Journal of Statistical Planning and Inference 7 (1982) 131-137
North-Holland Publishing Company

ON REGRESSING REGRESSION COEFFICIENTS

George J. BORJAS

Department of Economics, University of California, Santa Barbara, CA 93106, USA

Received 24 April 1978; revised manuscript received 27 July 1981
Recommended by G.A.F. Seber and J. Srivastava

Abstract: This paper considers a regression model in which coefficients obtained from a previous regression are themselves the object of analysis. It is shown that the parameters of interest can be obtained in two ways: pooling across observations and subsamples, or a two-stage process of first estimating the coefficients within each subsample, and then using these coefficients as dependent variables in a second stage regression. The relative properties of these estimators are analyzed, and the conditions under which the two estimators are equivalent are derived.

Keywords and phrases: Random coefficient models, Regressing regression coefficients.

AMS Subject Classification: Primary 62J05; Secondary 62F10.

1. Introduction

The recent availability of large scale data sets has led researchers to investigate variations in regression coefficients across observations. These models are, of course, a natural application of the literature on random coefficients [see, for example, Amemiya (1978), Cooley and Prescott (1973), Hildreth and Houck (1968), and Hsiao (1975)]. This paper deals with the case in which regression coefficients vary across subsamples (e.g., households or firms). Enough observations exist for each subsample to allow the researcher to estimate the regression coefficient within each subsample. It is then hypothesized that the variation in the regression coefficients across subsamples can be partly explained by systematic factors, and it is this equation which is the object of the analysis. In other words, the problem is one of estimating a regression model in which the dependent variable is a previously estimated regression coefficient.

There are several empirical examples of this type of model in the economics literature. Wachter (1970) estimated the wage-unemployment relationship within each industry and subsequently investigated why the unemployment effect on wage rates varied across industries. Hanushek (1973) analyzed how the returns to schooling (estimated within each metropolitan area) were affected by metropolitan area characteristics. Finally, Borjas and Mincer (1978) estimated the wage-experience profile for each individual in the Coleman–Rossi retrospective life history sample and then analyzed the determinants of slope differences across individuals.

In each of these studies, the two-stage procedure outlined above was used. First, the regression coefficient was estimated for each subsample. Secondly, these coefficients were regressed on the characteristics of the subsamples. Both Hanushek (1974) and Saxonhouse (1976) have noted that in this kind of procedure, a generalized least squares estimator is the appropriate estimator in the second stage. Saxonhouse also noted that an alternative method of estimation exists. The alternative is to pool the data across subsamples, and this pooled regression would still yield the parameters of interest. He was able to show that in the special case where the variation in the regression coefficients across subsamples was deterministic, the alternative estimators were identical.

This paper considers the more general case in which the variation in the regression coefficients across subsamples is partly stochastic. The objective of the study is to compare the properties of the alternative estimators, the pooled estimator and the two-stage estimator.

2. The pooled estimator and the two-stage estimator

Consider the regression model for the ith subsample ($i=1,\dots,n$):

$$\mathbf{y}_i=\mathbf{z}_i\beta_i+\varepsilon_i \tag{1}$$

where $\mathbf{y}_i$, $\mathbf{z}_i$, and ε_i are vectors of dimension $L\times 1$, β_i is a scalar, and the variables are measured in deviations from the mean. It is hypothesized that *across* subsamples β_i is determined by

$$\beta_i=\mathbf{s}_i\gamma+v_i \tag{2}$$

where $\mathbf{s}_i$ is $1\times k$, and γ is $k\times 1$. The vector of coefficients γ is assumed constant across observations, and it is the estimation of γ which is the object of the analysis. A simple method of estimating γ is to first estimate equation (1) within each subsample, and then use the n estimated coefficients as the dependent variables in (2). An alternative method of estimation can be obtained by substituting equation (2) into (1):

$$\mathbf{y}_i=\mathbf{z}_i\mathbf{s}_i\gamma+\varepsilon_i+\mathbf{z}_i v_i. \tag{3}$$

