Document Title:	**Detection and Prediction of Geographic Change in Crime Rates: Final Report**
Author(s):	**Peter Rogerson ; Rajan Batta ; Christopher Rump ; Alok Baveja**
Document No.:	**202974**
Date Received:	**11/21/2003**
Award Number:	**98-IJ-CX-K008**

Detection and Prediction of Geographic Changes in Crime Rates: Final Report

NIJ Grant Number: 98-IJ-CX-K008, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice

Principal Investigators: Peter Rogerson, Rajan Batta, and Christopher Rump
Consultant: Alok Baveja

Preface

An important element of effective law enforcement and community policing efforts is the quick identification of emergent "hot spots" of increasing criminal activity. Similarly, it is of interest to identify areas of declining activity in a timely manner, to aid in the development of appropriate and effective responses.

One objective of our research was to develop statistical methods and monitoring models for the quick detection of emerging and declining geographic clusters of criminal activity. Both "global" methods that monitor changes across an entire study area and "local" methods that focus upon smaller subareas were developed.

Clusters of criminal activity are often well-known, and current software may do little more than confirm what is already known about the existence of geographical patterns of crime. Our focus was upon the detection of clusters that occur in relation to some preexisting expectations (e.g., previous year's data). Thus only clusters that exist over and above what is expected will be detected. We also focused upon the monitoring of data as it becomes available, with the objective of detecting changes in geographic patterns as quickly as possible. The focus on clustering and changes in clustering is the subject of Chapters 1 and 2.

A second and related objective was to develop prediction models that forecast how the pattern of crime will change (i.e., geographic displacement) in response to deployments of resources. A focus on situational prevention calls for an evaluation of the effects of displacement and diffusion. Mounting evidence suggests that earlier assumptions that the displacement of crimes to other locations would be the natural result of enforcement may be overstated (Gabor 1990; Hesseling 1994). In addition, diffusion effects, whereby the benefits of enforcement spread to other areas, may be substantial (Sherman 1990; Weisburd and Green 1995a). Weisburd, in his development of a research agenda, suggests that "to better understand displacement and diffusion, studies should be initiated that are directed at these effects and not at the primary outcomes of crime prevention initiatives" (p. 15). Chapter 3 focuses upon the details of our socioeconomic model of geographical displacement and the spatial concentration of crime.

Using predictive models within a GIS context has implications for policing beyond fighting crime and disorder problems. Such models also have uses for strategic and budgetary planning, something that has to date been difficult to do in most police agencies that are often driven by crisis situations or political demands. Having predictive models available allows for planning and allows police to direct scant resources to an area before minor quality of life issues become chronic disorder problems, before they reach the "tipping point". If police can predict the movement of crime, they have the ability to plan with the community ways to prevent the destabilization of its neighborhoods. This gives police departments the ability to develop long range plans based not on conjecture or parochial interests but on solid information. It allows the police to apply a business model of forecasts and projections within policing and gives them the ability to project budgetary needs several years in advance. Currently, departments have little idea if the resources they are requesting will be sufficient – projections are usually based on past needs or information. Using predictive models, budget projections can be based on analytic data and not on mere conjecture.

References

Gabor, T. 1990. Crime displacement and situational prevention: toward the development of some principles. *Canadian Journal of Criminology* 32: 41-74.

Hesseling, R. 1994. Displacement: a review of the empirical literature. In *Crime and place: crime prevention studies, Volume 3*. Ed. R.V. Clarke. Monsey, NY: Willow Tree Press.

Sherman, L.W. 1990. Police crackdowns: Initial and residual deterrence. In *Crime and justice: a review of research*, eds. Tonry, M. & Morris, N., Chicago: Chicago University Press.

Weisburd, D. and Green, L. 1995. Policing drug hot-spots: the Jersey City drug market analysis experiment. *Justice Quarterly* 12: 711-35.

1. A Statistical Method for the Detection of Geographic Clustering

The purpose of this chapter is to describe methods that were developed to assess the significance of geographic clustering (with crime analysis as an intended application). These methods assess the maximum of a smoothed map of crime rates. A version of this chapter has been published in Rogerson (2002). The published version contains an illustration of the method. In addition, application of the method can be found in Rogerson (2003).

Kernel-based, smoothed estimates of spatial variables are useful in exploratory analyses because they yield a clear visual image of geographic variability in the underlying variable. In this chapter we suggest an approach for assessing the significance of peaks in the surface that results from the application of the smoothing kernel. The approach may also be thought of as a method for assessing the maximum among a set of suitably defined local statistics. Local statistics for data on a regular grid of cells are first defined by using a Gaussian kernel. Results from integral geometry are then used to find the probability that the maximum local statistic exceeds a given critical value. Approximations are provided that make implementation of the approach straightforward. For application of these methods to problems in crime analysis, see Rogerson (2003).

1.1. Introduction

A common problem in the study of geographic patterns is to determine whether there are local subregions that exhibit significantly high (or low) values on some variable of interest. Detecting areas of heightened criminal activity or disease incidence represent but two examples of such problems. Bailey and Gatrell (1995) and others have described the use of kernel-smoothing as one way to represent the spatial variability in the mean of a variable of interest. The value at any particular location is taken to be a weighted function of the values in the neighborhood of the location, with closer locations receiving higher weights. The result is a surface portraying the regional variation in the underlying value, smoothed enough to eliminate the roughness of the image that would result if the original data were used, but not so much that underlying geographic variability is eliminated. Although these images represent a useful visual way to explore data, one often wishes to assess the significance of peaks in the surface. Attempts at such hypothesis testing have been limited to Monte Carlo simulation (e.g., Kelsall and Diggle 1995) or to more formal statistical methods that do not control properly for the likelihood of a Type I error (e.g., Bowman and Azzalini 1997).

A second set of approaches to finding geographic clusters includes methods for scanning the study area to find subregions with atypical values. Openshaw's (1987) Geographical Analysis Machine (GAM), Kulldorff and Nagarwalla's (1994) spatial scan statistic, and the related methods of Turnbull et al. (1990), Fotheringham and Zhan (1996), and Besag and Newell (1991) all use (though are not necessarily confined to) circular scanning windows to search for subregions that contain regional values that would not have been expected to occur by chance. Some of these methods correct for the fact that multiple tests are being carried out (e.g, Kulldorff's scan statistic), while others do not (e.g, Openshaw's GAM). When multiple testing is accounted for, the significance of the most extreme result is evaluated using Monte Carlo simulation.

Finally, local statistics such as those developed by Getis and Ord (1992) and by Anselin (1995) may also be used to pick out local regions with values that are significantly higher or lower than expected. They are defined for individual regions as a function of the value of the variable in that region, and values of the variable in nearby regions. Local statistics are designed primarily for testing hypotheses of spatial association for particular localities; the issue of multiple testing arises when one wishes to test more than one local statistic for significance. The problems result from the correlation among tests of local statistics that are near to one another in space.

We developed a statistical method for the detection of geographic clustering that is based upon Worsley's (1996) work on the maxima of Gaussian random fields. The method provides a way to assess the significance of the maximum of a set of local statistics. It also may be viewed as a method that allows for the assessment of the statistical significance of a kernel-based, smoothed surface. Finally, the method is similar in concept to scan statistics, since like Kulldorff's scan statistic, it considers many possible subregions and evaluates the statistical significance of the most extreme value. In addition, the method yields a calculable critical value that may be derived without resorting to Monte Carlo simulation methods. A key question concerns the adequacy of approximating actual crime distributions with a Gaussian random field. For areal data comprised of observed and expected crime frequencies, some transformation will often be necessary to make the Gaussian assumption reasonable.

We focus on the specific case of regional values that are normally distributed, and then smoothed with a Gaussian kernel to create local statistics that are similar to the Getis-Ord G^* statistic (Getis and Ord 1992, 1996). Although crime incidence data rarely have a normal distribution, it is often possible to transform the data so that it does, approximately, satisfy this assumption. A partial justification for focusing upon a Gaussian kernel comes from the work of Siegmund and Worsley (1995), who suggest that box-shaped kernels are relatively less efficient at finding Gaussian-shaped clusters than are Gaussian-shaped kernels at finding box-shaped clusters.

The primary goal is to assess the significance of the maximum on a smoothed map of regional values (or, alternatively, the maximum on a map of suitably defined local statistics), where the underlying data are normal, and where a Gaussian kernel has been used to smooth individual observations. The steps involved may be foreshadowed as follows:

1. Construct local statistics z_i for each region using the standardized, original region-specific observations (denoted y_i), and weights defined below in Equation 1.3 as $w_{ij} = (\sqrt{\pi}\sigma)^{-1} \exp(-d_{ij}^{\ 2} / 2\sigma^2)$, where d_{ij} is the distance from cell i to cell j (for example, the distance from centroid to centroid) and σ is chosen as the standard deviation of a normal distribution that matches the size of the hypothesized cluster. Then define $z_i = \sum_j w_{ij} y_j$, assuming the subregions consist of a regular lattice of square cells, complete with a guard area defined at the edges of the study area. More generally, for either irregular subregions or regular grid cells near the edge of the study region when a guard area has not been defined, one should use $z_i = \sum_j w_{ij} y_j / \sqrt{\sum_j w_{ij}^2}$.

2. Find the critical value z^* such that $p(\max z_i > z^*) = \alpha$ by using that value of z^* that leaves probability $(1+.81\sigma^2)\alpha / A$ in the tail of the standard normal distribution, where the study region that has been subdivided into a grid of A square cells, each having side of unit length. Alternatively, z^* may be approximated by

$$z^* = \sqrt{-\sqrt{\pi} \ln\left(\frac{4\alpha(1+.81\sigma^2)}{A}\right)}.$$

The details can be found in section 1.5 of this chapter, and in Rogerson (2001).

1.2. The Geometry of Random Fields

Critical values for the maximum among a set of local statistics may be based upon the geometry of random fields. Let $\mathbf{x}$ be a location in *d*-dimensional space, and let $\mathbf{Y}(\mathbf{x})$ denote a random multivariate value observed at **x**. A random field is defined by the set of values $\mathbf{Y}(\mathbf{x})$ for some subset of interest within the *d*-dimensional space (Cressie 1993). Here we will confine our attention to univariate random fields in *d*=2 dimensions, though results are also available for cases where the number of dimensions is other than two. We will also pay particular attention to the special case of a Gaussian random field, where the values at each location are taken from a Gaussian distribution. Results for other types of random fields, including $\chi^2, t,$ and *F* fields are also available (see, for example Worsley 1994).

Recent developments have improved upon and generalized the pioneering work of Adler (1981), who derived an approximation for the probability that the maximum of a Gaussian random field would exceed a specified value. In particular, Worsley (1994) has used principles of integral geometry to derive the following, improved version of Adler's original expression for exceedance probabilities. In two dimensions, for the case where independent observations are observed at many points on a lattice, and then smoothed using a Gaussian kernel it is:

$$p(\max_i z_i > z^*) = \frac{Az^*\varphi(z^*)}{4\pi\sigma^2} + \frac{D\varphi(z^*)}{\sqrt{\pi}\sigma} + [1 - \Phi(z^*)] \qquad (1.1)$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the probability density and cumulative distribution function of a standard normal variate. *D* denotes the caliper diameter and *A* the area of the study region. The caliper diameter is the average of the diameter as measured through all rotations of the study area. For a rectangle with sides *a* and *b*, the caliper diameter is $(a+b)/2$; for a circular study region of radius *r*, the caliper diameter is equal to $2r$. Again, Equation 1.1 gives the probability that the maximum of a Gaussian random field, when smoothed by a Gaussian kernel, exceeds z^*.

The primary purpose of this chapter is to illustrate the use of these new results in problems involving the maximum among a set of particular local statistics (or alternatively, the maximum of a kernel-based surface). The reader interested in either more general results or the geometrical principles that form the foundation of the methods should consult the references.

1.3. Illustration

The method described above can be illustrated as follows. A 30x30 grid was filled with *y*-values generated from a normal distribution with mean 0 and variance 1. For this simulation of the null hypothesis of no local cluster, values of σ=1, 2 and 3 were used with the Gaussian kernel to smooth the initial *y* values, creating in the process a 30x30 grid of local statistics, z_i. To avoid edge effects, the 22x22 grid occupying the center of the 30x30 grid was searched for the maximum z_i value.

Using the values of A=484, D=22 in Equation 1.2, and setting the left-hand side equal to α=0.05 yields critical *z*-values of 3.779, 3.389, and 3.150 for the cases where $\sigma = 1$, 2, and 3, respectively. For comparison purposes, the 95th percentiles for the critical *z*-values were then found from 1,000 Monte Carlo simulations.

Results are shown in the first three columns of Table 1.1, along with the Type I error probabilities associated with using the critical value derived from Equation 1.1. Although the Type I error probabilities are close to their nominal value of 0.05 for the latter two cases, use of

Equation 1.2 would be overly conservative in the case where $\sigma = 1$. In fact, when $\sigma = 1$, Equation 1.2 is even more conservative than a Bonferroni adjustment (where the critical value of *z* is chosen using α / n instead of α, because there are *n* separate tests being carried out). With *n*=484 cells, the critical z-value following a Bonferroni adjustment is found to be 3.711 (using $.05/484$ as the area in the tail of the normal distribution).

Equation 1.2 is based upon the assumption of a continuous random field, and its poor performance for the case $\sigma = 1$ is due to the discreteness of the grid used to generate the observed values.

1.4. Approximation for discreteness of observations

Adjusted critical values may be found by first determining the amount of smoothing implicit in the initial discrete grid, which represents a set of aggregated or smoothed observations. We can represent the initial data as a smoothed Gaussian field in the following way. With *n*=484, the *z*-value associated with a Bonferroni adjustment is 3.711. By using *z*=3.711 in Equation (1.1) we can solve for the amount of smoothing that is imparted by the square grid ($\sigma = \sigma_0$). In particular, (1) may be rearranged so that σ_0 is the solution to the following quadratic equation:

$$(\alpha - 1 + \Phi(z))\sigma_0^{\,2} - \frac{D\varphi(z)}{\sqrt{\pi}}\sigma_0 - \frac{Az\varphi(z)}{4\pi} = 0. \qquad (1.5)$$

In our example, *A*=484, *D*=22, α =0.05, *z*=3.711, and solving for σ_0 yields σ_0=1.133. In section 6, an argument is presented that suggests that for most problems, this step is not necessary, and the value of σ_0 may always be taken as 10/9 = 1.111.

The total amount of smoothing in each case (σ_t) is then defined by combining the implied initial smoothing brought about by the discreteness of the grid (σ_0) with the smoothness chosen for the Gaussian kernel in defining the local statistics (say σ_l). Thus

$$\sigma_t = \sqrt{\sigma_0^2 + \sigma_l^2}$$

These choices imply respective σ_t values of 1.495, 2.288, and 3.199 for the cases where $\sigma_l = 1$, 2, and 3, respectively. Using these values in Equation (1.1) and setting the left-hand side equal to α =0.05 results in the critical values of z^* =3.556, 3.311, and 3.112 in the σ =1, 2, and 3 cases, respectively. These values are shown in column 4 of Table 1.1; their associated *p*-values are in close agreement with the nominal value of 0.05.

1.5. Approximations for the exceedance probability

Of the terms on the right-hand side of (1.1), the first term contributes most to the *p*-value on the left-hand side. Table 1.2 reveals the contributions of each term on the right-hand side of (1.1) to the nominal Type I error probability of 0.05 for the illustration above, which includes a correction for the discrete number of spatial units.

The table shows that the first term is by far the most important, and in each case the two-term approximation

$$p(\max z > z^*) \approx \frac{Az^*\varphi(z^*)}{4\pi\sigma^2} + \frac{D\varphi(z^*)}{\sqrt{\pi}\sigma} = \frac{\varphi(z^*)(Az^* + 4\sqrt{\pi}D\sigma)}{4\pi\sigma^2} \tag{1.6}$$

should suffice (since the sum of the first two terms is close to 0.05). When the amount of smoothing imparted by the kernel is sufficiently small, the one-term approximation

$$p(\max z > z^*) \approx \frac{Az^*\varphi(z^*)}{4\pi\sigma^2} \tag{1.7}$$

will be adequate.

The use of the approximations (1.6) and (1.7) will result in critical values of z^* that are slightly lower than those derived with the full, three-term expression in (1.1). Table 1.3 reveals that if $\sigma/\sqrt{A}$ is greater than about 0.05 or 0.10, the one-term approximation in (1.7) will be too liberal, and should be abandoned in favor of the approximation in (1.6).

1.5.1 An approach based on the effective number of independent resels

These approximations still require numerical solution for the desired critical value, z^*, and it is of interest to ask whether a simpler solution for z^* is possible. One possibility is to attempt an estimate of the effective number of spatial units or *resels*, upon which to base a Bonferroni adjustment[2]. The greater the amount of smoothing, the less accurate will be the Bonferroni adjustment which is based upon all n cells, and hence we seek a value for the number of resels, r, that will be less than n. Let us take

$$r = \frac{A}{(m\sigma)^2} \tag{1.8}$$

where m is an empirical constant of proportionality. For a grid of square cells, the idea here is to divide the study area into a number of resels that is directly proportional to the number of cells ($n = A$), and inversely proportional to the amount of smoothing, as measured by the variance of the Gaussian kernel.

A simple Bonferroni adjustment based on r turns out to be possible only because the value of m is approximately constant throughout a wide range of $\sigma/\sqrt{A}$ values. To illustrate, the value of m satisfying

$$\Phi^{-1}(1-\alpha/r) = \Phi^{-1}(1-\alpha(m\sigma)^2/A) = z^*$$

was determined for each of the rows in Table 1.4, using $\alpha = 0.05$, and using the values of $\sigma / \sqrt{A}$ and z^* given in each row (where the value of z^* is that determined from Equation (1.1). Table 1.4 shows that the value of *m*, as a function of $\sigma / \sqrt{A}$, is relatively flat over much of its range. This suggests that a very good approximation for z^* may be found by taking *m*=0.9, and therefore setting the number of resels equal to $r = A/(.9\sigma)^2$. A Bonferroni adjustment based upon this number of resels then yields the desired critical value, z^*:

$$z^* \approx \Phi^{-1}(1 - \frac{\alpha(.9\sigma)^2}{A}) \tag{1.9}$$

Thus one can determine the approximate critical value by finding the *z*-value that leaves $\alpha(.9\sigma)^2 / A$ in the tail of the standard normal distribution. Since this may require the use of a detailed *z*-table that provides areas for relatively high *z*-values, it is also of interest to find an approximation that does not require the use of such a detailed *z*-table. Using tight bounds for the cumulative distribution function of a normal variable (Sasvari 1999), z^* may also be approximated by

$$z^* \approx \sqrt{-\sqrt{\pi}\ln(\frac{4\alpha}{r})} = \sqrt{-\sqrt{\pi}\ln(\frac{4\alpha(1+.81\sigma^2)}{A})} \tag{1.10}$$

Note from Table 1.4 that as $\sigma / \sqrt{A}$ declines below about 0.02, this approximation will not work as well, since the value of *m* begins to decline away from 0.9. However, as column 5 of Table 1.3 shows, the use of *m*=0.9 when $\sigma / \sqrt{A}$ is as small as 0.01 results in critical values that are only slightly liberal.

Columns 5 and 6 of Table 1.1 demonstrate the adequacy of the approximations given by Equations 1.9 and 1.10 for the case of the maximum local statistic observed on the 22x22 grid.

The value of m=0.9 suggests that we can use $\sigma_0 = 1/0.9 = 1.111$ as a measure of the smoothing implicit in the discrete grid. This is because when there is no additional smoothing, (i.e., local statistics are based only on the value in the local cell, and weights associated with surrounding cells are zero), the local statistics in each cell are independent, and $r = A$. Using this in Equation 1.8 with m=0.9 requires a standard deviation of $\sigma_t = \sigma_0 = 1/0.9 = 1.111$.

