Inequality Constraints, Multicollinearity and Models of Police Expenditure*

ANDREW J. BUCK
Temple University and International Institute of Management
Philadelphia, Pennsylvania and Berlin, Germany

SIMON HAKIM
Temple University
Philadelphia, Pennsylvania

I. Introduction

Inequality constraints have been used very infrequently in the analysis of linear statistical models. This is not due to an inability to construct estimators. Rather, it is a result of an inability to say very much about the properties of the inequality constrained estimator. Hanson [15] proved the existence of inequality constrained maximum likelihood estimators. A year later Judge and Takayama [17] demonstrated the value of quadratic programming for obtaining estimates of the parameters in a model constrained by inequalities. More recently, Liew [20; 21] and Klemm and Sposito [19] have suggested closed form estimates. The work by Liew makes some reference to the properties of the inequality constrained least squares (ICLS) estimator. A more formal effort to determine the properties of the ICLS estimator was made by Lovell and Prescott [23]. Their paper considered only the case of one location parameter being restricted to either the positive or negative half of the real line. In a more recent paper Wardle [32] has considered a Bayesian approach to the problem of determining the small sample properties of the ICLS estimator. Our more classical results are a confirmation of his work.

The present paper generalizes the Lovell-Prescott constraint to a linear combination of the location parameters restricted to an interval. We will derive the mean and variance for the constrained estimator, and demonstrate the behavior of squared error risk for a particular example.

To illustrate the use of the ICLS estimator we constructed several models of community expenditure on police services. In the last decade the analysis of police expenditure has received considerable attention in the economics and public finance literature [2; 6; 27; 31]. A related question which has emerged following the pioneering work of Becker [3] analyzes the rational behavior of criminals and its effect on the supply of offenses. The interrelationship

* The authors wish to express their appreciation to Drs. Arnold Raphaelson and William Stull, and to an anonymous referee for their helpful comments.

between the supply of crimes and the level of police expenditure is a subject of an increasing number of studies [5; 11; 6; 29; 1; 25; 14]. Our illustrative model concerns the factors which determine the level of police services. The decision on the level of policing is based upon variables that measure the demand for police protection (wealth of the community and level of property crimes) and variables that determine the production of police services (effectiveness of policing and resource availability). Community disposable income seems to be the primary argument from both supply and demand attributes that explains police expenditure.

The cross section data on wealth, crime opportunity, police expenditure and property crime can be divided into two groups of communities. Separation of the groups is according to distance from the central city, which reflects accessibility for potential criminals and level of economic development. Using a dummy variable to indicate accessibility and development results in a high degree of multicollinearity between independent variables. Transforming the independent variables in a manner similar to that suggested by Searle [28] for use in the analysis of covariance accomplishes three purposes. First, it reduces the linear dependence between independent variables. Second, it allows slope parameters, in addition to intercepts, for the two groups to be different.[1] Finally, it allows us to use additional prior information about the effects of certain independent variables, in the form of inequality constraints, in our estimation procedure. By introducing sound prior information, we are able to improve our estimates of the remaining hypothesized coefficients.

The second section presents the general model and the assumptions to be used in the later sections. Also included in this section is a discussion of the choice of the constraint generalization that is considered in detail. The small sample properties of the ICLS estimators proposed in this section are explicitly derived.

In the third section we introduce the economic model which leads to the reduced form explaining police expenditure. The fourth section presents the various statistical models which differ in their specification of the same set of independent variables. The preferred model is selected on the basis of several statistical criteria and some of the independent variables are constrained based upon economic theory.

In the last section the interpretations and conclusions of the ICLS estimators are cast in the light of their usefulness in applied modelling. The results demonstrate the usefulness of the consideration of constraints in modelling police outlays and suggest other areas of analysis where these techniques may be similarly useful.

II. The Inequality Constrained Regression Model

Assumptions

On the basis of economic theory one may wish to estimate the parameters in the model

$$Y = X\beta + U \tag{1}$$

where X is an $n \times k$ matrix of full column rank, β is a $k \times 1$ vector and U is an $n \times 1$ vector of random errors. We make the additional assumption that the U_i, $i = 1, \cdots, n$ are distrib-

1. Heeding the warning of Green and Doll [10], we eschew the use of additional dummy variables for the purpose of estimating different slopes and intercepts as this exacerbates any existing collinearity.

uted independently and identically as $N(O, \sigma^2 I)$. As a consequence of these assumptions the ordinary least squares estimator for β is given by

$$\hat{\beta} = (X'X)^{-1}X'Y. \tag{2}$$

Not only is the estimator best linear unbiased but the $\hat{\beta}_i$, $i = 1, \cdots, k$ are distributed as $N(\beta, \sigma^2(X'X)^{-1})$.

In the following development we consider the general interval restriction

$$r_1 \leq H\beta \leq r_2 \tag{3}$$

where H is $1 \times k$ and r_1 and r_2 are real numbers. This specification permits a great deal of flexibility. The Lovell and Prescott [23] example is a special case of (3). It is the estimation of the parameters of (1) subject to the restriction of (3) that we consider here.

