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 $\{\mathbf{x}_t,\ t = 0, \dots, T\}$, 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:

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

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 model is revised. The problem is that equilibrium correction always corrects back to the *old* equilibrium, determined by $\boldsymbol{\gamma}$, $\boldsymbol{\mu}$, irrespective of whether or not the equilibrium has shifted. Even in the absence of model misspecification, systematic mis-forecasting will occur when there is a break: using the pre-break, in-sample DGP with the population values of the parameters, the equivalent of (6) becomes:

$$\Delta\widehat{\mathbf{x}}_{T+h|T+h-1} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+h-1} - \boldsymbol{\mu}\right) \tag{8}$$

with a forecast error of:

$$\mathsf{E}\left[\widehat{\boldsymbol{\epsilon}}_{T+h|T+h-1}\right] = \left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}\right) + \boldsymbol{\alpha}\left(\boldsymbol{\mu}^* - \boldsymbol{\mu}\right) \tag{9}$$

which is of the same form as (7). Consequently, (4) is a hazardous basis for forecasting after location shifts, even when it is a 'good' in-sample model, as the DGP is bound to be (compare Allen & Fildes, 2005).

2.1. Changes in dynamics

There are other possible shifts than changes in $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$. If instead, $\boldsymbol{\alpha}$ in $\boldsymbol{\Gamma} = (\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}')$ changes at time T to $\boldsymbol{\alpha}^*$, with $\boldsymbol{\gamma}$, $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ constant, the new DGP in levels is:

$$\mathbf{x}_{T+h} = \boldsymbol{\gamma} - \boldsymbol{\alpha}^*\boldsymbol{\mu} + \boldsymbol{\Gamma}^*\mathbf{x}_{T+h-1} + \boldsymbol{\epsilon}_{T+h} \tag{10}$$

which seems to involve a shift. However, (10) can be written as:

$$\Delta\mathbf{x}_{T+h} = \boldsymbol{\gamma} + \boldsymbol{\alpha}^*\left(\boldsymbol{\beta}'\mathbf{x}_{T+h-1} - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_{T+h}.$$

This formulation entails no permanent changes in either $\mathsf{E}\left[\Delta\mathbf{x}_{T+h}\right]$ or $\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x}_{T+h}\right]$ so there is no impact on the long-run of a change in dynamics when no location shift occurs. Consequently, the change in $\boldsymbol{\alpha}$ will be nearly undetectable. Similar considerations apply to changes in $\boldsymbol{\beta}$ when the equilibrium mean shifts accordingly to maintain a zero expected equilibrium deviation. Consequently, we take $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as constant with known population values in the remainder of the paper.

3. A class of robust forecasting devices

There are a number of possible solutions to the failure of VEqCMs to converge to a new equilibrium after a location shift, including rapid estimation updating (e.g., recursive or rolling windows) and intercept corrections. In this section, we develop then analyze a new class of robust forecasting devices that can avoid some of the systematic forecast failure described in the previous section. We investigate two members that differ in their smoothness, and compare how they react (i) under constant parameters in Section 3.1, then (ii) to measurement errors (Section 3.2), (iii) unknown impulses (Section 3.3), (iv) unanticipated location shifts (Section 3.4) at the forecast origin and (v) unknowingly omitted variables (Section 3.5). Then Section 3.6 considers the consequences of shifting the forecast origin through time, and Section 3.7 discusses the corresponding longer horizon forecasts of levels and growth rates. Section 3.8 summarizes the results on robustifying VEqCMs.

The basis of the approach can be illustrated by using the first difference of (2) for 1-step forecasts:

$$\widetilde{\Delta\mathbf{x}}_{T+h|T+h-1} = \Delta\mathbf{x}_{T+h-1} + \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'\Delta\mathbf{x}_{T+h-1} \tag{11}$$

as differencing eliminates the 'deterministic shifts' for the included variables, and converts location shifts to impulses. However, there is a deeper reason why such a differencing strategy might work. Eq. (11) is also:

$$\widetilde{\Delta\mathbf{x}}_{T+h|T+h-1} = \left(\mathbf{I}_n + \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'\right)\Delta\mathbf{x}_{T+h-1} \tag{12}$$

where from (1), $h > 2$ periods after the break, $\Delta\mathbf{x}_{T+h-1}$ is in fact generated by:

$$\Delta\mathbf{x}_{T+h-1} = \left[\boldsymbol{\gamma}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+h-2} - \boldsymbol{\mu}^*\right) + \boldsymbol{\phi}^{*\prime}(\mathbf{z}_{T+h-2} - \boldsymbol{\kappa}^*)\right] + \boldsymbol{\epsilon}_{T+h-1} \tag{13}$$

so that:

$$\widetilde{\Delta\mathbf{x}}_{T+h|T+h-1} = \left(\mathbf{I}_n + \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'\right)\left(\left[\boldsymbol{\gamma}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+h-2} - \boldsymbol{\mu}^*\right) + (\boldsymbol{\phi}^*)'(\mathbf{z}_{T+h-2} - \boldsymbol{\kappa}^*)\right] + \boldsymbol{\epsilon}_{T+h-1}\right). \tag{14}$$

Then the term in brackets $[\cdots]$ in (14) contains everything you ever wanted to know when forecasting—and you don't even need to ask. Specifically, without knowing anything about the location shifts or the omitted variables, the forecast given by (14):

(Ia) has the correct $\boldsymbol{\gamma}^*$, $\boldsymbol{\mu}^*$ and $\boldsymbol{\kappa}^*$, so adjusts to the new equilibrium;
(Ib) had $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ also shifted, (14) would still have the right adjustment speeds and cointegration;
(IIa,b) the correct population parameter values in $[\cdots]$ are incorporated without any estimation biases;
(III) but forecast origin uncertainty due to $\widehat{\mathbf{x}}_{T+h-2} \neq \mathbf{x}_{T+h-2}$ could be a problem (discussed below);
(IV) there is no estimation uncertainty in $[\cdots]$;
(V) $\mathbf{z}_{T+h-2}$ is included despite the omitted variables in the VEqCM;
(VI) innovation errors, $\boldsymbol{\epsilon}_{T+h-1}$ remain.

Robust devices are a form of automatic 'learning device', in that later forecasts reflect location shifts when forecasting after they have occurred, even when the modeler is unaware of their occurrence. Although (14) uses $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ need to be estimated and could shift, they are well-determined full-sample estimates. This analysis helps explain the past 'success' of random-walk type forecasts, $\widetilde{\Delta\mathbf{x}}_{T+h|T+h-1} = \Delta\mathbf{x}_{T+h-1}$ where $\widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}' = \mathbf{0}$.

An alternative expression for (11) complements this explanation, writing it as:

$$\begin{aligned}\widetilde{\Delta\mathbf{x}}_{T+h|T+h-1} &= \Delta\mathbf{x}_{T+h-1} + \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-1} - \widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-2}\right)\\ &= \widetilde{\boldsymbol{\gamma}} + \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-1} - \widetilde{\boldsymbol{\mu}}\right).\end{aligned} \tag{15}$$

In (15), $\Delta\mathbf{x}_{T+h-1}$ can be interpreted as the highly adaptive estimator $\widetilde{\boldsymbol{\gamma}}$ of the growth rate $\boldsymbol{\gamma}$ in (1), and the previous value of the cointegrating combination, $\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-2} = \widetilde{\boldsymbol{\mu}}$, estimates $\boldsymbol{\mu}$. We use $\widetilde{\boldsymbol{\gamma}}$ and $\widetilde{\boldsymbol{\mu}}$ as shorthand for these estimates, although they both depend on T and h. In this interpretation, both $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$ are replaced by instantaneous estimators that are unbiased both before and some time after the population parameters have shifted, since for $h > 2$, $\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x}_{T+h-2}\right] \simeq \boldsymbol{\mu}^*$ and $\mathsf{E}\left[\Delta\mathbf{x}_{T+h-1}\right] \simeq \boldsymbol{\gamma}^*$.

This reinterpretation of the approach in Hendry (2006) to robustifying forecasting models after location shifts defines a class of forecasting devices given by:

$$\overline{\Delta\mathbf{x}}_{T+h|T+h-1} = \frac{1}{r}\sum_{i=1}^{r}\Delta\mathbf{x}_{T+h-i} + \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-1} - \frac{1}{m}\sum_{i=1}^{m}\widehat{\boldsymbol{\beta}}'\mathbf{x}_{T+h-1-i}\right) \quad (16)$$

where the instantaneous estimates $\widetilde{\boldsymbol{\gamma}}$ and $\widetilde{\boldsymbol{\mu}}$ are replaced by local averages when $r > 1$ and $m > 1$, and we do not require $r = m$. Then $r = m = 1$ gives (11). The adaptation to unanticipated location or growth shifts is slowed by higher values of r and m, but the forecasts will be smoother, less inclined to overshoot at the end of a break, and consequently behave better when there are no further location shifts.

Four well-known forecasting devices are special cases of (16). First, the recursively updated full-sample estimated model in (1) essentially uses (for $r = m = T+1, \ldots, T+h-1$, and $K = 2$):

$$\widetilde{\boldsymbol{\gamma}}_{(r)} = \frac{1}{(r-K+1)}\sum_{t=K}^{r}\Delta\mathbf{x}_t \quad \text{and}$$
$$\widetilde{\boldsymbol{\mu}}_{(m)} = \frac{1}{(m-K+1)}\sum_{t=K}^{m}\widehat{\boldsymbol{\beta}}'\mathbf{x}_{t-1}. \quad (17)$$

The full-sample VEqCM will deliver the 'smoothest' forecasts, and will be best when there are no location shifts, but is the least robust to location shifts as noted in the previous section. Moreover, updating the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ after a location shift can lead to the loss of cointegration (see Castle et al., 2010), as the estimate of $\boldsymbol{\alpha}$ is driven towards zero to eliminate the adverse impact of the shift in $\boldsymbol{\mu}$.

Second, rolling windows use (17) with $K = r - s$ for a fixed $s > 1$, and $m = r$, so are more or less smooth or robust after a location shift depending on the length of the window, again noting that estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ may also be updated. Third, and at the other extreme to the VEqCM, (11) with $r = m = 1$ is the most robust after location shifts, but also produces the noisiest forecasts, with estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ unchanged. Finally, with quarterly data, a natural intermediate choice in (16) is $r = m = 4$. When Δx_t is the quarterly growth rate, then $\sum_{i=1}^{4}\Delta x_{T+h-i} = \Delta_4 x_{T+h-1}$ where $\Delta_4 = 1 - L^4$, which is the year-on-year growth rate. Although data are often seasonally adjusted, such a formulation should also be robust to seasonal variation.

