Regional Science and Urban Economics 11 (1981) 135–147. North-Holland

APPROPRIATE ROLES FOR STATISTICAL DECISION THEORY AND HYPOTHESIS TESTING IN MODEL SELECTION

An Exposition

Andrew J. BUCK and Simon HAKIM*

Temple University, Philadelphia, PA 19122, USA

Statistical studies in the social sciences often analyze a cross section of communities that are assumed to be homogeneous for certain characteristics. However, the researcher might be interested in investigating the general behavior of all communities as well as particular differences between groups of communities that are spatially separated. This paper suggests, and empirically examines, a procedure which utilizes both statistical decision theory and hypothesis testing and enables us to determine the best model for estimating behavioral relationships for the full sample and for subgroups. The paper examines property crime occurrences in 94 New Jersey suburban communities of Philadelphia utilizing the full sample as well as subgroups within it. The purpose of grouping is to further understand the possible interjurisdictional mobility of urban criminals. The study suggests possible factors that induce urban criminals to commit crimes in the suburbs.

1. Introduction

Statistical studies in the social sciences often analyze a sample drawn from a population that is assumed to be homogeneous with respect to certain characteristics. However, the population and the sample may have subgroups which exhibit different behavior for some of the characteristics. For example, data on property crime occurrence in contiguous localities might be distinguishable as two functionally and spatially distinct subgroups; an incremental increase in commercial development in areas accessible to the urban core may attract more property crime than less accessible areas. Quite often we ignore these differences and pool the data in order to gain more degrees of freedom from a limited number of observations [e.g., Balestra and Nerlove (1966)]. When pooling is not explicitly treated as a model selection problem we may overlook significant elements of analysis, e.g., intra-sample differences and possible explanations for them.

The paper illustrates the methodology to be used in the regression analysis of data with distinct subgroups, utilizing and further elaborating, in sequence, statistical decision theory and hypothesis testing.

The decision rule developed in this paper was first proposed by Akaike (1972, 1974, 1976). The Akaike Information Criterion (AIC) has been

*We would like to acknowledge the helpful comments of an anonymous referee.

proposed for variable selection and determination of the order of an ARIMA process. Here we expand his approach by showing that it can also be used for solving the data pooling problem. The decision rule provides us with a statistical criterion for model selection; one chooses that model which maximizes the empirical log likelihood function subject to a penalty for the indiscriminate inclusion of variables. Use of a decision rule imposes on the researcher consideration of the losses associated with both Type I and Type II errors when selecting from among alternative models. Quite often researchers use hypothesis testing for solving a problem in decision making. This is inappropriate methodology as it engages the use of pretest estimators for the purpose of refining the model that is finally selected. Frequently the final model for pooled data, for example, will be restricted to variables that are significant for the sample of pooled data and will omit variables that are not significant for subgroups. That is, on the basis of a test of hypothesis we pool the entire sample for all variables.

If data pooling is not appropriate for the selected model, on the basis of the decision theoretic outcome, then hypothesis testing should be used to determine the nature of intergroup differences. An empirical examination of a method suggested by Searle (1971) for the analysis of covariances is presented. The strength of Searle's method is in identifying group differences in the presence of multicollinearity. This procedure, outlined below, is seldom used and seems to be of special value to social scientists who examine group differences in regression analysis.

To illustrate the methodology proposed we use several models of criminal behavior. The data and a general model are explained in section 2. Also contained in section 2 is an explanation of the use of Akaike's Information Criterion (AIC) for model selection. The third section explores Searle's method as it deals with identifying the sources of group differences that resulted in the selection of a particular model under the AIC in section 2. In the final section we present our conclusions with regard to the appropriate use of statistical decision theory for purposes of model selection and hypothesis testing for purposes of deriving conclusions from the selected model.

2. Model selection

2.1. The AIC decision rule

The decision to combine two groups of communities for model estimation is felt by some to be a decision-theoretic problem. For that reason we consider the application of an information theoretic decision rule to the problem of pooling data.

