AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

stat.

GIANNINI FOUNDATION OF
AGRICULTURAL ECONOMICS
LIBRARY

~~WITHDRAWN~~
JAN 8 1976

ERASMUS UNIVERSITY ROTTERDAM

Netherlands school of economics
Econometric Institute

Report 7516/S

ASYMPTOTIC PROPERTIES OF A CORRELATION COEFFICIENT TYPE STATISTIC
CONNECTED WITH THE GENERAL LINEAR MODEL.

Laurens de Haan and Elselien Taconis - Haantjes

August, 1975

ASYMPTOTIC PROPERTIES OF A CORRELATION COEFFICIENT TYPE STATISTIC
CONNECTED WITH THE GENERAL LINEAR MODEL.

Laurens de Haan and Elselien Taconis - Haantjes

I. Introduction

The general linear model can be described as follows. The n -dimensional stochastic vector $\underline{y} = (y_i)_{i=1}^n$ has a normal distribution with $E\underline{y} = X\underline{\beta}$ and $\text{Cov } \underline{y} = \sigma^2 I$ where I is the $n \times n$ identity matrix and X a known $n \times k$ matrix of full rank ($k \leq n$); σ^2 and the components of $\underline{\beta}$ are unknown parameters.

We will be concerned with the situation where the first column of X is $(1, 1, \dots, 1)^T$. The maximum likelihood estimator of $\underline{\beta}$ is $\hat{\underline{\beta}} = (X^T X)^{-1} X^T \underline{y}$.

We follow Theil ([5] p. 163-179, see also [3]) and wish to compare the length of the vectors $N\underline{y}$ and $NX\hat{\underline{\beta}}$ where $N = I - \frac{1}{n} \mathbf{1}\mathbf{1}^T$ with $\mathbf{1}^T = (1, 1, \dots, 1)$. Thus e.g. $\underline{y}^T N \underline{y} = \sum y_i^2 - (\sum y_i)^2/n$. In an econometrical context it is usual to consider the quotient:

$$R_n^2 = \frac{\hat{\underline{\beta}}^T X^T N X \hat{\underline{\beta}}}{\underline{y}^T N \underline{y}}$$

It is wellknown [3] that if for some $\theta > 0$

$$(1) \quad \lim_{n \rightarrow \infty} \frac{\underline{\beta}^T X^T N X \underline{\beta}}{n\sigma^2} = \theta$$

then $\lim_{n \rightarrow \infty} \frac{R_n^2}{\theta + 1} = \frac{\theta}{\theta + 1}$ in probability. Observe that (1) holds if $\lim_{n \rightarrow \infty} \frac{X^T N X}{\sigma^2} = A$ a positive definite matrix. Write $\rho^2 = \frac{\theta}{\theta + 1}$.

Using the fact that R_n^2 has a non-central β -distribution, we prove (section 2) that R_n^2 is asymptotically normal provided

$$(2) \quad \lim_{n \rightarrow \infty} \sqrt{n} \left(\frac{\underline{\beta}^T X^T N X \underline{\beta}}{n\sigma^2} - \theta \right) = 0.$$

*] Underlined letters represent random variables (or random vectors).

In view of an adjustment of $\underline{R}_n^2$ for practical use ((5) p. 178-179) we derive an asymptotic expression for $\underline{ER}_n^2$ (section 3) under the stronger assumption

$$(3) \quad \lim_{n \rightarrow \infty} n \left(\frac{\beta^T X^T N X \beta}{n \sigma^2} - \theta \right) = c$$

for some real c . We were lead to this investigation by our failure to understand Barten's argument [1] leading to a different asymptotic expression for $\underline{ER}_n^2$.

II Asymptotic normality

We can write $\underline{R}_n^2$ as:

$$\underline{R}_n^2 = \frac{|\underline{NX}\hat{\underline{\beta}}|^2}{|\underline{NX}\hat{\underline{\beta}}|^2 + |\underline{y} - \underline{X}\hat{\underline{\beta}}|^2}.$$

Here $|\underline{y} - \underline{X}\hat{\underline{\beta}}|^2/\sigma^2$ has a chi-squared distribution with $n-k$ degrees of freedom. The distribution of $\underline{NX}\hat{\underline{\beta}} = \underline{NX}(\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{y}$ is normal with mean vector $\underline{NX}\beta$ and covariance matrix $\sigma^2 \underline{I}$. The matrix $\underline{NX}(\underline{X}^T \underline{X})^{-1} \underline{X}^T$ is symmetric and idempotent with rank $k-1$ so $|\underline{NX}\hat{\underline{\beta}}|^2/\sigma^2$ has a non-central chi-squared distribution with $k-1$ degrees of freedom and non-centrality parameter $\beta^T \underline{X}^T \underline{NX}\beta/\sigma^2 = \theta_n$. Furthermore $\underline{X}\hat{\underline{\beta}}$ and $|\underline{y} - \underline{X}\hat{\underline{\beta}}|^2$ are independent

so $\underline{R}_n^2$ has a non-central β -distribution with $k-1$ and $n-k$ degrees of freedom and non-centrality parameter θ_n ([2] p. 213).

