Inference and Forecasting for Fractional Autoregressive Integrated Moving Average Models, with an application to US and UK inflation

Marius Ooms* and Jurgen Doornik**

November 29, 1999

Econometric Institute Report 9947/A

Econometric Institute
Erasmus University Rotterdam
P.O. Box 1738
NL- 3000 DR Rotterdam
The Netherlands
email: ooms@few.eur.nl

Abstract

We discuss computational aspects of likelihood-based specification, estimation, inference, and forecasting of possibly nonstationary series with long memory. We use the ARFIMA(p, d, q) model with deterministic regressors and we compare sampling characteristics of approximate and exact first-order asymptotic methods. We extend the analysis using a higher-order asymptotic method, suggested by Cox and Reid (1987). Efficient computation and simulation allow us to apply parametric bootstrap inference as well. We investigate the relevance of the differences between the methods for the time-series analysis of monthly core consumer price inflation in the US and quarterly overall consumer price inflation in the UK. We concentrate on (stationarity) tests for the order of integration and on inference for out-of-sample forecasts of the price level.

* Econometric Institute, Erasmus University Rotterdam, The Netherlands.
** Nuffield College, New Road, Oxford OX1 1NF, UK.

1 Introduction

In this paper we discuss specification, estimation, simulation, and forecasting of, possibly nonstationary, time series with long memory. Our approach is essentially parametric, using the well-known Gaussian fractional ARIMA model. We concentrate on the computational efficiency and the sampling characteristics of approximate and exact maximum likelihood methods in settings which are relevant to the empirical analysis of inflation series. Finally, we perform a sensitivity analysis using the same model with GARCH errors.

1.1 Modelling

The ARFIMA(p, d, q) model allows one to model the long-run behaviour of the series in a flexible way. The empirical modelling process consists of three stages: identification, estimation and testing. The first stage is identification to determine the integer part of the order of differencing, d, and the orders of the autoregressive and the moving-average parts of the model, p, and q. The ARFIMA model enables interesting and easy-to-use tests for the null of short-memory stationarity ($d = 0$), as well as for the null of unit-root nonstationarity ($d = 1$). These tests complement more widely used KPSS and Dickey–Fuller type tests.

1.2 Estimation and computation

We consider three estimation methods of the candidate ARFIMA model. We compare nonlinear least squares (NLS) as in Beran (1995) with exact maximum likelihood methods of estimation (EML). EML, as proposed by Sowell (1992), has often been discarded as computationally unattractive, but the implementation by Doornik and Ooms (1999) has reduced requirements with regard to computer memory and computing power considerably. Furthermore, their package allows explanatory variables, enabling efficient testing for $d = 0$ and $d = 1$ conditional on the presence of trends and other permanent changes in mean. The simple EML estimator of the parameters of interest, d, can be severely biased in the presence of unknown nuisance parameters for regressor variables, even if there is only a constant to measure the unknown mean. Recently, higher order asymptotic methods have been shown to improve inference in the presence of those nuisance parameters, although applications in time series analysis have been scarce. The modified profile likelihood method (MPL), suggested by Cox and Reid (1987), is the third estimation method we consider. It works well, and is not difficult to implement given the algorithms for EML in the ARFIMA model.

The core of the EML method is the computation of the autocovariances as a function of the parameters of a stationary ARFIMA model. Following a comparison of several methods to compute the autocovariances, we discuss the advantages and disadvantages of the different estimation methods.

1.3 Forecasting

Next, we compute the optimal forecasts and corresponding confidence intervals. The treatment of pre-sample values can play an important role in forecasting with long-memory models. The optimal methods we employ are explicitly based on finite sample information sets; we also use 'naive' methods which ignore this issue.

1.4 Simulation

We compare the different estimators and forecasting methods using a parametric bootstrap analysis. Additional programming efforts required for such experiments are limited. The 'pseudo-sample generator' of this class uses the same exact autocovariance function of an ARFIMA model as required in EML estimation. We compare reliability of estimated confidence intervals for parameter estimates and forecasts across the different methods.

1.5 Estimation, inference and forecasts for inflation series

An important application of long-memory models is in forecasting consumer price inflation for developed economies. Inflation series display long non-periodic cycles typical of long-memory time series. This leads to significantly positive estimates of the order of integration, d, which is often the primary parameter of interest: it determines the rate of growth of out-of-sample forecast intervals of the log of the corresponding price level. This is of crucial importance for studies of long-run variation of e.g. indexed outlays.

We apply the different methods to inflation in consumer prices for the US and the UK. The parametric bootstrap analysis shows that inference is quite reliable in finite samples and that Gaussian inference based on MPL is clearly preferrable. Both US and UK inflation are long memory; US core inflation seems covariance stationary, but UK inflation does not.

We examine robustness by extending the model with conditional heteroskedasticity of the GARCH type. In this case, asymptotic approximations turn out to be less accurate.

2 Modelling

From the early sixties onwards, when Mandelbrot suggested the appropriateness of long-memory models for economic time series, there has been a steady growth in the literature on the subject. Robinson (1994) and Baillie (1996) provided useful surveys of the developments in the econometric modelling of long memory; Beran (1992) reviewed developments in long-memory modelling in other areas. Beran's monograph, Beran (1994), discusses most of the central issues, including forecasting.

The ARFIMA(p, d, q) model is used here for the statistical analysis of a univariate time series y_t with long memory. We write it as

$$\Phi(L)(1-L)^d(y_t-\mu_t)=\Theta(L)\varepsilon_t, \quad t=1,\ldots,T. \tag{1}$$

where $\Phi(L)=(1-\phi_1 L-\ldots-\phi_p L^p)$ is the autoregressive polynomial and $\Theta(L)=(1+\theta_1 L+\ldots+\theta_q L^q)$ is the moving average polynomial in the lag operator L; p and q are integers, d is real. $(1-L)^d$ is the fractional difference operator defined by the following binomial expansion:

$$(1-L)^d=\sum_{j=0}^{\infty}\delta_j L^j=\sum_{j=0}^{\infty}\left(\begin{array}{c}d\\j\end{array}\right)(-L)^j.$$

We assume $\varepsilon_t \sim \mathsf{NID}(0,\sigma_\varepsilon^2)$, and write μ_t for the mean of y_t. The ARMA-part of the model is invertible and stationary: all roots of $\Theta(z)=0$ and $\Phi(z)=0$ are outside the unit circle. In addition, $\Theta(z)=0$ and $\Phi(z)=0$ do not have common roots. We say that

$$z_t=y_t-\mu_t$$

is $I(d)$, integrated of order d. The zero mean ARFIMA(p, d, q) process z_t is covariance stationary if $d < 0.5$, see Hosking (1981). The autocovariance function γ_k of an ARFIMA(p, d, q) process decays hyperbolically: $\gamma_k \sim ck^{2d-1}, k \to \infty$,where c denotes a finite nonzero constant. We discuss computation of the autocovariance function in §3.3. The process is long memory in the case $0 < d < 0.5$. The spectral density $f_z(\omega)$ near zero is hyperbolic: $\lim_{\omega \to 0} f_z(\omega)\omega^{2d}$ exists and is finite. For $-0.5 < d < 0$ the process is called intermediate memory or 'overdifferenced', see Brockwell and Davis (1993) and e.g. Chang and Dickey (1994). In that case the inverse autocorrelations decay hyperbolically. We assume $d > -1$, which makes the process z_t invertible, see Odaki (1993).

Odaki used the following condition for invertibility: convergence of the mean squared error of the AR(∞)-based one-step-ahead prediction, $\mathsf{MSE}(\widehat{z}_{t|T})$, to the innovation variance σ_ε^2 as $T \to \infty$. The AR(∞) representation of z_t is defined as

$$z_t = \sum_{j=1}^{\infty} \pi_j z_{t-j} + \varepsilon_t. \tag{2}$$

In obvious notation: $\Pi(L) = \sum_{j=0}^{\infty} \pi_j L^j = \Theta(L)^{-1}\Phi(L)(1-L)^d$ and $\pi_0 = 1$. Note that $\Pi(1) = 0$ for $d > 0$: there is an AR unit root. Pre-sample values, i.e. z_j for $j < 0$, are set to zero in forecasting. We call the corresponding predictions 'naive' forecasts: these predictions are optimal if the observations are known into the infinite past. We call the corresponding one-step-ahead forecast errors naive residuals, denoted by $\tilde{e}_t$. The coefficients of $(1-L)^d$, δ_j, are easily computed: $\delta_0 = 1$ and

$$\delta_j = \prod_{0<k\leq j} \frac{k-1-d}{k}, \quad j = 1, 2, \ldots$$

The restriction $d > -1$ is imposed in EML and MPL estimation. Forecasting is discussed in more detail in §4.

Although z_t is invertible for the case $-1 < d \leq -0.5$, the MA representation of z_t:

$$z_t = 1 + \sum_{j=1}^{\infty} \psi_j \varepsilon_t = \Psi(L)\varepsilon_t = \Phi(L)^{-1}(1-L)^{-d}\Theta(L)\varepsilon_t \tag{3}$$

still has an MA-unit root: $\Psi(1) = 0$. The region $-1 < d \leq -0.5$ is excluded in NLS estimation, see also §3.5 below.

We see that $\psi_j \to 0, j \to \infty$ for $d < 1$. The process is therefore mean-reverting in this case: innovations ε_t only have a transitory effect on the time series process. In fact: $\psi_k \sim ck^{d-1}$ for $-0.5 < d < 1$.

2.1 Empirical analysis of $I(d)$ processes and unit root tests

In empirical work it is useful to inspect major sample characteristics of the data before employing ARFIMA models. In particular, it is essential to inspect time-series plots of the data, (inverse) sample autocorrelation functions, sample spectra and variance plots.

Suppose first that y_t is not trending so that constancy of $\mu_t = \mu$ is a reasonable assumption. A time-series plot does not reveal a clear trend, but mean-reversion is slow if $0 < d < 0.5$. The slow convergence of γ_k of long-memory processes with $d > 0$ is reflected in the sample autocorrelation function (SACF), which dies off slowly if $0 < d < 0.5$. For the nonstationary cases ($d > 0.5$), one can even expect a linear decline, see e.g. Hassler (1997) for a recent study. The frequency-domain representation of the sample autocorrelation function is the sample spectral density function (SDF). The (log) SDF has a high power near zero if $d > 0$, indicating the pole in the population density.

One can also look at the variance of (sub)sample means $\overline{y}_T = T^{-1}\sum_{t=1}^{T} y_t$ as a function of the (sub)sample size T. It holds that $\mathsf{var}(\overline{y}_T) \sim cT^{2d-1}$ for $T \to \infty$. A plot of $T\mathsf{var}(\overline{y}_T)$ versus T is known as the variance-time function; $\log(\mathsf{var}(\overline{y}_T))$ versus $\log(T)$ is a variance plot. The variance-time function converges to the so-called long-run variance of y_t: $\sigma^2 = \{\Phi(1)\}^{-2}\Theta(1)^2\sigma_\varepsilon^2$ only if d=0; it diverges for $d > 0$. The variance plot has a slope of -1 for $d = 0$. The slope is less negative for $d > 0$.

The slow convergence of the sample mean to the population mean for $d > 0$ has a number of consequences. First, the mean is hard to estimate with satisfactory precision, even when the process is stationary and the sample size is large. Second, sample autocorrelations are biased towards zero, see e.g. Agiakloglou and Newbold (1994). Third, in contrast to the ARMA(p, q) case, estimation of the mean is important for efficient estimation of the other parameters in the model if the sample size is not extremely large, see e.g. Smith Jr, Sowell, and Zin (1997). However, recent advances in higher order asymptotic theory seem to provide an adequate and computationally efficient solution for this problem, see §3 below. Fourth, inference on long-run forecasts of the mean is nontrivial, c.f §6.

Suppose next that y_t is trending, so that a linear trend in mean might be a reasonable hypothesis: $\mu_t = \beta_0 + \beta_1 t$. In this case one can look at the sample autocorrelations, spectra and variance-time functions in deviation from the OLS estimate of μ_t: $\widehat{z}_t = y_t - \widehat{\mu}_t$. The OLS estimator of $\beta = (\beta_0, \beta_1)'$ is not efficient if $0 < d < 0.5$, see Dahlhaus (1995), but $\widehat{\beta}_1$ still converges rather quickly. In the analysis of economic time series, attention is often focused on the $I(1)$ case, and it may be interesting to test this against $I(d)$, for $0 < d < 1$ or $1 < d < 2$. This also requires examination of the sample autocorrelations, log spectrum and variance-time functions of the first differences. If $d < 1$, then $(1 - L)y_t$ is overdifferenced: the log SDF has a deep trough near zero, and the sample inverse ACF dies off slowly.

Note that standard unit root tests, like the Augmented Dickey–Fuller (ADF) test and the Phillips–Perron test, are not useful to test the $I(1)$ hypothesis against $I(d)$ for $d < 1$, see Hassler and Wolters (1994). The ADF-test examines whether $\Pi(1) = 0$ in (2), which is true for all $d \in (0, 1]$. Robust AR unit root tests like the Philips–Perron test require an estimate of the long-run variance σ^2, to make the null distribution free of nuisance parameters, see e.g. Stock (1994). However, σ^2 does not exist for $d > 0$. The null distributions of these unit root test are not robust with respect to (stationary) long memory.

Well known stationary tests for the $I(0)$ hypothesis, like the KPSS-test, do have power against fractional alternatives, see Lee and Schmidt (1996). These tests check the convergence of the variance-time function. They compare the variance of $T^{-1}\sum_{i=1}^{T} y_i$ with the variance of y_t, and reject against $I(d)$, $d > 0$, if the variance of the cumulative sum is too large. The size of these tests is difficult to control in the presence of additional AR or MA-type short-memory correlation in finite samples, see e.g. Hobijn, Franses, and Ooms (1998).

In the context of ARFIMA modelling, testing for a fixed order of $I(d)$ using a 'full information' parametric analysis is straightforward. Null-limiting distributions are chi-squared and do not depend on the presence or absence of (breaking) trends in μ_t, either in the DGP or in the testing model. That is in sharp contrast with standard AR unit-root tests and stationarity tests. Gil-Alaña and Robinson (1997) presented an overview of methods to test $I(d)$

hypotheses in econometrics. They employed semiparametric LM-tests for fractional hypotheses, which are locally asymptotically as efficient as the full information Wald-tests that we use below.

