2001s-11

Autoregression-Based Estimators for ARFIMA Models

John W. Galbraith, Victoria Zinde-Walsh

Série Scientifique
Scientific Series

CIRANO
Centre interuniversitaire de recherche
en analyse des organisations

Montréal
Février 2001

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du ministère de la Recherche, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de la Recherche, de la Science et de la Technologie, and grants and research mandates obtained by its research teams.

Les organisations-partenaires / The Partner Organizations

- •École des Hautes Études Commerciales
- •École Polytechnique
- •Université Concordia
- •Université de Montréal
- •Université du Québec à Montréal
- •Université Laval
- •Université McGill
- •MEQ
- •MRST
- •Alcan Aluminium Ltée
- •AXA Canada
- •Banque du Canada
- •Banque Laurentienne du Canada
- •Banque Nationale du Canada
- •Banque Royale du Canada
- •Bell Québec
- •Bombardier
- •Bourse de Montréal
- •Développement des ressources humaines Canada (DRHC)
- •Fédération des caisses populaires Desjardins de Montréal et de l'Ouest-du-Québec
- •Hydro-Québec
- •Imasco
- •Industrie Canada
- •Pratt & Whitney Canada Inc.
- •Raymond Chabot Grant Thornton
- •Ville de Montréal

ISSN 1198-8177

Autoregression-Based Estimators for ARFIMA Models*

John W. Galbraith†, *Victoria Zinde-Walsh*‡

Résumé / Abstract

Nous décrivons une méthode d'estimation pour les paramètres des modèles ARFIMA stationnaires ou non-stationnaires, basée sur l'approximation auto-régressive. Nous démontrons que la procédure est consistante pour $-½ < d < 1$, et dans le cas stationnaire nous donnons une approximation Normale utilisable pour inférence statistique. La méthode fonctionne bien en échantillon fini, et donne des résultats comparables pour la plupart des valeurs du paramètre d, stationnaires ou non. Il y a aussi des indications de robustesse à la mauvaise spécification du modèle ARFIMA à estimer, et le calcul des estimations est simple.

This paper describes a parameter estimation method for both stationary and non-stationary ARFIMA (p,d,q) models, based on autoregressive approximation. We demonstrate consistency of the estimator for $-½ < d < 1$, and in the stationary case we provide a Normal approximation to the finite-sample distribution which can be used for inference. The method provides good finite-sample performance, comparable with that of ML, and stable performance across a range of stationary and non-stationary values of the fractional differencing parameter. In addition, it appears to be relatively robust to mis-specification of the ARFIMA model to be estimated, and is computationally straightforward.

Mots Clés : Modèle ARFIMA, autorégression, intégration fractionnelle, mémoire longue

Keywords: ARFIMA model, autoregression, fractional integration, long memory

JEL: C12, C22

---

* Corresponding Author: John Galbraith, Department of Economics, McGill University, 888 Sherbrooke Street West, Montréal, Qc, Canada H3A 2T7 Tel.: (514) 985-4008 Fax: (514) 985-4039 email: galbraij@cirano.qc.ca
We thank Javier Hidalgo, Aman Ullah and seminar participants at Alberta, Meiji Gakuin, Montréal, Queen's and York universities, and the 1997 Econometric Society European Meeting, for valuable comments. Thanks also to the Fonds pour la formation de chercheurs et l'aide à la recherche (Québec) and the Social Sciences and Humanities Research Council of Canada for financial support. The program for computing GPH estimates was written by Pedro de Lima; the code for ML estimation of ARFIMA models was based on programs to compute the autocovariances of ARFIMA models written by Ching-Fan Chung. Nigel Wilkins kindly provided additional code for ML estimation.

† McGill University and CIRANO

‡ McGill University

1. Introduction

Long memory processes, and in particular models based on fractional integration as in Granger and Joyeux (1980) and Hosking (1981), have come to play an increasing role in time series analysis as longer time series have become available; financial time series have yielded an especially large number of applications. There is a correspondingly substantial literature on the estimation of such models, both in the frequency domain and the time domain; these contributions can be further sub-divided into those which estimate a fractional integration (FI) parameter jointly with the standard ARMA parameters of an ARFIMA model, and those which estimate the FI parameter alone, leaving any ARMA or other parameters for possible estimation in a second stage.

Important contributions to frequency-domain estimation of the FI parameter include those of Geweke and Porter-Hudak (1983) and Robinson (1995). In the time domain, Hosking (1981), Li and MacLeod (1986) and Haslett and Rafferty (1989) suggested estimation strategies based on first-stage estimation of the long-memory parameter alone. A potential finite-sample problem arises in processes that have a short-memory component, which can lead to bias as the short-memory components project onto the long-memory parameter in the first stage; see Agiakloglou et al. (1992) for a discussion of this bias in the context of the Geweke and Porter-Hudak (hereafter GPH) estimator.

Sowell (1992a) and Tieslau et al. (1996) treat joint estimation of fractional integration and ARMA parameters, in the former case via Maximum Likelihood, and in the latter via a minimum-distance estimator based on estimated autocorrelations and the theoretical autocorrelations for a given ARFIMA(p, d, q) process. Because they are based on

autocorrelations, these methods require that the process be stationary, or (assuming prior knowledge of the non-stationarity) be transformed to stationarity. Martin and Wilkins (1999) use the indirect inference estimator, which uses simulation to obtain the function to which distance is minimized.

The present paper offers an alternative estimation method for joint estimation of a set of ARFIMA parameters which is applicable to both stationary and non-stationary processes. The method is based on autoregressive approximation, as considered by Galbraith and Zinde-Walsh (1994, 1997), and has several advantages in addition to ease of computation. First, it offers stable performance across a range of stationary (including 'anti-persistent') and non-stationary values of the long-memory parameter d, that is $-\frac{1}{2} < d < 1$, and therefore does not require prior knowledge of the non-stationarity or transformation to the stationarity region. In general, the autoregressive method performs well in the finite-sample cases that we examine, yielding root mean squared errors comparable with those of exact time-domain ML in the cases for which that estimator is applicable. Perhaps most importantly, this method appears to be relatively robust to mis-specification, being based on AR approximations which can represent quite general processes. We offer simulation evidence on each of these points.

In Section 2 we briefly review estimation of ARFIMA models in the time domain and autoregressive approximation of ARFIMA processes, and present estimators based on this approximation. Section 3 considers both asymptotic and finite-sample properties of this estimation strategy, and Section 4 concludes.

2. ARFIMA process representation and time-domain estimation

2.1 Assumptions and the autoregressive representation

Consider an ARFIMA(p, d, q) process, defined as

$$P(L)(1-L)^d y_t = Q(L)\varepsilon_t, \tag{2.1.1}$$

where $\{y_t\}_{t=1}^T$ is a set of observations on the process of interest and $\{\varepsilon_t\}_{t=1}^T$ forms a stationary innovation sequence such that for the $\sigma-$ algebra $\mathcal{F}_{t-1}$ generated by $\{\varepsilon_\tau,\ \tau \leq t-1\}$, $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ a.s., $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2 > 0$ a.s., and $E(\varepsilon_t^4) < \infty$. Let $P(L), Q(L)$ be polynomials of degrees p and q, such that $P(L) = 1-\rho_1 L - \ldots \rho_p L^p,\ Q(L) = 1+\theta_1 L + \ldots + \theta_q L^q$. Assume that $Q(L)$ has all roots outside the unit circle, so that the moving average part of the process is invertible; we can then write

$$[Q(L)]^{-1} P(L)(1-L)^d y_t = \varepsilon_t. \tag{2.1.2}$$

We will assume also that $Q(L)$ and $P(L)$ have no common factors, and that the roots of $P(L)$ are outside the unit circle. However, we will consider non-stationarity arising through values of d greater than $\frac{1}{2}$, and provide both analytic and simulation results for those cases. We treat the process as having zero mean, although we will note below the effect of using an estimated mean, and will use an estimated mean in simulation evaluation of the estimator.

The term $(1-L)^d$ may be expanded as

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k = \sum_{k=0}^{\infty} b_k L^k, \tag{2.1.3}$$

with $b_0 = 1,\ b_1 = -d,\ b_2 = \frac{1}{2}d(1-d),\ b_j = \frac{1}{j}b_{j-1}(j-1-d),\ j \geq 3$. If $|d| < \frac{1}{2}$, then $\sum_{k=0}^{\infty}\binom{d}{k}^2 < \infty$, and (2.1.3) defines a stationary process. For $d > -\frac{1}{2}$, $(1-L)^d$ is invertible, and expression (2.1.2) can therefore be used to obtain the coefficients of the infinite autoregressive representation of the ARFIMA process y :

$$(1 - \sum_{i=1}^{\infty}\delta_i L^i)y_t = \varepsilon_t, \tag{2.1.4}$$

with $\delta_i = b_i - \sum_{j=1}^{q}\theta_j\delta_{i-j} + \sum_{j=1}^{p}\rho_j b_{i-j}$ and $\sum\delta_i^2 < \infty$ where $|d| < \frac{1}{2}$.

2.2 Time-domain estimation

Time-domain estimators have been proposed by, among others, Hosking (1981), Li and MacLeod (1986), Sowell (1992a) and Tieslau et al. (1996). The latter two allow joint estimation of all ARFIMA parameters.

Sowell (1992a) gives an exact Maximum Likelihood algorithm for stationary ARFIMA models with distinct roots in the AR polynomial. As Baillie (1996) notes, ML is computationally burdensome here, since substantial calculation (including $T \times T$ matrix inversion) is required at each iteration of the numerical optimization. Below we compare the properties of exact ML with those of an estimator based on the coefficients of an autoregressive approximation, analogous to the ARMA estimation methods of Saikkonen (1986) and Galbraith and Zinde-Walsh (1997). To obtain the estimator we will use estimators of the coefficients of a long autoregression, and the autoregressive expansion of the ARFIMA(p,d,q) model, to obtain the estimates of the long memory parameter d together with parameters which characterize the "short memory" parts of the model. Below we will give a rule of thumb, based on $\ln(T)$, for lag length selection.

Before defining the ARFIMA parameter estimators, we therefore begin by considering three estimators for the coefficients of an autoregressive approximation, on which ARFIMA estimates can later be based. Each of the three has the same asymptotic limit in the stationarity region.[1] One of the three, the OLS estimator, can be used in the non-stationarity region as well; for that reason OLS would be used in applications, where stationarity is not normally known *a priori* to apply. However other estimators, in particular the Yule-Walker, are convenient for obtaining theoretical properties in the stationarity region.

The OLS estimator $\tilde{a}_{OLS}$ solves

$$\tilde{a}_{OLS} = \text{ argmin } \left(\frac{1}{T}\right) \left[\sum_{t=k+1}^{T} (y_t - \alpha_1 y_{t-1} - \cdots - \alpha_k y_{t-k})^2 \right]. \tag{2.2.1}$$

The second estimator is the spectral estimator, asymptotic properties of which were examined by Yajima (1992) for Gaussian errors. We denote it $\tilde{a}_{sp}$; it follows from Yajima (1992) that, in the present notation,

$$\begin{aligned} \tilde{a}_{sp} = \text{ argmin } \left(\frac{1}{T}\right) \Big[& y_1^2 + (y_2 - \alpha_1 y_1)^2 + \cdots + (y_k - \alpha_1 y_{k-1} - \cdots - \alpha_{k-1} y_1)^2 \\ & + \sum_{t=k+1}^{T} (y_t - \alpha_1 y_{t-1} - \cdots - \alpha_k y_{t-k})^2 \Big]. \end{aligned} \tag{2.2.2}$$

The third is the Yule-Walker estimator. Define $\gamma_j = E(y_t y_{t-j})$ and $\hat{\gamma}_j = \frac{1}{T-k} \sum_{t=k+1}^{T} y_t y_{t-j}$, $\Sigma(k)$ as the $k \times k$ matrix with $\{\Sigma(k)\}_{ij} = \gamma_{|i-j|}$, and $\hat{\Sigma}(k)$ as the matrix with elements $\{\hat{\Sigma}(k)\}_{ij} = \hat{\gamma}_{|i-j|}$. The Yule-Walker estimator is then $\tilde{a}_{YW}$, which

[1] Each of course depends on the AR order parameter, k, but to simplify notation this dependence will not be indicated explicitly.

solves

$$\hat{\Sigma}(k)\tilde{a}_{YW} = \hat{\gamma}(k), \tag{2.2.3}$$

where $\hat{\gamma}(k) = (\hat{\gamma}_1, \hat{\gamma}_2, \ldots \hat{\gamma}_k)$. Consider also the population analogue of (2.2.3),

$$\Sigma(k)a(k) = \gamma(k), \tag{2.2.4}$$

where $a(k)$ is the solution to the system.

