Stat.

ERASMUS UNIVERSITY ROTTERDAM

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 *] $\underline{y} = (\underline{y}_i)_{i=1}^n$ has a normal distribution with $E\underline{y} = X\beta$ and Cov $\underline{y} = \sigma^2 I$ where I is the nxn identity matrix and X a known nxk matrix of full rank $(k \leq n)$; σ^2 and the components of β are unknown parameters. We will be concerned with the situation where the first column of X is $(1,1,\ldots,1)^T$. The maximum likelihood estimator of β is $\hat{\underline{\beta}} = (X^TX)^{-1}X^T\underline{y}$.

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 $N = I - \frac{1}{n}\iota\iota^T$ with $\iota^T = (1,1,\ldots,1)$. Thus e.g. $\underline{y}^TN\underline{y} = \Sigma\underline{y}_i^2 - (\Sigma\underline{y}_i)^2$. In an econometrical context it is usual to consider the quotient:

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

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

$$(1) \qquad \lim_{n\to\infty} \frac{\beta^T X^T N X \beta}{n\sigma^2} = \Theta$$

then $\lim_{n\to\infty} \underline{R}_n^2 = \frac{\Theta}{\Theta + 1}$ in probability. Observe that (1) holds if $\lim_{n\to\infty} \frac{X^TNX}{\sigma^2} = A$ a positive definite matrix. Write $\rho^2 = \frac{\Theta}{\Theta + 1}$.

Using the fact that $\underline{R}_n^2$ has a non-central β-distribution, we prove (section 2) that $\underline{R}_n^2$ is asymptotically normal provided

$$(2) \qquad \lim_{n\to\infty} \sqrt{n}\left(\frac{\beta^T X^T N X \beta}{n\sigma^2} - \Theta\right) = 0.$$

*] 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 $E\underline{R}_n^2$ (section 3) under the stronger assumption

$$(3) \qquad \lim_{n\to\infty} n \left(\frac{\beta^T X^T N X \beta}{n\sigma^2} - \Theta \right) = c$$

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 $E\underline{R}_n^2$.

II Asymptotic normality

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

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

Here $|\underline{y} - X\hat{\underline{\beta}}|^2/\sigma^2$ has a chi-squared distribution with n-k degrees of freedom. The distribution of $NX\hat{\underline{\beta}} = NX(X^TX)^{-1}X^T\underline{y}$ is normal with mean vector $NX\beta$ and covariance matrix $\sigma^2 I$. The matrix $NX(X^TX)^{-1}X^T$ is symmetric and idempotent with rank k-1 so $|NX\hat{\underline{\beta}}|^2/\sigma^2$ has a non-central chi-squared distribution with k-1 degrees of freedom and non-centrality parameter $\beta^T X^T N X \beta/\sigma^2 = \Theta_n$. Furthermore $X\hat{\underline{\beta}}$ and $|\underline{y}-X\hat{\underline{\beta}}|^2$ 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 $\underline{B}^2 \backsim \chi^2(k-1,\Theta_n)$, $\underline{C}^2 \backsim \chi^2(n-k)$ and $\lim_{n\to\infty} n^{-1}\,\Theta_n = \Theta = \rho^2/(1-\rho^2)$.

Now

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

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 $\frac{1}{n}\underline{B}^2 = \frac{1}{n}|\underline{w}-\sqrt{\Theta_n}|^2$ with $|\underline{w}|^2 = \sum_{i=1}^{k-1} \underline{w}_i^2$ and $\underline{w}_i$ are independent and standard normal. It follows $\lim_{n\to\infty} \frac{1}{n}\underline{B}^2 = \lim_{n\to\infty} \frac{1}{n}\Theta_n = \Theta$. Hence $\lim_{n\to\infty} \frac{1}{n}(\underline{B}^2+\underline{C}^2)(\Theta+1) = (\Theta+1)^2$ in probability.

The numerator of (4) can be written as

$$(5) \qquad \frac{1}{\sqrt{n}}(\underline{B}^2-\Theta_n) - \frac{\Theta}{\sqrt{n}}(\underline{C}^2-n) + \sqrt{n}\left(\frac{\Theta_n}{n}-\Theta\right).$$

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

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

The first part tends to zero and the second part is $N(0,\frac{4\Theta_n}{n})$. 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 $E\underline{R}_n^2$.

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]

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

$$\bullet\ (r^2)^{\frac{k-1}{2}+\beta-1} (1-r^2)^{\frac{n-k}{2}-1} .$$

The expectation of $\underline{R}_n^2$ can be written as

$$E\underline{R}_n^2 = \sum_{\beta=0}^{\infty} e^{-\frac{1}{2}\Theta_n} \frac{(\frac{1}{2}\Theta_n)^\beta}{\beta!} \frac{\Gamma(\frac{n-1}{2}+\beta)}{\Gamma(\frac{k-1}{2}+\beta)\Gamma(\frac{n-k}{2})} \bullet$$

$$\int_0^1 (r^2)^{\frac{k-1}{2}+\beta} (1-r^2)^{\frac{n-k}{2}-1} dr^2 \bullet$$

$$= \sum_{\beta=0}^{\infty} e^{-\frac{1}{2}\Theta_n} \frac{(\frac{1}{2}\Theta_n)^\beta}{\beta!} \bullet \frac{\Gamma(\frac{n-1}{2}+\beta)}{\Gamma(\frac{k-1}{2}+\beta)\Gamma(\frac{n-k}{2})} \frac{\Gamma(\frac{k-1}{2}+\beta+1)\Gamma(\frac{n-k}{2})}{\Gamma(\frac{n-1}{2}+\beta+1)}$$