If the amount of smoothing (σ) is greater than or equal to one, the fact that the resel approach can be used when $\sigma / \sqrt{A} > .01$ implies that the approach may be adopted when $A <$ 10,000. If the grid is finer than 100x100 = 10,000 or $\sigma < 1$, then one should ensure that the total amount of smoothing yields $\sigma_t / \sqrt{A} > 0.01$ before proceeding with the critical z-value based on resels.

1.6. Discussion

In this chapter, we have considered a local statistic based upon a Gaussian kernel. The statistic has the desirable feature that one may easily find the critical value (via Equation 1.9 or 1.10) necessary for testing the significance of the maximum of the local statistics defined over a study area. The statistic relies on the assumption that the underlying data come from a normal distribution. In choosing kernels for smoothing, it is commonly noted that the choice of a kernel function is much less important than choosing the bandwidth. Since estimates are relatively robust with respect to the form of the kernel function, the Gaussian kernel is a good choice when one is interested in assessing the significance of maxima, since it lends itself readily to such testing. In addition, one should be aware that the choice of bandwidth should be made to match the hypothesized cluster size; in the different context of optimal estimation of kernel surfaces, bandwidth choice could be quite different.

The distribution of local statistics is also affected by global spatial autocorrelation. In particular, the presence of global spatial autocorrelation will make it more difficult to detect significant local statistics. Recent developments in the study of random fields (e.g., Worsley et al. 1999) suggest that the approach described above might also be modified to find an approximate critical value associated with the maximum local statistic in the presence of global spatial autocorrelation.

There are situations where one may be interested in trying different amounts of smoothing (i.e., choose various values for σ) to see at which scale local statistics are most significant. Kulldorff's spatial scan statistic handles this case using a Monte Carlo approach. Siegmund and Worsley (1995) provide details on how critical values may be derived analytically when one wishes to test a range of σ values.

Finally, this chapter has focused upon the development of the method; applications of the ideas summarized here to problems in crime analysis may be found in Rogerson (2003). An S-Plus computer program for carrying out the approach outlined here (in the context of an application to disease clusters) is available in Han and Rogerson (2003).

Footnotes

[1]Note that this kernel is a scaled version of the more commonly used Gaussian kernel

$$k_2(\mathbf{x}'\mathbf{x}) = \frac{1}{2\pi}\exp(-\mathbf{x}'\mathbf{x}/2)$$

With the first kernel, k_1, we have

$$\int_{x_2=-\infty}^{\infty}\int_{x_1=-\infty}^{\infty} k_1(\mathbf{x}'\mathbf{x})dx_1dx_2 = 2\sqrt{\pi},$$

which is *not* equal to the more usual

$$\int_{x_2=-\infty}^{\infty}\int_{x_1=-\infty}^{\infty} k_2(\mathbf{x}'\mathbf{x})dx_1dx_2 = 1.$$

We use $k_1(\cdot)$ rather than $k_2(\cdot)$ to satisfy the Siegmund and Worsley condition -- that is,

$$\int_{x_2=-\infty}^{\infty}\int_{x_1=-\infty}^{\infty} [k_1(\mathbf{x}'\mathbf{x})]^2 dx_1dx_2 = 1.$$

[2]This definition of resels is similar in concept, though different in detail, when compared with that used by Worsley et al. (1992).

Appendix: The Gaussian kernel

For the Gaussian random field, the variable of interest, Y, is defined at each location in space, $\mathbf{x}$, and at each location, $Y(\mathbf{x})$ has a normal distribution. In practice we will have observed values for a finite number of points in space (for example, observed as grid points on a regular lattice). Following Siegmund and Worsley (1995), we will define the kernel, $k(\cdot)$, such that it is a square integrable function, that, without loss of generality, satisfies

$$k(t)^2\,dt = 1.$$

Furthermore, we will focus on the special case of the Gaussian kernel in two dimensions:

$$k_1(\mathbf{x}'\mathbf{x}) = \pi^{-1/2}\exp(-\mathbf{x}'\mathbf{x}/2),$$

where $\mathbf{x}' = \{x_1\ x_2\}$ represents the two-vector containing the coordinates of location x.

In practice, the local statistic at location $\mathbf{x}_i$, $z(\mathbf{x}_i)$, is a weighted sum of the variable values at other locations, with the weights equal to the kernel value:

$$z(\mathbf{x}_i) = \sum_{j=1}^{n} \frac{1}{\sigma} k_1\left(\frac{(\mathbf{x}_j - \mathbf{x}_i)'(\mathbf{x}_j - \mathbf{x}_i)}{\sigma^2}\right) y_j = \sum_{j=1}^{n} \frac{1}{\sqrt{\pi}\sigma} \exp[(-(\mathbf{x}_j - \mathbf{x}_i)'(\mathbf{x}_j - \mathbf{x}_i)/2\sigma^2] y_j \qquad (1.2)$$

where $(\mathbf{x}_j - \mathbf{x}_i)'(\mathbf{x}_j - \mathbf{x}_i)$ is the squared distance from point i to point j, and σ is the bandwidth (and standard deviation) of the Gaussian kernel, k_1.[1]

The definition used for k_1 has the desirable consequence of making the smoothed estimates/local statistics, $z(\mathbf{x}_i)$, standard normal variables, when the original variables, Y, are also expressed as standard normal variables. Since $z(\mathbf{x}_i)$ will have a standard normal distribution, it may be tested for statistical significance. To see this, recognize first that the local statistic is a weighted sum of the other observations:

$$z(\mathbf{x}_i) = \sum_{j} \frac{1}{\sqrt{\pi}\sigma} e^{-d_{ij}^2/2\sigma^2} y_j = \sum_{j} w_{ij} y_j. \qquad (1.3)$$

If $Y \sim N(0,1)$, the distribution of $z(\mathbf{x}_i)$ is normal and its expectation is

$$\mathrm{E}[z(\mathbf{x}_i)] = \sum_j w_{ij}\mathrm{E}[y_j] = 0.$$

The variance of $z(\mathbf{x}_i)$ is

$$\mathrm{V}[z(\mathbf{x}_i)] = \sum_j w_{ij}^2 \mathrm{V}[y_j] = \sum_j w_{ij}^2.$$

With the definitions of z_1 and $k_1(\cdot)$ the sum of the squared weights is equal to one.and so the variance of $z(\mathbf{x}_i)$ is approximately equal to one. This approximation does not hold for cells near the edge of study regions consisting of regular cells, nor does it hold for irregular lattices. A more general definition that can be used for these cases is

$$z_i(\mathbf{x}_i) = \frac{\sum_j w_{ij} y_j}{\sqrt{\sum_j w_{ij}^2}}. \qquad (1.4)$$

The term in the denominator ensures that the variance of the resulting local statistic, z_i, will be equal to one.

References

Adler, R.J. 1981. *The geometry of random fields*. New York: Wiley.

Anselin, L. 1995. Local indicators of spatial analysis -- LISA. *Geographical Analysis* 27: 93-115.

Bailey, A. and Gatrell, A. 1995. *Interactive spatial data analysis*. London: Longman.

Besag, J. and J. Newell 1991. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society Series A* 154: 143-55.

Bowman, A.W. and Azzalini, A. 1997. *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*. Oxford: Clarendon Press.

Cressie, N. 1993. *Statistics for spatial data.* New York: Wiley.

Fotheringham, A. S. and F. B. Zhan 1996. A comparison of three exploratory methods for cluster detection in spatial point patterns. *Geographical Analysis* 28: 200-18.

Getis, A. and Ord, J. K. 1992. The analysis of spatial association by the use of distance statistics. *Geographical Analysis* 24: 190-206.

Getis, A. and Ord, J. K. 1996. Local spatial statistics: an overview. In *Spatial analysis: modelling in a GIS environment*. Eds. P. Longley and M. Batty. Cambridge: Geoinformation International, pp. 261-77.

Han, D. and Rogerson, P. Application of a GIS-based statistical method to assess spatio-temporal changes in breast cancer clustering in the northeastern United States. In Geographic information systems and health applications. Eds. O. Khan and R. Skinner. London: Idea Group Publishing.

Kelsall, J. and Diggle, P. 1995. Non-parametric estimation of spatial variation in relative risk. *Statistics in Medicine* 14: 2335-42.

Kulldorff, M. and N. Nagarwalla 1994. Spatial disease clusters: detection and inference. *Statistics in Medicine* 14: 799-810.

Mercer, W.B. and Hall, A.D. 1911. The experimental error of field trials. *Journal of Agricultural Science* 4: 107-32.

Openshaw, S., Charlton, M., Wymer, C., and Craft, A. 1987. A mark 1 geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems* 1: 335-58.

Rogerson, P. 2001. A statistical method for the detection of geographic clustering. *Geographical Analysis* 33: 215-27.

Rogerson, P. 2003. The Application of New Spatial Statistical Methods to the Detection of Geographical Patterns of Crime. *GIS and Crime Analysis* Eds. G. Clarke and J. Stillwell. Forthcoming.

Sasvari, Z. 1999. Tight bounds for the normal distribution. *American Mathematical Monthly* 106: 76.

Siegmund, D. O. and Worsley, K.J. 1995. Approximate tail probabilities for the maxima of some random fields. *Annals of Statistics* 23: 608-39.

Turnbull, B., Iwano, EJ, Burnett, WS, Howe, HL, and Clark, LC. 1990. Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *American Journal of Epidemiology* 132: S136-43.

Worsley, K.J., Evans, A.C., Marrett, S., and Neelin, P. 1992. A three dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism* 12: 900-18.

Worsley, K.J. 1994. Local maxima and the expected Euler characteristic of excursion sets of χ^2, F, and t fields. *Advances in Applied Probability* 26: 13-42.

Worsley, K.J. 1996. The geometry of random images. *Chance* 9,1: 27-40.

Worsley, K.J., Andermann, M., Koulis, T., MacDonald, D., and Evans, A.C. 1999. Detecting changes in non-isotropic images. Unpublished manuscript.

(1)	(2)	(3)	(4)		(5)	(6)
σ	Simulated 95th percentile	Critical value (Eq. 1.2)	Critical value adjusted for discreteness		Eq. 1.9	Eq. 1.10
1	3.575	3.779 (.029)	3.556	(.052)	3.556	3.572
2	3.342	3.389 (.043)	3.311	(.053)	3.328	3.354
3	3.110	3.150 (.041)	3.112	(.050)	3.136	3.172

Table 1.1. Simulated and approximate critical values ($\alpha = 0.05$) for the maximum local statistic when $n = A = 484$.

σ_t	first term	second term	third term
1.495	.0439	.0059	.0002
2.288	.0405	.0090	.0005
3.199	.0369	.0122	.0009

Table 1.2. Contribution of terms in Equation 1.2 to the Type I error probability

(1)	(2)	(3)	(4)	(5)
$\sigma / \sqrt{A}$	z^* (Eq. 1.2)	z^* (Eq. 1.3) two-term approximation	z^* (Eq. 1.4) one-term approximation	Resels m=0.9
0.01	4.535	4.535 (.050)	4.532 (.051)	4.463 (.068)
0.02	4.205	4.205 (.050)	4.196 (.052)	4.160 (.060)
0.05	3.727	3.727 (.050)	3.700 (.055)	3.716 (.052)
0.10	3.334	3.331 (.050)	3.267 (.061)	3.349 (.047)
0.15	3.094	3.087 (.050)	2.977 (.069)	3.118 (.047)
0.20	2.922	2.909 (.052)	2.748 (.079)	2.944 (.047)
0.25	2.790	2.768 (.053)	2.552 (.090)	2.803 (.048)

Table 1.3. Approximations for the critical value z^*

$\sigma / \sqrt{A}$	z*	*m*
.01	4.535	0.76
.02	4.205	0.81
.05	3.727	0.88
.10	3.334	0.92
.15	3.094	0.94
.20	2.922	0.93
.25	2.790	0.92
.30	2.683	0.90
.35	2.596	0.88
.40	2.523	0.85

Table 1.4. The relative flatness of *m* as a function of $\sigma / \sqrt{A}$.

2. Spatial Monitoring of Geographic Patterns: An Application to Crime Analysis

This chapter describes a new procedure for detecting changes over time in the spatial pattern of point events, combining the nearest neighbor index and cumulative sum methods. The method results in the rapid detection of deviations from expected geographic patterns. It may also be used for various subregions and may be implemented using time windows of differing length to search for any changes in spatial pattern that may occur at particular time scales. The method is illustrated using 1996 arson data from the Buffalo, NY police department. A published version of this account is available in Rogerson and Sun (2002).

2.1 Introduction

Statistical methods for detecting clusters in spatial point patterns are almost always applied retrospectively, in the sense that the statistical test is applied at a single, given point in time using observed (and possibly aggregate) data on point locations. In many situations, it is desirable to carry out such tests repeatedly as new point location data are collected, with the objective of detecting change as quickly as possible. For example, it is of interest to detect changes in the spatial pattern of disease rapidly (Farrington and Beale 1998; Rogerson 1997). This interest is part of a more general, longstanding interest in the monitoring of public health (see, e.g., Chen 1978). Monitoring the residential locations of new customers is important for businesses to assess their markets and competition. Quick detection of changes in the pattern of criminal activity may lead to improved allocation of police resources.

Standard methods of point pattern analysis are not applicable to these problems, and new methods are required for the rapid detection of changes in spatial patterns. In this chapter we develop and evaluate a method based on the synthesis of the nearest neighbor index with the cumulative sum methods used in industrial process control. Of course many alternatives to the nearest neighbor index are available for the detection of clusters in spatial point patterns, some of which have been used specifically in the area of crime analysis (see, e.g., Ripley 1976; Openshaw et al. 1987; Block 1995, among a large list of others). Some of these focus explicitly upon space-time interactions, such as the Knox method (1964) and Kulldorff's space-time scan statistic (2001). The choice of the nearest neighbor index is based upon its simplicity and its common use in crime analysis, and the monitoring methods presented are general enough that they may be adapted to other statistics and methods aimed at cluster detection.

Section 2.2 provides a brief summary of the nearest neighbor index and cumulative sum methods. Section 2.3 suggests how these methods may be combined, and provides illustrative examples for point patterns that are simulated in the unit square. Finally, section 2.4 applies the method to 1996 crime data from the Buffalo, NY Police Department.

2.2 Background

This section provides a brief review of the nearest neighbor index and cumulative sum methods.

2.2.1. *Nearest-neighbor statistic*

The nearest neighbor index (Clark and Evans 1954) compares the observed mean of the distances between points and their nearest neighbors with the distance *expected* between nearest neighbors in a random pattern. The nearest neighbor index, R, is the ratio of the observed to the expected distance. The expected distance is given by

$$r_e = \frac{1}{2\sqrt{\rho}}, \qquad (2.1)$$

where ρ=*N*/*A* is the density of points and is equal to the number of points (*N*) divided by the size of the study area (*A*). Thus

$$R = \frac{\overline{r}_{obs}}{r_e} \qquad (2.2)$$

R values less than 1 indicate clustering, since the observed mean distance between neighboring points is less than that expected in a random pattern. The minimum value of *R* is zero, which occurs when all points are at a single location. The theoretical maximum of *R* is 2.149, which occurs when points are maximally dispersed in the plane. The standard deviation of the mean distance between nearest neighbors in a random pattern is

$$\sigma_{r_e} = \frac{0.26}{\sqrt{N\rho}}. \qquad (2.3)$$

This allows the use of a statistical test using the quantity

$$z = \frac{r_o - r_e}{\sigma_{r_e}}, \qquad (2.4)$$

where r_o is the observed mean distance between nearest neighbors. Under the null hypothesis of a random point pattern, *z* has, approximately, a standard normal distribution. An observed *z*-score that is less than the critical value of *z* would lead to rejection of the null hypothesis in favor of the conclusion that significant clustering exists. Users need to be aware that the statistic can depend upon the shape of the study area -- highly rectangular areas produce relatively low values of *R* since randomly located points are more likely to be close to their neighbors. Also, since only the nearest neighbor (and not, for example, second- and third-nearest neighbors) is considered, detection of clustering is limited to clustering that occurs on relatively small spatial scales. For additional discussion of the nearest neighbor index, see, for example, King (1969), Griffith and Amrhein (1991), and Bailey and Gatrell (1995). The nearest neighbor statistic is but one of a large number of methods for looking at spatial clustering. References to other methods, as used in the context of crime analysis, are available in the references to the CrimeStat manual (http://www.icpsr.umich.edu/NACJD/crimestat/CrimeStatReferences.pdf).

2.2.2. *Cumulative sum methods*

Cumulative sum (or cusum) methods are designed to detect changes in the mean value of a quantity of interest (see, for example, Ryan 1989; Wetherill and Brown 1991; Montgomery 1996). These methods are widely used in industrial process control to monitor the quality of production characteristics. They rely upon the assumption that the quantity being monitored is a normally distributed variable that exhibits no serial autocorrelation. Without loss of generality, let the variable be converted to a *z*-score with mean 0 and variance 1. Then the cumulative sum, following observation *t*, is defined as

$$S_t = \max(0, S_{t-1} + z - k), \qquad (2.5)$$

where *k* is a parameter. A change in mean is signaled if $S_t > h$, where *h* is another parameter to be defined.

Thus values of *z* in excess of *k* are cumulated. The parameter *k* in this instance, where a standardized variable is being monitored, is often chosen to be equal to ½; in the more general case, *k* is often chosen to be equal to ½ the standard deviation associated with the variable being monitored. The parameter *h* is chosen in conjunction with an acceptable rate of “false alarms”; high values of *h* will lead to a low probability of a false alarm, but also a lower probability of detecting a real change. Table 2.1 depicts the values of *h* associated with given average times until a false alarm. These times are called the "in-control" average run length, and are designated by the notation ARL_0. When k=1/2, an approximation for the in-control average run length (ARL_0) may be derived from:

$$\mathrm{ARL}_0 = 2(e^a - a - 1), \tag{2.6}$$

where a=h+1.166. One can make practical use of this approximation to choose the parameter h. To do so, one first decides upon a value of ARL_0, and then solves the approximation for the corresponding value of h. In the more general situation where a non-standardized variable is being monitored, the critical value of the cusum is determined by multiplying the value of h by the standard deviation of the variable being monitored.

The value of k is often set equal to 1/2 because this choice tends to minimize the average out-of -control run length (that is, the time until a signal of change is sent when a real pattern change has occurred) for a given value of ARL_0.

2.3. Monitoring changes in point patterns

There are at least two reasons why it is not desirable to repeat statistical tests that use the nearest neighbor index. First, one must account for the fact that an adjustment should be made for the number of tests being carried out. Consider the following simulation. Fifty points were successively located in the unit square. Following the location of each point (beginning with the second point, since a single point can not be thought of as a cluster), a nearest neighbor index was calculated and a z-statistic computed using the means and standard deviations given in Table 2.2.[1] This z-score was then compared with the critical value of z = -1.96 (corresponding to a one-tailed test with α = 0.025, or a two-tailed test with α = 0.05). In 23% of the 10,000 replications, the null hypothesis of no clustering was rejected before 50 points had been located in the square. This high percentage is due to the fact that more than one hypothesis has been tested.

An adjustment may be used to account for the fact that 49 separate tests are being conducted; such an adjustment uses the fact that we want the probability that no significant result has been found after all 49 tests have been carried out to be equal to, say, 0.975 (i.e., $(1-x)^{49}=.975$), where x is the probability of rejection in a single test. In this case, solving for x yields $x=0.00052$, and the corresponding critical value of z is

$z = -3.28$. This adjustment is conservative (since the separate tests are not actually independent); in 10,000 simulations of the null scenario of fifty randomly located points in the unit square, only 0.6% of the time was the null hypothesis rejected (compared to the nominal value of 2.5%). Less conservative adjustments that account for the correlation between tests are not straightforward to derive.

