Discussion Paper 34

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis
250 Marquette Avenue
Minneapolis, Minnesota 55480

November 1990

# ON MAXIMUM-LIKELIHOOD ESTIMATION OF THE DIFFERENCING PARAMETER OF FRACTIONALLY INTEGRATED NOISE WITH UNKNOWN MEAN

Yin-Wong Cheung
Crown College

Frank X. Diebold
Institute for Empirical Macroeconomics and
University of Pennsylvania

Discussions with Larry Christiano, Christian Gourieroux, Bobby Mariano, Marc Nerlove, Peter Robinson, and Fallaw Sowell were extremely valuable. Useful comments were also received from participants at the 1990 NSF/NBER Time Series Conference at the University of California, San Diego, and from seminar participants at New York University, the University of British Columbia, and the University of Southern California. All remaining errors and shortcomings are ours alone. Diebold gratefully acknowledges financial support from the National Science Foundation (grant SES 89-21715), the University of Pennsylvania Research Foundation (grant 3-71441) and the Institute for Empirical Macroeconomics.

This material is based on work supported by the National Science Foundation under Grant No. SES-8722451. The Government has certain rights to this material.

## ABSTRACT

There are two approaches to maximum likelihood (ML) estimation of the parameter of fractionally-integrated noise: approximate frequency-domain ML (Fox and Taqqwu, 1986) and exact time-domain ML (Solwell, 1990a). If the mean of the process is <u>known</u>, then a clear finite-sample mean-squared error (MSE) ranking of the estimators emerges: the exact time-domain estimator has smaller MSE. We show in this paper, however, that the finite-sample efficiency of approximate frequency-domain ML relative to exact time-domain ML rises dramatically when the mean result is unknown and instead must be estimated. The intuition for our result is straightforward: The frequency-domain ML estimator is invariant to the true but unknown mean of the process, while the time-domain ML estimator is not. Feasible time-domain estimation must therefore be based upon de-meaned data, but the long memory associated with fractional integration makes precise estimation of the mean difficult. We conclude that the frequency-domain estimator is an attractive and efficient alternative for situations in which large sample sizes render time-domain estimation impractical.

## 1. Introduction

The literature on long-memory time series processes, and in particular, fractionally-integrated ARMA (ARFIMA) processes, has grown rapidly since the early contributions of Granger and Joyeux (1980), Hosking (1981) and Geweke and Porter-Hudak (1983). Recent theoretical work includes Fox and Taqqu (1986), Robinson (1988, 1990), Sowell (1990a, b), Gourieroux et al. (1987) and Yajima (1985, 1988), among others.

The theory is beginning to be used in applied econometric work, in which flexible characterization of long-run, or low-frequency, dynamics is often of crucial importance. Examples include Diebold and Rudebusch (1989), Haubrich and Lo (1988) and Sowell (1990c) (real output dynamics and the unit-root hypothesis), Diebold and Rudebusch (1990a) (disposable income dynamics and the permanent-income hypothesis), Lo (1988) (predictability of stock returns and the efficient-markets hypothesis), Shea (1990) (variance bounds for the interest-rate term structure and Hicks' expectations hypothesis), Diebold, Husted and Rush (1990) (real exchange rate dynamics and the purchasing power parity hypothesis), Cheung (1990) (nominal exchange rate dynamics and the efficient markets hypothesis) and Hassett (1990) (real wage dynamics and the intertemporal substitution hypothesis).

Most such applied work, however, makes use of estimation procedures whose properties are incompletely understood, and whose properties are likely to be suboptimal relative to maximum-likelihood (ML) under correct model specification.[1] Hence the interest in recent work on exact time-domain and approximate frequency-domain ML estimation of fractionally-integrated models. Examples include Fox and Taqqu (1986), who construct an asymptotic approximation to the likelihood of an ARFIMA process in the frequency domain, Sowell (1990a), who constructs the exact likelihood function of an ARFIMA process in the time domain, and Li and McLeod (1986), who study the asymptotic properties of the ML estimator[2].

Monte Carlo analyses, in particular Sowell (1990a), have shown that (in finite samples) the time-domain ML estimator is substantially more efficient than the frequency-domain ML estimator, when the mean of the process is known. Thus, in spite of the fact that time-domain ML estimation is more tedious than frequency-domain ML (due to the (TxT) covariance matrix that must be inverted at each evaluation of the likelihood function), time-domain ML appears to be an attractive estimator.[3]

In practice, of course, the mean is not known, so that existing Monte Carlo results correspond to an infeasible estimator. What happens when a feasible time-domain ML estimator, obtained by removing an estimated mean from the time

---

[1] The leading such estimation technique is the two-step procedure of Geweke and Porter-Hudak (1983), which we shall call the GPH procedure.

[2] Li and Mcleod work under the assumption of a known mean. Relation of that assumption is the subject of this paper.

[3] Here and throughout, T denotes sample size.

series prior to estimation, is used? In this paper, we motivate this question in light of some important differences underlying the construction of the time- and frequency-domain ML estimators, and we provide an answer.

The paper proceeds as follows. In section 2, we discuss the details of our Monte Carlo experiment, in which we explore the efficiency of the frequency-domain ML estimator (which does not depend on the mean) relative to that of the time-domain ML estimator (with the population mean assumed known and, alternatively, with the arithmetic sample mean removed prior to analysis). In section 3, we report the results of the Monte Carlo analysis; the efficiency of frequency-domain ML relative to time-domain ML with estimated mean is strikingly different from its efficiency relative to time-domain ML with known mean. In section 4, we offer additional discussion. Section 5 concludes.

## 2. The Monte Carlo Experiment

### 2a. Data Generating Process

We work with the pure fractionally-integrated process

$$(1 - B)^d X_t = e_t \tag{1}$$

$$e_t \sim NID(0, 1), \tag{2}$$

$t = 1, 2, \ldots, T$, where B is the backshift operator and $-1/2 < d < 1/2$.[4]

[4] The restriction on d is sufficient for stationarity and invertibility, and can always be achieved by taking a suitable number of integer differences.

2b. Estimators

We first consider time-domain ML estimation. Assume first that the mean of the process is known, and without loss of generality assume that it is zero. Under the normality assumption, construction of the likelihood simply amounts to expressing the autocovariances of the process in terms of the underlying parameters (in this case d). Evaluation of the likelihood requires inversion of the (TxT) Toeplitz covariance matrix, $\Sigma(d)$, with ij-th entry,

$$\gamma_x(|i-j|) = (-1)^{|i-j|} \{\Gamma(1-2d) / [\Gamma(1-d+|i-j|)\ \Gamma(1-d-|i-j|)]\},$$

where $1/\Gamma(d) = 0$ when d is a non-positive integer.[5] The first estimator we explore is precisely the one that maximizes this likelihood, with the mean assumed known to be zero (and hence not estimated).[6] The estimator is denoted ML1. Formally,

$$\text{ML1} = \arg\max_{d} L(x-\mu;\ \Sigma(d)),$$

where $L(\cdot)$ denotes the likelihood function. ML1 is of course not feasible in practice, because the mean is never known, but it will serve as a useful benchmark.

The obvious feasible counterpart to ML1 is obtained by first removing the sample mean from the data,

$$\text{ML1a} = \arg\max_{d} L((x-\bar{x});\ \Sigma(d)).$$

---

[5] This result may be traced at least to Adenstedt (1974) for the ARFIMA(0,d,0) case, and is generalized by Sowell (1990a) to the ARFIMA(p,d,q) case.

[6] The FORTRAN code used to evaluate and maximize the likelihood was generously supplied by Fallaw Sowell.

As long as $\bar{x}$ is consistent, the feasible estimator will perform satisfactorily in large samples. The long memory associated with fractional integration may lead to extremely inefficient estimates of the mean, however, leading to poor performance of ML1a in samples of the size typically available in economics.

Now we consider frequency-domain ML estimation.[7] Following Fox and Taqqu (1986), we exploit the fact that maximization of the Gaussian likelihood is asymptotically equivalent to minimization of

$$\sum_{j=1}^{T-1} [I_x(2\pi j/T) \,/\, f_x(2\pi j/T,\ d)],$$

with respect to d, where $I_x(\lambda)$ is the periodogram of X at frequency $\lambda$, and

$$f_x(\lambda,\ d) = |1 - e^{-i\lambda}|^{-2d}$$

is proportional to the spectral density of X at frequency $\lambda$. We call the resulting estimator ML2; formally,

$$\text{ML2} = \arg\min_{d} \sum_{j=1}^{T-1} [I_x(2\pi j/T) \,/\, f_x(2\pi j/T,\ d)].$$

Some authors, such as Dahlhaus (1988) have argued that tapering may improve the finite-sample properties of the frequency-domain estimator; therefore, we also explore the properties of two frequency-domain ML estimators that make use of

---

[7] The frequency-domain procedure, which builds upon an important result of Whittle (1951), has received substantial attention in the estimation of short-memory ARMA and unobserved-components models. See, for example, Nerlove et al. (1979) and Harvey (1989).

trapezoidally tapered data. The first is

$$ML2a = \arg\min_{d} \sum_{j=1}^{T-1} [I_{xa}(2\pi j/T) \,/\, f_x(2\pi j/T,\, d)],$$

where $I_{xa}(\lambda)$ is the periodogram of Xa at frequency $\lambda$, and where

$$xa_t = k_t\, x_t,$$

with

$$k_t = \begin{cases} t/a & 1 \le t \le a \\ 1 & a+1 \le t \le T-a \\ (T+1-t)/a & T+1-a \le t \le T, \end{cases}$$

where a = .1T.