We will now prove that $\sqrt{n} (\underline{R}_n^2 - \rho^2)$ has a limiting normal distribution. Write

$$\underline{R}_n^2 = \frac{\underline{B}^2}{\underline{B}^2 + \underline{C}^2}$$

where $\underline{B}^2 \sim \chi^2(k-1, \theta_n)$, $\underline{C}^2 \sim \chi^2(n-k)$ and $\lim_{n \rightarrow \infty} n^{-1} \theta_n = \theta = \rho^2/(1-\rho^2)$.

Now

$$(4) \quad \sqrt{n} (\underline{R}_n^2 - \rho^2) = \sqrt{n} \frac{\underline{B}^2 - \underline{C}^2 \theta}{(\underline{B}^2 + \underline{C}^2) (1+\theta)} = \frac{\frac{1}{\sqrt{n}} \underline{B}^2 - \frac{1}{\sqrt{n}} \underline{C}^2 \theta}{\frac{1}{n} (\underline{B}^2 + \underline{C}^2) (1+\theta)}.$$

We first consider the denominator. According to the weak law of large numbers $\frac{1}{n} \underline{C}^2$ converges to 1 in probability. Write $\frac{1}{n} \underline{B}^2 = \frac{1}{n} |\underline{w} - \sqrt{\theta}|^2$ with

$$|\underline{w}|^2 = \sum_{i=1}^{k-1} \underline{w}_i^2 \text{ and } \underline{w}_i \text{ are independent and standard normal. It follows}$$

$\lim_{n \rightarrow \infty} \frac{1}{n} \underline{B}^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \theta_n = \theta$. Hence $\lim_{n \rightarrow \infty} \frac{1}{n} (\underline{B}^2 + \underline{C}^2) (\theta+1) = (\theta+1)^2$ in probability.

The numerator of (4) can be written as

$$(5) \quad \frac{1}{\sqrt{n}} (\underline{B}^2 - \theta_n) - \frac{\theta}{\sqrt{n}} (\underline{C}^2 - n) + \sqrt{n} \left(\frac{\theta_n}{n} - \theta \right).$$

Since $\frac{1}{\sqrt{n}} (\underline{B}^2 - \theta_n)$ and $\frac{\theta}{\sqrt{n}} (\underline{C}^2 - n)$ are independent, we have to prove that each part of (5) has a limiting distribution. Now

$$\frac{1}{\sqrt{n}} (\underline{B}^2 - \theta_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^k \underline{w}_i^2 - \frac{2}{\sqrt{n}} \sum_{i=1}^{k-1} \underline{w}_i \sqrt{\theta_n}.$$

The first part tends to zero and the second part is $N(0, \frac{4\theta}{n})$. Thus the distribution of $\frac{1}{\sqrt{n}} (\underline{B}^2 - \theta_n)$ tends to a normal distribution with mean zero and variance 4θ .

We apply the central limit theorem to find, that the second part $\frac{\theta}{\sqrt{n}} (\underline{C}^2 - n)$ of (5) is asymptotically $N(0, 2\theta^2)$.

Combining our results and applying (2) we have that $\sqrt{n}(\underline{R}_n^2 - \rho^2)$ converges in law to

$$N \left(0, \frac{2\theta^2 + 4\theta}{(\theta+1)^4} \right) = N(0, 2\rho^2(1-\rho^2)^2(2-\rho^2)).$$

III. The limiting behaviour of $\underline{ER}_n^2$.

We know that $\underline{R}_n^2$ has a non-central β -distribution with non-centrality parameter θ_n . So its density $g(r^2)$ is [4]

$$g(r^2) = \sum_{\beta=0}^{\infty} e^{-\frac{1}{2}\theta_n} \frac{(\frac{1}{2}\theta_n)^\beta}{\beta!} \frac{\Gamma(\frac{n-1}{2} + \beta)}{\Gamma(\frac{k-1}{2} + \beta)\Gamma(\frac{n-k}{2})} \cdot (r^2)^{\frac{k-1}{2} + \beta - 1} (1-r^2)^{\frac{n-k}{2} - 1}.$$