By pooling the L observations of each subsample across subsamples, an estimator of γ can be obtained. If we pool all subsamples and stack the nL observations equation (3) can be written as

$$\mathbf{y}=ZS\gamma+\varepsilon+Zv \tag{4}$$

where

$$\mathbf{y}=\begin{bmatrix}\mathbf{y}_1\\ \vdots\\ \mathbf{y}_n\end{bmatrix},\quad Z=\begin{bmatrix}\mathbf{z}_1 & & 0\\ & \ddots & \\ 0 & & \mathbf{z}_n\end{bmatrix},\quad S=\begin{bmatrix}\mathbf{s}_1\\ \vdots\\ \mathbf{s}_n\end{bmatrix},\quad \varepsilon=\begin{bmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_n\end{bmatrix},\quad \upsilon=\begin{bmatrix}\upsilon_1\\ \vdots\\ \upsilon_n\end{bmatrix}.$$

Note that the dimensions of $\mathbf{y}$, Z, S, ε, and υ are $nL\times 1$, $nL\times n$, $n\times k$, $nL\times 1$, and $n\times 1$, respectively.

The stochastic properties of the disturbances in the model are given by

$$E(\varepsilon)=0, \tag{5a}$$

$$E(\upsilon)=0, \tag{5b}$$

$$E(\varepsilon_i\varepsilon_j')=\sigma_\varepsilon^2\Psi_{ii},\quad i=j, \tag{5c}$$

$$E(\varepsilon_i\varepsilon_j')=0,\qquad i\neq j \tag{5d}$$

$$\mathrm{Var}(\upsilon)=\sigma_\upsilon^2 I, \tag{5e}$$

$$E(\varepsilon\upsilon')=0. \tag{5f}$$

Equation (5c) allows for the disturbances to be correlated *within* each subsample. This is likely to occur when, for example, a subsample is composed of time-series observations on an individual, firm, or industry.

Given these assumptions, the error term in (4) has mean zero and a variance–covariance matrix given by

$$\Sigma=\sigma_\varepsilon^2\Psi+\sigma_\upsilon^2 ZZ' \tag{6}$$

where

$$\Psi=\begin{bmatrix}\Psi_{11} & 0 & \\ & \vdots & \\ 0 & & \Psi_{nn}\end{bmatrix}.$$

Consider the case where σ_ε^2, σ_υ^2, and Ψ are known so that Σ is a known matrix (the estimation procedure in random coefficients models when Σ is unknown is discussed in Hanushek (1974) and Hsiao (1975)). The pooled estimator of γ is then given by the generalized least squares estimator:

$$\hat{\gamma}=(S'Z'\Sigma^{-1}ZS)^{-1}S'Z'\Sigma^{-1}\mathbf{y}. \tag{7}$$

It can be easily seen by the properties of the error term that $\hat{\gamma}$ is an unbiased estimator of γ and that $\mathrm{Var}(\hat{\gamma})=(S'Z'\Sigma^{-1}ZS)^{-1}$. A more convenient term for the variance of $\hat{\gamma}$ can be derived by noting that due to the uncorrelatedness of ε and υ

$$\Sigma^{-1}=\begin{bmatrix}\sigma_\varepsilon^2\Psi_{11}+\sigma_\upsilon^2\mathbf{z}_1\mathbf{z}_1' & & 0\\ & \ddots & \\ 0 & & \sigma_\varepsilon^2\Psi_{nn}+\sigma_\upsilon^2\mathbf{z}_n\mathbf{z}_n'\end{bmatrix}^{-1}. \tag{8}$$

Using the definition of Σ^{-1} in (8) and substituting the definitions of Z and S into $(S'Z'\Sigma^{-1}ZS)^{-1}$, we obtain

$$\operatorname{Var}(\hat{\gamma})=\left[\sum_{i=1}^{n} s_i' z_i'(\sigma_\varepsilon^2 \Psi_{ii}+\sigma_v^2 z_i z_i')^{-1} z_i s_i\right]^{-1}. \tag{9}$$