Perhaps more importantly, there is a great deal of "inertia" in the nearest neighbor index when it is calculated repeatedly, after each new point has been located. If points begin to cluster, the nearest neighbor index may not decline quickly, since it will always be based upon an *average* of the distances to *all* nearest neighbors, and not just the distance to nearest neighbors for the most recent points. Thus it may take a long time for changes to appear in the statistic.

2.3.1. *A cusum approach for the nearest neighbor index*

Here we combine the nearest neighbor and cumulative sum methods as follows. At each stage in the evolution of an observed point pattern (e.g., when t-1 points have been observed to date), we locate a point at random on the map, and the distance from this point to its nearest neighbor is calculated. This is repeated a large number of times, and the mean ($\bar{d}$) and variance

(σ_d^2) of the distances from the randomly located points to their nearest neighbors are found. Then a z-score is assigned to observation t as follows:

$$z = \frac{d_{\text{obs}} - \bar{d}}{\sigma_d}, \tag{2.7}$$

where d_{obs} is the observed distance from point t to its nearest neighbor. These quantities may then be cumulated in a cusum scheme. The cusum scheme described by Equation 2.5 would be used to detect departures from randomness in the direction of uniformity; such a scheme would signal a change when observed distances between neighbors began to exceed the distances expected in a random pattern. To detect departures from randomness in the direction of clustering, one would use

$$S_t = \max(0, S_{t-1} - z - k), \tag{2.8}$$

where S_t is the cumulative sum at time t, and k is a parameter usually set equal to ½, and more generally set equal to ½ the size of the change (in terms of standard deviational units) that one is trying to detect. Again, a signal of change in pattern is sent when S_t exceeds the thredhold parameter h.

Because distances to nearest neighbors do not follow a normal distribution, the assumption of normality, required by the cusum approach, is violated. That is, the z-values do not have a normal distribution. A solution is to aggregate successive, observations; we can define a new z-score, $z_{(b)}$ that is simply the average of b successive observations. The mean of $z_{(b)}$ will still be equal to zero, and the variance of $z_{(b)}$ is equal to $1/b$. We then replace z in Equation 2.8 with the quantity $(z_{(b)}-0)/(1/\sqrt{b})$. Usually the value of b can be quite small for the assumption of normality to be acceptable; for the simulations in the unit square, a batch size of b=3 was found to be acceptable.

Locating points on a map at random is questionable, since actual points will occur on a network. For a recent suggestion on locating points randomly on a network, along with subsequent point pattern analysis, see Okabe and Yamada (2001).

2.3.2 *Simulations of clustering in the unit square*

The simulation scenario described in the beginning of this section (where 50 points are located successively, at random, in a square having both x and y-coordinates ranging from 0 to 1) was modified to generate clustering as follows. After observation t=20, points were located randomly with x- and y-coordinates in the interval (0,0.25) with probability 0.2, and located randomly within the entire (0,1) square with probability 0.8. Sequential use of the nearest neighbor index resulted in detection of clustering in 54.3% of the 10,000 simulations on or before the 50th observation (in 40.3% of the simulations clusters were detected after observation 20). With the conservative adjustment for multiple testing, clusters were found on or before the 50th observation in only 10.1% of the simulations (in 9.8% of the simulations clusters were found after observation 20).

Using the combined cusum-nearest neighbor approach described in section 2.3.1, clusters were detected in 97.7% of the 10,000 simulations on or before the 50th observation. The mean observation number at which a clustering signal was received was 38.5 – a bit more than 18 observations after clustering began.

Note the substantial improvement in cluster detection in comparison with the sequential use of the nearest neighbor test. Sequential use of the nearest neighbor test is hampered by the inertia associated with the first twenty observations, which follow the null hypothesis of no clustering. Even after a change in process occurs after observation 20, the nearest neighbor index calculated after subsequent observations will contain information based upon the first twenty observations, and hence it declines only slowly.

2.4. Application to crime analysis and data from the Buffalo Police Department

Our focus is upon modifying a statistic which has been widely used in crime analysis (the nearest neighbor index) to detect *changes* in the spatial patterns of criminal activities. Crime analysts, in addition to being interested in the identification of existing crime hot spots, are also interested in methods that can quickly detect new, emerging "hot spots", so that policing efforts can be allocated more efficiently.

Data on the locations and times of 379 arsons were available for 1996 from the Buffalo Police Department (BPD). The data represent actual incidents; the data are to be distinguished from emergency calls for service (which would include false alarms) and arrest data. Figure 2.1 shows how the nearest neighbor index changes throughout the year for arsons. There appears to be a fairly steady decline in R for arsons during 1996. Because the statistic changes only slowly over time, we next turn to a cumulative sum approach in the next subsection to determine if and when significant changes take place in the underlying geographic pattern.

2.4.1 *Cumulative sum approach for 1996 arson data*

The cusum nearest neighbor approach described in section 2.3 was used with the 1996 BPD arson data. Presumably this approach will be more sensitive in identifying points in time where the pattern has changed, in comparison with the trends shown in Figure 2.1 (which has a great deal of embedded inertia).

The first 200 arsons of 1996 (occurring during the period from January to July) were used as a "base" pattern. For this base period, it was determined that the location of each successive point was an average of 0.216 times the distance from its nearest neighbor than was expected. This yields a baseline measure of clustering that exists in the population; arsons cluster during the base period because population is clustered and because crimes such as arsons may tend to occur more in some areas than in others. The reader should keep in mind that here we are interested in deviations from this baseline amount of clustering, and not in the detection of clustering itself. Recall that an original estimate of expected distance is determined by locating a point at random within the study area, computing the distance to its nearest neighbor, and then repeating this many times; we now wish to scale this expected distance downward in accordance with the observed clustering). Thus we use

$$z = \frac{d_{\text{obs}} - .216\bar{d}}{.216\sigma_d}, \tag{2.9}$$

in place of Equation (2.7). An alternative would be to use the baseline locations to estimate a kernel density estimate of arson occurrences. One could then sample from this as a way of generating points to estimate $\bar{d}$ and σ_d (see, e.g., Brunsdon 1995). The correlation between successive values of z (i.e., the correlation between z_t and z_{t+1}) was found to be insignificant, and thus the underlying assumption of no serial autocorrelation is satisfied. A more complete assessment would also include examination higher order correlations such as the correlation between z_t and z_{t+2}).

Surveillance of the pattern then began in August. Values of k=1/2 and h=4.12 were used. The value of h=4.12 was arrived at by using equation 2.2, after choosing a ARL_0 value of 380 (corresponding to one false clustering alarm per year). Figures 2.2 and 2.3 show that the cusum statistic becomes critically negative, indicating clustering relative to the base pattern, after about 80 observations (in October). Note from Figure 2.2 that the cusum statistic remains in the critical region for most of the remainder of the year. Changing the definition of the base period from the first 200 observations to either the first 100 or first 150 observations also leads to a cluster signal at this same time. This stability of the signalling with respect to changes in the base period provides reassuring evidence that results are not overly sensitive to minor changes in definition of the base period. The cusum statistic is commonly reset to zero following a signal (especially in industrial process control, where the change, often a defect, can be noted and the equipment or process appropriately modified). Figure 2.3 demonstrates that when the cusum is reset to zero, the signal of clustering is given two additional times before the end of the year, indicating that the cause of the increased clustering has persisted.

Figures 2.4 and 2.5 depict the spatial pattern of arsons during 1996. Figure 2.4 contains twelve black triangles, representing those 1996 arsons that occurred just prior to the first cluster signal. Figure 2.5 contains triangles, representing those 1996 arsons that followed the first cluster signal. The maps show that the triangles are nearer to neighboring arsons than the dots are to their neighbors -- arsons occurring later in the year were more likely to belong to clusters.

A natural question to ask is *why* the pattern has changed. One possibility is that there is seasonal variation in the pattern; in future work we intend to examine data from other years. Another possibility is that it was the base period that was unusual; perhaps the spatial pattern during the first half of 1996 was *less* clustered and more uniform and spread out than is usual. Again, study of data in adjacent years should help to shed additional light on this question.

In this example, surveillance took place across the entire study region, and changes were detected in the citywide pattern during October 1996. It is straightforward to modify this procedure when only a subregion is of interest. A GIS (in this case, Arcview 3.0 was used) may be used to select the set of points of interest, and to create a new coverage and corresponding table that contains information only on those arsons within the subregion.

It is important to note that other types of surveillance may also be desired. For example, we may wish to detect deviations from the base period that occur in the opposite direction of what we have been considering here -- namely, distances from new arsons to their nearest neighbors that are *greater* than expected. This would perhaps indicate that arsons were beginning to occur at new locations (which in turn might be the result of geographic displacement following an enforcement effort). Or we may wish to find periods of time where *recent* arsons are located nearer to one another in comparison with some base period. This latter example is treated in the next section.

2.4.2 *Surveillance using a moving window of observations*

One of the characteristics of the surveillance method as described to this point is that the nearest neighbor distance for a newly observed point is calculated as the minimum of the distance to *all previous observations*. For the case of surveillance of arsons in the City of Buffalo (section 2.4.1), an arson cluster alarm was sounded in October of 1996. This implied that recent observations were locating nearer to previous arsons than expected. But this could mean simply that the October 1996 arsons were located close to other arsons that were quite removed in a temporal sense (for example, perhaps the October 1996 arsons were located close to the location of January or February arsons). While this type of monitoring will sometimes be of interest, it will also be of interest to monitor changes in the pattern of arsons that occur over specified windows of time.

Suppose for example that we wish to implement spatial surveillance using a temporal window containing the previous 10 observations. Thus we would be looking for an increase (or conceivably a decrease) in the degree of clustering, where the definition of clustering is based upon the minimum distance from a newly observed point to any of the 10 previous observations. To implement this, we first find the minimum distance from a point, observed during the base period, to its nearest neighbor (where the set of nearest neighbors includes only the ten most recent observations). We next compare that distance with the distance expected in a random pattern (again obtained by taking the mean of a large number of minimum distances from randomly chosen points to sets of ten successive points that have been observed during the base period). When surveillance begins, this process is continued, with the observed distance to the previous ten observations being compared to the distance that would be expected if the base pattern did not change.

To illustrate, we define a subregion of the city of Buffalo where arson density appears the highest, and start surveillance in that subregion at observation 101 (after establishing a base pattern with the first 100 observations) with a moving window of ten observations. Figure 2.6 indicates that there are two cluster signals over the remainder of the year. Figure 2.7 displays the locations of the observed arsons that occurred Oct 5-11, just prior to the second of these cluster signals. These arsons are located nearer to one another than would be expected, given the usual distances observed between sequences of ten arsons observed during the base period. Sensitivity to changes in the base period were investigated by defining the first 50 observations and the first 150 observations as the base period. In each of these alternatives, cluster signals were noted at the same times as those displayed in Figure 2.6.

Figure 2.8 displays the results of surveillance for *increases* in the distances between neighbors. Two primary signals are sent -- the first after the 63rd observation following commencement of monitoring (on October 17th), and the second after the 107th observation on December 10th). Finally, Figure 2.9 portrays, using black triangles, the observations leading up to the second of these two signals. These arson locations tended to occur farther from their nearest neighbors (where nearest neighbors are defined by the minimum distance to the ten previous observations) than would be expected. Indeed the black triangles appear in this figure in relatively scattered locations within the subregion. This change might have been either temporary or more long-lasting. The fact that the cusum returns to less than critical values before observation 120 suggests that it was temporary. It is interesting that the signals for uniformity (large distances between arsons) occurred immediately following the cluster signals. Perhaps the temporary change was the result of increased patrol in the area that has a high density of arsons.

2.5 Summary and discussion

In this chapter, we have described a procedure for monitoring changes in spatial patterns over time. The method results in the rapid detection of deviations from expected patterns. It may be used for various subregions of the study area, and it may be implemented using time windows of differing length to search for changes in spatial pattern that may occur at particular time scales. Although the method has been used here in conjunction with the nearest neighbor index, it may be adapted for use with other spatial statistics. In particular, if X_t denotes any measure of spatial pattern at time t, we may use the following z-scores in a monitoring system that employs cumulative sums:

$$z = \frac{X_t - \mathrm{E}[X_t \mid X_{t-1}]}{\sqrt{\mathrm{V}[X_t \mid X_{t-1}]}} \tag{2.10}$$

An important issue concerns the choice of a base pattern. Ideally the analyst should be able to specify with confidence some prior period of time that was in some sense stable with respect to the evolution of spatial pattern and that could serve as a basis for comparison. One would not likely want to choose an odd or unusual period of time as a base period, any subsequent changes that were detected might simply signal a return to normalcy. In this application we only had access to one year of data; in general it would be important to have several years of data to be able to account for seasonal trends.

It should be clear that this method does not give the analyst answers to the all-important question of *why* the change in pattern has occurred. It does provide, however, a way of signaling *when* a significant spatial change occurred, and this should lead to both better short-term, strategic plans, and further hypotheses and investigations regarding the cause of change. In addition to signaling unexpected changes in patterns, it should also be of interest to detect changes in spatial patterns such as displacement that can be expected following targeted enforcement efforts. Although it would clearly have been interesting to investigate the possible causes of the changes in arson patterns in Buffalo (described in section 2.4), this was unfortunately not possible.

The monitoring approach described here focuses upon changes in *geographical* patterns only. Thus it does not signal either increases or decreases in the volume of criminal activity that may have taken place. This should not be viewed as a weakness of the method; the spatial monitoring method is designed to do exactly what its name implies, and should of course be combined with other appropriate analytic tools that achieve other objectives.

Although used here on a set of past data, it should be reemphasized that this monitoring system is designed primarily for handling new data as they become available. Although retrospective detection of pattern changes will certainly be of interest to the crime analyst in many situations, it is the rapid detection of changes in *current* patterns that are of most interest.

Finally, we are not suggesting that investigators wait for changes in patterns to develop before they begin their investigations. For some crime types, it will be important to follow any potential information, and a high "false-alarm" rate in the monitoring system may therefore be tolerable. The system described here provides just one of many important pieces of information and is designed to complement, rather than replace, other methods of crime analysis. Clearly, changes in spatial patterns may occur for many reasons. In some cases, investigators and crime analysts will possess "expert knowledge" that will be far more useful than a statistical analysis. But there are many other cases where statistical monitoring of pattern changes should prove beneficial to crime analysts. Individuals are notoriously poor at detecting whether significant clusters exist on a map -- there is a tendency to see clusters where none exist. It is therefore not a good idea to rely simply on visual interpretation. In addition, crimes that take place with high frequency may lead to such a high stream of data that it would be easy to overlook changes in pattern.

Footnotes

[1]One difficulty in implementing the statistical test described in section 2.2.1 concerns boundary effects -- the nearest neighbor of a point inside of the study area may lie outside of the study area. To obviate the difficulties caused by boundary effects, a simulation was performed to find critical values of the nearest neighbor index. N points were randomly located in the unit square and the nearest neighbor index R was calculated. For each value of N, this was repeated 10,000 times. The resulting mean values of R and the standard deviations of R associated with tests for clustering in the unit square are shown in Table 2.2. Note that in a bounded region such as the square used here, the observed distance between nearest neighbors will be somewhat greater than that expected in a random pattern, yielding a mean value of R slightly greater than one (since distances to near neighbors lying outside of the bounded region are discarded).

Table 2.1. In-Control ARLs (False Alarm Rates) for Various Values of *h*

h	ARL_0
2.5	68.9
2.6	76.9
2.7	85.8
2.8	95.6
2.9	106.5
3.0	118.6
3.1	131.9
3.2	146.7
3.3	163.1
3.4	181.2
3.5	201.2
3.6	223.4
3.7	247.9
3.8	275.0
3.9	304.9
4.0	338.1
4.1	374.7
4.2	415.3
4.3	460.1
4.4	509.6
4.5	564.4
4.6	625.0
4.7	691.9
4.8	766.0
4.9	847.8
5.0	938.2
5.1	1038.2
5.2	1148.7
5.3	1270.9
5.4	1405.9

Table 2.2

Simulated Mean and Critical Values of *R* in a Random Point Pattern

Number of Points (*N*)	Mean *R*	Std. Dev *R*
2	1.470	.7120
3	1.347	.4924
4	1.286	.4041
5	1.248	.3436
6	1.222	.2994
7	1.203	.2745
8	1.188	.2511
9	1.180	.2311
10	1.164	.2202
11	1.152	.2035
12	1.144	.1954
13	1.138	.1837
14	1.135	.1767
15	1.132	.1700
16	1.121	.1659
17	1.118	.1590
18	1.117	.1521
19	1.111	.1486
20	1.111	.1453
21	1.106	.1384
22	1.101	.1353
23	1.100	.1313
24	1.099	.1283
25	1.096	.1257
26	1.094	.1235
27	1.092	.1197
28	1.088	.1170
29	1.085	.1138
30	1.085	.1133
31	1.084	.1108
32	1.084	.1096
33	1.082	.1067
34	1.079	.1065
35	1.080	.1045

Number of Points	Mean *R*	Std. Dev. *R*
36	1.078	.1016
37	1.076	.1018
38	1.075	.0989
39	1.075	.0972
40	1.076	.0967
41	1.072	.0950
42	1.071	.0941
43	1.070	.0919
44	1.068	.0917
45	1.070	.0899
46	1.067	.0888
47	1.067	.0874
48	1.068	.0874
49	1.065	.0853
50	1.064	.0853

References

Bailey, T.C. and Gatrell, A.C. (1995). *Interactive Spatial Data Analysis*. Longman, Essex, England.

Block, R. C. 1995. STAC hot-spot areas: a statistical tool for law enforcement decisions. In R.C. Block, M. Dadboub, and S. Fregly, eds. *Crime Analysis Through Computer Mapping*. Washington, DC: Police Executive Research Forum, pp. 15-32.

Brunsdon, C. 1995. Estimating probability surfaces for geographical points data: an adaptive algorithm. *Computers and Geosciences* 21: 877-94.

Chen R. A surveillance system for congenital malformations, *Journal of the American Statistical Association* 1978; **73**: 323-327.

Clark, P.J. and Evans, F.C. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, 35, 445 453.

Farrington, C.P. and Beale, A.D. 1998. The detection of outbreaks of infectious disease. In *GEOMED '97, International Workshop on Geomedical Systems*. Eds. L. Gierl, A. D. Cliff, A. Valleron, P. Farrington, and M. Bull. pp. 97-117. Stuttgart: B.G. Teubner.

Griffith, D.A. and Amrhein, C.A. (1991). *Statistical Analysis for Geographers.* Prentice-Hall, Englewood Cliffs, New Jersey.

King, L. J. (1969). *Statistical Analysis in Geography*. Prentice-Hall, Englewood Cliffs, New Jersey.

Knox, G. (1964) The detection of space-time interactions. *Appl Statist*., **13**, 25-29.

Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society Series A* 2001; **164**: 61-72.

Levine, N. 1999. *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations*. Annandale, VA: Ned Levine and Associates.

Montgomery, D. (1996). *Introduction to Statistical Quality Control*. John Wiley, New York.

Okabe, A. and Yamada, I. 2001. The K-Function Method on a Network and Its Computational Implementation. *Geographical Analysis* 33: 271-90.

Openshaw, S., Charlton, M., Wymer, C., and Craft, A.W. 1987. A mark I geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems* 1: 335-58.

Ripley, B. 1976. The second order analysis of stationary point processes. *Journal of Applied Probability* 13: 255-66.

Rogerson, P. 1997. Surveillance systems for monitoring the development of spatial patterns. *Statistics in Medicine*, 16: 2081-93.

