BIAS IN AN ESTIMATOR OF THE FRACTIONAL DIFFERENCE PARAMETER

By Christos Agiakloglou, Paul Newbold and Mark Wohar

University of Illinois and University of Nebraska

First version received May 1991

Abstract. An estimator of the difference parameter in a class of long-memory time series models is examined. It is shown that, in particular circumstances, the estimator can be badly biased, and tests based on it consequently seriously misleading. The source of this bias is identified, and it is shown that its magnitude can readily be predicted through straightforward analytical arguments.

Keywords. Fractional difference; long memory; periodogram.

1. INTRODUCTION

Over the last few years there has been considerable interest in the application of long-memory time series models in a number of fields, including hydrology and economics. Let X_t be an observed time series, and consider the model

$$\phi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t \tag{1.1}$$

where ε_t is white noise. In (1.1), d is not necessarily an integer, so that the possibility of fractional differencing is allowed. In this case, d can be viewed as another model parameter, and issues of estimation and testing for the parameter arise. Difficulties occur since, in practice, the structure of the autoregressive moving-average (ARMA) part of the generating model will be unknown.

Geweke and Porter-Hudak (1983) proposed an attractively simple nonparametric procedure for the estimation of the fractional difference parameter d. The GPH approach follows from the fact that the spectral density of X_t is

$$f_x(\lambda) = \{4\sin^2(\lambda/2\}^{-d} f_u(\lambda)$$

where $f_u(\lambda)$ is the spectral density of $u_t = (1 - B)^d X_t$. Taking logarithms and evaluating at the harmonic ordinates $\lambda_j = 2\pi j/T$, where T is the sample size, then yields

$$\ln\{f_x(\lambda_j)\} = \ln\{f_u(0)\} - d\ln\{4\sin^2(\lambda_j/2)\} + \ln\left\{\frac{f_u(\lambda_j)}{f_u(0)}\right\}. \tag{1.2}$$

The crucial step in the development of the GPH estimator of d follows from the assertion that, over sufficiently low frequency ordinates, the last term in

(1.2) is negligible, or at least approximately constant. Then, if $I(\lambda_j)$ denotes the periodogram at ordinate λ_j, we can write approximately

$$\ln\{I(\lambda_j)\} = \ln\{f_u(0)\} - d\ln\{4\sin^2(\lambda_j/2)\} + \ln\left\{\frac{I(\lambda_j)}{f_x(\lambda_j)}\right\}.$$

This suggests the regression formulation

$$\ln\{I(\lambda_j)\} = \alpha + \beta\ln\{4\sin^2(\lambda_j/2)\} + e_j, \qquad j = 1, \ldots, K, \tag{1.3}$$

where $\beta = -d$ and the error terms e_j are independent with zero mean and variance $\pi^2/6$.

This suggests a simple ordinary least squares (OLS) estimator for the fractional difference parameter d. Tests and confidence intervals for that parameter can be based on the usual regression statistics, or the prior knowledge that the variance of the error terms is $\pi^2/6$ can be imposed. Of course, this simple estimator is possible as a result of dropping the last term in (1.2), which involves the unknown autoregressive and moving-average parameters in (1.1). Geweke and Porter-Hudak establish consistency and asymptotic normality of the estimator provided that in (1.3) the number of ordinates is taken as $K = g(T)$, where $\lim g(T) = \infty$ and $\lim g(T)/T = 0$. In practical applications, the usual prescription has been to take $K = T^{1/2}$, and we shall follow that convention in the remainder of this paper.

The GPH procedure has been applied to a collection of economic time series by Diebold and Rudebusch (1989) and extended to the case of seasonal time series by Porter-Hudak (1990). In the typical application, the sample size is in the neighborhood of 100 observations, so that the regression (1.3) will be based on about 10 values. This suggests that the standard error of the estimator of d will be quite large. (It is in fact 0.29 for $K = 10$, when an error variance of $\pi^2/6$ is assumed in the regression (1.3)). What seems to be less well understood is the potential for bias in this estimator in moderate sample sizes when the ARMA structure in (1.1) does not justify the dropping of the last term in (1.2). In the remainder of this paper we will analyze this bias and its consequences.

