Finite sample efficiency of OLS in linear regression models with long-memory disturbances [1]

Christian Kleiber [2]

Fachbereich Statistik, Universität Dortmund,

D-44221 Dortmund, Germany.

Abstract

OLS is as efficient as GLS in the linear regression model with long-memory errors as the long-memory parameter approaches the boundary of the stationarity region, provided the model contains a constant term. This generalizes previous results of Samarov & Taqqu (*Journal of Time Series Analysis* 9, 1988, pp. 191-200) to the regression case and gives a further example of the 'high-correlation asymptotics' of Krämer & Baltagi (*Economics Letters* 50, 1996, pp. 13-17).

Keywords: Efficiency of OLS, linear regression, long memory.

JEL classification: C13, C22.

[1] Research supported by Deutsche Forschungsgemeinschaft (DFG), Sonderforschungsbereich 475.

[2] Phone: +49-231-755 5419, Fax: +49-231-755 5284, E-mail: Kleiber@statistik.uni-dortmund.de

1 Introduction

Consider the linear regression model

$$y = X\,\beta + u, \tag{1}$$

where X is nonstochastic and has full rank, the vector of errors, u, has mean zero, and $cov(u) = V$ is some positive definite matrix. The efficiency of the ordinary least squares (OLS) estimator $\hat{\beta} = (X'X)^{-1}X'y$ relative to the generalized least squares (GLS) estimator $\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$ has been studied for many covariance structures V, the bulk of this literature being devoted to large sample efficiency. There is also a smaller literature on 'high-correlation asymptotics' dealing with finite sample efficiency of OLS for varying V; the most prominent case is the linear regression model with AR(1) errors. It is well known that here OLS is efficient if the autocorrelation parameter tends to one, provided the model contains a constant term (Krämer, 1980). For additional examples see Krämer and Baltagi (1996). The present note shows that, in the linear regression model (1) with a constant term, OLS is also fully efficient in finite samples if $\{u_t\}$ is a long-memory process and the long-memory parameter approaches the boundary of the stationarity region.

2 Relative efficiency of OLS

The most prominent examples of long-memory models are fractionally integrated ARMA models (ARFIMA(p,d,q)) and fractional Gaussian noise. The former is defined by

$$\phi(B)\,(1-B)^d\,u_t = \theta(B)\,\epsilon_t, \tag{2}$$

where $\{\epsilon_t\}$ is white noise (in the weak sense) and $\phi(B)$ and $\theta(B)$ are polynomials in the backshift operator B of orders p and q; the latter is defined as first differences of fractional Brownian motion. See Baillie (1996) or Beran (1994) for surveys of long-memory models. Process (2) is stationary and causal if and only if $d < 0.5$ and the AR polynomial $\phi(z)$ obeys the usual stationarity and causality conditions for autoregressive models. Long memory corresponds to $d > 0$.

Following Krämer (1980), the relative efficiency of OLS for a series of length T is defined as

$$\text{eff } (T, d) = \frac{\text{tr } cov(X\tilde{\beta})}{\text{tr } cov(X\hat{\beta})} = \frac{\text{tr } X(X'V^{-1}X)^{-1}X'}{\text{tr } X(X'X)^{-1}X'VX(X'X)^{-1}X'}, \tag{3}$$

where tr denotes trace. For a full discussion of large sample efficiency of OLS in model (1) with stationary long-memory errors see Yajima (1991) who, however, defines eff (T, d) in terms of the determinants of the covariance matrices. We are only concerned with finite sample efficiency here.

Since scalar functions of V cancel out in (3), V may be replaced by the correlation matrix R, as is done in the proof of the following theorem.

Theorem 1 *Let $\langle X \rangle$ denote the column space of X and suppose $e = (1, \dots, 1)' \in \langle X \rangle$. Then*

$$\lim_{d \to 0.5} \text{eff } (T, d) = 1 \quad \textit{for all } T \in \mathbb{N}$$

in model (1) if

(a) $\{u_t\}$ is a stationary ARFIMA process.

(b) $\{u_t\}$ is fractional Gaussian noise.