To derive the two-stage estimator, we first run a regression on the L observations of subsample i, yielding the estimated $\hat{\beta}_i$:

$$\hat{\beta}_i=(z_i' z_i)^{-1} z_i' y_i. \tag{10}$$

Given the estimated $\hat{\beta}_i$ for all subsamples the second stage of the process consists of estimating equation (2) with the n observations on $\hat{\beta}_i$. However, it is important to realize that by using $\hat{\beta}_i$ and not β_i as the dependent variable, the second stage regression is actually

$$\hat{\beta}=S\gamma+v+(Z'Z)^{-1}Z'\varepsilon \tag{11}$$

where $\hat{\beta}=(\hat{\beta}_1,\dots,\hat{\beta}_n)'$. Note that the error term in (11) has mean zero but that it is heteroscedastic with variance

$$\Omega=\sigma_v^2 \mathrm{I}+\sigma_\varepsilon^2(Z'Z)^{-1}Z'\Psi Z(Z'Z)^{-1}. \tag{12}$$

This fact implies that the minimum variance estimator of γ is the generalized least squares estimator

$$\tilde{\gamma}=(S'\Omega^{-1}S)^{-1}S'\Omega^{-1}\hat{\beta} \tag{13}$$

which is an unbiased estimator of γ with variance $(S'\Omega^{-1}S)^{-1}$. This expression for the variance can be rewritten as

$$\operatorname{Var}(\tilde{\gamma})=\left[\sum_{i=1}^{n} s_i'[\sigma_v^2+\sigma_\varepsilon^2(z_i'z_i)^{-1}z_i'\Psi_{ii}z_i(z_i'z_i)^{-1}]^{-1}s_i\right]^{-1}. \tag{14}$$

Theorem 1. *The pooled estimator, $\hat{\gamma}$, and the two-stage estimator, $\tilde{\gamma}$, both provide unbiased estimates of the coefficient vector γ. Further:*

(a) *if the variance matrix $\Psi_{ii}\neq k_i I$, where k_i is a scalar, then the pooled estimator has a lower variance;*

(b) *if the variance matrix $\Psi_{ii}=k_i I$, where k_i is a scalar, then the two methods of estimation are equivalent.*

Proof. To show unbiasedness of both estimators is trivial. To prove part (a) of the theorem note that the variance-covariance matrix of $\tilde{\gamma}$, $\operatorname{Var}(\tilde{\gamma})$, exceeds $\operatorname{Var}(\hat{\gamma})$ by a positive definite matrix if and only if $[\operatorname{Var}(\hat{\gamma})]^{-1}-[\operatorname{Var}(\tilde{\gamma})]^{-1}$ is positive definite. This expression can be written as

$$\begin{aligned}\Delta=\sum_{i=1}^{n} s_i'\{&z_i'[\sigma_\varepsilon^2\Psi_{ii}+\sigma_v^2 z_i z_i']^{-1}z_i\\ &-[\sigma_v^2+\sigma_\varepsilon^2(z_i'z_i)^{-1}z_i'\Psi_{ii}z_i(z_i'z_i)^{-1}]^{-1}\}s_i.\end{aligned} \tag{15}$$

The theorem is proven if this differential is a positive definite matrix when $\Psi_{ii}\neq k_i I$. To show this consider the jth term of (15):

$$\Delta_j = \mathbf{s}_j' \{ \mathbf{z}_j' [\sigma_\varepsilon^2 \Psi_{jj} + \sigma_v^2 \mathbf{z}_j \mathbf{z}_j']^{-1} \mathbf{z}_j - [\sigma_v^2 + \sigma_\varepsilon^2 (\mathbf{z}_j'\mathbf{z}_j)^{-1} \mathbf{z}_j' \Psi_{jj} \mathbf{z}_j (\mathbf{z}_j'\mathbf{z}_j)^{-1}]^{-1} \} \mathbf{s}_j. \tag{16}$$

