Journal of Econometrics 187 (2015) 345–357

Contents lists available at ScienceDirect

Journal of Econometrics

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

A flexible semiparametric forecasting model for time series

Degui Li [a], Oliver Linton [b,*], Zudi Lu [c]

[a] Department of Mathematics, University of York, York, YO10 5DD, UK

[b] Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge University, Cambridge, CB3 9DD, UK

[c] Southampton Statistical Sciences Research Institute & School of Mathematical Sciences, University of Southampton, Southampton, SO17 1BJ, UK

ARTICLE INFO

Article history:
Received 21 December 2012
Received in revised form
8 December 2014
Accepted 2 February 2015
Available online 25 March 2015

JEL classification:
C14
C22

Keywords:
Forecasting
Marginal regression
Model averaging
Kernel estimation
Near epoch dependence
Semiparametric estimation

ABSTRACT

In this paper, we propose a semiparametric procedure called the "Model Averaging MArginal Regression" (MAMAR) that is flexible for forecasting of time series. This procedure considers approximating a multivariate regression function by an affine combination of one-dimensional marginal regression functions. The weight parameters involved in the approximation are estimated by least squares on the basis of the first-stage nonparametric kernel estimates of the marginal regressions. Under some mild conditions, we have established asymptotic normality for the estimated weights and the regression function in two cases: Case I considers that the number of the covariates is fixed while Case II allows the number of the covariates depending on the sample size and diverging. As the observations are assumed to be stationary and near epoch dependent, the approach developed is applicable to both the estimation and forecasting issues in time series analysis. Furthermore, the method and result are augmented by a simulation study and illustrated by an application in forecasting the high frequency volatility of the FTSE100 index.

1. Introduction

In many situations of practical interest, we are faced with a large number of variables and uncertain functional forms. Linearity is widely adopted in macroeconometrics where data is limited, but for many relationships this linearity may imply absurd conclusions when covariates are pushed to extreme values. Moreover, in the regression settings, we may have to choose between a large number of covariates. In the case of time series, the problem can get even worse, since in both estimation and forecasting, all possible lags of all possible predictor variables may be the candidates and their influences are of unknown forms. One approach to deal with this problem is to use model selection tools that choose the best model according to some traditional criterion from a set of models. In some cases, such an approach can be very time consuming. Also, it may be neglecting features of the data that arrive through the models which are not selected but are almost as good as those which are selected. A popular method is to use model averaging whereby we fit a number of candidate models and then weight them according to some criterion (see, for example Hansen, 2007; Liang et al., 2011). Another approach that is now popular in statistics is to use some penalization device to force many weights to be zero. For instance, the least absolute shrinkage and selection operator (LASSO) proposed by Tibshirani (1996, 1997), is the penalized least squares estimate with the L_1 penalty. The penalized regression with general L_q penalty leads to a bridge regression (Frank and Friedman, 1993; Fu, 1998). Fan and Li (2001) used the smoothly clipped absolute deviation (SCAD) penalty in penalized likelihood estimation. For recent developments and surveys on model averaging and variable selection, the reader is referred to Claeskens and Hjort (2008), Fan and Lv (2008, 2010), Bühlmann and van de Geer (2011) and references therein. However, most of the literature regarding model averaging and selection has been concerned with parametric models, which assume some parametrically linear or nonlinear relationships among the variables considered. In this paper, we will consider nonparametric and semiparametric models and will focus on with model averaging, leaving the penalty model selection issue in the setting of this paper to our following future work.

