ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

FORECASTING FROM MIS-SPECIFIED MODELS IN THE PRESENCE OF UNANTICIPATED LOCATION SHIFTS

Michael P. Clements and David F. Hendry

Number 484
May 2010

Manor Road Building, Oxford OX1 3UQ

Forecasting from Mis-specified Models in the Presence of Unanticipated Location Shifts

Michael P. Clements
Department of Economics,
University of Warwick,
Coventry CV4 7AL.
m.p.clements@warwick.ac.uk

David F. Hendry*
Economics Department,
Oxford University
david.hendry@nuffield.ox.ac.uk

May 3, 2010

Abstract

This chapter describes the issues confronting any realistic context for economic forecasting, which is inevitably based on unknowingly mis-specified models, usually estimated from mis-measured data, facing intermittent and often unanticipated location shifts. We focus on mitigating the systematic forecast failures that result in such settings, and describe the background to our approach, the difficulties of evaluating forecasts, and the devices that are more robust when change occurs.

Contents

*We are grateful to Jennifer L. Castle and Kirstin Hubrich for helpful comments on an earlier draft. Prepared for the *Oxford Handbook of Economic Forecasting*.

JEL classifications: C51, C22.
KEYWORDS: Economic Forecasting; Location Shifts; Mis-specified Models; Robust Forecasts.

1 Introduction

In this chapter, we review the recent literature on forecasting when there are unanticipated structural breaks, or parameter non-constancies, and the forecasting model is mis-specified in unknown ways. We build on previous research into forecasting in the face of structural breaks, including our own recent contributions Clements and Hendry (2006, 2008), as well as Clements and Hendry (1998, 1999, 2002a). We consider some of the key material in earlier reviews, as well as recent developments, in order that the chapter can be read as a stand-alone survey of the field.

The key issue we address is why forecasts are sometimes inaccurate relative either to what might have been expected based on the past performance of the particular model in question, or relative to the forecasts produced by an alternative model or method. Once we understand what causes poor forecast performance, the possibility of improved forecasting models and methods arises. The central message is that unanticipated location shifts are the key culprit. On the positive side, some models are less susceptible to such breaks than others, at least after the break has occurred. The issues of forecasting breaks and during breaks are considered in Castle, Fawcett and Hendry (2010), and Castle, Fawcett and Hendry (this volume).

As argued in Clements and Hendry (2006), whether or not a structural break occurs is as much a feature of the model under consideration as of the process generating the data. The example they present is the structural change data generation process (DGP) of (e.g.,) Andrews (1993):

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

where $\varepsilon_t \sim \mathsf{IID}\left[0, \sigma_\varepsilon^2\right]$ (that is, *I*ndependently, *I*dentically *D*istributed, mean zero, variance σ_ε^2), and s_t is the indicator variable, $s_t \equiv 1_{(t>\tau)}$ which equals 1 when $t > \tau$ and zero when $t \leq \tau$. A constant-parameter p^{th}-order autoregression (AR(p)) for y_t of the form:

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

would experience a structural break when the parameter vector shifts. Let $\phi = (\mu_0\ \alpha_1 \ldots \alpha_p)'$, $\phi^* = \left(\mu_0^*\ \alpha_1^* \ldots \alpha_p^*\right)'$ and $\phi_1 = (\mu_{0,1}\ \alpha_{1,1} \ldots \alpha_{p,1})'$. Then the AR($p$) model parameters are $\phi_1 = \phi$ for $t \leq \tau$, but $\phi_1 = \phi + \phi^*$ for $t > \tau$. But if the forecasting model incorporated the terms which interact the existing regressors with a step dummy D_t defined by $D_t = s_t = 1_{(t>\tau)}$, then the extended model (letting $\mathbf{x}_t = (1\ y_{t-1} \ldots y_{t-p})'$):

$$y_t = \phi_{1,d}' \mathbf{x}_t + \phi_{2,d}' \mathbf{x}_t D_t + v_{t,d} \tag{3}$$

does not suffer a break as $(\phi_{1,d}'\ \phi_{2,d}') = (\phi'\ \phi^{*\prime})$ for all $t = 1, \ldots, T$, matching the data generating process, DGP (see e.g., Hendry, 1996).

Breaks may be precipitated by institutional, political, social, financial, legal, and technological change, and will generally be more comprehensible retrospectively than at the time of occurrence. In practice there will always be events which are essentially unknowable *ex ante*, which will then cause breaks in the parameters of a forecasting model. We typically assume that the modeller does not have knowledge of the process determining the break. In terms of the above example, the model is given by (2), whereas the DGP–which provides a complete description of the process up to a random disturbance term–is given by (3). Hence the forecasting models we consider will almost always be mis-specified for the DGP, and some will also be mis-specified for the 'local' DGP (LDGP, which is the derived DGP in the space of the variables under analysis: see Bontemps and Mizon, 2008). Importantly, breaks must change some aspects of distributions relative to their previous forms, so are intrinsically non-stationary. Consequently, the precise timing of every expectation is crucial to understanding the evolving means, variances etc., after a break. Strictly, expectations operators (denoted $\mathsf{E}[\cdot]$ below) should have time subscripts to denote the distribution averaged over as well as any conditioning information (e.g., $\mathsf{E}_{T+1}[y_{T+1}|z_T]$ say), but unless otherwise stated, will refer below to the variable whose expectation is being formed (y_{T+1} in the example).

Whilst we consider forecasting in the presence of breaks which are unpredictable and of unknown origin, the chapter by Castle, Fawcett and Hendry allows that there may be some advance information about a break–for example, early estimates of some of the components of an aggregate may be available which carry information about the break–and they discuss the problems that arise from such early estimates being less reliable measures than later estimates. The chapter by Reichlin discusses nowcasting.

The plan of our chapter is as follows. In section 2, we briefly review approaches to assessing forecast accuracy. This provides useful background material, and explains why we focus on the impact of breaks on the expected squared forecast error, and therefore frame much of the discussion of accuracy in terms of forecast bias and forecast-error variance. Section 3 describes a general framework for analyzing various aspects of the forecasting problem that may contribute to forecast error: at the most basic level, these are that the model may not capture important features of the DGP; the DGP may itself be non-constant; and the form and unknown coefficients of a forecasting model will need to be specified, selected and estimated from the available data. The key findings hold for general classes of models, and not just for specific DGPs and forecasting models. For example, any model which exhibits equilibrium-correcting behaviour will generate sequences of systematically biased forecasts consequent upon a shift in the equilibrium mean of the DGP. Section 4 focuses on the forecasting models that have been popular in time-series econometric modelling in the last quarter of a century, and fleshes out the possible sources of forecast error when there are various forms of structural break. We also consider forecasting devices that are likely to be less adversely affected by breaks.

The next two sections allow that macro-variables are usually aggregates (section 5), but may be available at a higher frequency (section 6), so we consider whether the effects of breaks on forecasts can be alleviated by forecasting the disaggregated components or by time disaggregation. Section 7 reviews recent empirical research on forecasting that averages over many different forecasts to try and obtain more accurate forecasts in the presence of general forms of instability and model mis-specification. Averaging takes place across forecasts from (i) different forecasting models and forecasting methods; (ii) different specifications of essentially the same forecasting model; and (iii) models estimated on different estimation samples, including recursive and rolling forecasting schemes. It is not always clear why such averaging might work beyond the usual arguments advanced in favour of forecast combination, which do not tend to stress instability, so we also revisit Hendry and Clements (2004) who analyze pooling and instability, albeit in simple models. We also review forecasting with factor models when there are instabilities. Finally, section 8 concludes the substance of this review with an analysis based on locations shifts coming from data revisions. A key reference is Patterson (2003), who explicitly considers the data measurement process along with the DGP, and shows that equilibrium-mean shifts and forecast failure can result from using different data vintages to estimate models and generate forecasts. Given the recent interest in real-time data analysis and forecasting, we include a review of this literature in so far as it relates to breaks: Croushore (this volume) provides a review of real-time forecasting. Finally, section 9 offers some concluding remarks.

2 Assessing forecast performance

2.1 Specific loss functions

In principle, forecasts should be evaluated in terms of their 'value'. Suppose, following Diebold, Gunther and Tay (1998), a user of a forecast has a loss function $L_1(a_t, y_{t+1})$, where a_t refers to a specific action, and $y_{t+1} \sim \mathsf{f}(y_{t+1})$ is the future value of the random variable drawn from the distribution $\mathsf{f}(\cdot)$. The optimal $a^*_{t,1}$ is chosen to minimize expected loss *given* the forecast density $\mathsf{p}_{t,1}(y_{t+1})$, so:

$$a^*_{t,1} = \underset{a_{t,1}\in A}{\arg\min} \int L_1(a_{t,1}, y_{t+1})\, \mathsf{p}_{t,1}(y_{t+1})\, \mathrm{d}y_{t+1}. \quad (4)$$

The user's expected loss is given by:

$$\mathsf{E}\left[L_1\left(a^*_{t,1}, y_{t+1}\right)\right] = \int L_1\left(a^*_{t,1}, y_{t+1}\right) \mathsf{f}(y_{t+1})\, \mathrm{d}y_{t+1}. \quad (5)$$

Given any loss function, $L_1(\cdot)$, the forecast $\mathsf{p}_{t,1}(y_{t+1})$ will be preferred to $\mathsf{p}_{t,2}(y_{t+1})$ if:

$$\mathsf{E}\left[L_1\left(a^*_{t,1}(\mathsf{p}_{t,1}(y_{t+1})), y_{t+1}\right)\right] < \mathsf{E}\left[L_1\left(a^*_{t,2}(\mathsf{p}_{t,2}(y_{t+1})), y_{t+1}\right)\right] \quad (6)$$

where the notation $a^*_{t,i}(\mathsf{p}_{t,i}(y))$ makes it plain that $a^*_{t,i}$ is the optimal action for forecast $\mathsf{p}_{t,i}(y)$.

Although attractive in principle, relatively common in meteorology (see, e.g., Katz and Murphy, 1997, and Katz and Lazo, this volume), and occurs in empirical finance (see, e.g., Leitch and Tanner, 1991, 1995), there has been little decision-based forecast evaluation in macroeconomics. The reasons for this are readily apparent when considering the assumptions (which include a constant distribution $\mathsf{f}(\cdot)$ over time) and the informational requirements that underpin the analysis in (6). The latter include knowledge of: the disutility emanating from actions taken today (a_t) in future states of nature (y_{t+1}); and the whole predictive distribution (as opposed to simply a point forecast of y_{t+1}). Moreover, the ranking between any two forecasts $\mathsf{p}_{t,1}(y_{t+1})$ and $\mathsf{p}_{t,2}(y_{t+1})$ usually depends on $L_1(.)$, so relative comparisons of forecast performance are ruled out unless the specific form of the loss function is agreed. Diebold *et al.* (1998) show that $\mathsf{f}(y_{t+1})$ will dominate for all loss functions, but this is far removed from the situation that confronts the practical forecaster. In the absence of well-defined mappings between forecast errors and their costs, macroeconomic forecasting has relied almost exclusively on general-purpose loss functions such as squared error loss, or absolute loss, or on relatively simple ways of allowing under-prediction and over-prediction of the same magnitude to be penalized at different rates (e.g., the 'linex' loss function of Varian, 1975).

2.2 Limitations of MSFE-based measures

Mean-squared forecast error (MSFE) based measures remain the dominant criteria for assessing accuracy.[1] For a vector of variables, the accuracy assessments are then based on the MSFE matrix:

$$\mathbf{V}_h \equiv \mathsf{E}\left[\mathbf{e}_{T+h}\mathbf{e}'_{T+h}\right] = \mathsf{V}\left[\mathbf{e}_{T+h}\right] + \mathsf{E}\left[\mathbf{e}_{T+h}\right]\mathsf{E}\left[\mathbf{e}'_{T+h}\right] \quad (7)$$

where $\mathbf{e}_{T+h}$ is a vector of h-step ahead forecast errors. Clements and Hendry (1993) show that this approach is not without its problems, because such measures may lack invariance to non-singular, scale-preserving, linear transformations for which the associated model class is invariant. Consequently, MSFE comparisons

[1] This section is based on Clements and Hendry (1993).

may result in inconsistent rankings of the forecast performance of different models for multi-step ahead horizons across different transformations of the variables (e.g., levels or differences). This is true whether we consider summary scalar measures of $\mathbf{V}_h$, such as the trace or the determinant, or the whole matrix $\mathbf{V}_h$ (e.g., whether the MSFE matrix differs from that of a rival model by a matrix that is positive definite).

The problems that arise can be illustrated with the following linear forecasting system:

$$\mathbf{x}_t = \tau + \mathbf{\Gamma x}_{t-1} + \nu_t \text{ where } \nu_t \sim \mathsf{IN}_n[\mathbf{0}, \mathbf{\Omega}_\nu], \tag{8}$$

which is similar to the models we consider in section 4. Here, $\mathbf{x}_t$ is an n-dimensional vector of variables, ν_t is an independent normal error with expectation $\mathsf{E}[\nu_t] = \mathbf{0}$ and variance matrix $\mathsf{V}[\nu_t] = \mathbf{\Omega}_\nu$. Equation (8) can be written in stacked form as:

$$\mathbf{\Phi s}_t = \nu_t \tag{9}$$

where $\mathbf{s}_t' = \left(\mathbf{x}_t', 1, \mathbf{x}_{t-1}'\right)$, and $\mathbf{\Phi} = (\mathbf{I}_n : -\tau : -\mathbf{\Gamma})$. Then the likelihood and generalized variance of the system in (9) are invariant under scale-preserving, non-singular transformations of the form:

$$\mathbf{M\Phi P}^{-1}\mathbf{Ps}_t = \mathbf{M}\nu_t$$

so:

$$\mathbf{\Phi}^*\mathbf{s}_t^* = \nu_t^* \quad \text{with} \quad \nu_t^* \sim \mathsf{IN}_n\left[\mathbf{0}, \mathbf{M\Omega}_\nu\mathbf{M}'\right]. \tag{10}$$

In (10), $\mathbf{s}_t^* = \mathbf{Ps}_t$, $\mathbf{M}$ and $\mathbf{P}$ are respectively $n \times n$ and $(2n+1) \times (2n+1)$ known non-singular matrices, where $|\mathbf{M}| = 1$, and $\mathbf{P}$ is the upper block-triangular matrix:

$$\mathbf{P} = \begin{pmatrix} \mathbf{I}_n & \mathbf{P}_{12} \\ \mathbf{0} & \mathbf{P}_{22} \end{pmatrix},$$

with $|\mathbf{P}_{22}| \neq 0$. Then:

$$\left|\mathbf{M\Omega}_\nu\mathbf{M}'\right| = |\mathbf{\Omega}_\nu|. \tag{11}$$

This means that forecasts and prediction intervals made in the original system (9) and then transformed after the event to $\mathbf{x}_t^*$ are identical to those made directly using the transformed system (10). This remains true when parameters first have to be estimated, provided the estimator is equivariant (e.g., maximum likelihood). For example, if a system is estimated for $\mathbf{x}_t$ on $\mathbf{x}_{t-1}$ by full-information maximum likelihood with $\widehat{\Delta\mathbf{x}}_t$ obtained by an identity, then the forecasts $\widehat{\Delta\mathbf{x}}_{T+h}$ of $\Delta\mathbf{x}_{T+h}$ are identical to those obtained from modelling $\Delta\mathbf{x}_t$ on $\mathbf{x}_{t-1}$ with $\widehat{\mathbf{x}}_t$ obtained by identity.

Consider now the properties of measures based on $\mathbf{V}_h$ for transformations $\mathbf{M}$ and $\mathbf{P}$. For transformations involving $\mathbf{M}$ only (so that $\mathbf{P} = \mathbf{I}_{2n+1}$), the matrix measure $\mathbf{V}_h$ and the determinant are invariant, but the trace is not: see e.g., Granger and Newbold (1986). For transformations using $\mathbf{P}$, and with $\mathbf{M} = \mathbf{I}_n$, neither the determinant nor the MSFE matrix are invariant for $h > 1$.

Invariance to $\mathbf{P}$ transformations can be achieved by accounting for covariances between different step-ahead errors. This gives the generalized forecast-error second-moment matrix (GFESM, which is close to predictive likelihood: see Bjørnstad, 1990) defined as:

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

where $\mathbf{E}_h$ stacks the forecast errors up to and including h-steps ahead:

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

Then, $|\boldsymbol{\Phi}_h|$ is also unaffected by $\mathbf{M}$ transforms, since denoting the vector of stacked forecast errors from the transformed model by $\widetilde{\mathbf{E}}'_h$:

$$\widetilde{\mathbf{E}}'_h = \left[\widehat{\nu}'_{T+1|T}\mathbf{M}', \widehat{\nu}'_{T+2|T}\mathbf{M}', \ldots, \widehat{\nu}'_{T+h-1|T}\mathbf{M}', \widehat{\nu}'_{T+h|T}\mathbf{M}'\right],$$

we have:

$$\left|\widetilde{\boldsymbol{\Phi}}_h\right| = \left|\mathsf{E}\left[\widetilde{\mathbf{E}}_h\widetilde{\mathbf{E}}'_h\right]\right| = \left|\mathsf{E}\left[\mathbf{E}_h\mathbf{E}'_h\right]\right|,$$

since $|\mathbf{I}_n \otimes \mathbf{M}| = 1$. The property of invariance provides a unique measure for a given model independently of its representation. However, beyond this, the GFESM is not especially compelling, and there may be computational issues, especially when there are only relatively small-samples of forecasts, as it requires the estimation of a square matrix of order nh. For further discussion, see Newbold, Harvey and Leybourne (1999). Consequently, we follow the literature in using simpler accuracy measures.