The second taper is identical, except that 25% of each end of the sample is tapered rather than 10%. The estimator is

$$ML2b = \arg\min_{d} \sum_{j=1}^{T-1} [I_{xb}(2\pi j/T) \,/\, f_x(2\pi j/T,\, d)],$$

where $I_{xb}(\lambda)$ is the periodogram of Xb at frequency $\lambda$, and where

$$xb_t = k_t'\, x_t,$$

with

$$k_t' = \begin{cases} t/b & 1 \le t \le b \\ 1 & b+1 \le t \le T-b \\ (T+1-t)/b & T+1-b \le t \le T, \end{cases}$$

where b = .25T.

## 2c. Experimental Design

Ten points in the parameter space are explored, corresponding to d = ±.05, ±.15, ±.25, ±.35, and ±.45. Sample sizes of T = 50, 100, 300 and 500 are explored. Realizations of

(2) for each T are generated by IMSL subroutine DRNNOA, and then the corresponding realizations of (1) for each {d, T} configuration are generated by multiplying vectors of N(0,1) deviates by the Choleski factor of the covariance matrix of X, as in Diebold and Rudebusch (1990b). The sampling properties of the various time- and frequency-domain ML estimators discussed earlier are explored. For each {d, T, estimator} configuration, 1000 Monte Carlo replications are performed, and the bias and mean squared error (MSE) across the replications are computed.

## 3. Results

The finite-sample biases of the various estimators are reported in table 1 and figures 1-3.[8] Bias is almost always negative; that is, d tends to be under-estimated. For fixed d, the bias of each estimator approaches 0 as T approaches $\infty$, as expected. For all fixed d and T, the absolute bias of ML1 is generally smallest, followed by ML1a, which generally has a smaller bias than any of ML2, ML2a and ML2b. The absolute biases of ML1a and ML2 are roughly comparable, however, particularly for large d. The data tapers often provide little bias reduction, although for the most persistent parameterizations (large positive d values) they do reduce bias, a finding that accords with Dahlhaus (1988). We summarize these results as:

$$B(ML1) < B(ML1a) \leq B(ML2) \approx B(ML2a) \approx B(ML2b), \qquad \text{(Result A)}$$

---

[8] To conserve space, we do not graph the bias for the tapered frequency-domain estimators.

for all fixed d and T, where B(·) denotes absolute bias.

Second, for all fixed T, the (absolute) bias of ML1 tends to increase slightly with d, while the bias of ML1a tends to increase <u>sharply</u> with d. The biases of the other estimators show no clear relationship to d. We summarize these results as:

$$d\uparrow \Rightarrow B(ML1)\uparrow$$

$$d\uparrow \Rightarrow B(ML1a)\uparrow\uparrow \qquad \text{(Result B)}$$

$$d\uparrow \Rightarrow B(ML2)\uparrow\downarrow,\ B(ML2a)\uparrow\downarrow,\ B(ML2b)\uparrow\downarrow,$$

for all fixed T. Additional discussion and interpretation of this result will be provided below, when we discuss the results for MSE.

Ultimately, we are not interested in small bias <u>per se</u>. In particular, the squared error loss with which we evaluate the estimators may be decomposed into the sum of squared bias and variance, so that high bias is acceptable if it is sufficiently compensated for by low variance. In short, we report the biases in order to enable the reader to mentally decompose the MSE's, to which we now turn, into their underlying components.

The finite-sample MSE's of the various estimators are reported in table 2 and figures 4-6. As expected, for fixed d, the MSE of each estimator decreases with T. A number of additional important results are apparent. First, for all fixed T and d, time-domain ML distinctly dominates frequency-domain ML when the mean is known. Second, for all fixed T and d, time-domain ML outperforms frequency-domain ML by a much smaller margin when the mean is estimated. Particularly for d > 0, there is little performance difference between the two estimators. Third, although neither taper consistently produces a reduction in the MSE of the frequency-domain estimator, both tapers reduce MSE when d is very close to 1/2. (Again, see Dahlhaus, 1988.) We summarize these results as:

$$\mathrm{MSE(ML1)} < \mathrm{MSE(ML1a)} \approx \mathrm{MSE(ML2)} \approx \mathrm{MSE(ML2a)} \approx \mathrm{MSE(ML2b)}, \qquad \text{(Result C)}$$

for all fixed d and T.

Now consider the effects of varying d. First, note that for each sample size, the MSE of ML1 is decreasing in d. Evidently the reduced variance of that estimator, which is obtained through the greater unconditional variation induced by higher d values, more than offsets the slightly increased squared bias associated with higher d values. Second, note that for each sample size, the MSE of ML1a tends to <u>increase</u> in d. Here, the reduced variance is not enough to offset the large bias increases associated with higher d. Third, note that the MSE's of the frequency-domain estimators are not particularly sensitive to d. We summarize these results as:

$$\begin{aligned} &d\uparrow \Rightarrow \mathrm{MSE(ML1)}\downarrow \\ &d\uparrow \Rightarrow \mathrm{MSE(ML1a)}\uparrow \qquad\qquad \text{(Result D)} \\ &d\uparrow \Rightarrow \mathrm{MSE(ML2)}\uparrow\downarrow,\ \mathrm{MSE(ML2a)}\uparrow\downarrow,\ \mathrm{MSE(ML2b)}\uparrow\downarrow, \end{aligned}$$

for all fixed T.

Understanding may be deepened, and all of the results conveniently summarized, by examining relative efficiencies. The efficiency of estimator i relative to that of estimator j, denoted Rij, is MSE(j)/MSE(i). In particular, we define:

R21 = MSE(ML1) / MSE(ML2)

R21a = MSE(ML1a) / MSE(ML2)

R2a1a = MSE(ML1a) / MSE(ML2a)

R2b1a = MSE(ML1a) / MSE(ML2b).

These relative efficiencies are reported in table 3, and graphed in figures 7 and 8, for the various T and d configurations. Here we shall focus on the figures, which conveniently summarize a large amount of information.

Consider first figure 7, in which the efficiency of frequency-domain ML relative to time-domain ML <u>with the mean known</u> (R21) is graphed as a function of d. Four relative efficiency curves are shown, each corresponding to a different sample size. Note that the absolute height of each of the four curves is low, indicating the low efficiency of frequency-domain ML relative to time-domain ML when the mean is known. Due to the asymptotic equivalence of the two estimators, relative efficiency tends to grow with T, approaching 1.0 in the limit. Also of interest is the fact that, for each sample size, the efficiency of frequency-domain ML relative to time-domain ML <u>decreases</u> with d. This is a manifestation of Result D.

Now consider figure 8, in which the efficiency of frequency-domain ML relative to time-domain ML <u>with the mean estimated</u> (R21a) is graphed as a function of d. First, note that the absolute height of each of the four curves is substantially higher than in figure 7, indicating the much improved efficiency of frequency-domain ML relative to time-domain ML when the mean is estimated. Again, due to the asymptotic equivalence of the two estimators, relative efficiency tends to grow with T, approaching 1.0 in the limit. Second, note that, for each sample size, the efficiency of frequency-domain ML relative to time-domain ML now <u>increases</u> with d--the more persistent the process, the better the relative performance of frequency-domain ML. Again, this is a manifestation of Result D.

Further insight into the deterioration of the relative performance of time-domain ML as d grows, when the mean is estimated, can be gained by recalling a result of Sowell (1990b, Theorem 1), who shows that for the pure fractional process (1)-(2), $\mathrm{var}(\Sigma_{t=1}^{T} x_t) = O(T^{1+2d})$. Thus, $T^{-1/2-d} \Sigma_{t=1}^{T} x_t$ has a stable limiting distribution, so that $T^{1/2-d} \bar{x}$ has a stable limiting distribution; that is, the convergence rate of the sample mean depends inversely on d. When d = 0 the usual root-T consistency obtains; convergence is faster or slower than root-T as d is less than or greater than 0. The larger is d, the more slowly the sample mean converges, and the poorer the performance of feasible time-domain ML.

## 4. Additional Discussion

Here we focus on certain aspects of the analysis that merit additional attention, with particular attention paid to directions for future research.

