International Journal of Forecasting 34 (2018) 408–430

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Macroeconomic forecasting using penalized regression methods

Stephan Smeekes, Etienne Wijler *

Maastricht University, Department of Quantitative Economics, The Netherlands

ARTICLE INFO

Keywords:
Forecasting
Lasso
Factor models
High-dimensional data
Cointegration

ABSTRACT

We study the suitability of applying lasso-type penalized regression techniques to macroeconomic forecasting with high-dimensional datasets. We consider the performances of lasso-type methods when the true DGP is a factor model, contradicting the sparsity assumptionthat underlies penalized regression methods. We also investigate how the methods handle unit roots and cointegration in the data. In an extensive simulation study we find that penalized regression methods are more robust to mis-specification than factor models, even if the underlying DGP possesses a factor structure. Furthermore, the penalized regression methods can be demonstrated to deliver forecast improvements over traditional approaches when applied to non-stationary data that contain cointegrated variables, despite a deterioration in their selective capabilities. Finally, we also consider an empirical applicationto a large macroeconomic U.S. dataset and demonstrate the competitive performance of penalized regression methods.

1. Introduction

In this paper we provide a thorough analysis of the forecasting capabilities of penalized regression in macroeconomic conditions. We study the performance of these methods in a simulation study when the true DGP is a factor model and when the data contain stochastic trends and may be cointegrated. We also provide a systematic comparison with factor models, the mainstream method used in macroeconomic forecasting, using both Monte Carlo simulations and an empirical application to macroeconomic data.

Despite the vast size of the forecasting literature, comprehensive comparisons between factor models and penalized regression remain scarce. Traditionally, the majority of the forecasting literature seems to have implicitly assumed the prevalence of a latent factor structure in economic datasets and therefore has mainly considered the performance of methods based on factor estimation. While very popular in statistics, only recently ℓ_1-penalized regression techniques, such as the lasso from Tibshirani (1996), are being explored as a viable alternative in macroeconometrics. Applications in forecasting in particular show that the use of penalized regression, potentially in combination with traditional techniques such as principal components (PC), delivers promising performance (e.g Kim & Swanson, 2014), though it is not yet really understood why. By providing a comprehensive study of penalized regression in "adverse" macroeconomic conditions, we complement the existing literature with a fresh perspective on these methods and a direct link to factor models.

Specifically, we address the apparent contradiction between the premise of forecasting with shrinkage estimators to identify a small subset of variables responsible for the variation in the dependent variable and the assumption that the variation in the dependent variable is best explained through aggregates of all available time series. The good empirical performance of penalized regression methods despite this contradiction gives rise to a number of practically relevant questions; (1) Is the common factor assumption really valid in practice? (2) Are the results due

* Correspondence to: Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands.
E-mail address: E.Wijler@maastrichtuniversity.nl (E. Wijler).

https://doi.org/10.1016/j.ijforecast.2018.01.001

to sample-dependent data idiosyncrasies? (3) Are other mechanisms at play such as an inherent robustness of shrinkage estimators to alternative DGP specifications?

We aim to shed light on these previously unexplored questions by conducting a detailed simulation study in which we compare the performance of a selection of the most popular and well understood variants of ℓ_1-shrinkage estimators and factor extraction methods. The novelty in these simulations comes from the wide range of DGPs considered, chosen such that no method is consistently favoured over another based on a priori expectations and to closely resemble the types of data that occur in empirical applications. The former goal is maintained through varying both the presence of common factors in the data as well as the degree of sparsity in the parameter space, while the latter goal is maintained through introducing levels of non-sphericity frequently encountered in empirical work.[1] In addition, we explore the potential of penalized regression in the non-stationary setting by generating a number of time series containing unit roots, some of which are cointegrated, and employ penalized regression directly on these series without any form of preprocessing. We complement the simulations with a comparison of the pseudo out-of-sample forecasting performance on a recently updated U.S. macroeconomic dataset available through the Fred-MD database (McCracken & Ng, 2015).

The results show that penalized regression performs remarkably well when there is at least some degree of sparsity in the parameter space and is relatively robust against alternative DGP specifications. Factor models perform slightly better than penalized regression when the predictors possess an approximate factor structure with low dependence in the errors, but their performance deteriorates substantially when increasing the level of non-sphericity in the idiosyncratic component. Penalized regression naturally does better than factor models on DGPs without factors, but more surprisingly also provides forecast improvements on DGPs containing a factor structure with strongly serially and cross-sectionally correlated idiosyncratic components. In addition, penalized regression shows promising results on cointegrated data, producing substantially lower forecast errors compared to standard OLS despite failing to identify the exact cointegrating vector at relatively high frequencies. Finally, the empirical application highlights that the forecast performance differentials between factor-based methods and shrinkage methods are sensitive to the target variable being forecast.