The inverse of the first bracketed term in (16) is given by

$$[\sigma_\varepsilon^2 \Psi_{jj} + \sigma_v^2 \mathbf{z}_j \mathbf{z}_j']^{-1} = \frac{1}{\sigma_\varepsilon^2} \Psi_{jj}^{-1} - \frac{\sigma_v^2 \Psi_{jj}^{-1} \mathbf{z}_j \mathbf{z}_j' \Psi_{jj}^{-1}}{\sigma_\varepsilon^2 (\sigma_v^2 \mathbf{z}_j' \Psi_{jj}^{-1} \mathbf{z}_j + \sigma_\varepsilon^2)}. \tag{17}$$

Similarly, the inverse of the second bracketed term in (17) can be shown to be

$$[\sigma_v^2 + \sigma_\varepsilon^2 (\mathbf{z}_j'\mathbf{z}_j)^{-1} \mathbf{z}_j' \Psi_{jj} \mathbf{z}_j (\mathbf{z}_j'\mathbf{z}_j)^{-1}]^{-1} = \frac{(\mathbf{z}_j'\mathbf{z}_j)(\mathbf{z}_j'\mathbf{z}_j)}{\sigma_v^2 (\mathbf{z}_j'\mathbf{z}_j)(\mathbf{z}_j'\mathbf{z}_j) + \sigma_\varepsilon^2 (\mathbf{z}_j' \Psi_{jj} \mathbf{z}_j)}. \tag{18}$$

Substituting equations (17) and (18) into (16) yields

$$\Delta_j = \mathbf{s}_j' \left[\frac{\mathbf{z}_j' \Psi_{jj}^{-1} \mathbf{z}_j}{\sigma_v^2 (\mathbf{z}_j' \Psi_{jj}^{-1} \mathbf{z}_j) + \sigma_\varepsilon^2} - \frac{(\mathbf{z}_j'\mathbf{z}_j)(\mathbf{z}_j'\mathbf{z}_j)}{\sigma_v^2 (\mathbf{z}_j'\mathbf{z}_j)(\mathbf{z}_j'\mathbf{z}_j) + \sigma_\varepsilon^2 (\mathbf{z}_j' \Psi_{jj} \mathbf{z}_j)} \right] \mathbf{s}_j. \tag{19}$$

Clearly, if equation (19) can be shown to be positive definite, then Δ_j is positive definite for all j, and hence Δ in (15) is positive definite. A necessary and sufficient condition for Δ_j to be positive definite is that the scalar inside the brackets be positive. This will be the case if and only if

$$(\mathbf{z}_j' \Psi_{jj} \mathbf{z}_j)(\mathbf{z}_j' \Psi_{jj}^{-1} \mathbf{z}_j) - (\mathbf{z}_j'\mathbf{z}_j)(\mathbf{z}_j'\mathbf{z}_j) > 0. \tag{20}$$

To prove that indeed equation (20) holds, introduce a transformation $\mathbf{z}_j = Q\mathbf{h}$, where the columns of Q are an orthonormal set of eigenvectors for Ψ_{jj}. Hence the following quadratic forms can be obtained.

$$\begin{aligned} \mathbf{z}_j' \Psi_{jj} \mathbf{z}_j &= \mathbf{h}'Q' \Psi_{jj} Q\mathbf{h} = \mathbf{h}'D\mathbf{h} = \sum_{i=1}^{L} \lambda_i h_i^2, \\ \mathbf{z}_j' \Psi_{jj}^{-1} \mathbf{z}_j &= \mathbf{h}'Q' \Psi_{jj}^{-1} Q\mathbf{h} = \mathbf{h}'D^{-1}\mathbf{h} = \sum_{i=1}^{L} (1/\lambda_i) h_i^2 \end{aligned} \tag{21}$$

where $D = Q'\Psi_{jj} Q$, a diagonal matrix whose diagonal elements are the eigenvalues (λ_i) of Ψ_{jj}, and where use has been made of the fact that since Ψ_{jj} is symmetric and positive definite the eigenvalues of Ψ_{jj}^{-1} are the reciprocals of the eigenvalues of Ψ_{jj}.

