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

ARTICLE INFO

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

ABSTRACT

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,

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

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,

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

where ϵ_t is a martingale difference sequence such that $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = \sigma^2$. Equivalently, the infinite *AR* representation is given by,

$$y_t = \sum_{i=1}^{\infty} \pi_i y_{t-i} + \epsilon_t. \tag{3}$$

For large lag i, the *IR* coefficients decay at the very slow hyperbolic rate given by $\psi_i \sim c_1 i^{d-1}$ and similarly the infinite autoregressive representation coefficients decay at the rate of $c_2 i^{-d-1}$. Finally, the autocorrelation coefficients decay at the rate of $c_3 i^{2d-1}$, where c_1, c_2 and c_3 are constants. If the short memory component is represented as an $ARMA(p, q)$ process, then Eq. (1) becomes the well known $ARFIMA(p, d, q)$ model, $\phi(L)(1-L)^d y_t = \theta(L)\epsilon_t$, where $\phi(L)$ and $\theta(L)$ are the polynomials in the lag operator of orders p and q respectively, with all their roots lying outside the unit circle to satisfy the requirements for stationarity and invertibility.

2.1. Frequency domain estimation of d

A substantial number of articles have been concerned with semiparametric estimation in the frequency domain to specifically estimate the long memory parameter. This work originally began with Geweke and Porter-Hudak (1983) and has now grown to include nonstationary long memory models; e.g. Phillips (2007). An early comparison of such estimation methods can be found in Bisaglia and Guegan (1998) and more recently in Haldrup and Nielsen (2007). Most of the estimators essentially solve a minimization problem of the form,

$$\widehat{d} = \arg_{d \in [d_1, d_2]} \min R(d),$$

where d_1 and d_2 are the lower and upper bounds respectively for the values of d such that $-\infty < d_1 < d_2 < \infty$ and $R(d)$ is the relevant objective function. The Local Whittle (*LW*) estimator, which is obtained by minimizing the objective function,

$$R^{LW}(d) = \ln\left[\frac{1}{m}\sum_{j=1}^{m} \omega_j^{2d} I\left(\omega_j\right)\right] - \frac{2d}{m}\sum_{j=1}^{m} \ln\left(\omega_j\right), \tag{4}$$

with respect to d, where $\omega_j = (2\pi j)/T$ for $j = 1, 2, \ldots T$ and $I(\omega_j)$ is the periodogram defined as,

$$I\left(\omega_j\right) = \frac{1}{2\pi T}\left|\sum_{j=1}^{T} y_t e^{i\omega_j t}\right|^2. \tag{5}$$

The estimator depends on the choice of bandwidth, m, which is generally chosen as $m = \lfloor T^{\alpha} \rfloor$ where $0 < \alpha < 4/5$; and where $\lfloor \cdot \rfloor$ denotes the integer part. The limiting distribution of the *LW* estimator has been derived under various assumptions concerning the short memory process by Robinson (1995), Dalla et al. (2005) and Shimotsu and Phillips (2006).

Several important extensions of the *LW* estimator have been introduced in the literature. In particular, Shimotsu and Phillips (2005) have proposed the Exact Local Whittle (*ELW*) approach using a "corrected" discrete Fourier transform of the series, where the objective function now becomes,

$$R^{ELW}(d) = \ln\left[\frac{1}{m}\sum_{j=1}^{m} I_{\nabla^d y}\left(\omega_j\right)\right] - \frac{2d}{m}\sum_{j=1}^{m} \ln\left(\omega_j\right), \tag{6}$$

where $\nabla^d = (1-L)^d$.

Abadir et al. (2007) have introduced the Fully Extended Local Whittle (*FELW*) where $d \in (p - 1/2, p + 1/2]$, for $p = 0, 1, 2, \ldots$ which has the particular attraction of covering the region of nonstationarity for long memory processes. A simpler version of the estimator, due to Abadir et al. (2011), redefines the periodogram as,

$$I^{FELW}\left(\omega_j\right) = \left|1 - e^{i\omega_j}\right|^{-2p} I_{\nabla^p y}\left(\omega_j\right),$$

Then the *FELW* is obtained by minimizing the quantity

$$R^{FELW}(d) = \ln\left[\frac{1}{m}\sum_{j=1}^{m} j^{2d} I^{FELW}\left(\omega_j\right)\right] - \frac{2d}{m}\sum_{j=1}^{m} \ln(j). \tag{7}$$

The *LW* and *LPW* are only known to be consistent estimators of d in the stationary region of $-1/2 < d < 1/2$; while the *ELW* and *FELW* estimators are known to be consistent for all values of d.

A particularly important issue concerns the choice of m, which is generally chosen in the range of $T^{1/2} \le m \le T^{4/5}$. In the usual case of ignorance of the short run dynamics, the bandwidth is generally selected in an ad hoc way and a popular choice is $m = T^{0.5}$. However, when there is substantial persistence in the short run dynamics, the value of m should potentially be reduced so that more weight is placed on ordinates of the periodogram associated with the low frequency components. On denoting the spectral density function (*SDF*) of y_t as $f(\omega)$, and the *SDF* of u_t is $f^*(\omega)$, then,

$$f(\omega) = |1-\exp(i\omega)|^{-2d} f^*(\omega).$$

The *SDF* of y_t can be approximated as $\omega \to 0+$, by $f(\omega) = \mathrm{L}(1/\omega)\omega^{-2d}\{1 + c\omega^{\beta} + o(\omega^{\beta})\}$ where $\mathrm{L}(1/\omega)$ is a slowly varying function with $0 < c < \infty$, usually chosen as one, and $\beta \in (0, 2]$. Henry (2001) has found the optimal bandwidth in terms of minimizing the *MSE* of d^*_{LW} to be,

$$m^*_{LW} = \left(\frac{3}{4\pi}\right)^{4/5} \left|\tau^* + \frac{d}{12}\right|^{-2/5} T^{4/5},$$

where $\tau^* = \left[\frac{f^{*\prime\prime}(0)}{2f^*(0)}\right]_{\omega=0}$ and τ^* has the interpretation of representing the degree of smoothness of the spectral density of the short memory component u_t, as the frequency approaches zero. Henry (2001) considers iterating between successive choices of $\hat{d}_{LW}$ and m^*_{LW} in order to estimate d.

2.2. Time domain estimation of d

Corresponding time domain estimation methods exist as alternatives to the *SPE*s in the frequency domain. Some initial work on estimating d from the large lag relationship that,

$$ln\left(\widehat{\rho}_k\right) = \left(2\widehat{d}-1\right)\ln(k),$$

is to be found in Hall, Koul and Turlach (1997). However, the generally preferred estimator is based on the minimum distance estimation method (*MDE*). This involves minimizing the generalized, or Mahalanobis, distance in terms of the population autocorrelations implied by a particular *ARFIMA* from the sample autocorrelations. The first article to suggest and to analyze this approach was by Tieslau et al. (1996), who also show that omission of the first autocorrelation at lag one leads to a considerable loss of asymptotic efficiency. Also the *MDE* is only $T^{1/2}$ consistent and asymptotically normally distributed for $-0.5 < d < 0.25$, a result which follows from the properties of the sample autocorrelations and was first derived by Hosking (1996). Subsequent articles by Mayoral (2007) and Zevallos and Palma (2013) have extended the range of the *MDE* for d. As the number of autocorrelations increases it appears that the *MDE* gains full asymptotic efficiency and this property appears to extend to stationary and invertible *ARMA* processes with non i.i.d. innovations; see Baillie and Chung (2001).

