Journal of Econometrics 186 (2015) 280–293

Contents lists available at ScienceDirect

Journal of Econometrics

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

Forecasting with factor-augmented regression: A frequentist model averaging approach

Xu Cheng [a], Bruce E. Hansen [b,*]

[a] University of Pennsylvania, United States
[b] University of Wisconsin, United States

ARTICLE INFO

Article history:
Available online 6 March 2015

JEL classification:
C52
C53

Keywords:
Cross-validation
Factor models
Forecast combination
Generated regressors
Mallows

ABSTRACT

This paper considers forecast combination with factor-augmented regression. In this framework, a large number of forecasting models are available, varying by the choice of factors and the number of lags. We investigate forecast combination across models using weights that minimize the Mallows and the leave-h-out cross validation criteria. The unobserved factor regressors are estimated by principle components of a large panel with N predictors over T periods. With these generated regressors, we show that the Mallows and leave-h-out cross validation criteria are asymptotically unbiased estimators of the one-step-ahead and multi-step-ahead mean squared forecast errors, respectively, provided that $N, T \to \infty$. (However, the paper does not establish any optimality properties for the methods.) In contrast to well-known results in the literature, this result suggests that the generated-regressor issue can be ignored for forecast combination, without restrictions on the relation between N and T.

Simulations show that the Mallows model averaging and leave-h-out cross-validation averaging methods yield lower mean squared forecast errors than alternative model selection and averaging methods such as AIC, BIC, cross validation, and Bayesian model averaging. We apply the proposed methods to the US macroeconomic data set in Stock and Watson (2012) and find that they compare favorably to many popular shrinkage-type forecasting methods.

1. Introduction

Factor-augmented regression has received much attention in high-dimensional problems where a large number of predictors are available over a long period. Assuming some latent factors generate the comovement of all predictors, one can forecast a particular series by the factors rather than by the original predictors, with the benefit of significant dimension reduction (Stock and Watson, 2002). In factor-augmented regression, the factors are determined and ordered by their importance in driving the covariability of many predictors, which may not be consistent with their forecast power for the particular series of interest, an issue discussed in Bai and Ng (2008, 2009). In consequence, model specification is necessary to determine which factors should be used in the forecast regression, in addition to specifying the number of lags of the dependent variable and the number of lags of the factors included. These decisions vary with the particular series of interest and the forecast horizon.

This paper proposes forecast combination based on frequentist model averaging criteria. The forecast combination is a weighted average of the predictions from a set of candidate models that vary by the choice of factors and the number of lags. The model averaging criteria are estimates of the mean squared forecast errors (MSFE). Hence, the weights that minimize these model averaging criteria are expected to minimize the MSFE. Two different types of model averaging methods are considered: the Mallows model averaging (MMA; Hansen, 2007) and the leave-h-out cross-validation averaging (CVA_h; Hansen, 2010). For one-step-ahead forecasting, the CVA_h method is equivalent to the jackknife model averaging (JMA) from Hansen and Racine (2012). The MMA and CVA_h methods were designed for standard regression models with observed regressors. However, dynamic factor models involve unobserved factors and their estimation creates generated regressors. The effect of generated regressors on model selection and combination has not previously been investigated. This paper makes this extension and provides a theoretical justification for frequentist model averaging methods in the presence of estimated factors.

We show that even in the presence of estimated factors, the Mallows and leave-h-out cross-validation criteria are asymp-

* Corresponding author.
E-mail address: behansen@wisc.edu (B.E. Hansen).

http://dx.doi.org/10.1016/j.jeconom.2015.02.010
0304-4076/

totically unbiased estimators of the one-step-ahead and multi-step-ahead MSFE, respectively, provided that $N, T \to \infty$. In consequence, these frequentist model averaging criteria can be applied to factor-augmented forecast combination without modification. Thus for model selection and combination, the generated-regressor issue can be safely ignored. This is in contrast to inference on the coefficients, where Pagan (1984), Bai and Ng (2009), Ludvigson and Ng (2011), and Gonçalves and Perron (2014) have shown that the generated regressors affect the sampling distribution. It is worth emphasizing that our result is not based on asymptotic rates of convergence (such as assuming $T^{1/2}/N \to 0$ as in Bai and Ng (2006)); instead it holds because the focus is on forecasting rather than parameter estimation. Indeed, in the context of a non-dynamic factor model (one without lagged dependent variables and no serial correlation) we show that the Mallows criterion is an unbiased estimate of the MSFE in finite samples, and retains the classic optimality developed in Li (1987), Andrews (1991), and Hansen (2007). In dynamic models our argument is asymptotic, and we do not establish any form of optimality, but our results do not rely on differing rates of convergence.

Our simulations demonstrate the superior finite-sample performance of the MMA and CVA_h forecasts in the sense of low MSFE. Our comparisons are quite thorough, comparing our procedures with AIC selection, BIC selection, Mallows selection, cross-validation selection, approximate Bayesian model averaging, equal weights, and the three-pass regression filter of Kelly and Pruitt (forthcoming). Our methods dominate the other procedures throughout the parameter space considered. These findings are consistent with the optimality of MMA and JMA in the absence of temporal dependence and generated regressors (Hansen, 2007; Hansen and Racine, 2012). In addition, the advantage of CVA_h is found most prominent in long-horizon forecasts with serially correlated forecast errors.