Finally, it seems easier to correct size distortions of the likelihood-based fractional tests for d in finite samples. The application of higher-order parametric likelihood-based inference leads to significant improvements for tests for d in the ARFIMA model with a constant and/or trend as we illustrate below. This improvement has not yet been established for the standard unit root tests, c.f. Hansen and Rahbek (1998).

3 Estimation and Computation

3.1 Exact and profile likelihood

Based on the normality assumption and with a procedure to compute the autocovariances in the $T \times T$ covariance matrix $\Sigma = \sigma_\varepsilon^2 \mathbf{R}$ of a $T \times 1$ vector of observations $\mathbf{y}$, the log-likelihood for the ARFIMA(p, d, q) model (1) with regressors is

$$\log L\left(d, \phi, \theta, \beta, \sigma_\varepsilon^2\right) = -\frac{T}{2}\log\left(2\pi\right) - \frac{T}{2}\log\sigma_\varepsilon^2 - \frac{1}{2}\log|\mathbf{R}| - \frac{1}{2\sigma_\varepsilon^2}\mathbf{z}'\mathbf{R}^{-1}\mathbf{z}, \tag{4}$$

where $\mathbf{z} = \mathbf{y} - \mathbf{X}\beta$ is a $T \times 1$ vector in an obvious notation and β is a k-vector. For example: $\mu_t = x_t'\beta = \beta_0 + \beta_1 t + \beta_2 t^2 + ... + \beta_{k-1}t^{k-1}$. When σ_ε^2 and β are concentrated out, the resulting normal profile likelihood function becomes:

$$\log L_P\left(d, \phi, \theta\right) = c - \frac{1}{2}\log|\mathbf{R}| - \frac{T}{2}\log\widehat{\mathbf{z}}'\mathbf{R}^{-1}\widehat{\mathbf{z}}, \tag{5}$$

where $\widehat{\mathbf{z}} = \mathbf{y} - \mathbf{X}\widehat{\beta}, \quad \widehat{\beta} = (\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}$.

There are two complicating factors when evaluating the likelihood: evaluation of the autocorrelation function to construct $\mathbf{R}$, and inverting this matrix of dimension $T \times T$. Initially, efforts to avoid these complications led to a range of approximate procedures. Indeed, it is still regularly claimed that exact maximum likelihood is prohibitively complex and slow, cf. Bollerslev and Jubinski (1999), Martin and Wilkins (1999), Ravishanker and Ray (1997). One of our purposes is to illustrate that EML is sufficiently fast to be used in larger samples or bootstrap methods.

3.2 Modified profile likelihood and related methods

The modified profile likelihood is a concept from higher order asymptotic theory, see Cox and Reid (1987). The aim of this theory is to develop more accurate inference on parameters of interest in the presence of (a large number) of nuisance parameters, see Barndorff-Nielsen and Cox (1994, Chapter 4) for a motivation and examples. This theory is well developed for the normal linear regression model with serial correlation of known functional form.

In the ARFIMA model, the fractional parameter d is often the parameter of interest. For example, when forecasting partial sums of the process z_t with known parameters, the growth of the variance of partial sums is proportional to H^{2d+1}, where H is the forecast horizon, cf. Beran (1994, §8.5).

A good estimate of d is also required in the estimation of the variance of the sample mean, especially for inference about the mean of y_t. In most cases, the unknown mean

μ_t is a function of nuisance parameters, whose presence has an adverse effect on the finite sample behaviour of the standard maximum likelihood estimator of d. When μ_t is estimated, either by simple regression, or jointly with d by maximizing the profile likelihood, $\widehat{d}$ can be severely biased, even in the simplest ARFIMA$(0, d, 0)$ model. Smith Jr, Sowell, and Zin (1997) suggested to overdifference the data to remove the constant term, and to directly estimate $d-1$ essentially without bias. However, this procedure is only effective when there are no additional regressors.

An and Bloomfield (1993) derived the modified profile likelihood, $\log L_M$, for the regression model with stationary ARFIMA-errors:

$$\log L_M\left(d, \phi, \theta\right) = c + \left(\frac{1}{T} - \frac{1}{2}\right) \log |\mathbf{R}| - \frac{1}{2} \log \left|\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right| - \left(\frac{T-k-2}{2}\right) \log \widehat{\mathbf{z}}'\mathbf{R}^{-1}\widehat{\mathbf{z}}. \quad (6)$$

They used higher order asymptotic theory to show that the expectation of the score of (6) for (d, ϕ, θ) at the true parameter values is $O(T^{-1})$, whereas this expectation is $O(1)$ for the score of (5). This higher order bias correction provides the main argument for the better behaviour of the maximum MPL estimator over the EML estimator. The EML and MPL estimators require Σ and its inverse to exist, and therefore require $d < 0.5$. The assumption of normality plays an important role in the derivation of the optimality of both estimators.

The Monte Carlo experiments of An and Bloomfield (1993), and the results in Hauser (1999), show that, for simple ARFIMA$(1, d, 1)$ models, MPL reduces bias for $\widehat{d}$, and that it leads to more accurate inference in finite samples. Hauser (1999) used an OLS-estimate of β in the MPL estimator in order to reduce the number of computations: $\widehat{\mathbf{z}} = \mathbf{y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

There are several related higher order asymptotic methods that lead to similar modifications of the likelihood analysis for the Gaussian linear regression model with correlated disturbances. Laskar and King (1998) provided an overview including the marginal likelihood following Tunnicliffe-Wilson (1989) and the approximate conditional profile likelihood following Cox and Reid (1993). The differences with the modified profile likelihood of Cox and Reid (1987) are minor.

3.3 Computing the autocovariances

Various techniques are available to evaluate the autocovariance function of a stationary ARFIMA process. A conceptually very simple procedure is to compute the autocovariances $\gamma_k = E(z_t z_{t-k})$, $k = \ldots, -1, 0, 1, \ldots$ from the MA representation (3), with $\psi_0 = 1$:

$$\gamma_k = \sum_{j=0}^{\infty} \psi_j^z \psi_{j+|k|}^z \sigma_\varepsilon^2. \quad (7)$$

Chan and Palma (1998) used this approach. The drawback is that, because ψ_j^z declines hyperbolically, many terms are needed for an accurate approximation.

Hosking (1981) expressed γ_k as a function of the autocovariances of the corresponding ARFIMA$(0, d, 0)$ process $w_t = \{\Theta(L)\}^{-1}\Phi(L)z_t$ and the ARMA(p, q) process $u_t = (1-L)^d z_t$:

$$\gamma_k = \sum_{l=-\infty}^{\infty} \gamma_j^u \gamma_{k-j}^w. \quad (8)$$

Both γ_j^w and γ_j^u are easily calculated using simple recursions, see e.g. Brockwell and Davis (1993, §3.3 and §13.2). Fast decay in the γ_j^u will help convergence, but when $\Phi(z)$ has roots close to the unit circle, it is still necessary to use many terms for accurate computation of γ_k. For pure ARFIMA$(0, d, q)$ processes, only $2q + 1$ terms in (8) are required, which is preferable to using the MA representation (7).

Ravishanker and Ray (1997) use:

$$\gamma_k = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \psi_i^u \gamma_{k+i-j}^w \psi_j^u \sigma_\varepsilon^2. \tag{9}$$

which alleviates the slow convergence of (7) due to a large d.

A seemingly simple alternative is to numerically integrate over the spectrum, see e.g. Harvey (1993, p. 229):

$$\gamma_k = \int_{-\pi}^{\pi} f_z(\omega) e^{i\omega k} d\omega, \tag{10}$$

where the spectrum of the ARFIMA-process, $f_z(\omega)$, is easily computed. However, numerical integration for each k does rapidly get prohibitively slow.

More specific methods improve speed and accuracy of the computation of γ_k: Hosking (1981) provided a closed form alternative to (8) for the ARFIMA$(1, d, 0)$ case. Sowell (1992) succeeded in extending and numerically implementing the results of Hosking for the ARFIMA(p, d, q) case, using the assumption of the unicity of the roots ρ_j^{-1}, $j = 1, \ldots, p$, of the AR polynomial $\Phi(z) = 0$. Each γ_k requires the computation of at least p hypergeometric function values $F(a, 1; c; \rho_j)$, where a and c are a function of d and j:

$$F(a, 1, c, \rho) = \sum_{i=1}^{\infty} \frac{a(a+1)\cdots(a+i-1)(i-1)!\rho^i}{c(c+1)\cdots(c+i-1)i!}.$$

This is a slowly converging series for $|\rho_i| \to 1$. Sowell (1992) achieved a major speed-up of the algorithm by evaluating the hypergeometric functions recursively: only the hypergeometric functions for γ_T have to be computed separately, with the remaining $\gamma_{T-1}, \ldots$ derived from the recursion. It is interesting that, while the backward recursion is numerically stable, the analogue forward version is highly unstable. In addition to the unicity of AR roots, Sowell's algorithm further requires that no roots are zero, although this was not mentioned in Sowell (1992). Doornik and Ooms (1999) adjusted the algorithm to make it applicable for $\rho_j = 0$, and to make it numerically stable for $\rho_j \to 0$.

Table 1 compares the methods for an ARFIMA$(1, d = 0.45, 1)$ model with $\phi = 0.8$ and $\theta = -0.5$. This is an example where the autocorrelations exhibit slow decay. First, we look at the accuracy of the approximations. All methods involve infinite sums, which must be terminated at some point. The first row in the table lists the 31th autocovariance, standardized with respect to the variance from Sowell's method: γ_{31}/γ_0^S. The number in parentheses for methods (7) and (8) gives the number of terms used in the summation. The numerical integration for (10) was implemented using QuadPack function QAGS, see Piessens, de Donker-Kapenga, Überhuber, and Kahaner (1983); QAGS required more than 200 function evaluations to attain the reported precision. The reported timings are for computing 1024 autocovariances, and are relative to Sowell's method.

Table 1 shows that neither the MA representation nor numerical integration are of practical use. The MA representation requires an infeasibly large number of terms to attain any precision when there is even moderately slow decay of the MA terms. The method is also slow,

Table 1: Comparison of ACF computation methods, $d = 0.45, \phi = 0.8, \theta = -0.5$

	MA (7)	Hosking (8)	Integration (10)	γ_k^S, Sowell
γ_{31}/γ_0^S	0.12543 (64 terms)	0.74707 (64)	0.74771	0.74771
	0.17978 (128)	0.74771 (128)		
	0.49223 (320 000)			
timing	> 1000	15	250	1

Listed are the values of the scaled 31st autocovariance of methods (7), (8), (10) and Sowell's method. The numbers in parentheses are the number of terms used in each approximation. The final row lists the computational time relative to Sowell's method.

despite using fast Fourier transforms to compute ψ_j and to evaluate the sum. Numerical integration is also too slow unless only a few autocovariances are required.

Sowell's method, on the other hand, is both fast and accurate. For Hosking's method we implemented the convolution using the fast Fourier transform, which made it a reasonably fast procedure. Its simplicity may occasionally be an argument for using it, e.g. in more complex models.

3.4 Computing the residual sums of squares

There are several ways to compute $\widehat{\mathbf{z}}'\mathbf{R}^{-1}\widehat{\mathbf{z}}$ and the determinant of the $T \times T$ matrix $\mathbf{R}$ in order to determine (5) or (6). Note first that $\mathbf{R}$ has a Toeplitz structure with only T different elements. We employ Durbin's algorithm, see e.g. Brockwell and Davis (1993, §5.2) for a time-series interpretation, or Golub and Van Loan (1989, §4.7.2) for a strictly numerical description. The algorithm solves

$$\mathbf{R} = \mathbf{LDL}' = \mathbf{PP}', \quad \mathbf{e} = \mathbf{D}^{-1/2}\mathbf{L}^{-1}\mathbf{z}. \tag{11}$$

This amounts to direct computation of the Choleski decomposition of $\mathbf{R}^{-1}$. By applying the factorization as it is computed, storage of the $\frac{1}{2}T(T+1)$ elements of the lower triangular matrix $\mathbf{L}$ is avoided. $\mathbf{D}$ is a diagonal matrix and $\mathbf{e}$ can be considered as a $T \times 1$ vector of residuals of the dynamic model. The same procedure is used to compute $\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}$. Sowell (1992) used a related algorithm which required storage of a triangular $T \times T$ matrix; this was an unnecessary part of his maximum-likelihood procedure. The jth row of $\mathbf{L}^{-1}$ consists of the partial linear regression coefficients $\tilde{\pi}_{j-1}^{j-1}, \tilde{\pi}_{j-2}^{j-1}, \dots, 1, 0, \dots$, where $\tilde{\pi}_j^j$ is the j-th partial autocorrelation. The diagonal matrix $\mathbf{D}$ contains the variances of the one-step ahead prediction errors for observation j conditional on the observations $1, \dots, j-1$, scaled by σ_ε^2. For large j (and T) the partial linear regressions coefficients converge to the first $j-2$ elements of the AR(∞)-representation (2), and therefore the residuals e_t converge to the naive residuals $\tilde{e}_t$ from the naive predictions.

3.5 Alternative computations and approximations of the likelihood

Two alternatives for the computation of the exact log-likelihood have been suggested, which attempt to avoid the computation of the T autocovariances for each (d, ϕ, θ). Chan and Palma (1998) use a prediction error decomposition of the likelihood, which is easily calculated using Kalman Filter recursions. Unfortunately, the required state dimension equals the sample size T and the computation of the covariance matrix of the initial state still requires the computation of Σ. Both with regard to storage requirements and number of computations this is not an attractive method for exact maximum likelihood estimation.

Ravishanker and Ray (1997) employ a factorization of the likelihood earlier suggested by Pai and Ravishanker (1996). Their method only requires computation of the first p and the last $2p+q-1$ autocovariances and is easily extended to multivariate models. They achieve this by introducing $p+q$ extra parameters which represent p pre-sample values of z_t and q pre-sample values of ε_t.

