

The Econometrics of Macroeconomic Forecasting

David F. Hendry*
Nuffield College, Oxford.

March 30, 2008

Abstract

When an econometric model coincides with the mechanism generating the data in an unchanging world, the theory of economic forecasting is reasonably well developed. However, less is known about forecasting when model and mechanism differ in a non-stationary and changing world. The paper addresses the basic concepts; the invariance of forecast accuracy measures to isomorphic model representations; the roles of causal information, parsimony and collinearity; a reformulated taxonomy of forecast errors; differencing and intercept corrections to robustify forecasts against biases due to shifts in deterministic factors; the removal of structural breaks by co-breaking; and forecasting using leading indicators.

1 Introduction

The theory of economic forecasting is reasonably well developed assuming the econometric model coincides with the mechanism generating the data in a (difference) stationary world: see, for example, Klein (1971) and Granger and Newbold (1986). Consider an n -dimensional stochastic process \mathbf{x}_t with density $D_{\mathbf{x}_t}(\mathbf{x}_t | \mathbf{X}_{t-1}, \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$, which is a function of past information $\mathbf{X}_{t-1} = (\dots \mathbf{x}_1 \dots \mathbf{x}_{t-1})$. A statistical forecast $\tilde{\mathbf{x}}_{T+h}$ for period $T + h$, conditional on information up to period T is given by $\tilde{\mathbf{x}}_{T+h} = \mathbf{f}_T(\mathbf{X}_T)$, where $\mathbf{f}_T(\cdot)$ reflects that a prior estimate of $\boldsymbol{\theta}$ may be needed. Forecasts calculated as the conditional expectation $\hat{\mathbf{x}}_{T+h} = E[\mathbf{x}_{T+h} | \mathbf{X}_T]$ are unbiased, and no other predictor conditional on only \mathbf{X}_T has a smaller mean-square forecast error (MSFE) matrix:

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

However, when the model is mis-specified for the mechanism in an unknown way, and requires estimation from available data, less is known about forecasting, particularly in a non-stationary economy subject to unanticipated structural breaks. In such a setting, not only is it extremely difficult to correctly model the underlying processes, the costs of failing to do so are large. Nevertheless, despite the lack of strong and specific assumptions, many useful insights can be derived, albeit usually articulated in special cases. Consequently, we consider some results that can be established, extending research reported in Clements and Hendry (1993, 1994, 1995a, 1995b, 1996b) and Hendry and Clements (1994a, 1994b).

There are many ways of making economic forecasts, including guessing; ‘informal models’; extrapolation; leading indicators; surveys; time-series models; and econometric systems. By focusing on statistical forecasting, we will not be concerned with the first three here, but section 11 discusses

*Royal Economic Society Presidential Address, 1993. Financial support from the ESRC under grants R000233447 and L11625015 is gratefully acknowledged. This paper is based on recent publications with Michael P. Clements. Thanks are due to Anindya Banerjee, Mike Clements, Jurgen Doornik, Rob Engle, Hans-Martin Krolzig, Grayham Mizon, Bent Nielsen, and Neil Shephard for many helpful discussions about the topics herein.

leading indicators based on Emerson and Hendry (1994). Scalar time-series models include Kalman (1960) and Box and Jenkins (1976), with the latter's autoregressive integrated moving-average models (ARIMAs) being a dominant class, based on the Wold decomposition theorem (Wold, 1938: any purely non-deterministic stationary time series can be expressed as an infinite moving average; see Cox and Miller, 1965, p.286–8, for a lucid discussion). Also see Harvey and Shephard (1992). The most common multivariate time-series form is the vector autoregression (VAR): see e.g., Doan, Litterman and Sims (1984). However, economic forecasting based on econometric models of multivariate time-series will be our primary focus, since such systems consolidate empirical and theoretical knowledge of how economies function, provide a framework for a progressive research strategy, and help explain their own failures as well as provide forecasts.

The success of econometric model-based forecasts depends upon:

- (a) there being regularities to be captured;
- (b) such regularities being informative about the future;
- (c) the proposed method capturing those regularities; and:
- (d) excluding non-regularities that swamp the regularities.

The first two are characteristics of the economic system; the last two of the forecasting method. The history of economic forecasting in the UK suggests that there are some regularities informative about future events, but also major irregularities as well (see e.g., Burns, 1986, Wallis, 1989, Pain and Britton, 1992, and Cook, 1995). The dynamic integrated systems with intermittent structural breaks that are formalized below seem consistent with such evidence. However, achieving (c) without suffering from (d) is difficult, and motivates the conceptual structure proposed below, as well as the emphasis on issues such as parsimony and collinearity, and the re-examination of the role of causal information when forecasting models are mis-specified. Several results transpire to be misleading once model mis-specification interacts with non-stationary data (denoting thereby the general sense of processes whose first two moments are not constant over time). Conversely, it becomes feasible to account for the empirical success of procedures that difference data, or use intercept corrections (see e.g., Theil, 1961, Klein, 1971, and Wallis and Whitley, 1991), although these methods have no rationale when models are correctly specified. Potential improvements also merit investigation, so co-breaking is considered, and shown to clarify some problems experienced with leading indicators.

The structure of the paper is as follows. The first three sections discuss the background concepts, models, and measures. The next three sections deduce their implications for causal information, parsimony, and collinearity, in an attempt to account for recent forecast failures, but only the first seems to be important. To resolve that problem, a reformulated taxonomy of forecast errors highlights forecast biases as due to unmodelled shifts in deterministic factors, providing a rationale for both differencing and intercept corrections to offset breaks. The alternative of removing regime shifts by co-breaking is proposed and applied to forecasting using leading indicators.

In more detail, section 2 considers the basic concepts needed to develop the analysis of forecasting when the model and mechanism differ. The notions of (un)predictable and forecastable are discussed: despite their close usage, the former is a property whereas forecasting is a procedure. Next, §3 describes the framework for economic forecasting using a vector autoregression in integrated-cointegrated variables. The issue of measuring forecast accuracy is described in §4, since evaluation may be dependent on the transformations examined for vector processes, or when forecasting more than 1-step ahead; that section focuses on the invariance, or otherwise, of putative measures to isomorphic representations of the model.

Then §5 discusses the role of causal information in economic forecasting, and shows that non-causal

variables may outperform in forecasting when the model and mechanism differ in a world subject to structural breaks. An example highlights the potential importance of excluding irrelevant, but changing, effects. The possible role of parsimony in h -step ahead forecasting is discussed in §6, but only a small effect is found for the constant-parameter cases examined. Section 7 shows that collinearity has little effect on forecast accuracy in constant worlds, but a larger effect if the collinearity alters, so parsimony may have a justification in non-constant processes.

Based on the analyses in Hendry and Clements (1994b) and Clements and Hendry (1996c), §8 shows that intercept corrections (non-zero values for a model's error terms over the forecast period) can help robustify forecasts against biases due to structural breaks. The taxonomy of forecast errors in §9 allows for structural change in the forecast period, the model and DGP to differ over the sample period, the parameters of the model to be estimated from the data, and the forecasts to commence from incorrect initial conditions. This re-emphasizes the possible role of non-causal variables, and warns of the potential dangers of selecting policy models by forecast-accuracy criteria. It also demonstrates the central role of shifts in deterministic factors, confirming the apt naming of intercept corrections, and the potential efficacy of differencing. The differential impact of structural breaks on models with and without cointegration feedbacks highlights that the latter equilibrium correct, but do not error correct between equilibria.

Next, §10 considers the removal of regime-shift non-stationarity by co-breaking, namely the cancellation of breaks across linear combinations of variables, analogous to cointegration removing unit roots. Such an outcome would allow a subset of variables to be forecast as anticipated. Finally, forecasting using leading indicators is discussed in §11 based on Emerson and Hendry (1996), raising issues of cointegration and co-breaking within the indices, and between their components and macro-economic variables. Since composite leading indicators (CLIs) are unlikely to co-break to mitigate regime shifts, the effects of adding CLIs to macro models may be potentially harmful. Given that most of the conventional results for constant-parameter cointegrated stationary processes do not seem to apply in realistic settings, §12 concludes that a formal theory of forecasting for mis-specified models under irregular and substantive structural breaks requires development, but is feasible.

2 Concepts

In this section, we define the predictability of a stochastic process relative to the available information, and the resulting forecastability of the series, then draw some implications.

2.1 Unpredictability

ν_t is an unpredictable process with respect to the information set \mathcal{I}_{t-1} if:¹

$$D_{\nu_t}(\nu_t | \mathcal{I}_{t-1}) = D_{\nu_t}(\nu_t), \quad (1)$$

so the conditional and unconditional distributions coincide. Unpredictability is invariant under non-singular contemporaneous transforms: e.g., if ν_t is unpredictable, so is $\mathbf{B}\nu_t$ where $|\mathbf{B}| \neq 0$. However, unpredictability is obviously not invariant under intertemporal transforms since if $\mathbf{u}_t = \nu_t + \mathbf{A}\mathbf{f}(\mathcal{I}_{t-1})$:

$$D_{\mathbf{u}_t}(\mathbf{u}_t | \mathcal{I}_{t-1}) \neq D_{\mathbf{u}_t}(\mathbf{u}_t)$$

¹The definition is equivalent to the statistical independence of ν_t from \mathcal{I}_{t-1} and does not connote 'wild': indeed, knowing $D_{\nu_t}(\nu_t)$ may be highly informative relative to not knowing it.

when $\mathbf{A} \neq \mathbf{0}$. The concept resolves the apparent ‘paradox’ that (e.g.) although the change in the log of real equity prices may be unpredictable, the level is predictable: since $x_t = \Delta x_t + x_{t-1}$, the ‘prediction’ of the current level is merely its immediate past value. Below, we assume the time series \mathbf{x}_t is of interest, and the information set \mathcal{I}_{t-1} includes at least the history of \mathbf{x}_t . When $\mathbf{x}_t = \boldsymbol{\nu}_t$, therefore, \mathbf{x}_t must be an innovation, and (weak) white noise when its second moment exists.

Unpredictability is relative to the information set used; e.g., it can happen that for $\mathcal{J}_{t-1} \subset \mathcal{I}_{t-1}$:

$$D_{\mathbf{u}_t}(\mathbf{u}_t | \mathcal{J}_{t-1}) = D_{\mathbf{u}_t}(\mathbf{u}_t) \quad \text{yet} \quad D_{\mathbf{u}_t}(\mathbf{u}_t | \mathcal{I}_{t-1}) \neq D_{\mathbf{u}_t}(\mathbf{u}_t).$$

However, $\mathcal{J}_{t-1} \subset \mathcal{I}_{t-1}$ does not preclude predictability. Unpredictability may also be relative to the time period, in that we could have:

$$D_{\mathbf{u}_t}(\mathbf{u}_t | \mathcal{I}_{t-1}) = D_{\mathbf{u}_t}(\mathbf{u}_t) \quad \text{for } t = 1, \dots, T \tag{2}$$

yet:

$$D_{\mathbf{u}_t}(\mathbf{u}_t | \mathcal{I}_{t-1}) \neq D_{\mathbf{u}_t}(\mathbf{u}_t) \quad \text{for } t = T + 1, \dots, T + H, \tag{3}$$

or vice versa. Finally, unpredictability may be relative to the horizon considered in that:

$$D_{\mathbf{u}_t}(\mathbf{u}_t | \mathcal{I}_{t-2}) = D_{\mathbf{u}_t}(\mathbf{u}_t) \quad \text{yet} \quad D_{\mathbf{u}_t}(\mathbf{u}_t | \mathcal{I}_{t-1}) \neq D_{\mathbf{u}_t}(\mathbf{u}_t).$$

However, the converse, that :

$$D_{\mathbf{u}_t}(\mathbf{u}_t | \mathcal{I}_{t-1}) = D_{\mathbf{u}_t}(\mathbf{u}_t) \quad \text{yet} \quad D_{\mathbf{u}_t}(\mathbf{u}_t | \mathcal{I}_{t-2}) \neq D_{\mathbf{u}_t}(\mathbf{u}_t)$$

is not possible as $\mathcal{I}_{t-2} \subseteq \mathcal{I}_{t-1}$ by definition.

Sequential factorization of the joint density of \mathbf{X}_T^1 yields the prediction representation:

$$D_{\mathbf{X}}(\mathbf{X}_T^1 | \mathcal{I}_0, \cdot) = \prod_{t=1}^T D_{\mathbf{x}_t}(\mathbf{x}_t | \mathcal{I}_{t-1}, \cdot). \tag{4}$$

Consequently, predictability requires combinations with \mathcal{I}_{t-1} : the ‘causes’ must be in train. Such causes need not be direct, and could be very indirect: e.g., a variable’s own lags may ‘capture’ actual past causes. Thus, when the relevant \mathcal{I}_{t-1} is known, structure is not necessary for forecasting, even under changed conditions. Unfortunately, that \mathcal{I}_{t-1} is known is most unlikely in economics, with important implications for understanding why ‘ad hoc’ methods can work well, as seen below.