Similarly, since $Q'Q = I$, it is easy to show that

$$(\mathbf{z}_j'\mathbf{z}_j)(\mathbf{z}_j'\mathbf{z}_j) = (\mathbf{h}'\mathbf{h})(\mathbf{h}'\mathbf{h}) = \left[\sum_{i=1}^{L} h_i^2 \right]^2. \tag{22}$$

Therefore, equation (20) can be rewritten as

$$\left[\sum_{i=1}^{L} \lambda_i h_i^2 \right] \left[\sum_{i=1}^{L} (1/\lambda_i) h_i^2 \right] - \left[\sum_{i=1}^{L} h_i^2 \right]^2 > 0. \tag{23}$$

The strict inequality in (23) follows from a straightforward application of the Schwarz inequality, $(\sum a_i b_i)^2 \leq \sum a_i^2 \sum b_i^2$, where $a_i^2 = \lambda_i h_i^2$ and $b_i^2 = (1/\lambda_i) h_i^2$. Equality will only be achieved if $\lambda_i = k_j$, implying that the variance-covariance matrix $\Psi_{jj} = k_j I$.

To prove part (b) of the theorem, note that the discussion shows that if the within subsample disturbances are homoscedastic, then $\mathrm{Var}(\hat{\gamma})=\mathrm{Var}(\tilde{\gamma})$. Recall that the pooled estimator is given by $\hat{\gamma}=(S'Z'\Sigma^{-1}ZS)^{-1}S'Z'\Sigma^{-1}\mathbf{y}$ and the two stage estimator is given by $\tilde{\gamma}=(S'\Omega^{-1}S)^{-1}S'\Omega^{-1}\hat{\boldsymbol{\beta}}$. The fact that the variances are identical means that $(S'Z'\Sigma^{-1}ZS)^{-1}=(S'\Omega^{-1}S)^{-1}$, thus to prove the equivalence of the two estimators it suffices to show $S'Z'\Sigma^{-1}\mathbf{y}=S'\Omega^{-1}\hat{\boldsymbol{\beta}}$. Note that

$$S'Z'\Sigma^{-1}\mathbf{y}=\sum_{i=1}^{n}\mathbf{s}_i'\mathbf{z}_i'\left[\frac{1}{k_i\sigma_\varepsilon^2}I-\frac{\sigma_v^2\mathbf{z}_i\mathbf{z}_i'}{k_i\sigma_\varepsilon^2(\sigma_v^2\mathbf{z}_i'\mathbf{z}_i+k_i\sigma_\varepsilon^2)}\right]\mathbf{y}_i \tag{24}$$

and

$$S'\Omega^{-1}\hat{\boldsymbol{\beta}}=\sum_{i=1}^{n}\frac{\mathbf{s}_i'\mathbf{z}_i'\mathbf{y}_i}{\sigma_v^2\mathbf{z}_i'\mathbf{z}_i+k_i\sigma_\varepsilon^2}. \tag{25}$$

The two estimators will be equivalent if each term in the sums (24) and (25) is identical. For example, consider the differential for the jth term in the sum

$$\Phi_j=\mathbf{s}_j'\left\{\frac{\mathbf{z}_j'\mathbf{y}_j}{k_j\sigma_\varepsilon^2}-\frac{\sigma_v^2(\mathbf{z}_j'\mathbf{z}_j)\mathbf{z}_j'\mathbf{y}_j}{k_j\sigma_\varepsilon^2(\sigma_v^2\mathbf{z}_j'\mathbf{z}_j+k_j\sigma_\varepsilon^2)}-\frac{\mathbf{z}_j'\mathbf{y}_j}{\sigma_v^2\mathbf{z}_j'\mathbf{z}_j+k_j\sigma_\varepsilon^2}\right\}. \tag{26}$$

By expanding equation (26) it can be easily shown that $\Phi_j=0$; hence $S'Z'\Sigma^{-1}\mathbf{y}=S'\Omega^{-1}\hat{\boldsymbol{\beta}}$ and the two estimators are identical.

