Journal of Public Economics 2 (1973) 289–318. North-Holland Publishing Company

# AN ECONOMETRIC MODEL OF THE SUPPLY AND CONTROL OF RECORDED OFFENCES IN ENGLAND AND WALES *

R.A. CARR-HILL
*University of Sussex*
and
N.H. STERN
*University of Oxford*

First version received January 1973, revised version received July 1973

## 1. Introduction

The purpose of this paper is to develop, test and interpret a simultaneous equations model of the processes generating criminal statistics in England and Wales. Specifically we are interested in the determination of the offence rate, the proportion of offences 'cleared-up'[1] and the number of policemen per capita. Our data refer to individual police districts in England and Wales in 1961 and 1966.

We give a brief description of the main components of our model so that we can introduce the theories and estimation problems involved; a more precise description of the variables we used is contained in sect. 2. Our first equation refers to the number of offences per capita (offence rate) in a district. Deterrence theories indicate that this offence rate depends on the proportion of crimes 'cleared-up' (or clear-up rate), if

* R.A.C.-H. is a criminologist and N.H.S. an economist. The partnership is equal as is the responsibility for errors. Special thanks are due to Art Goldberger, Al Klevorick, Graham Mizon and Nigel Walker. Helpful comments were received from David Downes, John Flemming, Jerry Hausman, David Hendry, Gordon Wasserman and seminars at Cambridge, Oxford, the LSE and the State University of New York at Albany. The helpful comments of the discussant, B.S. van der Laan, when an early version was presented to the 1971 Barcelona Meeting of the Econometric Society are also appreciated. Cliff Wymer kindly made the SIMUL programme available and the authors are grateful for the computational advice of Clive Payne.

The subject matter of this paper will form part of a book 'Analysis of Criminal Statistics' to be published by Seminar Press in 1974.

[1] An offence is cleared-up if the police is convinced that the identity of the offender is clear. This is not synonymous with a conviction or admission in court since there may be some barrier to formal proceedings. A complete definition of a 'clear-up' is in the data appendix available from the authors. In 1966 the average clear-up rate for England and Wales was 42%.

this reflects perceived probabilities of apprehension. Such theories might also focus on the number of policemen per capita and a measure of the equipment available to each officer. One would suppose that a theory which grants that a potential offender takes account of the possibility of apprehension, would also include, as determinants of the offence rate, the possible severity of punishment consequent on apprehension and the potential gains from the offence, i.e. possible 'swag'. Wilkins (1964) has pointed to the empirical importance of swag considerations in his study of car thefts. Becker (1968) has offered an expected gains theory of offences.

The major concern of criminologists, however, has been the differential offence rates between various social groups and classes. The most frequently mentioned are the lower socio-economic groups and the young. Our first equation therefore includes as explanatory variables, in addition to measures of the five factors already outlined, social indicators intended to capture the class and age structure of the population of the police district. Our procedure tests such theories if it can be supposed that most offences are committed locally.

Our second equation models the determination of the clear-up rate. The success rate at solving problems usually depends inter alia, on the number of problems, the number of solvers and the facilities at their disposal. We therefore included in the second equation, the offence rate, the number of policemen per capita and a measure of the equipment available to each officer.[2] Since some types of offence, particularly those against the person, are easier to solve than others we included a measure of crime-mix in this equation (based on the proportion of violent offences).[3] We attempted to capture scale effects by using a measure of the size of the police districts. Finally a social class indicator was used in this equation also, to capture possible differential immunity to suspicion of different groups.

The third equation was originally conceptualised as a model of the process by which the number of police per capita is established. This is partly central and partly local. We included the offence rate, the clear-up rate and the measure of crime-mix as variables intended to capture some of the pressures contributing to central and local feeling concerning the state of social control. Our measure of social class in this equa-

[2] Such an equation is also suggested in the work of Becker (1968).
[3] We suppose that we can take this as an exogenous variable for our system.

tion was based on the middle classes since, we felt, this group was more likely to exert pressure for the protection of property. We also included population density since the Home Office told us the ground to be covered was an important element in their considerations.

We have had to ignore, in the formal modelling, geographical interactions of the kind of which the following might be an example: an area with a high level of swag has a low offence rate because offenders are attracted to adjacent areas with still higher swag levels. This omission is less serious if, in our example, each area has a similar spread of swag opportunities in adjacent areas or if most offences are local in nature (as may be the case with many minor larcenies).

Our theories involved in the first equation usually refer to the 'real' number of offences. We conform to econometric terminology in calling this the 'true' number of offences although a more precise definition is the actual number of incidents which we suppose would have been regarded as offences if they had been reported. The processes captured in the second and third equations refer to the recorded offence rate. Hereafter we shall distinguish the recorded number of offences and the true number of offences.[4]

We have just described three equations of a model which can be written formally as

$$y^* = \alpha_1 p^* + \alpha_2 c + \sum_{i=3}^{7} \alpha_i x_i^1 + \alpha_0 + \epsilon_1 \tag{1}$$

$$p = \beta_1 y + \beta_2 c + \sum_{i=3}^{7} \beta_i x_i^2 + \beta_0 + \epsilon_2 \tag{2}$$

$$c = \gamma_1 y + \gamma_2 p + \sum_{i=3}^{5} \gamma_i x_i^3 + \gamma_0 + \epsilon_3 \, , \tag{3}$$

where $y$, $p$, $c$ are the logarithms of the offence rate, clear-up rate and number of policement per capita, asterisks denote true values, and $\alpha$, $\beta$, $\gamma$s are coefficients and $\epsilon$s are random. It is assumed (see sect. 3 for discussion) that the system is linear in the logarithms of the variables and, otherwise undefined, lower (upper) case letters will refer to the logarithm (exponential) of the corresponding upper (lower) case letters. $x_i^j$ is a pre-determined (exogenous) variable and is the $i$th explanatory

[4] In the first equation we suppose that individuals become aware of the 'true' probability of detection. There is some evidence for this – see Carr-Hill and Stern (1971).

variable in the $j$th equation. The data used to represent the variables described above will be specified in sect. 2.

We have so far described three equations of a system with five endogenous variables, $y$, $y^*$, $p$, $p^*$, $c$. We know in addition that $p^* + y^* \equiv p + y$ since both sides are equal to the logarithm of the number of clear-ups.[5]

Our last equation relates $y$ and $y^*$. This last equation is specified in more detail in sect. 3 when we discuss the econometric problems associated with our model and, in particular, the difficulties raised by the unobservable variables $p^*$ and $y^*$. Sect. 3 also contains a discussion of the econometric techniques appropriate for the model together with tests for correct specification, with special attention to a test for a possible structural difference between 1961 and 1966.

We have been, so far, cavalier with the sociological and criminological assumptions involved in setting up the model. Our excuse is that we have discussed these assumptions, and their place (together with the results) in the light of received theory and empirical work, elsewhere (see Carr-Hill and Stern (1971) and a forthcoming book).[6] In this paper we wish to concentrate on the econometric techniques and problems. We should, however, be clear as to the spirit in which this work is offered.

We are operating on a more aggregated scale as regards offences and populations than is usual in criminological research.[7] We are also using data to measure some of our variables that have been produced by police authorities who have a direct interest in many of the inferences which may be based on the data. Rather than detracting from the value of our exercise we feel that these aspects are central to the interest of it. The use of aggregate data can indicate which of the prima facie plausible microrelations actually appear as important at the 'macro' level. Macro indicators can also help in the selection of regions for micro study. Secondly, the way in which police–public relations and police practices affect the data will be central to the way in which we interpret our results and we hope that the results and explanations can increase our understanding of these processes.

[5] One reason why we have used the logarithmic form is to keep this equation linear – see sect. 3.

[6] See first footnote of paper.

[7] There are exceptions, of course. For example Willmer (1968). Also discussions of murder often focus on just one explanatory variable, the death penalty, and thus treat populations in a highly aggregated way.

The approach of the analysis is somewhat different from most criminological studies in another respect. The structural equations describe how the endogenous variables (e.g. offences) are generated by the exogenous. In other words we describe populations and then examine the number of (e.g.) offences that emerge rather than the criminologically more common practice of analysing and describing offences and convicted offenders.