2. ESTIMATION IN MODERATE-SIZED SAMPLES

To explore the behavior of the GPH estimator of the fractional difference parameter, we generated series of 100 observations from the first-order autoregressive model

$$(1 - \phi B)X_t = \varepsilon_t$$

so that the true value is $d = 0$. We estimated the regression (1.3) with $K = 10$. Results obtained over 1000 replications are set out in Table I. The most striking feature of this table is that, while the empirical standard deviations of the estimators are remarkably close to the asymptotic value of

TABLE I

SIMULATION RESULTS ON ESTIMATION AND TESTING FOR d FOR SERIES OF 100 OBSERVATIONS FROM AR(1) PROCESS (1000 REPLICATIONS)

	$\hat{d}$		Rejections of $d = 0$			
			$\pi^2/6$		OLS	
ϕ	Mean	Std dev.	$d < 0$	$d > 0$	$d < 0$	$d > 0$
0.9	0.71	0.30	0	702	0	707
			1	788	0	785
0.8	0.46	0.28	1	344	0	397
			2	473	0	510
0.6	0.18	0.30	19	70	18	104
			26	135	31	178
0.4	0.09	0.29	15	43	34	58
			30	77	51	107
0.2	0.03	0.29	30	21	47	42
			52	34	71	73
0	−0.01	0.30	40	22	58	27
			63	41	75	53
−0.2	−0.01	0.29	36	19	47	31
			59	37	75	47
−0.4	−0.03	0.30	40	17	73	30
			65	37	103	54
−0.6	−0.01	0.30	36	19	59	28
			61	46	87	56
−0.8	−0.02	0.28	38	7	55	21
			60	30	78	39
−0.9	−0.01	0.29	43	14	58	28
			58	33	96	60

Tests rejections against the indicated one-sided alternatives are respectively for nominal 2.5% and 5% significance levels.

0.29, there is a very substantial bias in the estimator of d for large positive values of the autoregressive parameter. This bias in turn renders tests of the null hypothesis that d is zero quite unreliable, whether the variance of the error term in (1.3) is set at its asymptotic value of $\pi^2/6$ or estimated from the OLS regression. For large positive ϕ, the null hypothesis $d = 0$ will very often be rejected against the alternative that d is positive at conventional significance levels. For other values of ϕ, it seems preferable to set the error variance at $\pi^2/6$ in carrying out these tests.

The cause of the bias in the estimator of the fractional difference parameter, noted in Table I, is clear. For large positive values of the autoregressive parameter, the log spectrum is not approximately constant in the relevant range, but strongly downward sloping. In fact, the bias is readily predicted on theoretical grounds. Directly from (1.3), the expected value of the least squares estimator of $d = -\beta$ is

$$E(\hat{d}) = \frac{\sum_{j=1}^{K}(Z_j - \bar{Z})E[\ln\{I(\lambda_j\}]}{\sum_{j=1}^{K}(Z_j - \bar{Z})^2} \tag{2.1}$$

where

$$Z_j = -\ln\{4\sin^2(\lambda_j/2)\}.$$

Evaluation of (2.1) requires an expression for the expectation of the logarithm of the periodogram. To obtain such an expression precisely is tedious, and we shall see that, for moderately large sample sizes, sufficiently accurate results can often be obtained from readily computed approximations. To obtain an approximation to (2.1), we can approximate the expectation of the logarithm of the periodogram by the logarithm of the spectrum or, with a modest additional effort, by the logarithm of the expected value of the periodogram, up to an additive constant. Thus we can set

$$E[\ln\{I(\lambda_j)\}] \doteq \ln\{f(\lambda_j)\} + \text{constant} \tag{2.2}$$

or

$$E[\ln\{I(\lambda_j)\}] \doteq \ln[E\{I(\lambda_j)\}] + \text{constant.} \tag{2.3}$$

