ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

FORECASTING BREAKS AND FORECASTING DURING BREAKS

Jennifer L. Castle, Nicholas W.P. Fawcett and David F. Hendry

Number 535
February 2011

Manor Road Building, Oxford OX1 3UQ

Chapter 11

Forecasting breaks and forecasting during breaks

Jennifer L. Castle, Nicholas W. P. Fawcett, David F. Hendry*
Economics Department and Institute for Economic Modelling, Oxford Martin School, Oxford University

February 23, 2011

Abstract

Success in accurately forecasting breaks requires that they are predictable from relevant information available at the forecast origin using an appropriate model form, which can be selected and estimated before the break. To clarify the roles of these six necessary conditions, we distinguish between the information set for 'normal forces' and the one for 'break drivers', then outline sources of potential information. Relevant non-linear, dynamic models facing multiple breaks can have more candidate variables than observations, so we discuss automatic model selection. As a failure to accurately forecast breaks remains likely, we augment our strategy by modelling breaks during their progress, and consider robust forecasting devices.

JEL classifications: C1, C53.
Keywords: Economic forecasting; Structural breaks; Information sets; Non-linearity.

1 Introduction

Given the wider challenges of structural breaks to economic forecasting discussed by Clements and Hendry (2011a) in this *Handbook*, the current chapter concentrates on methods of forecasting structural breaks themselves, either *ex ante* or during the progress of a break. A location shift occurs when the previous equilibrium mean changes to a new value. Unmodelled location shifts, such that the equilibrium mean shift is not known, are the most pernicious source of systematic forecast failure as shown in (e.g.) Clements and Hendry (1998, 2002b, 2006). In contrast, shifts in variables with zero means have smaller impacts on forecasts (see e.g., Hendry, 2000), but could lead to policy failures. Our approach sheds new light on the existing literature about breaks by distinguishing between 'conventional' information typically used in economic modelling and forecasting, and a wider information set that might help to explain why structural breaks happen. This broader range of variables need not be restricted to 'economic' phenomena such as inflation, and could encompass legislative changes, acts of terrorism, or other events. Our proposed approach offers a potential way to 'foresee' some breaks. It also provides a common framework to understand the impact of breaks on the three separate objectives of economic modelling over an existing sample of data; forecasting over a horizon; and policy making.

*This research was supported in part by grants from the Open Society Foundation and the Oxford Martin School. Contact details: jennifer.castle@magd.ox.ac.uk; nicholas.fawcett@lmh.ox.ac.uk; david.hendry@nuffield.ox.ac.uk. We thank Michael P. Clements, J. James Reade and Timo Teräsvirta for helpful comments on earlier drafts. Forthcoming, *Oxford Handbook of Economic Forecasting.*

Location shifts occur intermittently, and all too frequently: examples abound, as documented by Stock and Watson (1996), Barrell (2001), and Clements and Hendry (2001). The financial crisis of 2007–2010 is a further example of unanticipated equilibrium mean shifts. Figure 11.1 shows changes in world liquidity (panel a), UK mortgage lending (b), US Sub-prime loans (c) and US house prices changes (d): all show rapid, large falls. The assets of the world's largest financial institutions fell even more, precipitating a massive bailout. Both the timing and magnitude of the crash was not well anticipated, resulting in a significant global recession.

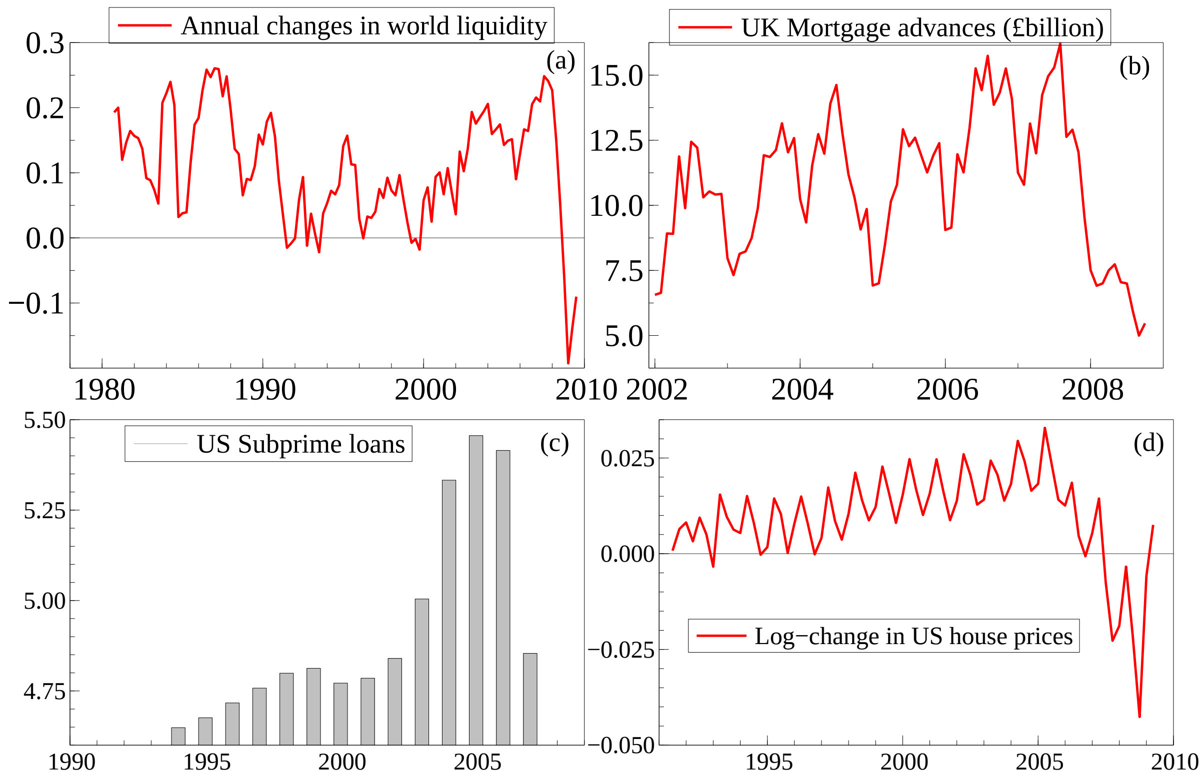

Figure 1: Some recent financial crisis variables

1.1 Necessary conditions for forecasting a break

In this section, we outline the set of requirements that are necessary to successfully forecast a break, elucidating these with reference to the 2004 Indian Ocean tsunami and the 2007–2010 financial crisis. Forecasting that a location shift is about to occur is the generic target for avoiding serious forecast and policy failures. One obvious route is to explain the intercept in a model of a variable, y (say), as a function of another observable variable (x), then use changes in x to forecast breaks in y. But 'breaks out' need 'breaks in' (following Cartwright, 1989), so if breaks in y are due to x, then x must have experienced a location shift. This would need to have been predicted, or to have happened sufficiently far in advance to allow y to be forecast.

In general, success at forecasting a break requires that:
(1) the break is predictable;
(2) there is information relevant to that predictability;
(3) such information is available at the forecast origin;
(4) the forecasting model embodies that information;
(5) there is an operational method for selecting the appropriate model; and

(6) the resulting forecasts are usefully accurate, with accurate measures of forecast-error uncertainty.
As some location shifts will not be anticipated, it would be useful if then:
(7) progress during a break transition could be forecast; and failing that
(8) there was a robust *ex post* forecasting strategy that mitigated systematic failure.
We first explain the roles of (1)–(6), using the failure to predict the 2004 Indian Ocean tsunami to highlight the issues with a physical example, where there is less controversy than most economic illustrations. We discuss (7) below, but only address (8) so far as needed.

First, whether or not a break is predictable depends on the 'real world' existence of causes preceding such a break. These may or may not exist, but usually do, albeit the lead time may be too small for practical purposes. The 2004 Indian Ocean tsunami was caused by an undersea earthquake off the west coast of Sumatra. In turn, that earthquake was triggered by a tension release at a subduction zone. Understanding how earthquakes reduce stress in a rupture zone, but increase stress in other parts of a fault, makes such events potentially predictable, albeit so far with a wide margin of timing, but improving estimates of magnitude. See Stein, Barka and Dieterich (1997) for predictions of, and Hubert-Ferrari, Barka, Jacques, Nalbant, Meyer, Armijo, Tapponnier and King (2000) for *ex post* discussions of, the 1999 earthquake at İzmit on the Anatolian fault.

Second, given that a break is indeed predictable before it happens, to actually forecast it requires there is information relevant to that predictability. Unfortunately, there was no relevant information for predicting the Sumatra earthquake at the time. However, once the earthquake had started a tsunami, there were potential devices that could have provided advance warning of its impact on neighbouring lands. These existed in the Pacific Ocean and now also exist in the Indian Ocean. Such information was not available at the time, although retrospectively Holliday, Rundle, Tiampo and Turcotte (2006) show that stress tension in that subduction region was measurable. This also highlights the importance of the third issue, namely that the relevant information must be available at the forecast origin.

Once a forecasting model for a tsunami embodies the information about its location, initiation, speed and the force of the shock wave, then the timings and locations of impacts become predictable, as has happened with Pacific Ocean tsunamis for some years. Again, the fifth aspect of an operational method for selecting the appropriate forecasting model is today clear for tsunamis, based on physical theory incorporating whatever data is delivered by the tsunami warning system in place: it would not have been clear a few centuries ago. Sixth, once tsunamis are accurately measured, the precise timings and locations of impacts are predictable within fairly small intervals. These may still vary with the detailed seabed and shore topography, but this variation is small, and doubtless forecast-error bands could be designed to reflect some unmodelled features, increasing the intrinsic uncertainty.

As a possible economic example, consider the 2007–2009 financial crisis. First, some aspects of the crisis had antecedent causes such as sub-prime loans, moral-hazard driven risk-taking, high leverage, increasing unsecured credit outstanding, growing income inequality, etc. Other causes depended on policy decisions which could have been made differently, and may not have been predictable *ex ante*. For example, letting Lehmann Brothers go bankrupt was a major negative shock, whereas earlier Bear Stearns was merged with JPMorgan, and later many of the OECD's largest financial institutions were bailed out. Consequently, some relevant information on the possible causes existed, and may even have been available at the time. However, important aspects such as agency problems and massive off-balance-sheet loans only came to light later.

The fourth issue, developing a forecasting model that embodies the relevant information is always problematic in economics, as discussed below. The fifth issue of model selection generates serious disagreements, although we now regard that as one of the least questionable aspects, as section 6 explains. The cumulative difficulties just noted for forecasting breaks in economics entail very wide margins of error for the occurrences, timings, signs and magnitudes of breaks, with no reliable measures of accuracy. Moreover, in any human-related study, reactions to forecasts can occur. If an incipient break was forecast by a trustworthy approach, actions would be taken to avoid its occurrence, which if successful, would apparently lead to a mis-prediction: see the discussion following Hendry (1986) and the alternative in Obstfeld (1996). However, an everyday example suggests that this is well understood in other contexts. Shortly after your car brakes are repaired, an emergency stop is needed and proves successful. You do not return to the garage to complain that the brakes worked, so avoiding the accident implicitly forecast by the decision to repair.

1.2 Why try if it is so difficult?

> 'Forecasting of hazardous volcanic phenomena is becoming more quantitative and based on understanding of the physics of the causative processes. Forecasting is evolving from empirical pattern recognition to forecasting based on models of the underlying dynamics. The coupling of highly non-linear and complex kinetic and dynamic processes leads to a rich range of behaviours. Due to intrinsic uncertainties and the complexity of non-linear systems, precise prediction is usually not achievable. Forecasts of eruptions and hazards need to be expressed in probabilistic terms that take account of uncertainties.' Sparks (2003)

Forecasting volcanic eruptions was once left to scrying, but like the earthquake analyses noted above, has advanced greatly over the last 50 years due to sustained research and data collection efforts. The practical difficulties predicting breaks are obviously immense, but the potential benefits suggest attempting a formalization. The contribution of this chapter is to assess how each of the 6 steps above can be translated into practice. While we have no positive and clear-cut economic examples to offer on forecasting breaks before they occur, we nevertheless believe the topic merits serious analysis, and below propose some tentative steps. An outline of our strategy follows.

We first re-examine the concept of unpredictability, and that leads us to distinguish two distinct roles for information:
(a) the regular forces which explain variations within conventional econometric models, denoted $\mathcal{J}$; and
(b) external information that might be relevant to accounting for shifts, denoted $\mathcal{K}$.
We delineate some new sources of high-frequency information such as *Google Trends* and prediction markets, potentially relevant to determining shifts. Although time disaggregation does not alter the misprediction impact of breaks on a given forecast target at the time of the break, earlier detection and so a quicker response may be feasible. The next step is to ascertain if the high-frequency information $\mathcal{K}$ can be used to model past shifts by a fractional lag, denoted $\mathcal{K}_{t-\delta}$ where $\delta < 1$, when the time period of observation is unity. This is similar to the approach that Castle, Fawcett and Hendry (2009) show can be applied to improve nowcasts.

The breadth of information in $\mathcal{I} = (\mathcal{J}, \mathcal{K})$ requires a strategy for examining a large number of candidate explanatory variables, of which an unknown subset are important and where that subset itself may change over time. To do so, we propose to include the potential variables from these new sources in general non-linear models, selected by the automatic modelling approach *Autometrics* (an Ox Package: see Doornik, 2009a, 2009c) as discussed in Castle and Hendry (2009).[1] We discuss the scope for such an automated approach when non-linearities and breaks can approximate each other, so need to be addressed jointly. Any selected model can be both simplified and evaluated by encompassing tests against plausible threshold mechanisms, such as those discussed in Granger and Teräsvirta (1993), to see if: (i) they simplify the selected specification; and (ii) are of a form that potentially can generate future 'jumps' akin to location shifts. Bontemps and Mizon (2008) provide a recent overview of encompassing and

[1] Castle, Doornik and Hendry (2010a) provide a general discussion of the properties of *Autometrics*, and Castle, Doornik and Hendry (2010b) discuss its application to detecting location shifts and outliers using impulse-indicator saturation (IIS), building on Hendry, Johansen and Santos (2008) and Johansen and Nielsen (2009).

Doornik (2008) discusses its role in *Autometrics*. Even when breaks, including those due to elements in $\mathcal{K}$, are unpredictable and hence impossible to forecast *ex ante*, there is still value in examining forecasting during a break. For example, further observations may be useful in learning about the break process, as in the empirical application by Castle, Fawcett and Hendry (2010). Of course, it may transpire that neither forecasting breaks nor learning about a break during its transition are viable, so a robust forecasting device needs to be ready when forecast failure is manifested.