2.3 Relative accuracy

The forecasts of a model of interest can be assessed relative to the forecasts of a rival model or a benchmark model, either formally as described by Clark and McCracken (this volume) and Giacomini (this volume), or informally by simply calculating the ratio of the MSFEs or RMSFEs. Assessments of accuracy of this sort are useful in determining the relative accuracy of two or more models when confronted by events, such as structural breaks, to which the forecasts are robust to different degrees.

Rather than assessing forecast performance relative to that of a rival, the accuracy of the forecasts can also be compared relative to what would have been expected based on the past fit of the model to the data. This idea underlies tests of predictive accuracy which compare an estimate of the forecast-error variance obtained from the past residuals with the actual MSFE: see, *inter alia*, Chow (1960), Christ (1966) and Hendry (1974, 1979) for early developments, and the recent contributions by Clements and Hendry (2002b) and Giacomini and Rossi (2009), who define a forecast breakdown as occurring when a model has a significantly worse out-of-sample performance than in-sample, judged by some loss function.

3 Forecast-error taxonomies

Before considering specific DGPs, local DGPs (LDGPs), and forecasting models such as vector autoregressions (VARs), we describe the forecast-error taxonomy in Clements and Hendry (2006). They start from the premise that models are mis-specified for DGPs which are subject to breaks, and that the models' parameters typically have to be estimated from the data. From this, they deduce the possible sources of forecast error that might occur. Their formulation is 'model free', in the sense that particular specifications of relationships are unnecessary. They obtain a taxonomy consisting of seven main sources of forecast error, and partition those by whether or not the source of error has an expectation that has a mean of zero over the relevant forecast horizon. Thus, as explained below, the taxonomy distinguishes between breaks affecting 'deterministic' and 'stochastic' variables, both in-sample and out-of-sample, as well as delineating the other possible sources of forecast error, including model mis-specification and parameter-estimation uncertainty, all of which might interact with breaks.

The interested reader is referred to Clements and Hendry (2006, p.611) for a formal development, while for setting the scene for the remainder of this chapter, a verbal description of the main sources will suffice. Firstly, consider structural changes to the deterministic components of the LDGP, specifically, location shifts. These include changes in intercepts or the parameters of linear trends, etc., and depending on the setup, can often be mapped into changes in equilibrium means and long-run growth rates (as in the following section). These will change the average future values of the outcomes, relative to the unchanged forecasts of the model, and so will induce forecast bias. Here, we refer exclusively to forecast-period location shifts, and assume that all in-sample shifts have been correctly modelled. When this is the case, a non-zero mean forecast error implies that a location shift has occurred. Hence location shifts play a crucial role in systematic forecast failure.

The other type of break is a 'stochastic break'. A simple example would be a shift in the autoregressive parameter of a first-order autoregression. Such breaks may or may not cause forecast bias. For example, if the process is zero-mean, then the shift in the autoregressive parameter will not induce bias, although higher moments will be affected. As shown in (e.g.) Hendry (2000), stochastic breaks can be hard to detect, and need not have any noticeable effects on the accuracy of forecasts, although they could lead to important distortions in economic policy if incorrect partial derivatives of policy variables were used following a stochastic break.

The first two sources of error reflect developments over the forecast period, with the model matching the LDGP in-sample. The third source of forecast error arises from mis-specification of the deterministic factors in the model relative to the LDGP in-sample. This form of mis-specification may arise from unmodelled in-sample shifts, possibly resulting from 'conventional' mis-specification such as the omission of a relevant variable that undergoes a mean shift. When all the in-sample deterministic terms, including all shifts in the LDGP, are correctly specified, then this component is zero. Mis-specification of in-sample deterministic factors ought to be detectable by rigorous testing (see Bai and Perron, 1998), or by the recent approach of impulse-indicator saturation (see Hendry, Johansen and Santos, 2008, and Johansen and Nielsen, 2009, summarized in Castle, Hendry and Fawcwett (this volume). Failure to correct for this potential source of error will cause systematic bias. The fourth source arises from mis-specification of the stochastic components, such as selecting the wrong lag length. In the example of the first-order autoregression, an AR(1) forecasting model would be stochastically mis-specified if the LDGP were an AR(2). Some forms of stochastic mis-specification will not cause forecast bias. This would be the case for an incorrect choice of lag for a zero-mean process, and generalizes to multivariate models.

The fifth forecast-error component results from data measurement errors, especially forecast-origin inaccuracy. When the data are accurate (especially important at the forecast origin), then that source is zero, but the converse is not entailed, as this term can be zero just because the data are mean zero. Related issues are discussed in section 8. The next source arises when unknown model parameters are replaced by estimates derived from the data. Biases in estimation could, but need not, induce a systematic forecast error. The main impact of estimation uncertainty will be on the forecast-error variance.

The last term is the unavoidable and unpredictable LDGP innovation error, which is the smallest achievable in the class given the information used. This final term is unlikely ever to be zero in any social science, although it will often have a zero mean by construction in-sample, and will usually be unpredictable from the past of the information in use. As with estimation uncertainty, the main practical impact is through forecast-error variances.

It is worth emphasizing a number of aspects of the discussion in this section. First, the key distinction is between whether the expectation of any particular component source is zero or non-zero. Systematic forecast bias will not result in the former case, and these sources of forecast error will affect higher moments of the forecast-error distribution, such as forecast-error variances. Conversely, if a non-zero mean error results from any source, systematic forecast errors will ensue. Secondly, the breaks which are most harmful from a forecasting perspective–those that induce non-zero mean errors–are ones which will be most easily detected.

This is true of in-sample deterministic mis-specifications, and is equally true after the event for forecast-period breaks. 'In-sample forecasts' will be poor immediately after the change has occurred and will be detectable using standard tests. Stochastic breaks or mis-specifications have little impact on mean forecast errors and will necessarily be harder to detect. Thirdly, the taxonomy suggests that to the extent that it is possible to 'design' models that have no deterministic terms, then systematic forecast errors emanating from changes in, or in-sample mis-specification of, deterministic components can be avoided, an insight that is developed in sections 4.5.2 and 4.5.3 below to yield robust forecasting devices. The taxonomy applies to any model form. Below we consider its application to the 'workhorse' model of macroeconomic forecasting.

4 Econometric models versus robust forecasting models

Acknowledging that many macroeconomic time series are integrated in their (log) levels, the standard paradigm in modern econometrics is to build models in the first differences of the series, whilst incorporating the long-run information, or equilibrium relationships, that may exist by virtue of linear combinations of the series being integrated of order zero ($\mathsf{I}(0)$). We believe that this is a realistic description of the types of forecasting models that have emerged over the last quarter of a century. However, economic processes will also typically be perturbed by periodic non-constancies, for the reasons discussed in the introduction. When breaks near the forecast origin take the form of location shifts, they have a pernicious effect on forecasts, often inducing systematic large errors, which we refer to as forecast failure. More generally, any models which have well-defined long-run or equilibrium solutions will be susceptible to that problem: as shown by. e.g., Clements and Hendry (2006), this is true not only of most widely-used models of the conditional mean, but also of popular volatility models, such as the autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982), and its generalizations (see *inter alia*, Engle and Bollerslev, 1987, Bollerslev, Chou and Kroner, 1992, and Shephard, 1996; and Bera and Higgins, 1993, and Baillie and Bollerslev, 1992 on forecasting).

We assume that the economy can be reasonably well approximated by vector equilibrium-correction models (with acronym VEqCMs), so variables can be modelled as being driven by a reduced number of common stochastic trends (see, e.g., Nelson and Plosser, 1982, Engle and Granger, 1987, and Johansen, 1988). We assume that breaks and non-constancies that have occurred sufficiently far in the past will have been detected and 'modelled', if only by the inclusion of dummy variables. For this reason, we concentrate on breaks at or around the time the forecasts are made, in order to sharpen the analysis. We recognize that gradual drift in parameters, or more evolutionary changes, may not fit well with these assumptions, and consider those in section 7.

Forecasting in cointegrated vector autoregressions (VARs) in the absence of breaks has been discussed by Engle and Yoo (1987), Clements and Hendry (1995), Lin and Tsay (1996) and Christoffersen and Diebold (1998), *inter alia*. Yet however well a model performs in-sample, nothing can preclude a location shift at, or after, the forecast origin, and that is the focus of this section. We consider the effects of such breaks on the properties of forecasts from models that incorporate equilibrium relations, namely, VEqCMs, and also consider a related forecasting model which is largely robust to the class of breaks that are shown to be particularly detrimental to VEqCMs. The analysis draws on Clements and Hendry (1996, 1998, 1999), and considers the 'differenced VEqCM' of Hendry (2006) as a way of robustifying the forecasts from the VEqCM class when there are structural breaks. The forecast-error taxonomy discussed in section 3 suggested that location shifts were likely to be the main determinant of systematic forecast failure in practice, and for that reason, we first consider this type of break. We then allow for more general breaks, and assess the likely impacts of these on the different forecasting models. First, section 4.1 describes the DGP and introduces the notation.

4.1 Data generating process

We write the in-sample DGP as an augmented first-order VAR for the n variables of interest $\mathbf{x}_t$, integrated of order one (denoted $\mathsf{I}(1)$, possibly after taking logs), with k stationary variables $\mathbf{z}_t$ (which are unknowingly omitted from all models) as:

$$\mathbf{x}_t = \eta_0 + \mathbf{\Gamma}_0 \mathbf{x}_{t-1} + \mathbf{\Upsilon}_0 \mathbf{z}_t + \nu_t \tag{12}$$

where $\nu_t \sim \mathsf{IN}_n [\mathbf{0}, \mathbf{\Omega}_\nu]$, denoting an independent, normally distributed innovation error with variance matrix $\mathbf{\Omega}_\nu$. Also, $\mathbf{\Gamma}_0$ and $\mathbf{\Upsilon}_0$ are $n \times n$ and $n \times k$ matrices of coefficients, and η_0 is an n dimensional vector of intercepts, where the zero subscript denotes a population parameter value. The specification in (12) is assumed constant in-sample over $t = 1, \ldots, T$, and the system is assumed to satisfy the $r < n$ cointegration relations:

$$\mathbf{\Gamma}_0 = \mathbf{I}_n + \alpha_0 \beta_0' \tag{13}$$

In (13), α_0 and β_0 are $n \times r$ full-rank matrices, no roots of $|\mathbf{I} - \mathbf{\Gamma}_0 L| = 0$ lie inside the unit circle (where $L^s \mathbf{x}_t = \mathbf{x}_{t-s}$), and $\alpha_{0,\perp}' \mathbf{\Gamma}_0 \beta_{0,\perp}$ is full rank $(n-r)$, where $\alpha_{0,\perp}$ and $\beta_{0,\perp}$ are full column rank $n \times (n-r)$ matrices, with $\alpha_0' \alpha_{0,\perp} = \beta_0' \beta_{0,\perp} = \mathbf{0}$ (see e.g., Johansen, 1992). Additional lags do not materially affect the analysis so are omitted for simplicity. Let:

$$\mathbf{\Gamma}_0 \alpha_0 = \left(\mathbf{I}_n + \alpha_0 \beta_0'\right) \alpha_0 = \alpha_0 \left(\mathbf{I}_n + \beta_0' \alpha_0\right) = \alpha_0 \mathbf{\Psi}_0. \tag{14}$$

In a cointegrated system, $\mathbf{\Psi}_0$ corresponds to the eigenvalues of $\mathbf{\Gamma}_0$ which are strictly less than unity in absolute value, so $\mathbf{\Psi}_0^h \to \mathbf{0}$ as $h \to \infty$.

Using (13), (12) is reparametrized as the $\mathsf{I}(0)$ VEqCM:

$$\Delta \mathbf{x}_t = \tau_0 + \alpha_0 \beta_0' \mathbf{x}_{t-1} + \mathbf{\Upsilon}_0 \left(\mathbf{z}_t - \psi_0\right) + \nu_t. \tag{15}$$

In (15), all of $\Delta \mathbf{x}_t$, $\beta_0' \mathbf{x}_t$ and $\mathbf{z}_t$ are $\mathsf{I}(0)$, but may have non-zero means. However, $\mathsf{E}[\Delta \beta_0' \mathbf{x}_t] = \mathsf{E}[\beta_0' \Delta \mathbf{x}_t] = \mathbf{0}$, as it is the average growth of an $\mathsf{I}(0)$ variable. Since $\mathsf{E}[\mathbf{z}_t - \psi_0] = \mathbf{0}$:

$$\tau_0 = \gamma_0 - \alpha_0 \mu_0 \tag{16}$$

then (15) can be rewritten as:

$$\Delta \mathbf{x}_t = \gamma_0 + \alpha_0 \left(\beta_0' \mathbf{x}_{t-1} - \mu_0\right) + \mathbf{\Upsilon}_0 \left(\mathbf{z}_t - \psi_0\right) + \nu_t \tag{17}$$

where $\mathsf{E}[\Delta \mathbf{x}_t] = \gamma_0$. Although (17) characterizes the in-sample DGP, the investigator is only modelling $\mathbf{x}_t$, unaware that $\mathbf{z}_t$ are relevant. Moreover, although (17) is constant in-sample, we allow for location shifts at (or near) the forecast origin at time T. Again for simplicity, we assume that $(\beta_0' \mathbf{x}_{t-1} - \mu_0)$ and $(\mathbf{z}_t - \psi_0)$ are orthogonal, although nothing material depends on that condition.

4.2 Forecasting models

The model postulated by the forecaster is first taken to be the in-sample VEqCM:

$$\Delta \mathbf{x}_t = \tau + \alpha \beta' \mathbf{x}_{t-1} + \epsilon_t \tag{18}$$

where $\epsilon_t \sim \mathsf{IN}_n[\mathbf{0}, \mathbf{\Omega}_\epsilon]$, and this is estimated from data (possibly mis-measured) up to time T as:

$$\widehat{\Delta \mathbf{x}}_T = \widehat{\tau}_{(T)} + \widehat{\alpha}_{(T)} \widehat{\beta}'_{(T)} \mathbf{x}_{T-1} \tag{19}$$

where subscripts in parentheses show the estimation sample. Then a sequence of 1-step forecasts is generated by:

$$\widehat{\Delta \mathbf{x}}_{T+h+1|T+h} = \widehat{\tau}_{(T)} + \widehat{\alpha}_{(T)} \widehat{\beta}'_{(T)} \mathbf{x}_{T+h} \tag{20}$$

for $h = 1, 2, \ldots$, with the possibility that parameter estimates are updated as the forecast origin increases:

$$\widetilde{\Delta \mathbf{x}}_{T+h+1|T+h} = \widehat{\tau}_{(T+h)} + \widehat{\alpha}_{(T+h)} \widehat{\beta}'_{(T+h)} \mathbf{x}_{T+h} \tag{21}$$

A multi-step forecast might be used, as noted below.

4.3 The impact of location shifts on VEqCMs

Within this framework, Clements and Hendry (1998, 1999, 2006), and the discussion in section 3, all point to a shift in the vector of equilibrium means, μ, denoted $\nabla \mu_0^* = \mu_0^* - \mu_0$, being the most pernicious change from a forecasting perspective. Here, μ_0^* denotes the post-break equilibrium mean. Although γ_0, α_0 and $\mathbf{\Omega}_\nu$ could alter as well in practice, reasonable changes to these rarely entail the same magnitude of forecast failure: see Hendry (2000). Shifts in γ_0, the underlying growth rate, could also be included, and this is discussed below.

Following a change to μ_0^* at the forecast origin (time T) the DGP becomes:

$$\Delta \mathbf{x}_{T+1} = \gamma_0 + \alpha_0 \left(\beta_0' \mathbf{x}_T - \mu_0^* \right) + \mathbf{\Upsilon}_0 \left(\mathbf{z}_{T+1} - \psi_0 \right) + \nu_{T+1} \tag{22}$$

so adding and subtracting $\alpha_0 \mu_0$ in (22):

$$\Delta \mathbf{x}_{T+1} = \gamma_0 + \alpha_0 \left(\beta_0' \mathbf{x}_T - \mu_0 \right) - \alpha_0 \nabla \mu_0^* + \mathbf{\Upsilon}_0 \left(\mathbf{z}_{T+1} - \psi_0 \right) + \nu_{T+1} \tag{23}$$

which isolates the parameter non-constancy in the term $-\alpha_0 \nabla \mu_0^*$ in (23). To establish the 'pure' effect of a location shift, we temporarily abstract from all other complications, including estimation, data measurement errors etc., so the constant-parameter forecast of $\Delta \mathbf{x}_{T+1}$ is assumed here to be given by (24) rather than (20):

$$\widehat{\Delta \mathbf{x}}_{T+1|T} = \gamma_0 + \alpha_0 \left(\beta_0' \mathbf{x}_T - \mu_0 \right). \tag{24}$$

Thus, the forecast uses known in-sample population parameters (estimation uncertainty is re-introduced below) and the model is correctly specified pre-break with $\mathbf{\Upsilon}_0 = \mathbf{0}$ in (17), and the only sources of errors are the location shift and the innovation ν_{T+1}. Under these assumptions, the bias (or expected error) of the 1-step ahead forecast error is given by:

$$\mathsf{E}\left[\Delta \mathbf{x}_{T+1} - \widehat{\Delta \mathbf{x}}_{T+1|T} \right] = -\alpha_0 \nabla \mu_0^* \tag{25}$$

which, from (23), corresponds to the unanticipated change in $\Delta \mathbf{x}_{T+1}$ relative to the constant-parameter setting. Importantly, one period later:

$$\Delta \mathbf{x}_{T+2} = \gamma_0 + \alpha_0 \left(\beta_0' \mathbf{x}_{T+1} - \mu_0 \right) - \alpha_0 \nabla \mu_0^* + \nu_{T+2} \tag{26}$$

so using:

$$\widehat{\Delta \mathbf{x}}_{T+2|T+1} = \gamma_0 + \alpha_0 \left(\beta_0' \mathbf{x}_{T+1} - \mu_0 \right) \tag{27}$$

the mean error remains the same as in (25), namely:

$$\mathsf{E}\left[\Delta\mathbf{x}_{T+2} - \widehat{\Delta\mathbf{x}}_{T+2|T+1}\right] = -\alpha_0\nabla\mu_0^* \tag{28}$$

Thus, systematic forecast errors result, and these are possibly very large, as μ relates to $\beta_0'\mathbf{x}$ which can have a large mean, whereas the errors relate to $\Delta\mathbf{x}$, which as a growth rate is bound to be relatively small.

