ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

Evaluating Multi-Step System Forecasts with Relatively Few Forecast-Error Observations

David F. Hendry and Andrew B. Martinez

Number 784
March 2016

Manor Road Building, Oxford OX1 3UQ

Evaluating Multi-Step System Forecasts with Relatively Few Forecast-Error Observations

David F. Hendry and Andrew B. Martinez

Department of Economics, and Institute for New Economic Thinking,

Oxford Martin School, University of Oxford, UK.*

March 7, 2016

Abstract

This paper develops a new approach for evaluating multi-step system forecasts with relatively few forecast-error observations. It extends Clements and Hendry (1993a) using 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 against these methods and provides correct forecast rankings.

Keywords: Invariance, Forecast Evaluation, Forecast Error, Moment Matrices, MSFE, GFESM

JEL classifications: C22, C32, C53

1 Introduction

Long-term forecasts are increasingly used in policy debates. For example, arguments in support of secular stagnation are tied to the evolution of long-term forecasts of potential GDP (e.g. Summers (2014)). Similarly, concerns about debt sustainability are often tied to long-term debt forecasts.[1] The use of long-term forecasts in policy debates is not just constrained to macroeconomics but has proliferated in a wide variety of other topics including climate change (e.g. Pretis and Roser (2016)). Despite the importance of long-term forecasts, there are relatively few analyses of how well they perform. This is partially due to a lack of forecast-error observations,

*This research was supported in part by grants from the Robertson Foundation and INET. We are indebted to Karim M. Abadir, Jennifer L. Castle, Søren Johansen, Jurgen A. Doornik, Bent Nielsen, Felix Pretis, participants at the 16th OxMetrics Users Conference and at the INET Oxford Research Seminar for helpful suggestions and comments on earlier drafts. All errors are our own. Contact: david.hendry@nuffield.ox.ac.uk; andrew.martinez@economics.ox.ac.uk.

[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.

which stems from the lag between when long-term forecasts are produced and when the actual outcome is observed.

Simultaneously, while forecast horizons have increased, forecasting systems have grown larger. Large macroeconomic forecasting models are regularly used by central banks.[2] Although forecasts of individual variables from large forecasting systems are evaluated regularly, relatively little work has been done to evaluate the accuracy of the whole system across all forecast horizons.One exception is Sinclair et al. (2012, 2015), who evaluate a vector of forecasts simultaneously. However, there are limitations in terms of how easily their method can be extended to simultaneously evaluate multiple horizons.

Many of the methods used to evaluate one-step-ahead forecasts are not sufficient to evaluate multi-step-ahead forecasts. Clements and Hendry (1993a) show that the mean square forecast error (MSFE) is not invariant to certain transformations when the forecast horizon is greater than one-step-ahead. Instead, they propose using the general matrix of the forecast-error second-moment and its determinant (GFESM) as an invariant measure of forecast accuracy. However, the GFESM deteriorates in relatively small samples. Thus, a new method for evaluating multi-step system forecasts when there are relatively few forecast-error observations is required.

This paper presents a solution to this problem by extending estimates of the GFESM to relatively small samples. It combines Clements and Hendry (1993a) with Abadir et al. (2014) to allow for 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 estimating the GFESM perform as KH gets closer to N? Is it possible to improve on the standard approach in settings where the forecast error 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 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 ranking of forecasts 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 ap-

[2]For example, see Bårdsen et al. (2012) and Burgess et al. (2013).

proach for estimating the GFESM when $KH < N$. Section 4 conducts Monte Carlo experiments for a known parameters model to examine 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 method to a vector of forecasts of the US economy. Section 7 concludes.

2 Existing Methods to Evaluate Forecast Accuracy

This section provides a brief introduction to the theory of forecasting as well as a common approach to evaluating forecasts, i.e. the mean square forecast error (MSFE). Subsequently, it introduces 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 as well as its limitations and how it has been applied in the literature.

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

$$\mathbf{Y}_t = \boldsymbol{\theta} + \sum_{j=0}^{p-1} \mathbf{\Pi}_j \mathbf{Y}_{t-1-j} + \mathbf{v}_t, \text{ where } \mathbf{v}_t \sim IN_K\left[\mathbf{0}, \mathbf{\Omega}\right] \text{ and } \left|\mathbf{\Pi}_j\right| < 1. \tag{2.1}$$

Bold terms represent vectors, and $\mathbf{v}_t$ is a $(K \times 1)$ vector of independent normal residuals with mean $\mathbf{0}$, variance $\mathbf{\Omega}$ and all the eigenvalues of the polynomial matrix $\left(\mathbf{I}_K - \sum_{j=0}^{p-1} \mathbf{\Pi}_j \mathbf{L}^{j+1}\right)$ where $\mathbf{L}$ denotes the lag operator, are inside the unit circle. Assuming the initial value is equal to its long-run mean, $\mathbf{\Theta} = \left(\mathbf{I}_K - \sum_{j=0}^{p-1} \mathbf{\Pi}_j\right)^{-1} \boldsymbol{\theta}$, then equation (2.1) can be rewritten as

$$\mathbf{X}_t = \sum_{j=0}^{p-1} \mathbf{\Pi}_j \mathbf{X}_{t-1-j} + \mathbf{v}_t, \tag{2.2}$$

where $\mathbf{X}_t = (\mathbf{Y}_t - \mathbf{\Theta})$ is demeaned by its long-run mean. The best possible h-step-ahead forecast at time T is the conditional expectation: $\mathbf{X}_{T+h|T} = \mathbb{E}_T\left[\mathbf{X}_{T+h}|\mathbf{X}_T\right] = \sum_{j=0}^{p-1} \mathbf{\Pi}_j^{h-j} \mathbf{X}_{T-j}$, where $h \in [1, ..., H]$. Therefore given estimates of $\left\{\widehat{\mathbf{\Pi}}_j\right\}$, the smallest forecast error is

$$\widetilde{\mathbf{u}}_{T+h|T} = \left(\mathbf{X}_{T+h} - \widetilde{\mathbf{X}}_{T+h|T}\right) = \sum_{j=0}^{p-1} (\mathbf{\Gamma}_{p,max(0,h-j)} \mathbf{\Pi}_j^{min(1,j)} - \widehat{\mathbf{\Gamma}}_{p,max(0,h-j)} \widehat{\mathbf{\Pi}}_j^{min(1,j)}) \mathbf{X}_{T-j} + \sum_{i=0}^{h-1} \mathbf{\Gamma}_{p,i} \mathbf{v}_{T+h-i}, \tag{2.3}$$

where $\mathbf{\Gamma}_{p,i} = \sum_{j=0}^{p-1} \mathbf{\Gamma}_{p,i-1-j} \mathbf{\Pi}_j$, $\mathbf{\Gamma}_{p,0} = \mathbf{I}_K$, $\mathbf{\Gamma}_{p,i} = \mathbf{0}$ when $i < 0$, and $\mathbf{\Pi}_j = \mathbf{0}$ when $j \geq p$. When the true parameters are known, equation (2.3) delivers unbiased forecast errors that have a variance of $\sum_{i=0}^{h-1} \mathbf{\Gamma}_{p,i} \mathbf{\Omega} \mathbf{\Gamma}'_{p,i}$. Thus, even the smallest possible multi-step forecast errors from a dynamic model are a moving average process.

The MSFE, which assumes a quadratic loss function, is commonly used 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):

$$\mathbb{E}_T\left[\left(\mathbf{X}_{T+h} - \widetilde{\mathbf{X}}_{T+h|T}\right)\left(\mathbf{X}_{T+h} - \widetilde{\mathbf{X}}_{T+h|T}\right)'\right] = \mathbb{E}_T\left[\widetilde{\mathbf{u}}_{T+h|T} \widetilde{\mathbf{u}}'_{T+h|T}\right] = \mathbf{V}_h = \sum_{i=0}^{h-1} \mathbf{\Gamma}_{p,i} \mathbf{\Omega} \mathbf{\Gamma}'_{p,i}, \tag{2.4}$$

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: $tr\,(V_h)$.

2.1 An invariant measure of forecast accuracy

Clements and Hendry (1993a, 1998) propose a more general and invariant measure of forecast accuracy termed the general matrix of the forecast-error second-moment (GMFESM) $\mathbf{\Phi}_H$ and its determinant $|\mathbf{\Phi}_H|$, the GFESM. The GMFESM is estimated by multiplying the stacked forecast errors across all horizons and variables. Following from equation (2.3) where the forecast error, $\widetilde{\mathbf{u}}_{T+h+n|T+n}$, is a $(K \times 1)$ vector of $h = 1, ..., H$ horizons from origin $T + n$, where $n = 0, ..., N - 1$ represents the number of forecast-error observations.[3] For $h = 1. . . . , H$ and $n = 0, . . . , N - 1$ it is possible to stack the forecast errors such that

$$\widetilde{\mathbf{W}}_{H,N|T} = \begin{pmatrix} \widetilde{\mathbf{u}}_{T+1|T} & \cdots & \widetilde{\mathbf{u}}_{T+N|T+N-1} \\ \vdots & \ddots & \vdots \\ \widetilde{\mathbf{u}}_{T+H|T} & \cdots & \widetilde{\mathbf{u}}_{T+H+N-1|T+N-1} \end{pmatrix} \tag{2.5}$$

is a $KH \times N$ matrix. Then the GMFESM is a $KH \times KH$ matrix

$$\widehat{\mathbf{\Phi}}_H = \tfrac{1}{N}\widetilde{\mathbf{W}}_{H,N|T}\widetilde{\mathbf{W}}'_{H,N|T} = \begin{pmatrix} \frac{1}{N}\sum_{n=0}^{N-1}\widetilde{\mathbf{u}}_{T+1+n|T+n}\widetilde{\mathbf{u}}'_{T+1+n|T+n} & \cdots & \frac{1}{N}\sum_{n=0}^{N-1}\widetilde{\mathbf{u}}_{T+1+n|T+n}\widetilde{\mathbf{u}}'_{T+H+n|T+n} \\ \vdots & \ddots & \vdots \\ \frac{1}{N}\sum_{n=0}^{N-1}\widetilde{\mathbf{u}}_{T+H+n|T+n}\widetilde{\mathbf{u}}'_{T+1+n|T+n} & \cdots & \frac{1}{N}\sum_{n=0}^{N-1}\widetilde{\mathbf{u}}_{T+H+n|T+n}\widetilde{\mathbf{u}}'_{T+H+n|T+n} \end{pmatrix}. \tag{2.6}$$

The GFESM is the determinant this equation. Equation (2.6) illustrates that each element along the main diagonal is the MSFE for all variables and horizons. Thus, the trace of this matrix is the trace MSFE over all K and H (TFESM). This illustrates the connection between the GFESM and more conventional forecast evaluation methods. It also shows that the GFESM encompasses all of the information available in the TFESM while also incorporating information from the covariances across variables and horizons.