In this section we present the general development of the AIC decision rule. In section 2.2 we extend the work to the problem of data pooling and in section 2.3 we consider the significance level implied by the use of an information decision rule in model selection.

The pioneering work in the area of statistical information theory has been done by Kullback (1959), Akaike (1972), and Sawa (1978). Their work has dealt with the development of a decision rule for the choice of additional variables to be included in a regression model. That is, they have considered the choice of a model from among alternatives that are subsets of a general model which contains all the possible variables. The alternative, or nested, models are derived by imposing zero restrictions on the general model.

For any statistical modelling problem researchers are faced with the problem of constructing a correct model. In general, the correct model exists over, say, an l-dimensional space, i.e., there are l independent variables. However, for lack of knowledge of the economic process or lack of data we restrict our search for an appropriate model to spaces of smaller dimension than l. Within the restricted space of possible models we would like to choose the model that is most similar to the correct model. Whatever measure of distance or model adequacy that we use should contain a penalty for indiscriminate inclusion of additional variables, i.e., those that cost us more in loss of efficiency than they yield in improved goodness of fit.

Let us begin the derivation of a measure of model adequacy by defining $g(Y)$ as the density function of the true probability distribution $G(Y)$ for a vector of continuous random variables $Y'=(y_1,\ldots,y_n)$ that we wish to model. Also let $f(Y\,|\,X\beta,\sigma^2)$ be a model for the unknown $g(Y)$, where β and σ^2 are unknown parameters and X is a set of independent variables. We adopt as our final model that which gives the maximum of the expected log likelihood, which is

$$S(g(Y):f(Y\,|\,X\beta,\sigma^2))=\int g(Y)\log f(Y\,|\,X\beta,\sigma^2)\mathrm{d}Y. \tag{1}$$

From the principle of maximum likelihood estimation we know that the log likelihood will be sensitive to deviations of $f(Y\,|\,X\beta,\sigma^2)$ from $g(Y)$. Thus, as a measure of model adequacy we use

$$I(G(Y):F(Y\,|\,X\beta,\sigma^2))=S(g:g)-S(g:f). \tag{2}$$

This may be written more explicitly as

$$I(G(Y):F(Y\,|\,X\beta,\sigma^2))=\int g(Y)\log\left(\frac{g(Y)}{f(Y\,|\,X\beta,\sigma^2)}\right). \tag{3}$$

The integral on the right-hand side is a measure of the distance between $g(Y)$ and $f(Y\,|\,X\beta,\sigma^2)$. While there are numerous rigorous arguments for choosing (3) as a measure of model adequacy, which we do not present here, the intuition is quite simple. We would like some measure of how surprised we are by the extent to which the sample data does not agree with the 'true' model: the further from the 'true' model the greater the surprise, since we thought that our specification was correct. Eq. (3) is a monotone increasing measure of 'surprise' that allows us to compare competing models. If two possible models are being entertained then one chooses the model for which the integral in (3) is the smallest.

An obvious shortcoming of (3) is that $g(Y)$ is unknown. However, Akaike has shown that under some fairly weak regularity conditions it is possible to estimate (3). An almost unbiased estimator of the measure presented in (3) is given by

$$\mathrm{AIC} = -2\log f(Y\,|\,X\beta,\sigma^2)+2k, \tag{4}$$

where k is the number of model parameters. The value of the empirical log likelihood $[-2\log f(y\,|\,X\beta,\sigma^2)]$ can be made smaller by adding more variables. The term $2k$ penalizes the researcher for the indiscriminate use of additional variables. In comparing two models one chooses the model with the numerically smallest AIC, regardless of the size of the absolute difference between the numbers.