In the case of the AR(1) model, (2.2) gives

$$E[\ln\{I(\lambda_j)\}] \doteq -\ln(1 - 2\phi\cos\lambda_j + \phi^2) + \text{constant}$$

and (2.3) leads to

$$E[\ln\{I(\lambda_j)\}] \doteq \ln\left\{1 + 2T^{-1}\sum_{t=1}^{T-1}(T-t)\phi^t\cos\lambda_j t\right\} + \text{constant.}$$

Substitution of these values in (2.1) then yields readily computed estimates of the bias in $\hat{d}$. Some values are shown in Table II for the case where the given time series follows a first-order autoregressive process. In the case of the first-order autoregressive generating model, both theoretical approximations are very close indeed to the empirical means of the sampling distributions of the estimators of the fractional difference parameter. In particular, our empirical observation that bias is a serious problem only for large *positive* values of the autoregressive parameter is predicted by this simple theory. Note also that the first of these theoretical approximations predicts the same

TABLE II

EXPECTED VALUES OF ESTIMATORS OF d FOR SERIES OF 100 OBSERVATIONS FROM AR(1) PROCESS

ϕ	$\ln(f)$	$\ln\{E(I)\}$	Sim.	ϕ	$\ln(f)$	$\ln\{E(I)\}$	Sim.
0.9	0.74	0.72	0.71	−0.1	−0.01	−0.01	
0.8	0.48	0.46	0.46	−0.2	−0.01	−0.01	−0.01
0.7	0.30	0.29		−0.3	−0.01	−0.01	
0.6	0.19	0.19	0.18	−0.4	−0.02	−0.02	−0.03
0.5	0.12	0.12		−0.5	−0.02	−0.02	
0.4	0.07	0.07	0.09	−0.6	−0.02	−0.02	−0.01
0.3	0.04	0.04		−0.7	−0.02	−0.02	
0.2	0.02	0.02	0.03	−0.8	−0.02	−0.02	−0.02
0.1	0.01	0.01		−0.9	−0.02	−0.02	−0.01

amount of bias whatever the true value of d, since the logarithm of the spectrum is simply a linear function of d.

Since the approximation based on (2.2) is reliable for samples of 100 observations, it seems likely that it will continue to be so for larger sample sizes. This permits easy assessment of the extent to which the problem of bias diminishes with increasing sample size. In Table III, where all calculations are based on taking the number of ordinates K in (1.3) equal to the square root of the sample size, we see that the bias decreases only very slowly as the sample size increases. Thus, for time series generated by an autoregressive process with parameter 0.9, for series of 900 observations the GPH estimator of d has mean 0.27, so that the consistency of this estimator is of cold comfort unless the sample size is very large indeed.

The GPH procedure can be used to test the null hypothesis $d = 0$ and, as our simulations in Table I indicate, this hypothesis will frequently be rejected for a first-order autoregressive process with large positive parameter. These findings can readily be predicted on theoretical grounds. Assuming that the estimator $\hat{d}$ is approximately normal, with mean derived from (2.2) and a standard deviation derived from setting the error variance in (1.3) to $\pi^2/6$ (which agrees well with the simulated values in Table I), probabilities for rejection of the null hypothesis follow from elementary calculations. Results for various sample sizes are shown for large positive ϕ in Table IV. Notice that, for $T = 100$, these approximate theoretical values agree well with the simulation results in Table I. Of course, as the sample size increases, the bias of $\hat{d}$ decreases, but so also does the standard error of the estimator. Hence,