First, we intentionally neglect the GPH estimator in our

Monte Carlo comparison. Such a comparison would be unfairly biased against the semi-parametric GPH estimator, because we work only under correct model specification. It might be desirable to study in future work the comparative properties of GPH, time-domain ML, and frequency-domain ML under model misspecification of various types. Such a study, however, would be very challenging in terms of experimental design.

Second, we intentionally focus on the case of pure fractional noise, that is, the ARFIMA(0,d,0) case. The pure fractional noise is of substantial interest in its own right, and moreover, it is best to attempt a thorough understanding of the pure fractional noise before proceeding to more complex processes, because the insights gained from its study are likely to provide useful guidance regarding behavior in more complex environments. The Monte Carlo analysis of Cheung (1990) deals with more complex models with AR, MA and ARMA components, with no change in the qualitative nature of the results reported here (but with a great increase in design complexity and computational burden).

Third, we use only the arithmetic sample mean to estimate the population mean. It is apparent that the performance of the time-domain ML estimator deteriorates significantly when the sample mean is removed from the data prior to estimation. The possibility arises, however, that alternative feasible time-domain estimators, based upon alternative estimators of the population mean that take account of the dynamic structure of the data, might produce better performance.

The obvious candidate estimator is Adenstedt's (1974) best linear unbiased (BLUE) estimator, which depends only on d, collapses to the sample mean in the d = 0 case, and can be made feasible by using a preliminary semi-parametric d estimate from the GPH procedure. Another candidate, which is clearly not a fully efficient estimator of the population mean, but might nevertheless provide substantial efficiency gains and is readily computed using standard software, is the least-squares estimator with a correction for first-order serial correlation.

It turns out, however, that the prospects for improving the performance of the feasible time-domain estimator by using alternative estimators of the mean are very limited. Samarov and Taqqu (1988) have shown analytically that, for a variety of sample sizes, the efficiency of the arithmetic mean estimator relative to Adenstedt's BLUE estimator is close to 100% over most of the parameter space and for a wide range of sample sizes. Similar results emerge from the Monte Carlo analyses of Mohr (1981) and Graf (1983).

Finally, we note that inclusion of the determinant term in the frequency-domain Gaussian likelihood is likely to improve the finite-sample performance of the frequency-domain estimator ML2, as suggested by Nerlove _et al._ for ARMA processes and Boes _et al._ for ARFIMA processes. Thus, the prospects for improving the performance of the ML2 estimator are _not_ so limited (in contrast to those for ML1a). Taken together, these insights make our main result even stronger: The efficiency of a simple variation of ML2 frequency-domain ML estimator relative to ML1a is likely to be even larger than that reported here for ML2.

## 5. Summary and Conclusions

We have examined the finite-sample performance of ML estimators of the parameter of a pure fractional process. We first showed that the efficiency of frequency-domain ML relative to time-domain ML is poor, when the mean of the process is known. These Monte Carlo results are in complete accord with those of Sowell (1990a). The time-domain ML estimator that assumes a known mean is not feasible, however, whereas the frequency-domain estimator is. We therefore compared the finite-sample efficiency of the leading feasible time-domain ML estimator to the frequency-domain estimator. The comparison is of key importance, because it is the one relevant to actual practice. The results were striking: the relative efficiency of approximate frequency-domain ML was much improved.

To determine the implications of our results for applied work, one must weigh costs and benefits. The feasible time-domain ML estimator, while much less efficient than its infeasible counterpart, nevertheless usually has somewhat lower MSE than the frequency-domain ML estimator for the sample sizes and parameter values examined here. (The biggest differences, of course, arise for the smallest samples.) Time-domain ML, however, requires tedious (TxT) covariance matrix inversion at each evaluation of the likelihood. Conversely, although the frequency-domain ML estimator has slightly higher MSE, it has the virtue of a light computational burden.

One might be tempted to conclude that the lighter computational burden associated with frequency-domain ML more than offsets its slightly higher MSE. We do not necessarily agree. Today's powerful computing environment makes Sowell's exact time-domain estimator viable for the small/medium sample sizes in which it can really make a difference. The good news provided by this paper is that, for the medium/large sample sizes in which time-domain ML is likely to be prohibitively tedious (or impossible), frequency-domain ML is likely to perform very well.

## References

Adenstedt, R.K. (1974), "On Large-Sample Estimation for the Mean of a Stationary Random Sequence," Annals of Statistics, 2, 1095-1107.

Boes, D.C., Davis, R.A. and Gupta, S.N. (1989), "Parameter Estimation in Low Order Fractionally Differenced ARMA Processes," Stochastic Hydrology and Hydraulics, 3, 97-110.

Cheung, Y.-W. (1990), Long Term Memory in Foreign Exchange Rates. Ph.D. Dissertation, University of Pennsylvania.

Dahlhaus, R. (1988), "Small Sample Effects in Time Series Analysis: A New Asymptotic Theory and a New Estimator," Annals of Statistics, 16, 808-841.

Diebold, F.X., Husted, S. and Rush, M. (1990), "Real Exchange Rates Under the Gold Standard," Manuscript, Department of Economics, University of Pennsylvania.

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

Diebold, F.X. and Rudebusch, G.D. (1990a), "Is Consumption too Smooth? Long Memory and the Deaton Paradox," Manuscript, Department of Economics, University of Pennsylvania. Forthcoming, Review of Economics and Statistics.

Diebold, F.X. and Rudebusch, G.D. (1990b), "On the Power Properties of Dickey-Fuller Tests Against Fractionally-Integrated Alternatives," Forthcoming, Economics Letters.

Fox, R. and Taqqu, M.S. (1986), "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time-Series," Annals of Statistics, 14, 517-532.

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

Gourieroux, C., Maurel, F. and Monfort, A. (1987), "Regression and Nonstationarity," Working Paper #8708, INSEE, Paris.

Graf, H.P. (1983), Long-Range Correlations and Estimation of the Self-Similarity Parameter. Ph.D. dissertation, Swiss Federal Institute of Technology, Zurich.

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

Harvey, A.C. (1989), Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge: Cambridge University Press.

Hassett, K. (1990), "Is the Aggregate Labor Market Exploding?," Manuscript, Graduate School of Business, Columbia University.

Haubrich, J.G. and Lo, A.W. (1988), "The Sources and Nature of Long-Term Memory in the Business Cycle," Manuscript, Department of Finance, Wharton School, University of Pennsylvania.

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

Li, W.K. and McLeod, A.E. (1986), "Fractional Time Series Modelling," Biometrika, 73, 217-21.

Lo, A.W. (1988), "Long-Term Memory in Stock Market Prices," Manuscript, Department of Finance, Wharton School, University of Pennsylvania.

Mohr, D.L. (1981), Modeling Data as a Fractional Gaussian Noise. Ph.D. dissertation, Princeton University.

Nerlove, M., Grether, D.M. and Carvalho, J.L. (1979), Analysis of Economic Time Series: A Synthesis. New York: Academic Press.

Robinson, P.M. (1988), "Nonparametric Function Estimation for Long-Memory Time Series," Manuscript, London School of Economics.

Robinson, P.M. (1990), "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression," Forthcoming, Journal of Econometrics.

Samarov, A. and Taqqu, M.S. (1988), "On the Efficiency of the Sample Mean in Long-Memory Noise," Journal of Time Series Analysis, 9, 191-200.

Shea, G.S. (1990), "Uncertainty and Implied Variance Bounds in Long-Memory Models of the Interest Rate Term Structure," Forthcoming, Empirical Economics.

Sowell, F.B. (1990a), "Maximum-Likelihood Estimation of Stationary Univariate Fractionally-Integrated Time-Series Models," Manuscript, Graduate School of Industrial

Administration, Carnegie-Mellon University.

Sowell, F.B. (1990b), "Fractional Unit-Root Distribution," Econometrica, 58, 495-506.

Sowell, F.B. (1990c), "Characterizing Long-Run Behavior," Manuscript, Graduate School of Industrial Administration, Carnegie-Mellon University.

Whittle, P. (1951), Hypothesis Testing in Time Series. Uppsala: Almqvist and Wiksell.

Yajima, Y. (1985), "On Estimation of Long-Memory Time Series Models," Australian Journal of Statistics, 27, 303-320.

Yajima, Y. (1988), "On Estimation of a Regression Model With Long-Memory Stationary Errors," Annals of Statistics, 16, 791-807.