2.2 Moments

Forecasting tends to focus on first and second moments assuming these exist. Then, $\boldsymbol{\nu}_t$ is unpredictable in mean at t if:

$$E[\boldsymbol{\nu}_t | \mathcal{I}_{t-1}] = E[\boldsymbol{\nu}_t].$$

Similarly, $\boldsymbol{\nu}_t$ is unpredictable in variance at t if:

$$V[\boldsymbol{\nu}_t | \mathcal{I}_{t-1}] = V[\boldsymbol{\nu}_t].$$

The converse of the latter includes (e.g.) autoregressive conditional heteroscedastic processes (ARCH or GARCH: see Engle, 1982, Bollerslev, Chou and Kroner, 1992, and Bollerslev, Engle and Nelson, 1994), or stochastic volatility schemes (see Shephard, 1996). Consequently, unpredictability in mean is not invariant under non-linear contemporaneous transforms, as in the weak white-noise ARCH process:

$$E[\boldsymbol{\nu}_t | \mathcal{I}_{t-1}] = E[\boldsymbol{\nu}_t] \quad \text{but} \quad E[\boldsymbol{\nu}_t \boldsymbol{\nu}_t' | \mathcal{I}_{t-1}] \neq E[\boldsymbol{\nu}_t \boldsymbol{\nu}_t'].$$

2.3 Forecastability

A forecasting rule is any systematic operational procedure for making statements about future events. We will focus on statistical forecasting using formal estimated econometric models. Whereas predictability is a property (of a stochastic process in relation to an information set), forecasting is a process. Moreover, forecasting is undertaken for a purpose, so its evaluation depends on how well it achieves that intent. Consequently, it is extremely difficult to define ‘forecastability’. One could perhaps define events as forecastable relative to a loss measure if the relevant procedure produced a lower expected loss than (say) the historical mean. This would be consistent with the change in the log of real equity prices being unforecastable, but the level forecastable using a random walk, on the criteria of bias or MSFE. Unfortunately, as shown in §4, MSFE rankings for multivariate, multi-step forecasts depend on the transformations used, so can alter in accuracy relative to the historical mean of the transform, rendering most definitions ambiguous.

2.4 Implications

These concepts have a number of important implications applicable to most statistical forecasting methods. First, from (1), since the conditional mean of an unpredictable process is its unconditional mean, predictability is necessary for forecastability. However, it is not sufficient, since the relevant information set may be unknown in practice. Further, there is a potential ambiguity in the use of the phrase ‘information set’ in the contexts of predictability and forecasting: \mathcal{I}_{t-1} denotes the conditioning set generated by the relevant events, whereas forecastability also requires knowledge of how \mathcal{I}_{t-1} enters the conditional density in (1). For example, \mathbf{v}_{t-1} may matter, but in an awkward non-linear way that eludes empirical modelling.

Secondly, translating ‘regularity’ as a systematic relation between the entity to be forecast and the available information, then conditions (a)–(d) above are sufficient for forecastability. They may not be necessary in principle (e.g., inspired guessing; precognition etc.), but for statistical forecasting, they seem close to necessary as can be seen by considering the removal of any one of them (e.g., if no regularities exist to be captured).

Thirdly, if the occurrence of large ex ante unpredictable shocks (such as earthquakes, or oil crises), induces their inclusion in later information sets (moving from (2) to (3) above), the past will be more explicable than the future is forecastable (cf. stock-market commentators?). Consequently, when the ‘true’ \mathcal{I}_{t-1} is unknown, to prevent the baseline innovation error variance being an underestimate, forecast-accuracy evaluation may require ‘unconditioning’ from within-sample rare events that have been modelled post hoc.

Fourthly, from (4), intertemporal transforms affect predictability, so no unique measure of predictability, and hence of forecast accuracy, exists. Linear dynamic econometric systems are invariant under linear transforms in that they retain the same error process, and transformed estimates of the original are usually the direct estimates of the transformed system: such transforms are used regularly in empirical research. But by definition, the predictability of the transformed variables is altered by any transforms that are intertemporal (e.g., switching from y_t on y_{t-1} to Δy_t on y_{t-1}).² This precludes unique generic rankings of methods, adding to the difficulty of theoretical analysis and practical appraisal.

Next, since new unpredictable components can enter in each period, forecast error variances could increase or decrease over increasing horizons from any given T , as a consequence of (2) versus (3).

²While 1-step MSFEs are invariant to that particular transform, measures such as R^2 are not.

For integrated processes, $V[x_{T+h}|\mathcal{I}_T]$ is non-decreasing in h when the innovation distribution is homoscedastic. Otherwise, when the initial forecast period T increases with real time, forecast uncertainty will be non-decreasing in h unless the innovation variance is ever-decreasing (since h -steps ahead from T becomes $h - 1$ from $T + 1$).³

Finally, and the focus of §5, when the ‘true’ \mathcal{I}_{t-1} is unknown one cannot prove that ‘genuinely’ relevant information must always dominate non-causal variables in forecasting. Rather, one can show in examples that the latter can be the ‘best available’ forecasting devices on some measures in the absence of omniscience (i.e., when the model is not the DGP). First, however, we need to explain the class of processes and models under analysis, and consider how forecast accuracy will be measured.

3 The framework

For an econometric theory of forecasting to deliver relevant conclusions about empirical forecasting, it must be based on assumptions that adequately capture the appropriate aspects of the real world to be forecast. Consequently, we consider a non-stationary (evolutionary) world subject to structural breaks, where the model differs from the mechanism, and requires estimation from available data. The present analysis considers integrated-cointegrated mechanisms which are linear in \mathbf{x}_t , but are also subject to shifts in the deterministic factors. Generalizations to longer lags, and non-linear relations seem feasible but await formal development.

3.1 The data generation process

For exposition, the data generation process (DGP) is defined over the period $t = 1, \dots, T$ by a first-order vector autoregressive process (VAR) in the n variables \mathbf{x}_t :

$$\mathbf{x}_t = \boldsymbol{\tau} + \boldsymbol{\Upsilon}\mathbf{x}_{t-1} + \boldsymbol{\nu}_t \text{ where } \boldsymbol{\nu}_t \sim \text{IN}_n[\mathbf{0}, \boldsymbol{\Omega}], \quad (5)$$

denoting an independent normal error with expectation $E[\boldsymbol{\nu}_t] = \mathbf{0}$ and variance matrix $V[\boldsymbol{\nu}_t] = \boldsymbol{\Omega}$. The DGP is integrated of order unity (I(1)), and satisfies $r < n$ cointegration relations such that:

$$\boldsymbol{\Upsilon} = \mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}', \quad (6)$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times r$ matrices of rank r .⁴ Then (5) can be reparameterized as the vector equilibrium-correction model (VEqCM):

$$\Delta\mathbf{x}_t = \boldsymbol{\tau} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\nu}_t, \quad (7)$$

where $\Delta\mathbf{x}_t$ and $\boldsymbol{\beta}'\mathbf{x}_t$ are I(0). Let:

$$\boldsymbol{\tau} = \boldsymbol{\gamma} - \boldsymbol{\alpha}\boldsymbol{\mu}, \quad (8)$$

where $\boldsymbol{\mu}$ is $r \times 1$ and $\boldsymbol{\beta}'\boldsymbol{\gamma} = \mathbf{0}$ so in deviations about means:⁵

$$(\Delta\mathbf{x}_t - \boldsymbol{\gamma}) = \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\nu}_t \quad (9)$$

where the system grows at the unconditional rate $E[\Delta\mathbf{x}_t] = \boldsymbol{\gamma}$ with long-run solution $E[\boldsymbol{\beta}'\mathbf{x}_t] = \boldsymbol{\mu}$.

³Chong and Hendry (1986) show that forecast confidence intervals may be non-monotonic in h when parameters are estimated: see §6.

⁴In (5), none of the roots of $|\mathbf{I} - \boldsymbol{\Upsilon}L| = 0$ lies inside the unit circle (where L is the lag operator, $L^s x_t = x_{t-s}$), and $\boldsymbol{\alpha}'\boldsymbol{\Phi}\boldsymbol{\beta}$ is rank $(n - r)$, where $\boldsymbol{\Phi}$ is the mean-lag matrix (here $\boldsymbol{\Upsilon}$), when $\boldsymbol{\alpha}_\perp$ and $\boldsymbol{\beta}_\perp$ are $n \times (n - r)$ matrices of rank $(n - r)$ such that $\boldsymbol{\alpha}'\boldsymbol{\alpha}_\perp = \boldsymbol{\beta}'\boldsymbol{\beta}_\perp = \mathbf{0}$.

⁵ $\boldsymbol{\gamma} = \boldsymbol{\beta}_\perp(\boldsymbol{\alpha}'_\perp\boldsymbol{\beta}_\perp)^{-1}\boldsymbol{\alpha}'_\perp\boldsymbol{\tau}$. The decomposition using $\boldsymbol{\tau} = \boldsymbol{\gamma} - \boldsymbol{\alpha}\boldsymbol{\mu}$ is not orthogonal since $\boldsymbol{\gamma}'\boldsymbol{\alpha}\boldsymbol{\mu} \neq \mathbf{0}$, but as a DGP, (9) is isomorphic to (7).

3.2 The model class

The form of the model coincides with (5) as a linear representation of \mathbf{x}_t , but is potentially mis-specified:

$$\mathbf{x}_t = \boldsymbol{\tau}_p + \boldsymbol{\Upsilon}_p \mathbf{x}_{t-1} + \mathbf{u}_t, \quad (10)$$

where the parameter estimates ($\hat{\boldsymbol{\tau}} : \hat{\boldsymbol{\Upsilon}} : \hat{\boldsymbol{\Omega}}$) are possibly inconsistent, with $\boldsymbol{\tau}_p \neq \boldsymbol{\tau}$ and $\boldsymbol{\Upsilon}_p \neq \boldsymbol{\Upsilon}$. Empirical econometric models like (10) are not numerically calibrated theoretical models, but have error processes which are derived, and so are not autonomous: see Gilbert (1986), Hendry (1995a), and Spanos (1986) *inter alia*. The theory of reduction explains the origin and status of such empirical models in terms of the implied information reductions relative to the process that generated the data. Some reductions, such as invalid marginalization, affect forecast accuracy directly, whereas others, such as aggregation, may primarily serve to define the object of interest.

Two specific models considered below are defined by ($\boldsymbol{\tau}_p = \boldsymbol{\tau}$, $\boldsymbol{\Upsilon}_p = \boldsymbol{\Upsilon}$) and ($\boldsymbol{\tau}_p = \boldsymbol{\gamma}$, $\boldsymbol{\Upsilon}_p = \mathbf{I}_n$). The first model is the DGP in sample. Although empirical econometric models are invariably not facsimiles of the DGP, they could match the data evidence in all measurable respects – i.e., be congruent; but as we allow for forecast-period structural change, the model will not coincide with the DGP in the forecast period. The second model is given by:

$$\Delta \mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\xi}_t$$

which is correctly specified only when $\boldsymbol{\alpha} = \mathbf{0}$ in (9), in which case $\boldsymbol{\xi}_t = \boldsymbol{\nu}_t$. It is a VAR in the differences of the variables (DVAR), and is mis-specified in sample by omitting the cointegrating vectors.

4 Measuring forecast accuracy

Although econometric analyses could begin by specifying a loss function from which the optimal predictor is derived, a well-defined mapping between forecast errors and their costs is not typical in macro-economics. Consequently, measures of forecast accuracy are often based on the MSFE matrix:

$$\mathbf{V}_h \equiv \mathbb{E} [\mathbf{e}_{T+h} \mathbf{e}'_{T+h}] = \mathbb{V} [\mathbf{e}_{T+h}] + \mathbb{E} [\mathbf{e}_{T+h}] \mathbb{E} [\mathbf{e}'_{T+h}] \quad (11)$$

where \mathbf{e}_{T+h} is a vector of h -step ahead forecast errors. Such measures may lack invariance to non-singular, scale-preserving, linear transformations for which the associated model class is invariant, so MSFE comparisons may yield inconsistent rankings between forecasting models on multi-step ahead forecasts depending on the particular transformations of variables examined (e.g., level or differences). Clements and Hendry (1993) show analytically that for multi-step forecasts, the trace, determinant, and the whole matrix \mathbf{V}_h lack invariance.