To facilitate the choice between members of the class in (16), we analytically investigate the 1-step ahead forecasts of (16) for $r = m = 1$ and $r = m = 4$ when: (i) the data generation process is unchanged, so that robustification of the forecasts is unnecessary; (ii) there are measurement errors; (iii) there is an impulse unbeknown to the forecaster; (iv) there is an unanticipated location shift; and (v) there are unknown omitted variables as in (1). For (i)–(iv) the in-sample DGP is given by:

$$\Delta\mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_t \quad \text{where } \boldsymbol{\epsilon}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \boldsymbol{\Omega}_\epsilon\right]. \quad (18)$$

We then consider (vi) shifting forward the forecast origin after a location shift, and (vii) longer horizon forecasts after the impact of a location shift.

3.1. Unchanged DGP

Although one would not need to use a robust forecasting device when the DGP is constant, the following analysis will nevertheless prove useful below. First, forecast errors from (18) are distributed as $\mathsf{IN}_n\left[\mathbf{0}, \boldsymbol{\Omega}_\epsilon\right]$, although that is an infeasible baseline.

When $r = m = 1$, and $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ are taken as known at their population values for analytical simplicity, (11) becomes:

$$\widetilde{\Delta\mathbf{x}}_{T+1|T} = \left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\Delta\mathbf{x}_T \quad (19)$$

leading to a forecast error for (18) of $\widetilde{\boldsymbol{\epsilon}}_{T+1|T} = \Delta\mathbf{x}_{T+1} - \widetilde{\Delta\mathbf{x}}_{T+1|T}$:

$$\begin{aligned}\widetilde{\boldsymbol{\epsilon}}_{T+1|T} &= \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_{T+1} \\ &\quad - \left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\left(\boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T-1} - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_T\right) \\ &= \boldsymbol{\alpha}\boldsymbol{\beta}'\left[\Delta\mathbf{x}_T - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T-1} - \boldsymbol{\mu}\right) - \boldsymbol{\epsilon}_T\right] + \Delta\boldsymbol{\epsilon}_{T+1} \\ &= \Delta\boldsymbol{\epsilon}_{T+1} \end{aligned} \quad (20)$$

noting that $\boldsymbol{\beta}'\boldsymbol{\gamma} = \mathbf{0}$. Then $\mathsf{E}\left[\widetilde{\boldsymbol{\epsilon}}_{T+1|T}\right] = \mathbf{0}$, but $\mathsf{V}\left[\widetilde{\boldsymbol{\epsilon}}_{T+1|T}\right] = 2\boldsymbol{\Omega}_\epsilon$, so that robustification has doubled the innovation-error variance.

Next, for $r = m = 4$, (16) is:

$$\widetilde{\widetilde{\Delta\mathbf{x}}}_{T+1|T} = \frac{1}{4}\sum_{i=0}^{3}\Delta\mathbf{x}_{T-i} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \frac{1}{4}\sum_{i=0}^{3}\boldsymbol{\beta}'\mathbf{x}_{T-1-i}\right) \quad (21)$$

so when $\widetilde{\widetilde{\boldsymbol{\epsilon}}}_{T+1|T} = \Delta\mathbf{x}_{T+1} - \widetilde{\widetilde{\Delta\mathbf{x}}}_{T+1|T}$:

$$\widetilde{\widetilde{\boldsymbol{\epsilon}}}_{T+1|T} = \left(\boldsymbol{\gamma} - \frac{1}{4}\sum_{i=0}^{3}\Delta\mathbf{x}_{T-i}\right) - \boldsymbol{\alpha}\left(\boldsymbol{\mu} - \frac{1}{4}\sum_{i=0}^{3}\boldsymbol{\beta}'\mathbf{x}_{T-1-i}\right) + \boldsymbol{\epsilon}_{T+1}. \quad (22)$$

The first term can be evaluated from (18) as:

$$\begin{aligned}\frac{1}{4}\sum_{i=0}^{3}\Delta\mathbf{x}_{T-i} &= \frac{1}{4}\sum_{i=0}^{3}\boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\frac{1}{4}\sum_{i=0}^{3}\boldsymbol{\beta}'\mathbf{x}_{T-i-1} - \frac{1}{4}\sum_{i=0}^{3}\boldsymbol{\mu}\right) + \frac{1}{4}\sum_{i=0}^{3}\boldsymbol{\epsilon}_{T-i} \\ &= \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\frac{1}{4}\sum_{i=0}^{3}\boldsymbol{\beta}'\mathbf{x}_{T-i-1} - \boldsymbol{\mu}\right) + \overline{\boldsymbol{\epsilon}}_T \end{aligned} \quad (23)$$

where $\overline{\boldsymbol{\epsilon}}_T = \frac{1}{4}\sum_{i=0}^{3} \boldsymbol{\epsilon}_{T-i}$, and substituting from (23) in (22) gives:

$$\widetilde{\widetilde{\boldsymbol{\epsilon}}}_{T+1|T} = \boldsymbol{\epsilon}_{T+1} - \overline{\boldsymbol{\epsilon}}_T. \tag{24}$$

Thus, $\mathsf{E}\left[\widetilde{\widetilde{\boldsymbol{\epsilon}}}_{T+1|T}\right] = \mathbf{0}$ with variance $\mathsf{V}\left[\widetilde{\widetilde{\boldsymbol{\epsilon}}}_{T+1|T}\right] = 1.25\boldsymbol{\Omega}_{\epsilon}$, so smoothing helps in this setting as expected. These outcomes compare to an error variance of $\boldsymbol{\Omega}_{\epsilon}$ when the in-sample DGP is used, so there is only a relatively small cost to this 'insurance policy' for r and m greater than unity.

3.2. Measurement errors at the forecast origin

National accounts data are subject to revision, so that the data on which forecasts are conditioned at the forecast origin will be measured with error, whereas assuming a relatively short revision process, the majority of the observations underlying the estimation of the model will constitute fully-revised or mature data.[4] A number of authors have tackled this problem, including Clements and Galvão (2013a,b), Garratt, Lee, Mise, and Shields (2008), Kishor and Koenig (2012) and Koenig, Dolmas, and Piger (2003), either by: estimating the revised values of recent-release data; using only lightly-revised data; or modeling the multiple-vintages available for any observation. Measurement error is most likely to be problematic in real-time forecasting exercises, when considering forecasts from a number of forecast origins, each time making use only of the vintage of data available at that forecast origin (see e.g., Patterson, 2003). Hence the data relating to the forecast origin observations will of necessity be first estimates. Our empirical forecasting exercise will be pseudo-real time, in that all observations are drawn from the same (recently-released) data vintage.[5] Nevertheless, we consider the performance of the robust class of model when there are measurement errors.

To begin with, suppose that only the latest observation is measured with error, so that $\overline{\mathbf{x}}_T = \mathbf{x}_T + \mathbf{v}_T$, but $\mathbf{v}_{T-1} = \mathbf{0}$. Then forecasting from (18) delivers:

$$\begin{aligned}\widehat{\Delta\mathbf{x}}_{T+1|T} &= \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\overline{\mathbf{x}}_T - \boldsymbol{\mu}\right) \\ &= \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right) + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{v}_T\end{aligned}$$

with a forecast error for (18) of $\widehat{\boldsymbol{\epsilon}}_{T+1|T} = \Delta\mathbf{x}_{T+1} - \widehat{\Delta\mathbf{x}}_{T+1|T}$ so:

$$\widehat{\boldsymbol{\epsilon}}_{T+1|T} = \boldsymbol{\epsilon}_{T+1} - \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{v}_T$$

where any additional bias depends on $\mathsf{E}[\mathbf{v}_T]$, and the variance is increased by $\mathsf{V}\left[\boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{v}_T\right] = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\Omega}_{v}\boldsymbol{\beta}\boldsymbol{\alpha}'$.

Next, consider forecasting with $r = m = 1$. Then the forecast is directly conditioned on the mis-measured observation, and (11) becomes:

$$\widetilde{\Delta\mathbf{x}}_{T+1|T} = \left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\Delta\overline{\mathbf{x}}_T = \boldsymbol{\Gamma}\Delta\mathbf{x}_T + \boldsymbol{\Gamma}\mathbf{v}_T$$

where $\boldsymbol{\Gamma} = \left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)$, again taking $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as known for simplicity, leading to a forecast error (18) of $\widetilde{\boldsymbol{\epsilon}}_{T+1|T} = \Delta\mathbf{x}_{T+1} - \widetilde{\Delta\mathbf{x}}_{T+1|T}$:

$$\begin{aligned}\widetilde{\boldsymbol{\epsilon}}_{T+1|T} &= \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_{T+1} \\ &\quad - \left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\left(\boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T-1} - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_T + \mathbf{v}_T\right) \\ &= \Delta\boldsymbol{\epsilon}_{T+1} - \boldsymbol{\Gamma}\mathbf{v}_T\end{aligned}$$

which augments (20) by $-\boldsymbol{\Gamma}\mathbf{v}_T$, so the bias depends on $\mathsf{E}[\mathbf{v}_T]$ and the variance is increased by $\mathsf{V}[\boldsymbol{\Gamma}\mathbf{v}_T] = \boldsymbol{\Gamma}\boldsymbol{\Omega}_{v}\boldsymbol{\Gamma}'$.

When the smoothed forecast device (21) is used with $r = m = 4$, the impact of the measurement error becomes $-\left(\frac{1}{4}\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\mathbf{v}_T$, so averaging attenuates both any bias due to the measurement error, and its variance component.