The known-parameters case offers further insight into the properties of the GFESM. When the true parameters are known, but the error variance is estimated by the assumptions of equation (2.1) and $\widehat{\mathbf{\Omega}}_{N+h} = \frac{1}{N+h}\sum_{n=0}^{N-1+h}\mathbf{v}_{T+n}\mathbf{v}'_{T+n}$, then plugging equation (2.3) into equation (2.6), simplifying and then taking the determinant gives the GFESM

$$\left|\widehat{\mathbf{\Phi}}_H\right| = \begin{vmatrix} \mathbf{\Gamma}_{p,0}\widehat{\mathbf{\Omega}}_{1+N}\mathbf{\Gamma}'_{p,0} & \cdots & \mathbf{\Gamma}_{p,0}\widehat{\mathbf{\Omega}}_{1+N}\mathbf{\Gamma}'_{p,H-1} \\ \vdots & \ddots & \vdots \\ \mathbf{\Gamma}_{p,H-1}\widehat{\mathbf{\Omega}}_{1+N}\mathbf{\Gamma}'_{p,0} & \cdots & \sum_{i=0}^{H-1}\mathbf{\Gamma}_{p,i}\widehat{\mathbf{\Omega}}_{H+N-i}\mathbf{\Gamma}'_{p,i} \end{vmatrix} = \begin{vmatrix} \mathbf{\Gamma}_{p,0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Gamma}_{p,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{\Gamma}_{p,H-1} & \cdots & \mathbf{\Gamma}_{p,1} & \mathbf{\Gamma}_{p,0} \end{vmatrix} \begin{vmatrix} \widehat{\mathbf{\Omega}}_{1+N} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \widehat{\mathbf{\Omega}}_{H+N} \end{vmatrix} \begin{vmatrix} \mathbf{\Gamma}_{p,0} & \mathbf{\Gamma}_{p,1} & \cdots & \mathbf{\Gamma}_{p,H-1} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{\Gamma}_1 \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{\Gamma}_{p,0} \end{vmatrix} = \prod_{h=1}^{H}\left|\widehat{\mathbf{\Omega}}_{h+N}\right|. \tag{2.7}$$

The second equality follows since the determinant of a product is the product of the determinants. The third follows since the determinant of the matrix of coefficients is unity since $\mathbf{\Gamma}_{p,0} = \mathbf{I}_K$. Thus, in this special case the GFESM is the product of the estimated error variances across horizons.

[3] Note that in the literature forecast origins and forecast-error observations are used interchangeably. Here the latter term is used.

In finite samples when the true error variance is known and unchanging, or in large samples when the error variance converges to its true value such that $\boldsymbol{\Omega} = \widehat{\boldsymbol{\Omega}}_{N+1} = \ldots = \widehat{\boldsymbol{\Omega}}_{N+H}$, then the GFESM simplifies further, so standardized by the forecast horizon H is

$$|\boldsymbol{\Phi}_H|^{1/H} = \left(|\boldsymbol{\Omega}|^H\right)^{1/H} = |\boldsymbol{\Omega}| \,. \tag{2.8}$$

The standardized GFESM is useful since the GFESM increases with the forecast horizon. Clements and Hendry (1995, p. 135) note that this transformation "stabilizes estimates of the GFESM for large H when $|\boldsymbol{\Phi}_1|$ differs from unity" without any impact on forecast comparisons. However, it may also "downplay differences when parameters are estimated" (p. 139).

This case illustrates that when the true parameters are known, the one-step-ahead forecast errors specify the complete ranking across all horizons. Several studies, notably Engle (1993) and Gooijer and Vidiella-i-Anguera (2004), are critical of this result. They argue that the GFESM attributes more weight to the one-step forecast errors than other forecast evaluation methods. Clements and Hendry (1993b), Schmidt (1993) and Castle and Hendry (2008) note that although forecast uncertainty increases monotonically when all parameters are known, this does not hold when the error variance is not constant or when parameters are unknown in finite samples due to parameter estimation uncertainty. Further, this result is not unique to the GFESM. When all parameters including the error variance are known, then the TFESM relies on the one-step forecast errors.

While the GFESM is not unique in terms of the relative importance of the one-step forecast errors, it is unique in terms of its invariance properties under certain widely used transformations. Consider transforming the forecast errors by the matrix $\boldsymbol{M}_{K,h}$ such that the vector of forecast errors becomes $\boldsymbol{M}_{K,h}\widetilde{\mathbf{u}}_{T+h+n|T+n}$. Generalizing Clements and Hendry (1998, p. 73) to allow for transformations of variables and forecast horizons, the transformed matrix of stacked forecast errors in equation (2.5) becomes $\widetilde{\mathbf{W}}^*_{H,N|T} = (\boldsymbol{M}_H \otimes \boldsymbol{M}_K)\,\widetilde{\mathbf{W}}_{H,N|T}$ where $\boldsymbol{M}_H$ is an alternative weighting matrix across horizons while $\boldsymbol{M}_K$ is a scaling matrix across variables and $\otimes$ is the Kronecker product. Estimating the GFESM with the transformed stacked forecast errors gives

$$\left|\widehat{\boldsymbol{\Phi}}^*_H\right| = \left|\frac{1}{N}\widetilde{\mathbf{W}}^*_{H,N|T}\widetilde{\mathbf{W}}^{*\prime}_{H,N|T}\right| = \left|\widehat{\boldsymbol{\Phi}}_H\right| \times |\boldsymbol{M}_H|^{2H} \times |\boldsymbol{M}_K|^{2H} \,. \tag{2.9}$$

First consider the case where $abs\,(|\boldsymbol{M}_H|) = 1$ and $abs\,(|\boldsymbol{M}_K|) = 1$ such that transformations are scale preserving and non-singular. Included in this class of transformations are cointegrating combinations, differences, differentials and substitution of identities among others. Equation (2.9) shows that the GFESM is invariant to this class of transformation, $\left|\widehat{\boldsymbol{\Phi}}^*_H\right| = \left|\widehat{\boldsymbol{\Phi}}_H\right|$, and so the ranking of forecasts is unchanged. This result is not generally true for the TFESM, where $tr\left(\widehat{\boldsymbol{\Phi}}^*_H\right) \neq tr\left(\widehat{\boldsymbol{\Phi}}_H\right)$. For example, the choice between evaluating forecasts in levels or in

differences is arbitrary in terms of the GFESM but not in terms of the TFESM.[4]

Now consider the case where $|\boldsymbol{M}_K|$ (and/or $|\boldsymbol{M}_H|$) $\neq 1$. For example, when forecast errors have alternative scales across variables. The GFESM is not invariant to this transformation. However, as equation (2.9) illustrates, this will just rescale the original GFESM by the determinant of the respective weight matrix. Thus, transformations that are applied uniformly across all forecast models have no impact on the relative forecast rankings.

Interpretation of whether one forecast dominates the other in terms of the GFESM is similar to the MFESM and MSFE criteria. For example, consider the forecasts from two different models (A and B). Model B's forecasts dominate model A's forecasts in terms of the mean square forecast error if the MSFE for model B is smaller than that of model A. Similarly, this is true in the system case for the GMFESM, where the difference between the GMFESM of models A and B gives a positive definite matrix. It follows that GMFESM dominance is a sufficient but not a necessary condition for GFESM dominance where

$$\widehat{\boldsymbol{\Phi}}_H^A - \widehat{\boldsymbol{\Phi}}_H^B \succ \mathbf{0} \text{ implies } \left|\widehat{\boldsymbol{\Phi}}_H^A\right| > \left|\widehat{\boldsymbol{\Phi}}_H^B\right|. \tag{2.10}$$

Clements and Hendry (1998, p. 74) argue that "GFESM dominance is a weaker condition" than MFESM dominance. This can be rewritten as $\left|\widehat{\boldsymbol{\Phi}}_H^A\right| / \left|\widehat{\boldsymbol{\Phi}}_H^B\right| - 1 > 0$, which is close to $ln\left(\left|\widehat{\boldsymbol{\Phi}}_H^A\right| / \left|\widehat{\boldsymbol{\Phi}}_H^B\right|\right) > 0$ when $\left|\widehat{\boldsymbol{\Phi}}_H^A\right| \approx \left|\widehat{\boldsymbol{\Phi}}_H^B\right|$. Thus, it is common to report the log GFESM whose relation to the profile predictive likelihood is illustrated in Clements and Hendry (1993a, 1998).

2.2 Limitations of the GFESM

Despite its advantages relative to MSFE measures, the GFESM has an important limitation which hampers its applicability in practice. While the GMFESM is a positive definite moment matrix of the stacked forecast errors, estimates of the GMFESM are singular and so estimates of the GFESM go to zero when variables are collinear or when there are more variables times horizons (KH) than forecast-error observations (N).[5] While the former is solvable by excluding collinear variables or through other adjustments, the latter does not have a simple solution.

Since the GFESM tends to zero when $KH > N$, the literature tends to steer clear of this case. While Clements and Hendry (1995) study the "small sample properties" of the GFESM, their analysis remains well away from the singular GMFESM. Other studies limit their use of the GFESM by focusing on either variables or horizons. For example, Clements and Hendry (1997) focus on cross-variable forecast errors from five different transformations of the same variable. Paap et al. (1997) use the cross-horizon forecast errors of consumption data to compute the

[4] See Clements and Hendry (1995, 1997). Note that methods that rely on non-linear transformations, such as those discussed below, generate finite sample estimates of the GFESM that are not invariant.

