Efficiency as Correlation

E. H. Oksanen
Department of Economics
McMaster University
Hamilton, Ontario L8S 4M4
Canada

I. Introduction

The purpose of this note is to demonstrate that in the context of the linear regression model (with non-stochastic regressor matrix), the BLU (best linear unbiased) estimator of any coefficient has a squared correlation with any linear unbiased estimator equal to their relative efficiency (defined as the ratio of variances).

We first derive the relation between the variance of the difference in estimators, which appears in the denominator of Hausman's (1978) asymptotic test-statistic. Hausman derives the required result as a lemma in terms of two estimators, each consistent and asymptotically normal and one of them asymptotically reaching the Cramér-Rao bound. In a finite-sample context, and with Hausman's test as the motivation, Greene (1990) leaves as an exercise derivation of the covariance matrix of the difference between two coefficient vectors, one being the generalised least squares vector and the other being the difference between it and the ordinary least squares (OLS) coefficient vector. Obviously he is aware of derivations of the sort given below. Bacon (1988) makes a direct computation of the difference in variances in a simple illustration motivating Hausman's test-statistic. Given the result on the variance of differences, the correlation result follows trivially.

2. The Difference in Variances

Consider the linear regression model specified by

$$(1) \quad y = X\beta + u \quad u \sim (0, \sigma^2 I_N)$$

where X is $N \times k$, non-stochastic, with full column rank. The BLU (best linear unbiased) estimator of β is, according to the Gauss-Markov Theorem (GMT),

$$(2) \quad \hat{\beta} = (X'X)^{-1} X'y,$$

the OLS coefficient vector. The GMT is frequently proved (e.g., Goldberger, 1991) through specifying a linear unbiased estimator of β

$$(3) \quad \beta = Cy$$

where C is $k \times N$, non-stochastic, and defined by

$$(4) \quad C = (X'X)^{-1}X' + D.$$

The unbiasedness requirement leads to the set of linear restrictions

$$(5) \quad DX = 0.$$

Equations (3)-(5) imply

$$(6) \quad \tilde{\beta} = \hat{\beta} + Du,$$

and the GMT yields, where "Cov" denotes a variance-covariance matrix,

$$(7) \quad \text{Cov}(\tilde{\beta}) = \text{Cov}(\hat{\beta}) + \sigma^2 DD',$$

DD' being positive semi-definite.

From (6),

$$(8) \quad \text{Cov}(\tilde{\beta} - \hat{\beta}) = \sigma^2 DD'.$$

Combining (7) and (8) yields

$$(9) \quad \text{Cov}(\tilde{\beta} - \hat{\beta}) = \text{Cov}(\tilde{\beta}) - \text{Cov}(\hat{\beta}).$$

Specialising to the j th coefficient,

$$(10) \quad \text{Var}(\tilde{\beta}_j - \hat{\beta}_j) = \text{Var}(\tilde{\beta}_j) - \text{Var}(\hat{\beta}_j).$$

It follows upon expanding the left-hand side of (10) that

$$(11) \quad \text{Var}(\hat{\beta}_j) = \text{Cov}(\tilde{\beta}_j, \hat{\beta}_j).$$

3. Population Correlation Between Two Estimators

Consider two estimators of a particular coefficient, $\tilde{\beta}_j$ and $\hat{\beta}_j$. It follows using (11) that

$$(12) \quad \rho^2 = \text{Cov}^2(\tilde{\beta}_j, \hat{\beta}_j) / \text{Var}(\tilde{\beta}_j) \text{Var}(\hat{\beta}_j) = \text{Var}(\hat{\beta}_j) / \text{Var}(\tilde{\beta}_j).$$

That is, the relative efficiency of the two estimators equals their squared population correlation.

If the disturbances are assumed jointly normal then $\tilde{\beta}_j$ and $\hat{\beta}_j$ will be bivariate normal, and we can write a linear population regression function with $\hat{\beta}_j$ as regressor:

$$(13) \quad E[\tilde{\beta}_j | \hat{\beta}_j] = \tau + \delta \hat{\beta}_j.$$

The parameters in the population regression function are given by

$$(14) \quad \delta = \text{Cov}(\tilde{\beta}_j, \hat{\beta}_j) / \text{Var}(\hat{\beta}_j) \quad \text{and} \quad \tau = E(\tilde{\beta}_j) - \delta E(\hat{\beta}_j).$$

It follows from our results that

$$(15) \quad \delta = 1 \quad \text{and} \quad \tau = 0.$$

On average, in repeated sampling, an increase of one unit in the BLU estimator yields a unit increase in the value of the alternative estimator. There is of course another linear population regression function (the roles of regressor and regressand being reversed). It is linked to (13) through the product of the slopes being equal to the square of the population correlation coefficient.