Suppose, however, that the previous observation was also mis-measured by an independent error with the same variance, so $\overline{\mathbf{x}}_{T-1} = \mathbf{x}_{T-1} + \mathbf{v}_{T-1}$ but the impact of both measurement errors on parameter estimates can be ignored. For the VEqCM here, there is no additional impact, whereas when $r = m = 1$:

$$\widetilde{\Delta\mathbf{x}}_{T+1|T} = \left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\Delta\overline{\mathbf{x}}_T = \boldsymbol{\Gamma}\Delta\mathbf{x}_T + \boldsymbol{\Gamma}\Delta\mathbf{v}_T$$

leading to:

$$\widetilde{\boldsymbol{\epsilon}}_{T+1|T} = \Delta\mathbf{x}_{T+1} - \widetilde{\Delta\mathbf{x}}_{T+1|T} = \Delta\boldsymbol{\epsilon}_{T+1} - \boldsymbol{\Gamma}\Delta\mathbf{v}_T \tag{25}$$

so the forecast-error variance is increased by $\mathsf{V}[\boldsymbol{\Gamma}\Delta\mathbf{v}_T] = 2\boldsymbol{\Gamma}\boldsymbol{\Omega}_{v}\boldsymbol{\Gamma}'$, the doubling being due essentially to mis-estimating $\boldsymbol{\mu}$ by using $\boldsymbol{\beta}'\overline{\mathbf{x}}_{T-1}$. Although we have assumed serially uncorrelated measurement error, it is clear from (25) that positive correlation in the measurement errors will reduce the overall impact on forecast error variance, whereas negative correlation will have the opposite effect.

The smoothed forecasting device (21) is affected by the combined measurement errors:

$$\frac{1}{4}\mathbf{v}_T + \boldsymbol{\alpha}\boldsymbol{\beta}'\left(\mathbf{v}_T - \frac{1}{4}\mathbf{v}_{T-1}\right) \tag{26}$$

which augments the innovation-error variance of $1.25\boldsymbol{\Omega}_{\epsilon}$ (in the absence of measurement error) by the variance of (26), which is:

$$\left(\frac{1}{4}\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\boldsymbol{\Omega}_{\mathbf{v}}\left(\frac{1}{4}\mathbf{I}_n + \boldsymbol{\beta}\boldsymbol{\alpha}'\right) + \frac{1}{16}\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\Omega}_{\mathbf{v}}\boldsymbol{\beta}\boldsymbol{\alpha}'. \tag{27}$$

Smoothing over $r = m = 4$, therefore, again leads to a very small increase in the total error variance from the impact of the previous measurement error, and even over the DGP forecast error variance of $\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\Omega}_{\mathbf{v}}\boldsymbol{\beta}\boldsymbol{\alpha}'$.

The above expressions assume that the aim of the exercise is to forecast the true value. When data are subject to multiple regular and benchmark revisions a case can be made for targetting an earlier-vintage 'actual' value, as discussed in some of the references at the beginning of this section.

[4] See e.g., Fixler and Grimm (2005, 2008) and Landefeld, Seskin, and Fraumeni (2008) on the nature of the US Bureau of Economic Analysis (BEA) data releases and revisions.

[5] This is true of virtually all studies of factor model forecasting, partly due to the costs of assembling real-time data sets on the large numbers of variables typically employed in constructing factors. An exception is Bernanke and Boivin (2003), who calculate inflation forecasts based on factors constructed from the restricted set of variables for which they have real-time data. They show that the properties of the forecasts depend on the number of variables underlying the factors (more variables lead to more accurate forecasts) rather than whether real or pseudo-real-time data are used.

3.3. An unknown impulse at the forecast origin

Unlike a measurement error which only contaminates a model and not the underlying mechanism, an impulse at $T+1$ impacts on the DGP, which becomes:

$$\Delta\mathbf{x}_{T+1} = \boldsymbol{\gamma} + \boldsymbol{\delta}1_{\{T+1\}} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_{T+1} \tag{28}$$

leading to a forecast error of $\widehat{\boldsymbol{\epsilon}}_{T+1|T} = \boldsymbol{\delta}1_{\{T+1\}} + \boldsymbol{\epsilon}_{T+1}$.

Next, still using (19) to forecast, with $\widetilde{\boldsymbol{\epsilon}}_{T+1|T} = \Delta\mathbf{x}_{T+1} - \widetilde{\Delta\mathbf{x}}_{T+1|T}$ delivers:

$$\widetilde{\boldsymbol{\epsilon}}_{T+1|T} = \boldsymbol{\delta}1_{\{T+1\}} + \Delta\boldsymbol{\epsilon}_{T+1} \tag{29}$$

adding a bias of $\boldsymbol{\delta}$, but no additional variance component. In fact, the smoothed forecast will also have the additional error of $\boldsymbol{\delta}1_{\{T+1\}}$ so all three forecasting devices suffer the same additional error. Thus, in the first period after the forecast origin, the impacts of impulses differ from those of measurement errors, as impulses affect the DGP but not the model, whereas measurement errors do the opposite.

However, outcomes will differ in the second period, when the DGP reverts to:

$$\Delta\mathbf{x}_{T+2} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+1} - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_{T+2}.$$

The in-sample VEqCM is now unaffected, whereas the robust forecast ($r = m = 1$) of:

$$\widetilde{\Delta\mathbf{x}}_{T+2|T+1} = \left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\Delta\mathbf{x}_{T+1}$$

leads to the forecast error:

$$\widetilde{\boldsymbol{\epsilon}}_{T+2|T+1} = \Delta\boldsymbol{\epsilon}_{T+2} - \boldsymbol{\delta}1_{\{T+1\}} \tag{30}$$

which is increased by the full amount of the impulse at $T+1$, but with the opposite sign, as against the second-period doubling of the measurement error impact seen in (25).

In contrast, $r = m = 4$ will only partly reflect the impulse, for reasons similar to why it was less distorted by the measurement error, since:

$$\widetilde{\widetilde{\Delta\mathbf{x}}}_{T+2|T+1} = \frac{1}{4}\sum_{i=0}^{3}\Delta\mathbf{x}_{T+1-i} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+1} - \frac{1}{4}\sum_{i=0}^{3}\boldsymbol{\beta}'\mathbf{x}_{T-i}\right)$$

so using:

$$\frac{1}{4}\sum_{i=0}^{3}\Delta\mathbf{x}_{T+1-i} = \boldsymbol{\gamma} + \frac{1}{4}\boldsymbol{\delta}1_{\{T+1\}} + \boldsymbol{\alpha}\left(\frac{1}{4}\sum_{i=0}^{3}\boldsymbol{\beta}'\mathbf{x}_{T-i} - \boldsymbol{\mu}\right) + \overline{\boldsymbol{\epsilon}}_{T+1} \tag{31}$$

then $\widetilde{\widetilde{\boldsymbol{\epsilon}}}_{T+2|T+1} = \Delta\mathbf{x}_{T+2} - \widetilde{\widetilde{\Delta\mathbf{x}}}_{T+2|T+1}$ is:

$$\widetilde{\widetilde{\boldsymbol{\epsilon}}}_{T+2|T+1} = \boldsymbol{\epsilon}_{T+2} - \overline{\boldsymbol{\epsilon}}_{T+1} - \frac{1}{4}\boldsymbol{\delta}1_{\{T+1\}}.$$

Although the smoothing entailed by larger values of r and m is beneficial in all these cases, that advantage is not maintained for an unanticipated location shift at the forecast origin, as we now show.

3.4. Location shift at the forecast origin

In all three cases above, the smoothed robust predictor dominates that using $r = m = 1$, and is not much worse than the in-sample DGP. The outcomes differ considerably when an unanticipated location shift persists over the forecast horizon. The case of $r = m = 1$ above prompted the creation of the class, but we did not establish its benefits explicitly compared to using the in-sample DGP when $h > 2$ and $\Delta\mathbf{x}_{T+h}$ is generated by:

$$\Delta\mathbf{x}_{T+h} = \boldsymbol{\gamma}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+h-1} - \boldsymbol{\mu}^*\right) + \boldsymbol{\epsilon}_{T+h}. \tag{32}$$

The robust forecasts h-steps ahead are given by:

$$\begin{aligned}\widetilde{\Delta\mathbf{x}}_{T+h|T+h-1} &= \left(\mathbf{I}_n + \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'\right)\Delta\mathbf{x}_{T+h-1} \\ &= \widehat{\boldsymbol{\Gamma}}\left(\boldsymbol{\gamma}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+h-2} - \boldsymbol{\mu}^*\right) + \boldsymbol{\epsilon}_{T+h-1}\right).\end{aligned}$$

Simplifying with known $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, $\widetilde{\boldsymbol{\epsilon}}_{T+h|T+h-1} = \Delta\mathbf{x}_{T+h} - \widetilde{\Delta\mathbf{x}}_{T+h|T+h-1} = \Delta\boldsymbol{\epsilon}_{T+h}$, as in Section 3.1, precisely because the new DGP is unchanged from $T+h-1$ to $T+h$ for $h > 2$. Thus, the robust device dominates the in-sample DGP provided the squared combined shift is larger than the error variance.

The slower adjustment of $r = m = 4$ is not now beneficial, even taking the more favourable case of $h = 3$ to illustrate:

$$\begin{aligned}\widetilde{\widetilde{\Delta\mathbf{x}}}_{T+3|T+2} &= \frac{1}{4}\sum_{i=0}^{3}\Delta\mathbf{x}_{T+2-i} \\ &\quad + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+2} - \frac{1}{4}\sum_{i=0}^{3}\boldsymbol{\beta}'\mathbf{x}_{T+1-i}\right)\end{aligned}$$

so when $\widetilde{\widetilde{\boldsymbol{\epsilon}}}_{T+3|T+2} = \Delta\mathbf{x}_{T+3} - \widetilde{\widetilde{\Delta\mathbf{x}}}_{T+3|T+2}$:

$$\begin{aligned}\widetilde{\widetilde{\boldsymbol{\epsilon}}}_{T+3|T+2} &= \left(\boldsymbol{\gamma}^* - \frac{1}{4}\sum_{i=0}^{3}\Delta\mathbf{x}_{T+2-i}\right) \\ &\quad - \boldsymbol{\alpha}\left(\boldsymbol{\mu}^* - \frac{1}{4}\sum_{i=0}^{3}\boldsymbol{\beta}'\mathbf{x}_{T+1-i}\right) + \boldsymbol{\epsilon}_{T+3} \\ &= \frac{1}{2}\left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}\right) - \frac{3}{4}\boldsymbol{\alpha}\left(\boldsymbol{\mu}^* - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_{T+3} - \overline{\boldsymbol{\epsilon}}_{T+3}\end{aligned} \tag{33}$$

as:

$$\begin{aligned}\frac{1}{4}\sum_{i=0}^{3}\Delta\mathbf{x}_{T+2-i} &= \frac{1}{2}\left(\boldsymbol{\gamma}^* + \boldsymbol{\gamma}\right) + \boldsymbol{\alpha}\left(\frac{1}{4}\sum_{i=0}^{3}\boldsymbol{\beta}'\mathbf{x}_{T+1-i}\right. \\ &\quad \left. - \frac{1}{4}\left(\boldsymbol{\mu}^* + 3\boldsymbol{\mu}\right)\right) + \overline{\boldsymbol{\epsilon}}_{T+3}\end{aligned}$$

because $\Delta\mathbf{x}_{T+2-i}$ is generated by the post-shift parameters $\boldsymbol{\gamma}^*$ and $\boldsymbol{\mu}^*$ for $i = 0, 1$, whereas for $i = 2, 3$ the pre-shift values prevail. Although only a part of the shift is offset, that becomes increasingly so as h increases, and (33) will then dominate the in-sample DGP.

