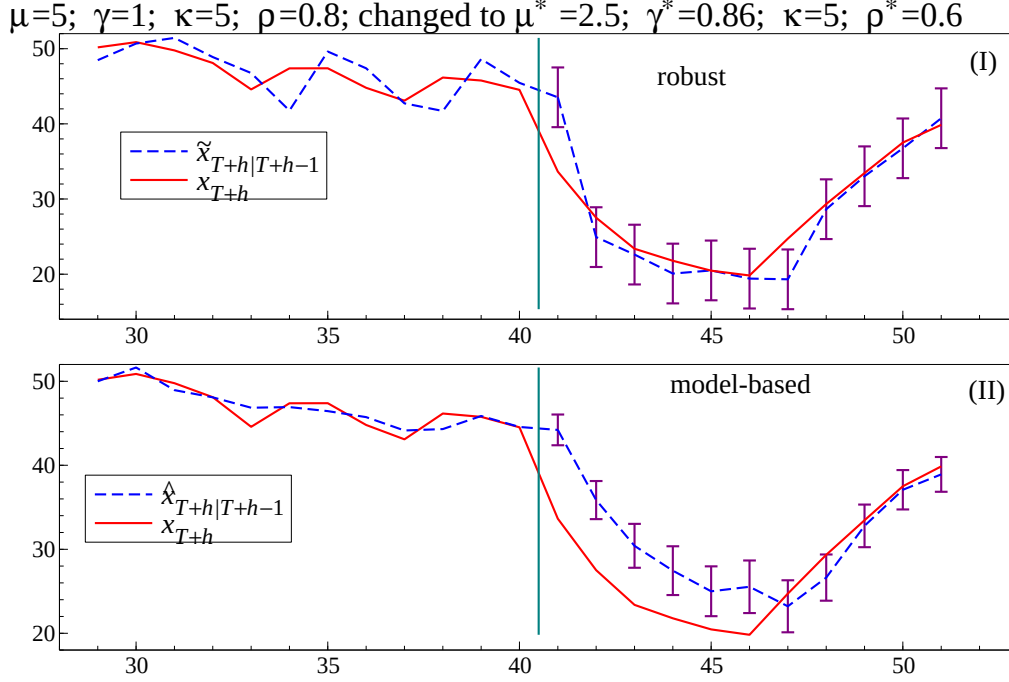


Figure 8: Differenced wrong model with changed ρ

where $1_{1990(3)}$ is an indicator for 1990(3). This model is congruent in-sample, but the forecasts over 2008(1)–2011(2) shown in figure 9 panel (i) reveal the forecast failure. The corresponding robust device here was, therefore:

$$\tilde{y}_t = y_{t-1} + 0.5\Delta y_{t-1} \quad \tilde{\sigma} = 0.021 \quad (27)$$

Panel (ii) records the multi-period, or dynamic, forecast, not unlike that in the first chart above (the robust dynamic forecasts would, of course, be similar). The forecasts from (27) are shown in panel (iii), with the squared forecast errors shown in panel (iv). The RMSFEs over the first 5 forecast errors are 0.122 versus 0.062, so essentially halved by the robust device, despite a much larger in-sample $\tilde{\sigma}$, albeit worse than that as the break was bigger than any in-sample.

8 Conclusion

Forecast failure is due to unanticipated location shifts. Location shifts induce systematic mis-forecasting in all forms of equilibrium-correction models, which comprise most macro-econometric systems in use. Conversely, every parameter in the data generation process can be shifted without any noticeable effect on the data or a model thereof when there is no location shift. Similarly, any location shift effect can be created by many different combinations of DGP parameters shifting, but which ones changed may not be discernable from the evidence till long after the occurrence of forecast failure. Thus, the verisimilitude of a model cannot be reliably checked by its forecasting success or failure. Moreover, many of the models that suffered massive forecast failure in the artificial data example forecast using the previous conditional expectation. What is needed for conditional expectations to be unbiased minimum variance predictors is that they are based on the *ex post* distribution, which requires a crystal ball to ‘see’ into the future. Even knowing the future values of unmodeled variables is insufficient to avoid forecast failure when there are shifts in the dynamics.