TABLE III

EXPECTED VALUES OF ESTIMATORS OF d, BASED ON (2.2), FOR SERIES GENERATED BY AR(1) PROCESS

ϕ	$T = 100$	$T = 225$	$T = 400$	$T = 625$	$T = 900$
0.9	0.74	0.56	0.43	0.34	0.27
0.8	0.48	0.29	0.20	0.14	0.10
0.7	0.30	0.16	0.10	0.06	0.05
0.6	0.19	0.09	0.05	0.03	0.02
0.5	0.12	0.05	0.03	0.02	0.01
0.4	0.07	0.03	0.02	0.01	0.01
0.3	0.04	0.02	0.01	0.01	0.00
0.2	0.02	0.01	0.00	0.00	0.00
0.1	0.01	0.00	0.00	0.00	0.00
−0.1	−0.01	−0.00	−0.00	−0.00	−0.00
−0.2	−0.01	−0.00	−0.00	−0.00	−0.00
−0.3	−0.01	−0.01	−0.00	−0.00	−0.00
−0.4	−0.02	−0.01	−0.00	−0.00	−0.00
−0.5	−0.02	−0.01	−0.00	−0.00	−0.00
−0.6	−0.02	−0.01	−0.00	−0.00	−0.00
−0.7	−0.02	−0.01	−0.00	−0.00	−0.00
−0.8	−0.02	−0.01	−0.00	−0.00	−0.00
−0.9	−0.02	−0.01	−0.00	−0.00	−0.00

TABLE IV

PROBABILITY OF REJECTING $d = 0$ AGAINST $d > 0$, AT 5% SIGNIFICANCE LEVEL BASED ON (2.2), FOR SERIES GENERATED BY AR(1) PROCESS

ϕ	$T = 100$	$T = 225$	$T = 400$	$T = 625$	$T = 900$
0.9	0.811	0.816	0.767	0.699	0.622
0.8	0.492	0.301	0.284	0.221	0.176
0.7	0.271	0.181	0.136	0.107	0.093
0.6	0.161	0.111	0.088	0.076	0.069

for large ϕ, the null hypothesis $d = 0$ will be rejected very often, even for long time series.

It should not be surprising that we should encounter difficulties with the GPH estimator, or any nonparametric estimator of d, for series generated by a first-order autoregressive process for large positive values of the autoregressive parameter. After all, for any value of ϕ strictly less than unity, the true d is zero, whereas $\phi = 1$ corresponds to $d = 1$. Nevertheless, the serious bias found here for $\phi \geqslant 0.7$, and the consequent unreliability of the associated hypothesis tests, renders the GPH procedure of little value unless the sample size is very large indeed.

In fact, exactly the same difficulty can arise even when the true generating process has no autoregressive component. That it should do so for series generated by a first-order moving-average process can be predicted from the approximation (2.2) since, apart from a multiplicative constant, the spectral density of an MA(1) process is the inverse of the spectral density of an AR(1) process. We generated series of 100 observations from the first-order moving-average process

$$X_t = \varepsilon_t - \theta\varepsilon_{t-1} \tag{2.4}$$

This is of some practical interest, since the GPH procedure is often applied to the first differences of economic time series, for which the model (2.4) with positive θ often appears to provide an adequate fit. The results of our simulation experiments are shown in Table V. It can be seen that the estimator of d is seriously biased, and tests based on this estimator are consequently unreliable, for large positive values of the moving-average parameter θ. For these values, the null hypothesis $d = 0$ is frequently rejected against the alternative that d is negative at low significance levels. Again, this difficulty is not encountered for negative values of that parameter.

As before, it is straightforward to use (2.1) to obtain theoretical approximations to the expected value of the estimators of d, based on (2.2) and (2.3). Specifically, for the MA(1) model, (2.2) gives

$$E\{\ln\{I(\lambda_j)\}] \doteq \ln(1 - 2\theta\cos\lambda_j + \theta^2) + \text{constant}$$

and (2.3) leads to

TABLE V

SIMULATION RESULTS ON ESTIMATION AND TESTING FOR d FOR SERIES OF 100 OBSERVATIONS FROM MA(1) PROCESS (1000 REPLICATIONS)