Our contribution complements the vast existing macroeconomic forecasting literature that is dominated by methods that exploit a latent factor structure, such as static factor models (e.g. Bai & Ng, 2008; Stock & Watson, 2002a,b), dynamic factor models (Doz, Giannone, & Reichlin, 2012; Eickmeier & Ziegler, 2008; Forni, Giovannelli, Lippi, & Soccorsi, 2016; Forni, Hallin, Lippi, & Reichlin, 2005), weighted principal components (Boivin & Ng, 2006), sparse principal components (Kristensen, 2017) or factor augmented vector autoregressions (Bai, Li, & Lu, 2016; Bernanke, Boivin, & Eliasz, 2005; Pesaran, Pick, & Timmerman, 2011). The conjecture that a small set of factors drives the variation in economic time series finds strong support through impressive forecasting performance of factor models on macroeconomic datasets from the U.S. (Stock & Watson, 2002a, 2012), the U.K. (Artis, Banerjee, & Marcellino, 2005) and the Euro area (Marcellino, Stock, & Watson, 2003). Spurred by theoretical developments such as the extension of the adaptive lasso to general time series frameworks by Medeiros and Mendes (2016), ℓ_1-penalized regression has gained more appeal and the body of applied literature taking into account these shrinkage estimators has grown considerably, with recent work covering penalized regression (De Mol, Giannone, & Reichlin, 2008; Gelper & Croux, 2008; Kim & Swanson, 2014; Li & Chen, 2014), reduced-rank vector autoregressions (Bernardini & Cubadda, 2015), Bayesian vector autoregressions (Bańbura, Giannone, & Reichlin, 2010) and penalized vector autoregressions (Barigozzi & Brownlees, 2017; Callot & Kock, 2014; Hsu, Hung, & Chang, 2008; Kascha & Trenkler, 2015). While some include a direct comparison between at least some form of factor models and penalized regression and demonstrate predictive capabilities of ℓ_1-penalized regression that is competitive to traditional factor models, the analysis is typically based on empirical data or simulations that do not provide detailed insights into the sensitivity of each method to its underlying assumptions.

The remainder of this paper is organized as follows. Section 2 describes the notation and reviews the methods considered. In Section 3 we perform the simulation based analysis of the forecasting performance, followed by the empirical application in Section 4. In Section 5 we conclude and suggest a number of interesting avenues for future research.

2. Methods

Suppose a researcher is interested in predicting an economic time series h-steps ahead with information available through time $t = 1, \ldots, T$. The researcher desires to include a pre-determined set of variables such as lags of the dependent variable or variables motivated through economic theory. In addition, she faces a large set of candidate variables that are potentially relevant to the dependent variable. This results in the following generic model:

$$y_{t+h} = \boldsymbol{w}_t'\boldsymbol{\beta}_w + \boldsymbol{x}_t'\boldsymbol{\beta}_x + \epsilon_{t+h} \tag{1}$$

where y_{t+h} is the scalar valued dependent variable to forecast and h is the forecast horizon. $\boldsymbol{w}_t$ is the $(p \times 1)$ predetermined vector of variables which the researcher requires to be in the model, $\boldsymbol{x}_t$ is the $(N \times 1)$ vector containing candidate variables that are potentially related to y_{t+h}, and ϵ_{t+h} is a disturbance term. The forecast of the response at time T is defined as $\hat{y}_{T+h|T} = \boldsymbol{w}_T'\hat{\boldsymbol{\beta}}_w + \boldsymbol{x}_T'\hat{\boldsymbol{\beta}}_x$. Letting $\boldsymbol{y} = (y_{1+h}, \ldots, y_{T+h})'$, $\boldsymbol{W} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_T)'$, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)'$ and $\boldsymbol{\epsilon} = (\epsilon_{1+h}, \ldots, \epsilon_{T+h})$ the model can be rewritten as

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{\beta}_w + \boldsymbol{X}\boldsymbol{\beta}_x + \boldsymbol{\epsilon}. \tag{2}$$

When the number of variables in the candidate set $\boldsymbol{X}$ is large relative to the number of available observations,

[1] Throughout this paper the term non-sphericity refers to the presence of cross-sectional and/or serial correlation in the idiosyncratic component of any data generating process.