The terms in the first-order conditions that $\tilde{a}_{OLS}$ and $\tilde{a}_{sp}$ solve differ from (2.2.3), for $\tilde{a}_{YW}$, by terms of order in probability at most $O_p(kT^{-1})$; it follows that all of the estimators above differ by at most $O_p(k^2T^{-1})$.[2] Therefore each one has the same asymptotic limit $a(k)$, and the choice of k as $O(\ln T)$ implies that all of the estimators have the same asymptotic distributions. We will therefore use the notation $\tilde{a}$ to denote any one of these estimators; no distinction among them need be made in considering asymptotic properties of ARFIMA parameter estimators based on such autoregressive estimates, for $-\frac{1}{2} < d < \frac{1}{2}$. Below we use the Yule-Walker estimator in derivations for the stationary case; we use OLS for derivations in non-stationary cases and in the Monte Carlo experiments reported in subsection 3.3.

The values $a(k)$ of (2.2.4) are related to the coefficients of the infinite autoregressive representation of the stationary ARFIMA process. If we denote by $\delta_{[1,\infty)}$ the vector of coefficients of the infinite autoregression (2.1.4), then that vector solves

$$\Sigma(\infty)\delta_{[1,\infty)} = \gamma(\infty). \tag{2.2.5}$$

[2]The only non-standard case here is that of $-\frac{1}{2} < d < 0$, where the eigenvalues of Σ^{-1} could grow at a rate as high as k^{-2d}, as shown in Appendix A.

Partition $\delta'_{[1,\infty)}$ as $(\delta'_{[1,k]} : \delta'_{[k+1,\infty)})$, and denote the top-right sub-matrix of Σ_∞, partitioned conformably, by $\Sigma_{[k+1,\infty)}$. Then

$$\Sigma(k)\delta_{[1,k]} + \Sigma_{[k+1,\infty)}\delta_{[k+1,\infty)} = \gamma(k),$$

and therefore

$$a(k) - \delta_{[1,k]} = (\Sigma(k))^{-1}\Sigma_{[k+1,\infty)}\delta_{[k+1,\infty)}. \tag{2.2.6}$$

Thus an autoregressive estimator of $\delta_{[1,k]}$ will include a deterministic bias represented by (2.2.6). We demonstrate below that this bias goes to zero as $k, T \to \infty$.

2.3 ARFIMA parameter estimation

Now that we have introduced estimators of autoregressive coefficients for the ARFIMA process, we can proceed to define estimators for the full set of ARFIMA parameters based on any of the above estimators. We define the vector of all the parameters of ARFIMA to be estimated by $\omega = (d, \rho', \theta')$, where $\rho' = (\rho_1, \ldots, \rho_p), \theta' = (\theta_1, \ldots, \theta_q)$. Let $\tilde{\alpha}$ be any of the autoregressive coefficient estimators introduced above, and let $\delta(\omega)$ denote the vector containing the coefficients of the infinite autoregressive representation of the process, viewed as functions of ω.

We will examine a minimum-distance estimator of the form

$$\tilde{\omega}_\delta = \text{ argmin } (\tilde{a} - \delta(\omega))'\Omega(\tilde{a} - \delta(\omega)), \tag{2.2.7}$$

constructed using any of the estimators $\tilde{a}$ of (2.2.1-2.2.3), where $\delta(\omega)$ is given explicitly in (2.1.4); Ω represents a $k \times k$ weighting matrix. In all simulations below, Ω is chosen as the inverse of the estimated covariance matrix of $\tilde{a}$. The use of such a weighting matrix

implies giving greater emphasis to the relatively precisely-estimated coefficients on lower lags, which contributes to the efficiency of the estimator.[3]

It is the fact that this estimator uses autoregressive parameters, instead of autocorrelations, that allows its use for non-stationary processes.[4] Note finally that estimates can be obtained for multiple specifications of an ARFIMA model from a single autoregressive fit to the data, which is convenient in model selection (information criteria, for example, can be computed based on the residuals from each of several models).

3. Properties and performance of the estimators

In this section we consider asymptotic and finite-sample properties of ARFIMA parameter estimates based on (2.2.7), obtained using any of the various preliminary autoregressive estimators (OLS, spectral, Yule-Walker) discussed in Section 2. We will also note the corresponding properties of some of the existing ML, MDE and spectral estimators.

Maximum likelihood produces asymptotically Normal estimates of ARFIMA parameters which converge at the usual $T^{\frac{1}{2}}$ rate for stationary Gaussian processes (Dahlhaus 1989). The Quasi-MLE has this property as well for a range of assumptions on the error

[3]If $\Omega = I$, the $k \times k$ identity matrix, we have an unweighted form of the estimator. In simulations we found substantial contributions to finite-sample efficiency from the use of $\Omega = \text{cov}(\tilde{a})^{-1}$, and this form alone is used in the results presented below.

[4]It is noteworthy that for the pure ARFIMA(0,d,0) case, an estimator can be based on the first coefficient of the approximating AR model, since that first coefficient converges to $-d$ in this case, which corresponds to (2.1.3). This point has a parallel in the use of the same estimator by Galbraith and Zinde-Walsh (1994) for the pure MA(1) model (and also in the fact, noted by Tieslau et al., that a consistent estimator of d can be based on the first autocorrelation; that is, $\hat{d} = \frac{\hat{\rho}_1}{(1+\hat{\rho}_1)}$). The fact that the same estimate can serve for either of two different models underlines, of course, the importance of model selection in determining the character of the estimated representation.

process; see Giraitis and Surgailis (1990). The MDE of Tieslau et al. converges at the standard rate to an asymptotic normal distribution for $d \in (-\frac{1}{2}, \frac{1}{4})$; at $d = \frac{1}{4}$, convergence is to the normal distribution at rate $(\frac{T}{log(T)})^{\frac{1}{2}}$, and for $d \in (\frac{1}{4}, \frac{1}{2})$ convergence is at rate $T^{(\frac{1}{2}-d)}$ and the limiting distribution is non-normal. For the Geweke and Porter-Hudak estimator based on least squares estimates of a regression with dependent variable given by the harmonic ordinates of the periodogram, and for a generalized version which discards a number of lower frequencies, the asymptotic properties are obtained by Robinson (1995). Asymptotic properties of the indirect inference estimator for long memory models (Martin and Wilkins 1999) have not been established.

Before considering properties of ARFIMA parameter estimates based on autoregression, we examine the estimates $\tilde{a}$ of the autoregressive parameters themselves. In subsection 3.1 we show consistency of $\tilde{\alpha}$ in both stationary and non-stationary cases, and show that a Normal approximation to the asymptotic distribution can be used. Consistency and distributional results for the estimates $\tilde{\omega}_\delta$ of the ARFIMA parameters are given in 3.2. Simulation results describing Finite sample performance appear in 3.3.

3.1 Asymptotic properties of estimators $\tilde{a}$

The first set of results concerns consistency of the autoregressive parameter estimates, from which parametric model parameter estimates are later deduced, in both stationary and non-stationary cases. As noted above we use the Yule-Walker estimator for $\tilde{a}$ in the stationary case; the same properties then hold for OLS and spectral estimates. Theorem 1 establishes that $\tilde{\alpha}$ is a consistent estimator of $\delta_{[1,k]}$ as $T \to \infty,\ k \to \infty$ in the stationarity region; the cases $0 < d < \frac{1}{2}$ and $-\frac{1}{2} < d < 0$ (antipersistence) require somewhat different

treatment. Theorem 2 establishes the result, using OLS, in the non-stationary case $\frac{1}{2} < d < 1$. For $d = 1$, the process contains a unit root; see for example Phillips (1987) for asymptotic results. For brevity we omit the case where $d = \frac{1}{2}$. A proof of consistency can be constructed similarly to that of Theorem 2, using the rates appropriate to the $d = \frac{1}{2}$ case.

Theorem 1. For $d \in (-\frac{1}{2}, \frac{1}{2})$ and $a(k)$ as defined in (2.2.4), $\|a(k) - \delta_{[1,k]}\| = O(k^{|d|-\frac{1}{2}})$ as $k \to \infty$. Under the assumptions in Section 2.1, for any ν, ε there exist $k, \tilde{T}$ such that $\Pr(\|a(k) - \delta_{[1,k]}\| > \varepsilon) < \nu \quad \forall T > \tilde{T}$.

Proof: See Appendix A.

Next consider the nonstationary case $\frac{1}{2} < d < 1$. Shimotsu and Phillips (1999) discuss two possible characterizations of non-stationary I(d) processes for $\frac{1}{2} < d < \frac{3}{2}$: one defines y_t as a partial sum of a stationary fractionally integrated process z_t, so that $y_t = y_0 + \sum_{i=1}^{t} z_i$, while the other defines y_t via the expansion of the function of the lag operator, $(y_t - y_0) = \sum_{i=0}^{t-1} \left(\frac{1}{i!}\right) \frac{\Gamma(d+i)}{\Gamma(d)} \varepsilon_{t-i}$. Each leads to an expression of the form $(1-L)^d(y_t - y_0) = \varepsilon_t$; Shimotsu and Phillips show that the essential difference lies in the fact that the first definition involves the presence of pre-sample values. Here we treat y_t as a partial sum and assume that y_0 and all pre-sample values are zeroes; the same results would hold if we assumed that $\max\{y_t : t \leq 0\} = O_p(1)$. Note that while Theorems 2 and 4 use the partial sum representation, this estimator does not rely on differenced data in the non-stationary case, so that a *priori* knowledge of the range in which d lies is not necessary.

Theorem 2. For $d \in (\frac{1}{2}, 1)$, let the differenced process $z_t = y_t - y_{t-1},\ t = 2, \ldots, T$, be a

stationary ARFIMA process satisfying the assumptions of Section 2.1, with $d' = 1 - d < 0$, and let $y_{-i} = 0 \quad \forall i \geq 0$. Then as $T \to \infty, k \to \infty, T^{\frac{1}{2}} k \to 0$, we have

$$\|\tilde{\alpha} - \delta_{[1,k]}\| = \begin{cases} O_p(k^{-3d+\frac{3}{2}}) & \text{if } -\frac{1}{2} < d < \frac{3}{4}, \\ O_p(k^{-d}) & \text{if } d \geq \frac{3}{4}. \end{cases}$$

Proof: See Appendix B.

Next we characterize the asymptotic distributions of the estimators $\tilde{\alpha}(k)$ of δ, in the stationarity region, as $T, k \to \infty$.

For fixed k and $-\frac{1}{2} < d < \frac{1}{4}$, the estimator $\tilde{a}(k)$ has an asymptotic Normal distribution with the usual convergence rate of $T^{\frac{1}{2}}$ and asymptotic mean of $a(k)$ (Yajima 1992). For the coefficients δ of the infinite autoregression, the difference $\tilde{a}(k) - \delta_{[1,k]}$ can be represented as the sum of $(\tilde{a}(k) - a(k))$ and $(a(k) - \delta_{[1,k]})$; the first of these terms has an asymptotic Normal distribution, and the second goes to zero, by Theorem 1, as $k \to \infty$. The Normal distribution can therefore be used for inference in large samples, for $-\frac{1}{2} < d < \frac{1}{4}$.

For $\frac{1}{4} < d < \frac{1}{2}$, the asymptotic distribution of $\tilde{\alpha}(k)$ is not Normal, but is related to the Rosenblatt distribution; the convergence rate is non-standard (Yajima 1992), and if an estimated mean is subtracted the asymptotic distribution changes. Nonetheless, we can again represent $\tilde{a}(k) - \delta_{[1,k]}$ as a sum of two terms, one of which has an asymptotic Normal distribution, and the second of which is a 'correction' term which can be made arbitrarily small in probability as $T, k \to \infty$. This representation applies to all $-\frac{1}{2} < d < \frac{1}{2}$, and is based on Hosking (1996), where asymptotic normality is established for differences of sample covariances, for all stationary ARFIMA processes, under the conditions given in Section 2.1. Hosking's results apply to cases where the mean is estimated. Our method

of proof uses *sequential asymptotics*, and thus does not provide an asymptotic Normal distribution, but rather a Normal *approximation* to the finite-sample distribution for which the covariance and closeness to the true distribution are governed by the choice of k for sufficiently large T. The result is therefore not in conflict with Yajima's asymptotic result, but nonetheless indicates that it is possible to conduct approximate inference using the Normal distribution, for d in this range.

Theorem 3. Under the conditions of Theorem 1, the difference $\tilde{a}(k) - \delta_{[1,k]}$ can be represented as the sum $\zeta(k) + \eta(k)$, where for any positive $\varepsilon,\ \nu$ there exists k sufficiently large that $\Pr(\|\eta(k)\| < \varepsilon) > 1 - \nu$, and $T^{\frac{1}{2}}\zeta(k)$ has a limiting Normal distribution of the form $N(0, W(k))$.