Finally, an interesting question concerns the relationship between the GLS estimators presented in this paper and the inefficient ordinary least squares estimators. It follows from the properties of the error term that the OLS estimator of γ using the pooled method, $\hat{\gamma}^0$, is unbiased and that $\mathrm{Var}(\hat{\gamma}^0)=(S'Z'ZS)^{-1}(S'Z'\Sigma ZS)(S'Z'ZS)^{-1}$. Similarly, the OLS estimator of γ using the two-stage method, $\tilde{\gamma}^0$, is also unbiased with variance $\mathrm{Var}(\tilde{\gamma}^0)=(S'S)^{-1}S'\Omega S(S'S)^{-1}$. A simple relationship can be established between the two OLS estimators and the GLS estimators discussed earlier in an empirically relevant special case.

Theorem 2. *If the vector of explanatory variables* $\mathbf{z}_i$ *is the same for all subsamples, and if* $\Psi_{ii}=I$ *for all subsamples, then the use of OLS either in the pooled method or in the second stage of the two-stage method yields identical estimators. Further, these estimators are equivalent to the GLS estimators.*

Proof. To first establish the equivalence of the two estimators under the conditions of the theorem, note that

$$\hat{\gamma}^0=(S'Z'ZS)^{-1}S'Z'\mathbf{y}, \tag{27}$$

$$\tilde{\gamma}^0=(S'S)^{-1}S'(Z'Z)^{-1}Z'\mathbf{y}, \tag{28}$$

$$Z'Z=\sigma_z^2I, \tag{29}$$

where σ_z^2 is the sum of squares $\mathbf{z}_i'\mathbf{z}_i$, which is constant across subsamples. It can then be easily verified that

$$\hat{\gamma}^0=[1/\sigma_z^2](S'S)^{-1}S'Z'\mathbf{y}=\tilde{\gamma}^0 \tag{30}$$

establishing that the two OLS estimators provide the same estimates.

To prove the equivalence of OLS and GLS in this case, part (b) of Theorem 1 has shown that both GLS methods are identical, thus we need only consider one of the two methods of estimation. In particular, consider the two-stage estimator. To prove the equivalence of GLS and OLS we must establish that $(S'\Omega^{-1}S)^{-1}S'\Omega^{-1}\hat{\beta} = (S'S)^{-1}S'\hat{\beta}$. Note that

$$\Omega = [(\sigma_z^2 \sigma_v^2 + \sigma_\varepsilon^2)/\sigma_z^2]I. \tag{31}$$

Using (31) it is easily shown that both GLS and OLS provide identical estimators.

References

Amemiya, T. (1978). A note on a random coefficients model. *Internat. Econom. Rev.* 19, 793-796.

Borjas, G.J. and J. Mincer (1978). The distribution of earnings profiles in longitudinal data. In: Z. Griliches et al., eds., *Income Distribution and Economic Inequality*. Halsted Press, New York, 175-197.

Cooley, T. and E. Prescott (1973). Varying parameter regression, a theory and some applications. *Ann. Econom. Social Measurement* 2, 463-474.

Hanushek, E.A. (1973). Regional differences in the structure of earnings. *Rev. Econom. Statist.* 55, 204-213.

Hanushek, E.A. (1974). Efficient estimators for regressing regression coefficients. *Amer. Statist.* 28, 66-67.

Hildreth, C. and J. Houck (1968). Some estimators for a linear model with random coefficients. *J. Amer. Statist. Assoc.* 63, 584-595.

Hsiao, C. (1975). Some estimation methods for a random coefficients model. *Econometrica* 43, 305-325.

Saxonhouse, G.R. (1976). Estimated parameters as dependent variables. *Amer. Econom. Rev.* 66, 178-183.

Wachter, M.L. (1970). Relative wage equations for United States manufacturing industries 1947-1967. *Rev. Econom. Statist.* 52, 405-410.