OLS–based Asymptotic Inference in Linear Regression Models with Trending Regressors and AR(p)–Disturbances[1]

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

Francesc Marmol
Departamento de Estadistica y Econometria, Universidad Carlos III, Madrid, Spain

Summary

We show that OLS and GLS are asymptotically equivalent in the linear regression model with $AR(p)$–disturbances and a wide range of trending regressors, and that OLS–based statistical inference is still meaningful after proper adjustment of the test–statistics.

1 Notation and assumptions

We consider the standard 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 u_t is a stationary, zero mean $AR(p)$–process,

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

[1]Research supported by Deutsche Forschungsgemeinschaft.

with $iid(0,\sigma^2)$ ε_t's and all roots of the polynomial $1+\rho_1 z+\ldots+\rho_p z^p$ outside the unit circle. Our main concern is OLS–based statistical inference when the regressors x_t are independent of the disturbances and "trending", by which we mean that they satisfy an invariance principle

$$\frac{1}{g_i(T)} x_{[Tr],i} \xrightarrow{d} B_i(r) \quad \text{as} \quad T \to \infty, \tag{3}$$

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, possibly degenerate random element in $D[0,1]$ (the set of all real–valued functions on the unit interval who are right continuous and have left–hand–limits, endowed with the Skorohod–Topology; see Billingsley 1968, chapter 3). Also, we assume that

$$g(T)^{-1} x_{[Tr]} \xrightarrow{d} B(r), \tag{4}$$

where $g(T) = diag(g_1(T), \ldots, g_k(T))$ and where $B(r)$ is a random element in $D[0,1]^k$ with components $B_i(r)$, and that $\int_0^1 B(r)B(r)'dr$ is invertible with probability 1.

The crucial condition (3) covers various special cases: (i) Stochastic $I(1)$–regressors, where $g_i(T) = \sqrt{T}$ and where (under suitable regularity conditions) $B_i(r)$ is Brownian Motion. (ii) Nonstochastic polynomial regressors, where $x_{it} = t^i$ and $g_i(T) = T^i$, and where $B_i(r) = r^i$. (iii) Nonstationary fractionally integrated regressors, where $(1-L)^d x_{ti} = \varepsilon_{ti}$ with $d > \frac{1}{2}$ and stationary ARMA ε_{ti}'s, where $g_i(T) = \sqrt{T^{2d-1}}$ and where $B_i(r)$ is fractional Brownian Motion (Sowell 1990, Chung 1995, Dolado and Marmol 1998). It does not cover exponential trends, as it is easily seen that invariance principles like (3) do then no longer hold.

The topic of the paper is the asymptotic performance of the OLS–estimator

$$\hat{\beta} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t y_t, \tag{5}$$

both relative to GLS and as regarding inference, generalizing Grenander (1954), Rosenblatt (1956), Krämer (1985, 1998), Phillips and Park (1988), Krämer and Hassler (1998) or Dolado and Marmol (1998), who either consider only special cases of trend or focus on the asymptotic efficiency of OLS, disregarding inference. We show that OLS is asymptotically efficient, thus establishing the invariance principle (3) as the heart of the well known efficiency results in the papers above, and show that OLS–based F–tests are still asymptotically valid in the context of autocorrelated disturbances if the OLS–based variance estimator is divided by an estimator of the long–term variance of the disturbances. This was first noted by Krämer (1987) and Phillips and Park (1988) in the context of polynomial and $I(1)$–regressors, but extends to all types of trend comprised by (3).

2 Asymptotic properties of OLS–based coefficient estimates

We first compare the properties of OLS to those of the OLS–estimator $\tilde{\beta}$, which in the present context is obtained by applying OLS to

$$\begin{aligned} \tilde{y}_t &= \tilde{x}_t'\beta + \varepsilon_t, \quad \text{where} & (6)\\ \tilde{x}_t &= x_t + \rho_1 x_{t-1} + \ldots + \rho_p x_{t-p} \quad \text{and} & (7)\\ \tilde{y}_t &= y_t + \rho_1 y_{t-1} + \ldots + \rho_p y_{t-p} \quad (t > p) & (8)\end{aligned}$$

and where observations $t = 1, \ldots, p$, which are asymptotically irrelevant, are ignored.

