Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Econometrics 131 (2006) 539–578
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series

Geetesh Bhardwaj, Norman R. Swanson*

Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901-1248, USA

Available online 25 February 2005

Abstract

This paper addresses the notion that many fractional $I(d)$ processes may fall into the "empty box" category, as discussed in Granger (Aspects of research strategies for time series analysis, Presentation to the Conference on New Developments in Time Series Economics, Yale University, 1999). We present ex ante forecasting evidence which suggests that ARFIMA models estimated using a variety of standard estimation procedures yield "approximations" to the true unknown underlying DGPs that sometimes provide significantly better out-of-sample predictions than AR, MA, ARMA, GARCH, and related models, based on analysis of point mean-square forecast errors (MSFEs), and based on the use of predictive accuracy tests. The strongest evidence in favor of ARFIMA models arises when various transformations of 5 major stock index returns are examined. Additional evidence based on analysis of the Stock and Watson (J. Bus. Econom. Stat. 20 (2002) 147–162) data set, the returns series data set examined by Ding et al. (J. Empirical Finance 1 (1993) 83–106), and based on a series of Monte Carlo experiments is also discussed.

JEL classification: C15; C22; C53

Keywords: Fractional integration; Long memory; Parameter estimation error; Stock returns; Long horizon prediction

*Corresponding author. Tel.: +1 732 932 7432; fax: +1 732 932 7416.
E-mail addresses: bhardwaj@econ.rutgers.edu (G. Bhardwaj), nswanson@econ.rutgers.edu (N.R. Swanson).

0304-4076/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2005.01.016

1. Introduction

The last 2 decades of macro- and financial economic research has resulted in a vast array of important contributions in the area of long-memory modelling, both from a theoretical and an empirical perspective. From a theoretical perspective, much effort has focussed on issues of testing and estimation, and a very few important contributions include Granger (1980), Granger and Joyeux (1980), Hosking (1981), Geweke and Porter-Hudak (1983), Lo (1991), Sowell (1992a, b), Ding et al. (1993), Cheung and Diebold (1994), Robinson (1995a,b), Engle and Smith (1999), Diebold and Inoue (2001), Breitung and Hassler (2002), and Dittman and Granger (2002). The empirical analysis of long-memory models has seen equally impressive treatment, including studies by Diebold and Rudebusch (1989, 1991a, b), Hassler and Wolters (1995), Hyung and Franses (2001), Bos et al. (2002), Chio and Zivot (2002), and van Dijk et al. (2002), to name but a few.[1] The impressive array of papers on the subject is perhaps not surprising, given that long-memory models in economics is one of the many important areas of research that has stemmed from seminal contributions made by Clive W.J. Granger (see e.g. Granger, 1980; Granger and Joyeux, 1980). Indeed, in the write-up disseminated by the Royal Swedish Academy of Sciences upon announcement that Clive W.J. Granger and Robert F. Engle had won the 2003 Nobel Prize in Economics, it was stated that[2]

> Granger has left his mark in a number of areas. [Other than in the development of the concept of cointegration]. His development of a testable definition of causality (Granger, 1969) has spawned a vast literature. He has also contributed to the theory of so-called long-memory models that have become popular in the econometric literature (Granger and Joyeux, 1980). Furthermore, Granger was among the first to consider the use of spectral analysis (Granger and Hatanaka, 1964) as well as non-linear models (Granger and Andersen, 1978) in research on economic time series.

This paper attempts to add to the wealth of literature on the topic by asking a number of questions related to prediction using long-memory models, and by presenting some new empirical evidence.

First, as pointed out by many authors, including Diebold and Inoue (2001), Engle and Smith (1999), and Granger and Hyung (1999), so-called spurious long-memory (i.e. when in-sample tests find long-memory even when there is none) arises in many contexts, such as when there are (stochastic) structural breaks in linear and non-linear models, in the context of regime switching models, and when forming models using variables that are (simple) non-linear transformations of underlying "short-memory" variables. The spurious long-memory feature has been illustrated

[1] Many other empirical and theoretical studies are referenced in the extensive survey paper by Baillie (1996).

[2] See list of references under "Bank of Sweden (2003)" for a reference to the document.

convincingly using theoretical, empirical, and experimental arguments in the above papers. Bhardwaj and Swanson (2003) add to the evidence by showing, via Monte Carlo experimentation, that spurious long-memory may arise if reliance is placed on any of 5 standard tests of short-memory, even if the true data generating processes (DGPs) are linear with no data transformation, structural breaks, and/or regime switching properties. In the current paper, we confirm these finding via predictive analysis. In particular, three different data sets due to Ding et al. (1993), Stock and Watson (2002) and Leybourne et al. (2003) are examined, and it is shown that standard short-memory tests find ample evidence of long-memory, even when ex ante prediction analysis indicates that ARFIMA models constructed using 4 different estimators of the differencing parameter, d, are inferior to various AR, MA, ARMA, GARCH, and related models, where the term inferior is meant to denote that one model outperforms another, based on point mean square out-of-sample forecast error (MSFE) comparison (using Diebold and Mariano (DM), 1995) predictive accuracy tests).

Second, there has been little evidence in the literature supporting the usefulness of long-memory models for prediction. In a discussion of this and related issues, for example, Granger (1999) acknowledges the importance of outliers, breaks, and undesirable distributional properties in the context of long-memory models, and concludes that there is a good case to be made for $I(d)$ processes falling into the "empty box" category (i.e. ARFIMA models have stochastic properties that essentially do not mimic the properties of the data). We attempt to stem the tide of negative evidence by presenting ex ante forecasting evidence based on various financial and macroeconomic data sets. One is an updated version of the absolute returns series examined by Ding et al. (DGE, 1993) and Granger and Ding (1996). Evidence based on analysis of this very large data set suggests that ARFIMA models estimated using a variety of standard estimation procedures yield "approximations" to the true unknown underlying DGP that can sometimes provide significantly better out-of-sample predictions than simple linear non-ARFIMA models of the type mentioned above, based on analysis of point mean-square forecast errors (MSFEs) as well as based on application of Diebold and Mariano (1995) predictive accuracy tests and Clark and McCracken (2001) encompassing t-tests. Furthermore, the samples used in the DGE data set appear to be sufficiently large so as to remedy finite sample bias properties of 4 standard d-estimators (including Geweke and Porter-Hudak, Whittle, rescaled range and modified rescaled range estimators) that have been so widely discussed in the literature.

Interestingly, similar results arise even when much smaller samples of data are examined, such as our second data set which includes daily stock index returns for 5 major indices, as examined by Leybourne et al. (LHM, 2003). For example, based on sequences of recursive ex ante 1-day, 1-week, and 1-month ahead predictions an ARFIMA model is preferred to a non-ARFIMA model 13, 15, and 21 times, respectively. These results are based on application of DM tests (using a 10% significance level) to a single ARFIMA and a single non-ARFIMA model, where the ARFIMA and non-ARFIMA models have previously been selected based on an initial ex ante predictive evaluation of the first-half of the sample. The largest number of "wins" here is thus 21, which is actually around half of the time, as there are 45 models in total for each estimation scheme and forecast horizon (i.e. there are 5 different stock indexes times 9 different data transformation and sample period combinations).[3] This sort of evidence does not carry over to much shorter macroeconomic time series, however. In particular, there are only a limited number of significant findings in favor of ARFIMA models when comparing truly ex ante predictions of the 215 macroeconomic variables examined by Stock and Watson (SW, 2002). This finding, as well as many of the other findings discussed above are validated via a series of Monte Carlo experiments which assess, in a real-time context, the predictive ability of various ARFIMA and non-ARFIMA models.[4]

Third, we pose a number of related questions, such as the following: What is the impact of forecast horizon on predictive performance of various ARFIMA and non-ARFIMA models? How quickly do empirical estimates of the difference operator deteriorate in settings where the number of available observations may be limited? Does the parsimony of ARIMA models relative to related ARFIMA models ensure that ARIMA models will yield more precise predictions, on average? With regard to the first question, we present evidence suggesting that long-memory models may be particularly useful at longer forecast horizons. With regard to the second question, we find that samples of 5000 or more observations yield very stable rolling and recursive estimates of d, while samples of 2500 or fewer observations lead to substantial increases in estimator standard errors. Finally, with regard to the third question, it appears that parsimony is not always necessary to produce accurate predictions, as our less parsimonious ARFIMA models sometimes dominate their more parsimonious ARMA counterparts, even for moderately sized samples of around 2500 observations.

The rest of the paper is organized as follows. In Section 2 we briefly review ARFIMA processes, and outline the empirical estimation and testing methodology used in the rest of the paper. In Section 3, we present the results of an empirical investigation of the 17,054 observation DGE data set, the 4950 observation LHM data set, and the 215 variable macroeconomic observation SW data set. Section 4 contains the results of a series of Monte Carlo experiments that were designed to yield further evidence on a number of issues and findings based on our empirical analysis. Section 5 concludes.

[3]Our empirical results thus support the conjecture made by two anonymous referees that misspecification of long-memory features is likely to be more important for multi-step ahead forecasts.

[4]It should be noted that we do not compare our ARFIMA models to models with breaks. In a working paper version of this paper (see Bhardwaj and Swanson, 2003) simple regime switching models were estimated in our empirical experiments. These models were outperformed by the simple linear non-ARFIMA models based on ex ante analysis of the first-half of all sample periods examined, and hence no regime switching models are included in this version of the paper. However, it is clear that more complicated regime switching models, as well as linear and non-linear models with structural breaks, might fare better from a predictive perspective. Analysis of this possibility is discussed elsewhere in the literature, and is left to future research.

2. Empirical methods

The prototypical ARFIMA model examined in the literature is

$$\Phi(L)(1-L)^d y_t = \Theta(L)\varepsilon_t, \tag{1}$$

where d is the fractional differencing parameter, ε_t is white noise, and the process is covariance stationary for $-0.5<d<0.5$, with mean reversion when $d<1$. This model is a generalization of the fractional white-noise process described in Granger (1980), Granger and Joyeux (1980), and Hosking (1981), where for the purpose of analyzing the properties of the process, $\Theta(L)$ is set equal to unity (Baillie, 1996 surveys numerous papers that have analyzed the properties of the ARFIMA process). Given that many time series exhibit very slowly decaying autocorrelations, the potential advantage of using ARFIMA models with hyperbolic autocorrelation decay patterns when modelling economic and financial times series seems clear (as opposed to models such as ARMA processes that have exponential or geometric decay). The potential importance of the hyperbolic decay property can be easily seen by noting that

$$\begin{aligned}(1-L)^d &= \sum_{j=0}^{\infty}(-1)^j\binom{d}{j}(L)^j = 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 \\ &\quad + \cdots = \sum_{j=0}^{\infty} b_j(d)\end{aligned} \tag{2}$$

for any $d>-1$.[5] As a simple illustration, Table 1 reports the values of the coefficients associated with different lags in the expansion of $(1-L)^d$ given in Eq. (2). The last column of the table gives the lag after which coefficients of the polynomial become smaller than $1.0\text{e}-004$. It is interesting to note that by this crude yardstick the coefficients are included even after 400 lags, in the case when $d=0.2$.

There are currently dozens of estimation methods for and tests of long-memory models. Perhaps one of the reasons for the wide array of tools for estimation and testing is that the current consensus suggests that good estimation techniques remain elusive, and many of the tests used for long-memory have been shown via finite sample experiments to perform quite poorly. Much of this evidence has been reported in the context of comparing one or two classes of estimators/tests, such as rescaled range (RR)-type estimators (as introduced by Hurst, 1951 and modified by Lo, 1991, for example) and log periodogram regression estimators due to Geweke and Porter-Hudak (GPH, 1983). In the face of all of the negative publicity, it a bit surprising that few papers seem to compare more that one or two different (classes of) estimators and/or tests. Our approach, while still far from exhaustive, is to use a variety of the most widely used estimators and tests in our subsequent empirical

[5]For $d>0$, the differencing filter can also be expanded using hypergeometric functions, as follows: $(1-L)^d = \Gamma(-d)\sum_{j=0}^{\infty} L^k \Gamma(j-d)/\Gamma(j+1) = F(-d,1,1,L)$, where $F(a,b,c,z) = \Gamma(c)/[\Gamma(a)\Gamma(b)]\sum_{j=0}^{\infty} z^j\ \Gamma(a+j)\Gamma(b+j)/[\Gamma(c+j)\Gamma(j+1)]$.

Table 1
The long-memory filter $(1-L)^{d}$[a]

d	lag = 5	lag = 10	lag = 20	lag = 25	lag = 50	lag = 75	lag = 100	Lag truncation
0.2	−0.0255	−0.0110	−0.0047	−0.0036	−0.0016	−0.001	−0.0007	496
0.3	−0.0297	−0.0118	−0.0048	−0.0035	−0.0014	−0.0008	−0.0006	387
0.4	−0.0300	−0.0110	−0.0041	−0.0030	−0.0011	−0.0006	−0.0004	281
0.6	−0.0228	−0.0071	−0.0023	−0.0016	−0.0005	−0.0003	−0.0002	139
0.7	−0.0173	−0.0050	−0.0015	−0.0010	−0.0003	−0.0002	−0.0001	96

[a]*Notes*: Values taken by the filter $(1-L)^d$ are reported in columns 2–8. The last column gives the lag after which the absolute value of coefficients of the polynomial become smaller than $1.0e-004$.

investigation and experimental analysis. In particular, we consider 4 quite widely used estimation methods and 5 different long-memory tests.[6]

2.1. Long-memory model estimation

2.1.1. GPH estimator

The GPH estimation procedure is a two-step procedure, which begins with the estimation of d, and is based on the following log-periodogram regression:[7]

$$\ln[I(\omega_j)] = \beta_0 + \beta_1 \ln\left[4\sin^2\left(\frac{\omega_j}{2}\right)\right] + v_j, \tag{3}$$

where

$$\omega_j = \frac{2\pi j}{T}, \quad j = 1, 2, \ldots, m.$$

The estimate of d, say $\widehat{d}_{\mathrm{GPH}}$, is $-\widehat{\beta}_1$, ω_j represents the $m=\sqrt{T}$ Fourier frequencies, and $I(\omega_j)$ denotes the sample periodogram defined as

$$I(\omega_j) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} y_t \mathrm{e}^{-\omega_j t}\right|^2. \tag{4}$$

The critical assumption for this estimator is that the spectrum of the ARFIMA(p,d,q) process is the same as that of an ARFIMA$(0,d,0)$ process (the spectrum of the ARFIMA(p,d,q) process in (1), under some regularity conditions, is given by $I(\omega_j) = z(\omega_j)\left(2\sin(\frac{\omega_j}{2})\right)^{-2d}$, where $z(\omega_j)$ is the spectrum of an ARMA process). We use $m=\sqrt{T}$, as is done in Diebold and Rudebusch (1989), although the choice of m when ε_t is autocorrelated can heavily impact empirical results (see Sowell,

[6]Perhaps the most glaring omission from our list of estimators is the full information maximum-likelihood estimator of Sowell (1992a). While his estimator is theoretically appealing, it is computationally demanding as it requires inversion of $T \times T$ matrices of non-linear functions of hypergeometric functions. For evidence on the finite sample performance of this estimator, the reader is referred to Cheung and Diebold (1994). For an updated discussion of the maximum-likelihood estimator and its properties, see Doornik and Ooms (2003).

[7]The regression model is usually estimated using least squares.

1992b for discussion). Robinson (1995a) shows that $(\pi^2/24m)^{-1/2}(\widehat{d}_{\mathrm{GPH}} - d) \to \mathrm{N}(0,1)$, for $-\frac{1}{2}<d<\frac{1}{2}$, and for $j = l, \ldots, m$ in the equation for ω above, where l is analogous to the usual lag truncation parameter. As is also the case with the next two estimators, the second step of the GPH estimation procedure involves fitting an ARMA model to the filtered data, given the estimate of d. Agiakloglou et al. (1992) show that the GPH estimator has substantial finite sample bias, and is inefficient when ε_t is a persistent AR or MA process. Many authors assume normality of the filtered data in order to use standard estimation and inference procedures in the analysis of the final ARFIMA model (see e.g. Diebold and Rudebusch, 1989,1991a). Numerous variants of this estimator continue to be widely used in the empirical literature.[8]

2.1.2. WHI estimator

Another semiparametric estimator, the Whittle estimator, is also often used to estimate d. Perhaps one of the more promising of these is the local Whittle estimator proposed by Künsch (1987) and modified by Robinson (1995b). This is another periodogram-based estimator, and the crucial assumption is that for fractionally integrated series, the autocorrelation (ρ) at lag l is proportional to l^{2d-1}. This implies that the spectral density which is the Fourier transform of the autocovariance γ is proportional to $(\omega_j)^{-2d}$. The local Whittle estimator of d, say $\widehat{d}_{\mathrm{WHI}}$, is obtained by maximizing the local Whittle log likelihood at Fourier frequencies close to zero, given by

$$\Gamma(d) = -\frac{1}{2\pi m}\sum_{j=1}^{m}\frac{I(\omega_j)}{f(\omega_j;d)} - \frac{1}{2\pi m}\sum_{j=1}^{m} f(\omega_j;d), \tag{5}$$

where $f(\omega_j;d)$ is the spectral density (which is proportional to $(\omega_j)^{-2d}$). As frequencies close to zero are used, we require that $m \to \infty$ and $\frac{1}{m} + \frac{m}{T} \to 0$, as $T \to \infty$. Taqqu and Teverovsky (1997) show that $\widehat{d}_{\mathrm{WHI}}$ can be obtained by maximizing the following function:

$$\widehat{\Gamma}(d) = \ln\left(\frac{1}{m}\sum_{j=1}^{m}\frac{I(\omega_j)}{\omega_j^{-2d}}\right) - 2d\frac{1}{m}\sum_{j=1}^{m}\ln(\omega_j). \tag{6}$$

Robinson (1995b) shows that for estimates of d obtained in this way, $(4m)^{1/2}(\widehat{d}_{\mathrm{WHI}} - d) \to \mathrm{N}(0,1)$, for $-\frac{1}{2}<d<\frac{1}{2}$. Taqqu and Teverovsky (1997) study the robustness of standard, local, and aggregated Whittle estimators to non-normal innovations, and find that the local Whittle estimator performs well in finite samples. Shimotsu and Phillips (2002) develop an exact local Whittle estimator that applies throughout the stationary and non-stationary regions of d, while Andrews and Sun (2002) develop an adaptive local polynomial Whittle estimator in order to address the slow rate of convergence and associated large finite sample bias associated with the local Whittle estimator. In this paper, we use the local Whittle estimator discussed in Taqqu and Teverovsky (1997).

[8] For a recent overview of frequency domain estimators, see Robinson (2003, Chapter 1).

2.1.3. RR estimator

The rescaled range estimator was originally proposed as a test for long-term dependence in the time series. The statistics is calculated by dividing range with standard deviation. In particular, define

$$\widehat{Q}_T = \frac{\widehat{R}_T}{\widehat{\sigma}_T}, \tag{7}$$

where $\widehat{\sigma}_T^2$ is the usual maximum-likelihood variance estimator of y_t, and $\widehat{R}_T = \max_{0<i\leqslant T}\sum_{t=1}^{i}(y_t-\bar{y}) - \min_{0<i\leqslant T}\sum_{t=1}^{i}(y_t-\bar{y})$. The estimate of d, say $\widehat{d}_{\mathrm{RR}}$, is obtained using the result that $\operatorname{plim}_{T\to\infty}\left(T^{-d-\frac{1}{2}}\frac{\widehat{R}_T}{\widehat{\sigma}_T}\right) = constant$ (see Hurst, 1951; Lo, 1991, and the references cited therein), and is

$$\widehat{d}_{\mathrm{RR}} = \frac{\ln(\widehat{Q}_T)}{\ln(T)} - \frac{1}{2}. \tag{8}$$

Lo (1991) shows that $T^{-1/2}\widehat{Q}_T$ is asymptotically distributed as the range of a standard Brownian bridge. With regard to testing in this context, note that there are extensively documented deficiencies associated with long-memory tests based on $T^{-1/2}\widehat{Q}_T$, particularly in the presence of data generated by a short-memory processes combined with a long-memory component (see e.g. Cheung, 1993). For this reason, Lo (1991) suggests the modified RR test, whereby $\widehat{\sigma}_T^2$ is replaced by a heteroskedasticity and autocorrelation consistent variance estimator, namely

$$\widehat{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})^2 + \frac{2}{T}\sum_{j=1}^{q} w_j(q)\left\{\sum_{t=j+1}^{T}(y_t-\bar{y})(y_{t-j}-\bar{y})\right\}, \tag{9}$$

where

$$w_j(q) = 1 - \frac{j}{q+1}, \quad q<T.$$

It is known from Phillips (1987) that $\widehat{\sigma}_T^2$ is consistent when $q = \mathrm{O}(T^{1/4})$, at least in the context of unit root tests, although choosing q in the current context is a major difficulty. This statistic still weakly converges to the range of a Brownian bridge.

