Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares with Trending Regressors and Stationary Autoregressive Disturbances[1]

by

Walter Krämer
Fachbereich Statistik, Universität Dortmund
D - 44221 Dortmund, Germany

Abstract

This note generalizes previous results on the asymptotic equivalence of Ordinary and Generalized Least Squares estimates in Linear Regression models with trending data.

This note considers the relative efficiency of OLS versus GLS in the linear regression model

$$y_t = x_t'\beta + u_t \quad (t = 1, 2, \ldots), \tag{1}$$

where x_t and β are $K \times 1$ and where the unobservable disturbances u_t are autocorrelated but independent of the regressors $x_t = (x_{t1}, \ldots, x_{tK})'$. It is well known that, given the regressors, OLS is in general no longer BLUE when disturbances are correlated, but as GLS (the BLUE) is often only of academic interest due to lack of knowledge of the disturbance correlation structure, there has been an enormous interest in statistics and econometrics in the relative efficiency of OLS (Watson 1968, Krämer 1980, Krämer and Donninger 1986, Busse et al. 1994 among many others).

[1] Research supported by Deutsche Forschungsgmeinschaft through SFB 475

One strand of this literature, originating with Grenander (1954), is concerned with conditions on regressors and disturbances which guarantee that OLS is at least asymptotically efficient (Rosenblatt 1956, Chipman 1979, Krämer 1982, 1986, Phillips and Park 1988, Krämer and Hassler 1997). One sufficient condition for the asymptotic efficiency of OLS that emerges in this literature is that the regressors are in some sense trending" (for a precise definition see below). In conjunction with stationarity, in particular stationary autoregressive disturbances, this is then shown to imply that the respective limiting distributions of OLS and GLS are identical.

The present note extends and unifies this literature by suggesting a generic form of "trend", and by showing that it is this generic property of trending data which implies the asymptotic equivalence of OLS and GLS.

In what follows the disturbances u_t from (1) are assumed stationary $AR(p)$,

$$u_t + \rho_1 u_{t-1} + \ldots + \rho_p u_{t-p} = \varepsilon_t, \tag{2}$$

where the ε_t's are i.i.d. $(0, \sigma^2)$, and where stationarity implies that all roots of the polynomial $1 + \rho_1 z + \ldots + \rho_p z^p$ are outside the unit circle.

Ignoring observations $1, \ldots, p$, which are asymptotically irrelevant, the GLS–estimator $\tilde{\beta}$ for β is obtained by applying OLS to

$$\begin{aligned} \tilde{y}_t &= \tilde{x}_t'\beta + \varepsilon_t, \quad \text{where} && (3)\\ \tilde{x}_t &= x_t + \rho_1 x_{t-1} + \ldots + \rho_p x_{t-p} \quad \text{and} && (4)\\ \tilde{y}_t &= y_t + \rho_1 y_{t-1} + \ldots + \rho_p y_{t-p} \quad (t > p) && (5) \end{aligned}$$

i.e. $\hat{\beta} = (\sum \tilde{x}_t \tilde{x}_t')^{-1} \sum \tilde{x} \hat{y}_t$ and

$$\tilde{\beta} - \beta = \left(\sum_{t=p+1}^{T} \tilde{x}_t \tilde{x}_t' \right)^{-1} \sum_{t=p+1}^{T} \tilde{x}_t \varepsilon_t. \tag{6}$$

This compares with

$$\hat{\beta} - \beta = \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} \sum_{t=1}^{T} x_t u_t \tag{7}$$

for the OLS–estimator $\hat{\beta}$.