Our work is therefore offered in the following spirit. The approach to modelling and statistical techniques is fairly new to criminology (although not to other subjects). In order to put the problem in our form and interpret the results we have had to make assumptions about relations which seem plausible at the current state of knowledge but which have not yet been thoroughly established. In economics, at least, it is often the case that the assumptions for, and interpretations of the results from, macro-models can motivate micro-research as well as being based on previous micro-research. In a similar fashion we hope that our macro analysis of the criminal statistics will lead to and be tested by further research at the micro level.

## 2. The variables

We describe, in this section, the measures we used to capture the concepts outlined in the introduction. We shall be brief, and a precise and complete description of data sources is available on application to the authors.

The data are for a cross-section of police districts in 1961 and 1966. These years were chosen since they were census and sample[8] census years, respectively, so that data on social class and age structure are available. Our other main sources, apart from published census data, were the Home Office publications 'Supplementary Statistics Relating to Crime and Criminal Proceedings' (Supp. Stats.) for 1961 and 1966 and the annual publication 'Police Force Statistics' (PFS) put out by the local authority treasurers.

We had two main sets of data: for 64 urban police[9] districts in 1961

[8] We suppose that the errors involved in the census of 1966 being only a 10% sample are small compared with errors elsewhere in the model.

[9] We were assisted with the 1961 data by a report prepared by W.F. Greenhalgh (1966) and kindly made available by the Home Office.

Table 1
Included exogenous variables in model for 1961 urban and 1966 urban.

| | Variable | Definition | Source[a] |
|---|---|---|---|
| Equation 1 | | | |
| | *F* | The proportion of convicted given custodial treatment[b] | Supp. Stats. |
| | *A* | Proportion of young (15–24) males in population | Census |
| | *S* | Proportion of population that is working-class, specifically socio-economic groups 7, 10, 11, 15 | Census |
| | *T* | Total rateable value per area[c] | Rates and Rateable Values[d] |
| | *E* | Total police expenditure per officer | PFS |
| Equation 2 | | | |
| | *A* | See Eq. 1 | |
| | *S* | See Eq. 1 | |
| | *N* | Population | Census 1961 PFS 1966 |
| | *E* | See Eq. 1 | |
| | *V* | Proportion of offences that are violent[b] | Supp. Stats. |
| Equation 3 | | | |
| | *M* | Proportion of population that is middle-class, specifically socio-economic groups 5, 6, 8, 9, 12, 14 | Census |
| | *V* | See Eq. 2 | |
| | *D* | Population density | PFS |

[a] See beginning of sect. 2 for definition of these sources.
[b] Custodial treatment and violent offences defined in data appendix available from authors. We should have liked to incorporate in *F* data on the length of offences but these proved inadequate.
[c] The base on which local property taxes are levied.
[d] Published by HMSO for Ministry of Housing and Local Government.
Note: The natural logarithms of the above variables were used in the estimation and these logarithms are denoted by the corresponding lower case letters.

and for 66 urban police districts in 1966.[10] Two further data sets were obtained by pooling the 44 rural districts with the urban districts in 1966 and pooling the 1961 and 1966 urban sets. We call these four sets

[10] These were all the urban police districts in England and Wales except for the Metropolitan Police District (MPD) and some adjacent authorities. The MPD and these authorities were omitted since the overlap with several local authority districts made census data difficult to ascribe.

respectively 1961 urban, 1966 urban, 1966 urban and rural pooled, and 1961 and 1966 urban pooled.

The system that we eventually estimated was a three equation partial reduced form of the five equation model described in sect. 1 where the unobservable variables $y^*$ and $p^*$ were eliminated from equation (1), and (2, and (3) retain their form (see sect. 3). These equations were for the recorded offence rate, recorded clear-up rate and the number of policemen per capita. The number of recorded offences was the total number of recorded indictable offences and was taken from Supp. Stats. Indictable offences are those offences deemed to be serious enough to warrant the posibility of trial in front of a jury. Roughly 65% of recorded indictable offences are larcenies (and similar offences), 20% are breaking-and-entering type offences and 7 or 8% are offences against the person. District clear-up rates are unpublished and were kindly made available by the Home Office. The number of policemen per capita was taken from PFS.

In table 1 above we state the exogenous variables that were used in each of the three equations for the main data sets: 1961 urban and 1966 urban.

For the 1966 urban and rural pooled set the model was modified as follows. The percentage of the area that was urbanised was included in each equation which acts, in part, like a dummy variable distinguishing town and country. The natural logarithm was used and the source was PFS. In the first equation this variable replaced $T$ and in the third equation $D$ since it was fairly highly correlated with both.

## 3. Econometric techniques and problems

### *3.1. Unobservable variables, identification, and the estimated model*[11]

The five equation model outlined in sect. 1 contains the unobservable variables $y^*$ and $p^*$. We now describe the procedure we used to meet this problem. Our model is

[11] This section owes much to the comments of Art Goldberger. General references for terms, theorems and techniques unfamiliar to the reader are Johnston (1971) and Malinvaud (1971).

$$y^* = \alpha_1 p^* + \alpha_2 c + \sum_{i=3}^{7} \alpha_i x_i^1 + \alpha_0 + \epsilon_1 \tag{1}$$

$$p = \beta_1 y + \beta_2 c + \sum_{i=3}^{7} \beta_i x_i^2 + \beta_0 + \epsilon_2 \tag{2}$$

$$c = \gamma_1 y + \gamma_2 p + \sum_{i=3}^{5} \gamma_i x_i^3 + \gamma_0 + \epsilon_3 \tag{3}$$

$$p^* + y^* \equiv p + y \tag{4}$$

$$y = y^* + k(\;) + u \;. \tag{5}$$

We have described eqs. (1)–(4) in sect. 1 and table 1. Eq. (5) states that the number of reported offences is a fraction of the number of true offences where this fraction has logarithm $k(\;) + u$ where $k$ is a function of as yet unspecified variables and $u$ is a random term mean zero. We may as well assume that $k$ is linear in the logarithms of its arguments. The problem, of course, is to know what the arguments of the $k(\;)$ function are. Let us for the moment, however, remain agnostic, and substitute from eqs. (4) and (5) into (1). We have

$$y = \alpha_1 p + \alpha_2 c + \sum_{i=3}^{7} \alpha_i x_i^1 + \alpha_0 + (1+\alpha_1)\, k(\;) + (1+\alpha_1)\, u + \epsilon_1 \;. \tag{1$'$}$$

We now see that if any of the variables on the r.h.s. of (1)′ are among the arguments of $k(\;)$ then the corresponding parameter $\alpha_i$ is unidentified as a parameter of (1). For example, suppose $k$ depends on $c$ (and it is reasonable to suppose that it does) we write $k(\;) = \eta\, c + k_1(\;)$ and (1)′ becomes

$$y = \alpha_1 p + (\alpha_2 + (1+\alpha_1)\,\eta)\, c + \sum_{i=3}^{7} \alpha_i x_i^1 + \alpha_0 + (1+\alpha_1)\, k_1(\;)$$
$$+ (1+\alpha_1)\, u + \epsilon_1 \;. \tag{1$''$}$$

Suppose we now assume $k_1$ is constant. We can estimate the system (1)″, (2) and (3) since it is a 3-equation system in 3 endogenous variables $y$, $p$, $c$. It is easy to check that the system (1)″, (2), (3) with exogenous variables as in table 1 satisfies the rank condition for identification.[12] The problem is that we estimate $\alpha_1$ and $(\alpha_2 + (1 + \alpha_1)\eta)$ and cannot isolate $\alpha_2$ and $\eta$.

The estimation procedure we, in fact, used was to include as explanatory variables in the first equation $p$, $c$ and the five exogenous variables of table 1. In other words we have

$$y = \alpha_1' p + \alpha_2' c + \sum_{i=3}^{7} \alpha_i' x_i^1 + \alpha_0' + \epsilon_1 \quad . \tag{1'''}$$

The system estimated was (1)‴, (2) and (3). We thus have a three equation system with three endogenous variables $y$, $p$, $c$ and nine exogenous variables. There are 22 coefficients (including 3 constant terms) and 6 distinct entries in the variance–covariance matrix. The rank condition for identification is satisfied.