	$\hat{d}$		Rejections of $d = 0$			
			$\pi^2/6$		OLS	
θ	Mean	Std dev.	$d < 0$	$d > 0$	$d < 0$	$d > 0$
0.9	−0.60	0.32	510	0	580	0
			639	0	667	0
0.8	−0.42	0.30	275	2	345	1
			394	4	435	1
0.6	−0.18	0.29	80	3	140	8
			134	11	201	11
0.4	−0.07	0.28	50	4	87	16
			84	17	115	25
0.2	−0.02	0.30	37	16	54	39
			60	33	76	59
−0.2	0.00	0.29	29	16	52	25
			54	44	74	52
−0.4	0.02	0.28	20	22	48	37
			38	40	66	70
−0.6	0.02	0.30	33	22	52	39
			53	43	68	71
−0.8	0.01	0.31	30	28	55	43
			56	46	80	66
−0.9	0.02	0.29	24	18	35	41
			45	43	58	64

Tests rejections against the indicated one-sided alternatives are respectively for nominal 2.5% and 5% significance levels.

$$E[\ln\{I(\lambda_j)\}] \doteq \ln\{1 - 2T^{-1}(T - 1)\theta\cos\lambda_j + \theta^2\} + \text{constant}$$

The results, shown in Table VI, differ somewhat from the first-order autoregressive case. While both approximations to the mean of the sampling distribution of the GPH estimator of d predict substantial downward bias for large positive θ, the values based on (2.2) differ somewhat from the

TABLE VI

EXPECTED VALUES OF ESTIMATORS OF d FOR SERIES OF 100 OBSERVATIONS FROM MA(1) PROCESS

θ	$\ln(f)$	$\ln\{E(I)\}$	Sim.	θ	$\ln(f)$	$\ln\{E(I)\}$	Sim.
0.9	−0.74	−0.57	−0.60	−0.1	0.01	0.01	
0.8	−0.48	−0.41	−0.42	−0.2	0.01	0.01	0.00
0.7	−0.30	−0.28		−0.3	0.01	0.01	
0.6	−0.19	−0.18	−0.18	−0.4	0.02	0.02	0.02
0.5	−0.12	−0.12		−0.5	0.02	0.02	
0.4	−0.07	−0.07	−0.07	−0.6	0.02	0.02	0.02
0.3	−0.04	−0.04		−0.7	0.02	0.02	
0.2	−0.02	−0.02	−0.02	−0.8	0.02	0.02	0.01
0.1	−0.01	−0.01		−0.9	0.02	0.02	0.02

simulation results in that case. (They are, of course, just the negatives of the entries in the corresponding column of Table II.) Intuitively, the poorness of the approximation comes about because, for large θ, the spectral density of an MA(1) process is close to zero for low frequencies. Then on taking logarithms, the small bias in the periodogram as an estimator of the spectrum is exaggerated. The theoretical approximation based on (2.3) does yield values that are satisfactorily close to the simulation results.

Given the superiority here of the approximation (2.3), we shall use it to explore the behavior of the bias in the estimator of d as the sample size increases. The results of these calculations are set out in Table VII. As in the autoregressive case, where there is serious bias, it decreases only very slowly as the sample size increases. Of course, the fact that the bias in the GPH estimator is somewhat smaller in the MA(1) case than in the AR(1) case is entirely fortuitous. If an unbiased estimator for the log spectrum could be employed in place of the log periodogram as the dependent variable in (1.3), the magnitudes of the biases in the MA(1) case would be the same as those for the AR(1) case (as given in Table III) as follows by inserting such an estimator, for which (2.2) would be an equality, in (2.1).

In Table VIII we give theoretical probabilities for rejection of the null hypothesis $d = 0$ against the alternative that the difference parameter is negative. These results also are based on the approximation (2.3). Again, for large values of the moving-average parameter, the probabilities of misleading test conclusions remain disturbingly high, even for large sample sizes.