2.2. *Data pooling and the AIC*

From the above comments, the extension of the AIC to the problem of data pooling is quite straightforward. With the proper choice of restrictions, a model which allows all the parameters of the two groups to be different can be turned into a model that is equivalent to pooling the data. To show how this is done suppose we have observations on k variables for each of two groups of size n_a and n_b. If the two groups are thought to be dissimilar then one should specify the following model:

$$\begin{bmatrix} Y^a \\ Y^b \end{bmatrix} = \begin{bmatrix} X^a & 0 \\ 0 & X^b \end{bmatrix} \begin{bmatrix} \beta^a \\ \beta^b \end{bmatrix} + \begin{bmatrix} U^a \\ U^b \end{bmatrix}. \tag{5}$$

The superscripts identify the group. Note that there are $n_a + n_b$ observations and $2k$ coefficients. When the two groups are thought to be part of the same population then (5) should be estimated subject to the restriction

$$I_k\beta^a - I_k\beta^b = 0, \tag{6}$$

where I_k is a k-dimensional identity matrix. This would be equivalent to estimating the parameters of the pooled model

$$\begin{bmatrix} Y^a \\ Y^b \end{bmatrix} = \begin{bmatrix} X^a \\ X^b \end{bmatrix} \beta + \begin{bmatrix} U^a \\ U^b \end{bmatrix}. \tag{7}$$

Now there are only k coefficients to be estimated from $n_a + n_b$ observations. Thus, with the proper zero restrictions, given in (6), the pooled model (7) is a nested alternative to the model in (5). The model in (7) would correspond to a null hypothesis that the two groups are homogeneous. The corresponding alternate hypothesis would be that of heterogeneity, implying model (5).

For illustrative purposes we propose a model which explains the import of property crime to suburban localities. Offenders are attracted to communities in direct relation to their accessibility and the crime opportunities they offer, and inversely to any deterrent characteristics that may be present. The linear model can be specified as

$$Y_i = \alpha_0 + \alpha_1 d_i + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}, \tag{8}$$

where Y is the ratio of the number of property crimes to acres of developed area and i denotes the community. Crime opportunity, X_1, is measured by commercial land use as a percent of total developed area. A second crime opportunity variable, X_2, is equalized assessed real estate valuation per developed acre. The deterrent efforts of a community, X_3, are measured by per capita police expenditure. A dummy variable, d, is used to classify communities according to their accessibility to an urban core.

The sample consists of 94 suburban and rural jurisdictions in southern New Jersey which are part of the Philadelphia metropolitan area. The sample could be split into two parts according to the communities' location with respect to the urban center and their economic attributes. The first group (Group a) consists of 26 urbanized suburbs of Philadelphia and Camden, contiguous with each other and located along major arterial roads. The remaining 68 communities (Group b) are mostly rural, inaccessible and/or in early stages of their economic development.

The model choice in our example involves specifications that either treat all communities as similar in all respects, or the communities may be separated into groups; the difference is the use of a four slope parameter model or a model with eight slope parameters respectively. For the pooled model, i.e., eq. (1) of table 1, the AIC computed from eq. (4) is 0.00435. For the eight parameter model the matrix of observations on the independent variables is block diagonal, corresponding to eq. (5). The estimation of the eight parameters is reported as eqs. (2) and (3) of table 1, which is done in

Table 1
Estimate of alternative model parameters.[a]