Table 1
Finite-Sample Bias

| | ML1 | ML1a | ML2 | ML2a | ML2b |
|---|---|---|---|---|---|
| T=50 | | | | | |
| d=-.45 | -.0193 | -.0389 | -.0781 | -.0933 | -.0997 |
| d=-.35 | -.0291 | -.0630 | -.0846 | -.0932 | -.0983 |
| d=-.25 | -.0206 | -.0695 | -.0991 | -.1016 | -.1081 |
| d=-.15 | -.0237 | -.0740 | -.0979 | -.1002 | -.1044 |
| d=-.05 | -.0329 | -.0893 | -.1022 | -.1014 | -.1035 |
| d=.05 | -.0283 | -.0950 | -.1030 | -.1007 | .0989 |
| d=.15 | -.0213 | -.0916 | -.0972 | .0979 | .0971 |
| d=.25 | -.0318 | -.1068 | -.1031 | -.0966 | -.0906 |
| d=.35 | -.0365 | -.1182 | -.0929 | -.0796 | .0664 |
| d=.45 | -.0433 | -.1339 | -.0886 | -.0473 | .0121 |
| T=100 | | | | | |
| d=-.45 | -.0134 | -.0230 | -.0349 | -.0474 | -.0499 |
| d=-.35 | -.0156 | -.0331 | -.0437 | -.0529 | -.0570 |
| d=-.25 | -.0072 | -.0308 | -.0392 | -.0436 | -.0468 |
| d=-.15 | -.0165 | -.0451 | -.0524 | -.0555 | -.0583 |
| d=-.05 | -.0091 | -.0410 | -.0472 | -.0480 | -.0494 |
| d=.05 | -.0150 | -.0494 | -.0530 | -.0543 | -.0540 |
| d=.15 | -.0143 | -.0503 | -.0502 | -.0490 | -.0504 |
| d=.25 | -.0137 | -.0528 | -.0471 | -.0437 | -.0418 |
| d=.35 | -.0188 | -.0594 | -.0435 | -.0353 | -.0299 |
| d=.45 | -.0271 | -.0789 | -.0412 | -.0119 | .0049 |
| t=300 | | | | | |
| d=-.45 | -.0028 | -.0053 | -.0057 | -.0178 | -.0185 |
| d=-.35 | -.0044 | -.0110 | -.0119 | -.0183 | -.0184 |
| d=-.25 | -.0036 | -.0136 | -.0148 | -.0172 | -.0205 |
| d=-.15 | -.0029 | -.0143 | -.0160 | -.0182 | -.0141 |
| d=-.05 | -.0017 | -.0167 | -.0177 | -.0168 | -.0207 |
| d=.05 | -.0048 | -.0173 | -.0184 | -.0195 | -.0204 |
| d=.15 | -.0041 | -.0174 | -.0170 | -.0166 | -.0182 |
| d=.25 | -.0041 | -.0186 | -.0156 | -.0141 | -.0136 |
| d=.35 | -.0064 | -.0224 | -.0147 | -.0121 | -.0101 |
| d=.45 | -.0133 | -.0349 | -.0253 | -.0088 | .0082 |
| T=500 | | | | | |
| d=-.45 | -.0036 | -.0048 | -.0041 | -.0104 | -.0104 |
| d=-.35 | -.0016 | -.0075 | -.0070 | -.0100 | -.0126 |
| d=-.25 | -.0033 | -.0091 | -.0092 | -.0111 | -.0120 |
| d=-.15 | -.0022 | -.0100 | -.0107 | -.0139 | -.0103 |
| d=-.05 | -.0030 | -.0122 | -.0121 | -.0120 | -.0117 |
| d=.05 | -.0019 | -.0095 | -.0135 | -.0186 | -.0193 |
| d=.15 | -.0013 | -.0094 | -.0116 | -.0090 | -.0108 |
| d=.25 | -.0020 | -.0119 | -.0097 | -.0090 | -.0086 |
| d=.35 | -.0018 | -.0113 | -.0057 | -.0054 | -.0057 |
| d=.45 | -.0056 | -.0195 | -.0105 | -.0018 | -.0010 |

Note to table 1: The mnemonics are defined as follows. ML1: time-domain ML, true mean removed; ML1a: time-domain ML, arithmetic mean removed; ML2: frequency-domain ML; ML2a: frequency-domain ML, taper a; ML2b: frequency-domain ML, taper b. The bias is the mean estimate over the Monte Carlo replications, less the true value.

**Table 2**
**Finite-Sample Mean Squared Error**

| | ML1 | ML1a | ML2 | ML2a | ML2b |
|---|---|---|---|---|---|
| T=50 | | | | | |
| d=-.45 | .0175 | .0191 | .0294 | .0320 | .0352 |
| d=-.35 | .0190 | .0235 | .0310 | .0329 | .0370 |
| d=-.25 | .0173 | .0246 | .0333 | .0337 | .0386 |
| d=-.15 | .0163 | .0241 | .0329 | .0354 | .0389 |
| d=-.05 | .0169 | .0278 | .0338 | .0340 | .0388 |
| d=.05 | .0156 | .0280 | .0335 | .0342 | .0380 |
| d=.15 | .0127 | .0274 | .0342 | .0353 | .0378 |
| d=.25 | .0128 | .0296 | .0343 | .0344 | .0362 |
| d=.35 | .0102 | .0297 | .0311 | .0295 | .0316 |
| d=.45 | .0079 | .0302 | .0307 | .0254 | .0287 |
| T=100 | | | | | |
| d=-.45 | .0078 | .0086 | .0107 | .0114 | .0130 |
| d=-.35 | .0080 | .0094 | .0114 | .0126 | .0149 |
| d=-.25 | .0075 | .0092 | .0106 | .0117 | .0143 |
| d=-.15 | .0074 | .0104 | .0117 | .0135 | .0159 |
| d=-.05 | .0074 | .0102 | .0112 | .0117 | .0138 |
| d=.05 | .0076 | .0110 | .0121 | .0133 | .0152 |
| d=.15 | .0062 | .0103 | .0111 | .0118 | .0139 |
| d=.25 | .0060 | .0105 | .0112 | .0113 | .0125 |
| d=.35 | .0049 | .0103 | .0109 | .0104 | .0110 |
| d=.45 | .0035 | .0114 | .0113 | .0102 | .0120 |
| T=300 | | | | | |
| d=-.45 | .0024 | .0023 | .0027 | .0030 | .0036 |
| d=-.35 | .0026 | .0024 | .0027 | .0032 | .0036 |
| d=-.25 | .0022 | .0025 | .0026 | .0032 | .0035 |
| d=-.15 | .0024 | .0024 | .0025 | .0029 | .0032 |
| d=-.05 | .0021 | .0028 | .0029 | .0030 | .0040 |
| d=.05 | .0022 | .0027 | .0027 | .0030 | .0036 |
| d=.15 | .0021 | .0026 | .0027 | .0029 | .0035 |
| d=.25 | .0020 | .0026 | .0027 | .0029 | .0034 |
| d=.35 | .0020 | .0027 | .0028 | .0029 | .0034 |
| d=.45 | .0014 | .0028 | .0026 | .0021 | .0023 |
| T=500 | | | | | |
| d=-.45 | .0013 | .0014 | .0015 | .0014 | .0018 |
| d=-.35 | .0013 | .0015 | .0016 | .0017 | .0020 |
| d=-.25 | .0013 | .0015 | .0016 | .0015 | .0020 |
| d=-.15 | .0012 | .0015 | .0015 | .0017 | .0020 |
| d=-.05 | .0013 | .0016 | .0015 | .0016 | .0020 |
| d=.05 | .0014 | .0016 | .0016 | .0021 | .0023 |
| d=.15 | .0013 | .0014 | .0015 | .0017 | .0022 |
| d=.25 | .0012 | .0015 | .0015 | .0016 | .0018 |
| d=.35 | .0011 | .0014 | .0015 | .0016 | .0019 |
| d=.45 | .0008 | .0013 | .0012 | .0012 | .0014 |

Note to table 2: The mnemonics are defined as in table 1.