Denote the linear forecasting system by:

$$\boldsymbol{\Gamma} \mathbf{s}_t = \mathbf{u}_t \quad \text{with} \quad \mathbf{u}_t \sim \text{IN}_{n+k} [\mathbf{0}, \boldsymbol{\Sigma}] \quad (12)$$

where $\mathbf{s}'_t = (\mathbf{x}'_t : \mathbf{z}'_t)$, \mathbf{z}_t are the m available predetermined variables and $\boldsymbol{\Sigma}$ is symmetric, positive semi-definite: for example, in (5), $\mathbf{z}'_t = (1, \mathbf{x}'_{t-1})$ and $\boldsymbol{\Gamma} = (\mathbf{I}_n : -\boldsymbol{\tau} : -\boldsymbol{\Upsilon})$. Then the likelihood and generalized variance of the system in (12) are invariant under scale-preserving, non-singular transformations of the form:

$$\mathbf{M} \boldsymbol{\Gamma} \mathbf{P}^{-1} \mathbf{P} \mathbf{s}_t = \mathbf{M} \mathbf{u}_t$$

so:

$$\boldsymbol{\Gamma}^* \mathbf{s}_t^* = \mathbf{u}_t^* \quad \text{with} \quad \mathbf{u}_t^* \sim \text{IN}_{n+k} (\mathbf{0}, \mathbf{M} \boldsymbol{\Sigma} \mathbf{M}') . \quad (13)$$

In (13), $\mathbf{s}_t^* = \mathbf{P}\mathbf{s}_t$, \mathbf{M} and \mathbf{P} are respectively $n \times n$ and $(m+n) \times (m+n)$ known non-singular matrices where $|\mathbf{M}| = 1$, and \mathbf{P} is the upper block-triangular matrix:

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

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

$$|\mathbf{M}\boldsymbol{\Sigma}\mathbf{M}'| = |\boldsymbol{\Sigma}|, \quad (14)$$

so the systems (12) and (13) are isomorphic. Forecasts and forecast confidence intervals made in the original system and transformed after the event to \mathbf{x}_t^* , or made directly from the transformed system, are identical; and this remains true when parameters are estimated by any method that is invariant (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 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. A point of potential confusion is that the differences of the forecasts of $\widehat{\mathbf{x}}_{T+h}$ may not equal $\widehat{\Delta\mathbf{x}}_{T+h}$ despite using $\Delta\widehat{\mathbf{x}}_{T+h} = \widehat{\mathbf{x}}_{T+h} - \widehat{\mathbf{x}}_{T+h-1}$ if at $h = 1$, the actual initial condition \mathbf{x}_T is subtracted: this adds an intercept correction setting the model back on track, often markedly improving the forecasts as shown below. Nevertheless, the forecasts themselves are invariant to linear transforms; the present issue is the lack of invariance of some measures of their accuracy.

For transformations involving \mathbf{M} only (i.e., $\mathbf{P} = \mathbf{I}_{n+m}$), the matrix measure \mathbf{V}_h and determinant are invariant, but the trace is not: see Granger and Newbold (1986). When $\mathbf{M} = \mathbf{I}_n$, for transformations using \mathbf{P} , neither the determinant nor the MSFE matrix are invariant for $h > 1$, even though the distribution of the \mathbf{u}_t is unaffected by (13): see Clements and Hendry (1996a).

Invariance to \mathbf{P} transformations in a measure requires accounting for covariances between different step-ahead errors, leading to a generalized forecast-error second-moment matrix (GFESM, which is close to predictive likelihood: see Bjørnstad, 1990):

$$\boldsymbol{\Phi}_h = \mathbf{E} [\mathbf{E}_h \mathbf{E}_h'],$$

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

$$\mathbf{E}_h' = [\mathbf{e}'_{T+1}, \mathbf{e}'_{T+2}, \dots, \mathbf{e}'_{T+h-1}, \mathbf{e}'_{T+h}].$$

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

$$\tilde{\mathbf{E}}_h' = [\mathbf{e}'_{T+1}\mathbf{M}', \mathbf{e}'_{T+2}\mathbf{M}', \dots, \mathbf{e}'_{T+h-1}\mathbf{M}', \mathbf{e}'_{T+h}\mathbf{M}'],$$

we have:

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

since $|\mathbf{I}_n \otimes \mathbf{M}| = 1$.

Although invariance is useful to determine a unique measure for a fixed model independently of its representation, it is not compelling, and often several forecast-accuracy indices are reported.

5 Causal information in economic forecasting

We now consider the role of causal information in economic forecasting first when the model coincides with the mechanism, then when it does not; the mechanism is allowed to be non-constant over time.

In the first case, causal information is always useful, and produces better forecasts than non-causal. Adding further variables produces no improvement. Even when the model is mis-specified, causally-relevant information generally improves forecasts providing the mechanism generates stationary data. Such a result cannot be shown for a mis-specified model of a non-constant mechanism, and non-causal additional variables potentially can be more useful than causally-relevant ones so long as the model remains mis-specified.

To demonstrate these claims, we assume all parameters are known: estimation uncertainty would reinforce the main conclusion. While sufficiently poor estimates would weaken any conclusions from the first case, our concern is to establish that causally-relevant variables cannot be relied upon to produce the ‘best’ forecasts when the model is mis-specified, and parameter uncertainty would strengthen this finding.

5.1 Model coincides with the mechanism

Consider the DGP in (5) for the n $l(1)$ variables \mathbf{x}_t . Here, (5) is both the model and the DGP, although it could be written in a lower-dimensional parameter space in terms of $l(0)$ transformations of the original variables as in (9) above. The notation is simplest when the mechanism is constant, so we prove the result for 1-step forecasts in that setting first.

The in-sample conditional expectation of \mathbf{x}_{T+1} given \mathbf{x}_T is:

$$E[\mathbf{x}_{T+1} | \mathbf{x}_T] = \boldsymbol{\tau} + \boldsymbol{\Upsilon} \mathbf{x}_T$$

and this delivers the (matrix) minimum MSFE. Under the present assumptions, the resulting forecast error is a homoscedastic innovation against all further information:

$$E[\boldsymbol{\nu}_{T+1} | \mathbf{x}_T] = \mathbf{0} \text{ and } V[\boldsymbol{\nu}_{T+1} | \mathbf{x}_T] = \boldsymbol{\Omega}. \quad (15)$$

Consequently, adding any further variables \mathbf{z}_{t-1} to (5) will not improve the forecast accuracy of mean or variance.

Conversely, replacing any $x_{i,t-1}$ by any or all elements from \mathbf{z}_{t-1} will lead to inefficient forecasts unless there is perfect correlation between $x_{i,t}$ and \mathbf{z}_t . Denote the resulting regressor vector by $\bar{\mathbf{x}}_{t-1}$, then, forecasting from:

$$\mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\Gamma} \bar{\mathbf{x}}_{t-1} + \mathbf{e}_t,$$

where $E[\mathbf{e}_t | \bar{\mathbf{x}}_{t-1}] = \mathbf{0}$ using:

$$\tilde{\mathbf{x}}_{T+1} = \boldsymbol{\gamma} + \boldsymbol{\Gamma} \bar{\mathbf{x}}_T,$$

the forecast error is:

$$\mathbf{e}_{T+1} = \mathbf{x}_{T+1} - \tilde{\mathbf{x}}_{T+1} = (\boldsymbol{\tau} - \boldsymbol{\gamma}) + \boldsymbol{\Upsilon} \mathbf{x}_T - \boldsymbol{\Gamma} \bar{\mathbf{x}}_T + \boldsymbol{\nu}_{T+1}.$$

Let $\mathbf{x}_t = \boldsymbol{\zeta} + \boldsymbol{\Psi} \bar{\mathbf{x}}_t + \mathbf{w}_t$ (say) with $E[\mathbf{w}_t | \bar{\mathbf{x}}_t] = \mathbf{0}$ and $V[\mathbf{w}_t | \bar{\mathbf{x}}_t] = \boldsymbol{\Phi}$, so:

$$\mathbf{e}_{T+1} = (\boldsymbol{\tau} - \boldsymbol{\gamma} + \boldsymbol{\Upsilon} \boldsymbol{\zeta}) + (\boldsymbol{\Upsilon} \boldsymbol{\Psi} - \boldsymbol{\Gamma}) \bar{\mathbf{x}}_T + \boldsymbol{\Upsilon} \mathbf{w}_T + \boldsymbol{\nu}_{T+1}$$

with mean:

$$E[\mathbf{e}_{T+1} | \bar{\mathbf{x}}_T] = (\boldsymbol{\tau} - \boldsymbol{\gamma} + \boldsymbol{\Upsilon} \boldsymbol{\zeta}) + (\boldsymbol{\Upsilon} \boldsymbol{\Psi} - \boldsymbol{\Gamma}) \bar{\mathbf{x}}_T = \mathbf{0} \quad (16)$$

so that $\boldsymbol{\gamma} = \boldsymbol{\tau} + \boldsymbol{\Upsilon} \boldsymbol{\zeta}$ and $\boldsymbol{\Upsilon} \boldsymbol{\Psi} = \boldsymbol{\Gamma}$; and variance:

$$V[\mathbf{e}_{T+1} | \bar{\mathbf{x}}_T] = \boldsymbol{\Omega} + \boldsymbol{\Upsilon} \boldsymbol{\Phi} \boldsymbol{\Upsilon}'. \quad (17)$$

Thus, the forecasts are conditionally unbiased (16), but inefficient (17).

Next, in a non-constant DGP, §9 shows that the main non-constancies of interest concern direct or indirect changes in the deterministic components of (5). Either τ can change, or if Υ changes, the unconditional means of the $l(0)$ components alter. We only consider the former. Let τ change to τ^* , so the DGP in the forecast period becomes:

$$\mathbf{x}_{T+1} = \tau^* + \Upsilon \mathbf{x}_T + \nu_T \quad (18)$$

Since the model also switches to (18) by being the mechanism, the forecast errors have the same properties as in (15), and the previous result is unchanged. Its converse, that (18) will dominate incorrect models, is more tedious to show, but follows from a generalization of the argument in (16) and (17).

Such powerful results are not surprising; but the assumption that the model coincides with the mechanism is extremely strong and not empirically relevant.

5.2 Model does not coincide with the mechanism

First, we show that if the process is stationary, predictive failure is unconditionally unlikely, irrespective of how badly the model is specified (see Hendry, 1979), but that causal information dominates non-causal. Even so, non-causal might help, if it acts as a proxy for the omitted causal variables.. Then we provide an example where causal information does not help once structural breaks are introduced.

Reparameterize the system as in (9):

$$\Delta \mathbf{x}_t = \gamma + \alpha (\beta' \mathbf{x}_{t-1} - \mu) + \nu_t. \quad (19)$$

There are many ways in which a model could be mis-specified for the mechanism in (19), but we only consider omission of the $l(0)$ cointegrating components. Denote the model by:

$$\Delta \mathbf{x}_t = \delta + \rho (\beta'_1 \mathbf{x}_{t-1} - \mu_1) + \eta_t \quad (20)$$

where β'_1 is (perhaps a linear transform of) a subset of the r cointegrating vectors in (19), and μ_1 is the unconditional expectation of $\beta'_1 \mathbf{x}_t$. Then, as $E[\beta'_1 \mathbf{x}_{t-1}] = \mu_1$, $\delta = \gamma$, and hence for known parameters in (20) and forecast $\widehat{\Delta \mathbf{x}_{T+1}} = \gamma + \rho(\beta'_1 \mathbf{x}_T - \mu_1)$:

$$E \left[\widehat{\Delta \mathbf{x}_{T+1}} \right] = \gamma$$

so forecasts are unconditionally unbiased, though inefficient. Adding any omitted $l(0)$ linear combinations of \mathbf{x}_{t-1} will improve forecasts, as will adding any $\Delta \mathbf{x}_{t-1}$ which proxy for omitted $\beta'_2 \mathbf{x}_{t-1}$.

Thus, the notion of basing forecasting on ‘causal models’ still has substance, perhaps qualified by the need to estimate parameters from small samples of badly-measured data. However, once the model is not the mechanism and the mechanism is non-constant, the dominance of causal information over non-causal cannot be shown. We consider a counter example where non-causal information dominates causal on at least one forecast criterion, unless omniscience is assumed. The result may help explain some of the apparent success of the approach in Box and Jenkins (1976).

Consider a world in which GNP (Y) is ‘caused’ by the exchange rate (E):

$$Y_t = \alpha E_{t-1} + \epsilon_t \quad \text{with } \epsilon_t \sim \text{IN} [0, \sigma_\epsilon^2] \quad \text{for } t \in \mathcal{T}_1 \quad (21)$$

where $\mathcal{T}_1 = [1, T_1]$. Then at $T_1 + 1$, the DGP changes to:

$$Y_t = \beta R_{t-1} + v_t \quad \text{with } v_t \sim \text{IN} [0, \sigma_v^2] \quad \text{for } t \in \mathcal{T}_2 \quad (22)$$

where $\mathcal{T}_2 = [T_1 + 1, T]$. The collapse of Bretton Woods, leaving the ERM, or entering EMU are potential examples, albeit that the model is overly simplistic. We assume that E_t and R_t are driftless random walks, but are always positive (otherwise, shifts in intercepts would be needed), given by $\Delta E_t = e_t$ and $\Delta R_t = r_t$ with mean-zero, white-noise innovations e_t and r_t , and variances σ_e^2 and σ_r^2 .

Throughout, Y_t is predictable in mean from the universal information set, since (setting $D_t = 1$ when $t \in \mathcal{T}_1$, zero otherwise):

$$\mathbb{E}[Y_t | \mathcal{I}_{t-1}] = \alpha D_t E_{t-1} + \beta (1 - D_t) R_{t-1}.$$

However, using the criterion of unbiasedness of forecasts, Y_t is forecastable from (E_{t-1}, R_{t-1}) after T_1 only if the switch point is known. For example, the model based on regressing Y_t on E_{t-1} and R_{t-1} , namely:

$$Y_t = \psi_1 E_{t-1} + \psi_2 R_{t-1} + u_t \quad (23)$$

will not suffice as:⁶

$$\mathbb{E} \left[\begin{pmatrix} \hat{\psi}_1 \\ \hat{\psi}_2 \end{pmatrix} \right] = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \simeq \begin{pmatrix} K\alpha \\ (1-K)\beta \end{pmatrix},$$

where $K = T_1(T_1 - 1) / (T(T - 1))$. Forecasting using $\hat{Y}_t = \psi_1 E_{t-1} + \psi_2 R_{t-1}$ for $t \in \mathcal{T}_2$ yields:

$$\mathbb{E}[\hat{Y}_t | \mathcal{I}_{t-1}] \simeq K\alpha E_{t-1} + (1-K)\beta R_{t-1} = \beta R_{t-1} + K(\alpha E_{t-1} - \beta R_{t-1}),$$

so that:

$$\mathbb{E}[Y_t - \hat{Y}_t | \mathcal{I}_{t-1}, D_t = 0] = -K(\alpha E_{t-1} - \beta R_{t-1}).$$

As E is the relative price of two currencies, it could move wildly due to the other country's behaviour, swamping any predictability from R to produce badly biased forecasts.