4.4 Forecasting by a double-differenced device

There are two closely related approaches to avoiding forecast failure by improving robustness to location shifts:

- forecasting from a double-differenced device (denoted DDD) which adjusts quickly to breaks;
- differencing the VEqCM in (17) to eliminate the equilibrium mean and growth intercept.

In this section, we only consider the DDD. The DDD does not attempt to fit the data in-sample and consequently will have a larger forecast-error variance than (18), but this will be partially offset by lower parameter estimation uncertainty (cf. Clements and Hendry, 1998, chap. 12). However, when a break occurs after forecasts are announced, the DDD will fare no better than the VEqCM in terms of forecast bias (see Clements and Hendry, 1996, 1999). The key difference between the two lies in their performance when forecasting at any point after a break has already occurred, in which case the VEqCM continues to perform just as badly in terms of forecast bias as shown in (28), whereas the DDD forecasts are less affected by the earlier break, as we show below. Differencing the VEqCM is shown in section 4.5.3 to also reduce forecast biases when there are shifts in μ_0.

The DDD is the so-called ‘naive’ forecasting rule:

$$\overline{\Delta\mathbf{x}}_{T+h+1|T+h} = \Delta\mathbf{x}_{T+h} \tag{29}$$

reflecting that most economic time series do not continuously accelerate, so:

$$\mathsf{E}\left[\Delta^2\mathbf{x}_t\right] = \mathbf{0} \tag{30}$$

(hence the epithet DDD). Thus, despite its simplicity, (29) delivers unconditionally unbiased, but noisy, forecasts when $\mathbf{x}_t$ is $\mathsf{I}(1)$, or is $\mathsf{I}(2)$ with a zero mean (when $\mathbf{x}_t$ is $\mathsf{I}(0)$, only single additional differencing is required). In comparison to (28), one period after the break, the forecast error using the DDD forecast $\overline{\Delta\mathbf{x}}_{T+2|T+1} = \Delta\mathbf{x}_{T+1}$, when the DGP is given by (22) with $\mathbf{\Upsilon}_0 = \mathbf{0}$, is:

$$\begin{aligned}\overline{\mathbf{u}}_{T+2|T+1} &= \Delta\mathbf{x}_{T+2} - \overline{\Delta\mathbf{x}}_{T+2|T+1} = \gamma_0 + \alpha_0\left(\beta_0'\mathbf{x}_{T+1} - \mu_0^*\right) + \nu_{T+2} - \Delta\mathbf{x}_{T+1} \\ &= \alpha_0\beta_0'\Delta\mathbf{x}_{T+1} + \Delta\nu_{T+2},\end{aligned}$$

where we have used $\Delta\mathbf{x}_{T+1} = \gamma_0 + \alpha_0\left(\beta_0'\mathbf{x}_T - \mu_0^*\right) + \nu_{T+1}$. As $\beta_0'\gamma_0 = \mathbf{0}$, the forecast bias is now given by:

$$\mathsf{E}\left[\overline{\mathbf{u}}_{T+2|T+1}\right] = \mathsf{E}\left[\alpha_0\beta_0'\Delta\mathbf{x}_{T+1} + \Delta\nu_{T+2}\right] = \alpha_0\beta_0'\mathsf{E}\left[\alpha_0\left(\beta_0'\mathbf{x}_T - \mu_0^*\right)\right] = -\alpha_0\left(\beta_0'\alpha_0\right)\nabla\mu_0^*. \tag{31}$$

Since $\beta_0'\alpha_0$ corresponds to the stationary roots of the system, the expected 1-step error of the VEqCM from (28) is necessarily larger than (31), although the bias of the DDD forecasts is not zero. However, the sequence of 1-step DDD errors that would be made h-periods after the shift occurred is:

$$\mathsf{E}\left[\Delta\mathbf{x}_{T+h} - \overline{\Delta\mathbf{x}}_{T+h|T+h-1}\right] = -\alpha_0\left(\beta'\alpha\right)\mathbf{\Psi}_0^{h-2}\nabla\mu_0^* \tag{32}$$

where $\mathbf{\Psi}$ is defined in (14). This outcome compares increasingly favourably to the persistent VEqCM bias $-\alpha_0\nabla\mu_0^*$ as the forecast horizon h increases.

There are a number of related ways of viewing the forecasting success of DDD relative to the VEqCM even though the DDD is mis-specified for the DGP both pre and post-break, and is a non-congruent model of the process (in the sense of, e.g., Hendry, 1995). These include that double differencing:
(a) removes up to two unit roots, any intercepts and linear trends;
(b) changes location shifts to 'blips' and breaks in trends to impulses, thereby almost ensuring the absence of deterministic terms;
(c) does not introduce any estimation uncertainty.
Although (29) will still suffer forecast failure when large changes in μ_0 occur, it adjusts to breaks, whereas the VEqCM fails to do so. In the next section, we consider the properties of forecasts from the VEqCM and the DDD when the LDGP is subject to general forms of non-constancy, and in so doing shed further light on the reasons for the success of differencing as a robust forecasting device.

4.5 General non-constancy

Having highlighted the differential effect of a location shift on the forecasts of a VEqCM and the DDD, we now consider the case where all the parameters shift, to parallel the taxonomy discussion in section 3, so for $h = 1, 2 \ldots$, after the break, the DGP becomes:

$$\Delta\mathbf{x}_{T+h} = \gamma_0^* + \alpha_0^*\left((\beta_0^*)'\mathbf{x}_{T+h-1} - \mu_0^*\right) + \mathbf{\Upsilon}_0^*\left(\mathbf{z}_{T+h} - \psi_0^*\right) + \nu_{T+h}. \tag{33}$$

4.5.1 VEqCM forecasts

The forecasting model omits relevant variables, because $\mathbf{\Upsilon}_0 \neq \mathbf{0}$, although that is unknown to the investigator, but as the omitted variables are orthogonal to the included variables, the in-sample forecasting model errors are given by $\epsilon_t = \mathbf{\Upsilon}_0\left(\mathbf{z}_t - \psi_0\right) + \nu_t$. Let the VEqCM forecast be:

$$\widehat{\Delta\mathbf{x}}_{T+h|T+h-1} = \widehat{\gamma}_{(T)} + \widehat{\alpha}_{(T)}\left(\widehat{\beta}'_{(T)}\mathbf{x}_{T+h-1} - \widehat{\mu}_{(T)}\right) \tag{34}$$

so that if $\mathbf{w}_{T+h|T+h-1}$ denotes the forecast errors, $\mathbf{w}_{T+h|T+h-1} = \Delta\mathbf{x}_{T+h} - \widehat{\Delta\mathbf{x}}_{T+h|T+h-1}$ then:

$$\begin{aligned}\mathbf{w}_{T+h|T+h-1} = {} & \gamma_0^* + \alpha_0^*\left((\beta_0^*)'\mathbf{x}_{T+h-1} - \mu_0^*\right) + \mathbf{\Upsilon}_0^*\left(\mathbf{z}_{T+h} - \psi_0^*\right) + \nu_{T+h} \\ & - \widehat{\gamma}_{(T)} - \widehat{\alpha}_{(T)}\left(\widehat{\beta}'_{(T)}\mathbf{x}_{T+h-1} - \widehat{\mu}_{(T)}\right)\end{aligned} \tag{35}$$

Thus, the forecast error could consist of all the main sources of error set out in section 3:
(i) stochastic and deterministic breaks–changes in parameters that multiply stochastic terms (α and β), and changes in deterministic terms (γ and μ);
(ii) omitted variables–the forecast omits any influence from $(\mathbf{z}_{T+h} - \psi_0^*)$;
(iii) inconsistent parameter estimates for both stochastic and deterministic terms–if the omitted variables

were not orthogonal to the included variables, or in-sample data measurement errors occurred;
(iv) estimation uncertainty–the forecast is based on $(\widehat{\gamma}_{(T)}, \widehat{\alpha}_{(T)}, \widehat{\beta}_{(T)}, \widehat{\mu}_{(T)})$ rather than the population values of these parameters;
(v) innovation errors–ν_{T+h}; and
(vi) forecast origin measurement errors–if the data the forecast is conditioned on are also measured with error: see section 8.

To highlight the impact of the break, we again replace all in-sample estimates by their corresponding pseudo-true parameter values, assuming no in-sample data mis-measurement, and orthogonal omitted variables, so in fact (e.g.) $\mathsf{E}[\widehat{\gamma}_{(T)}] = \gamma_0$ etc., leading to the sequence of 1-step ahead forecast errors as the forecast origin increases after the break:[2]

$$\mathbf{w}_{T+h|T+h-1} \simeq \gamma_0^* + \alpha_0^* \left((\beta_0^*)' \mathbf{x}_{T+h-1} - \mu_0^* \right) + \boldsymbol{\Upsilon}_0^* \left(\mathbf{z}_{T+h} - \psi_0^* \right) + \nu_{T+h} - \gamma_0 - \alpha_0 \left(\beta_0' \mathbf{x}_{T+h-1} - \mu_0 \right) \quad (36)$$

To calculate the average forecast error, we set $\beta_0^* = \beta_0$, and use $\mathsf{E}\left[\mathbf{z}_{T+h} - \psi_0^*\right] = \mathbf{0}$ so that:[3]

$$\mathsf{E}\left[\mathbf{w}_{T+h|T+h-1}\right] = \nabla\gamma_0^* - \alpha_0 \nabla\mu_0^* + \nabla\alpha_0^* \mathsf{E}\left[\beta_0' \mathbf{x}_{T+h-1}\right] \quad (37)$$

where $\nabla\alpha_0^* = \alpha_0^* - \alpha_0$, and:

$$\mathsf{E}\left[\beta_0' \mathbf{x}_{T+h-1}\right] = \left(\boldsymbol{\Psi}_0^*\right)^{h-1} \mu_0 - \sum_{s=0}^{h-2} \left(\boldsymbol{\Psi}_0^*\right)^s \beta_0' \alpha_0^* \mu_0^*.$$

This last expression comes from:

$$\beta_0' \mathbf{x}_{T+h-1} = \left(1 + \beta_0' \alpha_0^*\right) \beta_0' \mathbf{x}_{T+h-2} - \beta_0' \alpha_0^* \mu_0^* + \beta_0' \nu_{T+h-1},$$

with $\boldsymbol{\Psi}_0^* = 1 + \beta_0' \alpha_0^*$, and $\mathsf{E}\left[\beta_0' \mathbf{x}_T\right] = \mu_0$, using $\beta_0' \gamma_0^* = \beta_0' \gamma_0 = \mathbf{0}$.

The expression for the bias of the VEqCM forecasts, (37), contains additional terms relative to (28). One comes from the change in the growth rate, while the final term will be zero when the weights on the cointegrating vectors remain unchanged over the break. All the other problems in (*ii*)–(*vi*), especially data mis-measurement and parameter estimation variances, will further increase forecast error uncertainty.

4.5.2 DDD forecasts

In comparison, for the same DGP, consider forecasting from the DDD two or more periods after the break occurred. Hence, for $h > 1$ we use $\overline{\Delta^2 \mathbf{x}}_{T+h|T+h-1} = \mathbf{0}$, or:

$$\overline{\Delta \mathbf{x}}_{T+h|T+h-1} = \Delta \mathbf{x}_{T+h-1} \quad (38)$$

to forecast $\Delta \mathbf{x}_{T+h}$ with the resulting forecast error $\overline{\mathbf{u}}_{T+h|T+h-1} = \Delta \mathbf{x}_{T+h} - \overline{\Delta \mathbf{x}}_{T+h|T+h-1}$, which is (even keeping β_0^* fixed at β_0):

$$\begin{aligned} \overline{\mathbf{u}}_{T+h|T+h-1} &= \gamma_0^* + \alpha_0^* \left(\beta_0{}' \mathbf{x}_{T+h-1} - \mu_0^* \right) + \boldsymbol{\Upsilon}_0^* \left(\mathbf{z}_{T+h} - \psi_0^* \right) + \nu_{T+h} \\ &\quad - \left[\gamma_0^* + \alpha_0^* \left(\beta_0{}' \mathbf{x}_{T+h-2} - \mu_0^* \right) + \boldsymbol{\Upsilon}_0^* \left(\mathbf{z}_{T+h-1} - \psi_0^* \right) + \nu_{T+h-1} \right] \\ &= \alpha_0^* \beta_0{}' \Delta \mathbf{x}_{T+h-1} + \boldsymbol{\Upsilon}_0^* \Delta \mathbf{z}_{T+h} + \Delta \nu_{T+h}. \end{aligned} \quad (39)$$

[2]By fixing the model's parameters at their in-sample pseudo-true values, we are not updating the model as the forecast origin moves further past the break: Castle *et al.* (2010) discuss updating in that context.

[3]As $(\beta_0^*)' \mathbf{x}_{T+i-1}$ and $\beta_0' \mathbf{x}_{T+i-1}$ cannot both be $\mathsf{I}(0)$ when the cointegrating vector changes substantively, unconditional expectations are then not well defined, so we set $\beta_0^* = \beta_0$ here.

Using $\mathsf{E}\left[\Delta \mathbf{z}_{T+h-1}\right] = \mathbf{0}$, we derive the expected forecast error as:

$$\mathsf{E}\left[\overline{\mathbf{u}}_{T+h|T+h-1}\right] = \alpha_0^* \beta_0{}' \mathsf{E}\left[\Delta \mathbf{x}_{T+h-1}\right] = -\alpha_0^* \beta_0' \alpha_0^* \mathbf{\Psi}_0^{h-2} \nabla \mu_0^* \tag{40}$$

where we have used the expression for $\mathsf{E}[\beta_0 \mathbf{x}_{T+h-1}]$ derived in the previous section. Inspection of (40) reveals that $\overline{\Delta \mathbf{x}}_{T+h|T+h-1}$ suffers decreasingly small forecast biases as h increases, whereas the corresponding VEqCM faces systematic forecast failure from (37).

The crucial insight, and the reason why DDD fares so well relative to the VEqCM, is that the DDD forecast value, $\Delta \mathbf{x}_{T+h-1}$, for $h > 1$, is given by the DGP itself–*irrespective of what is known by the economist*–since $\Delta \mathbf{x}_{T+h-1}$ is:

$$\Delta \mathbf{x}_{T+h-1} = \gamma_0^* + \alpha_0^* \left(\beta_0{}' \mathbf{x}_{T+h-2} - \mu_0^*\right) + \mathbf{\Upsilon}_0^* \left(\mathbf{z}_{T+h-1} - \psi_0^*\right) + \nu_{T+h-1} \tag{41}$$

From (38), $\Delta \mathbf{x}_{T+h-1}$ must equal all the right-hand side terms, so already embodies all the breaks, unknowingly omitted variables etc., so automatically incorporates model mis-specification and parameter change, without the forecaster needing to know the causal variables or the structure or contents of the DGP. Thus, the use of $\Delta \mathbf{x}_{T+h-1}$ avoids all of the problems *(i)–(iv)* above:
(i) all stochastic and deterministic breaks are included;
(ii) it contains the unknown influences, $\mathbf{z}_{T+h-1}$;
(iii) the new population parameters are correctly embodied;
(iv) there are no parameters to estimate, precluding estimation uncertainty.
Forecasting is a fundamentally different activity from either modelling, where a formal specification that is close to the LDGP is essential, or policy, where causal knowledge and constant (indeed invariant) identified parameters are both paramount: delineating such influences is not necessary for forecasting.

Against these major advantages, the drawbacks to using (38) are:
(a) the unwanted presence of ν_{T+h-1} in (41), leading to the differencing of the innovation error and a doubling of that contribution to the forecast-error variance; and
(b) all variables in (41) are lagged one period relative to the DGP, adding 'noise' from many $\mathsf{I}(-1)$ effects.
Thus there is a trade-off between using a carefully modelled VEqCM like (34) (which might nevertheless still be mis-specified and non-constant), and the 'naive' predictor (38): when forecasting in the face of many location shifts for DGPs that are more complicated than the model, $\Delta \mathbf{x}_{T+h-1}$ can often be a dominant forecasting device. Sufficiently long after a break (so h is large), (40) converges to zero, so that forecast failure is a temporary phenomenon in this regard, and (38) will only fail to outperform the VEqCM if the variance effects are sufficiently large. Nevertheless, a DDD does not really 'forecast': rather it closely tracks outcomes after the event.

4.5.3 The Differenced VEqCM

Since developing a VEqCM for forecasting involves considerable investment, and unlike a DDD has both potential policy applications and could forecast, we consider forecasting not from (17) itself, but from the difference of the estimated congruent representation, namely:

$$\widehat{\Delta^2 \mathbf{x}}_{T+h|T+h-1} = \Delta \widehat{\gamma}_{(T)} + \widehat{\alpha}_{(T)} \Delta \left(\widehat{\beta}'_{(T)} \mathbf{x}_{T+h-1} - \widehat{\mu}_{(T)}\right) = \widehat{\alpha}_{(T)} \widehat{\beta}'_{(T)} \Delta \mathbf{x}_{T+h-1} \tag{42}$$

with the rank restriction from cointegration imposed. (42) can be interpreted as augmenting the DDD forecast (38) by 'adding back' the main observable component omitted by using just the lagged first difference, namely $\widehat{\alpha}_{(T)}\widehat{\beta}'_{(T)}\Delta\mathbf{x}_{T+h-1}$. Since zero-mean shifts in parameters do not have a major effect on forecast accuracy, such a strategy is likely to have some benefits even if α or β change.