**Table 3**
**Finite-Sample Relative Efficiency**

| | R21 | R21a | R2a1a | R2b1a |
|---|---|---|---|---|
| <u>T=50</u> | | | | |
| d=-.45 | .5952 | .6497 | .5969 | .5461 |
| d=-.35 | .6129 | .7581 | .7143 | .6351 |
| d=-.25 | .5195 | .7387 | .7300 | .6373 |
| d=-.15 | .4954 | .7325 | .6808 | .6195 |
| d=-.05 | .5000 | .8225 | .8176 | .7165 |
| d=.05 | .4657 | .8358 | .8187 | .7368 |
| d=.15 | .3713 | .8012 | .7762 | .7249 |
| d=.25 | .3732 | .8630 | .8605 | .8177 |
| d=.35 | .3280 | .9550 | 1.0068 | .9399 |
| d=.45 | .2573 | .9837 | 1.1890 | 1.0523 |
| <u>T=100</u> | | | | |
| d=-.45 | .7290 | .8037 | .7544 | .6615 |
| d=-.35 | .7018 | .8785 | .7460 | .6309 |
| d=-.25 | .7075 | .8679 | .7863 | .6434 |
| d=-.15 | .6325 | .8889 | .7704 | .6541 |
| d=-.05 | .6607 | .9107 | .8718 | .7391 |
| d=.05 | .6281 | .9091 | .8271 | .7237 |
| d=.15 | .5586 | .9279 | .8729 | .7410 |
| d=.25 | .5357 | .9459 | .9292 | .8400 |
| d=.35 | .4495 | .9450 | .9904 | .9364 |
| d=.45 | .3097 | 1.0088 | 1.1176 | .9500 |
| <u>T=300</u> | | | | |
| d=-.45 | .8889 | .8519 | .7667 | .6389 |
| d=-.35 | 1.0000 | .9231 | .7500 | .6667 |
| d=-.25 | .8462 | .9615 | .7812 | .7143 |
| d=-.15 | .9600 | .9600 | .8276 | .7500 |
| d=-.05 | .7241 | .9655 | .9333 | .7000 |
| d=.05 | .8148 | 1.0000 | .9000 | .7500 |
| d=.15 | .7778 | .9630 | .8966 | .7429 |
| d=.25 | .7407 | .9630 | .8966 | .7647 |
| d=.35 | .7143 | .9643 | .9310 | .7941 |
| d=.45 | .5385 | 1.0769 | 1.3333 | 1.2174 |
| <u>T=500</u> | | | | |
| d=-.45 | .8667 | .9333 | 1.0000 | .7777 |
| d=-.35 | .8125 | .9375 | .8823 | .7500 |
| d=-.25 | .8125 | .9375 | 1.0000 | .7500 |
| d=-.15 | .8000 | 1.0000 | .8824 | .7500 |
| d=-.05 | .8667 | 1.0670 | 1.0000 | .8000 |
| d=.05 | .8750 | 1.0000 | .7619 | .6956 |
| d=.15 | .8667 | .9333 | .8235 | .6364 |
| d=.25 | .8000 | 1.0000 | .9375 | .8333 |
| d=.35 | .7333 | .9333 | .8750 | .7368 |
| d=.45 | .6667 | 1.0833 | 1.0833 | .9286 |

Note to table 3: Rij denotes the efficiency of estimator j relative to that of estimator i. Thus, R21 = MSE(ML1) / MSE(ML2), R21a = MSE(ML1a) / MSE(ML2), R2a1a = MSE(ML1a) / MSE(ML2a), and R2b1a = MSE(ML1a) / MSE(ML2b).

FIGURE 1

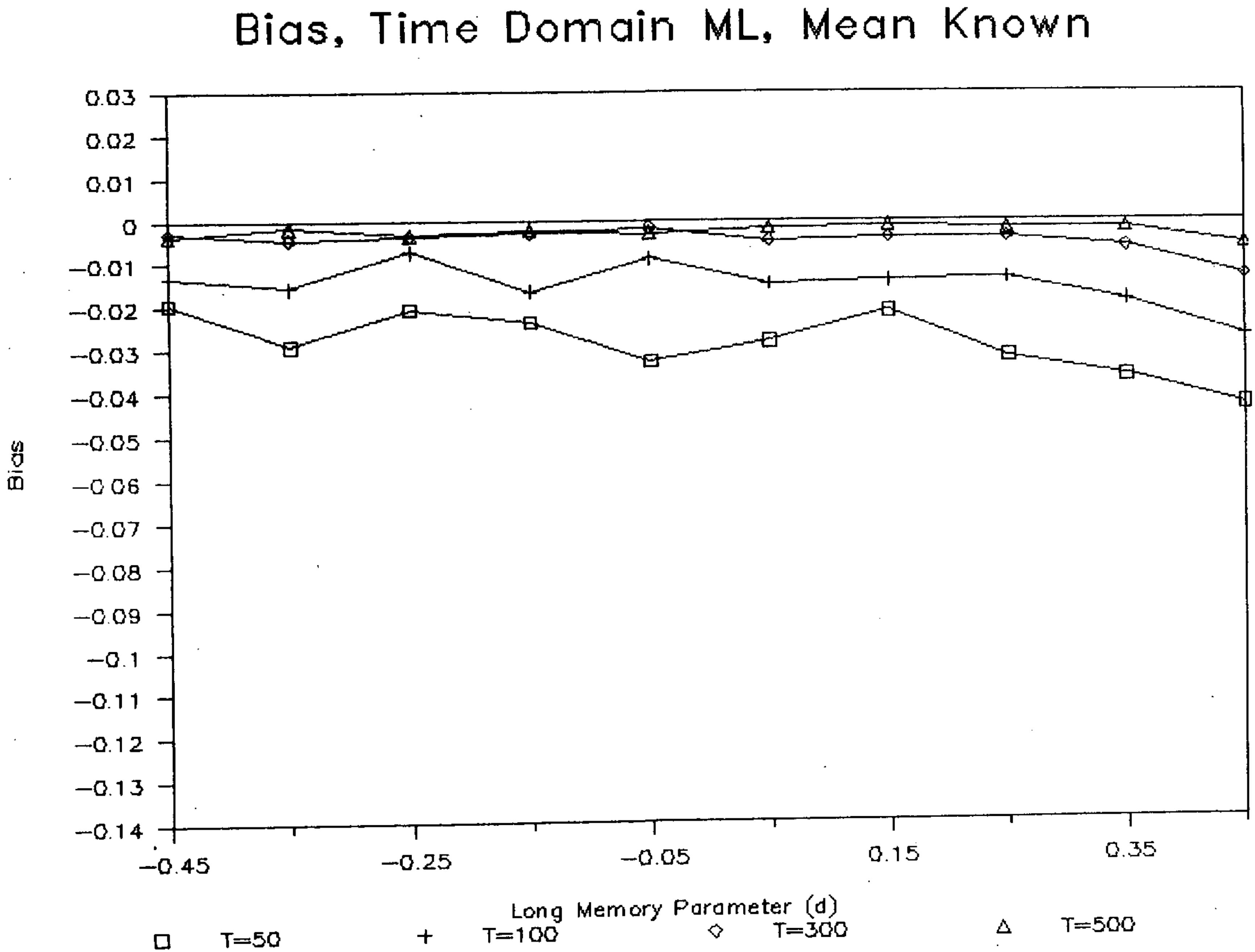

Bias, Time Domain ML, Mean Known
0.03
0.02
0.01
0
−0.01
−0.02
−0.03
−0.04
−0.05
−0.06
−0.07
−0.08
−0.09
−0.1
−0.11
−0.12
−0.13
−0.14
Bias
−0.45
−0.25
−0.05
0.15
0.35
Long Memory Parameter (d)
T=50
T=100
T=300
T=500