Proof: See Appendix C. The asymptotic covariance matrix $W(k)$ is given in the proof of Theorem C in Appendix C.

As noted earlier, we suggest choosing $k = O(\ln T)$; a particular rule is given below.

Figure 1(a–d) illustrates the approximation provided by the Normal by depicting the simulated densities of the first and second autoregressive coefficients in estimated AR representations of ARFIMA$(0, d, 0)$ models, $d = \{0.4, 0.5, 0.7, 0.9\}$, $T = 100$ $(k = 8)$. In each case the Normal approximation is very close to the true distribution of the AR coefficients; small but clearly discernible departures from the Normal are visible for larger values of d, particularly near the means of the distributions. Note that these simulations include values of $d \geq 0.5$, to which Theorem 3 does not apply; nonetheless the Normal provides a reasonable finite-sample approximation.

3.2 Asymptotic properties of estimators of ARFIMA parameters ω

In order to discuss consistency of the ARFIMA parameter estimates $\tilde{\omega}_\delta$ given in (2.2.7) above, we need an additional condition.

Condition 3.2. There exists a non-stochastic infinite-dimensional matrix Π corresponding to a bounded norm operator on the space L_2 such that $\|\Omega - \Pi_k\| \to 0$, where Π_k is the $k \times k$ principal sub-matrix of Π (convergence is in probability if Ω is stochastic).

If the estimator is used in unweighted form ($\Omega = I$), then all of the matrices involved are identity matrices and this condition is trivially satisfied. If the weighting matrix is an inverse covariance matrix (known or estimated) of a stationary and invertible process, then Π corresponds to the Toeplitz form of the inverse process.

The next theorem shows that $\tilde{\omega}_\delta$ is a consistent estimator of the true vector of parameters ω_0, and Theorem 5 shows that the difference between the two can again be represented as a sum of two terms, one with an asymptotically Normal distribution, and one which goes to zero in probability as $k, T \to \infty$. We denote the quadratic form $(\tilde{a} - \delta(\omega))'\Omega(\tilde{a} - \delta(\omega))$ by $Q_{T,k}(\tilde{\alpha}, \omega)$.

Theorem 4. Under the conditions of Theorems 1 and 2 and condition 3.2, and where $T \to \infty$, $k \to \infty$ and $T^{\frac{1}{2}}k \to 0$, the minimum-distance estimator $\tilde{\omega}_\delta$ is consistent for ω_0, the true parameter vector of a correctly-specified ARFIMA model. The corresponding distance function $\min Q_{T,k}(\tilde{\alpha}, \omega)$ goes to zero in probability.

Proof: See Appendix C.

Again it is possible to conduct inference using an approximate Normal distribution, as Theorem 5 indicates.

Theorem 5. Under the conditions of Theorem 1, the difference $\tilde{\omega}_\delta - \omega_0$ can be represented as the sum $\xi(k) + b(k)$ where for each k the correction term is

$$b(k) = \left(\frac{\partial \delta'(\omega_0)}{\partial \omega} \Omega \frac{\partial \delta(\omega_0)}{\partial \omega'} \right)^{-1} \frac{\partial \delta'(\omega_0)}{\partial \omega} \Omega \eta(k),$$

with $\eta(k)$ as given in Theorem 3, and $T^{-\frac{1}{2}}\xi(k)$ is asymptotically distributed as $N(0, V(k))$ as $T \to \infty$, for each k.

Proof: See Appendix C. The asymptotic covariance matrix $V(k)$ is given in the proof.

3.3 Performance in finite samples

To investigate the performance of these estimators in small samples, we generate simulated samples from a selection of ARFIMA$(0, d, 0)$, $(1, d, 0)$ and $(1, d, 1)$ processes, and compare the results by parameter root mean squared error (RMSE). ML and autoregressive estimators, along with GPH in the $(0, d, 0)$ cases, are compared. In order to examine the impact of mis-specification we consider several cases of $(0, d, 0)$ models of $(1, d, 0)$ processes; in these cases performance is evaluated by the out-of-sample RMSE of 1-step forecasts.

In all of the following simulations, the mean is treated as unknown and the sample mean is removed from the process prior to ARFIMA parameter estimation. In the frequency domain, removal of an estimated mean is not required; the zero frequency is excluded. However, Cheung and Diebold (1994) show that feasible (i.e., with estimated

mean) time-domain ML and frequency domain ML nonetheless perform similarly in moderately large samples, and in fact that time-domain ML shows lower MSE in small samples. Simulated fractional noise is obtained by transformation of (Gaussian) white noise using the Cholesky decomposition of the exact covariance of the ARFIMA process, obtained using the results of Chung (1994); see also Diebold and Rudebusch (1989). In each case estimators are compared using the same sets of starting values (zero in $(0, d, 0)$ cases; the better of two alternatives is chosen in multiple-parameter cases). Optimization is performed using the Powell algorithm. Throughout, we use the rule of thumb that AR lag length can be chosen as $8 + 3\ln(\frac{T}{100}),\ T \geq 100$, rounded to the nearest integer. This rule is approximately equivalent to $3\ln(T) - 6$; we express it in the former way to emphasize $T = 100$ as the 'base' sample size.

At least 1000 replications are used in each case, more in the $(0, d, 0)$ cases. For multiple-parameter models we use a sample size of 100, in common with much of the literature containing results on exact and approximate ML; the computational cost of the repeated inversion of $T \times T$ covariance matrices associated with ML becomes prohibitive for simulation at large sample sizes. In the $(0, d, 0)$ cases we report results for $T = \{100, 400\}$.

(i) ARFIMA$(0, d, 0)$

We begin by comparing estimators in the pure fractionally-integrated case, which has been the most thoroughly examined in the literature to date. Figure 2a/b plots the RMSE's from ML, GPH and the AR estimator, on the grid of values from $d = -0.4$ to $d = 0.9$ at an interval of 0.1.[5] Both ML and AR show lower RMSE than GPH throughout the stationarity

[5]ML is computed only in the stationarity region where it can be obtained without prior

and invertibility region and, for the AR estimator, this holds for the non-stationary values considered as well. The AR estimator has lower RMSE than ML at moderate absolute values of d, slightly higher RMSE at $d = 0.3, 0.4$, and substantially higher at $d = -0.4$. However, ML is constrained by the optimization algorithm to lie in $d \in (-0.5, 0.5)$, whereas AR is not constrained in this way, being usable outside this interval; ML therefore benefits near -0.5 and 0.5 from being constrained to lie in a region around the true value which is fairly tightly bounded on one side.

Qualitative results do not differ greatly between the two sample sizes of Figures 2a and 2b. At $T = 400$, the RMSE of the AR estimator is almost completely insensitive to the value of d, except at -0.4 where it rises to almost the value produced by GPH.

(ii) ARFIMA$(1, d, 0)$ *and ARFIMA*$(1, d, 1)$

The first multiple-parameter cases that we consider are ARFIMA$(1, d, 0)$, a simple model which allows short-run and long-run components. In these cases, one element of the estimation error arises through the difficulty in discriminating these two components at small sample sizes, since AR and fractional integration parameters can imply similar patterns of autocorrelation for the first few lags. Table 1 gives results for a few cases similar to the ARFIMA$(1, d, 1)$ processes that we will address below: $d = \{-0.3, 0.3\}$ and $\rho = \{0.7, -0.2\}$.

knowledge of the need for differencing. The maximum number of periodogram ordinates is included for GPH, which is optimal for this case in which there is no short memory component.

Table 1

RMSE's of ARFIMA(1,d,0) parameter estimates

$T = 100$; 2000 replications (ML), 10000 replications (AR)

d	ρ	$\hat{d}_{AR}$	$\hat{d}_{ML}$	$\hat{\rho}_{AR}$	$\hat{\rho}_{ML}$
-0.3	0.7	0.266	0.186	0.287	0.195
-0.3	-0.2	0.110	0.120	0.137	0.139
0.3	0.7	0.179	0.237	0.150	0.116
0.3	-0.2	0.209	0.233	0.217	0.241

Neither estimator dominates in these examples; ML is markedly better at (-0.3, 0.7), AR is better at (0.3, -0.2). The two are very similar at (-0.3, 0.2). For the case (0.3, 0.7), possibly the most interesting in combining positive d with positive short-memory autocorrelation, d is better estimated by the AR estimator, and ρ by ML.

The ARFIMA$(1, d, 1)$ parameterizations examined in Table 2 are those used by Chung and Baillie (1993). Note that in each case the short-memory parameters ρ and θ have the same sign, so that the corresponding terms $(1-\rho L)$ and $(1+\theta L)$ in the lag polynomials do not approach cancellation; at $\rho = 0.5, \theta = -0.5$, for example, cancellation would take place and the apparent ARFIMA$(1, d, 1)$ process would in fact be ARFIMA$(0, d, 0)$. In cases of near-cancellation, the process may be well approximated by a process with substantially different parameter values, but similar roots of the short-memory polynomials, leading to difficulty in evaluating the estimates. Such cases are ruled out by parameter values such as these.

Table 2

RMSE's of ARFIMA(1,d,1) parameter estimates

$T = 100$; 1000 replications (ML), 5000 replications (AR)

(d, ρ, θ)	$\hat{d}_{AR}$	$\hat{d}_{ML}$	$\hat{\rho}_{AR}$	$\hat{\rho}_{ML}$	$\hat{\theta}_{AR}$	$\hat{\theta}_{ML}$
(-0.3, 0.5, 0.2)	0.251	0.163	0.350	0.220	0.213	0.163
(-0.3, 0.2, 0.5)	0.207	0.151	0.295	0.237	0.152	0.154
(-0.3,-0.5,-0.2)	0.339	0.162	0.136	0.139	0.391	0.254
(-0.3,-0.2,-0.5)	0.333	0.179	0.183	0.173	0.291	0.250
(0, 0.5, 0.2)	0.285	0.298	0.307	0.235	0.198	0.154
(0, 0.2, 0.5)	0.291	0.285	0.332	0.313	0.144	0.139
(0,-0.5,-0.2)	0.228	0.187	0.154	0.142	0.328	0.265
(0,-0.2,-0.5)	0.243	0.263	0.220	0.202	0.384	0.381
(0.3, 0.5, 0.2)	0.375	0.408	0.285	0.271	0.178	0.152
(0.3, 0.2, 0.5)	0.375	0.390	0.378	0.381	0.132	0.124
(0.3,-0.5,-0.2)	0.231	0.209	0.156	0.149	0.324	0.285
(0.3,-0.2,-0.5)	0.286	0.314	0.238	0.207	0.434	0.430

The RMSE's of ML estimates are generally somewhat better than those reported by Chung and Baillie for the approximate CSS estimator. The pattern of relative RMSE's for the AR and ML estimators is broadly similar to that observed in lower-order cases. Neither estimator dominates; ML gives lower RMSE for d in seven of the twelve cases, AR in five. Four of the cases in which ML is superior are the four cases with negative d, where ML's performance is markedly better. For non-negative values, performance of the two for d is similar. In estimation of the short-memory parameters, ML shows an advantage throughout the parameter space, although the differences in RMSE between the techniques are usually quite small.

(iii) mis-specified cases

In the final set of simulations we consider ML and AR estimates of d in processes that are ARFIMA$(1, d, 1)$, but where the model used is ARFIMA$(0, d, 0)$. These results serve to illustrate the point that the ARFIMA models estimators need not be seen as purely 'parametric' estimators, in the sense of depending crucially on a correct parameterization of the process; this is true in particular of the AR estimator, where the autoregressive structure which extracts statistical information from the data is capable of fitting a very general class of processes, and for some purposes useful estimates can also be obtained from mis-specified models. In this sense this estimator may be thought of as semi-parametric: the long-memory component is captured via the parametric fractionally-integrated model with parameter d, and the short-memory component via a variable number of AR (or ARMA) terms, which may be increased with sample size to detect increasingly subtle short-memory features as sample information accumulates.

In mis-specified cases the obvious criterion of evaluation, accuracy of parameter estimates, is not applicable; some parameters are missing, and in such cases it will typically be optimal to deviate from the 'true' values of parameters for some purposes. For this exercise, we therefore evaluate the mis-specified models by the accuracy (RMSE) of 1-step out-of-sample forecasts of the true process. We consider $d = \{-0.3, 0., 0.3\}$ and $(\rho, \theta) = \{(0.7, 0.5), (0., 0.), (0.2, 0.2)\}$, for a total of nine cases (cases in which $(\rho, \theta) = \{(0, 0)$ are of course correctly specified, and are included for comparison). We again use $T = 100$ in each case, and noise variance is unity. Results are presented in Table 3. This small set of results is of course not intended to be definitive, but to indicate the possibility that

the flexible AR form may have advantages, for some purposes such as forecasting, where model form is unknown.