Consider an alternative forecasting procedure that ignores the information on E_{t-1} , R_{t-1} and simply uses the time series on Y_t . For example, differencing once yields:

$$\begin{aligned} \Delta Y_t &= \alpha e_{t-1} + \Delta \epsilon_t \quad t \in \mathcal{T}_1 \\ \Delta Y_{T_1+1} &= \beta R_{T_1} - \alpha E_{T_1-1} + v_{T_1+1} - \epsilon_{T_1} \\ \Delta Y_t &= \beta r_{t-1} + \Delta v_t \quad t > T_1 + 1 \end{aligned} \quad (24)$$

so for $t \in \mathcal{T}_2$:

$$\mathbb{E}[\Delta Y_t] = \mathbb{E}[\beta r_{t-1} + \Delta v_t] = 0$$

and hence the forecast $\tilde{Y}_t = Y_{t-1}$ is unconditionally unbiased (albeit 'inefficient'). Thus, even though Y_{t-1} does not directly enter the DGP, $\tilde{Y}_t = Y_{t-1}$ is better in terms of bias than \hat{Y}_t , which included the correct causal variable R_{t-1} . However, \tilde{Y}_t is slightly biased conditionally on Y_{t-1} , since after the break:⁷

$$\mathbb{E}[Y_t - \tilde{Y}_t | Y_{t-1}] = \mathbb{E}[\Delta Y_t | Y_{t-1}] = \mathbb{E}[\beta r_{t-1} + \Delta v_t | Y_{t-1}] = \mathbb{E}[v_{t-1} | Y_{t-1}] \neq 0.$$

The inability to prove that causally-relevant variables will dominate for mis-specified models has important implications. First, there exist methods of robustifying forecasts against structural breaks

⁶The analytical calculations depend on what is assumed about the time-series properties of E_t and R_t . If these are not independent driftless random walks, but (e.g.) are correlated, or have heteroscedastic errors after the regime shift, then a somewhat different, but related, analysis is needed. Appendix A provides the derivation.

⁷At the switch point, since $K \simeq 1$, the relative biases conditional on \mathcal{I}_{T_1} could go either way as $\mathbb{E}[Y_{T_1+1} - \hat{Y}_{T_1+1} | \mathcal{I}_{T_1}] \simeq \beta R_{T_1} - \alpha E_{T_1-1}$, whereas $\mathbb{E}[Y_{T_1+1} - \tilde{Y}_{T_1+1} | \mathcal{I}_{T_1}] = \beta R_{T_1} - \alpha E_{T_1-1} - \mathbb{E}[\epsilon_{T_1} | \mathcal{I}_{T_1}]$, and hence could be constructed to be smaller for \tilde{Y}_{T_1+1} .

that have occurred: the differencing in (24) is one example, and others are noted below. Alternatively, other forms of non-causal information may prove relevant, such as intercept corrections. Secondly, the differenced process in (24) is close to an ARIMA with a large negative moving-average root and an outlier at the break point, although that is not the DGP (see the analysis of multi-step estimators in Clements and Hendry, 1996d). Thirdly, for large enough unmodelled breaks, ‘causal’ models will lose on MSFE (and related criteria) to models that are robustified against breaks; this is examined further in §9 below. Finally, while the example is specifically constructed to demonstrate the possibility of dominating causal information, §9 provides a general class that also does so, and highlights the distinction between error-correction and equilibrium-correction mechanisms.

6 Parsimony

It is not easy to find formal reasons for the advantages of parsimony in forecasting, despite a general folklore that it matters. Estimated parameter variances decline with sample size, whereas inconsistencies do not, so any trade-off rapidly moves against parsimony: we show below that in correct model specifications, large parameter variances due to ‘collinearity’ cannot be a justification. The origins of the arguments for parsimony in the methods of Box and Jenkins (1976) are because of lack of identification in ARIMA models when there are redundant common factors, so do not generalize to other model classes. There is some evidence that empirically, parsimony may help (see, e.g., the basis for ‘Bayesian DVARS’ in Doan *et al.*, 1984). In practice, model mis-specification and structural change seem likely to be more important than parameter uncertainty, leading to an alternative justification noted below, but we first consider a case where parsimony can be shown to matter. We focus on h -step ahead scalar forecasts under correct specification, drawing on Hendry and Clements (1993).

Consider a stationary first-order autoregression defined by:

$$y_t = \rho y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim \text{IN} [0, \sigma_\epsilon^2],$$

with $|\rho| < 1$. Then $E[y_t] = 0$ and $E[y_t^2] = \sigma_y^2 = \sigma_\epsilon^2 / (1 - \rho^2)$.

We examine the effect on forecasting h -steps ahead of estimating ρ relative to imposing it at zero. Since:

$$y_{T+h} = \rho^h y_T + \sum_{j=0}^{h-1} \rho^j \epsilon_{T+h-j}, \quad (25)$$

the h -step conditional forecast error from the estimated model using $\hat{y}_{T+h} = \hat{\rho}^h y_T$ in (25) is:

$$\hat{\epsilon}_{T+h} = y_{T+h} - \hat{y}_{T+h} = (\rho^h - \hat{\rho}^h) y_T + \sum_{j=0}^{h-1} \rho^j \epsilon_{T+h-j}, \quad (26)$$

so that for forecasting the level y_{T+h} on a MSFE basis (see Schmidt, 1974, Baillie, 1979, Chong and Hendry, 1986, and Campos, 1992, noting the caveats in Clements and Hendry, 1993):

$$\begin{aligned} M[\hat{\epsilon}_{T+h} | y_T] &\simeq \frac{(1 - \rho^{2h}) \sigma_\epsilon^2}{(1 - \rho^2)} + T^{-1} h^2 \rho^{2(h-1)} (1 - \rho^2) y_T^2 \\ &= \sigma_y^2 \left[(1 - \rho^{2h}) + T^{-1} h^2 \rho^{2(h-1)} (1 - \rho^2) y_T^{\dagger 2} \right], \end{aligned} \quad (27)$$

where $y_T^\dagger = y_T / \sigma_y$. When a forecast of zero (the unconditional mean) is used instead, so $\tilde{y}_{T+h} = 0 \forall h$, then (again for levels):

$$M[\tilde{\epsilon}_{T+h} | y_T] = \frac{(1 - \rho^{2h}) \sigma_\epsilon^2}{(1 - \rho^2)} + \rho^{2h} y_T^2$$

$$= \sigma_y^2 \left[(1 - \rho^{2h}) + \rho^{2h} y_T^{\dagger 2} \right]. \quad (28)$$

Hence, the relative MSFE, denoted $R(\cdot)$, is:

$$\begin{aligned} R(\tilde{\varepsilon}, \hat{\varepsilon}, h) &= \frac{M[\tilde{\varepsilon}_{T+h} | y_T] - M[\hat{\varepsilon}_{T+h} | y_T]}{\sigma_y^2} \\ &= T^{-1} \rho^{2(h-1)} (1 - \rho^2) [\phi_{\rho=0}^2 - h^2] y_T^{\dagger 2}, \end{aligned} \quad (29)$$

where $\phi_{\rho=0}^2 = T\rho^2 / (1 - \rho^2)$ is the non-centrality parameter of the F-test of $H_0: \rho = 0$ in (25). Thus, the condition for retaining the estimated coefficient rather than imposing it arbitrarily at zero becomes increasingly stringent as h increases, crudely expressed as needing $t_\rho > h$. Although the term as a whole is tending to zero, this provides some basis for parsimony in estimation for forecasting. A cross-over of sign in $R(\cdot)$ must occur at some h : there always exists an h at which negative $R(\cdot)$ values occur. The formula in (29) also explains why forecast confidence bands are non-monotonic in h , and can exceed the unconditional forecast uncertainty (see Chong and Hendry, 1986, and Ericsson and Marquez, 1989). Further, when $\phi_{\rho=0}^2 > 1$, a weighted average of the estimated and imposed models will outperform either, matching the optimal weight for pooling the models obtained in Hendry and Clements (1994b). This leads in turn to a scientific basis for intercept corrections using zero as the long-run outcome (e.g., for second-differenced data). At first sight, therefore, a formula such as (29) seems promising.

Unfortunately, Hendry and Clements (1993) show that the result does not generalize easily, nor is the effect large in Monte Carlo studies. Single parameters in vector systems cannot be selected by such a criterion, since powering a matrix has very indirect effects on its elements. More generally, parameter uncertainty is of order $O(T^{-1})$ in stationary processes, and even smaller for unit-root processes, as against other errors of $O(1)$. Consequently, sampling uncertainty does not seem to be the most serious problem, particularly as we now show that collinearity cannot justify parsimony for forecasting unless structural breaks occur.

However, as noted earlier, one argument for parsimony is excluding non-constant aspects. If an ‘irrelevant’ variable with a non-zero mean is nevertheless estimated as significant in a model, perhaps because it proxies another omitted effect (or acts like E_{t-1} in (23)), biased forecasts will result when that variable undergoes a change in its time-series behaviour. After a break, re-estimation on a least-squares criterion will reveal its irrelevance, so the model will revert to a relatively constant fit despite the forecast failure. However, a ‘more parsimonious’ model that excluded that effect would have produced better forecasts. One route that such effects may act along is collinearity, so we next look at that.

7 Collinearity in forecasting

One might anticipate that parsimony would have a more important role when there was substantial ‘collinearity’ in the explanatory variables. Despite the non-uniqueness of collinearity (see Hendry, 1995a), we investigate such a possibility in this section using the static regression model :

$$y_t = \beta' \mathbf{x}_t + \nu_t \text{ where } \nu_t \sim \text{IN} [0, \sigma_\nu^2] \quad (30)$$

with $\mathbf{x}_t \sim \text{IN}_k [\mathbf{0}, \mathbf{\Omega}]$ independently of $\{\nu_t\}$. For large T :

$$\sqrt{T} (\hat{\beta} - \beta) \underset{d}{\sim} N_k [\mathbf{0}, \sigma_\nu^2 \mathbf{\Omega}^{-1}].$$

Then for k regressors with estimated coefficients and known future values of \mathbf{x} :

$$\hat{y}_{T+1} = \hat{\beta}' \mathbf{x}_{T+1},$$

so that the forecast error:

$$\hat{\nu}_{T+1} = y_{T+1} - \hat{y}_{T+1} = v_{T+1} - \mathbf{x}'_{T+1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}),$$

and hence:

$$M[\hat{\nu}_{T+1} | \mathbf{x}_{T+1}] \simeq \sigma_v^2 (1 + T^{-1} \mathbf{x}'_{T+1} \boldsymbol{\Omega}^{-1} \mathbf{x}_{T+1}). \quad (31)$$

Factorize $\boldsymbol{\Omega}$ as $\mathbf{H}'\boldsymbol{\Lambda}\mathbf{H}$ where $\boldsymbol{\Lambda}$ is diagonal, $\mathbf{H}'\mathbf{H} = \mathbf{I}_k$ and $\mathbf{H}\mathbf{x}_t = \mathbf{z}_t$, so that $\mathbf{z}_t \sim \text{IN}_k[\mathbf{0}, \boldsymbol{\Lambda}]$, and:

$$y_t = \boldsymbol{\gamma}'\mathbf{z}_t + \nu_t \text{ where } \boldsymbol{\gamma} = \mathbf{H}\boldsymbol{\beta}. \quad (32)$$

Then $\mathbf{x}'_{T+1} \boldsymbol{\Omega}^{-1} \mathbf{x}_{T+1}$ equals:

$$\mathbf{x}'_{T+1} (\mathbf{H}'\boldsymbol{\Lambda}\mathbf{H})^{-1} \mathbf{x}_{T+1} = \mathbf{x}'_{T+1} \mathbf{H}' \boldsymbol{\Lambda}^{-1} \mathbf{H} \mathbf{x}_{T+1} = \mathbf{z}'_{T+1} \boldsymbol{\Lambda}^{-1} \mathbf{z}_{T+1} = \sum_{i=1}^k \frac{z_{i,T+1}^2}{\lambda_i}. \quad (33)$$

On average, $E[z_{i,T+1}^2] = \lambda_i$, and therefore, $E[\mathbf{x}'_{T+1} \boldsymbol{\Omega}^{-1} \mathbf{x}_{T+1}] = k$. Hence:

$$M[\hat{\nu}_{T+1}] \simeq \sigma_v^2 (1 + T^{-1}k). \quad (34)$$

This shows that any ‘collinearity’ in \mathbf{x}_t is irrelevant to forecasting so long as the marginal process remains constant. Alternatively, the model is invariant under linear, and therefore orthogonal, transforms as shown in (32), so 1-step forecasts are unaffected.

