International Journal of Forecasting 31 (2015) 99–112

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Robust approaches to forecasting

Jennifer L. Castle[a,*], Michael P. Clements[b,1], David F. Hendry[c,2]

[a] Magdalen College and Institute for New Economic Thinking, Oxford Martin School, University of Oxford, UK

[b] ICMA Centre, University of Reading, and Institute for New Economic Thinking, Oxford Martin School, University of Oxford, UK

[c] Economics Department and Institute for New Economic Thinking, Oxford Martin School, University of Oxford, UK

ARTICLE INFO

Keywords:
Forecast biases
Smoothed forecasting devices
Factor models
GDP forecasts
Location shifts

ABSTRACT

We investigate alternative robust approaches to forecasting, using a new class of robust devices, contrasted with equilibrium-correction models. Their forecasting properties are derived facing a range of likely empirical problems at the forecast origin, including measurement errors, impulses, omitted variables, unanticipated location shifts and incorrectly included variables that experience a shift. We derive the resulting forecast biases and error variances, and indicate when the methods are likely to perform well. The robust methods are applied to forecasting US GDP using autoregressive models, and also to autoregressive models with factors extracted from a large dataset of macroeconomic variables. We consider forecasting performance over the Great Recession, and over an earlier more quiescent period.

1. Introduction

In recent times, there has been increased interest in forecasting with diffusion indices and factor models: see e.g., Castle, Clements, and Hendry (2013), Forni, Hallin, Lippi, and Reichlin (2000), Peña and Poncela (2004), Schumacher and Breitung (2008) and Stock and Watson (1989, 1999, 2009).[3] In Castle, Clements, and Hendry (2013), we investigated which approach to forecasting output levels and growth using factors, variables, both, or neither performed best on quarterly data over the Great Recession to 2011. After updating and extending the dataset from Stock and Watson (2009), we used *Autometrics* (as described in Doornik (2009), and Doornik and Hendry (2013)) for in-sample modeling, which allowed all the principal components of the variables as well as the original variables to be included jointly, while also tackling multiple breaks by impulse-indicator saturation (IIS: see Castle, Doornik, & Hendry, 2012, and Johansen & Nielsen, 2009). Forecasting US GDP growth over 1-, 4- and 8-step horizons showed that factor models were somewhat more useful for short-term forecasting, but their relative performance declined as the forecast horizon increased. We found (like many other investigators) that it was difficult to beat scalar autoregressions: Fildes and Stekler (2002) provide a survey of macroeconomic forecasting before the Great Recession, which the follow up in Stekler and Talwar (2011) show was not well predicted. Our own forecasts for GDP levels highlighted the need for robust strategies (such as intercept corrections) when location shifts (i.e., shifts in the previous unconditional mean) occurred. The empirical results were consistent with the forecast-error taxonomy

* Corresponding author. Tel.: +44 1865 276067.
E-mail addresses: jennifer.castle@magd.ox.ac.uk (J.L. Castle), m.p.clements@icmacentre.ac.uk (M.P. Clements), david.hendry@nuffield.ox.ac.uk (D.F. Hendry).

1 Tel.: +44 1183 784377.

2 Tel.: +44 1865 616640.

3 It is a great pleasure to contribute a paper on economic forecasting to an issue of *International Journal of Forecasting* in honor of Herman Stekler, whose many research findings have so greatly advanced our understanding and practice of this most treacherous topic.

http://dx.doi.org/10.1016/j.ijforecast.2014.11.002
0169-2070/

for factor models, which highlighted the impacts of location shifts on systematic forecast-error biases.

In this paper, we develop the theory of forecasting for members of a robust class of forecasting model motivated by Hendry (2006). The robust approach proposed is applicable to factor models and models with variables (as well as hybrids), almost all of which are variants of equilibrium-correction models (EqCMs), but seeks to avoid the systematic forecast failure symptomatic of EqCMs after a location shift. The class of robust forecasting devices we introduce includes that proposed in Hendry (2006) as the most flexible, and hence the most volatile, at one end, with the least flexible, conventional full-sample vector equilibrium-correction model (VEqCM) at the other, having recursive updating, rolling windows, and smoothed robust devices as intermediate cases. We then compare and contrast findings for forecasting US GDP over both a quiescent period (2000(1)–2006(4)) and a period including the Great Recession (2007(1)–2011(2)) to see how well robust forecasting devices motivated by Hendry (2006) perform in the face of breaks.