Beran (1995) developed an approximate maximum likelihood estimator based on minimizing the sum of squared naive residuals, which is also applicable for nonstationary ARFIMA-processes with $d > 0.5$. The approximate log likelihood is:

$$\log L_A\left(d,\phi,\theta,\beta\right) = c - \frac{1}{2}\log\frac{1}{T-k}\sum_{t=2}^{T}\tilde{e}_t^2, \tag{12}$$

where $\tilde{e}_t$ are the one-step-prediction errors from the naive predictions defined near (2). Beran proved asymptotic efficiency and normality of the resulting estimators for (d,ϕ,θ). Beveridge and Oickle (1993) and Chung and Baillie (1993) presented Monte Carlo evidence which suggest it to be a good estimator for ARFIMA(0,d,0) models with unknown mean. Chung and Baillie (1993) called this estimator the conditional sum-of-squares estimator. We call it the nonlinear least-squares (NLS) estimator.

EML, MPL and NLS are all implemented in the ARFIMA-package of Doornik and Ooms (1999), which is a class of procedures in the programming language Ox, see Doornik (1998). The package also implements the forecasting and bootstrap methods discussed below.

4 Forecasting

The residuals e_t of EML and $\tilde{e}_t$ of NLS estimation are the results of two different methods of prediction: best linear unbiased prediction and naive prediction, which can also be applied out-of-sample.

The best linear prediction of z_{T+H}, given the information in $\mathbf{z}$ and knowing the parameters of the ARFIMA process, is given by

$$\widehat{z}_{T+H} = \left(\gamma_{T-1+H}\cdots\gamma_H\right)\left(\Sigma_T\right)^{-1}\mathbf{z} = \mathbf{q}_H'\mathbf{z} \tag{13}$$

see e.g. Beran (1994, §8.7) or Brockwell and Davis (1993, §5.1). Again, this can be viewed as a regression of z_{T+H} on $\mathbf{z}$, and $(\Sigma)_T^{-1}\mathbf{z}$ can be computed efficiently using a Durbin type algorithm. Let $\mathbf{Q}_{H|T}\mathbf{z}$ denote the optimal forecast for $\widehat{\mathbf{z}}_{H|T} = (\widehat{\mathbf{z}}_{T+1}\cdots\widehat{\mathbf{z}}_{T+H})'$, then $\mathsf{var}(\widehat{\mathbf{z}}_{H|T}) = \Omega_{H|T} = \Sigma_H - \mathbf{Q}_{H|T}\Sigma_T\mathbf{Q}_{H|T}'$. The diagonal elements give the mean squared errors of the optimal j-step ahead forecasts $j=1,\dots,H$. It is often of interest to forecast partial sums of z_t, e.g. when log price-level predictions are constructed as partial sums of inflation forecasts. The variance matrix of cumulative prediction errors is then easily computed as $\mathbf{C}\Omega_{H|T}\mathbf{C}'$ where $\mathbf{C}$ is a lower triangular matrix of only ones.

The naive method recursively predicts $\tilde{z}_{T+1}, \tilde{z}_{T+2},\dots$ using AR-representation (2) up to order $T, T+1,\dots$. Corresponding variances of $\tilde{z}_{T+H}$ are computed using the MA-coefficients of (3):

$$\mathsf{var}(\tilde{z}_{T+H}) = \sigma_\varepsilon^2 \left(1 + \sum_{j=1}^{H-1} \psi_j^2\right). \tag{14}$$

Again, $\tilde{z}_{t+H|T}$ converges to $\widehat{z}_{t+H|T}$ as $T \to \infty$.

5 First order, higher order and simulation-based inference

5.1 Wald, LM and LR type tests

The asymptotic efficiency and normality of the maximum likelihood estimators for (d, ϕ, θ) and β have been established by Dahlhaus (1989) and Dahlhaus (1995).

We use numerical differentiation to compute the covariance matrix of the $(\hat{d}, \hat{\phi}, \hat{\theta})$ from the Hessian, of the (modified) profile log likelihood function at the (approximate) M(P)L estimates. The covariance matrix of $\widehat{\beta}$ is of the familiar GLS type:

$$\mathsf{var}\widehat{\beta} = \widehat{\sigma}_\varepsilon^2(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}.$$

To our knowledge, an MPL estimator for $\widehat{\sigma}_\varepsilon^2$, has not been derived, except for special cases like the standard linear model without dynamics, where it equals the familiar unbiased OLS-estimator: $\widehat{\sigma}_\varepsilon^2 = (T-k)^{-1}\mathbf{e}'\mathbf{e}$, see e.g. Barndorff-Nielsen and Cox (1994, example 4.9). For "MPL-inference" on β we employ the OLS-formula for $\widehat{\sigma}_\varepsilon^2$. Kiviet and Phillips (1998) discuss recent developments in bias corrections of $\widehat{\sigma}_\varepsilon^2$ in AR-models and suggest to increase the degrees-of-freedom correction, k, also with the number of estimated AR parameters, p.

Wald-tests on well identified null-hypotheses in the interior have limiting chi-squared null distributions. We compute the covariance matrix for the nonlinear least squares estimates $(\hat{d}, \hat{\phi}, \hat{\theta}, \hat{\beta})$ using a numerical Hessian evaluated at the NLS estimates.

By prefiltering z_t using the naive filter (2) with a fixed d, one can test every $I(d)$ null hypotheses against higher or lower orders of integration using chi-squared asymptotic inference. Gil-Alaña and Robinson (1997) followed this approach using LM tests for $I(0.5)$ up to $I(2.25)$. Their semiparametric tests have poor finite sample characteristics, making the outcomes hard to interpret.

Prefiltering with $(1-L)^d$ can also be used to obtain test statistics for AR unit-root non-stationarity, where null distributions usually do not have limiting chi-squared distributions. Testing $\Phi(1) = 0$ against $\Phi(1) < 0$ without assuming a fixed value for d under the null is difficult. The null distribution of the test statistic will then depend on the unknown d, see Chan and Terrin (1995), and Jeganathan (1999). Ling and Li (1999) considered AR unit root testing using joint NLS estimates of ARMA parameters and d in semi-long memory ARFIMA models. They obtained limiting distributions, free of nuisance parameters, comparable to the limiting distributions for OLS based unit root tests in AR models obtained by Chan and Wei (1988). However, semi-long memory ARFIMA models have starting conditions that differ essentially from the standard ARFIMA model, also when $\Phi(1) < 0$ and $d < 0.5$.

As we mentioned above, standard asymptotic inference does not always work well for ARFIMA models in finite samples. The EML estimator can be severely biased, and empirical confidence levels for Wald tests may deviate substantially from the nominal levels. Hauser (1999) showed in extensive Monte Carlo simulations for low-order ARFIMA models that empirical confidence levels for EML based tests on (d, ϕ, θ, β) may deviate substantially from their theoretical counterparts, especially when parameters are near the boundary of the parameter space and when $\Phi(z)$ and $\Theta(z)$ are close to having common roots, so that the model is approximately non-identified.

5.2 Bootstrap inference

In light of the previous section it is therefore important to check whether one is dealing with a problematic parameter region for inference. If an estimator or test is not well behaved it may be necessary to transform the parameters or change the estimation procedure to get more reliable inference. A parametric bootstrap analysis is a very useful way to examine the appropriateness of the different estimators and their corresponding tests. Efficient computations now allow us to perform serious bootstrap exercises for empirically relevant cases in minutes or hours, rather than days.

All that is required for the parametric bootstrap, in addition to efficient implementations of estimators and tests, are exact drawings from the DGP of an ARFIMA process. Let $\mathbf{PP}'$ be a Choleski factorization of Σ then drawings $\mathbf{y}$ are conveniently generated as $\mathbf{y} = \mathbf{P}\varepsilon + \mathbf{X}\beta$ where ε is a vector of independent standard normal drawings. For large samples sizes, storage of the triangular matrix $\mathbf{P}$ may be problematic. In that case, an inverse version of Durbin's algorithm can be applied.

We use the bootstrap samples for several purposes. First, we check the distribution of the estimators and corresponding tests empirically for completely specified data generating processes within our model class using the estimates for our data. We look both at marginal and at joint distributions for the parameter estimates. This bootstrap analysis also allows us to get an idea of the effect of parameter uncertainty on the forecast error variances of feasible forecasts. This is especially important for forecasts of nonstationary processes, where the rate at which the forecast error variance grows with the horizon, depends on the unknown parameter d.

Next, we use bootstrap samples for parametric bootstrap tests: the time series is resampled using the estimates of parameters under a null hypothesis of interest. We compute the test statistics for the observed sample, as well as for each bootstrap sample. If the estimated bootstrap p-value, i.e. the proportion of simulated test statistics that exceeds the observed test statistic, is smaller than our significance level, the null hypothesis is rejected. In many circumstances one may expect bootstrap inference to be more accurate than standard asymptotic inference, see Davidson and MacKinnon (1999b). Finally, we use parametric bootstrap samples for a DGP outside the basic model class, in our case an ARFIMA model with GARCH one-step-ahead prediction errors defined by

$$\varepsilon_t = h_t^{1/2}\eta_t, \tag{15}$$

$$h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \tilde{\beta}_i h_{t-i}, \tag{16}$$

where η_t is i.i.d zero mean random variable with $E(\eta_t^2)$=1.

This enables us to check the robustness of our inference with respect to the model assumptions. Since an exact likelihood method is not available for the ARFIMA–GARCH model, we use an extension of the NLS estimator, cf. Ling and Li (1997):

$$\begin{aligned}
\log L_A\left(d, \phi, \theta, \beta, \alpha, \tilde{\beta}\right) &= c - \tfrac{1}{2}\textstyle\sum_{t=2}^{T} \log h_t - \tfrac{1}{2}\log \tfrac{1}{T-1}\textstyle\sum_{t=2}^{T} \dfrac{\tilde{e}_t^2}{h_t} \\
h_t &= \alpha_0 + \textstyle\sum_{i=1}^{r} \alpha_i \tilde{e}_{t-i}^2 + \textstyle\sum_{i=1}^{s} \tilde{\beta}_i h_{t-i},
\end{aligned} \tag{17}$$

where $\alpha = (\alpha_0, \ldots, \alpha_r), \tilde{\beta} = (\beta_1, \ldots, \beta_s)$.

Asymptotic robustness of approximate Gaussian estimators for ARFIMA models in the presence of GARCH errors has recently been investigated. Robinson and Henry (1999) showed the widely used frequency domain semi-parametric of d to be robust to GARCH errors under some additional moment conditions on the innovations. Ling and Li (1997) showed the information matrix corresponding to (17) to be block-diagonal with respect to the ARFIMA and GARCH parameters, which means that the simple NLS-estimator based on (12) is asymptotically efficient in the presence of GARCH errors. Ling and Li (1997) also assumed moment conditions on the innovations but they did not need to assume $d < 0.5$.

6 Modelling Inflation in the US and UK

In this section we apply ARFIMA-models to two aggregate postwar inflation series, one for the US, and one for the UK. These data possess typical long-memory properties: sample autocorrelation functions decay only very slowly, but there is no clear long-run trend indicating nonstationarity. Indeed, many postwar aggregate inflation series of OECD countries have been successfully modelled using a variety of long-memory models and associated methods of inference. These studies regularly found values for d that differed significantly from zero, see e.g. Hassler and Wolters (1995), Delgado and Robinson (1993), Baillie, Chung, and Tieslau (1996), and Bos, Franses, and Ooms (1999). Seasonality also matters, see e.g. Hassler and Wolters (1995) and Ooms and Hassler (1997).

The first series is a monthly price index series for core inflation in the US, with Bureau of Labor Statistics code SA0L1E, i.e. the urban consumer price index for all items except food and energy, which is one of the series analyzed by Ooms (1996). The sample period is 1959.1–1995.12. The second series is a quarterly UK price index series for total consumption used in the analysis of Scott (1995), for the period 1959.1–1992.4. We construct the corresponding inflation series by taking logs and first differences. Both inflation series show long memory and clear seasonality. We first adjust the series for deterministic seasonality by regression on centered seasonal dummies. In order to avoid unnecessary complications, we assume the mean seasonal pattern to be constant over time, i.e. the changes in the seasonal factors in prices are not persistent. There is no clear evidence against this assumption e.g. when we apply tests for the order of integration at seasonal frequencies.

Figure 1a presents a time-series plot of seasonally-adjusted US core inflation. Figure 2a contains the corresponding sample autocorrelation function (SACF) $\widehat{\gamma}_k/\widehat{\gamma}_0$, for lags of 1 up to 120 months. All figures were made using GiveWin 1.2, see Doornik and Hendry (1999).

The time-series plot shows slow mean reversion of inflation. There is a pronounced additive outlier in July 1980 which has to be taken into account. We model it with a single dummy variable for a temporary change in mean in 1980.7. The decay in the SACF is very slow, indicating a d close to, or inside the nonstationary region $[0.5, \infty)$. There is also some evidence of remaining (short-memory) seasonality in the SACF with local maxima at lags that are a

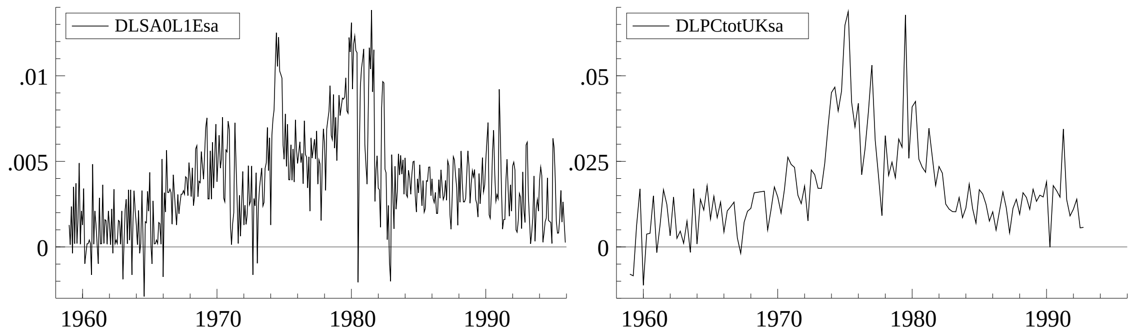

Figure 1: US and UK Seasonally Adjusted Inflation Rates.