Rogerson, P. and Sun, Y. 2001. Spatial monitoring of geographic patterns: an application to crime analysis. *Computers, Environment, and Urban Systems*, 25/6: 539-56.

Ryan, T. P. (1989). *Statistical Methods for Quality Improvement*. John Wiley, New York.

Sherman, Lawrence and David Weisburd. (1995). General deterrent effects of police patrol in crime 'hot spots': a randomized, controlled trial. *Justice Quarterly*, 12 (4), 625-648.

Wetherill, G.W. and Brown, D.W. 1991. *Statistical Process Control: Theory and Practice*. Chapman and Hall, New York.

Figure Captions and Notes

Figure 2.1. 1996 Arsons: Nearest neighbor index Over Time

Figure 2.2. Cumulative Sum for 1996 Arsons

Note: Cusum not reset to zero following alarm

Figure 2.3. Cumulative Sum for 1996 Arsons

Note: Cusum reset to zero following alarm

Figure 2.4. 1996 Arsons (Triangles Represent Arsons Leading to Cluster Signal in Early October)

Figure 2.5. Arson Locations Before and After Signal ("triangle" indicates after)

Figure 2.6. Cumulative Sum with Window of Ten Observations

Figure 2.7. Arson Locations Leading to Clustering Signal with Window of Ten Observations

Figure 2.8. Cumulative Sum with Window of Ten Observations

Figure 2.9. Arson Locations Leading to Dispersion Signal With Window of Ten Observations

Chapter 3. Optimal Police Enforcement Allocation: A Socio-Economic Model of Geographical Displacement and Spatial Concentration of Crime

3.1. Introduction

A considerable amount of money from federal as well as state and local government has been allocated to reduce crime. Although varying in type and scope, crime prevention programs have been widely advocated to supplement proactive policing. Evaluations of preventive programs have been predicated on the unstated assumption that the offender population is rigid and fixed in space. However, empirical evidence suggests that crime mobility is a very critical issue to evaluate the effectiveness of crime reduction programs. Indeed, a burglary prevention program implemented in a community would be hailed as successful if burglary rates in that community dropped following implementation. However, consideration must be given to the possibility that burglaries increased in adjacent communities or that the community was experiencing increases in other crimes (Gabor 1990). Thus, an efficient program considers not only where the law enforcement resources are applied but also their impact on the surrounding environment in terms of criminal mobility, interjurisdictional spillovers of police, and potential interactions among adjacent neighborhoods.

In the 1970s, the first concrete evidence of a crime displacement effect began to emerge from two studies conducted under the auspices of the Rand Institute in New York City. In one (Press 1971), a 40 percent increase of police manpower in one precinct of New York occasioned a reduction of street crimes therein, but also appeared to produce a compensating increase in these crimes in adjacent precincts. The second study (Chaiken, Lawless and Stevenson 1974) revealed that the introduction of an exact-fare system to curtail robberies on New York City's buses achieved a dramatic decline in these stick-ups, but not without magnifying the problem of robbery on the city's subway system. Other notable examples of crime mobility in the early part of the 1970s were identified in Columbus, Ohio, and Newark, New Jersey. In Columbus (Lateef 1974), a police helicopter patrol program appeared to displace robberies, burglaries, and auto thefts to precincts not covered by the patrols. In Newark (Tyrpak 1975), the intensification of street lighting in several high-crime precincts seemed to result in a shift of some of these crimes to bordering precincts. A Canadian study (Gabor 1981) revealed that the Operation-Identification property-making program might have moved some break-ins from the homes of participants to those not participating in the program.

A practical problem that police officers frequently encounter is a question of where criminal activities are most likely to displace when some of the neighborhoods in their patrol area receive extra enforcement. This problem is especially important when the police are planning crackdown programs to apply to some of their neighborhoods, or when new officers are to be assigned to some neighborhoods.

Recently, researchers have begun to mathematically model this observed criminal behavior. One advantage of studying crime by creating a mathematical model is that it provides police officers with some useful quantitative information. Caulkins (1993) creates a crime model specifically designed to study a crackdown program on a drug market. This influential paper has led to much additional work to better understand drug markets, e.g., Baveja et al. (1993, 1997), Naik et al. (1996) and Kort et al. (1998). Deutsch, Hakim and Weinblatt (1984) apply a time series technique to forecast crime numbers in the area of interest. They then create a criminal transportation problem with impedance costs to predict the crime displacement effects. Wortley (1998) described a two-stage situational prevention model that attempts to give fuller recognition to the complexity of the person-environment relationship, and, in doing so, also seeks to address the theoretical problems of crime displacement on locations. We refer the reader to surveys of the criminology literature of operations research and management science that are available in Maltz (1994), Barnett, Caulkins and Maltz (2001) and Blumstein (2002).

Our work develops a mathematical model of criminal behavior among several adjacent neighborhoods. In this model, criminals are assumed to make rational choices. In the rational choice approach (e.g., Cornish and Clarke 1987), offenders are assumed to seek benefit from their criminal behavior. They decide whether or not to displace their attentions elsewhere based on the characteristics of particular offenses, in particular, their opportunities and profits. We apply maximum utility theory to describe how criminals might respond to enforcement pressure.

Using our socio-economic model of criminal behavior, our purpose is to develop an optimal enforcement allocation policy. The "best" allocation, of course, depends on the objective involved. We examine two plausible objectives: (i) minimizing the total number of crimes among the neighborhoods, and (ii) minimizing the difference in the number of crimes between neighborhoods. Since the allocation policies for these two objectives may not coincide, we also explore policies that yield a comprising solution. For this purpose, we establish the existence of so-called non-dominated solutions, which, although not necessarily the best solution for either objective, are not worse under both objectives when compared to any other solution.

It is important to note that our model does *not* assume a constant total criminal activity. That is, the effect of crime displacement does not necessarily result in the total amount of crime remaining constant. If this were the case, there would be little net benefit to police enforcement strategies. In fact the specific goal we have is to seek an enforcement strategy that will minimize the total net crime of all neighborhoods. The difference is that we recognize the fact that crime displacement will occur as part of how criminals respond to changing enforcement pressures.

The rest of this report is organized as follows. In Section 3.2, we establish a mathematical expected return function for crime that demonstrates the relationship between the number of crimes and some socio-economic conditions. Section 3.3 discusses some natural properties of criminal behavior that are suggested by the model. Section 3.4 attempts to explain why both poor and affluent neighborhoods experience high crime rates. In Section 3.5, we analyze crime displacement effects and provide a measurement to evaluate the efficiency of a crackdown program. Section 3.6 predicts where criminals will tend to displace their activity when facing the pressure of intense enforcement. Section 3.7 then applies the model to study enforcement allocation policies between two neighborhoods. Section 3.8 generalizes the results for two neighborhoods to multiple neighborhoods. In Section 3.9 a case study involving a burglary dataset from the Buffalo Police Department is analyzed.

3.2. Crime Expected Return Function

Suppose the wealth level in a neighborhood is w and the amount of law enforcement it receives is E. The wealth level can be the median or mean of household incomes, and the law enforcement level can be measured by the patrol hours or police monetary budget applied in the neighborhood. If a criminal commits a successful crime in the neighborhood, he acquires reward; if arrested, he forfeits this take. Therefore, the expected monetary return from committing a crime in a neighborhood is the product of the probability of not being arrested and the reward (Freeman *et al*. 1996). Wang *et al.* (2000) create exponential functions to describe the arrest probability function and the reward function. The arrest probability, a function of the enforcement per crime, is defined as

$$P_A(E/n) = 1 - exp(-\alpha(E/n)), \tag{3.1}$$

where n is the number of crime incidents and α is a positive constant. Note that under a fixed level of enforcement, a crime has a lesser probability of arrest in an area which has a larger amount of crime incidents due to the low average enforcement devoted to it. Greenwood et al. (1977) appears to be the first to document this inverse relationship between the number of crimes and the probability of arrest.

The reward function is defined as

$$R(n) = c\, w\, exp(-\beta n), \tag{3.2}$$

where c is a proportionality constant and β is a positive constant. This expression implies the return to a successful crime incident decreases exponentially with the number of crime incidents via an appropriate positive constant and is proportional to the wealth of the neighborhood. It assumes the total monetary return to a crime in a neighborhood is limited by the wealth of the neighborhood: the more incidents in the neighborhood, the less wealth that remains for others.

Equations (3.1) and (3.2) give us an expression for the expected monetary return from committing a crime in a neighborhood as:

$$f(n) = R(n)*(1-P_A(E/n)) = c\omega\, exp(-\alpha E/n - \beta n). \tag{3.3}$$

As we have seen, equation (3.3) depends on several parameters. The α value reflects the effectiveness of the per-incident enforcement E/n in making an arrest. This parameter may vary for different crimes and neighborhoods. Since we will be concentrating on a single crime type, α will reflect the variability of arrest effectiveness among the neighborhoods. Different types of crimes might have different c and β values. A crime that requires a higher level of skill to commit, and hence commands a higher return, should have a higher c value. Crimes that are more peer-competitive (easily affected by the number of criminals) have a higher β value. Since our focus is on a single type of crime, the values of c and β are assumed to be identical for all neighborhoods.

Suppose criminals have an opportunity cost, m, for committing a specific type of crime. This cost may reflect foregone opportunities such as gainful employment or crime of a different type. If the expected return in the neighborhood is greater than the opportunity cost, crime in that neighborhood is attractive. On the other hand, if the expected return is less than the opportunity cost, criminals might quit committing crimes or displace their criminal activity. In equilibrium, the expected return from committing a crime is equal to the opportunity cost, and criminals are indifferent between committing crimes or not in the neighborhood. Mathematically, we represent equilibrium as $f(n) = m$.

Solving $f(n) = m$, two solutions for n are found:

$$n = n^{-} := (\sqrt{k} - \sqrt{k - 4\alpha\beta E})/2\beta, \tag{3.4}$$

$$n = n^{+} := (\sqrt{k} + \sqrt{k - 4\alpha\beta E})/2\beta, \tag{3.5}$$

provided that $E \leq k/4\alpha\beta$ where $k = [ln(cw/m)]^2$. As illustrated in Figure 2.1, in the case that the number of crimes is less than n^-, the expected return is below the opportunity cost. The neighborhood is unattractive to criminals and they will exit the neighborhood or quit committing crimes. However, this decrease in crime increases the amount of enforcement per crime and might further encourage more criminals to leave. This phenomenon is called *positive feedback* (Kleiman, 1988), and will tend to collapse criminal activities in the neighborhood. If the number of incidents lies between n^- and n^+, the neighborhood is attractive to criminals and the number of crime incidents is then expected to increase. However, as the number of crimes reaches saturation at n^+, the expected return of committing a crime no longer provides an economic incentive to attract additional crime. Thus the number of crime incidents will stay at n^+ in equilibrium. If the number of crime incidents is greater than n^+, the neighborhood is over-saturated and crime incidents will drop until the number reaches equilibrium at n^+. From these observations, we call n^- and n^+ the *unstable equilibrium* and the *stable equilibrium*, respectively.

In the case that $E > k/4\alpha\beta$, i.e., a relatively large police pressure, $f(n) = m$ does not have a solution since the expected return never reaches the opportunity cost. Hence the criminal activity eventually collapses. We can summarize the discussion of this section into the following proposition.

Proposition 2.1. The number of crime incidents in equilibrium depends on the initial number of crime incidents when the law enforcement is first applied. If the initial number of crime incidents is less than the unstable equilibrium n^-, the criminal activity will collapse and no crimes survive in the neighborhood. Otherwise, the number of crime incidents will reach the stable equilibrium n^+.

3.3. Some Properties of Criminal Activity

To further investigate criminal activity, we first define another function $S(n) := n\, f(n)$, the total expected monetary amount supplied by the crime victims when the number of crime incidents is n. We now consider some properties of $f(n)$ and $S(n)$. It is worth noting that

$$\lim_{n\to 0} S(n) = \lim_{n\to\infty} S(n) = \lim_{n\to 0} f(n) = \lim_{n\to\infty} f(n) = 0\,.$$

Also, the first order derivatives of S(n) and $f(n)$ are

$$S'(n) = f(n)(1+ \alpha E/n - \beta n), \text{ and}$$

$$f'(n) = f(n)(\alpha E/n^2 - \beta).$$

Solving $S'(n) = 0$ and $f'(n) = 0$ for $n > 0$, we have $n^{**} = [1+ (1 + 4\alpha\beta E)^{1/2}]/2\beta$ and $n^{*} = (\alpha E/\beta)^{1/2}$, respectively. Furthermore, $S'(n) > 0$ for $0 < n < n^{**}$ and $S'(n) < 0$ for $n > n^{**}$; $f'(n) > 0$ for $0 < n < n^{*}$ and $f'(n) < 0$ for $n > n^{*}$. Therefore, we can conclude the following proposition.

Proposition 3.1. S(n) and $f(n)$ share the following similar properties:

1. These two functions are positive and approach zero as n approaches either zero or infinity.
2. $S(n)$ has a unique maximum at $n^{**} = [1+ (1 + 4\alpha\beta E)^{1/2}]/2\beta$ and $f(n)$ has a unique maximum at $n^{*} = (\alpha E/\beta)^{1/2}$. We denote the maximum total expected supply level as $S^{**} = S(n^{**})$ and the maximum expected return per crime as $f^{*} = f(n^{*}) = cw\, exp(-2(\alpha\beta E)^{1/2})$.
3. $S(n)$ and $f(n)$ are increasing with n as $0 < n < n^{**}$ and $0 < n < n^{*}$, respectively. Also, they are decreasing with n as $n > n^{**}$ and $n > n^{*}$, respectively.

Property 1 provides the natural statement that whatever the number of crimes in a neighborhood, the expected monetary return of a crime and the total amount are positive. Further, when n is very large, these values are small because of the limited wealth in the neighborhood and the potential victims' awareness of crimes that will encourage them to add more security to protect their wealth.

In property 2, n^{*} can be interpreted as the ideal crime level for individual criminals. When the number of incidents reaches n^{*}, criminals can expect the largest return $f^{*} = c\ w\ exp[-2(\alpha\beta E)^{1/2}]$ from committing a crime. At this level, the expected return per crime is at its highest. Thus, n^{*} is the ideal crime level for individually optimizing criminals such as a corner drug dealer, a burglar or petty thief.

It is worth noting that although f^* decreases, n^* increases with an increase in enforcement. This is because high enforcement increases the probability of arrest and reduces the maximum return criminals might get. However, from the perspective of criminals, in a high law enforcement environment they prefer more crimes to occur in the neighborhood to share the arrest pressure and reduce the average enforcement upon them. This decreases the possibility of their being arrested. Hence, the n^* value is larger in an environment of higher enforcement.

When the number of crime incidents reaches n^{**}, criminals enjoy their maximum total expected return. Thus, if there is a group of criminals that can organize the criminal activity in this neighborhood, n^{**} is the crime level they want to maintain. We can thus think of n^{**} as the organized crime equilibrium level. At this level, the organization will try to maintain crime level n^{**} by protecting their territory from outside criminals and asking their members to keep up their current activities. However, if the neighborhood is open to outside criminals, the success of current criminals tends to attract more crimes into the neighborhood. This continues until there are too many crimes in the neighborhood and eventually the number of crimes will reach the stable equilibrium n^+. Hereafter, if not mentioned specifically, we assume that no criminal organizations control a neighborhood.

Property 3 says that $f(n)$ is an increasing function for $n < n^*$. This is due to the fact that for a small number of n the average enforcement for each potential crime incident is large and hence the arrest rate is relatively high. In this situation, more crime incidents can reduce the arrest possibility and then increase the expected return. On the other hand, $f(n)$ is a decreasing function for $n > n^*$. That is, due to the wealth limitation of the neighborhood, the expected return eventually falls down as the number of crimes increases. After the number of incidents in the neighborhood reaches a certain level, the wealth to be shared per incident decreases. A similar explanation holds for $S(n)$. As $n < n^{**}$, the number of crimes is not high enough to warrant the attention of potential victims. At this time, they are more willing to give (to criminals) than to add more security to protect their property. However, once we have $n > n^{**}$, the curve decreases with the number of crimes since the loss to criminals exceeds the endurance of potential victims.

Proposition 3.1 also concludes that the shape of the expected return function (total monetary return curve) (*i*) has a unique maximum, (*ii*) approaches to zero asymptotically at the two end points, and (*iii*) is unimodal, increasing with n as $n < n^*$ (n^{**}) and decreasing with n as $n > n^*$ (n^{**}).

The next proposition provides decision makers information about how much enforcement should be allocated in a neighborhood to collapse the criminal activity. Note that to collapse criminal activity in the neighborhood, it suffices to assign $\mathrm{E} = [k/4\alpha\beta]^+$, which we define as exceeding $k/4\alpha\beta$ by an infinitesimal amount of enforcement.

Proposition 3.2. Suppose the initial number of crimes in a neighborhood is n_0. The minimum enforcement resources required to collapse criminal activity in the neighborhood is

(i) $[k/4\alpha\beta]^+$, if $n_0 \geq k^{1/2}/2\beta$;

(ii) $[n_0(k^{1/2}-\beta n_0)/\alpha]^+$, if $n_0 < k^{1/2}/2\beta$.

Proof:

From Proposition 2.1, in case that the initial number of crimes is less than the unstable equilibrium, the criminal activity finally collapses. Therefore, we are looking for E such that

$$n_0 < [k^{1/2} - (k - 4\alpha\beta\mathrm{E})^{1/2}]/2\beta.$$

Now, if $k^{1/2} - 2\beta n_0 > 0$, i.e., $n_0 < k^{1/2}/2\beta$,

$$n_0 < [k^{1/2} - (k - 4\alpha\beta\mathrm{E})^{1/2}]/2\beta \Leftrightarrow (k - 4\alpha\beta\mathrm{E})^{1/2} < k^{1/2} - 2\beta n_0 \Leftrightarrow \mathrm{E} > n_0(k^{1/2}-\beta n_0)/\alpha.$$

Hence, it is required to allocate E = $[n_0(k^{1/2}-\beta\, n_0\,)/\alpha]^+$ to collapse criminal activity.

If $k^{1/2}$ - $2\beta n_0 \leq 0$, i.e., $n_0 \geq k^{1/2}/2\beta$, there is no solution for E. But, from the discussion of the previous section, no matter what the initial number of crime incidents is, if the law enforcement applied to the neighborhood is greater than $k/4\alpha\beta$, the criminal activity will collapse. Therefore, to collapse criminal activity, it is required to assign E = $[k/4\alpha\beta]^+$.

The proposition is proved.

Corollary 3.1. The amount of law enforcement required to collapse criminal activity is increasing in the initial number of crimes.

Proof:

From Proposition 3.2, if $n_0 < k^{1/2}/2\beta$, it requires E = $[n_0(k^{1/2}-\beta\, n_0\,)/\alpha]^+$ to collapse criminal activity. Now, assume that $n_0 \leq k^{1/2}/2\beta$. We get

$$\partial/\partial n_0(\, n_0(k^{1/2}-\beta\, n_0\,)/\alpha\,) = k^{1/2}/\alpha - 2\beta\, n_0/\alpha \geq k^{1/2}/\alpha - 2\beta(k^{1/2}/2\beta)/\alpha = 0.$$

Hence, $n_0(k^{1/2}-\beta\, n_0)/\alpha$ is an increasing function of n_0 and the upper bound happens when n_0 = $k^{1/2}/2\beta$ with function value $k/4\alpha\beta$. This proves the corollary.

Corollary 3.1 gives guidance on the cost of implementing a crackdown program. The smaller the number of crimes the less the required enforcement level is for achieving a crackdown. If we wait till the criminal activity in a neighborhood matures, it will cost us more to collapse the activity.