The expectation of R_n^2 can be written as

$$\begin{aligned} ER_n^2 &= \sum_{\beta=0}^{\infty} e^{-\frac{1}{2}\theta_n} \frac{(\frac{1}{2}\theta_n)^\beta}{\beta!} \frac{\Gamma(\frac{n-1}{2} + \beta)}{\Gamma(\frac{k-1}{2} + \beta)\Gamma(\frac{n-k}{2})} \int_0^1 (r^2)^{\frac{k-1}{2} + \beta - 1} (1-r^2)^{\frac{n-k}{2} - 1} dr^2 \\ &= \sum_{\beta=0}^{\infty} e^{-\frac{1}{2}\theta_n} \frac{(\frac{1}{2}\theta_n)^\beta}{\beta!} \cdot \frac{\Gamma(\frac{n-1}{2} + \beta)}{\Gamma(\frac{k-1}{2} + \beta)\Gamma(\frac{n-k}{2})} \frac{\Gamma(\frac{k-1}{2} + \beta + 1)\Gamma(\frac{n-k}{2})}{\Gamma(\frac{n-1}{2} + \beta + 1)} \\ &= \sum_{\beta=0}^{\infty} e^{-\frac{1}{2}\theta_n} \frac{(\frac{1}{2}\theta_n)^\beta}{\beta!} \frac{\frac{k-1}{2} + \beta}{\frac{n-1}{2} + \beta}. \end{aligned}$$

So $1 - ER_n^2 = \frac{n-k}{n-1} \sum_{\beta=0}^{\infty} e^{-\frac{1}{2}\theta_n} \frac{(\frac{1}{2}\theta_n)^\beta}{\beta!} \frac{1}{1 + \frac{2\beta}{n-1}}$

We write $\int_0^{\infty} e^{\frac{-2\beta}{n-1}x} dx$ for $(1 + \frac{2\beta}{n-1})^{-1}$ and obtain

$$\begin{aligned} 1 - ER_n^2 &= \frac{n-k}{n-1} \int_0^{\infty} e^{-x} \sum_{\beta=0}^{\infty} e^{-\frac{1}{2}\theta_n} \frac{(\frac{1}{2}\theta_n)^\beta}{\beta!} e^{\frac{-2\beta x}{n-1}} dx \\ &= \frac{n-k}{n-1} \int_0^{\infty} e^{-x} e^{-\frac{1}{2}\theta_n} \exp \left[\frac{1}{2}\theta_n e^{\frac{-2x}{n-1}} \right] dx \end{aligned}$$

First we prove that $\lim_{n \rightarrow \infty} n \{ (1 - \rho_n^2) - (1 - \overline{ER}_n^2) \} = \text{constant}$

where $1 - \rho_n^2 = \frac{n-k}{n-1} \left(1 + \frac{\theta_n}{n-1} \right)^{-1} = \frac{n-k}{n-1} \int_0^\infty e^{-x(1 + \frac{\theta_n}{n-1})} dx.$

Now
$$\begin{aligned} n \{ (1 - \rho_n^2) - (1 - \overline{ER}_n^2) \} &= n \left\{ \frac{n-k}{n-1} \int_0^\infty e^{-x(1 + \frac{\theta_n}{n-1})} dx + \right. \\ &\quad \left. - \frac{n-k}{n-1} \int_0^\infty e^{-x} e^{-\frac{1}{2}\theta_n} \exp \left[\frac{1}{2}\theta_n e^{\frac{-2x}{n-1}} \right] dx \right\} \\ &= n \frac{n-k}{n-1} \int_0^\infty e^{-x} \left[e^{\frac{-x\theta_n}{n-1}} e^{-\frac{1}{2}\theta_n} \cdot \exp \left[\frac{1}{2}\theta_n e^{\frac{-2x}{n-1}} - 1 + \frac{2x}{n-1} \right] \right] dx \\ (6) \quad &= n \frac{n-k}{n-1} \int_0^\infty e^{-x(1 + \frac{\theta_n}{n-1})} \left[1 - \exp \frac{1}{2}\theta_n \left(e^{\frac{-2x}{n-1}} - 1 + \frac{2x}{n-1} \right) \right] dx. \end{aligned}$$

The next step is to compare this integral with the following expression

(7)
$$-n \frac{n-k}{n-1} \int_0^\infty e^{-x(1 + \frac{\theta_n}{n-1})} \frac{1}{2}\theta_n \left[e^{\frac{-2x}{n-1}} - 1 + \frac{2x}{n-1} \right] dx$$

which can be evaluated explicitly and converges to $\frac{-2\theta}{(1+\theta)^3}$ as $n \rightarrow \infty$.

We will show that the difference between (7) and (6) tends to zero.