Properties of the Estimator

The estimator we propose to use is

$$\begin{aligned} \beta^* = \hat{\beta} &+ \{(X'X)^{-1}H'(r_1 - H\hat{\beta})/H(X'X)^{-1}\}\,\psi_{(-\infty,r_1)}(H\hat{\beta}) \\ &+ \{(X'X)^{-1}H'(r_2 - H\hat{\beta})/H(X'X)^{-1}H'\}\,\psi_{(r_2,\infty)}(H\hat{\beta}) \end{aligned} \tag{4}$$

where $\psi_{(\;)}(\cdot)$ is an indicator function that takes the value of one if the random variable falls in the subscripted interval and zero otherwise. This particular estimator is the result of the application of the Kuhn-Tucker theorem to the estimation problem outlined above. In essence, our ICLS estimator revises the least squares estimates by the middle term if $H\hat{\beta} < r_1$, and by the last term if $H\hat{\beta} > r_2$. If one were to construct the Lagrange expression for the minimization of the residual sum of squares subject to the inequality constraints of (3), then the second and third terms of (4) are seen to be the product of the Lagrange multipliers and $(X'X)^{-1}H'$.

The ICLS estimator is implemented through the use of quadratic programming, i.e., the residual sum of squares is minimized subject to (3). Whenever the constraints are binding the result is to project the least squares estimates obliquely onto the constraint.

The first moment of the ICLS estimator is found by integrating a weighted average of the terms of (4). The weights correspond to the probability density below r_1 and above r_2. Thus, the mean of the ICLS estimator is found to be

$$\begin{aligned} E\beta^* = \beta &- \{\phi[(r_2 - H\beta)/\sigma(H(X'X)^{-1}H')^{1/2}] - \phi[(r_1 - H\beta)/\sigma(H(X'X)^{-1}H')^{1/2}]\}\,(X'X)^{-1}H' \\ &+ \{[(r_2 - H\beta)/\sigma]P(H\hat{\beta} > r_2) - [(r_1 - H\beta)/\sigma]P(H\hat{\beta} < r_1)\}\,(X'X)^{-1}H' \end{aligned} \tag{5}$$

where ϕ is the standard normal density. Note that the terms in curly brackets are scalars. As the true value of $H\beta$ moves outside the interval (r_1, r_2) the ICLS estimator becomes more seriously biased. That is, if your prior information is correct and $r_2 - H\beta = H\beta - r_1$ then β^* is unbiased. The extent of the bias depends directly on the degree to which one's prior information is incorrect. As in the case of exact linear restrictions, it is unlikely that one's prior information is exactly correct. In the instance where r_1 and r_2 have been chosen so that $r_1 \leq H\beta \leq r_2$ is true, the choice of estimator should be based on mean square error or risk, both of which consider bias and variance.

Using Φ and F to denote the cumulative normal probability function, the variance of the estimator proposed in (4) may be written as

$$\mathrm{Var}(\beta^*) = \sigma^2(X'X)^{-1} + \sigma^2 \Big\{ 1 - \Phi(H\hat{\beta} - r_2) - \Phi(H\hat{\beta} - r_1) + \int_{H\beta<r_2} (r_2 - H\beta)dF(H\hat{\beta})$$

$$\cdot \int_{H\beta>r_2} (r_2 - H\beta)dF(H\hat{\beta}) + \int_{H\beta<r_1} (H\beta - r_1)dF(H\hat{\beta}) \int_{H\beta>r_1} (H\beta - r_1)dF(H\hat{\beta})$$

$$- 2(\phi(H\beta - r_1)\phi(r_2 - H\beta) + (H\beta - r_1)(r_2 - H\beta)\Phi(H\hat{\beta} - r_1)\Phi(r_2 - H\hat{\beta})$$

$$- (r_2 - H\beta)\phi(H\beta - r_1)\Phi(r_2 - H\hat{\beta})$$

$$- (H\beta - r_1)\phi(r_2 - H\beta)\Phi(H\hat{\beta} - r_1))\Big\} (X'X)^{-1}H'H. \tag{6}$$

The variance of the ICLS estimator is less than that of the ordinary least squares estimator. In curly brackets, the second through fifth terms are negative. The sixth through ninth terms may be thought of as a quadratic and must have a sum greater than zero, multiplication by -2 makes the sum negative. Although not readily apparent, the sum of all the terms in curly brackets will be negative. Intuitively, the ICLS estimator has a smaller variance because the density is concentrated in a smaller range than for OLS.

To summarize the analysis thus far, the ICLS estimator is biased, but with smaller variance. Of course, all restricted estimators will have smaller variances than comparable unrestricted estimators. A better means of comparison would be on the basis of squared error risk, defined by

$$\mu(\hat{\beta}, \beta) = E\|\hat{\beta} - \beta\|^2. \tag{7}$$

For purposes of exposition we consider the risk associated with a particular example. For the example, we make the additional assumption that $X'X = I$ and define the restriction vector to be $H = (0 \cdots 0\ 1\ 0 \cdots 0)$. The prior restriction is of the form $H\beta > r$. In this case the risk of the ICLS estimator is

$$\mu(\beta^*, \beta) = k\sigma^2 - \sigma^2\Phi[(r - \hat{\beta}_i)/\sigma] + (r - \beta_i)^2\Phi[(r - \hat{\beta}_i)/\sigma] + \sigma\Phi[(r - \beta_i)/\sigma]. \tag{8}$$

Figure 1 shows the relationship between ordinary least squares, equality restricted least squares (RLS) and the ICLS estimator in terms of risk. When the constraint is close to being binding, then the exact restriction estimator has the smallest risk. The ICLS estimator dominates only where the restriction is well below the value of the parameter.