3.4. Crime Rates and Wealth Level

This section studies the relationship between crime rates and wealth level. We use our model and the theory of deprivation and strain (for an overview see Belknap 1989) to solve the paradox that both affluent and poor societies may have high crime rates. A reasonable explanation for the inconsistent relationship is provided.

According to deprivation and strain theories, in a poor society with gross income disparities, more persons will feel the need to compensate for their perceived or actual deprivation through criminal activity (Jongman *et al*. 1991; van Dijk 1994). The potential criminals in such societies tend to be risky and accept low return from committing a crime. In terms of our model, the opportunity costs in poor societies will be shifted downwards in comparison to elsewhere. Also, in poor societies there are less viable targets available than in more affluent societies where people can better afford losses. For this reason, the monetary return curves shift downward as well. Whether the number of crime incidents will eventually increase or decrease in societies with different wealth levels depends on the strength of factors affecting the opportunity cost and expected return from a crime. If the opportunity costs are more strongly shifted downwards than expected returns are, then the intersection point will be moved to the right (as shown in Figure 4.1). Therefore, the number of crimes will eventually increase.

The findings of the International Crime Survey indicate that the levels of most types of property criminals are in fact relatively high in many countries with low GNPs and/or massive unemployment. By far the highest rates of property crime rates were measured in cities in developing countries in Africa and South America, such as Kampala, Dar es Salam, Tunis, Rio de Janeiro, and Buenos Aires (van Dijk and Zvekic 1993). These results suggest that opportunity costs are indeed placed lower in underdeveloped socio-economic societies and, in equilibrium, the downward shift of the opportunity costs seem to be larger than the downward shift of the expected return curves. This explains why some cities in the third World suffer extremely high crime rates, although criminals do not have higher expected returns from committing crimes.

In more affluent societies the presence of luxury goods yields opportunity for large profits from committing crimes. As a consequence, expected return curves are shifted upwards considerably. At the same time, the effect of deprivation and strain is not very strong in affluent societies. Along with welfare provisions and lower unemployment rates, the opportunity costs are shifted upwards too. If the expected return curves are more strongly shifted upwards than the opportunity costs are, then the intersection point will be moved to the right and it causes the number of crimes to increase (see Figure 4.2).

Recall that the number of crimes in equilibrium is $[k^{1/2} + (k - 4\alpha\beta E)^{1/2}]/2\beta$ where $k = [ln(cw/m)]^2$. The number of crime incidents in equilibrium has a positive relationship with the expected return (which is positively proportional to the wealth level of the neighborhood *w*), but has an inverse relationship with the opportunity cost (*m*) for committing crimes. Furthermore, affluent neighborhoods which have higher *w* values also have higher *m* values by the theory of deprivation and strain. Generally, neighborhoods with higher values of *w*/*m* attract more crimes, and the net effects on the opportunity costs and expected return curves can help us explain why both some developing and some of the most affluent communities experience relatively high levels of property crimes. For the communities in developing (less affluent) countries, the high crime rates are caused by the low opportunity costs for criminal wages. For communities in affluent countries, high expected returns for criminal behavior are responsible for high crime rates.

3.5. Geographical Displacement Phenomenon

In this section we demonstrate how a crackdown program being applied to a neighborhood impacts an adjacent neighborhood. We start with the case of two adjacent neighborhoods, neighborhood 1 and neighborhood 2, and study how a crackdown program applied in neighborhood 2 affects the criminal activities in both neighborhoods.

We presume that criminals tend to commit crime in the neighborhood with larger amount of expected return and they have an opportunity cost, *m*, for specific type of crime. If the expected return in a neighborhood is greater than the opportunity cost, crimes in that neighborhood are attractive and criminals in other neighborhoods will tend to move in. On the other hand, if the expected return is less than the opportunity cost, criminals might quit committing crimes or commit crimes in other neighborhoods. In equilibrium, a person of criminal mindset is indifferent between committing crime in any neighborhood. Thus, both neighborhoods have the same expected return from committing a crime, equal to the opportunity cost. That is, in equilibrium, $f_1(n_1) = f_2(n_2) = m$ so that the equilibrium numbers of crime incidents in the neighborhoods are

$$n_1 = (k_1 + \sqrt{k_1 - 4\alpha_1 \beta E_1})/2\beta \text{, and}$$

$$n_2 = (k_2 + \sqrt{k_2 - 4\alpha_2 \beta E_2})/2\beta \text{,}$$

where $k_i = [ln(cw_i/m)]^2$.

When the crackdown is first applied to Neighborhood 2, the expected return of a crime in Neighborhood 2 suddenly drops to $m_2 = f_2(E_2+\Delta E, n_2)$, which is less than *m* since the expected return function decreases with the amount of enforcement. At this stage, under the enforcement level of $E_2+\Delta E$, criminal activity in neighborhood 2 (with the number of crimes n_2) is too self-competitive. Criminals who used to commit crimes in Neighborhood 2 have now three alternatives (*i*) quit committing crimes because of higher enforcement pressure, (*ii*) move to neighborhood 1 to pursue a higher expected return, or (*iii*) stay in neighborhood 2 and accept a lower expected return.

The first two alternatives cause the number of crimes in neighborhood 2 to decrease. Losing crimes in neighborhood 2 makes the criminal activity less self-competitive and gradually increases the expected return in the neighborhood. At the same time, since some criminals move from neighborhood 2 to neighborhood 1, the number of crimes in neighborhood 1 is then over-saturated and the expected return in the neighborhood drops. As we have seen, the expected return of committing a crime gradually increases from m_2 in neighborhood 2 and decreases from m in neighborhood 1. When the expected returns of the two neighborhoods equalize, a new equilibrium between m_2 and m is reached. We denote the new equilibrium by m', which is now the new opportunity cost of a crime in the two neighborhoods. This scenario, called a *geographical displacement phenomenon,* is illustrated in Figure 5.1.

Under the new equilibrium, we should have $f_1(n_1') = f_2(n_2') = m'$ and the equilibrium numbers of crime incidents in the two neighborhoods are

$$n_1' = (k_1' + \sqrt{k_1' - 4\alpha_1 \beta E_1}) / 2\beta \text{, and}$$

$$n_2' = (k_2' + \sqrt{k_2' - 4\alpha_2 \beta (E_2 + \Delta E)}) / 2\beta \text{,}$$

where $k_i' = [ln(cw_i/m')]^2$.

Note that the opportunity cost decreases from m (before a crackdown is applied to neighborhood 2) to m' (after a crackdown is applied to neighborhood 2). This is because criminals in neighborhood 1 perceive the pressure *indirectly* from the enforcement *directly* applied to neighborhood 2 and the opportunity costs have an inverse relationship with the total enforcement. Here, we assume the criminals have no geographical preference in which neighborhood to commit crimes. Their only concern is the amount of the (illegal) expected return.

If the crackdown program applied in neighborhood 2 only forces criminals to move to neighborhood 1 but does not reduce the criminals' opportunity cost in the neighborhood, the criminal activity in neighborhood 1 will be too self-competitive and eventually the number of crimes in the neighborhood will drop back to n_1. In this situation, criminals in neighborhood 1 suddenly face many competitors moving in from neighborhood 2 and find that the expected return is not as high as before. Instead of reducing their opportunity cost, they choose to quit committing crimes in neighborhood 1. The displacement effect only happens when the crackdown program is first applied. However, in a long-term point of view, the crackdown program in neighborhood 2 has an absolute success.

3.6. Prediction on Crime Movement

In this section, we predict the direction of crime movement when the enforcement allocation policy is changed; specifically, when the crackdown program is applied in one of several neighborhoods.

We first consider the situation where decision makers would like to decrease the crime number in one of their patrol neighborhoods, the target neighborhood; however, they are not supplied with extra enforcement from outside resources. That is, they have to collect some resources from neighborhoods and apply the resources to their target neighborhood. Suppose the interested area consists of n neighborhoods and neighborhood i receives the amount of law enforcement E_i for $i = 1...n$. We would like to add the amount of enforcement ΔE to the target neighborhood l. Let S denote the set of neighborhoods from which some enforcement will be removed to obtain these resources. The enforcement levels of the neighborhoods other than l and those in S are kept the same. Assume the new allocation policy is

$E_l' = E_l + \Delta E$;

$E_j' = E_j - \Delta E_j$, for $j \in S$, where $\quad {}_{j \in S}\Delta E_j = \Delta E$;

$E_k' = E_k$ for $k \notin S \cup \{l\}$.

Recall our assumption that the criminals' opportunity cost depends on the total amount of enforcement in the neighborhood. In this case, since the total amount of enforcement does not change, the value of the opportunity cost should be the same after the new enforcement allocation. Hence, we should have $n_l' < n_l$ and $n_j' > n_j$ for $j \in S$, and the number of crimes is kept at the same level for the other neighborhoods.

This result is very intuitive. Without receiving extra enforcement resources, any crackdown program applied in one neighborhood by reducing enforcement in other neighborhoods does not really help the crime control in a global sense. We cannot reduce the number of crimes in the target neighborhood without increasing the numbers of crimes in the neighborhoods which supply enforcement to the target neighborhood.

Now we consider the case that extra enforcement resources can be obtained. Suppose that we assign the entire extra enforcement to the target neighborhood, i.e., let

$E_l' = E_l + \Delta E$;

$E_j' = E_j$ for $j \neq l$.

Since the total amount of enforcement is increased by ΔE, the criminals' opportunity cost decreases. Therefore, we should have $n_l' < n_l$ and $n_j' > n_j$ for $j \neq l$. To check which neighborhoods criminals are most likely to displace their criminal activities, taking the first derivative of the stable equilibrium $n = [k^{1/2} + (k - 4\alpha\beta E)^{1/2}]/2\beta$ on m, we have

$$\partial n/\partial m = - [\ 1 + (\ 1 - 4\alpha\beta E/k)^{-1/2}\ ]/(2\beta m).$$

The absolute value of $\partial n/\partial m$ is increasing with $\alpha E/k$. Hence, for neighborhoods with larger $\alpha E/k$ values, the number of crimes tends to increase more rapidly as m decreases. Therefore, when extra enforcement resources are introduced to one of the neighborhoods and the other neighborhoods remain at the original levels of enforcement, criminals tend to displace their criminal activity to the neighborhoods with larger $\alpha E/k$ values.

Note that for multiple neighborhoods in which criminals have the same level of opportunity cost, the neighborhoods with larger $\alpha E/k$ values will tend to have smaller numbers of crime incidents in equilibrium. Therefore, if a crackdown program is applied to a neighborhood and the displacement effect does occur, criminals will eventually displace most of their activities to the neighborhoods that had less numbers of crimes before the crackdown program was applied. This implies that when criminal activity among the neighborhoods reaches steady state again, the disparity of the number of crimes among the neighborhoods decreases.

3.7. Optimal Allocation Policies in Two Neighborhoods

In this section, we study the optimal allocation policies with the case of two neighborhoods. Total enforcement resources are assumed fixed and the objectives are to determine the proportion of enforcement that should be applied to each neighborhood.

The optimal allocation policies are developed based on two alternative objectives: (*i*) minimizing the weighted sum of crime numbers, $\lambda_1 n_1 + \lambda_2 n_2$, and (*ii*) minimizing the difference of the number of crime incidents between two neighborhoods, $|\gamma_1 n_1 - \gamma_2 n_2|$, where λ_1, λ_2, γ_1 and γ_2 are positive constants. Again, $n_i = n_i^+$ if allocated $x_i < (k_i/l_i)^+$ or $n_i = 0$ if $x_i = (k_i/l_i)^+$.

Minimizing the weighted sum is a general global concern. If all the crime numbers in different neighborhoods are equally weighted, the objective is to minimize the total number of crimes among all neighborhoods. If interested in per capita crime, one might use the population of the neighborhoods as an index to weight the neighborhoods. However, there is no guarantee that minimizing the total weighted number of crimes will lead to an equitable distribution of resources. There may be, instead, a desire to balance the crime numbers among all neighborhoods. Hence, the alternative objective of minimizing the difference in the numbers of crimes between the neighborhoods is a possible remedy.

Consider the trivial case that the enforcement E is sufficient to collapse criminal activities in both neighborhoods simultaneously, i.e., $E > k_1/4\alpha_1\beta + k_2/4\alpha_2\beta$, or equivalently $k_1/l_1 + k_2/l_2 < 1$. Then by setting $x_i = (k_i/l_i)^+$, it is possible to achieve the optimal solutions $\lambda_1 n_1 + \lambda_2 n_2 = 0$ and $|\gamma_1 n_1 - \gamma_2 n_2| = 0$. Therefore, we shall only consider the situation that E is not sufficient to collapse criminal activities in the neighborhoods simultaneously. Let $x_1 = x$ and $x_2 = 1\text{-}x$, so that all enforcement is allocated. Also, let $h = \sqrt{k_1/l_1 + k_2/l_2 - 1}$.

3.7.1 Minimizing Total Number of Crime Incidents

In this section we wish to minimize the weighted sum of crime numbers $\lambda_1 n_1 + \lambda_2 n_2$ of the two neighborhoods, where λ_1 and λ_2 are positive constants.

The allocation problem can be formulated mathematically as

<Problem 7.1> Minimize $F_1(x) = \lambda_1 n_1 + \lambda_2 n_2$

subject to $\max\{0, 1 - (k_2/l_2)^+\} \le x \le \min\{1, (k_1/l_1)^+\}$.

The optimal enforcement allocation policy happens to be a *one-neighborhood crackdown policy*, which allocates as much enforcement as possible to collapse the criminal activities in one neighborhood and then any remaining enforcement to the other neighborhood. This is due to the fact that the objective function, $F_1(x)$, is concave in the proportion of enforcement allocated in neighborhood 1, x, and is discontinuous at the endpoints $(k_1/l_1)^+$ and $1 - (k_2/l_2)^+$. The details are summarized in Theorem 7.1.

Theorem 7.1: Suppose enforcement E is not sufficient to collapse criminal activities in the two neighborhoods simultaneously. The optimal allocation policy that minimizes the weighted sum of crime numbers $\lambda_1 n_1 + \lambda_2 n_2$ is a one-neighborhood crackdown policy as follows.

1. If E is sufficient to collapse criminal activities in both of the two neighborhoods but not simultaneously, the crackdown should be made in the neighborhood with highest $\lambda(\sqrt{k} + h\sqrt{l})$ value.

2. If E is sufficient to collapse criminal activities in only one of the two neighborhoods, say neighborhood 1, then the crackdown should be made in neighborhood 1 if $2\lambda_1\sqrt{k_1} \leq \lambda_2(h\sqrt{l_2} - \sqrt{k_2 - l_2})$; otherwise, neighborhood 2 will receive all enforcement. The first case holds if $\lambda_1 = \lambda_2$ and neighborhood 1 has a better arrest ability ($\alpha_1 > \alpha_2$).

3. If E is not sufficient to collapse criminal activities in either neighborhood, the crackdown should be made in the neighborhood with highest $\lambda(\sqrt{k} - \sqrt{k - l})$ value.

Proof: See Wang et al. (2000b).

The following two corollaries to Theorem 7.1 provide police management some intuitive tips for making allocation policies under some specific circumstances. As in Theorem 7.1, we assume in both corollaries that the total enforcement of a crackdown program is not sufficient to collapse the criminal activities of the two neighborhoods simultaneously. The proof of the corollaries can be derived directly from Theorem 7.1.

Corollary 7.1: If the objective equally weights the number of crimes, and the wealth of the two neighborhoods is equal, i.e., $\lambda_1 = \lambda_2$, and $w_1 = w_2$ or $k_1 = k_2$, then the crackdown should be made in the neighborhood with larger α value.

Corollary 7.2: If the objective equally weights the number of crimes, and the efficiency of making arrest in the two neighborhoods is equal, i.e., $\lambda_1 = \lambda_2$ and $\alpha_1 = \alpha_2$ or $l_1 = l_2$, then

1. If E is sufficient to collapse criminal activities in both of the two neighborhoods but not simultaneously, the crackdown should be made in the wealthier neighborhood.
2. Otherwise, the crackdown should be made in the less wealthy neighborhood.

3.7.2 Minimizing Crime Disparity

In this section we wish to minimize the weighted difference of crime numbers $|\gamma_1 n_1 - \gamma_2 n_2|$ between the two neighborhoods, where γ_1 and γ_2 are positive constants.The allocation problem can be formulated mathematically as

<Problem 7.2> Minimize $|F_2(x)| = |\gamma_1 n_1 - \gamma_2 n_2|$

subject to $\max\{0, 1 - (k_2/l_2)^+\} \leq x \leq \min\{1, (k_1/l_1)^+\}$.

If the crime numbers are equally weighted for the two neighborhoods, the objective is to minimize the difference in crime numbers. If γ_i is the inverse number of the population in neighborhood *i*, the objective is to minimize the disparity of per capita crime between the neighborhoods. The optimal enforcement allocation policy is summarized in Theorem 7.2.

Theorem 7.2: If $\gamma_1\sqrt{k_1} \geq \gamma_2\sqrt{k_2}$ and E is not sufficient to collapse criminal activities in the two neighborhoods at the same time, the weighted difference of the numbers of crime incidents between the two neighborhoods can be minimized to zero except for the following two situations.

1. $k_1 < l_1$ and $\gamma_1\sqrt{k_1} > \gamma_2(\sqrt{k_2} + h\sqrt{l_2})$.
2. $k_1 \geq l_1$ and $2\gamma_2\sqrt{k_2} < \gamma_1(\sqrt{k_1} + \sqrt{k_1 - l_1})$.

Proof: See Wang et al. (2000b).

These two cases occur when one or two of the following situations: (1) the difference between the wealth of the two neighborhoods is relatively large ($k_1 >> k_2$); (2) more attention is paid to the crime level in neighborhood 1 ($\gamma_1 >> \gamma_2$). Specifically, in the case that the total enforcement E is sufficient to collapse criminal activities in the neighborhood with highest $\gamma\sqrt{k}$ value ($k_1 < l_1$), the difference of the numbers of crime incidents between the two neighborhoods might not be able to be minimized to zero when it is inefficient to making arrests in neighborhood 2 (l_2 is very small). On the other hand, in the case that the total enforcement E is not sufficient to collapse criminal activities in the neighborhood with highest $\gamma\sqrt{k}$ value ($k_1 \geq l_1$), the difference of the numbers of crime incidents between the two neighborhoods might not be able to be minimized to zero when it is inefficient to make arrests in neighborhood 1 (l_1 is very small).

We note that, the optimal solution in problem 7.2, which minimizes the difference, also minimizes the maximum of weighted crime numbers between the two neighborhoods. This can be seen clearly as the following three cases: (1) If both neighborhoods in the optimal solution of Problem 7.2 are partially filled, the objective value of Problem 7.2 is zero. This solution also minimizes the maximum since relocating enforcement from one neighborhood to the other will increase the weighted crime number in the neighborhood that loses enforcement and then thereby increases the maximum. (2) If neighborhood 1 is partially filled (receives all enforcement) and neighborhood 2 receives no enforcement in the optimal solution of Problem 7.2, then the maximum weighted number of crimes between the two neighborhoods must happen in neighborhood 1. Otherwise, relocating some enforcement from neighborhood 1 to neighborhood 2 will reduce the objective value of Problem 7.2. Hence since all enforcement is in neighborhood 1, then this solution minimizes the maximum weighted number of crimes between the neighborhoods. (3) If one neighborhood is fully filled in the optimal solution of Problem 7.2 then the difference of the weighted number of crimes between the two neighborhoods is equivalent to the weighted number of crimes of the other neighborhood. Hence minimizing the difference is equivalent to minimizing the maximum of weighted crime numbers between the two neighborhoods.