We apply the proposed methods to the US macroeconomic data set in Stock and Watson (2012) and find that they compare favorably to many popular shrinkage-type forecasting methods.

The frequentist model averaging approach adopted here extends the large literature on forecast combination, see Granger (1989), Clemen (1989), Diebold and Lopez (1996), Hendry and Clements (2002), Timmermann (2006), and Stock and Watson (2006), for reviews. Stock and Watson (1999, 2004, 2012) provide detailed empirical evidence demonstrating the gains of forecast combination. The simplest forecast combination is to use equal weights. Compared to simple model averaging, MMA and CVA_h are less sensitive to the choice of candidate models. Alternative frequentist forecast combination methods are proposed by Bates and Granger (1969), Granger and Ramanathan (1984), Timmermann (2006), Buckland et al. (1997), Burnham and Anderson (2002), Hjort and Claeskens (2003), Elliott et al. (2013), among others. Hansen (2008) shows that MMA has superior MSFE in one-step-ahead forecasts when compared to many other methods.

Another popular model averaging approach is the Bayesian model averaging (BMA; Min and Zellner, 1993). The BMA has been widely used in econometric applications, including Sala-i-Martin et al. (2004), Brock and Durlauf (2001), Brock et al. (2003), Avramov (2002), Fernandez et al. (2001a,b), Garratt et al. (2003), and Wright (2008, 2009). Geweke and Amisano (2011) propose optimal density combination for forecast models. Compared to BMA, the frequentist model averaging approach here does not reply on priors and allows for misspecification through the balance of misspecification errors against overparameterization. Furthermore, our frequentist model averaging approach explicitly deals with generated-regressors, while BMA has no known adjustment.

As an alternative to the model averaging approach, forecasts can be based on one model picked by model selection. Numerous model selection criteria have been proposed, including the Akaike information criterion (AIC; Akaike, 1973), Mallows' C_p (Mallows, 1973), Bayesian information criterion (BIC; Schwarz, 1978), and cross-validation (CV; Stone, 1974). Bai and Ng (2009) argue that these model selection criteria are unsatisfactory for factor-augmented regression because they rely on the specific ordering of the factors and the lags, where the natural order may not work well for the forecast of a particular series. This issue is alleviated in forecast combination by the flexibility of choosing candidate models. In addition, the above model selection procedures have not been investigated in the presence of generated regressors; ours is the first to make this extension.

This paper complements the growing literature on forecasting with many regressors. In addition to those discussed above, many papers consider forecast in a data rich environment. Forni et al. (2000, 2005) consider the generalized dynamic factor model and frequency domain estimation. Bernanke et al. (2005) propose forecast with factor-augmented vector autoregressive (FAVAR) model. Bai and Ng (2008) form target predictors associated with the object of interest. Bai and Ng (2009) introduce the boosting approach. A factor-augmented VARMA model is suggested by Dufour and Stevanovic (2010). Pesaran et al. (2011) also investigate multi-step forecasting with correlated errors and factor-augmentation, but in a multivariate framework. Stock and Watson (2012) describe a general shrinkage representation that covers special cases like pretest, BMA, empirical Bayes, and bagging (Inoue and Kilian, 2008). Kelly and Pruitt (forthcoming) propose a three-pass-regression filter to handle many predictors. Tu and Lee (2012) consider forecast with supervised factor models. Dobrev and Schaumgurg (2013) propose using regularized reduced rank regression models for multivariate forecasting with many regressors. A comprehensive comparison among many competing methods is available in Kim and Swanson (2014). The dynamic factor model is reviewed in Stock and Watson (2011). Ng (2011) provides an excellent review on variable selection and contains additional references.

The rest of the paper is organized as follows. Section 2 introduces the dynamic factor model and describes the estimators and combination forecasts. Section 3 provides a detailed description of forecast selection and combination procedures based on the Mallows and leave-h-out cross-validation criteria. Section 4 provides theoretical justification by showing the Mallows and leave-h-out cross-validation criteria are asymptotically unbiased estimators of the MSFE. Monte Carlo simulations and an empirical application to US macroeconomic data are presented in Sections 5 and 6. Summary and discussions are provided in Section 7.

Matlab and Gauss code for the simulation and empirical work reported in the paper is posted at www.ssc.wisc.edu/~bhansen.

2. Model and estimation

Suppose we have observations (y_t, X_{it}) for $t = 1, \ldots, T$ and $i = 1, \ldots, N$, and the goal is to forecast y_{T+h} using the factor-augmented regression model

$$y_{t+h} = \alpha_0 + \alpha(L)y_t + \beta(L)'F_t + \varepsilon_{t+h} \tag{2.1}$$

where $h \geq 1$ is the forecast horizon and $F_t \in \mathbb{R}^r$ are unobserved common factors satisfying

$$X_{it} = \lambda_i' F_t + e_{it}. \tag{2.2}$$

The vectors $\lambda_i \in \mathbb{R}^r$ are called the factor loadings, e_{it} is called an idiosyncratic error, and $\alpha(L)$ and $\beta(L)$ are lag polynomials of order p and q, respectively, for some $0 \leq p \leq p_{\max}$ and $0 \leq q \leq q_{\max}$. We assume that a sufficient number of initial observations are available in history so that the variables in (2.1) are available for T time series observations.