III. A Model of Police Expenditure

Our model of police expenditure considers a community which both produces and consumes an abstract good which we shall call security, denoted S. The model we shall develop is adapted from Lindahl [22] and is an economy with personalized price for the public commodity. Prices that are unique to each consumer could be established following the procedures suggested by Malinvaud [24]. One obvious requirement is that consumers correctly reveal their preferences.[2]

2. One might argue that preferences are correctly revealed by one's choice of residence, both between communities and within a particular community [30]. The fact that security varies within the community and we do not assume identical utility curves necessitates the use of personalized prices.

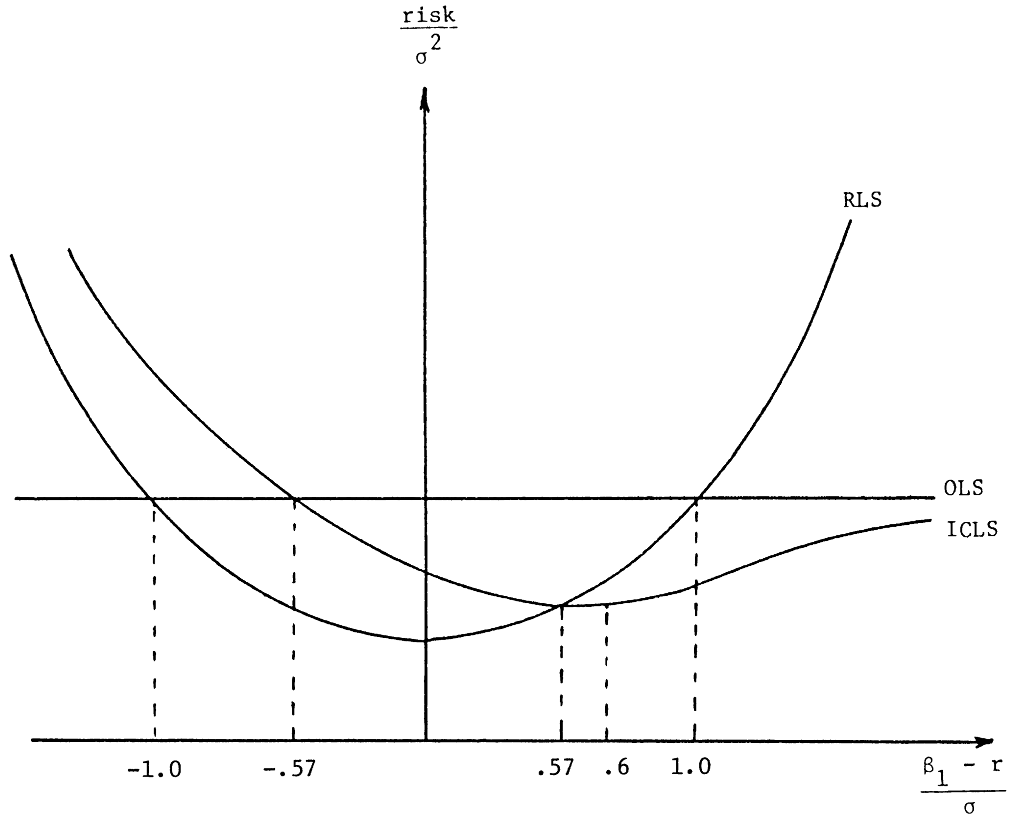

Figure 1. Risk Functions of RLS, OLS and ICLS Estimators

The model developed here differs from the traditional police expenditure models [1; 25; 29] in that we do not deal with the supply of and demand for criminal offenses. It seems to us that although security is a very abstract service, it is more encompassing and more accurately represents what the community believes it is buying from its police department.

Unless otherwise noted, our model of the supply of and demand for security makes all the usual regularity, continuity and convexity assumptions for preferences and production.

As consumers of security, the members of the community wish to maximize the utility gains from the consumption of S as well as all other goods (Z) subject to a budget constraint. Thus, for each of n individuals in the community we have

$$\underset{s,z}{\text{Max}}\ U_i(S_i, Z_i) \quad i = 1, \cdots, n \tag{9}$$

subject to the budget constraint

$$B_i = p_s^i S_i + P_z Z_i \tag{10}$$

where B is income and the P's are prices.