We can assert that $\alpha_i'$ is a parameter of equation (1) only if we are convinced that the corresponding variable is not an argument of $k(\ )$. If we are not so convinced then we should interpret the parameter as a product of two processes: the first determining the 'true' level of offences, the second determining the number of these offences which reach the record book. We shall make liberal use of such interpretations in sect. 4. It should be noted that if $y$ is an argument of $k(\ )$ then none of the $\alpha'_i$ will be equal to $\alpha_i$, and that we have assumed that $k(\ )$ does not depend on any of the exogenous variables that were excluded from eq. (1) (see table 1).

We have used the logarithmic form partly since we want to keep (4) linear and partly since it is reasonable to discuss the relationships in terms of the effect of proportional changes in the 'right-hand side' variables on those of the left-hand side. Further we tried two alternative specifications of the functional form – reported in an earlier version of this paper presented to the 1971 Econometric Society Meeting in Barcelona. One of the alternatives was to use $\log P/1-P$ for $\log P$, for those variables that are proportions. The other was to use all variables in their linear (unlogged form). Neither exhibited superiority over the logarith-

[12] It is easy to check the order condition – $n,v,m,d$ are excluded from the first equation, $f,t,m,d$ are excluded from the second, and $a,s,t,f,n,e$ from the third. The order condition is only necessary. The rank condition, necessary and sufficient for identification, is also clearly satisfied since each equation has an exogenous variable unique to it.

mic form, so that our version performs as well as the others and has the linearity advantages described.

It is natural to ask whether there is a way of eliminating $p^*$ and $y^*$ from the system that does not have the problems associated with eq. (5). It is commonly held[13] that the reporting rate for breaking-and-entering offences is very high (mainly for insurance reasons). Suppose we add to our system the following equations

$$y^*_{BE} = y_{BE} \tag{6}$$

$$y^* = y^*_{BE} + b + v \tag{7}$$

where $y^*_{BE}$ is the logarithm of the true number of breaking-and-entering offences, $y_{BE}$ is the logarithm of the reported number, $b$ is a constant and $v$ a random term mean zero. So we assume all breaking-and-entering offences are reported and that true breaking-and-entering offences are a fraction (with a random component) of total indictable offences. Suppose also that we decompose (4) into

(4a) $p \equiv g - y$ and (4b) $p^* \equiv g - y^*$,

where $g$ is the logarithm of the number of clear-ups. Then the system (1), (2), (3), (4a), (4b), (5), (6), (7) is an eight equation system in the 8 endogenous variables $y$, $y^*$, $y_{BE}$, $y^*_{BE}$, $p$, $p^*$, $c$, $g$. We can form the partial reduced form

$$(1+\alpha_1)y_{BE} = \alpha_1 g + \alpha_2 c + \sum_{i=3}^{7} \alpha_i x_i^1 + \alpha_0 - (1+\alpha_1)b - (1+\alpha_1)v + \epsilon_1 \tag{8}$$

$$g = (1+\beta_1)y + \beta_2 c + \sum_{i=3}^{7} \beta_i x_i^2 + \beta_0 + \epsilon_2 \tag{9}$$

$$c = (\gamma_1 + \gamma_2)y + \gamma_2 g + \sum_{i=3}^{5} \gamma_i x_i^3 + \gamma_0 + \epsilon_3 \tag{10}$$

$$y = y_{BE} + b + k(\,) + u + v \tag{11}$$

[13] See Willmer (1968).

which is four equations in four endogenous variables $y$, $y_{BE}$, $g$, and $c$. Suppose we further assume that $k(\ )$ has no arguments apart from those variables included in equations (8)–(11). Then a two stage least squares estimate of (8) would enable us to identify $\alpha_1$. All this depends however on our accepting eqs. (6) and (7). Early attempts[14] with a procedure using breaking-and-entering offences lead us to believe that either the reporting rate for breaking-and-entering is not sufficiently high or that crime patterns are not sufficiently fixed for the purposes of the above procedure. The view that reporting rates for breaking-and-entering offences may be low is corroborated, at least for the USA, by the research of the National Opinion Research Centre (Chicago) who found that only 31% of the burglaries (similar to our breaking-and-entering offences) of their sample were reported – see p. 8 of Ennis (1967).

### 3.2. The estimation technique

The results given in sect. 4 are full information maximum likelihood (FIML) estimates.[15] The computing programme used was SIMUL 7 due to Cliff Wymer of the LSE and kindly made available by him. Readers unfamiliar with the simultaneous equations techniques of econometrics will find good descriptions in Malinvaud (1971) or Johnston (1971).

FIML estimates are thought to be sensitive to complete model specification, see Malinvaud (1971) or Johnston (1971). (Clearly, the specification of a particular equation of interest is important to that equation, using any technique.) The SIMUL programme also gives two stage least square estimates (2SLS) which are known to be more robust to specification error elsewhere in the model and it is therefore reassuring that 2SLS and FIML results for our models are similar (see the first draft of this paper presented to the 1971 Barcelona Econometric Society Meeting). We use FIML since it is efficient and, since the programme gives us likelihood values, it enables us to perform likelihood ratio tests rather easily,[16] and to carry out tests on the a priori restrictions as a whole – see next sub-section.

[14] See the earlier version of this paper presented to the 1971 Econometric Society Meeting in Barcelona.

[15] We make the usual assumption that $(\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t})$ is $N(0, \Sigma)$ and $COV(\epsilon_{it}, \epsilon_{it'}) = 0$, $t \neq t'$ where $\epsilon_{it}$ is the error term in the $t$th observation on the $i$th equation.

[16] In principle likelihood ratio tests can be constructed for limited information maximum likelihood and 2SLS. This would have required a good deal of extra programming.

### 3.3. Tests of model specification and for structural change

For each of the results given in sect. 4 we give the 'chi-square value of the log-likelihood ratio' and its corresponding degrees of freedom. The null hypothesis for this chi-square test is that the overidentifying restrictions (taken as a group) involved in the particular model are correct and this is being tested against the hypothesis that an exactly identified model is correct (all exactly identified models have the same maximum likelihood i.e. that derived from reduced form estimation).

We have specified our model in sect. 3.2 and we make the assumption that the errors for each observation are distributed $N(0,\Sigma)$ and that errors for different observations are uncorrelated (see footnote 15). Suppose $L_u$ is the maximum of the joint probability distribution of the observations on the endogenous variables $Y$ when $Y$ take the observed values, the exogenous variables $X$ are given, when the coefficients $\theta$ vary in some linear subspace $\Lambda$ and the variance–covariance matrix, $\Sigma$, is unrestricted. Suppose now we put $m$ linear restrictions on $\Lambda$ and compute the new maximum value $L_r$. It is well-known that $2 \log L_u/L_r$ is asymptotically chi-square with $m$ degrees of freedom.[17] The 'chi-square value of the log-likelihood ratio' is $2 \log L_u/L_r$ and we can compare this statistic with the entry at the appropriate significance level in the chi-square tables.

Our test for a structural break between 1961 and 1966 is based on similar reasoning. This test is due to Grayham Mizon and is described and justified in detail in his paper (1972). The test is based on the maximised log-likelihood values from the data sets 1961 and 1966 urban pooled, 1961 urban and 1966 urban ($l_3$, $l_1$, $l_2$ respectively). The log-likelihood value for the model (pooling the 1961 urban and 1966 urban data) where we do not restrict the parameters ($\theta$ and $\Sigma$) in 1961 and 1966 to be the same is $l_1 + l_2$. The log-likelihood value for the model (1961 and 1966 urban pooled) where we do restrict the parameters to be the same is $l_3$. Thus $2(l_1 + l_2 - l_3)$ is asymptotically chi-square on the null hypothesis of no structural change. The number of degrees of freedom is equal to the extra number of restrictions in the more constrained model, which equals the number of coefficients plus the number of distinct entries in the variance–covariance matrix. In

[17] An early description of the test is in Hood and Koopmans (1952) p. 178, and versions of the stated property of the likelihood ratio can be found in text books on multivariate analysis.

this case – see model of sect. 3.1 – we have 22 coefficients (counting constant terms) and 6 distinct entries in the variance–covariance matrix. We thus have 28 degrees of freedom.