2.1.4. AML estimator

The fourth estimator that we use is the approximate maximum-likelihood estimator of Beran (1995). For any ARFIMA model given by Eq. (1), $d = m + \delta$, where $\delta \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, and m is an integer (assumed known) denoting the number of times the series must be differenced in order to attain stationarity, say

$$x_t = (1-L)^m y_t. \tag{10}$$

To form the estimator, a value of δ is fixed, and an ARMA model is fitted to the filtered x_t data, yielding a sequence of residuals. This is repeated over a fine grid of $d = m + \delta$, and $\widehat{d}_{\mathrm{AML}}$ is the value which minimizes the sum squared residuals. The choice of m is critical, given that the method only yields asymptotically normal estimates of the parameters of the ARFIMA model if $\delta \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, for example (see Robinson, 2003, Chapter 1 for a critical discussion of the AML estimator).

In summary, three of the estimation methods described in the preceding paragraphs for ARFIMA models require first estimating d. Thereafter, an ARMA model is fitted to the filtered data by using maximum likelihood to estimate parameters, and via the use of the Schwarz Information Criterion for lag selection. The maximum number of lags was picked for each of the data sets examined in our empirical section by initially examining the first-half of the sample to ascertain what sorts of lag structures were usually chosen using the SIC. The exception to the above approach is the AML estimator, for which a grid of d values is searched across, with a new ARMA model fitted for each values of d in the grid, and resulting models compared using mean square error.

2.2. Short memory tests

Four of the five tests that we use when evaluating our time series are based on the above discussion, including the GPH, RR, MRR, and WHI tests, where the MRR is the modified RR test due to Lo (1991). Notice that of these, only the GPH and WHI tests are based directly upon examination of the d estimator, while the RR and MRR tests do not involve first estimating d. The fifth test that we use is the non-parametric short-memory test of Leybourne et al. (LHM, 2003). Their test is based on the rate of decay of the autocovariance function. In particular, the null hypothesis of the test is that the data are short-memory (i.e. that $\sum_{j=0}^{\infty}|\gamma_j| < \infty$, where γ_j is the autocovariance of y_t at lag j), and the test is based on the notion that one can distinguish between short and long-memory via knowledge of the rate at which $\gamma_j \to 0$, as $j \to \infty$. The test statistics is

$$S_{k,T} = \frac{T^{1/2}\widehat{\gamma}_{k_T}}{\widehat{\sigma}_\infty}, \tag{11}$$

where $\widehat{\sigma}_\infty^2 = \widehat{\gamma}_0^2 + 2\sum_{j=1}^{l_T}\widehat{\gamma}_j^2$, $\widehat{\gamma}_j = T^{-1}\sum_{t=j+1}^{T} y_t y_{t-j}$, y_t in this case is the demeaned series, and k_T, l_T are chosen such that $k_T, l_T \longrightarrow \infty$, as $T \longrightarrow \infty$ and $\frac{k_T}{l_T} \longrightarrow 0$, $k_T < l_T$. The values which we use, as suggested by LHM, are $k_T = \frac{5.5T^{1/2}}{\ln(T)}$ and $l_T = 4(\frac{T}{100})^{1/4}$. In this context, $S_{k_T} \longrightarrow \mathrm{N}(0, 1)$, under the null hypothesis. There are many other important tests available in the literature which are not examined here, including but not limited to the KPSS test (see Lee and Schmidt, 1996) and the augmented Dickey–Fuller test (see Diebold and Rudebusch, 1991b).

2.3. Predictive accuracy testing

If, as is often the case, the ultimate goal of an empirical investigation is the specification of predictive models, then a natural tool for testing for the presence of long-memory is the predictive accuracy test. In this case, if an ARFIMA model can be shown to yield predictions that are superior to those from a variety of alternative linear (and non-linear) models, then one has direct evidence of long-memory, at least in the sense that the long-memory model is the best available "approximation" to the true underlying DGP. Conversely, even if one finds evidence of long-memory via application of the tests discussed above, then there is little use specifying long-memory models if they do not outpredict simpler alternatives. There is a rich recent literature on predictive accuracy testing, most of which draws in one way or another on Granger and Newbold (1986), where simple tests comparing mean-square forecast errors (MSFEs) of pairs of alternative models under assumptions of normality are outlined. Perhaps the most important of the predictive accuracy tests that have been developed over the last 20 years is the Diebold and Mariano (DM, 1995) test. The statistic is

$$\widehat{d}_P = P^{-1/2}\frac{\sum_{t=R-h+1}^{T-1}(f(\widehat{v}_{0,t+h}) - f(\widehat{v}_{1,t+h}))}{\widehat{\sigma}_P}, \tag{12}$$

where R denotes the estimation period, P is the prediction period, f is some generic loss function, $h\geqslant 1$ is the forecast horizon, $\widehat{v}_{0,t+h}$ and $\widehat{v}_{1,t+h}$ are h-step ahead prediction errors for models 0 and 1 (where model 0 is assumed to be the ARFIMA model), constructed using consistent estimators, and $\widehat{\sigma}_P^2$ is defined as

$$\begin{aligned}\widehat{\sigma}_P^2 = {} & \frac{1}{P}\sum_{t=R-h+1}^{T-1}(f(\widehat{v}_{0,t+h}) - f(\widehat{v}_{1,t+h}))^2 \\ & + \frac{2}{P}\sum_{j=1}^{l_P} w_j \sum_{t=R-h+1+j}^{T-1}(f(\widehat{v}_{0,t+h}) - f(\widehat{v}_{1,t+h}))(f(\widehat{v}_{0,t+h-j}) - f(\widehat{v}_{1,t+h-j})),\end{aligned} \tag{13}$$

where $w_j = 1 - \frac{j}{l_P+1}$, $l_P = \mathrm{o}(P^{1/4})$. The hypotheses of interest are

$$\mathrm{H}_0 : E(f(v_{0,t+h}) - f(v_{1,t+h})) = 0$$

and

$$\mathrm{H}_\mathrm{A} : E(f(v_{0,t+h}) - f(v_{1,t+h})) \neq 0.$$

The DM test, when constructed as outlined above for non-nested models, has a standard normal limiting distribution under the null hypothesis.[9] West (1996) shows that when the out-of-sample period grows at a rate not slower than the rate at which the estimation period grows (i.e. $\frac{P}{R} \to \pi$, with $0<\pi\leqslant\infty$), parameter estimation error generally affects the limiting distribution of the DM test in stationary contexts. On the other hand, if $\pi = 0$, then parameter estimation error has no effect. Additionally, Clark and McCracken (2001) point out the importance of addressing the issue of nestedness when applying DM and related tests.[10] Other recent papers in this area include Christoffersen (1998), Christoffersen and Diebold (1997), Clements and

[9]We assume quadratic loss in our applications, so that $f(v_{0,t+h}) = v_{0,t+h}^2$, for example.

[10]Chao et al. (2001) address not only nestedness, by using a consistent specification testing approach to predictive accuracy testing, but also allow for misspecification amongst competing models; an important feature if one is to presume that all models are approximations, and hence all models may be (dynamically) misspecified. White (2000) further extends the Diebold and Mariano framework by allowing for the joint comparison of multiple models, while Corradi and Swanson (2005a,b,c) extend White (2000) to predictive density evaluation with parameter estimation error.

Smith (2000,2002), Corradi and Swanson (2002), Debold et al. (1998,1999), Harvey et al. (1997), and the references therein, to name but a few. Although, the DM test does not have a normal limiting distribution under the null of non-causality when nested models are compared, the statistic can still be used as an important diagnostic in predictiveaccuracy analyses. Furthermore, the non-standard limit distribution is reasonably approximated by a standard normal in many contexts (see McCracken, 1999 for tabulated critical values). For this reason, and as a rough guide, we use critical values obtained from the N(0, 1) distribution when carrying out DM tests. A final caveat that should be mentioned is that the work of McCracken (and that of Clark and McCracken discussed below) assumes stationarity, assumes correct specification under the null hypothesis, and often assumes that estimation is via least squares. Of course, if we are willing to make the strong assumption of correct specification under the null, then the ARFIMA model and the non-ARFIMA models are the same, implying for example that $d = 0$, so that only the common ARMA components in the models remain, and hence errors are short-memory. Nevertheless, it is true that in general some if not many of the assumptions may be broken in our context, and extensions of their tests and related tests to more general contexts is the subject of ongoing research by a number of authors.[11] This is another reason why the critical values used in this paper should be viewed only as rough approximations.

We also report results based on the application of the Clark and McCracken (CM, 2001) encompassing test, which is designed for comparing nested models. The test statistic is

$$ENC - t = (P-1)^{1/2} \frac{\overline{c}}{\left(P^{-1}\sum_{t=R}^{T-1}(c_{t+h} - \overline{c})\right)^{1/2}},$$

where $c_{t+h} = \widehat{v}_{0,t+h}(\widehat{v}_{0,t+h} - \widehat{v}_{1,t+h})$ and $\overline{c} = P^{-1}\sum_{t=R}^{T-1} c_{t+1}$. This test has the same hypotheses as the DM test, except that the alternative is $\mathrm{H_A}: E(f(v_{0,t+h}) - f(v_{k,t+h}))>0$. If $\pi = 0$, the limiting distribution is N(0, 1) for $h = 1$. The limiting distribution for $h>1$ is non-standard, as discussed in CM. However, as long as a Newey and West (1987)-type estimator (of the generic form given above for the DM test) is used when $h>1$, then the tabulated critical values are quite close to the N(0, 1) values, and hence we use the standard normal distribution as a rough guide for all horizons (see CM for further discussion).

2.4. Predictive model selection

In the sequel, forecasts are 1-step, 5-steps and 20-steps ahead, when daily stock market data are examined, corresponding to 1-day, 1-week and 1-month ahead predictions. For the long DGE data set, we also examine 120- and 240-step ahead forecasts, corresponding to 6-month and 1-year ahead predictions. Additionally, forecasts are 1-step, 3-steps and 12-steps ahead, when monthly U.S. macroeconomic

[11] For example, for further discussion of the ramifications of using non-stationary variables when constructing tests of predictive ability, see Corradi et al. (2001) and Rossi (2003).

data are examined, corresponding to 1-month, 1-quarter and 1-year ahead predictions. Estimation is carried out as discussed above for ARFIMA models, and using maximum likelihood for non-ARFIMA models. More precisely, each sample of T observations is first split in half. The first-half of the sample is then used to produce $0.25T$ rolling (and recursive) predictions (the other $0.25T$ observations are used as the initial sample for model estimation) based on rolling (and recursively) estimated models (i.e. parameters are updated before each new prediction is constructed). These predictions are then used to select a "best" ARFIMA and a "best" non-ARFIMA model, based on point out-of-sample mean-square forecast error comparison. At this juncture, the specifications of the ARFIMA and non-ARFIMA models to be used in later predictive evaluation are fixed. Parameters in the models may be updated, however. In particular, recursive and rolling ex ante predictions of the observations in the second half of the sample are then constructed, with parameters in the ARFIMA and non-ARFIMA "best" models updated before each new forecast is constructed. Additionally, different models are constructed for each forecast horizon, as opposed to estimating a single model and iterating forward when constructing multiple step ahead forecasts. Reported DM and encompassing t-tests are thus based on the second-half of the sample, and involve comparing only two models. Results for mean absolute deviation and mean absolute percentage error loss functions have also been tabulated, and are available upon request from the authors.

It should be stressed that the methodology presented above is often used in 'horse-races' of the type that we are carrying out, so as not to "cherry-pick" the forecast-best models (see e.g. Swanson and White (1995, 1997) and the references cited therein). However, there are many other ways to avoid issues that arise when comparing many models, such as the prevalence of sequential test-bias and overfitting. For recent important papers that address model these and related issues, the reader is referred to White (2000), Inoue and Kilian (IK, 2003), Corradi and Swanson (2005b), and Hansen et al. (HLN, 2004). IK suggest the use of information criterion (such as the Schwarz Information Criterion—SIC) for choosing the best forecasting model, while HLN propose a model confidence set approach to the same problem. Of note, though, is that the BIC-based approach of IK is not applicable under near stationarity and non-linearity, and is not consistent when non-nested models are being compared. HLN takes a different approach, as they are concerned with narrowing down from a larger set of models to a smaller set that includes the best forecasting model. When their approach is used, for example, it is found that ARFIMA volatility models do not outperform simpler non-ARFIMA volatility models.

3. Empirical evidence

In our empirical (and subsequent Monte Carlo) investigation, the following models are used:

(1) $ARFIMA(p,d,q)$: $\Phi(L)(1-L)^d y_t = \alpha + \Theta(L)\varepsilon_t$, where d can take fractional values;

(2) *Random Walk*: $y_t = y_{t-1} + \varepsilon_t$;
(3) *Random Walk with Drift*: $y_t = \alpha + y_{t-1} + \varepsilon_t$;
(4) *AR*(*p*): $\Phi(L)y_t = \alpha + \varepsilon_t$;
(5) *MA*(*q*): $y_t = \alpha + \Theta(L)\varepsilon_t$;
(6) *ARMA*(*p*, *q*): $\Phi(L)y_t = \alpha + \Theta(L)\varepsilon_t$;
(7) *ARIMA*(*p*, *d*, *q*): $\Phi(L)(1 - L)^d y_t = \alpha + \Theta(L)\varepsilon_t$, where d can take integer values;
(8) *GARCH*: $\Phi(L)y_t = \alpha + \varepsilon_t$, where $\varepsilon_t = h_t^{1/2} v_t$ with $\mathrm{E}(\varepsilon_t^2|\Im_{t-1}) = h_t = \varpi + \alpha_1\varepsilon_{t-1}^2 + \cdots + \alpha_q\varepsilon_{t-q}^2 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p}$, and where $\Im_{t-1}$ is the usual filtration of the data; and

In these models, ε_t is the disturbance term, $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$, and $\Theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q$, where L is the lag operator. All models (except ARFIMA models) are estimated using (quasi) maximum likelihood, with values of p and q chosen via use of the Schwarz Information Criterion (SIC), and integer values of d in ARIMA models selected via application of the augmented Dickey–Fuller test at a 5% level. Errors in the GARCH models are assumed to be normally distributed. ARFIMA models are estimated using the four estimation techniques discussed above (GPH, RR, WHI, and AML). When fitting ARFIMA models, we used an arbitrary cut-off of $1.0e - 004$. Terms in the polynomial expansion with coefficients smaller in absolute value than this cut-off were truncated.[12]

In the proceeding subsections, we carry out our empirical investigation by examining the long-memory and ARFIMA predictive properties of the S&P500 series used by Ding et al. (1993) and Granger and Ding (1996), the 5 stock index returns used by Leybourne et al. (2003), and the 215 Stock and Watson (2002) macroeconomic variables. Before discussing the results, however, some comments concerning the data are in order. Our first data set is an updated version of the long historical S&P500 returns data set of DGE. The period covered is January 4, 1928–September 30, 2003 (20,105 observations), so that our data set is somewhat longer than the 17,054 observations (ending on August 30, 1990) examined by DGE. Our second data set consists of the returns data used in Leybourne et al. (2003), where strong evidence of long-memory is found via application of their short-memory test. In particular, we model 4950 (or more, depending on the particular index) daily returns for the following stock indices: S&P500, FTSE100, DAX, Nikkei225, and the Hang Seng.[13] We consider absolute returns, squared returns, and log-squared returns, thus nesting a variety of different data transformations that have been shown in earlier papers (see e.g. Granger and Ding, 1996) to have long-memory properties. All series span the period 01/04/1981–01/18/2002, and the ex ante predictive samples used in our analysis include the entire second-half of the

[12]The exception to the rule is the case of the SW data, for which sufficient observations were not available, and for which, after some experimentation, the arbitrary cut-off was set at 120 lags.

[13]It should be stressed, however, that our sample is first split in half for initial model selection. Thus, our predictive analyses carried out in order to compare ARFIMA and non-ARFIMA models are based on less than 2500 observations.

sample, as well as periods in the second-half of the sample pre- and post-1987 October crash. Finally, we examine the Stock–Watson data set, which consists of the 215 variables used in their well-known diffusion index paper. In the paper, they examine multi-step ahead predictions of 8 key US macroeconomic variables, in a simulated real-time forecasting environment, using all 215 US series to construct diffusion indices. Their data were collected in 1999, and so represent a snapshot of the vintages and releases of data available at that point in time. The data series vary in length, span the period 1959–1998, and are generally 400–500 observations in length. All series are monthly. Appendix 2 of Stock and Watson (2002) contains definitions of all of the series (which are omitted here for the sake of brevity), and discusses the data transformations applied to each series. Note that all series were "differenced to stationarity" in Stock and Watson (2002), prior to model fitting. We use the same data transformations as they did, so that many of the series are expressed in growth rates, and some series are even differenced twice. In summary, our approach is to use exactly the same data set as used in Stock and Watson (2002). However, rather than focusing on predictions of 8 series, we consider predictions of all 215 variables, using estimated versions of the models outlined above.

Before turning to our empirical results, it should be pointed out that we found very little connection between rejection of short-memory test null hypotheses and whether or not the ARFIMA model outperformed the ARIMA model, when comparing predictions. Results confirming this finding have been tabulated and are available upon request. However, there is little that one can discern from the results, other than the lack of correlation between findings in favor of long-memory and incidence of ARFIMA forecasting model 'wins'. One possible explanation of this is the poor finite sample properties of the short-memory tests used in this paper, as pointed out in numerous papers in the literature on fractionally integrated models. Other possible explanations for this finding exist, but discussion of them is left to future research.

3.1. S&P500 Returns: January 4, 1928– September 30, 2003

Table 2 summarizes results based on analysis of our long returns data set. Before discussing these results, however, it is first worth noting that the four alternative estimators of d yield quite similar estimates, as opposed to the types of estimates obtained when our other much shorter data sets are examined. In particular, note that if one were to use the first-half of the sample for estimation, one would find values of d equal to 0.49 (GPH), 0.41 (AML), 0.31 (RR) and 0.43 (WHI).[14] Furthermore, all methods find one AR lag, and all but one method finds 1 MA lag. This is as expected for large samples. In the next subsection, we show that our 4 estimators yield radically different values even when the in-sample period used is moderately large, with approximately 2500 observations, so that the convergence of

[14]These estimates of d are very close to those obtained by Ding et al. (1993) and by Granger and Ding (1996) using their fractionally integrated ARCH model.

Table 2
Analysis of U.S. S&P500 daily absolute returns[a]

Estimation scheme and forecast horizon	ARFIMA model	d	Non-ARFIMA model	DM	ENC-t	DM best vs. RW
1-day ahead, recursive	WHI (1,1)	0.41 (0.0001)	ARMA(4,2)	−1.18	0.47	−13.64
5-day ahead, recursive	GPH (1,2)	0.57 (0.0011)	ARMA(4,2)	−0.71	1.75	−10.10
20-day ahead, recursive	GPH (1,2)	0.57 (0.0011)	ARMA(4,2)	−0.68	2.91	−5.96
120-day ahead, recursive	GPH (1,2)	0.57 (0.0011)	ARMA(4,2)	0.38	7.52	−6.33
240-day ahead, recursive	GPH (1,2)	0.57 (0.0011)	ARMA(4,2)	0.52	10.22	−6.16
1-day ahead, rolling	RR (1,1)	0.25 (0.0009)	ARMA(4,2)	2.02	4.56	−12.44
5-day ahead, rolling	GPH (1,2)	0.55 (0.0044)	ARMA(4,2)	−2.28	0.26	−10.24
20-day ahead, rolling	GPH (1,2)	0.55 (0.0044)	ARMA(4,2)	−2.44	0.79	−5.91
120-day ahead, rolling	GPH (1,2)	0.55 (0.0044)	ARMA(4,2)	−4.07	0.09	−6.32
240-day ahead, rolling	RR (1,1)	0.25 (0.0009)	ARMA(4,2)	−2.62	2.72	−5.90

[a]Models are estimated as discussed above, and model acronyms used are as outlined in Section 3 and Table 7. Data used in this table correspond to those used in Ding et al. (1993), are daily, and span the period 1928–2003. Reported results are based on predictive evaluation using the second-half of the sample. The 'ARFIMA Model' and the 'non-ARFIMA Model' are the models chosen using MSFEs associated with ex ante recursive (rolling) estimation and 1-, 5-, 20-, 120- and 240-step ahead prediction of the different model/lag combinations using the first 50% of sample. The remaining 50% of sample is used for subsequent ex ante prediction, the results of which are reported in the table. Further details are given in Section 2.4. In the second column, entries in brackets indicate the number of AR and MA lags chosen for the ARFIMA model. The third column lists the average (and standard error) of the estimated values of d across the entire ex ante sample, thus these entries are conditional on the selected ARFIMA model. Diebold and Mariano (DM) test statistics are based on MSFE loss, and application of the test assumes that parameter estimation error vanishes and that the standard normal limiting distribution is asymptotically valid, as discussed in Section 2.3. Negative statistic values for DM statistics indicate that the point MSFE associated with the ARFIMA model is lower than that for the non-ARFIMA model, and the null hypothesis of the test is that of equal predictive accuracy. ENC-t statistics reported in the sixth column of the table, are normally distributed for $h = 1$, and correspond to the null hypothesis that the ARFIMA model encompasses the non-ARFIMA model. The last column of the table reports the MSFE loss-based DM test statistics, where the best ARFIMA/non-ARFIMA model is taken as model 0 (see discussion in Section 2.3) which is compared to RW-based forecast.

the estimators is extremely slow, although they do eventually converge. This yields credence to Granger's (1999) observation that estimates of d can vary greatly across different sample periods and sample sizes, and are generally not robust at all (see next section for further evidence of this).