Eq.	Intercept	P^a	P^b	S^a	S^b	V^a	V^b	X_1	X_2	X_3	d	RSS^b	F	R^2	Degrees of freedom
(1)	−0.00846							0.03726[c]	1.5×10^{-6c}	0.00119	0.2208	0.3551	41.41	0.65	89
	(−0.53)							(6.2)	(3.01)	(1.10)	(1.20)				
(2)	0.0595							0.0253[c]	3.3×10^{-6c}	−0.0056		0.1851	5.94	0.45	22
	(0.67)							(2.00)	(2.9)	(−1.36)					
(3)	0.0016							0.0442[c]	3.3×10^{-7}	0.0027[c]		0.1184	47.11	0.69	64
	(0.13)							(7.48)	(0.78)	(3.40)					
(4)	0.03675[c]	0.033[c]	0.0414[c]						1.4×10^{-6c}	0.00096	0.08765[c]	0.3529	33.05	0.65	88
	(2.04)	(3.9)	(4.96)						(2.94)	(0.85)	(4.8)				
(5)	0.3034[c]			2.6×10^{-6c}	0.8×10^{-6}			0.03803[c]		0.00167	0.05302[c]	0.3396	35.05	0.67	88
	(1.93)			(3.51)	(1.36)			(6.43)		(1.53)	(2.94)				
(6)	0.00754					−0.0036	0.00199	0.03476[c]	1.5×10^{-6c}		0.03169	0.3406	34.89	0.66	88
	(0.47)					(−1.33)	(1.74)	(5.74)	(3.11)		(1.65)				
(7)	0.05375[c]			3.1×10^{-6c}	0.56×10^{-6}	−0.00493	0.00301[c]	0.03470			0.06697[c]	0.3129	32.57	0.69	87
	(5.50)			(4.19)	(1.0)	(−1.86)	(2.59)	(5.96)			(3.93)				

[a]Numbers in parentheses are t-statistics.
[b]RSS denotes residual sum of squares.
[c]Coefficient significantly different from zero at 5% level.

order to preserve compactness of the table. The AIC for the eight parameter model is found to be 0.00413.

The AIC for the pooled model [eq. (1), table 1] is larger than that for the eight parameter model [eqs. (2) and (3), table 1]. According to the development of section 2.1 we should conclude that the general specification more closely represents the true state of the world.

2.3. The level of significance of the AIC

In this section we consider the level of significance associated with the AIC decision rule. Although the decision rule presented here is a mechanical rule for model selection it does not preclude the possibility of choosing incorrectly, i.e., it is statistically possible to select a model that is not closest to the true model. The error is similar to the rejection of a true null in hypothesis testing. As a consequence there is an implied significance level for the AIC decision rule that is determined by the structure of the problem. The level of significance implied by the minimum AIC procedure is calculated as follows: Let the subscript 1 denote the pooled model, given by (7), and 2 denote the model given by (5). Then comparing the AICs gives

$$\mathrm{AIC}_1 - \mathrm{AIC}_2 = (n_a + n_b)\ln(\hat{\sigma}_1^2/\hat{\sigma}_2^2) - 2(k_2 - k_1), \tag{9}$$

where k denotes the number of independent variables in the respective models. When the difference in (9) is less than zero we choose the pooled model.

The difference in (9) is equivalent to checking the inequality

$$\frac{\hat{\sigma}_1^2 - \hat{\sigma}_2^2}{\hat{\sigma}_2^2} \gtrless \exp\{2(k_2 - k_1)/(n_a + n_b)\} - 1. \tag{10}$$

If the left-hand side is less than the term on the right-hand side then the pooled model is selected.

Under the usual assumption of a normally distributed error term, the left-hand side of (10) can be multiplied by $(n_a + n_b - k_2)/(k_2 - k_1)$ to give an F statistic. Multiplying the right-hand side of (10) by the same term gives the implied critical point:

$$\frac{(n_a + n_b - k_2)}{k_2 - k_1}\{\exp(2(k_2 - k_1)/(n_a + n_b)) - 1\}. \tag{11}$$

Thus, for our example, the significance level is given by

$$P\Big\{F(k_2-k_1,n_a+n_b-k_2)$$
$$>[\exp(2(k_2-k_1)/(n_a+n_b-k_2))-1]\cdot\left(\frac{n_a+n_b-k_2}{k_2-k_1}\right)\Big\}$$
$$=P\{F(4,86)>1.9099\}\doteq 0.11.$$

Thus, the implied significance level for our example is less conservative than the significance level normally chosen by social scientists when using hypothesis testing to answer the same type of question. For smaller degrees of freedom the significance level implied by the procedure outlined above becomes larger, i.e., less conservative. The result of this section suggests that treating the two groups separately is statistically preferred to pooling the sample.