[5] Granger (1993) first pointed this out for the GFESM. Strictly speaking the former is considered singular while the latter is more precisely defined as ill-conditioned. However here they are referred to interchangeably.

Table 2.1: Overview of Previous Studies that apply the GFESM

Study	K (max)	H (max)	KH	N (min)
Clements and Hendry (1995)	2	20	40	50
Hoffman and Rasche (1996)	5	16	80	28
Clements and Hendry (1997)	5	8	40	22
Paap et al. (1997)	1	8	8	16
Kunst and Franses (1998)	3	8	24	28
Zeng and Swanson (1998)	5	5	25	1,400
Anderson et al. (2002)	6	16	96	36
Vahid and Issler (2002)	3	16	48	100
Gooijer and Ray (2003)	2	5	10	10
Clements and Galvão (2004)	2	24	48	100
Gooijer and Vidiella-i-Anguera (2004)	2	5	10	40
Veloce (2004)	1	8	8	13
Castle and Hendry (2008)	1	8	8	20
Bowsher and Meeks (2008)	6	1	6	72
Baillie and Morana (2009)	1	12	12	940
Athanasopoulos et al. (2011)	3	16	48	90
Simionescu (2013)	4	1	4	40

GFESM. Other studies that take this approach include Kunst and Franses (1998) , Gooijer and Ray (2003), Gooijer and Vidiella-i-Anguera (2004), Veloce (2004), Castle and Hendry (2008), Athanasopoulos et al. (2011) and Simionescu (2013). While the GMFESM is non-singular in most of these studies, in some cases there are only just enough forecast-error observations.

The GFESM is not limited in all applications, as some (particularly in finance), have relatively large samples of forecast-error observations. For example, see Zeng and Swanson (1998), Clements and Galvão (2004), Bowsher and Meeks (2008) and Baillie and Morana (2009). See Table 2.1 for how close previous studies get to $KH > N$.

Hoffman and Rasche (1996) is one of the only studies to apply the GFESM when $KH > N$. They impose restrictions on the estimated GMFESM based on assumptions about its structure. The crucial assumption is that the forecast errors are uncorrelated across variables. While the "cross-variable implications of the system forecast ranking" are lost by imposing these restrictions, they justify this by claiming that the rankings do not change substantially at shorter horizons when comparisons between the constrained and unconstrained GFESM are feasible. Anderson et al. (2002) also use a similar approach.

When there are relatively few observations, the focus is either on cross-horizon or cross-variable forecast errors, rather than both. Studies that do tackle this case, notably Hoffman and Rasche (1996) and Anderson et al. (2002), make assumptions about the structure of the GMFESM. Several other studies, including Kunst and Franses (1998) and Gooijer and Ray (2003) are close to the boundary of the singular GMFESM, but do not analyze the performance of the estimates. While the GFESM is estimable in such cases, it is unclear how outcomes are affected by assumptions about the error structure or as observations get close to the singular case. The rest of this paper addresses these questions.

3 Methodological Approach

While few approaches exist for estimating moment matrices with relatively few observations, there is an extended literature for covariance matrices. Several methods in the literature ensure estimates of covariance matrices remain well-conditioned (i.e. positive definite). However, they typically require imposing relatively stringent assumptions about the structure. As Abadir et al. (2014) notes, there are three main avenues in the literature to ensure positive definiteness. These include imposing restrictions, using factor models, and a class of methods known as shrinkage.

Few of these approaches have been applied in the context of the GFESM. The only exception is the method proposed in Hoffman and Rasche (1996), which imposes zero restrictions by assuming that forecast errors are uncorrelated across variables and allows the estimated GMFESM to be positive definite as long as $\boldsymbol{H} \leq \boldsymbol{N}$. This suggests that there are potentially large improvements in estimates of the GFESM by using more flexible methods when $\boldsymbol{KH} > \boldsymbol{N}$.

One method developed explicitly for time series applications and potentially applicable to the GFESM is proposed in McMurry and Politis (2010) and extended in Jentsch and Politis (2015). This method is classified as a shrinkage approach in that it down weights off-diagonal elements of the covariance matrix.[6] However, while the benefits of this method are clear in terms of standard autocovariance matrices, they are less clear for the GMFESM where forecast error covariances are less likely to taper off slowly.

This paper focuses instead on a more flexible approach to ensure the GMFESM remains positive definite. A "design-free" method was proposed by Abadir et al. (2014) which does not impose any assumptions about the distribution of the sample or the parametric structure of the matrix. This gives the method much more flexibility and Abadir et al. (2014) demonstrate that it performs well relative to a variety of approaches in different settings.

The crucial assumption imposed by this method is that the data are IID. However, since multi-step forecast errors from any dynamic model are serially correlated, this assumption is not satisfied. Abadir et al. (2014) suggest that one way to avoid this issue is to filter the data with an appropriate model. Instead, this paper transforms the stacked forecast errors to eliminate serial correlation. However, this approach requires knowledge of the autoregressive coefficient.

In the known parameters case, if equation (2.3) is substituted into equation (2.5), Clements

[6] Even so, this approach only generates estimates that are asymptotically positive definite. Thus, in finite samples additional constraints are imposed on estimates of the eigenvalues.

and Hendry (1998, p. 315) show that the stacked forecast errors can be decomposed as

$$\widetilde{\mathbf{W}}_{H,N|T} = \begin{pmatrix} \mathbf{\Gamma}_{p,0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Gamma}_{p,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{\Gamma}_{p,H-1} & \cdots & \mathbf{\Gamma}_{p,1} & \mathbf{\Gamma}_{p,0} \end{pmatrix} \begin{pmatrix} \mathbf{v}_{T+1} & \cdots & \cdots & \mathbf{v}_{T+N} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{v}_{T+H} & \cdots & \cdots & \mathbf{v}_{T+H+N-1} \end{pmatrix} = \mathbf{\Psi}_{p,H}\mathbf{Z}_{H,N}. \quad (3.1)$$

This result also follows from equation (2.7). $\mathbf{Z}_{H,N}$ no longer contains terms that are serially-correlated. When this decomposition is applied to the GMFESM it becomes

$$\widehat{\mathbf{\Phi}}_H = \mathbf{\Psi}_{p,H}\mathbf{\Sigma}_{H,N}\mathbf{\Psi}'_{p,H}, \quad (3.2)$$

where $\mathbf{\Sigma}_H = \frac{1}{N}\mathbf{Z}_{H,N}\mathbf{Q}_N\mathbf{Z}'_{H,N}$, and $\mathbf{Q}_N$ is a demeaning matrix. In the known parameters case the mean is zero so the moment and covariance matrices are the same. However, in misspecified models, the squared bias needs to be added to get the GMFESM.

When the forecast model is known, then under the assumption that it is well-specified the autoregressive coefficient can be estimated. Alternatively, when the underlying forecast model is not known, then estimates of the autoregressive coefficient can be be derived from the forecast errors. This is done by taking the h-step-ahead forecast error at $T+n$ as in equation (2.3) and subtracting the $h-1$-step-ahead forecast error at $T+n+1$. Assuming that the forecast model and the DGP do not change substantially across observations, this transformation gives

$$\Delta\widetilde{\widetilde{\mathbf{u}}}_{T+h|T+n} = \widetilde{\mathbf{u}}_{T+h|T+n} - \widetilde{\mathbf{u}}_{T+h-1|T+n+1} = \mathbf{\Gamma}_{p,h-1}\mathbf{v}_{T+1}. \quad (3.3)$$

$\mathbf{\Gamma}_{p,h-1}$ can be estimated by regressing the 1-step-ahead forecast errors, $\{\mathbf{u}_{T+h|T+n}\}_{n=0}^{N-2}$, on the quasi-differenced h-step-ahead forecast errors, $\{\Delta\widetilde{\mathbf{u}}_{T+h|T+n}\}_{n=0}^{N-2}$. Repeating this analysis for $h = \{2, ..., H\}$ traces out $\mathbf{\Psi}_{p,H}$, which is necessary when the forecast model is unknown. Note that the estimation sample is $N-1$, so estimates of $\mathbf{\Gamma}_{p,h-1}$ become less reliable in small samples.

Once an estimate of $\mathbf{\Psi}_{p,H}$ is constructed, then the forecast errors can then be transformed to eliminate serial correlation: $\widetilde{\mathbf{Z}}_{H,N} = \mathbf{\Psi}_{p,H}^{-1}\widetilde{\mathbf{W}}_{H,N|T}$. As long as the parameters are estimated so that $\widehat{\mathbf{\Psi}}_{p,H}^{-1} \approx \mathbf{\Psi}_{p,H}^{-1}$ and the forecast model is reasonably well specified, then this transformation will eliminate most serial correlation. The transformed covariance matrix becomes $\widehat{\mathbf{\Sigma}}_{H,N} = \frac{1}{N}\widetilde{\mathbf{Z}}_{H,N}\widehat{\mathbf{Q}}_N\widetilde{\mathbf{Z}}'_{H,N}$, which remains singular when $KH > N$.[7] Even after this transformation, in the context of forecasting $\widetilde{\mathbf{Z}}_{H,N}$ is only quasi-IID. While this transformation ensures independence, there is no guarantee that the residuals are identically distributed over time. Hendry and Mizon (2014) argue that shifts in the distribution are a major source of forecast failure.

Let $\widehat{\mathbf{\Sigma}}_{H,N}$ be the $KH \times KH$ positive definite covariance matrix of the transformed stacked forecast errors. This generalizes the Abadir et al. (2014) approach, which is a special case where

[7] For the "standard" approach in the known parameters case: $\left|\widehat{\mathbf{\Sigma}}_H\right| = \left|\widehat{\mathbf{\Phi}}_H\right|$ as seen from equation (2.7).