Corollary 7.3: If $\gamma_1 = \gamma_2$, and the wealth of the two neighborhoods is equal, i.e., $w_1 = w_2$ or $k_1 = k_2 = k$, then difference in weighted number of crimes between the two neighborhoods can always be minimized to zero, and the optimal amount of enforcement a neighborhood receives is in reverse proportion to its efficiency of making arrest, i.e., $x_1/x_2 = \alpha_2/\alpha_1$.

Proof: From the results of Theorem 7.2, the two exceptions cannot happen. Hence, the objective value can always to be minimized to zero. Also, solving $G_2(x) = 0$ (see proof of Theorem 7.2) with the assumption that $k_1 = k_2 = k$ yields $x_1 = x = l_2/(l_1+ l_2) = \alpha_2/(\alpha_1+ \alpha_2)$ and $x_2 = 1 - x = l_1/(l_1+ l_2) = \alpha_1/(\alpha_1+ \alpha_2)$.

3.7.3 Non-Dominated Solutions

In previous sections we have found the optimal solutions for both of the objectives, minimizing the weighted sum of crime numbers and minimizing the weighted difference of crime numbers between two neighborhoods. In this section, we are interested in *non-dominated solutions*, solutions for which no other solution is better under both objectives. We may formally define non-dominated solutions as follows. A feasible solution x is non-dominated by another feasible solution y if $F_1(x) < F_1(y)$ or $|F_2(x)| < |F_2(y)$ or $F_1(x) = F_1(y)$ and $|F_2(x)| = |F_2(y)|$. Feasible solution

x is called non-dominated if it is non-dominated by all feasible solutions. Also, a feasible solution is called dominated if it is not a non-dominated solution. Non-dominated solutions become important when multiple objectives are considered and it is difficult to decide which objectives are particularly most critical. In this situation, decision makers should eliminate dominated solutions, for which there exist solutions better under both objectives, and consider only non-dominated solutions.

In some circumstances, there may not be any non-dominated solutions. For example, without loss of generality, we assume neighborhood 1 is the neighborhood with highest $\gamma\sqrt{k}$ value. No non-dominated solutions exist when it is impossible to equalize the weighted crime numbers of the neighborhoods in Problem 7.2 and the optimal solution in Problem 7.1 gives neighborhood 1 highest priority to receive enforcement. In this case, the optimal solutions for both of the problems are exactly the same: a crackdown on neighborhood 1. Therefore, other solutions are dominated by this optimal solution (see Figure 7.1, the case that no non-dominated solutions exist).

On the other hand, in some cases, every feasible solution can be a non-dominated solution. These situations may happen especially when the optimal solution of Problem 7.1 is the worst solution of Problem 7.2 and vise versa. One such situation occurs when $\lambda_1 = \lambda_2$, $\gamma_1 = \gamma_2$ and two neighborhoods have exactly the same conditions, i.e., $k_1 = k_2$ and $l_1 = l_2$. In this special case, the optimal policy to minimize the difference of crimes between the two neighborhoods is to evenly split the enforcement ($x = 0.5$). However, this natural policy maximizes the total number of crimes of the two neighborhoods. Also, either one of the two one-neighborhood crackdown policies (cracking down neighborhood 1 or neighborhood 2) is optimal for Problem 7.1 but the worst for Problem 7.2. Along with the symmetry of both objective functions and their monotone property on the two wings (see Figure 7.1, the case that every feasible solution is non-dominated), it can be easily seen that none of the feasible solutions dominates each other. Therefore, we conclude that every solution is non-dominated

As illustrated in Figure 7.1, we have seen that under certain circumstances there may be no non-dominated solutions at all. On the other hand, for some circumstances, every feasible solution can be a non-dominated solution. Generally, there is no analytic method that can easily find all non-dominated solutions. However, since we only consider two objective functions in two neighborhoods at this stage, the non-dominated solutions can be visually detected with the help of the plots of all objective values. It shows that if there are any non-dominated solutions, at least some of them are in an interval adjacent to the optimal solution of Problem 7.1.

3.8. Optimal Allocation Policies in Multiple Neighborhoods

We have completed the discussion of the enforcement allocation problem in the case of two neighborhoods. Usually, of course, there are more than two neighborhoods of interest. This section generalizes from the case of two neighborhoods to that of multiple neighborhoods. We develop computationally efficient algorithms to generate optimal allocation policies. Some results formed for the case of two neighborhoods are helpful in the situation of multiple neighborhoods. However, since the objective functions are highly non-linear, they cannot be applied directly.

3.8.1 Minimizing Total Number of Crime Incidents

If E is not sufficient to collapse criminal activities in all neighborhoods at the same time, i.e., $E < \quad {}_{i=1..n}(k_i/4\alpha_i\beta)^{+}$, then the mathematical formulation can be written as:

<Problem 8.1> Minimize $\sum_{i=1}^{n} \lambda_i n_i$

subject to $\sum_{i=1}^{n} x_i = 1;$

$$0 \le x_i \le (k_i/l_i)^+, \text{ for } i = 1, 2, \ldots, n.$$

where $k_i = [ln(cw_i/m)]^2$, $l_i = 4\alpha_i\beta E$, for $i = 1, 2, \ldots, n$.

We assume that at most one of the neighborhoods is not able to be collapsed, i.e., satisfies $k/l \ge 1$. This assumption can be made without loss of generality by recognizing that if there is more than one neighborhood with $k/l \ge 1$, then it suffices to select only the neighborhood with largest $\lambda(\sqrt{k} - \sqrt{k-l})$ for the analysis. The other such neighborhoods will be allocated zero enforcement under the optimal policy (see case 3 of Theorem 7.1).

For any pair of neighborhoods, Theorem 7.1 suggests a one-neighborhood crackdown policy, where sufficient resources are allocated to one of the two neighborhoods to attempt to collapse criminal activities there. The other neighborhood then receives consideration for any additional resources. Therefore, considering all n neighborhoods we know that an optimal allocation policy has all but one of the x_i's either equal to 0 ("empty") or $(k_i/l_i)^+$ ("filled"). The "partially filled" neighborhood, call it q^* with $0 < x_{q^*} \le k_{q^*}/l_{q^*}$, results from the remaining resources after collapsing the filled neighborhoods. We can now assume that we need only k_i/l_i resources to fill any other neighborhood i, as the infinitesimal resources to complete the fill can be viewed as part of the "excess" resources assigned to q^*. Since the optimal solution consists of a set of fully filled neighborhoods, I^*, one partially filled neighborhood, q^*, and a set of empty neighborhoods, O*, we focus our attention on the subproblem of selecting the fully filled and empty neighborhoods, given the partially filled neighborhood q^*.

Given a designated partially filled neighborhood q^*, we seek a binary allocation policy for the remaining neighborhoods. Here y_i = 0 represents leaving neighborhood i empty with number of crimes $n_i = \sqrt{k_i} / \beta$, and y_i = 1 represents fully filling neighborhood i so that n_i = 0. Thus, the objective of our subproblem is to maximize the collapsed criminal activities subject to a constraint on enforcement resources via

<Problem 8.1’> Maximize $\displaystyle {}_{i\in C'} \lambda_i \sqrt{k_i}\, y_i$

subject to $\displaystyle {}_{i\in C'} \frac{k_i}{l_i^{'}} y_i \le 1;$

$y_i \in \{0, 1\}$ for $i \in C'$.

where $C' = \{1, 2,\ldots,n\} \setminus \{q^*\}$, $E' = E - E_q^*$ (where E_q^* represents the currently unknown amount of enforcement allocated in neighborhood q^* under an optimal policy) and $l_i' = 4\,\alpha_i \beta E'$.

Here we have transformed the original Problem 8.1 via the binary variable $y_i = x_i l_i'/k_i$. The resulting Problem 8.1' is the well-known knapsack problem (cf. Taha1975). However, since all excess resources will be assigned to q^*, at optimality l_i' will choose to ensure that the knapsack constraint ${}_{i\in C'} \frac{k_i}{l_i^{'}} y_i \le 1$ is tight. Without loss of generality, we renumber the neighborhoods such that

$$\frac{\lambda_1\sqrt{k_1}}{k_1 / l_1^{'}} \geq \dots \geq \frac{\lambda_p\sqrt{k_p}}{k_p / l_p^{'}} \geq \dots \geq \frac{\lambda_{n-1}\sqrt{k_{n-1}}}{k_{n-1} / l_{n-1}^{'}}, \qquad \text{i.e.,}$$

$$\lambda_1\alpha_1 / \sqrt{k_1} \geq \dots \geq \lambda_p\alpha_p / \sqrt{k_p} \geq \dots \geq \lambda_{n-1}\alpha_{n-1} / \sqrt{k_{n-1}},$$

where p is the integer ($0 \leq p \leq n$) such that $\sum_{i=1}^{p} \frac{k_i}{l_i^{'}} = 1$. Dantzig (1957) has shown that this ratio ordering provides the optimal *fractional* solution to the knapsack problem, where at most one variable is assigned the fractional slack of the knapsack constraint. Since there is no slack in our constraint at optimality, this ratio ordering provides the optimal *integer* solution

$$y_j^* = \begin{cases} 1 & \text{if} \quad j = 1,\dots,p; \\ 0 & \text{if} \quad j = p+1,\dots,n-1. \end{cases}$$

Recall our assumption that the optimal solution is $x_i = k_i/l_i^{'}$ if $x_i \in I^*$ and $x_i = 0$ if $x_i \in O^*$, this implies $I^* = \{1, 2, ..., p\}$and $O^* = \{p+1, p+2, ..., n-1\}$.

From the discussion so far, we can easily determine I^* and O^* given the neighborhood that should be partially filled and how much enforcement should be allocated in it are known. This leads to the following polynomial algorithm for finding the optimal allocation policy for Problem 8.1.

Algorithm 8.1

1. Number neighborhoods such that $\lambda_1\alpha_1 / \sqrt{k_1} \geq \lambda_2\alpha_2 / \sqrt{k_2} \geq \dots \geq \lambda_n\alpha_n / \sqrt{k_n}$; let $C = \{1,2,\dots,n\}$.

2. Let $q = 1$ and $f^* = \infty$ for $i = 1,\dots,n$.

3. Find s and p, where s is the smallest integer such that $\sum_{i=1..s,i\neq q} \frac{k_i}{l_i} + \frac{k_q}{l_q} \geq 1$ and p is the smallest integer such that $\sum_{i=1..p,i\neq q} \frac{k_i}{l_i} \geq 1$.

4. Let $E_i = k_i/4\alpha_i\beta$ for $i = 1, \ldots, s$ and $E_i = 0$ for $i = s+1, \ldots, n$.

5. Let $E_q = E - \sum_{i=1..s,i\neq q} E_i$.

6. Calculate $f = \sum_{i=s+1...n,i\neq q} \frac{\lambda_i \sqrt{k_i}}{\beta} + \lambda_q \frac{\sqrt{k_q} + \sqrt{k_q - l_q x_q}}{2\beta}$. If $f < f^*$, let $f^* = f$; $I^* = \{1, 2, ..., s\}\backslash\{q\}$; and $O^* = \{s+1, s+2, ..., n\}\backslash\{q\}$.

7. Let $s = s + 1$; if $s < p$, go to Step 4.

8. Let $q = q + 1$; if $q \leq n$, go to Step 3.

9. f^* is the optimal objective value and the optimal policy is to collapse criminal activities in the neighborhoods of I^* and allocate the remaining enforcement to neighborhood q.

Step 1 in Algorithm 8.1 sorts the neighborhoods in a decreasing order on $\lambda\alpha/\sqrt{k}$. The neighborhoods with larger $\lambda\alpha/\sqrt{k}$ value have higher priority to receive enforcement if the partially filled neighborhood is known. The complexity of this sorting process is O(n ln(n)). The loop starting from Step 2 and ending at Step 8 enumerates all the possibilities of the partially filled neighborhood from 1 to n. The complexity of this loop is O(n). Step 3 determines s and p for the choice of q. Here, if q is partially filled, all neighborhoods $i \leq s$ must be fully filled because of their higher priority to receive enforcement; and all neighborhoods $i \geq p$ must be empty because of their lower priority to receive enforcement and the limited amount of enforcement. The loop starting from Step 4 and ending at Step 7 enumerates all the possibilities of fully filled neighborhoods under the assumption that q is the partially filled neighborhood. The complexity of this loop is O(n). Since we have a loop of enumerating all possibilities of fully filled neighborhoods under the loop which enumerates all possibilities of the partially filled neighborhood. The complexity of running these two loop is O(n^2). Step 5 computes the amount of enforcement which neighborhood q should receive in the case that the fully filled neighborhoods have been determined. Step 6 calculates the objective function of the current allocation policy and updates the optimal solution if necessary. When we run out of all possibilities, Step 9 concludes the algorithm. Those steps only need one calculation time. Hence the complexity of Algorithm 8.1 is O(n^2).

In some special circumstances, the problem can be easy. The following corollaries identify two such cases.

Corollary 8.1: If the objective equally weights crime numbers among neighborhoods, and the neighborhoods have equal wealth level, then the neighborhood where enforcement is more efficiently applied to arrest has higher priority to receive enforcement. That is, the optimal allocation policy is to allocate enforcement to the neighborhood with largest α as much as possible until the criminal activities in the neighborhood can be collapsed. If there is still enforcement available, allocate the remaining enforcement to the neighborhood with second largest and so on until all enforcement is allocated.

Proof: Directly followed from Corollary 7.1.

Corollary 8.2: If the objective equally weights crime numbers among neighborhoods and the efficiency of making arrest in the neighborhoods is equal, then the poorer neighborhood has higher priority to receive enforcement. That is, the optimal allocation policy is to allocate enforcement to the poorest neighborhood until the criminal activities in the neighborhood can be collapsed. If there is still enforcement available, allocate the remaining enforcement to the next poorest neighborhood and so on until all enforcement is allocated.

Proof: Directly followed from Algorithm 8.1 and property 2 of Corollary 7.2.

3.8.2 Minimizing Crime Disparity

This section discusses the optimal policy of minimizing the disparity of weighted crime numbers among multiple neighborhoods. The purpose here is to fairly distribute enforcement to the neighborhoods. The problem can be formulated mathematically as:

<Problem 8.2> Minimize $\max_{i,j}\{\gamma_i n_i - \gamma_j n_j\}$

subject to $\sum_{i=1}^{n} x_i = 1;$

$0 \le x_i \le (k_i / l_i)^+$, for $i = 1, 2, \ldots, n$.

The optimal solution in Problem 8.2 can be zero if and only if there exists $0 \le x_i \le (k_i / l_i)^+$ and $0 \le x_j \le (k_j / l_j)^+$ such that $\gamma_i(\sqrt{k_i} + \sqrt{k_i - l_i x_i}) = \gamma_j(\sqrt{k_j} + \sqrt{k_j - l_j x_j})$ for each pair of i and j. After some basic algebraic calculations,

$$\gamma_i(\sqrt{k_i}+\sqrt{k_i-l_i x_i}\,)=\gamma_1(\sqrt{k_1}+\sqrt{k_1-l_1 x_1}\,)$$

$$\Leftrightarrow \qquad x_i=\frac{k_i-(\gamma_1\sqrt{k_1}\,/\,\gamma_i+\gamma_1\sqrt{k_1-l_1x_1}\,/\,\gamma_i-\sqrt{k_i}\,)^2}{l_i} \text{ for } i=2,\ldots,n.$$

Hence, to check if there is a possibility to equalize the crime numbers among the neighborhoods, we should first solve the following quadratic equation in x_1

$$x_1+\sum_{i=2}^{n}\frac{k_i-(\gamma_1\sqrt{k_1}\,/\,\pi_i+\gamma_1\sqrt{k_1-l_1x_1}\,/\,\gamma_i-\sqrt{k_i}\,)^2}{l_i}=1.$$

If the solution satisfies $0 \le x_i \le k_i/l_i$ for i = 1, 2,...,n, then we may equalize the crime numbers of the neighborhoods by letting x_i be the proportion of enforcement that neighborhood i receives.

If the solution does not satisfy the constraints $0 \le x_i \le k_i/l_i$ for each i, then there is no possibility to minimize the objective value to zero. We shall now focus on solving the problem in this case.

Recall that, according to our model, when no enforcement is allocated in a neighborhood, the crime activities of the neighborhood are limited by the wealth level, and the crime number reaches $\sqrt{k}\,/\,\beta$ in equilibrium. Also, if we can increase enforcement pressure to the level that reduces the crime number below $\sqrt{k}\,/\,2\beta$ in equilibrium, then the criminal activities will collapse. Therefore, a non-zero equilibrium must be between $\sqrt{k}\,/\,2\beta$ and $\sqrt{k}\,/\,\beta$ depending on the amount of enforcement applied. We call $\sqrt{k}\,/\,\beta$ and $\sqrt{k}\,/\,2\beta$ the *saturated crime number* and *marginal crime number* of the neighborhood, respectively.

Consider two neighborhoods i and j with $\gamma_i\sqrt{k_i}>2\gamma_j\sqrt{k_j}$. The two neighborhoods never have the same weighted crime number unless both of them are free of crimes since the weighted marginal crime number in neighborhood i ($\gamma_i \sqrt{k_i} / 2\beta$) is greater than the weighted saturated crime number in neighborhood j ($\gamma_j \sqrt{k_j} / \beta$).

Without loss of generality, it is assumed that $\gamma_1 \sqrt{k_1} \geq \gamma_2 \sqrt{k_2} \geq \ldots \geq \gamma_n \sqrt{k_n}$. Suppose p is the neighborhood with the maximum weighted crime number, n_p^*, among the neighborhoods under the optimal policy. Then the neighborhoods with $\gamma \sqrt{k} > 2\gamma_p \sqrt{k_p}$, say neighborhoods 1, ..., s, must be free of crime in the optimal policy. This is because the weighted marginal crime number in these neighborhoods is greater than the weighted saturated crime number of neighborhood p. Hence, we should have $x_i^* = (k_i/l_i)^+$ for $i = 1, ..., s$. Furthermore, any partially filled neighborhood must have the same weighted crime number as neighborhood p, since if a neighborhood is partially filled and has weighted crime number less than $\gamma_p n_p^*$, then moving a small amount of enforcement resources from it to the neighborhoods with the largest crime numbers will decrease our objective value. Also, the neighborhoods with zero enforcement allocated must have $\gamma \sqrt{k} / \beta < \gamma_p n_p^*$. We now can conclude that there exist s and p such that the optimal solution is

$x_i^* = (k_i/l_i)^+$ for $i = 1, ..., s$,

x_i^* for $i = s+1, ..., p$, where x_i^* is the fraction such that $\gamma_i n^*_i = \gamma_p n_p{}^*$,

$x_i^* = 0$ for $i = p+1, ..., n$.

The following greedy algorithm gives the optimal solution.

Algorithm 8.2

1. Sort the neighborhoods by their weighted wealth level: $\gamma_1 \sqrt{k_1} \geq \gamma_2 \sqrt{k_2} \geq \ldots \geq \gamma_n \sqrt{k_n}$.

2. Let i = 1; $E^o = E$.

3. Find the greatest integer m such that $\sum_{j=i}^{m} \dfrac{\gamma_m(\sqrt{k_j k_m} - \gamma_m k_m / \gamma_j)}{\gamma_j \alpha_j \beta} \leq E^o$.

4. If $\gamma_i \sqrt{k_i} > 2\gamma_m \sqrt{k_m}$, let $E^0 = E^0 - \dfrac{k_i}{4\alpha_i \beta}$; $x_i = \left(\dfrac{k_i}{l_i}\right)^+$; $i = i+1$ and go to Step 3. Otherwise,

   solve $$x_i + \sum_{j=i+1}^{m} \frac{k_j - (\gamma_i \sqrt{k_i} / \gamma_j + \gamma_i \sqrt{k_i - l_i x_i} / \gamma_j - \sqrt{k_j}\,)^2}{l_j} = 1;$$

   $$x_j = \frac{k_i - (\gamma_i \sqrt{k_i} / \gamma_j + \gamma_i \sqrt{k_i - l_i x_i} / \gamma_j - \sqrt{k_j}\,)^2}{l_j} \text{ for } i = i, \ldots, m.$$

   for x_i, ..., x_m; let x_{m+1} = ... = x_n = 0, and stop.