THEOREM 1: *Let $W(r)$ be Brownian Motion, independent of $B(r)$, with variance $\tilde{\sigma}^2 = \sigma^2/(1 + \rho_1 + \ldots + \rho_p)^2$. The limiting distributions as $T \to \infty$ of $\sqrt{T}g(T)(\tilde{\beta} - \beta)$ and $\sqrt{T}g(T)(\hat{\beta} - \beta)$ are then identical and given by*

$$[\int_0^1 B(r)B(r)'dr]^{-1} \int_0^1 B(r)dW(r). \quad (9)$$

PROOF: We have

$$\hat{\beta} - \beta = \sum_{t=1}^{T} (x_t x_t')^{-1} \sum_{t=1}^{T} x_t u_t \ , \tag{10}$$

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

$$\frac{1}{T} g(T)^{-1} \sum_{t=1}^{T} x_t x_t' g(T)^{-1} \xrightarrow{d} \int_0^1 B(r)B(r)'dr \quad \text{and} \tag{12}$$

$$\frac{1}{\sqrt{T}} g(T)^{-1} \sum_{t=1}^{T} x_t u_t \xrightarrow{d} \int_0^1 B(r)dW(r), \tag{13}$$

where (12) follows from (4) and the continuous mapping theorem (Billingsley 1968, p. 30) and where (13) 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). Taken together, (12) and (13) give (9) as the limiting distribution of OLS.

As to GLS, we have

$$\begin{aligned} g(T)^{-1}\tilde{x}_{[Tr]} &= g(T)^{-1}(1+\rho_1+\ldots+\rho_p)x_{[Tr]} + o_p(1) \quad \text{and} & (14) \\ T^{-\frac{1}{2}} \sum_{s=1}^{[Tr]} \varepsilon_s &= T^{-\frac{1}{2}}(1+\rho_1+\ldots+\rho_p) \sum_{s=1}^{[Tr]} u_s + o_p(1), & (15) \end{aligned}$$

which implies, emulating the proof of Theorem 2.2 in Phillips and Park (1988, p. 114) that

$$\begin{pmatrix} g(T)^{-1}\tilde{x}_{[Tr]} \\ (T)^{-1} \sum_{s=1}^{[Tr]} \varepsilon_s \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \tilde{B}(r) \\ \tilde{u}(r) \end{pmatrix} . \tag{16}$$

However,

$$\begin{aligned} \tilde{B}(r) &= (1+\rho_1+\ldots+\rho_p)B(r) \quad \text{and} & (17) \\ \tilde{W}(r) &= (1+\rho_1+\ldots+\rho_p)W(r), & (18) \end{aligned}$$

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

$$\tilde{\beta} - \beta = \left(\sum_{t=1}^{T} \tilde{x}_t \tilde{x}_t' \right)^{-1} \sum_{t=1}^{T} \tilde{x}_t \varepsilon_t, \tag{19}$$

this implies that

$$\begin{aligned} \sqrt{T} g(T)(\tilde{\beta} - \beta) & \xrightarrow{d} [\int_0^1 \tilde{B}(r)\tilde{B}(r)' dr]^{-1} \int_0^1 \tilde{B}(r) d\tilde{W}(r) \\ & = [\int_0^1 B(r)B(r)' dr]^{-1} \int_0^1 B(r) dW(r), \end{aligned} \tag{20}$$

as the term $1 + \rho_1 + \ldots + \rho_p$ cancels out. □

Theorem 1 shows also, in view of $g(T) \to \infty$, that OLS and GLS are consistent and converge to the true parameter vector faster than in the case of nontrending regressors, confirming well known results from regression analysis ("superconsistency"). One can also extend Theorem 1 to include the feasible GLS–estimator, which is obtained by plugging estimated ρ's into (7) and (8). It is easy to show that these estimates, if based on OLS–residuals $y_t - x_t'\hat{\beta}$, are consistent, and that the limiting distribution (9) obtains for feasible GLS as well.

To derive the limiting null distribution of the F-test, which will be the concern of section 3, it is more useful to normalize the estimation errors $\hat{\beta} - \beta$ differently, as is done in our next result.

THEOREM 2: *Assume that $B(r)$ can be expressed as a uniformly continuous functional of a K–dimensional Brownian Motion. Then, as $T \to \infty$, both $(\Sigma x_t x_t')^{-\frac{1}{2}}(\hat{\beta} - \beta)$ and $(\Sigma(x_t x_t'))^{-\frac{1}{2}}(\tilde{\beta} - \beta)$ tend in distribution to $N(0, \tilde{\sigma}^2 I)$.*

PROOF: From Theorem 1 and the continuous mapping theorem, we deduce that

$$(\Sigma x_t x_t')^{\frac{1}{2}}(\hat{\beta} - \beta) \xrightarrow{d} (\int_0^1 B(r)B(r)' dr))^{-\frac{1}{2}} \int_0^1 B(r) dW(r). \tag{21}$$

As $B(r)$ is by assumption a continuous functional of Brownian Motion $\tilde{B}$, we deduce from Phillips and Park (1988, p. 114) that

$$\int_0^1 B(r)dW(r)|_{\sigma(\tilde{B})} \sim N\left(0, \tilde{\sigma}^2 \int_0^1 B(r)B(r)'dr\right), \tag{22}$$

from which (21) follows.