3.5. Unknown omitted variables

Reverting to a constant DGP, but where a set of k variables $\mathbf{z}_t$ is unknowingly omitted both in-sample and over the forecast horizon so:

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

the main difference between the two robust devices is that for $r = m = 1$, $\boldsymbol{\phi}'\Delta\mathbf{z}_{T+h-1}$ is omitted as against $\boldsymbol{\phi}'\Delta_4\mathbf{z}_{T+h-1}$ when $r = m = 4$, and $\boldsymbol{\phi}'\mathbf{z}_{T+h-1}$ for the VEqCM, so the impact will depend on the time-series properties of the $\{\mathbf{z}_t\}$. Within a general-to-specific automatic model selection strategy, searching over all variables and factors, problems related to omitted variables should be of secondary importance.

3.6. Shifting the forecast origin

Irrespective of the values of r and m in (16), all the forecasts made immediately after an unanticipated location shift will suffer from the same biases. However, when forecasting from origins sufficiently long after the shift has occurred, the results for a constant parameter DGP will again hold, assuming the earlier in-sample breaks have been modeled and no further shifts occur. For forecast origins shortly after the break, the outcomes for the different approaches are less clearcut, especially for the smoothed robust forecasting device. The relevance of these origins for our understanding of the outcomes of empirical forecast exercises, where forecasts are produced from a relatively large number of origins, and forecast accuracy is estimated by averaging errors across origins, will depend on the frequency of shifts over the forecast period. We next illustrate this for forecasts of levels and growth rates made at time $T+2$ and later, when the DGP is (18).

3.7. Longer horizon forecasts of levels and growth rates

First, VEqCMs like (6) continue to cumulate the error in (5) when forecasting levels. Let $\boldsymbol{\Gamma} = \left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)$ and $\boldsymbol{\psi} = (\boldsymbol{\gamma} - \boldsymbol{\alpha}\boldsymbol{\mu})$, then the forecast of $T+h$ from $T+2$ for $h \geq 3$ using the in-sample DGP, even for known in-sample parameter values and correct initial conditions, is:

$$\begin{aligned}\widehat{\mathbf{x}}_{T+h|T+2} &= \boldsymbol{\Gamma}\widehat{\mathbf{x}}_{T+h-1|T+2} + \boldsymbol{\psi} \\ &= \boldsymbol{\Gamma}^2\widehat{\mathbf{x}}_{T+h-2|T+2} + \boldsymbol{\psi} + \boldsymbol{\Gamma}\boldsymbol{\psi} = \cdots \\ &= \sum_{i=0}^{h-3}\boldsymbol{\Gamma}^i\boldsymbol{\psi} + \boldsymbol{\Gamma}^{h-2}\mathbf{x}_{T+2} \end{aligned} \quad (34)$$

leading to ever larger forecast errors since the post-shift DGP generates:

$$\mathbf{x}_{T+h} = \sum_{i=0}^{h-3}\boldsymbol{\Gamma}^i\boldsymbol{\psi}^* + \boldsymbol{\Gamma}^{h-2}\mathbf{x}_{T+2} + \sum_{i=0}^{h-3}\boldsymbol{\Gamma}^i\boldsymbol{\epsilon}_{T+h-i}. \quad (35)$$

Because $\boldsymbol{\Gamma}^h\boldsymbol{\alpha} = \boldsymbol{\alpha}\left(\mathbf{I}_r + \boldsymbol{\beta}'\boldsymbol{\alpha}\right)^h = \boldsymbol{\alpha}\boldsymbol{\Lambda}^h$ where $\boldsymbol{\Lambda}$ has all its eigenvalues inside the unit circle, and $\boldsymbol{\Gamma}^h\boldsymbol{\gamma} = \boldsymbol{\gamma}$, $\boldsymbol{\Gamma}^h\boldsymbol{\gamma}^* = \boldsymbol{\gamma}^*$ then:

$$\begin{aligned}\mathsf{E}\left[\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T+2}\right] &= (h-2)\left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}\right) \\ &\quad - \boldsymbol{\alpha}\sum_{i=0}^{h-3}\boldsymbol{\Lambda}^i\left(\boldsymbol{\mu}^* - \boldsymbol{\mu}\right). \end{aligned} \quad (36)$$

Although $\boldsymbol{\Lambda}^i \to 0$ as $i \to \infty$, so the second terms eventually converges, the sum continually increases.

However, for forecasting growth rates, since:

$$\widehat{\mathbf{x}}_{T+h-1|T+2} = \boldsymbol{\Gamma}^{h-3}\mathbf{x}_{T+2} + \sum_{i=0}^{h-4}\boldsymbol{\Gamma}^i\boldsymbol{\psi}$$

then for $h > 4$:

$$\begin{aligned}\Delta\widehat{\mathbf{x}}_{T+h|T+2} &= \left(\boldsymbol{\Gamma}^{h-2} - \boldsymbol{\Gamma}^{h-3}\right)\mathbf{x}_{T+2} + \left(\sum_{i=0}^{h-3}\boldsymbol{\Gamma}^i - \sum_{i=0}^{h-4}\boldsymbol{\Gamma}^i\right)\boldsymbol{\psi} \\ &= \boldsymbol{\Gamma}^{h-3}\left[\boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+2} - \boldsymbol{\mu}\right)\right]. \end{aligned}$$

The conditional expectation of the outcome from the DGP is:

$$\mathsf{E}\left[\Delta\mathbf{x}_{T+h|T+2}|\mathbf{x}_{T+2}\right] = \boldsymbol{\Gamma}^{h-3}\left[\boldsymbol{\gamma}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+2} - \boldsymbol{\mu}^*\right)\right]$$

so the forecast error of $\widehat{\boldsymbol{\epsilon}}_{T+h|T+2} = \Delta\mathbf{x}_{T+h|T+2} - \Delta\widehat{\mathbf{x}}_{T+h|T+2}$ in the growth rate from using the pre-shift DGP has an expected value of:

$$\begin{aligned}\mathsf{E}\left[\widehat{\boldsymbol{\epsilon}}_{T+h|T+2}\right] &= \boldsymbol{\Gamma}^{h-3}\left[\left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}\right) - \boldsymbol{\alpha}\left(\boldsymbol{\mu}^* - \boldsymbol{\mu}\right)\right] \\ &= \left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}\right) - \boldsymbol{\alpha}\boldsymbol{\Lambda}^{h-3}\left(\boldsymbol{\mu}^* - \boldsymbol{\mu}\right) \end{aligned} \quad (37)$$

indicating a persistent error from the shift in $\boldsymbol{\gamma}$, whereas the effect of the shift in $\boldsymbol{\mu}$ on the forecast growth rate dies out as h increases. Consequently, it matters considerably whether forecast errors are evaluated for levels or growth rates.

Consider now the robust forecast device with $r = m = 1$. The growth-rate forecast is given by:

$$\widetilde{\Delta\mathbf{x}}_{T+h|T+2} = \boldsymbol{\Gamma}\widetilde{\Delta\mathbf{x}}_{T+h-1|T+2} = \cdots = \boldsymbol{\Gamma}^{h-2}\Delta\mathbf{x}_{T+2} \quad (38)$$

so when $\widetilde{\boldsymbol{\epsilon}}_{T+h|T+2} = \Delta\mathbf{x}_{T+h|T+2} - \widetilde{\Delta\mathbf{x}}_{T+h|T+2}$:

$$\begin{aligned}&\mathsf{E}\left[\widetilde{\boldsymbol{\epsilon}}_{T+h|T+2}|\mathbf{x}_{T+2}\right] \\ &\quad = \boldsymbol{\Gamma}^{h-3}\left[\boldsymbol{\gamma}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+2} - \boldsymbol{\mu}^*\right) - \boldsymbol{\Gamma}\Delta\mathbf{x}_{T+2}\right] \\ &\quad = \boldsymbol{\Gamma}^{h-3}\left[\boldsymbol{\gamma}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+2} - \boldsymbol{\mu}^*\right)\right. \\ &\qquad \left. - \boldsymbol{\Gamma}\left(\boldsymbol{\gamma}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+1} - \boldsymbol{\mu}^*\right)\right)\right]. \end{aligned} \quad (39)$$

Consequently, assuming $\boldsymbol{\beta}'\boldsymbol{\gamma}^* = \mathbf{0}$ as above:

$$\begin{aligned}\mathsf{E}\left[\widetilde{\boldsymbol{\epsilon}}_{T+h|T+2}\right] &= \boldsymbol{\Gamma}^{h-3}\boldsymbol{\alpha}\boldsymbol{\beta}'\mathsf{E}\left[\Delta\mathbf{x}_{T+2} - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+1} - \boldsymbol{\mu}^*\right)\right] \\ &= \mathbf{0}. \end{aligned}$$

Hence forecasting later than two periods after the break, when there are no further location shifts, the robust device returns unbiased longer-run forecasts of the growth rate, compared to the recurring biases of the non-robust model's forecasts in (37). Moreover, this robust method will then deliver unbiased forecasts of the levels. For example:

$$\begin{aligned}\widetilde{\mathbf{x}}_{T+4|T+2} &= \widetilde{\mathbf{x}}_{T+3} + \boldsymbol{\Gamma}\Delta\widetilde{\mathbf{x}}_{T+3} = \left(\mathbf{I}_n + \boldsymbol{\Gamma} + \boldsymbol{\Gamma}^2\right)\mathbf{x}_{T+2} \\ &\quad - \boldsymbol{\Gamma}\left(\mathbf{I}_n + \boldsymbol{\Gamma}\right)\mathbf{x}_{T+1} \end{aligned} \quad (40)$$

whereas:

$$\begin{aligned}\mathbf{x}_{T+4} &= \boldsymbol{\psi}^* + \boldsymbol{\Gamma}\mathbf{x}_{T+3} + \boldsymbol{\epsilon}_{T+4} = \left(\mathbf{I}_n + \boldsymbol{\Gamma}\right)\boldsymbol{\psi}^* \\ &\quad + \boldsymbol{\Gamma}^2\mathbf{x}_{T+2} + \boldsymbol{\Gamma}\boldsymbol{\epsilon}_{T+3} + \boldsymbol{\epsilon}_{T+4} \end{aligned}$$

so using $\mathbf{x}_{T+2} = \boldsymbol{\psi}^* + \boldsymbol{\Gamma}\mathbf{x}_{T+1} + \boldsymbol{\epsilon}_{T+2}$:

$$\mathbf{x}_{T+4} - \widetilde{\mathbf{x}}_{T+4|T+2} = \boldsymbol{\epsilon}_{T+4} + \boldsymbol{\Gamma}\boldsymbol{\epsilon}_{T+3} - \left(\mathbf{I}_n + \boldsymbol{\Gamma}\right)\boldsymbol{\epsilon}_{T+2} \quad (41)$$

which has an unconditional expectation of zero, but a large variance.