## 4. Results and interpretations

### 4.1. Testing the model

The chi-square value of the log-likelihood ratio in the test for a structural break between 1961 and 1966 was 79.3 with 28 degrees of freedom. This leads us to reject the null hypothesis of no structural change between 1961 and 1966 even at the 0.1% level. The estimates of the 22-coefficient model, described in sects. 1 and 3.1, for the four data sets (see sect. 2) are given in the appendix to this paper, together with the likelihood values.

Table 2 (i)

**1966 Urban and rural pooled restricted**

| Variable to be explained | Explanatory variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *y* | *p* | *c* | *f* | *a* | *s* | %urbn. | *e* | const. |
| | −0.59 | +0.74 | −0.17 | +0.63 | +0.11 | +0.45 | +0.40 | −1.64 |
| | (0.24) | (0.22) | (0.09) | (0.15) | (0.13) | (0.09) | (0.14) | (2.28) |
| | 2.50 | 3.41 | 1.86 | 4.29 | 0.88 | 4.92 | 2.81 | 0.72 |
| *p* | *c* | *n* | %urbn. | const. | | | | |
| | −1.15 | −0.16 | +0.31 | −8.82 | | | | |
| | (0.34) | (0.04) | (0.11) | (2.62) | | | | |
| | 3.39 | 4.31 | 2.76 | 3.37 | | | | |
| *c* | *p* | *m* | *v* | %urbn. | const. | | | |
| | +1.22 | −0.86 | −0.44 | +0.40 | −3.98 | | | |
| | (0.41) | (0.39) | (0.11) | (0.08) | (2.18) | | | |
| | 2.98 | 2.23 | 3.88 | 5.22 | 1.82 | | | |

| Covariance matrix | | | Variance | | | Chi-square value of log-likelihood ratio 17.057 with 10 d.o.f. |
|---|---|---|---|---|---|---|
| 0.039 | | | | *y* | 0.175 | |
| 0.001 | 0.040 | | | *p* | 0.028 | |
| −0.005 | −0.007 | 0.053 | | *c* | 0.042 | |

Table 2 (i)

1966 Urban restricted

| Variable to be explained | Explanatory variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $y$ | $p$ | $c$ | $f$ | $a$ | $s$ | $t$ | $e$ | const. |
| | −0.26 | +0.44 | −0.17 | +0.45 | +0.31 | +0.16 | +0.44 | −2.67 |
| | (0.32) | (0.62) | (0.10) | (0.17) | (0.16) | (0.06) | (0.15) | (4.62) |
| | 0.81 | 0.71 | 1.72 | 2.63 | 1.97 | 2.70 | 2.94 | 0.58 |
| $p$ | $c$ | $s$ | $n$ | const. | | | | |
| | −1.54 | +0.21 | −0.12 | −10.62 | | | | |
| | (0.52) | (0.12) | (0.04) | (3.44) | | | | |
| | 2.96 | 1.70 | 2.91 | 3.09 | | | | |
| $c$ | $p$ | $m$ | $v$ | $d$ | const. | | | |
| | −0.00 | −0.86 | −0.12 | +0.03 | −1.27 | | | |
| | (0.17) | (0.27) | (0.05) | (0.03) | (1.66) | | | |
| | 0.03 | 3.16 | 2.42 | 1.01 | 0.77 | | | |

| Covariance matrix of residuals | | | Variance | | Chi-square value of log-likelihood ratio 17.769 with 13 d.o.f. |
|---|---|---|---|---|---|
| 0.039 | | | $y$ | 0.173 | |
| 0.010 | 0.064 | | $p$ | 0.038 | |
| 0.008 | 0.024 | 0.014 | $c$ | 0.018 | |

Having seen that it is reasonable to suppose that there was a structural break between 1961 and 1966 we looked for the best model in each year. This was a fairly short search with no rigid objective function. We had three main guides however. First, our views on those variables that were likely to be insignificant in each year given the various organisational/social changes between 1961 and 1966. Secondly, the significance level of the chi-square value in the test of all the over-identifying restrictions in the model and finally, the significance levels of the coefficients. We give the results of this search in table 2(i) for the three data sets 1961 urban, 1966 urban, 1966 urban and rural pooled with the added title 'restrictions'. Comparisons between the more restricted and less restricted (22-coefficient) models, incorporating comments on the search just described are contained in the Appendix.

For each of the 1961 urban and 1966 urban data sets the null hypothesis that the over-identifying restrictions for the models of table 2 are correct is accepted at the 10% level (by a comfortable margin). The estimated variance–covariance matrix of residuals is discussed at the end of sect. 4.2.

Table 2 (i)

1961 Urban restricted

| Variable to be explained | Explanatory variables | | | | | |
|---|---|---|---|---|---|---|
| $y$ | $p$ | $f$ | $a$ | $t$ | const. | |
| | −0.66 | −0.28 | −0.06 | +0.18 | +3.61 | |
| | (0.26) | (0.08) | (0.19) | (0.05) | (1.80) | |
| | 2.55 | 3.41 | 0.30 | 3.65 | 2.00 | |
| $p$ | $c$ | $a$ | $s$ | $n$ | $v$ | const. |
| | −1.01 | −0.36 | +0.24 | −0.12 | +0.12 | +5.67 |
| | (0.37) | (0.10) | (0.10) | (0.02) | (0.07) | (0.66) |
| | 2.68 | 3.80 | 2.51 | 4.81 | 1.65 | 8.58 |
| $c$ | $p$ | $m$ | $v$ | const. | | |
| | +0.03 | −0.92 | −0.06 | +6.10 | | |
| | (0.12) | (0.21) | (0.15) | (1.28) | | |
| | 0.28 | 4.38 | 1.19 | 4.76 | | |

| Covariance matrix of residuals | | | Variance | | Chi-square value of log-likelihood ratio 8.834 with 9 d.o.f. |
|---|---|---|---|---|---|
| 0.041 | | | $y$ | 0.075 | |
| 0.005 | 0.020 | | $p$ | 0.028 | |
| 0.009 | 0.009 | 0.010 | $c$ | 0.013 | |

Notes (i) Number in brackets below coefficient is the asymptotic standard error. Below this is the $T$-value.
(ii) Variables defined in table 1.
(iii) Chi-square values discussed in sects. 3.3 and 4.1.
(iv) Programme SIMUL 7 by Cliff Wymer run on Harwell Atlas.
(v) The constant terms are affected by the scaling of variables for computational reasons – the variables in the results throughout table 2(i) are all in the same units.

In table 2(ii) we give the reduced forms[18] derived from the structures 1961 urban restricted and 1966 urban restricted of table 2(i). The coefficient of an explanatory variable in (e.g.) the first equation of a model in table 2(ii) tells us the percentage change in $y$ that would be predicted by the model for a one per cent in the corresponding (unlogged) exogenous variable (holding other exogenous variables constant) *taking into account* the changes that would come about in $p$ and $c$ also. These coefficients are sometimes called 'impact multipliers'.

[18] See Malinvaud (1971) and Johnston (1971) for a discussion of reduced forms.

Table 2(ii)
Derived reduced forms.