TABLE VII

EXPECTED VALUES OF ESTIMATORS OF d, BASED ON (2.3), FOR SERIES GENERATED BY MA(1) PROCESS

θ	$T = 100$	$T = 225$	$T = 400$	$T = 625$	$T = 900$
0.9	−0.57	−0.37	−0.26	−0.19	−0.14
0.8	−0.41	−0.24	−0.15	−0.11	−0.08
0.7	−0.28	−0.14	−0.09	−0.06	−0.04
0.6	−0.18	−0.09	−0.05	−0.03	−0.02
0.5	−0.12	−0.05	−0.03	−0.02	−0.01
0.4	−0.07	−0.03	−0.02	−0.01	−0.01
0.3	−0.04	−0.02	−0.01	−0.01	−0.00
0.2	−0.02	−0.01	−0.00	−0.00	−0.00
0.1	−0.01	−0.00	−0.00	−0.00	−0.00
−0.1	0.01	0.00	0.00	0.00	0.00
−0.2	0.01	0.00	0.00	0.00	0.00
−0.3	0.01	0.01	0.00	0.00	0.00
−0.4	0.02	0.01	0.00	0.00	0.00
−0.5	0.02	0.01	0.00	0.00	0.00
−0.6	0.02	0.01	0.00	0.00	0.00
−0.7	0.02	0.01	0.00	0.00	0.00
−0.8	0.02	0.01	0.00	0.00	0.00
−0.9	0.02	0.01	0.00	0.00	0.00

TABLE VIII

PROBABILITY OF REJECTING $d = 0$ AGAINST $d < 0$, AT 5% SIGNIFICANCE LEVEL BASED ON (2.3), FOR SERIES GENERATED BY MA(1) PROCESS

θ	$T = 100$	$T = 225$	$T = 400$	$T = 625$	$T = 900$
0.9	0.606	0.520	0.421	0.330	0.264
0.8	0.405	0.291	0.215	0.163	0.136
0.7	0.242	0.161	0.121	0.098	0.085
0.6	0.151	0.106	0.085	0.073	0.067

3. EXAMPLES

In the previous section we noted serious bias in the GPH estimator of d for models with large positive autoregressive or moving-average parameters. In practical data analysis, these are precisely the cases where it is difficult to establish with much confidence the appropriate degree of differencing, i.e. it is difficult to distinguish between $d = 0$ and $d = 1$, even disregarding intermediate possibilities. If an AR(1) model is fitted and a large positive parameter estimate results, there is the possibility that differencing is needed. On the other hand, if a large positive value is estimated for the parameter of an MA(1) model, the analyst might suspect over-differencing. It may then be that, when GPH estimates between 0 and 1 have been found, these are signals of the difficulty in distinguishing between the extremes $d = 0$ and $d = 1$ rather than indicating the presence of an intermediate value. We do not intend to doubt the existence of generating processes with fractional d, but we doubt the ability of the GPH procedure to detect them.

Here we shall consider two of the time series analyzed, through the GPH approach, by Diebold and Rudebusch (1989). In all cases these authors began by taking first differences, applying the regression (1.3) to the periodogram of the differenced data. Thus, if the notation (1.1) is preserved for the original series, the parameter directly estimated is $d^* = d - 1$.

As a first example, we analyze a series of 158 observations on unemployment in the United States. Applying the regression (1.3), with $K = 12$, to the series of first differences yields the estimate $\hat{d}^* = -0.26$, with standard error 0.26. This might be taken as modest, but hardly compelling, evidence against the hypothesis $d^* = 0$; i.e. $d = 1$. (It is interesting to note a regression not reported by Diebold and Rudebusch. Application of (1.3) to the undifferenced series yields $\hat{d} = 0.9997$.) In fact, there is some disagreement among economists as to whether the appropriate value for d is 0 or 1 if a traditional autoregressive integrated moving-average (ARIMA) model is to be fitted to unemployment data. The balance of opinion in the profession appears to favor stationarity. We fitted to these unemployment data ARIMA(p, 0, q) and ARIMA(p, 1, q) models. In both cases, all models with $p + q \leq 6$ were tried, estimation was through full maximum likelihood using SPSS, and within the two sets a model was selected through the SBC criterion. The two estimated models that resulted were