Table 3

RMS Forecast Errors in ARFIMA(0,d,0) models of ARFIMA(1,d,1) processes
$T = 100$, *5000 replications*

d	ρ	θ	AR	ML
-0.3	0.	0.	1.0026	1.0016
0.	0.	0.	1.0028	1.0036
0.3	0.	0.	1.0103	1.0103
-0.3	0.2	0.2	1.0263	1.0262
0.	0.2	0.2	1.0265	1.0256
0.3	0.2	0.2	1.0280	1.0351
-0.3	0.7	0.5	1.1378	1.1566
0.	0.7	0.5	1.1382	1.2728
0.3	0.7	0.5	1.1381	1.5823

In the correctly-specified comparison cases, these RMS forecast error results mirror the parameter RMSE results of Figure 2a; ML is superior at d =-0.3, AR at d =0, and the two are very close at $d = 0.3$. Where $(\rho, \theta) = (0.2, 0.2)$, ML produces slightly better results at $d = -0.3$ and 0. However, the AR forecasts are substantially better at $d = 0.3$, a phenomenon which appears more strongly at $(\rho, \theta) = (0.7, 0.5)$. In all of the latter cases, AR estimates produce markedly better 1-step forecasts. Note that in cases such as these where there is substantial autocorrelation from the short-memory parameters, the AR estimator is able to compensate for the lack of short-memory components in the estimated model by raising the estimate of d beyond what is possible for ML because of the stationarity

requirement. Of course, time-domain ML forecasts with non-stationary values of d can also be produced by imposing a differenced representation; however, to do so it is again necessary to make a determination of an optimal degree of differencing, either *a priori* or based on features of the data, which the AR estimator does not require.

4. Concluding remarks

Estimation of long-memory models by autoregressive approximation is feasible even in quite small samples, and is consistent across a range of stationary and non-stationary values of the long-memory parameter, so that estimation does not require knowledge of an appropriate transformation to apply.

Autoregressive estimation, and model selection based on these estimates, is often convenient in choosing starting values for techniques such as ML; estimation based on this principle has been used for ARMA models as far back as Durbin (1960). While this is one potential use of autoregressive estimates, such estimates seem to have desirable features beyond their ability to produce good starting values quickly. Their relative insensitivity to the stationarity of the process, and ability for a wide class of processes to provide an arbitrarily good statistical approximation as sample size and order increase, make them well suited to the treatment of processes of unknown form. This may be especially valuable in economic data for which the ARFIMA class itself will typically be only an approximate representation.

References

Agiakloglou, C., P. Newbold and M. Wohar (1992) Bias In an Estimator of The Fractional Difference Parameter. *Journal of Time Series Analysis* 14, 235-246.

Amemiya, T. (1985) *Advanced Econometrics*. Harvard University Press, Cambridge, MA.

Berk, K.N. (1974) Consistent Autoregressive Spectral Estimates. *Annals of Statistics* 2, 489–502.

Cheung, Y.W. and F.X. Diebold (1994) On Maximum-Likelihood Estimation of the Differencing Parameter of Fractionally-Integrated Noise with Unknown Mean. *Journal of Econometrics* 62, 311-316.

Chung, C.-F. (1994) A note on calculating the autocovariances of the fractionally integrated ARMA models. *Economics Letters* 45, 293-297.

Chung, C.-F. and R.T. Baillie (1993) Small Sample Bias in Conditional Sum of Squares Estimators of Fractionally-Integrated ARMA Models. *Empirical Economics* 18, 791-806.

Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes. *Annals of Statistics* 17, 1749-1766.

Davies, R.B. and D.S. Harte (1987) Tests for Hurst Effect. *Biometrika* 74, 95-101.

Diebold, F.X. and G.D. Rudebusch (1989) Long Memory and Persistence in Aggregate Output. *Journal of Monetary Economics* 24, 189-209.

Durbin, J. (1960). The Fitting of Time Series Models. *Review of the International Statistical Institute* 28, 233-243.

Galbraith, J.W. and V. Zinde-Walsh (1994) A Simple, Non-Iterative Estimator for Moving-Average Models. *Biometrika* 81, 143-155.

Galbraith, J.W. and V. Zinde-Walsh (1997) On Some Simple, Autoregression-based Estimation and Identification Techniques for ARMA Models. *Biometrika* 84, 685-696.

Geweke, J. and S. Porter-Hudak (1983) The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis* 4, 221-238.

Giraitis, L. and D. Surgailis (1990) A Central Limit Theorem for Quadratic Forms in Strongly Dependent Linear Variables and its Application to Asymptotic Normality of Whittle's Estimate. *Probability Theory and Related Fields* 86, 87-104.

Granger, C.W.J. (1980) Long Memory Relationships and the Aggregation of Dynamic Models. *Journal of Econometrics* 14, 227-238.

Granger, C.W.J., and R. Joyeux (1980) An Introduction to Long-Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis* 1, 15-39.

Grenander, U. and G. Szegö (1958) *Toeplitz Forms and Their Applications*. University of California Press, Berkeley.

Hosking, J.R.M. (1981) Fractional Differencing. *Biometrika* 68, 165-176.

Hosking, J.R.M. (1996) Asymptotic Distributions of the Sample Mean, Autocovariances, and Autocorrelations of Long-memory Time Series. *Journal of Econometrics* 73, 261-284.

Hurst, H.E. (1951) Long-Term Storage Capacity of Reservoirs.*Transactions of the American Society of Civil Engineers* 116, 770-799.

Janacek, G.J. (1982) Determining the Degree of Differencing for Time Series via the Long Spectrum. *Journal of Time Series Analysis* 3, 177-183.

Li, W.K. and A.I. McLeod (1986) Fractional Time Series Modelling. *Biometrika* 73, 217-221.

Martin, V.L. and N.P Wilkins (1999) Indirect Estimation of ARFIMA and VARFIMA Models. *Journal of Econometrics* 93, 149-175.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277-301.

Priestley, M.B. (1981) *Spectral Analysis and Time Series.* Academic Press, New York.

Robinson, P.M. (1991) Testing for Strong Serial Correlation and Dynamic Heteroskedasticity in Multiple Regression. *Journal of Econometrics* 47, 67-84.

Robinson, P.M. (1995) Log-periodogram Regression of Time Series with Long-Range Dependence. *Annals of Statistics* 23, 1048-1072.

Saikkonen, P. (1986). Asymptotic Properties of some Preliminary Estimators for Autoregressive Moving Average Time Series Models. *Journal of Time Series Analysis* 7, 133-155.

Shimotsu, K. and P.C.B. Phillips (1999) Modified Local Whittle Estimation of the Memory Parameter in the Nonstationary Case. Working paper, Yale University.

Sowell, F.B. (1992) Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models. *Journal of Econometrics* 53, 165-188.

Tanaka, K. (1999) Nonstationary Fractional Unit Roots. *Econometric Theory* 15, 549-582.

Tieslau, M.A., P. Schmidt and R.T. Baillie (1996) A Minimum-Distance Estimator for Long-Memory Processes. *Journal of Econometrics* 71, 249-264.

Yajima, Y. (1992) Asymptotic Properties of Estimates in Incorrect ARMA Models for Long-memory Time Series. Proc. IMA, Springer-Verlag, 375-382.

old stuff

We also consider estimators that allow us to use Hosking's (1996) results on asymptotic normality of the differences between sample covariances. An estimator with these properties that is similar in form to $\tilde{a}_{YW}$ can be constructed in the following way. We define the vector $\gamma(k,-) = (\gamma_1 - \gamma_2, \gamma_2 - \gamma_3, \ldots \gamma_k - \gamma_{k+1})$ and its sample analogue $\hat{\gamma}(k,-)$, and the matrix $\Sigma(k,-)$ with typical element $\{\Sigma(k,-)\}_{ij} = \gamma_{|i-j|} - \gamma_{|i-j+1|}$, and its sample analogue $\hat{\Sigma}(k,-)$. Define the estimator $\hat{a}(k,-)$ as the solution to

$$\hat{\Sigma}(k,-)\hat{a}(k,-) = \hat{\gamma}(k,-), \tag{2.2.5}$$

and let $a(k,-)$ be the solution to

$$\Sigma(k,-)a(k,-) = \gamma(k,-). \tag{2.2.6}$$

As $\text{plim}(\tilde{a})$ is denoted by $a(k)$, here $a(k,-) = \text{plim}(\hat{a}(k,-))$.

An estimator with a simpler form was proposed by Hosking, who suggested that ratios of differences of autocorrelations can be used. Indeed, consider the vector $\hat{a}_H$ with components

$$\hat{a}_{H(i)} = \frac{1-\hat{\rho}_i}{1-\hat{\rho}_{k+1}},$$

for $i = 1, \ldots, k$. This vector converges at rate $T^{-\frac{1}{2}}$ to its population analogue a_H and has an asymptotically normal distribution for all $d \in (-\frac{1}{2}, \frac{1}{2})$.

Appendix A.

Proof of Theorem 1.

Consider the expression in (2.2.6).

For the norm of the left-hand-side vector we can write

$$\left\|a(k) - \delta_{[1,k]}\right\| \leq \left\|\Sigma(k)^{-1}\right\| \left\|\Sigma_{[k+1,\infty)}\delta_{[k+1,\infty)}\right\|.$$

We begin by evaluating the norm of $\Sigma_{[k+1,\infty)}\delta_{[k+1,\infty)}$. Denote the j-th component of this vector by s_j where

$$s_j = \sum_{i=1}^{\infty} \gamma_{|k+i-j|}\delta_{k+i},\ j = 1, 2, ...k.$$

Recall that we can bound $|\gamma_l|$ and $|\delta_l|$ as follows:

$$|\gamma_l| < c_\gamma l^{2d-1} \text{ and } |\delta_l| < c_\delta l^{-d-1} \text{ for some } c_\gamma, c_\delta > 0.$$

Then for $-\frac{1}{2} < d < 0$ we have

$$s_j \leq c_\gamma c_\delta \sum_{i=1}^{\infty}(k-j+i)^{2d-1}(k+i)^{-d-1} \leq \tilde{c}k^{-d-1}(k-j)^{2d},$$

where $\tilde{c} > 0$ is some constant. The last inequality is obtained by noting that $\sum_{i=1}^{\infty}(k-j+i)^{2d-1}$ is $O((k-j)^{2d})$.

For $0 < d < \frac{1}{2}$,

$$s_j \leq c_\gamma c_\delta \sum_{i=1}^{\infty}(k-j+i)^{2d-1}(k+i)^{-d-1} \leq \tilde{c}k^{-d}(k-j)^{2d},$$

where $\tilde{c} > 0$ is some constant. Here the last inequality similarly follows from $\sum_{i=1}^{\infty}(k+i)^{-d-1} = O(k^{-d})$.

The norm of the vector with components s_j is thus $\|s\| = \left(\sum_{j=1}^{k} s_j^2\right)^{\frac{1}{2}}$ and

$$\begin{aligned} \|s\| &\leq \left(\tilde{c}k^{-2d-2}\textstyle\sum_{j=1}^{k} j^{4d}\right)^{\frac{1}{2}} = O(k^{\frac{2d-1}{2}}),\ \text{if } -\tfrac{1}{2} < d < 0; \\ \|s\| &\leq \left(\tilde{c}k^{-2d}\textstyle\sum_{j=1}^{k} j^{4d-2}\right)^{\frac{1}{2}} = O(k^{\frac{2d-1}{2}}),\ \text{if } 0 \leq d < \tfrac{1}{2}. \end{aligned} \tag{A.1}$$

Next consider $\|\Sigma(k)^{-1}\|$. It is well known that if all the eigenvalues of $\Sigma(k)$ are greater than some $\phi(k) > 0$ then $\|\Sigma(k)^{-1}\| < (\phi(k))^{-1}$.

Consider the spectral density function associated with Σ; denote by $f(x)$ the spectral density multiplied by 2π. Here

$$f(x) = [P(e^{ix})P(e^{-ix})]^{-1}(2 - 2\cos x)^{-d}Q(e^{ix})Q(e^{-ix}).$$

The polynomials P, Q with none of the roots on the unit circle are bounded there from above and below, providing bounds on $f(x)$;

$$B_L(2 - 2\cos x)^{-d} < f(x) < B_U(2 - 2\cos x)^{-d}.$$

By Grenander and Szego (p.64) this implies that the eigenvalues associated with $\Sigma(k)$ are bounded from below by the eigenvalues associated with the function $f_L(x) = B_L(2 - 2\cos x)^{-d}$.

In the case $0 < d < \frac{1}{2}$ the lower bound of $f_L(x)$ is $B_L 4^{-d}$; it follows that

$$\|a(k) - \delta\| = O(k^{d-1/2}).$$

When $d < 0$ the lower bound of $f_L(x)$ is zero. We use a different approach to obtain a tighter bound for the eigenvalues of $\Sigma(k)$ from below and evaluate the rate at which this lower bound goes to zero. To do this we use properties of eigenvalues of symmetric matrices and of circulant matrices.

