FINAL PROGRESS REPORT

PHASE 1 OF GRANT #NI-71-041

I. INTRODUCTION AND SUMMARY

This report summarizes the effort on this grant during the past quarter, the last quarter of the first phase of the grant. The research has been focused on three major directions leading toward the development and implementation of the feedback version of the JUSSIM interactive model for criminal justice planning. The areas emphasized are:

1) Further development and reporting of the analytical feedback model to help shape the interactive model;
2) Development of a structure for the feedback version of the interactive JUSSIM model;
3) Development of a procedure for estimating virgin arrest rates from total arrest data in order to help users interact with the feedback model.

II. ANALYTICAL FEEDBACK MODEL

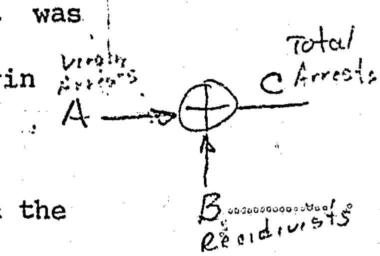
The paper entitled "An Analytical and Empirical Study of the Recidivism Process", submitted with the previous progress report, represented a first attempt to deal comprehensively with the feedback and recidivism process. One of the critical questions in exploring that process is the sensitivity of the results of the model to varying forms of virgin-arrest input data. These are important

$A_1 + B_1 = C_1$ $\quad$ if $B = cst$

$A_2 + B_2 = C_2$ $\quad$ $A_1 - A_2 = C_1 - C_2$

or $\Delta A = C_1 - C_2$

because of the inevitable uncertainty that will be involved in user's estimates of these virgin arrest rates. The model was extended to test its sensitivity to various forms of virgin arrest input rates.



The paper has been revised somewhat, as reflected in the enclosed paper. Although they are not yet reported in the draft paper, additional explorations were conducted to explore the sensitivity of total arrest (C) rates to reductions in virgin (ΔA) arrest rates compared to recidivism (B) rates.

III. STRUCTURING OF THE FEEDBACK MODEL

We have begun to formulate specific computer algorithms for introducing the feedback process into the JUSSIM model. Initially this is being accomplished with the DEMO model reported in the JUSSIM User's Manual.*

The feedback structure is being created by associating each drop-out point of the linear model with one of a small set (about 3-6) of "release points". At each release point each crime type will have an associated recidivism probability (α) and time lag until recidivism (τ). The flow from each release point will then be fed through a crime-switch matrix and returned to the total arrest process after an appropriate time lag. This process is illustrated in Figure 1.

A feedback model, in contrast to the linear model, operates as a "closed system" with the exception only of the continuing input of

---

*Belkin, J., A. Blumstein and W. Glass, "JUSSIM, An Interactive Computer Program for Analysis of Criminal Justice Systems", Urban Systems Institute, School of Urban and Public Affairs, Carnegie-Mellon University, July, 1971.

virgin arrestees, and the drop-out of the non-recidivists. As a result, there are new complexities in handling any program, such as decriminalization of some crime types, which changes the flow through the system, other than by influencing recidivism probabilities or virgin arrest rates. A separate "holding loop" is created to handle such programs. This loop will continue to process through the crime-switch matrix those previously arrested individuals who may be repeating the newly decriminalized offense, since they may also commit other offenses and so should reappear in the arrest stage as recidivists. This holding loop is illustrated by the dotted lines of Figure 1.

Since time is largely irrelevant to the linear model, one of the new technical issues to be dealt with in the feedback model is the choice of the appropriate time interval over which to measure recidivism. For the first time, we are concerned here with the explicit problem of multiple arrests per year by the same individual. If we examine the system only once per year, then it would be difficult to introduce those multiple arrests. Calculating flow at very short intervals would lead to excessive computational time. Thus, exploration of the sensitivity of calculated recidivism results to the computation interval have been undertaken, leading to an assessment that about 0.0 years is about an appropriate time interval.

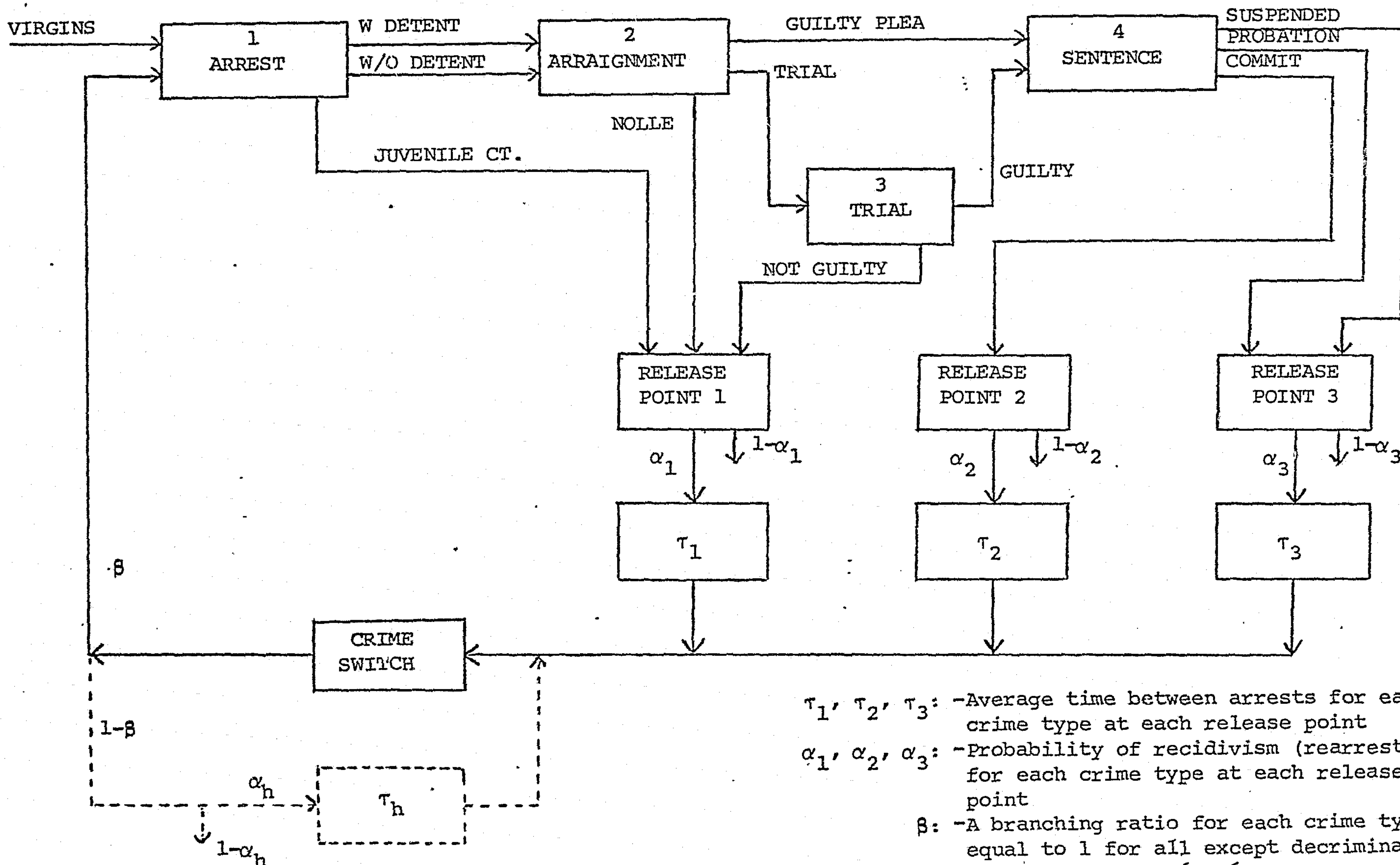
IV. ESTIMATION OF INITIAL CONDITIONS ON RECIDIVISTS

One of the requirements in operating a feedback model is the problem of providing an initial supply of "recidivists" who had been processed by the system in prior years, and who will be recidivating in the current and future years. Computational procedures have been developed to permit a user to provide recent historical data on total arrest rates (N(t)) and recidivism parameters α and τ, from which the program provides an initial set of future recidivist arrestees whose last arrest was experienced previously to the initiation of the model run.

V. PLANS FOR THE FUTURE

The plans for the immediate future upon extension of the grant include the formal programming of the feedback model to obtain an operating DEMO version by April 1. Initial user trials will be initiated shortly thereafter in parallel. Detailed data analysis will be undertaken to develop the base case recidivism parameters and crime-switch matrix for implementation in Philadelphia and/or Allegheny County. The detailed programming for the complete JUSSIM feedback model will be initiated as soon as critical user problems have been identified in preliminary user trials during the early spring. We are aiming for a full operational model to be used with representative CJS planners during the summer of 1972.

FIGURE 1



τ_1, τ_2, τ_3: -Average time between arrests for each crime type at each release point

$\alpha_1, \alpha_2, \alpha_3$: -Probability of recidivism (rearrest) for each crime type at each release point

β: -A branching ratio for each crime type equal to 1 for all except decriminalized offenses where $0 \leq \beta \leq 1$.

α_h, τ_h: -Recidivism parameters in holding loop set equal to representative values for a decriminalized offense

AN ANALYTICAL AND EMPIRICAL STUDY OF THE RECIDIVISM PROCESS

by

Jacob Belkin
Alfred Blumstein
William Glass

URBAN SYSTEMS INSTITUTE
SCHOOL OF URBAN AND PUBLIC AFFAIRS
CARNEGIE-MELLON UNIVERSITY

AN ANALYTICAL AND EMPIRICAL STUDY OF THE RECIDIVISM PROCESS

INTRODUCTION AND BACKGROUND

The future load on the criminal justice system will be the sum of a continuing input of people being arrested for the first time, virgin arrests, and the recycling of individuals who have been arrested previously, recidivists. While the forces that influence an individual to commit criminal acts, whether or not that individual had had prior contact with the CJS, are not generally under the control of the CJS, the individuals who are once in the system present a special responsibility, and provide a special opportunity, for the CJS to take some action that will reduce recidivism and thus reduce future loads.

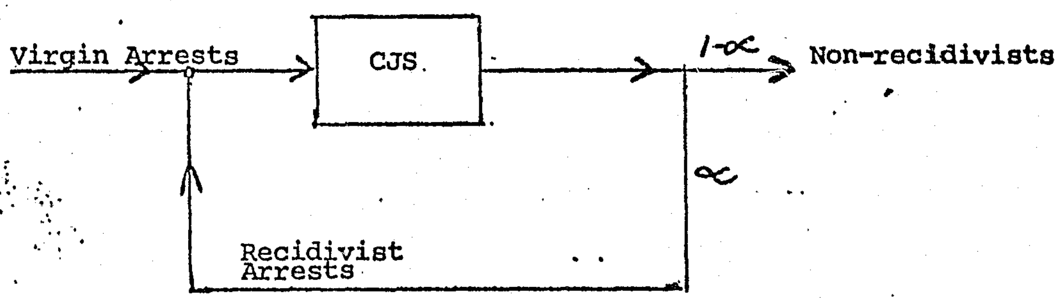
If there are opportunities for reducing recidivism through rehabilitation programs which do not require the CJS to incur additional costs, then clearly such programs should be implemented. However, it is more likely that the implementation of programs designed to reduce recidivism will add to the average system processing costs per individual. Such costs can be viewed as a current investment with an expected return equal to the reduction in future costs resulting from a reduction in the number of recidivists. To evaluate this investment and its possible future return, it is desirable to have a model which can be easily manipulated to determine the effect of a reduction in the probability of re-arrest on the future total load on the CJS comprehending the continuing input of virgin arrests.