From the first order conditions from the maximization of (9) subject to (10) we can derive the individual's demand curve for security. Each individual is confronted with his personalized price of a unit of security and he announces truthfully his desired level of consumption. The security provided to the community is determined by the largest quantity demanded by any individual in the jurisdiction. Since security is a public good we are actually summing the individual demand curves vertically. The market demand curve for security will be given by:

$$S^d = S^d\left(\sum_{i=1}^{n} P_s^i, P_z, B_i\right) \tag{11}$$

We know additionally that the total amount spent on security must equal a given fraction of tax payments. That is,

$$S \sum_{i=1}^{n} P_s^i = k \sum_{i=1}^{n} (t_w W_i + t_c COM_i) \tag{12}$$

where t_w is the tax rate on private property (W), t_c is the tax rate on commercial property (COM) and k is the exogenously determined fraction of the community budget designated for security.[3] Thus, when either W or COM increase, expenditure on security also increases.

We now consider the community as a producer of security. As a monopoly producer of security, the police department faces a zero profit constraint imposed by the residents. From particular choices of capital (K) and labor (L) inputs the police department allows a certain level of crime (PC) and recovers some percentage of stolen property (PPR).[4,5] These tangible results of the police function in combination with a set of demographic and economic characteristics that assigns the community to either one of two groups (A), produces a level of security, S. Thus, the maximization problem at hand is as follows:

$$\max_{K,L} S = f(PPR, PC, A) \tag{13}$$

subject to the profit and production constraints

$$S(\Sigma P_s^i) = P_K K + P_L L \tag{14}$$

$$PPR = g(K, L) \tag{15}$$

$$PC = h(K, L) \tag{16}$$

From the maximization of (13) subject to (14) to (16) we can determine the quantity of security supplied. That is

$$S^s = S^s\left(P_K, P_L, \sum_{i=1}^{n} P_s^i\right) \qquad i = 1, \cdots, n. \tag{17}$$

3. The community continues to spend on police services up to the level at which the marginal benefit of the last dollar spent is equal for all local public services. Since this issue is only marginally related to our subject, the determination of k is not being treated explicitly.

4. Several researchers have examined the complexity of defining and measuring police output. Security is a joint product which is a function of both inputs and other socioeconomic and physical attributes of the community [13, 41–44; 26; 18, 8; 16, 351].

5. PC and PPR may be thought of as intermediate outputs of the police department. In any case, they are more easily measured and observed than K and L, hence their introduction here. We assume

$$\partial S/\partial PC < 0, \qquad \partial^2 S/\partial PC^2 > 0.$$

The literature [5; 9; 13, 50–52] presents several reporting problems associated with PC and PPR. Most of the problems can be resolved by improved reporting or econometric estimation of the unreported crimes. On the other hand, the theoretical problems in measuring capital and labor are much more complex.

It should be noted that $S\Sigma P_s^i$ is equal to the right hand side of (12). In equilibrium we have equality between (11) and (17).[6]

Statistical modelling of the market for security presents several difficulties. The obvious problem is that one cannot measure the number of units of security consumed by a community, nor can we observe the unit price. Furthermore, there are considerable difficulties in observing P_K and P_L.

We have resolved these problems by recognizing that while we cannot observe P_s and S we can observe the total expenditure on security, ($S\Sigma P_s^i$). Further, while we cannot easily observe P_K or P_L we can observe PC and PPR, the intermediate outputs of police departments which depend on the quantities of K and L employed, and in turn on P_K and P_L.

Thus we specify a reduced form equation that characterizes equilibrium in the market for security as:

$$S \sum_{i=1}^{n} P_s^i = F(PC, W, PPR, COM, A). \tag{18}$$

The reduced form model is analyzed using 1970 data on sixty-one suburban and rural incorporated communities in the New Jersey part of the Philadelphia Standard Metropolitan Area which have their own police department. Generally speaking, the municipalities range from established, wealthy suburban centers, in very close proximity to the major urban areas—Philadelphia and Camden—to remote rural farmlands and forest areas.

The region can be easily divided into two groups of communities which differ in their housing conditions and age, socioeconomic characteristics of their population and other relevant crime attraction attributes. Table I defines the variables and identifies the data source. Police expenditure is standardized by the acreage of the community's developed area, and not by population size. Our approach allows us to model the comparative intensity (e.g. patrolling) of protection provided by local police departments, while an expenditure per capita model focuses on the factors which explain the per capita fiscal burden communities choose to undertake. $S\Sigma P_s^i$ is a proxy for the intensity of policing in the part of the municipality where most of the policing takes place.

Standardization by acreage was statistically motivated as well. Examination of the residuals from our regression model before standardization by acreage showed a very strong relationship between geographic size of the community and dispersion of the residuals. Thus, all variables in the model, with the exception of the dummy variable A, have been standardized by community size.

PC is the property crime variable. The residents of a community recognize PC as an intermediate output of their police department. If, in a given class of community, property crime level is increasing, the residents will try to buy more security by spending a greater amount on police protection.

We have omitted violent crime from our behavioral model and the reduced form on the basis of previous empirical work [7; 11]. It would appear that crimes of passion do not respond to economic incentives. Thus, we assume that consumers of security recognize this and make their budgetary decisions on the basis of the prevailing rate of property crime.