The regressors x_{ti} are assumed to be "trending", by which I mean that they satisfy an invariance principle

$$\frac{1}{g_i(T)} x_{[Tr],i} \xrightarrow{d} B_i(r) \quad \text{as} \quad T \to \infty, \tag{8}$$

where $\xrightarrow{d}$ denotes convergence in distribution, $[Tr]$ is the integer part of Tr, $g_i(T) \to \infty$ and $B_i(r)$ is some non–zero random element in $D[0,1]$ (which might be degenerate; $D[0,1]$ is the set of all real valued functions on the unit interval, endowed with the Skorohod topology, see Billingsley 1968, chapter 3). This definition includes the well known cases of stochastic $I(1)$–regressors, where $g_i(T) = \sqrt{T}$ and where (under suitable regularity conditions) $B_i(r)$ is Brownian Motion. It also includes the case of nonstochastic polynomial regressors, e.g. $x_{it} = t^i$, where $g_i(T) = T^i$ and where $B_i(r)$ is the function $f_i(r) = r^i$. And it also includes the case of nonstationary fractionally integrated regressors, where

$$(1 - L)^d x_{ti} = \varepsilon_{ti} \tag{9}$$

with $d > \frac{1}{2}$ and stationary ARMA ε's, and where $g_i(T) = \sqrt{T^{2d-1}}$ and $B_i(r)$ is fractional Brownian Motion (Sowell 1990, Chung 1995). However, the definition does not allow for exponential trends, as it is easily seen that invariance principles like (8) do then no longer hold.

The main theorem is concerned with the case where <u>all</u> regressors in the model (1) are trending (extensions to the case where there are also nontrending regressors, in particular an intercept, are discussed in section 4). To that purpose, let $g(T) := diag\,(g_1(T), \ldots, g_K(T))$ and assume in addition to (8) that

$$[g(T)]^{-1} x_{[Tr]} \xrightarrow{d} B(r), \tag{10}$$

where $B(r)$ is a random element in $D[0,1]^K$ (with components $B_i(r)$ defined in (8)) and that $\int_0^1 B(r)B(r)'dr$ is invertible with probability 1. Since u_t and x_ρ are independent for all t and s, this implies that

$$\begin{pmatrix} [g(T)]^{-1} & x_{[Tr]} \\ T^{-\frac{1}{2}} & \sum_{s=1}^{[Tr]} u_s \end{pmatrix} \xrightarrow{d} \begin{pmatrix} B(r) \\ W(r) \end{pmatrix}, \tag{11}$$

where $W(r)$ is Brownian Motion, independent of $B(r)$, with variance

$$\bar{\sigma}^2 = \frac{\sigma^2}{(1 + \rho_1 + \ldots + \rho_p)^2}. \tag{12}$$

THEOREM: *Under the assumptions made above, we have*

$$\sqrt{T} g(T)(\hat{\beta} - \beta) \xrightarrow{d} \left[\int_0^1 B(r)B(r)'dr\right]^{-1} \left[\int_0^1 B(r)dW(r)\right] \tag{13}$$

and the same applies to GLS.

CORROLLARY: Since $g(T) \to \infty$, we have

$$\hat{\beta} - \beta = o_p \frac{1}{\sqrt{T}}, \tag{14}$$

i.e. the OLS-estimator converges to the true parameter vector faster than in the standard case with no trend in the regressors.

PROOF OF THE THEOREM: The relationship (13) follows immediately from

$$\frac{1}{T}[g(T)]^{-1} \sum_{t=1}^{T} x_t x_t' [g(T)]^{-1} \xrightarrow{d} \int_0^1 B(r)B(r)'dr \tag{15}$$

and

$$\frac{1}{\sqrt{T}}[g(T)]^{-1}\sum_{t=1}^{T} x_t u_t \xrightarrow{d} \int_0^1 B(r)dW(r), \tag{16}$$

where the first relationship follows from (10) and the continuous mapping theorem (Billingsley 1968, p. 30), and where the latter convergence follows from the independence of $W(r)$ and $B(r)$ and a general theorem on the convergence to stochastic integrals in Hansen (1992, p. 491).