The considerable leverage that a reduction in the probability of rearrest has on total future arrests for a cohort was demonstrated in work done with a feedback model of the criminal justice system.[1] In runs made with that model, a 1/3 reduction in re-arrest probabilities was shown to reduce total arrests for a cohort by a factor of about 2. However, the complexity of that model which included a crime-switch matrix and re-arrest functions with parameters dependent on disposition, age, and crime at last arrest, makes it difficult to identify the parameters critical to that result. In addition, this model did not comprehend the continuing input of virgins.

Baker[2] formulated a feedback model that comprehended a continuing input of virgin arrests and did some explorations of the impact of changes in recidivism probabilities but used no empirical data to estimate virgin arrests. Christensen[3], while dealing with the question of the cumulative probability of an individual being arrested sometime during his lifetime, estimated the probability of first arrest as a function of age but made no attempt to formulate an estimate of the number of virgin arrests per year, nor did he treat the recidivism process. Thus, there has been no successful analysis based on empirical data that deals with both virgin arrests and recidivists. Such an analysis would allow for an examination of the effect of changes in either component on the future load of the criminal justice system and would identify the role that each component has played in the sharp increase in total arrests during the last decade.

II. MODEL STRUCTURE

The simplest feedback model of the CJS can be represented by the following diagram:



Here, we aggregate the entire CJS into a single overall processing stage with one point of release and one flow back to arrest.* However, the other input to the CJS are those people arrested for the first time, "virgin" arrests. Thus, total arrests, which in general would be a reference point in the development of a feedback model (or a point at which the effects of changes in recidivism rates would be measured) becomes the sum of two components, the virgin arrests and the flow of recidivists.

Merely establishing the branching ratio or rearrest probability, is not sufficient to describe the recidivism process totally even in this simple model. The time between arrests can vary from just a few weeks or months for those individuals who are released (e.g, on bail,

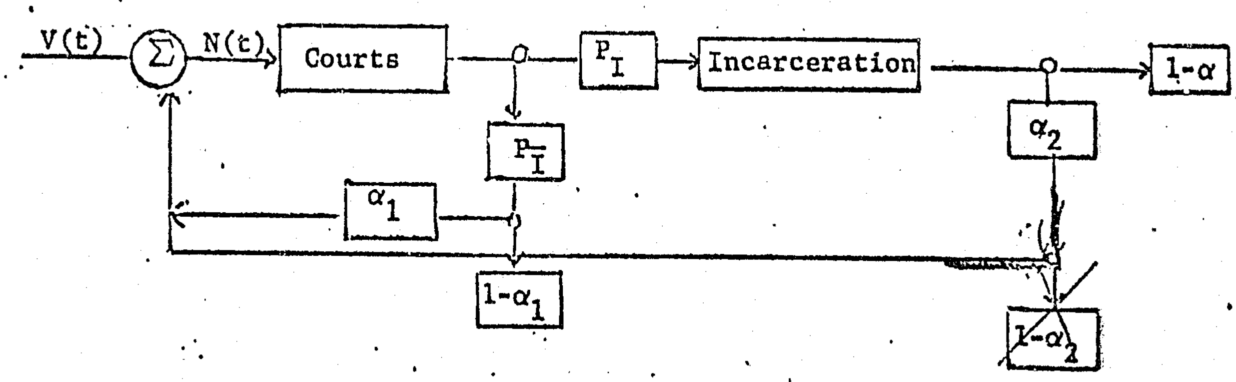
---

* In the discussion here we use rearrest as the basic system characterization of recidivism. Rearrest is an inappropriate definition of an individual's recidivism, since only conviction can be used as evidence of a commission of a crime. However, we are here concerned with the total flow through the CJS and from this perspective arrest defines the point of entering the system and thus rearrest is the appropriate recidivism definition. In our use of arrest data, we deal only with arrests for non-traffic violations including, however, arrests for drunken driving.

for the previous charge) to several years (e.g., individuals who for a long time either do not commit a crime, or successfully evade detection or are incarcerated). Obviously, those individuals who are arrested more than once during the course of a year contribute in a significantly different manner to the total arrest rates than those individuals who are not arrested again for several years.

This time lag between arrests also means that any change in recidivism parameters cannot be reflected immediately in changes in total arrests. Thus, the time lag between arrests must be incorporated in a feedback model to accurately characterize the recidivism process and to obtain realistic estimates of the effect of changing recidivism parameters. In addition, of course, the time variation of the other component of total arrests, namely virgin arrests, must be incorporated in a feedback model.

With only a small increase in the complexity of the model, it is possible to comprehend the fact that the average time between arrests for individuals incarcerated must be greater than the average time between arrests for individuals released without incarceration from all other points in the system. We thus have the following representation of the CJS and the recidivism process:

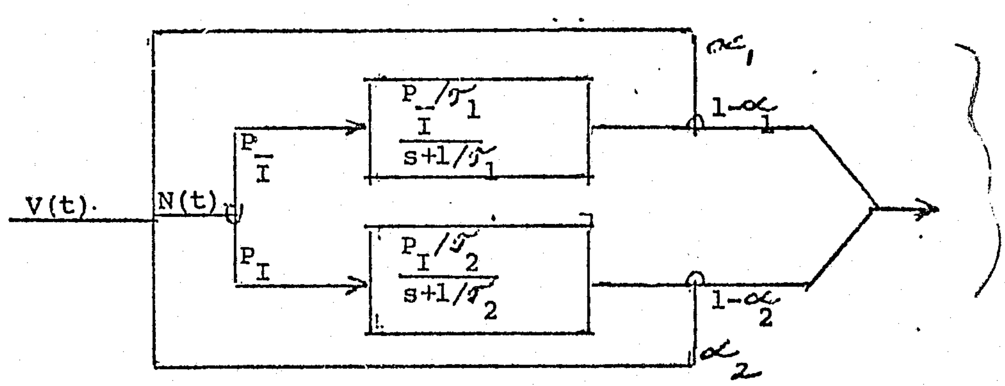


where p_I = probability of incarceration

$p_{\bar{I}}$ = probability of all dispositions other than incarceration (e.g., release prior to trial, acquittal, probation)

α_1, α_2 = proportion of individuals rearrested after release without incarceration and with incarceration respectively

The differential equation governing a positive feedback loop such as represented in the above diagram can be solved using Laplace transforms.* In terms of transforms, the model can be represented as follows:



where V(t) = virgin arrests per unit time at time t

N(t) = total arrests per unit time " " "

τ_1 = rearrest time constant for persons arrested but not incarcerated

τ_2 = rearrest time constant for persons arrested and incarcerated

P_I, $P_{\bar{I}}$, α_1, α_2 } as defined previously

*The authors wish to acknowledge the help of Professor R. Jandrisevits, Mechanical Engineering Department, Carnegie-Mellon University in formulating this solution to the differential equation.

In this simplified model, we have a single input, V(t), the rate of virgin arrests as a function of time; a single output, the total arrest rate, N(t) as a function of time; and six parameters, the branching ratios P_I, $P_{\bar{I}}$, α_1, and α_2 and the time constants τ_1 and τ_2.

III. MODEL ASSUMPTIONS

A. Branching Ratios - P_I, $P_{\bar{I}}$

In this model there is a single value for the probability of incarceration given arrest. Of course, the probability of incarceration given arrest varies considerably by crime type. For example, in a study of the Connecticut CJS,[4] the model results indicate the following probability of commitment for adults arrested and charged with the following offenses:

OFFENSES	% ADULT ARRESTEES COMMITTED
Burglary	17%
Larceny	7%
Narcotics	13%

In a study of a CJS in one county in Pennsylvania, data show that the probability of incarceration given arrest for juveniles varies by crime type and is different for new referrals and for referrals of juveniles already being supervised.

OFFENSES	% OF REFERRALS SENT TO INSTITUTIONS	
	New Referrals	Active Referrals
Burglary	3%	20%
Larceny	1%	9%
Narcotics	2%	16%

However, the proportion of total arrests resulting in incarceration is significantly changing over time; estimating P_I as a constant without regard to crime type should introduce no significant distortions in this macro-study of the recidivism process.

B. Recidivism Probabilities, α_1 and α_2

In this model, the recidivist proportions are explicitly identified only for two groups: those released after incarceration and those released at any other point in the system (e.g., prior to trial, acquittal, probation). In addition, the recidivism proportion for each group is assumed to be constant over time.

Sellin and Wolfgang[5] collected data on all of the police contacts generated by a cohort of almost 10,000 males born in 1945 in the city of Philadelphia over the period 1945 to 1963 (from birth through age 17). Data from this study has been used extensively throughout this work. At this point, the important results of the study are the estimates of the proportion of recidivists as a function of the number of prior arrests. In Chapter 11 of their forthcoming book, Sellin and Wolfgang investigated the transition matrix, as defined below, for the k to k+1 offense for $k = 1,\ldots 8$.

	Nonindex	Injury	Theft	Damage	Combination	Desist
Non index						
Injury						
Theft						
Damage						
Combination						
Desist						

Sellin and Wolfgang found that obtaining an average summary matrix was an adequate description of each of the transition matrices for $k > 2$. They further found that if the desist state was eliminated, then the average summary matrix was an adequate representation for all transitions $k \geq 1$. Their conclusion was that there was a statistically significant difference in rate of recidivism between transitions $\overline{12}$, $\overline{23}$, and all other offenses. Their data showed that the proportion of recidivism after the first offense was approximately 54% and the proportion after the second offense was approximately 65%, and the proportion of recidivism after subsequent offenses scattered between 70% and 80%, but appeared to be independent of the actual number of previous offenses.

Since the number of previous arrests is not a part of the present model, estimates of recidivism probabilities derived from the model must be interpreted to be average across the number of previous offenses. In light of the results obtained by Sellin and Wolfgang, this average will be higher than the actual recidivist proportions for offenders with one or two previous arrests and lower than the actual recidivist proportion for offenders with three or more previous arrests.

There are still other factors that apparently correlate with the probability of recidivism. As part of its Careers in Crime Program, the FBI has been publishing figures in its Annual Uniform Crime Reports on the follow-up of offenders after their release from the Federal criminal justice system. This follow-up was originally begun with a group of offenders released in 1963. Another group of offenders released in 1965 were followed for new arrests through 1969. In the 1970 issue of Uniform Crime Reports, results on this four-year follow-up show the effect on recidivism probabilities of type of release in 1965, the type of crime in 1965, and the age of the offender at release.

These results seem to indicate significant differences in recidivism rates by type of release in 1965. For example, of those people released after a sentence of fine or probation, only 37% were rearrested by 1969, while 85% of those people who were acquitted or had their cases dismissed in 1965 were rearrested by 1969. Similarly, significant differences appear by type of crime in 1965. For example, while 80% of those people arrested in 1965 for auto theft were rearrested by 1969, only 45% of those violating liquor laws in 1965 were rearrested by 1969.

The study further shows a somewhat significant effect of age on the probability of rearrest. For example, in the 1970 UCR, the follow-up indicates that 74% of the offenders released in 1965 under the age of 20 were rearrested by 1969, while 65% of those between the ages of 25 and 29 after release in 1965 were rearrested by 1969. For ages 60 and over, the recidivism rate has dropped to 38%.