6. We also find that

$$\sum_{i=1}^{n} P_s^i/P_z = \sum_{i=1}^{n} MRS_{s,z}{}^i = MRT_{s,z}.$$

Table I. Definition of Variables and Sources of Data

Notation	Description of Variable	Source of Data
$S\Sigma P_s^i$	Police expenditure per acre developed area	Division of Local Government Services: 1970 Statements of Financial Conditions of Counties and Municipalities, N.J. Department of Community Affairs.
PC	Property crimes per acre developed area	New Jersey, Uniform Crime Reporting Section, 1970 Uniform Crime Report, N.J. Attorney General.
W	State equalized real estate valuation density	Same as $S\Sigma P_s^i$.
PPR	Proportion of property recovered of total value property stolen	U.S. Federal Bureau of Investigation, 1970 National Uniform Crime Report, Supplement Data forms UCR26400.
COM	Proportion of commercial land of total developed land area	Delaware Valley Regional Planning Commission, 1970 Land Use File -- by Acres, DVRPC, Philadelphia, PA.
A	A dummy variable indicating the location of the community with respect to the central cities	Determined by the researchers. A=1 places are adjacent to or on arterial roads leading to the regional central cities.

W is real estate valuation per developed acre, which indicates the wealth of the community and the intensity of development. Real estate categories chosen to express the income effect on the community's decision on the amount spent on police service include all residential, commercial, and industrial land uses. The total dollar figure of real estate valuation is standardized by the total acreage of the developed part of the community. The proportional value of real estate that is used in municipal assessment for taxation purposes differs over communities. Thus, each municipality's assessment is standardized by an equalization ratio developed by the state of New Jersey. By so doing we determine comparable market values of real estate density across communities. It is expected that the wealthier the community, the higher its demand for all public services, including police, assuming that police is a normal good [7; 31].

PPR is the proportion of stolen property which is recovered. This output dimension of policing measures the effectiveness of police expenditure. The higher the output of policing, assuming efficient use of resources, the greater is the amount spent on police protection [31]. Thus, we hypothesize a positive relationship between *PPR* and $S\Sigma P_s^i$.

COM is the amount of land devoted to commercial establishments as a proportion of the total developed area. We know from prior information that commercial establishments attract property crime, and as a result their owners demand more intensive policing. Also, commercial establishments require proportionately more protection than do other land uses. Prior information requires a positive relationship between *COM* and $S\Sigma P_s^i$ [2; 25].

A is a dummy variable used to indicate the higher level of police expenditure in more accessible communities. The first ring suburbs experience more vehicular traffic and have more calls for police service. They usually have to serve residents, transients and tourists. Also, first ring suburbs have a lower population turnover, are wealthier, and are enriched with commercial activity. The region under study was divided into two groups of commu-

Table II. Sample Correlation of Variables Used in the Analysis[a]

	$S\Sigma P_s^i$	W	PC	PPR	COM	A	W1	W2	PC1	PC2	COM1	COM2
$S\Sigma P_s^i$	1.000											
W	.760*	1.000										
PC	.651*	.570*	1.000									
PPR	.387*	.142	.268*	1.000								
COM	.496*	.422*	.687*	.239	1.00							
A	.407*	.504*	.457*	.120	.434*	1.000						
W1	.492*	.641*	.306*	.084	.129	.000	1.000					
W2	.414*	.580*	.248*	.047	.209	.000	.000	1.000				
PC1	.250*	.347*	.652*	.184	.357*	.000	.490*	.000	1.000			
PC2	.514*	.226	.633*	.163	.445*	.000	.000	.391*	.000	1.000		
COM1	.070	.128	.316*	.132	.653*	.000	.197	.000	.505*	.000	1.000	
COM2	.441*	.196	.456*	.163	.622*	.000	.000	.336*	.000	.505*	.000	1.000

[a]Sample size is 61 communities in New Jersey portion of Philadelphia's S.M.S.A.
*Coefficient is significantly different from zero at 0.05 probability level.

nities which are different with respect to their accessibility from the central cities. Group 1 includes twenty five communities along major arterial roads and/or are adjacent to the two central cities Camden and Philadelphia ($A = 1$).

Group 2 includes thirty six localities which are not accessible by major roads from the two central cities ($A = 0$). A expresses different attributes of the two groups which are not reflected in the other independent variables. Given all explanatory variables at the same level, we expect the level of police activity to be higher for group 1 communities.

IV. Analysis

In order to examine the model, and illustrate the use of ICLS estimators we use multiple regression analysis, adjusted for the multicollinearity which is built into the model. We know that the accessible places are where the wealthier residents live and they generate more commercial activities. Thus, we expect high collinearity between A, COM and W. Improvement of our understanding of the separate effects of these variables on the dependent variable for each group requires reduction of the multicollinearity between A and these variables. Only if multicollinearity between W, COM, and A is reduced can we reveal the statistically significant effect of these variables on $S\Sigma P_s^i$.

Linearly transforming the original values of all cases in each group by subtracting the group's mean (which is constant for all cases within each group) retains all the explanatory power of the original variables, including that of the dummy variable. This actually shifts the means of the two groups to zero. The distribution within each group remains unchanged.

Multiplication of the transformed, continuous independent variables by the dummy variable, A for group 1 cases, or by its complement, $1 - A$ for group 2 cases, enables testing for differences in the slopes of the two groups, while maintaining independence with respect to the dummy variable. Note that the coefficient of each restated variable applies either to the cases in group 1, or to the cases in group 2.