The structure of the paper is as follows. Section 2 reviews the problems confronting VEqCMs as non-robust forecasting models when unanticipated location shifts occur at or near the forecast origin. Section 2.1 considers changes in dynamics to show they will not by themselves generate systematic forecast failure. Section 3 describes the new class of robust forecasting devices, explains why they can avoid systematic forecast failure, then investigates two members that differ in their smoothness, and compares how they react under: (i) constant parameters in Section 3.1, then to (ii) measurement errors in Section 3.2, (iii) unknown impulses in Section 3.3, (iv) unanticipated location shifts at the forecast origin in Section 3.4, (v) unknowingly omitted variables in Section 3.5, (vi) changing forecast origins after shifts in Section 3.6 and (vii) making longer-horizon forecasts in Section 3.7: Section 3.8 draws some conclusions on robustifying VEqCMs. Section 4 applies that analysis to factor-based models for forecasting facing a location shift. Section 4.1 considers location shifts induced by changes over time in the relevance of 'explanatory variables'. Section 5 presents the empirical analysis forecasting US GDP over the period 2000(1)–2011(2), divided as noted above. Section 6 concludes.

2. Vector equilibrium-correction models

This section introduces our notation and establishes a benchmark by showing that VEqCMs are not robust when forecasting after an unanticipated location shift, so can suffer systematic forecast failure (see e.g., Clements & Hendry, 1999, 2006). Explaining why equilibrium-correction models are susceptible to forecast failure in such circumstances leads us to introduce the new class that is robust.

Consider a data generation process (DGP) given by the first-order open VEqCM for an n-dimensional time series $\{\mathbf{x}_t,\ t = 0, \ldots, T\}$, integrated of first order, denoted $\mathsf{I}(1)$:

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

where $\boldsymbol{\epsilon}_t \sim \mathsf{IN}_n[\mathbf{0}, \boldsymbol{\Omega}_\epsilon]$, denoting an independent normal random vector with mean $\mathsf{E}[\boldsymbol{\epsilon}_t] = \mathbf{0}$ and variance $\mathsf{V}[\boldsymbol{\epsilon}_t] = \boldsymbol{\Omega}_\epsilon$. In addition to lags of the $\mathbf{x}_t$'s, we allow that $\mathbf{x}_t$ may depend on a set of k explanatory variables $\mathbf{z}_t$, which may be $\mathsf{I}(0)$ individual variables and/or factors. As indicated by (1), $\Delta\mathbf{x}_t$ responds to disequilibria between $\mathbf{z}_{t-1}$ and its mean $\mathsf{E}[\mathbf{z}_{t-1}] = \boldsymbol{\kappa}$, so the DGP is also equilibrium-correcting in the $\mathbf{z}_t$'s. However, the omission of $\mathbf{z}_{t-1}$ is not known to the investigator. In (1), both $\Delta\mathbf{x}_t$ and $\boldsymbol{\beta}'\mathbf{x}_t$ are $\mathsf{I}(0)$, with equilibrium mean $\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x}_t\right] = \boldsymbol{\mu}$ and average growth $\mathsf{E}[\Delta\mathbf{x}_t] = \boldsymbol{\gamma}$ in-sample. Then (1) is incorrectly estimated as:

$$\Delta\widehat{\mathbf{x}}_t = \widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_{t-1} - \widehat{\boldsymbol{\mu}}\right) \tag{2}$$

where $\mathsf{E}[\widehat{\boldsymbol{\gamma}}] = \boldsymbol{\gamma}_e$ and $\mathsf{E}[\widehat{\boldsymbol{\mu}}] = \boldsymbol{\mu}_e$, and usually $\boldsymbol{\gamma}_e \neq \boldsymbol{\gamma}$ and $\boldsymbol{\mu}_e \neq \boldsymbol{\mu}$ because of the model mis-specification. In general, there will also be small-sample biases in these estimates, but we ignore these to sharpen the analysis. We also ignore biases and variances in estimating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, as well as changes therein (as discussed in Section 2.1).