In the table, the "best" ARFIMA and non-ARFIMA models are first chosen as discussed above. As d is re-estimated prior to the construction of each new forecast, means and standard errors of the sequence of d values are reported in the table. As might be expected, different d mean values, which are calculated for each estimation scheme (i.e. recursive or rolling) and each forecast horizon, are all quite close to one another, with the exception of the RR estimator. Additionally, all standard errors are extremely small. Interestingly, though, the means are always above 0.5 except in the case of RR estimator. This is in contrast to the usual finding that $d < 0.5$. Although various explanations for these seemingly large values of d are possible, a leading explanation might be as follows. If, as suggested by Clive Granger and others, long-memory arises in part due to various sorts of misspecification, then it may be the case that greater accumulation of misspecification problems leads to greater "spurious" long-memory. In the sense that our multiple step ahead prediction models may be more poorly specified than our 1-step ahead models (given that we construct a new prediction model for each horizon, and that greater horizons involve using more distant lags of the dependent variable on the RHS of the forecasting model), we have indirect evidence that more severe misspecification, in the form of missing dynamic information, may lead to larger estimated values for d. This finding, if true, has implications for empirical research, as it may help us to better understand the relative merits of using different approaches for constructing multiple-step ahead forecasting models. Finally, it should be stressed that the best ARFIMA/non-ARFIMA models yield significantly better forecasts, when compared to a 'naive' forecasts based on a random walk (or no-change) model. The DM test favors best ARFIMA/non-ARFIMA at all the forecast horizons, though the p-values are slightly higher for longer multi-step ahead forecasts.

Turning next to the DM and encompassing test results reported in the table, notice that the DM statistics are negative in all but three cases. As the ARFIMA model is taken as model 0 (see discussion in Section 2.3), this means that the point MSFEs are lower for the ARFIMA model than the non-ARFIMA model. One exception is the case where the rolling estimation scheme is used and $h = 1$ (this is the case where the RR estimator is used, and where the average d value across the out-of-sample period is 0.25). Interestingly the other two exceptions are when we use recursive estimation and significantly longer forecast horizons (120 and 240-step ahead predictions). However, in all three of these exceptional cases, the DM statistics does not significantly favor either model. On the other hand, use of the rolling estimation scheme results in significantly superior multiple-step ahead predictions for the ARFIMA model, at standard significance levels. This finding is relevant, given that the MSFEs are quite similar when comparing recursive and rolling estimation schemes. The encompassing t-test yields somewhat similar results. In particular, the null hypothesis is most clearly rejected in favor of the alternative that the non-ARFIMA model is the more precise predictive model for the rolling estimation scheme with $h = 1$, and the two recursive estimation schemes with the longest forecasting horizons ($h = 120$ and 240). Interestingly, the null may also be rejected for $h = 20$ when recursive estimation is used (the statistic value is 2.91) and $h = 240$ when rolling estimation is used (the statistic value is 2.72), although in this case using critical values from the $\mathrm{N}(0, 1)$ is only a rough approximation, as the distribution is non-standard, and contains nuisance parameters (so that, in principle, bootstrap methods need to be should be valid and need to be used in order to obtain valid critical values, for example).

While these results are somewhat mixed, they do constitute evidence that long-memory models may actually be useful in certain cases, when constructing forecasting models. Furthermore, as long as the in-sample period is very large, then all of our differencing operator estimators perform adequately (with the possible exception of the RR estimator), and any one of them can be successfully used to estimate "winning" prediction models. Put differently, no model from amongst those considered performs better than our simple ARFIMA models, at least based on point MSFE (with the one exception that is noted above). It should, however, be stressed that structural breaks, regime switching, etc. have not been accounted for in any of our models, and it remains to see whether the types of results obtained here will also hold when structural breaks and regime switching are allowed for in both our short- and long-memory models. Some results in this regard are given in the next subsection, where different return series are examined both pre- and post-1987.

3.2. International stock index returns: January 4, 1981–January 18, 2002

Table 3 contains a summary of empirical results for 5 different stock market indices. Absolute, squared, and log-squared returns are evaluated using the 5 short-memory tests discussed above, an ARFIMA and a non-ARFIMA model is estimated, with these models chosen based on prior ex ante analysis of the first-half of the sample, and rolling and recursive ex ante predictions are made and compared using the second-half of the sample. A number of conclusions can be made based on the analysis reported in the table. First, note that the short-memory null hypothesis (given in brackets in the first column of the table) is rejected most of the time, for most of the indices, regardless of how returns are transformed prior to test statistic construction. At face value, this might be taken as strong evidence of the potential usefulness of ARFIMA models for these data. However, it is well known that the 5 tests used in our study have poor finite sample size properties when faced with non-linear models, such as regime switching models. Indeed, results reported in a working paper version of this paper (see Bhardwaj and Swanson, 2003) suggest that size is very poor even when data are generated according to linear models, such as AR processes with reasonably large roots (e.g. an AR(1) with slope $= 0.75$). Thus, the tests are probably unreliable for the types of data usually examined by macroeconomic and financial economists. This is one of the reasons why we focus on out-of-sample forecast evaluation.

A second conclusion concerns the reported DM test results. Negative entries in the "DM" columns in the table indicate cases for which point MSFEs are lower when the ARFIMA model is used.[15] Starred entries correspond to rejections based on 10% level tests using the $N(0,1)$ distribution (see above for further discussion). Consider recursive forecasts. In this case, the ARFIMA model is preferred 13, 15, and 21 times at the 1, 5, and 20 day ahead horizons, respectively. Notice that the largest number of "wins" for the ARFIMA model is 21, which is around half of the time, as there are 45 models in total for each estimation scheme and forecast horizon (i.e. 5 different stock indexes times 9 different data transformation and sample period

[15]More detailed tables of results for this and the next empirical example that include specifics on which ARFIMA and non-ARFIMA models are compared for each estimation scheme and forecast horizon have been tabulated and are available upon request from the authors.

Table 3
Analysis of international stock market data[a]

Series	SM rejec	d	Recursive estimation scheme						Rolling estimation scheme					
			1-day ahead		5-day ahead		20-day ahead		1-day ahead		5-day ahead		20-day ahead	
			DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t
S & P 500														
$\|r_t\|$	5	0.64 (0.05)	-2.07^{*}	-4.28	-0.58	-3.73	-2.12^{*}	-1.14	-1.17	-3.80	-0.23	-2.75	-6.53^{*}	-5.26
r_t^2	3	0.07 (0.01)	-4.19^{*}	-3.36	-3.90^{*}	-1.42	-7.00^{*}	-8.95	-4.03^{*}	-1.48	-4.23^{*}	-0.37	-4.75^{*}	-3.88
$\log(r_t^2)$	5	0.97 (0.04)	-8.09^{*}	0.56	-7.67^{*}	0.76	-8.11^{*}	0.09	-7.02^{*}	0.57	0.59	0.14	-0.80	0.96
$\|r_t\|$ Pre-87	5	0.52 (0.02)	-1.68^{*}	0.97	-1.94^{*}	0.24	-1.83^{*}	0.05	-1.53	0.86	-1.43	0.48	-1.73^{*}	0.18
r_t^2 Pre-87	5	0.58 (0.04)	$2.49^{\dagger}$	$3.64^{\dagger}$	1.44	$3.27^{\dagger}$	0.40	$1.94^{\dagger}$	1.24	$3.71^{\dagger}$	0.83	0.40	1.30	$2.78^{\dagger}$
$\log r_t^2$ Pre-87	5	0.12 (0.03)	-1.29	0.30	-1.14	0.95	-1.28	0.83	0.90	0.53	0.01	0.61	-0.78	0.24
$\|r_t\|$ Post-87	5	0.46 (0.01)	0.20	$1.75^{\dagger}$	-0.04	0.15	-0.63	0.90	0.32	0.03	-1.32	0.15	-3.31^{*}	-1.92
r_t^2 Post-87	4	0.53 (0.08)	0.51	0.81	-0.10	1.06	-1.32	0.13	-0.11	0.45	-1.14	0.61	$2.86^{\dagger}$	$4.95^{\dagger}$
$\log r_t^2$ Post-87	4	0.23 (0.02)	$2.69^{\dagger}$	$5.70^{\dagger}$	$3.59^{\dagger}$	$7.16^{\dagger}$	$5.96^{\dagger}$	$11.07^{\dagger}$	$4.46^{\dagger}$	$7.62^{\dagger}$	$5.16^{\dagger}$	$9.40^{\dagger}$	$1.91^{\dagger}$	$4.20^{\dagger}$
FTSE														
$\|r_t\|$	4	0.19 (0.03)	$1.72^{\dagger}$	$3.60^{\dagger}$	-2.52^{*}	-0.70	-3.35^{*}	-0.53	-0.63	1.04	-1.26	0.72	-2.34^{*}	0.19
r_t^2	4	0.21 (0.03)	-5.52^{*}	0.04	-4.76^{*}	0.16	-4.96^{*}	1.21	-2.70^{*}	0.98	-4.18^{*}	-2.15	1.37	$5.14^{\dagger}$
$\log r_t^2$	3	0.15 (0.04)	$1.92^{\dagger}$	$4.05^{\dagger}$	$2.08^{\dagger}$	$4.67^{\dagger}$	$2.29^{\dagger}$	$5.68^{\dagger}$	-0.70	0.32	-0.04	0.79	-3.37^{*}	-0.48
$\|r_t\|$ Pre-87	4	0.68 (0.06)	-0.10	0.37	-1.67^{*}	-0.92	-3.69^{*}	-3.18	$2.36^{\dagger}$	$3.98^{\dagger}$	0.11	0.32	-0.92	0.40
r_t^2 Pre-87	3	0.34 (0.01)	-6.07^{*}	-1.19	-5.36^{*}	-1.66	-4.39^{*}	0.58	$2.22^{\dagger}$	$3.74^{\dagger}$	$1.80^{\dagger}$	$3.29^{\dagger}$	-7.05^{*}	0.12
$\log r_t^2$ Pre-87	4	0.15 (0.01)	0.80	0.50	0.93	0.33	$3.70^{\dagger}$	$10.01^{\dagger}$	$4.35^{\dagger}$	$6.97^{\dagger}$	0.72	$1.54^{\dagger}$	0.62	0.31
$\|r_t\|$ Post-87	4	0.47 (0.01)	-1.18	-0.60	-1.64	-1.52	-1.99^{*}	-1.79	-1.27	-0.16	-1.59	-1.10	-1.80^{*}	-1.42
r_t^2 Post-87	3	0.17 (0.05)	$1.69^{\dagger}$	$3.97^{\dagger}$	1.60	$2.71^{\dagger}$	-3.50^{*}	-2.60	$1.76^{\dagger}$	$4.02^{\dagger}$	0.05	0.73	-0.79	1.06
$\log r_t^2$ Post-87	4	0.17 (0.03)	1.43	$3.95^{\dagger}$	-2.49^{*}	0.05	-3.33^{*}	0.28	$2.55^{\dagger}$	$6.84^{\dagger}$	-3.00^{*}	0.76	-3.00^{*}	0.09
DAX														
$\|r_t\|$	5	0.47 (0.01)	0.08	0.18	-0.22	0.24	-0.64	0.66	-0.49	0.96	0.49	1.15	-1.43	0.77
r_t^2	5	0.16 (0.02)	0.74	$1.96^{\dagger}$	-3.21^{*}	-3.43	-3.82^{*}	-4.07	0.26	0.55	-3.00^{*}	-2.60	-3.88^{*}	-2.90
$\log r_t^2$	5	0.20 (0.02)	0.51	0.83	0.34	0.12	1.13	$4.71^{\dagger}$	0.79	0.60	1.55	$4.58^{\dagger}$	1.63	$5.42^{\dagger}$
$\|r_t\|$ Pre-87	5	0.18 (0.04)	1.47	$3.73^{\dagger}$	0.98	0.31	-1.76^{*}	0.24	1.26	$3.63^{\dagger}$	-0.22	0.58	-0.83	0.13
r_t^2 Pre-87	5	0.17 (0.04)	-2.93^{*}	-1.51	-0.44	0.91	-1.50	0.78	-2.54^{*}	-0.04	-1.49	-0.33	-0.55	0.72

$\log r_t^2$ Pre-87	4	0.14 (0.04)	−0.22	0.66	0.35	$2.40^{\dagger}$	0.73	$3.73^{\dagger}$	−1.10	0.16	0.31	0.33	−0.33	0.71
$\lvert r_t\rvert$ Post-87	5	0.68 (0.07)	0.78	0.01	−1.61	−0.36	-1.78^{*}	0.24	−1.53	−0.46	-1.76^{*}	−0.76	−1.63	−0.40
r_t^2 Post-87	4	0.16 (0.03)	0.36	0.29	−1.10	0.12	-1.75^{*}	0.69	0.02	0.05	−0.48	0.45	−0.48	0.35
$\log r_t^2$ Post-87	5	0.20 (0.02)	$2.83^{\dagger}$	$5.80^{\dagger}$	$3.39^{\dagger}$	$6.73^{\dagger}$	$3.45^{\dagger}$	$6.79^{\dagger}$	0.97	0.93	1.36	0.36	$3.06^{\dagger}$	$6.92^{\dagger}$
Nikkei														
$\lvert r_t\rvert$	5	0.47 (0.01)	0.29	0.20	-2.33^{*}	0.14	-3.03^{*}	0.41	$3.77^{\dagger}$	$6.32^{\dagger}$	−1.21	0.19	−1.09	0.90
r_t^2	5	0.18 (0.02)	-6.73^{*}	−2.10	−0.84	0.82	−0.92	0.69	1.33	$2.56^{\dagger}$	-4.65^{*}	−3.30	0.98	$4.32^{\dagger}$
$\log r_t^2$	5	0.94 (0.04)	0.03	0.15	0.81	$2.94^{\dagger}$	0.24	$3.78^{\dagger}$	0.13	0.59	0.48	0.39	−1.43	0.93
$\lvert r_t\rvert$ Pre-87	5	0.15 (0.03)	0.02	0.62	−0.52	0.80	−1.09	0.92	−0.31	0.38	−0.96	0.77	−1.10	0.75
r_t^2 Pre-87	4	0.11 (0.02)	-2.56^{*}	−1.92	1.02	$2.04^{\dagger}$	−0.68	0.63	-2.48^{*}	−1.19	−0.49	0.27	−0.17	0.42
$\log r_t^2$ Pre-87	4	0.67 (0.19)	-3.92^{*}	−0.75	−1.40	0.95	-1.66^{*}	0.08	-3.30^{*}	0.11	-2.24^{*}	0.62	-2.45^{*}	0.41
$\lvert r_t\rvert$ Post-87	5	0.22 (0.02)	0.67	0.50	-3.01^{*}	−0.98	−0.19	0.12	$2.53^{\dagger}$	$4.89^{\dagger}$	$2.03^{\dagger}$	$4.07^{\dagger}$	$2.15^{\dagger}$	$4.75^{\dagger}$
r_t^2 Post-87	4	0.17 (0.02)	0.84	0.62	0.73	$2.85^{\dagger}$	0.50	$2.10^{\dagger}$	0.18	0.27	$1.75^{\dagger}$	$3.78^{\dagger}$	0.79	$2.16^{\dagger}$
$\log r_t^2$ Post-87	5	0.78 (0.14)	$2.44^{\dagger}$	$3.78^{\dagger}$	$3.10^{\dagger}$	$4.75^{\dagger}$	$3.54^{\dagger}$	$6.30^{\dagger}$	0.83	0.57	0.30	0.18	$3.93^{\dagger}$	$6.56^{\dagger}$
Hang Sang														
$\lvert r_t\rvert$	5	0.21 (0.02)	-2.47^{*}	0.30	-2.56^{*}	−0.38	-2.67^{*}	−1.72	−1.06	0.32	-2.59^{*}	−0.12	-3.07^{*}	−1.76
r_t^2	0	0.16 (0.05)	-2.75^{*}	0.96	-3.57^{*}	−2.91	-2.46^{*}	−3.06	-2.71^{*}	−0.91	-3.61^{*}	−2.41	-2.41^{*}	−2.19
$\log r_t^2$	4	0.22 (0.01)	$3.19^{\dagger}$	$5.05^{\dagger}$	$2.27^{\dagger}$	$5.55^{\dagger}$	$2.55^{\dagger}$	$6.92^{\dagger}$	$3.07^{\dagger}$	$5.84^{\dagger}$	$2.52^{\dagger}$	$5.62^{\dagger}$	$2.34^{\dagger}$	$6.77^{\dagger}$
$\lvert r_t\rvert$ Pre-87	4	0.15 (0.05)	−1.47	−0.42	-3.19^{*}	−3.24	-4.02^{*}	−5.77	−0.26	0.33	−0.65	0.29	−0.73	0.40
r_t^2 Pre-87	3	0.13 (0.03)	-6.96^{*}	−4.85	-5.78^{*}	−9.27	$13.15^{\dagger}$	$14.94^{\dagger}$	-2.86^{*}	0.01	-4.94^{*}	−5.90	$4.15^{\dagger}$	$8.26^{\dagger}$
$\log r_t^2$ Pre-87	2	1.07 (0.07)	−0.21	0.71	−0.60	0.52	-2.42^{*}	0.16	$1.89^{\dagger}$	$4.99^{\dagger}$	−0.06	0.86	−1.51	0.63
$\lvert r_t\rvert$ Post-87	5	0.19 (0.04)	0.40	0.74	0.33	0.04	−1.47	1.07	−0.60	0.80	−0.14	0.61	−0.92	0.22
r_t^2 Post-87	1	0.26 (0.13)	0.67	0.32	−0.99	−0.12	−0.92	1.12	−1.11	0.79	−1.14	−1.27	−1.36	0.38
$\log r_t^2$ Post-87	5	0.18 (0.02)	$2.61^{\dagger}$	$4.81^{\dagger}$	1.05	$3.09^{\dagger}$	0.70	$3.77^{\dagger}$	$2.74^{\dagger}$	$5.25^{\dagger}$	$1.68^{\dagger}$	$4.19^{\dagger}$	$2.19^{\dagger}$	$5.67^{\dagger}$

[a] *Notes*: See notes to Table 2. Data used in this table correspond to those used in Leybourne et al. (2003), and the variables are daily, spanning the period 1981–2002. Reported results are based on predictive evaluation using the second-half of the sample. The number in brackets appearing beside the series name reports the number of short-memory test rejections based on application of all 5 SM tests discussed above, at a 10% nominal significance level. The second column of entries reports the average and standard error of estimated d values for the case of one step ahead recursive forecasting. Starred and daggered DM and ENC-t test statistics indicate rejection of the tests' null hypothesis at a 10% nominal significance level, based on MSFE loss, with starred entries indicating rejection in favor of ARFIMA models and daggered entries indicating rejection in favor of non-ARFIMA models (see notes to Table 2 for further details).

combinations). Thus, at least at the 20-day ahead horizon, the empirical findings can hardly be accounted for by chance.[16] Analogous numbers corresponding the number of times the non-ARFIMA model is preferred are 9, 5, and 7. Thus, the ARFIMA models are preferred around twice as frequently, and the number of ARFIMA "wins" increases with forecast horizon, while the number of non-ARFIMA wins stays the same or decreases with forecast horizon. Although, the ARFIMA model no longer wins twice as often under the rolling estimation scheme, the pattern of increasing wins with horizon remains. In particular, in the rolling case, the corresponding numbers for ARFIMA wins are 8, 10, and 13; and those for non-ARFIMA wins are 11, 6, and 8. Indeed, the only case for which the non-ARFIMA model appears to consistently dominate the ARFIMA model is the post-1987 crash period when $\log r_t^2$ is modelled.