$H = 1$, $\mathbf{B}_H = \mathbf{0}$, and $\mathbf{\Pi} = \mathbf{0}$ so that $\widetilde{\mathbf{W}}_{N|T} = \widetilde{\mathbf{Z}}_N$ and where $\mathbf{v}_T \sim IN_K[\mathbf{0}, \mathbf{\Omega}]$.[8] The covariance matrix of the transformed forecast errors can be decomposed as

$$\widehat{\mathbf{\Sigma}}_{H,N} = \frac{1}{N}\widetilde{\mathbf{Z}}_{H,N}\widehat{\mathbf{Q}}_N\widetilde{\mathbf{Z}}'_{H,N} = \widehat{\mathbf{P}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{P}}', \tag{3.4}$$

where $\widehat{\mathbf{P}}$ is an orthogonal matrix of eigenvectors with a typical column $\widehat{\mathbf{p}}_i$ $(i = 1, ..., KH)$ and $\widehat{\mathbf{\Lambda}}$ is the diagonal matrix of eigenvalues. Since $\widehat{\mathbf{P}}$ is orthogonal, it is well-conditioned for any value of N and KH. Therefore, the only source of ill-conditioning in $\widehat{\mathbf{\Sigma}}_H$ when $KH > N$ stems from the estimation of the matrix of eigenvalues, $\widehat{\mathbf{\Lambda}}$, since the estimated eigenvalues beyond N tend to zero. This illustrates why estimation is limited by the relative number of observations for the GFESM but not for the TFESM, since the determinant of a symmetric matrix is equal to the product of its eigenvalues, $\prod_{i=0}^{K\times H}\lambda_i$, while the trace is equal the sum of its eigenvalues, $\sum_{i=0}^{K\times H}\lambda_i$, see Abadir and Magnus (2005, Chapter 7).

The "design-free" method exploits the well-conditioning of the eigenvectors to generate an alternative estimator of the eigenvalues, $\widehat{\mathbf{\Lambda}}$. Using the fact that the eigenvectors are normalized so that $\widehat{\mathbf{P}}'\widehat{\mathbf{P}} = \boldsymbol{I}_{K\times H}$, equation (3.4) can be rearranged as

$$\widehat{\mathbf{\Lambda}} = \widehat{\mathbf{P}}'\widehat{\mathbf{\Sigma}}_{H,N}\widehat{\mathbf{P}} = diag(\widehat{var}(\widehat{\mathbf{p}}'_1\widetilde{Z}_{H,N}), \ldots, \widehat{var}(\widehat{\mathbf{p}}'_{K\times H}\widetilde{Z}_{H,N})), \tag{3.5}$$

where $\widehat{var}(\widehat{\mathbf{p}}'_i\widetilde{Z}_{H,N}) = \frac{1}{N}p_i\widetilde{\mathbf{Z}}_{H,N}\widehat{\mathbf{Q}}_N\widetilde{\mathbf{Z}}'_{H,N}p'_i$. The second equality follows since $\widehat{\mathbf{\Lambda}}$ is diagonal. The innovation in this method is to base the estimates of the eigenvectors, $\widehat{\mathbf{P}}$, on a subsample of the forecast-error observations. The IID assumption for $\widetilde{\mathbf{Z}}_{H,N}$ implies that it is efficient to use $s < N$ observations. Splitting the transformed stacked forecast errors into two samples gives

$$\widetilde{\mathbf{Z}}_{H,N} = (\widetilde{\mathbf{Z}}_{H,s} : \widetilde{\mathbf{Z}}_{H,N-s}), \tag{3.6}$$

where $\widetilde{\mathbf{Z}}_{H,s}$ and $\widetilde{\mathbf{Z}}_{H,N-s}$ are $KH \times s$ and $KH \times (N-s)$, respectively. The first-step estimator of the eigenvectors, $\widehat{\mathbf{P}}_s$, is obtained by estimating $\widehat{\mathbf{\Sigma}}_H$ from s observations

$$\widehat{\mathbf{\Sigma}}_{H,s} = \widehat{\mathbf{P}}_s\widehat{\mathbf{\Lambda}}_s\widehat{\mathbf{P}}'_s. \tag{3.7}$$

The first-step estimator can then be used to orthogonalize the remaining $N - s$ observations in order to derive an estimate of the eigenvalues

$$\widetilde{\mathbf{\Lambda}}_s = diag(\widehat{\mathbf{P}}_s\widehat{\mathbf{\Sigma}}_{H,N-s}\widehat{\mathbf{P}}'_s). \tag{3.8}$$

Now the IID assumption for $\widetilde{\mathbf{Z}}_{H,N}$ implies that it is efficient to use $N - s$ observations for this second step. Furthermore, Abadir et al. (2014) show that reusing any of the s observations worsens the estimate of $\widetilde{\mathbf{\Lambda}}_s$. This new estimate replaces $\widehat{\mathbf{\Lambda}}$ in equation (3.4). Multiplying the covariance matrix by $\mathbf{\Psi}_H$ in equation (3.2) undoes the error transformation and gives the

[8] Abadir et al. (2014) considers alternative distributions. However, only the Gaussian case is required here.

"design-free" method for estimating the GMFESM,

$$\widetilde{\mathbf{\Phi}}_{H,s} = \mathbf{\Psi}'_{p,H}\widehat{\mathbf{P}}\widetilde{\mathbf{\Lambda}}_s\widehat{\mathbf{P}}'\mathbf{\Psi}_{p,H} = \mathbf{\Psi}'_{p,H}\widehat{\mathbf{P}}diag(\widehat{\mathbf{P}}'_s\widehat{\mathbf{\Sigma}}_{H,N-s}\widehat{\mathbf{P}}_s)\widehat{\mathbf{P}}'\mathbf{\Psi}_{p,H}, \tag{3.9}$$

which is positive definite and non-singular as long as $N > s$ even when $KH > N$.[9]

There are many possible ways to choose subsamples of size s from N. Given the IID assumption, the "design-free" method can be generalized by averaging over R subsample re-samples of s to "reduce the variability that comes with the choice of any one specific combination of s observations" (p. 167). As a result, the generalized "design-free" method is

$$\widetilde{\mathbf{\Phi}}_{H,s,R} = \mathbf{\Psi}'_{p,H}\widehat{\mathbf{P}}\left(\frac{1}{R}\sum_{r=1}^{R}\widetilde{\mathbf{\Lambda}}_{s,r}\right)\widehat{\mathbf{P}}'\mathbf{\Psi}_{p,H}. \tag{3.10}$$

The general "design-free" approach adds two additional parameters to the estimation of the GMFESM: the size of the first-step estimator, s, and the number of subsample re-samples, R. First, consider how to choose the optimal number of subsample re-samples. In theory, R can be as large as $\binom{N}{s}$. However, given the IID assumption, each of these $\binom{N}{s}$ possibilities are equivalent so that the choice of R is of minor importance.

Choosing the optimal size of the first-step estimator is more complicated. As Abadir et al. (2014, p. 169) note, there is a trade off between the estimation of $\widehat{\mathbf{P}}_s$, which improves with large s, and the estimation of $\widehat{\mathbf{\Sigma}}_{H,N-s}$, which improves with a small s. For example, assume that $\{\mathbf{v}_T\}$ has the same known variance for all K variables, and observations, $H + N$, but is uncorrelated across variables so that $\widehat{\mathbf{\Omega}}_{1+N} = ... = \widehat{\mathbf{\Omega}}_{H+N} = \mathbf{\Omega} = \sigma^2\boldsymbol{I}_K$. Then, the covariance matrix of the transformed stacked forecast errors is a scalar matrix

$$\mathbf{\Sigma}_H = \sigma^2\boldsymbol{I}_{KH}, \tag{3.11}$$

where σ^2 is the same across variables and horizons. Abadir et al. (2014, p. 167) argue that in this case, the precision of the estimation of $\widetilde{\mathbf{\Phi}}_{H,s,R}$ is "invariant to [the eigenvectors], and the optimal choice of s is as small as possible to increase the precision of the eigenvalues estimated in the second-step". Relaxing the assumption on variance invalidates this result. However, the special case provides insight into the choice of the first-step estimator and also illustrates why alternative approaches such as zero restrictions or shrinkage are effective.

In practice, there is considerable variation in the optimal choice of s. Abadir et al. (2014) argue that "choosing small s delivers the largest improvements in the conditioning of the estimated variance matrix" (p. 174) but also suggest that choosing s to be around $0.5N$ provides the best estimate of the covariance matrix (p. 177). This supports the choice of a range of s to evaluate the stability of the performance of the "design-free" method.

[9] Since $\widetilde{\mathbf{Z}}_{H,s}$ is de-meaned, then the GMFESM will be non-singular as long as $N - 1 > s$. Note that in misspecified models, the squared bias term needs to be added to equation (3.9).

By choosing the optimal first-step estimator to orthogonalize the unused observations, the approach proposed in Abadir et al. (2014, p. 166) reformulates a multivariate problem from one of "ill-conditioning and imprecision to a univariate problem" in each diagonal element in equation (3.5). While this approach generates well-defined eigenvalues, and therefore positive definite matrices, it is important to note that there is no additional information beyond N.

Once $KH > N$, then the estimated eigenvalues beyond N tend to zero. The additional eigenvectors beyond N are then arbitrary and can be normalized to any value. Thus, while positive definite estimates of the GMFESM are generated, these estimates, and therefore the GFESM, may deviate substantially from the true underlying values. However, in principle as long as the normalization of the eigenvectors is applied uniformly across forecast models then it is possible that this approach would still generate invariant rankings of the forecasts. Furthermore, averaging across the subsample re-samples may help reduce substantial deviations.

4 Simulations

In this section, Monte Carlo experiments are run to study the small-sample properties of various estimators of the GFESM across different horizons, variables and DGPs. The simulation design is as follows: the DGP is a first-order vector-autoregressive process (VAR(1)):

$$\mathbf{X}_t = \mathbf{\Pi}\mathbf{X}_{t-1} + \mathbf{v}_t, \text{ where } \mathbf{v}_t \sim IN_K\left[\mathbf{0}, \mathbf{\Omega}\right], \ \mathbf{\Omega} = \mathbf{I}_K \text{ and } \mathbf{\Pi} = 0.5 * \mathbf{I}_K. \tag{4.1}$$

where $\mathbf{X}_t = (\mathbf{Y}_t - \mathbf{\Theta})$ is a vector of K variables that are demeaned by a vector of their long-run means. In the baseline case, the system is composed of two variables ($K = 2$) forecast across four horizons ($H = 4$). The known parameters forecast model is considered so that $\mathbf{\Pi}$ is known, there is no bias and there is no parameter estimation uncertainty. This implies that the alternative methods for transforming the stacked forecast errors are equivalent, and so any serial correlation is eliminated using $\mathbf{\Psi}_{1,H}^{-1}$ (see equation (3.1)).