Thus, this chapter proposes a new approach to forecasting breaks by focusing on the role of information. The chapter builds on both the older literature on non-linear forecasting models and the more recent literature on robust forecasting models. Our approach aims to encompass non-linear forecasting models such as Markov-switching models, threshold and smooth transition models and neural networks, which attempt to forecast regimes changes that have previously occurred. See, e.g., Krolzig (1997) for Markov switching VARs; Hamilton (1990, 1992) for regime-switching models; Bacon and Watts (1971), Quandt (1983), Granger and Teräsvirta (1993), Teräsvirta (1994), Chan and Tong (1986), Luukkonen, Saikkonen and Teräsvirta (1988) and Priestley (1981) for various forms of smooth-transition regression (STR) model and Zhang, Patuwo and Hu (1998) and Rech (2002) for neural networks. These models rely on regularities in the break process to capture the transition from one regime to another. However, our approach also aims to handle breaks that were not anticipated. In this setting, the origin of the break may be missed but the non-linear forecasting model could capture the transition to the new equilibrium. Section 6 distinguishes between systematic location shifts that have occurred previously and irregular shifts.

With this background, the chapter's structure can be understood. Section 2 considers the concepts of unpredictability and information to address the first necessary condition. The former is considered more formally in §2.1 and the latter in §2.2. Then section 3 separates information into two distinct sets, namely the regular economic forces affecting agents' behaviour, and potentially very different information from politics, law, financial innovation or technology, about events that cause sudden shifts in the regular determinants. This is followed in section 4 by an overview of some potential sources for this second source of information in the context of economics: §4.1 notes leading indicators and survey data; §4.2 discusses disaggregating data over time and/or variables; §4.3 looks at a recent addition to available high-frequency information in *Google Trends* data; §4.4 discusses the possible role for prediction markets; and §4.5 considers the potential to improve contemporaneous data at the forecast origin by better nowcasts.

Having established the potential satisfaction of the first three conditions, section 5 considers how to formulate a relevant (non-linear) model class, leading to the formulation of forecasting models in §5.1, the role of non-linear functions in §5.2, threshold models in §5.3, approximations for an operational approach in §5.4, and testing for non-linearity in §5.5. Section 6 discusses model selection, with the salient ideas outlined in §6.1, emphasizing the inter-related nature of breaks and non-linearities, and the need to avoid non-linear functions aligning with outliers or breaks. The former are tackled by impulse-indicator saturation (IIS) in §6.2, and the latter by cubic polynomial approximations in §6.3. These are then reduced to threshold-type models using encompassing checks in §6.4. Section 7 investigates forecasting breaks and forecasting during breaks. §7.1 considers forecasting a repeated break based on a threshold model, then §7.2 investigates forecasting during a break, noting two approaches: (a) forecasting facing a shift in the intercept of a model of a variable y induced by another observable variable z; and (b) when the shift is of a known exponential form and timing. The latter also allows an evaluation of the likely accuracy of break forecasts. Section 8 draws some implications for economic modelling, forecasting and policy analysis, and section 9 concludes. Robust forecasting methods are considered in Clements and Hendry (2011a), so are not addressed here.

2 Unpredictability and information

In this section, we consider the role of information in determining whether a process is at least partly predictable. The necessary conditions for predicting breaks outlined in §1.1 are extended to incorporate the two information sets, $\mathcal{J}_T$ and $\mathcal{K}_T$.

One can never establish in advance whether a specific break will, or will not, be predictable, as that depends on the information in existence. As Captain Cook sailed towards the (to him) unknown New Zealand in 1769, he was searching for land, so was not totally surprised by its discovery. A Maori on the shore looking out to sea on October 6th that year could never have predicted the break that Cook's arrival entailed, and indeed did not even have words for many of the items on board the *Endeavour*, such as cannons. However, an observer on the Moon with suitable equipment could have predicted Cook's 'discovery' and the meeting up with Maori. The observer would need the 6 conditions in §1.1 to be satisfied in order for the 'break' to be fully predictable, even if it was not to the other parties involved. For example, if it was a cloudy day, the trajectory of the *Endeavour* could not have been followed by an optical telescope, but could by a 'x-ray' type. Alternatively, if a hidden underwater rock wrecked the vessel between the outside observer's forecast and Cook's arrival, the forecast would have failed from an unanticipated 'break'. Hence, the observer must have information that is relevant and available at the forecast origin, along with an appropriately selected model embodying the information, such that the forecasts of Captain Cook's arrival are usefully accurate (e.g., strong winds do not blow the ship off course).

Unpredictability and the role of information were analyzed in Clements and Hendry (2005). Our new formulation distinguishes two information sets, which potentially might be from very different sources. The first information set, $\mathcal{J}_T$, derives from economics, and reflects the regular forces of agents' behaviour. The other information set, $\mathcal{K}_T$, could be from politics, law, financial innovation, technology, climate, geology, or war, and reflects the causes of sudden shifts. This distinction is illustrated by Hendry and Ericsson (1991), who model money demand with current and lagged prices, incomes, and the opportunity costs of holding money versus other alternatives. A break in this relationship, though, is triggered by the Banking Act of 1984, which legalized interest-bearing sight deposits. That radically shifted the opportunity costs of holding money, so altering demand relative to the prevailing levels of all the economic determinants. The Banking Act would be in the information set $\mathcal{K}$, with a resultant shift in the own interest rate, which is in $\mathcal{J}$. Section 3 shows that it is helpful to distinguish these two sets, even though both are part of 'information'. In practice, the partitioning between $\mathcal{J}$ and $\mathcal{K}$ may not be well defined, and could depend on how empirically relevant the economic theory embodied in $\mathcal{J}$ is.

We make no claim that information to predict breaks will actually exist in any given instance, nor that it will be known *ex ante*. Nevertheless, it is crucial to take account of both information sets when the second exists, which may require using much wider information than entailed by a 'conventional' economic analysis of agents' behaviour. A classic example is that one set of forces may lead to the outbreak of a civil war (say), yet very different factors facilitate its continuation: see e.g., Collier and Hoeffler (2007).

2.1 Unpredictability

A non-degenerate random variable, $\boldsymbol{\nu}_t$, with distribution $\mathsf{D}_{\boldsymbol{\nu}_t}(\cdot)$ is unpredictable with respect to an information set $\mathcal{I}_{t-1}$ over a time interval $\mathcal{T}$ if:

$$\mathsf{D}_{\boldsymbol{\nu}_t}(\boldsymbol{\nu}_t \mid \mathcal{I}_{t-1}) = \mathsf{D}_{\boldsymbol{\nu}_t}(\boldsymbol{\nu}_t) \;\; \forall t \in \mathcal{T} \tag{1}$$

where $\mid$ denotes conditioning, here on $\mathcal{I}_{t-1}$. Unpredictability is a property of $\boldsymbol{\nu}_t$ in relation to $\mathcal{I}_{t-1}$ intrinsic to $\boldsymbol{\nu}_t$, where $\mathcal{T}$ may be the singleton, $\{t\}$. When (1) holds, then no aspect of the behaviour of $\boldsymbol{\nu}_t$

is better predicted by knowing $\mathcal{I}_{t-1}$, than unconditionally. That includes shifts in the distribution $\mathsf{D}_{\boldsymbol{\nu}_t}(\cdot)$, so breaks would not be predictable if $\mathcal{I}_{t-1}$ was the 'universal' information set. A simple example of an unpredictable random variable is $\boldsymbol{\nu}_t \sim \mathsf{IN}[0, \sigma_\nu^2]$, denoting independent normally-distributed with mean $\mathsf{E}[\boldsymbol{\nu}_t] = 0$, where $\mathsf{E}[\cdot]$ denotes an expectation, and variance $\mathsf{V}[\boldsymbol{\nu}_t] = \sigma_\nu^2$. Then $\boldsymbol{\nu}_t$ is unpredictable from $\mathcal{I}_{t-1}$ as (1) holds, although well behaved. Since we are concerned with breaks, throughout we allow all distributions and entailed relationships to change, perhaps abruptly. Thus, we define expectations more precisely. The unconditional expectation of $\boldsymbol{\nu}_t$ formed at time t for time t is $\mathsf{E}_t[\boldsymbol{\nu}_t] = 0$, whereas the unconditional expectation of $\boldsymbol{\nu}_t$ formed at time $t-1$ for time t is $\mathsf{E}_{t-1}[\boldsymbol{\nu}_t] = 0$, which is also well defined. The subscript on E denotes the distribution over which expectations are calculated, which could be changing through time.

Relative to $\mathcal{I}_{t-1}$, a variable becomes predictable only if it depends on that information. However, predictability does not entail that the behaviour of $\{\boldsymbol{\nu}_t\}$ is fully accounted for, even when $\mathcal{I}_{t-1}$ is the 'universal' information set. Only a more limited information set $\mathcal{J}_{t-1} \subset \mathcal{I}_{t-1}$ is usually available, which here we take to be the standard variables analyzed in economic time series. It is convenient to take $\mathsf{E}_t[\boldsymbol{\nu}_t] = \mathbf{0}\ \forall t$ in (1). Then, predictability requires combinations of $\boldsymbol{\nu}_t$ with $\mathcal{J}_{t-1}$ as in:

$$\mathbf{y}_t = \boldsymbol{\phi}_t\left(\mathcal{J}_{t-1}, \boldsymbol{\nu}_t\right) \tag{2}$$

From (2), $\mathbf{y}_t$ depends on both the information set and the innovation process, so:

$$\mathsf{D}_{\mathbf{y}_t}\left(\mathbf{y}_t \mid \mathcal{J}_{t-1}\right) \neq \mathsf{D}_{\mathbf{y}_t}\left(\mathbf{y}_t\right) \quad \forall t \in \mathcal{T} \tag{3}$$

A common special case of (2) is:

$$\mathbf{y}_t = \mathbf{h}_t\left(\mathcal{J}_{t-1}\right) + \mathbf{u}_t \ \text{ where } \ \mathbf{u}_t = \boldsymbol{\nu}_t \odot \boldsymbol{\varphi}_t\left(\mathcal{J}_{t-1}\right) \tag{4}$$

and $\odot$ denotes element by element multiplication, so $u_{i,t} = \nu_{i,t}\varphi_{i,t}\left(\mathcal{J}_{t-1}\right)$. Then, $\mathbf{y}_t$ in (4) is predictable even if $\boldsymbol{\nu}_t$ is not, as in general:

$$\mathsf{E}_t\left[\mathbf{y}_t \mid \mathcal{J}_{t-1}\right] = \mathbf{h}_t\left(\mathcal{J}_{t-1}\right) \neq \mathsf{E}_t\left[\mathbf{y}_t\right].$$

Conversely, the variance of $\mathbf{y}_t$ may be predictable even if $\mathbf{y}_t$ is not, although both (and other aspects) may or may not be. However, the formulations in (2) and (4) do not address why there are shifts, denoted by the subscripts of $\boldsymbol{\phi}_t(\cdot)$ and $\mathbf{h}_t(\cdot)$ respectively. We return to that issue in section 3.

2.2 Information

By definition of information, $\mathcal{J}_{t-1} \subseteq \mathcal{J}_t\ \forall t$. Let the universal set determining the data generating process (DGP) of y at T be $\mathcal{I}_T^* = (\mathcal{J}_T, \mathcal{K}_T, \mathcal{M}_T)$ where $\mathcal{M}_T$ is unknown but relevant. To obtain $\mathcal{I}_T = (\mathcal{J}_T, \mathcal{K}_T)$ requires marginalizing $\mathcal{I}_T^*$ with respect to $\mathcal{M}_T$, so discarding $\mathcal{M}_T$. Then $\mathcal{I}_T$ no longer fully characterizes the DGP of y, and instead the reduced distribution is called the local data generating process (LDGP). Marginalizing with respect to relevant information is almost never without loss of information.

However, if there were no distributional shifts, reduced information about $\mathcal{J}$ would increase the uncertainty with which $\mathbf{y}_t$ could be determined from (4), but not induce biases as follows. Let $\mathcal{J}_{t-1}^* \subset \mathcal{J}_{t-1}$ and consider:

$$\mathsf{E}\left[\mathbf{y}_t \mid \mathcal{J}_{t-1}^*\right] = \mathbf{h}^*\left(\mathcal{J}_{t-1}^*\right) \tag{5}$$

then we can write:

$$\mathbf{y}_t = \mathbf{h}^*\left(\mathcal{J}_{t-1}^*\right) + \mathbf{e}_t$$

say, where:

$$\mathsf{E}\left[\mathbf{e}_t \mid \mathcal{J}_{t-1}^*\right] = \mathsf{E}\left[\mathbf{y}_t \mid \mathcal{J}_{t-1}^*\right] - \mathbf{h}^*\left(\mathcal{J}_{t-1}^*\right) = \mathbf{0}.$$

Then:

$$\mathsf{E}\left[\mathbf{e}_t \mid \mathcal{J}_{t-1}\right] = \mathbf{h}\left(\mathcal{J}_{t-1}\right) - \mathbf{h}^*\left(\mathcal{J}^*_{t-1}\right) \neq \mathbf{0}$$

so information is lost, leading to a larger variance.

Thus, the key cost of marginalizing with respect to $\mathcal{M}_T$ is when it accounts in part for distributional shifts, such that even knowing $\mathcal{K}_{t-1}$ does not lead to a constant LDGP.

When distributions shift, care is required in defining expectations to reflect the probability distribution integrated over, $\mathsf{D}_{\mathbf{y}_t}\left(\mathbf{y}_t|\cdot\right)$, the timing of the expectation (e.g., t), and the timing and contents of the available information that is conditioned on (e.g., $\mathcal{J}_{t-1}$). For example:

$$\mathsf{E}_{T+1}\left[y_{T+1} \mid \mathcal{J}_T\right] = \int y_{T+1}\mathsf{D}_{y_{T+1}}\left(y_{T+1}|\mathcal{J}_T\right)\mathrm{d}y_{T+1} \tag{6}$$

As the available information changes, the conditional distribution alters, leading to a different expectational function. Forecasting the mean of y_{T+1} from a forecast origin at T by (6) would require a 'crystal ball' to know the complete future distribution $\mathsf{D}_{y_{T+1}}(y_{T+1}|\mathcal{J}_T)$. Given the definition of $\mathcal{J}$, conditioning on $\mathcal{J}_{T+1}$ is infeasible. However, when distributions shift:

$$\mathsf{E}_{T}\left[y_{T+1} \mid \mathcal{J}_T\right] = \int y_{T+1}\mathsf{D}_{y_{T}}\left(y_{T+1}|\mathcal{J}_T\right)\mathrm{d}y_{T+1} \tag{7}$$

need not even be unbiased for $\mathsf{E}_{T+1}[y_{T+1}]$, see Hendry and Mizon (2009). Indeed, after a location shift, that bias is the cause of forecast failure.