Alternatively, (42) can be written as:

$$\widehat{\Delta\mathbf{x}}_{T+h|T+h-1} = \Delta\mathbf{x}_{T+h-1} + \widehat{\alpha}_{(T)}\widehat{\beta}'_{(T)}\Delta\mathbf{x}_{T+h-1} = \left(\mathbf{I}_n + \widehat{\alpha}_{(T)}\widehat{\beta}'_{(T)}\right)\Delta\mathbf{x}_{T+h-1} \quad (43)$$

Thus, $\Delta\mathbf{x}_{T+h-1}$ in (43) can also be interpreted as a highly adaptive estimator of γ in (17), where μ is then also approximated by the previous value of the cointegrating combination, $\beta'\mathbf{x}_{T+h-2}$, so both parameters are replaced by 'instantaneous unbiased estimators'. Given (33), at $h > 1$ when $\beta_0^* = \beta_0$:

$$\Delta^2\mathbf{x}_{T+h} = \alpha_0^*\beta_0{}'\Delta\mathbf{x}_{T+h-1} + \boldsymbol{\Upsilon}_0^*\Delta\mathbf{z}_{T+h} + \Delta\nu_{T+h} \quad (44)$$

then from (44), letting $\widehat{\upsilon}_{T+h|T+h-1} = \Delta^2\mathbf{x}_{T+h} - \widehat{\Delta^2\mathbf{x}}_{T+h|T+h-1}$:

$$\widehat{\upsilon}_{T+h|T+h-1} = \left(\alpha_0^*\beta_0' - \widehat{\alpha}_{(T)}\widehat{\beta}'_{(T)}\right)\Delta\mathbf{x}_{T+h-1} + \boldsymbol{\Upsilon}_0^*\Delta\mathbf{z}_{T+h} + \Delta\nu_{T+h}$$

revealing that the forecast error consists of $\mathsf{I}(-1)$ terms. To clarify the outcome, take the case where $\beta_0 = \widehat{\beta}_{(T)}$ and $\widehat{\alpha}_{(T)} = \alpha_0$ (so estimation uncertainty is ignored), leading to:

$$\mathsf{E}\left[\widehat{\upsilon}_{T+h|T+h-1}\right] = \nabla\alpha_0^*\beta_0'\mathsf{E}\left[\Delta\mathbf{x}_{T+h-1}\right] = -\left(\nabla\alpha_0^*\right)\left(\beta_0'\alpha_0^*\right)\boldsymbol{\Psi}_0^{h-2}\nabla\mu_0^* \quad (45)$$

Changes in α_0 are unlikely to be large, so (45) will generally be small, and in any case declines as the forecast origin moves past the break point. Since the change in the 'loadings', $\nabla\alpha_0^*$, is likely to be smaller than their post-break value, α_0^*, we would expect a smaller bias from (45) than from (40)

Like a DDD, the DVEqCM in (42) has no deterministic terms (e.g., constants or linear trends, t), and hence does not equilibrium correct, thereby reducing the risks attached to VEqCMs. However, although it will produce noisy forecasts, smoothed variants are easily formulated if desired. When there are no locations shifts, the 'insurance' of differencing must worsen forecast accuracy and precision, so is a clear cost to using this strategy. Conversely, if location shifts do occur, differencing can pay handsome dividends, as is evident from comparing (45) to (37).

5 Disaggregation over variables

In this section, building on Hendry and Hubrich (2009), we consider whether forecast failure can be mitigated when there are parameter non-constancies by making use of any available disaggregate information, extending the literature that considers these issues in the absence of structural breaks (see, e.g., the literature cited by Hendry and Hubrich, 2009, especially Giacomini and White, 2006, and the review by Lütkepohl, 2006). The variables of interest in macroeconomic forecasting will often be aggregates: e.g., GDP is an aggregate of components, most of which are in fact themselves aggregates (e.g., disaggregated consumer's expenditure, or output components). In recent years, aggregates over geographical regions, such as Euro-area aggregate measures across individual member states, have received increasing attention, although there has been little direct work on the best forecasting strategy in that context when models are unknowingly mis-specified and there are structural breaks. Two forecasting strategies are of particular interest:
i) aggregating the forecasts of the disaggregated components;
ii) forecasting the aggregate directly.

We first deduce the implications of changing the information set by adding disaggregates, then investigate the extent to which the two forecasting strategies differ when confronted by an unanticipated break.

Unpredictability is relative to the information used, so when the process to be predicted is $\mathbf{y}_t = \mathbf{h}_t(\mathcal{I}_{t-1}) + \nu_t$, with ν_t a non-degenerate vector random variable unpredictable with respect to the information set $\mathcal{I}_{t-1}$, predictions of no greater accuracy will result from a reduced information set $\mathcal{J}_{t-1} \subset \mathcal{I}_{t-1}$, although the predictions will remain unbiased (see Clements and Hendry, 2005). The converse holds for increasing information.

When the conditional distribution $\mathsf{D}_{y^a_{T+1}}(y^a_{T+1}|\cdot)$ is the target of interest, one form of increasing the information set is to disaggregate a variable y^a_T into its components $y_{i,T}$, and then to add those disaggregates to the information set. Disaggregating the aggregate y^a_t into its components $y_{i,t}$ extends the information set from $\mathcal{J}_{t-1}$ to $\mathcal{I}_{T-1}$ by including both the past aggregate and the disaggregates, so cannot lower, predictability of the aggregate y^a_T. Consider a scalar y^a_t to be predicted, composed of:

$$y^a_{T+1} = w_{1,T+1}y_{1,T+1} + w_{2,T+1}y_{2,T+1} \tag{46}$$

with weights $w_{1,T+1}$ and $w_{2,T+1} = (1 - w_{1,T+1})$ for each of the two components, where the weights may change over time, but could be fixed.

Let the DGP for the aggregate be:

$$y^a_t = \mathbf{x}'_{t-1}\delta_t + u_t$$

and its components:

$$y_{i,t} = \mathbf{x}'_{t-1}\gamma_{i,t} + e_{i,t}$$

so:

$$\mathsf{E}_{T+1}\left[y_{i,T+1} \mid \mathbf{x}_T\right] = \mathbf{x}'_T\gamma_{i,T+1} \tag{47}$$

which is the conditional expectation of each component $y_{i,T+1}$ and hence is the minimum mean-square error (MSE) predictor. The subscript $T+1$ in E_{T+1} denotes that the expectation might be time-varying, for example, from structural change in the prediction period. $\mathbf{x}_T$ denotes the general set of variables up to time T, including lags of the aggregate and disaggregate components of the aggregate, as well as other variables.

Then, taking conditional expectations in (46), aggregating the two terms in (47) delivers $\mathsf{E}_{T+1}[y^a_{T+1}|\mathbf{x}_T]$:

$$\begin{aligned}\mathsf{E}_{T+1}\left[y^a_{T+1} \mid \mathbf{x}_T\right] &= \sum_{i=1}^{2} w_{i,T+1}\mathsf{E}_{T+1}\left[y_{i,T+1} \mid \mathbf{x}_T\right] \\ &= \sum_{i=1}^{2} w_{i,T+1}\mathbf{x}'_T\gamma_{i,T+1} \\ &= \sum_{i=1}^{2} \mathbf{x}'_T\gamma^*_{i,T+1} = \lambda'_{T+1}\mathbf{x}_T.\end{aligned} \tag{48}$$

By way of comparison, consider predicting y^a_{T+1} directly from $\mathbf{x}_T$:

$$\mathsf{E}_{T+1}\left[y^a_{T+1} \mid \mathbf{x}_T\right] = \delta'_{T+1}\mathbf{x}_T \tag{49}$$

Then as the left-hand sides of (48) and (49) are the same, so are the right, hence $\delta'_{T+1}\mathbf{x}_T = \lambda'_{T+1}\mathbf{x}_T\ \forall\mathbf{x}_T$. Thus, the two approaches have the same prediction error:

$$y_{T+1} - \mathsf{E}_{T+1}\left[y^a_{T+1} \mid \mathbf{x}_T\right] = u_{T+1} \tag{50}$$

which is unpredictable from $\mathbf{x}_T$: once the general information set $\mathbf{x}_T$ is used, predicting y_{T+1} directly is the same as aggregating the component predictions.

In practice, if either the weights $w_{i,T+1}$ or the coefficients of the component models $\delta_{i,T+1}$ change more than the coefficients of the aggregate model λ_{T+1}, predicting the aggregate directly including disaggregate information can be more accurate than aggregating the component predictions. This might be the case if some of the disaggregate components are volatile and therefore difficult to predict, or covariances between disaggregates lead to a more stable aggregate. On the other hand, if the weights and the coefficients of the disaggregate models do not change much, the disaggregates may be easier to predict. Then aggregating the disaggregate predictions could improve over predicting the aggregate directly even when disaggregate information is included. Thus, the key issue in aggregate prediction is not predicting the component changes, but including lags of those components in $\mathbf{x}_T$. This result implies that weights are not needed for aggregating the component predictions, and also saves the additional effort of specifying disaggregate models for the components, at least for the purpose of forecasting the aggregate (the disaggregate models and predictions might be of interest in themselves).

5.1 Disaggregating when there are unanticipated breaks

Let $\mathbf{y}_t$ denote the vector of disaggregate variables, with typical element $y_{i,t}$. We assume that $\mathbf{y}_t$ is $\mathsf{I}(0)$, and that the in-sample DGP is the stationary first-order VAR:

$$\mathbf{y}_t = \varphi + \mathbf{\Pi}\mathbf{y}_{t-1} + \epsilon_t \text{ for } t = 1, \ldots T-1 \tag{51}$$

where $\epsilon_t \sim \mathsf{IID}[\mathbf{0}, \mathbf{\Omega}]$ and the unknown parameters $(\varphi, \mathbf{\Pi})$ are constant, but have to be estimated. It is convenient to write the DGP in mean-deviation form:

$$\mathbf{y}_t - \mathsf{E}[\mathbf{y}_t] = \mathbf{y}_t - \phi_y = \mathbf{\Pi}(\mathbf{y}_{t-1} - \phi_y) + \epsilon_t \tag{52}$$

where ϕ_y is defined by $\mathsf{E}[\mathbf{y}_t] = \varphi + \mathbf{\Pi}\mathsf{E}[\mathbf{y}_{t-1}] = \varphi + \mathbf{\Pi}\phi_y = \phi_y$. However, as above there is an (unknown) break just before the forecast origin at $T-1$, so the following analysis extends the derivations in Hendry and Hubrich (2009) to forecasting from an origin one period after the break:

$$\mathbf{y}_{T+h} = \varphi^* + \mathbf{\Pi}^*\mathbf{y}_{T+h-1} + \epsilon_{T+h} \text{ for } h = 0, \ldots H \tag{53}$$

although the process stays $\mathsf{I}(0)$. Note that $\mathbf{y}_t$, $t = T, T+1, \ldots$ is non-stationary, with $\mathsf{E}[\mathbf{y}_T] = \varphi^* + \mathbf{\Pi}^*\phi_y$ and $\mathsf{E}[\mathbf{y}_{T+1}] = (\mathbf{I}_n + \mathbf{\Pi}^*)\varphi^* + (\mathbf{\Pi}^*)^2\phi_y$ etc., converging on $\mathsf{E}[\mathbf{y}_{T+h}] = \phi_y^* = (\mathbf{I}_n - \mathbf{\Pi}^*)^{-1}\varphi^*$ as $h \to \infty$.

Assuming a break at $T-1$, rather than a forecast origin break as in Hendry and Hubrich (2009), may at first sight appear a minor change, but turns out to have important consequences for the use of disaggregate information compared to forecasting the aggregate. When the break occurs at the forecast origin, structural breaks (both unmodelled location shifts and slope changes) affect both methods of forecasting the aggregate equally. Hence the relative forecast accuracy of aggregating the disaggregate forecasts, and forecasting from the model for the aggregate, is unaffected. As shown below, this is not the case for the earlier break. In retrospect, this is similar to what was shown above when comparing the DDD and VEqCM: they produced identical forecast errors for a break at T when forecasting $T+1$, but not when forecasting $T+2$ one period later.

5.2 Aggregating disaggregated forecasts

We assume the disaggregated forecasts are based on an estimated version of the in-sample DGP using data up to $T-1$: other possibilities are that each disaggregate could be forecast using univariate time-series models; or indeed from models with additional explanatory variables. Our assumption focuses the analysis on structural breaks. The forecasts of the disaggregates from the estimated version of (52) based on the forecast origin $\widehat{\mathbf{y}}_T$ (possibly measured with error, relative to $\mathbf{y}_T$) are given by:

$$\widehat{\mathbf{y}}_{T+1|T} = \widehat{\phi}_y + \widehat{\mathbf{\Pi}}\left(\widehat{\mathbf{y}}_T - \widehat{\phi}_y\right) \tag{54}$$

The variable of interest is the aggregate, denoted y_t^a, where $y_t^a = \omega' \mathbf{y}_t$, with ω the vector of weights, so that the error in forecasting the aggregate is $\omega' \widehat{\epsilon}_{T+1|T} = \omega' \mathbf{y}_{T+1} - \omega' \widehat{\mathbf{y}}_{T+1|T}$ which can be written as:

$$\omega' \widehat{\epsilon}_{T+1|T} = \left[\omega' \phi_y^* - \omega' \widehat{\phi}_y\right] + \left[\omega' \mathbf{\Pi}^* \left(\mathbf{y}_T - \phi_y^*\right) - \omega' \widehat{\mathbf{\Pi}}\left(\widehat{\mathbf{y}}_T - \widehat{\phi}_y\right)\right] + \omega' \epsilon_{T+1} \tag{55}$$

Assuming the relevant moments exist, we let $\mathsf{E}[\widehat{\mathbf{\Pi}}] = \mathbf{\Pi}_e$ and $\mathsf{E}[\widehat{\phi}_y] = \phi_{y,e}$. The forecast-error taxonomy is formed by decomposing the forecast error in (55) into terms that reflect parameter shifts, parameter mis-specifications, and the estimation uncertainty of the parameters–see table 1. The decomposition uses $\varphi = (\mathbf{I}_n - \mathbf{\Pi})\phi_y$ and $\varphi^* = (\mathbf{I}_n - \mathbf{\Pi}^*)\phi_y^*$; and from (53), $\mathsf{E}[\mathbf{y}_T] = \varphi^* + \mathbf{\Pi}^* \phi_y = \phi_y + (\mathbf{I}_n - \mathbf{\Pi}^*)(\phi_y^* - \phi_y) = \phi_y^* - \mathbf{\Pi}^*(\phi_y^* - \phi_y)$. Terms with non-zero means will contribute to the bias of the forecast and are shown in bold. All entries will be discussed below, where they are compared to their counterparts from a decomposition of the forecast error resulting from a model for the aggregate.[4]

Table 1: Aggregated disaggregate forecast-error taxonomy

$$\begin{array}{ll}
\omega' \widehat{\epsilon}_{T+1|T} \simeq & \\
\omega' \left[\mathbf{I}_n - (\mathbf{\Pi}^*)^2 - \mathbf{\Pi}_e\left(\mathbf{I}_n - \mathbf{\Pi}^*\right)\right]\left(\phi_y^* - \phi_y\right) & (ia)\ \textbf{equilibrium mean change} \\
+\omega' \left(\mathbf{\Pi}^* - \mathbf{\Pi}\right)\left(\mathbf{y}_T - \mathsf{E}\left[\mathbf{y}_T\right]\right) & (ib)\ \text{slope change} \\
+\omega' \left(\mathbf{I}_n - \mathbf{\Pi}_e\right)\left(\phi_y - \phi_{y,e}\right) & (iia)\ \textbf{equilibrium mean misspecification} \\
+\omega' \left(\mathbf{\Pi} - \mathbf{\Pi}_e\right)\left(\mathbf{y}_T - \mathsf{E}\left[\mathbf{y}_T\right]\right) & (iib)\ \text{slope mis-specification} \\
+\omega' \left(\mathbf{I}_n - \mathbf{\Pi}_e\right)\left(\phi_{y,e} - \widehat{\phi}_y\right) & (iiia)\ \text{equilibrium mean estimation} \\
-\omega' \left(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}_e\right)\left(\mathbf{y}_T - \phi_{y,e}\right) & (iiib)\ \text{slope estimation} \\
-\omega' \mathbf{\Pi}_e\left(\mathbf{y}_T - \widehat{\mathbf{y}}_T\right) & (iv)\ \textbf{forecast-origin mis-measurement} \\
+\omega' \left(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}_e\right)\left(\widehat{\phi}_y - \phi_{y,e}\right) & (va)\ \text{covariance interaction} \\
-\omega' \left(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}_e\right)\left(\widehat{\mathbf{y}}_T - \mathbf{y}_T\right) & (vb)\ \text{mis-measurement interaction} \\
+\omega' \epsilon_{T+1} & (vi)\ \text{innovation error.}
\end{array} \tag{56}$$

[4]As the DGP is $\mathsf{I}(0)$, the dependence of the estimated parameters on the last observation is $\mathsf{O}_p(T^{-1})$, as can be seen by ending estimation at $T-2$, so is omitted below.