However, when $\boldsymbol{\beta}$ stays constant, but $\boldsymbol{\Omega}$ changes to $\boldsymbol{\Omega}^*$, with $\boldsymbol{\Lambda}$ changing to $\boldsymbol{\Lambda}^*$ then:

$$E\left[\sum_{i=1}^k \frac{z_{i,T+1}^2}{\lambda_i}\right] = \sum_{i=1}^k \frac{\lambda_i^*}{\lambda_i}, \quad (35)$$

so:

$$M[\hat{\nu}_{T+1}] \simeq \sigma_v^2 \left(1 + T^{-1} \sum_{i=1}^k \frac{\lambda_i^*}{\lambda_i}\right)$$

Thus, changes in the eigenvalues of the least-well determined β_i corresponding to the smallest λ_i will induce the biggest relative changes in $M[\hat{\nu}_{T+1}]$. For example, when $\lambda_1 = 0.001$ but $\lambda_1^* = 0.1$ then even for $T = 100$, the error variance is doubled. Thus, simplification could pay dividends for ex ante forecasting in such a state of nature.

Generalizations to dynamic systems are less clear cut, since the collinearity can alter only if the system specification changes, which anyway induces a structural break in some equations. However, (30) might be the only equation of interest in a system, in which case the analysis applies. Again we see a potential case for tighter parameterization once breaks are allowed.

8 Intercept corrections

Intercept corrections (ICs) are non-zero values for a model’s error terms added over a forecast period to adjust a model-generated forecast to prior beliefs, allow for anticipated future events that are not explicitly incorporated in a model, or ‘fix-up’ a model for perceived mis-specification over the past: see, for example, Wallis and Whitley (1991). A general theory of the role of intercept corrections in macroeconomic forecasting is provided in Hendry and Clements (1994b): here we establish the effects of one commonly-used IC following Clements and Hendry (1996c).

A typical IC for a 1-step ahead forecast is to add in the residual from the final sample observation to the first period forecast value. To see the consequences, consider the simple example of a stationary first-order autoregressive process:

$$y_t = \rho y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim \text{IN} [0, \sigma_v^2] \quad (36)$$

and $|\rho| < 1$. Let $\hat{\rho}$ denote the estimate of ρ , then the conventional forecast of period $T + 1$ given T is:

$$\hat{y}_{T+1} = \hat{\rho} y_T,$$

with forecast error:

$$\hat{\epsilon}_{T+1} = y_{T+1} - \hat{y}_{T+1} = (\rho - \hat{\rho}) y_T + \epsilon_{T+1}.$$

The intercept-correcting forecast is:

$$\hat{y}_{i,T+1} = \hat{y}_{T+1} + \hat{\epsilon}_T = \hat{\rho} y_T + \hat{\epsilon}_T,$$

with a forecast error $\hat{\epsilon}_{i,T+1}$ given by

$$\hat{\epsilon}_{i,T+1} = y_{T+1} - \hat{y}_{i,T+1} = (\rho - \hat{\rho}) y_T + \epsilon_{T+1} - \hat{\epsilon}_T = \Delta \hat{\epsilon}_{T+1},$$

so that the IC differences the original forecast error. We use this result in the next section to help correct for forecast biases arising from unanticipated structural breaks.

9 A taxonomy of forecast errors

To clarify the impact of structural breaks on forecasting, we reconsider the taxonomy of forecast errors in Clements and Hendry (1994). Despite allowing general structural change in the forecast period, the model and DGP to differ over the sample period in any way, the parameters of the model to be estimated (perhaps inconsistently) from the data, and the forecasts to commence from initial conditions (denoted by $\hat{\mathbf{x}}_T$) which may differ from the ‘true’ values \mathbf{x}_T , nevertheless some useful results can be established. The present analysis will highlight the role of changes in deterministic terms.

Write the closed system (5) in $I(0)$ space using the n variables \mathbf{y}_t (r of which are $\beta' \mathbf{x}_{t-1}$ and $n - r$ are $\Delta \mathbf{x}_t$) as:⁸

$$\mathbf{y}_t = \boldsymbol{\phi} + \boldsymbol{\Pi} \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t \text{ with } \boldsymbol{\epsilon}_t \sim \text{IN}_n [\mathbf{0}, \boldsymbol{\Omega}_\epsilon], \quad (37)$$

where by construction, the unconditional mean of \mathbf{y}_t is:

$$\text{E} [\mathbf{y}_t] = (\mathbf{I}_n - \boldsymbol{\Pi})^{-1} \boldsymbol{\phi} = \boldsymbol{\varphi} \quad (38)$$

so:

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

This formulation is convenient for distinguishing between changes that induce biased forecasts, and those that do not.

The h -step ahead forecasts at time T for $h = 1, \dots, H$ are (using $\hat{\boldsymbol{\varphi}} = (\mathbf{I}_n - \hat{\boldsymbol{\Pi}})^{-1} \hat{\boldsymbol{\phi}}$):

$$\hat{\mathbf{y}}_{T+h} - \hat{\boldsymbol{\varphi}} = \hat{\boldsymbol{\Pi}} (\hat{\mathbf{y}}_{T+h-1} - \hat{\boldsymbol{\varphi}}) = \hat{\boldsymbol{\Pi}}^h (\hat{\mathbf{y}}_T - \hat{\boldsymbol{\varphi}}), \quad (39)$$

⁸The identities determining the future values of the cointegrating vectors are omitted for simplicity: see Hendry and Doornik (1994). The analysis is conditional on assuming such vectors remain unchanged, so the transformed system remains $I(0)$ after the structural change.

where ‘ $\hat{\cdot}$ ’s on parameters denote estimates, and on random variables, forecasts. Although the initial condition is uncertain, we assume $E[\hat{y}_T] = \varphi$ so on average it is unbiased.

Prior to forecasting, $(\phi : \Pi)$ changes to $(\phi^* : \Pi^*)$ where Π^* still has all its eigenvalues less than unity in absolute value, so from $T + 1$ the data are generated by:

$$\mathbf{y}_{T+1} = \phi^* + \Pi^* \mathbf{y}_T + \epsilon_{T+1}.$$

Letting $\phi^* = (\mathbf{I}_n - \Pi^*) \varphi^*$:

$$\mathbf{y}_{T+h} - \varphi^* = \Pi^* (\mathbf{y}_{T+h-1} - \varphi^*) + \epsilon_{T+h} = (\Pi^*)^h (\mathbf{y}_T - \varphi^*) + \sum_{i=0}^{h-1} (\Pi^*)^i \epsilon_{T+h-i}. \quad (40)$$

Many other factors could induce serious forecast errors such as large blips (e.g., 1968(1) and (2) for consumers’ expenditure in the UK), but we construe the first blip to be the above change in the mean, followed by a second shift superimposed thereon. From (39) and (40), the h -step ahead forecast error $\hat{\epsilon}_{T+h} = \mathbf{y}_{T+h} - \hat{y}_{T+h}$ is:

$$\hat{\epsilon}_{T+h} = \varphi^* - \hat{\varphi} + (\Pi^*)^h (\mathbf{y}_T - \varphi^*) - \hat{\Pi}^h (\hat{y}_T - \hat{\varphi}) + \sum_{i=0}^{h-1} (\Pi^*)^i \epsilon_{T+h-i}. \quad (41)$$

Deviations between sample estimates and population parameters are denoted by $\delta_\varphi = \hat{\varphi} - \varphi_p$, where $\varphi_p = (\mathbf{I}_n - \Pi_p)^{-1} \phi_p$, and $\delta_\Pi = \hat{\Pi} - \Pi_p$, with $(\hat{y}_T - \mathbf{y}_T) = \delta_y$. To obtain a clearer interpretation of the various sources of forecast errors, we ignore powers and cross-products in the δ s for parameters, but not those involving parameters interacting with initial conditions. Appendix B details the derivation, and defines \mathbf{C}_h and \mathbf{F}_h .

Table I combines the possible sources of forecast errors that arise from the above decompositions.

Table I: Forecast error taxonomy

$$\begin{aligned} \hat{\epsilon}_{T+h} \simeq & \left((\Pi^*)^h - \Pi^h \right) (\mathbf{y}_T - \varphi) && (ia) \text{ slope change} \\ & + \left(\mathbf{I}_n - (\Pi^*)^h \right) (\varphi^* - \varphi) && (ib) \text{ equilibrium-mean change} \\ & + \left(\Pi^h - \Pi_p^h \right) (\mathbf{y}_T - \varphi) && (iia) \text{ slope mis-specification} \\ & + \left(\mathbf{I}_n - \Pi_p^h \right) (\varphi - \varphi_p) && (iib) \text{ equilibrium-mean mis-specification} \\ & - \mathbf{F}_h \delta_\Pi^\nu && (iiia) \text{ slope estimation} \\ & - \left(\mathbf{I}_n - \Pi_p^h \right) \delta_\varphi && (iiib) \text{ equilibrium-mean estimation} \\ & - \left(\Pi_p^h + \mathbf{C}_h \right) \delta_y && (iv) \text{ initial condition uncertainty} \\ & + \sum_{i=0}^{h-1} (\Pi^*)^i \epsilon_{T+h-i} && (v) \text{ error accumulation.} \end{aligned}$$

In the present formulation, the second and fourth rows alone induce biases, whereas the remainder only affect forecast-error variances. The role of econometrics in reducing each of the forecast errors is discussed in Clements and Hendry (1994) and Hendry and Clements (1994a): here we concentrate on the consequences and properties of breaks.

First, we establish the approximate unbiasedness of forecasts based on estimates of Π in equation (37) when the model is correctly specified and $\phi = \mathbf{0}$, even though $E[\delta_\Pi^\nu] \neq \mathbf{0}$. Since $\epsilon_t \sim \text{IN}_n[\mathbf{0}, \Omega_\epsilon]$, an antithetic-variate argument based on normality (or more generally, any symmetric error distribution: see e.g., Hendry and Trivedi, 1972) can exploit $P(\epsilon_t) = P(-\epsilon_t)$. Since $\mathbf{y}_t(\epsilon_t) = -\mathbf{y}_t(-\epsilon_t)$ in VARs,

whereas $\widehat{\Pi}$ is an even function of ϵ_t , so $\widehat{\Pi}(\{\epsilon_t\}) = \widehat{\Pi}(\{-\epsilon_t\})$, denoting forecast errors by $\hat{\epsilon}$ and $\tilde{\epsilon}$ respectively when the generating process is $\{\epsilon_t\}$ ($\{-\epsilon_t\}$), for 1-step ahead:

$$\hat{\epsilon}_{T+1} = (\Pi - \widehat{\Pi}) y_T + \epsilon_{T+1} \quad \text{and} \quad \tilde{\epsilon}_{T+1} = -(\Pi - \widehat{\Pi}) y_T - \epsilon_{T+1} \quad (42)$$

and these average to zero for every possible error process drawing. Although the system is a dynamic process, the parameter estimate bias does not necessarily bias the forecasts. When $E[\delta_\phi] \neq \mathbf{0}$, or for more complicated processes, the result is not exact, but suggests that finite-sample forecast error biases are unlikely to be the most serious problem.

Next, taking expectations in Table I, assuming such finite-sample biases are negligible:

$$E[\hat{\epsilon}_{T+h}] = (\mathbf{I}_n - (\Pi^*)^h) (\varphi^* - \varphi) + (\mathbf{I}_n - \Pi_p^h) (\varphi - \varphi_p). \quad (43)$$

However, almost all estimation methods ensure that residuals have zero means in-sample, in which case the second term is zero by construction. Then, forecasts will be biased only to the extent that the long-run mean shifts from the in-sample population value. The forecast-error bias is zero for mean-zero processes ($\varphi_p = \varphi^* = \mathbf{0}$), or when shifts in ϕ^* offset those in Π^* to leave φ unaffected ($\varphi^* = \varphi$). Only direct or induced shifts in the deterministic factors lead to serious forecast biases. Moreover, such effects do not die out as the horizon increases, but converge to the full impact of the shift. There are variance effects as well, which are detrimental on a MSFE basis; and the ex ante forecast-error variance estimates will mis-estimate those ruling ex post, but these problems seem likely to be dominated by mean-shift structural breaks. Although non-linearities, asymmetric errors, or roots moving onto the unit circle could generate more complicated outcomes, the basic finding points up the requirement for eliminating systematic forecast-error biases.

By way of contrast, changes in the dynamics, and dynamic parameter mis-specifications, are both multiplied by mean-zero terms, so vanish on average: indeed, they would have no effect whatever on the forecast errors if the initial condition equalled the equilibrium mean. Conversely, the larger the disequilibrium at the time of a shift in slope, the larger the resulting impact; but as Π^* , Π and Π_p (by assumption) have all their roots inside the unit circle, these effects die out as the horizon expands.

