VOLUME 28, ISSUE 3
JULY–SEPTEMBER 2012
ISSN 0169-2070
ELSEVIER
international journal of forecasting
IIF
International Institute of Forecasters

International Journal of Forecasting 28 (2012) 739–740

Contents lists available at SciVerse ScienceDirect

International Journal of Forecasting

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

Discussion

Comment on ‘Fast sparse regression and classification’ by J.H. Friedman

George Kapetanios [a], M. Hashem Pesaran [b,c,*]

[a] Queen Mary, University of London, United Kingdom
[b] Cambridge University, United Kingdom
[c] University of Southern California, USA

1. Introduction

The paper by Friedman (2012) discusses a general framework for estimating the vector of coefficients associated with a large number of variables at a given point in time for forecasting a particular target variable one or more periods ahead. The issue is that the number of variables available is large relative to the number of observations available, and therefore standard estimation methods are not applicable. The paper then goes on to formulate the problem as one where standard estimation methods need to be modified by using appropriate penalty terms in the optimization criterion.

The use of the penalty terms considered by Friedman (2012) takes the form of a bridge regression, which provides a unified paradigm for dealing with the high dimensional estimation problem being considered. As the author argues, the main issue with this type of regression is the cost of the computations involved in the minimisation problem needed to carry out a bridge regression, and the author provides a general suggestion for reducing this cost via the use of direct search methods, and generalised path seeking procedures in particular. This direct search methodology is then analysed and expounded with the help of a number of examples.

In this discussion, we focus on two main issues that concern us. The first issue relates to the determination of the parameter $\boldsymbol{\beta}$ associated with the specification of the Generalised Elastic Net of Section 2.3.2 of Friedman (2012). As he notes, the need for regularisation methods arises from the difficulty of estimating the expected loss (risk) in his eq. (3) using estimators such as those in his eq. (8). Reading the paper, one wonders about the theoretical properties of regularisation associated with this estimation, especially in the case of the Generalised Elastic Net, with its dependence on the data generating mechanism for the vector of $\mathbf{x}_i$ and the effect of the choice of an estimator of $\boldsymbol{\beta}$ on these properties. It is clear that estimating his eq. (3) for large sets of predictors is extremely challenging, as one needs to (implicitly or explicitly) integrate over the large dimensional object $\mathbf{x}_i$, with all of the problems associated with such an integration.

This issue is best exemplified by discussing Example 4.3 of the paper. The example has a very specific generating mechanism for $\mathbf{x}_i$, in the sense that the covariance matrix for $\mathbf{x}_i$ has the same value for all non-diagonal elements. Clearly, this is a rather restricted specification which is essentially a single factor model with homogeneous loadings. One wonders to what extent the choice of the covariance matrix affects the performance of the Generalised Elastic Net. Further, given the reported sensitivity of the Generalised Elastic Net to the value of $\boldsymbol{\beta}$, it would be nice to have a clearer feel for the uncertainty associated with a data dependent method for selecting $\boldsymbol{\beta}$, such as cross-validation, particularly if the underlying data generating process turns out to be non-stationary, which is the second issue that concerns us.

Friedman (2012) assumes that both the target variable and the predictor variables are stationary, and that they all remain stationary during the period over which the target variable is being forecast. Such an assumption may be appropriate for biological or other data forms, but it is almost certainly not satisfied for economic data. Evidence of this is widespread in applied econometric research, and it is therefore of paramount importance to delineate the implications of the failure of the stationarity assumption for the methods considered by Friedman, if they are to be applied fruitfully to economic and financial data.

[*] Correspondence to: Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, United Kingdom. Tel.: +44 0 1223 335216; fax: +44 0 1223 338403.
E-mail address: mhp1@cam.ac.uk (M.H. Pesaran).

0169-2070/$ – see front matter © 2012 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2012.04.002

2. Structural change and its implications for sparse regression and classification methods

In order to set the scene, we recast the framework of Friedman (2012) given in his eq. (2) by allowing for structural change, in principle. We have

$$F(\mathbf{x}_i, \mathbf{a}_i) = a_{0,i} + \sum_{j=1}^{n} a_{j,i} x_{i,j}, \tag{1}$$

where $\mathbf{a}_i = (a_{0,i}, a_{1,i}, \ldots, a_{n,i})'$ is a vector of parameters. $a_{0,i}$ is a constant, whereas $a_{j,i}, j = 1, \ldots, n$, represent slope coefficients. The distinction between the two is crucial in the context of structural change. The only difference between this and eq. (2) of Friedman (2012) is the presence of the i subscript associated with the parameter vector $\mathbf{a}$, indicating that the parameters may depend on the observation considered. Forecasting in this framework is clearly hopeless unless considerable structure is imposed on the evolution of the parameter vector, $\mathbf{a}_i$, over i. The parameter change can be modelled assuming that break points are either discrete or continuous. A prominent example of the latter is the random walk formulation, $\mathbf{a}_i = \mathbf{a}_{i-1} + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{\varepsilon}_i$ is an *i.i.d.* sequence with a relatively small variance.