5.3 Forecasting the aggregate directly by its past

The model for the aggregate is assumed to consist solely of an explanation based on its own past. Given the DGP (51), the implied model for the aggregate is:

$$y_t^a = \omega'\phi_y + \omega'\mathbf{\Pi}\left(\mathbf{y}_{t-1} - \phi_y\right) + \omega'\epsilon_t = \tau + \kappa\left(y_{t-1}^a - \tau\right) + \nu_t \tag{57}$$

where $\tau = \omega'\phi_y$, such that (τ, κ) orthogonalize $\left(1, \left(y_{t-1}^a - \omega'\phi_y\right)\right)$ with respect to ν_t, so:

$$\nu_t = \omega'\left(\mathbf{\Pi} - \kappa\mathbf{I}_n\right)\left(\mathbf{y}_{t-1} - \phi_y\right) + \omega'\epsilon_t \tag{58}$$

which is not an innovation with respect to $\mathbf{y}_{t-1}$, a potential loss from using the aggregate model (57), possibly offset by the reduction in parameter estimation uncertainty from having only 2 parameters to estimate compared to $n(n+1)$.

The forecasts from (57) are:

$$\widetilde{y}_{T+1|T}^a = \widetilde{\tau} + \widetilde{\kappa}\left(\widehat{y}_T^a - \widetilde{\tau}\right)$$

where $\widehat{y}_T^a = \omega'\widehat{\mathbf{y}}_T$ at the forecast origin, whereas the forecast-period aggregated DGP is:

$$y_{T+1}^a = \omega'\phi_y^* + \omega'\mathbf{\Pi}^*\left(\mathbf{y}_T - \phi_y^*\right) + \omega'\epsilon_{T+1}.$$

Letting $\omega'\phi_y^* = \tau^*$, and $\widetilde{\nu}_{T+1|T} = y_{T+1}^a - \widetilde{y}_{T+1|T}^a$ denote the aggregate model forecast error, then:

$$\widetilde{\nu}_{T+1|T} = \left(\tau^* - \widetilde{\tau}\right) + \omega'\mathbf{\Pi}^*\left(\mathbf{y}_T - \phi_y^*\right) - \widetilde{\kappa}\left(\widehat{y}_T^a - \widetilde{\tau}\right) + \omega'\epsilon_{T+1} \tag{59}$$

The derivations of the taxonomy are similar to those for table 1. Terms in bold letters again denote those that need not be zero under unconditional expectations, but would be zero if no shift in the mean occurred over the forecast period when a well-specified model was used from accurate forecast-origin measurements.

Table 2: Aggregate forecast-error taxonomy

$\widetilde{\nu}_{T+1\|T} \simeq$		
$\omega'\left[\mathbf{I}_n - \left(\mathbf{\Pi}^*\right)^2 - \kappa\left(\mathbf{I}_n - \mathbf{\Pi}^*\right)\right]\left(\phi_y^* - \phi_y\right)$	(*Ia*) **equilibrium mean change**	
$+\omega'\left(\mathbf{\Pi}^* - \mathbf{\Pi}\right)\left(\mathbf{y}_T - \mathsf{E}\left[\mathbf{y}_T\right]\right)$	(*Ib*) slope change	
$+\left(1-\kappa\right)\left(\tau - \tau_e\right)$	(*IIa*) **equilibrium mean mis-specification**	
$+\omega'\left(\mathbf{\Pi} - \kappa\mathbf{I}_n\right)\left(\mathbf{y}_T - \mathsf{E}\left[\mathbf{y}_T\right]\right)$	(*IIb*) slope mis-specification	
$+\left(1-\kappa\right)\left(\tau_e - \widetilde{\tau}\right)$	(*IIIa*) equilibrium mean estimation	(60)
$-\left(\widetilde{\kappa} - \kappa\right)\left(y_T^a - \tau_e\right)$	(*IIIb*) slope estimation	
$-\kappa\left(y_T^a - \widehat{y}_T^a\right)$	(*IV*) **forecast-origin mis-measurement**	
$+\left(\kappa - \widetilde{\kappa}\right)\left(\tau_e - \widetilde{\tau}\right)$	(*Va*) covariance interaction	
$+\left(\kappa - \widetilde{\kappa}\right)\left(\widehat{y}_T^a - y_T^a\right)$	(*Vb*) mis-measurement interaction	
$+\omega'\epsilon_{T+1}$	(*VI*) innovation error.	

The key finding from comparing the sources of forecast error in tables 1 and 2 is that the impact of the equilibrium mean shift is not the same in the two cases, unlike for a break at T. It will only have the same effect when $\mathbf{\Pi}_e = \kappa\mathbf{I}_n$: when this condition holds, there is no loss of information from using the aggregate model, as is evident from the expression for the aggregate model error term (58). Whenever the aggregate model is mis-specified relative to aggregating the disaggregates, the equilibrium mean change will affect the two differently. Since the impact is mediated by the new slope parameter, the relative size of (*ia*) and (*Ia*) cannot be determined without putting more structure on the problem. Secondly, the impact of a slope change (absent an equilibrium mean change) is the same in the two taxonomies, (*ib*) and (*Ib*). Hence, there may or may not be a benefit to combining disaggregate forecasts for forecasting an aggregate when there are breaks.

Consider now the other terms that might be expected to contribute to forecast bias. The first of these is mis-specification of the long-run mean: (*IIa*) and (*iia*). Provided the in-sample DGP is constant and the model is well specified, these terms ought to be zero in expectation whether we aggregate forecasts from the disaggregate model, or if we forecast using a model for the aggregate. The other term is data mis-measurement at the forecast origin (*iv*) and (*IV*). This would impact differentially in the two taxonomies if, say, the disaggregate variables, but not the aggregate, were subject to non-zero-mean measurement errors.

In the absence of equilibrium mean shifts, the choice between the two approaches will depend on the trade off between the lower estimation uncertainty of the aggregate model (which as here will typically have few parameters to be estimated relative to the disaggregate model) and the degree to which the aggregate model is mis-specified for the DGP.

5.4 Using robust methods

In this section, we consider whether there is any benefit to using robust methods such as DDD in this context. Firstly, consider the aggregated disaggregated forecast. As above, the shift occurs at $T-1$, so that:

$$\mathbf{y}_T = \varphi^* + \mathbf{\Pi}^* \mathbf{y}_{T-1} + \epsilon_T$$

and the DDD forecast is given by:

$$\overline{\mathbf{y}}_{T+1|T} = \widehat{\mathbf{y}}_T. \tag{61}$$

Denote the forecast error of the disaggregates using DDD as $\overline{\mathbf{u}}_{T+1|T} = \mathbf{y}_{T+1} - \overline{\mathbf{y}}_{T+1|T}$:

$$\begin{aligned}\overline{\mathbf{u}}_{T+1|T} &= \varphi^* + \mathbf{\Pi}^* \mathbf{y}_T + \epsilon_{T+1} - \widehat{\mathbf{y}}_T \\ &= \varphi^* + \mathbf{\Pi}^* \mathbf{y}_T + \epsilon_{T+1} - (\varphi^* + \mathbf{\Pi}^* \mathbf{y}_{T-1} + \epsilon_T) - (\widehat{\mathbf{y}}_T - \mathbf{y}_T) \\ &= \mathbf{\Pi}^* \Delta \mathbf{y}_T + \Delta \epsilon_{T+1} - (\widehat{\mathbf{y}}_T - \mathbf{y}_T).\end{aligned}$$

Then the error in forecasting the aggregate is:

$$\begin{aligned}\omega' \overline{\mathbf{u}}_{T+1|T} &= \omega' \mathbf{\Pi}^* \Delta \mathbf{y}_T - \omega' (\widehat{\mathbf{y}}_T - \mathbf{y}_T) + \omega' \Delta \epsilon_{T+1} \\ &= \omega' \mathbf{\Pi}^* (\mathbf{\Pi}^* - \mathbf{I}_n) \nabla \phi_y^* + \omega' \mathbf{\Pi}^* (\Delta \mathbf{y}_T - \mathsf{E}[\Delta \mathbf{y}_T]) - \omega' (\widehat{\mathbf{y}}_T - \mathbf{y}_T) + \omega' \Delta \epsilon_{T+1}\end{aligned}$$

where the decomposition in the second line comes from:

$$\Delta \mathbf{y}_T = (\mathbf{\Pi}^* - \mathbf{I}_n) \left(\mathbf{y}_{T-1} - \phi_y^* \right) + \epsilon_T$$

and so:

$$\mathsf{E}[\Delta \mathbf{y}_T] = (\mathbf{\Pi}^* - \mathbf{I}_n) \mathsf{E}\left[\left(\phi_y - \phi_y^* \right) \right] = (\mathbf{\Pi}^* - \mathbf{I}_n) \nabla \phi_y^*.$$

The resulting taxonomy is given in table 3.

Table 3: Aggregated-disaggregate robust forecast-error taxonomy

$$\begin{array}{lll}
\omega'\overline{\mathbf{u}}_{T+1|T} = & & \\
\omega'\mathbf{\Pi}^*\left(\mathbf{\Pi}^* - \mathbf{I}_n\right)\left(\phi_y^* - \phi_y\right) & (ia)\ \textbf{equilibrium mean change} & \\
+\omega'\mathbf{\Pi}^*\left(\Delta\mathbf{y}_T - \mathsf{E}\left[\Delta\mathbf{y}_T\right]\right) & (ib)\ \text{slope change} & \\
+0 & (iia)\ \textbf{equilibrium mean misspecification} & \\
+0 & (iib)\ \text{slope mis-specification} & \\
+0 & (iiia)\ \text{equilibrium mean estimation} & (62) \\
+0 & (iiib)\ \text{slope estimation} & \\
-\omega'\left(\widehat{\mathbf{y}}_T - \mathbf{y}_T\right) & (iv)\ \textbf{forecast-origin mis-measurement} & \\
+0 & (va)\ \text{covariance interaction} & \\
+0 & (vb)\ \text{mis-measurement interaction} & \\
+\omega'\Delta\epsilon_{T+1} & (vi)\ \text{innovation error.} &
\end{array}$$

Relative to the earlier taxonomy using the in-sample model for the disaggregates (table 1), (*iia*), (*iib*), (*iiia*), (*iiib*), (*va*) and (*vb*) all become zero, despite the 'mis-specification' of the DDD for the LDGP. Conversely, the LDGP disturbance is differenced (doubling the contribution of this source of error to the forecast-error variance), (*vi*). Consider now the effects of breaks. If we assume that $\mathbf{\Pi}^* \simeq \mathbf{\Pi}$, so that the change in the slope is 'small', the equilibrium mean change ($\phi_y^* - \phi_y$) is multiplied by approximately $(\mathbf{I}_n - \mathbf{\Pi})$ in table 1, but in table 3 involves the product $\mathbf{\Pi}^*\left(\mathbf{\Pi}^* - \mathbf{I}_n\right)$, which is more than halved [see (*ia*)]. Unsurprisingly, the robust forecast device reduces the impact of location shifts. Next, (*ib*) (slope effect), now involves the actual new parameter $\mathbf{\Pi}^*$ rather than the change, but again is multiplied by a zero-mean term, which will be small in general as $\mathbf{y}_t$ is I(0).[5] Finally, (*iv*) (forecast-origin mis-measurement) enters unweighted by the coefficient matrix. Castle, Fawcett and Hendry (2009) and Castle, Hendry and Fawcett (this volume) consider ways of improving the accuracy of initial estimates for forecasting when location shifts have just occurred.

Suppose instead that the aggregate is forecast directly using a robust forecasting device such as DDD:

$$\overline{y}^a_{T+1|T} = \widehat{y}^a_T \tag{63}$$

Because $\widehat{y}^a_T = \omega'\widehat{y}_T$, this is identical to aggregating the DDD forecasts of the disaggregates, although for other robust forecasting devices, such as the DVEqCM, this equivalence will not hold.

6 Time disaggregation

In this section, we next allow for the possibility that data may be available at a higher frequency than usually modelled, and consider whether that can be exploited to offset the effects of breaks on forecast performance. Castle and Hendry (2008) show that although higher-frequency data may enable the earlier detection of breaks, it is otherwise unhelpful in terms of mitigating the effects of breaks one period later, when the target variable is measured at the lower frequency. By more rapidly 'rolling ahead', the forecast origin passes the break for the higher-frequency data before it does so on the temporally-aggregated data, which should

[5] $\mathbf{\Pi}^*$ can be decomposed into the sequence $\mathbf{\Pi}^* = (\mathbf{\Pi}^* - \mathbf{\Pi}) + (\mathbf{\Pi} - \mathbf{\Pi}_e) - \left(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}_e\right) + \widehat{\mathbf{\Pi}}$ but little insight is added as there always remains a 'level' component, and none of the intermediate terms enter (61).

be advantageous unless the 'signal' is swamped by measurement error in the most recent higher-frequency data. However, we show that when the break has already occurred the use of time-disaggregated data can improve forecast accuracy.

In what follows, we assume the DGP is at the level of the time disaggregate, or higher-frequency, data, as otherwise there could be no benefit from temporal disaggregation, and that the forecasting model either matches the DGP, or only uses lower frequency data–Andreou, Ghysels and Kourtellos (this volume) consider forecasting models that mix data frequencies.

Let $\overline{y}_t$ (y_τ) denote the variable of interest and $\overline{\mathbf{z}}_t$ ($\mathbf{z}_\tau$) denote the vector of k explanatory variables with elements $\overline{z}_{i,t}$ ($z_{i,\tau}$), where $\tau = 2t$ for $\tau = 1, \ldots, \Upsilon + H$ (e.g., τ denotes semi-annual as against annual frequency), although the implications hold for any even τ. When y is a flow variable, the (equally scaled) time-aggregated variable $\overline{y}_t = (y_\tau + y_{\tau-1})/2$, for $\tau = 2, 4, 6, \ldots$. We consider an $\mathsf{I}(0)$ autoregressive distributed-lag DGP at the level of the time-disaggregated variables, which has constant parameters in-sample, so that for y_τ:

$$y_\tau = \mu + \rho y_{\tau-1} + \beta' \mathbf{z}_{\tau-1} + \epsilon_\tau \;\text{ for }\; \tau = 1, \ldots, \Upsilon - 1 \tag{64}$$

where $\epsilon_\tau \sim \mathsf{IN}[0, \sigma^2_\epsilon]$, and $|\rho| < 1$. Our analysis extends that of Castle and Hendry (2008) by introducing the break one (half) period before the forecast origin, rather than at the forecast origin, such that:

$$y_{\Upsilon+h} = \mu^* + \rho^* y_{\Upsilon+h-1} + (\beta^*)' \mathbf{z}_{\Upsilon+h-1} + \epsilon_{\Upsilon+h} \;\text{ for }\; h = 0, 1, \ldots, H \tag{65}$$

although the process is again assumed to remain $\mathsf{I}(0)$. Also, let:

$$\mathbf{z}_\tau = \gamma_z + \mathbf{\Gamma} \mathbf{z}_{\tau-1} + \epsilon_{z,\tau} \;\text{ for } \tau = 1, \ldots, \Upsilon - 1,$$

with $\phi_z = (\mathbf{I}_k - \mathbf{\Gamma})^{-1} \gamma_z$, and:

$$\mathbf{z}_{\Upsilon+h} = \gamma^*_z + \mathbf{\Gamma}^* \mathbf{z}_{\Upsilon+h-1} + \epsilon_{z,\Upsilon+h} \;\text{ for } h = 0, 1, \ldots, H,$$

with $\phi^*_z = (\mathbf{I}_k - \mathbf{\Gamma}^*)^{-1} \gamma^*_z$.

In section 6.1, we compare forecasts based on the disaggregated data with forecasts when only the time-aggregated data are available.

6.1 Forecasting using disaggregates versus aggregate data alone

We can write the in-sample DGP in mean-deviation form by taking expectations in (64) to give:

$$\mathsf{E}[y_\tau] = \mu + \rho \mathsf{E}[y_{\tau-1}] + \beta' \mathsf{E}[\mathbf{z}_{\tau-1}] = \mu + \rho \phi_y + \beta' \phi_z = \phi_y,$$

which defines the long-run in-sample mean as:

$$\phi_y = (1 - \rho)^{-1} \left(\mu + \beta' \phi_z \right) \tag{66}$$

By direct substitution:

$$y_\tau - \phi_y = \rho \left(y_{\tau-1} - \phi_y \right) + \beta' \left(\mathbf{z}_{\tau-1} - \phi_z \right) + \epsilon_\tau \;\text{ for }\; \tau = 1, \ldots, \Upsilon - 1 \tag{67}$$

Letting $\phi^*_y = (1 - \rho^*)^{-1} \left(\mu^* + (\beta^*)' \phi^*_z \right)$ denote the post-break equilibrium mean, then from (65):

$$y_{\Upsilon+h} - \phi^*_y = \rho^* \left(y_{\Upsilon+h-1} - \phi^*_y \right) + (\beta^*)' \left(\mathbf{z}_{\Upsilon+h-1} - \phi^*_z \right) + \epsilon_{\Upsilon+h} \;\text{ for }\; h = 0, \ldots, H \tag{68}$$

As we wish to forecast the time-aggregated outcome, $\overline{y}_{T+1} = (y_{\Upsilon+2} + y_{\Upsilon+1})/2$, consider $y_{\Upsilon+1}$ and $y_{\Upsilon+2}$:

$$y_{\Upsilon+1} = \phi_y^* + \rho^* \left(y_\Upsilon - \phi_y^*\right) + (\beta^*)' \left(\mathbf{z}_\Upsilon - \phi_z^*\right) + \epsilon_{\Upsilon+1} \tag{69}$$

$$\begin{aligned} y_{\Upsilon+2} &= \phi_y^* + \rho^* \left(y_{\Upsilon+1} - \phi_y^*\right) + (\beta^*)' \left(\mathbf{z}_{\Upsilon+1} - \phi_z^*\right) + \epsilon_{\Upsilon+2} \\ &= \phi_y^* + (\rho^*)^2 \left(y_\Upsilon - \phi_y^*\right) + \left(\rho^* (\beta^*)' + (\beta^*)' \mathbf{\Gamma}^*\right) \left(\mathbf{z}_\Upsilon - \phi_z^*\right) \\ &+ \epsilon_{\Upsilon+2} + \rho^* \epsilon_{\Upsilon+1} + (\beta^*)' \epsilon_{z,\Upsilon+1} \end{aligned} \tag{70}$$

where we have substituted $\mathbf{z}_{\Upsilon+1} - \phi_z^* = \mathbf{\Gamma}^* \left(\mathbf{z}_\Upsilon - \phi_z^*\right) + \epsilon_{z,\Upsilon+1}$.