The simulations explore the performance of various method of computing the GMFESM for a range of forecast-error observations, including where the standard approach is singular. Four separate approaches are considered: the "standard", "constrained", "tapered", and the "design-free" methods. By doing so, this analysis assesses the performance of a range of methods as they relate to the GFESM. See Abadir et al. (2014) for comparisons of the "design-free" method against an alternative range of approaches in a different context.

The "standard" approach was originally proposed by Clements and Hendry (1993a). Adapting equation (2.6) to illustrate the transformed stacked forecast errors from equation (3.1), the "standard" approach is computed as

$$\left|\widehat{\mathbf{\Phi}}_H\right|^{1/H} = \left|\frac{1}{N}\mathbf{\Psi}_{1,H}\widetilde{\mathbf{Z}}_{H,N}\widetilde{\mathbf{Z}}'_{H,N}\mathbf{\Psi}'_{1,H}\right|^{1/H} = \left|\frac{1}{N}\widetilde{\mathbf{Z}}_{H,N}\widetilde{\mathbf{Z}}'_{H,N}\right|^{1/H}, \tag{4.2}$$

where $\widetilde{\mathbf{W}}_{H,N|T} = \Psi_{1,H}\widetilde{\mathbf{Z}}_{H,N}$ is estimated from equation (2.5) and the forecast errors, $\widetilde{\mathbf{u}}_{T+h+n|T+n}$ are generated from the known parameters model (see equation (2.3) when $p = 1$ and $\widehat{\mathbf{\Pi}} = \mathbf{\Pi}$). The second equality follows from the fact that the determinant of the coefficient matrix is unity, so that the "standard" approach is invariant to the use of the transformed or non-transformed stacked forecast errors. Given that the baseline system consists of two variables and four forecast horizons ($KH = 8$), then the minimum number of forecast-error observations, N, for the "standard" approach to be well-conditioned is 8.

Next, the "design-free" method is considered. This approach computes the GFESM as

$$\left|\widetilde{\mathbf{\Phi}}_{H,s,R}\right|^{1/H} = \left|\widehat{\mathbf{P}}\left(\frac{1}{R}\sum_{r=1}^{R}\widetilde{\mathbf{\Lambda}}_{s,r}\right)\widehat{\mathbf{P}}'\right|^{1/H}, \tag{4.3}$$

where the "standard" approach, $\widehat{\mathbf{\Phi}}_H$, is replaced by the general "design-free" method from equation (3.10). Unlike the "standard" approach in equation (4.2), the "design-free" method is not invariant to transformations of the forecast errors. The "design-free" method requires additional choices in terms of R and s. Setting $R = 20$ is sufficient to reap most benefits.[10] In terms of s, Abadir et al. (2014, p. 173) find "robust performance as s varies around its optimal value, more specifically around approximately $s \in [0.2N, 0.8N]$". Thus, a range is used $s = \{s_1, s_2, s_3\} = \{0.2N, 0.5N, 0.8N\}$ where the resulting number is then rounded to the nearest integer less than $N - 1$. For example, when $N = 20$ then $s_1 = 4$, $s_2 = 10$, $s_3 = 16$. For simplicity, s_1 is referred to as "Min", s_2 is referred to as the "Mid", and s_3 is referred to as "Max".[11] These three values give a range for how the "design-free" method varies with s.

The "constrained" and "tapered" methods are considered for comparison. They are constructed following Hoffman and Rasche (1996) and Jentsch and Politis (2015) respectively. The "constrained" method is estimated exactly as in equation (4.2) except that restrictions on the GMFESM are imposed such that all of the cross-variable covariances are set to zero prior to taking the determinant. When these restrictions are valid the "constrained" method is invariant to transforming the stacked forecast errors. Since the baseline simulation experiment assumes $\mathbf{\Omega} = \mathbf{I}_K$, then the imposed restrictions are valid. This may skew the results in favour of the "constrained" method. The minimum N required for the "constrained" method to be non-singular is 4, which is equal to the longest forecast horizon.

The "tapered" method down-weights elements of the covariance matrix as a function of distance from the main diagonal. The maximum between the estimated eigenvalues and some

[10] See Abadir et al. (2014, p. 169). Simulations show that there is a trade-off in that there are marginal benefits to increasing R when the estimator is singular, but that with higher R performance declines as N increases.

[11] Abadir et al. (2014) also propose a 'grand average' method which averages over the different subsamples. However, it is excluded here since its performance generally lies between the subsamples considered. Their Ox code is available at: `http://bit.ly/1cqMpUR`.

Table 4.1: Largest to Smallest Eigenvalues by Method, when $KH = 8$ and $N = 6$

Eigenvalue	"true"	"standard"	"const."	"tapered"	"design-free"		
					"Min"	"Mid"	"Max"
1	2.84	4.82	3.95	3.31	2.50	1.97	1.60
2	2.23	2.24	2.07	1.67	1.78	1.51	1.11
3	1.33	1.27	1.29	1.08	1.06	0.88	0.67
4	1.04	0.75	0.89	0.77	0.79	0.66	0.47
5	0.76	0.39	0.61	0.56	0.54	0.48	0.34
6	0.61	0.13	0.41	0.34	0.41	0.38	0.25
7	0.54	0.00	0.26	0.19	0.31	0.29	0.18
8	0.43	0.00	0.13	0.11	0.23	0.20	0.11
GFESM	1.21	0.00	1.64	0.33	1.09	0.35	0.02

Note: Estimates are averaged across 1,000 replications

convergence factor is used to ensure positive definiteness. Here any element more than two positions away from the main diagonal are set equal to zero and the convergence factor is specified as $\frac{1}{N}$.[12] See Jentsch and Politis (2015, pp. 7-9). Similar to the "design-free" approach, the "tapered" method is not invariant to transformations of the forecast errors.

In order to ascertain how close the methods get to the actual GFESM, it is necessary to have a baseline against which to compare. In a Monte Carlo setting, the DGP is known, so it is also possible to estimate the "true" GFESM. From equation (2.7) when $p = 1$, the "true" GFESM for the known parameters forecast model is

$$|\mathbf{\Phi}_H|^{1/H} = \left| \begin{array}{ccc} \widehat{\mathbf{\Omega}}_{1+N} & \cdots & \widehat{\mathbf{\Omega}}_{1+N}\mathbf{\Pi}'^{H-1} \\ \vdots & \ddots & \vdots \\ \mathbf{\Pi}^{H-1}\widehat{\mathbf{\Omega}}_{1+N} & \cdots & \sum_{i=0}^{H-1}\mathbf{\Pi}^{i}\widehat{\mathbf{\Omega}}_{H+N-1-i}\mathbf{\Pi}'^{i} \end{array} \right|^{1/H}, \tag{4.4}$$

where $\widehat{\mathbf{\Omega}}$ is estimated from the correctly specified model such that

$$\widehat{\mathbf{\Omega}}_{N+h} = \frac{1}{T+N+h}\sum_{i=0}^{T+N+h-1}\mathbf{v}_i\mathbf{v}_i'. \tag{4.5}$$

The "true" GFESM is used as the baseline against which the various methods are compared. The analysis first examines how different methods estimate eigenvalues of the GMFESM and then the GFESM itself. Simulations are run with 1,000 replications.[13]

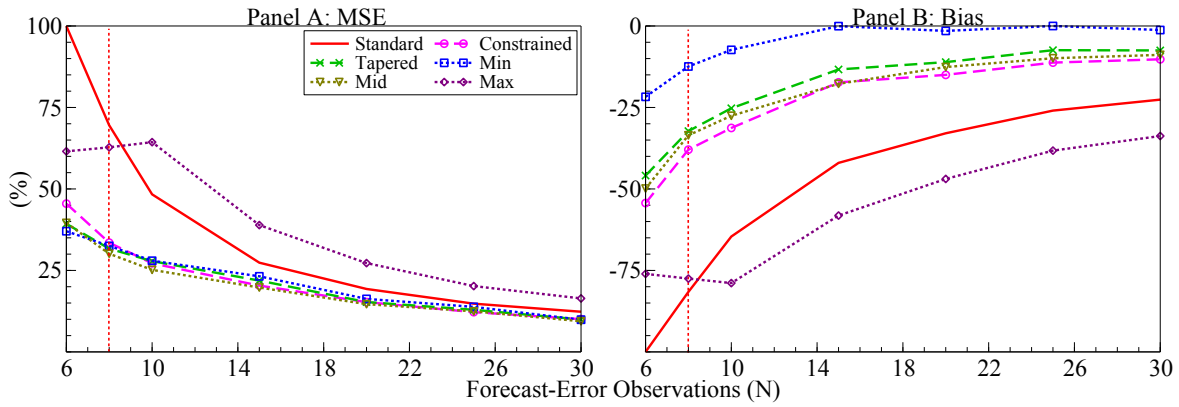
4.1 Eigenvalues

An initial view of the relative performance of the various methods is given by examining how well they estimate the eigenvalues of the GMFESM. Since the GFESM is the product of the eigenvalues, these estimates are important. As Table 4.1 illustrates, when $KH > N$ the "standard" approach overstates the largest eigenvalues while substantially under reporting the smallest eigenvalues. In fact, the "standard" approach goes to zero because it severely under estimates any eigenvalues greater than the number of forecast-error observations.

[12] This implies that the user defined constants following Jentsch and Politis (2015) notation are all set equal to $\epsilon = \ell = \beta = 1$. Their R code is available at: http://www.math.ucsd.edu/~politis/SOFT/function_MLPB.R.

[13] All numerical results were obtained using OxMetrics Version 7.10; see Doornik (2013); Doornik and Hendry (2013a).

Figure 4.1: Comparing estimator performance for the known parameters model



Each of the other methods attempts to smooth this dynamic by reducing the over/under reporting of the largest/smallest eigenvalues. The smoothing done by the "constrained" method is the least disruptive when its restrictions are valid. As a result, the net changes to the eigenvalues relative to the "standard" approach balance out such that the reduction in over reporting of the largest eigenvalue is exactly offset by a reduction in under reporting of the smallest eigenvalues.