The *MDE* is relatively simple to compute for the *ARFIMA*$(0, d, 0)$ process, but far more complicated for higher order *ARFIMA*(p, d, q) models. This study uses the estimator introduced by Mayoral (2007) which has a relatively simple form and is based on the truncated (*AR*) which is a finite version of Eq. (3). The residuals can then be estimated by,

$$\hat{\epsilon}_t = y_t - \sum_{i=1}^{t-h-1} \hat{\pi}_i (y_{t-i} - \hat{\mu}), \quad t = h+1, \ldots, T, \tag{8}$$

where $h = d - \zeta$ and $\zeta = d - \lfloor d + 0.5 \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part, and μ is generally estimated by using the sample mean defined as $\bar{y}(h) = \frac{1}{T-h}\sum_{h+1}^{T} y_t$. The definitions of both h and ζ allow for inference in both the stationary and nonstationary cases. The generalized *MDE* is defined as the vector of model parameters, $\hat{\lambda}_k$, that minimizes the following objective function,

$$V_{K\epsilon}(\lambda, y) = \sum_{i=1}^{k} \hat{\rho}_{\epsilon(\lambda)}(i)^2, \tag{9}$$

where $\hat{\rho}_{\epsilon(\lambda)}(i)$ denotes the sample *i-th* autocorrelation associated with the residuals.

Please cite this article as: Baillie, R.T., et al., Bandwidth selection by cross-validation for forecasting long memory financial time series, J. Empir. Finance (2014), http://dx.doi.org/10.1016/j.jempfin.2014.04.002

2.3. Filtering long memory components

If the long memory parameter d is known, then the observed y_t series can be fractionally filtered to obtain,

$$u_t = y_t - \sum_{l=1}^{t-1} \pi_l(d) y_{t-l},$$

where,

$$(1-L)^d y_t = y_t - \sum_{l=1}^{\infty} \pi_l(d) y_{t-l},$$

and $\pi_l(d)$ are the coefficients of the infinite AR representation of y_t in terms of u_t, so that,

$$\pi_l(d) = \Gamma(l-d)\Gamma(-d)^{-1}\Gamma(l+1),$$

where $\Gamma(\bullet)$ denotes the Gamma function. In practice, d is unknown and can be replaced by the LW estimate, $\hat{d}_{LW}$. Then, the Feasible Fractionally Filtered (FFF) series based on observable quantities is,

$$\hat{u}_t = y_t - \sum_{l=1}^{t-1} \hat{\pi}_l\left(\hat{d}_{LW}\right) y_{t-l}, \tag{10}$$

where $\hat{\pi}_l\left(\hat{d}_{LW}\right) = \Gamma\left(l-\hat{d}_{LW}\right)\Gamma\left(-\hat{d}_{LW}\right)^{-1}\Gamma(l+1)$. Surprisingly little has been done on the properties of the filtered series apart from Baillie and Kapetanios (2013) who show that the LW two step estimator ($LWTSE$) of the short memory parameters will only be $O_p(m^{-1/2})$ with $m < T^{1/2}$.

3. Choice of optimal bandwidth

As previously discussed, a critical issue in the application of the frequency domain SPEs of d concerns the appropriate choice of the bandwidth m. Similarly, the corresponding issue with the MDE in the time domain is the choice of the number of autocorrelations k to be used. The method of Henry (2001) requires knowledge of τ^* which is the smoothness of the spectral density function of the short memory component. For simpler short memory processes, exact analytic expressions are available and for the $ARFIMA(1, d, 0)$ process, Henry (2001) finds that,

$$\tau^* = \frac{-\phi}{(1-\phi)^2}. \tag{11}$$

Similar derivations are possible for higher order models and after some lengthy algebra, for the $ARFIMA(2, d, 0)$ it can be shown that,

$$\tau^* = \frac{\phi_1\phi_2 - \phi_1 - 4\phi_2}{(1-\phi_1-\phi_2)^2}. \tag{12}$$

Clearly, a major difficulty of this method concerns the assumption that the exact form of the short memory dynamics and its associated parameter values are completely known. Henry (2001) suggests an iterative method for the estimation of d and the bandwidth. This method is also quite ad hoc in terms of the selection of the bandwidth and is further investigated in the simulation section of this paper.

3.1. The new cross-validation method

This paper proposes a new approach for the selection of the bandwidth for the estimation of the long memory parameter. The cross-validation approach provides a clear objective function to optimize for the estimation of the model; and in this instance, is based on the minimization of the $RMSFE$ for out of sample forecasts from a maintained $ARFIMA$ structure. First, an initial SPE of d is applied, the series is then fractionally filtered, and the residuals are modeled in terms of a short memory model, which is a first order linear approximation to the short memory component. Clearly, more complicated nonlinear models could be used in conjunction with the long memory representation; see van Dijk et al. (2002), and Baillie and Kapetanios (2008). The estimated short memory model is then used to produce forecasts of the short memory component which are cumulated to produce forecasts for the observed time series. This procedure is repeated for each forecast horizon and the $RMSFE$ is computed for each possible value for the bandwidth using a grid search.

The initial stage of the cross-validation procedure is based on the observed sample of T; and the number of observations used for cross-validation is denoted by T_{TS}. This notation is fairly standard in the cross-validation literature and T_{TS} denotes the number of observations in the "training set". The maximum forecast horizon is given by h_{max}, and hence the sample size used in the estimation has $T_{IN} = T - T_{TS} - h_{max}$ observations. For each possible value of the bandwidth, m, the estimation and forecasting are based on the sample y_t, $t = 1, \ldots, T_{IN}$. These forecasts, are denoted by $\hat{y}_{t-T_{IN}|T_{IN}}(m)$, with corresponding forecast errors of $\hat{e}_{t-T_{IN}|T_{IN}} = y_t - \hat{y}_{t-T_{IN}|T_{IN}}$, for $t = T_{IN}, \ldots, T_{IN} + h_{max}$. This algorithm is repeated for each forecast horizon to produce $\hat{e}_{t-s|s}(m)$, for $s = T_{IN}, \ldots, T - h_{max}$, and for $t = s + 1, \ldots, s + h_{max}$. The corresponding quantities of $RMSE_h(m) = \left(\frac{1}{T_{TS}}\sum_{s=T_{IN}}^{T-h_{max}} \hat{e}_{h|s}(m)\right)$, are then determined for $h = 1, \ldots, h_{max}$. The cross-validation technique then selects the value of the bandwidth m which minimizes the *RMSE*. Clearly, many different choices of horizon are possible; and alternatively criteria based on some linear combination of the $RMSE_h(m)$.

It is also worth noting that the practical application of the method does not require making assumptions about the structure of the underlying process. A researcher could use a model selection tool based on information criteria in the absence of theoretical parameterizations. Baillie, Kapetanios and Papailias (2014) provide modified order selection criteria that can be used for this purpose in the presence of long memory.

4. Simulation results

The performance of the cross-validation procedure in the context of long memory models was assessed through the following simulation study, which was based on the following data generating processes (*DGP*s):

(A): $ARFIMA(0, d, 0)$ with $d = \{0.4, 0.8\}$,
(B): $ARFIMA(1, d, 0)$ with $d = \{0.4, 0.8\}$ and $\phi = \{0.5, 0.9\}$,
(C): $ARFIMA(2, d, 0)$ with $d = \{0.4, 0.8\}$ and $(\phi_1, \phi_2) = (0.9, -0.5)$.

In all the experiments the parameters were set as $T_{TS} = 12$ and $h_{max} = 24$, with a sample size of $T = 200$ and with 200 replications. In each realization, or replication, all three of the frequency domain based *SPE*s, namely *LW*, *ELW* and *FELW* were implemented, along with the time domain *MDE* method of Mayoral (2007). For each *SPE* the automatic bandwidth choices were $m = \lfloor T^{0.4} \rfloor$, $m = \lfloor T^{0.5} \rfloor$ and $m = \lfloor T^{0.65} \rfloor$; and the maximum lag of the autocorrelation for the *MDE* was $k = \lfloor T^{0.25} \rfloor$, and $k = \lfloor T^{0.5} \rfloor$. Additionally, the bandwidth selection method of Henry (2001) was also used.

In terms of the cross-validation method the bandwidth selection was achieved by a grid search with the range of bandwidths m determined from

$$m \epsilon \left\{ \left[T^{0.3}\right], \left[T^{0.3}\right] + 1, \ldots, \left[T^{0.8}\right] - 2, \left[T^{0.8}\right] - 1 \right\}.$$

Table 1
RMSFE for different forecast horizons and different *SPE*s of the long memory parameter.

