International Journal of Forecasting 33 (2017) 359–372

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Evaluating multi-step system forecasts with relatively few forecast-error observations

David F. Hendry, Andrew B. Martinez *

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

ARTICLE INFO

Keywords:
Invariance
Forecast evaluation
Forecast error
Moment matrices
MSFE
GFESM

ABSTRACT

This paper develops a new approach for evaluating multi-step system forecasts with relatively few forecast-error observations. It extends the work of Clements and Hendry (1993) by using that of Abadir et al. (2014) to generate "design-free" estimates of the general matrix of the forecast-error second-moment when there are relatively few forecast-error observations. Simulations show that the usefulness of alternative methods deteriorates when their assumptions are violated. The new approach compares well with these methods and provides correct forecast rankings.

1. Introduction

Both short- and long-term forecasts are important for policy debates. For example, short-term debt forecasts are linked to concerns about growth, while long-term debt forecasts are tied to debt sustainability (e.g., Martinez, 2015).[1] The joint importance of short- and long-term forecasts is not limited to the area of macroeconomics, but is found in a variety of topics, including climate change (e.g., Pretis & Roser, 2016). While forecasts are usually evaluated separately across horizons, Clements and Hendry (1993) illustrate that it is important to evaluate forecasts jointly in order to capture any dependence in the forecast errors across horizons. This allows for the assessment of forecast performance in both the short- and long-term.

While the joint evaluation of multiple forecast horizons is useful, so too is the evaluation of large forecasting systems. Large macroeconomic forecasting models which try to capture the dependencies across variables are used regularly by central banks.[2] Although individual variables from large forecasting systems are evaluated regularly, relatively little work has been done to evaluate the accuracy of the whole system jointly. One exception is the work of Sinclair, Stekler, and Carnow (2012, 2015), who evaluate a vector of forecasts simultaneously. However, there are limitations as to how easily these methods can be extended to evaluate multiple horizons jointly.

Despite the importance of analyzing forecasts jointly across horizons and variables, the available methods have limitations, due largely to a shortage of forecast-error observations. For example, Clements and Hendry (1993) propose the use of the general matrix of the forecast-error second-moment and its determinant (GFESM) as an invariant measure of the forecast accuracy. While the GFESM allows for the joint evaluation of forecast errors across variables and horizons, it deteriorates rapidly in relatively small samples. Thus, an improved approach for the evaluation of multi-step system forecasts with relatively few forecast-error observations is required.

* Corresponding author.
E-mail addresses: david.hendry@nuffield.ox.ac.uk (D.F. Hendry), andrew.martinez@economics.ox.ac.uk (A.B. Martinez).

1 See Standard & Poor's, "United States of America long-term rating lowered to 'AA+' on political risks and rising debt burden: outlook negative", August 2, 2011.

2 For example, see Bårdsen, den Reijer, Jonasson, and Nymoen (2012) and Burgess et al. (2013).

http://dx.doi.org/10.1016/j.ijforecast.2016.08.007
0169-2070/

This paper presents a solution to this problem by extending estimates of the GFESM to relatively small samples. It combines the work of Abadir, Distaso and Žikeš (2014) and Clements and Hendry (1993) to allow for the estimation of the GFESM when there are more variables (K) times forecast horizons (H) than forecast-error observations (N). This extends the GFESM to large forecasting systems with long horizons even when there are relatively few forecast-error observations.

The paper seeks to answer the following questions: How well does the standard approach for calculating the GFESM perform as KH approaches N? Is it possible to improve on the standard approach when the forecast-error second-moment matrix is singular or non-singular? Comparing several approaches, this paper extends estimates of the GFESM beyond the non-singular case to the singular case where $KH > N$. It yields several important findings. First, the standard approach is increasingly biased and imprecise when there are relatively few forecast-error observations, which can distort the forecast rankings. Second, the proposed method outperforms the standard approach across a variety of forecast models and data generation processes (DGPs). Third, the proposed method typically produces the correct forecast ranking even when there are relatively few observations.

The rest of the paper is structured as follows. The next section reviews different forecast accuracy evaluation methods, with a focus on the GFESM. Section 3 lays out a new analytical approach for calculating the GFESM when $KH > N$. Section 4 conducts Monte Carlo experiments for a known-parameters model in order to determine how well the proposed method performs relative to the standard approach in small-sample settings and across various DGPs. Section 5 extends the Monte Carlo experiments to alternative forecast models. Section 6 applies the methods to a vector of forecasts of the US economy. Section 7 concludes.

2. Existing methods for evaluating forecast accuracy

This section provides a brief introduction to the theory of forecasting, as well as a common approach to the evaluation of forecasts, namely the mean square forecast error (MSFE). It then goes on to introduce an alternative approach, the general matrix of the forecast-error second-moment and its determinant (GFESM). The advantages of the GFESM over MSFE measures are illustrated, together with its limitations and the ways in which it has been applied in the literature.