Actually, the effects of all of these variables, point of release, crime type, and age, are entirely confounded. As pointed out in the 1970 UCR, "this type of sentence (fine and probation) is generally found in connection with violations such as income tax, fraud, and embezzlement." Similarly, age and crime type are confounded. Again examining the 1970 UCR (Table 28), less than 10,000 of the over 125,000 arrestees for auto theft were individuals 30 years old and over, while approximately 45% of the arrestees for fraud were in that age group. However, despite these difficulties, it does appear that the probability of recidivism is correlated with age and perhaps also with previous crime and previous disposition as well.

Because these factors are not comprehended in the present model the recidivism proportions are averages for all ages and for all types of previous crimes and previous dispositions. It is not apparent at this time how the effect of age can be incorporated in the model. However, identification of previous crime and previous disposition can be comprehended by linking the feedback model to JUSSIM[6], a linear model of the CJS that identifies by crime type the flow of individuals through the various stages of the CJS to points of release.

C. Time Constants, τ_1 and τ_2

The nature of the time constants as defined in this model, τ_1 and τ_2, are such that the proportion of the flow passing through a processing stage in an elapsed time t is $1 - e^{\frac{-(t)}{\tau}}$. Thus, the assumption is made that the time between arrest follows the negative exponential distribution with a mean time equal to τ.

---

*Effects of age could be comprehended at the expense of recognizing the effect of crime type by using the crime categories as age categories and using a "crime-switch" matrix as a means of advancing the age of each group in the arrestee population. Further consideration will have to be given to the implications of this trade-off between age and crime type.

Because in this model the court process is duplicated in the two paths of the circuit, the time from arrest to disposition in court is included in each of the two time constants τ_1 and τ_2.

A feedback model requires data on virgin arrests as distinct from recidivist arrests and data on recidivism probabilities. Generally speaking, data on virgin arrests is not available. There are many reasons, of course, why data on virgin arrests is difficult to obtain. For instance, the mobility of population, the sanctity of juvenile arrest records, and the absence of any routine examination of prior arrest history for adults.*

Routine collection of recidivism data is hampered by much the same reasons. The limited recidivism data that is available is generally developed as part of an evaluation of an experimental correctional treatment program and this may not reflect typical recidivism experience. Thus, there is little likelihood that potential users will have adequate recidivism data.

Typically, however, CJS planners will have available reasonably accurate data of <u>total arrests</u> for their jurisdiction over a number of years. We thus confront again the question of separating total arrest data into its two components.

Consider the relationship between virgin arrests and total arrests. For a group of individuals arrested for the first time in year i, V(i), a subset $V(i) * \alpha$ will eventually be arrested at least one more time. The number of these individuals arrested at least a third time will be $v(i) * \alpha^2$.

*Data on previous adult arrests such as that compiled by the California Bureau of Criminal Statistics and that reported in a special study of a Washington, D. C. criminal court are apparently limited to the arrest histories of felons, presumably because this data is obtained at the point of pre-sentence investigation.

Then the total number of arrests generated by this group of individuals will be

$$V(i) + V(i)\alpha + V(i)\alpha^2 + \ldots = \sum_{m=0}^{M} V(i)\alpha^m = V(i)\frac{(1-\alpha^M)}{(1-\alpha)}$$

which converges to $\frac{V(i)}{1-\alpha}$ as M becomes large, since $0 > \alpha > 1$.

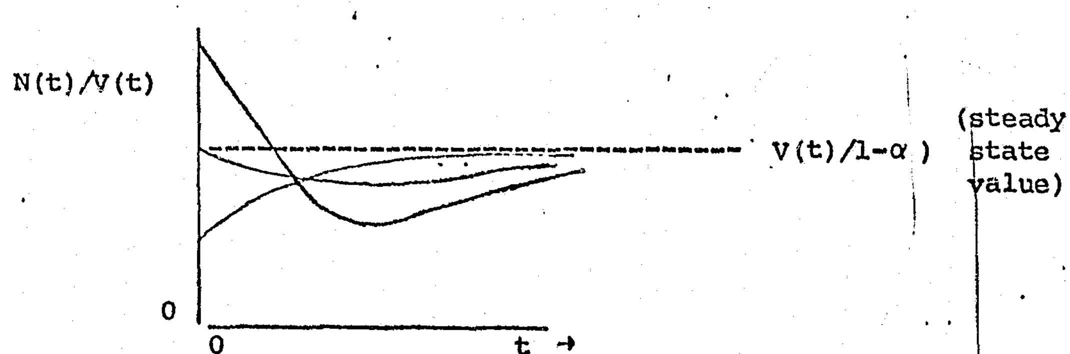
If we assume steady state, that is

$V(1) = V(2) = V(3) = v(n) = K$, and α and τ also independent of time, then the total number of arrests, N_1, generated by the arrests of virgins over a period of n years will equal $nK/(1-\alpha)$, and the annual number of total arrests, $N(t)$, is

$$N(t) = \frac{nK}{(1-\alpha)n} = \frac{K}{1-\alpha}$$

Thus, $1/(1-\alpha)$ can be taken as the "steady state" value for $N(t)/V(t)$.

The model has been solved repeatedly with a virgin arrest input of the form $V(t) = V(o) + Gt$, for a variety of initial values for total arrests, $N(0)$, for G, the slope of the virgin arrest curve, and for τ_1 and τ_2, the time constants governing the distribution of the time between arrests. In all cases, $\alpha_1 = \alpha_2$. The results of these runs for $G > 0$, and reasonable values for τ_1 and τ_2 (>6 months) may be generalized as follows:



While the shape of the curve depends on whether or not $N(0) > \frac{V(0)}{1-\alpha}$ in all cases $N(t)/V(t) < \frac{V(t)}{1-\alpha}$ for large t. Thus, if we can obtain an estimate of typical values of $N(t)/V(t)$, we can calculate a lower bound on the actual value of the aggregate recidivism branching ratio α.

We now generate the virgin arrests per unit time $v(t)$, and use data available from the Uniform Crime Reports, as an estimate of total arrests per unit time, $N(t)$, which we shall call UCR(t).

A. Virgin Arrest Rate

In 1967, Ronald Christensen[6] developed a model to estimate the probability that a person born in 1960 would be arrested during his lifetime. Although Christensen's objectives were different than ours, the portion of his model involving virgin arrests is of great interest. His model is based on the assumption that the probability of an individual being arrested for the first time (i.e., a virgin arrest) is a function of age only. This probability, p(a), was derived by estimating the number of virgin arrests by age for the entire U.S. and dividing by the total population of that age.

For our purposes, we wish to calculate the number of virgin arrests for the period 1930 to 1970 with the major emphasis on the years 1960 to 1970 for which reasonably good total arrest data is available. To do this, we assumed that the fraction of the population of age a that is virgin has been constant at least over the period 1960-1970. Since there may be a significant probability of first arrest even to age 30 it is necessary to invoke this assumption over the lifetime of such individuals, i.e., since 1930. This is equivalent to assuming that p(a) has been effectively stable since 1930. While there is little data currently available to indicate how p(a) may have been changing over this period, future refinements of this work will try to deal with this problem.

B. Population Estimation

To use the Christensen model, it is first necessary to obtain the age distribution of the U.S. population. These data were obtained for the years 1930, 1940, 1950, and 1960 from the U.S. census.[7] In addition, the yearly population estimates by five-year age categories were obtained from the Statistical Abstract for the years 1961 to 1969. The process for estimating the population by age for non-census years then proceeded as follows.

First, for the period 1930 to 1960, the population for each age was assumed to change linearly between census years. If T represents the year of the last census (i.e., 1930, 1940, or 1950), then

$$POP(a + t, T + t) = POP(a, T) + \frac{POP(a + 10, T + 10) - POP(a, T)}{10} t \qquad \text{for } t = 1,2,\ldots.9 \text{ and } a = 1,2,\ldots$$

In addition, the population of children born between census years was computed according to

$$POP(a, T + t) = POP(a + 10 - t, T + 10) \quad \text{for } a < t \leq 9$$

which assumes that there were no deaths between birth and the next census.

To obtain a more accurate estimate of the population during the period 1961 to 1969, we used the yearly census estimates of the number of children under one year of age and the remainder of the population in roughly five year age categories.

Thus it was possible to correct the population for each age forecast on the assumption of no deaths and zero net immigration by the ratio of the census bureau's estimate of total population by age category to the forecasted estimate of population for that age category.

$$POP(a,\ t) = POP(a - 1,\ t - 1) * \frac{EPOP(i,\ t)}{\sum_{l=j'-1}^{k-1} POP(l, t - 1)}$$

$$\text{for } j \leq a \leq k$$

where:

EPOP(i, t) is the estimated population of age group i in year t

and

age group i covers ages j through k, inclusive.

Finally, for the projection beyond 1969, the 1967 mortality rates by age[8], M(a), were used and net immigration was assumed to be zero

$$POP(a,\ t) = POP(a - 1,\ t - 1) * (1 - M(a - 1))$$

The number of births was calculated by assuming that they increased linearly from the 1969 level to the estimated 1980 level*

$$POP(0,\ t) = POP(0,\ 1969 + \frac{POP(0,\ 1980) - POP(0,\ 1969)}{11} * (t - 1969)$$

This procedure was used to project the population through the year 1978. The final results for the total U. S. population are shown in Figure 1.

*This value was obtained by using an approximate value of the mean of the series B, C, D, and E estimates of the number of births in 1980, as provided by the Bureau of the Census. However, the value used in this calculation is not important since only those born in 1970-1972 will be old enough by 1978 to have a non-zero value of p(a), and so few will be arrested that they will have little effect on the final result.

C. Virgin Arrest Rate Estimation

The other data necessary for the model is $p(a)$. This parameter has proved to be extremely hard to estimate, especially for ages 18 and over. Christensen derived $p(a)$ from estimating the fraction of all arrests that were virgin arrests, $r(a)$, and knowing the total number of arrests by age. However, very little hard data was available at that time to estimate $r(a)$. Fortunately, since 1967, more data became available which is a great help in obtaining $p(a)$ for juveniles. In Sellin and Wolfgang's birth cohort of 9,945 juvenile males in the city of Philadelphia covering the years 1945 to 1963 (birth through 17 years of age), 3,475 of the juveniles were arrested at least once. Since the data includes the age at first arrest and by assuming that the sample size is constant, we can obtain an estimate of $p(a)$ which we will call $p_1(a)$.

The first obvious problem with $p_1(a)$ is that it assumes an all male population. We next derive a revised estimate, called $p_2(a)$, which is for a male-female population. Now,

$$p_2(a) \approx \frac{\text{total virgin arrests}}{\text{total population}}$$

$$\approx \frac{\text{total virgin arrests}}{\text{total population}} * \frac{\text{male virgin arrests}}{\text{male virgin arrests}} * \frac{\text{male population}}{\text{male population}}$$

$$= \frac{\text{male virgin arrests}}{\text{male population}} * \frac{\text{male population}}{\text{total population}} * \frac{\text{total virgin arrests}}{\text{male virgin arrests}}$$

$$= p_1(a) * \frac{\text{male fraction of population}}{\text{male fraction of virgin arrests}}$$

The numerator is easily evaluated for the city of Philadelphia from the 1960 census data.[9] The denominator is slightly more troublesome, but reasonable estimates may be made from Philadelphia Juvenile Court data.[10] Since roughly 80% of referrals to the court are a result of an arrest,[11] the approximation should be reasonably good.