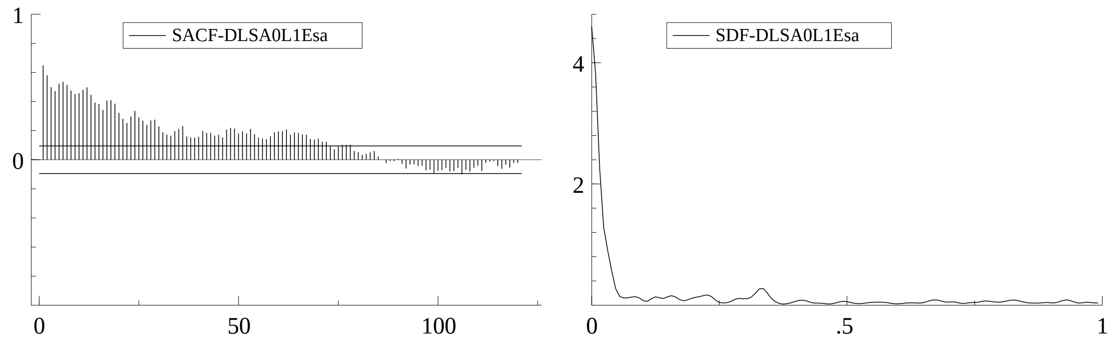

Figure 2: Sample Autocorrelation Functions and Spectral Densities of US Inflation Rates.

multiple of 6. The nonparametric estimate of the spectral density function (SDF) in Figure 2b shows a clear peak near frequency zero.

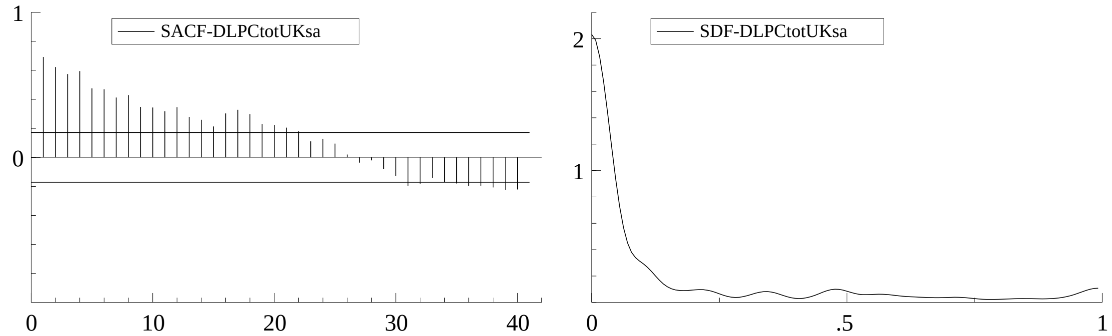

Figure 3: Sample Autocorrelation Functions and Spectral Densities of UK Inflation Rates.

Figure 1b shows the corresponding time-series plot for seasonally-adjusted UK consumer price inflation, DLPCtotUKsa, which seems more erratic. Here we treat 1979.3 as an outlier. The outliers in 1990.2 and 1991.2 are clearly connected, which is why we introduce one dummy variable taking value 1 in 1990.2 and value -1 in 1991.2. The SACF and SDF are presented in Figure 3. The basic pattern is the same as for the US series.

6.1 Inference

An adequate, yet parsimonious, approximation to the second-order characteristics for both time series is given by the following ARFIMA model with zero-frequency long-memory and short-memory seasonality:

$$(1 - \phi_p L^p)(1 - L)^d (y_t - x_t'\beta) = (1 + \theta_q L^q)\varepsilon_t,$$

where $p = q = 12$ for the monthly US data, and $p = q = 4$ for the quarterly UK data, and x_t consists of a constant and dummy variables. Tables 2 and 3 present estimation results for both countries using EML, MPL and NLS. The NLS-GARCH results are discussed in §8. The US sample is larger, and we expect a closer similarity of EML, MPL and NLS for the US series. This is confirmed by the US values for $\widehat{d}$ close to 0.4 obtained for all three methods.

The UK estimates for d vary from 0.47 to 0.64. The standard errors for the parameter estimates are also much larger for the UK series than for the US series. The EML standard error for $\widehat{d}$ is an exception, but there $\widehat{d}$ is too close to 0.5 to attach too much meaning to its standard error as we show below. Both the US and UK series exhibit clear evidence of long memory. The US series seems covariance stationary, and the UK seems nonstationary, but one cannot easily reject $d = 0.5$ for the UK data.

The differences between the estimators are more pronounced and therefore more interesting for the UK. The results in Table 3 illustrate an important practical problem of estimating a d near 0.5. EML delivers an estimate smaller than 0.5, whereas MPL did not converge. Therefore we difference the data and re-estimate. The tabulated value for d is $1 + \widehat{d}_\Delta$. As a consequence we can no longer identify a value for the mean of inflation: the constant drops out of the regressor set. Note that MPL and EML differ only if there are regressors in the model, so except for the small effect of the dummy variables, the MPL results can also be viewed as EML estimates obtained by modelling the first differences.

Table 2: Estimation results for US inflation

	EML	MPL	NLS	NLS-GARCH
$\widehat{d}$	0.400 (0.033)	0.411 (0.034)	0.409 (0.036)	0.316 (0.038)
$\widehat{\phi}_{12}$	0.769 (0.101)	0.783 (0.091)	0.620 (0.110)	0.678 (0.077)
$\widehat{\theta}_{12}$	-0.605 (0.123)	-0.621 (0.113)	-0.447 (0.121)	-0.460 (0.085)
Constant	0.00313 (0.0024)	0.00309 (0.0028)	0.00432 (0.00174)	0.00362 (0.00095)
DUM807	-0.0114 (0.0017)	-0.0114 (0.0017)	-0.0115 (0.0017)	-0.0105 (0.0024)
$\widehat{\alpha}_1$				0.201 (0.064)
$\widehat{\beta}_1$				0.762 (0.068)
$10^6\widehat{\sigma}^2$	3.6260	3.6250	3.5476	3.7889

Sample: 1959.1–1995.12 ($T = 444$). All methods directly applied to inflation series.

There is no need for a priori differencing when applying NLS. There is, however, a change in the interpretation of the constant term for different values of d. For $-0.5 < d < 0.5$ the constant term represents the mean of inflation, but for $0.5 < d < 1$ it should be interpreted as the mean growth rate of inflation. The constant is of course unidentified (cannot be estimated) if $d = 1$. In the UK case the mean growth estimate of -0.07 is apparently not well identified in the data.

Table 3: Estimation results for UK inflation

	EML	MPL	NLS	NLS-GARCH
$\widehat{d}$	0.471 (0.035)	0.581 (0.078)	0.642 (0.074)	0.520 (0.074)
$\widehat{\phi}_4$	0.767 (0.13)	0.809 (0.11)	0.740 (0.078)	0.789 (0.062)
$\widehat{\theta}_4$	-0.487 (0.18)	-0.544 (0.15)	-0.487 (0.12)	-0.555 (0.102)
Constant	0.0116 (0.034)	—	-0.0695 (0.040)	-0.0500 (0.018)
DUM793	0.0411 (0.0063)	0.0411 (0.0060)	0.0409 (0.0057)	0.0405 (0.0072)
DUM902912	0.0185 (0.0040)	0.0187 (0.0039)	0.0187 (0.0037)	0.0187 (0.0021)
$\widehat{\alpha}_1$				0.242 (0.109)
$\widehat{\beta}_1$				0.681 (0.108)
$10^5\widehat{\sigma}^2$	5.4399	5.4769	5.1171	5.2354

Sample: 1959.1–1992.4 ($T = 136$). EML, NLS applied to inflation; MPL applied to Δinflation.

As discussed in §3, the difference between MPL and NLS is mainly due to the residual sum of squares term in the likelihoods of both methods. Figure 4 compares the residuals of MPL and NLS. The residuals in the first part of the sample are indicated by small circles. There the naive residuals of the NLS method are clearly smaller than the MPL residuals. The information in the first part of the sample is used more efficiently by MPL. It is clear that NLS effectively reduces the weight of the first observations on the inference. Although we obtain a large d, we observe a close similarity of the residuals of the two methods in the remaining part of the sample. The partial linear regression coefficients mimick the coefficients of the AR(∞)-representation quite closely for, say, $j > 40$, cf. §3.4.

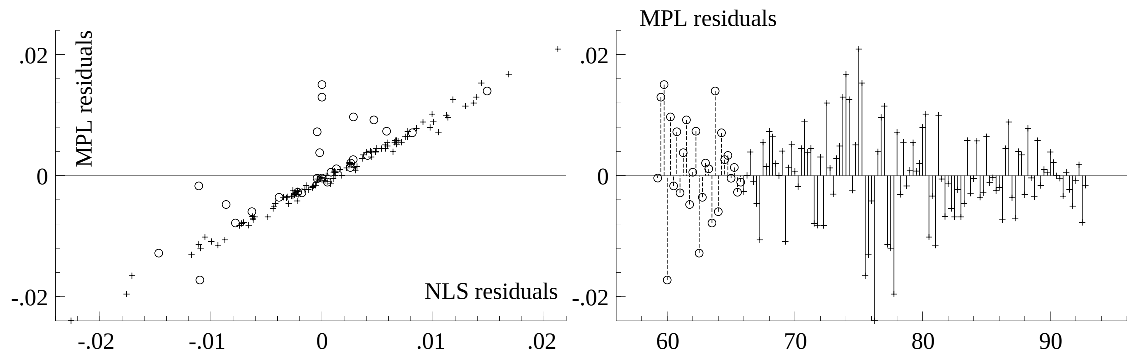

Figure 4: Residuals for MPL and NLS methods for the UK model.

Table 4 presents diagnostics from PcGive (Hendry and Doornik (1999)), applied to the EML and MPL residuals e_t and to the naive NLS residuals $\tilde{e}_t$. There is no clear evidence against the white noise assumption but some non-normality and ARCH effects are detected. The results for the EML-e_t, MPL-e_t and $\tilde{e}_t$ are qualitatively similar. The UK results for EML and MPL match more closely than MPL and NLS, although the latter residuals are based on approximately the same $\widehat{d}$. This is most clear in the normality tests. It seems that the effective downweighting of the first observations by NLS also happens to remove evidence against non-normality.

The LM-type test results must be treated with some caution since we did not correct the test statistics for the prefitting of dynamic parameters and the removal of the outliers.

Table 4: Diagnostic tests for residuals

UK residuals	AR 1-5		ARCH 4		Normality	
EML	0.793	[0.556]	5.578	[0.000]	8.299	[0.016]
MPL	0.569	[0.724]	5.031	[0.001]	7.387	[0.025]
NLS	0.264	[0.932]	5.395	[0.001]	5.178	[0.075]
US residuals	AR 1-48		ARCH 11		Normality	
EML	1.248	[0.134]	4.539	[0.000]	4.929	[0.085]
MPL	1.245	[0.136]	4.410	[0.000]	5.158	[0.076]
NLS	1.295	[0.099]	4.388	[0.000]	7.696	[0.021]

LM tests for AR-type serial correlation in the conditional mean and ARCH-type serial correlation in the conditional variance; Doornik and Hansen (1994) tests for normality. p-values in brackets.

Li and McLeod (1986) discussed the behaviour of time domain serial correlation tests of naive residuals of long-memory models; Harvey and Koopman (1992) discussed the effect of prefiltering on normality tests. As long as the dynamics are well specified, standard normality tests can be applied. As an additional specification test we tried to fit extra ARMA parameters. These turned out not to be significant. The bootstrap analysis presented below shows no problems of non-identification because of cancelling roots in the AR and MA part of the model, so the models do not seem to be overspecified either.

6.2 Forecasting

An important goal of long-memory time-series modelling is to perform inference on long-range forecasts. How do the differences in estimates of d and β translate into the location and width of forecast intervals for inflation and the log price level? We present the main picture in Figure 5. We give detailed results in Tables 5, 6 and 7.

Figure 5 displays the last 7 years of the UK sample, together with nominal 95% confidence forecast intervals up to a horizon of 8 years. The different panels clearly show how the rate of growth of the forecast interval depends on $\widehat{d}$. The effective orders of integration are about 0.47, -0.42 and 0.64 for the top graphs and 1.47, 1.58 an 1.64 for the bottom graphs.

The EML estimate of 0.47 indicates stationarity for the level of UK inflation. With a d so close to 0.5 we observe that the forecast and its variance converge only very slowly to the unconditional mean and variance. The (short-memory) deviations in the stationary seasonal pattern are apparently predictable. This shows most clearly in the forecasts for the changes in inflation ($d = -0.42$) which form the basis of the MPL projections. Here the forecasts converge must faster to the unconditional (zero) mean and variance of the process.

The NLS forecasts for inflation are based on $d = 0.64$, which is clearly above 0.5. The regression constant in the model now determines a (downward) trend inflation. The variance of the H-step ahead forecasts grows proportionally to H^{2d-1} for nonstationary $d > 0.5$, assuming $H/T \to 0$, c.f. Beran (1994, §8.6). Therefore, the width of the NLS forecast interval for inflation is proportional to the corresponding square root: $cH^{0.64-0.5}$. Eventually the entire forecast interval for inflation will be strictly smaller than zero, but not until far into the third millennium.

The different forecasts for the log price level are displayed in the bottom graphs of Figure 5. All these forecasts show clearly that the UK price level is not mean reverting: $d \gg 1$.