Nevertheless, there may be additional information that could explain some of the changes from $\mathsf{D}_{y_T}\left(\cdot\right)$ to $\mathsf{D}_{y_{T+1}}\left(\cdot\right)$: we denote that information by $\mathcal{K}$. When $\mathcal{K}$ exists, that ensures the first necessary condition is met. We now assume that $\mathcal{K}_{T+1-\delta}$ is available at a higher frequency when seeking to forecast $\mathbf{y}_{T+1}$, and so is closer to the forecast than $\mathcal{J}_T$, and address this timing issue in §4. Consequently, building on §1.2, and allowing information to be dated within sub-intervals of the time unit of observation of $\{\mathbf{y}_t\}$, we re-express the necessary conditions for predicting breaks as:
(1) the existence of $\mathcal{K}_{T+1-\delta}$;
(2) knowledge of the elements in $\mathcal{K}_{T+1-\delta}$;
(3) observing (or accurately forecasting), $\mathcal{K}_{T+1-\delta}$ in time to forecast $\mathbf{y}_{T+1}$;
(4) knowledge of how $\mathcal{K}_{T+1-\delta}$ shifts $\mathbf{h}_{T+1}\left(\cdot\right)$;
(5) selecting a model of $\mathbf{h}_t\left(\cdot\right)$ assuming $\mathcal{I}_{t-\delta} = \left(\mathcal{J}_{t-1}, \mathcal{K}_{t-\delta}\right)$, and estimating its parameters prior to T;
(6) producing forecasts of $\mathbf{y}_{T+1}$ that are usefully accurate, with accurate forecast-error uncertainty.

3 Two information sets

Assuming that a break is predictable, we now address the second condition: does there exist information relevant to that predictability? We have distinguished two information sets determining changes in $\mathbf{y}_{T+1}$:

(A) $\mathcal{J}_T$: the information set which enters $\mathbf{h}_{T+1}\left(\mathcal{J}_T\right)$.

(B) $\mathcal{K}_{T+1-\delta}$: the information set explaining shifts in $\mathbf{h}_{T+1}\left(\cdot\right) \neq \mathbf{h}_T\left(\cdot\right)$.

The former are standard economic forces. For example, for money demand, $\mathbf{h}_t(\mathcal{J}_{t-1})$ depends on incomes, prices, interest rates and lags thereof. 'Conventional econometrics' concerns modelling $\mathsf{E}_t[\mathbf{y}_t|\mathcal{J}_{t-1}] = \mathbf{h}_t(\mathcal{J}_{t-1})$, and forecasting $\mathbf{y}_{T+1}$ by $\widehat{\mathbf{h}}_{T+1}(\mathcal{J}_T)$. Such a formulation takes $\mathbf{h}_t(\cdot)$ as given (usually deterministic), and often assumed constant. To understand changes in $\mathbf{h}_{T+1}(\cdot)$ and their relation to $\mathcal{I}_T$, we need to deconstruct (4), and will do so for the special case of a constant conditional variance, $\boldsymbol{\varphi}_t(\mathcal{J}_{t-1}) = \boldsymbol{\varphi}$.

$\mathcal{K}_{T+1-\delta}$ shifts the relationship between T and $T+1$ due to changes in (say), legislation, financial innovation, technology, and policy regime switches. Such a shift in $\mathbf{h}_{T+1}(\cdot)$ then alters (e.g.) money demand at the same levels of incomes, prices, and interest rates. In general, $\mathbf{h}_{T+1}(\cdot)$ changes with $\mathcal{K}_{T+1}$, and, as we are primarily concerned with location shifts, we write that dependence as:

$$\mathsf{E}_{T+1}\left[\mathbf{h}_{T+1}\left(\mathcal{J}_T\right) \mid \mathcal{J}_T, \mathcal{K}_{T+1}\right] = \mathbf{h}_0\left(\mathcal{J}_T\right) + \mathbf{h}_1\left(\mathcal{K}_{T+1}\right) \quad (8)$$

where $\mathbf{h}_0(\cdot)$ is constant. The second term is generally zero, but accounts for shifts by a step function when they happen. At $T+1-\delta$, however, an investigator at best knows ($\mathcal{J}_T$, $\mathcal{K}_{T+1-\delta}$), in which case:

$$\mathsf{E}_{T+1-\delta}\left[\mathbf{h}_{T+1}\left(\mathcal{J}_T\right) \mid \mathcal{J}_T, \mathcal{K}_{T+1-\delta}\right] = \mathbf{h}_0\left(\mathcal{J}_T\right) + \mathsf{E}_{T+1-\delta}\left[\mathbf{h}_1\left(\mathcal{K}_{T+1}\right) \mid \mathcal{K}_{T+1-\delta}\right].$$

If a location shift is unpredictable, so $\mathsf{E}_{T+1-\delta}\left[\mathbf{h}_1\left(\mathcal{K}_{T+1}\right) | \mathcal{K}_{T+1-\delta}\right] = \mathbf{0}$, there will be no perceptible difference from the information set which enters $\mathbf{h}_{T+1}\left(\mathcal{J}_T\right)$. The aim, therefore, is to ascertain available information $\mathcal{K}_{T+1-\delta}$ such that $\mathsf{E}_{T+1-\delta}\left[\mathbf{h}_1\left(\mathcal{K}_{T+1}\right) | \mathcal{K}_{T+1-\delta}\right] \simeq \mathbf{h}_1\left(\mathcal{K}_{T+1}\right)$, a daunting, but not impossible task, to which we now turn.

4 Available information

The above analysis suggests monitoring a wide set of sources of information for potential changes, including those beyond the usual economic variables. There are a number of possibilities for improving the use of information, as well as considering information not normally included in $\mathcal{J}_{t-1}$ that might be helpful in modelling and available for forecasting breaks:

1. leading indicators and survey data (§4.1);
2. disaggregating data over time and/or variables (§4.2);
3. *Google Trends* data (§4.3);
4. prediction markets data (§4.4);
5. improved data at the forecast origin (§4.5).

We consider these in turn.

4.1 Leading indicators

Leading indicators of a break would solve the problem, but seem elusive when needed. Putative leading indicators have an unimpressive historical record when incorporated as part of the first information set, $\mathcal{J}_{t-1}$, see e.g., Diebold and Rudebusch (1991), and Emerson and Hendry (1996). However, they might be more effective as 'break predictors', within the second information set, $\mathcal{K}_{T+1-\delta}$. Marcellino (2006) provides an overview and Camba-Mendez, Kapetanios, Weale and Smith (2002) gives an empirical example.

Historical analyses sometimes include survey data from consumers and businesses about their plans and expectations. Such surveys take time to conduct and process, and are rarely available sufficiently far ahead of breaks to help forecast their appearance. However, they may be useful during the course of a break. Other correlated contemporaneous variables that might help include high-frequency data on road traffic and air passenger numbers, energy consumption etc.

4.2 Data disaggregated over time and variables

Higher-frequency data could help to detect breaks earlier than from a time-aggregated process, and thereby facilitate faster adaptation following a break. Although Castle and Hendry (2008) show that time disaggregation does not change the impact of breaks on forecasts of outcomes at the lower frequency from a forecast origin when the break happens, Clements and Hendry (2011a) show time disaggregation can improve forecasts one period later, namely during the break. With regard to disaggregation across variables, Hendry and Hubrich (2009) consider the possible benefits from forecasting aggregate variables of interest via their disaggregate components. They also find that disaggregation across variables does not reduce the impact of breaks on aggregate forecasts at the time of the break, but again Clements and Hendry (2011a) show that result changes during the break. Relative performance depends on a trade-off between estimation uncertainty and mis-specification deriving from aggregation.

4.3 Google Trends data

Google Trends data could potentially improve forecasts of many variables including car, home, and retail sales, travel behaviour, and the spread of diseases such as flu. Choi and Varian (2009) add Google query variables to linear seasonal autoregressions to measure their contribution to improving nowcasts. An *Autometrics* approach allows for more variables than observations as well as impulse-indicator saturation to detect outliers and location shifts. Hence, large relevant subsets of Google query variables can be added to the candidate set. Doornik (2009b) shows that forecasts of swine flu' cases can be markedly improved by adopting such a general approach. Castle and Hendry (2010b) suggest their inclusion in a general framework for nowcasting the missing components of aggregates.

4.4 Data from prediction markets

Prediction markets, like the Iowa electronic market (http://www.biz.uiowa.edu/iem/index.cfm) or intrade (http://www.intrade.com/jsp/intrade/trading/t_index.jsp), are a recent source of information about future events, based on the probabilities of outcomes evolving as the weight of betting changes over time. Participants presumably draw information from many sources, including developments in the economy, opinion polls and surveys, so the betting outcome at any point is a weighted average of the strengths of their beliefs. Thus, prediction markets essentially 'integrate' the many different forecasting models used by individual market participants. Each forecast is weighted by a metric, $w_{m,t}$, based on the weight of betting, to produce a probability forecast of the event outcome: see Hendry and Reade (2009). A combination of individual forecasts can outperform any individual forecast by delivering a smaller mean-square forecast error (MSFE) than any one alone as first shown by Bates and Granger (1969) (see Aiolfi, Capistrán and Timmermann, 2011). When models are partial explanations, a combination of them might improve by cancelling offsetting biases as in Hendry and Clements (2004). If all relevant explanatory variables were orthogonal, and models used subsets thereof, then their combination would reflect all available information, though not necessarily optimally.

Prediction markets inevitably have a termination point, by which date, the probability must have converged on either 0 or 1. Passage of time matters *per se*, and affects betting on the probability of the event being forecasted. Thus, it is almost inevitable that interval forecasts will narrow, and most 'prediction markets' graphs drift towards the actual outcome *ex post*. However, it is still possible to get a 'last second' major shift, such as an overtime goal in soccer, leading to a large switch in probability. Thus, like most economic forecasts, future location shifts seem to be assumed absent by participants in such markets, although the intrinsic uncertainty of such a possibility may affect the size of bets made and the bid/offer spread. Nevertheless, prediction markets are a potential source of information that might help forecast breaks.

4.5 Improved data at the forecast origin

Castle and Hendry (2010b) analyze the four problems of 'missing data', 'measurement error', 'changing database', and 'breaks' intrinsic to nowcasting. They show that when breaks occur in that setting, 'nowcasting' can be improved by adopting a general information set which includes the sources just discussed. Data are analyzed by an *Autometrics* approach embodying IIS, and using robust forecasting devices, similar to the approach considered here in the context of forecasting. An obvious benefit of more accurate data at the forecast origin is knowledge of where an economy actually is, allowing a better assessment of the sources and magnitudes of breaks.

5 Model formulation for forecasting breaks

In this section, we outline the non-linear forecasting framework. The aim is to forecast a vector of n stochastic variables $\{\mathbf{y}_{T+h}\}$ over a forecast horizon $h = 1, \ldots, H$, from a forecast origin at T, when the possibility of a break is omnipresent. The joint density of $\mathbf{y}_t$ at time t is $\mathsf{D}_{\mathbf{y}_t}(\mathbf{y}_t|\mathbf{Y}_{t-1}^{t-r}, \mathbf{W}_t^{t-s}, \mathbf{q}_t, \boldsymbol{\kappa}_t)$ conditional on $\mathbf{Y}_{t-1}^{t-r} = (\mathbf{y}_{t-r}, \ldots, \mathbf{y}_{t-1})$ and $\mathbf{W}_t^{t-s} = (\mathbf{w}_{t-s}, \ldots, \mathbf{w}_t)$, which are variables that become relevant at breaks, where $\mathbf{q}_t$ denotes the deterministic terms such as intercepts, trends and known indicators, and $\{\boldsymbol{\kappa}_t\}$ are the agents' decision parameters.

We first consider the formulation of forecasting models in §5.1, then address forecasting using non-linear models in §5.2. Next, we describe a setting in §5.3 where a break is treated as an 'absorbing barrier', which when 'hit' irreversibly alters the process. We then consider a smooth transition to such a barrier. Fifth, to facilitate the model selection procedures discussed in section 6, and avoid identification problems under the null of linearity, we describe an approximation based on a Taylor-series expansion in §5.4. Finally, we discuss testing for non-linearity in §5.5.

5.1 Forecasting models

A 'conventional' model for $\mathbf{y}_t$ can be written as $\mathsf{M}_{\mathbf{y}}(\mathbf{y}_t|\mathbf{Y}_{t-1}^{t-k}, \widetilde{\mathbf{q}}_t, \boldsymbol{\theta})$, with a subset of deterministic terms, $\widetilde{\mathbf{q}}_t$, lag length k and stochastic specification defined by the parameters $\boldsymbol{\theta}$. This model is fitted over the sample $t = 1, \ldots, T$ to produce the forecast sequence $\{\widehat{\mathbf{y}}_{T+h|T}\}$. The parameter estimates are functions of the measured data, which is denoted $\widetilde{\mathbf{Y}}_T^1$, and the in-sample set of deterministic terms $\widetilde{\mathbf{Q}}_T^1$, so that using the full sample:

$$\widehat{\boldsymbol{\theta}}_{(T)} = \mathbf{f}_T\left(\widetilde{\mathbf{Y}}_T^1, \widetilde{\mathbf{Q}}_T^1\right) \tag{9}$$

where the parenthetical subscript on $\widehat{\boldsymbol{\theta}}$ in (9) denotes using data till time T. Letting the model be denoted by $\mathbf{y}_t = \mathbf{g}(\cdot) + \boldsymbol{\varepsilon}_t$, then forecasts ignoring breaks are given by:

$$\widehat{\mathbf{y}}_{T+h|T} = \mathbf{g}_h\left(\widetilde{\mathbf{Y}}_T^{T-k+1}, \widetilde{\mathbf{Q}}_{T+h}^1, \widehat{\boldsymbol{\theta}}_{(T)}\right) \tag{10}$$

leading to the sequence of forecast errors $\widehat{\boldsymbol{\varepsilon}}_{T+h|T} = \mathbf{y}_{T+h} - \widehat{\mathbf{y}}_{T+h|T}$. Clements and Hendry (2006) provide a taxonomy of sources of forecast errors.

Since shifts in the density function of $\mathbf{y}_t$ depend on $\{\mathbf{w}\}$ (elements of $\mathcal{K}$ above), investigators would want to include such variables in the model if known. Although it may not be possible to know, or obtain, all the constituents of $\mathbf{w}_t$ for any particular analysis, we allow that a subset might be available at a higher frequency, denoted by $\mathbf{z}_{t-\delta}$. This then entails estimating an additional parameter vector $\boldsymbol{\rho}$. In practice, there may not be a history to include, and the relevant information may not initially be numerical, for example a terrorist attack is a one-off observation probably represented later by an indicator. Therefore, inclusion of $\mathbf{z}_{t-\delta}$ may be via an adjustment rather than direct modelling (see §7.2 below). The model now is $\mathsf{M}_{\mathbf{y}_t}(\mathbf{y}_t|\mathbf{Y}_{t-1}^{t-k}, \mathbf{Z}_{t-\delta}^{t-s}, \widetilde{\mathbf{q}}_t, \boldsymbol{\theta}, \boldsymbol{\rho}_t)$, so we need to consider its possible formulation, and do so as a non-linear relationship. Clements and Galvão (2004), Sensier, Artis, Osborn and Birchenhall (2004) and Kock and Teräsvirta (2010) demonstrate the relevance of non-linear models for forecasting. Here, the emphasis is on incorporating components that facilitate 'jumps' to capture location shifts.