The next problem to correct for is the fact that we have based $p_2(a)$ on an urban population. To get p(a), we would like to do the following:

$$p(a) = p_2(a) * \frac{\text{U.S. virgin arrest rate}}{\text{Philadelphia virgin arrest rate}}$$

Unfortunately, evaluating the numerator is totally out of the question, since it is the answer that we are seeking. In light of this, we have had to settle for the following correction factor:

$$p(a) = p_2(a) * \frac{\text{U.S. arrest rate for juveniles}}{\text{city arrest rate for juveniles}}$$

$$= p_2(a) * \frac{3866}{4830}$$ [12]

Using this correction assumes that the fraction of arrests that are virgin arrests is the same for city and rural and that Philadelphia has a juvenile arrest rate of a typical city of 2,500 people or more.

This gives an estimate of p(a) through age 17. Estimates for age 18 and above have proven extremely difficult. Based on the 1965 study by the Stanford Research Institute of convicted felons in the District of Columbia[13], Christensen estimated that 7.6% of adult arrests were first arrests and used this estimate to calculate p(a) for ages 18 and over. The SRI study has several serious problems:

1. Only convicted felons are included, so that the sample is probably biased toward the repeater;
2. The sample is primarily drawn from a major city;
3. Previous arrests for drunkenness or disorderly conduct were not included in this portion of the study.

Thus, some errors bias the virgin fraction of arrests upward (#3) while others bias it downwards (#1, 2). This limits our ability to determine the true number. Since all similar studies have more serious drawbacks (e.g., omitting all juvenile arrests)[14], there is no sound reason for substituting another set of estimates for Christensen's. For this reason, we have used Christensen's p(a) for ages 18 and over. A plot of the final p(a) curve is shown in Figure 2. Note that there is no sharp discontinuity between ages 17 and 18. This implies that Christensen's estimates are reasonably consistent with those obtained from the Wolfgang study.

At this point, let us pause to derive an estimate of the probability that a person will have an arrest record by the time he is A years old. This is given by:

Male:
$$P_m(A) \approx \sum_{a=0}^{A} \frac{0.881}{0.492} \; p(a)\, L(a)$$

Female:
$$p_f(A) \approx 2 \sum_{a=0}^{A} \frac{0.119}{0.508} \; p(a)\, L(a)$$

where:

L(a) is the probability of a five-year-old surviving to age a.[15]

The results for this calculation are shown in Figure 3. These curves show the probability of arrest at the life expectancy age to be 0.60 for males and 0.16 for females as compared with Christensen's estimates of 0.52 and 0.13 respectively.

It is now a simple matter to multiply p(a) by the population estimates for each year and total across age to obtain the total virgin arrests for each year. These results for the years 1930 to 1978 are shown in Figure 3 .

We now have an estimate of virgin arrests that is based on the following assumptions:

1) p(a) is a time independent function of age;

2) the male fraction of first referrals to the Philadelphia Juvenile Court is the same as the male portion of virgin arrests;

3) r(a) is the same for cities over 2,500 in population and for the entire U.S.;

4) the juvenile arrest rate for Philadelphia is the average of all cities over 2,500 in population;

5) the estimate of p(a) for $a \geq 18$ as derived by Christensen is correct;

6) the arrest rate for the entire U.S. is the same as the rate observed in the portion of the population covered by the UCR.

It is hoped that in future developments of this model that these assumptions will not be necessary. By obtaining adequate arrest and population data for the city of Philadelphia, it will be possible to generate estimates of virgin arrests for Philadelphia. It should also be possible to obtain some data on the time variation of p(a). These improvements will eliminate the need for most of the current assumptions.

V. PARAMETER ESTIMATION

We first calculate a lower bound on α based on an estimate of the ratio $N(t)/V(t)$. This ratio can be estimated by developing an estimate of total arrests from UCR data, UCR(t), and dividing each year's estimate by the virgin arrests calculated for that year, as shown in Figure 4. The total arrests for the U.S. must be estimated since the FBI Uniform Crime Reports do not cover the entire U.S. population. To get these estimates, it was assumed that the arrest rate for the U.S. as a whole was the same as for the UCR population. Although this procedure overestimates total arrests,* it is necessary to use it since this assumption was used before in deriving p(a) (in the urban correction factor and by Christensen). The total arrests obtained by this process for the years 1960 to 1970 are plotted in Figure 5 and the ratio of total arrests to virgin arrests is plotted in Figure 6.

As these results show, the ratio of total arrests based on UCR data to the calculated value of virgin arrests over the period 1960-1970 was approximately 7. Thus, $\alpha \geq 1 - \frac{1}{7} = .857$, where $\alpha = p_{\bar{I}} \times \alpha_1 + (p_I) \times \alpha_2$.

We can improve our estimates of the recidivism branching ratios as well as the time constants by using the virgin arrest curve, as discussed above, as input to the simplified model of the recidivism process, and repeatedly solving for total arrests, N(t), for various values of α_1, α_2, τ_1, τ_2, until the best fit is obtained to the total arrests estimated from UCR data.

In performing these calculations, three assumptions were made. First, the branching ratios p_I and $p_{\bar{I}}$ were not estimated, but rather were assumed constant throughout the procedures used in exploring the other parameters.

*The UCR typically covers a higher percentage of the urban areas than of the rural areas. This biases the population towards cities, which typically have a higher than average arrest rate.

Since P_I has been explicitly identified only as a means of reflecting the effect of incarceration on the time between arrests, incarceration here reflects sentence to prison only. On this basis, $P_I = .03$ from data for Connecticut[16] and Allegheny County.[17]

Secondly, α_1 was assumed equal to α_2. This assumption was made because first, there were no hard data available to separately estimate α_1 and α_2. Further, since $P_{\bar{I}} * \alpha_1 + P_I * \alpha_2 > .85$ and $P_{\bar{I}} = .97$, α_1 must be very close to .85, and thus α_2 which is very likely to be in the range $\alpha_1 \leq \alpha_2 \leq 1$ cannot be significantly different from α_1.

The third assumption was that $\tau_2 = \tau_1 + 1$ year. As indicated above, the branching ratios $P_{\bar{I}}$ and P_I were assumed to be .97 and .03 respectively, on the basis that incarceration in the model should represent only individuals sent to prison. Based on the study of the Connecticut criminal justice system, the average time served for individuals in prison was approximately 1 year.*

With these assumptions, we calculated total arrests per year, N(t) over the period 1930-1970 for α over the range .865 to .885 in steps of .005 and for $\tau_1 = .5$ to 1.5 in steps of .1. Using as a measure of fit

$$\sum_{t=1960}^{1970} [N(t) - UCR(t)]^2$$

the curves plotted in Figure 7 show that at the lowest value of α, .865, the best fit is obtained at the minimum value of τ_1, .5. At the highest value

---

* Ref. 4, p. 64. The average daily population for the four Connecticut institutions housing adults serving extended sentences was about 1650. The results of applying JUSSIM to the Connecticut CJS showed about 1600 individuals sentenced to prison from the Connecticut Superior Court, the point at which the bulk of these prisoners are sentenced. Assuming steady state this yields an average time served of 1 year.

** As mentioned previously the data to which the model output was compared probably overestimates total arrests for the U.S. because of the necessity of scaling the UCR population to the U.S. population.

of α, .88, the best fit is obtained at the highest value of τ, 1.5. However, at the intermediate values of α, .87 and .875, the sums of squares is at a minimum within the range of τ_1 investigated, namely .8 and 1.1 years respectively.

These results show that an equally good fit can be obtained with very small changes in α, accompanied by significant shifts in τ. Thus, the most conclusive results obtained from these parametric studies is that α must be within the very small range of .87 to .88 provided that τ_1 is not less than .5 or greater than 1.5 years.***

Because the time between arrests have been assumed to follow an exponential distribution, the parameter τ_1 represents the average time between arrests for individuals not sentenced to prison. The average time between arrests estimated from the Sellin and Wolfgang cohort, without regard to any periods of incarceration, is 444 days, or 1.25 years, well within the range examined. While little additional hard data exists to make an estimate of the average time between arrests, Christensen[18] estimated that the average number of arrests per year per adult is 1.7 and for juveniles, 1.36. This would imply an average time between arrests of about .6 years for adults and about .7 years for juveniles. Thus, the range used in this experimentation, .5 to 1.5 years, seems to be sufficiently large. While this experimentation cannot give a more precise estimate of τ_1, this seems of little importance in light of the fact that these results provide a very narrow range on the estimate of α, namely that the recidivism probability averaged over all points of release in the system, over all ages of offenders, and over all values of previous arrests is between .87 and .88.

This estimate of the probability of recidivism is not out of line with other estimates. As indicated earlier, Sellin and Wolfgang found that for 3 or more arrests, the probability of recidivism was not affected by the number of previous arrests and was in the range of .70 to .80. Considering that the data from the juvenile cohort provides no information on criminal careers beyond age 18, and that we are here concerned with individuals arrested at least twice, the truncation in the data is bound to give only a lower bound on the probability of rearrest. Further, as pointed out by Blumstein and Larson[19], a recidivism rate of .875 for arrest to arrest is consistent with a corrections to corrections recidivism of .3, which is in the range of the typical rates reported for corrections to corrections recidivism.

While the estimates of the recidivism probability seems reasonable, a troubling assumption in the model is that the probability of first arrest remains constant for all ages at least through the last decade, and in fact in the calculations, throughout the period 1930 - 1970. As Figure 8 shows the total arrests, $N(t)$, as calculated with $\alpha = .875$, and $\tau_1 = 1.1$, $n(t)$, does not show the sharp increase during the period 1966 to 1970 that is exhibited by UCR (t), and thus the calculated value of total arrests underestimates the arrests estimated from UCR data increasingly in 1969 and 1970. This discrepancy in the most recent years could be accounted for by an increase in virgin arrests beyond that accounted for by changes in the population. In order to test the possibility that our estimates of the recidivism probability might be significantly affected by a more sharply increasing rate of virgin arrests, a new virgin arrest function $v'(t)$, was postulated. This new function is shown in Figure 9 along with the originally calculated value of virgin arrests, $v(t)$.

This new $v'(t)$ was not entirely arbitrarily chosen. Manipulation of UCR and population data indicates that while the rate of arrests for the 15-20 year-old group was essentially the same in 1963 as it was in 1960, it has been steadily increasing from a value of 7,990 per 100,000 in 1963 to 11,660 arrests per 100,000 in 1970. Despite the fact that this increase in the rate of arrests of the 15-20 year-old group probably over estimates changes in virgin arrests, since many in this age group are undoubtedly not being arrested for the first time, we estimated $v'(t)$ on the basis of this rate of increase.

With this new input of virgin arrests, the same grid search over α and τ was performed seeking the best fit on total arrests as estimated from UCR data for the period 1960-1970. The output from the model using this new virgin arrest function called $n'(t)$, is shown in Figure 10 for $\alpha = .86$, and $\tau_1 = 1.2$. Also shown in Figure 10 are the total arrests estimated from UCR data, UCR(t).