$ARFIMA(0, d, 0)$, $d = 0.4$, $T = 200$							
Estimator	Bandwidth	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$
LW	$m = \lfloor T^{0.4} \rfloor$	1.099	1.031	1.027	1.028	1.027	1.025
	$m = \lfloor T^{0.65} \rfloor$	0.954	0.976	0.977	0.976	0.975	0.975
	Automatic	0.946	0.970	0.972	0.973	0.972	0.971
	*Grid*1	0.964	0.976	0.978	0.978	0.978	0.979
	*Grid*2	0.959	0.975	0.975	0.974	0.972	0.971
	*Grid*3	0.964	0.970	0.973	0.982	0.974	0.979
ELW	$m = \lfloor T^{0.4} \rfloor$	1.049	1.010	1.012	1.005	0.998	0.994
	$m = \lfloor T^{0.65} \rfloor$	0.964	0.982	0.987	0.990	0.994	0.993
	Automatic	0.959	0.978	0.985	0.987	0.992	0.992
	*Grid*1	0.974	0.984	0.987	0.985	0.987	0.987
	*Grid*2	0.974	0.984	0.987	0.985	0.987	0.987
	*Grid*3	0.989	0.984	0.987	0.988	0.989	0.992
FELW	$m = \lfloor T^{0.4} \rfloor$	1.165	1.054	1.047	1.045	1.045	1.050
	$m = \lfloor T^{0.65} \rfloor$	0.960	0.976	0.983	0.987	0.987	0.992
	Automatic	0.952	0.969	0.977	0.982	0.982	0.988
	*Grid*1	0.979	0.981	0.984	0.986	0.986	0.994
	*Grid*2	0.965	0.977	0.982	0.982	0.982	0.988
	*Grid*3	0.952	0.969	0.979	0.983	0.983	0.991
MDE	$k = \lfloor T^{0.25} \rfloor$	1.000	1.000	1.000	1.000	1.000	0.999
	$k = \lfloor T^{0.5} \rfloor$	1.001	1.000	1.001	1.001	1.001	1.001
	*Grid*1	1.001	1.000	1.001	1.001	1.001	1.001
	*Grid*2	1.001	1.001	1.001	1.001	1.001	1.000
	*Grid*3	1.001	0.999	1.000	1.000	1.000	1.000

RMSFE relative to $m = \lfloor T^{0.5} \rfloor$ for the Local Whittles and $k = \lfloor T^{1/3} \rfloor$ for the MDE. *Automatic* denotes Henry (2001) bandwidth choice. *Grid*1, *Grid*2 and *Grid*3 denote grid search methods.

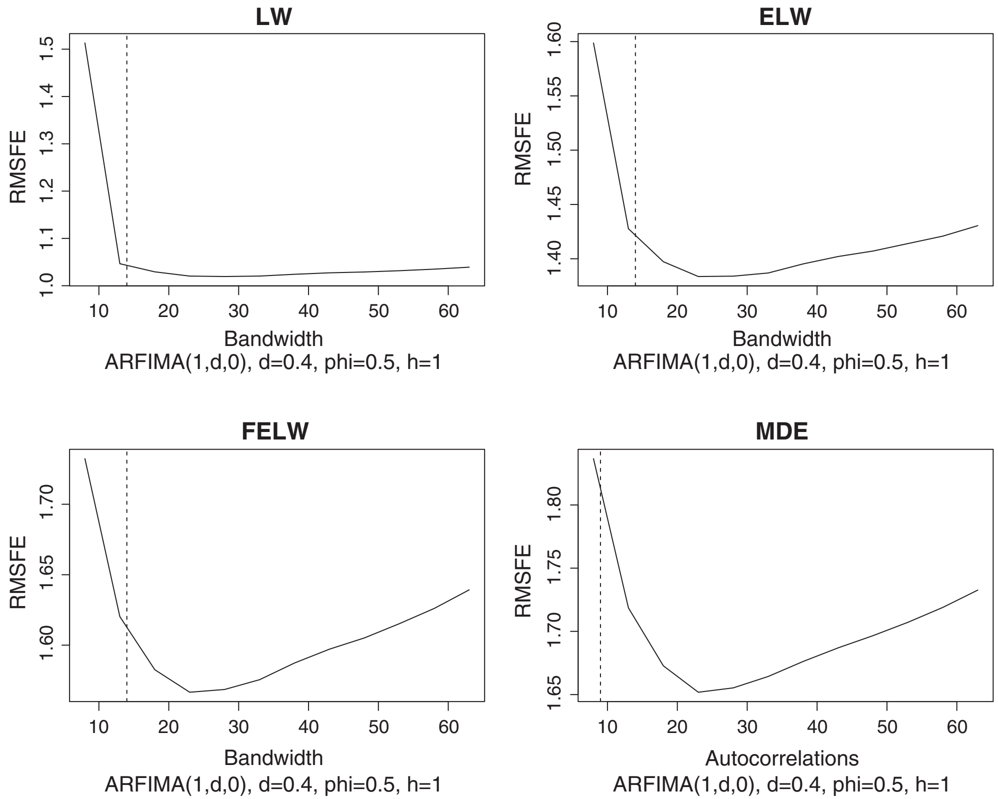

Fig. 1. *ARFIMA*(1, d, 0), $\phi = 0.5$, $d = 0.4$. RMSFE as a function of the bandwidth. Notes. The dashed line denotes the $m = \lfloor T^{0.5} \rfloor$ and $k = \lfloor T^{1/3} \rfloor$ benchmarks. The forecast varies depending on the bandwidth used in the estimation. The benchmarks are shown not to return results with the smallest error.