Property 1 (Separation Theorem; Wilkinson (1965), pp.103-104). If S_n is a symmetric matrix and is partitioned as

$$S_n = \begin{bmatrix} S_{n-1} & s \\ s' & 1 \end{bmatrix},$$

then the eigenvalues $\lambda_i(S_{n-1})$ of S_{n-1} numbered in order of magnitude separate the eigenvalues $\lambda_i(S_n)$ of S_n:

$$\lambda_1(S_n) \leq \lambda_1(S_{n-1}) \leq \lambda_2(S_n) \leq ... \leq \lambda_n(S_n).$$

Property 2 (Priestley(1981), p.261). If C_k is a circulant matrix (and also

a Hermitian Toeplitz matrix) of the form

$$C_k = \begin{bmatrix} \gamma_o & \gamma_1 & \cdots & \gamma_{k-1} & \gamma_k & \gamma_k \cdots & \gamma_1 \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{k-1} & \gamma_k & \cdots \\ \vdots & \ddots & \ddots & \ddots & \cdots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \cdots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \gamma_0 & \gamma_1 \\ \gamma_1 & \cdots & \cdots & \cdots & \cdots & \gamma_1 & \gamma_0 \end{bmatrix},$$

then its eigenvalues (here not subscipted by order of magnitude) are given by

$$\lambda_l(C_k) = \sum_{r=-k}^{k} \gamma_r exp(-i\omega_l r), \text{ where } \omega_l = \frac{2\pi l}{2k+1}. \tag{A.2}$$

We shall evaluate the eigenvalues of $\Sigma(k)$ by embedding this matrix in a circulant C_k. By Property 1 the smallest eigenvalue of $\Sigma(k)$ is bounded from below by the smallest eigenvalue of C_k, which will be given by $\lambda_l(C_k)$ in (A.2) for an appropriate ω_l, moreover, by applying the separation theorem repeatedly we can see that the smallest eigenvalue of $\Sigma(k)$ is bounded from below by any of k smallest eigenvalues of C_k.

The $\lambda_l(C_k)$ for large k is approximated by $f(\omega_l) = \sum_{r=-\infty}^{\infty} \gamma_r exp(-i\omega_l r)$, which without loss of generality we can assume to equal $(1-\cos\frac{2\pi}{k})^{-d}$, moreover for $l = 1$ we have that

$$\lambda_1(C_k) - f(\omega_1) = -2 \sum_{r=k+1}^{\infty} \gamma_r \cos(\omega_1 r) > 0$$

by comparing magnitudes of groups of positive and negative terms in the sum, thus

$$\lambda_1(C_k) \geq f(\omega_1) = (1 - \cos\frac{2\pi}{k})^{-d} \approx O(k^{2d}),$$

where $\approx$ indicates that $\lambda_1(C_k)$ declines at exactly the rate $O(k^{2d})$. Next we show that $\lambda_1(C_k)$ is among the k smallest eigenvalues. Indeed, for any l

$$|\lambda_l(C_k) - f(\omega_l)| = \left| 2 \sum_{r=k+1}^{\infty} \gamma_r \cos(\omega_l r) \right| = O(k^{2d}).$$

On the other hand, if $\frac{k}{2} > l > k^{\alpha}$ with $\alpha < 1$

$$f(\omega_l) = (1 - \cos\frac{2\pi l}{k})^{-d} \approx O(k^{2d(1-\alpha)}),$$

thus for large enough k the corresponding $\lambda_l(C_k) > \lambda_1(C_k)$.

It follows that $\|\Sigma(k)^{-1}\| = O(k^{-2d})$. Combining this with (A.1) we get that

$$\left\|a(k) - \delta_{[1,k]}\right\| = O(k^{-2d+d-\frac{1}{2}}) = O(k^{-d-\frac{1}{2}}).$$

Finally, for any ν, ε choose k such that $\left\|a(k) - \delta_{[1,k]}\right\| < \frac{\varepsilon}{2}$; then for that value of k, by consistency of $\tilde{a}$ for $a(k)$ (Yajima, 1992), find $\tilde{T}$ such that $\Pr(\|\tilde{a} - a(k)\| > \frac{\varepsilon}{2}) < \nu$ for samples of size $T > \tilde{T}$.

Appendix B.

We begin with a lemma which will be useful for the proof of Theorem 2.

Lemma B. *Under the assumptions in Theorem 2*

.

$$T^{-2d}\sum_{t=k}^{T} y_t^2 \ = O_p(1), \text{ but not } o_p(1); \tag{B.1}$$

$$Ez_i y_j \quad = \begin{cases} O(1) & \text{if } i \le j, \\ O(i-j)^{2d-2} & \text{if } i > j; \end{cases} \tag{B.2}$$

$$E(z_i y_j)^2 \ = O(j); \tag{B.3}$$

$$E(z_i y_{i-l} z_j y_{j-l}) = O(1) \text{ for } i \neq j; \tag{B.4}$$

$$E(T^{-1}\Sigma z_{t-l} y_t)^2 = O(1); \tag{B.5}$$

$$T^{-1}\Sigma z_{t-l} y_t = O_p(1). \tag{B.6}$$

Proof.

For y_t defined here as the partial sum of a stationary fractionally integrated process $y_t = \sum_{i=1}^{t} z_i$ it was shown by Shimotsu and Phillips (1999) that the order of magnitude of y_t is similar to that obtained when the process is defined directly via the expansion of the operator $(1-L)^d$. Thus similarly by Lemma 2.13 in Shimotsu and Phillips we have that $Ey_T^2 = O(T^{2d-1})$ and from Tanaka (1999) $T^{-2d}\sum_{t=k}^{T} y_{t-1}^2$ converges in law to a functional of an

integrated Brownian motion and is thus bounded in probability, which gives (B.1).

Since y_t is a partial sum of z where $E(z_i z_j) = \gamma_{i-j}$, we have that $Ez_i y_j = \sum_{l=1}^{j} \gamma_{i-l}$. Recall that γ_m for the process z is $O(m^{2d-3})$; (B.2) follows.

Next we consider the MA representation $z_t = \Sigma_0^\infty \xi_l e_{t-l}$ and recall that $\xi_l = O(l^{d-2})$, and that e has bounded fourth moments by the assumptions in 2.1. We can express

$$\begin{aligned}
&E(z_{i_1} z_{i_2} z_{i_3} z_{i_4}) = E(e^4) \textstyle\sum_{l=0}^{\infty} \xi_l \xi_{l+i_1-i_2} \xi_{l+i_1-i_3} \xi_{l+i_1-i_4} + \\
&(E(e^2))^2 \left[\gamma_{i_1-i_2}\gamma_{i_3-i_4} + \gamma_{i_1-i_3}\gamma_{i_2-i_4} + \gamma_{i_1-i_4}\gamma_{i_2-i_3}\right] \\
&= O(|i_1 - i_2|^{d-2} |i_1 - i_3|^{d-2} |i_1 - i_4|^{d-2} + |i_1 - i_2|^{2d-3} |i_3 - i_4|^{2d-3} \\
&+ |i_1 - i_3|^{2d-3} |i_2 - i_4|^{2d-3} + |i_1 - i_4|^{2d-3} |i_2 - i_3|^{2d-3}).
\end{aligned}$$

for all different subscripts on z, and similarly

$$E(z_i^2 z_l^2) = O(1);\ E(z_j^2 z_l z_m) = O(|j-l|^{3d-3} |j-m|^{2d-3}).$$

Then (B.3) is obtained by expressing

$$E(z_i y_j)^2 = \sum_{l=1}^{j} E(z_i z_l)^2 + 2 \sum_{l=1}^{j} \sum_{m=l+1}^{j} E({z_i}^2 z_l z_m)$$

substituting the expectation for each term from the relation above and evaluating the orders. Note that it is terms $E(z_i^2 z_j^2)$ that provide the largest contribution to the sum.

A similar substitution into the expression

$$E(z_i y_{i-l} z_j y_{j-l}) = \sum_{n=1}^{i-l} \sum_{m=1}^{j-l} z_i z_n z_j z_m$$

provides (B.4); note that this expression has no terms of the form $E(z_i^2 z_j^2)$.

The expression in (B.5) immediately follows and thus (B.6) follows by Chebyshev's inequality.

Proof of Theorem 2.

Consider the AR(∞) representation and write

$$y_t - \delta_1 y_{t-1} - \cdots - \delta_k y_{t-k} = \sum_{i=1}^{\infty} \delta_{k+i} y_{t-k-i} + \varepsilon_t, t = k+1, ..., T \qquad \text{(B.7)}$$

Recall that y with a non-positive subscipt is zero, thus on the right-hand side we have

$$\sum_{i=1}^{t-k} \delta_{k+i} y_{t-k-i} + \varepsilon_t;$$

denote this quantity, which represents the error in the k-th order AR representation, by u_t.

Compare the OLS estimator $\hat{a}' = (\hat{a}_1, ...\hat{a}_k)$ of the model

$$y_t = \delta_1 y_{t-1} + \cdots + \delta_k y_{t-k} + u_t \tag{B.8}$$

with the OLS estimator $\hat{\beta}' = (\hat{\beta}_1, ...\hat{\beta}_k)$ of

$$\begin{aligned} y_t &= X_t\beta + u_t \\ \text{with } X_t &= (Z_t, y_{t-1}) = (z_{t-1}, ...z_{t-k+1}, y_{t-1}),\ \beta' = (\beta_1, ...\beta_k). \end{aligned} \tag{B.9}$$

We note that the one-to-one relation $\delta_1 = \beta_k + \beta_1; \delta_l = \beta_{l-1} - \beta_l$ for $l = 2, ..., k$ produces the same residuals in each model thus we get $\hat{a}_1 = \hat{\beta}_k + \hat{\beta}_1; \hat{a}_l = \hat{\beta}_{l-1} - \hat{\beta}_l$ for $k \geqq l > 1$. If we can establish that $\hat{\beta}$ is a consistent estimator of β consistency of $\hat{a}$ as an estimator of δ will follow.

Consider (B.9); we can write

$$\hat{\beta} - \beta = \left[\sum_{t=k}^{T} X_t' X_t\right]^{-1} \left[\sum_{t=k}^{T} X_t' u_t\right].$$

.

Partition

$$\Gamma = \sum_{t=k}^{T} X_t' X_t = \begin{pmatrix} A & v \\ v' & w \end{pmatrix} \text{ with } A = \sum_{t=k}^{T} Z_t' Z_t; v = \sum_{t=k}^{T} Z_t' y_{t-1}; w = \sum_{t=k}^{T} y_{t-1}^2.$$

Define a diagonal $k \times k$ normalizing matrix $\Lambda = diag(T^{1/2}, ..., T^{1/2}, T^d)$; then

$$\Lambda\Gamma^{-1}\Lambda = \begin{pmatrix} T^{-1}A & T^{-1/2-d}v \\ T^{-1/2-d}v' & T^{-2d}w \end{pmatrix}^{-1}.$$

Since Z_t is stationary and as was demonstrated for $d' = 1 - d < 0$, the covariance matrix for the process $\Sigma(z)$ is such that $\|\Sigma(z)^{-1}\| = O(k^{-2d'})$, it can be shown from Lemma 3 of Berk (1974) that as $T \to \infty$ for k such that

$k^2T^{-1} \to 0$, $\|(T^{-1}A)^{-1} - \Sigma(z)^{-1}\| = o_p(k^{-2d'})$. From (B.6) $T^{-\frac{1}{2}-d}v = o_p(1)$; and from (B.1) $T^{-2d}\omega = O_p(1)$ and is bounded in probability away from zero. Thus the inverse of the partitioned matrix is

$$\Lambda\Gamma^{-1}\Lambda = \begin{pmatrix} \Sigma(z)^{-1}\,(1+o_p(1)) & o_p(1) \\ o_p(1) & O_p(1) \end{pmatrix}.$$

Finally evaluate the order of terms in suitably normalized vector $\sum_{t=k}^{T} X_t'u_t$. We can ignore the ε_t part in u_t and need to concentrate on evaluating the terms $T^{-1}\sum_{t=k}^{T} z_{t-l}\sum_{i=1}^{t-k-1}\delta_{k+i}y_{t-k-i}$ for the subvector $T^{-1}\sum_{t=k}^{T} Z_t'u_t$ of the first $k-1$ components and $T^{-2d}\sum_{t=k}^{T} y_{t-1}\sum_{i=1}^{t-k-1}\delta_{k+i}y_{t-k-i}$ for the kth component.