5. The final solution is the proportion of optimal enforcement allocation policy.

Step 1 in the algorithm sorts the neighborhoods in the order their weighted wealth level; neighborhoods that are more affluent and are highly weighted have higher priority to receive enforcement. Solving $\gamma_i(\sqrt{k_i} + \sqrt{k_i - 4\alpha_i \beta E_i}\,) / 2\beta = \gamma_j \sqrt{k_j} / \beta$ for E_i, we have $E_i = \gamma_j\left(\sqrt{k_i k_j} - \gamma_j k_j / \gamma_i\right) / (\gamma_i \alpha_i \beta)$. This represents the amount of enforcement that should be allocated in the higher priority neighborhood (neighborhood i) such that the weighted crime number in the neighborhood is equal to the saturated weighted crime number of the lower priority neighborhood (neighborhood j), if desirable. Hence, Step 3 evaluates how low the weighted crime numbers can go. Note that we might allocate too much enforcement in the neighborhoods since we only have to push the crime number of the neighborhood less than the marginal weighted crime number to collapse the criminal activities within it.

In Step 4, if $\gamma_i \sqrt{k_i} > 2\gamma_m \sqrt{k_m}$, we should allocate $(k_i/4\alpha_i\beta)^+ \approx k_i/4\alpha_i\beta$ in neighborhood i to collapse the criminal activities in the neighborhood. With this information, we let $x_i = (k_i/l_i)^+$ and run the algorithm again with the remaining enforcement $E^0 - k_i/4\alpha_i\beta$. On the other hand, if $\gamma_i \sqrt{k_i} \leq 2\gamma_m \sqrt{k_m}$ by the continuity of crime numbers in neighborhood j as a function of enforcement on $[0, k_j/4\alpha_j\beta)$ for $j = i, ..., m$, the equation system must have a unique solution and this provides the proportion of optimal enforcement allocation policy.

One important proposition, which can be generated from two-neighborhood case with similar as arguments in Section 7.2, is that the optimal solution - which minimizes the maximum difference of weighted crime numbers among the neighborhoods - is also the optimal solution which minimizes the maximum weighted number of crime incidents.

3.9. A Sample Case Study

In this section we apply our model to a burglary dataset in the City of Buffalo in the State of New York. We estimate the expected return of committing a burglary among the Buffalo Police Department's (BPD's) current five patrol districts and determine between which districts crime displacement is most likely to occur. We then suggest an allocation policy that equalizes the number of burglaries among the five districts.

Buffalo is the second-largest city in the State of New York with a population of about 330,000. The City of Buffalo is located on the shores of Lake Erie and comprises approximately 40 square miles. The Buffalo Police Department consists of 533 police officers and deploys between 33 and 54 police cars at any point of time.

According to the BPD administrators, the BPD has determined command district boundaries from experience. Typically, boundaries lie along major streets and are drawn so that the crime numbers are approximately the same in each district. Through a planning process that involved police chiefs and legislatures in the city government, the design of the BPD's geographical commands consists of the city's five police command districts named A-E (see Figure 9.1). The patrolling operations of each district are conducted essentially independent of other districts. That is, only very few instances of particular incidents require the patrol cars to cross the district boundaries.

There are several approaches that can be taken to equalize the number of crimes or, with the similar concern, to minimize the disparity of workload in the districts. One approach, which we use in this paper, is reallocating enforcement resources such that the criminals in the crime-ridden districts face more pressure and then we can hopefully reduce crime numbers in those districts. A second approach is redefining the districts. A third approach is a combination of both. D'Amico et al. (2000) apply simulated annealing method to redefine the boundaries and develop a new allocation policy in the City of Buffalo in order to minimize the workload disparity among the districts. Unavoidably, the new allocation policy tends to assign more enforcement resources to the crime-ridden districts. Without incorporating displacement effects, the potential problem we may face is the criminals in relatively crime-ridden districts might displace to the districts (originally relatively crime-free) that receive less enforcement in the new allocation policy.

Hence, it may have some unexpected results. From BPD's experience, this difficulty did arise when the current district design and allocation policy, which was designed to approximately equalize the crime number in each district, were implemented. As we can see from Table 2, it appears that the crime numbers, at least in the case of burglary, in the districts are not equal at all.

During the 17 weeks period between the first of February and the 31st of May in 1998, there were about 1,695 burglaries that took place in the city (around 100 burglaries per week). Table 2 lists all the burglary data from BPD dataset over a 17-week period. We use patrol hours as the measurement of enforcement level and the patrol hours are in a week-based unit since BPD's patrol schedule is different from day to day and it cycles weekly. The burglary crime numbers, however, are taken from a seventeen-week period. It should be noticed that no matter how we scale the variables, the values of parameters might differ for different scales, but the relative allocation results should not be affected.

According to our assumption that $P_A(E/n) = 1 - exp(-\alpha(E/n))$, the parameter α, which measures the arrest ability of a district, can be easily estimated from the available data for the five districts. They are .0300, .0612, .0462, .0288 and .0472 respectively. The result provides us with an intuition of geographical impact on arrest ability. It seems that districts with larger area have worse arrest ability (for relative district area, refer the map of the City of Buffalo in Figure 9.1). This outcome is expected since patrolling becomes inefficient if the patrol area is large. Also, District B, where the BPD Headquarters allocated, is most efficient to make arrests.

In equilibrium, the expected return in all the districts should reach the opportunity cost of a crime, i.e., $P_A\, c\omega\, exp(-\beta n) = m$. Here, P_A, w and n vary in different districts and are available from the dataset; however, c and β are neither available nor able to be estimated from the information we have. Since our study is limited on one type of crime, the values of the parameters c and β are the same for the five districts. If we assume c = .01 and β = .00085, the expected return in Districts A-E are $157, $79, $79, $140 and $137 respectively.

If a displacement effect does occur between two districts, according to our model, the expected return should be equal in equilibrium. Our results suggest that displacement effects occurred between Districts B-C and D-E. This conclusion seems reasonable since Districts D and E, located adjacent to each other in north Buffalo, are considered relatively crime-free and have similar economic conditions. Similarly, Districts B-C, located adjacent to each other in the central city, are considered relatively crime-ridden and have the same median income. District A has the largest expected return since criminals do not impede each other's efforts significantly. (The fewest burglaries happened in this district.)

Finally, we provide an optimal allocation policy in Table 3 which distributes total available enforcement resources, 6414.8 patrol hours, to the five districts such that the numbers of burglaries among them are as equal as possible. We assume displacement effect occurs between B-C and D-E and the opportunity cost of a crime for Districts A-E are $157, $79, $79, $ 138.5 and $138.5, respectively. Note that criminals in District A have the highest opportunity cost of a crime so that the saturated crime number in A is the lowest. This implies that A should have the lowest priority to receive enforcement.

