International Journal of Forecasting 32 (2016) 736–753

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Nonlinear forecasting with many predictors using kernel ridge regression

Peter Exterkate[a,b,*], Patrick J.F. Groenen[c,d], Christiaan Heij[c], Dick van Dijk[c,d,e]

[a] School of Economics, University of Sydney, Australia
[b] CREATES, Aarhus University, Denmark
[c] Econometric Institute, Erasmus University Rotterdam, The Netherlands
[d] Erasmus Research Institute of Management, Erasmus University Rotterdam, The Netherlands
[e] Tinbergen Institute, Erasmus University Rotterdam, The Netherlands

ARTICLE INFO

Keywords:
High dimensionality
Nonlinear forecasting
Ridge regression
Kernel methods

ABSTRACT

This paper puts forward kernel ridge regression as an approach for forecasting with many predictors that are related to the target variable nonlinearly. In kernel ridge regression, the observed predictor variables are mapped nonlinearly into a high-dimensional space, where estimation of the predictive regression model is based on a shrinkage estimator in order to avoid overfitting. We extend the kernel ridge regression methodology to enable its use for economic time series forecasting, by including lags of the dependent variable or other individual variables as predictors, as is typically desired in macroeconomic and financial applications. Both Monte Carlo simulations and an empirical application to various key measures of real economic activity confirm that kernel ridge regression can produce more accurate forecasts than traditional linear and nonlinear methods for dealing with many predictors based on principal components.

1. Introduction

Current practice involves forecasters in macroeconomics and finance facing a trade-off between model complexity and forecast accuracy. The uncertainty associated with model choice and parameter estimation means that a highly complex predictive regression model based on many variables or intricate nonlinear structures is often found to produce forecasts that are less accurate than those from a simpler model that discards some of the information that is at the researcher's disposal.

Various methods for working with many predictors while circumventing this *curse of dimensionality* in a linear framework have been applied in the recent forecasting literature, as was surveyed by Stock and Watson (2006). Most prominently, Stock and Watson (2002) advocate summarizing large panels of predictor variables into a small number of principal components, which are then used in a dynamic factor model for forecasting purposes. Alternative approaches include combining forecasts based on multiple models, where each includes only a relatively small number of variables (Aiolfi & Favero, 2005; Faust & Wright, 2009; Huang & Lee, 2010; Rapach, Strauss, & Zhou, 2010; Wright, 2009), partial least squares (Groen & Kapetanios, 2008), and Bayesian regression (Bańbura, Giannone, & Reichlin, 2010; Carriero, Kapetanios, & Marcellino, 2011; De Mol, Giannone, & Reichlin, 2008). Stock and Watson (2012) find that the dynamic factor model approach is preferable to these alternatives for forecasting macroeconomic time series; see also Çakmaklı and van Dijk (2010) and Ludvigson and Ng (2007, 2009) for successful applications in finance.

* Correspondence to: School of Economics, Merewether Building H04, The University of Sydney, NSW 2006, Australia. Tel.: +61 2 9351 8532.
E-mail address: peter.exterkate@sydney.edu.au (P. Exterkate).

http://dx.doi.org/10.1016/j.ijforecast.2015.11.017
0169-2070/

The possibility of the existence of nonlinear relationships among macroeconomic and financial time series has also received ample attention over the last two decades. The most popular nonlinear forecast methods include regime-switching models and neural networks, see the surveys by Teräsvirta (2006) and White (2006), respectively, and the comprehensive overview by Kock and Teräsvirta (2011). Alternative approaches include sieve estimation (Chen, 2007) and nonparametric regression (Pagan & Ullah, 1999). Typically, these approaches are suitable only for small numbers of predictors, and their ability to improve upon the predictive accuracy of linear forecasting techniques seems limited, see Medeiros, Teräsvirta, and Rech (2006), Stock and Watson (1999), and Teräsvirta, van Dijk, and Medeiros (2005), among others. Giovannetti (2013) proposes a hybrid approach, where a nonlinear model is estimated using principal components extracted from a large set of predictors.