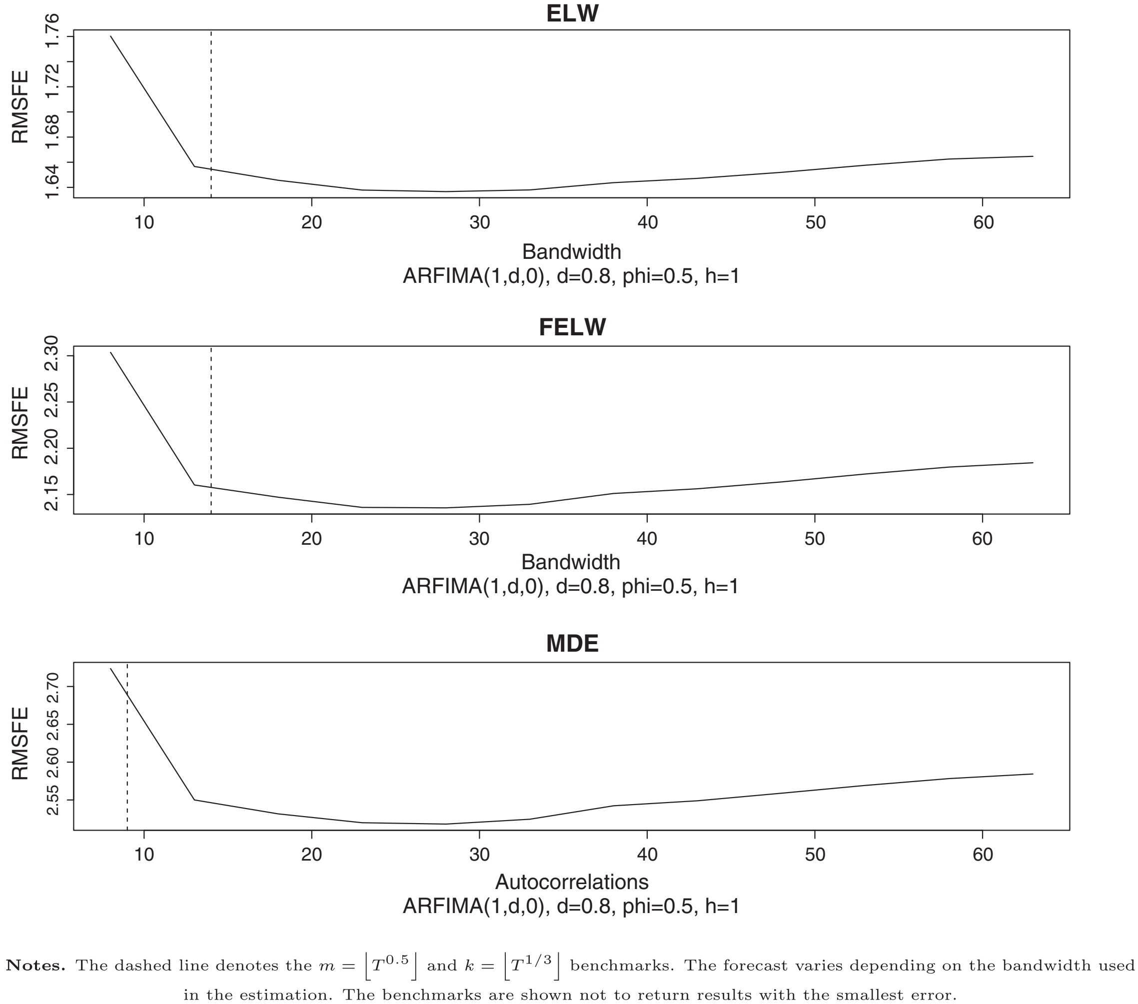

Fig. 2. *ARFIMA*(1, *d*, 0), $\phi = 0.5$, $d = 0.8$. RMSFE as a function of the bandwidth. Notes. The dashed line denotes the $m = \lfloor T^{0.5} \rfloor$ and $k = \lfloor T^{1/3} \rfloor$ benchmarks. The forecast varies depending on the bandwidth used in the estimation. The benchmarks are shown not to return results with the smallest error.

Table 2
RMSFE for different forecast horizons and different *SPE*s of the long memory parameter.

ARFIMA(0, *d*, 0), $d = 0.8$, $T = 200$							
Estimator	Bandwidth	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$
ELW	$m = \lfloor T^{0.4} \rfloor$	1.803	1.137	1.086	1.067	1.065	1.069
	$m = \lfloor T^{0.65} \rfloor$	0.963	0.963	0.963	0.962	0.958	0.959
	Automatic	0.958	0.957	0.958	0.958	0.953	0.953
	*Grid*1	1.001	0.994	0.978	0.964	0.956	0.950
	*Grid*2	1.001	0.994	0.978	0.964	0.956	0.950
	*Grid*3	0.989	0.989	0.962	0.955	0.955	0.956
FELW	$m = \lfloor T^{0.4} \rfloor$	1.543	1.182	1.146	1.133	1.133	1.139
	$m = \lfloor T^{0.65} \rfloor$	0.932	0.938	0.939	0.940	0.940	0.942
	Automatic	0.926	0.931	0.932	0.930	0.930	0.932
	*Grid*1	0.936	0.941	0.944	0.945	0.945	0.947
	*Grid*2	0.967	0.958	0.952	0.939	0.939	0.927
	*Grid*3	0.920	0.923	0.923	0.923	0.923	0.925
MDE	$k = \lfloor T^{0.25} \rfloor$	1.000	1.000	1.000	1.001	1.001	1.001
	$k = \lfloor T^{0.5} \rfloor$	1.000	1.000	1.000	1.000	1.000	0.999
	*Grid*1	0.999	0.998	0.997	0.996	0.996	0.996
	*Grid*2	1.000	0.999	0.998	0.996	0.996	0.994
	*Grid*3	0.999	0.997	0.996	0.995	0.995	0.993

RMSFE relative to $m = \lfloor T^{0.5} \rfloor$ for the Local Whittles and $k = \lfloor T^{1/3} \rfloor$ for the MDE. *Automatic* denotes Henry (2001) bandwidth choice. *Grid*1, *Grid*2 and *Grid*3 denote grid search methods.

Please cite this article as: Baillie, R.T., et al., Bandwidth selection by cross-validation for forecasting long memory financial time series, J. Empir. Finance (2014), http://dx.doi.org/10.1016/j.jempfin.2014.04.002

Table 3
RMSFE for different forecast horizons and different *SPEs* of the long memory parameter.

ARFIMA(1, *d*, 0), $d = 0.4$, $\phi = 0.5$, $T = 200$							
Estimator	Bandwidth	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$
LW	$m = \lfloor T^{0.4} \rfloor$	1.285	1.108	1.086	1.098	1.118	1.148
	$m = \lfloor T^{0.65} \rfloor$	0.959	0.973	0.981	0.980	0.981	0.981
	Automatic	0.966	0.977	0.983	0.982	0.984	0.984
	*Grid*1	0.968	0.977	0.980	0.978	0.983	0.983
	*Grid*2	0.977	0.984	0.985	0.980	0.983	0.986
	*Grid*3	0.968	0.976	0.975	0.977	0.985	0.990
ELW	$m = \lfloor T^{0.4} \rfloor$	1.049	1.043	1.050	1.052	1.062	1.070
	$m = \lfloor T^{0.65} \rfloor$	0.958	0.984	1.002	1.003	1.004	1.005
	Automatic	0.954	0.971	0.984	0.984	0.984	0.987
	*Grid*1	0.986	0.992	1.002	0.997	0.992	0.993
	*Grid*2	0.986	0.992	1.002	0.997	0.992	0.993
	*Grid*3	0.953	0.978	0.989	1.012	1.009	1.015
FELW	$m = \lfloor T^{0.4} \rfloor$	1.453	1.169	1.095	1.075	1.075	1.073
	$m = \lfloor T^{0.65} \rfloor$	0.956	0.975	0.990	0.994	0.994	0.995
	Automatic	0.956	0.969	0.981	0.984	0.984	0.985
	*Grid*1	0.953	0.967	0.978	0.980	0.980	0.984
	*Grid*2	0.982	0.977	0.982	0.975	0.975	0.969
	*Grid*3	0.971	0.978	0.984	0.981	0.981	0.977
MDE	$k = \lfloor T^{0.25} \rfloor$	1.003	1.005	1.004	1.003	1.003	1.000
	$k = \lfloor T^{0.5} \rfloor$	1.015	1.012	1.005	1.001	1.001	0.995
	*Grid*1	1.003	1.001	0.995	0.988	0.988	0.984
	*Grid*2	1.006	1.002	0.994	0.985	0.985	0.976
	*Grid*3	1.003	0.998	0.989	0.988	0.988	0.979

RMSFE relative to $m = \lfloor T^{0.5} \rfloor$ for the Local Whittles and $k = \lfloor T^{1/3} \rfloor$ for the MDE. *Automatic* denotes Henry (2001) bandwidth choice. *Grid*1, *Grid*2 and *Grid*3 denote grid search methods.

Such a grid search turns out to be more practically feasible than numerical optimization procedures. Finally, three different possible criteria were considered for the cross-validation procedure:

- the optimal bandwidth m^* is based on minimizing the *RMSFE* using only $h = 1$.
- the optimal bandwidth m^* is based on minimizing the *RMSFE* using, on average, all values of h.
- the optimal bandwidth m^* is based on minimizing the *RMSFE* for each particular value of h.

Forecasts were produced by each of the above methods over a period of $T_{EVAL} = 60$ observations, with *RMSFE* being calculated for each choice of bandwidth and lag of the autocorrelations.

The results for the fractional white noise in the *DGP* (A) are presented in Table 1 and show that the bandwidth selection makes little difference to the forecast criteria. This is not surprising given the nature of fractional white noise. The results are seen to carry over to nonstationary fractional white noise in (A) with $d = 0.8$; see Table 2.

Table 4
RMSFE for different forecast horizons and different *SPEs* of the long memory parameter.