Proof: (a) From Sowell (1992; see also Chung, 1994), the autocorrelation function of a stationary and causal ARFIMA process with distinct roots in the AR polynomial is given by

$$r(k) = r^*(k) \cdot \frac{\sum_{j=1}^{p} \xi_j \, A(d, k, \rho_j)}{\sum_{j=1}^{p} \xi_j \, A(d, 0, \rho_j)}, \quad k = 0, 1, 2, \dots \, , \tag{4}$$

where the ξ_j, $j = 1, \ldots, p$, are functions of the roots $1/\rho_j$ of $\phi(z)$, but not of d, and

$$A(d,k,\rho_j) = \sum_{i=-q}^{q} \psi_i \, B(d,k,p{+}i) \left[\rho_j^{2p} C(d, p+i-k, \rho_j) + C(d, k-p-i, \rho_j) - 1\right] .$$

Here

$$\psi_i = \sum_{m=0}^{q-|i|} \theta_m \theta_{m+|i|} \ ,$$

where the θ_m, $m = 0, \ldots, q$, are the MA coefficients,

$$B(d,k,h) = \frac{\Gamma(1-d-k)\,\Gamma(d-k+h)}{\Gamma(d-k)\,\Gamma(1-d-k+h)} \ ,$$

and

$$C(d,k,\rho_j) = \ {}_2F_1(d+k, 1; 1-d+k; \rho_j) \ ,$$

where ${}_2F_1$ is Gauss' hypergeometric series. Finally,

$$r^*(k) = \frac{\Gamma(1-d)\ \Gamma(k+d)}{\Gamma(d)\ \Gamma(k+1-d)} \tag{5}$$

is the autocorrelation function of fractionally integrated white noise.

Now $r^*(k) \to 1$ and $B(d,k,h) \to 1$ for $d \to 0.5$, whereas

$$\lim_{d \to 0.5} C(d,k,\rho_j) = \ {}_2F_1(1/2+k, 1; 1/2+k; \rho_j) = \ {}_1F_0(1;\rho_j) \ , \tag{6}$$

a geometric series independent of d and k (see Slater, 1966, chapter 1). Note that $|\rho_j| < 1$ by stationarity and causality, so this series converges. In (6), the interchange of the limits can be validated using a suitable version of the dominated convergence theorem (Billingsley, 1995, Theorem 16.8).

Overall, this gives $r(k) \to 1$ for $d \to 0.5$, hence the correlation matrix $R = (r(i-j))_{i,j}$ tends to $\bar{R} = ee'$, a matrix of ones. Now, $\bar{R}y \in \langle X \rangle$, $y \in \mathrm{I\!R}^T$, whenever $e \in \langle X \rangle$, and from Krämer and Baltagi (1996) the limiting efficiency is one in this case.

(b) From e.g. Baillie (1996, p. 16) the autocorrelation function of fractional Gaussian noise is given by

$$r(k) = \frac{1}{2}\left[(k+1)^{2H} - 2\,k^{2H} + (k-1)^{2H}\right],$$

which also implies $r(k) \to 1$ for $H \to 1$. Here, the 'Hurst coefficient' H is related to the fractional differencing parameter d of (a) as $d = H - 1/2$. Again, the theorem in Krämer and Baltagi (1996) yields a limiting efficiency of one if $e \in \langle X \rangle$. ∎

Part (a) generalizes earlier results of Samarov and Taqqu (1988, p. 198) in two directions. These authors obtained an optimal limiting efficiency of OLS for $X = e$ in finite samples when $\{u_t\}$ is fractionally integrated white noise (their α equals $-d$); as Theorem 1 shows, this carries over to all linear regressions with stationary ARFIMA errors that contain a constant term. Regarding (b), Beran (1994, pp. 149-151) found a high finite sample efficiency of the sample mean when errors are fractional Gaussian noise. Again, this carries over to all linear regressions that contain an intercept.

Although $e \in \langle X \rangle$ seems to be the most important case in empirical applications, it is of some interest to identify sets of regressors for which OLS should be avoided.

Corollary 2 *Suppose $e \notin \langle X \rangle \cup \langle X \rangle^{\perp}$, where $^{\perp}$ denotes the orthogonal complement. Then*

$$\lim_{d \to 0.5} \text{eff}\,(T, d) = 0 \quad \textit{for all } T \geq 2$$

in model (1) if

(a) $\{u_t\}$ is a stationary ARFIMA process.

(b) $\{u_t\}$ is fractional Gaussian noise.

Proof: In the proof of Theorem 1 it was shown that $r(1) = r_{i,i+1} \to 1$ if $d \to 0.5$ for both error structures, hence the result follows from the theorem in Krämer (1984). ∎

This implies, for example, that OLS is inefficient in the linear regression model with a linear trend, but no constant term, when the errors are fractionally integrated ARMA (or fractional Gaussian noise) and d (or H) is sufficiently large.