The actual value of $\overline{y}_{T+1}$ is, from averaging (69) and (70):

$$\begin{aligned} \overline{y}_{T+1} &= \phi_y^* + \rho^* (1 + \rho^*) \left(\overline{y}_T - \phi_y^*\right) \\ &+ \left[(1 + \rho^*) (\beta^*)' + (\beta^*)' \mathbf{\Gamma}^*\right] \left(\overline{\mathbf{z}}_T - \phi_z^*\right) \\ &- \frac{1}{2}\rho^* (1 + \rho^*) \left(y_{\Upsilon-1} - \phi_y^*\right) \\ &- \frac{1}{2} \left[(1 + \rho^*) (\beta^*)' + (\beta^*)' \mathbf{\Gamma}^*\right] \left(\mathbf{z}_{\Upsilon-1} - \phi_z^*\right) \\ &+ \frac{1}{2} \left[\epsilon_{\Upsilon+2} + (1 + \rho^*) \epsilon_{\Upsilon+1} + (\beta^*)' \epsilon_{z,\Upsilon+1}\right] \end{aligned} \tag{71}$$

This expression allows us to compare a number of strategies based on different assumptions about information availability.

Suppose that the DGP is known, i.e., that the break was anticipated and that the new post-break parameters are known. When data through $y_{\Upsilon+1}, \mathbf{z}_{\Upsilon+1}$ are known, the only source of forecast error is $\frac{1}{2}\epsilon_{\Upsilon+2}$, which is the error in forecasting $y_{\Upsilon+2}$. Secondly, when only data through y_Υ are known, the forecast of the aggregate is calculated by combining the 1-step forecast (of $y_{\Upsilon+1}$) and the 2-step forecast (of $y_{\Upsilon+2}$). This will worsen predictability, as the three error terms on the bottom line of (71) all contribute to the forecast error. Thirdly, when only time-aggregated data are available, the forecasting model necessarily omits the terms in the third and fourth lines of (71), and so is inferior to aggregating the disaggregated forecasts. However, the aggregate forecasting model will not have the same parameters as in the first two rows even under correct-specification, because the omitted (biannual) variables will be correlated with the annual variables.

In keeping with our general approach in this chapter, suppose now that the break is not anticipated, but the in-sample parameters are known (to highlight the break effect). In this case, the disaggregated 1 and 2-step ahead forecasts are given by:

$$\begin{aligned} \widetilde{y}_{\Upsilon+1|\Upsilon} &= \phi_y + \rho \left(y_\Upsilon - \phi_y\right) + \beta' \left(\mathbf{z}_\Upsilon - \phi_z\right) \\ \widetilde{y}_{\Upsilon+2|\Upsilon} &= \phi_y + \rho^2 \left(y_\Upsilon - \phi_y\right) + \left(\rho\beta' + \beta'\mathbf{\Gamma}\right) \left(\mathbf{z}_\Upsilon - \phi_z\right), \end{aligned}$$

so the combined forecast of the aggregate is:

$$\begin{aligned} \widetilde{\overline{y}}_{T+1|T} &= \phi_y + \rho (1 + \rho) \left(\overline{y}_T - \phi_y\right) \\ &+ \left[(1 + \rho) \beta' + \beta'\mathbf{\Gamma}\right] \left(\overline{\mathbf{z}}_T - \phi_z\right) \\ &- \frac{1}{2}\rho (1 + \rho) \left(y_{\Upsilon-1} - \phi_y\right) \\ &- \frac{1}{2} \left[(1 + \rho) \beta' + \beta'\mathbf{\Gamma}\right] \left(\mathbf{z}_{\Upsilon-1} - \phi_z\right) \end{aligned} \tag{72}$$

The forecast using only the time-aggregated data is given by:

$$\widehat{\overline{y}}_{T+1|T} = \phi_y^a + \rho^a \left(\overline{y}_T - \phi_y^a\right) + (\beta^a)' \left(\overline{\mathbf{z}}_T - \phi_z^a\right) \tag{73}$$

where the superscript ^{a} in (73) denotes a population parameter value in the in-sample model:

$$\overline{y}_t = \phi_y^a + \rho^a \left(\overline{y}_{t-1} - \phi_y^a\right) + (\beta^a)' \left(\overline{\mathbf{z}}_{t-1} - \phi_z^a\right) + u_t, \quad t = 1, \ldots, T-1,$$

which delivers the best fit (in population).

Consider the forecast errors $\widetilde{\overline{u}}_{T+1|T} = \overline{y}_{T+1} - \widetilde{\overline{y}}_{T+1|T}$ and $\widehat{\overline{u}}_{T+1|T} = \overline{y}_{T+1} - \widehat{\overline{y}}_{T+1|T}$. For the former:

$$\begin{aligned}
\widetilde{\overline{u}}_{T+1|T} = {} & \phi_y^* - \phi_y \\
& + \left[\rho^* (1+\rho^*) \left(\overline{y}_T - \phi_y^*\right)\right] - \left[\rho (1+\rho) \left(\overline{y}_T - \phi_y\right)\right] \\
& + \left[(1+\rho^*) (\beta^*)' + (\beta^*)' \boldsymbol{\Gamma}^*\right] \left(\overline{\mathbf{z}}_T - \phi_z^*\right) - \left[(1+\rho) \beta' + \beta' \boldsymbol{\Gamma}\right] \left(\overline{\mathbf{z}}_T - \phi_z\right) \\
& - \frac{1}{2} \left[\rho^* (1+\rho^*) \left(y_{\Upsilon-1} - \phi_y^*\right) - \rho (1+\rho) \left(y_{\Upsilon-1} - \phi_y\right)\right] \\
& - \frac{1}{2} \left\{\left[(1+\rho^*) (\beta^*)' + (\beta^*)' \boldsymbol{\Gamma}^*\right] \left(\mathbf{z}_{\Upsilon-1} - \phi_z^*\right) - \left[(1+\rho) \beta' + \beta' \boldsymbol{\Gamma}\right] \left(\mathbf{z}_{\Upsilon-1} - \phi_z\right)\right\} \\
& + \frac{1}{2} \left[\epsilon_{\Upsilon+2} + (1+\rho^*) \epsilon_{\Upsilon+1} + (\beta^*)' \epsilon_{z,\Upsilon+1}\right]
\end{aligned} \tag{74}$$

We write the term $y_{\Upsilon-1} - \phi_y^*$ as:

$$y_{\Upsilon-1} - \phi_y^* = y_{\Upsilon-1} - \mathsf{E}\left[y_{\Upsilon-1}\right] - \left(\phi_y^* - \phi_y\right)$$

to isolate the effect of the equilibrium mean shift. Similarly, for $\mathbf{z}_{\Upsilon-1} - \phi_z^*$, substitute:

$$\mathbf{z}_{\Upsilon-1} - \phi_z^* = \mathbf{z}_{\Upsilon-1} - \mathsf{E}\left[\mathbf{z}_{\Upsilon-1}\right] - \left(\phi_z^* - \phi_z\right).$$

Since:

$$\mathsf{E}\left[y_\Upsilon\right] = \mu^* + \rho^* \mathsf{E}\left[y_{\Upsilon-1}\right] + (\beta^*)' \mathsf{E}\left[\mathbf{z}_{\Upsilon-1}\right] = \mu^* + \rho^* \phi_y + (\beta^*)' \phi_z \tag{75}$$

whereas $\mathsf{E}\left[y_{\Upsilon-1}\right] = \phi_y$, so that:

$$\mathsf{E}\left[\overline{y}_T\right] = \frac{1}{2} \mathsf{E}\left(y_\Upsilon + y_{\Upsilon-1}\right) = \frac{1}{2} \left(\mu^* + (1+\rho^*) \phi_y + (\beta^*)' \phi_z\right),$$

then the terms $\left(\overline{y}_T - \phi_y\right)$ and $\left(\overline{y}_T - \phi_y^*\right)$ are:

$$\begin{aligned}
\left(\overline{y}_T - \phi_y\right) &= \left[\left(\overline{y}_T - \mathsf{E}\left[\overline{y}_T\right]\right) + \left(\mathsf{E}\left[\overline{y}_T\right] - \phi_y\right)\right] \\
\left(\overline{y}_T - \phi_y^*\right) &= \left[\left(\overline{y}_T - \mathsf{E}\left[\overline{y}_T\right]\right) + \left(\mathsf{E}\left[\overline{y}_T\right] - \phi_y\right) + \left(\phi_y - \phi_y^*\right)\right].
\end{aligned}$$

We decompose the term in $\overline{\mathbf{z}}_T$ in the same way, then write $\widetilde{\overline{u}}_{T+1|T}$ as:

$$\begin{aligned}
\widetilde{\overline{u}}_{T+1|T} = {} & \left[1 - \tfrac{1}{2} \rho^* (1+\rho^*)\right] \left(\phi_y^* - \phi_y\right) & (1) \\
& + \left[\rho^* (1+\rho^*) - \rho (1+\rho)\right] \left(\overline{y}_T - \mathsf{E}\left[\overline{y}_T\right]\right) & (2) \\
& + \left[\rho^* (1+\rho^*) - \rho (1+\rho)\right] \left(\mathsf{E}\left[\overline{y}_T\right] - \phi_y\right) & (3) \\
& - \tfrac{1}{2} \left[(1+\rho^*) (\beta^*)' + (\beta^*)' \boldsymbol{\Gamma}^*\right] \left(\phi_z^* - \phi_z\right) & (4) \\
& + \left[(1+\rho^*) (\beta^*)' - (1+\rho) \beta' + (\beta^*)' \boldsymbol{\Gamma}^* - \beta' \boldsymbol{\Gamma}\right] \left(\overline{\mathbf{z}}_T - \mathsf{E}\left[\overline{\mathbf{z}}_T\right]\right) & (5) \\
& + \left[(1+\rho^*) (\beta^*)' - (1+\rho) \beta' + (\beta^*)' \boldsymbol{\Gamma}^* - \beta' \boldsymbol{\Gamma}\right] \left(\mathsf{E}\left[\overline{\mathbf{z}}_T\right] - \phi_z\right) & (6) \\
& - \tfrac{1}{2} \rho^* (1+\rho^*) \left[y_{\Upsilon-1} - \mathsf{E}\left[y_{\Upsilon-1}\right]\right] & (7) \\
& + \tfrac{1}{2} \rho (1+\rho) \left(y_{\Upsilon-1} - \phi_y\right) & (8) \\
& - \tfrac{1}{2} \left[(1+\rho^*) (\beta^*)' + (\beta^*)' \boldsymbol{\Gamma}^*\right] \left[\mathbf{z}_{\Upsilon-1} - \mathsf{E}\left[\mathbf{z}_{\Upsilon-1}\right]\right] & (9) \\
& + \tfrac{1}{2} \left[(1+\rho) \beta' + \beta' \boldsymbol{\Gamma}\right] \left(\mathbf{z}_{\Upsilon-1} - \phi_z\right) & (10) \\
& + \tfrac{1}{2} \left[\epsilon_{\Upsilon+2} + (1+\rho^*) \epsilon_{\Upsilon+1} + (\beta^*)' \epsilon_{z,\Upsilon+1}\right] & (11)
\end{aligned} \tag{76}$$

The first term is the change in the equilibrium mean of y; the second is the zero-mean slope change; the third is the non-zero mean slope change (because $\mathsf{E}[\overline{y}_T] \neq \phi_y$); terms (4) to (7) are the matching effects for $\mathbf{z}$; and (8)–(11) are all zero in expectation.

Consider the aggregated model forecast error, $\widehat{\overline{u}}_{T+1|T} = \overline{y}_{T+1} - \widehat{\overline{y}}_{T+1|T}$, where the forecast is given by (73). To further focus the analysis on breaks and time-disaggregation, we assume that $\beta = \mathbf{0}$ in the DGP, so that the $\mathbf{z}_\tau$ variables are absent. We can then derive the in-sample population values of the parameters of the aggregate model as $\rho^a = \frac{1}{2}\rho(1+\rho)$ and $\phi_y^a = \phi_y$ (see Appendix). Then the forecast error becomes:

$$\begin{aligned}\widehat{\overline{u}}_{T+1|T} &= \phi_y^* - \phi_y \\ &\quad + \rho^*(1+\rho^*)\left(\overline{y}_T - \phi_y^*\right) - \frac{1}{2}\rho(1+\rho)\left(\overline{y}_T - \phi_y\right) \\ &\quad - \frac{1}{2}\rho^*(1+\rho^*)\left(y_{\Upsilon-1} - \phi_y^*\right) \\ &\quad + \frac{1}{2}\left[\epsilon_{\Upsilon+2} + (1+\rho^*)\,\epsilon_{\Upsilon+1}\right] \end{aligned} \tag{77}$$

As before, expand the term $y_{\Upsilon-1} - \phi_y^*$ as $y_{\Upsilon-1} - \phi_y^* = \left(y_{\Upsilon-1} - \mathsf{E}[y_{\Upsilon-1}]\right) - \left(\phi_y^* - \phi_y\right)$, and substitute:

$$\begin{aligned}\left(\overline{y}_T - \phi_y^*\right) &= \left[\left(\overline{y}_T - \mathsf{E}[\overline{y}_T]\right) + \left(\mathsf{E}[\overline{y}_T] - \phi_y\right) + \left(\phi_y - \phi_y^*\right)\right] \\ \left(\overline{y}_T - \phi_y\right) &= \left[\left(\overline{y}_T - \mathsf{E}[\overline{y}_T]\right) + \left(\mathsf{E}[\overline{y}_T] - \phi_y\right)\right].\end{aligned}$$

Then:

$$\begin{aligned}\widehat{\overline{u}}_{T+1|T} &= \left[1 - \frac{1}{2}\rho^*(1+\rho^*) + \frac{1}{4}\rho(1+\rho)(1-\rho^*)\right]\left(\phi_y^* - \phi_y\right) \\ &\quad + \left[\rho^*(1+\rho^*) - \frac{1}{2}\rho(1+\rho)\right]\left(\overline{y}_T - \mathsf{E}[\overline{y}_T]\right) \\ &\quad + \left[\rho^*(1+\rho^*) - \rho(1+\rho)\right]\left(\mathsf{E}[\overline{y}_T] - \phi_y\right) \\ &\quad - \frac{1}{2}\rho^*(1+\rho^*)\left[y_{\Upsilon-1} - \mathsf{E}[y_{\Upsilon-1}]\right] \\ &\quad + \frac{1}{2}\left[\epsilon_{\Upsilon+2} + (1+\rho^*)\,\epsilon_{\Upsilon+1}\right] \end{aligned} \tag{78}$$

where we have used $\mathsf{E}[\overline{y}_T] - \phi_y = \frac{1}{2}(1-\rho^*)\left(\phi_y^* - \phi_y\right)$, added $\frac{1}{2}\rho(1+\rho)\left(\mathsf{E}[\overline{y}_T] - \phi_y\right)$ to the first row, and subtracted it from the third.

Comparing the aggregated model forecast error decomposition to that from the disaggregated model (excluding $\mathbf{z}$ from the DGP and models in both cases), the key finding is that the impact of the equilibrium-mean shift now differs between the two. However, the third term–the non-zero mean slope change term–is identical in both cases. Suppose $\rho = \rho^*$, then we can write the first term of the aggregated error as:

$$\left[1 - \frac{1}{2}\rho(1+\rho)\left(1 - \frac{1}{2}(1-\rho)\right)\right]\left(\phi_y^* - \phi_y\right)$$

which exceeds the mean-shift term in the error of the disaggregated model. Thus, the change in the break date relative to the forecast origin means that disaggregation can help relative to the time-aggregated model.

6.2 DDD taxonomy

The DDD forecast of $\overline{y}_{T+1}$ from T in terms of the aggregated data is:

$$\widehat{\overline{y}}_{T+1|T} = \overline{y}_T \tag{79}$$

with corresponding forecast error $\overline{u}_{T+1|T} = \overline{y}_{T+1} - \widehat{\overline{y}}_{T+1|T} = \Delta\overline{y}_{T+1} = \frac{1}{2}[(y_{\Upsilon+2} + y_{\Upsilon+1}) - (y_{\Upsilon} + y_{\Upsilon-1})]$. The forecast error will be biased, as only y_{Υ} will reflect in part the parameter change, but the bias of the DDD forecasts will diminish as the forecast origin moves past the break point and the mean of the forecast value approaches the post-break mean. The existence of time-disaggregation at the level of the DGP does not raise any new issues for the DDD compared to previous sections.