5.2 Non-linear functions

There are two possible objectives when forecasting using non-linear models: either the non-linear model attempts to forecast the structural break as in §7.1 below, or it is used to update after a break has occurred as discussed in §7.2. Models in the former class attempt to directly model structural breaks in the form of location shifts, relying on regularities in the break process to capture the transition from one regime to another. Such models include switching regression models and the various forms of smooth-transition regression (STR) models. This assumes that information about breaks is embodied in the regular information set $\mathcal{J}_t$. Forecasts for this class of non-linear model approach the unconditional mean of the process as the horizon lengthens, see Clements and Krolzig (1998) and Clements and Hendry (2006). Models in the second class concern unpredicted breaks. In that case, the break was not forecast even with a non-linear model. However, once the break has occurred, a non-linear model may capture the transition to a new equilibrium. Models in this class include non-linear learning functions, such as the ogive function utilized in §8.

5.3 Threshold models

Consider a bank where its net worth, $N_t = A_{t-1} - L_{t-1}$, fluctuates with the value of its assets, A_{t-1}, mainly loans, less its liabilities, L_{t-1} (deposits) which are are relatively constant, when N_t/A_{t-1} is small, say 10%. When A_T falls sufficiently that $N_{T+1} < 0$, the bank is bankrupt and is closed, inducing a major regime or location shift. Such a large break may be predictable from the past 'regular' behaviour, and the change in A_T need not be much larger than usual: Lehman Brothers provide a possible example, where a fall of around 5% in A_T was sufficient for $N_{T+1} < 0$ as $N_T/A_{T-1} < 0.05$. This threshold effect could be captured by a threshold autoregressive (TAR) model, see Tong (1990), where:

$$\begin{aligned}\Delta N_t &= \Delta A_{t-1} - \Delta L_{t-1} \quad \text{if} \quad \Delta A_{t-1} + N_{t-1} > 0 \\ \Delta N_t &= B_t \quad\quad\quad\quad\quad\quad \text{if} \quad \Delta A_{t-1} + N_{t-1} \leq 0\end{aligned}$$

where B_t is degenerate, denoting bankruptcy.

The STR model generalizes the threshold model to allow for a smooth, rather than an abrupt, transition between regimes:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + (\boldsymbol{\theta}'\mathbf{x}_t)G(s_t;\gamma,c) + u_t \ \text{ where } \ u_t \sim \mathsf{IN}\left[0, \sigma_u^2\right] \tag{11}$$

for $t = 1, \ldots, T$, where $G(\cdot)$ is a transition function. Various distributional assumptions can be made for the transition function, and here we consider the logistic transition function:

$$G\left(s_t;\gamma,c\right) = \left[1 + \exp\left\{-\gamma\left(\frac{s_t - c}{\widehat{\sigma}_s}\right)\right\}\right]^{-1} = \left[1 + \exp\left\{-z_t\right\}\right]^{-1}. \tag{12}$$

In this monotonic transition, γ is the steepness parameter as a function of the transition variable, s_t, $\widehat{\sigma}_s$ is its standard deviation, and c is the switching threshold parameter. As $\gamma \to \infty$, the model becomes a two regime-switching regression model, and $\gamma > 0$ is an identifying restriction. These models focus on predictable shifts, and although their forecast performance to date is uninspiring, see *inter alia*, Granger and Teräsvirta (1993, ch.8), Tiao and Tsay (1994), Brooks (1997), Stock and Watson (1999), Clements, Franses and Swanson (2004) and Granger (2005), more favorable results are reported in Dahl and Hylleberg (2004) and Marcellino (2004). See Kock and Teräsvirta (2010) in this *Handbook* for a clear discussion and further references. Neural networks are another popular method of non-linear forecasting, see Swanson and White (1995, 1997) and Angstenberger (1996). Furthermore, Marcellino (2004) and Teräsvirta, van Dijk and Medeiros (2005) consider forecasting large numbers of macroeconomic variables using neural networks.

5.4 Approximating the logistic transition function

The model in (11) can be approximated by replacing the logistic transition function with a third-order Taylor expansion:[2]

$$y_t \simeq \boldsymbol{\beta}'\mathbf{x}_t + (\boldsymbol{\theta}'\mathbf{x}_t)\left[\frac{1}{2} + \frac{z_t}{4} - \frac{z_t^3}{48}\right] + v_t \tag{13}$$

which can be estimated as the linearized model:

$$y_t = \boldsymbol{\alpha}_1'\mathbf{x}_t + \boldsymbol{\alpha}_2'\mathbf{x}_t s_t + \boldsymbol{\alpha}_3'\mathbf{x}_t s_t^2 + \boldsymbol{\alpha}_4'\mathbf{x}_t s_t^3 + v_t, \quad v_t \underset{app}{\sim} \mathsf{IN}\left[0, \sigma_v^2\right] \tag{14}$$

For a scalar $\mathbf{x}_t$, the mappings from (11) to (14) are:

$$\alpha_1 = \beta + \frac{\theta}{2} - \frac{\theta\gamma c}{4\widehat{\sigma}_s} + \frac{\theta\gamma^3 c^3}{48\widehat{\sigma}_s^3};\ \alpha_2 = \frac{\theta\gamma}{4\widehat{\sigma}_s} - \frac{3\theta\gamma^3 c^2}{48\widehat{\sigma}_s^3};\ \alpha_3 = \frac{3\theta\gamma^3 c}{48\widehat{\sigma}_s^3};\ \alpha_4 = -\frac{\theta\gamma^3}{48\widehat{\sigma}_s^3}.$$

The key advantages of considering forms like (14) rather than (11) are:
(a) the former are always identified (whereas the latter are not under the null of linearity);
(b) the candidate regressor set $\mathbf{x}_t$ may be large, both from lagged values and many potential factors;
(c) the lags at which transition functions enter cannot be deduced theoretically so must be data based; and
(d) in-sample outliers, data contamination and location shifts still need handling.
Consequently, model selection is inevitably required. A structured approach that controls the null retention frequency of irrelevant variables while maintaining high efficiency for handling outliers under the null is essential, as addressed in section 6. The disadvantage of (14) is that polynomials are not bounded which could be particularly problematic in a forecasting context. Alternative expansions to (13) ensuring boundedness could be used. However, (14) is an intermediate step, after which encompassing tests will be used to obtain a bounded representation, so we do not anticipate using (14) directly for forecasting. First we consider testing for non-linearity in-sample as a precursor to selection.

5.5 Testing for non-linearity

A test of non-linearity is required to evaluate whether a given linear model is already sufficient, or whether a non-linear approximation must be utilized. Castle and Hendry (2010a) develop a low-dimension portmanteau test for use in a model selection framework. The eigenvectors of the data second-moment matrix of the $n > 1$ linear regressors $\{\mathbf{x}_t\}$ induce transformations to $\{\boldsymbol{\eta}_t\}$, then individual element squares and cubes, $\eta_{i,t}^2$ and $\eta_{i,t}^3$, are added to the linear model. For fixed regressors $\mathbf{x}_t$, the test that the added variables are irrelevant is an exact F-test with $2n$ degrees of freedom under the null (as there are

[2] The second derivative $G''(z)|_{z=0} = 0$, where $G(z) = \left[1 + e^{-z}\right]^{-1}$, so the z_t^2 term drops out of the Taylor expansion. There is still a quadratic component in s_t, as the cubic expansion of $z_t^3 = \gamma^3\left(\frac{s_t - c}{\widehat{\sigma}_s}\right)^3$ is included with $G'(z)|_{z=0} = \frac{1}{4}$ and $G'''(z)|_{z=0} = -\frac{1}{8}$. The fourth order term is $G''''(z)|_{z=0} = 0$, so the next relevant term is the fifth order term, $G'''''(z)|_{z=0} = \frac{1}{4}$. Inserting into the Taylor expansion yields $G(z) \approx \frac{1}{2} + \frac{z_t}{4} - \frac{z_t^3}{48} + \frac{z_t^5}{480}$. The orthogonal component of the fifth derivative relative to the third is likely to be small, so the third-order Taylor expansion in (13) seems sufficient.

n elements in $\boldsymbol{\eta}_t$), and is correctly sized. They show it has power against departures from linearity in many directions, including asymmetry or skewness. This test is an alternative to the functional-form, or heteroskedasticity, test in White (1980), but:
(a) tests for departures in up to the third derivative;
(b) reduces the dimensionality of such a test from $M = \frac{1}{3}n(n+1)(n+2)$ to $2n$;
(c) can be applied when $M >> T$ but $3n < T$; yet
(d) avoids potential high collinearity between functions of $\{\mathbf{x}_t\}$.
Generalizations to include additional functional forms, such as exponentials, are straightforward and result in a further n degrees of freedom per component. Importantly, the test can be applied after handling in-sample location shifts using IIS while retaining all linear regressors, as shown in Castle and Hendry (2009).

6 Selecting a forecasting model

In this section, we discuss automatic selection of a non-linear forecasting model, addressing requirement (5) of §1.1. Automatic model selection is an essential component of our strategy, since we envisage:
• a large set, n, of initial candidate variables, $\{\mathbf{x}_t\}$, some at high frequencies as discussed above;
• an even larger number of non-linear functions of $\{\mathbf{x}_t\}$, approximating ogive-type responses; and
• impulse-indicator saturation to handle data contamination, outliers and breaks, adding T additional candidates;
to make a total number of candidate variables denoted by N.

§6.1 outlines the salient ideas of the *Autometrics* approach. §6.2 describes IIS and explains the need to avoid non-linear functions aligning with outliers or breaks, since such 'spurious' relationships could prove detrimental when forecasting. §6.3 discusses the specific issues involved when modelling non-linearity, and §6.4 considers the reduction to a theory-based form, checked by encompassing.

6.1 Automatic model selection

A key feature of automatic multi-path search methods like *Autometrics* is that they can can handle more candidate variables, N, than observations, T, see Hendry and Krolzig (2005), and Doornik (2007). They achieve this by sequentially entering variables in blocks that are smaller than $T/2$, recording what variables are significant in that block, then dropping it to enter another block. The collected set of significant variables is then entered jointly, or the procedure is repeated if too many intermediate variables matter, and a reduction conducted to find the minimal congruent dominant subset. Expanding searches explore if any non-included variables may still matter.

Denoting the nominal rejection frequency of individual selection tests by α, then on average αN irrelevant variables will be retained by chance significance from the N initial candidates. By setting $\alpha \leq 1/N$, all but one of the irrelevant variables will be eliminated on average, so the costs of search are small. Also, even with many irrelevant candidate regressors, *Autometrics* has a null retention frequency close to its nominal size. Because IIS ensures near normality, the associated critical value, c_α, is not too large even for quite small values of α: e.g., $c_\alpha = 3.5$ at $\alpha = 0.0005$. It is also feasible to bias correct for selection, which reduces the mean-square errors (MSEs) of retained irrelevant variables at a small cost in increased MSEs for relevant variables. Finally, the retention of relevant variables is usually close to the theoretical power of the corresponding (one-off) significance test under the alternative.

6.2 Impulse-indicator saturation

Impulse-indicator saturation (IIS) adds an indicator for every observation to the candidate regressor set. After selection, the significant sub-set removes the impacts of breaks and outliers, as well as ensuring near normality, which is important for conducting conventional inference during search. Impulse indicators, denoted $1_{\{i=t\}}\forall t$, are generated for every observation, dividing these into J subsets, just $J = 2$ for the 'split sample' approach. Each subset forms an initial model to be searched, with *Autometrics* recording the significant indicators, and forming a joint model as the union of these. Then *Autometrics* re-selects the indicators that remain significant. Under the null that there are no outliers, αT indicators will be retained on average for a significance level α in a sample of size T. Consequently, Johansen and Nielsen (2009) establish that IIS is an efficient robust statistical method. For example, when $\alpha = 1/T$, only one observation is 'removed' on average, so IIS is 99% efficient for $T = 100$. Thus, despite adding greatly to the number of variables in any model selection search, IIS suffers little efficiency loss under the null of no contamination, as against a potentially large gain by avoiding unmodelled outliers and breaks. The important difference between outlier detection and IIS is illustrated by Hendry and Reade (2008).

Problems arise for non-linear model selection when fat-tailed distributions result in extreme observations. There is an increased probability that non-linear functions will align with outliers, in effect acting as indicators, and so being retained too often during search. For in-sample modelling this is not too problematic, as the representations will be similar, but for forecasting, can differ markedly. This can be illustrated even in a linear setting where an outlier is represented by $1_{\{t=s\}}$, which takes the value 1 in period s and 0 otherwise, for a regression between two unconnected variables:

$$y_t = \beta x_t + \delta 1_{\{t=s\}} + u_t \text{ where } u_t \sim \mathsf{IN}\left[0, \sigma_u^2\right] \tag{15}$$

$$x_t = \gamma 1_{\{t=s\}} + v_t \text{ with } v_t \sim \mathsf{IN}\left[0, \sigma_v^2\right] \tag{16}$$

where $\beta = 0$, $\mathsf{E}[v_t u_t] = 0$ and $\sigma_u = \sigma_v = 1$. If the indicator is omitted from the model $y_t = \lambda x_t + u_t$, the estimated coefficient $\widehat{\lambda}$ is:

$$\begin{aligned}\widehat{\lambda} = \frac{\sum x_t y_t}{\sum x_t^2} &= \frac{\delta\gamma \sum 1_{\{t=s\}}^2 + \sum (\delta v_t + \gamma u_t) 1_{\{t=s\}} + \sum v_t u_t}{\gamma^2 \sum 1_{\{t=s\}}^2 + 2\gamma \sum v_t 1_{\{t=s\}} + \sum v_t^2} \\ &= \frac{\delta\gamma + (\delta v_s + \gamma u_s) + \sum v_t u_t}{\gamma^2 + 2\gamma v_s + \sum v_t^2}\end{aligned} \tag{17}$$

as $\sum 1_{\{t=s\}}^2 = 1$. Since $\mathsf{E}[v_s] = \mathsf{E}[u_s] = 0$ and (as the denominator cannot be zero) we approximate by:

$$\mathsf{E}\left[\widehat{\lambda}\right] \simeq \frac{\delta\gamma}{\gamma^2 + T}$$

as:

$$\mathsf{V}\left[\widehat{\lambda}\right] = \frac{1}{\sum x_t^2} \simeq \frac{1}{\gamma^2 + T} \tag{18}$$

then:

$$\mathsf{E}\left[\mathsf{t}_{\widehat{\lambda}}^2\right] \simeq \frac{\delta^2\gamma^2}{\gamma^2 + T}. \tag{19}$$

To illustrate, suppose $\delta = 6$, $\gamma = 5$, and $T = 100$. Then:

$$\mathsf{t}_{\widehat{\beta}}^2 \simeq \frac{6^2 \times 5^2}{5^2 + 100} = 7.2 \tag{20}$$

While single outliers need to be quite large for this effect, that is plausible when considering non-linear transformations.

6.3 Modelling non-linearity

Econometric modelling of non-linear processes presents problems over and above those encountered when developing linear econometric models. Identifying a unique non-linear representation of an economic local data generating process (LDGP) can be difficult as there are an infinite number of potential functional forms. The three key concerns for the econometrician are specification of the functional form, selection of the relevant variables, lags etc., and identification of the parameters.

