Journal of Empirical Finance xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Empirical Finance

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

Bandwidth selection by cross-validation for forecasting long memory financial time series☆

Richard T. Baillie [a,b,c,d], George Kapetanios [e], Fotis Papailias [f,g,]*

[a] Department of Economics, MI State University, USA
[b] Department of Finance, MI State University, USA
[c] School of Economics and Finance, Queen Mary, University of London, UK
[d] Rimini Center for Economic Analysis, Italy
[e] School of Economics and Finance, Queen Mary, University of London, UK
[f] Queen's University Management School, Queens University Belfast, UK
[g] quantf research

A R T I C L E I N F O

Article history:
Received 19 August 2013
Received in revised form 29 March 2014
Accepted 1 April 2014
Available online xxxx

JEL classification:
C22
C53

Keywords:
Long memory
Fractional integration
Filtering
Cross-validation
Forecasting

A B S T R A C T

The paper addresses the issue of choice of bandwidth in the application of semiparametric estimation of the long memory parameter in a univariate time series process. The focus is on the properties of forecasts from the long memory model. A variety of cross-validation methods based on out of sample forecasting properties are proposed. These procedures are used for the choice of bandwidth and subsequent model selection. Simulation evidence is presented that demonstrates the advantage of the proposed new methodology.

1. Introduction

Many economic and financial time series possess the apparent long memory property with slow hyperbolically decaying impulse response (*IR*) coefficients and autocorrelations; see Granger and Joyeux (1980), Granger (1980) and Hosking (1981) for discussion and derivation of their properties, and a detailed survey by Baillie (1996). There are several different commonly applied approaches for dealing with these processes in a practical context. In some cases the presence of long memory may be an integral aspect of the feature of modeling. This generally leads to the developing of a fully articulated model, that incorporates long memory as well as all other features of the data and economic issues.

☆ The authors are very grateful for helpful comments from two anonymous referees, and from the editor, Christian Conrad. Any remaining errors are our responsibility.
* Corresponding author at: Queen's University Management School, Riddel Hall, 185 Stranmillis, BT9 5EE Belfast, UK. Tel.: +44 28 9097 4667; fax: +44 28 9097 4201.
E-mail addresses: f.papailias@qub.ac.uk, f.papailias@quantf.com (F. Papailias).
URL: http://www.quantf.com (F. Papailias).

http://dx.doi.org/10.1016/j.jempfin.2014.04.002
0927-5398/

An alternative approach is to regard long memory as merely an incidental feature of the data, that needs to be acknowledged and handled with another purpose in mind, such as impulse response (*IR*) analysis, or forecasting. One approach for dealing with inference for *IR* analysis is to use a sieve autoregression (*AR*) bootstrap approach and this method has been shown by Baillie and Kapetanios (2013) to have many desirable properties. Another approach, that makes use of the extensive theory of semiparametric estimation of long memory parameters, is to first fractionally filter the data, and then to model the short memory component and then finally to cumulate the short memory component back into the long memory space. For the purpose of forecasting a series, the final step could involve cumulating forecasts of the short memory component to produce forecasts for the original composite series. One reason for using this modeling approach is that first removing the long memory component greatly facilitates the modeling of the short memory part of the series. This wider class of models for the short memory component could include nonlinear models as in van Dijk, Frances and Paap (2002) and Baillie and Kapetanios (2008), or models with exogenous information, such as factor models. Other examples of this methodology are by Asai and McAleer (2012) who first extract the long memory component before estimating a stochastic volatility model following a Wishart distribution; and Koopmans, Carnera and Ooms (2007), who postulate a complex model involving various long memory components which, when filtered out, give estimates of the seasonality and cycles within daily electricity prices.

The forecasting approach discussed in this paper is strongly affected by the choice of the semiparametric estimator of the long memory parameter, denoted by d. The choice of the semiparametric estimate of d is fundamentally dependent on the choice of the bandwidth parameter, and is a critical issue in the analysis of long memory time series. In practice, this important decision is usually assumed away with a standard, or "automatic" choice for the bandwidth, which is a simple function of the sample size. This paper provides an alternative and concrete suggestion for determining the bandwidth based on the forecasting ability of the resulting model; and is therefore an entirely new approach to the existing paradigms. The method suggested in this paper appears to work well since it improves both the estimation of d and also the estimation of the short memory parameters; as well as the root mean squared forecast error, *RMSFE*, for some forecast horizon(s). Hence our new proposed cross-validation strategy is intrinsically designed to improve both the modeling and the forecasting of long memory time series. The properties of filters that are designed to extract and remove the long memory component of a time series are also clearly and inextricably linked to the choice of the estimator of d. Hence our proposed methodology is also directly relevant and applicable to determining the most appropriate method for filtering out long memory effects.

A further approach to the problem of bandwidth selection is the concept of "optimal bandwidth", due to Henry (2001). This idea is based on the minimization of the mean squared error (*MSE*) of the estimate of d given *complete knowledge* of the short memory component including parameter values and model specification. These assumptions are clearly unrealistic and Henry (2001) suggests iterating between the estimates of d and the short memory parameters which determine the bandwidth. The properties of Henry's optimal bandwidth method are compared with the automatic and also cross-validation techniques through an extensive simulation described later in this paper. The results of the simulation turn out to indicate the superiority of the suggested cross-validation technique for the determination of the bandwidth, where the criteria for cross-validation depend on out of sample forecasting performance.

A further idea also considered in this paper, follows a suggestion by Hall, Koul and Turlach (1997), which uses time domain semiparametric estimators, *SPE*s, which are based on the autocorrelations of the time series. For this reason the paper considers the time domain minimum distance estimators (*MDE*) as given by Mayoral (2007). These estimators are generally found to be quite competitive to the corresponding frequency domain *SPE*s.

The plan of this paper is the following: the next section presents the theoretical underpinnings of the analysis along with the frequency domain and time domain estimation methods for long memory processes. The basic ideas concerning filtering long memory features from a time series are also discussed in Section 2. The next section deals with the methodology available for the choice of bandwidth. Section 3 presents the details of a very detailed simulation study which investigates the various methods described above. Then, Section 4 provides an empirical application to several time series of real exchange rates. Finally, Section 6 provides a brief conclusion.

2. Basic theory

A long memory, fractionally integrated process has slow hyperbolic rates of decay associated with its impulse responses (*IR*s) and autocorrelations. Following Granger and Joyeux (1980), Granger (1980) and Hosking (1981), a univariate time series process with fractional integration in its conditional mean is represented by,

$$(1-L)^d(y_t-\mu) = u_t, \; t = 1, \ldots, T, \tag{1}$$

where L is the lag operator, μ is an intercept, and u_t is a short memory, $I(0)$ process. Then y_t is said to be a fractionally integrated process of order d, or $I(d)$. An $I(0)$ process is defined as having partial sums that converge weakly to Brownian motion. The parameter d represents the degree of "long memory", or persistence in the series. For $-1/2 < d < 1/2$ the process is stationary and invertible; while for $1/2 \le d \le 1$, the process does not have a finite variance, but still has a finite cumulative impulse response function. The Wold decomposition, or infinite order moving average representation, with coefficients given by the *IR*s, is given by,

$$y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \tag{2}$$