The results of attempting to fit the output of the model to the total arrests estimated from UCR data while using $v'(t)$ as input can be generalized as follows:

At no combination of α and τ was the absolute error over the entire period as low as could be obtained from using the $v(t)$ as input to the model. At $\alpha \geq .87$, the output of the model, $n'(t)$, even at $\tau_1 = 1.5$, overestimated the total arrests as estimated from UCR data for the most recent period, by as much as $n(t)$ had underestimated the total arrests in the same period. At $\alpha \leq .85$, $n'(t)$ underestimated total arrests based on UCR data for the period 1960-1965 by as much as a million and a half arrests in 1960.

As a result, there is no reason to believe that an overall recidivism probability for the entire period 1960 - 1970 could be significantly lower than our initial estimates of between .87 and .88 particularly since v'(t) probably overestimates the effect on virgin arrests of any increase in p)a), the probability of first arrest at age a , and this should drive down the value of α necessary to obtain a good fit with the total arrests estimated from UCR data. However, the tradeoff in the error between the first and second half of the 1960 - 1970 period does suggest that there may be a slightly lower α in the second five-year period than in the first five-year period. This would be consistent with the speculation that virgins with a lower than average probability of recidivism could be a higher proportion of total arrests in the second half of the decade than in the first half. However, while a combination of increasing virgin arrests and a lowering in the average recidivism probability might be plausible, it does not appear that an attempt to represent this process would result in any significant change in our estimates of the parameters governing the recidivism process.

Perhaps the most important implication of these results is that if we assume that recidivism after the first arrest is significantly lower than .875, then we must acknowledge a recidivism rate for offenders with a least two arrests of $>.9$. Since we are here dealing with recidivism rates averaged over age, and there is evidence, previously mentioned, that recidivism rates are lower for offenders older than 40, these results suggest that individuals arrested at least twice are almost certainly locked into a continuing criminal career.

Conclusions

This work demonstrates that it is possible to approximate the total arrests based on reported data from UCR from a curve of virgin arrests generated by estimating a single value of the probability of first arrest for every age category, while recognizing only the population changes by age over the period 1930 to 1970. While it is not entirely satisfying to assume that this probability of first arrest is independent of time, particularly in view of the fact that even a cursory examination of UCR data indicates a more and more youthful arrested population, it is significant that it is not necessary to propose sharply changing probabilities of first arrest to generate total arrest figures very close to those actually reported.

With this estimated probability of first arrest and a simplified model of the recidivism process, it becomes possible to quickly estimate reasonable recidivism parameters necessary to reproduce the total arrest figures for any jurisdiction. We thus have a method for providing users with the estimates of the parameters they need in working with a version of JUSSIM extended to include the feedback process. The next major step in providing the basic building blocks for such an extended model is the development of a matrix which indicates the switches in crime for recidivists from arrest i to arrest $i + 1$.

In addition to its use as a means of estimating the parameters of the recidivism process, a simplified model and the virgin arrest curve can be used to forecast total arrests and to examine the effects of changing recidivism rates. While previous work [20] has shown the significant effects that changes in recidivism probability can have on individual criminal careers, this model can demonstrate the effects of changes in recidivism on total arrests, which is, presumably, one of the major measures of effectiveness available to CJS planners. In addition, this model can demonstrate the importance of the time constants in recidivism which influence the immediacy of the effects that can be expected from changes in recidivism. In this way, some help can be provided to CJS planners in evaluating the effects of programs intended to reduce recidivism.

FIGURE 1

ESTIMATED TOTAL U. S. POPULATION 1930-1973

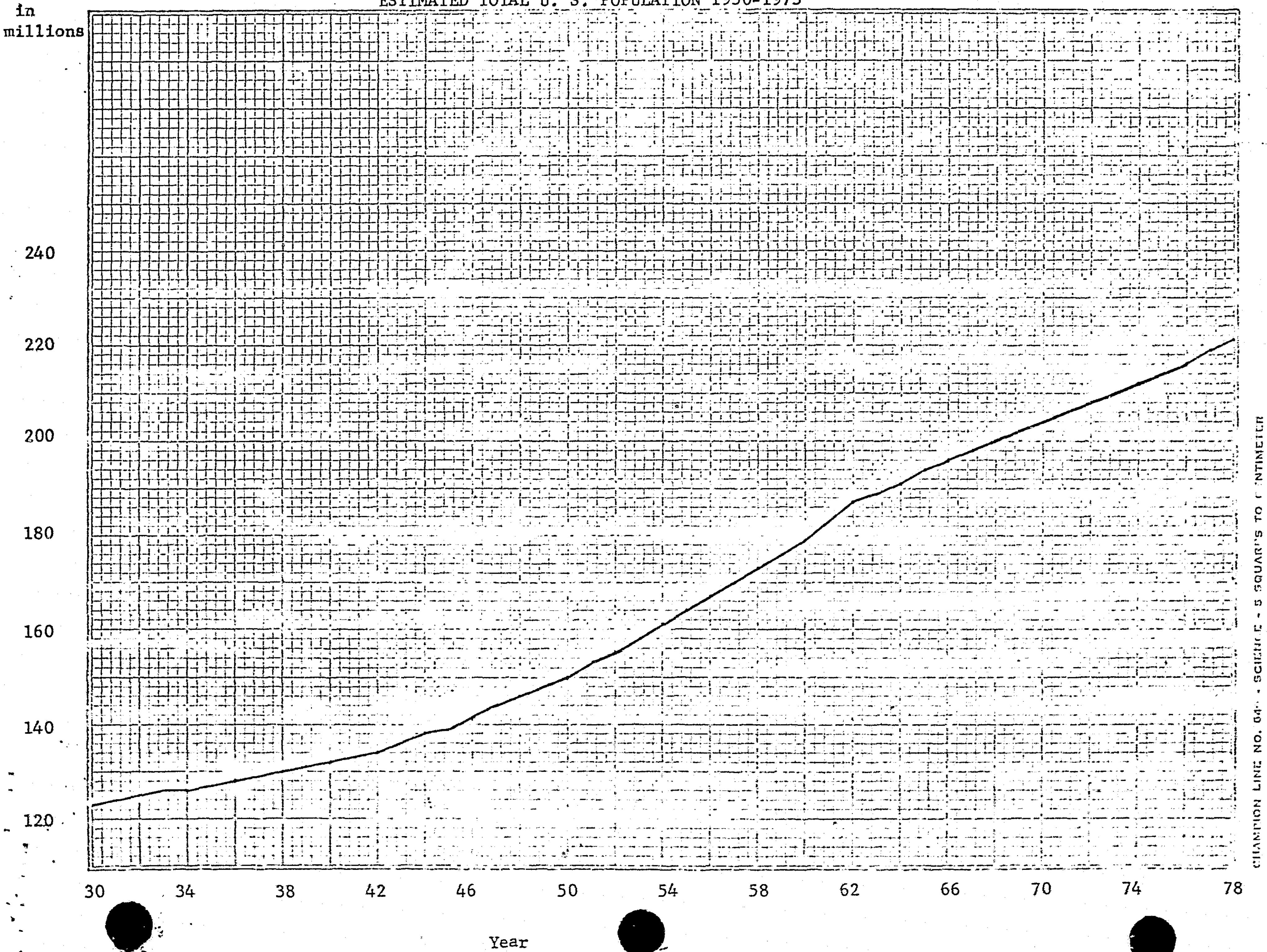


FIGURE 2
P(a)

