Efficient Estimators for Regressing Regression Coefficients*

ERIC A. HANUSHEK**

As a natural outgrowth of the proliferation of large-scale empirical analyses, there is an increased interest in unifying the results of related regression analyses. In particular, when similar models have been estimated under a variety of circumstances, it is often desirable to explain any differences in results. However, systematically analyzing such differences is often difficult because new statistical problems are introduced when one relates a series of different estimated relationships. The specific case of efficiently estimating a regression model which has an estimated regression coefficient as the dependent variable is developed here.

Examples of cases in which the results of previous regression analyses become the focal point of a second analysis are easy to find. Wachter (1970) first estimated the relationship between relative wage rates and aggregate unemployment rates for a series of 19 two-digit manufacturing industries and subsequently wished to test a series of hypotheses about why different industries would respond differently to varying unemployment situations. Askari and Cummings (1971) tested a series of models explaining variations in agricultural supply elasticities across countries and crops where the supply elasticities were gathered from the elasticity estimates made in several studies by different individuals. Hanushek has analyzed how differences in the monetary returns to schooling relate to characteristics of different metropolitan labor markets using the separate returns to schooling estimated for 121 metropolitan areas in Hanushek (1973). Each of these cases involves estimating a regression model where the dependent variable is itself a regression coefficient.

However, when interest centers upon variations in behavioral parameters which themselves have been estimated through prior applications of regression analysis, ordinary least squares is no longer the most appropriate estimation technique. The dependent variable—a regression coefficient—is observed with varying sampling errors, and this will lead to heteroscedastic errors in the second regression analysis. By using information about the estimated variance of this sampling error, it is possible to obtain more efficient parameter estimates than those of ordinary least squares in the second analysis.

Consider the case where T different parameters β_i are estimated for T different samples where the model in each sample is

$$\mathbf{Y}_i = \beta_i \mathbf{Z}_i + \mathbf{w}_i \tag{1}$$

($\mathbf{Y}_i$ and $\mathbf{Z}_i$ are vectors of observed variables within Sample i, $\mathbf{w}_i$ is a vector of stochastic terms for Sample i, and β_i is the scalar parameter of interest in Sample i). Further, the real interest centers upon a second stage model of the structure of the β_i's, and we wish to estimate a model such as

$$\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{v}. \tag{2}$$

where $\boldsymbol{\beta}$ is a $(T \times 1)$ vector of structural parameters from Equation 1, $\mathbf{X}$ is a $(T \times K)$ matrix of observations for K exogenous variables, $\boldsymbol{\gamma}$ is a $(K \times 1)$ vector of second stage parameters, and $\mathbf{v}$ is a $(T \times 1)$ vector of stochastic elements. However, $\boldsymbol{\beta}$ is not observed. Instead, we observe $\hat{\boldsymbol{\beta}}$, the set of estimated coefficients from Equation 1, where $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{u}$ and $\mathbf{u}$ is a vector of sampling errors in the regression estimates of Equation 1 over the T samples. Even if $E(\mathbf{vv}') = \sigma^2\mathbf{I}$, the error variance for the second stage regression analysis will not be homoscedastic, and, thus, ordinary least squares estimates will be inefficient.[1]

The problem is one of estimating parameters of the model

$$\hat{\boldsymbol{\beta}} = \mathbf{X}\boldsymbol{\gamma} + (\mathbf{u} + \mathbf{v}) = \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}. \tag{3}$$

Assuming that $\mathbf{u}$ and $\mathbf{v}$ are independent, that $E(\mathbf{vv}') = \sigma^2\mathbf{I}$, and that

$$E(u_i u_j) = \begin{cases} \omega_i^2, & i = j \\ \omega_{ij}, & i \neq j \end{cases}$$

we have

$$E(\boldsymbol{\varepsilon\varepsilon}') = \sigma^2\boldsymbol{\Omega} = \begin{bmatrix} 1 + \frac{\omega_1^2}{\sigma^2} & \frac{\omega_{12}}{\sigma^2} & \cdots & \frac{\omega_{1T}}{\sigma^2} \\ & \ddots & & \vdots \\ & & & 1 + \frac{\omega_T^2}{\sigma^2} \end{bmatrix} \sigma^2. \tag{4}$$

[1] The key element in this problem is interest in a dependent variable which is observed with some sampling error which differs among observations. There are other cases, generally less interesting, where this may also be important. For example, the first stage model (Equation 1), may be a single regression analysis which includes a large set of dummy variables with Equation 2 being a model explaining differences in the parameters of these qualitative variables. An example of this can be found in Hanushek (1972), where estimates of the value added in education of individual teachers are made and an attempt is made to explain why some teachers are better than others. This adds the complication that the u_i will not be independent of each other but can still be handled within the framework below. In general, however, this situation can be handled through direct estimation of a reduced form model for Equation 1 and 2. A second case arises when the dependent variable consists of sampled values as with Census data on individual characteristics such as income. The Census of Population, relying on 5 and 20 percent samples, has sampling variances related to number of observations in a given cell. The differences in the variance in these sampling errors will generally be small, thus making the problem less important.

* This problem was first brought to my attention by Franklin M. Fisher.