Consider first $T^{-1}\sum_{t=k}^{T} z_{t-l}\sum_{i=1}^{t-k-1}\delta_{k+i}y_{t-k-i}$ for $l = 1, ..., k-1$; denote this by $W(l)$. Then $W(l) = T^{-1}\sum_{i=1}^{T-k} w_i$ with $w_i = \delta_{k+i}\sum_{t=k+i}^{T} z_{t-l}y_{t-k-i}$. From (B.2)

$$Ew_i = \delta_{k+i}\sum_{t=k+i}^{T} O(k+i-l)^{2d-2} = O(k^{-d-1}T^{2d-1}).$$

Using (B.3)

$$Ew_i^2 = O(k^{-2d-2}T^2).$$

And from (B.4) for $i \neq j$

$$Ew_iw_j = O(k^{-2d-2}T).$$

Thus

$$E(T^{-1}\sum_{i=1}^{T-k} w_i)^2 = O(k^{-2d-2})$$

and for each l

$$W(l) = T^{-1}\sum_{i=1}^{T-k} w_i = O_p(k^{d-1}).$$

Then

$$\left\| T^{-1}\sum_{t=k}^{T} Z_t'u_t \right\| = (\Sigma W(l)^2)^{1/2} = O_p(k^{-d-1/2})$$

and

$$\left\|\Sigma^{-1}\right\| \left\| T^{-1}\sum_{t=k}^{T} Z_t'u_t \right\| = O_p(k^{-3d+3/2}).$$

Finally

$$T^{-2d}\sum \delta_{k+i}\sum_{t=k+i}^{T} y_{t-1}y_{t-k-i} \le T^{-2d}\sum \delta_{k+i}\sum_{t=k+i}^{T} y_{t-1}^2 = O_p(k^{-d})$$

and the statement of the Theorem obtains since $k^{-3d+3/2} < k^{-d}$ if $d \ge \frac{3}{4}$ and is $\ge k^{-d}$ for $\frac{1}{2} < d < \frac{3}{4}$.

Appendix C.

Proof of Theorem 3.

Define for some $1 < n < T$ the matrix $\hat{\Sigma}(-,k) = \hat{\Sigma}(k) - \hat{\gamma}_n \iota\iota'$ and the vector $\hat{\gamma}(-,k) = \hat{\gamma}(k) - \hat{\gamma}_n\iota$, where ι is a vector of ones, and consider $\hat{a}(-,k)$ that solves

$$\hat{\Sigma}(-,k)\hat{a}(-,k) = \hat{\gamma}(-,k).$$

Similarly defined matrices and vectors for the population quantities will be denoted by the same symbols but without the "hats". Also define $D_l = T^{1/2}[\hat{\gamma}_0 - \gamma_0 - \hat{\gamma}_l - \gamma_l]$. To prove Theorem 3 we prove the following Theorem that provides an explicit description of the asymptotics for the terms $\zeta(k) = \hat{a}(-,k) - a(-,k)$ and a bound for $\left\|a(-,k) - \delta_{[1,k]}\right\|$.

Theorem C. *Under the conditions of Theorem 3 as $T \to \infty$ for any fixed k, n the limiting distribution of $T^{\frac{1}{2}}(\hat{a}(-,k) - a(-,k))$ is multivariate normal with mean 0 and covariance matrix $W(k) = \Sigma(-,k)^{-1}AMA'\Sigma(-,k)^{-1}$. The $k \times (k+1)$ matrix A has elements*

$$\{A\}_{ij} = \begin{cases} a_{i-j} + a_{i+j} & \text{for } \ 1 \le i,j \le k, i \ne j, \\ a_{2i} - 1 & \text{for } i = j, \\ 1 - \Sigma a_i & \text{for } j = k+1, \end{cases} \tag{C.1}$$

where a_l are the components of the vector $a(-,k)$ for $1 \le l \le k$ and $a_0 = 1$, $a_l = 0$, if $l < 0$. The $(k+1) \times (k+1)$ matrix M is the limit covariance matrix of the vector G with the lth component $G_l = D_l$ for $l = 1, 2...k$ and $G_{k+1} = D_n$ for $l = k+1$; its elements are

$$\begin{aligned} \{M\}_{ij} &= limcov(D_m, D_l) \\ &= \tfrac{1}{2}\textstyle\sum_{s=-\infty}^{\infty}(\gamma_s - \gamma_{s-m} - \gamma_{s-l} + \gamma_{s-m+l})^2 + \kappa(\gamma_0 - \gamma_m)(\gamma_0 - \gamma_l) \end{aligned}$$

with $m = \begin{cases} n \text{ if } i = k+1 \\ i \text{ otherwise} \end{cases}$, $l = \begin{cases} n \text{ if } j = k+1 \\ j \text{ otherwise} \end{cases}$; κ *the kurtosis of* $\{\varepsilon_t\}$ *of* (2.1.1). *In addition, as* $k \to \infty, n \to \infty$

$$\left\|a(-,k) - \delta_{[1,k]}\right\| = \begin{cases} O(n^{2d-1}k^d) \text{ if } 0 \leq d < \frac{1}{2}; \\ O(n^{-1}k^{-d}) \text{ if } -\frac{1}{2} \leq d < 0, \end{cases}$$

and thus $\left\|a(-,k) - \delta_{[1,k]}\right\| \to 0$ *as* $k \to \infty$ *for* $-\frac{1}{2} \leq d < 0$, *and for* $0 \leq d < \frac{1}{2}$ *if* $n = O(k^\alpha)$ *for any* $\alpha > \frac{d}{1-2d}$.

Proof.

Consider

$$T^{\frac{1}{2}}\left(\hat{a}(-,k) - a(-,k)\right) = \hat{\Sigma}(-,k)^{-1}T^{\frac{1}{2}}[\hat{\gamma}(-,k) - \hat{\Sigma}(-,k)a(-,k)].$$

This equals

$$\begin{aligned} &\Sigma(-,k)^{-1}T^{\frac{1}{2}}[\hat{\gamma}(-,k) - \hat{\Sigma}(-,k)a(-,k)] \\ &+[\hat{\Sigma}(-,k)^{-1} - \Sigma(-,k)^{-1}]T^{\frac{1}{2}}[\hat{\gamma}(-,k) - \hat{\Sigma}(-,k)a(-,k)]. \end{aligned} \tag{C.2}$$

Examine the following factor in the terms in (C.2)

$$\begin{aligned} & T^{\frac{1}{2}}[\hat{\gamma}(-,k) - \hat{\Sigma}(-,k)a(-,k)] \\ = \; & T^{\frac{1}{2}}[\hat{\gamma}(-,k) - \gamma(-,k)] - T^{\frac{1}{2}}[\hat{\Sigma}(-,k) - \Sigma(-,k)]a(-,k). \end{aligned}$$

First,

$$T^{\frac{1}{2}}[\hat{\gamma}(-,k) - \gamma(-,k)] = (D_1 - D_n, D_2 - D_n, ...D_k - D_n)$$

where $D_l = T^{\frac{1}{2}}[\hat{\gamma}_0 - \gamma_0 - \hat{\gamma}_l - \gamma_l]$, as defined (in slightly different notation) by Hosking (1996). His Theorem 5 implies that as $T \to \infty$ for any fixed m the vector $D' = (D_1, D_2, ..., D_m)$ has as a limiting distribution a multivariate normal with mean zero and covariances given by

$$limcov(D_m, D_l) = \frac{1}{2}\sum_{s=-\infty}^{\infty}(\gamma_s - \gamma_{s-m} - \gamma_{s-l} + \gamma_{s-m+l})^2 + \kappa(\gamma_0 - \gamma_m)(\gamma_0 - \gamma_l)$$

where κ is the kurtosis of $\{\varepsilon_t\}$ of (2.1.1). This result holds for any $-\frac{1}{2} < d < \frac{1}{2}$ (Hosking, p.273) under the assumptions of Theorem C.

Next note that the element $T^{\frac{1}{2}}\{\hat{\Sigma}(-,k) - \Sigma(-,k)\}_{ij}$ is $D_n - D_{|i-j|}$. Consider the vector $\Delta(-)$ with elements $\{\Delta(-)\}_i = D_n - D_i$, the matrix $Q(-)$

with $\{Q(-)\}_{ij} = D_n - D_{|i-j|}$. Then the vector $\Delta(-) - Q(-)a(-,k)$ can be rewritten as AG, where the elements of A are given in (C.1) and the vector G is defined in the statement of Theorem C. Thus $T^{\frac{1}{2}}[\hat{\gamma}(-,k) - \hat{\Sigma}(-,k)a(-,k)]$ has a multivariate normal limit distribution with mean 0 and covariance matrix AVA'.

To deal with the second term in (C.2) examine $\left\|\hat{\Sigma}(-,k)^{-1} - \Sigma(-,k)^{-1}\right\|$. From Lemma 3 of Berk (1974)

$$\left\|\hat{\Sigma}(-,k)^{-1} - \Sigma(-,k)^{-1}\right\| \leq O_p(kT^{-\frac{1}{2}})$$

and the second term goes to zero in probability.

Thus for fixed k, $T^{\frac{1}{2}}[\hat{a}(-,k) - a(-,k)]$ has a limiting normal distribution with mean zero and covariance $\Sigma(-,k)^{-1}AVA\Sigma(-,k)^{-1}$.

Next to evaluate the norm of the difference $\|a(-,k) - \delta\|$ express

$$\Sigma(-,k)a(-,k) = (\Sigma(k) - \gamma_n \iota\iota')a(-,k) = \gamma(k) - \gamma_n\iota = \Sigma(k)a(k) - \gamma_n\iota.$$

We can write

$$a(-,k) = (\Sigma(k) - \gamma_n\iota\iota')^{-1}\Sigma(k)a(k) - (\Sigma(k) - \gamma_n\iota\iota')^{-1}\gamma_n\iota.$$

From the proof of Theorem 1 recall that $\|\Sigma(k)^{-1}\|$ is bounded if $d \geq 0$ and is $O(k^{-2d})$ if $d < 0$; we also know that $\gamma_n = O(n^{2d-1})$. Thus $\|\Sigma(k)^{-1}\gamma_n\iota\iota'\|$ is $O(n^{2d-1}k^{1/2})$ for $d \geq 0$ and $O(n^{-1}k^{1/2})$ for $d < 0$. Set n to grow faster than $max\{k^{\frac{1}{2(1-2d)}}, k^{1/2}\}$; then $\|\Sigma(k)^{-1}\gamma_n\iota\iota'\| = o(1)$ and we can expand

$$(I - \Sigma(k)^{-1}\gamma_n\iota\iota')^{-1} = I + \Sigma(k)^{-1}\gamma_n\iota\iota' + O(\left\|\Sigma(k)^{-1}\gamma_n\iota\iota'\right\|^2).$$

It follows that

$$a(-,k) - a(k) = \Sigma(k)^{-1}\gamma_n\iota\iota'a(k) - \Sigma(k)^{-1}\gamma_n\iota + o(\left\|\Sigma(k)^{-1}\gamma_n\iota\iota'\right\|);$$

$\|a(-,k) - a(k)\|$ is thus $O(k^{1/2}n^{2d-1})$ or $O(n^{-1}k^{1/2})$. Using the results in Theorem 1 we get that $\|a(-,k) - \delta\|$ is $O(k^d n^{2d-1})$ or $O(k^{-d}n^{-1})$ correspondingly. This proves Theorem C.

To conclude the proof of Theorem 3 we need to consider $\tilde{a}(-,k) - \tilde{a}(k)$; we have from the definitions of $\tilde{a}(-,k)$, $\tilde{a}(k)$

$$\tilde{a}(-,k) - \tilde{a}(k) = \hat{\Sigma}(k)^{-1}\hat{\gamma}_n\left(\iota\iota'\tilde{a}(-,k) - \iota\right). \tag{C.3}$$

Consider for $\tilde{T}$ selecting $k = O(ln\tilde{T})$, $n = \tilde{T}^v$ for some $1/2 > v > 0$. (The selection of n does not depend on d.) Then from e.g. Hosking (1996), Theorem 4 it follows that $\hat{\gamma}_n = O_p(n^{2d-1})$ and thus for large enough $\tilde{T}$ evalution of the right-hand side of (C.3) provides

$$\|\tilde{a}(-,k) - \tilde{a}(k)\| = O_p(\tilde{T}^{(2d-1)v} ln\tilde{T});$$

also from Theorem C

$$\|a(-,k) - \delta\| = o(\tilde{T}^{(2d-1)v} ln\tilde{T});$$

thus $\|\tilde{a}(-,k) - \tilde{a}(k)\| + \|a(-,k) - \delta\|$ can be made smaller than ε in probability by an appropriate choice of $\tilde{T}$; for the corresponding k define $\eta(k) = \tilde{a}(k) - \tilde{a}(-,k) + a(-,k) - \delta$; define $\zeta(k) = \tilde{a}(-,k) - a(-,k)$. Theorem C provides the limit normal distribution for $\zeta(k)$ and $\|\eta(k)\| < \varepsilon$ in probability. This concludes the proof of Theorem 3.