Table III. Police Expenditure per Acre Developed Area: No Constraints[a,b]

	Equation 1	Equation 2	Equation 3	Equation 4
Intercept	-126.780 (-.778)	871.720 (7.494)	-45.273 (-.279)	80.663 (.455)
A	-7.469 (-.578)	36.103 (2.865)	-1.483 (-.118)	9.166 (.720)
PPR	36.875 (3.051)	36.808 (3.018)	36.719 (3.256)	37.740 (3.452)
W	2.565 (6.289)		2.558 (6.721)	2.679 (7.236)
COM	21.384 (.426)	22.676 (.445)		3.204 (.070)
PC	171.83 (2.134)	170.350 (2.089)	138.960 (1.831)	
W1		2.639 (4.935)		
W2		2,477 (4.289)		
COM 1			-73.074 (-1.300)	
COM 2			154.290 (2.407)	
PC 1				-19.374 (-.216)
PC 2				371.040 (4.073)
Adjusted R^2	.68	.67	.72	.73
F Value	26.01	21.31	26.44	28.71
Haitovsky χ^2	14.455	19.859	17.762	18.327
Significance	.154	.178	.276	.246

a. The numbers in parentheses are "t" values.

b. The following coefficients were scaled by .001: W, COM, W1, W2, COM1, COM2, PPR.

As an illustration of the transformation, the wealth variable becomes $W1 = (W - \bar{W}1)(A)$ and $W2 = (W - \bar{W}2)(1 - A)$, where the variables with bars over them correspond to group means.

Two criteria are used in order to select the restated continuous variables: (1) variables which theoretically seem to be highly correlated with the (0,1) dummy, and (2) variables which the data suggests will improve the prediction of the dependent variable when they are restated. This method applies only to the case in which at least one continuous variable is

Table III, Continued. Police Expenditure per Acre Developed Area: No Constraints[a,b]

	Equation 5	Equation 6	Equation 7	Equation 8
Intercept	966.910 (7.876)	99.065 (.604)	1113.300 (11.197)	1131.200 (15.368)
A	43.096 (3.54)	9.847 (.859)	54.363 (5.002)	55.441 (5.880)
PPR	36.348 (3.234)	37.466 (3.429)	37.345 (3.519)	37.054 (3.500)
W		2.648 (7.139)		
COM			10.324 (.233)	
PC	127.360 (1.670)			
W1	2.944 (5.868)		3.396 (6.824)	3.373 (6.789)
W2	2.098 (3.849)		1.894 (3.637)	1.853 (3.559)
COM 1	-77.455 (-1.381)	-33.545 (-.584)		-27.906 (-.501)
COM 2	176.680 (2.651)	65.346 (.875)		75.194 (1.036)
PC 1		11.931 (.126)	-84.228 (-.910)	-52.357 (-.542)
PC 2		309.230 (2.853)	410.560 (4.541)	346.591 (3.254)
Adjusted R^2	.72	.74	.75	.75
F Value	23.02	24.82	26.75	23.69
Heitovsky χ^2	22.134	12.356	29.781	13.268
Significance	.391	.930	.096	.920

a. The numbers in parentheses are "t" values.

b. The following coefficients were scaled by .001: W, COM, W1, W2, COM1, COM2, PPR.

not multiplied by the dummy variable or by its complement. When all continuous independent variables are restated, then we actually observe a separate equation for each group, as if the cases in each group were derived from different populations.

This method reduces multicollinearity between the dummy variable and some of the independent variables in the equation, hence improving parameter estimation. However, this method does not eliminate collinearity among all the independent variables.

Table II demonstrates how the method based on Searle eliminates multicollinearity. For example, while $R(W \cdot A) = .504$, the restatement of W to $W1$ and $W2$ produces bivariate correlations of $R(W1 \cdot A) = R(W2 \cdot A) = 0$. Also, this method eliminates collinearity between the restated variables for the two groups, e.g. $R(W1 \cdot W2) = R(PC1 \cdot PC2) = 0$.

Table III reports the regression results for models with the original and transformed variables. This table presents eight equations which differ in their specification of the same independent variables. Equation (1) analyzes all independent variables in their original magnitudes. It requires that the slope coefficients be the same for both groups, although it does allow for different intercepts. Note that in this equation the independent variables have not been corrected for multicollinearity. As expected, *PPR*, *W*, *COM* and *PC* are all positively related to $S\Sigma P_s^i$ which is consistent with our previous expectations. Equations (2) through (8) express different possible transformations of *W*, *PC*, and *COM* with respect to their group association. In each equation we used "t" tests in order to test for significant differences between the two groups for these three variables. *A* becomes significant and positively related to $S\Sigma P_s^i$ only where its collinearity with *W* is eliminated (equations (2), (5), (7) and (8)). Consistent with the inferences made from Table II the Haitovsky [12] χ^2 test statistic for multicollinearity was larger than that for equation (1), in all but equations (6) and (8).