The smoothed device with $r = m = 4$ will only deliver unbiased growth rate forecasts once $h > 5$, but will converge on that outcome as h increases from 2. The forecasts of the levels will be biased, but not increasingly so once the growth forecasts are unbiased, and with a smaller forecast-error variance than (41).

3.8. Further discussion on robustifying VEqCMs

All the robust devices keep the in-sample VEqCM estimates, and would not work well if directly estimated, as such devices are manifestly non-congruent. Consequently, *Autometrics* style selection for r and m is also unlikely to work well, but 'in-sample forecast' evaluation of the extent that smoothing helps is possible. Equally, the lack of congruence from deliberately 'over-differencing' makes it difficult to calculate forecast standard errors. When data are poorly measured or are volatile, larger r and m seem advisable (the simplest case in Section 3.2 shows this); similarly for longer horizon forecasts when forecast-origin estimates are unreliable. Thus, trade-offs remain between robustness facing breaks and increased variance when there are no breaks. Moreover, we have omitted any effects from parameter estimation biases and variances, and treated each potential problem in isolation, although nothing precludes combinations thereof occurring empirically. Importantly, there is no need to robustify till after a forecast failure, because robust devices suffer the same initial forecast failure as non-robust. Improving on that performance would require such devices to be augmented by some method for forecasting shifts, although Castle, Fawcett, and Hendry (2011) show the difficulties in doing so.

There are other strategies available for improving the robustness of forecasting devices when location shifts occur, including both rolling windows and recursive updating, noted above as other members of the class in (16), as well as intercept corrections and averaging over a set of reasonable models. Castle et al. (2010) show that recursive updating in a VEqCM after a location shift can lead to the elimination of the equilibrium-correction terms, which are the source of the forecast failure, so the outcome ends close to an estimated robust device. We expect a similar outcome from rolling windows, so both would then lose any policy implications from equilibrium corrections, which can be recovered from (16) knowing its origins. Conversely, Castle et al. (2010) also show that both rolling windows and recursive updating help when previously high collinearities between explanatory variables are reduced by an 'external' break that nevertheless leaves the VEqCM constant. Since the above analysis shows that the impacts of the various problems differ and also affect each of the three exemplars in different ways, averaging over a number of these devices merits consideration.

The next section considers the application of robust methods to forecasting using factors. Note that forecasting models with factors are equilibrium-correction models when, as in (1) with the $\mathbf{z}_t$ interpreted as factors, the factors enter as $\mathsf{I}(0)$ variables with well-defined means.

4. Robust factor-based models for forecasting

In this section, we adapt the approach to allow for dynamic scalar equations dependent on factors estimated by the principal components of large numbers of unmodeled variables. Let Δy_t be the relevant transform of the dependent variable when the postulated factor model is:

$$\Delta y_t = \theta + \rho\left(\Delta y_{t-1} - \theta\right) + \boldsymbol{\delta}'\left(\mathbf{f}_{t-1} - \boldsymbol{\mu}\right) + \epsilon_t \tag{42}$$

where θ is the in-sample equilibrium mean, $\mathsf{E}\left[\Delta y_t\right] = \theta$, and $\mathbf{f}_t$ denotes a vector of $\mathsf{I}(0)$ factors or principal components with $\mathsf{E}\left[\mathbf{f}_t\right] = \boldsymbol{\mu}$, usually based on first or second differences of the original variables. The in-sample parameter estimates of (42) are denoted $\widehat{\theta}$, $\widehat{\rho}$, $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\delta}}$.

From the analysis in Section 3, the robust device for forecasting $h = 1$ step ahead over a horizon $k = 1, \ldots, K$ from an origin at T is:

$$\begin{aligned}\Delta\widetilde{y}_{T+k+1|T+k} &= \Delta y_{T+k} + \widehat{\rho}\Delta^2 y_{T+k} + \widehat{\boldsymbol{\delta}}\Delta\mathbf{f}_{T+k} \\ &= \Delta y_{T+k} + \widehat{\rho}\left(\Delta y_{T+k} - \Delta y_{T+k-1}\right) \\ &\quad + \widehat{\boldsymbol{\delta}}\left(\mathbf{f}_{T+k} - \mathbf{f}_{T+k-1}\right)\end{aligned} \tag{43}$$

where Δy_{T+k} is the 'instantaneous estimator' of the mean of Δy_{T+k+1}, and Δy_{T+k-1} is regarded as the 'estimator' of the mean of Δy_{T+k}, both of which are denoted 'θ' in the assumed constant-parameter model (42), and $\mathbf{f}_{T+k-1}$ 'estimates' $\boldsymbol{\mu}$. Correspondingly, the 1-step ahead smoothed robust variant becomes:

$$\begin{aligned}\Delta\overline{y}_{T+k+1|T+k} = &\ \frac{1}{r}\sum_{i=0}^{r-1}\Delta y_{T+k-i} \\ &+ \widehat{\rho}\left(\Delta y_{T+k} - \frac{1}{s}\sum_{i=1}^{s}\Delta y_{T+k-i}\right) \\ &+ \widehat{\boldsymbol{\delta}}'\left(\mathbf{f}_{T+k} - \frac{1}{m}\sum_{i=1}^{m}\mathbf{f}_{T+k-i}\right)\end{aligned} \tag{44}$$

so short moving averages are used to estimate the potentially changing values of θ and $\boldsymbol{\mu}$. When $r = s = m = 4$ say, $\frac{1}{r}\sum_{i=0}^{r-1}\Delta y_{T+k-i} = \frac{1}{4}\Delta_4 y_{T+k}$ denoted $\Delta_{(4)}y_{T+k}$ below, so is in the same units as θ (i.e., quarterly growth below).

As future values of large numbers of factors may be too difficult to forecast, direct multi-step estimation (DMS) is needed for (42) when $h > 1$, leading to (see e.g., Chevillon, 2007):

$$\Delta\widetilde{y}_{T+h+k|T+k} = \widetilde{\theta}_h + \widetilde{\rho}_h\left(\Delta y_{T+k} - \widehat{\theta}\right) + \widetilde{\boldsymbol{\delta}}_h'\left(\mathbf{f}_{T+k} - \widehat{\boldsymbol{\mu}}\right) \tag{45}$$

where $\widetilde{\theta}_h$, $\widetilde{\rho}_h$ and $\widetilde{\boldsymbol{\delta}}_h'$, denote the in-sample estimates from an h-step projection. Eq. (45) is also:

$$\Delta\widetilde{y}_{T+h+k|T+k} = \widetilde{\pi}_h + \widetilde{\rho}_h\Delta y_{T+k} + \widetilde{\boldsymbol{\delta}}_h'\mathbf{f}_{T+k} \tag{46}$$

where $\widetilde{\pi}_h$ is the estimated composite intercept, which is the formulation used in Section 5 and denoted PC1-4 (being based on the first four principal components).

There are several possible ways to robustify (46), corresponding to a robust version of that multi-step form using

h-period differences, or developing a multi-step robust device after first differencing. We chose a variant of the former (denoted R1PC1-4 below), namely:

$$\Delta\widetilde{y}_{T+h+k|T+k} = \Delta y_{T+k} + \widetilde{\rho}_h \Delta\Delta_{(h)} y_{T+k} + \widetilde{\boldsymbol{\delta}}_h' \Delta_{(h)} \mathbf{f}_{T+k}. \quad (47)$$

The other robust formulation would use first differences for the regressors. In both cases, the same parameter estimates as (46) are used. The robust versions of a first-order autoregression (AR1) reported below correspond to (47) with $\widetilde{\boldsymbol{\delta}}_h = \mathbf{0}$.

The multi-step smoothed robust variant is formulated as:

$$\begin{aligned}\Delta\overline{y}_{T+h+k|T+k} &= \frac{1}{r}\sum_{i=0}^{r-1} \Delta y_{T+k-i} \\ &\quad + \widetilde{\rho}_h \left(\Delta y_{T+k} - \frac{1}{s}\sum_{i=1}^{s} \Delta y_{T+k-i} \right) \\ &\quad + \widetilde{\boldsymbol{\delta}}_h' \left(\mathbf{f}_{T+k} - \frac{1}{m}\sum_{i=1}^{m} \mathbf{f}_{T+k-i} \right) \qquad (48)\end{aligned}$$

denoted R4PC1-4 below when $r = s = m = 4$. As before, (48) can be formulated with other values for r, s and m. First differences instead of $\Delta_{(h)}$ in (47) corresponds to $r = s = m = 1$ in (48).