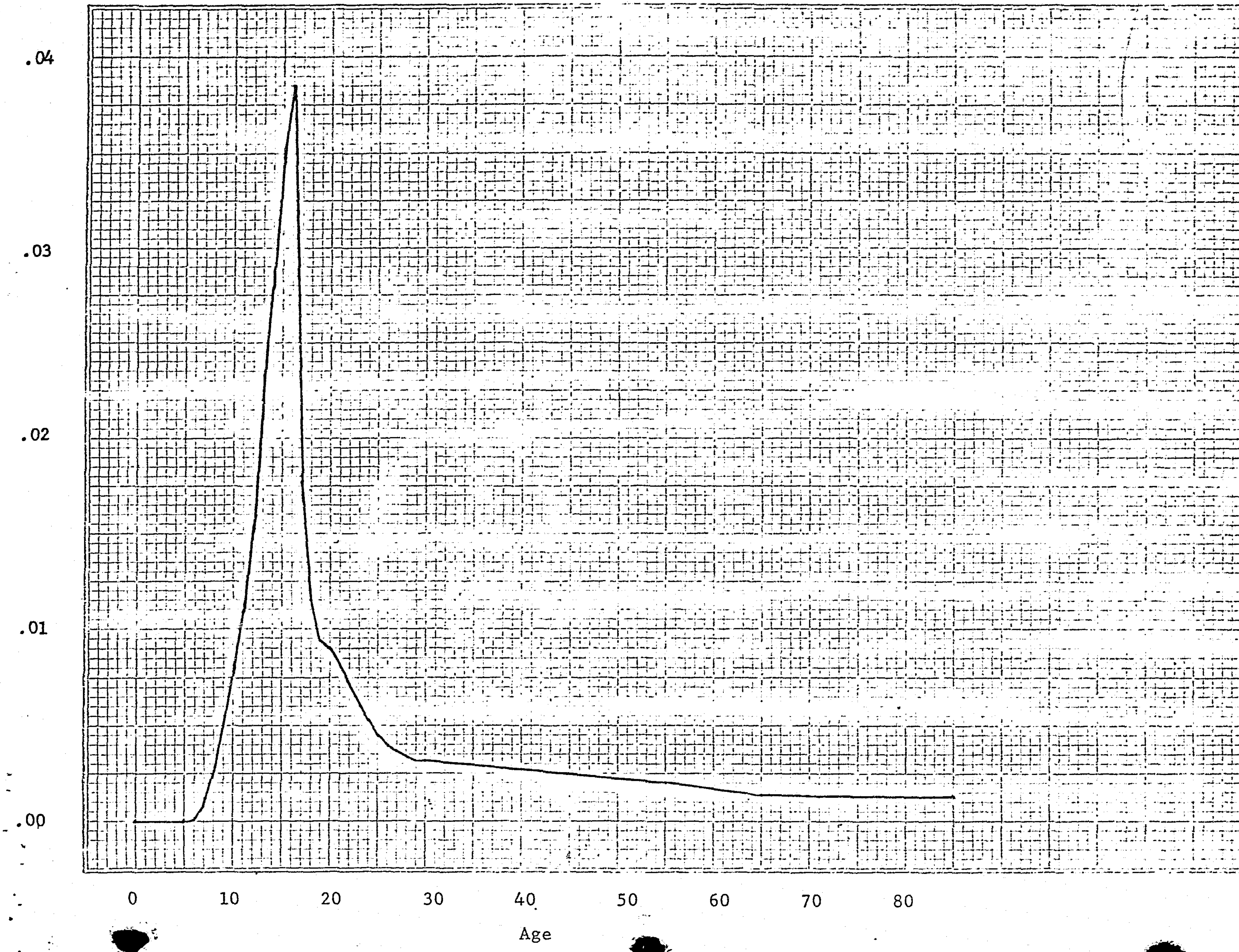


FIGURE 3

PROBABILITY OF BEING ARRESTED BY A GIVEN AGE

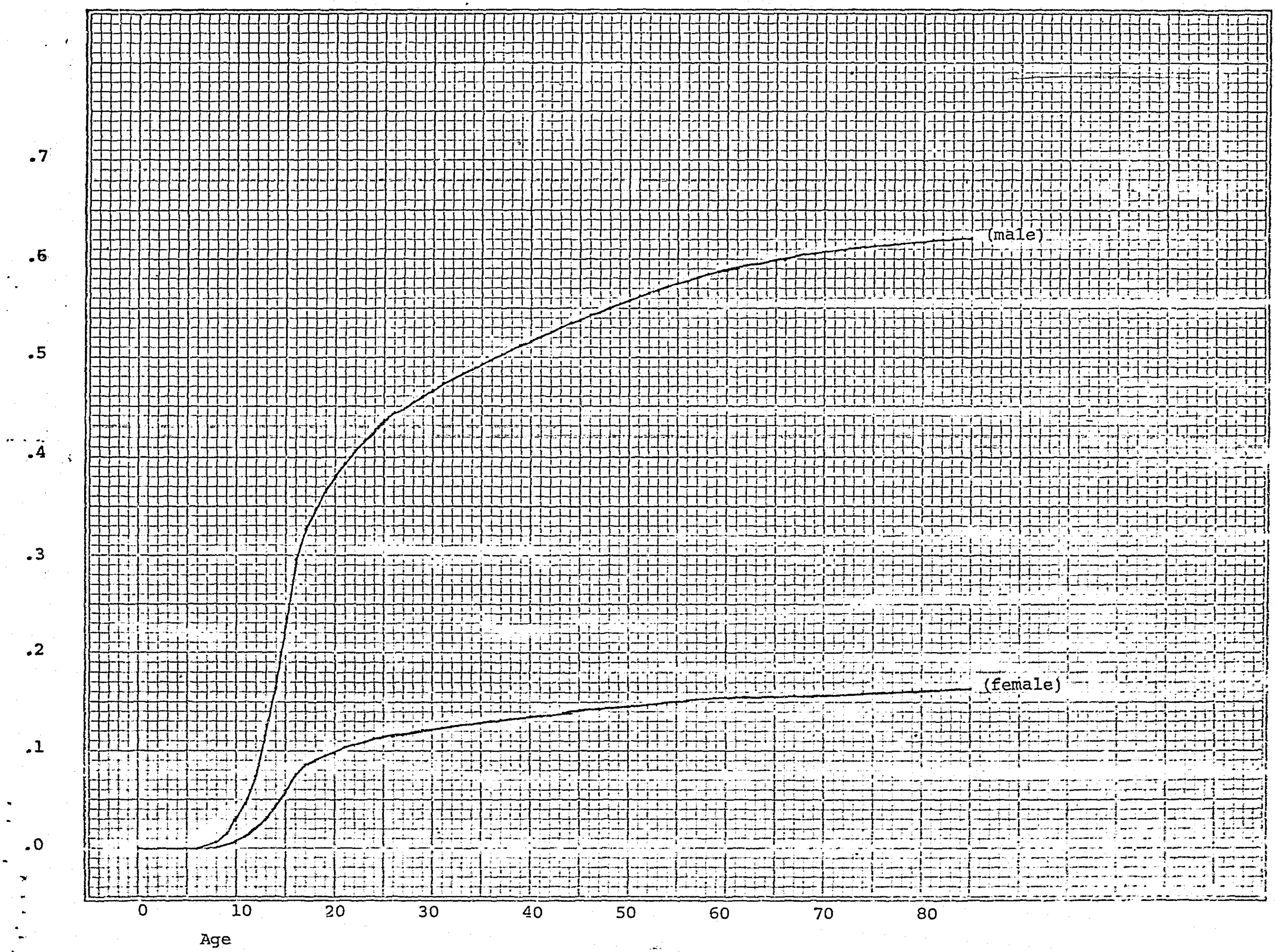


FIGURE 4

ESTIMATED VIRGIN ARRESTS 1930-1978

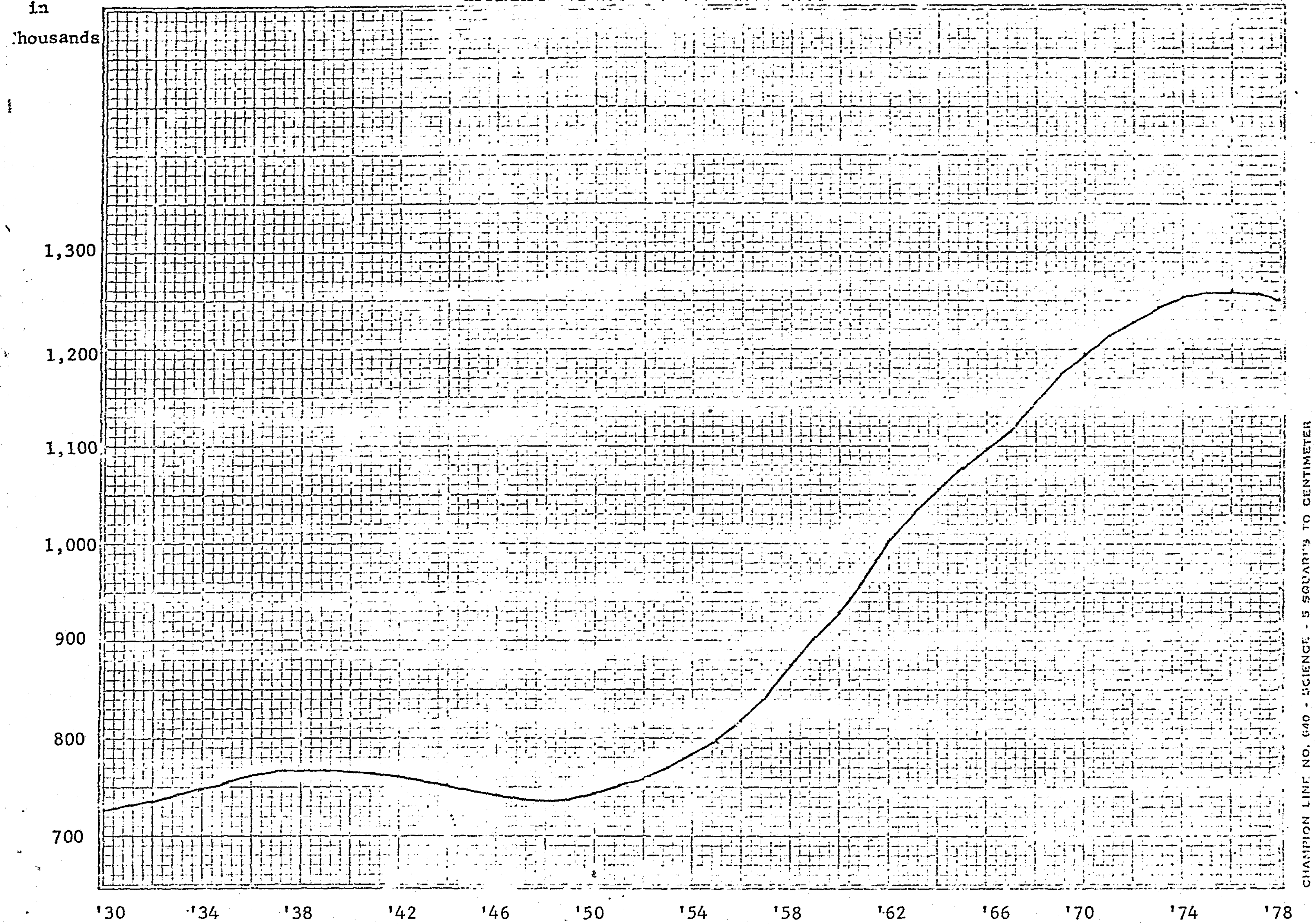
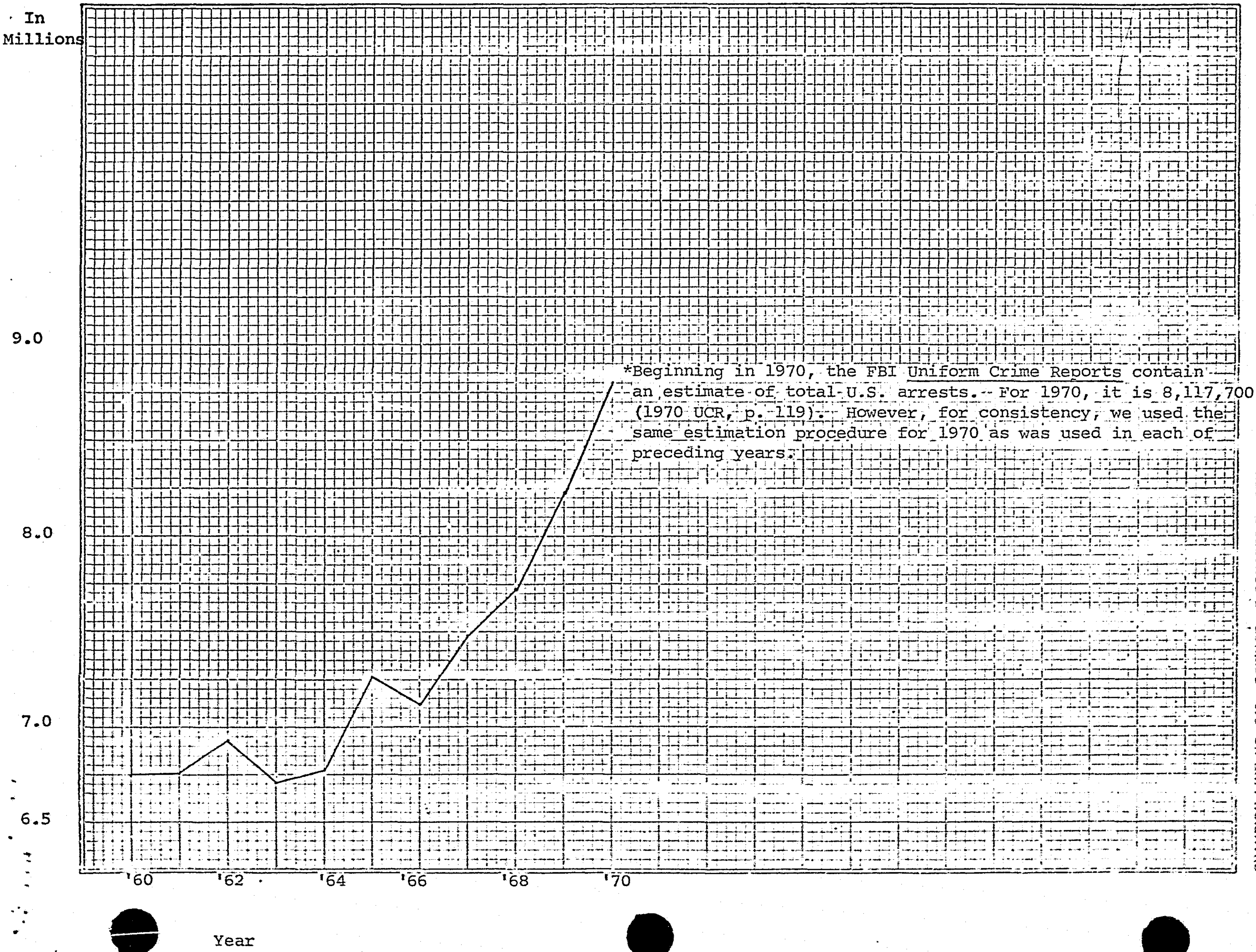