As to GLS, we have

$$(\Sigma x_t x_t')^{\frac{1}{2}}(\tilde{\beta}-\beta) = \left[(\Sigma x_t x_t')^{\frac{1}{2}}\,(\Sigma \tilde{x}_t \tilde{x}_t')^{\frac{1}{2}}\right]\left[(\Sigma \tilde{x}_t \tilde{x}_t')^{\frac{1}{2}}\,\Sigma \tilde{x}_t' \varepsilon_t\right], \tag{23}$$

where the first term tends to $(1+\rho_1+\ldots+\rho_p)^{-1}I_k$ and the second term tends to $N(0,\sigma^2 I)$, which completes the proof of the theorem. □

The additional requirement in Theorem 2 that B(r) can be written as a functional of Brownian Motion does not seem to be very restrictive. It is for instance satisfied for arbitrary $I(d)$ regressors $(d > 1/2)$, including d = 1, so the cases that are of interest in practice are covered. Also, an analogous version of Theorem 2 holds which establishes that both $(\Sigma \tilde{x}_t \tilde{x}_t')^{\frac{1}{2}}(\hat{\beta}-\beta)$ and $(\Sigma \tilde{x}_t' \tilde{x}_t)^{\frac{1}{2}}(\tilde{\beta}-\beta)$ tend in distribution to $N(0,\sigma^2 I)$.

3 Asymptotic inference

Next we consider the standard OLS–based F-Test of the hypothesis

$$H_0 : R\beta = r, \tag{24}$$

where R is $q \times k$ with rank $q(q < k)$. The test statistic is

$$F = (R\hat{\beta}-r)'[R(\Sigma x_t x_t')^{-1}R']^{-1}(R\hat{\beta}-r)/s^2, \tag{25}$$

where

$$s^2 = \sum_{t=1}^{T}(y_t - x_t'\hat{\beta})^2/(T-k). \tag{26}$$

It has long been known that the most serious implications of autocorrelated disturbances is not the resulting inefficiency of OLS but the misleading inference when standard tests are used. One way out of this dilemma are the well known autocorrelation–consistent covariance matrix estimates, but in the present context, the remedy is much simpler.

THEOREM 3: *Given H_0 and the assumptions from Theorem 2, we have, as* $T \to \infty$

$$F \xrightarrow{d} \frac{\tilde{\sigma}^2}{\sigma_u^2}\chi_q^2, \tag{27}$$

where $\sigma_u^2 = E(u_t^2) = \sigma^2(1 + \rho_1^2 + \ldots + \rho_p^2)$.

PROOF: We have

$$\begin{aligned}
&[R\left[\Sigma x_t x_t'\right)^{-1}R'\Big]^{-\frac{1}{2}}(R\hat{\beta} - r) \\
&= \left[R\left(\frac{1}{T}g(T)^{-1}\Sigma x_t x_t' g(T)^{-1}\right)^{-1}R'\right]^{-\frac{1}{2}}\sqrt{T}g(T)(R\hat{\beta} - r) \\
&= \left[R\left(\frac{1}{T}g(T)^{-1}\Sigma x_t x_t' g(T)^{-1}\right)^{-1}R'\right]^{-\frac{1}{2}}R\sqrt{T}g(T)(\hat{\beta} - \beta) \quad (\text{under} H_0).
\end{aligned}$$

Using (9), (12) and the continuous mapping theorem, we have that under H_0,

$$\begin{aligned}
&[R\left[\Sigma x_t x_t'\right)^{-1}R'\Big]^{-\frac{1}{2}}(R\hat{\beta} - r) \\
&\xrightarrow{d} \left[R\left(\int_0^1 B(r)B(r)'dr\right)^{-1}R'\right]^{-\frac{1}{2}} R\left(\int_0^1 B(r)B(r)'dr\right)^{-1}\int_0^1 B(r)dW(r) \\
&\equiv \left[R\left(\int_0^1 B(r)B(r)'dr\right)^{-1}R'\right]^{-\frac{1}{2}} R\left(\int_0^1 B(r)B(r)'dr\right)^{-\frac{1}{2}}\mathcal{N}(0, \tilde{\sigma}^2 I_k), \quad (28)
\end{aligned}$$

where "$\equiv$" denotes equality in distribution.