For models that have both variables $\mathbf{z}_{t-1}$ and factors $\mathbf{f}_{t-1}$ with in-sample means $\boldsymbol{\mu}_z$ and $\boldsymbol{\mu}_f$ respectively, then the initial general unrestricted model (GUM) would be the singular specification:

$$\begin{aligned}\Delta y_t &= \theta + \rho\left(\Delta y_{t-1} - \theta\right) + \boldsymbol{\lambda}'\left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_z\right) \\ &\quad + \boldsymbol{\delta}'\left(\mathbf{f}_{t-1} - \boldsymbol{\mu}_f\right) + \epsilon_t. \qquad (49)\end{aligned}$$

However, using combinations of contracting and expanding block searches, as in *Autometrics*, selection can be applied to deliver the 'non-robust' full-sample estimated model as shown in Castle, Clements, and Hendry (2013):

$$\begin{aligned}\Delta\widehat{y}_t &= \widehat{\theta} + \widehat{\rho}\left(\Delta y_{t-1} - \widehat{\theta}\right) + \widehat{\boldsymbol{\lambda}}_1'\left(\mathbf{z}_{1,t-1} - \widehat{\boldsymbol{\mu}}_{1,z}\right) \\ &\quad + \widehat{\boldsymbol{\delta}}_1'\left(\mathbf{f}_{1,t-1} - \widehat{\boldsymbol{\mu}}_{1,f}\right). \qquad (50)\end{aligned}$$

The corresponding smoothed robust approach for 1-step ahead forecasts becomes:

$$\begin{aligned}\Delta\overline{y}_{T+k+1|T+k} &= \frac{1}{r}\sum_{i=0}^{r-1} \Delta y_{T+k-i} \\ &\quad + \widehat{\rho}\left(\Delta y_{T+k} - \frac{1}{s}\sum_{i=1}^{s} \Delta y_{T+k-i} \right) \\ &\quad + \widehat{\boldsymbol{\lambda}}_1'\left(\mathbf{z}_{1,T+k} - \frac{1}{p}\sum_{i=1}^{p} \mathbf{z}_{1,T+k-i} \right) \\ &\quad + \widehat{\boldsymbol{\delta}}_1'\left(\mathbf{f}_{1,T+k} - \frac{1}{m}\sum_{i=1}^{m} \mathbf{f}_{1,T+k-i} \right). \qquad (51)\end{aligned}$$

Setting $r = s = m = p = 4$, for example, the smoothed robust 1-step forecasts are given by:

$$\begin{aligned}\Delta\overline{y}_{T+k+1|T+k} &= S_4 \Delta y_{T+k} + \widehat{\rho}\left(\Delta y_{T+k} - S_4 \Delta y_{T+k-1}\right) \\ &\quad + \widehat{\boldsymbol{\lambda}}_1'\left(\mathbf{z}_{1,T+k} - S_4 \mathbf{z}_{1,T+k-1}\right) \\ &\quad + \widehat{\boldsymbol{\delta}}_1'\left(\mathbf{f}_{1,T+k} - S_4 \mathbf{f}_{1,T+k-1}\right) \qquad (52)\end{aligned}$$

where $S_r = \frac{1}{r}\left(1 + L + L^2 + L^3 + \cdots L^{r-1}\right)$ and L is the lag operator.

To make matters concrete, when Δy_t denotes the quarterly GDP growth rate (defined as the first difference of the natural log of quarterly GDP), then the smoothed robust forecasting device has the interpretation that future quarterly GDP growth is predicted using lagged annual growth at quarterly rates, the previous quarterly growth rate Δy_{T+k-1} as a deviation from its past yearly average, and the relevant variables and factors entered as I(0) transformed deviations from their moving averages ($S_4\mathbf{z}_{1,T+k-1}$ and $S_4\mathbf{f}_{1,T+k-1}$). Thus, in this setting, we would use *Autometrics* model selection to get the best in-sample model, then generate robust forecasts therefrom as transformed in (52) as well as for $r = m = s = p = 1$ in (51).

However, the factors used below have already been differenced once or twice, so are already partly 'robustified'. Moreover, Chevillon (2007) shows that DMS also provides partial robustification against various forms of mis-specification, including location shifts. Thus, the extra differencing of the robust methods proposed for such factor formulations may be less helpful than for a VEqCM, and will be advantageous primarily when π changes, so is better approximated by recent lagged growth than a full-sample estimate. Beyond $h = 4$, the in-sample estimates $\widetilde{\rho}_h$ and $\widetilde{\boldsymbol{\delta}}_h$ below are tiny, so only the composite intercept mattered empirically. Consequently, we also consider a variant of (48) where the estimates of $\boldsymbol{\delta}$ are not constrained to those found from (46), but are re-estimated after imposing differencing: this is denoted R4PC1-4est below.

4.1. Change in the relevance of a variable

An alternative source of location shifts may be from changes over time in the relevance of 'explanatory variables' or factors. For example, Stock and Watson (2008) report a pseudo out-of-sample forecasting exercise for US inflation, and find that models with indicator variables only episodically enhance forecasts relative to a random walk model, at least over the period from 1985 onwards. Suppose the DGP is:

$$\begin{aligned} w_t &= \theta + \rho w_{t-1} + \varepsilon_t, \quad t < \tau \\ w_t &= \theta + \rho w_{t-1} + \gamma z_{t-1} + \varepsilon_t, \quad t \geq \tau \qquad (53)\end{aligned}$$

so that z_t helps to predict w_{t+1} only after time τ. Depending on the relative pre and post break sample sizes in (53), assuming that the coefficient of z_{t-1} was constant over the whole period might well result in z appearing to be insignificant when the full-sample model is estimated, and at the least its effect would be attenuated. Rolling windows of a length that did not span both regimes at the time of forecasting would perform well.

When the mean of z_{t-1} in the second period is μ, then (53) becomes:

$$\begin{aligned} w_t &= \left(\theta + \gamma\mu 1_{(t\geq\tau)}\right) \\ &\quad + \rho w_{t-1} + \gamma 1_{(t\geq\tau)}\left(z_{t-1} - \mu\right) + \varepsilon_t \qquad (54)\end{aligned}$$

where $1_{(t\geq\tau)} = 1$ for $t \geq \tau$. This extended model has constant parameters (as in Clements & Hendry, 1999, p. 260), but in practice, the dates of such breaks (as there may be several) and which variables (or factors) are affected are unknown. When there are a number of explanatory variables that might have changing effects, a strategy of interacting all possible such terms with step indicators for every date then selecting the optimal forecasting model could be infeasible. However, the most detrimental effect of omitting a variable such as z_{t-1} probably arises from the induced location shift, $\gamma\mu 1_{(t\geq\tau)}$. The strategy of step-indicator saturation (SIS) – saturating the model with step-shift indicators – then becomes a possible solution. Commencing at the first observation and continuing to every observation, one creates T indicator variables for T observations, of the form $\mathcal{S}_1 = \left\{1_{\{t\leq j\}}, j = 1, \ldots, T\right\}$, when $1_{\{t\leq j\}} = 1$ for observations up to j, and zero otherwise. Step indicators are the cumulations of impulse indicators up to each next observation. Such a general procedure is needed when the locations, durations, magnitudes and signs of location shifts are unknown. Castle, Doornik, Hendry, and Pretis (2013) investigate the theory of step-indicator saturation under the null of no location shifts, and for a variety of single and multiple shifts. SIS is included in *Autometrics* and is feasible because software like *Autometrics* can handle more candidate variables N than observations T, using a combination of expanding and contracting multiple block searches. SIS even provides a potential solution when the need to include z_{t-1} is not known (see Castle & Hendry, 2014). Location shifts when τ is near the forecast origin were considered above, and if found, SIS then provides a natural intercept correction.

Alternatively, suppose the DGP is:

$$w_t = \theta + \rho w_{t-1} + \gamma z_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T \tag{55}$$

but shifts to:

$$w_{T+k} = \theta + \rho w_{T+k-1} + \varepsilon_{T+k}, \quad k \geq 1 \tag{56}$$

so that z_{T+k-1} no longer predicts w_{T+k} after T. When the mean of z_{t-1} in-sample is μ_1 but changes magnitude to μ_2 in the forecast period, then the forecasts from the estimates of (55) are:

$$\widehat{w}_{T+k} = \widehat{\theta} + \widehat{\rho} w_{T+k-1} + \widehat{\gamma} z_{T+k-1} \tag{57}$$

so with $\widehat{\varepsilon}_{T+k} = w_{T+k} - \widehat{w}_{T+k}$:

$$\widehat{\varepsilon}_{T+k} = \left(\theta - \widehat{\theta}\right) + \left(\rho - \widehat{\rho}\right) w_{T+k-1} - \widehat{\gamma} z_{T+k-1} + \varepsilon_{T+k}$$

then (ignoring finite-sample biases of $\mathsf{O}(T^{-1})$ in (57)):

$$\mathsf{E}\left[\widehat{\varepsilon}_{T+k}\right] \simeq -\gamma\mu_2.$$

The most detrimental effect of including $\widehat{\gamma} z_{T+k-1}$ in (57) therefore arises from the induced location shift, $\gamma\mu_2$. As an empirical example (see Castle, Clements, & Hendry, 2013), consider the sudden and dramatic shift that occured in the monetary base following the financial crisis, which jumped from \$863bn in 2008(3) to \$1724bn in 2009(3). Models which retained a term in the monetary base would have experienced a forecast error shift of $\widehat{\gamma} z_{T+k-1}$ in the forecast period, resulting in forecast failure. Rolling windows of a length that eventually did not span both regimes would end performing well. In practice, intervention by the forecaster could also attenuate such effects, and we provide an *ex post* mimic of that below.

The simplest robust approach (e.g., the equivalent of (51) with $r = s = m = 1$) is:

$$\widetilde{w}_{T+k|T+k-1} = w_{T+k-1} + \widehat{\rho}\Delta w_{T+k-1} + \widehat{\gamma}\Delta z_{T+k-1}.$$

When $k > 1$, letting $\widetilde{\varepsilon}_{T+k|T+k-1} = w_{T+k} - \widetilde{w}_{T+k|T+k-1}$, and ignoring parameter estimation, we obtain:

$$\begin{aligned}\widetilde{\varepsilon}_{T+k|T+k-1} &= \Delta w_{T+k} - \rho\Delta w_{T+k-1} - \gamma\Delta z_{T+k-1} \\ &= -\gamma\Delta z_{T+k-1} + \Delta\varepsilon_{T+k}.\end{aligned}$$

Providing there are no further location shifts in z:

$$\mathsf{E}\left[\widetilde{\varepsilon}_{T+k|T+k-1}\right] \simeq -\gamma\mathsf{E}\left[\Delta z_{T+k-1}\right] = 0$$

and hence should outperform forecasts based on the in-sample DGP. In summary: our robust forecasting devices should work in a world in which the predictive ability of explanatory variables changes over time, and when these variables are themselves subject to large shifts, as occurred with monetary variables during the Great Recession.