3 Numerical illustration

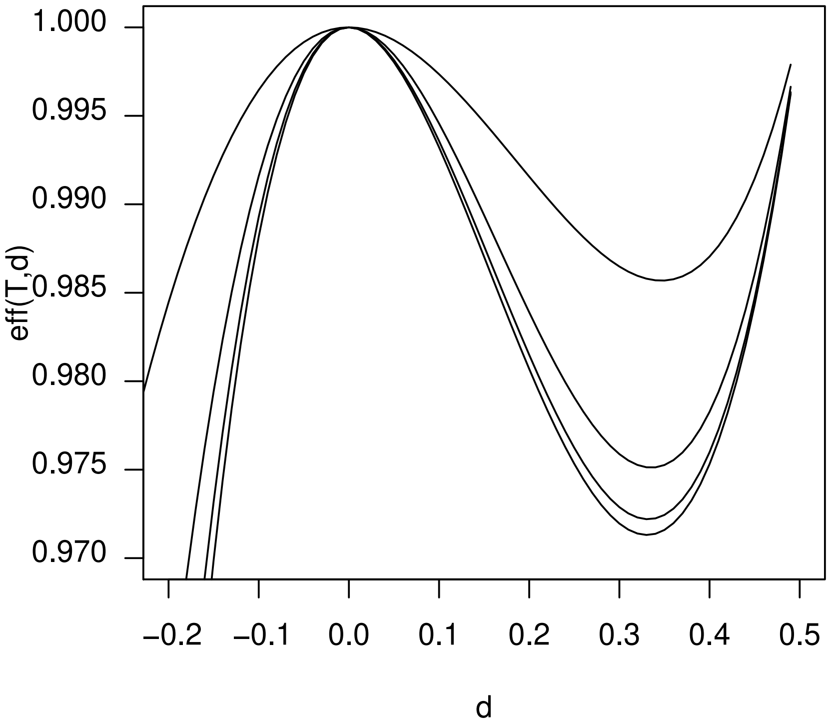
Table 1 presents relative efficiencies for a regression on a constant and a linear trend for $T = 10, 50, 100, 200, 500, 1000$ and selected values of $d \in [-0.5, 0.49]$, and Figure 1 graphs the efficiency function $\text{eff}(T, d)$ for $T = 10, 50, 200, 1000$ and $d \in [-0.2, 0.49]$, evaluated in steps of 0.01. The errors are assumed to be fractionally integrated white noise (cf. (5)).

Table 1: Efficiency of OLS relative to GLS

d	T=10	T=50	T=100	T=200	T=500	T=1000
-0.4	0.9304	0.8085	0.7641	0.7267	0.6871	0.6631
-0.3	0.9657	0.9001	0.8786	0.8614	0.8443	0.8348
-0.2	0.9845	0.9606	0.9531	0.9475	0.9425	0.9399
-0.1	0.9965	0.9916	0.9902	0.9893	0.9885	0.9881
0	1	1	1	1	1	1
0.1	0.9974	0.9946	0.9940	0.9936	0.9933	0.9932
0.2	0.9916	0.9839	0.9824	0.9815	0.9809	0.9807
0.3	0.9865	0.9759	0.9739	0.9729	0.9722	0.9720
0.4	0.9870	0.9783	0.9768	0.9760	0.9755	0.9753
0.45	0.9914	0.9861	0.9852	0.9847	0.9844	0.9843
0.49	0.9979	0.9966	0.9964	0.9963	0.9963	0.9962

Figure 1: Efficiency of OLS relative to GLS

T=10 , T=50 , T=200 , T=1000 (from top)



The results are quite similar to those of Samarov and Taqqu for $X = e$ (they cannot be identical as the relative efficiency depends on the regressors). In particular, eff (T, d) is decreasing in T for fixed d; also, there is a local minimum around $d = 0.32$. In general, the relative efficiency is uniformly above 0.97 for positive d, hence the loss of efficiency associated with OLS is not severe in the present context.

Acknowledgements

I am grateful to Walter Krämer for comments on a previous draft and to Achim Zeileis for programming the efficiency function.

References

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

Beran, J. 1994. Statistics for Long-Memory Processes. Chapman and Hall, London.

Billingsley, P. 1995. Probability and Measure, 3rd ed. John Wiley, New York.

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

Krämer, W. 1980. Finite sample efficiency of ordinary least squares in the linear regression model with autocorrelated errors. Journal of the American Statistical Association 75, 1005-1009.

Krämer, W. 1984. High correlation among errors and the relative efficiency of ordinary least squares in linear models. Statistical Papers 25, 135-142.

Krämer, W., Baltagi, B. 1996. A general condition for an optimal limiting efficiency of OLS in the general linear regression model. Economics Letters 50, 13-17.

Samarov, A., Taqqu, M.S. 1988. On the efficiency of the sample mean in long-memory noise. Journal of Time Series Analysis 9, 191-200.

Slater, L.J. 1966. Generalized Hypergeometric Functions. Cambridge University Press, Cambridge.

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

Yajima, Y. 1991. Asymptotic properties of the LSE in a regression model with long-memory stationary errors. The Annals of Statistics 19, 158-177.