The "tapered" and "design-free" methods do not impose constraints and so the resulting changes are less efficient. These methods require larger reductions in the largest eigenvalues to obtain similar increases in the smallest eigenvalues. This is especially clear for the "design-free" method where all of the eigenvalues are under reported but where "Min" and "Mid" under report the smallest eigenvalues the least. This has implications for the GFESM. While the GFESM for each of the methods is greater than zero, "Min" is the closest to the "true" GFESM.

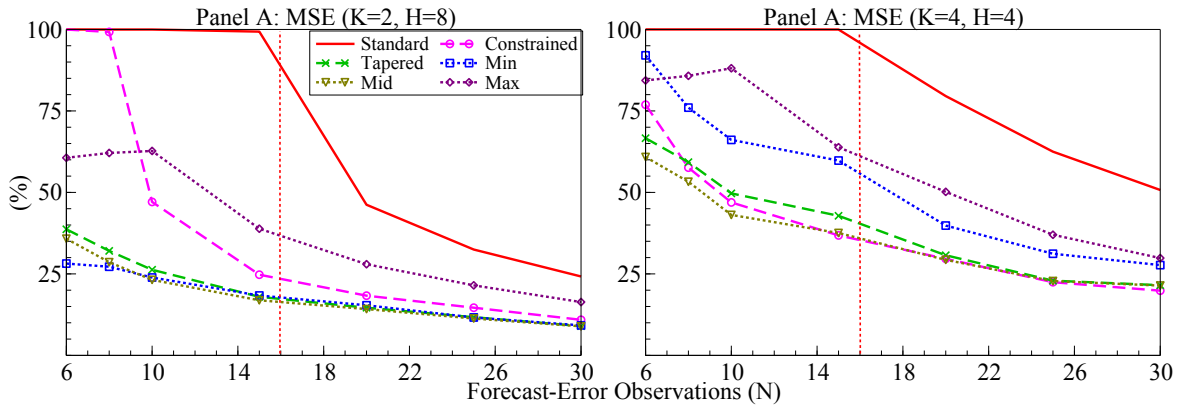
4.2 The GFESM

Now consider the impact of the different methods on estimates of the GFESM. Here bias and the relative mean square error (MSE) are used as judges of performance. Bias is calculated as the percentage difference between the estimated and the "true" GFESM. Relative MSE is calculated as the ratio of the mean square error of the estimated GFESM (compared against "true") to the mean square error of the singular GFESM (compared against "true"). This standardizes the MSE across methods. The results are presented in percent where 100 bias (−) or MSE (+) indicates that the estimate is no better than the singular case. Values less than $|\pm 100|$ represent an improvement relative to the singular case.

For the "standard" approach in Figure 4.1, the GMFESM is initially non-singular until $N = 8$ when the MSE and bias start to decrease. Put slightly differently, the "standard" approach deteriorates as N decreases relative to KH. Furthermore, while the "standard" approach is consistent for large N and fixed KH, it is biased in finite samples.

These results are of considerable interest given the relevance to previous applications of the

Figure 4.2: Comparing method performance for the known parameters model



GFESM. While most studies shy away from the GFESM when $KH > N$, several studies including Kunst and Franses (1998) and Gooijer and Ray (2003) come right up to this boundary using the "standard" approach. If these results are generalizable, then estimates of the GFESM using the "standard" approach at the borderline case are downward biased.

With the exception of "Max", each of the other methods improves upon the "standard" approach. Figure 4.1 shows that these methods remain non-singular when $KH > N$. In fact, they have lower biases and MSEs than the "standard" approach across all N. The largest gains are when $KH > N$ where the MSE and bias are reduced by 25 to 70 percent. While the methods cluster together, "Min" performs the best. This can be explained by the fact that this case is the same as equation (3.11) when $\sigma^2 = 1$. Panel B in Figure 4.1 illustrates that each of the methods underestimates the GFESM but that this bias decreases as N increases. The results are similar when the log GFESM is used instead of the standardized GFESM.

Figure 4.2 illustrates that the impact of an increase in K is not equivalent to the impact of an increase in H. In both panels, the minimum required number of forecast-error observations is 16. However, comparing Panel A in Figure 4.2 with Panel A in Figure 4.1 it is clear that despite a doubling of H and a decline in the relative size of N, the "design-free" and "tapered" methods perform equally well. On the other hand, the "standard" and "constrained" approaches perform worse as the forecast horizon increases. The "standard" approach now requires 16 forecast-error observations to be well-conditioned (compared to 8 before), whereas the "constrained" approach requires 8 forecast-error observations (compared to 4).

Comparing Panel B in Figure 4.2 with Panel A in Figure 4.1 illustrates that a doubling of K affects most methods similarly. The MSE for all methods relative to the baseline case in Figure 4.1 increases initially by 20 percent and remains elevated across all N. However, an increase in K also affects the dynamics such that "Min" does not consistently have the lowest MSE. Instead, "Mid" and the "constrained" methods tend to do best.

5 Comparing Estimator Performance Across Forecast Models

The previous section demonstrated the relative performance of the different methods for the known parameters model. However, it is unclear if these results hold for alternative forecast models. Furthermore, since the previous section only focused on one model, nothing could be said about how the performance of various methods relates to their ability to capture the actual rankings. This section extends the simulation results to alternative forecast models including the unknown parameters (VAR) and the random walk (RW) model. While the VAR, or unknown parameters model, does well in the case when there are no structural breaks, the RW is often used as a baseline and is more robust when there are structural breaks, albeit at short horizons.

In order to compare estimates of the GFESM across alternative models, the target values, which differ across the models, need to be computed. These "true" GFESM's are described before presenting the simulation results. This section focuses on models that are representative of two different cases: unknown, and fixed parameters.

Unknown parameters

The true parameters are unknown in the VAR model. Thus, the parameters (including a constant term) are estimated following each of the $T+n$ observations. Building on Schmidt (1974), Clements and Hendry (1998, p. 75) show that when the true parameters are unknown, the forecast error variance for the h-step-ahead forecast becomes

$$\mathbb{E}_T\left[\widehat{\mathbf{u}}_{T+h|T}\widehat{\mathbf{u}}'_{T+h|T}\right] = \sum_{i=0}^{h-1}\mathbf{\Pi}^i\mathbf{\Omega}\mathbf{\Pi}^{i'} + (\boldsymbol{I}_N \otimes \mathbf{X}'_T)\mathbf{Q}(h)'\mathbb{V}[\widehat{\mathbf{\Pi}}^v]\mathbf{Q}(h)(\boldsymbol{I}_N \otimes \mathbf{X}_T), \tag{5.1}$$

where $\mathbf{Q}(h)$ is the derivative of $(\mathbf{\Pi}^h)^v$ with respect to $\mathbf{\Pi}^v$ when $(.)^v$ denotes vectoring. When the system is mapped to an I(0) representation and there are no unrestricted variables, Doornik and Hendry (2013b) argue that parameter estimation uncertainty can be expressed as

$$\left(\sum_{i=0}^{h-1}\widehat{\mathbf{\Pi}}_1^i \otimes \widehat{\mathbf{b}}'_{T+h-i}\right)\mathbb{V}\left[\widehat{\mathbf{\Pi}}'^v\right]\left(\sum_{i=0}^{h-1}\widehat{\mathbf{\Pi}}_1^i \otimes \widehat{\mathbf{b}}'_{T+h-i}\right)', \tag{5.2}$$

where $\mathbf{b}'_{T+h-i} = \left(\mathbf{X}'_{T+h-i,h-1-i} : \mathbf{Z}'_{T+h-i}\right)$, $\mathbf{Z}$ denotes the vector of additional variables included in the model and $\widehat{\mathbf{\Pi}}_1$ denotes the estimated parameters for $\mathbf{X}$. Equation (5.2) is added to each term in the baseline GMFESM with modifications accounting for the covariances. When parameters are re-estimated after each additional observation, then if $\mathbf{A}_{h,n} = \left(\sum_{i=0}^{h-1}\widehat{\mathbf{\Pi}}_1^i \otimes \widehat{\mathbf{b}}'_{T+n+h-i}\right)$ and $\widetilde{\mathbf{V}}_N = \mathbb{V}\left[\widehat{\mathbf{\Pi}}'^v\right]$, the "true" GMFESM augments equation (4.4) by

$$\begin{pmatrix} \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{A}_{1,n}\widetilde{\mathbf{V}}_n\mathbf{A}_{1,n} & \cdots & \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{A}_{1,n}\widetilde{\mathbf{V}}_n\mathbf{A}_{H,n} \\ \vdots & \ddots & \vdots \\ \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{A}_{H,n}\widetilde{\mathbf{V}}_n\mathbf{A}_{1,n} & \cdots & \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{A}_{H,n}\widetilde{\mathbf{V}}_n\mathbf{A}_{H,n} \end{pmatrix}. \tag{5.3}$$

Fixed parameters

The RW model fixes the parameters. Thus, while there is no parameter estimation uncertainty, the estimate of the "true" GFESM needs to account for the difference between the actual and fixed parameters. The RW model does not change the VAR(1) other than to fix the parameter values, denoted as $\bar{\mathbf{\Pi}}$. When $\bar{\mathbf{\Pi}} = \boldsymbol{I}_K$, then the forecast error variance is

$$\mathbb{E}_T\left[\widehat{\mathbf{u}}_{T+h|T}\widehat{\mathbf{u}}'_{T+h|T}|\mathbf{X}_T\right] = \left(\boldsymbol{I}_K - \mathbf{\Pi}^h\right)\mathbf{X}_T\mathbf{X}'_T\left(\boldsymbol{I}_K - \mathbf{\Pi}^h\right)' + \sum_{i=0}^{h-1}\mathbf{\Pi}^i\mathbf{\Omega}\mathbf{\Pi}^{i'}. \tag{5.4}$$

Rather than taking into account the parameter uncertainty, this modified equation takes into account the difference between the values of the true and fixed parameters. As persistence increases towards a unit root ($\mathbf{\Pi} \to \boldsymbol{I}_K$), then the first term goes to zero and the forecast error decreases towards the known parameters model. Thus, if $\mathbf{J}_n = \mathbf{X}_{T+n}\mathbf{X}'_{T+n}$ and $\mathbf{F}_h = \left(\boldsymbol{I}_K - \mathbf{\Pi}^h\right)$ then the "true" GMFESM for the RW model augments equation (4.4) with