The methodology proposed here is to commence from very general models which nest a class of LDGPs suggested by theory. When the functional form and the set of relevant variables are both unknown, the non-linearity test in §5.5 has power to reject a false null, even for collinear linear variables and when there are potentially more non-linear functions than observations. Hence, it can be used in situations where other non-linearity tests are infeasible. The pre-test can establish whether there is a substantive non-linearity in-sample to merit modelling. If so, the non-linearity is first approximated by a general cubic, and model selection carried out with IIS, removing the identification problem of standard non-linear modelling by commencing with an identified approximation. Evidence on the forecast performance of non-linear models suggests that the non-linearity must be tightly specified to avoid poor forecast performance. Therefore, encompassing tests check the final selection against the theoretical functional form for further reductions. Thus, the algorithm tackles the specification, selection and identification problems.

The algorithm commences with the general polynomial approximation:

$$y_t = \boldsymbol{\beta}'\mathbf{w}_t + \boldsymbol{\tau}'\mathbf{g}(\mathbf{w}_t) + \sum_{i=1}^{T} \delta_i 1_{\{t=t_i\}} + v_t \quad \text{for} \quad t = 1, \ldots, T, \tag{21}$$

where $\mathbf{w}_t$ is the $n \times 1$ vector of potentially relevant variables and $\mathbf{g}(\mathbf{w}_t)$ is the $m \times 1$ vector of non-linear transformations. (21) nests (14), the polynomial approximation of the STR model, with $\mathbf{w}_t = (\mathbf{x}_t, \mathbf{s}_t)'$. It also allows for other breaks via impulse indicators, as well as predictable breaks captured by regimes parameterized by (γ, c) above, which could be vectors allowing for more than 2 regimes. The variables are reparameterized by double de-meaning all regressors as in Castle and Hendry (2009). Selection is undertaken using the technique for more variables than observations in *Autometrics*, with especially tight significance levels to control the null rejection frequency given the large number of non-linear functions. Mis-specification tests are applied to check the congruency of the reduction from (21).

6.4 Reduction to a theory-based form

Finally, the selected model is tested against the econometrician's preferred non-linear functional form, $f(\cdot)$ say. When that specification is a good representation of the data, the polynomial approximations should become insignificant in the augmented model, and be replaced by a significant $f(\cdot)$. This procedure avoids the identification problems associated with modelling specific classes of non-linearity, yet allows one to conclude with a specific functional form if that is the best representation.

The LDGP has two unknown components, the relevant variables and their functional forms. The polynomial approximation overcomes the latter and multi-path search solves the former, while the indicators capture outliers and breaks as well as avoiding over-retention of spurious non-linear functions. In practice, the precise set of relevant variables, transition variables, lag lengths, outliers, and breaks are not known, so the search commences with the general unrestricted model (GUM):

$$\begin{aligned} y_t &= \sum_{i=1}^{n}\beta_i x_{i,t} + \sum_{i=1}^{n}\sum_{j=i}^{n}\kappa_{i,j}x_{i,t}x_{j,t} + \sum_{i=1}^{n}\sum_{j=i}^{n}\sum_{k=j}^{n}\theta_{i,j,k}x_{i,t}x_{j,t}x_{k,t} + \sum_{i=1}^{n}\sum_{j=1}^{m}\gamma_{i,j}x_{i,t}s_{j,t} \\ &\quad + \sum_{i=1}^{n}\sum_{j=1}^{m}\lambda_{i,j}x_{i,t}s_{j,t}^2 + \sum_{i=1}^{n}\sum_{j=1}^{m}\phi_{i,j}x_{i,t}s_{j,t}^3 + \sum_{i=1}^{T}\delta_i 1_{\{t=t_i\}} + \epsilon_t \end{aligned} \tag{22}$$

where $\epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$. There are:
• n potential regressors $\mathbf{x}_t$, including lags of $\mathbf{w}_t$ and an intercept;
• m potential transition variables $\mathbf{s}_t$; and
• T indicators;
resulting in $M + 3nm + T$ variables in the GUM.
As the postulated model is a threshold/STR model, we separate out the potential transition variables, but in principle the transition variables could just be included in the set of potential regressors $\mathbf{w}_t$, resulting in a fourth-order polynomial GUM.

The model selected by an *Autometrics* procedure should capture the non-linearity inherent in an STR model, while enabling a more general specification to be selected if required. The approximation to the LSTR model can be tested by augmenting the final model selected by *Autometrics* with the non-linear component of the LSTR model. Eitrheim and Teräsvirta (1996) and Teräsvirta (1998) propose a test of no additive non-linearity as follows. Suppose k relevant variables were retained ($k \leq n$), and one transition variable was selected ($1 \leq m$) given by s_1, then the test of the approximation to the LSTR model would be:

$$\mathsf{H}_0\text{: } \boldsymbol{\pi} = \boldsymbol{\mu} = \boldsymbol{\psi} = \mathbf{0} \tag{23}$$

for the regression:

$$\begin{aligned} y_t &= \sum_{i=1}^{k}\widetilde{\tau}_i x_{i,t} + \sum_{i=1}^{k}\widetilde{\pi}_i x_{i,t}s_{1,t} + \sum_{i=1}^{k}\widetilde{\mu}_i x_{i,t}s_{1,t}^2 + \sum_{i=1}^{k}\widetilde{\psi}_i x_{i,t}s_{1,t}^3 \\ &\quad + \sum_{i=1}^{k}\left(\widetilde{\theta}_i x_{i,t}\right)\left[1 + \exp\left\{-\widetilde{\gamma}\left(\frac{s_{1,t} - \widetilde{c}}{\widehat{\sigma}_{s_1}}\right)\right\}\right]^{-1} + \widetilde{e}_t, \end{aligned} \tag{24}$$

where $\widetilde{}$ denotes the estimated parameters after including the LSTR model. Thus, while the polynomial approximation enables the econometrician to remain agnostic regarding the functional form at the stage of the GUM, the resulting specific model can be tested against the econometrician's preferred functional form as in (24). This provides a rigorous test for the validity of the postulated non-linear model.

7 Forecasting breaks and forecasting during breaks

In this section, we draw together the previous analyses discussing aspects of the six necessary conditions for successfully forecasting breaks. It is helpful to consider two potential settings. In the first, considered briefly in §7.1, we focus on a regime shift rather than an unanticipated break. A related switch has occurred before, and has enabled parameters to be estimated that are relevant to forecasting a repeat switch. In the second, the subject of §7.2, the forecast is having to take place during a new break, which may be a new location shift (§7.2.1) or a new exponential break (§7.2.3).

7.1 Forecasting a regime shift

A shift may have occurred in the past which provides evidence to facilitate a repeat forecast. The recent financial crisis does not make such a comment read too well, as there were many possible precursors, none of which enabled fore-warning of the precipitous collapses that ensued. Nevertheless, a variety of models have been proposed that could allow forecasts of shifts, of which we consider the threshold class below as an example. Others include the stop-break model in Engle and Smith (1999) and the ACR model of Bec, Rahbek and Shephard (2008), as well as regime-switching models and Markov switching VARs, where in both cases, past data enable the estimation of regime-switching probabilities. However, the forecasting performance of such models has not so far been too stellar, even in simulations where a switching process generates the data: see e.g., Clements and Krolzig (1998). The Bank of England use a suite of models to detect breaks, the pooling of which may proxy a non-linear transition function, see Labhard, Kapetanios and Price (2007).

Panel-data models for predicting crises are often couched in terms of indicators formed from observed data series: see among others, Kaminsky, Lizondo and Reinhart (1998b, 1998a), Berg and Pattillo (1999), Demirgüç-Kunt and Detragiache (2000), Kumar, Moorthy and Perraudin (2003), and Komulainen and Lukkarila (2003).[3] Most of the variables entering such analyses are part of the regular forces in economics, $\mathcal{J}_{t-1}$, albeit including cross-country variation, rather than different sources, but often have increased impacts at thresholds. As ever, to forecast the next break or crisis requires that the same forces operate in the same way, and that the model remains sufficiently constant for a viable forecast.

Hendry and Ericsson (1991) provide an example of forecasting a repeated break. They use a non-linear learning function for the UK based on functional form and parameter estimates previously obtained by Baba, Hendry and Starr (1992) for the US. The learning function was estimated non-linearly in subsequent analysis in Ericsson (1999). This offers an alternative to non-linear models that rely on past regime shifts within the data set in use. Instead, additional sources of information, such as whether a similar break has occurred elsewhere, may yield more rapid updating of forecasts during a break transition.

While the general polynomial form (22) can provide a good approximation in-sample, for forecasting purposes the model needs to be tightly specified and the final encompassing stage achieves this objective. Therefore it is sufficient to examine the forecast performance of specific non-linear functional forms as successful model selection will result in an identified specific non-linear model. Non-linear functional forms that are able to partly predict regime shifts must be of the form of a threshold type model, contingent on the requirements for predictability discussed in §2. A location shift is like a threshold effect where a 'jump' occurs when a variable exceeds a given level. Such a 'break' is then latent in the DGP. There are many possible threshold models, and as yet it is unclear which are most relevant, but it may be possible to approximate the class by one general member. At least one past shift is essential to identify the threshold reaction, which then requires that form to stay constant long enough to be the basis for forecasting a later shift.

We assess the forecast performance of the LSTAR model where the DGP is given by:

$$y_t = \mu + \rho y_{t-1} + \mu^* \left[1 - \exp\left(-\gamma\left(\frac{y_{t-1} - c}{\sigma_y}\right)\right)\right]^{-1} + \epsilon_t \tag{25}$$

where $\epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$ and $\gamma = 3$. We consider three break sizes of magnitude σ_y, $3\sigma_y$ and $5\sigma_y$, where $\mathsf{V}[y] = \sigma_y^2$, and four break probabilities, 1%, 5%, 10% and 20%, which are governed by the threshold parameter c. We compute twenty 1-step ahead forecasts for two sample sizes, $T = 100$ and 1000, discarding the initial 100 observations. $M = 1000$ replications were undertaken, discarding any

[3] One reason why we specifically call our procedure impulse-indicator saturation to minimize confusion with the many senses of 'indicator' and 'impulse'.

replications that did not converge. DGP parameter values were used as initial conditions for the LSTAR estimation.

In the first simulation, we consider forecasting at the forecast origin, regardless of which regime the process is currently in. In the second simulation, we select the process such that there is a break at the forecast origin, and compute just the 1-step forecast at $T+1$.[4] The forecasts are based on the in-sample parameter estimates, so are not recursive.

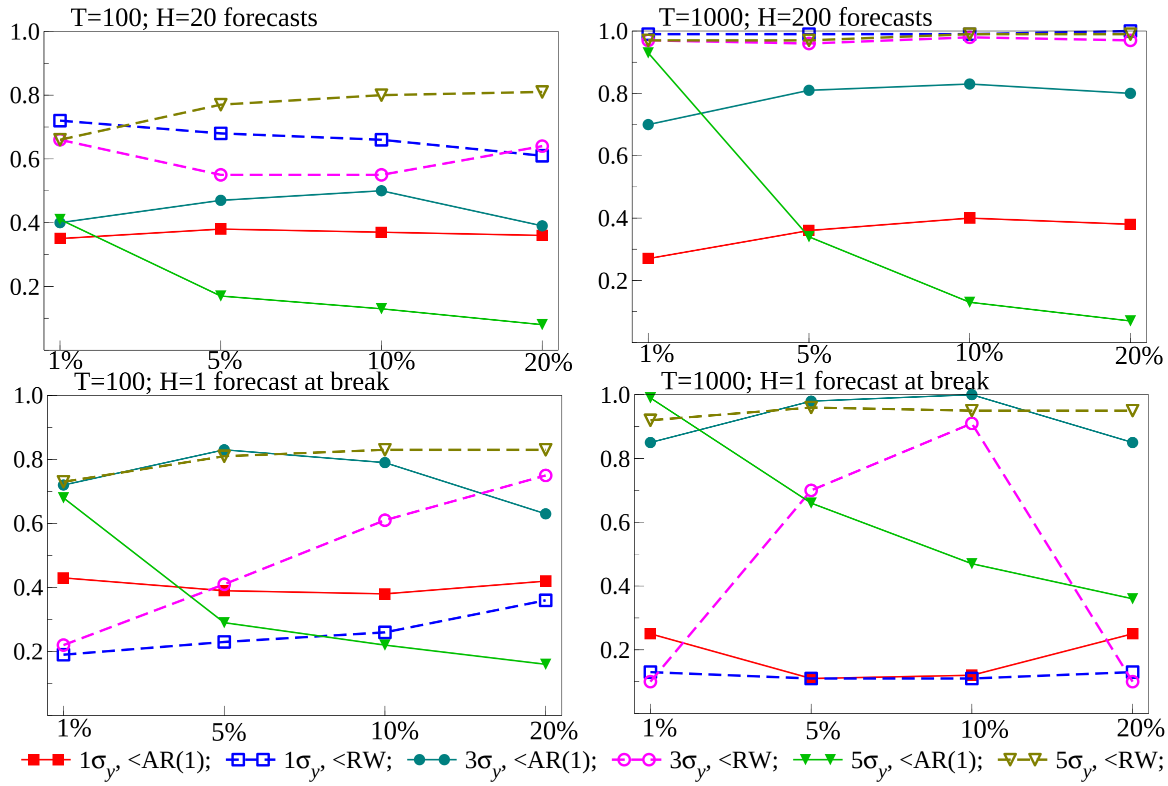

Figure 2: Fraction of draws in which LSTR model MSFE is lower than AR(1) and RW models

Results are recorded in figure 11.2, which reports the fraction of draws in which the LSTAR model MSFE is lower than those of benchmark AR(1) and random walk (RW) models. This is a more robust statistic than average MSFE. Panels (a) and (b) correspond to the first experiment, and panels (c) and (d) for forecasts at the break point. The horizontal axis records the percentage of breaks, controlled by c.

The results suggest that when the break is of a moderate size and occurs fairly frequently, the LSTAR forecasts are comparable to the benchmark AR(1) and RW forecasts. When breaks are small, or when they are large but they occur infrequently, a linear approximation is often preferable. For a break at the forecast origin, the information in the non-linear model does yield benefits over the benchmark forecasts. In this case, the smooth-transition model forecasts considerably better than the AR(1) and RW at moderate thresholds, but again, when breaks are large or when shifts occur frequently, the linear models are preferable. These results rely on a constant break process, such that a structural break at the forecast origin must have occurred previously for the LSTAR estimates to be accurate. Hence, there is evidence of a benefit in using non-linear models to forecast at a break origin, but this is dependent on the size and frequency of breaks in the in-sample process. These problems are mitigated, but not resolved, at larger sample sizes.

[4]The definition of a break is ambiguous in a smooth transition model, so we define the break as $y_T < \mathsf{E}[y_t^*]$ and $y_{T+1} \geq \mathsf{E}[y_t^*]$, where y_t^* is the process in the upper regime. For a break of $3\sigma_y$, $\mathsf{E}[y_t^*] = 5$.