ARFIMA(1, *d*, 0), $d = 0.8$, $\phi = 0.5$, $T = 200$							
Estimator	Bandwidth	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$
ELW	$m = \lfloor T^{0.4} \rfloor$	1.844	1.424	1.238	1.150	1.127	1.127
	$m = \lfloor T^{0.65} \rfloor$	1.004	1.012	1.014	1.013	1.009	1.005
	Automatic	1.042	1.048	1.051	1.054	1.058	1.059
	*Grid*1	0.946	0.935	0.941	0.938	0.922	0.919
	*Grid*2	0.946	0.935	0.941	0.938	0.922	0.919
	*Grid*3	0.860	0.910	0.925	0.954	0.960	0.948
FELW	$m = \lfloor T^{0.4} \rfloor$	1.422	1.254	1.170	1.130	1.130	1.119
	$m = \lfloor T^{0.65} \rfloor$	0.976	0.980	0.979	0.977	0.977	0.965
	Automatic	1.134	1.105	1.103	1.171	1.171	1.498
	*Grid*1	0.974	0.975	0.974	0.968	0.968	0.952
	*Grid*2	0.980	0.970	0.959	0.937	0.937	0.908
	*Grid*3	0.988	0.976	0.966	0.945	0.945	0.916
MDE	$k = \lfloor T^{0.25} \rfloor$	0.995	0.995	0.994	1.000	1.000	1.012
	$k = \lfloor T^{0.5} \rfloor$	1.016	1.009	1.000	0.994	0.994	1.001
	*Grid*1	0.992	0.982	0.976	0.968	0.968	0.967
	*Grid*2	1.009	0.998	0.985	0.966	0.966	0.958
	*Grid*3	0.992	0.992	0.980	0.970	0.970	0.960

RMSFE relative to $m = \lfloor T^{0.5} \rfloor$ for the Local Whittles and $k = \lfloor T^{1/3} \rfloor$ for the MDE. *Automatic* denotes Henry (2001) bandwidth choice. *Grid*1, *Grid*2 and *Grid*3 denote grid search methods.

Table 5
RMSFE for different forecast horizons and different *SPE*s of the long memory parameter.

ARFIMA(1, *d*, 0), $d = 0.4$, $\phi = 0.9$, $T = 200$							
Estimator	Bandwidth	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$
LW	$m = \lfloor T^{0.4} \rfloor$	1.016	1.005	1.003	1.024	1.055	1.091
	$m = \lfloor T^{0.65} \rfloor$	1.050	1.098	1.131	1.156	1.165	1.163
	Automatic	1.078	1.155	1.209	1.257	1.279	1.283
	*Grid*1	0.987	0.964	0.946	0.928	0.918	0.909
	*Grid*2	0.982	0.958	0.938	0.917	0.904	0.895
	*Grid*3	0.987	0.959	0.939	0.916	0.909	0.901
ELW	$m = \lfloor T^{0.4} \rfloor$	1.035	0.952	0.913	0.900	0.914	0.938
	$m = \lfloor T^{0.65} \rfloor$	1.072	1.143	1.197	1.255	1.284	1.297
	Automatic	1.133	1.277	1.402	1.557	1.653	1.713
	*Grid*1	0.969	0.925	0.892	0.856	0.836	0.823
	*Grid*2	0.969	0.925	0.892	0.856	0.836	0.823
	*Grid*3	0.963	0.921	0.888	0.850	0.840	0.822
FELW	$m = \lfloor T^{0.4} \rfloor$	0.997	0.964	0.948	0.950	0.950	0.981
	$m = \lfloor T^{0.65} \rfloor$	1.047	1.090	1.118	1.144	1.144	1.162
	Automatic	1.047	1.087	1.109	1.126	1.126	1.126
	*Grid*1	0.969	0.935	0.909	0.881	0.881	0.854
	*Grid*2	0.976	0.941	0.913	0.884	0.884	0.855
	*Grid*3	0.866	0.747	0.653	0.553	0.553	0.462
MDE	$k = \lfloor T^{0.25} \rfloor$	0.997	0.992	0.989	0.986	0.986	0.988
	$k = \lfloor T^{0.5} \rfloor$	0.998	0.996	0.995	0.991	0.991	0.989
	*Grid*1	0.992	0.985	0.984	0.983	0.983	0.982
	*Grid*2	0.992	0.986	0.983	0.976	0.976	0.971
	*Grid*3	0.992	0.985	0.983	0.976	0.976	0.969

RMSFE relative to $m = \lfloor T^{0.5} \rfloor$ for the Local Whittles and $k = \lfloor T^{1/3} \rfloor$ for the MDE. *Automatic* denotes Henry (2001) bandwidth choice. *Grid*1, *Grid*2 and *Grid*3 denote grid search methods.

The results from experiment (B) with the *ARFIMA*(1, *d*, 0) process with $\phi = 0.5$ and $d = 0.4$ are graphically displayed in Fig. 1. Now the minimum *RMSFE* for one-step ahead forecasts was obtained by using $m = 23$ compared with $m = 14$, which is obtained by the $T^{0.5}$ standard bandwidth selection approach. In this case the cross-validation methods are close in performance to the *Automatic* bandwidth selection method. Looking at Table 3 we see that the *LW* in *Grid*3 provides the same or slightly smaller *RMSFE* than the *Automatic* method for $h = 1, 3, 6, 12$; while *Grid*3 has a superior performance for the *ELW* and *Grid*1 provides better forecasts using the *FELW* estimator. It can also be seen that the cross-validation approach provides relatively better forecasting performance using the *MDE*. However, this is the case only in longer forecasting horizons, particularly for $h = \{12, \ 18, \ 24\}$.

For the *DGP* of an *ARFIMA*(1, *d*, 0) with $\phi = 0.5$ and $d = 0.8$ the results are reported in Table 4, and in this nonstationary long memory case, the superiority of the new cross-validation methodologies becomes clear. For one step ahead forecasts using *ELW*, the *Grid*3 method provides an *RSMFE* of 0.86 relative to the $m = \lfloor T^{0.5} \rfloor$ benchmark. Similarly, using the *FELW* the *Grid*1 provides a

Table 6
RMSFE for different forecast horizons and different *SPE*s of the long memory parameter.

ARFIMA(1, *d*, 0), $d = 0.8$, $\phi = 0.9$, $T = 200$							
Estimator	Bandwidth	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$
ELW	$m = \lfloor T^{0.4} \rfloor$	0.970	0.953	0.956	0.994	1.050	1.118
	$m = \lfloor T^{0.65} \rfloor$	1.078	1.144	1.197	1.252	1.283	1.309
	Automatic	1.156	1.311	1.454	1.634	1.754	1.852
	*Grid*1	0.961	0.927	0.898	0.868	0.848	0.836
	*Grid*2	0.961	0.927	0.898	0.868	0.848	0.836
	*Grid*3	0.967	0.924	0.894	0.863	0.848	0.844
FELW	$m = \lfloor T^{0.4} \rfloor$	0.979	0.964	0.965	0.980	0.980	1.023
	$m = \lfloor T^{0.65} \rfloor$	1.059	1.105	1.137	1.165	1.165	1.187
	Automatic	1.071	1.126	1.167	1.199	1.199	1.222
	*Grid*1	0.967	0.937	0.912	0.885	0.885	0.852
	*Grid*2	0.960	0.921	0.886	0.845	0.845	0.792
	*Grid*3	1.421	1.244	1.047	0.841	0.841	0.644
MDE	$k = \lfloor T^{0.25} \rfloor$	1.000	0.999	1.000	1.001	1.001	1.003
	$k = \lfloor T^{0.5} \rfloor$	0.954	0.959	0.964	0.973	0.973	0.982
	*Grid*1	0.843	0.856	0.876	0.906	0.906	0.941
	*Grid*2	0.855	0.865	0.879	0.897	0.897	0.919
	*Grid*3	0.843	0.856	0.873	0.897	0.897	0.947

RMSFE relative to $m = \lfloor T^{0.5} \rfloor$ for the Local Whittles and $k = \lfloor T^{1/3} \rfloor$ for the MDE. *Automatic* denotes Henry (2001) bandwidth choice. *Grid*1, *Grid*2 and *Grid*3 denote grid search methods.

Table 7
RMSFE for different forecast horizons and different *SPEs* of the long memory parameter.