**1961 Urban restricted**

| Endogenous variable | Exogenous variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *y* | *f* | *a* | *s* | | *n* | *v* | *m* | const. |
| | −0.28 | 0.18 | −0.15 | 0.18 | 0.07 | −0.59 | −0.12 | 3.90 |
| | (0.08) | (0.17) | (0.09) | (0.05) | (0.03) | (0.30) | (0.06) | (2.39) |
| | 3.41 | 1.02 | 1.79 | 3.65 | 2.23 | 1.98 | 1.96 | 1.63 |
| *p* | 0.00 | −0.35 | 0.23 | 0.00 | −0.11 | 0.89 | 0.17 | −0.45 |
| | (0.00) | (0.09) | (0.09) | (0.00) | (0.02) | (0.26) | (0.05) | (1.96) |
| | – | 4.12 | 2.64 | - | 5.70 | 3.49 | 3.43 | 0.23 |
| *c* | 0.00 | −0.01 | 0.01 | 0.00 | −0.00 | −0.89 | −0.06 | 6.09 |
| | (0.00) | (0.04) | (0.03) | (0.00) | (0.01) | (0.21) | (0.05) | (1.29) |
| | – | 0.28 | 0.28 | - | 0.28 | 4.30 | 1.25 | 4.71 |

Covariance matrix of residuals

| | | |
|---|---|---|
| 0.052 | | |
| −0.012 | 0.011 | |
| 0.009 | −0.001 | 0.010 |

| Variance | |
|---|---|
| *y* | 0.075 |
| *p* | 0.028 |
| *c* | 0.013 |

Table 2(ii)
Derived reduced forms.

**1966 Urban restricted**

| | Exogenous variables | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *y* | *f* | *a* | *s* | *t* | *n* | *v* | *e* | *m* | *d* | const. |
| | −0.17 | 0.45 | 0.26 | 0.16 | 0.03 | −0.10 | 0.44 | −0.73 | 0.02 | −0.95 |
| | (0.10) | (0.17) | (0.16) | (0.06) | (0.04) | (0.06) | (0.15) | (0.45) | (0.02) | (3.38) |
| | 1.72 | 2.63 | 1.61 | 2.70 | 0.74 | 1.59 | 2.94 | 1.64 | 0.98 | 0.28 |
| *p* | 0.00 | 0.00 | 0.21 | 0.00 | −0.12 | 0.18 | 0.00 | 1.34 | −0.04 | −8.73 |
| | (0.00) | (0.00) | (0.11) | (0.00) | (0.03) | (0.05) | (0.00) | (0.36) | (0.04) | (2.39) |
| | – | – | 1.86 | – | 4.14 | 3.58 | – | 4.00 | 1.09 | 3.65 |
| *c* | 0.00 | 0.00 | −0.00 | 0.00 | 0.00 | −0.12 | 0.00 | −0.87 | 0.03 | −1.23 |
| | (0.00) | (0.00) | (0.04) | (0.00) | (0.02) | (0.04) | (0.00) | (0.24) | (0.02) | (1.51) |
| | – | – | 0.03 | – | 0.03 | 3.16 | – | 3.59 | 1.09 | 0.81 |

Covariance matrix of residuals

| | | |
|---|---|---|
| 0.051 | | |
| −0.007 | 0.023 | |
| 0.014 | 0.002 | 0.015 |

| Variance | |
|---|---|
| *y* | 0.173 |
| *p* | 0.038 |
| *c* | 0.018 |

Note (i) See table 2(i).
(ii) A dash for a *T*-value denotes that the corresponding coefficient is identically zero.

### 4.2. Interpretation of structural coefficients

#### 4.2.1. Consistent results

We focus attention on 1961 urban restricted and 1966 urban restricted (see table 2) as our main set of results. We begin with a discussion of those coefficients that were significant[19] and behaved similarly in both years. Our discussion will be fairly brief as a more detailed analysis from the criminological and sociological point of view is available elsewhere, Carr-Hill and Sterns (1971), and in our forthcoming book.

In the first equation the coefficient of (severity of punishment) $f$ was significant and negative in both years.[20] The first equation coefficient of $p$ was also significant and negative in 1961 urban, and 1966 urban and rural pooled. It was also significant and negative in 1966 urban when we used a different measure of the probability of apprehension based on conviction rates.[21] The first equation coefficient of the 'swag' variable, $t$, was significant and positive in both years. All the variables that one would expect to enter a calculation of the expected gain from an offence, therefore, enter into the determination of the reported offence rate in the natural direction.

The elasticity ($e_p$) of the recorded offence rate with respect to the clear-up rate, $p$, was found to be higher than that ($e_f$) for the severity of punishment, $f$, a finding which conforms with common preconception.[22] It should be noted that in so far as percentage changes in our measures of $p$ and $f$ are larger (smaller) than perceived changes in probability of apprehension and severity of punishment, we under- (over-) estimate the elasticities of actual behaviour (assuming $k(\ )$ independent of $p, f$ – see sect. 3.1). There is a large fixed element in any punishment (the social consequences of court appearance) so it may well be the case that percentage changes in our $f$ measure are larger than perceived changes. The magnitude of the $p$, $f$ elasticities has dropped between 1961 and 1966 – in other words the population seems to have become less responsive to probabilities of being caught and the severity of punishment.[23]

[19] By significant we shall mean asymptotic $T$-values $\geqslant 2$.

[20] In 1966 the $T$-value was a little under the prescribed level (see table 2). This $T$-value was larger in runs where a different measure of $p$ was used (see 1971 version of this paper given to the Barcelona Meeting of the Econometric Society).

[21] See the earlier version of this paper presented at the 1971 Econometric Society Meeting in Barcelona.

[22] See e.g. Becker (1968).

[23] It seems unlikely that our measures of $p$, $f$ deviated more from the perceived values in 1961 than 1966.

The estimates of the coefficient of *t* for the two years are close. The constancy of this coefficient lends support to the view that the availability of 'swag' is the main[24] effect being captured here.

In the second equation there was just one coefficient (apart from that of *c*, discussed later) that was both significant and behaved similarly in both years – we found that clear-up rates vary inversely with size of police district and, further, the coefficient was the same in both years. This seems to indicate diseconomies of scale in crime-solving. If this is correct, the amalgamations of police forces of the later 1960s have to be justified rather carefully and vague appeals to economies of scale are not acceptable.

The third equation also has just one coefficient in this category. Perhaps surprisingly, however, we find the number of policemen per capita is inversely related to the proportion of the population that is middle-class (and again the coefficients for the two years are similar). We had thought that the opposite might hold as a result of a larger potential lobby for the defence of property. The explanation may be either that it is difficult to recruit policemen from middle-class areas (assuming police are recruited locally) or that the central authorities believe middle-class areas need less policing. We lean towards the former since most forces are below their 'established strength' and the recruitment hypothesis seems to concord rather better with the behaviour of other coefficients in this equation.[25] For example the proportion of violent offences in total offences is inversely related to the number of policemen (and significant in 1966) – it is hard to reconcile this as an allocation procedure, but it may be easier to recruit policemen where violent crime is low.[26] A similar argument applies to the behaviour of the coefficient of *p* in the third equation for 1966 urban and rural pooled.

We record those coefficients that were not significant in either year:

[24] Recall (beginning sect. 2) that *T* and *D* (population density) are correlated.

[25] Al Klevorick has pointed out to us that we should really have two equations instead of (3), one for supply (recruitment) and the other for demand (established recruiting targets) for policemen. We agree but feel that our identification problems are bad enough without specifying further equations.

[26] A referee has pointed out that we may have a simultaneity problem here too in that many policemen may imply low violent crime.

unemployment[27] and a measure of overcrowding within the home[28] (tried at an early stage) in the first equation, $y$, $v$ and the proportion of detectives in the force (tried at an early stage) in the second equation and $y$ in the third equation.

We remark in the Appendix that the urban and rural data are best treated separately. We note here the main differences in estimated coefficients between the 1966 urban set and the 1966 urban and rural pooled set. First the percentage of area urbanised affects $y$, $p$ and $c$ positively. Only the second of these effects is surprising, and our view is that the first equation coefficient reflects, at least in part, available swag. $p$ performs better with the pooled set in the first and third equations and $s$ worse in the first.