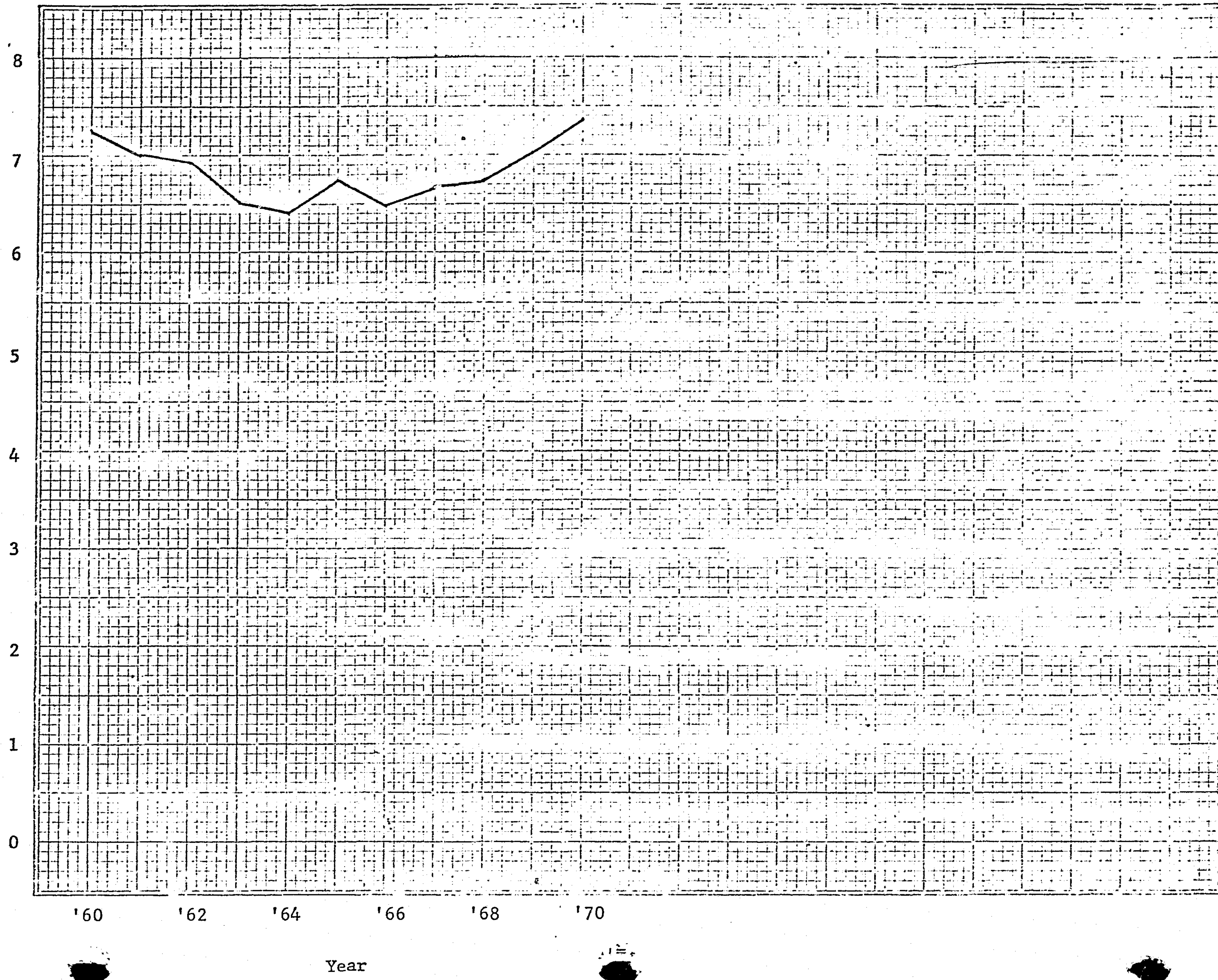
FIGURE 5
ESTIMATED TOTAL U.S. ARRESTS



*Beginning in 1970, the FBI Uniform Crime Reports contain an estimate of total U.S. arrests. For 1970, it is 8,117,700 (1970 UCR, p. 119). However, for consistency, we used the same estimation procedure for 1970 as was used in each of preceding years.

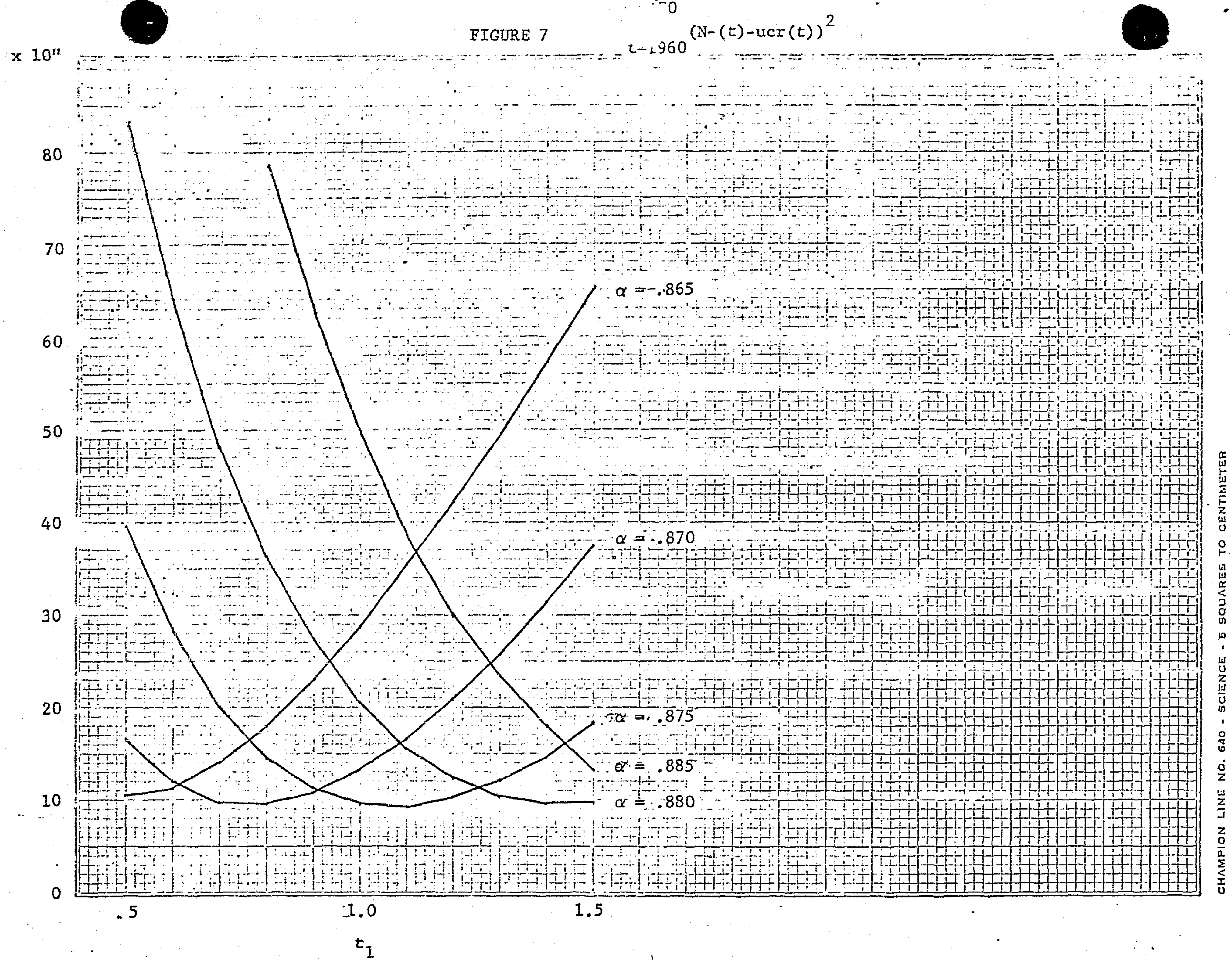
FIGURE 6

RATIO OF TOTAL ARRESTS TO VIRGIN ARRESTS



CHAMPION LINE NO. 640 - SCIENCE - 5 SQUARES TO CENTIMETER

FIGURE 7



0
t-1960 (N-(t)-ucr(t))^2
x 10"
80
70
60
50
40
30
20
10
0
.5
1.0
1.5
t1
α = -.865
α = -.870
α = .875
α = .885
α = .880
CHAMPION LINE NO. 640 - SCIENCE - 5 SQUARES TO CENTIMETER