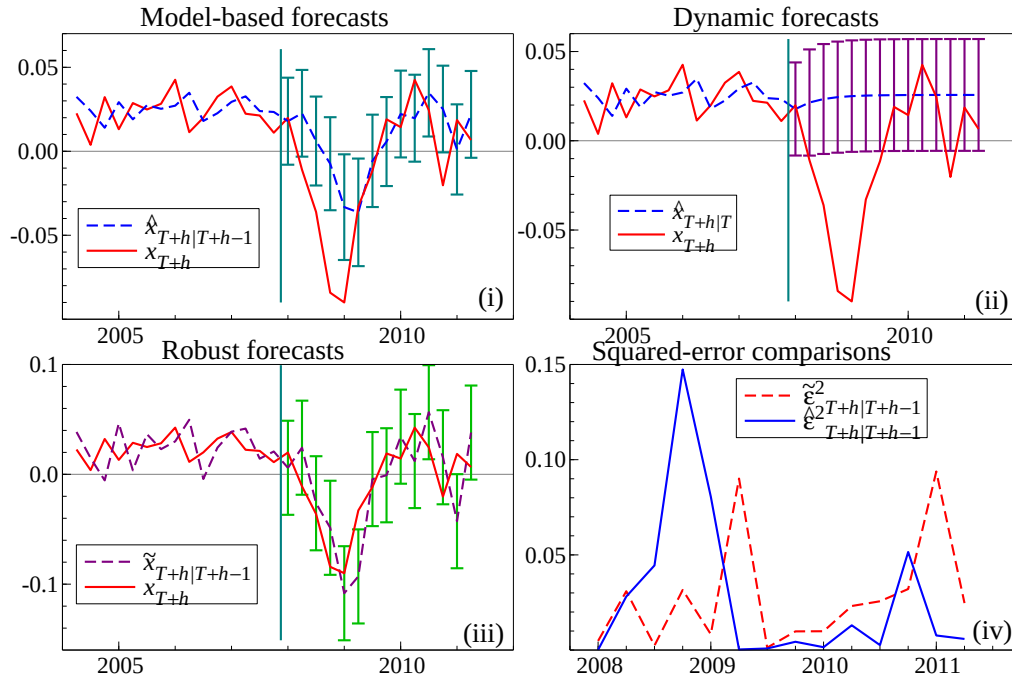


Figure 9: In-sample model, dynamic, and robust GDP forecasts with squared forecast errors

Systematic mis-forecasting can be mitigated by using the differences of the econometric system, retaining precisely the same estimates, even when the DGP parameters involved have changed. The costs of unnecessary differencing when there is no location shift are relatively small. In both cases, the policy implications of the structural system are the same, so may or may not be useful depending on the unknown source and form of the location shift. In neither case will the systems considered here, or their differences, forecast future location shifts: a different class of model seems to be needed for that, based on different information (see e.g., Castle, Fawcett and Hendry, 2010, 2011). Nevertheless, avoiding systematic forecast failure is crucial if policy is to be well based on what the future might bring forth, a conclusion Arnold Zellner would almost surely have supported.

References

- Barrell, R. (2001). Forecasting the world economy. In Hendry, D. F., and Ericsson, N. R. (eds.), *Understanding Economic Forecasts*, pp. 149–169. Cambridge, Mass.: MIT Press.
- Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011a). Evaluating automatic model selection. *Journal of Time Series Econometrics*, **3** (1), DOI: 10.2202/1941–1928.1097.
- Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011b). Model selection when there are multiple breaks. *Journal of Econometrics*, forthcoming.
- 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, M. P., and Hendry, D. F. (eds.), *Oxford Handbook of Economic Forecasting*, pp. 315–353. Oxford: Oxford University Press.