At the time when the parameter change occurs, forecasts will be incorrect from almost any statistical procedure. However, consider the period following the break. Forecasting $T + 2$ from $T + 1$ using (39) will generate the same bias as $T + 1$ (given by (43) for $h = 1$). However, some robustness to regime shifts which would otherwise bias forecasts can be obtained either by intercept corrections (ICs) that carry forward the shift from time $T + 1$; or by suitable differencing to eliminate the changed intercept in later periods. For the former, at time $T + 1$ to forecast time $T + 2$, subtracting the previous forecast error will on average produce unbiased forecasts (at a cost in forecast-error variance):

$$E[\hat{\epsilon}_{T+2} | \mathcal{I}_{T+1}] - E[\hat{\epsilon}_{T+1} | \mathcal{I}_T] = \mathbf{0}. \quad (44)$$

This IC again differences the forecast errors that would otherwise have been made, but possibly at a large cost in terms of increased forecast error variance for larger values of h : see Clements and Hendry (1996c).

Alternatively, first differencing the $I(0)$ data produces the naive forecast $\Delta \tilde{y}_{T+2} = \mathbf{0}$, or $\tilde{y}_{T+2} = y_{T+1}$, and as $E[y_{T+h}] = \varphi^* + (\Pi^*)^h (\varphi - \varphi^*)$, the mean forecast error is:

$$\begin{aligned} E[\tilde{\epsilon}_{T+2}] &= E[y_{T+2} - \tilde{y}_{T+2}] \\ &= (\Pi^*)^2 (\varphi - \varphi^*) - \Pi^* (\varphi - \varphi^*) \\ &= \Pi^* (\mathbf{I}_n - \Pi^*) (\varphi^* - \varphi). \end{aligned}$$

The bias vanishes at $\Pi^* = \mathbf{0}$ and is smaller than (43) in general. Thus, once again, a non-causal variable (\mathbf{y}_{t-1}) can dominate on some forecast-accuracy measure; and overdifferencing need not be disadvantageous (noting that some of the $\Delta \mathbf{y}_t$ are second-differenced \mathbf{x}_t).

Clements and Hendry (1996c) show that a VEqCM and DVAR for (5) have identical forecast-error biases when a forecast is made before a break occurs for a horizon that includes the break. This is so despite the former including, and the latter excluding, all the cointegration information; however, their forecast-error variances will differ. The biases for the VEqCM do not depend on whether the forecast starts pre or post the break: thus, there is no error correction after the break. However, the DVAR has different biases pre and post for breaks in α and μ , and these are usually smaller than the corresponding biases from the VEqCM. A forecasting model like $\Delta^2 \tilde{\mathbf{x}}_{T+1} = \Delta^2 \mathbf{x}_T$ seems robust to many of the shifts, but may ‘over-insure’ by not predicting any developments of interest, merely ‘tracking’ by never being badly wrong.

Although such devices can improve forecast accuracy, especially on bias measures, they entail nothing about the usefulness for other purposes of the forecasting model. Even if the resulting (intercept corrected or differenced) forecast is more accurate than that from (39), this does not imply choosing the ‘robustified’ model for policy, or later modelling exercises. If any policy changes were implemented on the basis of mechanistic forecasts, the latter would have the odd property of continuing to predict the same outcome however large the policy response. Thus, there may be benefits to pooling robust predictors with forecasts from econometric systems in the policy context, and this is a hypothesis to consider if encompassing fails: see Hendry and Mizon (1996).

10 Co-breaking

At this point in the analysis, econometric systems do not seem to be doing well relative to rather naive methods when the objective is forecasting. We cannot establish the relevance of causal information, yet can show cases where such systems perform poorly. Parsimony has not been demonstrated as uniformly beneficial, and ad hoc adjustments such as intercept corrections appear to be of value. We now introduce an alternative approach to handling structural breaks that seems well suited to econometric systems, and transpires to link back closely to cointegration.

Co-breaking is defined as the cancellation of breaks across linear combinations of variables. The breaks could be changes in any of α , β , γ , or μ as before, but again we focus on one-off shifts in the last two (so Υ stays constant). Thus, the relevant regime shifts are $\mu^* = \mu + \nabla \mu$ and $\gamma^* = \gamma + \nabla \gamma$ where:

$$\Delta \mathbf{x}_{T+1} = \gamma^* + \alpha (\beta' \mathbf{x}_T - \mu^*) + \nu_{T+1}. \quad (45)$$

Then from (45):

$$\begin{aligned} \Delta \mathbf{x}_{T+1} &= [\gamma + \alpha (\beta' \mathbf{x}_T - \mu) + \nu_{T+1}] + [\nabla \gamma - \alpha \nabla \mu] \\ &= \widetilde{\Delta \mathbf{x}_{T+1}} + [\nabla \gamma - \alpha \nabla \mu]. \end{aligned} \quad (46)$$

The first term is the constant-parameter value of $\Delta \mathbf{x}_{T+1}$ and the second term is the composite intercept shift. Form the m linear combinations $\phi' \Delta \mathbf{x}_{T+1}$:

$$\phi' \Delta \mathbf{x}_{T+1} = \phi' \widetilde{\Delta \mathbf{x}_{T+1}} + \phi' [\nabla \gamma - \alpha \nabla \mu]. \quad (47)$$

Then m -dimensional equilibrium-mean co-breaking requires that $\phi' \alpha \nabla \mu = \mathbf{0}$; whereas q -dimensional drift co-breaking requires $\phi' \nabla \gamma = \mathbf{0}$. We now establish that each is almost certain to occur, so a subsystem will be independent of the breaks (see Hendry, 1995c, for a general formulation and analysis of co-breaking).

First, ‘common trends’ are equilibrium-mean co-breaking. Since $\alpha'_{\perp} \alpha = \mathbf{0}$, when $\nabla \gamma = \mathbf{0}$:

$$\alpha'_{\perp} \Delta \mathbf{x}_{T+1} = \alpha'_{\perp} \gamma + \alpha'_{\perp} \nu_{T+1}. \quad (48)$$

Thus, the $n - r$ dimensional subset $\alpha'_{\perp} \Delta \mathbf{x}_{T+1}$ is unaffected by the shift in the equilibrium mean. This result comes close to explaining the effectiveness of differencing as a ‘solution’ to intercept shifts, noted above. Alternatively, it could be used to characterize ‘common trends’ as those combinations not affected by shifts in underlying equilibria.

Secondly, cointegrating vectors are drift co-breaking when they are trend free (i.e., $\beta' \gamma = \beta' \gamma^* = \mathbf{0}$). Thus, premultiply (45) by β' when $\nabla \mu = \mathbf{0}$:

$$\beta' \Delta \mathbf{x}_{T+1} = \beta' \alpha (\beta' \mathbf{x}_T - \mu) + \beta' \nu_{T+1}, \quad (49)$$

thereby eliminating the shift in the drift parameter. Providing $\beta' \alpha \neq \mathbf{0}$, the resulting subsystem is unaffected by changes in the trend rates of growth in the economy. Conversely, a vector that eliminated such breaks would look like a cointegration vector, so discrimination between cointegration and co-breaking may not be easy (and may be perhaps unnecessary). Thus, a subset of equations remains constant despite the break. If these were really ‘structural’ (see e.g., Hendry, 1995b) they would continue to be constant across further breaks. The remaining variables to be forecast would need extra differencing or intercept corrections pending further developments.

Consequently, there are concepts that suggest that econometric systems may yet prove a superior vehicle even in the forecasting context. We now apply this notion to help account for the potential benefits of econometric systems over leading indicators, precisely because the former involves causal information and the latter does not.

11 Forecasting using leading indicators

Indices of leading indicators are often used in both forecasting and in macro-economic modelling: see, inter alia, Artis, Bladen-Hovell, Osborn, Smith and Zhang (1993), Diebold and Rudebusch (1989, 1991a, 1991b), Lahiri and Moore (1991), Neftci (1979), Stock and Watson (1989, 1992), and Zarnowitz and Braun (1992). However, the procedures used to select the components of composite leading indicators (CLIs), and construct the resulting indices, are altered frequently. It is proposed in Emerson and Hendry (1994) that this phenomenon may be due to ignoring issues of cointegration within the indices, and between their components and macro-economic variables, and to the possibility that non-causal indicator systems are unlikely to co-break to mitigate regime shifts.

First, an indicator is any variable believed informative about another variable of interest; an index is a weighted average of a set of component indicators. A leading indicator is any variable whose known outcome occurs in advance of a variable that it is desired to forecast; a CLI is a combination of such variables. The Harvard A-B-C curves were the earliest construction meant to serve as a prediction system (see Persons, 1924).

Co-breaking provides conditions under which regime shifts vanish for linear combinations of variables. Such an outcome either entails a coincidentally equal effect, or a genuine relationship: across many breaks, the former seems unlikely, leading to poor forecasts. This highlights the key distinction between an indicator – which is non-causal of, and non-caused by, the target – and a causally-related variable. The former is unlikely to systematically experience the same shifts as the target. Thus, although CLIs and econometric models face similar problems for ex ante forecasting in a world of regime shifts, CLIs should suffer relatively when the mappings of the indicators to the outcomes are not causal relations.

Non-constant processes pose problems for all forecasting approaches. In leading-indicator methods, breaks can occur for reasons that would not affect an econometric model which embodied co-breaking relations. For example, the demand for money alters for portfolio reasons as well as transactions changes, so the demand function could be stable for a regime shift in interest rate policy, yet the correlation with GNP be unstable. Conversely, CLIs with a causal basis, that are both cointegrated with, and co-break with, the target, will maintain constant relationships with that target and hence provide a useful forecasting procedure, but one that is tantamount to an econometric model.

11.1 CLIs added to econometric systems

Next, consider adding a CLI to a DVAR for the DGP in (5):

$$\Delta \mathbf{x}_t = \boldsymbol{\theta} + \boldsymbol{\Theta} \Delta \mathbf{x}_{t-1} + \mathbf{v}_t. \quad (50)$$

As shown above, such approximations can be robust to regime shifts, particularly in $\boldsymbol{\mu}$ (see Clements and Hendry, 1995b). Denote the CLI by $c_{t-1} = \boldsymbol{\varphi}' \mathbf{x}_{t-1}$ and add it to (50), taking the special case $\boldsymbol{\Theta} = \mathbf{0}$ for simplicity, so that the augmented DVAR becomes:

$$\Delta \mathbf{x}_t = \boldsymbol{\delta} + \boldsymbol{\rho} c_{t-1} + \mathbf{u}_t. \quad (51)$$

When $\boldsymbol{\varphi}' = \mathbf{h}' \boldsymbol{\beta}'$ (where \mathbf{h} is $r \times 1$), the CLI is a cointegrating combination, so $\{\mathbf{u}_t\}$ will be stationary. Generally $\boldsymbol{\rho} \neq \mathbf{0}$, with $\boldsymbol{\rho} \boldsymbol{\varphi}'$ of rank 1 and $E[c_t] = \mathbf{h}' \boldsymbol{\mu}$ (if $\boldsymbol{\varphi}'$ does not cointegrate \mathbf{x}_t , $\{\mathbf{u}_t\}$ will be stationary only if $\boldsymbol{\rho} = \mathbf{0}$). When $\boldsymbol{\Theta} \neq \mathbf{0}$, both $\Delta \mathbf{x}_{t-1}$ and c_{t-1} will proxy the omitted cointegrating vectors, probably reducing the significance of the latter. However, if c_{t-1} does enter empirically, the intercept in (51) will depend on $\boldsymbol{\mu}$. Similar reasoning applies to a CLI based on differenced data.

There are interesting implications of this analysis under regime shifts in $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$. The CLI will improve forecasting performance relative to the DVAR only if co-breaking occurs using the CLI's weights. Since $\boldsymbol{\beta}$ is co-breaking for shifts in the growth rate $\boldsymbol{\gamma}$ when the cointegrating vectors do not trend (so $\boldsymbol{\beta}' \boldsymbol{\gamma} = \mathbf{0}$), whereas $\{\mathbf{u}_t\}$ depends on $\boldsymbol{\gamma}$, a CLI will be a poor proxy in such a state of nature relative to the correct cointegrating vectors. Further, although the DVAR is little affected by changes in $\boldsymbol{\mu}$, the CLI-based model in (51) depends on $\boldsymbol{\mu}$ and will experience similar predictive failure to the VEqCM, so the robustness of (50) to shifts in the equilibrium mean is lost as well. Emerson and Hendry (1994) offer some empirical evidence supporting this argument.

12 Conclusion

Despite the relative weak assumptions that the economy under analysis is non-stationary and subject to unanticipated structural breaks, that the model may differ from the mechanism in unknown ways, and that it requires estimation from available data, many useful insights can be derived. The resulting implications often differ considerably from those derived when the model is assumed to coincide with a constant mechanism. The fundamental concepts of predictability and forecastability point towards many of the general problems confronting successful forecasting. Causal information cannot be shown to uniformly dominate non-causal in such a setting, and there are no unique measures of forecast accuracy although some measures are not even invariant across isomorphic model representations. Also, intercept corrections have a theoretical justification in a world subject to structural breaks of unknown form, size, and timing by 'robustifying' forecasts against deterministic shifts. Such ICs reveal that the best forecasting model is not necessarily the best policy model.

There is a case for increased parsimony when making multi-step forecasts in constant-parameter worlds, but it is weak, particularly given the absence of any role for collinearity. However, even if the forecasting model remains constant, a break in the correlation structure of the regressors can induce poor forecasts due to variance effects from the least significant variables retained, consistent with the need to eliminate non-systematic effects.