Specifically, let $(Y_t, X_t^\intercal)$ be a stationary time series process, where $X_t = (X_{t1}, \ldots, X_{td})^\intercal$ is a d-dimensional random vector and the superscript $^\intercal$ stands for the transpose of a vector or matrix. In many applications, we need to consider estimating regression

* Corresponding author.
E-mail addresses: degui.li@york.ac.uk (D. Li), obl20@cam.ac.uk (O. Linton), Z.Lu@soton.ac.uk (Z. Lu).

http://dx.doi.org/10.1016/j.jeconom.2015.02.025
0304-4076/

function $\mathsf{E}(Y_t|X_t = x)$ with $x = (x_1, \ldots, x_d)^\tau$, which can be well estimated by nonparametric method when the dimension d is small, but very poorly if the dimension d is high (say larger than 3) owing to the so-called "curse of dimensionality". Various nonparametric and semiparametric models, such as additive models, varying coefficient models, partially linear models, have been studied to deal with the curse of dimensionality problem in the literature (see, for example Fan and Yao, 2003; Li and Racine, 2007; Teräsvirta et al., 2010). In the time series case, as mentioned above, the conditioning information may consist of an infinite number of lags, i.e., $d = \infty$. Linton and Sancetta (2009) established consistency of estimators of $\mathsf{E}(Y_t|Y_{t-1}, Y_{t-2}\ldots)$ under weak conditions without any functional form restrictions beyond some limited smoothness, but rates of convergence are not available and practical performance is likely to be poor without further restrictions. Instead, it makes sense to use lower dimensional predictors, but which one? We next consider some explicit semiparametric models that have been tried to address the issue in nonlinear time series.

Linton and Mammen (2005) considered the semiparametric (volatility) regression model

$$\mathsf{E}(Y_t^2|Y_{t-1}, Y_{t-2}\ldots) = \sum_{j=1}^{\infty} \psi_j(\theta) m(Y_{t-j}),$$

where $m(\cdot)$ is an unknown function and the parametric family $\{\psi_j(\theta), \theta \in \Theta, j = 1, \ldots, \infty\}$ satisfies some regularity conditions. This model includes the GARCH(1,1) as a special case and also includes an infinite set of lags. They assumed that $\{Y_t\}$ is stationary and geometrically mixing and thereby obtained a characterization of the function m as the solution of a linear integral equation with intercept of the form $m_\theta^*(x) = \sum_{j=1}^{\infty} \psi_j(\theta) m_j(x)$, where $m_j(x) = \mathsf{E}(Y_t^2|Y_{t-j} = x)$ for each j. They proposed an estimation strategy for the unknown quantities, which requires as input the estimation of $m_j(x)$ for $j = 1, 2, \ldots, J(T)$, where $J(T) = c \log T$ for some $c > 0$. They required to bound the estimation error of $m_j(x)$ uniformly over x and over $j = 1, 2, \ldots, J(T)$. However, they provided only a sketch proof of this result in the case where the process is assumed to have compact support and to be strong mixing with geometric decay. A recent paper by Li et al. (2012) provided a more rigorous and complete proof of this result. Linton and Mammen (2008) generalized this class of models to allow for exogenous regressors and more complicated dynamics. See Chen and Ghysels (2010) an application of these methods to volatility forecasting.

This general approach to modeling is promising but quite computationally demanding. In addition, the models considered thus far all have a finite number of unknown functions (for example, in Linton and Mammen (2005) only one unknown function was allowed), and so appear to be heavily over identified. In this paper, we aim at relaxing such restrictive assumptions and consider a semiparametric model that contains possibly infinitely many unknown functions all of which can enter into the prediction. This may be particularly useful in situations where there is a lot of nonlinearity and a rich dynamic structure. The most general version of our model is similar in some ways to the setting considered in Hansen (2007) except instead of observed covariates we have nonparametrically estimated ones. We call our method MAMAR (Model Averaging MArginal Regression). We obtain consistency and asymptotic normality of our procedure under general conditions. We further apply our method to volatility forecasting (where the time series is long and (log) linear models are predominant) and obtain some satisfactory results.

The rest of the paper is organized as follows. The approximation of MAMAR is presented in Section 2 and the semiparametric estimation method is presented in Section 3. The asymptotic properties for the estimators of the optimal weights and nonparametric estimators for finite covariates case are provided in Section 4.1, and Section 4.2 gives the theoretical results when the dimension of the covariates is diverging. Discussions of some related topics are given in Section 5. Numerical evidence of our methodology is given in Section 6. Section 7 concludes this paper. All the technical lemmas and the proofs of the main results are collected in the Appendix. A supplemental document gives the application of our method to Australian temperature data and the proofs of some auxiliary results (see Appendix B).