7.2 Forecasting during new breaks

Forecasting an imminent location shift is key to avoiding serious forecasting and policy failures. However, that goal remains elusive. We now consider forecasting during a break, noting two approaches:
(a) forecasting facing a shift in the intercept of a model of a variable y induced by another observable variable z; and
(b) when the shift is of a known exponential form and timing.
The accuracy of break forecasts will also be addressed in this section.

Even forecasting during a break is difficult. An 'external' break is defined as one in which the forecasting model stays constant, whereas an 'internal' break is one in which the model itself shifts. For an external break, the resultant changes in collinearity between explanatory variables (both included and incorrectly omitted) inevitably increase the forecast-error uncertainty, as measured by MSFE. When a break is internal, even when of a known form and timing, estimation uncertainty entails that robust devices perform as well. Nevertheless, if the 'outside' variable that enters the threshold function can trigger a crossing by its stochastic behaviour alone, and enters lagged, then forecasting a future shift becomes possible in principle, albeit with the converse danger of 'predicting' too many false breaks.

7.2.1 Forecasting during a new location shift

The idea explored in this section is to use changes in z to forecast breaks in y which have happened sufficiently far in advance, here one period. The DGP we consider is given by:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \phi_y \\ \phi_z \end{pmatrix} + \begin{pmatrix} \rho_y & 1_{\{t \geq T^*\}} \\ 0 & \rho_z \end{pmatrix} \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ v_t \end{pmatrix} \tag{26}$$

where $1_{\{t \geq T^*\}}$ is an indicator function which is zero until T^* and 1 after. In (26), $\epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2]$, $v_t \sim \mathsf{IN}[0, \sigma_v^2]$, and $\mathsf{E}[\epsilon_t v_s] = 0$, $\forall t, s$. Thus, y_t does not depend on z_{t-1} till $t = T^*$, at which point it experiences an induced location shift of γ_z as $\mathsf{E}[y_t] = \phi_y / \left(1 - \rho_y\right) = \gamma_y$, $\mathsf{E}[z_t] = \gamma_z = \phi_z / \left(1 - \rho_z\right)$ till T^*. At that stage, z_{t-1} could not usefully be included in the model of y_t as its coefficient in (28) was zero till then. Then $\mathsf{E}[y_{T^*}] = \phi_y + \rho_y \gamma_y + \gamma_z$ and:

$$\begin{aligned} \mathsf{E}[y_{T^*+1}] &= \left(1 + \rho_y\right)\left(\phi_y + \gamma_z\right) + \rho_y^2 \gamma_y, \\ \mathsf{E}[y_{T^*+2}] &= \left(1 + \rho_y + \rho_y^2\right)\left(\phi_y + \gamma_z\right) + \rho_y^3 \gamma_y, \\ \mathsf{E}[y_{T^*+h}] &= \sum_{i=0}^{h} \rho_y^i \left(\phi_y + \gamma_z\right) + \rho_y^{h+1} \gamma_y \end{aligned}$$

gradually converging to the new equilibrium mean of $(\phi_y + \gamma_z)/(1 - \rho_y)$. Figure 11.3 plots the evolution of $\mathsf{E}[y_{T^*+i}]$ for the DGP values given in §7.2.2.

Large and systematic forecast failure would ensue if no action was taken to correct the pre-break model, $y_t = \phi_y + \rho_y y_{t-1} + \epsilon_t$, or its forecasts. An intercept correction (IC) based on the first major error, with an average of γ_z, would be an effective robust device. If the usual calculation of doubling the error variance held, then the root mean square forecast error (RMSFE) would be $\sqrt{2\sigma_\epsilon^2}$ as against $\sqrt{\gamma_z^2 + \sigma_\epsilon^2}$ for not adjusting, which is larger whenever $\gamma_z > \sigma_\epsilon$. The differenced forecast, $\widetilde{y}_{T^*+h|T^*+h-1} = y_{T^*+h-1}$, would have average errors of $(\gamma_z + \rho_y \gamma_y)$, $\rho_y(\gamma_z + \rho_y \gamma_y)$, $\rho_y^2(\gamma_z + \rho_y \gamma_y)$, etc., so quickly become about half the uncorrected errors, but would also have an increased error variance. The estimated new model, which we write from $t = T^*$ as $y_t^* = y_t - \phi_y - \rho_y y_{t-1} = \lambda z_{t-1} + \epsilon_t$, assuming a sufficiently large prior sample that estimation uncertainty can be ignored for ϕ_y, ρ_y, yields at T^*:

$$\mathsf{E}\left[\breve{\lambda}_{T^*}\right] = \mathsf{E}\left[\frac{y_{T^*}^* z_{T^*-1}}{z_{T^*-1}^2}\right] = \mathsf{E}\left[\lambda + \frac{z_{T^*-1}\epsilon_{T^*}}{z_{T^*-1}^2}\right] \simeq \lambda \tag{27}$$

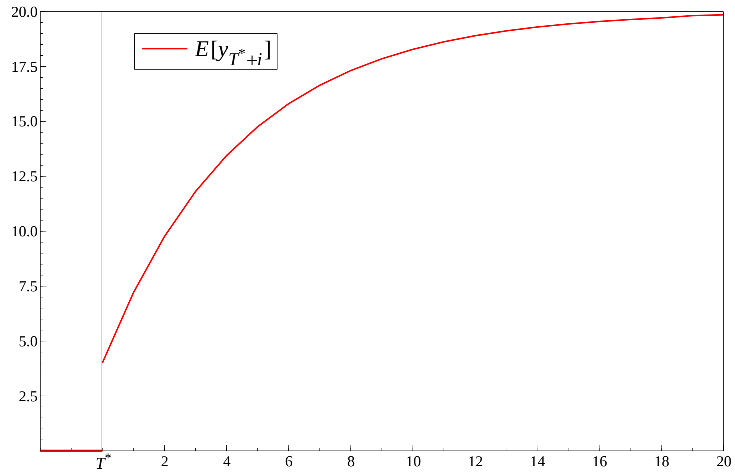

Figure 3: $\mathsf{E}\left[y_{T^*+i}\right]$ for $i = 0, \ldots, 20$ for DGP (26) with $\phi_y = 0$, $\phi_z = 2$, $\rho_y = 0.8$ and $\rho_z = 0.5$.

so acts like an intercept correction, and could well have small mean errors, but its RMSFE is hard to obtain analytically. Updating by extending the indicator would rapidly reduce the estimation variance, as we now show in a simulation study.

7.2.2 Monte Carlo

We set $T^* = 95$, $T = 100$, $\phi_y = 0$, $\phi_z = 2$, $\rho_y = 0.8$ and $\rho_z = 0.5$ with $M = 1000$ replications undertaken for the DGP:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix} + \begin{pmatrix} 0.8 & 1_{\{t \geq 95\}} \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ v_t \end{pmatrix} \tag{28}$$

Then $\mathsf{E}\left[y_{95}\right] = 4$, $\mathsf{E}\left[y_{96}\right] = 7.2$, $\mathsf{E}\left[y_{97}\right] = 9.76$, $\mathsf{E}\left[y_{98}\right] = 11.81$, converging towards the new equilibrium mean of 20, see fig. 11.3.

Figure 11.4 records the breakpoint Chow test rejection frequencies after the break has occurred for the estimated unadjusted model, which coincides with the DGP in-sample. The break is clearly detectable, even at T^*, as the mean has shifted by 4 standard deviations.

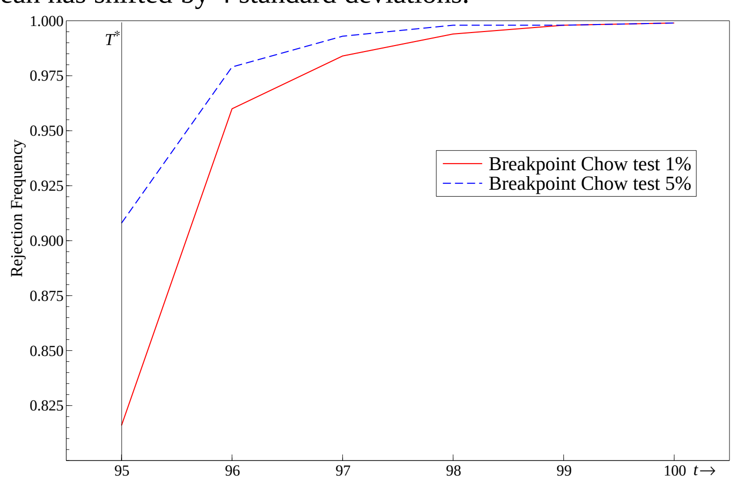

Figure 4: Breakpoint Chow test rejection frequencies

Table 11.1 reports the seven forecasting models considered, and fig. 11.5 records the mean error (ME) and RMSFE for 1-step ahead forecasts over $T^* + h$ for $h = 0, \ldots, 5$ for each model, calculated

	Model	$\widehat{y}_{T^*+h\|T^*+h-1} =$	Notes
[1]	DGP	$\widehat{\phi}_y + \widehat{\rho}_y y_{T^*+h-1} + z_{T^*+h-1}$	Impose $\lambda = 1$, i.e., infeasible benchmark
[2]	Unadjusted	$\widehat{\phi}_y + \widehat{\rho}_y y_{T^*+h-1}$	Estimated in-sample with no adjustment
[3]	Unadjusted recursive	$\widetilde{\phi}_y + \widetilde{\rho}_y y_{T^*+h-1}$	Recursive updating of unadjusted model
[4]	DGP recursive	$\widetilde{\phi}_y + \widetilde{\rho}_y y_{T^*+h-1} + \widetilde{\lambda} z_{T^*+h-1}$	Correct model but estimating λ recursively
[5]	DGP post-break	$\widehat{\phi}_y + \widehat{\rho}_y y_{T^*+h-1} + \check{\lambda} z_{T^*+h-1}$	Model estimated over post-break sample
[6]	DD	y_{T^*+h-1}	Differenced device
[7]	IC	$\widetilde{\phi}_y + \widetilde{\rho}_y y_{T^*+h-1} + \widetilde{u}_{T^*+h-1}$	Intercept correction, adding in last residual

Table 11.1: Forecasting models. Notes: $\widehat{\phi}_y$, $\widehat{\rho}_y$ are estimated over $t = 1, \ldots, T^* - 1$. $\widetilde{\phi}_y$, $\widetilde{\rho}_y$ are estimated over $t = 1, \ldots, T^* + h$. $\check{\lambda}$ is estimated over $t = T^*, \ldots, T^* + h$. $\widetilde{u}_{T+h-1}$ is the residual from the last in-sample observation.

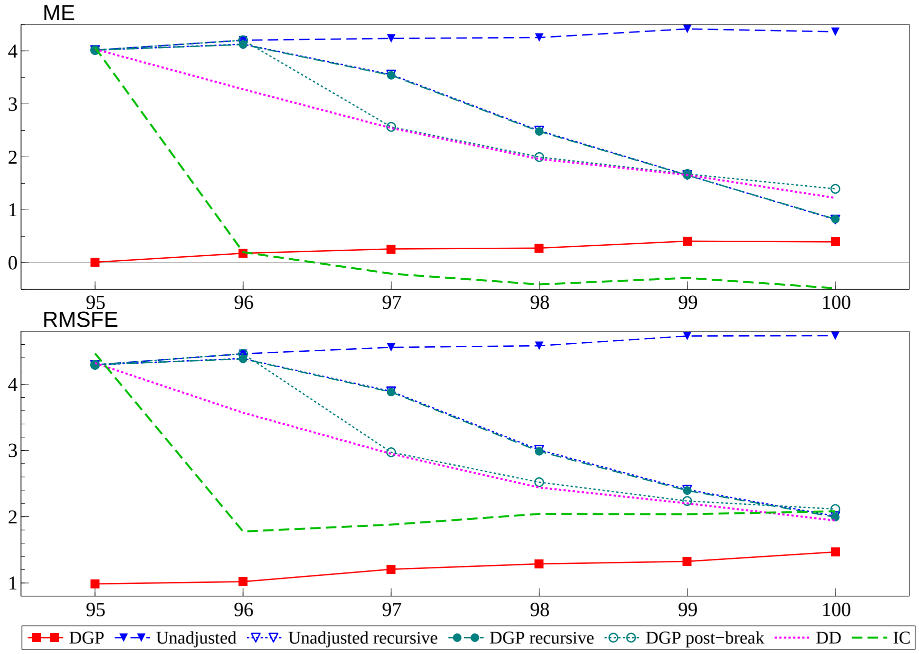

Figure 5: ME and RMSFE for each forecasting model when a break occurs at $T^* = 95$.

by:

$$\text{ME} = \frac{1}{M}\sum_{i=1}^{M}\left(y_{T^*+h} - \widehat{y}_{T^*+h|T^*+h-1}\right), \quad h = 0, \ldots, 5$$

$$\text{RMSFE} = \left(\frac{1}{M}\sum_{i=1}^{M}\left(y_{T^*+h} - \widehat{y}_{T^*+h|T^*+h-1}\right)^2\right)^{\frac{1}{2}}, \quad h = 0, \ldots, 5.$$

Monte Carlo uncertainty is measured by the standard deviation of the mean forecast error for each model at each horizon. The MC uncertainty is similar across all models and is 1.6 on average.

If the break was known to occur at T^*, and the weight on z_{t-1} were also known as in [1], there would be no systematic forecast error. However, if the break was unpredictable, all forecasting devices would make a forecast error of magnitude $\gamma_z = 4$ at the breakpoint. If the unadjusted model (which is correct in-sample) is retained with the coefficients estimated over the in-sample period, [2], there would be a systematic error of size γ_z for all forecast periods. If instead the unadjusted model was recursively estimated, [3], the bias begins to decline after two periods due to the recursively estimated coefficient $\widetilde{\rho}_y$ converging to unity, imposing a unit root to account for the break. The mis-specification due to omitting z_{t-1} is not directly causing the forecast failure: if the model included z_{t-1} for the whole sample, and the coefficients are again estimated recursively, [4], the estimate $\widetilde{\rho}_y$ is almost identical to that of the mis-specified model, and the estimate of λ (the coefficient on z_{t-1}) is close to zero despite the true parameter having a weight of unity after the break point.