FIGURE 2

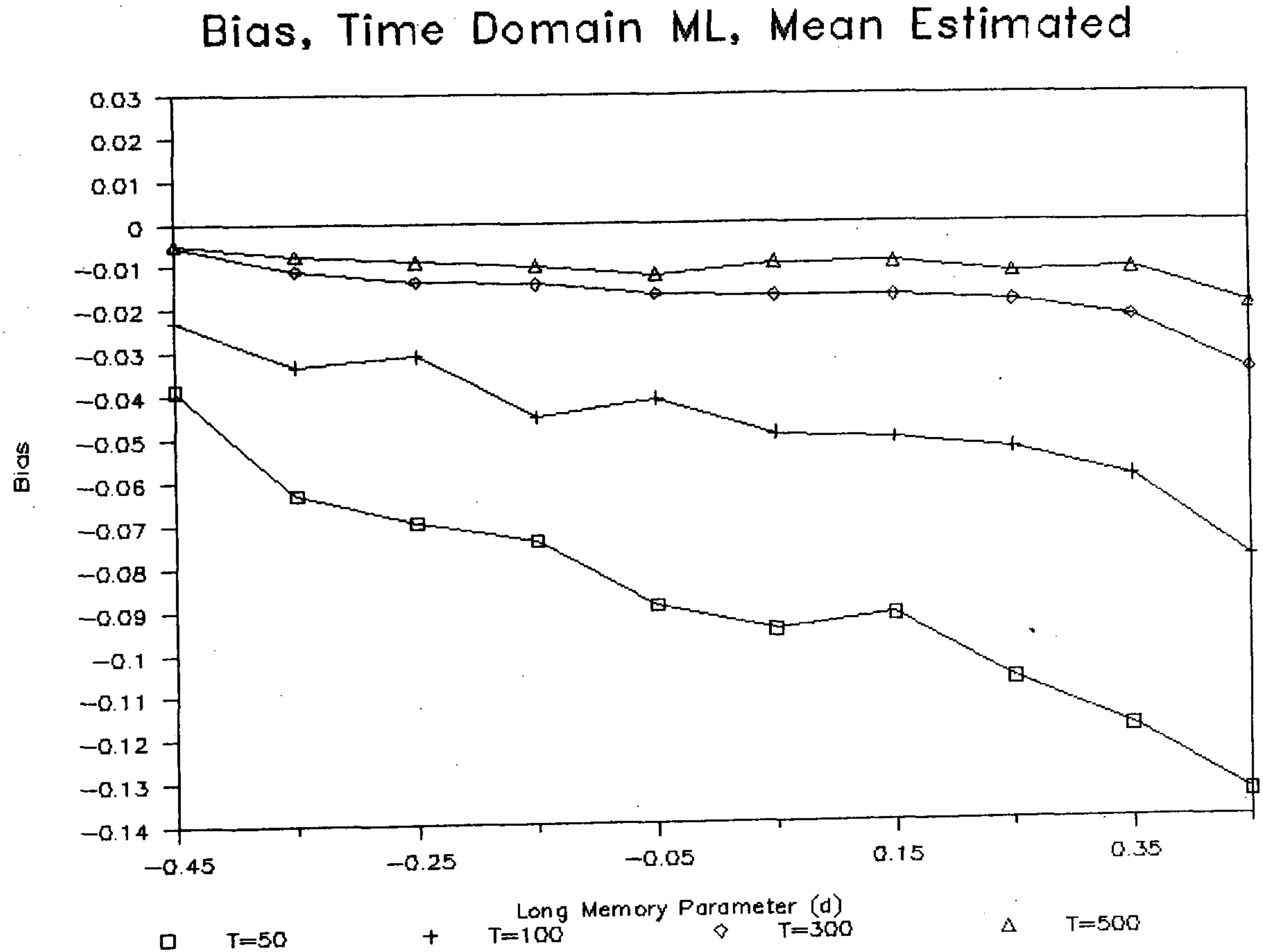

Bias, Time Domain ML, Mean Estimated
0.03
0.02
0.01
0
−0.01
−0.02
−0.03
−0.04
−0.05
−0.06
−0.07
−0.08
−0.09
−0.1
−0.11
−0.12
−0.13
−0.14
Bias
−0.45
−0.25
−0.05
0.15
0.35
Long Memory Parameter (d)
T=50
T=100
T=300
T=500

FIGURE 3

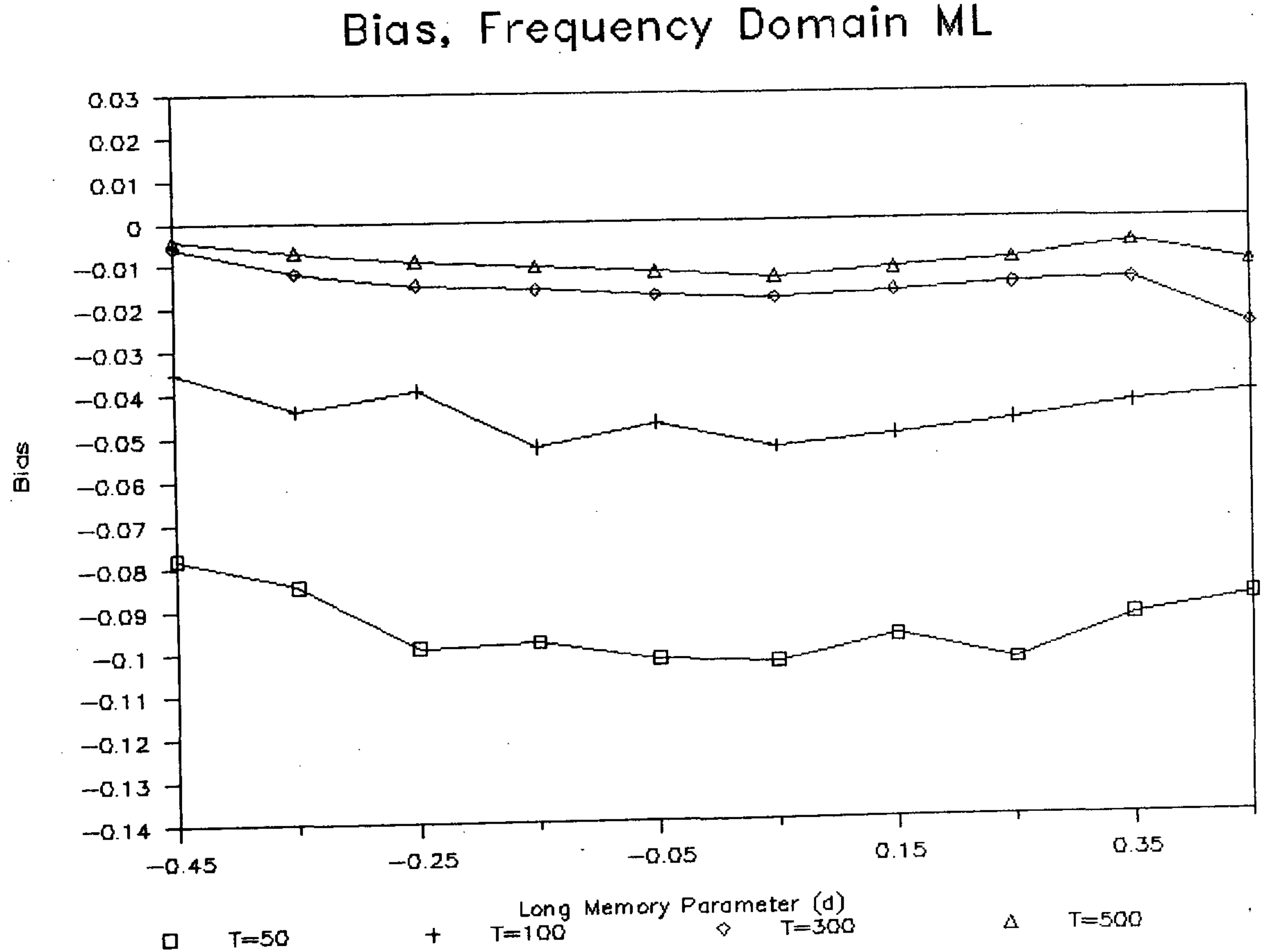

Bias, Frequency Domain ML
0.03
0.02
0.01
0
−0.01
−0.02
−0.03
−0.04
−0.05
−0.06
−0.07
−0.08
−0.09
−0.1
−0.11
−0.12
−0.13
−0.14
Bias
−0.45
−0.25
−0.05
0.15
0.35
Long Memory Parameter (d)
T=50
T=100
T=300
T=500