2. Approximation with MAMAR

We model or approximate the conditional regression function $\mathsf{E}(Y|X = x)$ by an affine combination of lower dimensional regression functions. Let S_ℓ denote the set of all subsets of $S = \{1, 2, \ldots, d\}$ of ℓ components, and this has cardinality $J_\ell = \binom{d}{\ell}$. For example, $S_2 = \{(1, 2), \ldots, (d-1, d)\}$ has cardinality $d(d-1)/2$. We model or approximate $m(x) = \mathsf{E}(Y|X = x)$ by

$$m_w(x) = w_0 + \sum_{j=1}^{J} w_j \mathsf{E}(Y|X_{(j)} = x_{(j)})$$

for some weights $w_j, j = 0, 1, \ldots, J$, where $X_{(j)} = (X_{i_1}, \ldots, X_{i_{k_j}})^\tau$ is a subset of X and $x_{(j)} = (x_{i_1}, \ldots, x_{i_{k_j}})^\tau$. In general, $X_{(j)}$ and $X_{(k)}$ could have different dimensions and of course overlapping members. The union of $X_{(j)}$ may exhaust one or more of S_ℓ or it may not. A simple special case that we focus on for much of the paper is where $J = d$ and $X_{(j)} = X_j$ is just the jth component and the covariates are non overlapping. This seems well suited to time series applications. In practice, one would not wish to take k_j to be too large, so as to avoid the curse of dimensionality.

We could be thinking of this as a family of models within which there is a true member that corresponds to the true regression function $m(x)$ or we could be thinking of this as an approximating or model averaging device. Either way, we are seeking $w = (w_0, w_1, \ldots, w_J)^\tau$ that minimizes

$$\mathsf{E}\left[Y - w_0 - \sum_{j=1}^{J} w_j \mathsf{E}(Y|X_{(j)})\right]^2. \tag{2.1}$$

In general, the minimizing weights may not be unique, but the minimization problem is a projection onto the space spanned by the functions $\{\mathsf{E}(Y|X_{(j)}), j = 1, \ldots, J\}$ and so there is a unique solution $m_w(x)$. We shall focus on the special case where there is a unique vector w (which is generally true for the special case where $J = d$ and $X_{(j)} = X_j$ is just the jth component and the covariates are non overlapping). In this case, the minimizer to (2.1), $w_o = (w_{o,0}, w_{o,1}, \ldots, w_{o,J})^\tau$, satisfies

$$w_{o,0} = \left(1 - \sum_{j=1}^{J} w_{o,j}\right)\mathsf{E}(Y), \qquad (w_{o,1}, \ldots, w_{o,J})^\tau = A^{-1}a, \tag{2.2}$$

where A is a $J \times J$ matrix whose (i, j)th component is $\mathsf{Cov}(\mathsf{E}(Y|X_{(i)}), \mathsf{E}(Y|X_{(j)}))$, and a is a J-dimensional vector whose ith component is $\mathsf{Cov}(\mathsf{E}(Y|X_{(i)}), Y)$. If the model is true or the approximation is perfect, (2.1) is equal to zero at the optimal weights but it need not be so. Obviously the conditional component regressions $\mathsf{E}(Y|X_{(j)} = x_j), j = 1, \ldots, J$, are unknown but low dimensional, so they can be well estimated by various nonparametric approaches. In Section 3, we will first estimate these conditional regression functions by the Nadaraya–Watson method and then use the least squares approach to obtain the estimator of w_o. We can consider this approach as a form of model averaging where we are averaging the "models": $\mathsf{E}(Y|X_{(j)} = x_{(j)}), j = 1, \ldots, J$, see Hansen (2007).