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

So

$$1 - E\underline{R}_n^2 = \frac{n-k}{n-1} \sum_{\beta=0}^{\infty} e^{-\frac{1}{2}\Theta_n} \frac{(\frac{1}{2}\Theta_n)^\beta}{\beta!} \frac{1}{1+\frac{2\beta}{n-1}}$$

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

$$1 - E\underline{R}_n^2 = \frac{n-k}{n-1} \int_0^\infty e^{-x} \sum_{\beta=0}^{\infty} e^{-\frac{1}{2}\Theta_n} \frac{(\frac{1}{2}\Theta_n)^\beta}{\beta!} e^{\frac{-2\beta x}{n-1}} dx$$

$$= \frac{n-k}{n-1} \int_0^\infty e^{-x} e^{-\frac{1}{2}\Theta_n} \exp\left[\tfrac{1}{2}\Theta_n e^{-\frac{2x}{n-1}}\right] dx$$

First we prove that $\lim_{n\to\infty} n\,\{(1-\rho_n^2) - (1-E\underline{R}_n^2)\} = \text{constant}$

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

Now $$n\,\{(1-\rho_n^2) - (1-E\underline{R}_n^2)\} = n\left\{\frac{n-k}{n-1}\int_0^\infty e^{-x\left(1+\frac{\Theta_n}{n-1}\right)}\,dx + {}\right.$$

$$\left. - \frac{n-k}{n-1}\int_0^\infty e^{-x}\, e^{-\frac{1}{2}\Theta_n} \exp\left[\tfrac{1}{2}\Theta_n e^{\frac{-2x}{n-1}}\right] dx\right\}$$

$$= n\,\frac{n-k}{n-1}\int_0^\infty e^{-x}\left[e^{\frac{-x\Theta_n}{n-1}} - e^{-\frac{1}{2}\Theta_n}\,.\,\exp\ \tfrac{1}{2}\Theta_n e^{\frac{-2x}{n-1}} - 1 + \frac{2x}{n-1}\right] dx$$

(6) $$= n\,\frac{n-k}{n-1}\int_0^\infty e^{-x\left(1+\frac{\Theta_n}{n-1}\right)}\left[1 - \exp\ \tfrac{1}{2}\Theta_n\left(e^{\frac{-2x}{n-1}} - 1 + \frac{2x}{n-1}\right)\right] dx.$$

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

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

which can be evaluated explicitly and converges to $\frac{-2\Theta}{(1+\Theta)^3}$ as $n\to\infty$.
We will show that the difference between (7) and (6) tends to zero.
This difference can be written as

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

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

We are going to use the following simple properties

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

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

(11) $0 < e^{-y} - 1 + y \leq \frac{1}{2}y^2$ 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))

$$0 < \int_{f_n}^\infty e^{-x(1+\frac{\Theta_n}{n-1})} n \left[e^{g_n(x)} - 1 - g_n(x) \right] dx$$

$$\leq \int_{f_n}^\infty e^{-x(1+\frac{\Theta_n}{n-1})} n \left[e^{\frac{\Theta_n x}{n-1}} - 1 - \frac{\Theta_n x}{n-1} \right] dx$$

$$\leq \int_{f_n}^\infty e^{-x(1+\frac{\Theta_n}{n-1})} n\, e^{\frac{\Theta_n x}{n-1}} dx = \left[-ne^{-x} \right]_{f_n}^\infty = n\, e^{-f_n}.$$

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

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

$$0 < \int_0^{f_n} e^{-x(1+\frac{\Theta_n}{n-1})} n \left[e^{g_n(x)} - 1 - g_n(x) \right] dx$$

$$\leq \int_0^{f_n} e^{-x(1+\frac{\Theta_n}{n-1})} n \left[e^{\frac{2\Theta_n x^2}{(n-1)^2}} - 1 - \frac{2\Theta_n x^2}{(n-1)^2} \right] dx$$

$$\leq \int_0^{f_n} n \left[e^{\frac{2\Theta_n x^2}{(n-1)^2}} - 1 - \frac{2\Theta_n x^2}{(n-1)^2} \right] dx$$

$$\leq \frac{e^{\frac{2\Theta_n f_n^2}{(n-1)^2}} - 1 - \frac{2\Theta_n f_n^2}{(n-1)^2}}{\frac{4\Theta_n^2 f_n^4}{(n-1)^4}} \int_0^{f_n} n . \frac{4\Theta_n^2 x^4}{(n-1)^4} \, dx$$

$$= \frac{e^{\frac{2\Theta_n f_n^2}{(n-1)^2}} - 1 - \frac{2\Theta_n f_n^2}{(n-1)^2}}{\frac{4\Theta_n^2 f_n^4}{(n-1)^4}} \bullet \frac{4n\Theta_n^2 f_n^5}{5(n-1)^4} .$$

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

$$\lim_{n\to\infty} n \{(1 - \rho_n^2) - (1 - \underline{E}R_n^2)\} = \frac{-2\Theta}{(1+\Theta)^3} = -2\rho^2(1-\rho^2)^2.$$

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

$$\lim_{n\to\infty} n \{(1 - \rho^2) - (1 - \underline{E}R_n^2)\} = (1-\rho^2)\{k-1 + (c-2\rho^2)\ (1-\rho^2)\}$$

provided assumption (3) holds i.e.

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

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.