In this paper, we introduce a forecasting technique that can deal with high-dimensionality and nonlinearity simultaneously. The central idea is to employ a flexible set of nonlinear prediction functions and to prevent overfitting by penalization, in a way that limits the computational complexity. In this approach, which is known as *kernel ridge regression*, the set of predictors is mapped into a high-dimensional (often even infinite-dimensional) space of nonlinear functions of the predictors. A forecast equation is estimated in this infinite-dimensional space, using a penalty (or shrinkage, or ridge) term to avoid overfitting. This allows kernel ridge regression to avoid the curse of dimensionality, which plagues alternative nonparametric approaches when allowing for flexible types of nonlinearity. Computational tractability is achieved by choosing the kernel in a convenient way, so that calculations in the infinite-dimensional space are actually prevented. This approach also avoids the computational difficulties that are encountered in standard linear ridge regression when the number of predictor variables is large relative to the number of time series observations. These properties mean that kernel ridge regression provides an attractive framework for estimating nonlinear predictive relationships in a data-rich environment.

The kernel methodology was developed in the machine learning community, an area which often involves large data sets. The terminology originates from operator theory, as computations are performed in terms of the kernel of a positive integral operator, see Vapnik (1995). We use the term *kernel* in this sense because it is the established term for this method in machine learning. This meaning should not be confused with other uses of the word, such as in kernel smoothing methods for local regression.

A typical application of kernel methods is classification; for example, in the optical recognition of pixel-by-pixel scans of handwritten characters. Schölkopf, Smola, and Müller (1998) document the outstanding performance of kernel methods for this classification task. Kernel ridge regression has also been found to work well in many other applications. Time series applications are scarce, and seem to be limited to deterministic (that is, non-stochastic) time series (Müller et al., 1997). On the other hand, linear penalized regression methods, including linear ridge regression, are used widely in economic forecasting; see the recent overview by Kim and Swanson (2014). To the best of our knowledge, kernel ridge regression has not yet been applied in the context of macroeconomic or financial time series forecasting.

This paper makes two methodological contributions to kernel ridge regression. First, we extend the approach to enable the use of models that include lags of the dependent variable or other individual variables as predictors, as is typically desired in economic and financial forecasting applications. Second, we derive a computationally efficient cross-validation procedure for selecting the tuning parameters involved in kernel ridge regression, including the shrinkage parameter that determines the strength of the penalization.

We provide simulation evidence demonstrating that kernel ridge regression delivers more accurate forecasts than conventional principal components methods in the presence of many predictors that are related nonlinearly to the target variable. These conventional methods include threshold autoregressions, extensions of principal component regressions to accommodate nonlinearity, as put forward by Bai and Ng (2008), sieves, and standard nonparametric regression techniques. The practical usefulness of kernel methods is confirmed in an empirical application to the forecasting of four key measures of US macroeconomic activity over the period 1970–2009: industrial production, personal income, manufacturing and trade sales, and employment. When traditional methods perform poorly, kernel ridge regression yields substantial improvements. This result holds for all series at a one-year horizon, and also at shorter horizons for production and income. When traditional forecasts are of good quality, as is the case for sales and employment, kernel-based forecasts remain competitive. We also find that kernel ridge regression is affected less by the 2008–09 financial and economic crisis than traditional methods.

The remainder of this paper is organized as follows. Section 2 describes the kernel methodology. The Monte Carlo simulation is presented in Section 3, and Section 4 discusses the empirical application. Conclusions are provided in Section 5. Details of the technical results are collected in Appendix A.

2. Methodology

The technique of kernel ridge regression (KRR) is based on ordinary least squares (OLS) regression and ridge regression. Therefore, we begin this section by providing a brief review of these methods, highlighting their limitations for dealing with nonlinearity and high-dimensionality. Next, we show how kernel ridge regression overcomes these drawbacks by means of the so-called *kernel trick*. We extend the KRR methodology to allow for "preferred" predictors, in order to enable it to be used in time series contexts. We also present the properties of some kernel functions that are popular because of their computational efficiency. As will become clear below, kernel ridge regression involves the use of tuning parameters. In Section 2.5 we propose a computationally efficient cross-validation procedure for selecting values for these parameters.