Notice also in the table that the mean and standard error (in brackets) of estimated d values are given. These correspond to estimates for the recursive estimation and 1-day ahead prediction models. Estimates for 5- and 20-day ahead models and for rolling estimation schemes are qualitatively similar and are not been included, for the sake of brevity (tabulated values are available upon request from the authors). It is important to note that even with the relatively large samples used in this example, the estimates of d clearly vary depending on which stock market index is used, how the data are transformed, and whether or not pre- or post-October 1987 data are examined. However, the estimates for daily S&P500 absolute returns are very close to those estimates reported in Table 2 for a much longer sample, suggesting that the spread of different d estimates may be as much due to data transformation as sample size. The standard errors are around two orders of magnitude greater, though, suggesting that parameter estimation error plays a great role. Nevertheless, in our context, as we are re-estimating the ARFIMA model many times, and constructing a new prediction each time, the parameter estimation error is likely mitigated somewhat, so that our prediction results are more indicative of what one might expect when using long-memory models than if one were to simply estimate the model a single time and construct a sequence of h-step ahead predictions without parameter updating.

A final finding from this empirical example is that the encompassing null hypothesis is only rejected, yielding evidence that the non-ARFIMA model dominates the ARFIMA, around 10 or fewer times, regardless of which estimation scheme and forecast horizon is considered (see each of the 6 columns with header "ENC-t" in the table).[17]

Overall, there is no clear evidence against the use of ARFIMA models for prediction in the context of stock market data, and indeed our evidence slightly

[16]It should be of interest to establish whether this result holds up when the set of possible non-ARFIMA models is augmented by various non-linear regime switching and related models.

[17]As is the case when the DGE data set is examined, best ARFIMA/non-ARFIMA models yield significantly better forecasts when compared with 'naive' no-change forecasts (i.e. when compared with random walk prediction models), when absolute returns are examined (full results on this are available upon request from the authors).

favors the ARFIMA models relative to simpler non-ARFIMA alternatives, particularly at multiple step ahead horizons.

3.3. Stock– Watson macroeconomic data set: 1959– 1998

Tables 4–6 collect results analogous to those reported in Table 3, but for the much shorter SW data set. These results are broken into three groupings: general macroeconomic variables (Table 4); financial variables (Table 5); and monetary variables (Table 6). As mentioned above, the 215 different time series have variously been differenced, log differenced, etc., according to the definitions given in Appendix 2 of SW. Perhaps the most important feature of the data set is that it contains variables with sample periods ranging from 1959–1998, so that only 400–500 monthly observations are available. Thus, we are subjecting the ARFIMA models to a very stringent test when using them to construct prediction models. Given that we know the estimates of d will be suspect, it would be very surprising if any ARFIMA models were shown to out-predict more parsimonious and precisely estimated AR, ARMA, and related models.

Turning to the results reported in the tables, notice that across all 215 variables, and for the recursive estimation scheme, the ARFIMA model is selected, based on application of 10% level DM tests, 30, 18, and 14 times, at the 1, 3, and 12 month horizons. Corresponding numbers for non-ARFIMA "wins" are 34, 14, and 22. Now consider the rolling estimation scheme. The numbers of wins corresponding to those mentioned above are 23, 9, and 12 for the ARFIMA model and 30, 14, and 24 for the non-ARFIMA model. Thus, each model wins around the same number of times. Furthermore, it is only at the 1-month ahead horizon that the total number of wins (63 for the recursive scheme and 50 for the rolling scheme) are substantially greater than 22 (i.e. 10% of the total number of models). Ultimately, then, one might expect that as the sample is decreased, the proportion of models rejecting the null will approach the size of the test. It is interesting, though, that even 400–500 observations seems to be enough to ensure that empirical findings favoring the ARFIMA and non-ARFIMA models around the same number of times may not simply be due to chance.

Of final note is that the encompassing test statistic suggests rejection of the encompassing null in around half of the models, regardless of variable, estimation scheme, and forecast horizon. This is again evidence that the two different models are faring equally well.

In summary, our analysis of the SW data set suggests two things. First, ARFIMA models may even be useful in rather small samples, particularly when the alternative linear models are of the variety we have considered here.[18] However, the number of times the ARFIMA model "wins" is clearly much greater when larger samples of data are available. Overall, this is a somewhat surprising finding, given that d is

[18]It should be stressed, however, that our 'small' samples begin with data in 1959, and are usually 400–500 observations long. It is far from clear that our findings would hold up for smaller samples of, say, quarterly data post-1980.

Table 4
Analysis of U.S. macroeconomic data (Stock and Watson data set)[a]

Variable	SM rejec	Recursive estimation scheme						Rolling estimation scheme					
		1-month ahead		3-month ahead		12-month ahead		1-month ahead		3-month ahead		12-month ahead	
		DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t
CONDO9	5	0.10	2.91†	0.10	2.96†	0.14	1.97†	0.72	2.87†	0.19	1.10	0.12	2.06†
CONPC	5	1.63	2.41†	1.67†	3.52†	1.69†	2.01†	0.19	2.89†	0.01	3.11†	0.61	2.32†
CONQC	5	2.50†	6.24†	2.53†	3.84†	2.58†	3.72†	0.37	2.17†	0.26	1.87†	0.45	3.33†
CONTC	5	1.93†	2.31†	1.97†	2.26†	2.00†	2.13†	0.06	2.53†	0.12	4.10†	1.05	2.84†
FTB	0	1.90†	2.63†	0.80	1.24	1.38	1.41†	0.60	1.52†	1.46	1.46†	0.11	0.43
FTMD	1	0.63	2.87†	−2.04*	−1.43	−2.06*	−1.43	1.02	2.59†	0.03	0.93	−1.84*	−1.10
FWAFIT	4	−0.26	1.49†	0.21	2.92†	1.11	9.68†	1.01	2.62†	0.82	2.40†	1.27	8.11†
GMCANQ	0	−1.34	−1.34	0.60	2.15†	−1.02	−0.61	0.62	1.99†	0.75	1.79†	1.00	1.89†
GMCDQ	2	−0.75	1.19	0.07	1.18	1.56	2.83†	0.23	1.21	0.01	1.32†	−0.56	0.92
GMCNQ	2	−0.76	0.60	−1.06	−1.15	−3.04*	−1.54	−0.75	0.71	0.08	0.55	−0.58	−0.27
GMCQ	2	1.54	2.76†	−1.83*	−0.81	1.31	0.71	−1.29	0.48	−1.95*	−0.71	1.38	1.57†
GMCSQ	2	−2.23*	−1.60	−1.32	−0.46	−0.36	0.58	2.37†	3.48†	0.27	1.24	0.81	1.33†
GMPYQ	2	1.21	1.57†	0.26	1.22	2.34†	1.86†	0.70	1.20	0.82	2.85†	2.64†	3.10†
GMYXPQ	3	−2.45*	−0.10	−1.49	−0.18	1.60	1.94†	−0.39	0.68	−0.18	1.25	1.87†	2.21†
HHSNTN	4	1.76†	2.15†	−1.66*	−1.87	−1.67*	−4.26	1.84†	2.15†	−1.05	−0.62	−1.25	−2.11
HMOB	5	0.82	2.76†	0.82	2.81†	0.84	2.45†	0.01	2.87†	0.02	2.22†	0.45	1.78†
HNIV	5	3.62†	6.08†	3.98†	3.57†	1.91†	1.94†	3.44†	2.35†	3.45†	2.79†	1.11	2.88†
HNR	5	1.72†	2.30†	0.63	3.67†	1.93†	2.58†	1.17	2.37†	1.14	3.55†	0.85	1.81†
HNS	5	0.80	3.16†	0.78	3.44†	0.75	4.78†	0.02	2.93†	0.31	1.44†	0.68	2.34†
HNSMW	5	0.81	2.82†	0.80	2.59†	0.78	1.94†	0.51	2.50†	0.29	3.75†	0.06	1.13
HNSNE	5	−0.10	−0.07	−2.93*	−1.31	0.26	2.64†	1.00	1.27	−2.62*	−0.62	0.49	4.21†
HNSSOU	5	0.94	2.85†	0.91	2.15†	0.89	3.42†	0.02	2.67†	0.03	1.02	0.18	1.18
HNSWST	5	0.36	2.89†	0.34	2.10†	0.33	3.04†	0.66	2.55†	0.35	2.78†	0.20	2.34†
HSBMW	4	0.73	3.54†	0.72	2.69†	0.67	3.08†	0.22	3.65†	0.15	2.91†	0.26	2.84†
HSBNE	4	−0.14	−0.10	−0.71	2.08†	−0.39	−0.70	1.14	1.26	−2.38*	−1.56	0.05	1.50†
HSBR	4	0.28	2.55†	0.21	2.85†	0.21	5.06†	0.28	2.97†	0.10	2.45†	0.09	2.61†
HSBSOU	4	0.52	3.85†	0.51	2.12†	0.50	4.22†	0.02	2.87†	0.05	2.26†	0.13	2.34†
HSBWST	4	0.15	6.66†	0.13	1.86†	0.14	3.33†	0.16	2.77†	0.09	2.10†	0.14	2.61†
HSFR	4	−1.73*	−0.87	0.10	0.41	0.45	3.50†	−1.65*	−0.83	−0.06	−0.02	0.32	2.89†
HSMW	4	−1.11	−1.46	−1.14	−1.99	−1.43	−2.88	2.05†	1.94†	1.62	3.21†	−1.65*	−0.65

HSNE	4	0.74	1.00	0.80	3.71†	−0.57	−2.59	0.85	1.01	0.65	3.08†	−0.26	−0.47
HSSOU	4	0.01	2.22†	0.02	2.22†	0.02	4.35†	0.40	2.93†	0.05	2.30†	0.21	2.96†
HSWST	4	0.02	2.81†	0.01	1.80†	0.01	3.45†	0.33	2.72†	0.01	2.99†	0.13	2.92†
IP	2	−0.73	0.67	0.07	1.46†	−0.61	0.38	−0.37	1.47†	0.42	2.15†	0.24	1.64†
IPC	1	1.49	1.73†	0.29	0.33	0.28	0.41	−0.35	0.13	0.80	0.70	−0.67	0.02
IPCD	1	−0.12	0.28	1.39	2.10†	−0.90	−0.57	0.10	2.54†	1.44	2.31†	0.23	0.65
IPCN	2	−1.64	−1.23	−1.7*	−1.31	−2.79*	−1.93	1.27	2.31†	0.08	0.14	−1.48	−1.47
IPD	2	0.22	1.09	−0.25	0.10	0.60	2.18†	−2.43*	−0.67	0.28	2.39†	−0.16	0.01
IPE	2	−2.85*	−2.08	−0.44	0.01	−0.05	0.69	−0.84	0.01	−0.22	0.60	1.06	1.99†
IPF	2	−2.91*	−2.35	−2.08*	−0.98	−2.03*	−1.14	−0.59	1.07	−0.18	−0.02	1.71†	2.76†
IPI	1	−2.20*	−1.29	−0.41	0.15	−0.97	0.07	−1.79*	−0.91	−0.02	1.74†	1.93†	1.80†
IPM	1	0.86	1.55†	−0.41	0.78	2.52†	2.91†	−0.55	1.09	−0.08	1.49†	−1.20	−0.87
IPMD	1	−0.80	0.31	1.31	3.92†	−0.42	−0.18	0.27	1.88†	0.67	2.61†	−1.23	−0.45
IPMFG	2	−1.60	0.09	0.35	2.02†	−0.28	1.03	−1.10	0.31	0.23	2.23†	0.14	1.68†
IPMIN	1	−1.38	−1.26	1.50	2.89†	1.27	1.89†	1.24	2.00†	1.16	2.48†	0.73	1.20
IPMND	1	−3.26*	−1.52	−0.41	0.20	−1.65*	−1.22	−0.34	0.72	−0.19	0.47	−0.03	0.03
IPN	1	0.27	1.48†	−0.83	−0.28	0.44	0.42	1.00	1.00	−0.17	0.34	1.40	1.25
IPP	2	−3.03*	−1.27	−2.09*	−0.82	−2.02*	−1.02	0.82	1.35†	−0.04	0.36	1.98†	1.90†
IPUT	2	−2.54*	−0.81	−0.36	0.12	−0.21	1.05	−2.72*	−0.72	−0.20	0.01	−0.41	−0.17
IPX	4	0.56	2.63†	0.29	3.80†	0.01	2.71†	1.06	2.28†	1.03	2.48†	0.48	2.92†
IPXDCA	4	0.14	2.15†	0.13	2.21†	0.11	2.43†	0.67	2.64†	0.19	1.02	0.06	2.18†
IPXMCA	3	0.28	2.94†	0.27	2.12†	0.28	2.41†	0.53	2.82†	0.60	1.17	0.25	2.32†
IPXMIN	4	0.44	2.80†	0.44	2.05†	0.48	2.26†	0.64	2.61†	0.41	2.94†	0.41	2.14†
IPXNCA	4	1.48	2.73†	0.12	3.24†	0.32	2.69†	1.16	8.21†	1.42	2.03†	1.09	2.35†
IPXUT	4	0.50	3.79†	0.49	1.86†	0.49	1.85†	0.40	2.65†	0.41	2.50†	0.00	2.14†
IVMFDQ	4	0.05	1.11	−1.00	−2.21	−0.84	−2.25	0.30	1.04	−0.46	−0.40	0.52	3.95†
IVMFGQ	3	0.43	1.26	−0.98	−2.30	−1.16	−2.71	0.55	1.47†	−0.22	0.12	−0.59	−0.84
IVMFNQ	2	6.97†	7.31†	0.32	0.99	−0.89	−1.30	0.38	2.26†	0.43	0.98	1.73†	2.10†
IVMTQ	3	−1.88*	−1.58	−1.92*	−3.16	−1.48	−3.48	−1.04	−0.39	−0.84	−0.89	−0.12	0.47
IVRRQ	3	−0.26	0.92	0.26	0.59	0.80	1.38†	−0.46	1.09	0.58	1.26	1.37	2.21†
IVSRMQ	1	1.45	2.04†	1.45	2.19†	0.59	1.16	2.01†	4.90†	1.29	2.39†	1.22	1.83†
IVSRQ	1	−0.70	0.90	1.00	2.26†	0.62	1.20	0.78	3.63†	1.12	2.72†	−0.53	0.62
IVSRRQ	1	4.40†	5.16†	0.40	1.97†	−0.86	−0.68	0.07	2.74†	1.94†	2.72†	0.24	0.43
IVSRWQ	0	−0.65	0.06	0.23	0.93	0.21	0.93	−0.74	−0.05	0.20	0.79	0.11	0.80
IVWRQ	2	−0.13	0.87	0.85	2.20†	−0.55	1.29†	−0.43	1.53†	−1.61	−0.15	2.02†	4.43†
LEH	4	1.69†	2.65†	−1.09	−0.45	−1.81*	−0.33	1.09	1.69†	−0.09	−0.05	1.19	1.86†
LEHCC	2	−0.40	−0.01	−0.15	−0.01	0.57	0.18	−0.20	0.82	−0.83	−0.76	−1.24	−1.22
LEHFR	3	−2.68*	−0.15	1.79†	2.81†	0.66	0.22	−2.33*	0.44	1.79†	2.81†	2.05†	2.44†
LEHM	2	−0.98	−1.29	−0.83	−0.46	0.76	0.14	−0.75	0.22	−0.93	−0.19	0.91	0.91
LEHS	3	1.70†	3.02†	1.05	2.05†	−1.52	4.27†	1.72†	3.77†	0.41	2.14†	−1.52	4.27†

Table 4 (*continued*)

Variable	SM rejec	Recursive estimation scheme						Rolling estimation scheme					
		1-month ahead		3-month ahead		12-month ahead		1-month ahead		3-month ahead		12-month ahead	
		DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t
LEHTT	4	1.70†	2.06†	1.34	2.59†	0.61	0.70	1.91†	2.81†	0.18	0.18	−0.20	−0.07
LEHTU	3	−2.22*	−2.23	1.56	0.31	−0.89	−0.47	1.97†	3.02†	0.92	0.76	1.34	1.08
LHEL	2	2.67†	4.45†	−1.54	2.11†	−1.33	2.07†	2.83†	2.82†	2.17†	2.74†	−1.25	2.66†
LHELX	4	1.30	4.10†	1.48	2.36†	−0.23	4.91†	3.08†	4.52†	0.66	5.16†	−0.18	4.77†
LHEM	1	−1.26	−0.64	−0.16	0.77	1.64	1.62†	−0.49	0.28	1.59	2.78†	2.59†	4.31†
LHNAG	2	−1.14	−1.84	−0.14	1.06	1.70†	3.10†	−0.49	−0.01	1.02	2.44†	2.54†	4.91†
LHU14	4	−2.29*	−1.44	−1.26	0.95	−0.65	−2.22	−1.84*	−0.50	0.10	2.88†	0.01	5.26†
LHU15	4	1.12	1.85†	0.82	4.15†	−0.63	1.19	0.93	1.70†	0.85	2.82†	−0.36	3.20†
LHU26	3	−0.38	0.71	−0.67	−0.22	−0.86	−1.04	0.14	1.22	−0.23	0.77	−0.45	0.62
LHU5	4	0.34	2.93†	0.77	3.21†	1.02	6.17†	0.15	2.04†	1.32	4.83†	1.28	2.45†
LHU680	4	−0.49	3.27†	0.49	2.85†	0.52	2.99†	−0.67	2.69†	0.32	5.65†	0.55	2.43†
LHUR	4	1.17	3.56†	0.81	2.68†	0.86	2.48†	1.11	1.62†	0.89	2.13†	1.19	2.36†
LP	1	1.58	2.11†	1.05	2.75†	1.37	6.66†	1.83†	2.03†	−0.86	−0.35	1.30	7.24†
LPCC	1	−1.98*	−0.34	−1.80*	−1.69	1.50	5.45†	−2.10*	−0.53	−1.80*	−1.49	−0.10	1.36†
LPED	1	1.65†	5.58†	0.98	2.87†	1.26	4.13†	−1.32	0.10	1.13	4.13†	1.18	3.91†
LPEM	2	1.23	6.11†	0.92	6.44†	0.36	3.93†	1.02	1.75†	0.85	4.45†	0.32	1.62†
LPEN	1	−3.24*	0.73	0.92	4.40†	0.81	4.31†	−3.70*	0.11	0.59	3.48†	0.11	1.70†
LPFR	4	0.29	0.92	0.54	3.09†	−0.10	2.63†	1.27	1.42†	1.01	4.29†	0.60	5.58†
LPGD	2	1.60	6.07†	−0.70	−1.41	0.84	4.00†	1.00	2.13†	0.97	4.81†	0.41	2.10†
LPGOV	4	0.50	1.06	1.74†	2.24†	1.58	1.99†	1.56	2.11†	1.08	2.06†	0.96	2.54†
LPHRM	3	2.09†	3.83†	1.81†	4.90†	0.17	2.26†	2.52†	2.58†	2.76†	2.30†	0.52	2.98†
LPMI	1	0.82	1.43†	0.12	0.29	0.59	2.11†	1.37	1.98†	1.50	2.04†	0.49	2.12†
LPMOSA	4	−0.43	2.25†	0.21	4.26†	0.34	2.76†	1.00	2.31†	−0.18	4.71†	0.04	2.52†
LPNAG	1	0.74	1.14	1.45	4.39†	1.38	8.33†	1.79†	1.70†	−0.72	−0.78	1.23	7.91†
LPS	2	−1.87*	−3.23	0.90	3.23†	1.03	5.64†	−1.91*	−2.58	1.00	3.98†	1.10	6.00†
LPSP	3	−1.60	−1.19	−0.76	−1.24	1.33	7.99†	2.22†	2.36†	0.93	1.93†	1.28	7.98†
LPT	3	−1.61	−3.94	1.46	4.73†	1.90†	8.17†	−2.18*	−1.90	−1.06	−2.31	1.88†	8.80†
LPTU	1	−0.86	1.62†	0.28	0.19	1.88†	0.85	−0.87	1.76†	−1.00	−0.93	1.41	1.30†
LW	1	−1.34	−0.64	1.21	1.63†	0.01	−0.60	−1.41	−0.30	1.32	1.77†	0.10	−0.15
MDO	1	−0.37	1.64†	0.06	1.55†	−1.42	−1.46	−0.45	1.37†	−0.36	0.83	−1.12	−1.25
MDOQ	0	−2.74*	−0.99	−0.26	0.98	1.53	2.84†	1.00	1.00	1.00	1.00	0.29	1.61†