7 Pooling and breaks

Forecast combination, or pooling, is often found to be beneficial, as seen in the empirical evidence cited by, e.g., Clemen (1989) and Timmermann (2006): see also the review articles by Diebold and Lopez (1996), Newbold and Harvey (2002), Clements and Harvey (2009), and Aiolfi, Capistràn andTimmermann (this volume). The leading explanation as to why this might be the case is the portfolio diversification argument in the original paper by Bates and Granger (1969), and recently discussed by Granger and Jeon (2004) and Timmermann (2006), amongst others. When a number of forecasts are each based on partial, and incompletely overlapping, information sets, combination exploits the information from the different sources. An explanation stressed by Hendry and Clements (2004) is that of forecasts based on mis-specified models when there are structural breaks. Given the focus of our chapter, we begin this section by briefly reviewing Hendry and Clements (2004), who provide some analytical results for a simple scenario in which there is a one-off break affecting two forecasting models, both of which are differently mis-specified for the DGP pre-break. There is also an empirical literature that addresses the efficacy of pooling over large numbers of models, estimation samples and forecasting schemes, in the face of perceived instability and model mis-specification of unknown forms, and we briefly review some of those findings in section 7.2. Finally, we briefly review some of the literature on factor models in so far as it relates to breaks.

7.1 Structural breaks and mis-specified models

This section draws on Hendry and Clements (2004). There are a number of ways in which structural breaks may cause forecast combinations to deliver more accurate forecasts than the constituent models. As an illustration, we consider a shift in the mean value of one of the explanatory variables during the forecast period. Suppose we wish to forecast the scalar y_t which is determined by:

$$y_t = \beta_1'\mathbf{w}_t + \beta_2'\mathbf{z}_t + e_t \text{ where } e_t \sim \mathsf{IN}\left[0, \sigma_e^2\right] \tag{80}$$

and:

$$\begin{pmatrix} \mathbf{w}_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{IN}_n\left[\begin{pmatrix} \phi_{w,t} \\ \phi_{z,t} \end{pmatrix}, \begin{pmatrix} \mathbf{\Omega}_{ww} & \mathbf{\Omega}_{wz} \\ \mathbf{\Omega}_{wz} & \mathbf{\Omega}_{zz} \end{pmatrix}\right] \tag{81}$$

with $\phi_{w,t} = \phi_{z,t} = \mathbf{0}$ in-sample. The in-sample relationship between the two sets of regressors is:

$$\mathbf{z}_t = \psi + \mathbf{\Pi}_{zw}\mathbf{w}_t + \eta_{zw,t} \text{ where } \mathsf{E}\left[\eta_{zw,t}\right] = \mathbf{0},\ \mathsf{E}\left[\mathbf{w}_t\eta_{zw,t}'\right] = \mathbf{0} \tag{82}$$

which implies the population parameters $(\psi, \mathbf{\Pi}_{zw})$ satisfy:

$$\begin{pmatrix} \psi' \\ \mathbf{\Pi}'_{zw} \end{pmatrix} \simeq \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_{ww} \end{pmatrix}^{-1} \begin{pmatrix} \phi'_{z,t} \\ \mathbf{\Omega}_{wz} \end{pmatrix} = \begin{pmatrix} \phi'_{z,t} \\ \mathbf{\Omega}_{ww}^{-1}\mathbf{\Omega}_{wz} \end{pmatrix} \tag{83}$$

where $\phi_{z,t} = \mathbf{0}$ for $t = 1, 2, \ldots, T$, but out-of-sample $\phi_{z,T+1} \equiv \mu_z \neq \mathbf{0}$, so the mean of one of the sets of determinants of y_t ($\mathbf{z}_t$) shifts.

Here, the forecasting models are mis-specified, in that investigators fit the distinct separate equations:

$$y_t = a_0 + \mathbf{a}'_1\mathbf{w}_t + u_t \tag{84}$$

$$y_t = b_0 + \mathbf{b}'_1\mathbf{z}_t + v_t \tag{85}$$

The 1-step ahead forecasts and forecast errors from these two models when $\mathbf{w}_{T+1}$ and $\mathbf{z}_{T+1}$ are known, are for (84):

$$\widehat{y}_{T+1} = \widehat{a}_0 + \widehat{\mathbf{a}}'_1\mathbf{w}_{T+1},$$

with forecast error:

$$\widehat{u}_{T+1} = y_{T+1} - \widehat{y}_{T+1} = (\beta_1 - \widehat{\mathbf{a}}_1)'\,\mathbf{w}_{T+1} - \widehat{a}_0 + \beta'_2\mathbf{z}_{T+1} + e_{T+1} \tag{86}$$

and from (85):

$$\widetilde{y}_{T+1} = \widetilde{b}_0 + \widetilde{\mathbf{b}}'_1\mathbf{z}_{T+1},$$

$$\widetilde{v}_{T+1} = y_{T+1} - \widetilde{y}_{T+1} = \beta'_1\mathbf{w}_{T+1} + \left(\beta_2 - \widetilde{\mathbf{b}}_1\right)'\mathbf{z}_{T+1} - \widetilde{b}_0 + e_{T+1} \tag{87}$$

Suppose the models are such that (84) provides a better fit to the data, and therefore more accurate forecasts in the absence of any breaks, other things being equal.

From the estimation sample, prior to any shifts, and assuming least-squares estimates of in-sample parameters:

$$\mathsf{E}\left[\widehat{a}_0\right] = 0 \text{ and } \mathsf{E}\left[\widehat{\mathbf{a}}_1\right] = \beta_1 + \mathbf{\Pi}'_{zw}\beta_2,$$

where the expression for $\mathsf{E}\left[\widehat{\mathbf{a}}_1\right]$ reflects ‘omitted variable bias’ in (84). If we ignore the impact of $\mathsf{O}_p\left(T^{-1}\right)$ terms on the MSFEs, then we can replace the estimates in (86) and (87) by their expected values. Doing so for (86), and substituting $\mathbf{z}_{T+1} = \mu_z + \mathbf{\Pi}_{zw}\mathbf{w}_{T+1} + \eta_{zw,T+1}$, we obtain:

$$\widehat{u}_{T+1} \simeq \beta'_2\mu_z + \beta'_2\eta_{zx,T+1} + e_{T+1},$$

with unconditional MSFE of:

$$\mathsf{E}\left[\widehat{u}^2_{T+1}\right] \simeq \sigma^2_e + \beta'_2\left(\mathbf{\Omega}_{\eta_{zw}} + \mu_z\mu'_z\right)\beta_2,$$

which reflects the effect of model mis-specification ($\beta'_2\mathbf{\Omega}_{\eta_{zw}}\beta_2$) and the effect of the shift in the mean of the omitted variable ($\beta'_2\mu_z\mu'_z\beta_2$).

But a break may also be induced in (85) when $\mathbf{z}_{T+1}$ shifts because:

$$\mathbf{w}_{T+1} = \kappa + \mathbf{\Pi}_{wz}\mathbf{z}_{T+1} + \eta_{wz,T+1} \text{ where } \mathsf{E}\left[\eta_{wz,T+1}\right] = \mathbf{0} \text{ and } \mathsf{E}\left[\mathbf{z}_{T+1}\eta'_{wz,T+1}\right] = \mathbf{0},$$

so $\kappa = -\mathbf{\Pi}_{wz}\mu_z$, whereas $\mathbf{\Pi}'_{wz} = \mathbf{\Omega}^{-1}_{zz}\mathbf{\Omega}_{zw}$, leading to a forecast error of:

$$\widetilde{v}_{T+1} \simeq \beta'_1\eta_{wz,T+1} - \beta'_1\mathbf{\Pi}_{wz}\mu_z + e_{T+1}$$

from substituting $\widehat{b}_0 = 0$, $\widehat{\mathbf{b}}_1 = \beta_2 + \mathbf{\Pi}'_{wz}\beta_1$ and for $\mathbf{w}_{T+1}$, in (87).

Then the approximate MSFEs (ignoring parameter estimation uncertainty) are:

$$\text{Model (84): } \mathsf{E}\left[\widehat{u}^2_{T+1}\right] \simeq \sigma_e^2 + \beta_2'\mathbf{\Omega}_{\eta_{zw}}\beta_2 + \left(\beta_2'\mu_z\right)^2$$
$$\text{Model (85): } \mathsf{E}\left[\widetilde{v}^2_{T+1}\right] \simeq \sigma_e^2 + \beta_1'\mathbf{\Omega}_{\eta_{wz}}\beta_1 + \left(\beta_1'\mathbf{\Pi}_{wz}\mu_z\right)^2 .$$

Hence the relative forecast accuracy ranking may be overturned in favour of (85) to the extent that the mean shift (μ_z) is large and the correlation between $\mathbf{z}$ and $\mathbf{w}$ is low (in the limit $\mathbf{\Pi}_{wz} = \mathbf{0}$).

However, our interest is in relative performance of the combined forecast. Consider the simplest form of forecast combination, namely equal weighting. This average forecast is:

$$\widehat{\widetilde{y}}_{T+1} = \frac{1}{2}\left(\widehat{y}_{T+1} + \widetilde{y}_{T+1}\right),$$

with error:

$$\begin{aligned}\widehat{\widetilde{e}}_{T+1} &= \widehat{u}_{T+1} + \frac{1}{2}\left(\widetilde{v}_{T+1} - \widehat{u}_{T+1}\right) \\ &= \frac{1}{2}\left(\beta_2' - \beta_1'\mathbf{\Pi}_{wz}\right)\mu_z + \frac{1}{2}\left(\beta_1'\eta_{wz,T+1} + \beta_2'\eta_{zw,T+1}\right) + e_{T+1},\end{aligned}$$

and approximate MSFE for $\mathsf{E}[\widehat{\widetilde{e}}^2_{T+1}]$ given by:

$$\sigma_e^2 + 0.25\left[\beta_1'\mathbf{\Omega}_{\eta_{wz}}\beta_1 + \beta_2'\mathbf{\Omega}_{\eta_{zw}}\beta_2 - 2\beta_1'\mathbf{\Omega}_{zw}\left(\mathbf{I}_{n_1} - \mathbf{\Pi}'_{zw}\mathbf{\Pi}'_{wz}\right)\beta_2 + \left[\left(\beta_2' - \beta_1'\mathbf{\Pi}_{wz}\right)\mu_z\right]^2\right] \quad (88)$$

Examination of (88) indicates that the combined forecast could beat both individual forecasts, and that rankings will depend on the size of the unmodelled shift in the $\mathbf{z}$ process, the error variances, and the correlations. Hendry and Clements (2004) show that sharper predictions can be drawn if, for example, each forecasting model contains a single explanatory variable.

The benefits of pooling in the above example derive from diversification, similar to portfolio theory: the idiosyncratic risk associated with each model is averaged when the 'shocks' are relatively uncorrelated. And like that theory, averaging fails when 'shocks' are highly correlated, as would be the case in our example if there were a location shift in the DGP (80), or in other variables, $\mathbf{x}_t$, unknowingly omitted. That last scenario may suggest averaging over many variables, so all relevant variables are included. Unfortunately, that can also be a bad idea: for 9 candidate variables, there are 10^9 models, but if only 4 variables matter, there are so many 'bad' models that they dilute the relevant, even if only a small weight is attached to each. Hendry and Reade (2008) show that Bayesian model averaging (BMA) does badly in such a setting, especially when there are large outliers, and that arbitrary rules like Occam's window do not correct that problem. Rather, it seems important to select the relevant variables, or at least eliminate many of the irrelevant. There are few useful analytical results on averaging versus selection with mis-specified models confronting breaks, as most studies consider small models with constant parameters where almost all the variables matter, and hence show that selection is dominated.

Estimation of (80) should dominate averaging the two sub-models in the above illustration, although, while it is an 'external break', the change in the mean of $\mathbf{z}_{T+1}$ would affect the accuracy of forecasts from an estimated version of (80) due to the change in collinearity when $\mathbf{\Pi}_{zw} \neq \mathbf{0}$ (see e.g., Castle *et al.*, 2010). In practice, general models tend not to outperform, suggesting that other factors are important, particularly the robustness of the various models to breaks. Indeed, a careful choice of the set to average over seems the most essential consideration, and should not only include the various 'structural' models that have been developed, but also their differences, intercept-corrected variants and other robust devices (see Hendry, 2004, for such an averaging application in a policy context).

7.2 Empirical comparisons

We focus on a relatively small number of recent empirical studies. The review articles cited in the introduction to this section cover much of the literature up to this point. As well as combining forecasts from models with different explanatory variables, a characteristic of the more recent literature is the combination of forecasting models which are essentially based on the same explanatory variables but differ in regard to the specification of the model and/or the estimation sample.

Clark and McCracken (2009a) consider VAR models of output, inflation and interest rates for the US. Because such models appear to be unstable, they calculate averages of individual VAR model forecasts, where the individual VAR models are designed to counter various forms of instability, and include: estimation using rolling windows of data; with intercept corrections (see Clements and Hendry, 1996); Bayesian VARs; time-varying parameters; and discounted least squares, as in Clark and McCracken (2008). The argument is that the form of instability in the output, inflation and interest rate DGP is unknown, while the efficacy of the different individual methods of dealing with the instability will depend on the form of instability. For example, slowly-changing parameter values may be picked up by the use of a rolling estimation window or discounted least squares, say, whereas a large structural break may favour methods such as intercept corrections. Hence averaging the forecasts from all these methods may produce forecasts that fare well, by and large, and this is their main finding.

Pesaran, Schuerman and Smith (2009) assess forecast combination in the context of the global vector autoregressive (GVAR) model of Dees, di Mauro, Pesaran and Smith (2007).[6] They consider combinations of forecasts from a range of different GVAR models (in part differentiated by the degree to which theory-based long-run restrictions are imposed, as well as different lag orders), as well as combinations of forecasts from the same specification estimated over different sample periods. Hence there is pooling in two dimensions–across models and estimation windows–giving rise to 'double-averaged' forecasts. They find these forecasts fare well for a number of variables, including the two key macro variables of output growth and inflation. Clark and McCracken (2009b) also consider pooling across estimation samples, specifically, they present analytical evidence pointing to the value of combining recursive and rolling forecasts (i.e., forecasts produced from models estimated on samples of increasing size and from models estimated on windows of data of fixed size) when there is instability.

There is also evidence that the more traditional combination of forecasts from models with different explanatory variables is beneficial when there are breaks. An example is Stock and Watson (2003), who show that the leading indicators that might have been expected to help predict the 2001 recession in the US (based on past performance) generally failed to do so (e.g., building permits and consumer confidence), while other indicators did better than expected (namely, stock market indicators). This general instability implicit in the failure of individual indicators to consistently 'lead' output suggest instead it might be advantageous to combine the forecasts from the individual-indicator models. For the period 1999Q1 to 2002Q3, Stock and Watson (2003) find some modest improvements of the order of 5% on MSFE from forecast combination relative to a benchmark autoregression in output growth. Clements and Galvão (2009a) revisit this issue using real-time data and exploiting the availability of monthly observations on the indicator variables (see also Clements and Galvão, 2008).

[6] See the comments and response published in the same issue of the *International Journal of Forecasting*.

7.3 Factor models

Dynamic factor models are becoming increasingly popular in economic forecasting: see, e.g., Stock and Watson (1999), Forni, Hallin, Lippi and Reichlin (2000) and Stock and Watson (this volume). Factor models are a particular way of 'pooling information' and contrast the combination of forecasts. Of interest is whether there are reasons to expect factor model forecasts to be more or less accurate than other approaches to forecasting when there are instabilities in relationships between variables. A key reference in this regard is Stock and Watson (2009), who also discuss the limited work on factor models and instability. Their main finding is that it might be possible to obtain good estimates of the factors themselves even when the relationships between the individual series and factors (the factor loadings) are unstable. This suggests that factor models are more likely to fail because of instability in the coefficients of forecasting models that use the factors, rather than because the factors themselves are poorly estimated. In the empirical exercise they report covering the period 1959–2006, with a putative break in 1984, this appears to be borne out as more accurate forecasts are obtained using the full-sample to estimate the factors but only the post-break sample to estimate the coefficients of the forecasting model.

Stock and Watson (2009) suggest that the reasons given by Hendry and Clements (2004) as to why pooling works when individual forecasting models are subject to location shifts might be essentially the explanation as to why factors can be reasonably accurately estimated even when the factor loadings are non-constant. Estimating the factors by using many series may provide reasonably accurate estimates because the instabilities 'average out'.

8 Data vintages and location shifts

The forecast-error taxonomies of the earlier sections in this chapter suggest data mis-measurement as a possible source of forecast error. The literature suggests that measurement errors and especially different data vintages may have pervasive effects on empirical research (see, *inter alia*, Hendry, 1994, and Croushore and Stark, 2003). There are issues to do with the validity of 'pseudo out-of-sample' forecast comparisons that use vintages of data that would not have been available to the forecaster at the time the forecast was made (see, e.g., Diebold and Rudebusch, 1991b, 1991a, Robertson and Tallman, 1998, Orphanides, 2001, Croushore and Stark, 2001, 2003, Faust, Rogers and Wright, 2003, and Orphanides and van Norden, 2005). A number of ways of accounting for data revisions when forecasting have been proposed, including the use of state-space models (e.g., Harvey, McKenzie, Blake and Desai, 1983, Howrey, 1984, Jacobs and van Norden, 2007, and Cunningham, Eklund, Jeffery, Kapetanios and Labhard, 2007), as well as the use of an early estimate of each observation, as in Koenig, Dolmas and Piger (2003) and Clements and Galvão (2008, 2009a, 2009b).