ARFIMA(2, d, 0), $d = 0.4$, $\phi_1 = 0.9$, $\phi_2 = -0.5$, $T = 200$							
Estimator	Bandwidth	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$
LW	$m = \lfloor T^{0.4} \rfloor$	1.114	1.085	1.045	1.059	1.071	1.079
	$m = \lfloor T^{0.65} \rfloor$	0.977	0.977	0.988	0.988	0.990	0.989
	Automatic	0.994	0.993	0.997	0.997	0.998	0.999
	*Grid*1	0.994	0.996	1.010	1.018	1.017	1.017
	*Grid*2	0.979	0.988	1.006	1.010	1.010	1.014
	*Grid*3	0.994	0.986	1.003	1.007	1.004	1.017
ELW	$m = \lfloor T^{0.4} \rfloor$	1.570	1.126	1.089	1.045	1.046	1.051
	$m = \lfloor T^{0.65} \rfloor$	0.976	0.985	0.998	0.999	0.997	0.996
	Automatic	1.177	1.024	1.018	0.998	0.996	0.998
	*Grid*1	1.004	1.068	1.110	1.178	1.228	1.282
	*Grid*2	1.004	1.068	1.110	1.178	1.228	1.282
	*Grid*3	1.001	1.044	1.266	1.423	1.503	1.607
FELW	$m = \lfloor T^{0.4} \rfloor$	1.109	1.085	1.036	1.045	1.045	1.054
	$m = \lfloor T^{0.65} \rfloor$	0.974	0.987	1.002	1.002	1.002	0.999
	Automatic	0.996	0.998	0.999	0.999	0.999	0.999
	*Grid*1	0.965	0.993	1.026	1.034	1.034	1.036
	*Grid*2	0.971	0.998	1.016	1.020	1.020	1.016
	*Grid*3	0.944	0.958	0.995	0.997	0.997	0.995
MDE	$k = \lfloor T^{0.25} \rfloor$	1.083	1.087	1.028	1.032	1.032	1.026
	$k = \lfloor T^{0.5} \rfloor$	0.992	0.997	0.999	1.002	1.002	1.003
	*Grid*1	0.993	0.994	0.996	0.997	0.997	0.998
	*Grid*2	0.994	0.994	0.995	0.995	0.995	0.997
	*Grid*3	0.993	0.992	0.995	0.995	0.995	0.998

RMSFE relative to $m = \lfloor T^{0.5} \rfloor$ for the Local Whittles and $k = \lfloor T^{1/3} \rfloor$ for the MDE. *Automatic* denotes Henry (2001) bandwidth choice. *Grid*1, *Grid*2 and *Grid*3 denote grid search methods.

relative *RSMFE* of 0.974 whereas the $m = \lfloor T^{0.65} \rfloor$ and *Automatic* frequencies give a relative *RMSFE* of 0.976 and 1.134 respectively. The better performance of the cross-validation approach is also apparent for other forecast horizons, for both *ELW* and *FELW*. The graphical illustration of the RMSFE as a function of the bandwidth can be seen in Fig. 2.

As seen in Table 5, the cross-validation approach also has a superior performance for the *ARFIMA*(1, d, 0) case with $\phi = 0.9$. For one step ahead prediction *Grid*3 using the *FELW* has a relative *RMSFE* of only 0.866 and also works extremely well for longer forecast horizons. Similar qualitative results are presented in Table 6 where $\phi = 0.9$ but now $d = 0.8$. These qualitative conclusions carry over to the experiment using the *DGP* in design (C), which is an *ARFIMA*(2, d, 0) with $(\phi_1, \phi_2) = (0.9, -0.5)$. Table 7 reports the results using $d = 0.4$ and Table 8 reports the results using $d = 0.8$.

Table 8
RMSFE for different forecast horizons and different *SPEs* of the long memory parameter.

ARFIMA(2, d, 0), $d = 0.8$, $\phi_1 = 0.9$, $\phi_2 = -0.5$, $T = 200$							
Estimator	Bandwidth	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$
ELW	$m = \lfloor T^{0.4} \rfloor$	1.235	1.161	1.075	1.048	1.073	1.103
	$m = \lfloor T^{0.65} \rfloor$	0.861	0.888	0.929	0.898	0.915	0.928
	Automatic	0.892	0.918	0.948	0.914	0.929	0.941
	*Grid*1	0.894	0.974	1.078	1.149	1.264	1.367
	*Grid*2	0.894	0.974	1.078	1.149	1.264	1.367
	*Grid*3	0.894	0.940	1.072	1.192	1.375	1.588
FELW	$m = \lfloor T^{0.4} \rfloor$	1.416	1.245	1.101	1.106	1.106	1.093
	$m = \lfloor T^{0.65} \rfloor$	0.956	0.961	0.975	0.981	0.981	0.989
	Automatic	0.993	0.993	0.994	0.993	0.993	0.994
	*Grid*1	0.939	0.957	0.995	1.030	1.030	1.051
	*Grid*2	0.965	0.981	0.996	1.002	1.002	0.996
	*Grid*3	0.933	0.937	0.959	0.968	0.968	0.966
MDE	$k = \lfloor T^{0.25} \rfloor$	1.229	1.216	1.137	1.136	1.136	1.119
	$k = \lfloor T^{0.5} \rfloor$	0.914	0.923	0.955	0.962	0.962	0.965
	*Grid*1	0.907	0.915	0.949	0.956	0.956	0.955
	*Grid*2	0.909	0.915	0.946	0.951	0.951	0.951
	*Grid*3	0.907	0.917	0.946	0.954	0.954	0.957

RMSFE relative to $m = \lfloor T^{0.5} \rfloor$ for the Local Whittles and $k = \lfloor T^{1/3} \rfloor$ for the MDE. *Automatic* denotes Henry (2001) bandwidth choice. *Grid*1, *Grid*2 and *Grid*3 denote grid search methods.

Please cite this article as: Baillie, R.T., et al., Bandwidth selection by cross-validation for forecasting long memory financial time series, J. Empir. Finance (2014), http://dx.doi.org/10.1016/j.jempfin.2014.04.002

Table 9
Root mean squared error of estimators using various frequencies relative to $m = \lfloor T^{0.5} \rfloor$.

		$m = \lfloor T^{0.4} \rfloor$	$m = \lfloor T^{0.65} \rfloor$	*Grid*1	*Grid*2	*Grid*3
ARFIMA(1, *d*, 0), $d = 0.4$, $\phi = 0.9$						
$\hat{d}$	*LW*	0.890	1.249	0.819	0.810	0.819
	ELW	0.870	1.289	0.838	0.819	0.838
	FELW	0.910	1.205	0.868	0.856	0.868
$\hat{\phi}$	*LW*	0.877	1.264	0.807	0.797	0.807
	ELW	0.855	1.304	0.830	0.815	0.830
	FELW	0.899	1.221	0.866	0.856	0.866
ARFIMA(1, *d*, 0), $d = 0.8$,$\phi = 0.9$						
$\hat{d}$	*ELW*	0.870	1.281	0.847	0.842	0.847
	FELW	0.888	1.244	0.851	0.842	0.851
$\hat{\phi}$	*ELW*	0.855	1.299	0.847	0.843	0.847
	FELW	0.866	1.263	0.846	0.838	0.846

*Grid*1, *Grid*2 and *Grid*3 denote grid search methods.

The reasons for the superior forecasting performance of the cross-validation methods can be seen in terms of the properties of the estimation of the parameters d and ϕ. In further simulation work with a sample size of $T = 200$ and 500 replications, Table 9 reports the root mean squared error, *RMSE*, for the estimators of d and ϕ using $m = \lfloor T^{0.4} \rfloor$, $m = \lfloor T^{0.65} \rfloor$ and the optimal frequencies obtained from the three cross-validation methods. The top panel of Table 9 considers an *ARFIMA*(1, *d*, 0) *DGP* with $d = 0.4$ and $\phi = 0.9$ whereas the bottom panel of Table 9 considers an *ARFIMA*(1, *d*, 0) DGP with $d = 0.8$ and $\phi = 0.9$. All results are relative to the *RMSE* of the estimators obtained using $m = [T^{0.5}]$. It is clear that the optimal bandwidths obtained by the three cross-validation methods provide parameter estimators with the smallest *RMSE*.