This difference can be written as

(8)
$$\frac{n-k}{n-1} \int_0^\infty e^{-x(1 + \frac{\theta_n}{n-1})} n \left[\frac{g_n(x)}{e} - 1 - g_n(x) \right] dx$$

with
$$g_n(x) = \frac{1}{2}\theta_n \left(e^{\frac{-2x}{n-1}} - 1 + \frac{2x}{n-1} \right).$$

We are going to use the following simple properties

$$(9) \quad \frac{e^x - 1 - x}{x^2} \text{ is an increasing function of } x \text{ when } x > 0,$$

$$(10) \quad 0 < e^{-y} - 1 + y \leq y \quad \text{if } y > 0 \text{ and}$$

$$(11) \quad 0 < e^{-y} - 1 + y \leq \frac{1}{2}y^2 \quad \text{if } y > 0.$$

We split the integral (8) into two parts $\int_0^\infty = \int_0^{f_n} + \int_{f_n}^\infty$

and will prove that both integrals converge to zero. First we consider the second part, then (use (10))

$$\begin{aligned} 0 < \int_{f_n}^\infty e^{-x(1 + \frac{\theta_n}{n-1})} \frac{1}{n} \left[e^{g_n(x)} - 1 - g_n(x) \right] dx \\ &\leq \int_{f_n}^\infty e^{-x(1 + \frac{\theta_n}{n-1})} \frac{1}{n} \left[e^{\frac{\theta_n x}{n-1}} - 1 - \frac{\theta_n x}{n-1} \right] dx \\ &\leq \int_{f_n}^\infty e^{-x(1 + \frac{\theta_n}{n-1})} \frac{\theta_n x}{n-1} dx = \left[-ne^{-x} \right]_{f_n}^\infty = ne^{-f_n}. \end{aligned}$$

Let for instance $f_n = 2 \log n$ then $\lim_{n \rightarrow \infty} ne^{-f_n} = \lim_{n \rightarrow \infty} e^{-(f_n - \log n)} = 0$

Next consider the first part, then (use (11) and (9) successively)

$$\begin{aligned} 0 < \int_0^{f_n} e^{-x(1 + \frac{\theta_n}{n-1})} \frac{1}{n} \left[e^{g_n(x)} - 1 - g_n(x) \right] dx \\ &\leq \int_0^{f_n} e^{-x(1 + \frac{\theta_n}{n-1})} \frac{1}{n} \left[\frac{2\theta_n x^2}{(n-1)^2} - 1 - \frac{2\theta_n x^2}{(n-1)^2} \right] dx \end{aligned}$$

$$\begin{aligned}
 & \leq \int_0^{f_n} n \left[e^{\frac{2\theta_n x^2}{(n-1)^2}} - 1 - \frac{2\theta_n x^2}{(n-1)^2} \right] dx \\
 & \leq \frac{\frac{2\theta_n f_n^2}{(n-1)^2} - 1 - \frac{2\theta_n f_n^2}{(n-1)^2}}{\frac{4\theta_n^2 f_n^4}{(n-1)^4}} \int_0^{f_n} n \cdot \frac{4\theta_n^2 x^4}{(n-1)^4} dx \\
 & = \frac{\frac{2\theta_n f_n^2}{(n-1)^2} - 1 - \frac{2\theta_n f_n^2}{(n-1)^2}}{\frac{4\theta_n^2 f_n^4}{(n-1)^4}} \cdot \frac{4n\theta_n^2 f_n^5}{5(n-1)^4} .
 \end{aligned}$$

Hence the first part also converges to zero as $n \rightarrow \infty$ and thus

$$\lim_{n \rightarrow \infty} n \{ (1 - \rho_n^2) - (1 - \frac{ER_n^2}{n}) \} = \frac{-2\theta}{(1+\theta)^3} = -2\rho^2(1-\rho^2)^2.$$

Up to now we only used assumption (1). We get

$$\lim_{n \rightarrow \infty} n \{ (1 - \rho^2) - (1 - \frac{ER_n^2}{n}) \} = (1-\rho^2) \{ k-1 + (c-2\rho^2)(1-\rho^2) \}$$

provided assumption (3) holds i.e.

$$\lim_{n \rightarrow \infty} n \left(\frac{\beta^T X^T N X \beta}{n\sigma^2} - \theta \right) = c.$$

References

- [1] Barten A.P. (1962). Note on Unbiased Estimation of the Squared Multiple Correlation Coefficient. *Statistica Neerlandica*, 16, 151-163.
- [2] Johnson N.L. and S. Kotz (1970). Continuous Univariate Distributions-2. Houghton Mifflin, Boston.
- [3] Koerts J. and A.P.J. Abrahamse (1969). On the Theory and Application of the General Linear Model. Rotterdam University Press.
- [4] Seber G.A.F. (1963). The Noncentral Chisquared and Beta Distribution. *Biometrika*, 50, 542-544.
- [5] Theil H. (1971). Principles of Econometrics. North Holland Publishing Company, Amsterdam.