Finally, Equation (7) was selected as the equation best expressing group differences. The criteria of choice included consistency with expected results, significance of coefficients, adjusted R^2, F-statistic for the model and Haitovsky's χ^2.

This model exhibits the simultaneous restatement of *W* and *PC*. Tests of significance on the coefficients for *W1* and *W2* reveal that they are different at the 5% level. The coefficients for *PC1* and *PC2* are also different at the 5% level.

Economic theory tells us that the coefficient on *W1* should be greater than the positive coefficient for *W2* and the coefficient on *PC1* should be greater than the positive coefficient for *PC2*. This may be due to the fact that *k* and/or t_w in equation (12) (both taken as exogenous in our model), are quite different for the two types of communities in our sample [13, 165–66]. Also, it has been shown elsewhere that wealthier communities attract additional crime more than in proportion to their greater wealth and thus the marginal effect of an additional dollar of wealth is greater in wealthier communities [8]. As for the differential effect of *PC* on police expenditure we rely on our assumption about the relationship between *PC* and *S* (see footnote 5).

Based upon these theoretical considerations we estimated equation (7) subject to two interval inequality constraints. The signs of *W1*, *W2*, *PC1*, and *PC2* were restricted to be positive as well as $W1 \geqq W2$ and $PC1 \geqq PC2$.

Estimation of the model parameters subject to the interval restrictions was implemented through the use of a quadratic programming routine. The ICLS estimates revise the OLS estimates by moving down the likelihood function as little as possible while still satisfying the constraints. The result of the constrained estimation follows:

$$S\Sigma P_s^i = \underset{}{1096.60} + \underset{(3.02)}{36.808\, PPR} + \underset{(.45)}{22.675\, COM} + \underset{(4.06)}{50.605\, A}$$

$$+ \underset{(4.93)}{2.639\, W1} + \underset{(4.29)}{2.478\, W2} + \underset{(2.09)}{170.350\, PC1} + \underset{(2.09)}{170.350\, PC2}$$

$$\text{Adjusted } R^2 = .67 \qquad (19)$$

Not surprisingly, the error sum of squares (adjusted R^2 is less) is greater for the constrained equation than for the unconstrained equation. The reason for the change in our estimates of the unconstrained parameters is a little less obvious. Table III indicates that for equation (7) the appropriate matrix of independent variables does not have mutually orthogonal columns. The result is that the revision of the OLS estimates is not an orthogonal projection back onto the constraints but represents moving down the likelihood contours as little as possible while still satisfying the constraints.

One of the principal results of Section II is that the moments of the ICLS estimator depend on some probabilities that cannot be known. As a consequence we cannot construct tests of hypothesis for the constrained parameters. More seriously, the constraints also affect the moments of the unconstrained parameters when the columns of the design matrix are not linearly independent. At best we can say that variances of the ICLS estimators are smaller than the variances for OLS estimators. Using "t" statistics calculated in the standard fashion (i.e. for an equation with a single exact linear restriction since the restriction on $W1$ and $W2$ is not binding), gives us conservative estimates of the true "t" statistic. These conservative estimates are presented in parentheses below the coefficients. The difference between the coefficients of $W1$ and $W2$ in equation (19) is not significant, however, their signs and order is as theoretically expected.

The only marked difference between the results reported in the constrained equation and equation (7) of Table III is the "significance" of the $PC1$ coefficient. We now find that it has the correct sign and is "significant".

Despite the drawbacks mentioned above, the ICLS estimator remains a valuable estimation method. As the Bayesians have been telling us for years, one should use all prior knowledge in estimation. The ICLS estimator is a fairly intuitive approach to the use of prior information, and one which most of us have been using implicitly all along.

V. Conclusion

In this paper we have derived the moments of some ICLS estimators. The moments of the ICLS estimator are shown to depend on unknowable probabilities. However, our results also show that for risk defined as $\mu(\hat{\beta}, \beta) = E||\hat{\beta}-\beta||^2$, the ICLS estimator is better than either the OLS or RLS estimator over some part of the parameter space. Lovell and Prescott have presented an example where this relationship may not be true. But, their example is a case that is seldom, if ever, encountered in economic research. The use of prior information in econometric research remains a valuable tool, as we have shown in this paper. Other examples of where one might use inequality constraints include analysis of consumption function or production function data.

We have also illustrated a valuable transformation for mitigating distortions induced by multicollinearity. By using a technique based on Searle's work, we are able to 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 estimating the coefficients on independent variables which do not exhibit different behavior for two groups.

As an application, a reduced form police expenditure function was derived from a behavioral model of the market for security and was empirically tested. The data were corrected for heteroscedasticity and multicollinearity. Our study reveals that, ceteris paribus, wealthy established communities adjacent to the central cities spend more, at the margin, on police than do other places, confirming prior expectations. This information was combined with prior information about the effect of property crime on police expenditure in the form of an inequality constrained model. By doing so we were able to incorporate prior information in our classical estimation routines. Thus, we would advocate the use of our constrained results for purposes of policy implementation or forecasting since they are theoretically correct.

References

1. Bahl, R. W., R. Gustely, and M. J. Wasylenko, "The Determinants of Local Government Police Expenditures: A Public Employment Approach." *National Tax Journal,* January 1978, 67–79.