The taxonomy of sources of forecast error clarifies the roles of model mis-specification, sampling variability, error accumulation, initial condition mis-measurement, intercept shifts, and slope-parameter changes. While structural breaks reduce forecastability, the consequences of many forms of break can be derived analytically. Further, models may be differentially susceptible to structural breaks, as shown analytically for VEqCMs and DVARs.

Co-breaking suggests the possibility of eliminating structural breaks by taking linear combinations of variables, which may help produce more robust subsystems. Finally, while leading indicators based on ex post correlations should forecast well in constant-parameter processes, they seem unlikely to provide a reliable forecasting approach under structural breaks as their intercepts are not likely to co-break when variables are genuine indicators that are not linked causally to the target. Thus, causal information retains a central role.

In many ways, this research aims to discover why methods such as Box–Jenkins or DVARs ‘work’ when econometric systems fail. Some potential answers have been proposed, namely, that they impose (so do not estimate) unit roots; that their formulation therefore retains the full values of the previous levels of the transformed variables which ensures a form of error correction; they restrict the information used by appealing to parsimony claims, and thereby happen to exclude non-systematic effects that might otherwise swamp forecastability; and overdifferencing removes permanent breaks in deterministic factors (perhaps at the cost of inducing negative moving-average residuals) demonstrated above to be central to avoiding systematic forecast error biases. Their very advantages as forecasting devices that are robust against deterministic shifts mitigate against their use in a policy setting.

The case for continuing to use econometric systems seems to depend in practice on their competing successfully in the forecasting arena. Cointegration, co-breaking, model selection procedures and rigorous testing help, but none of these ensures immunity to predictive failure from new forms of break. Thus, there is a powerful case for adopting more robust forecasting approaches than intercept corrections: a key development must be error-correction methods that do not eliminate other sources of information such as cointegration. An approach that incorporates causal information in the econometric system for co-breaking and policy, but operates with robustified forecasts merits development.

Many of the conclusions entail positive prescriptions for action; many are at first sight counter-intuitive (but then intuition is just the legacy of previously unquestioned beliefs); and many of the difficulties have possible remedies. In particular, we have only just begun to explore the implications of forecasting across structural breaks with mis-specified models, and many surprises undoubtedly await future researchers.

References

- Artis, M. J., Bladen-Hovell, R. C., Osborn, D. R., Smith, J. P., and Zhang, W. (1993). Turning point prediction in the UK: Preliminary results using CSO leading indicators. Presented to the Royal Economic Society Conference, York.
- Baillie, R. T. (1979). The asymptotic mean squared error of multistep prediction from the regression model with autoregressive errors. *Journal of the American Statistical Association*, **74**, 175–184.

- 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.
- Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994). ARCH models. In Engle, R. F., and McFadden, D. (eds.), *The Handbook of Econometrics, Volume 4*, pp. 2959–3038: North-Holland.
- Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.
- Burns, T. (1986). The interpretation and use of economic predictions. In *Proceedings of the Royal Society*, No. A407, pp. 103–125.
- Campos, J. (1992). Confidence intervals for linear combinations of forecasts from dynamic econometric models. *Journal of Policy Modeling*, **14**, 535–560.
- Chong, Y. Y., and Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *Review of Economic Studies*, **53**, 671–690. Reprinted in Granger C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.
- Clements, M. P., and Hendry, D. F. (1993). On the limitations of comparing mean squared forecast errors. *Journal of Forecasting*, **12**, 617–637. With discussion.
- Clements, M. P., and Hendry, D. F. (1994). Towards a theory of economic forecasting. In Hargreaves, C. (ed.), *Non-stationary Time-series Analyses and Cointegration*, pp. 9–52. Oxford: Oxford University Press.
- Clements, M. P., and Hendry, D. F. (1995a). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, **10**, 127–146.
- Clements, M. P., and Hendry, D. F. (1995b). Macro-economic forecasting and modelling. *Economic Journal*, **105**, 1001–1013.
- Clements, M. P., and Hendry, D. F. (1996a). An empirical study of seasonal unit roots in forecasting. *International Journal of Forecasting*. Forthcoming.
- Clements, M. P., and Hendry, D. F. (1996b). Forecasting in macro-economics. in Cox *et al.* (1996), pp. 101–141.
- Clements, M. P., and Hendry, D. F. (1996c). Intercept corrections and structural breaks. *Journal of Applied Econometrics*. Forthcoming.
- Clements, M. P., and Hendry, D. F. (1996d). Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics*. Forthcoming.
- Cook, S. (1995). Treasury economic forecasting. mimeo, Institute of Economics and Statistics, University of Oxford.
- Cox, D. R., Hinkley, D. V., and Barndorff-Nielsen, O. E. (eds.) (1996). *Time Series Models In econometrics, finance and other fields*. London: Chapman and Hall.
- Cox, D. R., and Miller, H. D. (1965). *The Theory of Stochastic Processes*: Chapman and Hall.
- Diebold, F. X., and Rudebusch, G. D. (1989). Scoring the leading indicators. *Journal of Business*, **62**, 369–391.
- Diebold, F. X., and Rudebusch, G. D. (1991a). Forecasting output with the composite leading index: An ex ante 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, and Moore (1991), pp. 231–256.

- Doan, T., Litterman, R., and Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, **3**, 1–100.
- Emerson, R. A., and Hendry, D. F. (1994). An evaluation of forecasting using leading indicators. Mimeo, Nuffield College, Oxford.
- Emerson, R. A., and Hendry, D. F. (1996). An evaluation of forecasting using leading indicators. *Journal of Forecasting*, **15**, 271–91.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflations. *Econometrica*, **50**, 987–1007.
- Ericsson, N. R., and Marquez, J. R. (1989). Exact and approximate multi-period mean-square forecast errors for dynamic econometric models. International finance discussion paper 348, Federal Reserve Board.
- Gilbert, C. L. (1986). Professor Hendry's methodology. *Oxford Bulletin of Economics and Statistics*, **48**, 283–307. Reprinted in Granger C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.
- Granger, C. W. J., and Newbold, P. (1986). *Forecasting Economic Time Series* 2nd edn. New York: Academic Press.
- Harvey, A. C., and Shephard, N. (1992). Structural time series models. In Maddala, G. S., Rao, C. R., and Vinod, H. D. (eds.), *Handbook of Statistics*, Vol. 11. Amsterdam: North-Holland.
- 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. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.
- Hendry, D. F. (1995a). *Dynamic Econometrics*. Oxford: Oxford University Press.
- Hendry, D. F. (1995b). Econometrics and business cycle empirics. *Economic Journal*, **105**, 1622–1636.
- Hendry, D. F. (1995c). A theory of co-breaking. Mimeo, Nuffield College, University of Oxford.
- Hendry, D. F., and Clements, M. P. (1993). On model selection when forecasting. Mimeo, Economics department, University of Oxford.
- Hendry, D. F., and Clements, M. P. (1994a). Can econometrics improve economic forecasting?. *Swiss Journal of Economics and Statistics*, **130**, 267–298.
- Hendry, D. F., and Clements, M. P. (1994b). On a theory of intercept corrections in macro-economic forecasting. In Holly, S. (ed.), *Money, Inflation and Employment: Essays in Honour of James Ball*, pp. 160–182. Aldershot: Edward Elgar.
- Hendry, D. F., and Doornik, J. A. (1994). Modelling linear dynamic econometric systems. *Scottish Journal of Political Economy*, **41**, 1–33.
- Hendry, D. F., and Mizon, G. E. (1996). Selecting econometric models for policy analysis by forecast accuracy. Mimeo, Nuffield College, University of Oxford.
- Hendry, D. F., and Trivedi, P. K. (1972). Maximum likelihood estimation of difference equations with moving-average errors: A simulation study. *Review of Economic Studies*, **32**, 117–145.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, **82**, 35–45.
- Klein, L. R. (1971). *An Essay on the Theory of Economic Prediction*. Chicago: Markham Publishing Company.

- Lahiri, K., and Moore, G. H. (eds.)(1991). *Leading economic indicators: New approaches and forecasting records*. Cambridge: Cambridge University Press.
- Neftci, S. N. (1979). Lead-lag relations, exogeneity and prediction of economic time series. *Econometrica*, **47**, 101–113.
- Pain, N., and Britton, A. (1992). The recent experience of economic forecasting in Britain: some lessons from national institute forecasts. Discussion paper (new series) 20, National Institute.
- Persons, W. M. (1924). *The Problem of Business Forecasting*. No. 6 in Pollak Foundation for Economic Research Publications. London: Pitman.
- Schmidt, P. (1974). The asymptotic distribution of forecasts in the dynamic simulation of an econometric model. *Econometrica*, **42**, 303–309.
- Shephard, N. G. (1996). Statistical aspects of ARCH and stochastic volatility. in Cox *et al.* (1996), pp. 1–67.
- Spanos, A. (1986). *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.
- Stock, J. H., and Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macro-Economic Annual*, 351–409.
- Stock, J. H., and Watson, M. W. (1992). A procedure for predicting recessions with leading indicators: Econometric issues and recent experience. Working paper 4014, NBER.
- Theil, H. (1961). *Economic Forecasts and Policy* 2nd edn. Amsterdam: North-Holland Publishing Company.
- Wallis, K. F. (1989). Macroeconomic forecasting: A survey. *Economic Journal*, **99**, 28–61.
- Wallis, K. F., and Whitley, J. D. (1991). Sources of error in forecasts and expectations: U.K. economic models 1984–8. *Journal of Forecasting*, **10**, 231–253.
- Wold, H. O. A. (1938). *A Study in The Analysis of Stationary Time Series*. Stockholm: Almqvist and Wicksell.
- Zarnowitz, V., and Braun, P. (1992). Major macroeconomic variables and leading indicators: Some estimates of their interrelations, 1886–1982. Working paper 2812, National Bureau of Economic Research, New York.

13 Appendices

13.1 A: Parameter estimates

From the properties of the $I(1)$ variables in the example, the expectations of the estimates of λ_1 and λ_2 in (23) are:

$$\begin{aligned}
 \mathbb{E} \left[\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix} \right] &= \mathbb{E} \left[\begin{pmatrix} T^{-2} \sum_{t=1}^T E_{t-1}^2 & T^{-2} \sum_{t=1}^T E_{t-1} R_{t-1} \\ T^{-2} \sum_{t=1}^T E_{t-1} R_{t-1} & T^{-2} \sum_{t=1}^T R_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} T^{-2} \sum_{t=1}^T E_{t-1} Y_t \\ T^{-2} \sum_{t=1}^T R_{t-1} Y_t \end{pmatrix} \right] \\
 &\simeq \left[\begin{pmatrix} T^{-2} \sum_{t=1}^T \mathbb{E} [E_{t-1}^2] & 0 \\ 0 & T^{-2} \sum_{t=1}^T \mathbb{E} [R_{t-1}^2] \end{pmatrix}^{-1} \begin{pmatrix} \alpha T^{-2} \sum_{t=1}^{T_1} \mathbb{E} [E_{t-1}^2] \\ \beta T^{-2} \sum_{t=T_1+1}^T \mathbb{E} [R_{t-1}^2] \end{pmatrix} \right] \\
 &= \left[\begin{pmatrix} T^{-2} \sum_{t=1}^T (t-1) & 0 \\ 0 & T^{-2} \sum_{t=1}^T (t-1) \end{pmatrix}^{-1} \begin{pmatrix} \alpha T^{-2} \sum_{t=1}^{T_1} (t-1) \\ \beta T^{-2} \sum_{t=T_1+1}^T (t-1) \end{pmatrix} \right]
 \end{aligned}$$

$$\begin{aligned}
&= \left[\begin{pmatrix} 2T(T-1)^{-1} & 0 \\ 0 & 2T(T-1)^{-1} \end{pmatrix} \begin{pmatrix} \frac{1}{2}\alpha T^{-2}T_1(T_1-1) \\ \frac{1}{2}\beta T^{-2}(T(T-1)-T_1(T_1-1)) \end{pmatrix} \right] \\
&= \begin{pmatrix} K\alpha \\ (1-K)\beta \end{pmatrix},
\end{aligned}$$

where:

$$K = \frac{T_1(T_1-1)}{T(T-1)}.$$

13.2 B: Taxonomy derivations

First, we make the approximation:

$$\hat{\Pi}^h = (\Pi_p + \delta_{\Pi})^h \simeq \Pi_p^h + \sum_{i=0}^{h-1} \Pi_p^i \delta_{\Pi} \Pi_p^{h-i-1} = \Pi_p^h + \mathbf{C}_h. \quad (52)$$

Also, using (52), letting $(\cdot)^\nu$ denote forming a vector and \otimes a Kronecker product:

$$\mathbf{C}_h(\mathbf{y}_T - \varphi_p) = (\mathbf{C}_h(\mathbf{y}_T - \varphi_p))^\nu = \left(\sum_{i=0}^{h-1} \Pi_p^i \otimes (\mathbf{y}_T - \varphi_p)' \Pi_p^{h-i-1} \right) \delta_{\Pi}^\nu = \mathbf{F}_h \delta_{\Pi}^\nu.$$

To highlight components due to different effects (parameter change, estimation inconsistency, and estimation uncertainty), we decompose the terms $(\Pi^*)^h(\mathbf{y}_T - \varphi^*)$ and $\hat{\Pi}^h(\hat{\mathbf{y}}_T - \hat{\varphi})$ in (41) into:

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

and:

$$\begin{aligned}
\hat{\Pi}^h(\hat{\mathbf{y}}_T - \hat{\varphi}) &= \left(\Pi_p^h + \mathbf{C}_h \right) (\delta_y - (\hat{\varphi} - \varphi_p) + (\mathbf{y}_T - \varphi) - (\varphi_p - \varphi)) \\
&= \left(\Pi_p^h + \mathbf{C}_h \right) \delta_y - \left(\Pi_p^h + \mathbf{C}_h \right) \delta_\varphi + \left(\Pi_p^h + \mathbf{C}_h \right) (\mathbf{y}_T - \varphi_p) \\
&= \left(\Pi_p^h + \mathbf{C}_h \right) \delta_y - \left(\Pi_p^h + \mathbf{C}_h \right) \delta_\varphi + \mathbf{F}_h \delta_{\Pi}^\nu + \Pi_p^h (\mathbf{y}_T - \varphi) - \Pi_p^h (\varphi_p - \varphi)
\end{aligned}$$

so $(\Pi^*)^h(\mathbf{y}_T - \varphi^*) - \hat{\Pi}^h(\hat{\mathbf{y}}_T - \hat{\varphi})$ yields:

$$\begin{aligned}
&\left((\Pi^*)^h - \Pi_p^h \right) (\mathbf{y}_T - \varphi) - \mathbf{F}_h \delta_{\Pi}^\nu - \left(\Pi_p^h + \mathbf{C}_h \right) \delta_y \\
&- (\Pi^*)^h (\varphi^* - \varphi) + \Pi_p^h (\varphi_p - \varphi) + \left(\Pi_p^h + \mathbf{C}_h \right) \delta_\varphi.
\end{aligned} \quad (53)$$

The interaction $\mathbf{C}_h \delta_\varphi$ is like a ‘covariance’, but is omitted from the table. Hence (53) becomes:

$$\begin{aligned}
&\left((\Pi^*)^h - \Pi_p^h \right) (\mathbf{y}_T - \varphi) + \left(\Pi_p^h - \Pi_p^h \right) (\mathbf{y}_T - \varphi) \\
&- (\Pi^*)^h (\varphi^* - \varphi) + \Pi_p^h (\varphi_p - \varphi) \\
&- \left(\Pi_p^h + \mathbf{C}_h \right) \delta_y - \mathbf{F}_h \delta_{\Pi}^\nu + \Pi_p^h \delta_\varphi.
\end{aligned}$$

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

Finally, for the term $\varphi^* - \hat{\varphi}$ in (41), we have (on the same principle):

$$(\varphi^* - \varphi) + (\varphi - \varphi_p) - \delta_\varphi.$$

14 Alternative derivation

This takes the initial parameterization and derives the result to check the answer. From:

$$\mathbf{y}_t = \boldsymbol{\phi} + \boldsymbol{\Pi} \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t \quad \text{with} \quad \boldsymbol{\epsilon}_t \sim \text{IN}_n[\mathbf{0}, \boldsymbol{\Omega}_\epsilon], \quad (54)$$

the h -step recursion is:

$$\begin{aligned} \mathbf{y}_{T+h} &= \boldsymbol{\phi} + \boldsymbol{\Pi} \mathbf{y}_{T+h-1} + \boldsymbol{\epsilon}_{T+h} \\ &= \sum_{i=0}^{h-1} \boldsymbol{\Pi}^i \boldsymbol{\phi} + \boldsymbol{\Pi}^h \mathbf{y}_T + \sum_{i=0}^{h-1} \boldsymbol{\Pi}^i \boldsymbol{\epsilon}_{T+h-i}. \end{aligned}$$

Now:

$$\sum_{i=0}^{h-1} \boldsymbol{\Pi}^i = \mathbf{I}_n + \boldsymbol{\Pi} + \dots + \boldsymbol{\Pi}^{h-1} = (\mathbf{I}_n - \boldsymbol{\Pi}^h) (\mathbf{I}_n - \boldsymbol{\Pi})^{-1}$$

as:

$$\sum_{i=0}^{h-1} \boldsymbol{\Pi}^i (\mathbf{I}_n - \boldsymbol{\Pi}) = (\mathbf{I}_n + \boldsymbol{\Pi} + \dots + \boldsymbol{\Pi}^{h-1}) (\mathbf{I}_n - \boldsymbol{\Pi}) = \mathbf{I}_n - \boldsymbol{\Pi}.$$

After break, get:

$$\begin{aligned} \mathbf{y}_{T+h} &= \sum_{i=0}^{h-1} (\boldsymbol{\Pi}^*)^i \boldsymbol{\phi}^* + (\boldsymbol{\Pi}^*)^h \mathbf{y}_T + \sum_{i=0}^{h-1} (\boldsymbol{\Pi}^*)^i \boldsymbol{\epsilon}_{T+h-i} \\ &= (\mathbf{I}_n - \boldsymbol{\Pi}^*)^{-1} (\mathbf{I}_n - (\boldsymbol{\Pi}^*)^h) \boldsymbol{\phi}^* + (\boldsymbol{\Pi}^*)^h \mathbf{y}_T + \sum_{i=0}^{h-1} (\boldsymbol{\Pi}^*)^i \boldsymbol{\epsilon}_{T+h-i}. \end{aligned}$$

Next, the model forecasts:

$$\hat{\mathbf{y}}_{T+h} = \sum_{i=0}^{h-1} \hat{\boldsymbol{\Pi}}^i \hat{\boldsymbol{\phi}} + \hat{\boldsymbol{\Pi}}^h \hat{\mathbf{y}}_T = (\mathbf{I}_n - \hat{\boldsymbol{\Pi}})^{-1} (\mathbf{I}_n - \hat{\boldsymbol{\Pi}}^h) \hat{\boldsymbol{\phi}} + \hat{\boldsymbol{\Pi}}^h \hat{\mathbf{y}}_T.$$

so forecast error is:

$$\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h} = \sum_{i=0}^{h-1} (\boldsymbol{\Pi}^*)^i \boldsymbol{\phi}^* - \sum_{i=0}^{h-1} \hat{\boldsymbol{\Pi}}^i \hat{\boldsymbol{\phi}} + ((\boldsymbol{\Pi}^*)^h - \hat{\boldsymbol{\Pi}}^h) \mathbf{y}_T + \hat{\boldsymbol{\Pi}}^h (\mathbf{y}_T - \hat{\mathbf{y}}_T) + \sum_{i=0}^{h-1} (\boldsymbol{\Pi}^*)^i \boldsymbol{\epsilon}_{T+h-i}.$$

Need a mechanism for approximating stochastic properties of $\hat{\boldsymbol{\Pi}}^h$. Use:

$$\hat{\boldsymbol{\Pi}}^h = (\boldsymbol{\Pi}_p + \boldsymbol{\delta}_\Pi)^h \simeq \boldsymbol{\Pi}_p^h + \sum_{i=0}^{h-1} \boldsymbol{\Pi}_p^i \boldsymbol{\delta}_\Pi \boldsymbol{\Pi}_p^{h-i-1} = \boldsymbol{\Pi}_p^h + \mathbf{C}_h. \quad (55)$$

Also, neglect interaction, so:

$$\mathbf{C}_h \mathbf{y}_T \simeq \mathbf{C}_h \boldsymbol{\varphi}$$

as $E[\mathbf{y}_T] = \boldsymbol{\varphi}$. Then:

$$\hat{\boldsymbol{\Pi}}^h (\mathbf{y}_T - \hat{\mathbf{y}}_T) = -(\boldsymbol{\Pi}_p^h + \mathbf{C}_h) \boldsymbol{\delta}_y.$$

and:

$$\begin{aligned} \left((\mathbf{\Pi}^*)^h - \widehat{\mathbf{\Pi}}^h \right) \mathbf{y}_T &= \left((\mathbf{\Pi}^*)^h - \mathbf{\Pi}_p^h - \mathbf{C}_h \right) \mathbf{y}_T \\ &\simeq \left((\mathbf{\Pi}^*)^h - \mathbf{\Pi}_p^h \right) (\mathbf{y}_T - \boldsymbol{\varphi}) + \left((\mathbf{\Pi}^*)^h - \mathbf{\Pi}_p^h \right) \boldsymbol{\varphi} - \mathbf{C}_h \boldsymbol{\varphi}. \end{aligned}$$

Further:

$$\mathbf{C}_h \boldsymbol{\varphi} = (\mathbf{C}_h \boldsymbol{\varphi})^\nu = \left(\sum_{i=0}^{h-1} \mathbf{\Pi}_p^i \otimes \boldsymbol{\varphi}' \mathbf{\Pi}_p^{h-i-1'} \right) \boldsymbol{\delta}_{\Pi}^\nu = \mathbf{F}_h \boldsymbol{\delta}_{\Pi}^\nu.$$

The most complex term is:

$$\begin{aligned} \sum_{i=0}^{h-1} (\mathbf{\Pi}^*)^i \boldsymbol{\phi}^* - \sum_{i=0}^{h-1} \widehat{\mathbf{\Pi}}^i \widehat{\boldsymbol{\phi}} &= \left(\mathbf{I}_n - (\mathbf{\Pi}^*)^h \right) \left(\mathbf{I}_n - (\mathbf{\Pi}^*) \right)^{-1} \boldsymbol{\phi}^* - \sum_{i=0}^{h-1} \left(\mathbf{\Pi}_p^i + \mathbf{C}_i \right) \left(\boldsymbol{\phi}_p + \boldsymbol{\delta}_\phi \right) \\ &= \left(\mathbf{I}_n - (\mathbf{\Pi}^*)^h \right) \boldsymbol{\varphi}^* - \sum_{i=0}^{h-1} \mathbf{\Pi}_p^i \boldsymbol{\phi}_p - \mathbf{D}_i \boldsymbol{\phi}_p - \sum_{i=0}^{h-1} \mathbf{\Pi}_p^i \boldsymbol{\delta}_\phi - \sum_{i=0}^{h-1} \mathbf{C}_i \boldsymbol{\delta}_\phi \\ &= \left(\mathbf{I}_n - (\mathbf{\Pi}^*)^h \right) \boldsymbol{\varphi}^* - \left(\mathbf{I}_n - \mathbf{\Pi}_p^h \right) \boldsymbol{\varphi}_p - \mathbf{D}_i \boldsymbol{\phi}_p - \sum_{i=0}^{h-1} \mathbf{\Pi}_p^i \boldsymbol{\delta}_\phi - \mathbf{D}_i \boldsymbol{\delta}_\phi \end{aligned}$$

where:

$$\mathbf{D}_i = \sum_{i=0}^{h-1} \mathbf{C}_i$$

as:

$$\widehat{\boldsymbol{\phi}} = \boldsymbol{\phi}_p + \widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_p = \boldsymbol{\phi}_p + \boldsymbol{\delta}_\phi.$$

Combining, first the terms in variance components:

$$- \sum_{i=0}^{h-1} \mathbf{\Pi}_p^i \boldsymbol{\delta}_\phi - \mathbf{F}_h \boldsymbol{\delta}_{\Pi}^\nu - \mathbf{D}_i (\boldsymbol{\delta}_\phi + \boldsymbol{\phi}_p)$$

where:

$$\sum_{i=0}^{h-1} \mathbf{\Pi}_p^i \boldsymbol{\delta}_\phi = \left(\mathbf{I}_n - \mathbf{\Pi}_p^h \right) \boldsymbol{\delta}_\phi.$$

The covariance terms $-\mathbf{D}_i (\boldsymbol{\delta}_\phi + \boldsymbol{\phi}_p)$ have been dropped in general: but may not be negligible.

Next, those in slope conditions:

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

and finally in the equilibrium mean factors:

$$\begin{aligned} &\left(\mathbf{I}_n - (\mathbf{\Pi}^*)^h \right) \boldsymbol{\varphi}^* - \left(\mathbf{I}_n - \mathbf{\Pi}_p^h \right) \boldsymbol{\varphi}_p + \left((\mathbf{\Pi}^*)^h - \mathbf{\Pi}_p^h \right) \boldsymbol{\varphi} \\ &= \left(\mathbf{I}_n - (\mathbf{\Pi}^*)^h \right) \boldsymbol{\varphi}^* - \left(\mathbf{I}_n - \mathbf{\Pi}_p^h \right) \boldsymbol{\varphi}_p - \left(\left(\mathbf{I}_n - (\mathbf{\Pi}^*)^h \right) - \left(\mathbf{I}_n - \mathbf{\Pi}_p^h \right) \right) \boldsymbol{\varphi} \\ &= \left(\mathbf{I}_n - (\mathbf{\Pi}^*)^h \right) (\boldsymbol{\varphi}^* - \boldsymbol{\varphi}) - \left(\mathbf{I}_n - \mathbf{\Pi}_p^h \right) (\boldsymbol{\varphi}_p - \boldsymbol{\varphi}). \end{aligned}$$

This delivers the above taxonomy table. Hence biased forecasts unless $\boldsymbol{\varphi}_p = \boldsymbol{\varphi}$. But $\boldsymbol{\varphi}_p \neq \boldsymbol{\varphi}$ would entail non-zero mean residuals in-sample, which seems unlikely.