N(5) vs. UCR(t) 1960-1970

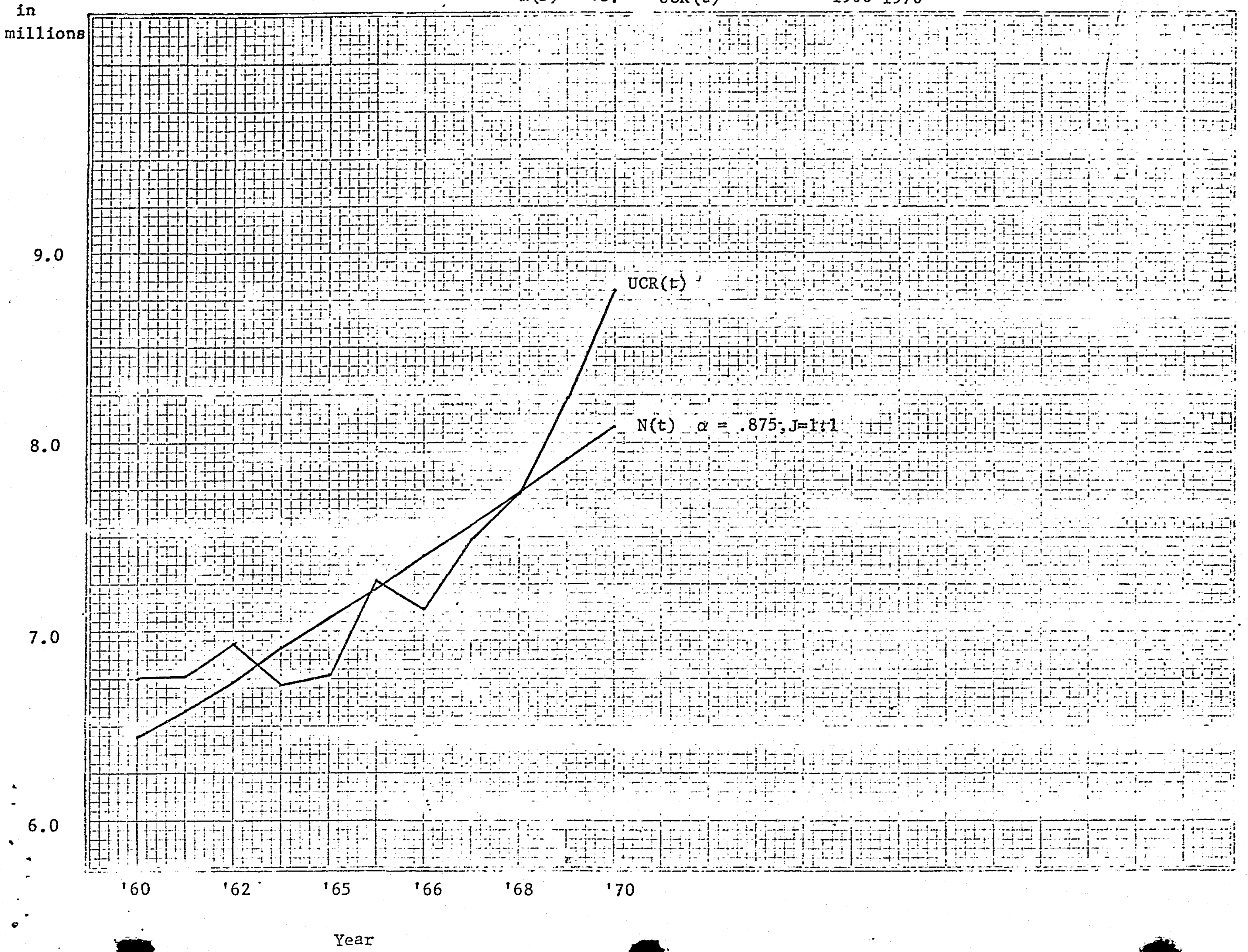


FIG. 9

V (t) and V'(t) 1960-1970

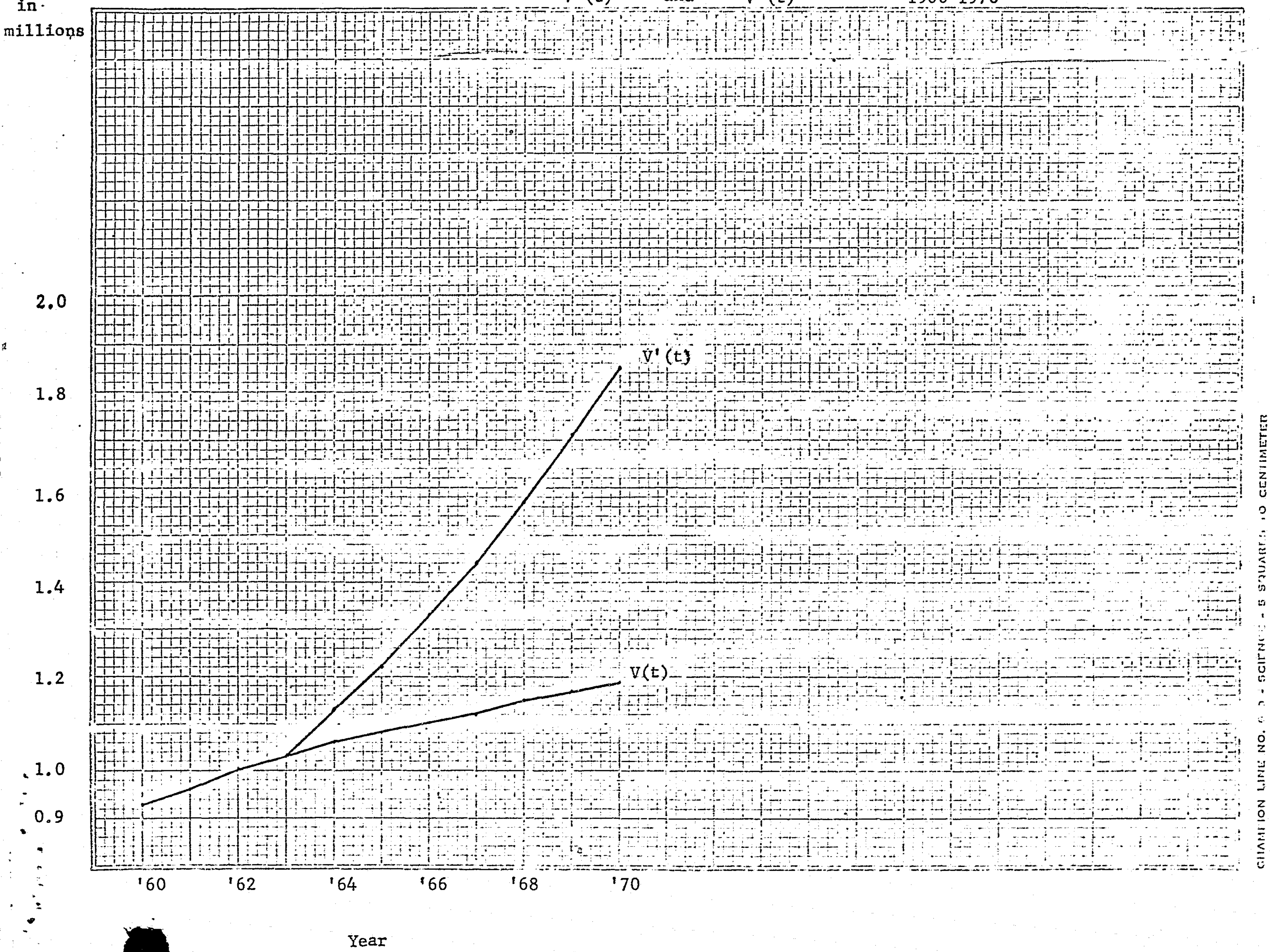
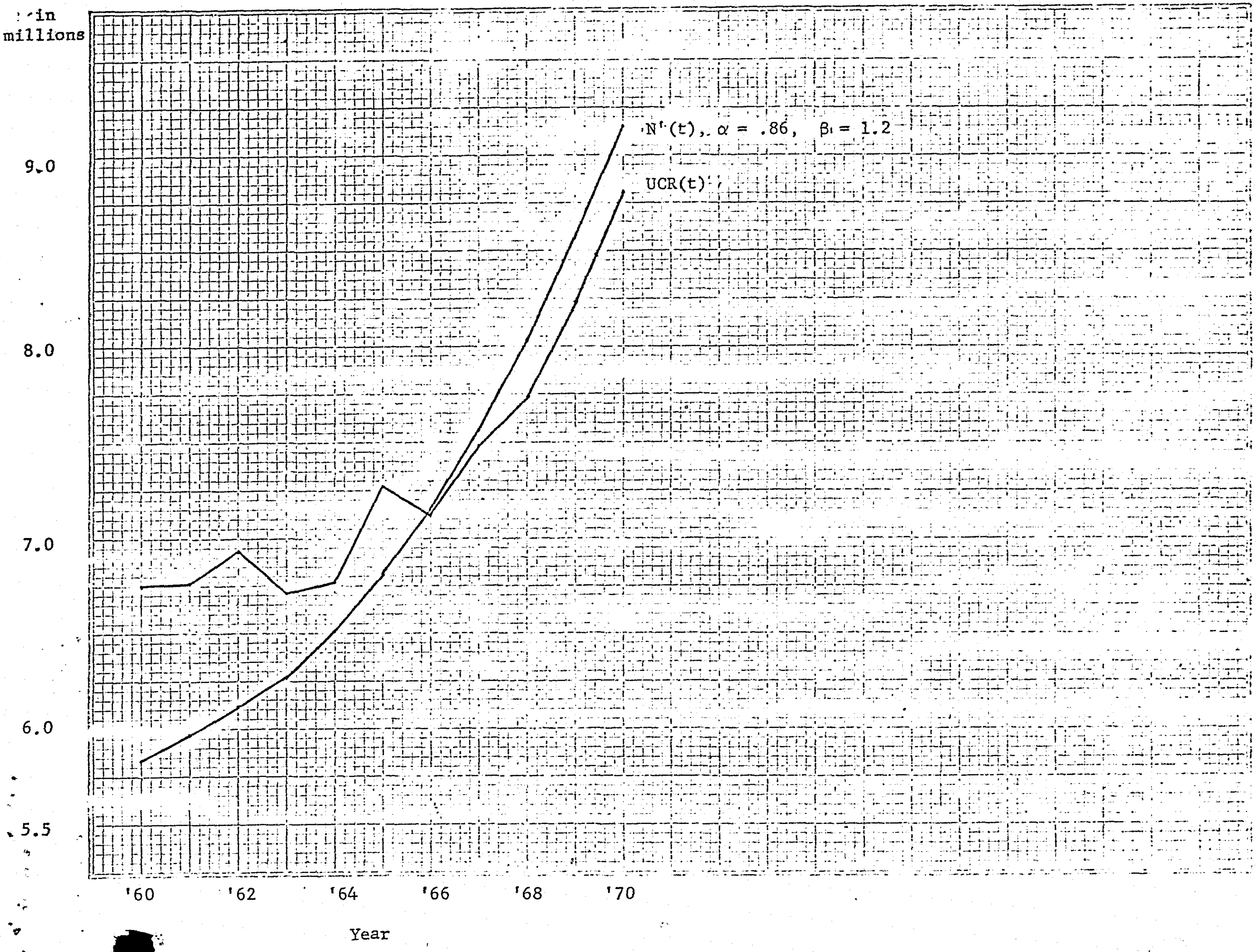


FIGURE 10

N'(t) vs. UCR(t) 1960-1970



REFERENCES

1. Blumstein, A., R. Larson, "Models of a Total Criminal Justice System", Operations Research, Vol. 17 #2, March-April, 1969.

2. Baker, James D., "On the Criminal Justice System", Law Enforcement Science Technology, Vol. II, S.I. Cohn, Editor, IIT Research Institute

3. Christensen, R., "Projected Percentage of U. S. Population with Criminal Arrest and Conviction Records", Appendix J., Task Force Report: Science and Technology, The President's Commission on Law Enforcement and Administration of Justice

4. Belkin, J., A. Blumstein, "Methodology for the Analysis of Total Criminal Justice Systems", Urban Systems Institute Working Paper, School of Urban and Public Affairs, Carnegie-Mellon University, November, 1970.

5. Sellin, T., and M. Wolfgang, Delinquency in a Birth Cohort, University of Chicago Press, 1972.

6. Belkin, J., A. Blumstein, W. Glass, "JUSSIM An Interactive Computer Program for Analysis of Criminal Justice Systems", Urban Systems Institute Working Paper, School of Urban and Public Affairs, Carnegie-Mellon University, July, 1971.

7. op cit. Ref. 3

8. "1960 Census of the Population: United States Summary", pp. 1-349, 1-350, 1-358.

9. Statistical Abstract of the United States - 1970, pp. 54-55

10. 1960 Census of the Population: Pennsylvania - General Population Characteristics, p. 40-103.

11. 48th Annual Report of the County Court of Philadelphia, 1961, p. 133. (The year 1960 was used for these calculations because of the reliance on the 1960 census. Since most virgin arrests for the cohort were between ages 13 and 17 (1958 to 1963), 1960 may be assumed to be an average year.

12. Ibid, pp. 52, 129.

13. Crime in the United States, Uniform Crime Reports - 1960, July 24, 1961, p. 89.

14. Report of the President's Commission on Crime in the District of Columbia, 1966, Appendix, pp. 511-644.

15. Crime and Delinquency in California - 1969, pp. 111-113. (This report shows that 23% of all felony defendants disposed of in California Superior Courts during 1969 had no prior record of arrest as adults. In the Stanford Research Institute study, 11.2% of the sample had a juvenile record, but no adult record.)(op.cit. p. 588.)

16. op.cit. Reference 3, pp 217-221.

17. op.cit. Reference 4.

18. Fields, Kenneth, M. Lettre, R. Stafford, "A Preliminary Description of the Allegheny County Criminal Justice System", Working Paper, Urban Systems Institute, School of Urban and Public Affairs, Carnegie-Mellon University, August, 1971.

19. op.cit. Reference 3

20. Blumstein, A., R. Larson, "Problems in Modeling and Measuring Recidivism" Journal of Research on Crime and Delinquency, (to be published, July 1972).

END