Consider a DGP that is defined by a stationary, pth-order vector autoregressive process (e.g., VAR(p)) for a vector of K variables $\mathbf{Y_t}$:

$$\mathbf{Y}_t = \boldsymbol{\theta} + \sum_{j=0}^{p-1} \boldsymbol{\Pi}_j \mathbf{Y}_{t-1-j} + \mathbf{v}_t, \quad \text{where } \mathbf{v}_t \sim IN_K[\mathbf{0}, \boldsymbol{\Omega}]. \tag{1}$$

The bold terms are vectors, and $\mathbf{v}_t$ is a $(K \times 1)$ vector of independent normal residuals with mean $\mathbf{0}$, positive definite variance $\boldsymbol{\Omega}$, and all of the eigenvalues of the polynomial matrix $\left(\mathbf{I}_K - \sum_{j=0}^{p-1} \boldsymbol{\Pi}_j L^{j+1}\right)$, where L denotes the lag operator, are inside the unit circle. Assuming that the initial value is equal to its long-run mean, $\boldsymbol{\Theta} = \left(\mathbf{I}_K - \sum_{j=0}^{p-1} \boldsymbol{\Pi}_j\right)^{-1} \boldsymbol{\theta}$, Eq. (1) can be rewritten as

$$\mathbf{X}_t = \sum_{j=0}^{p-1} \boldsymbol{\Pi}_j \mathbf{X}_{t-1-j} + \mathbf{v}_t, \tag{2}$$

where $\mathbf{X}_t = (\mathbf{Y}_t - \boldsymbol{\Theta})$ is demeaned by its long-run mean. Let $\boldsymbol{\Gamma}_{p,i} = \sum_{j=0}^{p-1} \boldsymbol{\Gamma}_{p,i-1-j} \boldsymbol{\Pi}_j$, $\boldsymbol{\Gamma}_{p,0} = \mathbf{I}_K$, $\boldsymbol{\Gamma}_{p,i} = \mathbf{0}$ when $i < 0$, and $\boldsymbol{\Pi}_j = \mathbf{0}$ when $j \geq p$. Then, the best possible h-step-ahead forecast at time T is the conditional expectation: $\mathbf{X}_{T+h|T} = \mathbb{E}_T[\mathbf{X}_{T+h}|\mathbf{X}_T] = \sum_{j=0}^{p-1} \boldsymbol{\Gamma}_{p,\max(0,h-j)} \boldsymbol{\Pi}_j^{\min(1,j)} \mathbf{X}_{T-j}$, where $h \in [1, \ldots, H]$. Therefore, given estimates of $\{\widehat{\boldsymbol{\Pi}}_j\}$, the forecast error is

$$\widetilde{\mathbf{u}}_{T+h|T} = \left(\mathbf{X}_{T+h} - \widetilde{\mathbf{X}}_{T+h|T}\right) = \sum_{j=0}^{p-1} \left(\boldsymbol{\Gamma}_{p,\max(0,h-j)} \boldsymbol{\Pi}_j^{\min(1,j)} - \widehat{\boldsymbol{\Gamma}}_{p,\max(0,h-j)} \widehat{\boldsymbol{\Pi}}_j^{\min(1,j)}\right) \mathbf{X}_{T-j} + \sum_{i=0}^{h-1} \boldsymbol{\Gamma}_{p,i} \mathbf{v}_{T+h-i}. \tag{3}$$

When the true parameters are known, Eq. (3) delivers unbiased forecast errors that have a variance of $\sum_{i=0}^{h-1} \boldsymbol{\Gamma}_{p,i} \boldsymbol{\Omega} \boldsymbol{\Gamma}'_{p,i}$. Thus, even the smallest possible multi-step forecast errors from a dynamic model are moving average processes.

The MSFE, which assumes a quadratic loss function, is used commonly to evaluate forecasts. In multivariate systems, the MSFE becomes the mean square forecast error matrix (or the matrix of the forecast-error second-moment, MFESM):

$$\mathbb{E}_T\left[\widetilde{\mathbf{u}}_{T+h|T} \widetilde{\mathbf{u}}'_{T+h|T}\right] = \mathbf{V}_h = \sum_{i=0}^{h-1} \boldsymbol{\Gamma}_{p,i} \boldsymbol{\Omega} \boldsymbol{\Gamma}'_{p,i}, \tag{4}$$

where the last equality holds when the true parameters are known and $\mathbf{v}_t$ is IID. Multivariate forecasts are often evaluated using the trace of the MFESM: $tr(\mathbf{V}_h)$.

2.1. An invariant measure of forecast accuracy

Clements and Hendry (1993, 1998) propose a more general and invariant measure of the forecast accuracy, termed the general matrix of the forecast-error second-moment (GMFESM), $\boldsymbol{\Phi}_H$, with its determinant $|\boldsymbol{\Phi}_H|$, the GFESM. The GMFESM is estimated by multiplying the stacked forecast errors across all horizons and variables. Following from Eq. (3), the forecast error $\widetilde{\mathbf{u}}_{T+h+n|T+n}$ is a $(K \times 1)$ vector of $h = 1, \ldots, H$ horizons from origin $T+n$, where $n = 0, \ldots, N-1$ represents the number of forecast-error observations.[3] Stack the forecast errors as

$$\widetilde{\mathbf{W}}_{H,N|T} = \begin{pmatrix} \widetilde{\mathbf{u}}_{T+1|T} & \cdots & \widetilde{\mathbf{u}}_{T+N|T+N-1} \\ \vdots & \ddots & \vdots \\ \widetilde{\mathbf{u}}_{T+H|T} & \cdots & \widetilde{\mathbf{u}}_{T+H+N-1|T+N-1} \end{pmatrix}, \tag{5}$$

[3] Note that the literature uses the terms 'forecast origins' and 'forecast-error observations' interchangeably. The latter term is used here.