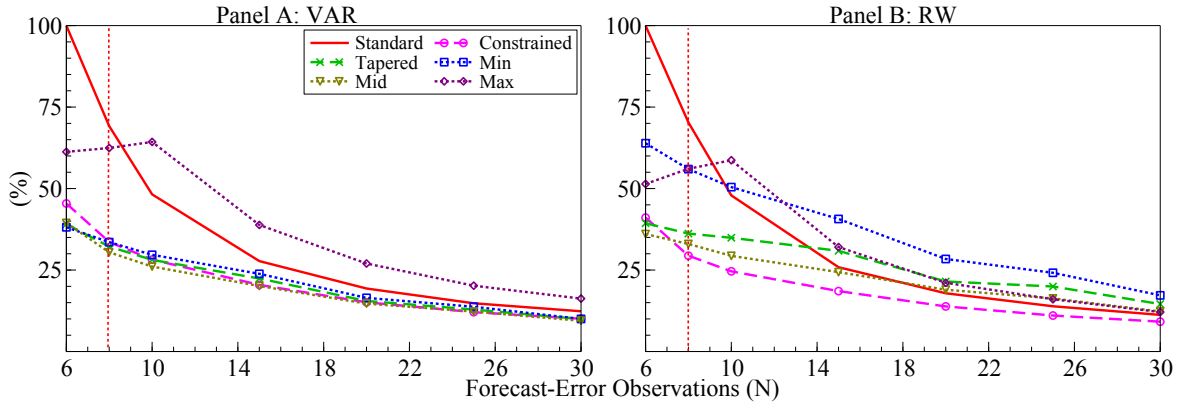
$$\begin{pmatrix} \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{F}_1\mathbf{J}_n\mathbf{F}_1 & \cdots & \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{F}_1\mathbf{J}_n\mathbf{F}_H \\ \vdots & \ddots & \vdots \\ \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{F}_H\mathbf{J}_n\mathbf{F}_1 & \cdots & \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{F}_H\mathbf{J}_n\mathbf{F}_H \end{pmatrix}. \tag{5.5}$$

Equations (5.3) and (5.5) augment (4.4) in different ways to give the "true" GMFESM for the respective forecast models, the determinants of which are the "true" GFESMs used as baselines to compare the various estimators. As in the known parameters case, the error variance in each of these cases is estimated from the correctly specified model in equation (4.5).

Estimates of the GFESM for the various models are generated using the forecast errors from equation (2.3) when $p = 1$, for the known, VAR and RW models with modifications for how $\widehat{\mathbf{\Pi}}$ is treated and whether a constant is estimated. The forecast errors for each model are stacked into matrices following equation (2.5) and then transformed by the parameter matrix in equation (3.1). The GFESM is then estimated by the methods described in the previous section using the respective transformed stacked forecast-error matrices for each model.

Equations (5.3) and (5.5) illustrate that when forecast models are misspecified, it is no longer straightforward to decompose the GMFESM as in equation (3.2). As a result, transforming the forecast errors may not remove all serial correlation from the forecast errors of the misspecified models. An IID test on the transformed forecast errors is useful in this case and also serves as a preliminary way to test for predictive failure. Assuming that the correct transformation was used, then under the null hypothesis of IID forecast errors the forecasts are well specified. Furthermore, since the forecast errors are no longer unbiased, the moment and covariance matrices are no longer equivalent. Thus, the squared bias term needs to be added back into the estimates of the GMFESM before taking the determinant.

Figure 5.1: Method performance across models using the MSE



5.1 Simulation Results

Figure 5.1 presents the simulation results for the alternative forecast models. Panel A presents the results for the VAR while panel B presents those for the RW. They illustrate that the various methods perform differently across models. The results for the VAR in Panel A of Figure 5.1 follow the known parameters model in Panel A of Figure 4.1 very closely. This stems from the fact that the VAR is well specified in this case. It also suggests that the parameter estimation uncertainty term from equation (5.3) does not substantially impact estimates of the GFESM. Note that, differences may be larger when the estimation period (T) is limited.

The results for the RW model in Panel B of Figure 5.1 illustrates that the previous results do not completely extend to misspecified models. In this case, the forecasts are misspecified such that the forecast error increases with the forecast horizon. Thus, heteroskedasticity increases along the diagonal of the GMFESM (see equation (5.5)) which means that the transformed stacked forecast errors are not IID.

In terms of the "design-free" method, a homogeneous estimate of the eigenvectors (i.e. smaller s) does poorly and so increasing the size of the first-step estimator (i.e. higher s) improves estimates, which is why "Max" sees relative improvements. The "tapered" method also fares worse which is likely because it downplays the covariance of the forecast errors at longer horizons. On the other hand, the "constrained" method performs well since there is no cross-variable or cross-residual correlation in the DGP. When cross-variable/residual correlations increase, then the "constrained" method no longer has the lowest MSE. Simulations for additional misspecified models suggest increasing differentiation across methods. These results continue to hold when structural breaks are included.[14]

[14] Additional simulation results are available upon request.

5.2 Ranking the Forecast Models

So far the analysis has focused on how well the GFESM can be estimated. However, it is also important that estimates of the GFESM provide a valid ranking of forecast models. In fact, it is possible that even when estimates are poor they can still provide informative rankings of the forecasts. The remaining analysis in this section focuses on how the various methods rank the forecasts. The DGP evaluated here contains 2 variables and 4 forecast horizons. Unlike the previous results, this analysis focuses on the actual estimates of the GFESM. The average of the RMSFE's across all variables and horizons (ATRMSFE) is calculated for comparison.

Table 5.1 presents estimates of the standardized GFESM across different sizes of N. For simplicity only three cases are considered: the singular case ($N = 6$), the borderline case ($N = 8$), and the non-singular case ($N = 20$). The rows present the various methods and the minor columns represent different forecast models. In addition to the known parameters, VAR and RW models, the differenced known parameters model is also included to illustrate that the invariance of the GFESM does not extend to the MSFE. The "true" GFESM ranking across all three cases from best to worst forecast model is: known parameters, differenced known parameters, VAR and RW. The known parameters model does best followed closely by the correctly specified VAR whereas the robustified models perform noticeably worse when there is no break. The ATRMSFE gives the same ranking as the "true" GFESM except that the ATRMSFE's are different for the known and Δ known parameters models. This difference is more pronounced when $\mathbf{\Pi} \neq 0.5$ and can help indicate the degree of serial correlation in the forecast errors.[15]

Now consider the rankings generated by the various GFESM estimators. In the first major column of Table 5.1, since $KH > N$, the "standard" approach is zero and therefore unable to distinguish between models. However, the "design-free" method generates a correct ranking. The "constrained" and "tapered" methods also generate correct rankings.

In the borderline case, the second major column of Table 5.1, each of the methods correctly ranks the forecasts. Now the "standard" approach also correctly ranks the forecast models. These results continue to hold as N increases (third major column). The results presented here are for the standardized GFESM. The specifics vary slightly for the log GFESM and the unstandardized GFESM. However, in general the results suggest that estimation biases enhance rather than distort the forecast rankings, Furthermore, when N is relatively small, a more important concern is the ability to distinguish between well-specified models. This is particularly true when there is a structural break at the end of the estimation sample (not shown), then it is hard to distinguish between a VAR model with an intercept correction term and an RW model.

[15] The known parameters model has a smaller ATRMSFE when $\mathbf{\Pi} < 0.5$ and a larger ATRMSFE when $\mathbf{\Pi} > 0.5$.

Table 5.1: Ranking forecast models using the GFESM

Method	$KH > N: N = 6$				$KH = N: N = 8$				$KH < N: N = 20$			
	Known	Δ Known	VAR	RW	Known	Δ Known	VAR	RW	Known	Δ Known	VAR	RW
"true"	0.99	0.99	1.03	1.44	0.99	0.99	1.03	1.47	0.99	0.99	1.03	1.49
"standard"	0.00	0.00	0.00	0.00	0.18	0.18	0.19	0.26	0.66	0.66	0.69	1.00
"Min"	0.78	0.78	0.80	1.57	0.87	0.87	0.90	1.68	0.98	0.98	1.02	1.77
"Mid"	0.50	0.50	0.52	1.01	0.66	0.66	0.69	1.27	0.87	0.87	0.90	1.56
"Max"	0.24	0.24	0.25	0.47	0.22	0.22	0.23	0.40	0.53	0.53	0.55	0.92
"constrained"	0.54	0.54	0.55	1.06	0.67	0.67	0.70	1.27	0.88	0.88	0.92	1.57
"tapered"	0.45	0.45	0.47	0.76	0.61	0.61	0.64	1.00	0.84	0.84	0.87	1.30
ATRMSFE	1.04	1.04	1.05	1.33	1.06	1.07	1.08	1.37	1.08	1.09	1.10	1.39

Notes:
1. Estimates are based on 1,000 replications.
2. ATRMSFE is calculated as the sum of the square roots of the diagonal of the GMFESM.

6 Evaluating a vector of forecasts for the US economy

The section applies the methods proposed by evaluating a vector of forecasts of the US economy generated by a vector equilibrium correction model (VEqCM) and a RW model. The data used and the forecast models are described in Anderson et al. (2002). Forecasts were generated for 6 variables up to 16 quarters ahead from 1989 through 1997. Thus 96 forecast-error observations are required to evaluate the forecasts using the "standard" approach, but a maximum of 36 forecast-error observations are available. Anderson et al. (2002) use the "constrained" method to compute reduction in the VEqCM's GFESM relative to the RW model.