FIGURE 4

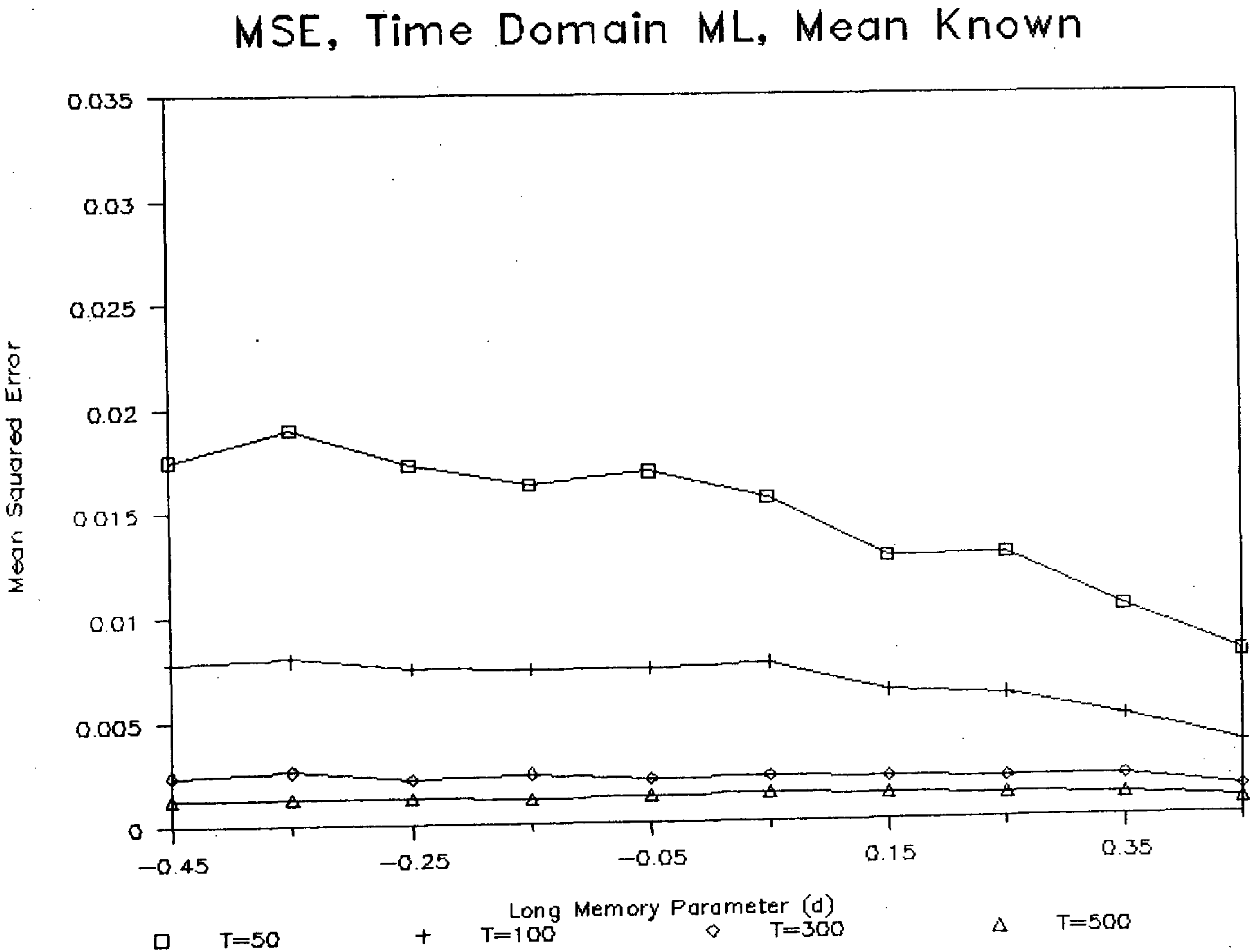

MSE, Time Domain ML, Mean Known
Mean Squared Error
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
−0.45
−0.25
−0.05
0.15
0.35
Long Memory Parameter (d)
T=50
T=100
T=300
T=500

FIGURE 5

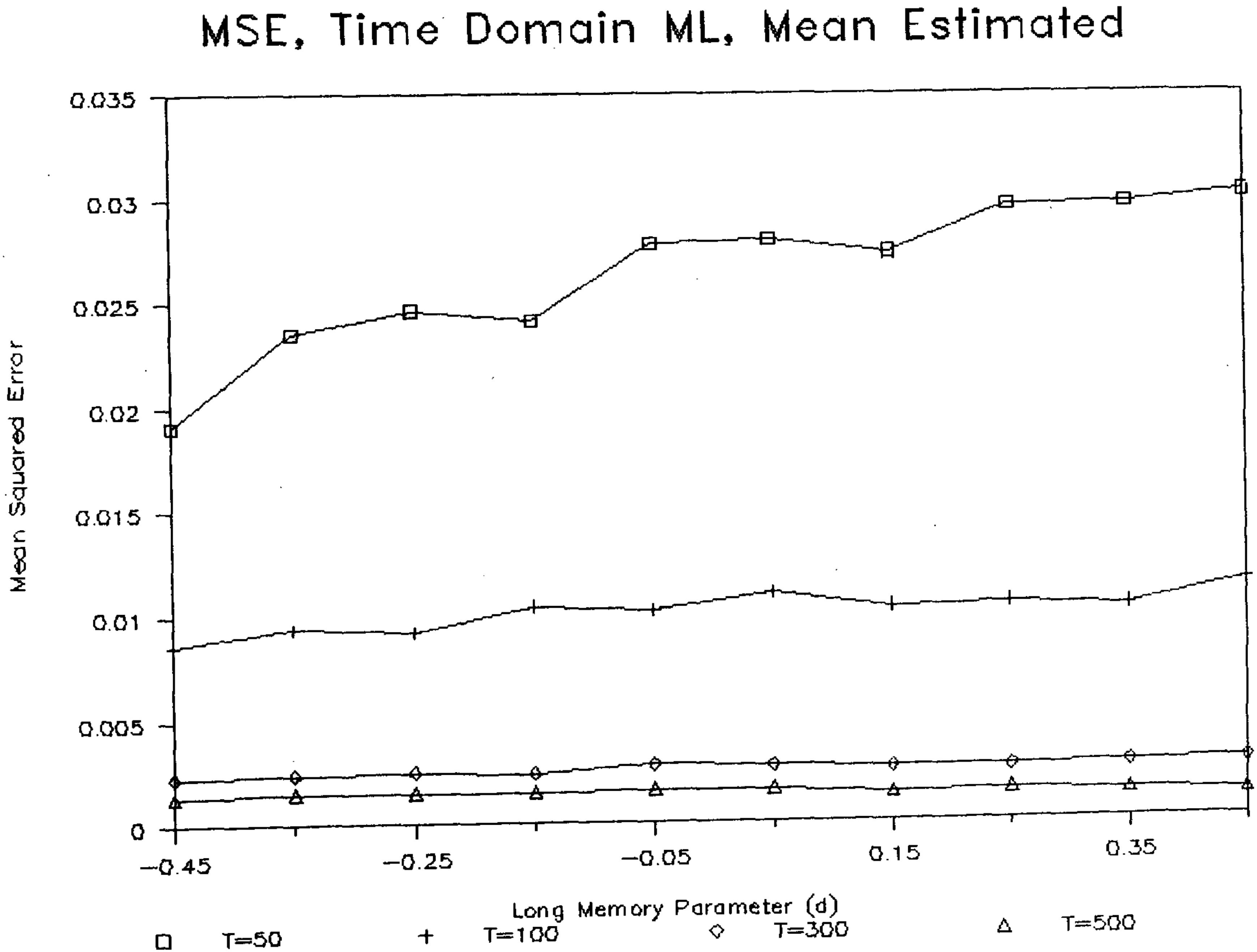
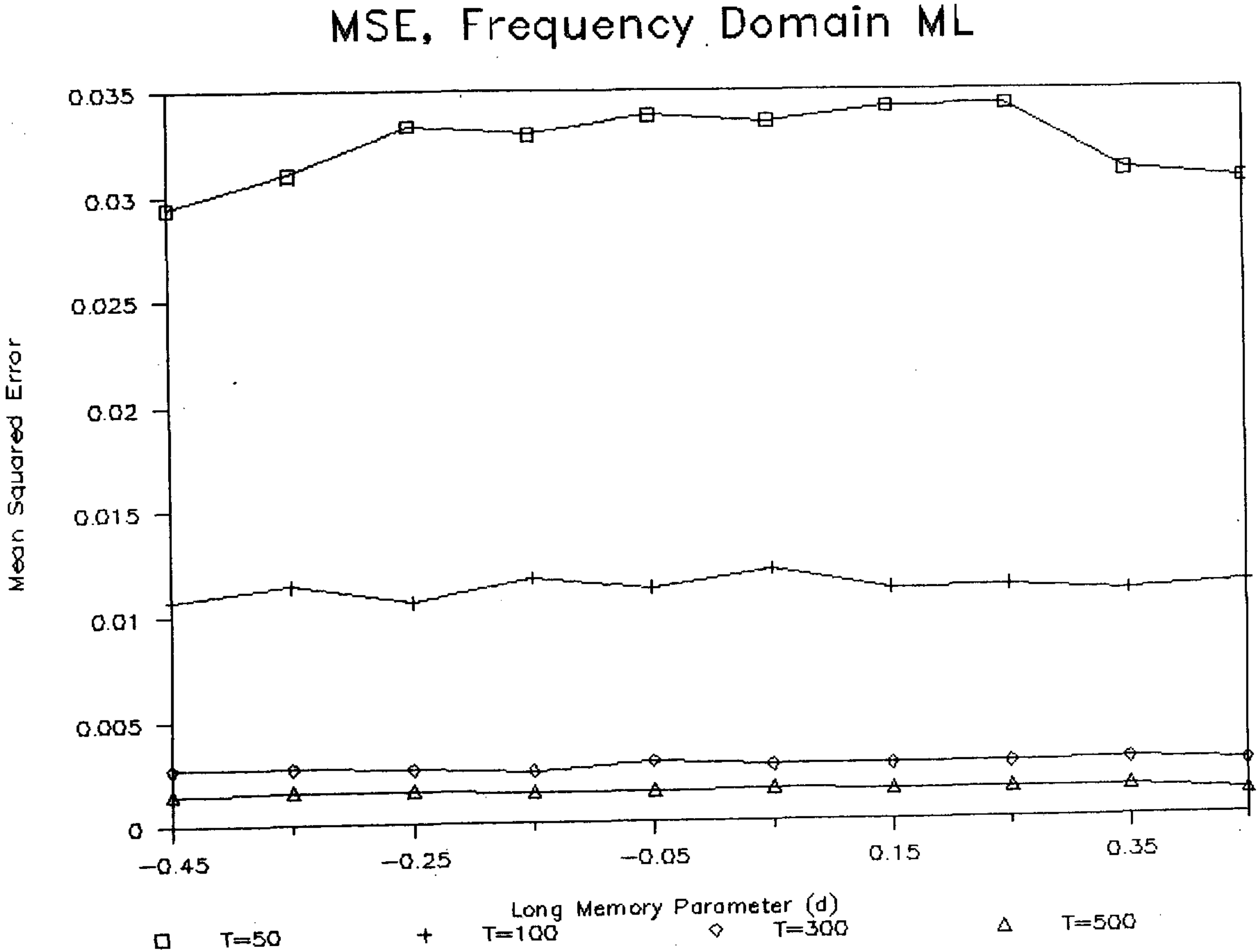