MDU	4	−5.18*	−4.08	0.58	1.96†	1.06	4.07†	−4.45*	−2.34	−0.63	−0.53	−1.38	−3.76
MDUWU	1	0.52	0.92	0.52	0.70	−1.86*	−1.30	0.15	0.79	−1.61	−0.88	−1.78*	−1.59
MNO	2	0.10	1.96†	0.38	0.76	1.83†	3.19†	−0.92	0.17	−0.33	−0.10	1.28	2.41†
MNOU	2	1.93†	3.00†	0.75	0.99	2.13†	1.31†	2.48†	3.24†	0.46	0.62	1.33	0.63
MNU	0	2.78†	2.34†	−1.56	−1.33	0.71	1.77†	1.17	1.55†	0.30	0.90	0.61	2.12†
MO	2	0.74	1.62†	−1.64	−0.59	0.63	1.88†	1.26	2.12†	−0.14	0.86	2.05†	3.23†
MOCMQ	0	−0.44	0.60	−0.01	−0.19	−0.67	−0.75	1.86†	3.39†	0.30	0.81	−0.60	0.07
MOWU	1	−0.77	1.55†	0.00	1.64†	−1.14	−1.19	0.06	0.88	0.42	0.63	−1.37	−1.94
MPCON	1	0.83	1.63†	0.75	1.08	2.04†	1.21	1.48	2.51†	1.16	1.80†	0.05	0.19
MPCONQ	0	0.95	1.05	0.60	0.97	1.35	0.82	1.12	1.49†	0.81	1.30†	0.42	0.18
MSDQ	1	−0.23	−0.19	0.56	1.10	1.67†	2.00†	1.48	2.13†	1.90†	2.23†	0.72	1.38†
MSMQ	1	1.85†	2.18†	0.45	0.89	0.28	0.78	2.49†	2.78†	−0.23	0.59	−0.70	−0.03
MSMTQ	1	1.14	1.94†	−0.56	−0.51	1.48	1.45†	1.21	2.00†	−1.18	−1.14	0.39	0.48
MSNQ	0	3.97†	4.06†	−0.22	0.30	−0.99	−0.73	1.45	1.51†	0.15	0.92	−2.45*	−1.44
MSONDQ	0	−1.39	−0.20	0.55	1.01	−0.62	−0.38	0.81	4.43†	0.40	0.79	−0.17	−0.08
MU	4	−4.93*	−3.38	0.70	2.11†	−1.69*	−3.44	−5.09*	−3.78	−0.22	0.17	−1.69*	−3.55
PMDEL	3	1.25	2.40†	1.51	5.05†	1.53	2.91†	0.88	2.38†	1.43	4.93†	1.48	8.70†
PMEMP	3	0.39	2.07†	−0.79	−1.14	−0.31	1.20	1.95†	2.52†	1.81†	3.25†	0.70	3.62†
PMI	3	−0.64	−0.42	−0.25	−0.09	0.12	1.58†	−0.44	0.02	1.27	3.75†	0.91	4.07†
PMNO	3	0.68	1.17	0.01	1.24	0.02	1.94†	1.65†	1.70†	0.87	2.54†	0.56	2.73†
PMNV	3	1.82†	1.64†	1.98†	3.36†	1.28	3.52†	−2.46*	−1.55	−0.14	0.87	−0.21	1.31†
PMP	3	0.37	0.55	−0.27	0.71	−0.25	0.89	−0.75	−0.38	1.01	2.67†	0.62	3.04†
RTNQ	2	−0.96	0.31	−0.61	−0.65	−1.45	−1.85	−1.21	0.11	0.39	0.50	1.25	0.73
RTQ	1	−0.82	0.52	−0.98	−0.61	1.23	0.56	−0.87	0.84	−0.57	−0.27	1.38	0.58
WTDQ	2	−2.4*	−1.00	−0.23	−0.06	1.44	1.19	−2.68*	−1.20	−0.39	0.48	3.02†	2.66†
WTNQ	1	−0.75	0.06	−0.96	−1.22	−1.26	−1.65	−0.47	0.18	0.31	0.79	1.63	2.27†
WTQ	1	−1.00	−0.78	−0.19	−0.20	−0.64	−0.91	−0.64	−0.45	0.22	0.17	0.62	0.42

[a] *Notes*: See notes to Tables 2 and 3. Data used in this table correspond to those used in Stock and Watson (2002), and the variables are monthly, spanning the period 1959–1998. Reported results are based on predictive evaluation using the second half of the sample.

Table 5
Analysis of U.S. financial data (Stock and Watson data set)[a]

Variable	SM rejec	Recursive estimation scheme						Rolling estimation scheme					
		1-month ahead		3-month ahead		12-month ahead		1-month ahead		3-month ahead		12-month ahead	
		DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t
FSDXP	4	$2.55^{\dagger}$	$2.74^{\dagger}$	$2.47^{\dagger}$	$-5.76^{\dagger}$	1.43	$9.12^{\dagger}$	$1.80^{\dagger}$	$2.87^{\dagger}$	$2.36^{\dagger}$	$5.73^{\dagger}$	$1.93^{\dagger}$	$2.06^{\dagger}$
FSNFI	3	−0.14	0.40	−1.12	−0.41	1.36	$1.39^{\dagger}$	0.11	0.76	0.86	$1.50^{\dagger}$	0.78	0.80
FSNIN	2	1.17	$2.28^{\dagger}$	-2.39^{*}	−1.40	-2.6^{*}	−1.65	1.09	$2.28^{\dagger}$	0.19	0.41	−1.07	−1.03
FSNTR	3	$1.89^{\dagger}$	$3.31^{\dagger}$	−1.13	−0.48	−1.39	−1.40	0.94	$2.73^{\dagger}$	−1.11	−0.39	−1.51	−0.90
FSNUT	2	-1.73^{*}	−0.93	-1.77^{*}	−1.33	−0.98	0.20	-1.98^{*}	−1.22	-1.85^{*}	−1.21	−1.06	0.08
FSNVV3	1	-3.58^{*}	−0.87	0.28	0.45	−0.37	−0.30	−0.90	−0.14	−1.32	−1.50	0.54	1.03
FSTB	0	−0.70	$1.99^{\dagger}$	1.23	$1.31^{\dagger}$	1.00	1.02	−0.85	$1.43^{\dagger}$	$2.19^{\dagger}$	$2.06^{\dagger}$	0.40	0.28
FSTE	2	0.27	−0.53	0.29	0.37	-2.65^{*}	0.48	0.52	1.20	−0.38	−0.15	-2.69^{*}	0.43
FSTM	1	-2.7^{*}	−0.48	0.81	$1.33^{\dagger}$	-2.38^{*}	−1.23	−1.08	$1.78^{\dagger}$	1.58	1.22	-2.16^{*}	−0.99
FSNCOM	1	$1.67^{\dagger}$	$2.28^{\dagger}$	-1.80^{*}	−0.43	$2.24^{\dagger}$	$3.47^{\dagger}$	0.66	1.00	−1.16	0.50	-2.04^{*}	−1.65
FSPCAP	0	$2.52^{\dagger}$	$3.35^{\dagger}$	−0.15	0.78	1.25	$1.57^{\dagger}$	0.99	1.19	−0.72	0.84	−0.77	0.34
FSPCOM	0	$2.27^{\dagger}$	$2.85^{\dagger}$	-2.16^{*}	−0.78	$2.66^{\dagger}$	$4.14^{\dagger}$	1.43	$2.70^{\dagger}$	-2.63^{*}	−1.62	-2.27^{*}	−1.05
FSPFI	1	$1.66^{\dagger}$	$2.43^{\dagger}$	0.32	0.50	0.23	0.37	$1.79^{\dagger}$	$2.32^{\dagger}$	−1.35	−0.40	−1.63	−1.04
FSPIN	0	$1.92^{\dagger}$	$2.42^{\dagger}$	-2.12^{*}	−0.75	1.43	$1.86^{\dagger}$	0.60	0.88	−1.38	−0.47	-2.14^{*}	−1.63
FSPTR	0	−0.28	0.32	0.15	0.50	0.72	0.66	1.23	$2.82^{\dagger}$	−1.16	−0.70	0.13	0.84
FSPUT	2	$2.58^{\dagger}$	$3.26^{\dagger}$	-1.70^{*}	−0.79	−1.58	−0.97	0.26	$1.71^{\dagger}$	−1.41	−0.49	0.79	$1.94^{\dagger}$
FSPXE	4	−0.54	$2.47^{\dagger}$	0.18	$3.20^{\dagger}$	−0.31	−1.05	1.31	$2.57^{\dagger}$	0.40	$3.07^{\dagger}$	−0.10	−0.01
EXRCAN	3	0.69	1.18	0.00	0.72	1.24	$1.58^{\dagger}$	0.44	0.94	−0.31	0.23	$1.89^{\dagger}$	$3.24^{\dagger}$
EXRGER	1	$2.47^{\dagger}$	$3.60^{\dagger}$	0.16	0.74	1.43	$2.65^{\dagger}$	$2.33^{\dagger}$	$3.44^{\dagger}$	0.37	1.08	$1.88^{\dagger}$	$3.10^{\dagger}$
EXRJAN	1	−1.05	−0.15	0.93	$1.50^{\dagger}$	1.14	$1.70^{\dagger}$	−0.82	0.02	1.04	$1.73^{\dagger}$	1.26	$1.96^{\dagger}$
EXRSW	3	−0.61	−0.18	0.68	$1.29^{\dagger}$	$1.74^{\dagger}$	$3.08^{\dagger}$	0.29	0.87	1.29	$2.27^{\dagger}$	$1.66^{\dagger}$	$2.75^{\dagger}$

EXRUK	0	−1.00	−1.00	−0.52	−0.29	−1.16	−1.11	−0.88	−0.94	−0.49	−0.27	0.53	1.01
EXRUS	0	0.28	1.01	0.36	1.02	1.24	2.10†	0.64	1.47†	0.64	1.45†	1.06	2.00†
FYAAAC	0	−0.09	1.00	0.87	1.13	1.78†	2.29†	1.44	3.77†	1.01	1.32†	1.52	1.94†
FYBAAC	2	1.08	2.96†	0.81	1.67†	1.54	2.34†	1.41	1.77†	0.84	1.80†	1.49	2.38†
FYCP90	1	0.32	1.61†	0.92	0.81	1.51	1.74†	0.40	1.61†	0.54	1.44†	1.48	1.65†
FYFF	1	0.31	0.83	−2.69*	−1.32	1.45	2.25†	0.32	1.46†	−1.31	1.00	1.54	2.22†
FYFHA	0	1.82†	2.11†	1.15	1.19	1.75†	2.10†	0.74	1.06	1.55	2.58†	1.72†	2.14†
FYGM3	0	0.22	1.81†	0.86	0.63	0.73	0.90	0.15	1.75†	0.94	0.74	0.80	0.93
FYGM6	0	0.60	1.25	1.81†	1.27	1.49	2.19†	0.32	2.08†	1.99†	1.31†	1.59	1.42†
FYGT1	0	1.39	1.40†	0.57	0.59	1.39	1.39†	0.48	0.89	0.64	0.75	1.63	2.13†
FYGT10	0	1.11	2.76†	1.81†	2.62†	0.48	0.59	0.30	1.50†	1.75†	2.75†	0.31	0.41
FYGT5	0	1.19	2.38†	1.34	1.91†	1.35	2.72†	1.41	2.64†	−1.49	−0.57	1.29	2.48†
SFYAAAC	3	−0.22	0.67	0.27	2.63†	−0.85	0.46	0.39	0.99	0.79	3.69†	−0.65	0.78
SFYBAAC	3	−0.40	0.85	−0.08	2.42†	−0.91	0.95	0.37	1.32†	0.83	3.81†	−0.42	1.65†
SFYFHA	3	−0.29	0.80	−0.23	2.53†	−0.27	1.51†	0.28	1.25	0.17	3.21†	−0.28	1.46†
SFYGM3	3	−1.54	0.14	0.50	3.03†	−0.35	1.66†	−0.71	0.74	0.17	1.03	−0.51	0.55
SFYGM6	2	−1.32	0.12	0.18	2.50†	−1.18	3.04†	−0.66	0.57	0.59	2.16†	0.99	4.62†
SFYGT1	2	−1.36	−0.23	0.77	2.62†	1.15	3.19†	−1.18	0.22	1.07	3.02†	2.07†	4.82†
SFYGT10	2	1.37	1.16	1.14	3.98†	−0.12	1.16	1.89†	1.49†	1.22	4.74†	−1.13	1.25
SFYGT5	2	1.54	1.47†	1.27	4.10†	0.28	1.46†	1.91†	1.74†	1.15	4.48†	−0.94	1.60†
SFYCP90	2	−1.77*	0.29	−1.54	−0.43	−1.12	2.42†	−2.10*	−1.72	−1.09	1.10	0.46	4.75†

[a] *Notes*: See notes to Table 4. Data used in this table correspond to those used in Stock and Watson (2002), and the variables are monthly, spanning the period 1959–1998, under the sub-categories; stock prices, exchange rates and interest rates.

Table 6
Analysis of U.S. monetary data (Stock and Watson data set)[a]

Variable	SM rejec	Recursive estimation scheme						Rolling estimation scheme					
		1-month ahead		3-month ahead		12-month ahead		1-month ahead		3-month ahead		12-month ahead	
		DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t	DM	ENC-t
CCI30M	5	-2.95^{*}	-0.49	$1.98^{\dagger}$	$4.54^{\dagger}$	$2.09^{\dagger}$	$7.64^{\dagger}$	-1.27	$1.71^{\dagger}$	-0.50	0.96	$2.03^{\dagger}$	$7.9^{\dagger}$
CCINRV	4	0.03	0.87	0.65	$2.85^{\dagger}$	-1.00	-4.99	-0.74	-0.27	-0.87	-3.08	-1.19	-3.26
CCINT	4	-1.20	-0.38	-1.16	-0.75	-1.04	-0.15	-0.44	$1.31^{\dagger}$	0.36	$3.55^{\dagger}$	0.15	1.22
CCINV	4	0.66	$1.97^{\dagger}$	0.06	$1.42^{\dagger}$	1.45	$6.90^{\dagger}$	1.60	$3.38^{\dagger}$	1.49	$3.61^{\dagger}$	1.40	$8.46^{\dagger}$
FCLBMC	2	0.85	$1.49^{\dagger}$	-2.16^{*}	0.63	-0.60	-0.75	0.94	$1.52^{\dagger}$	-1.68^{*}	0.82	-0.55	-0.58
FCLIN	3	0.64	$15.8^{\dagger}$	1.23	$2.10^{\dagger}$	1.60	$10.01^{\dagger}$	0.76	$3.21^{\dagger}$	0.23	1.07	0.97	$4.66^{\dagger}$
FCLNBF	3	0.41	0.83	-0.02	0.80	-0.27	0.36	-0.89	-0.30	0.24	0.88	-0.70	-0.33
FCLNQ	3	1.01	1.20	-0.52	$1.68^{\dagger}$	-0.55	1.06	1.22	1.21	-0.75	0.90	-0.90	0.31
FCLRE	3	$3.16^{\dagger}$	$6.42^{\dagger}$	-0.04	$1.78^{\dagger}$	0.02	$5.15^{\dagger}$	-1.25	-0.19	1.13	$3.02^{\dagger}$	0.74	$6.15^{\dagger}$
FCLS	3	0.59	$1.39^{\dagger}$	0.05	$2.02^{\dagger}$	$1.67^{\dagger}$	$5.05^{\dagger}$	-0.93	0.07	0.46	$1.78^{\dagger}$	$2.10^{\dagger}$	$4.89^{\dagger}$
FCSGV	2	0.10	1.11	1.01	$3.55^{\dagger}$	0.34	$1.88^{\dagger}$	-1.08	-0.58	-0.39	0.16	0.53	$3.11^{\dagger}$
FM1	2	-0.30	1.13	-0.80	0.28	0.07	0.04	0.15	$2.50^{\dagger}$	-0.91	0.24	0.41	0.41
FM2	2	1.16	$2.03^{\dagger}$	-0.15	0.39	0.53	0.36	0.97	$1.30^{\dagger}$	-0.70	0.22	0.46	0.32
FM2DQ	3	-3.07^{*}	-0.36	-1.04	-0.95	-1.33	-2.13	-2.46^{*}	-1.65	-1.14	-0.57	-1.26	-0.86
FM3	2	-0.70	1.19	0.44	0.41	0.27	-0.18	$2.21^{\dagger}$	$4.31^{\dagger}$	0.11	1.03	-0.67	-0.50
FMFBA	2	0.71	$2.40^{\dagger}$	-0.32	-0.56	-0.56	0.16	0.78	$1.89^{\dagger}$	0.41	0.23	0.88	0.52
FML	2	-0.93	$1.75^{\dagger}$	0.67	1.21	0.61	0.19	-1.70^{*}	0.19	0.89	$1.63^{\dagger}$	-0.25	-0.09
FMRNBC	2	0.00	1.13	0.45	-0.02	1.12	0.35	0.21	$1.58^{\dagger}$	0.47	$1.37^{\dagger}$	-1.20	-1.03
FMRRA	2	0.65	$1.98^{\dagger}$	1.12	0.75	0.08	0.07	0.40	$1.89^{\dagger}$	-1.12	-0.45	0.70	0.81
GMDC	3	-2.58^{*}	-1.15	1.57	0.66	0.29	0.30	-2.51^{*}	-1.06	0.57	0.72	0.91	0.63
GMDCD	2	-1.56	1.18	0.67	0.66	1.09	0.45	-2.2^{*}	0.35	$2.21^{\dagger}$	$2.33^{\dagger}$	0.92	$1.32^{\dagger}$
GMDCN	2	-1.39	0.81	-0.46	-0.26	0.64	0.33	-1.65^{*}	0.47	0.97	0.89	-1.72^{*}	-1.12

GMDCS	2	−1.24	−0.50	0.93	1.82†	1.77†	0.45	1.15	1.37†	0.93	0.72	0.58	0.32
PCGOLD	3	−1.68*	−0.45	0.02	0.22	0.61	0.33	−0.18	1.15	0.84	1.24	0.84	0.98
PMCP	3	−0.97	0.82	1.62	3.44†	1.34	2.99†	0.34	2.36†	1.62	5.11†	1.24	8.85†
PSM99Q	2	−0.41	1.45†	−0.61	0.46	−0.22	−0.23	−0.19	1.65†	−1.09	−0.04	−1.37	−1.06
PU81	3	1.23	2.47†	1.11	0.51	1.23	0.29	0.43	1.66†	0.20	0.19	1.02	1.07
PU83	3	0.48	0.47	−2.25*	−1.39	1.04	1.31†	0.84	1.79†	−1.92*	−0.95	−0.15	−0.11
PU84	3	−1.46	0.48	−0.18	0.06	0.30	0.20	−1.10	0.53	−0.88	−0.54	−1.27	−0.88
PU85	2	−0.25	1.15	−0.95	−0.27	−1.14	−0.46	−0.34	0.08	−0.27	−0.27	0.30	0.54
PUC	3	2.07†	3.85†	0.80	0.49	0.98	0.89	1.94†	3.96†	0.21	0.09	0.01	−0.17
PUCD	2	−0.90	−0.35	1.39	1.63†	−0.08	0.10	−2.70*	−0.20	0.74	0.65	−0.34	−0.10
PUH	3	1.40	2.74†	2.48†	1.68†	0.45	0.22	1.89†	2.95†	1.56	1.39†	1.22	0.86
PUNEW	2	1.59	2.30†	0.53	0.57	1.17	1.16	1.44	1.83†	1.10	1.28	0.28	0.53
PUS	2	0.11	1.74†	−0.42	0.88	−0.86	−0.74	1.10	2.28†	0.08	3.04†	−1.04	−1.07
PUXF	1	−2.04*	−0.92	0.20	0.24	−0.90	−0.65	−2.10*	−1.14	0.57	0.49	0.78	0.49
PUXHS	2	−1.10	0.64	2.71†	1.44†	−0.87	−0.16	2.04†	4.10†	1.36	1.14	−0.63	−0.39
PUXM	2	−0.52	0.39	0.66	0.63	0.12	0.32	−0.05	0.86	0.63	0.66	0.79	0.76
PW150A	4	3.23†	8.14†	0.99	0.67	−0.28	−0.13	−0.08	0.91	0.49	0.61	−0.12	0.12
PW160A	3	0.45	2.39†	−0.13	−0.10	1.39	1.14	1.12	1.74†	0.54	0.43	−2.21*	−1.67
PWCMSA	2	1.22	2.15†	−1.04	−0.64	−0.40	−0.09	0.33	1.46†	1.68†	1.75†	−0.84	−1.02
PWFCSA	2	1.96†	2.49†	−0.35	−0.13	1.22	0.54	2.35†	3.20†	−0.42	−0.25	2.15†	1.20
PWFSA	2	2.00†	2.83†	0.74	0.59	0.02	0.16	2.31†	2.47†	1.65†	1.51†	−0.73	−0.44
PWFXSA	4	−2.03*	−0.53	−1.30	−1.02	0.56	0.34	1.20	1.44†	−0.77	−0.31	1.79†	1.27
PWIMSA	2	−1.52	−0.77	−0.95	−0.49	−2.41*	−1.08	−0.71	0.42	−1.93*	−0.96	1.67†	2.64†

[a] *Notes*: See notes to Table 4. Data used in this table correspond to those used in Stock and Watson (2002), and the variables are monthly, spanning the period 1959–1998, under the sub-categories; money and credit quantity aggregates and price indexes.

estimated extremely imprecisely with such small samples. Second, small samples of data choose the less parsimonious ARFIMA model as often as the non-ARFIMA model. This is again surprising, given that all experiments are truly ex ante.