Proof of Theorem 4.

Under the assumptions of Theorems 1 and 2 and the assumption on Ω_k and for $k, T \to \infty$ so that $\tilde{a}(k) \to_p \delta(\omega_0)_{[1,k]}$, Q_T satisfies the conditions of Theorem 4.1.1 (Amemiya, 1985). Indeed, recall that the parameter vector ω is defined on a bounded set corresponding to the stationarity constraints under Theorem 1 and with $1/2 < d < 1$ under Theorem 2; we can consider the closure of this set. The function $Q_T(\tilde{\alpha}, \omega)$ converges in probability uniformly over all ω to a nonstochastic function $Q(\delta(\omega_0), \omega)$ since $\|\tilde{a} - \delta(\omega_0)\| = o_p(k^{-\alpha})$ with $\alpha > 0$ defined for the different processes in Theorem 1 and Theorem 2. For an ARFIMA process with the coefficients of the infinite autoregression given in 2.1 it is easy to see that $minQ(\delta(\omega_0), \omega) = 0$ if $\omega = \omega_0$ and is positive otherwise. Since $minQ_T(\tilde{\alpha}, \omega) \to_p minQ(\delta(\omega_0), \omega)$ the theorem follows.

Proof of Theorem 5.

Consider the first-order condition for the minimization of $Q_{T,k}(\tilde{a}, \omega)$:

$$-2\frac{\partial \delta(\omega)'}{\partial \omega}\Omega_k(\tilde{a} - \delta(\omega)) = 0.$$

Note that this condition holds at $\tilde{a} = \delta(\omega_0)$, $\omega = \omega_0$. Suppose that some $\tilde{a}$, $\tilde{\omega}$ solve the first-order condition. From Theorem 4 it follows that for

any neighbourhood $U(\delta(\omega_0), \omega_0)$ there exist large enough k, T such that $\Pr((\tilde{a}, \tilde{w}) \in U(\delta(\omega_0), \omega_0))$ is close to 1.

Then by the Implicit Function Theorem in the neighbourhood $U(\delta(\omega_0), \omega_0)$ of $(\delta(\omega_0), \omega_0)$ we have

$$\tilde{\omega}_\delta - \omega_0 = \frac{\partial \omega(\omega_0)}{\partial \tilde{a}}(\tilde{a} - \delta(\omega_0)) + o_p(\tilde{a} - \delta(\omega_0))$$

with

$$\frac{\partial \omega(\omega_0)}{\partial \tilde{a}} = \left[-2\frac{\partial^2 \delta'(\omega_0)}{\partial \omega \partial \omega'}\Omega(\tilde{a} - \delta(\omega_0)) + 2\frac{\partial \delta'(\omega_0)}{\partial \omega}\Omega\frac{\partial \delta(\omega_0)}{\partial \omega'}\right]^{-1}\left[-2\frac{\partial \delta'(\omega_0)}{\partial \omega}\Omega\right].$$

Substituting $\tilde{a}(k) - \delta(\omega) = \zeta(k) + \eta(k)$ from Theorem 3 we get that

$$\tilde{\omega}_\delta - \omega_0 = \xi(k) + b(k),$$

where $b = \left[\frac{\partial \delta'(\omega_0)}{\partial \omega}\Pi_k\frac{\partial \delta(\omega_0)}{\partial \omega'}\right]^{-1}\left[\frac{\partial \delta'(\omega_0)}{\partial \omega}\Pi_k\right]\eta(k)$ and $\xi(k)$ has a limiting normal distribution as $T \to \infty$ with the asymptotic covariance matrix

$$V(k) = \left(\frac{\partial \delta'(\omega_0)}{\partial \omega}\Omega\frac{\partial \delta(\omega_0)}{\partial \omega'}\right)^{-1}\frac{\partial \delta'(\omega_0)}{\partial \omega}\Omega W(k)\Omega\frac{\partial \delta(\omega_0)}{\partial \omega'}\left(\frac{\partial \delta'(\omega_0)}{\partial \omega}\Omega\frac{\partial \delta(\omega_0)}{\partial \omega'}\right)^{-1}. \tag{C.4}$$

Figure 1a:
Empirical density of first two AR coefficients vs. Normals
T=100, d=0.4, k=8

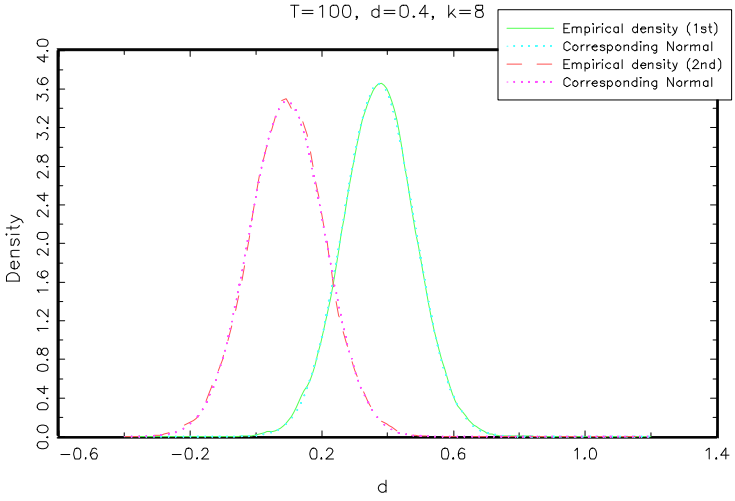


Figure 1b:
Empirical density of first two AR coefficients vs. Normals
T=100, d=0.5, k=8

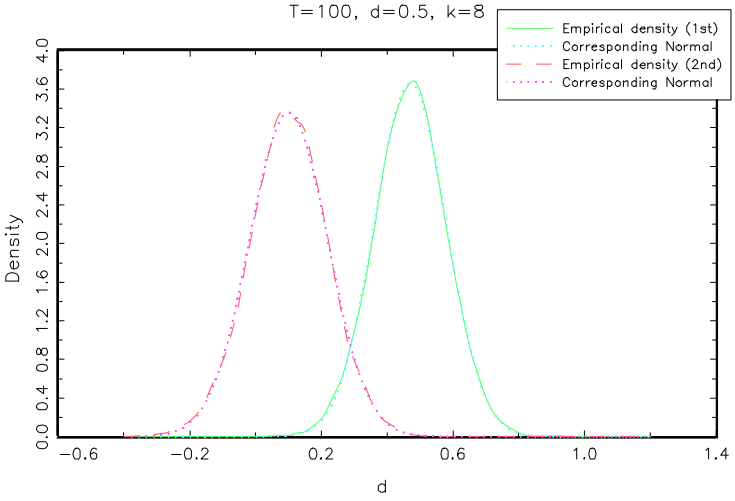


Figure 1c:
Empirical density of first two AR coefficients vs. Normals
T=100, d=0.7, k=8

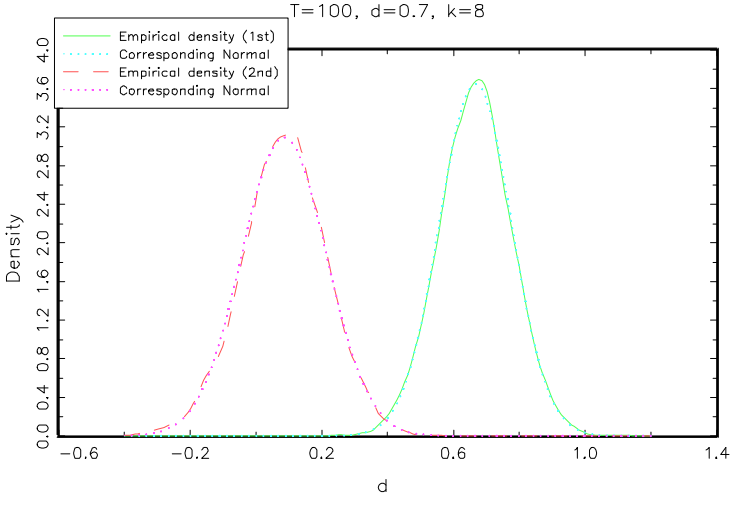


Figure 1d:
Empirical density of first two AR coefficients vs. Normals
T=100, d=0.9, k=8

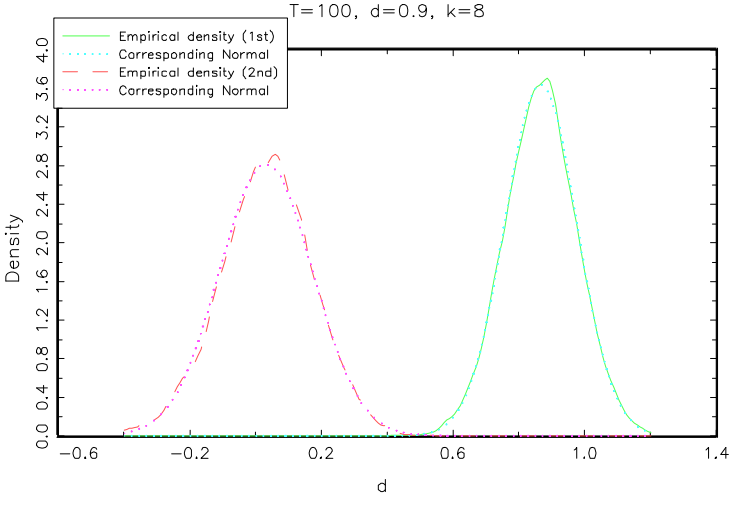


Figure 1e:
Empirical density of first two AR coefficients vs. Normals
T=400, d=0.4, k=12

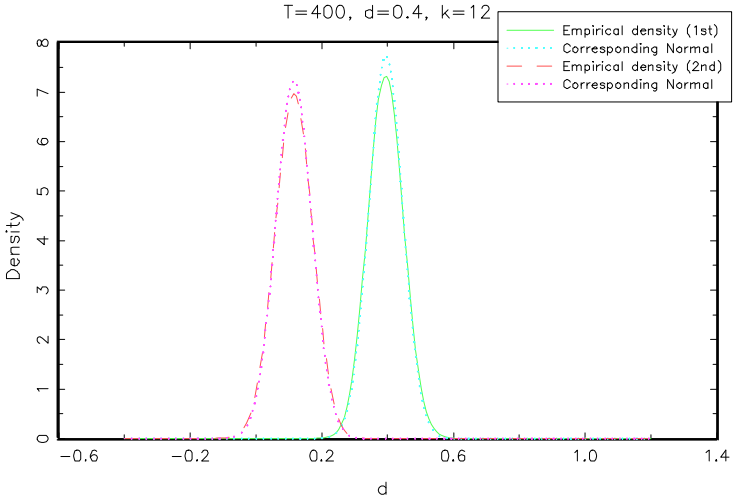


Figure 1f:
Empirical density of first two AR coefficients vs. Normals
T=400, d=0.5, k=12

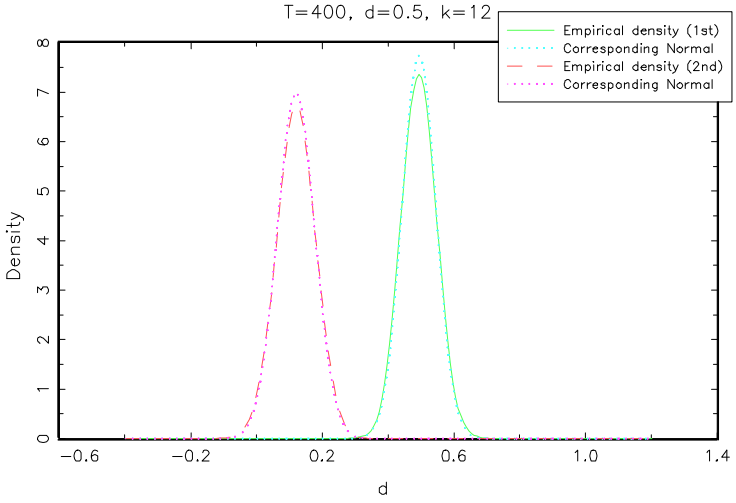


Figure 1g:
Empirical density of first two AR coefficients vs. Normals
T=400, d=0.7, k=12