Location shifts are the most pernicious problem for forecasting, since when $\boldsymbol{\gamma}$, $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$ shift to $\boldsymbol{\gamma}^*$, $\boldsymbol{\mu}^*$ and $\boldsymbol{\kappa}^*$ at time T, the DGP becomes:

$$\Delta\mathbf{x}_{T+1} = \boldsymbol{\gamma}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}^*\right) + \left(\boldsymbol{\phi}^*\right)'(\mathbf{z}_T - \boldsymbol{\kappa}^*) + \boldsymbol{\epsilon}_{T+1} \tag{3}$$

where we have allowed the coefficient of the omitted vector to change as well, so the 1-step ahead forecast errors from:

$$\Delta\widehat{\mathbf{x}}_{T+1|T} = \widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_T - \widehat{\boldsymbol{\mu}}\right) \tag{4}$$

have a mean of:

$$\left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}_e\right) - \boldsymbol{\alpha}\left(\boldsymbol{\mu}^* - \boldsymbol{\mu}_e\right) - \left(\boldsymbol{\phi}^*\right)'(\boldsymbol{\kappa}^* - \boldsymbol{\kappa}). \tag{5}$$

The extent of forecast failure depends on the magnitudes of the mean shifts in (5), but can be very large (see e.g., Castle, Fawcett, & Hendry, 2010). In fact, using (4) when the DGP is (3) leads to all of the following errors:

(ia) 'deterministic shifts' of $(\boldsymbol{\gamma},\ \boldsymbol{\mu},\ \boldsymbol{\kappa})$ to $(\boldsymbol{\gamma}^*,\ \boldsymbol{\mu}^*\ \boldsymbol{\kappa}^*)$;
(ib) 'stochastic breaks' of $\boldsymbol{\phi}$ to $\boldsymbol{\phi}^*$, although shifts in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are also perfectly possible;
(iia,b) inconsistent parameter estimates $\boldsymbol{\gamma}_e$ and $\boldsymbol{\mu}_e$ (and potentially also in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$) from the unknown omission of $\mathbf{z}_{t-1}$;
(iii) forecast origin uncertainty when $\widehat{\mathbf{x}}_T \neq \mathbf{x}_T$ (considered later though not explicitly included above);
(iva,b) estimation uncertainty from $\mathsf{V}[\widehat{\boldsymbol{\gamma}},\ \widehat{\boldsymbol{\alpha}},\ \widehat{\boldsymbol{\beta}},\ \widehat{\boldsymbol{\mu}}]$;
(v) omitted variables, $\mathbf{z}_T$;
(vi) innovation errors, $\boldsymbol{\epsilon}_{T+1}$.

When (4) is still used to forecast the outcomes from (3) even after several periods, so that:

$$\Delta\widehat{\mathbf{x}}_{T+h|T+h-1} = \widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-1} - \widehat{\boldsymbol{\mu}}\right) \tag{6}$$

then the forecast error $\widehat{\boldsymbol{\epsilon}}_{T+h|T+h-1} = \Delta\mathbf{x}_{T+h} - \Delta\widehat{\mathbf{x}}_{T+h|T+h-1}$ has a persistent bias (even assuming $\mathsf{E}[\mathbf{z}_{T+h-1}] = \boldsymbol{\kappa}^*$) of:

$$\mathsf{E}\left[\widehat{\boldsymbol{\epsilon}}_{T+h|T+h-1}\right] = \left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}_e\right) - \boldsymbol{\alpha}\left(\boldsymbol{\mu}^* - \boldsymbol{\mu}_e\right) \tag{7}$$

so the first two components in (5) continue to cause systematic mis-forecasting into the future until the estimated