As an exercise in Monte Carlo, students could be asked to verify (12). A simple drill would be to apply OLS with all N data points for the BLU estimator and OLS with only $N-m$ ($N-m \geq k$) data points for the alternative estimator.

REFERENCES

- Bacon, Robert (1988) A First Course in Econometric Theory, Oxford:
Oxford.
- Goldberger, Arthur S. (1991) A Course in Econometrics, Harvard:
Cambridge.
- Greene, William H. (1990), Econometric Analysis, Macmillan: New York.
- Hausman, J.A. (1978), "Specification Tests in Econometrics",
Econometrica, 1251-1271.

Recent McMaster University Economics Working Papers

(To obtain copies, write to: Secretary, Working Papers, Department of Economics, McMaster University, Hamilton, Ontario, Canada L8S 4M4. A charge of \$3 per paper will be levied on orders from institutions that do not have an arrangement for the exchange of working papers. Orders from individuals will be met free of charge, supplies permitting.)

- No. 91-04 Stephen R.G. Jones and W. Craig Riddell
The Measurement of Labour Force Dynamics with the Labour
Market Activity Survey: The LMAS Filter
- No. 91-05 Orazio Attanasio and Martin Browning
Consumption Over the Life Cycle and Over the Business Cycle
- No. 91-06 J.B. Burbidge and W.M. Scarth
Tariffs, Debt and Welfare
- No. 91-07 Gordon M. Myers and Yorgos Y. Papageorgiou
Homo Economicus in Perspective
- No. 91-08 Andre de Palma and Yorgos Y. Papageorgiou
Agglomeration as Local Instability of Spatially Uniform Steady-
States: A Note on Degeneracy
- No. 91-09 Andre de Palma and Yorgos Y. Papageorgiou
A Model of Rational Choice Behaviour Under Imperfect Ability
to Choose
- No. 91-10 Yorgos Y. Papageorgiou
Residential Choice When Place Utility is Not Precisely Known
in Advance
- No. 91-11 Gordon M. Myers and Yorgos Y. Papageorgiou
Fiscal Inequivalence, Incentive Equivalence and Pareto
Efficiency in a Decentralised Urban Context
- No. 91-12 E.H. Oksanen
Efficiency as Correlation
- No. 91-13 E.H. Oksanen
Simple Sampling Schemes for Calculating Joint Distributions
of Least Squares Coefficients and Least Squares Residuals
- No. 91-14 E.H. Oksanen and J.R. Williams
An Alternative Factor-Analytic Approach to Aggregation of
Input-Output Tables

OSEP Research Reports

<u>Number</u>	<u>Title</u>	<u>Author(s)</u>
273	Interprefectural Migration Patterns of Young Adults in Japan: An Explanation Using a Nested Logit Model	K.L. Liaw A. Otomo
274	Public Policy Analysis and the Impact of Interest Groups: The Contributions of Canadian Political Scientists	H.J. Jacek
275	Rent Control and Vacancy Rates in Canada	R.A. Muller
276	The Behavior of Oil Futures Prices Around OPEC Conferences	R. Deaves I. Krinsky
277	Currency Substitution and Exchange Rate Instability: Which Comes First	M.A. Elkhafif A.A. Kubursi
278	Agency Costs, Open-Ending and Closed-End Fund Discounts	R. Deaves I. Krinsky
279	Consumption Good and Risky Asset: The Dual Role of Housing in the Household Expenditure System	F.T. Denton A.L. Robb B.G. Spencer
280	Canadian Wealth Inequality in the Late Nineteenth Century: A Study of Wentworth County, Ontario, 1872-1902	L. DiMatteo P.J. George
281	The War Economy of Lebanon: The Confessional Links	A.A. Kubursi
282	Changing Patterns of Employment in Canada: Demographic and Social Factors	F.T. Denton P.C. Pineo B.G. Spencer
283	An Algorithm for Robust Regression	L. Magee
284	Contracting Out of Rent Control	T.J. Lewis R.A. Muller
285	Financial Loss Provisions and the Market Value of Residential Rental Property in Ontario	R.A. Muller
286	Interprefectural Migrations at Marriage in Japan: An Explanation by Personal Factors and Ecological Variables	H. Kawabe K-L. Liaw