MSE, Time Domain ML, Mean Estimated
Mean Squared Error
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
-0.45
-0.25
-0.05
0.15
0.35
Long Memory Parameter (d)
T=50
T=100
T=300
T=500

FIGURE 6



MSE, Frequency Domain ML
Mean Squared Error
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
-0.45
-0.25
-0.05
0.15
0.35
Long Memory Parameter (d)
T=50
T=100
T=300
T=500

FIGURE 7

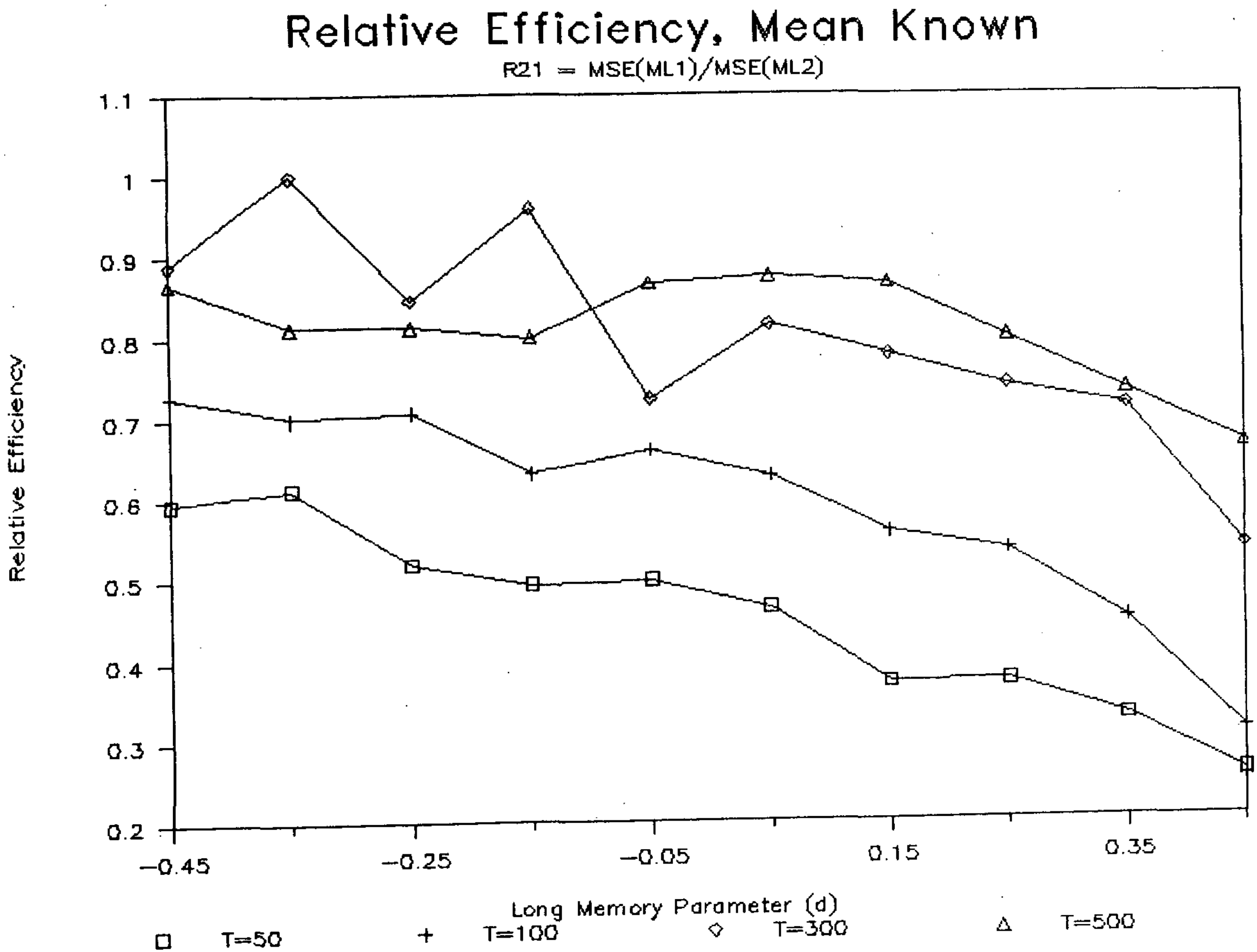
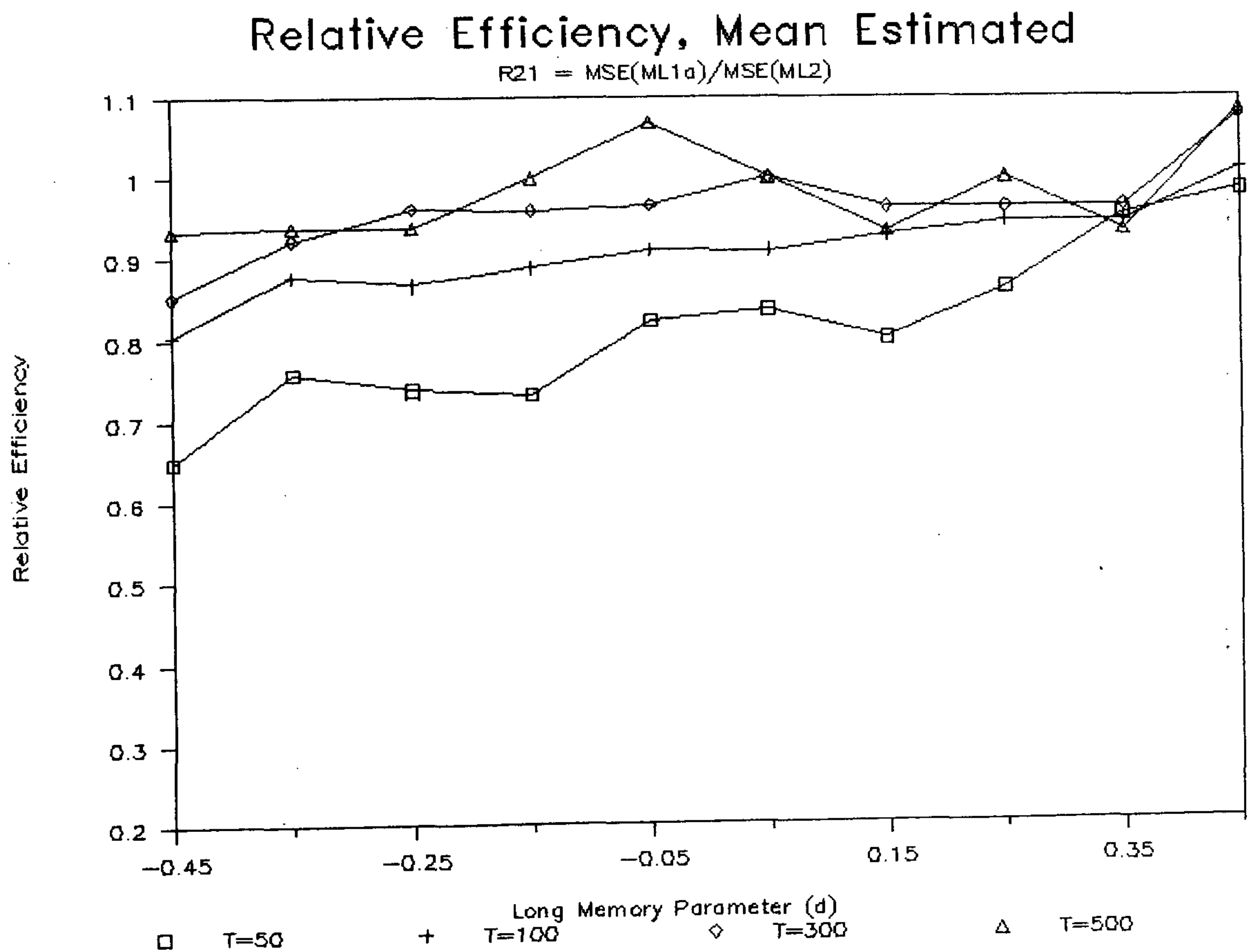

Relative Efficiency, Mean Known
R21 = MSE(ML1)/MSE(ML2)
Relative Efficiency
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
-0.45
-0.25
-0.05
0.15
0.35
Long Memory Parameter (d)
T=50
T=100
T=300
T=500

FIGURE 8


Relative Efficiency, Mean Estimated
R21 = MSE(ML1a)/MSE(ML2)
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Relative Efficiency
−0.45
−0.25
−0.05
0.15
0.35
Long Memory Parameter (d)
T=50
T=100
T=300
T=500