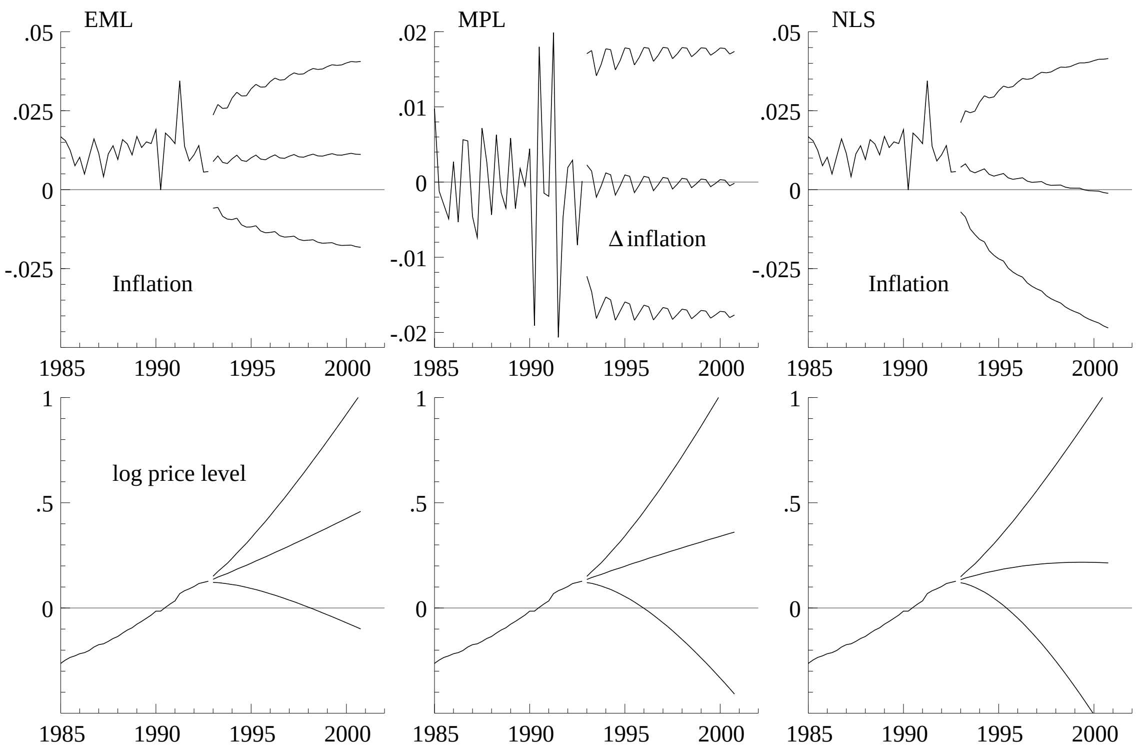

Figure 5: Forecasts of (Δ) inflation levels and log price levels with 95% confidence bands for the EML, MPL and NLS estimates for the UK model. Optimal forecasts for EML and MPL, naive forecasts for NLS.

The EML price level forecast is linearly trending and the width of the forecast interval is proportional to $H^{0.97}$. Therefore, the interval grows (nearly) at a linear rate as well. The forecast interval is eventually bounded from below, although this does not seem important for relevant forecast horizons. The MPL price-level forecasts are also linearly trending, since we have integrated the zero mean Δinflation series twice. The slope resembles the slope in the EML forecasts. The forecast interval grows proportionally to $H^{1.08}$. The NLS forecast approaches a quadratic trend and the growth of the forecast interval reflects the covariance nonstationarity most clearly.

Table 5: Forecast results for UK log-price level 1 to 32 quarters ahead

		EML		MPL				NLS	
		optimal: $\widehat{z}_{T+H}$		optimal: $\widehat{z}_{T+H}$		naive: $\tilde{z}_{T+H}$		naive: $\tilde{z}_{T+H}$	
H	Date	forecasts	RMSE	forec.	RMSE	forec.	RMSE	forec.	RMSE
1	1993.1	.13623	(.00738)	.13553	(.00740)	.13537	(.00740)	.13448	(0.00715)
2	1993.2	.14684	(.01314)	.14522	(.01386)	.14481	(.01385)	.14262	(0.01375)
3	1993.3	.15540	(.01879)	.15298	(.02052)	.15224	(.02049)	.14857	(0.02073)
4	1993.4	.16357	(.02440)	.16026	(.02735)	.15913	(.02730)	.15388	(0.02801)
12	1995.4	.24138	(.08474)	.22785	(.10408)	.22081	(.10353)	.19232	(0.11100)
24	1998.4	.36548	(.19902)	.33074	(.26306)	.30649	(.26043)	.21665	(0.28289)
32	2000.4	.45227	(.28430)	.40124	(.38961)	.36041	(.38457)	.21412	(0.41783)

Sample: 1959(1) - 1992(4). See Table 3 for model and parameter estimates.

Table 5 provides numerical evidence corresponding to Figure 5; it also contrasts the optimal and naive forecasts. The former involve formulae for projection into the finite past. The weight of past observations declines only slowly in long-memory models and one can expect a substantial difference with naive forecasts which use projection as if the infinite past is known. However, Table 5 shows that the differences are small. The results for MPL required us to integrate the forecasts for Δinflation twice. Evenso, the log price-level forecasts 8-years ahead differ only by .04, with estimated standard error of 0.390 and 0.385. In sum, we see that the choice of estimator matters much more than the choice of predictor. Naturally, the difference between optimal and naive forecasts becomes more pronounced as the sample size decreases.

Table 6: Forecast results for US inflation 1 to 24 months ahead

		EML		MPL				NLS	
		optimal: $\widehat{z}_{T+H}$		optimal: $\widehat{z}_{T+H}$		naive: $\tilde{z}_{T+H}$		naive: $\tilde{z}_{T+H}$	
H	Date	forecasts	RMSE	forec.	RMSE	forec.	RMSE	forec.	RMSE
1	1996.01	.29469	(.19045)	.29519	(.19086)	.29564	(.19082)	.27429	(0.18835)
2	1996.02	.31269	(.20514)	.31240	(.20640)	.31305	(.20634)	.29726	(0.20348)
3	1996.03	.26746	(.21197)	.26600	(.21373)	.26678	(.21364)	.26470	(0.21059)
6	1996.06	.17061	(.22160)	.16559	(.22418)	.16664	(.22403)	.19761	(0.22066)
12	1996.12	.17599	(.22940)	.17145	(.23277)	.17285	(.23251)	.20535	(0.22887)
24	1997.12	.20498	(.24560)	.19897	(.25020)	.20102	(.24967)	.25193	(0.24627)

Sample: 1959 (1) - 1995 (12). Model, see Table 2. Forecasts and standard errors multiplied by 100.

Table 7: Forecast results for US log-price level 1 to 24 months ahead

		EML		MPL				NLS	
		optimal: $\widehat{z}_{T+H}$		optimal: $\widehat{z}_{T+H}$		naive: $\tilde{z}_{T+H}$		naive: $\tilde{z}_{T+H}$	
H	Date	forecasts	RMSE	forec.	RMSE	forec.	RMSE	forec.	RMSE
1	1996.01	5.0949	(0.001904)	5.0949	(0.001909)	5.0949	(0.001908)	5.0947	(0.001883)
2	1996.02	5.0980	(0.003277)	5.0980	(0.003302)	5.0980	(0.003301)	5.0976	(0.003254)
3	1996.03	5.1007	(0.004581)	5.1006	(0.004632)	5.1007	(0.004630)	5.1003	(0.004561)
6	1996.06	5.1055	(0.008287)	5.1053	(0.008439)	5.1054	(0.008430)	5.1059	(0.008292)
12	1996.12	5.1187	(0.015232)	5.1183	(0.015627)	5.1185	(0.015597)	5.1203	(0.015316)
24	1997.12	5.1486	(0.030342)	5.1477	(0.031346)	5.1481	(0.031231)	5.1533	(0.030746)

Sample: 1959 (1) - 1995 (12). Model, see Table 2.

The US forecasts are only presented in numerical form. Table 6 shows the projected inflation rates and Table 7 gives the associated price-level forecasts. Note that the inflation forecasts are multiplied by 100. US core inflation is less erratic than UK inflation. Still, the one-month ahead standard forecast error is 0.19%. One quarter ahead the standard error is 0.46% and two years ahead it is 3%. The corresponding figures for UK inflation are 0.8% and 6%. The differences between estimation and forecast methods are much smaller for the US than for the UK data. The asymptotic equivalence of the estimators and projectors is apparent at the US sample size of 444. Two-year ahead price level point and interval forecasts differ at most 0.25% across methods.

7 Bootstrap Inference

Parametric bootstrap inference can be used to test for any value of d. As an important illustration, we take the parameter estimates of the different methods as our null hypotheses. In this way we can check the reliability of EML, MPL and NLS based inference in empirically relevant cases. In practice one can e.g. apply a parametric bootstrap test for $d = 1.01$ for inflation rates in the hope of rejecting the null of no mean reversion. Under each null hypothesis we draw 1000 pseudo-samples using the exact simulation algorithm described in §3. For each pseudo-sample we compute parameter estimates and associated t-values (or directed likelihoods) for tests on the true (DGP) value.

7.1 Bootstrap inference for the US model

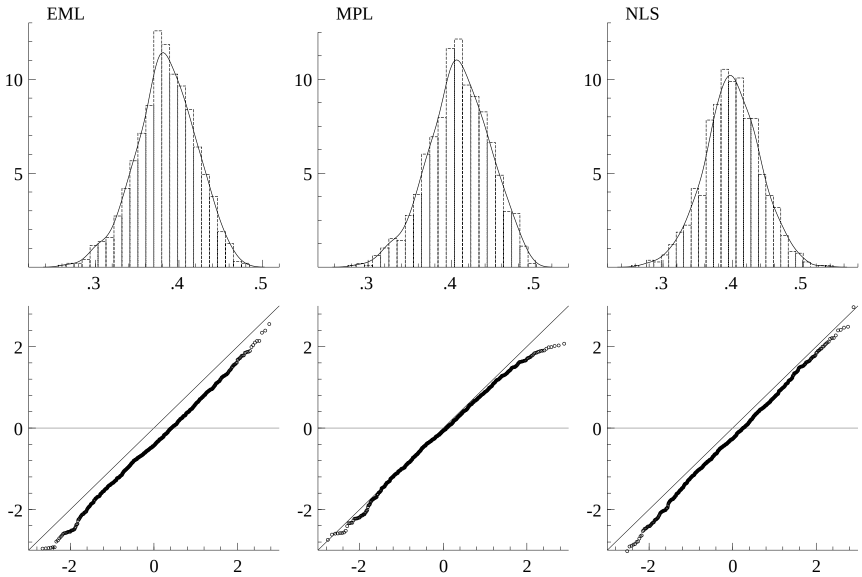

Figure 6: Estimates of parametric bootstrap densities for $\widehat{d}$ and QQ-plots for the corresponding directed likelihood or t-statistics for the US model of Table 2. See also Tables 8 and 9.

The US model bootstrap results in Figure 6 clearly show the downward bias of the EML estimator for d. The MPL estimator bias is much smaller; it is slightly skewed to the left, but this could be exaggerated because we excluded 17 estimates where the estimator did not converge to a $d < 0.5$. Estimation on first differences of inflation, as in our application to the UK data, leads to a more symmetric distribution. The NLS estimator also has a much smaller bias than EML, but is less precise than MPL.

The bias in the estimators leads to a corresponding bias in the t-statistics. The numerical results in Table 8 confirm that MPL-based inference is more accurate in this case. The scatterplots in Figure 7 show that d is well identified: $\widehat{d}$ is hardly correlated with $\widehat{\phi_{12}}$. This is an important check: empirical identification is a necessary condition for the successful application of standard asymptotic results in practice. The scatterplot of ϕ_{12} versus θ_{12} shows that both parameters are not well identified individually. Note that ϕ_{12} and $-\theta_{12}$ seem to differ about two standard errors if one looks at the NLS-output in Table 2 only. This would suggest no problems with regard to common roots in the AR and MA part of the model. However, especially the NLS-inference for these individual parameters is completely

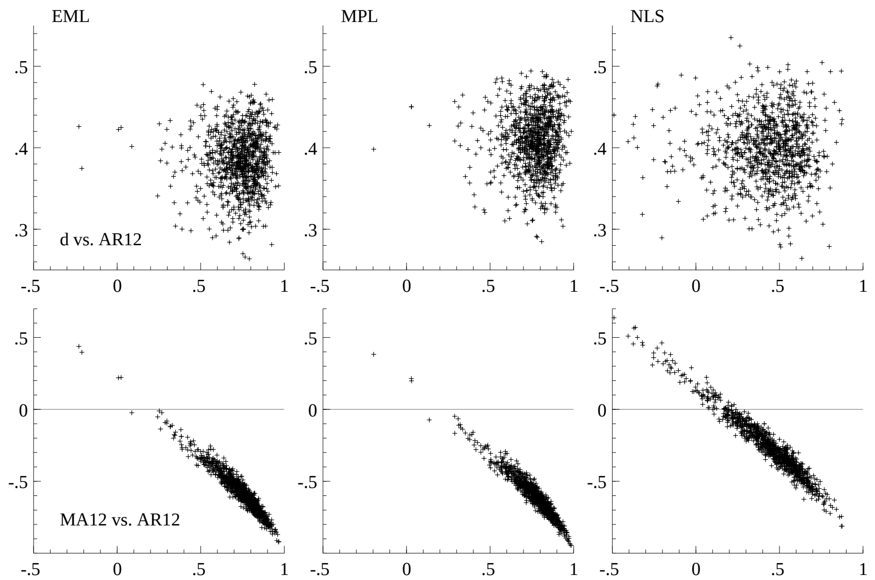

Figure 7: Cross plots of parametric bootstrap replications of three estimators for US model of Table 2.

Table 8: Parametric bootstrap estimator results for estimators US core inflation

	EML		MPL		NLS	
	bias	RMSE	bias	RMSE	bias	RMSE
$\widehat{d}$	-0.0156	0.039	-0.0031	0.037	-0.0099	0.040
$\widehat{\phi}_{12}$	-0.0386	0.143	-0.0208	0.125	-0.1865	0.284
$\widehat{\theta}_{12}$	0.0377	0.163	0.0194	0.147	0.1899	0.289
Constant	-0.00019	0.0025	-0.00021	0.0028	-0.00020	0.0025
DUM807	-0.00003	0.0017	-0.00004	0.0017	0.00004	0.0017
$10^6\widehat{\sigma}^2$	-0.0500	0.237	-0.0322	0.235	-0.0118	0.262

See Table 2 for the DGP, Fig. 6 for the density estimates. 1000 replications.

unreliable, even for this sample size of 444. The estimators are biased by about -0.2 and 0.2, they are not Gaussian and show a huge overdispersion. EML and especially MPL based inference works much better, although the EML and MPL estimators display some severe outliers as well.