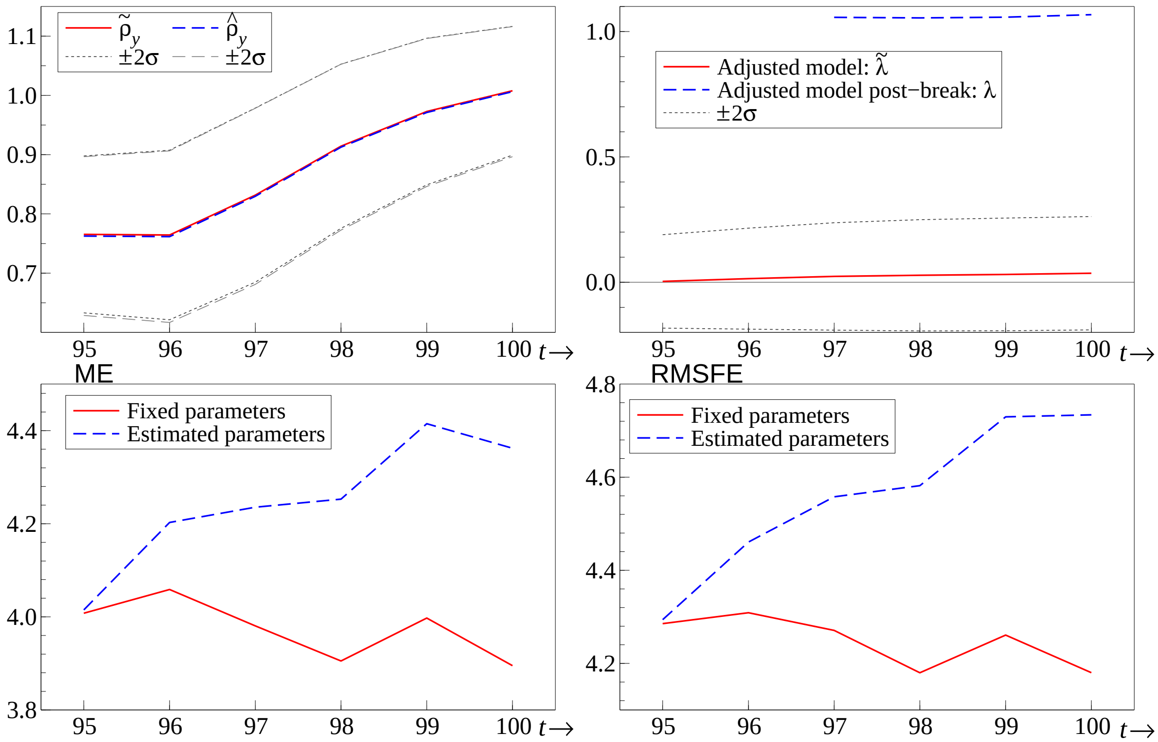

Figure 6: Parameter estimates for unadjusted and adjusted models, and impact of parameter estimation uncertainty on ME and RMSFE.

Fig. 11.6, panel a, records the recursively-estimated coefficient $\widetilde{\rho}_y$, for the unadjusted and adjusted models, both of which converge rapidly on unity, and panel b records the recursive estimates $\widetilde{\lambda}$ which are close to zero (solid line). The adjusted model would need to be estimated only over the post-break period in order to obtain an unbiased estimate of λ, [5]. This is infeasible at the break point as the $\{z\}$ process enters with a lag, but at $t = T^* + 1$, there is one observation from which to estimate λ by $\breve{\lambda}_{T^*}$ as in (27) Hence, we can obtain a forecast for y_{T^*+2} from:

$$\widehat{y}_{T^*+2|T^*+1} = \widehat{\phi}_y + \widehat{\rho}_y y_{T^*+1} + \breve{\lambda}_{T^*} z_{T^*+1} \tag{29}$$

making the forecasts for T^* and T^*+1 equal to the unadjusted model, then including z_{t-1} with coefficient estimate $\breve{\lambda}_{T^*}$. For the $T^* + 3$ forecast, there are two post-break data points from which to estimate λ, etc., so the estimate will become more accurate as time moves past the break point. The coefficient estimate $\breve{\lambda}_{T^*+i}$ is reported in fig. 11.6, panel b (dashed line: standard errors are not reported as $\breve{\lambda}_{T^*}$ is only based on one observation). Despite only having 1–4 observations from which to obtain an estimate, the coefficient estimates are close to λ, so the mean forecast error declines from period $T^* + 2$ onwards.

The forecasts from the adjusted model estimated post-break converge to the differenced device forecast. The mean error for DD, [6], declines at a rate $\phi_y^h \gamma_z$, so the average errors are 4, 3.2, 2.56, 2.05, but have an increased error variance. The intercept corrected device, [7], correctly adjusts one period after the break for the forecast error made at the break-point, and subsequent forecast errors are close to zero. The IC forecast-error uncertainty, using the mean absolute error (MAE), is approximately 1.5 for $h = 1, \ldots, 5$ which matches the theory ($\sqrt{2}\sigma_\epsilon$), although there is a small bias causing an increase in RMSFE. In contrast, the uncertainty for the unadjusted forecast $\sqrt{(\gamma_z^2 + \sigma_\epsilon^2)}$ matches the MAE of 4.1, which is substantially larger. We also examined an IC that used an average of the last 2 residuals but there was no improvement relative to the IC using 1 residual. Alternative forms of intercept correction would result in slightly different forecasts, but the conclusions would not change significantly. The IC forecast outperforms all other devices until $T = 100$.

In the above experiments, we set $\phi_y = 0$ but a non-zero mean for the y_t process has little effect on the forecasts, so results for $\phi_y \neq 0$ are not reported. The above simulations included the impact of in-sample parameter estimation uncertainty, but using fixed parameters instead results in little difference in forecast accuracy. Fig. 11.6 c and d compares ME and RMSFE for the unadjusted in-sample model with fixed and estimated parameters. Estimating the parameters recursively has an important impact when the model is mis-specified, as $\widehat{\rho}_y$ rapidly converges on unity to account for the break.

Finally we consider the impact of the break magnitude on forecast accuracy rankings. When $\gamma_z < \sigma_\epsilon$ the RMSFE for the IC would be larger than the unadjusted forecast, so for $\phi_z = 0.5$, then $\gamma_z = 1 = \sigma_\epsilon$. The sequence of outcomes is $\mathsf{E}[y_{95}] = 1$, $\mathsf{E}[y_{96}] = 1.8$, $\mathsf{E}[y_{97}] = 2.44$, converging to a new equilibrium of 5. Hence, the break at T^* is of size σ_ϵ. Figure 11.7 records the ME and RMSFE in the top panels. The rankings based on ME are similar to the previous case, although the estimated post-break model has a near-zero mean at $h = 2$, but a much larger RMSFE at $h = 2$ as $\breve{\lambda}$ is highly uncertain. When the break is small, it is difficult to identify the coefficient on $\breve{\lambda}$ until two or more observations are available. Other than model [5], all forecasting devices have similar RMSFEs here as expected given the theory, although the DD dominates in this example. When $\gamma_z < \sigma_\epsilon$ (panels c and d), with $\phi_z = 0.4$ so $\gamma_z = 0.8$, the IC has a higher RMSFE than the unadjusted model, with the DD and recursive forecasts outperforming.

To conclude, when the break is large, such that the mean of the process that enters at T^* is significantly larger than σ_ϵ, IC is preferable to many forecasting devices. It provides unbiased forecasts with the increased variance being a small price to pay. If the break is small, IC can be more costly than DD. All devices miss the break-point as the break was unpredictable, even knowing that z_{t-1} enters at T^* with a coefficient of unity. The known model with estimated parameters provides an upper bound on how (un)forecastable the effects of such breaks are. We next look at forecasting during a new exponential break and find that IC and DD are equivalent in that case.

7.2.3 Forecasting during a new exponential break

When a break is internal to a model, so it experiences a shift, estimation uncertainty can be so large even when the form and timing of the shift are known that robust devices perform as well as the estimated DGP. Castle *et al.* (2010) consider a simple DGP given by:

$$y_t = \alpha + \lambda \left[1 - \exp\left(-\psi \left[t - T + 1\right]\right)\right] 1_{\{t \geq T\}} + \epsilon_t \;\text{ with }\; \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{30}$$

where $\psi > 0$. The time T of the break and the form it takes are assumed known, which could be due to information in $\mathcal{K}_{t-\delta}$, but the parameter values are unknown. They consider 1-step ahead forecasts over $T + 1, \ldots, T + h$ for the four devices of an IC, a DD, an estimate of (30), and ignoring the break. Analytical MSFEs are shown in table 11.2.

The ranking varies with the parameter values in the DGP and with the size of the break relative to σ_ϵ^2. At 1-step ahead, the first three are identical, despite the second not estimating any parameters, and in general the fourth will be the largest for breaks bigger than σ_ϵ. The 2-step IC assumes an average across the two forecast errors, so only estimates one parameter. This 'smoothing' imparts a bias as the error

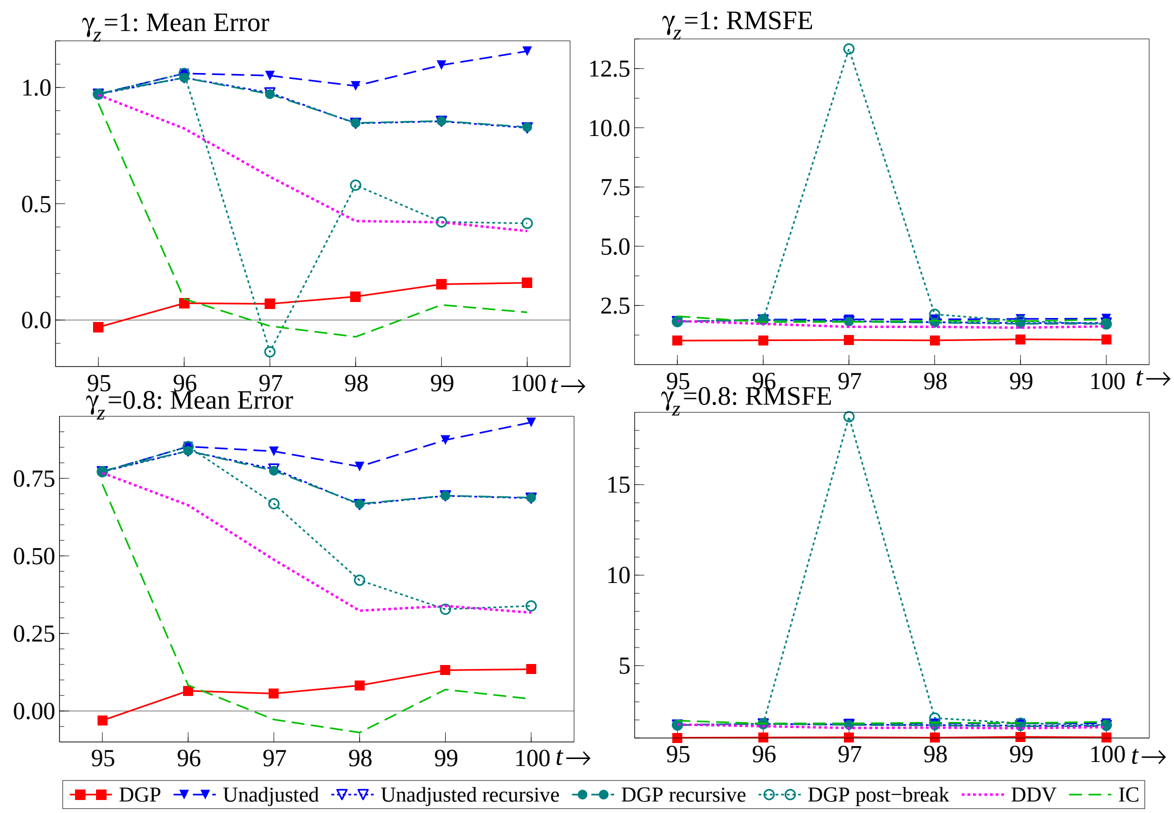

Figure 7: ME and RMSFE for smaller breaks in (28).

is increasing, and the DD MSFE is generally smaller than that of the IC. The DGP needs non-linear optimization to be estimated, so no analytic MSFE is presented, but simulation results show a lower MSFE than IC. However, the first three MSFEs are similar in the simulations up to 4 periods later, and all are substantially smaller than ignoring the break. Thus, there is little benefit in forecast error MSFE from knowing everything about the break other than its parameters relative to robust devices that make no such requirements and work well for other break forms, a conclusion supported by §7.2.1.

8 Implications for economic modelling, forecasting and policy

Unmodelled breaks have rather different impacts on the three objectives of empirical research, namely economic modelling, forecasting and policy. In particular, location shifts as we have focused on above, adversely affect empirical models when they are not taken into account. Parameter estimates become non-constant and can be far from the appropriate population value, see e.g., Hendry and Mizon (2010). Fortunately, such breaks can be handled when modelling in-sample: impulse-indicator saturation can detect and neutralize multiple breaks and outliers, reduce non-normality, and mitigate any spurious non-linearity that chances to align with breaks. Despite starting with large numbers of irrelevant candidate regressors, including more candidate variables than observations, *Autometrics* has a null retention frequency close to its nominal size, bias correction for selection is easily implemented, and greatly reduces the MSEs of any retained irrelevant variables. Then, provided no more location shifts occur, such a model will also work well for forecasting. Furthermore, if it satisfies super exogeneity of the conditioning variables the model will be appropriate for policy, having unbiased estimates of the relevant policy parameters. See Engle, Hendry and Richard (1983), Engle and Hendry (1993), and Hendry and Santos (2010).

1-step	
IC: $\mathsf{M}\left[\widetilde{\epsilon}_{T+1\vert T}\right]$	$\lambda^2 \exp\left(-2\psi\right)\left[1-\exp\left(-\psi\right)\right]^2 + 2\sigma_\epsilon^2$
DD: $\mathsf{M}\left[\overline{\epsilon}_{T+1\vert T}\right]$	$\lambda^2 \exp\left(-2\psi\right)\left[1-\exp\left(-\psi\right)\right]^2 + 2\sigma_\epsilon^2$
DGP: $\mathsf{M}\left[\widehat{\epsilon}_{T+1\vert T}\right]$	$\lambda^2 \exp\left(-2\psi\right)\left[1-\exp\left(-\psi\right)\right]^2 + 2\sigma_\epsilon^2$
No: $\mathsf{M}\left[\overleftarrow{\epsilon}_{T+1\vert T}\right]$	$\left(\lambda\left(1-\exp\left(-2\psi\right)\right) - \frac{1}{T}\lambda\left(1-\exp\left(-\psi\right)\right)\right)^2 + \sigma_\epsilon^2\left(1+\frac{1}{T}\right)$
2-step	
IC: $\mathsf{M}\left[\widetilde{\epsilon}_{T+2\vert T+1}\right]$	$\left(\lambda\exp\left(-\psi\right)\left(1+\exp\left(-\psi\right)-2\exp\left(-2\psi\right)\right)/2\right)^2 + 1.5\sigma_\epsilon^2$
DD: $\mathsf{M}\left[\overline{\epsilon}_{T+2\vert T+1}\right]$	$\left(\lambda\exp\left(-2\psi\right)\left(1-\exp\left(-\psi\right)\right)\right)^2 + 2\sigma_\epsilon^2$
DGP: $\mathsf{M}\left[\widehat{\epsilon}_{T+2\vert T+1}\right]$	NA
No: $\mathsf{M}\left[\overleftarrow{\epsilon}_{T+2\vert T+1}\right]$	$\left(\lambda\left(1-\exp\left(-3\psi\right)\right) - \frac{1}{T+1}\left[\lambda\left(1-\exp\left(-\psi\right)\right)+\lambda\left(1-\exp\left(-2\psi\right)\right)\right]\right)^2 + \sigma_\epsilon^2\left(1+\frac{2}{T+1}\right)$

Table 11.2: Internal break analytical MSFE comparisons.