5. Robust forecasts of US GDP and GDP growth

Our empirical forecasting exercise compares the forecast performance of regression models based on principal components with robust devices. We forecast quarterly GDP growth and the corresponding level over the period 2000–2011. The robust devices we consider are based on the class described in Section 4. We do not consider other possibilities, such as SIS as described in Section 4.1. Nor do we consider updating the model estimates over the forecast period as a way of mitigating the effects of breaks in order to focus on the new class of robust devices we have introduced, although in practice one might implement the robust devices with recursive or rolling forecasting schemes.

A number of authors have assessed the forecast performance of factor models over this period, and Stock and Watson (2011) review studies which explicitly consider the impact of breaks on factor-model forecasts, including Stock and Watson (2009) who suggest estimating the factors on the full historical period across the break (then, the Great Moderation around 1984, see, e.g., McConnell & Perez-Quiros, 2000), but only estimating forecasting models using factors as explanatory variables on the post-break period. Here we use the full estimation sample for factors, and the new robust devices to counter location shifts. AR benchmarks have typically been difficult to beat systematically, and Stock and Watson (2010) argue that simple univariate models, such as a random walk which is the simplest robust device, are competitive with models using explanatory variables, consistent with location shifts being important.

The data are discussed in detail in Castle, Clements, and Hendry (2013), so only briefly noted here in Section 5.1, and the forecasting models are described in Section 5.2. Section 5.3 presents the results.

5.1. Data

The data set, based on Stock and Watson (2009), consists of 144 quarterly time series for the United States over 1959(1)–2006(4), updated here to 2011(2). There are

$n = 109$ disaggregate variables, used both as the candidate set of regressors and the set for the principal components. All data are transformed to remove unit roots by taking first or second differences (usually in logs) as described in Stock and Watson (2009) Appendix Table A1. The data available for estimation span $T = 1962(3)$–$2011(2)$, so there are 150 in-sample observations after transformations and lags, with the forecast horizon of 2000(1)–2011(2), which is separated into the two periods 2000(1)–2006(4), and 2007(1)–2011(2), to assess the performance of the forecasting models over both a quiescent period and the financial crisis.

Three forecast horizons are recorded, for $h = 1, 4, 8$ step-ahead direct forecasts. Forecasts are evaluated over the full forecast sample of 2000(1)–2011(2) (46 observations), with the first forecast subsample of 2000(1)–2006(4) (28 observations), and the second of 2007(1)–2011(2) (18 observations).

5.2. Forecasting models

Our primary interest is in forecasting models which consist of the first four principal components. Castle, Clements, and Hendry (2013) found that such models performed well compared to models selected from factors and variables. This model, denoted PC1-4, is given by:

$$\Delta y_t = \gamma_0 + \rho_h \Delta y_{t-h} + \sum_{k=1}^{4} \gamma_{k,h} f_{k,t-h} + \epsilon_t \quad (58)$$

where Δy_t is the first difference of log real gross domestic product and $f_{k,t}$ is the kth PC. The PCs are extracted from the whole dataset. Although this may seem to favour such models, the effect could go either way. Information that is not available at the time of the forecast is used, but against that there were major changes in some of the disaggregated variables over the forecast period, consistent with the success of the robust device R4PC1-4est where coefficients of factors were re-estimated.

The usual benchmark forecasts of a random walk (RW):

$$\Delta \widehat{y}^{\mathrm{RW}}_{T+h+k|T+k} = \Delta y_{T+k} \quad (59)$$

where $T + k$ is the forecast origin, and a first-order autoregression are included for comparison, as is a robust version of the AR1, essentially (47) without factors. We evaluate the forecasts on root mean-square errors (RMSFEs) for both levels and growth rates. The level forecasts for GDP (in logs) are computed from the forecasts of the growth rates as:

$$\widehat{y}_{T+h+k|T+k} = \sum_{i=1}^{h} \Delta \widehat{y}_{T+i+k|T+k} + y_{T+k}$$

for $h = 4, 8$. Although 1-step ahead forecast errors are identical for levels and differences, results of both are reported for ease of comparison. Note that 4-step forecasts are evaluated from 2000:4 onwards and 8-step forecasts from 2001:4 (and similarly for the second subsample).

5.3. Results

Table 1 records the RMSFEs for log GDP and GDP growth for each of the forecasting models in the whole sample and two subperiods. For GDP growth, the four-period smoothed robust device that does not impose the original parameter estimates for the factors, but instead estimates them after differencing (R4PC1-4est) performs well, ranking second on all horizons for the full sample due to its favourable performance in the second, more volatile, sub-period. The unadjusted model that uses the first 4 PCs (PC1-4) tends to dominate at 1-step ahead and does well in the more quiescent first sub-period, as would be expected when there are no breaks. A simple AR(1) is difficult to beat at longer horizons, but there is no benefit to robustifying the AR(1) device, which in fact does worst overall. Both PC1-4 and R4PC1-4 outperform the non-smooth robust device and the random walk. The RMSFEs for 2007(1)–2011(2) are often twice as large as those for the first sub-sample. Robust devices are 'designed' for short-term rather than long-horizon forecasts, and the smoothing leads to an improvement over R1PC1-4 at almost all horizons, but R4PC1-4 also does surprisingly well at 8-steps ahead, further improved by R4PC1-4est.

The rankings are less clear for the levels forecasts, confirming that evaluation in levels and growth rates need not result in the same ranking. The unadjusted factor model obviously still performs best at the 1-step horizon, but at longer horizons the robust versions still do well. The non-smooth robust device (R1PC1-4), which did not rank in differences, is ranked second at the 8-step horizon across all sub-periods, with the smoothed variants often preferred at other horizons. The AR(1) device does well in the quiescent period in levels, but not in the more volatile later sub-period. There are very large increases at longer horizons in the RMSFEs for levels relative to growth, and between the two subsamples, which emphasizes the difficulty of forecasting the level of GDP especially when the impact of the crisis-period location shift was unanticipated: all the methods fail to 'foresee' the shift, even using full-sample PCs.

Fig. 1 illustrates this last feature for the 4-period ahead forecasts of changes and levels of log real GDP for PC1-4 and R4PC1-4est over the quiescent and turbulent periods. The systematic over-prediction of growth from mid 2007 to mid 2009 translates into increasing mis-forecasting of the levels. In many cases, averaging across the 'least poisonous' devices would have reduced RMSFEs, as panel d suggests visually here, confirmed by the average of those two levels forecasts having a RMSFE of 3.57.

6. Conclusion

The theory and evidence in Castle, Clements, and Hendry (2013) demonstrated the importance of robustifying forecasts to location shifts, a key source of forecast failure. Regression models, whether based on variables or factors, are equilibrium-correction formulations, so like all EqCMs, are not robust after location shifts, potentially facing systematic forecast failure. We presented a new class of forecasting devices that are robust after location shifts, and

Table 1
Forecast-error outcomes.

	PC1-4	R1PC1-4	R4PC1-4	R4PC1-4est	RW	AR1	R1AR1
$\Delta\widehat{y}_{T+h}$							
	0.56	0.83	0.63	*0.59*	0.75	0.73	0.86
Full sample	0.97	1.00	0.93	*0.85*	1.05	**0.84**	1.05
	1.01	1.08	0.92	*0.91*	1.10	**0.88**	1.10
	0.47	0.82	*0.52*	0.53	0.77	0.56	0.93
2000(1)–2006(4)	*0.63*	0.69	0.65	0.66	0.68	**0.54**	0.68
	0.64	0.85	0.74	0.74	0.86	**0.55**	0.85
	0.73	0.90	0.83	0.73	*0.73*	0.93	0.74
2007(1)–2011(2)	1.19	1.36	1.50	**1.08**	1.44	*1.14*	1.45
	1.30	1.37	*1.14*	**1.13**	1.40	1.20	1.40
$\widehat{y}_{T+h}$							
	0.56	0.83	0.63	*0.59*	0.75	0.73	0.86
Full sample	2.79	2.89	3.23	**2.61**	3.02	*2.75*	3.08
	5.96	*4.60*	**4.55**	4.78	4.70	5.28	4.69
	0.47	0.82	*0.52*	0.53	0.77	0.56	0.93
2000(1)–2006(4)	1.71	1.64	1.83	1.75	*1.62*	**1.34**	1.63
	3.62	*3.61*	3.69	3.92	3.68	**2.55**	3.66
	0.73	0.90	0.83	0.73	*0.73*	0.93	0.74
2007(1)–2011(2)	*3.89*	4.43	4.70	**3.77**	4.64	4.15	4.66
	8.92	*6.79*	**6.72**	6.98	6.93	8.54	6.92

RMSFEs × 100 for log GDP and quarterly GDP growth. Forecast devices are: PC1-4 (46); R1PC1-4 (47); R4PC1-4 (48); R4PC1-4est (48) with $\boldsymbol{\delta}$ re-estimated rather than imposed from (46); RW (59); AR1 (46), and R1AR1 (47) both with $\boldsymbol{\delta} = \mathbf{0}$. The three rows in each block correspond to 1, 4, and 8 step ahead forecast outcomes. The smallest RMSFE in each row is shown in bold, with the second smallest RMSFE in italics.