Expression (28) implies that

$$
\begin{aligned}
(28) \quad \equiv\ & \mathcal{N}\left[0, \tilde{\sigma}^2 \left(R\left(\int_0^1 B(r)B(r)'dr\right)^{-1} R'\right)^{-1}\right. \\
& \left. \times\ R\left(\int_0^1 B(r)B(r)'dr\right)^{-1} R'\right] \\
=\ & \mathcal{N}(0, \tilde{\sigma}^2 I_q). \qquad (29)
\end{aligned}
$$

On the other hand we have

$$
\begin{aligned}
s^2 &= \frac{1}{T-k}\sum_{t=1}^{T}(y_t - x_t'\hat{\beta})^2 \\
&= \frac{1}{T-k}\sum_{t=1}^{T} u_t^2 - \frac{1}{T-k}\left(\sum_{t=1}^{T} x_t u_t\right)'\left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \\
&\quad \times \left(\sum_{t=1}^{T} x_t u_t\right)\left(\frac{1}{\sqrt{T}} g(T)^{-1}\sum_{t=1}^{T} x_t u_t\right) \\
&= \frac{1}{T-k}\sum_{t=1}^{T} u_t^2 - \frac{1}{T-k}\left(\frac{1}{\sqrt{T}} g(T)^{-1}\sum_{t=1}^{T} x_t u_t\right)' \\
&\quad \times \left(\frac{1}{T} g(T)^{-1}\sum_{t=1}^{T} x_t x_t' g(T)^{-1}\right)^{-1}\left(\frac{1}{\sqrt{T}} g(T)^{-1}\sum_{t=1}^{T} x_t u_t\right) \\
&= \frac{1}{T-k} u_t^2 + o_p(T) \xrightarrow{p} E(u_t^2) = \sigma_u^2. \qquad (30)
\end{aligned}
$$

The theorem then follows from (25), (29), (30) and standard results. □

Theorem 3 immediately yields an operational test as follows: Let

$$
\hat{\sigma}^2 = \frac{1}{T-k}\sum_{t=1}^{T}(y_t - x_t'\tilde{\beta})^2 \qquad (31)
$$

be an estimator for σ^2 based on GLS–residuals; and let

$$
\tilde{s}^2 = \hat{\sigma}^2/(1 + \hat{\rho}_1 + \ldots \hat{\rho}_p^2)^2, \qquad (32)
$$

where $\hat{\rho}_i$, $i = \ldots, p$ denote the OLS–based estimates of ρ_i in (1) - (2).

Then, it is easy to show that

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2, \qquad (33)$$

$$\tilde{s}^2 \xrightarrow{p} \tilde{\sigma}^2. \qquad (34)$$

Together (27) and (33) - (34) imply that, under H_0

$$\frac{s^2}{\tilde{s}^2} F \xrightarrow{d} \chi_q^2, \qquad (35)$$

which gives an operational and asymptotically valid test.

Likewise, it is easy to show that the Wald statistics

$$F_1 = (R\tilde{\beta} - r)' \left[R \left(\sum_{t=1}^{T} \tilde{x}_t \tilde{x}_t' \right)^{-1} R' \right]^{-1} (R\tilde{\beta} - r)/\hat{\sigma}^2 \qquad (36)$$

and

$$F_2 = (R\hat{\beta} - r)' \left[R \left(\sum_{t=1}^{T} x_t x_t \right)^{-1} R' \right]^{-1} (R\hat{\beta} - r)/\hat{s}^2 \qquad (37)$$

are both asymptotically χ_q^2 under H_0.

References

Billingsley, P. (1968): *Convergence of Probability Measures.* Chicago.

Chung, Ch.-F. (1995): "Sample Variance, Sample Covariances, and Linear Regression of Stationary Multivariate Long Memory Processes." Mimeo, Michigan State University.

Dolado, J. and Marmol, F. (1998): "Asymptotic Inference for Nonstationary Fractionally Integrated Processes." Mimeo, Universidad Carlos III de Madrid.

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. (1987): "The F-Test with Polynomial Regressors and autocorrelation" (in German), *Allgemeines Statistisches Archiv* 71, 319 – 324.

Krämer, W. (1998): "Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares with Trending Regressors and Stationary Autoregressive Disturbances." In Galata/Küchenhoff (eds.): *Econometrics in Theory and Practice (Festschrift for Hans Schneeweiß)*, 137 – 142.

Krämer, W. and Hassler, U. (1998): "Limiting Efficiency of OLS vs. GLS when Regressors are fractionally integrated." To appear in *Economics Letters*

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.