The analogue for GLS follows along similar lines. Emulating the proof of theorem 2.2 in Phillips and Park (1988, p. 114), one verifies, using

$$g(T)^{-1}\tilde{x}_{[Tr]} \quad = \quad g(T)^{-1}(1+\rho_1+\ldots+\rho_p)x_{[Tr]} + o_p(1) \tag{17}$$

and

$$T^{-\frac{1}{2}}\sum_{s=1}^{[T_r]}\varepsilon_s \quad = \quad T^{-\frac{1}{2}}(1+\rho_1+\ldots+\rho_p)\sum_{s=1}^{[T_r]} u_t + o_p(1) \tag{18}$$

that the regressors $\tilde{x}_t$ and the disturbances ε_t in (3) satisfy the invariance principle

$$\begin{pmatrix} g(T)^{-1}\tilde{x}_{[Tr]} \\ \\ T^{-\frac{1}{2}}\sum_{s=1}^{[T_r]}\varepsilon_s \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \tilde{B}(r) \\ \\ \tilde{W}(r) \end{pmatrix} \tag{19}$$

where

$$\begin{aligned} \tilde{B}(r) &= (1+\rho_1+\ldots+\rho_p)B(r) \\ \tilde{W}(r) &= (1+\rho_1+\ldots+\rho_p)W(r) \end{aligned} \tag{20}$$

and where $\tilde{W}(r)$ is independent of $\tilde{B}(r)$. In conjunction with

$$\frac{1}{T}[g(T)]^{-1}\sum_{t=1}^{T}\tilde{x}_t\tilde{x}_t'[g(T)]^{-1} \xrightarrow{d} \int_0^1 \tilde{B}(r)\tilde{B}(r)'dr, \tag{21}$$

and

$$\frac{1}{\sqrt{T}}[g(T)]^{-1}\sum_{t=1}^{[Tr]}\tilde{x}_t\varepsilon_t \xrightarrow{d} \tilde{B}(r)d\tilde{W}(r), \tag{22}$$

this implies (as the term $\sum_{i=1}^{p}\rho_i$ cancels out) that

$$\begin{aligned}\sqrt{T}g(T)(\tilde{\beta}-\beta) \quad &\xrightarrow{d} \quad \left[\int_0^1 \tilde{B}(r)\tilde{B}(r)'dr\right]^{-1}\left[\int_0^1 \tilde{B}(r)d\tilde{W}(r)\right] \\ &= \quad \left[\int_0^1 B(r)B(r)'dr\right]^{-1}\left[\int_0^1 B(r)dW(r)\right], \end{aligned} \tag{23}$$

proving the theorem.

A central requirement for the theorem above is that <u>all</u> regressors are trending, and that the limiting matrix $\int_0^1 B(r)B(r)'dr$ is invertible with probability 1. Both conditions are often violated in empirical applications.

Consider for instance the case of a singular limiting matrix $\int_0^1 B(r)B(r)'dr$, for example $K=2$, $x_{t1}=t$, $x_{t2}=t+(-1)^t$. Here we have $g_i(T)=T$,

$$\sum_{t=n}^{T} x_t x_t' = \begin{bmatrix} \sum_{t=1}^{T} t^2 & \sum_{t=1}^{T} t^2+(-1)^t \\ \sum_{t=1}^{T} t^2+(-1)^t & \sum_{t=1}^{T} t^2+2(-1)^t+T \end{bmatrix} \tag{24}$$

and

$$\frac{1}{T^3}\sum_{t=1}^{T} x_t x_t' \to \int_0^1 B(r)B(r)'dr = \frac{1}{3}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \tag{25}$$

The limiting matrix $\int_0^1 B(r)B(r)'dr$ is therefore noninvertible with probability 1, and the right–hand side of (13) does not exist.

The limiting distribution of both OLS and GLS is in such cases best derived by a reparametrization of the model:

$$
\begin{aligned}
y_t &= \beta_1 t + \beta_2(t + (-1)^t) + u_t \qquad (26)\\
&= (\beta_1 + \beta_2)t + \beta_2(-1)^t + u_t\\
&= \alpha_1 t + \alpha_2(-1)^t + u_t
\end{aligned}
$$

Now we have one regressor with trend and one without, and, from a general result in Rosenblatt (1956), we know that the limiting distribution of $\sqrt{T}(\hat{\alpha}_2 - \alpha_2)$ has a larger variance than the limiting distribution of $\sqrt{T}(\tilde{\alpha}_2 - \alpha_2)$. As the limiting distribution of both $\hat{\beta}_1 = (\hat{\alpha}_1 + \hat{\alpha}_2)$ and $\tilde{\beta}_1 = (\tilde{\alpha}_1 + \tilde{\alpha}_2)$ is dominated by $\hat{\alpha}_1$ and $\tilde{\alpha}_1$, respectively, this implies that the existence of a second trend destroys both the superconsistency and the asymptotic efficiency of OLS.