4. Experimental evidence

Table 7 summarizes the DGPs and parameterizations considered in our experiments. The generic DGPs are the same as those used in our empirical analysis. For the ARFIMA models, data are generated using fractional values of d over the interval $(0, 1)$, including $d = \{0.20, 0.30, 0.40, 0.60, 0.90\}$. Additionally, MA(1) and AR(1) coefficients were specified, including $\{0.0, -0.5\}$ (MA) and $\{0.3, 0.6, 0.9\}$ (AR). Thus, we examine 35 different ARFIMA specifications. When generating ARFIMA data, we used an arbitrary cut-off of $1.0e - 004$. Terms in the polynomial expansion with coefficients smaller in absolute value than this cut-off were truncated. All DGPs include at most one lag, so that AR models have one autoregressive lag and MA models have one moving average lag, etc. All variables are generated using standard normal errors. In the non-ARFIMA models, autoregressive slope parameters considered include $\{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, the MA models have coefficients equal to $\{-0.7, -0.4, -0.1, 0.2, 0.3, 0.4, 0.5, 0.9\}$, and values of d equal to 0 and 1 are considered. For the GARCH DGP, 8 different specifications were considered. All of the parameterizations in this case were chosen to mirror the types of parameters observed when estimating the models using our stock market and macroeconomic variables. Samples of $T = \{1000, 4000\}$ were used. Given the generated data, all analysis was carried out in exactly the same way as for our empirical examples. In particular, a "best" ARFIMA and non-ARFIMA model was first selected using point MSFE comparison of recursive (rolling) predictions based on the first-half of the sample, for 3 prediction horizons (1-, 5-, and 20-step ahead). Then, the second-half of the sample was used for ex ante comparison of the 2 models, again using either recursive or rolling estimation schemes, and for all three horizons. All results are based on 500 Monte Carlo replications.

A summary of our experimental findings is given in Table 8(1) and (2) (ARFIMA DGPs) and Table 9(1) and (2) (non-ARFIMA DGPs). The tables report the proportion of times that the ARFIMA models win a forecasting competition, based on direct comparison of point MSFE first entry in each bracketed trio of numbers and based on 10% level DM tests (second entry). The last entry in each bracketed group of numbers reports the proportion of times that the encompassing null hypothesis fails to reject. Thus, all entries report various measures of the proportional of ARFIMA model "wins". Columns in the tables refer to the estimation scheme used, and to the forecast horizon. The clear pattern that emerges when comparing Tables 8(1) and (2) is that the proportion of times that the ARFIMA model wins (when the true DGP is an ARFIMA process) increases rather substantially when the sample is increased from 1000 to 4000 observations. Note also that the first-half of the sample is used to select the ARFIMA and non-ARFIMA

Table 7
Data generating processes used in the Monte Carlo experiments[a]

Name	DGP, ARFIMA(p,d,q)	Name	DGP,Non-ARFIMA
ARFIMA1	$p=0,\ d=0.2,\ q=0$	RW	Random walk without drift
ARFIMA2	$p=1,\ d=0.2,\ q=0$, AR coef $=0.3$	AR1	AR$(p), p=1$, AR coef $=0.2$
ARFIMA3	$p=1,\ d=0.2,\ q=1$, AR coef $=0.3$, MA coef $=-0.5$	AR2	AR$(p), p=1$, AR coef $=0.3$
ARFIMA4	$p=1,\ d=0.2,\ q=0$, AR coef $=0.6$	AR3	AR$(p),\ p=1$, AR coef $=0.4$
ARFIMA5	$p=1,\ d=0.2,\ q=1$, AR coef $=0.6$, MA coef $=-0.5$	AR4	AR$(p), p=1$, AR coef $=0.5$
ARFIMA6	$p=1,\ d=0.2,\ q=0$, AR coef $=0.9$	AR5	AR$(p),\ p=1$, AR coef $=0.6$
ARFIMA7	$p=1,\ d=0.2,\ q=1$, AR coef $=0.9$, MA coef $=-0.5$	AR6	AR$(p), p=1$, AR coef $=0.7$
ARFIMA8	$p=0,\ d=0.3,\ q=0$	AR7	AR$(p), p=1$, AR coef $=0.8$
ARFIMA9	$p=1,\ d=0.3,\ q=0$, AR coef $=0.3$	AR8	AR$(p), p=1$, AR coef $=0.9$
ARFIMA10	$p=1,\ d=0.3,\ q=1$, AR coef $=0.3$, MA coef $=-0.5$	MA1	MA$(q), q=1$, MA coef $=-0.7$
ARFIMA11	$p=1,\ d=0.3,\ q=0$, AR coef $=0.6$	MA2	MA$(q), q=1$, MA coef $=-0.4$
ARFIMA12	$p=1,\ d=0.3,\ q=1$, AR coef $=0.6$, MA coef $=-0.5$	MA3	MA$(q), q=1$, MA coef $=-0.1$
ARFIMA13	$p=1,\ d=0.3,\ q=0$, AR coef $=0.9$	MA4	MA$(q), q=1$, MAcoef $=0.2$
ARFIMA14	$p=1,\ d=0.3,\ q=1$, AR coef $=0.9$, MA coef $=-0.5$	MA5	$MA(q), q=1$, MA coef $=0.3$
ARFIMA15	$p=0,\ d=0.4,\ q=0$	MA6	MA$(q), q=1$, MA coef $=0.4$
ARFIMA16	$p=1,\ d=0.4,\ q=0$, AR coef $=0.3$	MA7	MA$(q), q=1$, MA coef $=0.5$
ARFIMA17	$p=1,\ d=0.4,\ q=1$, AR coef $=0.3$, MA coef $=-0.5$	MA8	MA$(q), q=1$, MA coef $=0.9$
ARFIMA18	$p=1,\ d=0.4,\ q=0$, AR coef $=0.6$	GARCH1	AR-GARCH AR coef $=-0.3$ ARCH coef $=0.20$, GARCH Coef $=0.45$
ARFIMA19	$p=1,\ d=0.4,\ q=1$, AR coef $=0.6$, MA coef $=-0.5$	GARCH2	AR-GARCH AR coef $=-0.2$ ARCH coef $=0.25$, GARCH Coef $=0.10$
ARFIMA20	$p=1,\ d=0.4,\ q=0$, AR coef $=0.9$	GARCH3	AR-GARCH AR coef $=-0.1$ ARCH coef $=0.35$, GARCH Coef $=0.30$
ARFIMA21	$p=1,\ d=0.4,\ q=1$, AR coef $=0.9$, MA coef $=-0.5$	GARCH4	AR-GARCH AR coef $=0.2$ ARCH coef $=0.15$, GARCH Coef $=0.80$
ARFIMA22	$p=0,\ d=0.6,\ q=0$	GARCH5	AR-GARCH AR coef $=0.3$ ARCH coef $=0.15$, GARCH Coef $=0.75$
ARFIMA23	$p=1,\ d=0.6,\ q=0$, AR coef $=0.3$	GARCH6	AR-GARCH AR coef $=0.5$ ARCH coef $=0.20$, GARCH Coef $=0.70$
ARFIMA24	$p=1,\ d=0.6,\ q=1$, AR coef $=0.3$, MA coef $=0.5$	CARCH7	AR-GARCH AR coef $=0.7$ ARCH coef $=0.20$, GARCH Coef $=0.70$
ARFIMA25	$p=1,\ d=0.6,\ q=0$, AR coef $=0.6$	GARCH8	AR-GARCH AR coef $=0.9$ ARCH coef $=0.10$, GARCH Coef $=0.80$
ARFIMA26	$p=1,\ d=0.6,\ q=1$, AR coef $=0.6$, MA coef $=-0.5$	ARIMA1	ARIMA$(p,d,q), p=1,\ d=0,\ q=1$, AR coef $=0.1$,MA coef $=-0.52$
ARFIMA27	$p=1,\ d=0.6,\ q=0$, AR coef $=0.9$	ARIMA2	ARIMA$(p,d,q), p=1, d=0, q=1$, AR coef $=0.3$, MA coef $=-0.9$
ARFIMA28	$p=1,\ d=0.6,\ q=1$, AR coef $=0.9$, MA coef $=-0.5$	ARIMA3	ARIMA$(p,d,q), p=1,\ d=0,\ q=1$, AR coef $=0.6$, MA coef $=-0.2$
ARFIMA29	$p=0,\ d=0.9,\ q=0$	ARIMA4	ARIMA$(p,d,q),\ p=1,\ d=0,\ q=1$, AR coef $=0.9$, MA coef $=-0.7$
ARFIMA30	$p=1,\ d=0.9,\ q=0$, AR coef $=0.3$	ARIMA5	ARIMA$(p,d,q), p=1,\ d=1,\ q=1$, AR coef $=0.1$, MA coef $=-0.2$
ARFIMA31	$p=1,\ d=0.9,\ q=1$, AR coef $=0.3$, MA coef $=-0.5$	ARIMA6	ARIMA$(p,d,q),\ p=1,\ d=1,\ q=1$, AR coef $=0.3$, MA coef $=-0.9$
ARFIMA32	$p=1,\ d=0.9,\ q=0$, AR coef $=0.6$	ARIMA7	ARIMA$(p,d,q),\ p=1,\ d=1,\ q=1$, AR coef $=0.6$, MA coef $=-0.2$
ARFIMA33	$p=1,\ d=0.9,\ q=1$, AR coef $=0.6$, MA coef $=-0.5$	ARIMA8	ARIMA$(p,d,q), p=1, d=1, q=1$, AR coef $=0.9$, MA coef $=-0.7$
ARFIMA34	$p=1,\ d=0.9,\ q=0$, AR coef $=0.9$		
ARFIMA35	$p=1,\ d=0.9,\ q=1$, AR coef $=0.9$, MA coef $=-0.5$		

[a] *Notes*: This table summarizes all of the DGPs used in the Monte Carlo experiments reported in Section 4 of the paper. Parameters of non-ARFIMA DGPs are chosen on the basis of fitting these linear models to selected SW, LHM and DGE data, as discussed in Section 3. Parameters for the ARFIMA DGPs are based on values reported in the literature.

Table 8
Relative forecasting performance ARFIMA DGP—sample size: (1) 1000 and (2) 4000[a]

Model	Rec-1	Rec-5	Rec-20	Rol-1	Rol-5	Rol-20
ARFIMA1	(0.57, 0.19, 0.58)	(0.60, 0.21, 0.60)	(0.41, 0.11, 0.43)	(0.60, 0.29, 0.60)	(0.64, 0.19, 0.60)	(0.31, 0.10, 0.30)
ARFIMA2	(0.46, 0.15, 0.40)	(0.69, 0.25, 0.58)	(0.45, 0.12, 0.39)	(0.45, 0.11, 0.33)	(0.61, 0.21, 0.48)	(0.31, 0.08, 0.25)
ARFIMA3	(0.50, 0.15, 0.54)	(0.61, 0.23, 0.66)	(0.42, 0.12, 0.50)	(0.55, 0.12, 0.51)	(0.61, 0.19, 0.63)	(0.33, 0.10, 0.49)
ARFIMA4	(0.57, 0.16, 0.65)	(0.47, 0.18, 0.36)	(0.41, 0.02, 0.27)	(0.44, 0.14, 0.57)	(0.51, 0.20, 0.41)	(0.31, 0.02, 0.19)
ARFIMA5	(0.45, 0.12, 0.42)	(0.62, 0.19, 0.56)	(0.40, 0.09, 0.40)	(0.43, 0.16, 0.40)	(0.60, 0.22, 0.55)	(0.33, 0.09, 0.24)
ARFIMA6	(0.60, 0.21, 0.69)	(0.41, 0.12, 0.39)	(0.49, 0.09, 0.26)	(0.56, 0.18, 0.68)	(0.50, 0.17, 0.44)	(0.45, 0.10, 0.13)
ARFIMA7	(0.61, 0.19, 0.71)	(0.40, 0.13, 0.40)	(0.53, 0.09, 0.23)	(0.66, 0.20, 0.72)	(0.40, 0.11, 0.42)	(0.51, 0.08, 0.15)
ARFIMA8	(0.65, 0.22, 0.35)	(0.78, 0.26, 0.57)	(0.61, 0.13, 0.48)	(0.35, 0.17, 0.35)	(0.65, 0.39, 0.57)	(0.61, 0.22, 0.48)
ARFIMA9	(0.55, 0.28, 0.53)	(0.78, 0.40, 0.60)	(0.47, 0.11, 0.36)	(0.49, 0.24, 0.44)	(0.74, 0.39, 0.48)	(0.46, 0.16, 0.30)
ARFIMA10	(0.41, 0.15, 0.42)	(0.54, 0.23, 0.54)	(0.48, 0.16, 0.48)	(0.53, 0.32, 0.53)	(0.53, 0.23, 0.50)	(0.46, 0.18, 0.47)
ARFIMA11	(0.61, 0.27, 0.73)	(0.70, 0.24, 0.50)	(0.46, 0.10, 0.37)	(0.63, 0.18, 0.67)	(0.60, 0.24, 0.44)	(0.49, 0.08, 0.24)
ARFIMA12	(0.60, 0.23, 0.52)	(0.68, 0.23, 0.49)	(0.49, 0.16, 0.38)	(0.47, 0.12, 0.46)	(0.66, 0.29, 0.50)	(0.50, 0.18, 0.37)
ARFIMA13	(0.63, 0.29, 0.71)	(0.55, 0.15, 0.52)	(0.55, 0.08, 0.33)	(0.68, 0.29, 0.77)	(0.51, 0.10, 0.42)	(0.59, 0.11, 0.30)
ARFIMA14	(0.65, 0.27, 0.74)	(0.53, 0.14, 0.50)	(0.59, 0.12, 0.38)	(0.66, 0.28, 0.76)	(0.52, 0.10, 0.46)	(0.60, 0.14, 0.34)
ARFIMA15	(0.50, 0.25, 0.38)	(0.75, 0.38, 0.50)	(0.88, 0.50, 0.50)	(0.50, 0.25, 0.38)	(0.75, 0.38, 0.38)	(0.88, 0.25, 0.38)
ARFIMA16	(0.75, 0.51, 0.62)	(0.83, 0.43, 0.54)	(0.60, 0.33, 0.39)	(0.77, 0.42, 0.63)	(0.80, 0.48, 0.52)	(0.66, 0.32, 0.39)
ARFIMA17	(0.29, 0.07, 0.36)	(0.57, 0.21, 0.43)	(0.57, 0.32, 0.46)	(0.39, 0.18, 0.36)	(0.43, 0.21, 0.32)	(0.57, 0.36, 0.36)
ARFIMA18	(0.79, 0.43, 0.80)	(0.71, 0.35, 0.50)	(0.66, 0.28, 0.44)	(0.72, 0.39, 0.78)	(0.69, 0.32, 0.46)	(0.66, 0.22, 0.41)
ARFIMA19	(0.79, 0.41, 0.68)	(0.74, 0.35, 0.53)	(0.66, 0.33, 0.49)	(0.65, 0.27, 0.55)	(0.75, 0.32, 0.54)	(0.67, 0.29, 0.42)
ARFIMA20	(0.55, 0.15, 0.61)	(0.63, 0.24, 0.52)	(0.64, 0.16, 0.44)	(0.57, 0.21, 0.66)	(0.53, 0.19, 0.41)	(0.49, 0.08, 0.27)
ARFIMA21	(0.40, 0.20, 0.40)	(0.80, 0.40, 0.80)	(1.00, 0.40, 0.60)	(0.60, 0.40, 0.40)	(0.60, 0.40, 0.60)	(0.40, 0.40, 0.20)
ARFIMA22	(0.73, 0.45, 0.64)	(0.81, 0.47, 0.59)	(0.84, 0.56, 0.49)	(0.71, 0.41, 0.64)	(0.67, 0.34, 0.42)	(0.83, 0.48, 0.48)
ARFIMA23	(0.78, 0.40, 0.67)	(0.85, 0.57, 0.50)	(0.75, 0.48, 0.45)	(0.65, 0.31, 0.45)	(0.77, 0.46, 0.38)	(0.64, 0.34, 0.34)
ARFIMA24	(0.71, 0.21, 0.61)	(0.79, 0.32, 0.46)	(0.79, 0.46, 0.36)	(0.57, 0.25, 0.54)	(0.61, 0.25, 0.39)	(0.79, 0.29, 0.29)
ARFIMA25	(0.63, 0.30, 0.65)	(0.61, 0.32, 0.50)	(0.78, 0.50, 0.56)	(0.55, 0.23, 0.58)	(0.50, 0.24, 0.32)	(0.65, 0.24, 0.36)
ARFIMA26	(0.66, 0.37, 0.67)	(0.78, 0.48, 0.56)	(0.77, 0.47, 0.46)	(0.64, 0.30, 0.55)	(0.66, 0.35, 0.42)	(0.77, 0.35, 0.43)
ARFIMA27	(0.46, 0.19, 0.49)	(0.52, 0.16, 0.33)	(0.54, 0.16, 0.32)	(0.46, 0.27, 0.26)	(0.52, 0.14, 0.34)	(0.46, 0.19, 0.25)
ARFIMA28	(0.37, 0.09, 0.39)	(0.42, 0.08, 0.27)	(0.40, 0.11, 0.22)	(0.36, 0.07, 0.30)	(0.42, 0.04, 0.24)	(0.28, 0.09, 0.15)
ARFIMA29	(0.63, 0.27, 0.59)	(0.61, 0.32, 0.40)	(0.65, 0.41, 0.43)	(0.59, 0.27, 0.47)	(0.61, 0.25, 0.36)	(0.59, 0.27, 0.39)
ARFIMA30	(0.61, 0.26, 0.49)	(0.68, 0.28, 0.39)	(0.48, 0.29, 0.28)	(0.54, 0.14, 0.40)	(0.52, 0.22, 0.27)	(0.37, 0.16, 0.24)
ARFIMA31	(0.72, 0.41, 0.71)	(0.70, 0.36, 0.52)	(0.70, 0.44, 0.46)	(0.73, 0.34, 0.60)	(0.64, 0.29, 0.47)	(0.73, 0.42, 0.41)
ARFIMA32	(0.34, 0.04, 0.37)	(0.29, 0.09, 0.21)	(0.42, 0.21, 0.27)	(0.35, 0.10, 0.41)	(0.33, 0.08, 0.18)	(0.33, 0.15, 0.24)
ARFIMA33	(0.51, 0.26, 0.50)	(0.63, 0.26, 0.40)	(0.55, 0.26, 0.33)	(0.52, 0.20, 0.45)	(0.61, 0.24, 0.35)	(0.45, 0.21, 0.28)
ARFIMA34	(0.43, 0.19, 0.16)	(0.43, 0.17, 0.28)	(0.49, 0.17, 0.21)	(0.24, 0.12, 0.21)	(0.26, 0.15, 0.07)	(0.24, 0.07, 0.08)
ARFIMA35	(0.23, 0.09, 0.20)	(0.23, 0.07, 0.18)	(0.39, 0.07, 0.11)	(0.14, 0.02, 0.11)	(0.16, 0.05, 0.07)	(0.23, 0.05, 0.07)
ARFIMA1	(0.61, 0.33, 0.50)	(0.78, 0.44, 0.56)	(0.67, 0.44, 0.61)	(0.56, 0.33, 0.44)	(0.94, 0.56, 0.83)	(0.83, 0.44, 0.83)
ARFIMA2	(0.78, 0.41, 0.62)	(0.84, 0.59, 0.70)	(0.81, 0.35, 0.68)	(0.70, 0.46, 0.68)	(0.89, 0.49, 0.65)	(0.76, 0.27, 0.54)