$$(1 - 1.54B + 0.60B^2)(X_t - \mu) = \varepsilon_t \tag{3.1}$$

and

$$(1 - 1.48B + 0.62B^2)(1 - B)X_t = (1 - 0.89B)\varepsilon_t \tag{3.2}$$

We now ask how the GPH procedure would behave when applied to first differences of series of 158 observations generated from these models, and using 12 ordinates in (1.3). For the model (3.1), use of the approximation (2.2) in (2.1) suggests that the estimator $\hat{d}^*$ has mean -0.46 and use of (2.3) gives -0.37. In fact, for 1000 replications the empirical mean of $\hat{d}^*$ was -0.39 and the standard deviation was 0.25. Using a 5% level test, the null hypothesis $d^* = 0$ was rejected 426 times against the alternative $d^* < 0$ when the error variance in (1.3) was set at $\pi^2/6$, and 466 times when it was estimated from the OLS regression. For the model (3.2), use of (2.2) yields an approximate mean of -0.55 for $\hat{d}^*$ and (2.3) gives -0.50. For 1000 simulated series the empirical mean of $\hat{d}^*$ was -0.51 and the standard deviation was 0.25. With a 5% level test, the null hypothesis $d^* = 0$ was rejected 621 times against the alternative $d^* < 0$ when the error variance was set at $\pi^2/6$, and 648 times when the OLS estimate was used.

It is ironic that the GPH procedure produces, on average, a smaller estimate of d and more rejections of $d = 1$ for the model (3.2) where true d is 1 than for the model (3.1) with true d equal to zero. Given its large standard error, the estimate of d^* found by Diebold and Rudebusch does not seem very surprising in view of our simulation results. Oddly, it is somewhat more compatible with model (3.1) than with model (3.2), and suggests less evidence against $d = 1$ than might be expected on the average from data generated by either model.

Our second example, again taken from Diebold and Rudebusch, is a series of 119 observation on the logarithm of US real gross national product. Application, as by Diebold and Rudebusch, of (1.3), with $K = 11$, to the first differences gives $\hat{d}^* = -0.50$, with standard error 0.27, suggesting strong evidence against $d^* = 0$.

We fitted all ARIMA(p, 1, q) models with $p + q \leq 6$. The model selected by the SBC criterion was

$$(1 - 1.232B + 0.381B^2)\{(1 - B)X_t - \mu\} = (1 - 0.999B)\varepsilon_t \tag{3.3}$$

The estimate of the moving-average parameter is, of course, perplexing. It could result from the well-known 'pile-up' effect of maximum likelihood estimates of moving-average parameters on the boundary of the parameter region. On the other hand it may be a result of over-differencing. (It is interesting to note that, on the surface, an AR(1) model seemed not implausible. This was the third choice of the SBC criterion, and very close to the second. The estimated autoregressive parameter was 0.29. As we have seen, for time series generated by such a model the expected value of $\hat{d}^*$ would be *positive*.) Certainly it is possible to entertain some doubts about the adequacy of the model (3.3). Nevertheless, such processes are perfectly respectable, and it is straightforward to generate data from them. Of course, virtually identical results would follow if the moving-average parameter were set to unity.

Based on the analysis of the first differences of 119 observations from (3.3), and 11 ordinates in (1.3), the approximation (2.2) gives a mean of -0.64 for $\hat{d}^*$ while (2.3) yields -0.54. Over 1000 replications the empirical mean of $\hat{d}^*$ was -0.56, which is remarkably close to the estimate obtained by Diebold and Rudebusch for a single series, and the standard deviation was 0.28. With a 5% level test, the null hypothesis $d^* = 0$ was rejected 653 times against $d^* < 0$ when the error variance was set at $\pi^2/6$, and 686 times when it was estimated through OLS.