However, there is still hope that in a model with nontrending regressors, OLS retains both superconsistency and efficiency at least for the coefficients of the trending regressors. Consider the model:

$$
y_t = x_t'\beta + z_t'\alpha + u_t, \qquad (27)
$$

where the components of x_t are trending and the components of z_t are not. In view of the Frisch–Waugh–Theorem (see Fiebig et al. 1996), the OLS–estimate for β in (27) is numerically identical to the OLS–estimate for β in

$$
y_t^* = x_t^*\beta + u_t, \qquad (28)
$$

where y_t^* and x_t^* are the residuals from a regression on z_t. Therefore, if the x_t^* satisfy the invariance principle (10), the OLS estimate for β in (28) is asymptotically as efficient as the GLS–estimate in (28).

Unfortunately, the GLS–estimate for β in (28) need not be numerically identical to the GLS–estimate for β in (27). Fiebig et al. (1996) give conditions for this be the case, from which it is seen that at least for the most important case of a regression with an intercept, i.e. $z_t = 1$ and $x_t^* = x_t - \bar{x}$, the superconsistency and limiting efficiency of OLS prevails.

Billingsley, P. (1968): *Convergence of probability measures*, Chicago.

Busse, R.; Jeske, R. and Krämer, W. (1994): "Efficiency of least–squares estimation of polynomial trend when residuals are autocorrelated", *Economics Letters* 45, 267–271.

Chipman, J.S. (1979): "Efficiency of least–squares estimation of linear trend when residuals are autocorrelated." *Econometrica* 47, 115 – 127.

Chung, Ch.-F. (1995): " Sample variance, sample covariances, and linear regression of stationary multivariate long memory processes." Mimeo, Michigan State University.

Fiebig, D. G.; Bartels, R. and Krämer, W. (1996): "The Frish–Waugh–Theorem and Generalized Least Squares", *Econometric Reviews* 15, 431–443.

Grenander, U. (1954): "On the estimation of regression coefficients in the case of an autocorrelated disturbance." *Annals of Mathematical Statistics* 25, 252 – 272.

Hansen, B.E. (1992): "Convergence to stochastic integrals for dependent heterogeneous processes." *Econometric Theory* 8, 489 – 500.

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. (1982): "Note on estimating linear trend when residuals are autocorrelated." *Econometrica* 50, 1065 – 1067.

Krämer, W. (1985): "Least squares regression when the independent variable follows an ARIMA–process." *Journal of the American Statistical Association* 81, 150 – 154.

Krämer, W. and Donninger, C. (1986): "Spatial autocorrelation among errors and the relative efficiency of OLS in the Linear Regression Model." *Journal of the American Statistical Association* 82, 577 – 579.

Krämer, W. and Hassler, U. (1997): "Limiting efficiency of OLS vs. GLS when regressors are fractionally integrated." Technical Report No. 1, SFB 475, Fachbereich Statistik, Universität Dortmund.

Phillips, P.C.B. and Park, J.Y. (1988): "Asymptotic Equivalence of OLS and GLS in Regressions with Integrated Regressors." *Journal of the American Statistical Association* 83, 111 – 115.

Rosenblatt, H.M. (1956): "Some regression problems in time series analysis." *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 165 – 186.

Sowell, F. (1990): "The fractional unit root distribution." *Econometrica* 58, 495 – 505.

Watson, G. (1968): "Linear Least Squares regression." *Annals of Mathematical Statistics* 38, 1679 – 1698.