Fortunately, inference for the individual AR and MA parameters is not as important as inference on d. The sum, $\phi_{12}+\theta_{12}$, can be interpreted as the impulse response of the short-memory component of inflation after one year. This statistic has a much better behaved distribution, also when applying the NLS estimator. The short-memory impulse response after 2 years, $\phi_{12}(\phi_{12}+\theta_{12})$, is also well behaved, see Figure 8. Again, we see that MPL leads to more accurate inference.

The estimation of the dummy parameters presents no problems. The distribution of the estimator for the constant is also approximately Gaussian and the empirical RMSE corresponds quite closely to the estimated standard error, compare Tables 8 and 2. However, Table 9 shows that the t-test for the constant is overdispersed, especially if one applies NLS. The overdispersion is a consequence of the parameter uncertainty in d which is not properly taken into account using the present methods. The uncertainty in $\widehat{d}$ is too large to use con-

Table 9: Empirical rejection probabilities at 10% and 5% nominal level for two-sided t-tests. US core inflation

	EML		MPL		NLS	
	10%	5%	10%	5%	10%	5%
$t-\widehat{d}$	.122	.067	.083	.041	.115	.058
$t-\widehat{\phi}_{12}$	.103	.075	.131	.092	.270	.174
$t-\widehat{\theta}_{12}$	.118	.069	.132	.094	.288	.192
$t-$Constant	.187	.125	.142	.094	.328	.255
$t-$DUM807	.109	.056	.111	.057	.113	.061

See Figure 6 for the QQ-plots. 1000 replications.

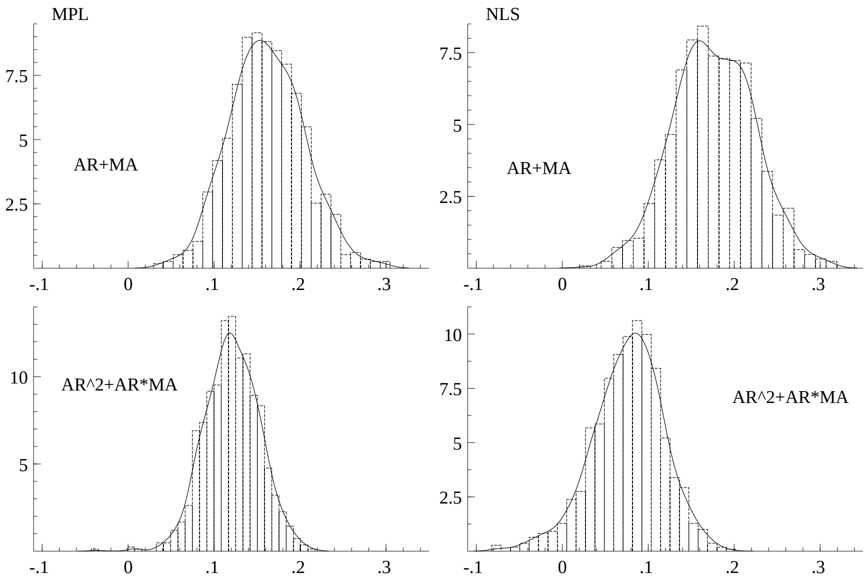

Figure 8: Estimates of parametric bootstrap densities impulse responses short-memory part of the US model at seasonal lags 12, AR+MA parameter, and 24, AR^2+AR×MA.

ventional GLS-type standard errors for inference on the mean of the process, illustrated by the fact that the empirical coverage probability at a nominal 90% level is only 67%.[1] This indicates that NLS-based Wald tests for structural stability, which test for changes in the mean, may also suffer from overrejection in similar situations; also see Bos, Franses, and Ooms (1999), who experienced the same problem with LM-tests for structural stability. Note that the problem of estimating the standard error of the mean of inflation is related to the problem of estimating the standard error of the forecasts of longer horizon price levels, which is largely equivalent to estimating the standard error of the mean of future inflation. Forecast analysis is discussed in §7.3.

Summarizing the bootstrap results for the US model, we see that MPL based inference is clearly more efficient and reliable than EML and NLS. The differences between EML, MPL and NLS are larger than one would expect from the simple point estimates.

[1]Beran (1994, §8.6) provides more accurate (non-Gaussian) approximations to the finite sample distribution of the test statistic for the mean of an ARFIMA(0,d,0) process.

7.2 Bootstrap inference for the UK model

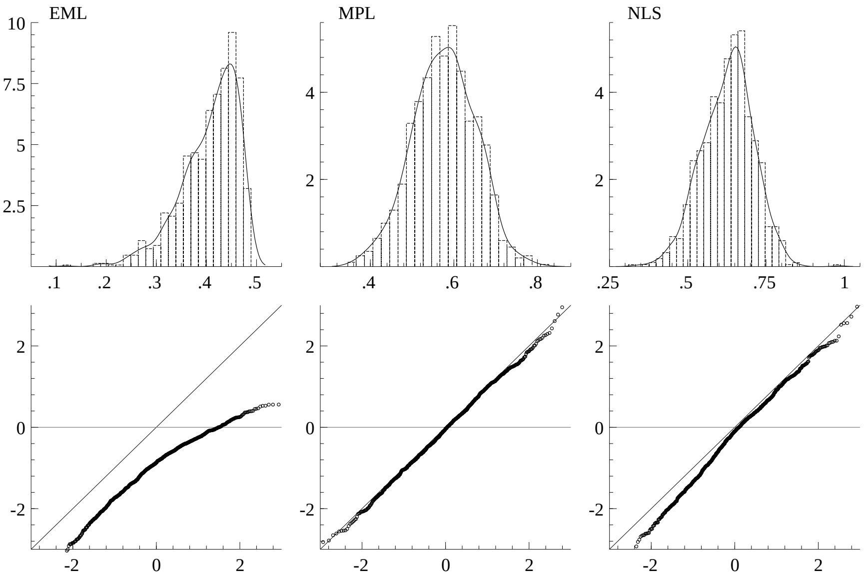

Figure 9: Estimates of parametric bootstrap densities for $\widehat{d}$ and QQ-plots for the corresponding 'directed' likelihood or t-statistics for the UK model of Table 3. See also Tables 10 and 11.

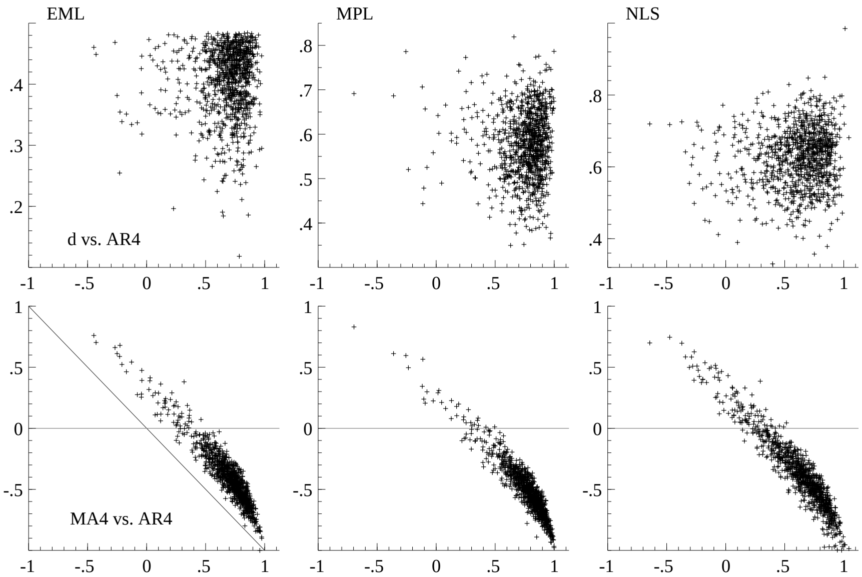

Figure 10: Cross plots of parametric bootstrap replications of three estimators for UK model of Table 3.

The bootstrap results for the UK model tell a similar story. On the whole, the estimators are of course less precise than in the US model: the sample is smaller and covers a shorter period. The problems with the EML-estimator for d close to 0.5 are very clear from Figures 9 and 10. The QQ-plot for the t-statistics of the MPL estimator (here applied to first differences of inflation) shows that Sowell's solution, which is to use first differences (estimating $d-1$ instead of d), works well, even in this small sample. Direct application of NLS to the inflation levels does not work as well, but it is not nearly as flawed as applying EML to the levels.

Table 10: Parametric bootstrap estimator results for UK consumer price inflation

	EML		MPL		NLS	
	bias	RMSE	bias	RMSE	bias	RMSE
$\widehat{d}$	-0.0644	0.086	-0.0048	0.076	-0.0146	0.083
$\widehat{\phi}_4$	-0.0857	0.213	-0.0540	0.191	-0.1259	0.283
$\widehat{\theta}_4$	0.0964	0.257	0.0451	0.225	0.1188	0.314
Constant	0.00073	0.033			-0.0014	0.048
DUM793	0.00013	0.006	0.00026	0.006	0.00020	0.006
DUM902912	-0.00004	0.004	0.00002	0.004	-0.00003	0.004
$10^5\widehat{\sigma}^2$	-0.2171	0.655	-0.1415	0.665	-0.2669	0.649

See Table 3 for the corresponding DGP. See Figure 9 for the density estimates.

Table 11: Empirical rejection probabilities at 10% and 5% nominal level for two-sided t-tests. UK consumer price inflation

	EML		MPL		NLS	
	10%	5%	10%	5%	10%	5%
$t-\widehat{d}$	.189	.114	.090	.051	.138	.078
$t-\widehat{\phi}_4$	.120	.080	.137	.105	.284	.203
$t-\widehat{\theta}_4$	.150	.104	.148	.102	.288	.214
$t-$Constant	.460	.393			.295	.206
$t-$DUM793	.103	.053	.099	.049	.113	.062
$t-$DUM902912	.103	.049	.129	.076	.103	.054

See Fig. 9 for the QQ-plots. 1000 replications.

Again we see that ϕ_4 and θ_4 are not well identified individually. There are more outliers in the empirical distribution in Figure 10 than in the corresponding Figure 7 for the US model. However, it is also clear that $\phi_4 + \theta_4 = 0$, i.e. cancelling roots and no seasonal parameters can be rejected. The line $\phi_4 + \theta_4 = 0$ in the scatter plot indicates this. Note that EML and MPL restrict ϕ_4 to an interval $[-1+\delta, 1-\delta]$, where we chose $\delta = 0.001$. NLS does not explicitly impose this restriction. Table 10 summarizes the UK bootstrap results for the parameter estimation. MPL is again clearly preferrable: it is both more efficient and more reliable. NLS is second best, and the EML application is very dangerous. Table 11 summarizes the behaviour of individual t-tests. Only the t-tests for the dummies and the MPL-based t-test for d are reliable. Testing for the overall mean of UK inflation does not seem to make sense, UK inflation being nonstationary. The NLS-based test for a trend in inflation is also very inaccurate.

7.3 Bootstrap forecasts

We use extended bootstrap samples to check the reliability of our asymptotic likelihood-based inference on the forecasts. Note that the derivations of forecast intervals neglect the estimation uncertainty in the parameters and treat them effectively as known. It is interesting to see whether this is appropriate in the current context, where we re-estimate the model in each bootstrap sample.

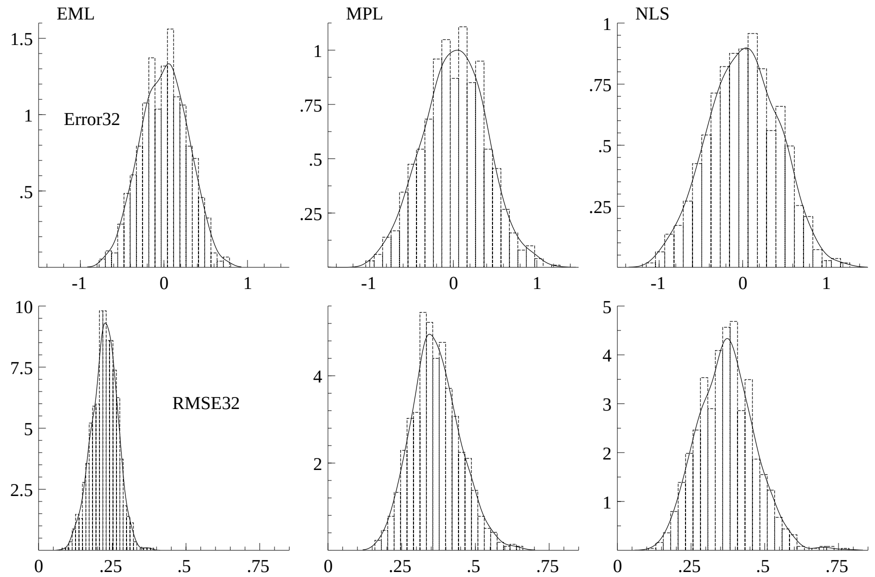

Figure 11: Estimates of parametric bootstrap densities of the 8 year ahead forecast error for the UK log price level, for three estimators and estimates of the density of the corresponding estimators of root mean square error of 8 year ahead forecasts.

Table 12: Parametric Bootstrap Forecast results for US log-price level up to 24 months ahead

		EML (optimal)			MPL (optimal)			NLS (naive)		
		RMSE	$\sqrt{\widehat{\mathsf{var}}(\widehat{z}_{T+H})}$		RMSE	$\sqrt{\widehat{\mathsf{var}}(\widehat{z}_{T+H})}$		RMSE	$\sqrt{\widehat{\mathsf{var}}(\widetilde{z}_{T+H})}$	
h	Date		mean	RMSE		mean	RMSE		mean	RMSE
1	1996.01	.0190	.0189	.00061	.0190	.0190	.00061	0.0187	0.0185	0.00063
2	1996.02	.0323	.0323	.00118	.0326	.0328	.00121	0.0326	0.0318	0.00124
3	1996.03	.0450	.0449	.00193	.0457	.0460	.00200	0.0451	0.0445	0.00206
12	1996.12	.1512	.1465	.01161	.1558	.1546	.01256	0.1525	0.1479	0.01284
24	1997.12	.3002	.2890	.02939	.3115	.3094	.03220	0.3079	0.2963	0.03304

Sample: 1959 (1) - 1995 (12). 1000 replications. All numbers are multiplied by 10.
See Table 2 for the DGP, Table 7 for the corresponding sample estimates of forecast errors.