3. Determination of group differences

In this section we introduce a method based upon the work of Searle, in order to determine significant differences between groups in regression analysis.

The common method of identifying group differences is by introducing an intercept and/or slope dummy variable(s), a la Chow (1960). Multicollinearity between the dummy variable and other independent variables is often a problem in such a setup. For example, in our New Jersey data set wealthier communities also tend to be more economically developed and more accessible from the urban centers. Wealthier localities are more urbanized than are poorer localities. Also, careful observation reveals that the wealthy communities support more commercial establishments, by dollar value, than do poorer communities. The result is diminished t-statistics of the coefficients of the variables suspected of being highly collinear and erroneous inference.

Searle's method can be used for mitigating the impact of multicollinearity on the standard tests of hypothesis, and enables testing differences of both intercepts and slopes between groups. Essentially it entails measuring the independent variables as deviations from the group means and weighting by the value of the dummy variable. We adopt the following notation:

$$P^a=(X_1-\bar{X}_1^a)'d,$$
$$P^b=(X_1-\bar{X}_1^b)'(1-d),$$
$$S^a=(X_2-\bar{X}_2^a)'d,$$
$$S^b=(X_2-\bar{X}_2^b)'(1-d),$$
$$V^a=(X_3-\bar{X}_3^a)'d,$$
$$V^b=(X_3-\bar{X}_3^b)'(1-d),$$

where all variables represent 94 element column vectors corresponding to the original variables. $\bar{X}_j^t$ denotes the mean of the jth variable for the tth group. The pairs of new variables will have zero inner products. Also, the inner product between the group dummy and any transformed variable will be zero. The consequence is that the corresponding correlation coefficients will be zero.

The advantage to be gained from constructing the transformed variables P^t, S^t, and V^t, where superscript t defines the group association, $t=a,b$, is that we may now examine the different effect that any given independent variable might have in the two groups. In other words, in the regression form [e.g., eq. (4) in table 1] we might observe $\partial Y/\partial P^{\mathrm{a}} \gtreqless \partial Y/\partial P^{\mathrm{b}}$. At the same time the transformed variables have zero or reduced simple correlations. Thus we may test the coefficients without the distortions introduced by multicollinearity.

At least one of the original independent variables, X_j, should not be transformed in order to avoid splitting the observations into two independent samples, which will yield parameter estimates for two independent equations. Referring to eq. (5) it becomes clear that by introducing P^t, S^t, and V^t we are estimating separately the equations for groups a and b. If, however, we maintain one (or more) of the original variables X_j, which does not exhibit different group behavior, then we can still estimate the parameters of a single equation, without unnecessary loss of degrees of freedom.

In order to analyze the different group behavior for the two independent variables we tested three separate equations. In each one we transformed the original variables into P^t, S^t, and V^t, respectively [eqs. (4) through (6) in table 1]. The three tests of hypotheses outlined below were used to determine the independent variables which exhibit different group behavior. Eq. (7) illustrates the final model obtained from the three hypothesis tests.

Let the general model be

$$Y=\beta_0+\beta_1 P^{\mathrm{a}}+\beta_2 P^{\mathrm{b}}+\beta_3 S^{\mathrm{a}}+\beta_4 S^{\mathrm{b}}+\beta_5 V^{\mathrm{a}}+\beta_6 V^{\mathrm{b}}+\beta_7 X_1$$
$$+\beta_8 X_2+\beta_9 X_3+\beta_{10} d+U.$$

Three one-tail tests of the group coefficients for the variables P^t, S^t, and V^t are outlined below.