5. Empirical application

There have been several different modeling strategies to assess the persistence of real exchange rates, including forms of long memory models; for example see Diebold, Husted and Rush (1991), Baum, Barkulas and Caglayan (1999), Dufrenot et al. (2008), Chortareas, Kapetanios and Shin (2002) and Chortareas and Kapetanios (2004). This paper uses data on real effective exchange rates (*REER*) for the *UK* and the *US*. The data is provided by *EUROSTAT* and the *REER* series measures the nominal effective exchange rate series deflated by nominal unit labor costs. The data consists of 229 monthly observations from January, 1994 through December, 2012; and in this study the last 60 observations are used for evaluating the forecasting methods in the same manner as in the simulations. The *Automatic* bandwidth choice and the *LW* are excluded from the empirical work. Similarly to the simulations, the method uses $T_{TS} = 20$[1] and $h_{\max} = 24$ in the grid searches and $T_{EVAL} = 60$ and $h_{max} = 24$ in the forecasting exercise. The possible models of an *ARFIMA*(0, *d*, 0), *ARFIMA*(1, *d*, 0) and *ARFIMA*(2, *d*, 0) are all considered, and Table 10 reports results indicating that the cross-validation methods and the standard choice of $m = \lfloor T^{0.65} \rfloor$ result in better forecasts compared to the $m = \lfloor T^{0.5} \rfloor$ benchmark using both *ELW* and *FELW*. For longer forecasting horizons of $h = 12, 18, 24$, the cross-validation methods are clearly superior compared to the automatic methods. *Grid*2 provides an *RMSFE* as low as 0.877. The use of other models such as the *ARFIMA*(1, *d*, 0) or *ARFIMA*(2, *d*, 0) exhibit the same qualitative results associated with the *ELW* and *FELW* estimators. *Grid*1, *Grid*2, *Grid*3 and $m = [T^{0.65}]$ methods perform better compared to the benchmark, while the *MDE* case fails to provide any significant improvements.

Analogous results for the *US REER* series are provided in Table 11 and the *ELW* estimator together with the cross-validation bandwidths provide better forecasts compared to the benchmark and in general to other choices of *m*. This is particularly true for the higher order *ARFIMA*(2, d, 0) model. There is also some evidence of further improvement by using the *FELW* estimator.

6. Conclusions

This paper has addressed the issue of bandwidth selection for semiparametric estimators of the long memory parameter in a univariate time series process. The primary focus is on the properties of forecasts from the long memory model. We have proposed a variety of cross-validation methods for the selection of the bandwidth and the subsequent model selection and forecasting. It is found that the use of common, automatic bandwidth choices for the estimation of the long memory parameter generally leads to suboptimal model selection and hence subsequent forecasting. Our proposed cross-validation methods appear, in both simulations and empirical experiments, to generally improve, and certainly never deteriorate, the quality of model selection and out of sample forecasting performance across most forecast horizons. We conclude that the cross-validation technique proposed in this paper can be a valuable method for analyzing long memory time series.

Please cite this article as: Baillie, R.T., et al., Bandwidth selection by cross-validation for forecasting long memory financial time series, J. Empir. Finance (2014), http://dx.doi.org/10.1016/j.jempfin.2014.04.002

Table 10
UK real effective exchange rate relative to $m=\lfloor T^{0.5}\rfloor$ for the Local Whittles and $k=\lfloor T^{1/3}\rfloor$ for the MDE.

Real effective exchange rate *UK*

Model	Bandwidth	*ELW*						*FELW*						*MDE*					
		1	3	6	12	18	24	1	3	6	12	18	24	1	3	6	12	18	24
ARFIMA(0, *d*, 0)	$m=\lfloor T^{0.4}\rfloor$	1.004	1.003	1.014	1.020	1.019	1.013	1.009	1.008	1.059	1.041	1.054	1.046	1.003	1.004	1.003	1.002	1.000	0.999
	$m=\lfloor T^{0.65}\rfloor$	0.992	0.967	0.982	0.955	0.926	0.899	0.978	0.934	0.925	0.917	0.873	0.848	1.002	1.002	1.002	1.003	1.004	1.005
	*Grid*1	0.985	0.965	0.990	0.939	0.951	0.941	0.978	0.936	0.936	0.902	0.895	0.878	1.004	0.999	0.989	0.989	0.968	0.959
	*Grid*2	0.988	0.960	0.987	0.921	0.923	0.906	0.974	0.931	0.941	0.894	0.877	0.859	1.002	0.996	0.985	0.983	0.961	0.948
	*Grid*3	0.985	0.965	0.980	0.919	0.920	0.906	0.978	0.929	0.919	0.894	0.880	0.859	1.004	0.998	0.987	0.985	0.961	0.948
ARFIMA(1, *d*, 0)	$m=\lfloor T^{0.4}\rfloor$	0.993	0.998	1.008	1.020	1.019	1.015	1.009	1.017	1.064	1.052	1.062	1.054	0.998	0.989	0.962	0.926	0.897	0.871
	$m=\lfloor T^{0.65}\rfloor$	0.972	0.951	0.966	0.944	0.921	0.896	0.950	0.908	0.899	0.898	0.863	0.841	1.011	1.014	1.008	1.013	1.020	1.018
	*Grid*1	0.987	0.966	0.985	0.926	0.933	0.912	0.968	0.922	0.922	0.883	0.882	0.868	1.025	1.036	1.032	1.058	1.077	1.082
	*Grid*2	0.980	0.957	0.981	0.916	0.923	0.907	0.951	0.907	0.918	0.879	0.870	0.854	1.022	1.033	1.029	1.055	1.075	1.081
	*Grid*3	0.987	0.964	0.985	0.915	0.918	0.906	0.968	0.917	0.915	0.882	0.872	0.854	1.025	1.035	1.030	1.057	1.075	1.081
ARFIMA(2, *d*, 0)	$m=\lfloor T^{0.4}\rfloor$	0.991	0.994	1.004	1.019	1.020	1.015	1.011	1.024	1.074	1.072	1.077	1.069	0.999	0.993	0.973	0.955	0.934	0.925
	$m=\lfloor T^{0.65}\rfloor$	0.972	0.946	0.957	0.937	0.919	0.896	0.948	0.896	0.879	0.878	0.853	0.835	1.015	1.023	1.022	1.024	1.032	1.034
	*Grid*1	0.982	0.961	0.986	0.931	0.937	0.915	0.968	0.916	0.907	0.870	0.876	0.865	1.018	1.019	1.005	1.012	0.996	0.981
	*Grid*2	0.981	0.956	0.975	0.914	0.925	0.911	0.951	0.899	0.903	0.864	0.863	0.850	1.016	1.016	1.002	1.009	0.994	0.980
	*Grid*3	0.982	0.963	0.979	0.913	0.919	0.904	0.968	0.903	0.906	0.870	0.865	0.849	1.018	1.018	1.004	1.011	0.994	0.980

*Grid*1, *Grid*2 and *Grid*3 denote grid search methods.

Please cite this article as: Baillie, R.T., et al., Bandwidth selection by cross-validation for forecasting long memory financial time series, J. Empir. Finance (2014), http://dx.doi.org/10.1016/j.jempfin.2014.04.002

Table 11
US real effective exchange rate relative to $m = \lfloor T^{0.5} \rfloor$ for the Local Whittles and $k = \lfloor T^{1/3} \rfloor$ for the MDE.