ARFIMA3	(0.54, 0.27, 0.39)	(0.63, 0.37, 0.63)	(0.63, 0.39, 0.73)	(0.68, 0.34, 0.51)	(0.83, 0.41, 0.76)	(0.71, 0.39, 0.73)
ARFIMA4	(0.90, 0.80, 0.85)	(0.90, 0.70, 0.80)	(0.90, 0.60, 0.82)	(0.88, 0.70, 0.84)	(0.87, 0.54, 0.51)	(0.89, 0.50, 0.70)
ARFIMA5	(0.92, 0.71, 0.82)	(0.89, 0.65, 0.75)	(0.85, 0.63, 0.84)	(0.85, 0.62, 0.75)	(0.94, 0.55, 0.85)	(0.91, 0.51, 0.70)
ARFIMA6	(0.72, 0.29, 0.55)	(0.78, 0.35, 0.48)	(0.78, 0.31, 0.44)	(0.87, 0.39, 0.74)	(0.75, 0.27, 0.35)	(0.78, 0.25, 0.48)
ARFIMA7	(0.71, 0.21, 0.55)	(0.75, 0.25, 0.45)	(0.79, 0.45, 0.64)	(0.58, 0.30, 0.47)	(0.78, 0.32, 0.35)	(0.87, 0.35, 0.43)
ARFIMA8	(0.78, 0.60, 0.70)	(0.85, 0.65, 0.75)	(0.90, 0.75, 0.70)	(0.75, 0.54, 0.55)	(0.85, 0.71, 0.67)	(0.95, 0.65, 0.66)
ARFIMA9	(0.78, 0.45, 0.67)	(0.72, 0.40, 0.61)	(0.86, 0.64, 0.71)	(0.65, 0.56, 0.60)	(0.80, 0.56, 0.65)	(0.85, 0.65, 0.76)
ARFIMA10	(0.42, 0.15, 0.25)	(0.65, 0.32, 0.55)	(0.82, 0.55, 0.72)	(0.31, 0.13, 0.25)	(0.53, 0.25, 0.41)	(0.82, 0.61, 0.53)
ARFIMA11	(0.97, 0.80, 0.85)	(0.94, 0.79, 0.89)	(0.95, 0.82, 0.80)	(0.98, 0.94, 0.95)	(0.91, 0.80, 0.65)	(0.94, 0.88, 0.68)
ARFIMA12	(0.90, 0.85, 0.85)	(0.95, 0.75, 0.7)	(0.90, 0.75, 0.65)	(0.89, 0.80, 0.85)	(0.90, 0.71, 0.80)	(0.95, 0.88, 0.74)
ARFIMA13	(0.88, 0.63, 0.75)	(0.75, 0.35, 0.65)	(0.70, 0.35, 0.45)	(0.80, 0.31, 0.68)	(0.85, 0.35, 0.50)	(0.75, 0.35, 0.55)
ARFIMA14	(0.85, 0.67, 0.74)	(0.83, 0.54, 0.58)	(0.89, 0.52, 0.38)	(0.86, 0.48, 0.78)	(0.82, 0.40, 0.56)	(0.80, 0.44, 0.54)
ARFIMA15	(0.85, 0.65, 0.65)	(0.90, 0.60, 0.55)	(0.85, 0.55, 0.50)	(0.91, 0.62, 0.65)	(0.94, 0.60, 0.55)	(0.88, 0.60, 0.45)
ARFIMA16	(0.65, 0.43, 0.44)	(0.64, 0.34, 0.35)	(0.84, 0.45, 0.55)	(0.54, 0.45, 0.50)	(0.60, 0.40, 0.45)	(0.65, 0.45, 0.35)
ARFIMA17	(0.78, 0.55, 0.54)	(0.85, 0.45, 0.60)	(1.00, 0.65, 0.69)	(0.65, 0.45, 0.55)	(0.75, 0.45, 0.55)	(0.85, 0.65, 0.55)
ARFIMA18	(0.95, 0.85, 0.89)	(1.00, 0.95, 0.55)	(1.00, 0.85, 0.77)	(1.00, 0.85, 0.95)	(1.00, 0.9, 0.70)	(1.00, 0.85, 0.70)
ARFIMA19	(0.95, 0.80, 0.82)	(0.95, 0.70, 0.55)	(0.94, 0.78, 0.65)	(0.95, 0.80, 0.75)	(0.94, 0.70, 0.65)	(0.96, 0.65, 0.75)
ARFIMA20	(0.75, 0.38, 0.68)	(0.75, 0.44, 0.64)	(0.84, 0.42, 0.54)	(0.65, 0.41, 0.60)	(0.74, 0.45, 0.65)	(0.75, 0.30, 0.55)
ARFIMA21	(0.84, 0.42, 0.74)	(0.82, 0.46, 0.78)	(0.99, 0.45, 0.68)	(0.76, 0.40, 0.64)	(0.86, 0.40, 0.66)	(0.94, 0.44, 0.62)
ARFIMA22	(0.95, 0.55, 0.75)	(0.94, 0.58, 0.67)	(1.00, 0.64, 0.76)	(0.76, 0.40, 0.46)	(0.86, 0.32, 0.40)	(0.85, 0.36, 0.48)
ARFIMA23	(0.80, 0.51, 0.65)	(0.76, 0.45, 0.52)	(0.80, 0.35, 0.65)	(0.72, 0.51, 0.72)	(0.63, 0.45, 0.51)	(0.65, 0.32, 0.41)
ARFIMA24	(0.77, 0.31, 0.60)	(0.77, 0.32, 0.46)	(0.89, 0.56, 0.66)	(0.77, 0.35, 0.60)	(0.68, 0.30, 0.44)	(0.84, 0.32, 0.48)
ARFIMA25	(0.91, 0.79, 0.95)	(0.90, 0.80, 0.76)	(0.94, 0.76, 0.52)	(0.95, 0.86, 0.95)	(0.94, 0.76, 0.62)	(0.92, 0.62, 0.57)
ARFIMA26	(0.64, 0.67, 0.83)	(0.83, 0.67, 0.83)	(0.81, 0.50, 0.65)	(0.67, 0.50, 0.64)	(0.67, 0.43, 0.31)	(0.83, 0.33, 0.50)
ARFIMA27	(0.64, 0.30, 0.60)	(0.66, 0.38, 0.60)	(0.82, 0.32, 0.70)	(0.53, 0.31, 0.40)	(0.65, 0.32, 0.42)	(0.58, 0.31, 0.36)
ARFIMA28	(0.55, 0.32, 0.48)	(0.63, 0.34, 0.45)	(0.64, 0.30, 0.31)	(0.54, 0.25, 0.50)	(0.63, 0.31, 0.41)	(0.55, 0.25, 0.31)
ARFIMA29	(0.65, 0.32, 0.52)	(0.68, 0.35, 0.42)	(0.65, 0.44, 0.42)	(0.67, 0.35, 0.53)	(0.70, 0.41, 0.61)	(0.64, 0.34, 0.44)
ARFIMA30	(0.60, 0.28, 0.50)	(0.66, 0.30, 0.41)	(0.50, 0.24, 0.30)	(0.52, 0.20, 0.38)	(0.48, 0.24, 0.36)	(0.42, 0.26, 0.30)
ARFIMA31	(0.64, 0.32, 0.56)	(0.62, 0.36, 0.44)	(0.66, 0.40, 0.42)	(0.63, 0.30, 0.56)	(0.64, 0.30, 0.40)	(0.62, 0.40, 0.34)
ARFIMA32	(0.54, 0.24, 0.30)	(0.49, 0.19, 0.28)	(0.44, 0.20, 0.32)	(0.45, 0.19, 0.38)	(0.40, 0.17, 0.22)	(0.41, 0.16, 0.24)
ARFIMA33	(0.45, 0.22, 0.36)	(0.58, 0.24, 0.37)	(0.50, 0.22, 0.30)	(0.48, 0.21, 0.40)	(0.56, 0.21, 0.33)	(0.41, 0.19, 0.30)
ARFIMA34	(0.40, 0.17, 0.18)	(0.37, 0.20, 0.27)	(0.44, 0.23, 0.29)	(0.21, 0.13, 0.20)	(0.23, 0.14, 0.11)	(0.20, 0.10, 0.11)
ARFIMA35	(0.21, 0.10, 0.15)	(0.20, 0.09, 0.14)	(0.35, 0.15, 0.18)	(0.16, 0.07, 0.10)	(0.17, 0.09, 0.11)	(0.25, 0.13, 0.14)

[a] *Notes*: Model acronyms used are outlined in Section 3 and Table 7. All estimation, model fitting, and prediction mirrors that used in the empirical analysis (see notes to Table 2 and 3 and well as the discussion in Sections 2 and 3 of the paper). Rec-1 (Rol-1) refers to one step ahead forecasts based on recursive (rolling) estimation schemes. Additionally, 5- and 20-step ahead results are reported in the 3rd, 4th, 6th, and 7th columns of entries. The first entry in each bracketed triple shows the proportion of times the ARFIMA model is preferred based on comparison of its' point MSFE with that of the non-ARFIMA model, where each model is 'pre-selected' using the first-half of the sample, as discussed above and in the footnote to Table 2. The second entry reported the number of ARFIMA 'wins' based on application of the DM test at a nominal 10% level of significance. The third entry is analogous to the DM test entry, except that rejection incidence based on application of the ENC-t test is reported. All results are based on 500 Monte Carlo iterations.

Table 9
Relative forecasting performance non-ARFIMA DGP—sample size: (1) 1000 and (2) 4000[a]

Model	Rec-1	Rec-5	Rec-20	Rol-1	Rol-5	Rol-20
RW	(0.02, 0.00, 0.02)	(0.06, 0.00, 0.04)	(0.18, 0.02, 0.10)	(0.04, 0.00, 0.02)	(0.08, 0.02, 0.02)	(0.12, 0.08, 0.08)
AR1	(0.24, 0.04, 0.34)	(0.24, 0.02, 0.56)	(0.28, 0.02, 0.64)	(0.24, 0.08, 0.44)	(0.16, 0.00, 0.42)	(0.24, 0.02, 0.46)
AR2	(0.34, 0.08, 0.52)	(0.24, 0.04, 0.40)	(0.32, 0.02, 0.64)	(0.34, 0.12, 0.42)	(0.16, 0.00, 0.30)	(0.24, 0.02, 0.46)
AR3	(0.24, 0.03, 0.24)	(0.13, 0.00, 0.21)	(0.29, 0.03, 0.47)	(0.21, 0.08, 0.26)	(0.16, 0.00, 0.24)	(0.21, 0.03, 0.45)
AR4	(0.08, 0.00, 0.22)	(0.26, 0.02, 0.34)	(0.28, 0.02, 0.40)	(0.10, 0.00, 0.14)	(0.20, 0.02, 0.24)	(0.24, 0.02, 0.30)
AR5	(0.26, 0.00, 0.42)	(0.26, 0.06, 0.26)	(0.26, 0.00, 0.38)	(0.14, 0.02, 0.28)	(0.14, 0.02, 0.12)	(0.22, 0.00, 0.28)
AR6	(0.32, 0.02, 0.50)	(0.32, 0.02, 0.24)	(0.26, 0.00, 0.32)	(0.14, 0.04, 0.28)	(0.26, 0.06, 0.20)	(0.20, 0.00, 0.18)
AR7	(0.26, 0.04, 0.44)	(0.32, 0.10, 0.30)	(0.24, 0.02, 0.24)	(0.18, 0.04, 0.34)	(0.30, 0.02, 0.20)	(0.24, 0.00, 0.14)
AR8	(0.21, 0.05, 0.50)	(0.16, 0.05, 0.18)	(0.37, 0.03, 0.08)	(0.21, 0.03, 0.32)	(0.18, 0.05, 0.18)	(0.39, 0.00, 0.11)
MA1	(0.26, 0.04, 0.48)	(0.22, 0.11, 0.96)	(0.33, 0.00, 0.93)	(0.26, 0.04, 0.59)	(0.37, 0.11, 0.89)	(0.52, 0.11, 0.93)
MA2	(0.46, 0.08, 0.68)	(0.40, 0.06, 0.90)	(0.48, 0.02, 0.96)	(0.50, 0.10, 0.68)	(0.32, 0.08, 0.84)	(0.30, 0.04, 0.90)
MA3	(0.20, 0.05, 0.37)	(0.27, 0.00, 0.71)	(0.41, 0.02, 0.80)	(0.27, 0.12, 0.44)	(0.20, 0.05, 0.71)	(0.39, 0.05, 0.78)
MA4	(0.22, 0.02, 0.40)	(0.32, 0.02, 0.60)	(0.32, 0.02, 0.62)	(0.28, 0.06, 0.40)	(0.20, 0.00, 0.40)	(0.22, 0.04, 0.54)
MA5	(0.30, 0.08, 0.50)	(0.28, 0.02, 0.58)	(0.32, 0.02, 0.60)	(0.24, 0.04, 0.44)	(0.18, 0.02, 0.40)	(0.24, 0.02, 0.48)
MA6	(0.32, 0.08, 0.42)	(0.28, 0.06, 0.54)	(0.28, 0.00, 0.56)	(0.24, 0.06, 0.42)	(0.22, 0.02, 0.42)	(0.24, 0.02, 0.46)
MA7	(0.28, 0.10, 0.40)	(0.28, 0.04, 0.46)	(0.38, 0.02, 0.58)	(0.24, 0.06, 0.34)	(0.22, 0.00, 0.34)	(0.22, 0.00, 0.44)
MA8	(0.20, 0.02, 0.24)	(0.30, 0.04, 0.42)	(0.32, 0.00, 0.52)	(0.10, 0.00, 0.18)	(0.18, 0.00, 0.34)	(0.22, 0.04, 0.44)
GARCH1	(0.26, 0.04, 0.38)	(0.44, 0.06, 0.74)	(0.32, 0.08, 0.86)	(0.18, 0.02, 0.36)	(0.28, 0.08, 0.84)	(0.20, 0.02, 0.78)
GARCH2	(0.30, 0.08, 0.54)	(0.20, 0.00, 0.68)	(0.32, 0.06, 0.80)	(0.30, 0.08, 0.54)	(0.32, 0.06, 0.64)	(0.24, 0.04, 0.76)
GARCH3	(0.28, 0.12, 0.46)	(0.26, 0.02, 0.60)	(0.32, 0.06, 0.70)	(0.26, 0.06, 0.48)	(0.32, 0.04, 0.66)	(0.32, 0.04, 0.68)
GARCH4	(0.42, 0.08, 0.52)	(0.34, 0.04, 0.50)	(0.24, 0.02, 0.62)	(0.54, 0.12, 0.66)	(0.28, 0.00, 0.42)	(0.20, 0.00, 0.54)
GARCH5	(0.26, 0.04, 0.46)	(0.26, 0.04, 0.42)	(0.20, 0.04, 0.42)	(0.28, 0.02, 0.46)	(0.18, 0.02, 0.30)	(0.14, 0.02, 0.38)
GARCH6	(0.08, 0.04, 0.20)	(0.20, 0.02, 0.22)	(0.26, 0.06, 0.34)	(0.16, 0.06, 0.26)	(0.16, 0.02, 0.22)	(0.16, 0.02, 0.30)
GARCH7	(0.30, 0.08, 0.52)	(0.44, 0.04, 0.38)	(0.24, 0.02, 0.24)	(0.28, 0.04, 0.50)	(0.42, 0.02, 0.38)	(0.22, 0.02, 0.22)
GARCH8	(0.32, 0.04, 0.44)	(0.24, 0.08, 0.20)	(0.22, 0.02, 0.12)	(0.16, 0.06, 0.36)	(0.24, 0.08, 0.16)	(0.20, 0.06, 0.12)
ARIMA1	(0.28, 0.06, 0.46)	(0.34, 0.00, 0.76)	(0.30, 0.00, 0.80)	(0.46, 0.10, 0.58)	(0.20, 0.04, 0.68)	(0.34, 0.02, 0.68)
ARIMA2	(0.52, 0.08, 0.62)	(0.44, 0.06, 0.94)	(0.36, 0.10, 0.96)	(0.56, 0.10, 0.74)	(0.52, 0.16, 0.94)	(0.44, 0.08, 0.94)
ARIMA3	(0.36, 0.04, 0.46)	(0.26, 0.02, 0.32)	(0.34, 0.02, 0.46)	(0.14, 0.02, 0.26)	(0.22, 0.00, 0.20)	(0.26, 0.02, 0.30)
ARIMA4	(0.26, 0.10, 0.32)	(0.38, 0.16, 0.38)	(0.36, 0.04, 0.28)	(0.16, 0.08, 0.20)	(0.42, 0.14, 0.46)	(0.24, 0.04, 0.16)
ARIMA5	(0.12, 0.00, 0.10)	(0.04, 0.00, 0.04)	(0.14, 0.02, 0.06)	(0.08, 0.02, 0.04)	(0.10, 0.00, 0.02)	(0.12, 0.00, 0.02)
ARIMA6	(0.48, 0.20, 0.36)	(0.42, 0.16, 0.24)	(0.52, 0.22, 0.18)	(0.50, 0.22, 0.28)	(0.50, 0.22, 0.28)	(0.48, 0.30, 0.16)
ARIMA7	(0.30, 0.10, 0.34)	(0.26, 0.08, 0.16)	(0.20, 0.02, 0.12)	(0.24, 0.08, 0.18)	(0.38, 0.06, 0.12)	(0.26, 0.00, 0.02)
ARIMA8	(0.60, 0.32, 0.68)	(0.68, 0.32, 0.62)	(0.74, 0.16, 0.52)	(0.26, 0.16, 0.22)	(0.66, 0.26, 0.30)	(0.70, 0.04, 0.24)