### *4.2.2. Differences between years*

We now discuss the second group of coefficients, i.e., those that behave in a markedly different way in the two years. These are $c$, $e$, $a$ and $s$ in the first two equations. We lean heavily on our interpretation of the model as one of recorded offences in the explanation of the behaviour of these coefficients (see sect. 3.1). We take $c$ first. We distinguish three effects on recorded offences of more policemen in an area. We call these the creating, reporting and preventing effects. The first of these refers to the 'creation' of recorded offences by a policeman in the sense that he sees and records a minor crime that otherwise go unreported. Secondly more policemen in evidence may mean that more members of the public report minor crimes than they otherwise would. An alternative interpretation of this effect is that the presence of police in greater numbers leads to an increased public awareness of legality and illegality. Further it would affect the public's view of whether it was worth the effort of reporting since it would affect an estimate of the likelihood of action being taken on the report. We

[27] We might have done better with unemployment disaggregated according to age or with participation rates – see (e.g.) Phillips et al. (1972). On the other hand Grieson (1972) found that unemployment was unimportant in his cross-section study of juvenile arrests, a parallel study to this based on US data. Unemployment is probably rather better correlated with participation rates in time series than cross-section and thus should be expected to work better in the former. There may also be some age-unemployment interaction effects – see Carr-Hill and Stern (1971).

[28] The variable used was the proportion of households with less than half a person per room.

suspect that, with increased mechanisation and deteriorating police–public relations these last two interpretations will be of increasing importance. Thirdly, the presence of a policeman may deter (even for a given level of $p$). The first two would increase the recorded offence rate and the third reduce it.

These effects also operate on the clear-up rate since they affect offence 'mix' and so are relevant for the $p$-equation. The creating effect would act positively on the clear-up rate since the clear-up rate for crimes recorded in this way must be close to one (or, at least, above average). The reporting effect would act negatively on the clear-up rate since crimes, reported to policemen which would not have been reported had a policeman not been around, are probably either less easy to solve than average or less worth the effort of the police than average.[29] The prevention effect would also act negatively since we assume that crimes deterred by the existence of more policemen are easier to solve than average i.e. with a given offence rate and extra police we expect the offence mix to be more 'professional'. We summarise these effects with the following table:

| | Offence Rate | Clear-up Rate |
|---|---|---|
| Creating Effect | + | + |
| Reporting Effect | + | – |
| Preventing Effect | – | – |

The total effect of more policemen on the offence and clear-up rates is a combination of these effects. The major determinants of the relative strength of these three effects are the state of police–public relations, the mobility of the police force, and the formality of police procedures.

Senior police officers have pointed to the apparent deterioration in relations with the public in the 1960s (see below). They have suggested that this was a consequence of the increased mobility and formalisation of procedures[30] that occurred in this period following the Royal Commission (Willink) report of 1962 and the increase in emphasis on management and methods. Examples of such changes are the institution of the Research and Planning Branch (1963), computerisation of Offenders Index (1963), and Regional Crime Squads (early 1960s).

[29] It should be remembered that many indictable offences are thefts of articles of low value

[30] See the quotes below. It is unclear that the causation is only in the direction described.

We should therefore expect an increase in the importance of the first effect from 1961 to 1966 and a decrease in the importance of the second effect. If the last two effects were the most important in 1961 we should expect a total negative effect on the clear-up rate and an insignificant effect on the level of offences. As the creating effect is relatively stronger in 1966 we should expect a total positive effect for 1966 on the level of offences, but still a negative effect on the clear-up rate. An examination of table 2(i) shows that these expectations are justified (the significance of $c$ in the $y$-equation is improved in the 1966 pooled set). In fact the coefficient of $c$ in the $p$-equation increased in absolute magnitude from 1961 to 1966 and we conclude that the preventing effect on the clear-up rate became stronger between 1961 and 1966. The explanation would seem to be that the increased mobility of the police force meant that those crimes deterred in 1966 had clear-up rates higher above the average for that year than was the case in 1961. In other words one effect of increased mobility and formalisation is the 'weeding out' of easily solved crimes leaving the more difficult ones. The large size of the $c$ coefficient in the second equation seems to indicate that this deterrent effect is strong.

The only coefficient concerning $e$ that was significant was in the first equation in 1966. Effects of increases in $c$ and $e$ are probably similar. Applying the above analysis it seems that we have a combination of the creating and preventing effects involved.[31] In 1961 these effects cancelled each other out. Although both increased in importance, the creating effect was dominant for the offence equation producing the result observed. Again the preventing effect for the $p$-equation seems to have increased in importance from 1961 to 1966 since the preventing and creating effects cancel in the $p$-equation producing an insignificant coefficient. It seems, therefore, that the effects of $c$-increases (more policemen) and $e$-increases (more expenditure per policeman) can be understood in similar ways. The major difference is that with $e$ the reporting effect is lost. In other words, it does seem as some chief constables have suggested, that the problems of loss of contact with the public, and increased formalisation and mechanisation are serious. Indeed, Chief Inspector Brooks stated on television that 'We found that through increased mobilisation and mechanisation that they were

[31] The reporting effect is presumably unimportant for $e$ since the public are less likely to approach a motorised policemen – see quotes below.

getting out of touch with the public...' (Feb. 15th 1971, BBC2, Crimes and the Criminal No. 6.) In the same programme the Chief Constable of Hampshire indicated the increased importance of the recording effect by saying 'The motorised policeman means more criminals are caught and probably sets up a deterrent'.

The results obtained for the '*a*' variable (proportion of males aged 15–24 in the population) in the first two equations can be understood in terms of deteriorating police relations with youth. It may be the case that areas with high '*a*' have higher actual offence rates but also that these will not necessarily be reflected in the recorded statistics.

Suppose the young do commit more offences but the police turn a blind eye to many and pursue less aggressive prosecution and TIC[32] policies for younger offenders. Then there is no reason to suppose that we should observe higher recorded offence rates in youthful areas and we might observe lower clear-up rates. The following – purely illustrative – numerical example clarifies the point. Suppose we have 100 young people and 8 of them commit 2 offences each. Suppose also that only 10 of the 16 offences are reported and 3 young people are charged with one offence each. For this group offences $Y = 10$ and clear-up rate $P = 3/10$. Now take 100 older people and suppose 6 persons commit 2 offences each, 10 of the 12 are reported and 2 people are charged with 2 offences each; $Y = 10$, $P = 4/10$.

We suggest that in 1961 police-practices might have been as just described and this would explain our result of '*a*' being insignificant in the first equation and significant with a negative coefficient in the second. If relations deteriorated with the young in the 1960s we should expect the effects just outlined to weaken and it transpires that in 1966 we find '*a*' significant with positive coefficient in the first equation and insignificant in the second equation.

It should be noted that there is a difference between the observation that young males contribute a higher than average number of convictions per capita (which we observe from the age breakdown of convictions statistics) and the findings here that (in 1966) areas with higher proportions of young males had higher recorded offences. Note that in our numerical example it is true that 3 young appear in court but only 2 old. It is not true that recorded offence rates are higher for the younger group, however.

[32] TIC stands for 'taken into consideration'.

The best explanation of the results concerning $s$ (the percentage of population which is working class) in the first two equations would seem to be in working class areas – where we suppose a possibly higher offence rate (or, alternatively, differential immunity or police discrimination) – that there has been a loosening of informal community controls and that there is a relative ease in obtaining convictions. If in 1961 working class informal controls were still strong (so that the police were less frequently involved) we need not expect '$s$' to be significant in the $y$ equation. If, however, these controls weakened in the first half of the 1960s, we should expect more minor offences to have been reported. It may also have been the case that such extra offences were difficult to solve. We therefore expect, and find, in 1961 a positive coefficient in the $p$-equation, but an insignificant coefficient in the $y$-equation and, further, in 1966, an insignificant coefficient in the $p$-equation but a positive coefficient in the $y$-equation.

### *4.3. Variance–covariance matrix of residuals*

We again concentrate on 1961 urban restricted and 1966 urban restricted – see tables 2(i) and 2(ii). We have not used any precise measure of goodness-of-fit for the structure as a whole, and it is not always the case that the estimated variance of the residual in any particular equation in a simultaneous equations context is less than the variance of the variable the equation is intended to 'explain'.[33] Any comments from a comparison of variance of structural residuals with the variances of the corresponding endogenous variable should, therefore, be oriented towards understanding the structure rather than examining goodness-of-fit.