Unfortunately, location shifts still happen intermittently close to the forecast origin and out of sample, and remain an ever-present issue for forecasting and policy. We have outlined some steps that might help predict breaks; if not, then to model and forecast during breaks; and if neither of those, then mitigate the impacts of breaks on forecasting by using a robust device. Hendry and Mizon (2005) discuss additional approaches with the potential to help.

Location shifts could, but need not, entail that policy parameters have altered. Conversely, policy parameters may have altered without a location shift having occurred, and may be difficult to detect till long after they have changed: see Hendry (2000) for example. However, most policy changes involve level shifts in variables like taxes, interest rates etc., which become location shifts in mis-specified models, and hence will reveal previous changes in policy parameters. Rescuing the forecasts from a model by a robustifying device, so there is no apparent forecast failure, does not imply that the resulting policy analysis is valid–nor does it vitiate that analysis. For example, the UK communications industries regulator and competition authority, Ofcom, had to forecast discounted net TV advertising revenues 40 quarterly ahead to calculate the fee for renewing ITV3's licence to broadcast advertising. The models and forecasts are reported in http://www.ofcom.org.uk/research/tv/reports/tvadvmarket.pdf. Structural change just before the forecast origin in 2003(2) was marked, with new non-terrestrial TV channels, bar-code based and internet advertising etc., affecting both the supply of and demand for TV advertising time. Forecasts from a differenced vector equilibrium-correction model, designed to offset any location shifts from such changes, revealed a downward trend over the decade ahead despite the then benign economic environment, by 'capturing' some of the breaks' effects.

Thus, the three different activities necessitate different evaluation criteria, and success or failure on any one does not entail success or failure on any of the others.

9 Conclusions

Whether breaks are predictable from relevant information available at the forecast origin remains unknown as yet. But we see progress in developing forecasting models; and methods of testing for and selecting such models. The theory of unpredictability suggests considering two information sets, namely regular forces and shift drivers, where one might model the latter as a non-linear ogive. Recent developments in high-frequency data collection, such as *Google Trends*, suggests relevant information may be available.

The approach proposed in this chapter builds on the existing literature using non-linear models to forecast regime shifts, such as smooth transition or Marov-switching models. However, the approach also considers whether breaks that have not occurred previously can be predicted by use of an additional information set that captures break drivers. Embedding the analysis in the context of available information allows for a broader set of models to be considered and we view this as the dominant approach if we are to have any success in forecasting breaks.

Commencing with a test of functional form with power against a range of non-linearities, the *Autometrics* approach can tackle data contamination, multiple breaks and non-normality by impulse-indicator saturation. It can also handle more variables than observations, and so is a feasible approach to non-linear modelling. *Autometrics* maintains the null rejection frequency at the desired level and can bias correct the final estimates for selection.

The analyses and simulations suggest that even if a break is not predicted, its course can be tracked quite well soon after its occurrence, albeit not much better than robust devices.

We conclude that it is feasible to extend automatic model selection algorithms for non-linear functions and location shifts as a starting basis for potentially modelling and eventually perhaps even forecasting breaks.

References

Aiolfi, M., Capistrán, C., and Timmermann, A. (2011). Forecast Combinations. In Clements, and Hendry (2011b), Ch. 8.

Angstenberger, J. (1996). Prediction of the S&P 500 index with neural networks. In Taylor, J. G. (ed.), *Neural Networks and Their Applications*, pp. 143–152. Chichester: John Wiley & Sons.

Baba, Y., Hendry, D. F., and Starr, R. M. (1992). The demand for M1 in the U.S.A., 1960–1988. *Review of Economic Studies*, **59**, 25–61.

Bacon, D. W., and Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, **58**, 525–534.

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.

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

Bec, F., Rahbek, A., and Shephard, N. (2008). The ACR model: A multivariate dynamic mixture autoregression. *Oxford Bulletin of Economics and Statistics*, **70**, 583–618.

Berg, A., and Pattillo, C. (1999). Predicting currency crises: The indicators approach and an alternative. *Journal of International Money and Finance*, **18**, 561–586.

Bontemps, C., and Mizon, G. E. (2008). Encompassing: Concepts and implementation. *Oxford Bulletin of Economics and Statistics*, **70**, 721–750.

Brooks, C. (1997). Linear and nonlinear (non-) forecastability of daily sterling exchange rates. *Journal of Forecasting*, **16**, 125–145.

Camba-Mendez, G., Kapetanios, G., Weale, M. R., and Smith, R. J. (2002). The forecasting performance of the OECD composite leading indicators for France, Germany, Italy and the UK. In Clements, and Hendry (2002a), pp. 386–408.

Cartwright, N. (1989). *Nature's Capacities and their Measurement*. Oxford: Clarendon Press.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2010a). 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. (2010b). Model selection when there are multiple breaks. *Journal of Econometrics*, forthcoming.

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., and Hendry, D. F. (2008). Forecasting UK inflation: Empirical evidence on robust forecasting devices. In Rapach, D. E., and Wohar, M. E. (eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty: Frontiers of Economics and Globalization Volume 3*, pp. 41–92. Bingley, UK: Emerald Group.

Castle, J. L., and Hendry, D. F. (2010). Automatic selection of non-linear models. In Wang, L., Garnier, H., and Jackman, T. (eds.), *System Identification, Environmental Modelling and Control*, forthcoming. New York: Springer.

Castle, J. L., and Hendry, D. F. (2010a). A low-dimension portmanteau test for non-linearity. *Journal of Econometrics*, **158**, 231–245.

Castle, J. L., and Hendry, D. F. (2010b). 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.

Chan, K. S., and Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, **7**, 179–194.

Choi, H., and Varian, H. (2009). Predicting the present with Google Trends. Unpublished paper, Economics Research Group, Google.

Clements, M. P., Franses, P. H., and Swanson, N. R. (2004). Forecasting economic and financial time series with non-linear models. *International Journal of Forecasting*, **20**, 169–183.

Clements, M. P., and Galvão, A. B. (2004). A comparison of tests of non-linear cointegration with an application to the predictability of US interest rates using the term structure. *International Journal of Forecasting*, **20**, 219–236.

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

Clements, M. P., and Hendry, D. F. (2001). An historical perspective on forecast errors. *National Institute Economic Review*, **177**, 100–112.

Clements, M. P., and Hendry, D. F. (eds.)(2002a). *A Companion to Economic Forecasting*. Oxford: Blackwells.

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

Clements, M. P., and Hendry, D. F. (2005). Guest editors' introduction: Information in economic forecasting. *Oxford Bulletin of Economics and Statistics*, **67**, 713–753.

Clements, M. P., and Hendry, D. F. (2006). Forecasting with breaks in data processes. 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. (2011a). Forecasting from Mis-specified Models in the Presence of Unanticipated Location Shifts. In Clements, and Hendry (2011b), Ch. 10.

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

Clements, M. P., and Krolzig, H.-M. (1998). A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP. *Econometrics Journal*, **1**, C47–75.

Collier, P., and Hoeffler, A. (2007). Civil war. Chapter 22, *Handbook of Defense Economics*, T. Sandler and K. Hartley (eds), Elsevier.

Dahl, C. M., and Hylleberg, S. (2004). Flexible regression models and relative forecast performance. *International Journal of Forecasting*, **20**, 201–217.

Demirgüç-Kunt, A., and Detragiache, E. (2000). Monitoring banking sector fragility: A multivariate logit approach with an application to the 1996-97 banking crises. *World Bank Economic Review*, **14**, 287–307.

Diebold, F. X., and Rudebusch, G. D. (1991). Forecasting output with the composite leading index: An ex ante analysis. *Journal of the American Statistical Association*, **86**, 603–610.

Doornik, J. A. (2007). Econometric model selection with more variables than observations. Working paper, Economics Department, University of Oxford.

Doornik, J. A. (2008). Encompassing and automatic model selection. *Oxford Bulletin of Economics and Statistics*, **70**, 915–925.

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

Doornik, J. A. (2009b). Improving the timeliness of data on influenza-like illnesses using Google Trends. Typescript, Department of Economics, University of Oxford.

Doornik, J. A. (2009c). *Object-Oriented Matrix Programming using Ox* 7th edn. London: Timberlake Consultants Press.

Eitrheim, Ø., and Teräsvirta, T. (1996). Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, **74**, 59–75.

Emerson, R. A., and Hendry, D. F. (1996). An evaluation of forecasting using leading indicators. *Journal of Forecasting*, **15**, 271–91.

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.

Engle, R. F., and Smith, A. D. (1999). Stochastic permanent breaks. *Review of Economics and Statistics*, **81**, 553–574.

Ericsson, N. R. (1999). Empirical modeling of money demand. *Empirical Economics*, **23**, 295–315.

Granger, C. W. J. (2005). Preface: Some thoughts on the future of forecasting. *Oxford Bulletin of Economics and Statistics*, **67**, 707–711.

Granger, C. W. J., and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, **45**, 39–70.

Hamilton, J. D. (1992). Estimation, inference and forecasting of time series subject to changes in regime. In Maddala, G. S., Rao, C. R., and Vinod, H. D. (eds.), *Handbook of Statistics*, Vol. 11. Amsterdam: North-Holland.

Hendry, D. F. (1986). The role of prediction in evaluating econometric models. In *Proceedings of the Royal Society*, Vol. A407, pp. 25–33.

Hendry, D. F. (2000). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics*, **11**, 45–65.

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

Hendry, D. F., and Ericsson, N. R. (1991). Modeling the demand for narrow money in the United Kingdom and the United States. *European Economic Review*, **35**, 833–886.

Hendry, D. F., and Hubrich, K. (2009). Combining disaggregate forecasts versus disaggregate information to forecast an aggregate. *Journal of Business and Economic Statistics*, DOI:10.1198/jbes.2009.07112.

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., and Mizon, G. E. (2005). Forecasting in the presence of structural breaks and policy regime shifts. In Andrews, D. W. K., and Stock, J. H. (eds.), *Identification and Inference for Econometric Models*, pp. 480–502. Cambridge: Cambridge University Press.

Hendry, D. F., and Mizon, G. E. (2009). Economic policy analysis in a rapidly changing world. Unpublished paper, Economics Department, Oxford.

Hendry, D. F., and Mizon, G. E. (2010). 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 Reade, J. J. (2008). Elusive return predictability: Discussion. *International Journal of Forecasting*, **24**, 22–28.

Hendry, D. F., and Reade, J. J. (2009). How should we make economic forecasts?. *VoxEU*, **11 June**, http://www.voxeu.org/index.php? q=node/3647.

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.

Holliday, J. R., Rundle, J. B., Tiampo, K. F., and Turcotte, D. L. (2006). Using earthquake intensities to forecast earthquake occurrence times. *Nonlinear Processes in Geophysics*, **13**, 585–593.

Hubert-Ferrari, A., Barka, A., Jacques, E., Nalbant, S. S., Meyer, B., Armijo, R., Tapponnier, P., and King, G. C. P. (2000). Seismic hazard in the Marmara Sea region following the 17 August 1999 Izmit earthquake. *Nature*, **404**, 269–273.

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.

Kaminsky, G., Lizondo, S., and Reinhart, C. (1998a). Currency and banking crises: the early warnings of distress. International Finance DP 629, Board of Governors of the Federal Reserve System, Washington, DC.

Kaminsky, G., Lizondo, S., and Reinhart, C. (1998b). Leading indicators of currency crises. IMF Staff Papers No 1, IMF, Washington, DC.

Kock, A. B., and Teräsvirta, T. (2010). Forecasting with nonlinear time series models. In Clements, M. P., and Hendry, D. F. (eds.), *Oxford Handbook of Economic Forecasting*. Oxford: OUP.

Komulainen, T., and Lukkarila, J. (2003). What drives financial crises in emerging markets?. DP 5, Institute for Economies in Transition, Bank of Finland.

Krolzig, H.-M. (1997). *Markov Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis*. Berlin: Springer Verlag.

Kumar, M., Moorthy, U., and Perraudin, W. (2003). Predicting emerging market currency crashes. *Journal of Empirical Finance*, **10**, 427–454.

Labhard, V., Kapetanios, G., and Price, S. (2007). Forecast combination and the Bank of England's suite of statistical forecasting models. Working Paper No. 323, Bank of England, London.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988). Testing linearity in univariate time series models. *Scandinavian Journal of Statistics*, **15**, 161–175.

Marcellino, M. (2004). Forecasting EMU macroeconomic variables. *International Journal of Forecasting*, **20**, 359–372.

Marcellino, M. (2006). Leading Indicators. In Elliot, G., Granger, C. W. J., and Timmermann, A. (eds.), *Handbook of Economic Forecasting*, Ch. 16. Amsterdam: Elsevier.

Obstfeld, M. (1996). Models of currency crises with self-fulfilling features. *European Economic Review*, **40**, 1037–1048.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. New York: Academic Press.

Quandt, R. E. (1983). Computational problems and methods. In *Computational Problems and Methods*, pp. 699–746. Amsterdam: North Holland.

Rech, G. (2002). Forecasting with artificial neural network models. SEE/EFI Working Paper Series in Economics and Finance 491, Stockholm School of Economics, Sweden.

Sensier, M., Artis, M., Osborn, D. R., and Birchenhall, C. (2004). Domestic and international influences on business cycle regimes in Europe. *International Journal of Forecasting*, **20**, 343–357.

Sparks, R. S. J. (2003). Forecasting volcanic eruptions. *Earth and Planetary Science Letters*, **210**, 1–15. Concerned mainly with Montserrat.

Stein, R., Barka, A., and Dieterich, J. (1997). Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophysical Journal International*, **128**, 594–604.

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. (1999). A comparison of linear and nonlinear models for forecasting macroeconomic time series. In Engle, R. F., and White, H. (eds.), *Cointegration, Causality and Forecasting*, pp. 1–44. Oxford: Oxford University Press.

Swanson, N. R., and White, H. (1995). A model selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business and Economic Statistics*, **13**, 265–275.

Swanson, N. R., and White, H. (1997). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics*, **79**, 540–550.

Teräsvirta, T. (1994). Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, **89**, 208–218.

Teräsvirta, T. (1998). Modelling economic relationships with smooth transition regressions. In Ullah, A., and Giles, D. E. A. (eds.), *Handbook of Applied Economic Statistics*, pp. 507–552. New York: Marcel Dekker.

Teräsvirta, T., van Dijk, D., and Medeiros, M. C. (2005). Smooth transition autoregressions, neural networks, and linear models in forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting*, **21**, 755–774.

Tiao, G. C., and Tsay, R. S. (1994). Some advances in non-linear and adaptive modelling in time-series. *Journal of Forecasting*, **13**, 109–131.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, **14**, 35–62.