We present the results in Tables 12 and 13, and in Figure 11. For each estimation method and each horizon we present three numbers. The first is the empirical RMSE of the actual forecast errors across all bootstrap samples. Overall these empirical RMSEs compare well with the RMSE that we estimated directly using the standard formulae for our actual sample (note that the results have been multiplied by 10). The next two statistics describe the empirical distribution of the estimated RMSEs. For one-step ahead forecasts these are mainly determined by the empirical distribution of $\widehat{\sigma}_\varepsilon$. For longer multi-step forecasts they are also influenced by $\widehat{d}$. The uncertainty in the estimated RMSE, which we denote by $\sqrt{\widehat{\mathsf{var}}(\widehat{z}_{T+H})}$ and $\sqrt{\widehat{\mathsf{var}}(\widetilde{z}_{T+H})}$, is quite substantial. This is most clearly shown in Figure 11 for the eight-year ahead forecasts of the UK price level, where we used values for d of 0.47, 0.58 and 0.64 in the data generating process. The difference between those values is also reflected in the mean

Table 13: Parametric Bootstrap Forecast results for UK log-price level up to 32 quarters ahead

		EML (optimal)			MPL (optimal)			NLS (naive)		
		RMSE	$\sqrt{\widehat{\mathsf{var}}(\widehat{z}_{T+H})}$		RMSE	$\sqrt{\widehat{\mathsf{var}}(\widehat{z}_{T+H})}$		RMSE	$\sqrt{\widehat{\mathsf{var}}(\tilde{z}_{T+H})}$	
h	Date		mean	RMSE		mean	RMSE		mean	RMSE
1	1993.1	.0072	.0072	.00043	0.0077	.0073	.00044	0.00696	0.00695	0.00042
2	1993.2	.0133	.0125	.00082	0.0140	.0136	.00094	0.01383	0.01328	0.00094
4	1993.4	.0247	.0223	.00186	0.0271	.0268	.00255	0.02840	0.02690	0.00267
12	1995.4	.0854	.0730	.00950	0.1012	.1009	.01525	0.1130	0.1047	0.01739
24	1998.4	.1986	.1612	.02791	0.2544	.2498	.04989	0.2876	0.2592	0.05885
32	2000.4	.2793	.2236	.04329	0.3801	.3665	.08194	0.4207	0.3784	0.09674

Sample: 1959(1) - 1992(4). Results based on 1000 replications for EML and MPL.
NLS results exclude 4 outlying sets of forecasts. See Table 3 for the DGP.
See Table 5 for the corresponding sample estimates of forecast errors.

and variance of the RMSE estimates in Table 13. Note that the EML-estimator of the RMSE is severely downward biased, the empirical RMSE for EML is about 0.28. This does not mean that EML leads to more efficient inference, it merely reflects the lower value of d in the parametric bootstrap DGP. As indicated above, MPL and NLS inference are more reliable in this case. The MPL and NLS estimators of the RMSE are well behaved and centered around the empirical values of 0.38 and 0.42 respectively.

In sum, there is substantial uncertainty about the variance of future price levels. Suppose the data are generated with $d = 0.6$ and a true 8-year ahead RMSE of 0.36, then an approximate symmetric 95% confidence interval for the MPL-based RMSE estimator is given by [0.20-0.52].

8 Extension with GARCH errors

The Gaussian homoskedastic ARFIMA-model leads to convenient chi-squared inference for d, and it also leads to normal forecast intervals. It gives an adequate characterization of the second order properties of the data and it provides optimal linear forecasts. These inference and optimality properties may not be robust. The diagnostics of Table 4 revealed signs of conditional heteroskedasticity. Therefore we need to see how the methods perform in such a context. After all, ARCH error models were first introduced by Engle (1982) to model the UK inflation series for the 1970s.

To start, we estimate a Gaussian ARFIMA-GARCH(1,1) model using the approximate NLS-type likelihood (17), following Ling and Li (1997) and Baillie, Chung, and Tieslau (1996). The estimation results are presented in the last columns of Tables 2 and 3 to facilitate comparison with the Gaussian models. Standard errors are based on the second derivatives of the log-likelihood, using zero for the cross-derivative between estimated ARFIMA (with regressors) and GARCH parameters.

The GARCH parameters are clearly significant and indicate substantial persistence in the volatility of inflation: $\alpha_1 + \beta_1 > 0.9$ for both series. Additional parameters to allow for t-distributed innovations or for higher order GARCH turned out to be insignificant using traditional likelihood methods.

Although Ling and Li (1997) showed that the information with regard to d and the ARMA parameters on the one hand and the GARCH parameters on the other hand is asymptotically orthogonal, it turns out that the introduction of GARCH errors substantially reduces the estimate of d in this case, whereas the other parameters are largely unchanged. There is no qualitative difference, for the US-series we still find $d < 0.5$, and for the UK we have $d > 0.5$.

The parameter σ^2 in Tables 2 and 3 denotes the unconditional variance of the innovations, or 'average' one-step-ahead forecast error variance, and is comparable between the homoskedastic models and the model with GARCH innovations. It is computed as $\sigma^2 = \alpha_0/(1-\alpha_1-\tilde{\beta}_1)$.

8.1 Effects of neglecting GARCH errors on inference

We use the US estimates for the ARFIMA-GARCH model to examine the effect of neglected GARCHeffects on the performance of the different estimators. We generate 1000 samples following the procedure in §5.2 where we now replace the standard normal errors by GARCH(1,1) errors. The parameters for the DGP were based on the data multiplied by 1000, which scales the coefficient on regressors by the same amount, and the GARCH parameter α_0 by 10^6. With $\alpha_1+\beta_1$ close to unity, it is desirable to estimate the GARCH model imposing the restriction $\alpha_1+\beta_1 < 1$. The results reported below were obtained under that restriction, but, as expected from the block-diagonality of the information matrix, the ARFIMA estimates are not much affected whether or not this restriction is imposed on the GARCH parameters.

Table 14: Parametric bootstrap results for estimators US core inflation in the presence of GARCH(1,1) errors.

	EML		MPL		NLS		NLS-GARCH	
	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
$\widehat{d}$	-0.017	0.057	-0.006	0.055	-0.014	0.06	-0.006	0.04
$\widehat{\phi}_{12}$	-0.040	0.162	-0.027	0.159	-0.065	0.18	-0.123	0.20
$\widehat{\theta}_{12}$	0.033	0.189	0.020	0.188	0.066	0.20	0.132	0.21
Constant	-0.066	1.030	-0.062	1.018	-0.117	1.53	-0.034	1.41
DUM807	0.126	2.086	0.136	2.075	0.129	2.09	0.027	1.89

See last column of Table 2 for the DGP, 1000 replications. Data multiplied by 1000.
In all cases the DGP has GARCH errors: $\alpha_0 = 0.179, \alpha_1 = 0.201, \tilde{\beta}_1 = 0.762$.
Note that the DGP is different from that in Table 8.

The results are presented in Table 14, with corresponding rejection frequencies of t-tests in Table 15. It is clear that standard Wald tests do not work very well for this sample size. Therefore we consider some bootstrap tests on the value of d, see e.g. Davidson and MacKinnon (1999b). The hypothesis of interest is $H_0 : d = 0.5$ in the ARFIMA-GARCH model. For US inflation, unlike the UK, the estimated d is again significantly below 0.5. The parametric bootstrap test procedure is based on the parameters estimated under H_0. Using these estimates, we generate $B = 199$ bootstrap samples, on which the unrestricted model is estimated. Following Davidson and MacKinnon (1999a) we use the H_0-parameters as starting values for the iterative estimation in each bootstrap sample, but unlike Davidson and MacKinnon (1999a) we iterate until convergence. Consequently, we have B t-values on d: $P(|\hat{d}_b - 0.5| < t_b) = 0.05$ for $b = 1,\dots,B$. The relative position of the t-value from the original model within these bootstrapped t-values gives us the bootstrap p-value. For the US

Table 15: Empirical rejection probabilities at 10% and 5% nominal level for two-sided t-tests. DGP with GARCH(1,1) errors. US core inflation

	EML		MPL		NLS		NLS-GARCH	
	10%	5%	10%	5%	10%	5%	10%	5%
$\widehat{d}$	.249	.183	.244	.168	.245	.189	.129	.086
$\widehat{\phi}_{12}$	.153	.085	.167	.111	.152	.098	.317	.224
$\widehat{\theta}_{12}$	.179	.123	.188	.128	.195	.126	.313	.236
Constant	.108	.060	.084	.047	.198	.138	.310	.226
DUM807	.106	.072	.105	.073	.109	.074	.101	.061

See last column of Table 2 for the DGP. 1000 replications.

we find a one-sided bootstrap p value of less than 1%, and for the UK about 50%, confirming the model outcomes of the last columns of Tables 2 and 3.

8.2 Effects of neglecting GARCH errors on forecast inference

Finally, we look at the effect of introducing GARCH errors on the inference regarding forecasts. Baillie and Bollerslev (1992, §4) and Bera and Higgins (1993, §5) derived forecasting formulae for the conditional mean and variance of ARMA-GARCH models. The GARCH model is a nonlinear time series model and the out-of-sample forecast error variance is time-dependent, i.e. it depends on T, here interpreted as the index of the last observation of our sample.

The formula for the conditional variance of the naive optimal $H-$step ahead forecast error of an ARFIMA-GARCH(1,1) is a straightforward extension of (14)

$$\begin{aligned} \mathsf{var}(\tilde{z}_{T+H|T}) &= \sum_{j=0}^{H-1} \psi_j^2 h_{T+j+1|T}, && (18) \\ h_{T+j|T} &= \alpha_0 + (\alpha_1 + \tilde{\beta}_1) h_{T+j-1|T} - \tilde{\beta}_1 E(\nu_{T+j-1|T}), && (19) \\ E(\nu_{T+j-1|T}) &= \nu_T, \qquad j = 1, \\ E(\nu_{T+j-1|T}) &= 0, \qquad j > 1. \end{aligned} \tag{20}$$

Formula (19) is derived by applying (14) to the ARMA representation of the squared innovations ε_t^2,

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{m} \left(\alpha_i + \tilde{\beta}_i\right) \varepsilon_{t-i}^2 - \sum_{i=1}^{s} \beta_i \nu_{t-i} + \nu_t$$

where $m = \max(r, s)$, $\alpha_i = 0, i > s$, $\tilde{\beta}_i = 0, i > r$, $\nu_t = \varepsilon_t^2 - h_t$. In our case $r = s = m = 1$.

As the sample size T and horizon H tend to infinity, forecasts of GARCH processes, ε_t, converge to the unconditional distribution of a GARCH process, which is non-Gaussian in general, even if the innovations ε_t are Gaussian. So far, there are few analytical results available on the distribution of ARFIMA-GARCH forecasts for long horizons. Beran and Feng (1999, Theorem 3), using results by Ling (1999) and Hosking (1996) showed that $T^{\frac{1}{2}-d}$ times the sample mean of a covariance stationary ARFIMA-GARCH process with $d \in (-0.5, 0.5)$ has a Gaussian limiting distribution. Noting the equivalence between inference on the mean of inflation and estimating uncertainty of the long-run mean of future inflation, one may expect a Gaussian distribution for (very) long horizon forecasts of the log price level, as long as inflation is $I(d)$, with $d \in (-0.5, 0.5)$. However, Gaussian distributions for the price level forecasts are not expected for the forecast horizon and sample size that we consider.

We repeat the bootstrap forecast analysis of §7.3 using the ARFIMA-GARCH model for the US series. Again we generate 1000 bootstrap ARFIMA-GARCH samples, where we use the first parts of the samples for estimation and the last parts of the samples for naive forecasting. We estimate both the ARFIMA and the ARFIMA-GARCH model by NLS. We did not impose covariance stationarity under estimation. Equations (18)-(19) are used to compute RMSE-estimates in each bootstrap sample. Note that we do not condition on the last observations of our actual observations: h_T and ν_t vary across bootstrap samples. The bootstrap forecast error distribution should not be interpreted as an approximation to the conditional forecast error distribution, as in e.g. Geweke (1989). We present the empirical RMSEs of the actual forecast errors across all bootstrap samples in Table 16 and compare these with the empirical distribution of the RMSE-estimates for which we present the mean and the RMSE. This allows us to examine the effect of parameter uncertainty on ARFIMA-GARCH forecast interval estimates.

Table 16: Parametric Bootstrap Forecast results for US log-price level up to 24 months ahead using ARFIMA-GARCH DGP.

	NLS (naive)			NLS-GARCH (naive)		
	RMSE	$\sqrt{\widehat{\mathsf{var}}(\tilde{z}_{T+H})}$		RMSE	$\sqrt{\widehat{\mathsf{var}}(\tilde{z}_{T+H})}$	
h		mean	RMSE		mean	RMSE
1	0.0220	0.0206	0.00539	0.0204	0.0213	0.00627
2	0.0355	0.0338	0.00893	0.0342	0.0363	0.01046
3	0.0479	0.0458	0.01227	0.0469	0.0493	0.01438
12	0.1435	0.1354	0.04081	0.1396	0.1468	0.04605
24	0.2766	0.2609	0.08655	0.2870	0.2846	0.09404

1000 replications. All results multiplied by 10.
See Table 2 for the DGP, Fig. 12 for the corresponding estimates of densities of forecast errors and forecast RMSE estimators.
See Table 12 for corresponding results for the homoskedastic DGP.

Comparing Table 16 with Table 12 we see that the mean of the forecast RMSE and the mean of the corresponding estimated forecast RMSE do not differ too much between the ARFIMA-DGP and the ARFIMA-GARCH-DGP, which is encouraging. The RMSEs for the ARFIMA-GARCH-DGP of Table 16 are slightly higher for short horizons, probably owing to the conditional heteroskedasticity, and slightly lower for long horizons, owing to the lower d in the ARFIMA-GARCH-DGP. The largest difference between the results of Table 16 and Table 12 occurs for the RMSE of the estimator of the RMSE. Naturally, the RMSE for the RMSE estimator for short run forecasts error is much larger, i.e nearly 10 times as large, in the GARCH case. The difference becomes much smaller for longer run forecasts, where the RMSE for the RMSE-estimator is only 3 times as large. Note that the parameters of the GARCH part are close to those of an IGARCH-model, leading to very large empirical variances of error variances.