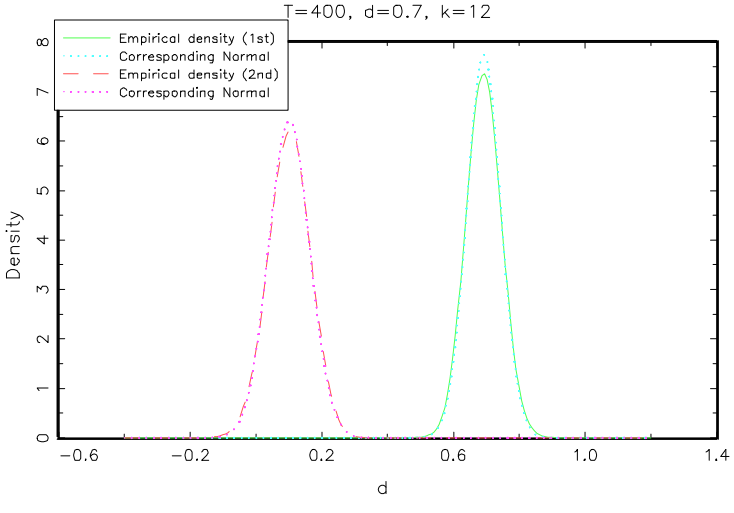


Figure 1h:
Empirical density of first two AR coefficients vs. Normals
T=400, d=0.9, k=12

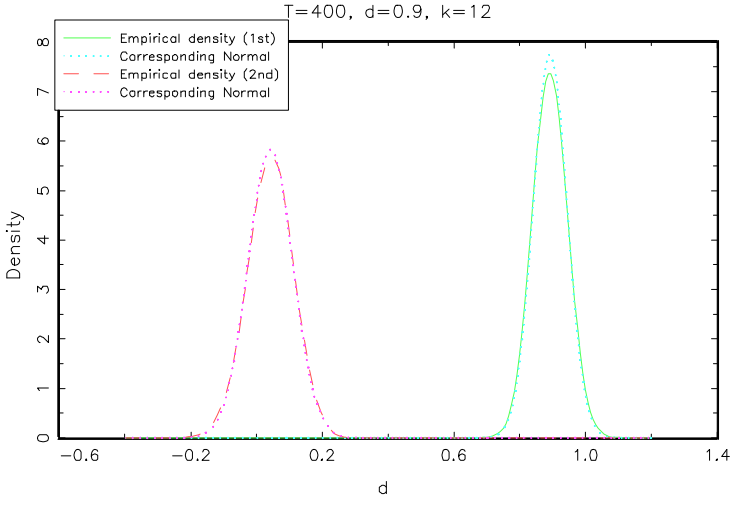


Figure 2a
Root mean squared errors of estimates of d, ARFIMA(0,d,0)
T=100

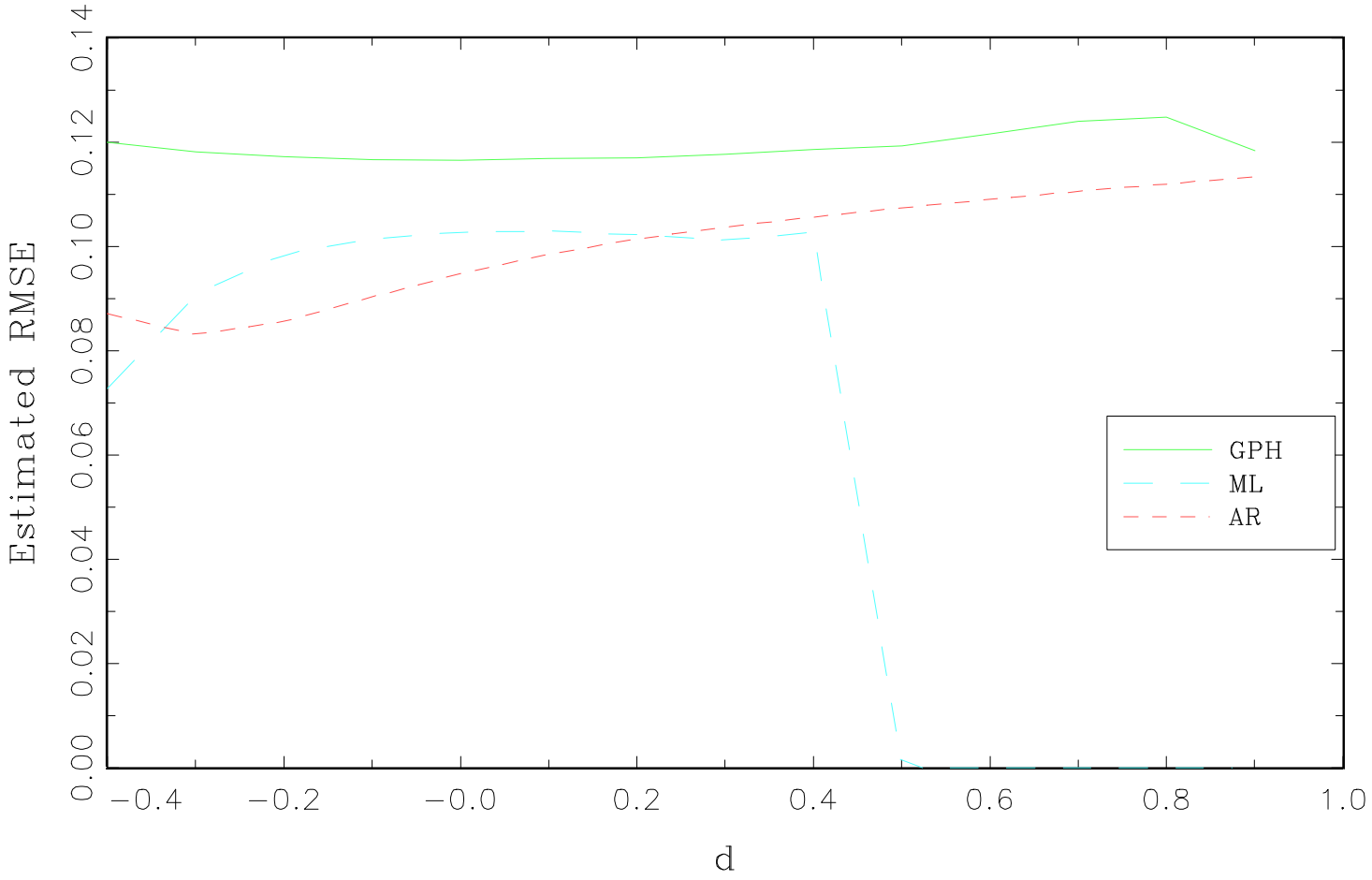
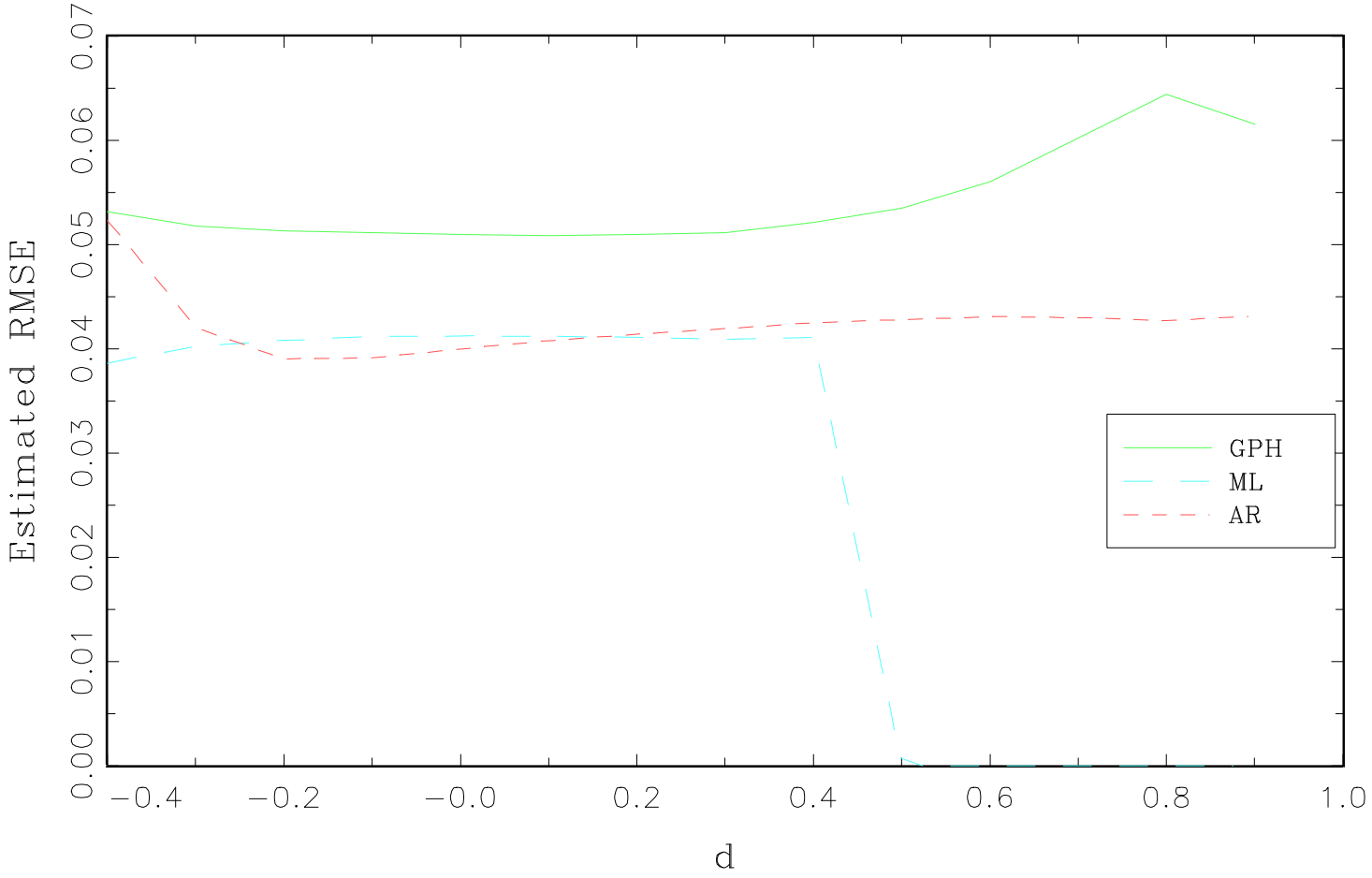


Figure 2b
Root mean squared errors of estimates of d, ARFIMA(0,d,0)
T=400



Liste des publications au CIRANO *

Cahiers CIRANO / *CIRANO Papers* (ISSN 1198-8169)

99c-1 Les Expos, l'OSM, les universités, les hôpitaux : Le coût d'un déficit de 400 000 emplois au Québec — Expos, Montréal Symphony Orchestra, Universities, Hospitals: The Cost of a 400,000-Job Shortfall in Québec / Marcel Boyer

96c-1 Peut-on créer des emplois en réglementant le temps de travail? / Robert Lacroix

95c-2 Anomalies de marché et sélection des titres au Canada / Richard Guay, Jean-François L'Her et Jean-Marc Suret

95c-1 La réglementation incitative / Marcel Boyer

94c-3 L'importance relative des gouvernements : causes, conséquences et organisations alternative / Claude Montmarquette

94c-2 Commercial Bankruptcy and Financial Reorganization in Canada / Jocelyn Martel

94c-1 Faire ou faire faire : La perspective de l'économie des organisations / Michel Patry

Série Scientifique / *Scientific Series* (ISSN 1198-8177)

2001s-10 Heterogeneous Returns to Human Capital and Dynamic Self-Selection / Christian Belzil et Jörgen Hansen

2001s-09 Return to a High School Diploma and the Decision to Drop Out: New Evidence from Canada / Daniel Parent

2001s-08 Leader and Follower: A Differential Game Model / Hassan Benchekroun et Ngo Van Long

2001s-07 Emission Taxes and Standards for an Asymmetric Oligopoly / Ngo Van Long et Antoine Soubeyran

2001s-06 Risque de modèle de volatilité / Ali Alami et Éric Renault

2001s-05 The Effect of Pay-for-Performance Contracts on Wages /Daniel Parent

2001s-04 Incentive Pay in the United States: Its Determinants and Its Effects / Daniel Parent

2001s-03 Testing and Comparing Value-at-Risk Measures / Peter Christoffersen, Jinyong Hahn et Atsushi Inoue

2001s-02 Empirical Assessment of an Intertemporal Option Pricing Model with Latent Variables / René Garcia, Richard Luger et Éric Renault

2001s-01 Asymmetric Smiles, Leverage Effects and Structural Parameters / René Garcia, Richard Luger et Éric Renault

2000s-60 Technological Paradigms and the Measurement of Innovation / Bernard Sinclair-Desgagné

2000s-59 Taxpayers' Response to Tax Rate Changes: A Canadian Panel Study / Robert Gagné, Jean-François Nadeau et François Vaillancourt

---

* Vous pouvez consulter la liste complète des publications du CIRANO et les publications elles-mêmes sur notre site Internet à l'adresse suivante :

http://www.cirano.umontreal.ca/publication/documents.html