The first test of hypothesis is for the difference between the coefficients on P^{a} and P^{b}. We hypothesize that commercialization is a stronger crime attractant in Group a communities than in Group b communities:

$$H_0:\beta_1-\beta_2=0,$$

$$H_1:\beta_1-\beta_2>0.$$

Note

$$\operatorname{var}(\beta_1-\beta_2)=\operatorname{var}(\beta_1)+\operatorname{var}(\beta_2)-2\operatorname{cov}(\beta_1,\beta_2),$$

but since P^a and P^b are orthogonal, by construction, the third term is zero. Therefore the observed t-statistic is

$$t=\frac{\hat{\beta}_1-\hat{\beta}_2}{\sqrt{\operatorname{var}(\hat{\beta}_1-\hat{\beta}_2)}}=\frac{0.033-0.0414}{\sqrt{0.008461^2+0.008346^2}}\approx -0.707.$$

For a one tail test the critical value at the 5% level is 1.69. Thus, the coefficient on P^a and P^b are not significantly different from each other.

We hypothesize that more developed communities which are wealthier exhibit relatively greater opportunities to criminals than poorer communities. The effect of wealth concentration on crime is higher in more developed places than in less developed communities. In a similar fashion the observed t for the test on the S^a and S^b coefficients:

$$H_0:\beta_3-\beta_4=0,$$
$$H_1:\beta_3-\beta_4>0,$$

is given by

$$t=\frac{(2.6-0.8)\times 10^{-6}}{\sqrt{(0.7407\times 10^{-6})^2+(0.5882\times 10^{-6})^2}}\approx 1.903$$

For a one tail test the coefficients of S^a and S^b are significantly different from each other at the 5% confidence level.

We hypothesize a deterrent effect of policing on the crime level, i.e., a negative sign for the coefficients of V^a and V^b. The level of policing which is exhibited in the less developed localities might have low or even no deterrent effect. On the other hand, in the wealthy more developed places which exhibit higher concentrations of police, activities might experience a greater deterrent effect of policing. Thus, this test, for the difference between the coefficients on V^a and V^b is

$$H_0:\beta_5-\beta_6=0,$$
$$H_1:\beta_5-\beta_6<0,$$

the observed t is

$$t=\frac{-0.0036-0.00199}{\sqrt{0.0027^2+0.00114^2}}\approx -1.908$$

Again, for a one tail test, the coefficients of V^a and V^b are significantly different from each other at the 5% confidence level.

As a consequence of the three tests of hypothesis, we specified a final form of the model shown in eq. (7) of table 1. In eq. (7) it is found that the two groups have significantly different intercepts. That is, the mean level of property crime per developed acre is higher for the more accessible communities, a result which is not revealed otherwise.[1] The measure of crime opportunity, commercial land use, is found to be significant, although the two groups are not statistically different.

As tested above, commercial activities do attract property crimes. However the higher concentration of commercial land use in Group a does not yield a stronger impetus to criminals. The wealth variable appears significant for the accessible communities (Group a) but not for the others (Group b). This latter result may be due to the fact that the less accessible communities (Group b) have low levels of crime attractions (rather than commercial opportunities), hence wealth is not sufficient to attract the criminal. For the accessible communities the police expenditure variable appears to deter criminals, hence has the correct sign, and in a one tail test we can reject the null hypothesis at the 5% level of significance. For the less accessible communities (Group b), the police expenditure coefficient is positive and significant. [For a theoretical explanation of this phenomenon see Allison (1972), Zipin (1974), McPheters and Stronge (1974), and Baumol (1967).]

In the presence of high multicollinearity a great deal of information is revealed by Searle's transformation. By using this technique we reduce multicollinearity and improve estimation without losing any information contained in the variables which are transformed. It also allows us to utilize the full sample in order to estimate independent variables which do not exhibit different behavior for the two groups. In particular, the level of property crime per developed acre is significantly higher for Group a communities even when the effects of wealth, commercial concentration and police performance are equalized for all communities.