Real effective exchange rate *US*																			
		ELW						*FELW*						*MDE*					
Model	Bandwidth	1	3	6	12	18	24	1	3	6	12	18	24	1	3	6	12	18	24
ARFIMA(0, *d*, 0)	$m = \lfloor T^{0.4} \rfloor$	0.991	0.994	1.006	1.058	1.105	1.093	1.038	1.090	1.116	1.266	1.337	1.298	1.004	1.004	1.005	1.007	1.012	1.016
	$m = \lfloor T^{0.65} \rfloor$	1.011	0.995	0.957	0.911	0.874	0.827	1.012	0.988	0.937	0.852	0.803	0.755	0.998	0.999	1.002	1.012	1.013	1.025
	*Grid*1	1.013	1.007	1.015	1.043	1.061	1.028	1.026	1.011	0.995	0.985	0.979	0.987	1.008	1.030	1.010	1.028	1.051	1.054
	*Grid*2	1.002	0.977	0.953	0.935	0.953	0.919	1.018	0.990	0.933	0.855	0.839	0.806	1.001	1.001	0.999	0.998	1.004	1.017
	*Grid*3	1.013	1.011	0.967	0.935	0.960	0.926	1.026	0.980	0.941	0.860	0.837	0.816	1.008	1.000	0.991	0.998	1.002	1.008
ARFIMA(1, *d*, 0)	$m = \lfloor T^{0.4} \rfloor$	1.025	1.057	1.052	1.109	1.242	1.223	1.083	1.167	1.186	1.332	1.456	1.409	1.060	1.126	1.164	1.377	1.920	2.008
	$m = \lfloor T^{0.65} \rfloor$	1.001	0.987	0.953	0.909	0.870	0.826	0.994	0.968	0.928	0.847	0.794	0.751	1.005	1.017	1.046	1.105	1.119	1.098
	*Grid*1	0.998	0.992	0.979	0.939	0.924	0.953	0.998	0.970	0.926	0.843	0.822	0.799	1.033	1.093	1.171	1.397	1.564	1.505
	*Grid*2	0.999	0.983	0.954	0.917	0.902	0.876	0.995	0.968	0.923	0.844	0.820	0.796	1.043	1.096	1.181	1.406	1.578	1.532
	*Grid*3	0.998	0.986	0.956	0.925	0.913	0.853	0.998	0.964	0.921	0.851	0.821	0.808	1.036	1.095	1.188	1.412	1.579	1.520
ARFIMA(2, *d*, 0)	$m = \lfloor T^{0.4} \rfloor$	1.022	1.047	1.047	1.105	1.223	1.209	1.080	1.147	1.162	1.310	1.410	1.364	0.992	0.971	0.938	0.861	0.739	0.772
	$m = \lfloor T^{0.65} \rfloor$	1.003	0.992	0.954	0.909	0.872	0.826	0.997	0.977	0.932	0.849	0.800	0.754	1.003	1.006	1.018	1.043	1.062	1.088
	*Grid*1	0.999	0.998	0.983	0.941	0.924	0.952	1.001	0.975	0.927	0.848	0.835	0.806	1.017	1.034	1.052	1.103	1.151	1.179
	*Grid*2	0.998	0.978	0.949	0.926	0.935	0.892	0.999	0.978	0.927	0.849	0.831	0.803	1.019	1.034	1.052	1.097	1.148	1.194
	*Grid*3	0.999	0.986	0.956	0.931	0.917	0.855	1.001	0.972	0.929	0.851	0.831	0.814	1.017	1.034	1.057	1.099	1.147	1.186

*Grid*1, *Grid*2 and *Grid*3 denote grid search methods.

References

Abadir, K.M., Distaso, W., Giraitis, L., 2007. Non-stationarity extended local Whittle estimation. J. Econ. 141, 1353–1384.
Abadir, K.M., Distaso, W., Giraitis, L., 2011. An I(d) model with trend and cycles. J. Econ. 163, 186–199.
Asai, M., McAleer, M., 2012. A fractionally integrated Wishart stochastic volatility model. Soka University Working Paper.
Baillie, R.T., 1996. Long memory processes and fractional integration in econometrics. J. Econ. 73, 5–59.
Baillie, R.T., Chung, H., 2001. Estimation of GARCH models from the autocorrelations of the squares of a process. J. Time Ser. Anal. 22, 631–650.
Baillie, R.T., Kapetanios, G., 2008. Nonlinear models for strongly dependent processes with financial applications. J. Econ. 147, 60–71.
Baillie, R.T., Kapetanios, G., 2013. Estimation and inference for impulse response functions from univariate strongly persistent processes. Econ. J. 16, 373–399.
Baillie, R.T., Kapetanios, G., Papailias, F., 2014. Modified information criteria and selection of long memory time series models. Comput. Stat. Data Anal. (forthcoming) (http://www.sciencedirect.com/science/article/pii/S0167947313001539).
Baum, C., Barkulas, J.T., Caglayan, M., 1999. Long memory or structural breaks: can either explain nonstationary real exchange rates under the current float? J. Int. Financ. Mark. Inst. Money 9–4, 359–376.
Bisaglia, L., Guegan, M., 1998. A comparison of techniques of estimation in long-memory processes. Comput. Stat. Data Anal. 27, 61–81.
Chortareas, G., Kapetanios, G., Shin, Y., 2002. Nonlinear mean reversion in real exchange rates. Econ. Lett. 77, 411–417.
Chortareas, G., Kapetanios, G., 2004. The yen real exchange rate may be stationary after all: evidence from nonlinear unit root tests. Oxf. Bull. Econ. Stat. 66, 113–131.
Dalla, V., Giraitis, L., Hidalgo, J., 2005. Consistent estimation of the long memory parameter for nonlinear time series. J. Time Ser. Anal. 27, 211–251.
Diebold, F.X., Husted, S., Rush, M., 1991. Real exchange rates under the gold standard. J. Polit. Econ. 99, 1252–1271.
Dufrenot, G., Lardic, S., Mathieu, L., Mignon, V., Peguin-Feissolle, A., 2008. Explaining the European exchange rates deviations: long memory or non-linear adjustment? J. Int. Financ. Mark. Inst. Money 207–215 (18-3?).
Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. J. Time Ser. Anal. 4, 221–238.
Granger, C.W.J., 1980. Long memory relationships and the aggregation of dynamic models. J. Econ. 14, 227–238.
Granger, C.W.J., Joyeux, R., 1980. An introduction to long memory time series models and fractional differencing. J. Time Ser. Anal. 1, 15–39.
Haldrup, N., Nielsen, M.O., 2007. Estimation of fractional integration in the presence of data noise. Comput. Stat. Data Anal. 51, 3100–3114.
Hall, P., Koul, H.L., Turlach, B.A., 1997. Note on convergence rates of semiparametric estimators of dependence index. Ann. Stat. 25, 1725–1739.
Henry, M., 2001. Robust automatic bandwidth for long memory. J. Time Ser. Anal. 22–3, 293–316.
Hosking, J.R.M., 1981. Fractional differencing. Biometrika 65, 165–176.
Hosking, J.R.M., 1996. Asymptotic distribution of the sample mean, autocovariances and autocorrelations of long memory time series. J. Econ. 73, 261–284.
Koopmans, S.J., Carnera, A., Ooms, M., 2007. Periodic seasonal Reg-*ARFIMA-GARCH* models for daily electricity prices. J. Am. Stat. Assoc. 102, 16–27.
Mayoral, L., 2007. Minimum distance estimation of stationary and non-stationary *ARFIMA* processes. Econ. J. 10, 124–148.
Phillips, P.C.B., 2007. Unit root periodogram. J. Econ. 138, 104–124.
Robinson, P.M., 1995. Gaussian semiparametric estimation of long range dependence. Ann. Stat. 23, 1630–1661.
Shimotsu, K., Phillips, P.C.B., 2005. Exact local Whittle estimation of fractional integration. Ann. Stat. 33, 1890–1933.
Shimotsu, K., Phillips, P.C.B., 2006. Local Whittle estimation of fractional integration and some of its variants. J. Econ. 130, 209–233.
Tieslau, M.A., Schmidt, P., Baillie, R.T., 1996. A minimum distance estimator for long-memory processes. J. Econ. 71, 249–264.
van Dijk, D., Frances, P.H., Paap, R., 2002. A nonlinear long memory model with applications to US unemployment. J. Econ. 110, 135–165.
Zevallos, M., Palma, W., 2013. Minimum distance estimation of ARFIMA processes. Comput. Stat. Data Anal. 58, 242–256.

Please cite this article as: Baillie, R.T., et al., Bandwidth selection by cross-validation for forecasting long memory financial time series, J. Empir. Finance (2014), http://dx.doi.org/10.1016/j.jempfin.2014.04.002