- Castle, J. L., and Shephard, N. (eds.)(2009). *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.
- Christ, C. F. (1966). *Econometric Models and Methods*. New York: John Wiley.
- 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. (2001). An historical perspective on forecast errors. *National Institute Economic Review*, **177**, 100–112.
- Doornik, J. A. (2007). Econometric model selection with more variables than observations. Working paper, Economics Department, University of Oxford.
- Doornik, J. A. (2009). Autometrics. In Castle, and Shephard (2009), pp. 88–121.
- Engle, R. F., and Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, **55**, 251–276.
- Engle, R. F., and Hendry, D. F. (1993). Testing super exogeneity and invariance in regression models. *Journal of Econometrics*, **56**, 119–139.
- Engle, R. F., Hendry, D. F., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, **51**, 277–304.
- Frisch, R. (1938). Statistical versus theoretical relations in economic macrodynamics. Mimeograph dated 17 July 1938, League of Nations Memorandum. Reproduced by University of Oslo in 1948 with Tinbergen's comments. Contained in Memorandum 'Autonomy of Economic Relations', 6 November 1948, Oslo, Universitets Økonomiske Institutt. Reprinted in Hendry D. F. and Morgan M. S. (1995), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.
- Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, **16**, 121–130.
- Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, **12**, 1–118. Supplement.
- Hendry, D. F. (1975). The consequences of mis-specification of dynamic structure, autocorrelation and simultaneity in a simple model with an application to the demand for imports. In Renton, G. A. (ed.), *Modelling the Economy*, pp. 286–322. London: Heinemann Educational Books.
- Hendry, D. F. (1976). The structure of simultaneous equations estimators. *Journal of Econometrics*, **4**, 51–88. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.
- Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.
- Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, **135**, 399–426.
- Hendry, D. F., and Doornik, J. A. (2009). *Empirical Econometric Modelling using PcGive: Volume I*. London: Timberlake Consultants Press.
- Hendry, D. F., and Johansen, S. (2011). Model selection when forcing retention of theory variables. Unpublished paper, Economics Department, University of Oxford.
- 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 Krolzig, H.-M. (2005). The properties of automatic Gets modelling. *Economic Journal*, **115**, C32–C61.
- Hendry, D. F., Lu, M., and Mizon, G. E. (2009). Identification and nonunique structure. In Castle, and Shephard (2009), pp. 345–366.
- Hendry, D. F., and Mizon, G. E. (1993). Evaluating dynamic econometric models by encompassing the VAR. In Phillips, P. C. B. (ed.), *Models, Methods and Applications of Econometrics*, pp. 272–300.

Oxford: Basil Blackwell.

- Hendry, D. F., and Mizon, G. E. (2010). On the mathematical basis of inter-temporal optimization. Discussion paper 497, Economics Department, Oxford.
- Hendry, D. F., and Mizon, G. E. (2011a). Econometric modelling of time series with outlying observations. *Journal of Time Series Econometrics*, **3** (1), DOI: 10.2202/1941-1928.1100.
- Hendry, D. F., and Mizon, G. E. (2011b). What needs rethinking in macroeconomics?. *Global Policy*, **2**, 176–183. DOI: 10.1111/j.1758-5899.2011.00082.x.
- Hendry, D. F., and Mizon, G. E. (2012). Open-model forecast-error taxonomy. In Chen, X., and Swanson, N. R. (eds.), *Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions*. New York: Springer. forthcoming.
- Hendry, D. F., Neale, A. J., and Srba, F. (1988). Econometric analysis of small linear systems using Pc-Fiml. *Journal of Econometrics*, **38**, 203–226.
- Hendry, D. F., and Santos, C. (2010). An automatic test of super exogeneity. In Watson, M. W., Bollerslev, T., and Russell, J. (eds.), *Volatility and Time Series Econometrics*, pp. 164–193. Oxford: Oxford University Press.
- Hood, W. C., and Koopmans, T. C. (eds.)(1953). *Studies in Econometric Method*. No. 14 in Cowles Commission Monograph. New York: John Wiley & Sons.
- 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.
- Koopmans, T. C. (ed.)(1950). *Statistical Inference in Dynamic Economic Models*. No. 10 in Cowles Commission Monograph. New York: John Wiley & Sons.
- Lucas, R. E. (1976). Econometric policy evaluation: A critique. In Brunner, K., and Meltzer, A. (eds.), *The Phillips Curve and Labor Markets*, Vol. 1 of *Carnegie-Rochester Conferences on Public Policy*, pp. 19–46. Amsterdam: North-Holland Publishing Company.
- Renfro, C. G. (2004). Econometric software: The first fifty years in perspective. *Journal of Economic and Social Measurement*, **29**, 9–107.
- 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.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, **57**, 348–368.
- Zellner, A. (1986). Biased predictors, rationality and the evaluation of forecasts. *Economics Letters*, **21**, 45–48.
- Zellner, A., and Palm, F. I. (1974). Time series analysis and simultaneous equation models. *Journal of Econometrics*, **2**, 17–54.
- Zellner, A., and Theil, H. (1962). Three-stage least-squares: Simultaneous estimation of simultaneous equations. *Econometrica*, **30**, 54–78.
- Zellner, A., and Tobias, J. (1992). A note on aggregation, disaggregation and forecasting performance. *Journal of Forecasting*, **19**, 457–469.