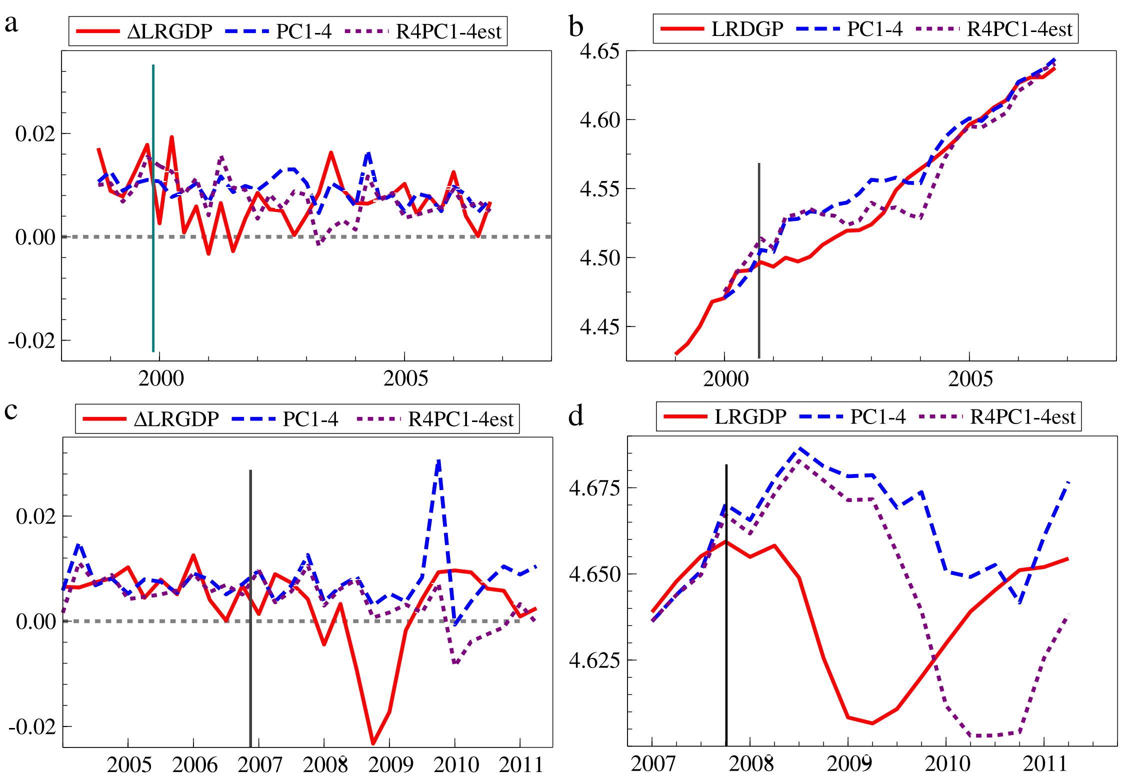

Fig. 1. 4-period ahead forecasts of changes (column 1) and levels (column 2) of log real GDP by PC1-4 and R4PC1-4est over both quiescent (row 1) and turbulent periods (row 2).

analyzed their properties in a variety of settings. For large location shifts, the most adaptable should prove advantageous, but if other problems are present, such as measurement errors at the forecast origin, a smoothed variant may perform better.

The empirical application considered log GDP and GDP growth, computing forecasts using robust devices based around the first four principal components, which had performed as well as selection over factors and variables in Castle, Clements, and Hendry (2013). Only counting 1-step outcomes once, the four-period smoothed robust device re-estimating the factor coefficients after differencing, produced either the smallest or second smallest RMSFEs on 7 occasions, as against the AR1 8 times and the non-robust device based on the levels of the PCs 6 times. The success of the last may be because the principal components are themselves averages over many variables, or may be due to using full-sample estimates of PCs. Alternatively its relative success may be because the PCs are differenced (sometimes twice), so are already partly robustified. These three approaches dominated a random walk, the more highly adaptive robust device (although that did relatively well at longer horizons for levels), and a robust AR(1) model, providing some support for smoothing robust methods.

The forecast performances of all these devices would probably be improved by recursive updating, combining information between devices, both by averaging and only switching to a robust method after forecast failure then switching back once 'normality' returned, but worsened by having only preliminary data available at successive forecast horizons.

Acknowledgments

This research was supported in part by grants from the Open Society Foundations and the Oxford Martin School. We are grateful to Fred Joutz, Tara Sinclair and two anonymous referees for helpful comments on an earlier draft.

References

Allen, P. G., & Fildes, R. A. (2005). Levels, differences and ECMs—principles for improved econometric forecasting. *Oxford Bulletin of Economics and Statistics*, *67*, 881–904.
Bernanke, B. S., & Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, *50*, 525–546.
Castle, J. L., Clements, M. P., & Hendry, D. F. (2013). Forecasting by factors, by variables, by both or neither? *Journal of Econometrics*, *177*, 305–319.
Castle, J. L., Doornik, J. A., & Hendry, D. F. (2012). Model selection when there are multiple breaks. *Journal of Econometrics*, *169*(2), 239–246.
Castle, J.L., Doornik, J.A., Hendry, D.F., & Pretis, F. (2013). *Detecting location shifts by step-indicator saturation*. Working paper, Economics Department, Oxford University.
Castle, J. L., Fawcett, N. W. P., & Hendry, D. F. (2010). Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics*, *158*, 25–36.
Castle, J.L., Fawcett, N.W.P., & Hendry, D.F. (2011). Forecasting breaks and during breaks. See Clements and Hendry (2011) (pp. 315–353).
Castle, J. L., & Hendry, D. F. (2014). Model selection in under-specified equations with breaks. *Journal of Econometrics*, *178*, 286–293.
Castle, J. L., & Shephard, N. (Eds.) (2009). *The methodology and practice of econometrics*. Oxford: Oxford University Press.
Chevillon, G. (2007). Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, *21*, 746–785.
Clements, M. P., & Galvão, A. B. (2013a). Forecasting with vector autoregressive models of data vintages: US output growth and inflation. *International Journal of Forecasting*, *29*(4), 698–714.
Clements, M. P., & Galvão, A. B. (2013b). Real-time forecasting of inflation and output growth with autoregressive models in the presence of data revisions. *Journal of Applied Econometrics*, *28*(3), 458–477.
Clements, M. P., & Hendry, D. F. (1999). *Forecasting non-stationary economic time series*. Cambridge, Mass.: MIT Press.
Clements, M. P., & Hendry, D. F. (2006). Forecasting with breaks. In G. Elliott, C. Granger, & A. Timmermann (Eds.), *Handbook of economics: Vol. 24. Handbook of economic forecasting, Vol. 1* (pp. 605–657). Elsevier, North-Holland.
Clements, M. P., & Hendry, D. F. (Eds.) (2011). *Oxford handbook of economic forecasting*. Oxford: Oxford University Press.
Doornik, J.A. (2009). Autometrics. See Castle and Shephard (2009) (pp. 88–121).
Doornik, J. A., & Hendry, D. F. (2013). *Empirical econometric modelling using PcGive: Volume I* (7th ed.). London: Timberlake Consultants Press.
Fildes, R. A., & Stekler, H. O. (2002). The state of macroeconomic forecasting. *Journal of Macroeconomics*, *24*, 435–468.
Fixler, D. J., & Grimm, B. T. (2005). Reliability of the NIPA estimates of US economic activity. *Survey of Current Business*, *85*, 9–19.
Fixler, D. J., & Grimm, B. T. (2008). The reliability of the GDP and GDI estimates. *Survey of Current Business*, *88*, 16–32.
Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized factor model: identification and estimation. *Review of Economics and Statistics*, *82*, 540–554.
Garratt, A., Lee, K., Mise, E., & Shields, K. (2008). Real time representations of the output gap. *Review of Economics and Statistics*, *90*, 792–804.
Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, *135*, 399–426.
Johansen, S., & Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. See Castle and Shephard (2009) (pp. 1–36).
Kishor, N. K., & Koenig, E. F. (2012). VAR estimation and forecasting when data are subject to revision. *Journal of Business and Economic Statistics*, *30*(2), 181–190.
Koenig, E. F., Dolmas, S., & Piger, J. (2003). The use and abuse of real-time data in economic forecasting. *The Review of Economics and Statistics*, *85*(3), 618–628.
Landefeld, J. S., Seskin, E. P., & Fraumeni, B. M. (2008). Taking the pulse of the economy. *The Journal of Economic Perspectives*, *22*, 193–216.
McConnell, M. M., & Perez-Quiros, G. (2000). Output fluctuations in the United States: what has changed since the early 1980s? *American Economic Review*, *90*, 1464–1476.
Patterson, K. D. (2003). Exploiting information in vintages of time-series data. *International Journal of Forecasting*, *19*, 177–197.
Peña, D., & Poncela, P. (2004). Forecasting with nonstationary dynamic factor models. *Journal of Econometrics*, *119*, 291–321.
Schumacher, C., & Breitung, J. (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, *24*, 386–398.
Stekler, H.O., & Talwar, R. (2011). *Economic forecasting in the great recession*. Working paper, 2011-005, Department of Economics, George Washington University.
Stock, J. H., & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. In *NBER macro-economic annual* (pp. 351–409).
Stock, J. H., & Watson, M. W. (1999). A comparison of linear and nonlinear models for forecasting macroeconomic time series. In R. F. Engle, & H. White (Eds.), *Cointegration, causality and forecasting* (pp. 1–44). Oxford: Oxford University Press.
Stock, J.H., & Watson, M.W. (2008). *Phillips curve inflation forecasts*. Working paper 14322, NBER, Cambridge, MA.
Stock, J.H., & Watson, M.W. (2009). *Forecasting in dynamic factor models subject to structural instability*. See Castle and Shephard (2009) (Chapter 7).
Stock, J.H., & Watson, M.W. (2010). *Modelling inflation after the crisis*. NBER Working paper series 16488.
Stock, J.H., & Watson, M.W. (2011). Dynamic factor models. See Clements and Hendry (2011) (Chapter 2).

Jennifer L. Castle is an Official Fellow in Economics at Magdalen College, Oxford, and Fellow of the Institute for New Economic Thinking at the Oxford Martin School, Oxford University. She previously held a British Academy postdoctoral research fellowship at Nuffield College, Oxford. Her research interests lie in the fields of model selection and forecasting, in which she has published more than 15 articles.

Michael P. Clements is Professor of Econometrics at the ICMA Centre, Henley Business School, University of Reading, and a member of the Institute for New Economic Thinking at the Oxford Martin School, Oxford University. He was previously Professor of Econometrics, Warwick University, and was an Editor of the IJF from 2001 to 2012, where he is currently an Associate Editor. He has published widely in the areas of time-series econometrics and forecasting, is a Distinguished Author *Journal of Applied Econometrics*, and an Honorary Fellow, International Journal of Forecasters.

David F. Hendry, Kt, is Professor of Economics Oxford University, Fellow of Nuffield College, and Director, Program in Economic Modeling, INET Oxford. He was previously Professor of Econometrics, London School of Economics. He was Knighted in 2009; is an Honorary Vice-President and past President Royal Economic Society; Fellow, British Academy, Royal Society of Edinburgh, Econometric Society and *Journal of Econometrics*; Foreign Honorary Member, American Economic Association and American Academy of Arts and Sciences; Honorary Fellow, International Institute of Forecasters and an Academician of the Academy of Social Sciences. He has received eight Honorary Doctorates, as well as the Guy Medal in Bronze from the Royal Statistical Society. He is listed by the ISI as one of the world's 200 most cited economists, is a Thomson Reuters Citation Laureate, and has published more than 200 papers and 25 books on econometric methods, theory, modelling, and history; numerical techniques; computing; empirical economics; and forecasting, for which he received the Isaac Kerstenetzky Scholarly Achievement Award.