The most interesting aspect of the variance–covariance matrix, however, is that the variance of the residual in the $p$-equation in 1966 urban is higher than the variance of $p$. The (technical) explanation lies in the large negative coefficient of $c$ in the $p$-equation together with the high positive correlation between the residuals in the second and third equations (0.80). Thus a positive random disturbance in $p$ would go with a positive disturbance in $c$ which would act through the second equation in the opposite direction to the original random disturbance on $p$ (in the 1966 pooled set the random disturbance on $p$ would raise $c$ directly

[33] The covariance of the fitted value and the error term may not vanish.

through the coefficient in the $p$-equation rather than through a correlation of the random terms). An explanation of the behaviour that produces this result could be as follows. A chief constable who is more successful in recruiting than would be predicted by the third equation, also achieves higher clear-up rates than predicted by equation two. This could be a morale effect in both equations, or that such a chief constable feels an obligation to show high clear-up rates and concentrates more on creating functions than deterring functions (there is also a positive correlation between the residuals in the first and third equation). It could alternatively simply mean that more efficient recruiters are more efficient offence-solvers.

The analyses of the previous paragraph conform to the common view that clear-up rates come under a strong chief constable influence. This random element (in the sense that it is not in our equations) can account only for a large residual in the second equation and not, without further elaboration, for the residual being larger than that of $p$. To explain this we seem to need a story along the lines just given.

It is more reasonable to use comparisons of the residual variances from the derived reduced form with the variables to be explained to examine goodness-of-fit since they tell us how well the variations in the exogenous variables account for those of the endogenous. The residual variance as a fraction of total variance for 1961 urban restricted for $y$, $p$, $c$ respectively is 0.69, 0.39, 0.77 and for 1966 0.29, 0.61, 0.80. The fit for $y$ and $p$ is better than for $c$ (problems with the $c$-equation were discussed in sect. 4.2). It is natural to expect random elements in this social process to be larger than for the more traditional economic processes; thus the fits for $p$ in 1961 and especially $y$ in 1966 must be regarded as rather good. The increased formalisation of reporting procedures in 1966 might account for the better explanation of $y$. A differential enthusiasm for the new emphasis on efficiency amongst chief constables would account for the worse explanation of $p$ in 1966.

### 4.4. Policy

Our comments will be mostly negative, partly since we do not feel that a model at our level of aggregation should lead directly to specific positive policy recommendations. We can, however, on the basis of what we have learned from our study, question some of the arguments used in favour of particular policies and thus, perhaps, the policies themselves.

The variables in our model that the authorities might try to influence directly are $f$, $c$, and $e$, presumably with a view to affecting the number of offences and the proportion solved. We have seen how the effects of $c$ and $e$ on the 'true' number of offences is obscured by the creating and reporting effects. In particular it seems that in 1966 the creating effect is rather important. If, therefore, more policemen provide an increase in recorded offences, it is illegitimate to use this increase as an argument for still more policemen (there may be other arguments). It is reasonable to ask proclaimers of a crime wave to make a judgment about recording effects when they use statistics.

The most promising variable for those interested in reducing offence rates would seem to be the clear-up rate. The severity of punishment seems to yield rather weak responses. For example, custodial treatment for 24% as opposed to 20% of offenders found guilty would imply a 3.4% reduction in offences (using 1966 estimate and calculating only the direct effect through the first structural equation). This would involve a 20% increase in the population in custody.

Unfortunately, we can give little guidance on how $p$ can be affected. The main consistent variable in the second structural equation, a negative effect of size of district on $p$, argues against the concentration on amalgamations. Thus more specific justifications of this policy than general appeals to economies of scale are necessary. It is difficult to go further than this except to note the apparent importance of local force practices (which presumably means chief constable behaviour) in determining $p$.

## 5. Conclusions and further research

Our analysis has lent some support to theories of offending behaviour that concentrate on probability of detection, severity of punishment if caught and the availability of swag, in so far as our measures of these variables proved significant at the aggregate level. We have also seen that the effect of the working class and the young on the crime figures seem to depend in an important way on their relations with the police. Further, the effect of the police themselves on these figures depends on their formal recording practices.

We have suggested plausible specific modes of individual, group and police behaviour which would lead to the effects we have observed at the aggregate level. This is, in part, the function of such aggregate analysis. The next step would be to examine these modes of behaviour at the micro-level. One line of attack would be to disaggregate into offence sub-classes. A major problem here, however, is to disaggregate the variables measuring police inputs and severity of punishment.[34]

Our application of the formal techniques of econometrics has been experimental and we wish it to be viewed as such. We hope we have shown, however, that problems of unobservable variables, identification and simultaneous causation in this area can be made more clear if they are handled in a formal manner and further that these problems need not be insuperable. Finally, we hope, in our interpretation of results and their policy implications not only to have increased our understanding of the generation of aggregate data but also to have thereby raised serious questions about the way in which these data are used in current policy discussion.

## Appendix

This appendix presents the results for the full 22-coefficient model described in sects. 1 and 3.1. These estimates were used mainly to show that there was a structural break between 1961 and 1966. We compare briefly these results with the estimates of table 2 which come from adding extra restrictions to the equations. Sect. 4.1 should be recalled at this point.

For 1961 urban and 1966 urban the extra restrictions had little effect on the chi-square values for the significance of the over-identifying restrictions taken as a group, but succeeded in tightening up the significance levels a little. For both data sets the null hypothesis that the over-identifying restrictions are correct is accepted at the 10% level (by a comfortable margin). The estimated variance–covariance matrix of the residuals is discussed at the end of sect. 4.2.

We also conducted likelihood ratio tests for the models of table 2 against the less restricted models of the appendix. The null hypothesis is that the extra restrictions of table 2 over the appendix are correct. The chi-square value for this test for 1961 urban was 3.46 with 7

[34] Martin and Wilson (1969) show that the breaking down of police time by activity on different types of offence is difficult.

degrees of freedom. We would accept the null hypothesis at the 80% level (Prob. $[\psi_7^2 > 3.82] = 0.80$). For 1966 urban we have a chi-square value of 7.66 with 5 degrees of freedom. We accept the null hypothesis at the 10% level. For 1966 urban and rural pooled we have a chi-square value of 1.43 with 5 degrees of freedom. We would accept the null hypothesis at the 90% level. We can therefore regard the extra restrictions as having some justification with that for 1961 urban and 1966 urban and rural pooled being strong.

A large improvement in the chi-square value is obtained when extra restrictions are put on for the 1966 urban and rural pooled set. Even so we could only reach a situation where the over-identifying restrictions

Table A(i)

**1961 Urban**

| Variable to be explained | Explanatory variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *y* | *p* | *c* | *f* | *a* | *s* | *t* | *e* | const. |
| | −0.77 | −0.04 | −0.24 | −0.10 | +0.18 | +0.17 | +0.03 | +13.4 |
| | (0.29) | (0.54) | (0.09) | (0.19) | (0.18) | (0.05) | (0.34) | (6.3) |
| | 2.68 | 0.08 | 2.58 | 0.51 | 1.01 | 3.27 | 0.09 | 2.11 |
| *p* | *y* | *c* | *a* | *s* | *n* | *e* | *v* | const. |
| | −0.03 | −0.99 | −0.36 | +0.26 | −0.13 | +0.28 | +0.12 | −0.91 |
| | (0.18) | (0.38) | (0.11) | (0.10) | (0.04) | (0.23) | (0.07) | (4.78) |
| | 0.15 | 2.57 | 3.33 | 2.55 | 3.48 | 1.23 | 1.70 | 0.19 |
| *c* | *y* | *p* | *m* | *v* | *d* | const. | | |
| | −0.01 | +0.04 | −0.99 | −0.06 | +0.01 | +0.22 | | |
| | (0.15) | (0.16) | (0.24) | (0.05) | (0.04) | (3.01) | | |
| | 0.10 | 0.23 | 4.12 | 1.17 | 0.25 | 0.07 | | |

| Covariance matrix of residuals | | | Variance | | Chi-square value of log-likelihood ratio 9.112 with 8 d.o.f. |
|---|---|---|---|---|---|
| 0.040 | | | *y* | 0.075 | |
| 0.007 | 0.020 | | *p* | 0.028 | |
| 0.010 | 0.009 | 0.011 | *c* | 0.013 | |

Notes (i) Number in brackets below coefficient is the asymptotic standard error. Below this is the *T*-value.