** Cost of Living Council, Office of Price Stabilization, Analysis Div., 2000 M St., N.W., Washington, D.C. 20508.

When $E(\boldsymbol{\varepsilon\varepsilon}')$ is known, Aitken, or generalized least squares, estimators can be used to obtain efficient estimators of $\boldsymbol{\gamma}$. In this particular situation, $E(\boldsymbol{\varepsilon\varepsilon}')$ is only partially known, but it is possible to use a two-round estimation procedure which estimates the missing elements of $E(\boldsymbol{\varepsilon\varepsilon}')$ in order to apply generalized least squares.

Since we normally have estimates of $E(\mathbf{uu}')$ from the estimated coefficient variables in the regression analysis of Equation 1, we need only an estimate of σ^2 in order to apply generalized least squares. Consider estimating Equation 2 with ordinary least squares. The expected value of the estimated error variance using the calculated residuals (e_i) is[2]

$$E(s^2) = E\left(\frac{\Sigma e_i^2}{T-K}\right) = \sigma^2 \frac{\mathrm{tr}\boldsymbol{\Omega} - \mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X})}{T-K} \tag{5}$$

If $\boldsymbol{\Omega} = \mathbf{I} - \mathbf{F}$, we find that

$$\mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}) = \mathbf{K} - \mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{FX}, \tag{6}$$

and

$$\mathrm{tr}\boldsymbol{\Omega} = T + \frac{\Sigma\omega_i^2}{\sigma^2}. \tag{7}$$

If we further define $\mathbf{G} = \sigma^2\mathbf{F}$, and substitute (6) and (7) into (5), we have

$$E(s^2) = \sigma^2 \frac{T + \dfrac{\Sigma\omega_i^2}{\sigma^2} - \left(K - \dfrac{1}{\sigma^2}\mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{GX})\right)}{T-K}. \tag{8}$$

[2] The development of Equations 5–7 can be found in Goldberger (1964), pp. 238–239. His normalization of Ω is not used here.

The first stage model(s) provide an estimate of $-\mathbf{G}$ and $\Sigma\omega_i^2$, and the first round OLS estimates provide s^2. Thus, we can estimate σ^2 by

$$\hat{\sigma}^2 = \frac{s^2(T-K) - \Sigma\omega_i^2 - \mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{GX}}{T-K}. \tag{9}$$

By substituting estimated values of σ^2, ω_i^2 and ω_{ij} into (4), a second-round estimate of (2) can be made using Aitken estimators. When the observations of $\boldsymbol{\beta}$ come from parameter estimates in a series of different equations, $\mathbf{G}$ reduces to a diagonal matrix. This two round estimation will then provide asymptotically efficient estimates of the second stage parameters.[3]

REFERENCES

Askari, Hassein and John Cummings, *Comparison of Supply Response in Agriculture Across Crops and Countries*, 1971 (Mimeographed).

Goldberger, Arthur S., *Econometric Theory*, New York: John Wiley and Sons, 1964.

Hanushek, Eric A., "Regional Differences in the Structure of Earnings," *Review of Economics and Statistics*, 55 (May 1973), 204–213.

——, *Education and Race*, Lexington, Mass.: D. C. Heath, 1972.

Maddala, G. S., "Generalized Least Squares with an Estimated Variance Covariance Matrix," *Econometrica*, 39 (January 1971), 23–34.

Wachter, Michael L., "Relative Wage Equations for United States Manufacturing Industries 1947–1967," *Review of Economics and Statistics*, 52 (November 1970), 405–410.

[3] The estimates are asymptotically best except when there are lagged endogenous variables in Equation 2; cf. G. S. Maddala (1971). However, this exception is not very relevant for the type of problem considered here.

A Note on Estimating Covariance Components*

S. R. SEARLE** and T. R. ROUNSAVILLE***

Abstract

A well-known formula for expressing a covariance in terms of variances is shown to hold true for estimating components of covariance.

Covariance components are defined by Rao [1971] as the off-diagonal elements of a variance-covariance matrix that is not diagonal. The more usual way of looking at covariance components, that familiar to geneticists for example, is of having two random variables X_1 and X_2, observed in pairs, with components of covariance being the components of the covariance between X_1 and X_2 defined in the same way as are the components of variance of X_1 and of X_2. A simple illustration is the one-way classification. If x_{1ij} and x_{2ij} are measurements on the jth unit of the ith class, the customary random effects models are

$$x_{1ij} = \mu_1 + \alpha_i + e_{ij} \quad \text{and} \quad x_{2ij} = \mu_2 + \beta_i + \epsilon_{ij} \tag{1}$$

where μ_1 and μ_2 are overall means, α_i and β_i are the random effects due to the ith class and e_{ij} and ϵ_{ij} are the error terms, the variance components being σ_α^2, σ_β^2 and σ_e^2, σ_ϵ^2 respectively. The covariance components are then $\sigma_{\alpha\beta} = \mathrm{cov}(\alpha_i, \beta_i)$ for all i and $\sigma_{e\epsilon} = \mathrm{cov}(e_{ij}, \epsilon_{ij})$ for all i and j, so that $\sigma_{x_{1ij}x_{2ij}} = \sigma_{\alpha\beta} + \sigma_{e\epsilon} \equiv \sigma_{xy}$ for all i and j.

* Paper BU-269 in the Biometrics Unit.

** Biometrics Unit, 337 Warren Hall, Cornell Univ., Ithaca, N.Y. 14850.

*** Animal Science Dept., Cornell Univ., Ithaca, N.Y. 14850.