4. Conclusions

In this paper we have demonstrated and further elaborated the use of

[1]The coefficient of d in eq. (1) table 1 is not significantly different from zero at the 5% level. We have also tested for significant group differences for the other independent variables using the method of Chow (1960). [See also Fisher (1970), Kuh (1963), and Booms (1966).] This method is often used incorrectly in deciding on the appropriateness of data pooling since it involves hypothesis testing without standardized criterion for the choice of Type I and Type II errors. Also, since collinearity between the dummy variable and other independent variables is not controlled under this method, the true group differences are not revealed. For example, in our New Jersey analysis the Chow test resulted in conflicting conclusions from the appropriate hypotheses. The Chow test conclusions were also contrary to what we had observed in the communities used here.

Akaike's information criterion for model selection. The result of that selection procedure was the choice, for our case study, of a model that treated the two groups of communities as distinctly different. The AIC does not permit identification of the source of those differences, but importantly, does show that there are differences. Given that there are differences an extension of hypothesis testing permits identification of those variables which differed between groups.

Turning to statistical hypothesis testing we considered a technique based upon Searle's work. The method suggested by Searle, with further implementation on our part, through its linear transformation of the independent variables, eliminates the collinearity between the group dummy and the independent variables, and between pairs of independent variables, thus improving all parameter estimates. Using Searle's approach we can now distinguish the sources of the differences between the groups, sources that would be hidden under other more commonly used procedures.[2]

The study suggests the use of an extension of Akaike's information criterion in order to decide whether to pool sample data or to classify data by sample subgroups. Once such a decision is made, we suggest the use of Searle's technique in order to test hypotheses concerning the different effects of the independent variables for the two groups. We believe that in contrast to current practices, the use of both statistical decision theory techniques and hypotheses testing is necessary for better explanation of significantly different group behavior in the context of multivariate regression analysis.

[2]Note that the Searle methodology is merely the Chow test nested within it. If multicollinearity were not present then the two approaches would yield the same conclusions.

References

Akaike, H., 1972, Information theory and an extension of the maximum likelihood principle, Proceedings of the Second International Symposium on Information Theory, 267–281.

Akaike, H., 1974, A new look at the statistical model identification, IEEE Transactions on Automatic Control, AC-19, 716–723.

Akaike, H., 1976, On entropy maximization principle, in: P.R. Krishraiah, ed., Proceedings of the Symposium on Applications of Statistics (North-Holland, New York) 99–123.

Allison, J.P., 1972, Economic factors and the rate of crime, Land Economics 48, 193–196.

Balestra, P. and M. Nerlove, 1966, Pooling cross section and time-series data in the estimation of a dynamic model: The demand for natural gas, Econometrica 34, 585–612.

Baumol, W.J., 1967, Macroeconomics of unbalanced growth: The anatomy of urban crisis, American Economic Review 57, 415–426.

Booms, B.H., 1966, City government form and public expenditure levels, National Tax Journal 19, 187–199.

Chow, G.C., 1960, Tests of equality between sets of coefficients in two linear regressions, Econometrica 28, 591–605.

Fisher, F.M., 1970, Tests of equality between sets of coefficients: An expository note, Econometrica 38, 361–366.

Kuh, E., 1963, Capital stock growth: A micro-econometric approach (North-Holland, Amsterdam) 115–157.

Kullback, S., 1959, Information theory and statistics (Wiley, New York).
McPheters, L.R. and W.R. Stronge, 1974, Law enforcement expenditures and urban crime, National Tax Journal 27, 633–644.
Sawa, T., 1978, Selection of variables in regression analysis, Econometrica 46, 1273–1292.
Searle, S.R., 1971, Linear models (Wiley, New York) 355–358.
Zipin, P.M. et al., 1974, Crime rates and public expenditures for police protection: A comment, Review of Social Economy 32, 222–225.