(ii) Variables defined in table 1.

(iii) Chi-square values discussed in sects. 3.3 and 4.1.

(iv) Programme SIMUL 7 by Cliff Wymer run on Harwell Atlas.

(v) The constant terms are affected by the scaling of variables for computational reasons – the variables in the results of tables A(i)–A(iv) are all in the same units.

## Table A (ii)

**1966 Urban**

| Variable to be explained | Explanatory variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *y* | *p* | *c* | *f* | *a* | *s* | *t* | *e* | const. |
| | −0.03 | +0.87 | −0.14 | +0.53 | +0.25 | +0.17 | +0.35 | +5.9 |
| | (0.28) | (0.56) | (0.08) | (0.18) | (0.16) | (0.06) | (0.17) | (4.3) |
| | 0.10 | 1.57 | 1.64 | 2.89 | 1.51 | 2.96 | 2.10 | 1.37 |
| *p* | *y* | *c* | *a* | *s* | *n* | *e* | *v* | const. |
| | +0.45 | −2.68 | −0.68 | +0.27 | −0.15 | +0.11 | −0.19 | −11.8 |
| | (0.55) | (1.41) | (0.49) | (0.20) | (0.07) | (0.24) | (0.21) | (10.40) |
| | 0.82 | 1.87 | 1.37 | 1.38 | 2.23 | 0.46 | 0.89 | 1.14 |
| *c* | *y* | *p* | *m* | *v* | *d* | const. | | |
| | −0.02 | +0.04 | −0.83 | −0.15 | +0.05 | 0.49 | | |
| | (0.05) | (0.16) | (0.31) | (0.06) | (0.04) | (2.66) | | |
| | 0.49 | 0.25 | 2.69 | 2.55 | 1.29 | 0.19 | | |

| Covariance matrix of residuals | | | Variance | | | Chi-square value of log-likelihood ratio 10.137 with 8 d.o.f. |
|---|---|---|---|---|---|---|
| 0.038 | | | | *y* | 0.173 | |
| −0.025 | 0.119 | | | *p* | 0.038 | |
| 0.002 | 0.033 | 0.015 | | *c* | 0.018 | |

## Table A (iii)

**1961 and 1966 Urban pooled**

| Variable to be explained | Explanatory variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *y* | *p* | *c* | *f* | *a* | *s* | *t* | *e* | const. |
| | −0.52 | +0.40 | −0.28 | +0.32 | +0.24 | +0.17 | +0.55 | +7.37 |
| | (0.24) | (0.42) | (0.07) | (0.13) | (0.13) | (0.04) | (0.14) | (3.50) |
| | 2.23 | 0.84 | 3.95 | 2.48 | 1.82 | 4.10 | 3.87 | 2.10 |
| *p* | *y* | *c* | *a* | *s* | *n* | *e* | *v* | const. |
| | −0.03 | −1.11 | −0.29 | +0.25 | −0.12 | +0.27 | +0.06 | −1.95 |
| | (0.12) | (0.33) | (0.11) | (0.08) | (0.02) | (0.09) | (0.06) | (2.90) |
| | 0.23 | 3.41 | 2.60 | 3.26 | 4.80 | 2.91 | 0.91 | 0.67 |
| *c* | *y* | *p* | *m* | *v* | *d* | const. | | |
| | +0.00 | −0.08 | −1.03 | −0.09 | +0.01 | +1.65 | | |
| | (0.04) | (0.11) | (0.21) | (0.04) | (0.02) | (2.06) | | |
| | 0.08 | 0.67 | 4.78 | 2.13 | 0.39 | 0.80 | | |

| Covariance matrix of residuals | | | Variance | | | Chi-square value of log-likelihood ratio 15.156 with 8 d.o.f. |
|---|---|---|---|---|---|---|
| 0.052 | | | | *y* | 0.171 | |
| 0.013 | 0.036 | | | *p* | 0.037 | |
| 0.013 | 0.019 | 0.016 | | *c* | 0.021 | |

Chi-square value of log-likelihood ratio in the rest for a structural break between 1961 and 1966: 79.30 with 28 d.o.f.

Table A(iv)

1966 Urban and rural pooled

| Variable to be explained | Explanatory variables | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *y* | *p* | *c* | *f* | *a* | *s* | %urbn. | *e* | const. | |
| | −0.58 | +0.76 | −0.17 | +0.62 | +0.11 | +0.45 | +0.41 | −1.60 | |
| | (0.24) | (0.22) | (0.09) | (0.15) | (0.13) | (0.09) | (0.14) | (2.29) | |
| | 2.45 | 3.46 | 1.87 | 4.27 | 0.87 | 4.95 | 2.84 | 0.70 | |
| *p* | *y* | *c* | *a* | *s* | *n* | *e* | *v* | %urbn. | const. |
| | −0.26 | −1.28 | +0.11 | +0.15 | −0.18 | +0.08 | −0.09 | +0.50 | −12.25 |
| | (0.53) | (1.53) | (0.38) | (0.20) | (0.13) | (0.28) | (0.27) | (0.41) | (12.02) |
| | 0.49 | 0.75 | 0.29 | 0.75 | 1.39 | 0.29 | 0.36 | 1.21 | 1.02 |
| *c* | *y* | *p* | *m* | *v* | %urbn. | const. | | | |
| | −0.01 | +1.24 | −0.91 | −0.45 | +0.41 | −3.76 | | | |
| | (0.09) | (0.41) | (0.42) | (0.12) | (0.11) | (2.33) | | | |
| | 0.12 | 3.03 | 2.11 | 3.83 | 3.80 | 1.61 | | | |

Covariance matrix of residuals

| | | |
|---|---|---|
| 0.039 | | |
| 0.010 | 0.049 | |
| −0.004 | 0.001 | 0.055 |

| Variance | | |
|---|---|---|
| | *y* | 0.175 |
| | *p* | 0.028 |
| | *c* | 0.042 |

Chi-square value of log-likelihood ratio 15.661 with 4 d.o.f.

are accepted at the 5% level. This, together with the high significance in each equation of the coefficient of the variable measuring percentage of the area that is urbanised leads us to suppose that urban and rural data are better treated separately.[35] We felt, however, that the rural observations constituted too small a sample to use to estimate a model of our size.

[35] A formal likelihood ratio test of this hypothesis could be conducted in a parallel way to our test for a structural break between 1961 and 1966.

## References

Becker, G., 1968, Crime and punishment: an economic approach, Journal of Political Economy 76, March–April.

Carr-Hill, R.A. and N.H. Stern, 1971, Variations in recorded statistics of police work, offender behaviour, and judicial activity in England and Wales in the early 1960s, mimeo.

Ennis, P.H., 1967, Criminal victimisation in the US, NORC University of Chicago.

Greenhaigh, W.F., 1966, Police regression analysis, SA/POL 6, Home Office, Scientific Adviser's Branch Police Report.

Grieson, R., 1972, The determinants of juvenile arrests, MIT working paper 87.
Hood, William C. and T.C. Koopmans, eds., 1952, Studies in econometric method, Cowles Commission Monograph 14, Wiley, New York.
Johnston, J., 1971, Econometric methods, 2nd ed., McGraw–Hill, New York.
Malinvaud, E., 1971, Statistical methods of econometrics, 2nd ed., North-Holland, Amsterdam.
Martin, J.P. and Gail Wilson, 1969, The police: a study in manpower, Heinemann, London.
Mizon, G., 1972, A test for structural change in a simultaneous equations model, St. Catherine's College, Oxford.
Phillips, H.L., D. Votey and Maxwell, 1972, Crime, youth and the labour market, Journal of Political Economy 80, 3, Part 1, May/June.
Wilkins, L.T., 1964. Social deviance: social policy, action and research, Tavistock, London.
Willmer, M., 1968, Is the battle being won?, Policy Research Bulletin No. 7.