Whatever its merits as a generator of the given time series, the model (3.3) does remarkably well in predicting the Diebold–Rudebusch estimate of d. It is difficult to know whether to assert that (3.3) is a model in which $d = 1$—which technically it is—or one for which $d = 0$ (with a linear trend term), from which it is surely indistinguishable. However, it is *not* a model with $d = 0.5$—a compromise that seems to have arisen here through uncertainty about the two extreme possibilities. It is certainly possible that these data were generated by a process with fractional d. However, application of the GPH procedure has provided no evidence on which to base such a claim.

4. SUMMARY

Our analysis of the GPH estimator of the fractional difference parameter and of hypothesis tests based on that estimator suggests cause for concern. Taking the AR(1) and MA(1) models as illustrations, we have shown that, in moderate-sized samples, the estimator behaves as predicted by asymptotic theory for a wide range of parameter values. However, in some cases, where the generating process has a large positive autoregressive or moving average root, inference based on this estimator can be seriously misleading, even in very large samples. For such generating processes the GPH estimator of the fractional difference parameter can have serious bias, and associated hypothesis tests are likely to yield incorrect inferences. It appears that reliance on this procedure is justified only if the analyst has some knowledge about the autoregressive and moving-average components of the generating model. But, as a practical matter, such knowledge will not be available, and the attractive feature of a truly nonparametric inferential procedure is that it should not be necesssary. It appears that, in practice, routinely setting $K = T^{1/2}$ in (1.3) can be a poor strategy. Reducing K can decrease the bias somewhat, but only at the expense of increasing the standard error, which is already high for the sample sizes to which this methodology has been applied.

Although our results do not establish the conclusion, we doubt the possibility of developing any satisfactory truly nonparametric estimator of the fractional difference parameter. Rather, it appears necessary to estimate jointly all of the parameters, of which d is one, of the model (1.1). Fox and Taqqu (1986) have discussed such a procedure in the frequency domain, while Sowell (1986, 1990) has derived an algorithm for the computation of the covariance matrix, and hence the likelihood function, allowing estimation through full maximum likelihood. Of course these two approaches, particularly the latter, are computationally far more burdensome than the GPH procedure, even if the autoregressive and moving-average orders are assumed to be known. In practice, these orders will not be known, presumably necessitating the estimation of the model for a range of values of (p, q). Not only would this be computationally expensive, but it raises difficult and perhaps intractable questions on inference about d. While valid inference appears possible for a specific choice of (p, q), quite different results are likely to arise from some other choice. In these circumstances it is not appropriate to base inference on d on the fiction that (p, q) is known. For example, given a data set, an analyst might establish a parsimonious model that appears to be satisfactory, at least as a potential generator of short-run forecasts. However, the possibility is always open to add both an autoregressive and a moving-average term to this model. We shall not know that the additional parameters are redundant, although we may well find that they have little impact on short-run forecasting. They could, however, have a strong impact on inference about d. Their values are likely to be poorly estimated, and values close to unity may not be strongly contradicted by the data. It is precisely in these circumstances that there will be considerable uncertainty about d, whatever estimation procedure is used. Of course, the GPH estimator was developed to circumvent such intractable issues, and it is unfortunate that it is incapable of reliably doing so.

REFERENCES

Diebold, F. X. and Rudebusch, G. D. (1989) Long memory and persistence in aggregate output. *J. Monet. Econ.* 24, 189–209.

Fox, R. and Taqqu, M. S. (1986) Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* 14, 517–32.

Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *J. Time Series Anal.* 4, 221–38.

Porter-Hudak, S. (1990) An application of the seasonal fractionally differenced model to the monetary aggregates. *J. Am. Statist. Assoc.* 85, 338–44.

Sowell, F. B. (1986) Fractionally integrated vector time series. Ph.D. Thesis, Duke University.

— (1990) Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Working Paper, Graduate School of Industrial Administration, Carnegie Mellon University.