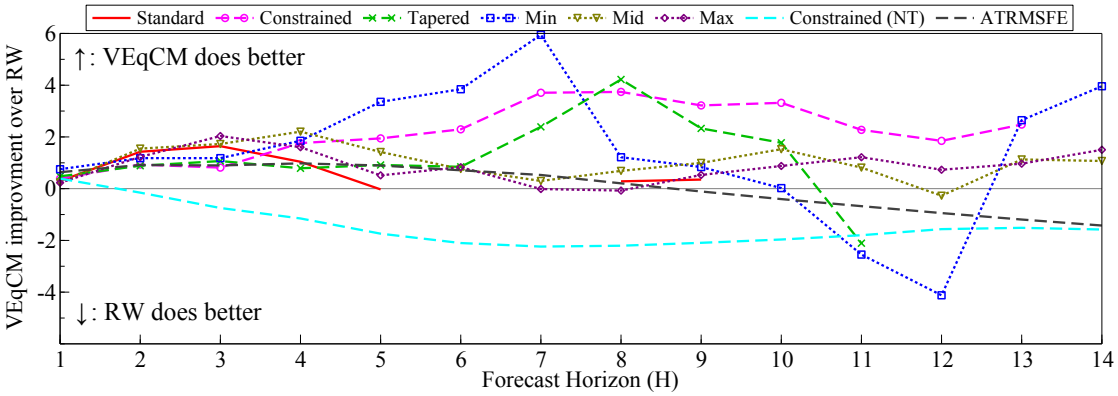
While this analysis estimated a VEqCM and generated forecasts that exhibit the same general features, the original GFESM results could not be replicated (see Anderson et al. (2002, Table 6)).[16] Using the methods described above, this analysis computes the improvement of the VEqCM over the RW in terms of the log GFESM and the ATRMSFE for $1-14$ quarters ahead forecasts. This allows for comparisons across various methods.

Figure 6.1 presents the results from this analysis. Although the "standard" approach is only available through 5 horizons, most of the alternative methods can be estimated through 14 horizons. The results illustrate that while each of the methods are initially quite close, the "Mid" and "Max" methods appear to follow the standard approach most closely. The volatility of the "Min" method suggests that both models are misspecified and therefor more observations are necessary to estimate the eigenvectors.

Most of the GFESM methods suggest that the VEqCM dominates the RW across all 14 horizons, albeit to varying degrees and with notable exceptions between $10-12$ quarters ahead. The "constrained" approach tends to underestimate the performance of the VEqCM when the forecast errors are not transformed. This suggests that the assumptions this method imposes are invalid. The ATRMSFE suggests that the VEqCM only dominates the RW through 8 quarters ahead, after which the RW dominates the VEqCM. This implies that the VEqCM forecasts benefit when accounting for the error covariance across variables and longer horizons.

[16] The full results are available upon request.

Figure 6.1: VEqCM forecast performance for the US economy, 1989-97



Note: Improvement in the GFESM methods is calculated as the difference between the standardized log GFESM's. Improvement in the ATRMSFE is calculated as the percentage difference between the ATRMSFE's scaled by 10. Constrained (NT) is the non-transformed version of the "constrained" approach.

7 Conclusions

This paper develops a new approach for evaluating multi-step system forecasts when there are relatively few forecast-error observations. In the process, it proposes a solution to one of the long-standing limitations of the GFESM. Section 3 shows how the approach developed in Abadir et al. (2014) can be generalized to estimate the moment matrix of the stacked forecast errors. Combining this method with Clements and Hendry (1993a) allows for precise estimation of the GFESM even when $KH > N$.

The simulations in sections 4 and 5 show that the "standard" approach is downwards biased and deteriorates rapidly when there are relatively few forecast observations. On the other hand, simulations show that the alternative methods can greatly reduce this bias in relatively well-specified forecast models. While there is limited differentiation between alternative methods in well-specified models, this changes for misspecified models. The "constrained" approach does consistently well in simulations since its restrictions are valid for the DGP designs considered, its performance deteriorates when the imposed restrictions are invalid or when the forecast horizon increases. Furthermore, while "Min" does better in the baseline situations, "Mid" appears more robust to different types of model misspecification. Despite these differences, each of the methods correctly rank the models.

The simulation results also extend to an application of the methods in evaluating forecasts of the US economy. In general the GFESM methods consistently rank a VEqCM over a RW across all horizons, which differs substantially from the results for the trace MSFE statistic. This illustrates the benefits of being able to extend estimates of the GFESM to cases when there are more variables times horizons than forecast-error observations.

Bibliography

Abadir, K. M., Distaso, W., and Žikeš, F. (2014). Design-free estimation of variance matrices. *Journal of Econometrics*, 181(2):165 – 180.

Abadir, K. M. and Magnus, J. R. (2005). *Matrix algebra*, volume 1. Cambridge University Press.

Anderson, R. G., Hoffman, D. L., and Rasche, R. H. (2002). A vector error-correction forecasting model of the US economy. *Journal of Macroeconomics*, 24(4):569 – 598.

Athanasopoulos, G., de Carvalho Guillén, O. T., Issler, J. V., and Vahid, F. (2011). Model selection, estimation and forecasting in VAR models with short-run and long-run restrictions. *Journal of Econometrics*, 164(1):116 – 129.

Baillie, R. T. and Morana, C. (2009). Modelling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach. *Journal of Economic Dynamics and Control*, 33(8):1577 – 1592.

Bowsher, C. G. and Meeks, R. (2008). The dynamics of economic functions: Modeling and forecasting the yield curve. *Journal of the American Statistical Association*, 103(484):1419–1437.

Bårdsen, G., den Reijer, A., Jonasson, P., and Nymoen, R. (2012). MOSES: Model for studying the economy of Sweden. *Economic Modelling*, 29(6):2566 – 2582.

Burgess, S., Fernandez-Corugedo, E., Groth, C., Harrison, R., Monti, F., Theodoridis, K., and Waldron, M. (2013). The Bank of England's forecasting platform: COMPASS, MAPS, EASE and the suite of models. *The Bank of England Working Paper Series*, (No. 471).

Castle, J. L. and Hendry, D. F. (2008). Forecasting UK inflation: The roles of structural breaks and time disaggregation. In Rapach, D. E. and Wohar, M. E., editors, *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, volume 3 of *Frontiers of Economics and Globalization*, chapter 2, pages 41–92. Bingley, UK: Emerald Group.

Clements, M. P. and Galvão, A. B. (2004). A comparison of tests of nonlinear cointegration with application to the predictability of US interest rates using the term structure. *International Journal of Forecasting*, 20(2):219 – 236.

Clements, M. P. and Hendry, D. F. (1993a). On the limitations of comparing mean square forecast errors. *Journal of Forecasting*, 12(8):617–637.

Clements, M. P. and Hendry, D. F. (1993b). On the limitations of comparing mean square forecast errors: A reply. *Journal of Forecasting*, 12(8):669–676.

Clements, M. P. and Hendry, D. F. (1995). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, 10(2):127–146.

Clements, M. P. and Hendry, D. F. (1997). An empirical study of seasonal unit roots in forecasting. *International Journal of Forecasting*, 13(3):341 – 355.

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

Doornik, J. A. (2013). *Ox 7: An Object-Oriented Matrix Programming Language*. Timberlake Consultants Ltd.

Doornik, J. A. and Hendry, D. F. (2013a). *Empirical Econometric Modelling*, volume I of *PcGive 14*. Timberlake Consultants Ltd.

Doornik, J. A. and Hendry, D. F. (2013b). *Modelling Dynamic Systems*, volume II of *PcGive 14*. Timberlake Consultants Ltd.

Engle, R. F. (1993). On the limitations of comparing mean square forecast errors: Comment. *Journal of Forecasting*, 12(8):642–644.

Gooijer, J. G. D. and Ray, B. K. (2003). Modeling vector nonlinear time series using POLYMARS. *Computational Statistics & Data Analysis*, 42(1–2):73 – 90.

Gooijer, J. G. D. and Vidiella-i-Anguera, A. (2004). Forecasting threshold cointegrated systems. *International Journal of Forecasting*, 20(2):237 – 253.

Granger, C. W. J. (1993). On the limitations of comparing mean square forecast errors: Comment. *Journal of Forecasting*, 12(8):651–652.

Hendry, D. F. and Mizon, G. E. (2014). Unpredictability in economic analysis, econometric modeling and forecasting. *Journal of Econometrics*, 182(1):186–195.

Hoffman, D. L. and Rasche, R. H. (1996). Assessing forecast performance in a cointegrated system. *Journal of Applied Econometrics*, 11(5):495–517.

Jentsch, C. and Politis, D. N. (2015). Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *The Annals of Statistics*, 43(3):1117–1140.

Kunst, R. M. and Franses, P. H. (1998). The impact of seasonal constants on forecasting seasonally cointegrated time series. *Journal of Forecasting*, 17(2):109–124.

McMurry, T. L. and Politis, D. N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis*, 31(6):471–482.

Paap, R., Franses, P. H., and Hoek, H. (1997). Mean shifts, unit roots and forecasting seasonal time series. *International Journal of Forecasting*, 13(3):357 – 368.

Pretis, F. and Roser, M. (2016). Decoupling projections from reality: Co2 emission-intesnity growth in climate scenarios. Department of Economics Working Paper, University of Oxford.

Schmidt, P. (1974). The asymptotic distribution of forecasts in the dynamic simulation of an econometric model. *Econometrica*, 42(2):pp. 303–309.

Schmidt, P. (1993). On the limitations of comparing mean square forecast errors: Comment. *Journal of Forecasting*, 12(8):660–662.

Simionescu, M. (2013). The use of VARMA models in forecasting macroeconomic indicators. *Economics & Sociology*, 6(2):94–102.

Sinclair, T. M., Stekler, H. O., and Carnow, W. (2012). A new approach for evaluating economic forecasts. *Economics Bulletin*, 32(3):2332–2342.

Sinclair, T. M., Stekler, H. O., and Carnow, W. (2015). Evaluating a vector of the Fed's forecasts. *International Journal of Forecasting*, 31(1):157 – 164.

Summers, L. H. (2014). Reflections on the 'new secular stagnation hypothesis'. *Secular Stagnation: Facts, Causes and Cures*, pages 27 – 38.

Vahid, F. and Issler, J. V. (2002). The importance of common cyclical features in VAR analysis: A Monte-Carlo study. *Journal of Econometrics*, 109(2):341 – 363.

Veloce, W. (2004). Forecasting inbound Canadian tourism: An evaluation of error corrections model forecasts. *Tourism Economics*, 10(3):263–280.

Zeng, T. and Swanson, N. R. (1998). Predictive evaluation of econometric forecasting models in commodity futures markets. *Studies in Nonlinear Dynamics & Econometrics*, 2(4):159–177.