The first approach requires the estimation of both the points and sizes of the breaks, and, as was argued by Pesaran and Timmermann (2007) and Pesaran, Pick, and Pranovich (2011), is likely to involve an important trade-off between the use of a shorter sample with a smaller bias and the use of a longer sample which is likely to yield forecast errors with smaller variances. The second approach requires the estimation of the coefficient process, usually by setting up and estimating a state space model. This approach has been discussed by Cogley and Sargent (2001) and Cogley and Sargent (2005), among others. A useful survey is provided by Hyndman, Koehler, Ord, and Snyder (2008). Under fairly general conditions, both approaches reduce to the down-weighting of observations before estimation and forecasting. See, for example, Eklund, Kapetanios, and Price (2010), Giraitis, Kapetanios, and Price (2012), and Pesaran et al. (2011). This is only a very brief description of a vast body of literature whose impact on forecasting in the presence of large data sets has, so far, been minimal.

Clearly, when structural change is present, the forecasting problem addressed by Friedman (2012) becomes much more difficult, since, unlike in the stationary case, one cannot assume that more data will make the forecasting problem easier to solve. It is important to note that the above-quoted work on down-weighting past data deals only with finite (and quite small) sets of regressors. Given the need to allow for structural change, at least for economic and financial data, it is important to consider what aspects of Friedman's analysis need to be qualified. In particular, it would be interesting to see whether the bridge regression approach can be combined fruitfully with the down-weighting of observations.

Another important aspect of Friedman's approach is the need to orthogonalise the predictors before they are used for forecasting. This is important, as orthogonality is a condition for the direct search methods to be equivalent to the more theoretically justified methods like bridge regressions. However, the presence of structural change means that orthogonalising variables appropriately is not straightforward. This in turn raises important issues concerning the equivalence of direct search methods and bridge regressions. Perhaps the starkest way to illustrate the problem is simply to note that the main aim of the methods advocated by Friedman (2012) is to efficiently minimise

$$E_{y,x}(y_i - \mathbf{a}'\mathbf{x}_i)^2 = E_y(y_i)^2 + \mathbf{a}'E_x(\mathbf{x}_i\mathbf{x}_i')\mathbf{a} - 2\mathbf{a}'E_{y,x}(\mathbf{x}_i y_i).$$

However, since, under structural change, neither $E_x(\mathbf{x}_i\mathbf{x}_i')$ nor $E_{y,x}(\mathbf{x}_i y_i)$ remain fixed over i during either the estimation or forecasting periods, the minimisation problem will be much more complicated to implement.

A second issue posed by the presence of structural change is related to the fact that structural change may be more problematic when the number of variables is large, since the stability of the correlation matrix is less likely to hold. When the number of variables is small and they have been chosen based on theoretical criteria, one can provide arguments under which correlation structures are reasonably stable, or if they are unstable then their instability may be easier to surmise. However, when the number of variables is large, such accounts are more difficult to support.

To conclude, we believe that the question of which variables to consider and that of the handling of structural change have to be considered jointly. In certain cases where structural change is an important factor, one may wish to use a small data set and sacrifice the benefit of a large set of predictors, given the potential cost of accommodating them inappropriately in a changing environment.

References

Cogley, T., & Sargent, T. J. (2001). Evolving post-World War 2 U.S. inflation dynamics. *NBER Macroeconomics Annual*, *16*, 331–337.

Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post-WWII US. *Review of Economic Dynamics*, *8*, 262–302.

Eklund, J., Kapetanios, G., & Price, S. (2010). *Forecasting in the presence of recent structural change*. Bank of England Working Paper No. 406.

Friedman, J. H. (2012). Fast sparse regression and classification. *International Journal of Forecasting*, *28*(3), 722–738.

Giraitis, L., Kapetanios, G., & Price, S. (2012). *Adaptive forecasting in the presence of recent and ongoing structural change*. Working Paper, Queen Mary, University of London.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Heidelberg: Springer-Verlag.

Pesaran, M. H., Pick, A., & Pranovich, M. (2011). *Optimal forecasts in the presence of structural breaks*. Working paper, University of Cambridge. Available at SSRN: http://ssrn.com/abstract=1977191 or http://dx.doi.org/10.2139/ssrn.1977191.

Pesaran, M. H., & Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, *137*(1), 134–161.