Although data are typically subject to various rounds of 'regular' revisions, as well as infrequent benchmark revisions, so that at any forecast origin the available data set will include a first estimate of the most recent data observation, first and second estimates of the previous observation, and so on to multiple estimates of the earlier observations, our exposition allows for just two vintages.[7] Following Patterson (2003), we show that estimating the forecasting model on 'fully-revised' data, and then generating forecasts from first-estimates of the latest data, will show up as an equilibrium mean shift when the revisions (between the first estimates and fully-revised data) are non-zero mean for any of the variables in the cointegrating

[7]For example, the national income and product account data of the U.S. Bureau of Economic Analysis are subject to a complicated revision process that includes three annual revisions in the July of each year (see Fixler and Grimm, 2005, 2008, and Landefeld, Seskin and Fraumeni, 2008).

relationship. To bring out the impact of data revisions, we also abstract from other sources of forecast error. Thus, we assume that the DGP is the VEqCM given by (18), reproduced here as:

$$\Delta \mathbf{x}_t^{(2)} = \tau + \alpha\beta' \mathbf{x}_{t-1}^{(2)} + \epsilon_t, \quad t = 1, \ldots, T-1 \tag{89}$$

where the superscript denotes the fully-revised data. We assume that the forecasting model matches the DGP, that its parameters are known, and that there are no breaks. The error in forecasting period $T+1$ based on information up to T would simply be the innovation error ϵ_{T+1} *if* the forecast was conditioned on $\mathbf{x}_T^{(2)}$, namely:

$$\Delta \mathbf{x}_{T+1|T} = \tau + \alpha\beta' \mathbf{x}_T^{(2)} \tag{90}$$

Suppose instead only an early estimate of $\mathbf{x}_T$ is available, $\mathbf{x}_T^{(1)}$, so that the forecast is of necessity conditioned on this vintage of the data. Suppose that $\mathbf{x}_T^{(2)} = \mu^{(2)} + \mathbf{x}_T^{(1)} + \eta_T$. When $\mu^{(2)} = \mathbf{0}$ and $\mathsf{E}[\eta_T \mathbf{x}_T^{(1)}] = \mathbf{0}$, then the first estimates are efficient for the later vintage, in that the revisions are unpredictable and the later vintage adds news. When $\mu^{(2)} \neq \mathbf{0}$ then the first release is unconditionally biased for the later release. Some early findings of Mankiw and Shapiro (1986) suggested US output growth revisions were largely news, whilst more recent research by Aruoba (2008), Corradi, Fernandez and Swanson (2009) (who allow for a non-linear relationship when they test the rationality of the early release) and Clements and Galvão (2010), has been more mixed.

We write $\tau = \gamma - \alpha\mu$, as before, then conditioning the forecast (90) on $\mathbf{x}_T^{(1)}$ results in:

$$\begin{aligned} \Delta \mathbf{x}_{T+1|T,(1)} &= \gamma + \alpha\beta' \left(\mathbf{x}_T^{(2)} - \mu \right) - \alpha\beta' \mu^{(2)} - \alpha\beta' \eta_T \\ &= \Delta \mathbf{x}_{T+1|T} - \alpha\beta' \mu^{(2)} - \alpha\beta' \eta_T. \end{aligned} \tag{91}$$

In general, the mean of the forecast $\Delta \mathbf{x}_{T+1|T,(1)}$ has shifted relative to (90), which by construction is unbiased. This will only not be the case when either (i) data revisions are zero mean, or (ii) $\beta' \mu^{(2)} = \mathbf{0}$ so that β is contemporaneous mean co-breaking: see Hendry and Massmann (2007). In any event, the stochastic component of the revision ($\alpha\beta' \eta_T$) will inflate the forecast-error variance relative to when the forecast is conditioned on the final data. Hence, even though there may be a constant-parameter LDGP for the final data (as assumed here), the use of earlier release data to generate forecasts may induce a location shift in the forecasts relative to the actuals. This adds to the likely benefits from improved initial estimates (nowcasts), especially when location shifts occur in some of the disaggregates (see e.g., Castle *et al.*, 2009, and Castle and Hendry, 2010).

9 Conclusion

Economic forecasting is inevitably carried out using (unknowingly) mis-specified models, in the face of unanticipated location shifts. In Clements and Hendry (1998, 1999), we investigated the consequences for forecast inaccuracy, and showed that much of the existing empirical evidence on forecast performance could be accounted for by a theory which allowed for both features, using parameter estimates based on mis-measured data. Here, we have summarized approaches to ameliorating the adverse effects of breaks when models may be incorrectly specified, and considered robustifying forecasts by additional differencing, disaggregation over variables and time, pooling of forecasts–including factor methods–and the impacts of changing data vintages.

We explained why additional differencing confers substantial benefits when location shifts occur, without excessive costs when they do not, while retaining the policy-relevant parameters. We showed that disaggregation by variable and by time may be fruitful once the break has already occurred. Specifically, disaggregating variables does not alter the impacts of a break at the forecast origin, but differentially affects aggregate forecasts and forecasts of disaggregates one period later. This suggests that in some circumstances it may pay to forecast the disaggregate variables, and then aggregate those forecasts to obtain a forecast of the disaggregate, rather than forecasting the aggregate variable directly using only aggregate information. Similarly, time disaggregation does not ameliorate the effects of breaks that occur at the forecast origin, but we were able to show that one (disaggregated) time period later, the forecasts from the time-disaggregated model will be less adversely affected by breaks than direct forecasts of the aggregate. Pooling can help reduce the effects of breaks, and factor methods can be seen as an exemplar.

Overall, unanticipated breaks remain a major problem for forecasters. The chapter by Castle, Hendry and Fawcett considers a number of related issues. These include: can breaks be forecasted, and if not, are there methods of improving forecasts during breaks?

10 Appendix

Here, we derive the in-sample population parameter values of the aggregate model of section 6.1 for the process given by equation (64) with $\beta = \mathbf{0}$. The aggregate model is defined by:

$$\overline{y}_t = \phi_y^a + \rho^a \left(\overline{y}_{t-1} - \phi_y^a\right) + u_t \quad t = 1, \ldots, T-1.$$

The population value of ρ^a is given by $\rho^a = Cov\left(\overline{y}_t, \overline{y}_{t-1}\right)/Var\left(\overline{y}_t\right)$. From:

$$Cov\left(\overline{y}_t, \overline{y}_{t-1}\right) = cov\left(\frac{y_\tau + y_{\tau-1}}{2}, \frac{y_{\tau-2} + y_{\tau-3}}{2}\right) = \frac{1}{2}\left\{2\mathsf{E}\left[y_\tau y_{\tau-2}\right] + \mathsf{E}\left[y_\tau y_{\tau-1}\right] + \mathsf{E}\left[y_\tau y_{\tau-3}\right]\right\} - \phi_y^2$$

where $\mathsf{E}\left[y_\tau y_{\tau-i}\right] = \rho^i \sigma_u^2\left(1-\rho^2\right) + \phi_y^2$,$i = 1, 2, \ldots$, it follows that $\rho^a = \frac{1}{2}\rho\left(1+\rho\right)$, and $\phi_y^a = \mathsf{E}\left[\frac{y_\tau + y_{\tau-1}}{2}\right] = \phi_y$.

References

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

Aruoba, S. B. (2008). Data revisions are not well-behaved. *Journal of Money, Credit and Banking*, **40**, 319–340.

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

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

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.

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

Bjørnstad, J. F. (1990). Predictive likelihood: A review. *Statistical Science*, **5**, 242–265.

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

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

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*, **forthcoming**, –.

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). Nowcasting from disaggregates in the face of location shifts. *Journal of Forecasting*, **forthcoming**, –.

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

Christ, C. F. (1966). *Econometric Models and Methods*. New York: John Wiley.

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

Clark, T. E., and McCracken, M. W. (2008). Forecasting with small macroeconomic VARs in the presence of instabilities. 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. 93–147: Emerald.

Clark, T. E., and McCracken, M. W. (2009a). Averaging forecasts from VARs with uncertain instabilities. *Journal of Applied Econometrics*. Forthcoming.

Clark, T. E., and McCracken, M. W. (2009b). Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review*. Forthcoming.

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

Clements, M. P., and Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business and Economic Statistics*, **26**, 546–554. No. 4.

Clements, M. P., and Galvão, A. B. (2009a). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics*, **24**, 1187–1206.

Clements, M. P., and Galvão, A. B. (2009b). Real-time forecasting of inflation and output growth in the presence of data revisions. Discussion paper, Department of Economics, University of Warwick.

Clements, M. P., and Galvão, A. B. (2010). First announcements and real economic activity. *European Economic Review*. Forthcoming.

Clements, M. P., and Harvey, D. I. (2009). Forecasting combination and encompassing. In Mills, T. C., and Patterson, K. (eds.), *Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics*, pp. 169–198: Palgrave MacMillan.

Clements, M. P., and Hendry, D. F. (1993). On the limitations of comparing mean squared forecast errors (with discussion). *Journal of Forecasting*, **12**, 617–637. Reprinted in T.C. Mills (ed.), *Economic Forecasting*. Edward Elgar, 1999.

Clements, M. P., and Hendry, D. F. (1995). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, **10**, 127–146.

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

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

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

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

Clements, M. P., and Hendry, D. F. (2002b). Modelling methodology and forecast failure. *The Econometrics Journal*, **5**, 319–344.

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. (2008). Forecasting annual UK inflation using an econometric model over 1875–1991. 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. 3–39: Emerald.

Corradi, V., Fernandez, A., and Swanson, N. R. (2009). Information in the revision process of real-time datasets. *Journal of Business and Economic Statistics*, **27**, 455–467.

Croushore, D., and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, **105**, 111–130.

Croushore, D., and Stark, T. (2003). A real-time data set for macroeconomists: Does the data vintage matter?. *The Review of Economics and Statistics*, **85**, 605–617.

Cunningham, A., Eklund, J., Jeffery, C., Kapetanios, G., and Labhard, V. (2007). A state space approach to extracting the signal from uncertain data. Working paper no. 336, The Bank of England.

Dees, S., di Mauro, F., Pesaran, M. H., and Smith, L. V. (2007). Exploring the international linkages in the EURO area: A global VAR analysis. *Journal of Applied Econometrics*, **22**, 1–38.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts: With applications to financial risk management. *International Economic Review*, **39**, 863–883.

Diebold, F. X., and Lopez, J. A. (1996). Forecast evaluation and combination. In Maddala, G. S., and Rao, C. R. (eds.), *Handbook of Statistics*, Vol. 14, pp. 241–268: Amsterdam: North–Holland.

Diebold, F. X., and Rudebusch, G. D. (1991a). Forecasting output with the composite leading index: A real-time analysis. *Journal of the American Statistical Association*, **86**, 603–610.

Diebold, F. X., and Rudebusch, G. D. (1991b). Turning point prediction with the composite leading index: An ex ante analysis. In Lahiri, K., and Moore, G. H. (eds.), *Leading economic indicators: New* *approaches and forecasting records*, pp. 231–256. Cambridge: Cambridge University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.

Engle, R. F., and Bollerslev, T. (1987). Modelling the persistence of conditional variances. *Econometric Reviews*, **5**, 1–50.

Engle, R. F., and Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, **55**, 251–276.

Engle, R. F., and Yoo, B. S. (1987). Forecasting and testing in co-integrated systems. *Journal of Econometrics*, **35**, 143–159.

Faust, J., Rogers, J. H., and Wright, J. H. (2003). Exchange rate forecasting: The errors we've really made. *Journal of International Economic Review*, **60**, 35–39.

Fixler, D. J., and Grimm, B. T. (2005). Reliability of the nipa estimates of U.S. economic activity. *Survey of Current Business*, **85**, 9–19.

Fixler, D. J., and Grimm, B. T. (2008). The reliability of the GDP and GDI estimates. *Survey of Current Business*, **88**, 16–32.

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

Giacomini, R., and Rossi, B. (2009). Detecting and predicting forecast breakdown. *Review of Economic Studies*, **76**, 669–705.

Giacomini, R., and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, **74**, 1545 – 1578.

Granger, C. W. J., and Jeon, Y. (2004). Thick modeling. *Economic Modelling*, **21**, 323–343.

Granger, C. W. J., and Newbold, P. (1986). *Forecasting Economic Time Series*, 2nd edn. New York: Academic Press.

Harvey, A. C., McKenzie, C. R., Blake, D. P. C., and Desai, M. J. (1983). Irregular data revisions. In Zellner, A. (ed.), *Applied Time Series Analysis of Economic Data*. Washington, DC: Bureau of the Census.

Hendry, D. F. (1974). Stochastic specification in an aggregate demand model of the United Kingdom. *Econometrica*, **42**, 559–578. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.

Hendry, D. F. (1979). Predictive failure and econometric modelling in macro-economics: The transactions demand for money. In Ormerod, P. (ed.), *Economic Modelling*, pp. 217–242. London: Heinemann. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000; and in J. Campos, N.R. Ericsson and D.F. Hendry (eds.), *General to Specific Modelling*. Edward Elgar, 2005.

Hendry, D. F. (1994). HUS revisited. *Oxford Review of Economic Policy*, **10**, 86–106.

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

Hendry, D. F. (1996). On the constancy of time-series econometric equations. *Economic and Social Review*, **27**, 401–422.

Hendry, D. F. (2000). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics*, **11**, 45–65. Reprinted in *The Economics of Structural Change*, Hagemann, H. Landesman, M. and Scazzieri, R. (eds.), Edward Elgar, Cheltenham, 2002.

Hendry, D. F. (2004). Forecasting long-run TV advertising expenditure in the UK. Commissioned report, Ofcom, London. http://www.ofcom.org.uk/research/tv/reports/tvadvmarket.pdf.

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

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

Hendry, D. F., 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 Massmann, M. (2007). Co-breaking: Recent advances and a synopsis of the literature. *Journal of Business and Economic Statistics*, **25**, 33–51.

Hendry, D. F., and Reade, J. J. (2008). Modelling and forecasting using model averaging. Working paper, Economics Department, Oxford University.

Howrey, E. P. (1984). Data revisions, reconstruction and prediction: an application to inventory investment. *The Review of Economics and Statistics*, **66**, 386–393.

Jacobs, J. P. A. M., and van Norden, S. (2007). Modeling data revisions: Measurement error and dynamics of 'true' values. Technical report cref 07-09, HEC, Montreal.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254. Reprinted in R.F. Engle and C.W.J. Granger (eds), *Long-Run Economic Relationships*, Oxford: Oxford University Press, 1991, 131–52.

Johansen, S. (1992). A representation of vector autoregressive processes integrated of order 2. *Econometric Theory*, **8**, 188–202.

Johansen, S., and Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, J. L., and Shephard, N. (eds.), *The Methodology and Practice of Econometrics*, pp. 1–36. Oxford: Oxford University Press.

Katz, R. W., and Murphy, A. H. (eds.)(1997). *Economic Value of Weather and Climate Forecasts*. Cambridge: Cambridge University Press.

Koenig, E. F., Dolmas, S., and Piger, J. (2003). The use and abuse of real-time data in economic forecasting. *The Review of Economics and Statistics*, **85(3)**, 618–628.

Landefeld, J. S., Seskin, E. P., and Fraumeni, B. M. (2008). Taking the pulse of the economy. *Journal of Economic Perspectives*, **22**, 193–216.

Leitch, G., and Tanner, J. E. (1991). Economic forecast evaluation: profits versus the conventional error measures. *American Economic Review*, **81**, 580–590.

Leitch, G., and Tanner, J. E. (1995). Professional economic forecasts: are they worth their costs?. *Journal of Forecasting*, **14**, 143–157.

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

Lütkepohl, H. (2006). Forecasting with VARMA models. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 288–325: Elsevier, Horth-Holland.

Mankiw, N. G., and Shapiro, M. D. (1986). News or noise: An analysis of GNP revisions. *Survey of Current Business (May 1986), US Department of Commerce, Bureau of Economic Analysis*, 20–25.

Nelson, C. R., and Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics*, **10**, 139–162.

Newbold, P., and Harvey, D. I. (2002). Forecasting combination and encompassing. In Clements, M. P., and Hendry, D. F. (eds.), *A Companion to Economic Forecasting*, pp. 268–283: Oxford: Blackwells.

Newbold, P., Harvey, D. I., and Leybourne (1999). Ranking competing multi-step forecasts. In Engle, R. F., and White, H. (eds.), *Cointegration, Forecasting and Causality*. Oxford: Oxford University Press.

Orphanides, A. (2001). Monetary policy rules based on real-time data. *American Economic Review*, **91(4)**, 964–985.

Orphanides, A., and van Norden, S. (2005). The reliability of inflation forecasts based on output gaps in real time. *Journal of Money, Credit and Banking*, **37**, 583–601.

Patterson, K. D. (2003). Exploiting information in vintages of time-series data. *International Journal of Forecasting*, **19**, 177–197.

Pesaran, M. H., Schuerman, T., and Smith, L. V. (2009). Forecasting economic and financial variables with global VARs. *International Journal of Forecasting*, **25**, 642–675.

Robertson, J. C., and Tallman, E. W. (1998). Data vintages and measuring forecast model performance. *Federal Reserve Bank of Atlanta Economic Review, Fourth Quarter*, 4–20.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In Cox, D. R., Hinkley, D. V., and Barndorff-Nielsen, O. E. (eds.), *Time Series Models: In Econometrics, Finance and other Fields*, pp. 1–67. London: Chapman and Hall.

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.

Stock, J. H., and Watson, M. W. (2003). How Did Leading Indicator Forecasts Perform During the 2001 Recession. *Federal Reserve Bank of Richmond, Economic Quarterly*, **89/3**, 71–90.

Stock, J. H., and Watson, M. W. (2009). Forecasting in dynamic factor models subject to structural instability. In Castle, J. L., and Shephard, N. (eds.), *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry. Chapter 7*, pp. 173–205. Oxford: Oxford University Press.

Timmermann, A. (2006). Forecast combinations. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 135–196: Elsevier, Horth-Holland.

Varian, H. R. (1975). A Bayesian approach to real estate assessment. In Fienberg, S. E., and Zellner, A. (eds.), *Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage*, pp. 195–208. Amsterdam: North Holland.