2. Beaton, P. W., "The Determinants of Police Protection Expenditures." *National Tax Journal,* June 1974, 335–49.

3. Becker, G., "Crime and Punishment—An Economic Approach." *Journal of Political Economy,* March-April 1968, 169–217.

4. Buck, A., and Simon Hakim, "Model Selection in Analyzing Spatial Groups in Regression Analysis." *Geographical Analysis,* October 1980, 392–98.

5. Carr-Hill, R. A. and N. H. Stern, "An Econometric Model of the Supply and Control of Recorded Offenses in England and Wales." *Journal of Public Economics,* November 1973, 289–318.

6. Chapman, J. I., W. Z. Hirsch, and S. Sonenblum, "Crime Prevention, the Police Production Function, and Budgeting." *Public Finance,* June 1975, 197–215.

7. Chapman, J. I., "An Economic Model of Crime and Police: Some Empirical Results." *Journal of Research in Crime and Delinquency,* January 1976, 48–63.

8. Fabrikant, Richard, "Interjurisdictional Spillovers of Urban Police Services: Comment." *Southern Economic Journal,* January 1980, 955–61.

9. Goldberg, I. *"True" Crime Rates: The Determance Hypothesis Revised.* Palo Alto: Center for the Econometric Studies of Crime and the Criminal Justice System, Hoover Institute, Stanford University, 1977. (Technical Report EMCRD-2-77.)

10. Green, R. D. and J. P. Doll, "Dummy Variables and Seasonality—A Curio." *American Statistician,* May 1974, 60–62.

11. Greenwood, M. J. and W. J. Wadycki, "Crime Rates and Public Expenditure for Police Protection: Their Interaction." *Review of Social Economy,* October 1973, pp. 138–51.

12. Haitovsky, Y., "Multicollinearity in Regression Analysis: A Comment." *The Review of Economics and Statistics,* November 1969, 486–89.

13. Hakim, S. "Spatial and Temporal Relationships of Property Crime, Police Expenditure and Suburban Development." Ph.D. dissertation, University of Pennsylvania, 1976.

14. ———, et al., "Interjurisdictional Spillover of Crime and Police Expenditure." *Land Economics,* May 1979, 200–212.

15. Hanson, M. A., "Inequality Constrained Maximum Likelihood Estimation." *Annals of the Institute of Mathematical Statistics,* September 1965, 311–22.

16. Hirsch, W. Z. *Urban Economic Analysis.* New York: McGraw-Hill, 1973.

17. Judge, G. G. and T. Takayama, "Inequality Restrictions in Regression Analysis." *Journal of the American Statistical Association,* March 1966, 161–81.

18. Kakalik, J. S. and S. Wilkhorn. *Aids to Decisionmaking in Police Patrol.* Santa Monica: RAND Corp., February 1971. (Report R-593-HUD/RC.)

19. Klemm, R. J. and V. A. Sposito, "Closed Form Solutions for Least Squares when the Parameters are Subject to Inequality Constraints," American Statistical Association, *1976 Proceedings of the Business and Economic Statistics Section,* 1976, 392–96.

20. Liew, C. K., "Inequality Constrained Least Squares Estimation." *Journal of the American Statistical Association,* September 1976, 746 (A).

21. ———, "A Two Stage Least Squares Estimation with Inequality Restrictions on Parameters." *Review of Economics and Statistics,* May 1976, 234 (B).

22. Lindahl, E. "Just Taxation: A Positive Solution," in *Classics in the Theory of Public Finance,* edited by R. Musgrave and A. Peacock. New York: Macmillan, 1958.

23. Lovell, M. C. and E. Prescott, "Multiple Regression with Inequality Constraints: Pretesting Bias, Hypothesis Testing and Efficiency." *Journal of the American Statistical Association,* June 1970, 913–25.

24. Malinvaud, E., "A Planning Approach to the Public Good Problem." *Swedish Journal of Economics,* January 1971, 96–112.

25. McPheters, L. R. and W. B. Stronge, "Law Enforcement Expenditures and Urban Crime." *National Tax Journal,* December 1974, 631–44.

26. Ostrom, E., "On the Meaning and Measurement of Output and Efficiency in the Provision of Urban Police Services." *Journal of Criminal Justice,* March 1973, 93–111.

27. Popp, D. O. and F. D. Sebold, "Quasi Returns-to-Scale in the Provision of Police Service." *Public Finance,* March 1972, 46–61.

28. Searle, R. L. *Linear Models.* New York: John Wiley & Sons, 1971, 355–58.

29. Thaler, R., "An Econometric Analysis of Property Crime." *Journal of Public Economics,* March 1977, 37–51.

30. Tiebout, C., "A Pure theory of Local Expenditures." *Journal of Political Economy,* October 1956, 416–24.

31. Walzer, N., "Economies of scale and Municipal Police Services: The Illinois Experience." *Review of Economics and Statistics,* November 1972, 431–38.

32. Wardle, W. D., "Simple Linear Regression with Inequality Constraints on the Parameters: A Bayesian Approach." *Communications and Statistics,* June 1974, 535–547.