Since the moment estimates in Table 16 correspond to distributions which are not expected to be Gaussian, we also present some of their estimated densities in Figure 12. The first column of plots corresponds to the last row of the last three columns of Table 12, the last two columns of plots correspond to Table 16. The most striking difference is in the scale and the shape in the distributions of the forecast-RMSE estimator. This distribution is much more spread out under the ARFIMA-GARCH DGP, with occasional outliers. Estimating and forecasting using the extra GARCH part leads to a larger variation under this DGP. The parameter uncertainty in the GARCH model and the conditional heteroskedasticity (even for 24-period ahead forecasts) add considerable variation to the forecast-RMSE estimator.

Note that the mean of the 3 bottom row of distributions in Figure 12 is approximately the same, namely 0.26, corresponding to a 2-year ahead standard forecast error of 2.6%.

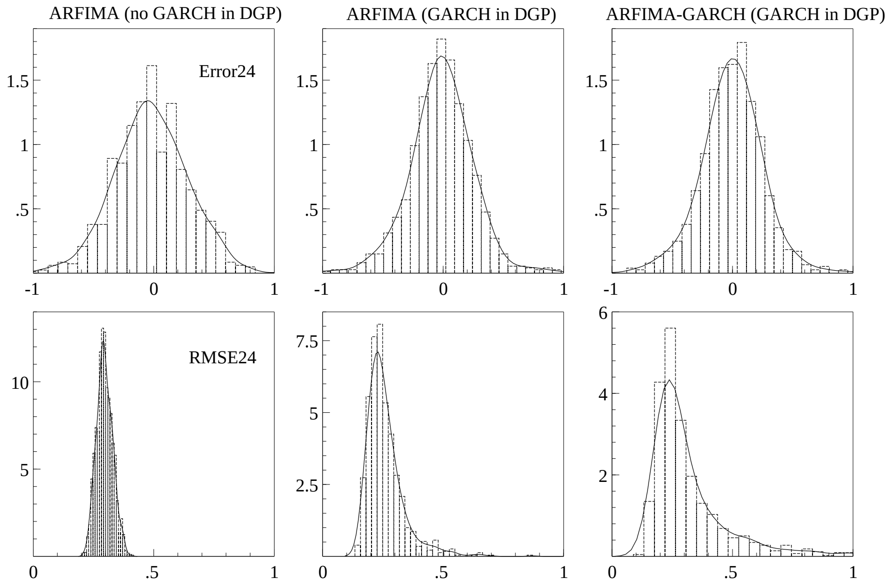

Figure 12: Estimates of parametric bootstrap densities, under ARFIMA-DGP (first column) and under ARFIMA-GARCH DGP, of the 2 year ahead (unconditional) forecast error for the US log price level, for the NLS and the NLS-GARCH estimators, and estimates of the density of the corresponding estimators of root mean square error of 2 year ahead forecasts. All forecast errors are multiplied by 10. Also see Table 12 and Table 16.

9 Conclusions

We considered several practical issues when using ARFIMA models to forecast inflation. We compared three methods to estimate Gaussian ARFIMA models: exact maximum likelihood (EML), modified profile likelihood (MPL), and nonlinear least squares (NLS). We discussed computation of the exact and modified profile likelihood. An adaptation of Sowell's method to compute the autocovariance function is the only method that is sufficiently accurate and fast to perform bootstrap analysis for the likelihood based methods.

For models that are relevant for postwar monthly US core inflation and for quarterly overall UK consumer price inflation, it was shown that MPL is clearly the preferred method of inference. A parametric bootstrap analysis revealed that it is both more reliable and more efficient. Inference on the integration parameter d is especially satisfactory. However, currently available (higher order) asymptotic approximations did not lead to reliable inference on the (trending) mean of inflation.

We also compared optimal and naive forecasts. For the sample sizes at hand the difference between the two forecasting methods turned out to be of little importance. The bootstrap forecast analysis allowed us to estimate the uncertainty in root mean-squared estimates of the conditional long-range forecasts errors for the overall price level. This uncertainty is mainly influenced by the estimation uncertainty on the fractional parameter d. We extended the bootstrap analysis to allow for GARCH errors, which were found to be empirically relevant.

Although some of the asymptotic results seem to give a good guidance for finite sample inference in our model, we recommend the application of parametric bootstrap tests. Not only can bootstrap tests be expected to provide more accurate inference, the bootstrap samples also deliver additional evidence on the adequacy of parameterizations, estimators, and corresponding Wald tests for the parameters and forecasts of ARFIMA processes for inflation series. This was illustrated in §6.3 and §6.4.

The parametric bootstrap analysis of the estimators of the multistep-forecast RMSE shows these to have nonneglible variation in empirically relevant cases. This analysis is an important step in further efforts to take the uncertainty about the fractional order of integration into account in the construction of multi-step-forecast intervals in ARFIMA modelling.

References

Agiakloglou, C. and P. Newbold (1994). Langrange multiplier tests for fractional difference. *Journal of Time Series Analysis 15*, 253–262.

An, S. and P. Bloomfield (1993). Cox and Reid's modification in regression models with correlated errors. Discussion paper, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, U.S.A.

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

Baillie, R. T. and T. Bollerslev (1992). Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics 52*, 91–113.

Baillie, R. T., C.-F. Chung, and Tieslau (1996). Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of Applied Econometrics 11*, 23–40.

Barndorff-Nielsen, O. E. and D. R. Cox (1994). *Inference and Asymptotics.* London, UK: Chapman and Hall.

Bera, A. K. and M. L. Higgins (1993). ARCH models: Properties, estimation and testing. *Journal of Economic Surveys 7*, 305–362.

Beran, J. (1992). Statistical methods for data with long-range dependence. *Statistical Science 7*, 404–416.

Beran, J. (1994). *Statistics for Long-Memory Processes.* London: Chapman and Hall.

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 57*, 659–672.

Beran, J. and Y. Feng (1999). Local polynomial estimation with a FARIMA-GARCH error process. Discussion Paper 99/08, Center of Finance and Econometrics, University of Konstanz.

Beveridge, S. and C. Oickle (1993). Estimating fractionally integrated time series models. *Economics Letters 43*, 137–142.

Bollerslev, T. and D. Jubinski (1999). Equity trading volume and volatility: Latent information arrivals and common long-run dependencies. *Journal of Business & Economic Statistics 17*, 9–21.

Bos, C. S., P. H. B. F. Franses, and M. Ooms (1999). Long memory and level shifts: Reanalyzing inflation rates. *Empirical Economics 24*, 427–449.

Brockwell, P. J. and R. A. Davis (1993). *Time Series: Theory and Methods (2nd ed.).* USA: Springer-Verlag, New-York.

Chan, N. H. and W. Palma (1998). State space modeling of long-memory processes. *Annals of Statistics 26*, 719–740.

Chan, N. H. and N. Terrin (1995). Inference for unstable long-memory processes with applications to fractional unit root regressions. *The Annals of Statistics 23*, 1662–1683.

Chan, N. H. and C. Z. Wei (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics 16*(1), 367–401.

Chang, M. C. and D. A. Dickey (1994). Recognizing overdifferenced time series. *Journal of Time Series Analysis 15*, 1–18.

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.

Cox, D. R. and N. Reid (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society Series B 49*, 1–39.

Cox, D. R. and N. Reid (1993). A note on the calculation of adjusted profile likelihood. *Journal of the Royal Statistical Society, Series B 49*, 467–471.

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

Dahlhaus, R. (1995). Efficient location and regression estimation for long range dependent regression models. *The Annals of Statistics 23*, 1029–1047.

Davidson, R. and J. MacKinnon (1999a). Bootstrap testing in nonlinear models. *International Economic Review 40*, 487–508.

Davidson, R. and J. MacKinnon (1999b). The size distortion of bootstrap tests. *Econometric Theory 15*, 361–376.

Delgado, M. A. and P. M. Robinson (1993). New methods for the analysis of long-memory time-series: Application to Spanish inflation. *Journal of Forecasting 13*, 97–107.

Doornik, J. A. (1998). *Object-Oriented Matrix Programming using Ox 2.0.* London: Timberlake Consultants Press.

Doornik, J. A. and H. Hansen (1994). A practical test for univariate and multivariate normality. Discussion paper, Nuffield College.

Doornik, J. A. and D. F. Hendry (1999). *GiveWin: An Interface to Empirical Modelling* (2nd ed.). London: Timberlake Consultants Press.

Doornik, J. A. and M. Ooms (1999). A package for estimating, forecasting and simulating ARFIMA models: Arfima package 1.0 for Ox. Discussion paper, Nuffield College, Oxford.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1008.

Geweke, J. (1989). Exact predictive densities for linear models with ARCH disturbances. *Journal of Econometrics 40*, 63–88.

Gil-Alaña, L. A. and P. M. Robinson (1997). Testing of unit root and other nonstationary hypotheses in macroeconomic time series. *Journal of Econometrics 80*, 241–268.

Golub, G. H. and C. F. Van Loan (1989). *Matrix Computations.* Baltimore: The Johns Hopkins University Press.

Hansen, H. and A. Rahbek (1998). Approximate conditional unit root inference. Discussion Paper Preprint no. 8, Institute of Theoretical Statistics, University of Copenhagen, `http://www.stat.ku.dk`.

Harvey, A. C. (1993). *Time Series Models, second edition.* New York, London: Harvest Wheatsheaf.

Harvey, A. C. and S. J. Koopman (1992). Diagnostic checking of unobserved components time series models. *Journal of Business and Economic Statistics 10*, 377–389.

Hassler, U. (1997). Sample autocorrelations of nonstationary fractionally integrated series. *Statistical Papers 38*, 43–62.

Hassler, U. and J. Wolters (1994). On the power of unit root tests against fractional alternatives. *Economics Letters 45*, 1–6.

Hassler, U. and J. Wolters (1995). Long memory in inflation rates: International evidence. *Journal of Business and Economic Statistics 13*, 37–46.

Hauser, M. A. (1999). Maximum likelihood estimators for ARMA and ARFIMA models: A Monte Carlo study. *Journal of Statistical Planning and Inference 80*, 229–255.

Hendry, D. F. and J. A. Doornik (1999). *Empirical Econometric Modelling using PcGive: Volume I.* London: Timberlake Consultants Press.

Hobijn, B., P. H. Franses, and M. Ooms (1998). Generalizations of the KPSS test. Discussion Paper 9802, Econometric Institute, Erasmus University Rotterdam.

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.

Jeganathan, P. (1999). On asymptotic inference in cointegrated time series with fractionally integrated errors. *Econometric Theory 15*, 583–621.

Kiviet, J. F. and G. D. A. Phillips (1998). Degrees of freedom adjustment for variance estimators in dynamic regression models. *The Econometrics Journal 1*, 44–70.

Laskar, M. R. and M. L. King (1998). Modified likelihood and related methods for handling nuisance parameters in the linear regression model. Discussion Paper 5/98, Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3168, Australia.

Lee, D. and P. Schmidt (1996). On the power of the KPSS test of stationarity against fractionally-integrated alternatives. *Journal of Econometrics 73*, 285–302.

Li, W. K. and A. I. McLeod (1986). Fractional time series modelling. *Biometrika 73*, 217–221.

Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *Journal of Applied Probability 36*, 301–313.

Ling, S. and W. K. Li (1997). On Fractionally Integrated Autoregressive Moving-Average models with Conditional Heteroskedasticity. *Journal of the American Statistical Association 92*, 1184–1194.

Ling, S. and W. K. Li (1999). Asymptotic inference for nonstationary fractionally integrated autoregressive moving-average models. Discussion paper, Department of Economics, University of Western Australia.

Martin, V. and N. Wilkins (1999). Indirect estimation of ARFIMA and VARFIMA models. *Journal of Econometrics 93*, 149–175.

Odaki, M. (1993). On the invertibility of fractionally differenced ARIMA processes. *Biometrika 80*, 703–709.

Ooms, M. (1996). Long memory and seasonality in U.S. consumer price inflation: An empirical investigation at varying levels of aggregation. Discussion paper, Econometric Institute, Erasmus University Rotterdam, presented at North American Winter Meeting Econometric Society, New Orleans, January 1997.

Ooms, M. and U. Hassler (1997). On the effect of seasonal adjustment on the log-periodogram regression. *Economics Letters 56*, 135–141.

Pai, J. S. and N. Ravishanker (1996). Exact likelihood function forms for an ARFIMA process. In *Data Analysis and Information Systems, Statistical and Conceptual Approaches. Proceedings of the 19th Annual Conference of the Gesellschaft für Klassifikation e.V., University of Basel, March 8-10, 1995.*, pp. 323–331. Springer Verlag, Berlin.

Piessens, R., E. de Donker-Kapenga, C. W. Überhuber, and D. K. Kahaner (1983). *QUADPACK, A Subroutine Package for Automatic Integration*. Heidelberg: Springer-Verlag.

Ravishanker, N. and B. K. Ray (1997). Bayesian analysis of vector ARFIMA processes. *Australian Journal of Statistics 39*, 295–312.

Robinson, P. (1994). Time series with strong dependence. In C. A. Sims (Ed.), *Advances in Econometrics, Sixth World Congress*, pp. 47–95. Cambridge: Cambridge University Press.

Robinson, P. M. and M. Henry (1999). Long and short memory conditional heteroskedasticity in estimating the memory parameter of levels. *Econometric Theory 15*, 299–336.

Scott, A. (1995). Why is consumption so seasonal? Discussion paper, Institute of Economics and Statistics, Oxford University, Oxford, OX1 3UL, UK.

Smith Jr, A. A., F. Sowell, and S. E. Zin (1997). Fractional integration with drift: Estimation in small samples. *Empirical Economics 22*, 103–116.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics 53*, 165–188.

Stock, J. H. (1994). Unit roots, structural breaks and trends. In *Handbook of Econometrics, Volume IV*, pp. 2739–2841. Elsevier Science.

Tunnicliffe-Wilson, G. (1989). On the use of marginal likelihood in time series model estimation. *J. Royal Statistical Society B 51*, 15–27.