RW	(0.03, 0.00, 0.02)	(0.07, 0.01, 0.04)	(0.12, 0.01, 0.09)	(0.03, 0.01, 0.02)	(0.07, 0.01, 0.02)	(0.09, 0.05, 0.06)
AR1	(0.23, 0.05, 0.38)	(0.23, 0.02, 0.63)	(0.26, 0.02, 0.72)	(0.23, 0.09, 0.51)	(0.15, 0.00, 0.47)	(0.23, 0.02, 0.52)
AR2	(0.32, 0.09, 0.59)	(0.23, 0.05, 0.45)	(0.3, 0.02, 0.72)	(0.32, 0.14, 0.47)	(0.15, 0.00, 0.34)	(0.23, 0.02, 0.52)
AR3	(0.23, 0.03, 0.27)	(0.12, 0.01, 0.24)	(0.27, 0.03, 0.53)	(0.22, 0.09, 0.29)	(0.15, 0.01, 0.27)	(0.21, 0.03, 0.51)
AR4	(0.08, 0.00, 0.25)	(0.24, 0.02, 0.38)	(0.26, 0.02, 0.45)	(0.09, 0.01, 0.16)	(0.19, 0.02, 0.27)	(0.23, 0.02, 0.34)
AR5	(0.24, 0.01, 0.47)	(0.24, 0.07, 0.29)	(0.24, 0.03, 0.43)	(0.13, 0.02, 0.32)	(0.13, 0.02, 0.14)	(0.21, 0.02, 0.32)
AR6	(0.31, 0.02, 0.57)	(0.29, 0.02, 0.27)	(0.24, 0.00, 0.36)	(0.13, 0.05, 0.32)	(0.24, 0.07, 0.23)	(0.19, 0.00, 0.19)
AR7	(0.24, 0.05, 0.35)	(0.31, 0.11, 0.34)	(0.23, 0.02, 0.27)	(0.17, 0.05, 0.38)	(0.28, 0.02, 0.23)	(0.23, 0.03, 0.16)
AR8	(0.21, 0.06, 0.57)	(0.15, 0.06, 0.21)	(0.35, 0.03, 0.09)	(0.19, 0.03, 0.36)	(0.17, 0.06, 0.19)	(0.37, 0.05, 0.12)
MA1	(0.27, 0.04, 0.45)	(0.22, 0.11, 0.14)	(0.34, 0.10, 0.87)	(0.27, 0.04, 0.55)	(0.38, 0.11, 0.84)	(0.53, 0.10, 0.87)
MA2	(0.47, 0.08, 0.64)	(0.41, 0.06, 0.85)	(0.49, 0.02, 0.79)	(0.51, 0.09, 0.64)	(0.33, 0.08, 0.79)	(0.31, 0.04, 0.85)
MA3	(0.21, 0.05, 0.35)	(0.28, 0.01, 0.67)	(0.42, 0.02, 0.75)	(0.28, 0.11, 0.41)	(0.21, 0.05, 0.67)	(0.41, 0.05, 0.73)
MA4	(0.22, 0.02, 0.38)	(0.33, 0.02, 0.56)	(0.33, 0.02, 0.58)	(0.29, 0.06, 0.38)	(0.18, 0.02, 0.38)	(0.22, 0.04, 0.51)
MA5	(0.31, 0.08, 0.47)	(0.29, 0.02, 0.55)	(0.33, 0.02, 0.56)	(0.24, 0.04, 0.41)	(0.18, 0.02, 0.38)	(0.24, 0.02, 0.45)
MA6	(0.33, 0.08, 0.39)	(0.29, 0.06, 0.51)	(0.29, 0.05, 0.53)	(0.24, 0.06, 0.39)	(0.22, 0.02, 0.39)	(0.24, 0.02, 0.43)
MA7	(0.29, 0.09, 0.38)	(0.29, 0.04, 0.43)	(0.39, 0.02, 0.55)	(0.24, 0.06, 0.32)	(0.22, 0.05, 0.32)	(0.22, 0.05, 0.41)
MA8	(0.20, 0.02, 0.23)	(0.31, 0.04, 0.39)	(0.33, 0.06, 0.49)	(0.10, 0.01, 0.17)	(0.18, 0.03, 0.32)	(0.22, 0.04, 0.41)
GARCH1	(0.22, 0.04, 0.39)	(0.37, 0.06, 0.76)	(0.27, 0.08, 0.89)	(0.15, 0.02, 0.37)	(0.24, 0.08, 0.87)	(0.17, 0.02, 0.79)
GARCH2	(0.26, 0.08, 0.56)	(0.17, 0.01, 0.70)	(0.27, 0.06, 0.82)	(0.26, 0.08, 0.56)	(0.27, 0.06, 0.66)	(0.20, 0.04, 0.78)
GARCH3	(0.24, 0.12, 0.47)	(0.22, 0.02, 0.62)	(0.27, 0.06, 0.72)	(0.22, 0.06, 0.49)	(0.27, 0.04, 0.68)	(0.27, 0.04, 0.70)
GARCH4	(0.36, 0.08, 0.54)	(0.29, 0.04, 0.52)	(0.21, 0.02, 0.64)	(0.46, 0.12, 0.68)	(0.24, 0.02, 0.43)	(0.17, 0.01, 0.56)
GARCH5	(0.22, 0.04, 0.47)	(0.22, 0.04, 0.43)	(0.17, 0.04, 0.43)	(0.24, 0.02, 0.47)	(0.15, 0.02, 0.31)	(0.12, 0.02, 0.39)
GARCH6	(0.07, 0.04, 0.21)	(0.17, 0.02, 0.23)	(0.22, 0.06, 0.35)	(0.14, 0.06, 0.27)	(0.14, 0.02, 0.23)	(0.14, 0.02, 0.31)
GARCH7	(0.26, 0.08, 0.54)	(0.37, 0.04, 0.39)	(0.20, 0.02, 0.25)	(0.24, 0.04, 0.52)	(0.36, 0.02, 0.39)	(0.19, 0.02, 0.23)
GARCH8	(0.27, 0.04, 0.45)	(0.20, 0.08, 0.21)	(0.19, 0.02, 0.12)	(0.14, 0.06, 0.37)	(0.21, 0.08, 0.16)	(0.17, 0.06, 0.12)
ARIMA1	(0.28, 0.05, 0.36)	(0.34, 0.02, 0.60)	(0.30, 0.01, 0.63)	(0.46, 0.08, 0.46)	(0.20, 0.03, 0.54)	(0.34, 0.02, 0.54)
ARIMA2	(0.51, 0.06, 0.49)	(0.44, 0.05, 0.74)	(0.36, 0.08, 0.76)	(0.55, 0.08, 0.58)	(0.51, 0.13, 0.74)	(0.44, 0.06, 0.74)
ARIMA3	(0.36, 0.03, 0.36)	(0.26, 0.02, 0.25)	(0.34, 0.02, 0.36)	(0.14, 0.02, 0.21)	(0.22, 0.01, 0.16)	(0.26, 0.02, 0.24)
ARIMA4	(0.26, 0.08, 0.25)	(0.38, 0.13, 0.30)	(0.36, 0.03, 0.22)	(0.16, 0.06, 0.16)	(0.42, 0.11, 0.36)	(0.24, 0.03, 0.13)
ARIMA5	(0.10, 0.00, 0.07)	(0.03, 0.00, 0.03)	(0.12, 0.01, 0.04)	(0.07, 0.01, 0.03)	(0.08, 0.00, 0.01)	(0.10, 0.00, 0.01)
ARIMA6	(0.48, 0.15, 0.26)	(0.42, 0.12, 0.18)	(0.51, 0.16, 0.13)	(0.50, 0.16, 0.20)	(0.50, 0.16, 0.20)	(0.48, 0.22, 0.12)
ARIMA7	(0.30, 0.07, 0.25)	(0.26, 0.06, 0.12)	(0.20, 0.01, 0.09)	(0.24, 0.06, 0.13)	(0.38, 0.04, 0.09)	(0.26, 0.00, 0.01)
ARIMA8	(0.59, 0.23, 0.50)	(0.67, 0.23, 0.45)	(0.73, 0.12, 0.38)	(0.26, 0.12, 0.16)	(0.65, 0.19, 0.22)	(0.69, 0.03, 0.18)

[a] *Notes*: See notes to Table 8.

model to use in the subsequent "horse-race", and hence results reported in Table 8(1), for example, are based on sequences of only 500 predictions. Given that estimation of d in this table is thus also carried out with far fewer than 1000 observations, it is perhaps noteworthy that the ARFIMA model still outperforms the non-ARFIMA model around 50% of the time, and sometimes as much as 70–80% of the time. These numbers increase dramatically when the sample is 4000 observations, with ARFIMA models "wins" occurring around 70–100% of the time in most cases. Thus, moderately sized samples may be enough to achieve gains from using ARFIMA models. This finding is in accord with the findings reported in the empirical part of the paper.

Not surprisingly, the ARFIMA model wins very little of the time, when the true DGP is a non-ARFIMA model. Furthermore, the incidence of ARFIMA "wins" decreases when the sample is 4000 rather than 1000 observations (compare Table 9(1) with (2)).

Although the above results appear somewhat promising, it should be stressed that parameter estimation error does play an important role. To illustrate this point, note that in Table 10 two different ARFIMA models are compared using the modelling approach discussed above. One is an ARFIMA model with d estimated, and the other assumes that d is known—so that parameters estimated each time predictions are constructed are only the ARMA parameters. Numerical values in the table are percentages, and are extremely high, as expected, as they measure the percentage of times that models with all parameters known outperform models with d estimated, based on point MSFE comparison. What is perhaps surprising is that the impact of estimating d remains essentially unchanged when the sample size is increased from

Table 10
The impact of estimating[a]

Model	d	Sample size 1000						Sample size 4000					
		Rec-1	Rec-5	Rec-20	Rol-1	Rol-5	Rol-20	Rec-1	Rec-5	Rec-20	Rol-1	Rol-5	Rol-20
ARFIMA2	0.2	87.0	89.0	96.0	88.0	90.0	91.0	89.0	92.0	95.0	89.0	92.0	94.0
ARFIMA7		99.0	99.0	100.0	97.0	96.0	99.0	98.0	99.0	99.0	98.0	99.0	100.0
ARFIMA9	0.3	88.0	92.0	97.0	90.0	94.0	96.0	92.0	94.0	98.0	91.0	96.0	97.0
ARFIMA14		100.0	99.0	100.0	99.0	100.0	99.0	100.0	100.0	100.0	100.0	100.0	100.0
ARFIMA16	0.4	89.0	92.0	98.0	89.0	91.0	95.0	93.0	94.0	99.0	94.0	94.0	98.0
ARFIMA21		98.0	100.0	100.0	90.0	96.0	94.0	99.0	100.0	100.0	100.0	100.0	100.0
ARFIMA23	0.6	92.0	94.0	97.0	89.0	92.0	94.0	94.0	95.0	99.0	94.0	96.0	99.0
ARFIMA28		97.0	100.0	99.0	95.0	96.0	94.0	99.0	100.0	100.0	98.0	99.0	100.0
ARFIMA30	0.7	93.0	94.0	97.0	91.0	98.0	97.0	95.0	96.0	99.0	95.0	96.0	99.0
ARFIMA35		99.8	100.0	100.0	97.0	99.0	100.0	99.0	100.0	100.0	100.0	100.0	100.0

[a]*Notes*: Entries report the results of point MSFE comparison between ARFIMA models with all parameters known and models where all parameters except d are known. In particular, the percentage of times that the ARFIMA model with d known 'wins' is reported. See Sections 2.4 and 3 as well as the notes to Tables 2 and 8 for a discussion of how models are estimated, predictions are formed, acronyms used, etc. Results are based on 500 Monte Carlo iterations.

1000 to 4000 observations, again affirming that very long samples are needed before the impact of parameter estimation error begins to diminish.[19]

5. Concluding remarks

We present the results of an empirical and Monte Carlo investigation of the usefulness of ARFIMA models in practical prediction-based applications, and find evidence that such models may yield reasonable approximations to unknown underlying DGPs, in the sense that the models often significantly outperform a fairly wide class of the benchmark non-ARFIMA models, including AR, ARMA, ARIMA, random walk, GARCH, and related models. This finding is particularly apparent with longer samples of data such as an international stock index return data set with around 5000 observations. Another finding of our analysis is that more parsimonious ARIMA models are clearly not always preferred when predicting financial data—a rather surprising result given the large body of research suggesting that more parsimonious models often outperform more heavily parameterized models.[20] Finally, there appears little to choose between various estimators of d when samples are as large as often encountered in financial economics. For shorter samples such as those encountered in macroeconomics, parameter estimation error appears to plague estimates of d, and predictive performance of ARFIMA models is appreciably worsened, relative to the longer financial data sets examined in this paper. Overall, we conclude that long-memory processes, and in particular ARFIMA processes, might not fall into the "empty box" category after all, although much further research is needed before overwhelmingly conclusive evidence in either direction can be given. For example, it should be of interest to investigate whether our finding that ARFIMA models most frequently outperform simpler linear models at longer prediction horizons hold up when the alternatives considered also include various types of regime switching, threshold, and related non-linear models. On a related note, alternative estimators of d may be useful when building forecasting models using smaller data sets, such as estimators based on predictive error loss minimization (see e.g. Bhardwaj and Swanson, 2003). These and related issues are left to future research.

[19]It should be noted that the performance of ARFIMA models where d is known always improves when the sample size is increased from 1000 to 4000 observations. Thus, any parameter estimation error reduction (and hence improvement in performance of our ARFIMA models which use estimates of d) due to increasing sample size appears to be more than offset by improved performance of our arbitrary cut-off for polynomial expansion (as this is the only 'approximation' inherent to the ARFIMA models used to construct predictions when d is known), at least when samples are between 1000 and 4000 observations.

[20]It should be noted that our ARFIMA models are only slightly less 'parsimonious' than our ARIMA models, in the sense that the only 'new' parameter is d, when one moves from an ARIMA to an ARFIMA model. However, it should also be recalled that there is another sense in which what one might refer to loosely as 'parsimony' is reduced when ARFIMA models are estimated. Namely, ARFIMA models involve ad hoc application of truncation filters, while ARIMA models do not.

Acknowledgements

This paper has been prepared for the special issue of the Journal of Econometrics on "Empirical Methods in Macroeconomics and Finance", and the authors are grateful to the organizers and participants of the related conference held at Bocconi University in October 2003. The many stimulating papers presented at the conference, and the ensuing discussions, have served in large part to shape this paper. The authors are particularly grateful to Frank Schorfheide and three anonymous referees, all of whom provided invaluable comments and suggestions on an earlier version of this paper. Finally, thanks are owed to Valentina Corradi and Clive W.J. Granger for stimulating discussions, and Zhuanxin Ding, Steve Leybourne, and Mark Watson for providing the financial and macroeconomic data sets used in the empirical section of the paper. Swanson has benefited from the support of Rutgers University in the form of a Research Council grant.

References

Agiakloglou, C., Newbold, P., Wohar, M., 1992. Bias in an estimator of the fractional difference parameter. Journal of Time Series Analysis 14, 235–246.

Andrews, D.W.K. Sun, Y., 2002. Adaptive local Whittle estimation of long-range dependence. Working Paper, Yale University.

Baillie, R.T., 1996. Long memory processes and fractional integration in econometrics. Journal of Econometrics 73, 5–59.

Bank of Sweden, 2003. Time-series econometrics: cointegration and autoregressive conditional heteroskedasticity. Advanced Information on the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel, The Royal Swedish Academy of Sciences.

Beran, J., 1995. Maximum likelihood estimation of the differencing parameter for invertible short and long-memory autoregressive integrated moving average models. Journal of the Royal Statistical Society Series B 57, 659–672.

Bhardwaj, G., Swanson, N.R., 2003. An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. Working Paper, Rutgers University.

Bos, C.S., Franses, P.H., Ooms, M., 2002. Inflation, forecast intervals and long-memory regression models. International Journal of Forecasting 18, 243–264.

Breitung, J., Hassler, U., 2002. Inference on the cointegration rank in fractionally integrated processes. Journal of Econometrics 110, 167–185.

Chao, J.C., Corradi, V., Swanson, N.R., 2001. An out of sample test for Granger causality. Macroeconomic Dynamics 5, 598–620.

Cheung, Y.-W., 1993. Tests for fractional integration: a Monte Carlo investigation. Journal of Time Series Analysis 14, 331–345.

Cheung, Y.-W., Diebold, F.X., 1994. On maximum likelihood estimation of the difference parameter of fractionally integrated noise with unknown mean. Journal of Econometrics 62, 301–316.

Chio, K., Zivot, E., 2002. Long memory and structural changes in the forward discount: an empirical investigation. Working Paper, University of Washington.

Christoffersen, P.F., 1998. Evaluating interval forecasts. International Economic Review 39, 841–862.

Christoffersen, P., Diebold, F.X., 1997. Optimal prediction under asymmetric loss. Econometric Theory 13, 808–817.

Clark, T.E., McCracken, M.W., 2001. Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics 105, 85–110.

Clements, M.P., Smith, J., 2000. Evaluating the forecast densities of linear and nonlinear models: applications to output growth and unemployment. Journal of Forecasting 19, 255–276.

Clements, M.P., Smith, J., 2002. Evaluating multivariate forecast densities: a comparison of two approaches. International Journal of Forecasting 18, 397–407.

Corradi, V., Swanson, N.R., 2002. A consistent test for out of sample nonlinear predictive ability. Journal of Econometrics 110, 353–381.

Corradi, V., Swanson, N.R., 2005a. Bootstrap procedures for recursive estimation schemes with applications to forecast model selection. Working Paper, Rutgers University.

Corradi, V., Swanson, N.R., 2005b. Predictive density and conditional interval accuracy tests. Working Paper, Rutgers University.

Corradi, V., Swanson, N.R., 2005c. Predictive density evaluation. In: Elliott. G., Granger, C.W.J., Timmerman, A. (Eds.), Handbook of Economic Forecasting. Elsevier, Amsterdam, forthcoming.

Corradi, V., Swanson, N.R., Olivetti, C., 2001. Predictive ability with cointegrated variables. Journal of Econometrics 104, 315–358.

Diebold, F.X., Gunther, T., Tay, A.S., 1998. Evaluating density forecasts with applications to finance and management. International Economic Review 39, 863–883.

Diebold, F.X., Hahn, J., Tay, A.S., 1999. Multivariate density forecast evaluation and calibration in financial risk management: high frequency returns on foreign exchange. Review of Economics and Statistics 81, 661–673.

Diebold, F., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics 105, 131–159.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

Diebold, F.X., Rudebusch, G.D., 1989. Long memory and persistence in aggregate output. Journal of Monetary Economics 24, 189–209.

Diebold, F.X., Rudebusch, G.D., 1991a. Is consumption too smooth? Long memory and the Deaton paradox. Review of Economics and Statistics 73, 1–9.

Diebold, F.X., Rudebusch, G.D., 1991b. On the power of the Dickey–Fuller test against fractional alternatives. Economics Letters 35, 155–160.

Ding, Z., Granger, C.W.J., Engle, R.F., 1993. A long-memory property of stock returns and a new model. Journal of Empirical Finance 1, 83–106.

Dittman, I., Granger, C.W.J., 2002. Properties of nonlinear transformations of fractionally integrated processes. Journal of Econometrics 110, 113–133.

Doornik, J.A., Ooms, M., 2003. Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. Computational Statistics and Data Analysis 42, 333–348.

Engle, R.F., Smith, A.D., 1999. Stochastic permanent breaks. Review of Economics and Statistics 81, 553–574.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long-memory time series models. Journal of Time Series Analysis 4, 221–238.

Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.

Granger, C.W.J., 1980. Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14, 227–238.

Granger, C.W.J., 1999. Aspects of research strategies for time series analysis. Presentation to the Conference on New Developments in Time Series Economics, Yale University.

Granger, C.W.J., Andersen, A.P., 1978. Introduction to Bilinear Time Series Models. Vandenhoeck and Ruprecht, Göttingen.

Granger, C.W.J., Ding, Z., 1996. Varieties of long-memory models. Journal of Econometrics 73, 61–77.

Granger, C.W.J., Hatanaka, M., 1964. Spectral Analysis of Economic Time Series. Princeton University Press, Princeton.

Granger, C.W.J., Hyung, N., 1999. Occasional structural breaks and long-memory. Working Paper, University of California, San Diego.

Granger, C.W.J., Joyeux, R., 1980. An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis 1, 15–30.

Granger, C.W.J., Newbold, P., 1986. Forecasting Economic Time Series. Academic Press, San Diego.

Hansen, P.R., Lunde, A., Nason, J.M., 2004. Model confidence sets for forecasting models. Working Paper, Brown University.
Harvey, D.I., Leybourne, S.J., Newbold, P., 1997. Tests for forecast encompassing. Journal of Business and Economic Statistics 16, 254–259.
Hassler, U., Wolters, J., 1995. Long memory in inflation rates: international evidence. Journal of Business and Economic Statistics 13, 37–45.
Hosking, J., 1981. Fractional differencing. Biometrica 68, 165–176.
Hurst, H.E., 1951. Long-term storage capacity of reservoirs. Transactions of the American Society of Civil Engineers 116, 770–799.
Hyung, N., Franses, P.H., 2001. Structural breaks and long-memory in US inflation rates: do they matter for forecasting? Working Paper, Erasmus University.
Inoue, A., Kilian, L., On the selection of forecasting models. Working Paper, University of Michigan.
Künsch, H.R., 1987. Statistical aspects of self-similar processes. In: Prohorov, Y., Sasanov, V.V. (Eds.), Proceedings of the First World Congress of the Bernoulli Society. VNU Science Press, Utrecht.
Lee, D., Schmidt, P., 1996. On the power of the KPSS test of stationarity against fractionally integrated alternatives. Journal of Econometrics 73, 285–302.
Leybourne, S., Harris, D., McCabe, B., 2003. A robust test for short-memory. Working Paper, University of Nottingham.
Lo, A., 1991. Long-term memory in stock market prices. Econometrica 59, 1279–1313.
McCracken, M.W., Asymptotics for out of sample tests of causality. Working Paper, Louisiana State University.
Newey, W.K., West, K.D., 1987. A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Phillips, P.C.B., 1987. Time series regression with a unit root. Econometrica 55, 277–301.
Robinson, P., 1995a. Log-periodogram regression of time series with long range dependence. The Annals of Statistics 23, 1048–1072.
Robinson, P., 1995b. Gaussian semiparametric estimation of long range dependence. The Annals of Statistics 23, 1630–1661.
Robinson, P., 2003. Time Series with Long Memory. Oxford University Press, Oxford.
Rossi, B., 2003. Testing long-horizon predictive ability with high persistence, and the Meese-Rogoff puzzle. Working Paper, Duke University.
Shimotsu, K., Phillips, P.C.B., 2002. Exact local Whittle estimation of fractional integration. Working Paper, University of Essex.
Sowell, F.B., 1992a. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Journal of Econometrics 53, 165–188.
Sowell, F.B., 1992b. Modelling long-run behavior with the fractional ARIMA model. Journal of Monetary Economics 29, 277–302.
Stock, J., Watson, M., 2002. Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics 20, 147–162.
Swanson, N.R., White, H., 1995. A model selection approach to assessing the information in the term structure using linear models and artificial neural networks. Journal of Business and Economic Statistics 13, 265–279.
Swanson, N.R., White, H., 1997. A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. Review of Economics and Statistics 79, 540–550.
Taqqu, M., Teverovsky, V., 1997. Robustness of Whittle-type estimators for time series with long-range dependence. Stochastic Models 13, 723–757.
van Dijk, D., Franses, P., Paap, R., 2002. A nonlinear long-memory model, with an application to US unemployment. Journal of Econometrics 110, 135–165.
West, K., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.
White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.