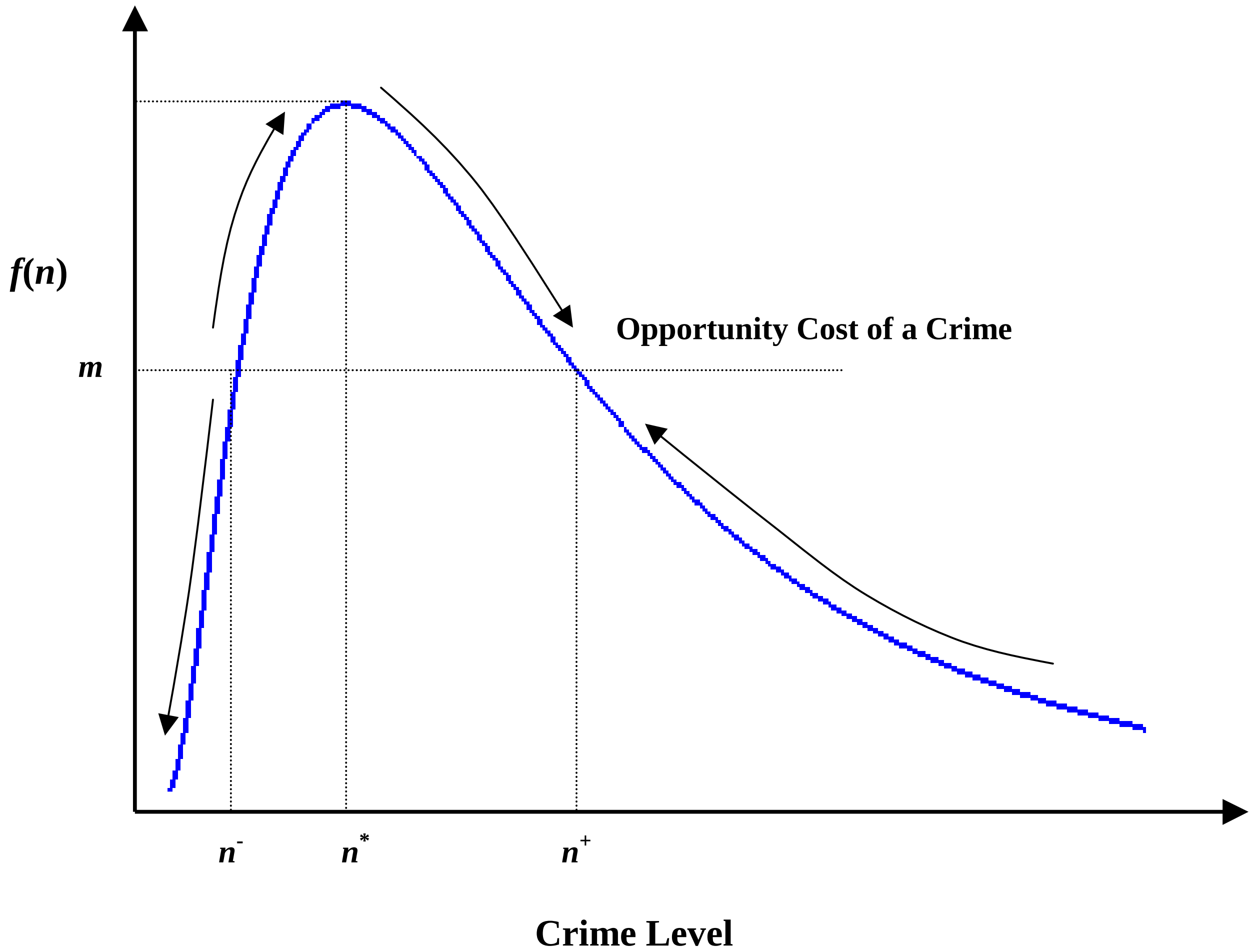


Figure 2.1. Plot of Expected Monetary Return as a Function of the Crime Number

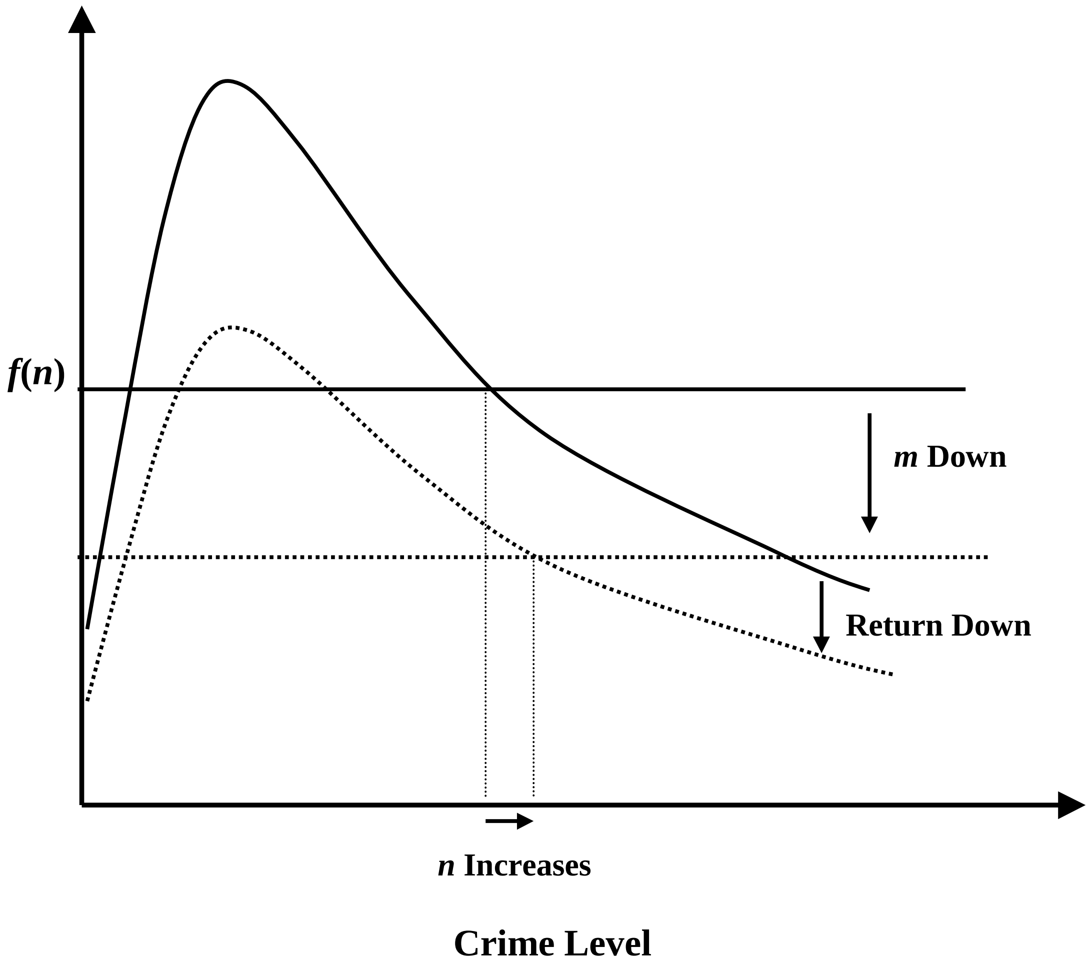


Figure 4.1. The Effect of Poverty on the Crime Numbers

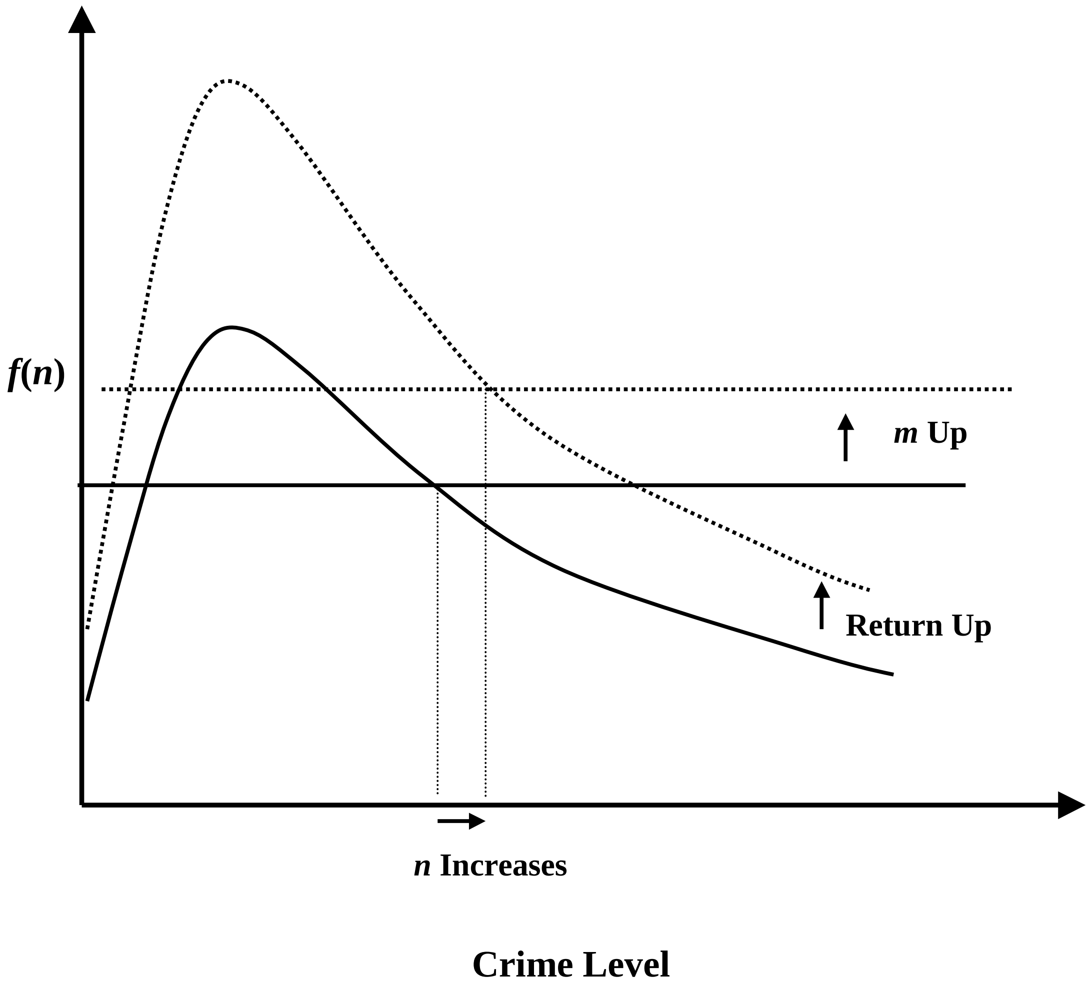


Figure 4.2. The Effect of Affluence on the Crime Numbers

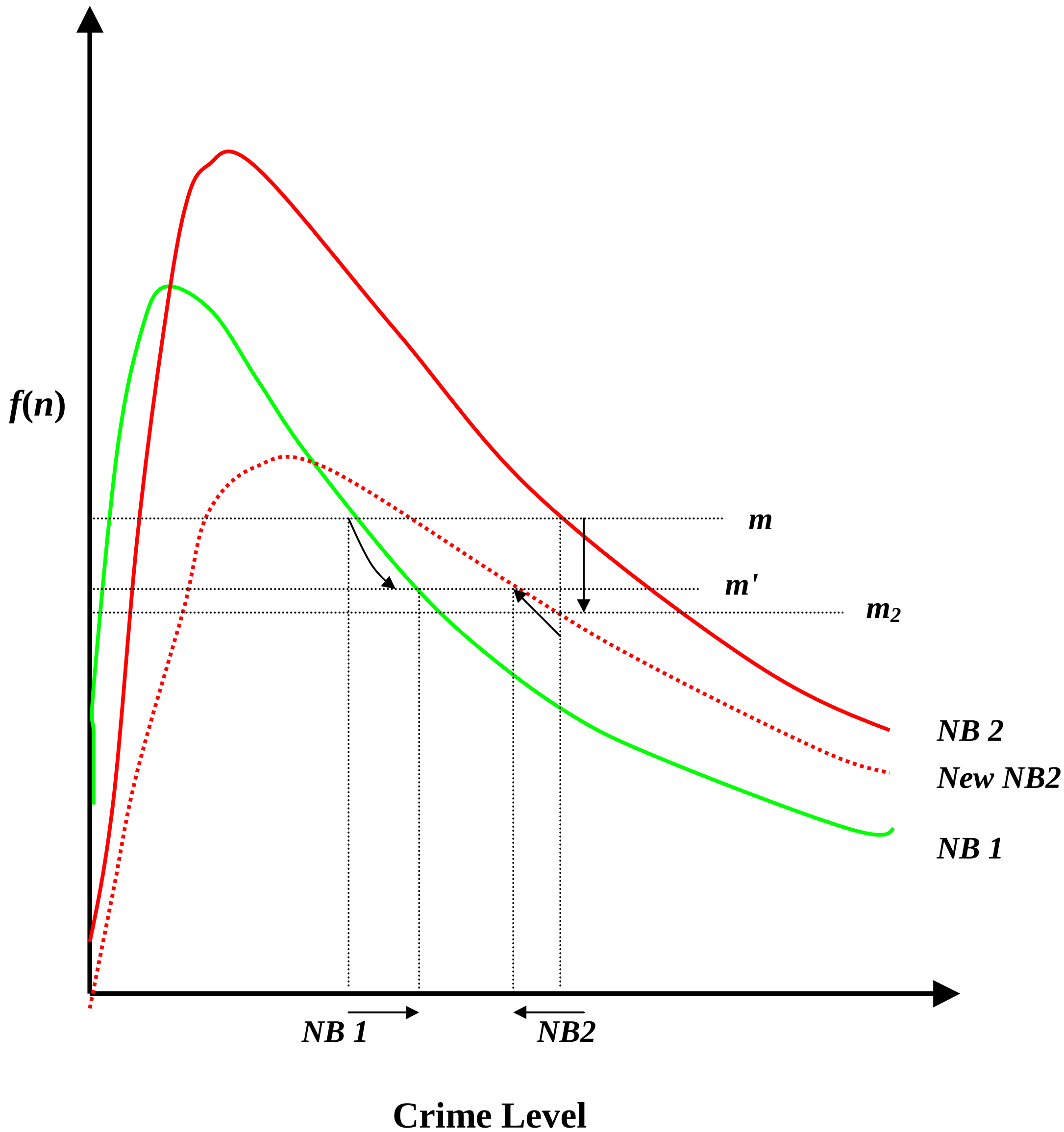
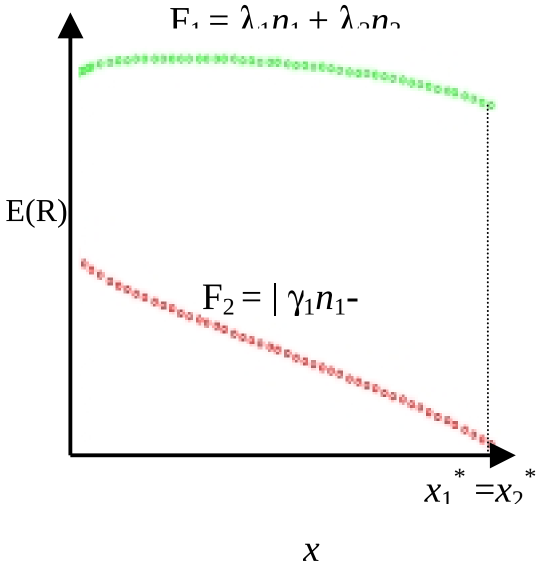
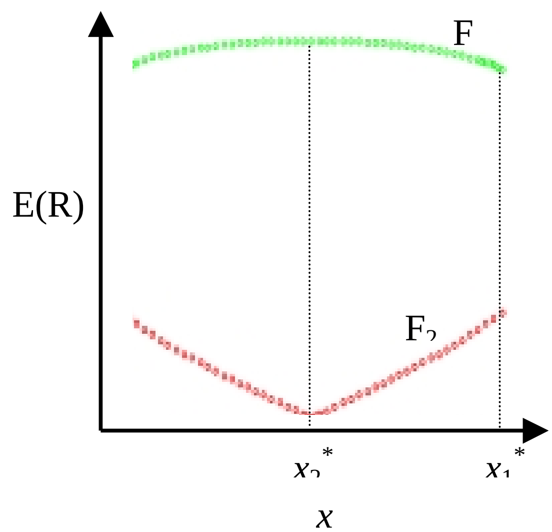


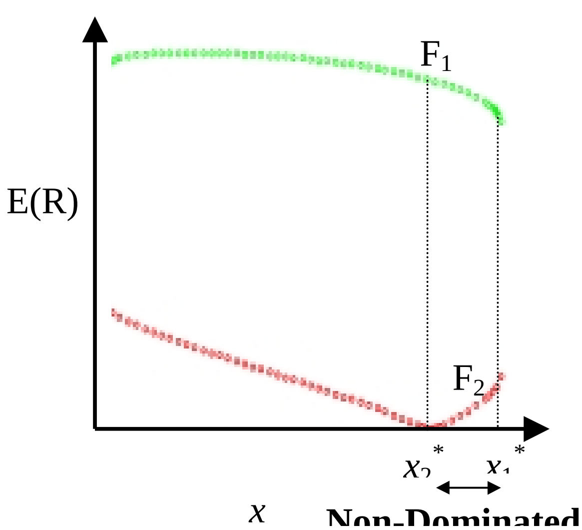
Figure 5.1. Geographical Displacement Phenomenon



Situation Where No Non-Dominated Solutions Exist



Situation Where Every Feasible Solution is Non-Dominated



Non-Dominated Solutions form an Interval Adjacent to
*

Figure 7.1. Non-Dominated Solutions in Different Circumstances

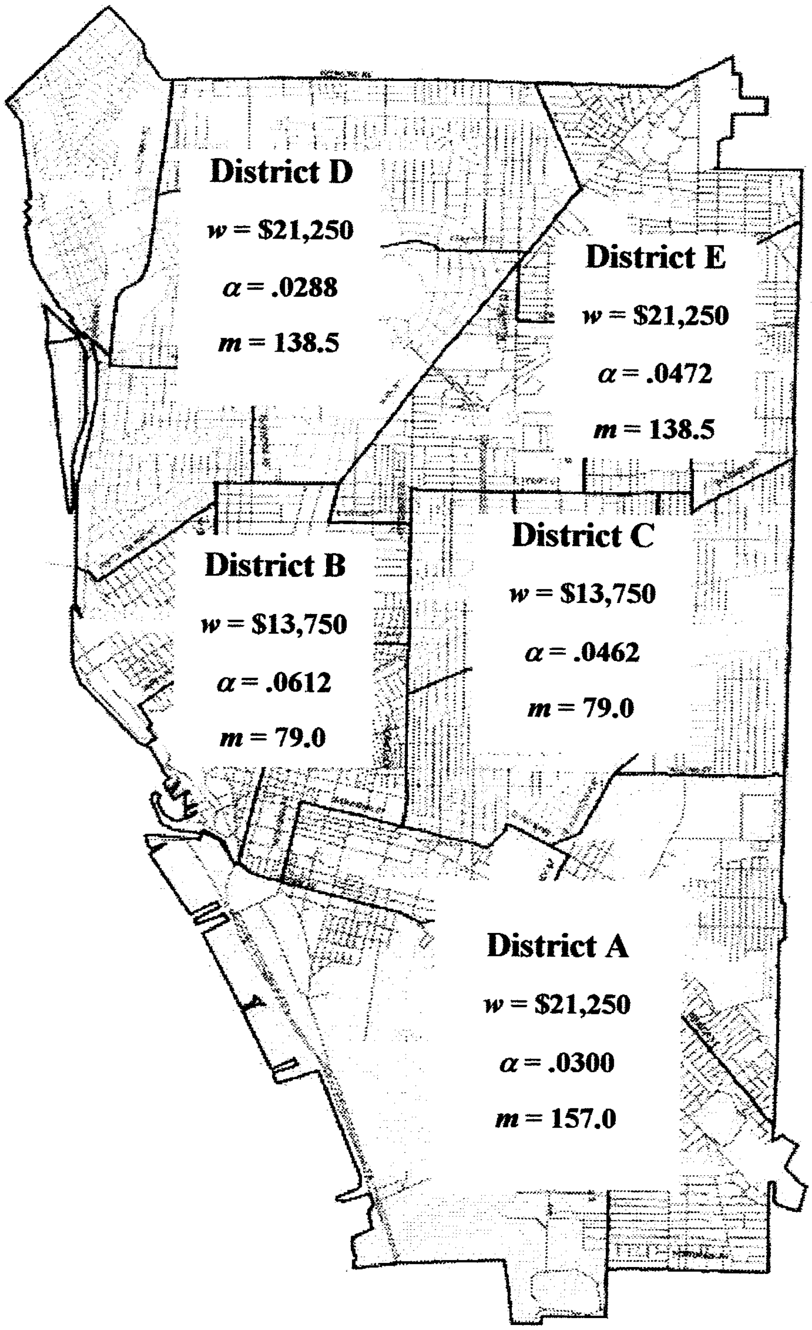


Figure 9.1. The Five Districts in the City of Buffalo

References

A. Baveja, R. Batta, J.P. Caullkins and M.H. Karwan. Modeling the Response of Illicit Drug Markets to Local Enforcement. *Journal of Socio-Economic and Planning Sciences*, 27, 73-89 (1993).

A. Barnett, J.P. Caulkins and M.D. Maltz. Crime and Justice. In *Encyclopedia of OperationsResearch and Management Science*, 2nd edition, S.I. Gass and C.M. Harris (ed.), Kluwer Academic Publishers, Boston (2001).

J. Belknap. The Economic-Crime Link. *Criminal Justice Abstracts*, 21, 140-54 (1989).

A. Blumstein. Crime Modeling. *Operations Research*, 50 (1), 16-24 (2002).

J. P. Caulkins. The Distribution and Consumption of Illicit Drugs: Some Mathematical Models and Their Policy Implications. Unpublished *Ph.D. dissertation*, Massachusetts Institute of Technology, Cambridge, Mass. (1990).

D. B. Cornish, and R. V. Clarke. Understanding Crime Displacement: An Application of Rational Choice Theory". *Criminology*, 25, 933-947 (1987).

J. Deutsch, S. Hakim, and J. Weinblatt. Interjurisdictonal Criminal Mobility: A Theoretical Perspective. *Urban Studies*, 21, 451-458 (1984).

S. J. Deutsch, J. J. Jarvis, and R. G. Parker. A Network Flow Model for Forecasting and Evaluating Criminal Displacement. *Evaluation Quarterly*, 3, 219-235 (1979).

S. Freeman, J. Grogger, and J. Sonstelie. The Spatial Concentration of Crime. *Journal of Urban Economics*, 40, 216-231 (1996).

T. Gabor. Crime Displacement and Situational Prevention: Toward the Displacement of Some Principles. *Canadian Journal of Criminology*, 32, 41-73 (1990).

T. Gabor, The crime displacement hypothesis: An Empirical Examination. *Crime and Delinquency*, 26, 390-404 (1981).

P. W. Greenwood, J. M. Chaiken and J. Petersilia. *The Criminal Investigation Process*. D.C. Health, Lexington. MA. (1977).

R. W. Jongman, F. Weerman and L. Krees. Maatschappelijke Tweedeling Criminaliteit. *Tijdschrift voor Criminologie*. 33, 29-41 (1991).

M. A. R. Kleiman. Crackdowns: The Effects of Intensive Enforcement on Retail Heroin Dealing. In *Street Level Drug Enforcement: Examining the Issues*, M. R. CHAIKEN (ed.), National Institute of Justice, Washington, D.C. (1988).

B. Lateef. Helicopter Patrol in Law Enforcement-- An Evaluation. *Journal of Police Science and Administration*, 2, 62-65 (1974).

M.D. Maltz. Operations Research in Studying Crime and Justice: Its History and Accomplishments. In *Operations Research and Public Systems*, S.M. Pollock, A. Barnett and M. Rothkopf (eds.), Elsevier, Amsterdam (1994).

T. Reppetto. Crime Prevention and the Displacement Phenomenon. *Crime and Delinquency*, Vol. 22,166-177 (1976).

S. Tyrpak. *Newark High-Impact Anti-Crime Program: Street Lighting Project Interim Evaluation Report*. Newark, NJ: Office of Criminal Justice Planning (1975).

J. J. M. van Dijk. Understanding Crime Rates: On the Interactions between the Rational Choices of Victims and Offenders. *The British Journal of Criminal*, 34, 105-121 (1994).

J. J. M. van Dijk, and U. Zvekic. Surveying Crime in the Global Village: Assumptions, Experiences and Ultimate Goals, in A. Alvazzi del Frate, U. Zvekic, and J. J. M. van Dijk. eds, *Understanding Crime: Experiences of Crime and Crime Control*: 365-77. Rome:UNICRI. (1993).

S. J. Wang, R. Batta and C. M. Rump. Optimal Allocation of Police Enforcement among Adjacent Neighborhoods (2000). Under Review.

Chapter 4. Data-Driven Framework for Understanding Criminal Activity

4.1 Introduction

There has been a recent resurgence in use of OR/MS models to manage and control criminal activity. However, much of this research is either (1) purely analytical – i.e. difficult to validate through real data, or (2) it is entirely statistical in nature that is constrained severely by our knowledge of the data pieces, their accuracy and availability. This component of the project fills in a pressing need for developing a comprehensive, data-driven modeling framework that furthers our understanding of the factors that measure and affect criminal activity.

This work was undertaken with help from police departments in Camden, NJ and Philadelphia, PA. The work studied and modeled criminal activity at both macro and micro levels. The micro-level component of this project looked at enforcement issues at the street level based on a probabilistic framework. On the other hand, the macro-level module looked at strategic policing issues confronting city police departments using a case-based reasoning, artificial intelligence framework. Output of this project can therefore be categorized as follows:

1. **Sequential Illicit-Drug Enforcement Module –** In this part of the project, a sequential decision-making model is developed for assisting enforcement officials in allocating resources during a crackdown operation on illicit drug markets. The Sequential Crackdown Model considers a probabilistic framework, where the probability of incarceration of a dealer and the probability of dealing are modeled as a function of the size of a drug market, crackdown enforcement level, drug dealer's financial hardship, and other market characteristics. The displacement of drug dealers to "other" drug markets is modeled through the fact that the dealer is "not" dealing in that particular market. Therefore the displacement effect is modeled via the probability of "not dealing" in this model.

   The model was developed and tested in consultation with enforcement officials from Philadelphia, PA and Camden, NJ. An implementation scheme is developed for updating parameters on each day of the crackdown operation. Guidelines are provided for enforcement officials to improve the chances of success during a crackdown operation.

   Results show that using maximum enforcement for a significant number of days during a crackdown may be optimal in neighborhoods with a severe drug problem. A cyclic crackdown-backoff strategy may be optimal where residual deterrence dominates financial hardship. Nonetheless, for all markets, a much quicker and less costly collapse could be implemented if the daily enforcement availability is increased. The model also provides rules of thumb for identifying markets where crackdowns would be unsuccessful in eliminating a drug market.

2. **Data-Driven Strategic Policing Decision-Support System –** This component was aimed at providing a practical, data-driven decision support tool for police departments. This work was undertaken in partnership with Philadelphia and Camden, NJ police departments. Input factors were identified through a series of interviews and subsequently aggregated into three dimensions of enforcement (data mainly from Law Enforcement Management and Administrative Statistics - LEMAS), crime (Uniform Crime Reports –UCR, and FBI sources) and environment (U.S. Census). The data was extracted, coalesced and then normalized resulting in a master file to be used as input to the model. The software and model was

upgraded to incorporate the police department goals of benchmarking more efficient/effective similar police departments. The model was then tested both with help from members of the partner police departments and via a controlled experiment. The results showed the effectiveness of this software in: (1) developing a comprehensive database that incorporates environmental, enforcement factors along with crime statistics, (2) understanding and measuring criminal activity based on a comparative, data-driven modeling framework, (3) encouraging meaningful communication among similar police departments. Besides the practical usefulness of this project in enhancing the capability of police department in making more informed decisions, the project makes two theoretical contributions. First, it provides a unique modeling framework classifying and aggregating input data into three dimensions – environment, enforcement and crime. Relevant factors are identified and new measures (e.g. racial match index) developed that help define these dimensions. Second, this project identifies four important strategic model goals that assist police departments in moving toward a direction of proactive management.

4.2 Sequential Illicit-Drug Enforcement Module

Illicit drugs, and their control, continue to impose a significant cost on our society both in the short and long term. This trend can be inferred through DEA statistics that show their budget increasing from $74.9 million in 1973 to $1,550 million in 2000. While there is no consensus on the choice of strategies to control the problem, i.e., Local vs. Border Control, and Enforcement vs. Treatment, and Education, a significant amount of money continues to be spent. The complexity of issues makes finding the "right" strategy mix difficult, and sometimes even counter-intuitive. Quantitative and data analysis can often help grapple with some of these issues systematically. Some policy makers argue that a focus on a national strategy is misplaced because it ignores the local nature of the drug problem and suggest putting a greater emphasis on local enforcement. On the other hand, predatory crime theory warns that local enforcement may lead to unnecessary levels of property crime unless local enforcement is limited. Recent models however argue that, in most cases, drug enforcement is almost always a worthwhile strategy in an overall sense after incorporating cost of drug-related crimes. In-depth street-level ethnographic studies can be a useful tool in identifying patterns and strategies effective in localized environment with possible generalizations beyond the region of study. More efforts are under way to bring newer methods of accurately evaluating policing enforcement. These trends in better methods of measurement/evaluation coupled with support from theoretical models make a strong case for local enforcement as a viable option to control illicit drug sales on the street.

Even strong opponents of local enforcement tend to agree on the necessity of street enforcement for controlling the illicit drug problem. In this module a restrictive definition of local enforcement is used, where we focus on retail-level dealing on street corners alone rather than focusing on surveillance and tracking systems. One local-enforcement strategy that continues to be used by police departments across the country is that of crackdowns, i.e., concentration of resources in a geographic area for a limited time or for targeting specific types of crime. There have been studies done on evaluating the impact of crackdowns and some on modeling the effect of a crackdown on street illicit-drug markets.

For the most part, past modeling literature on drug crackdowns has looked at drug dealing and arrests as a deterministic activity, while in reality both events are probabilistic in nature. During a crackdown operation, drug dealers make decisions of whether to deal or not based on their past experience and the financial hardship they face. Similarly, the chance of getting arrested depends on the level of enforcement and the number of dealers *actually* dealing in the market.

While dynamic control models are not new to illicit drugs or to law enforcement, this work looks at discrete time interval which better represents the realistic time window of a "day", instead of the continuous-time framework considered in past literature thus far.

Another factor that needs to be addressed, to make a model that better represents reality, is the sequential and adaptive response of a drug dealer to police enforcement. A drug dealer makes decisions on a daily basis that adapts in response to their *expectation of risk* of being caught. It is therefore appropriate that a more realistic model should explicitly consider sequential decision-making during a crackdown capturing the "learning effect" of a drug dealer.

Therefore, a probabilistic framework where sequential decision making is considered.

Motivation for the Model

The model developed here was motivated by a dialog with enforcement officials in Camden, NJ and Philadelphia, PA, culminating a list of unanswered questions. Both cities are in close geographical proximity and despite their difference in size, share similar severity of the illicit drug problem. Based on a series of discussions with experienced enforcement officials, it was clear that the financial hardship of a neighborhood has a very significant impact on a drug dealer's response to enforcement and needed to be explicitly incorporated within the modeling framework. While there are several forms of drug crackdowns, this project restricted attention to the one targeting "open-air" street drug dealers. Based on these considerations, the following questions were raised:

Q.1. Under what circumstance is it worthwhile to undertake a crackdown operation on an illicit drug market?
Q.2. Will maximum enforcement on each day of a crackdown result in maximum reduction in violence per unit enforcement used?
Q.3. Is it optimal to use more enforcement for larger markets and less for smaller markets?
Q.4. If there are two markets of similar size, A and B, where market A is more conductive to arrests, is it optimal to spend more resources in market A than market B?
Q.5. What impact does financial hardship of a neighborhood have on enforcement strategies during a crackdown?

A model has been developed that is explicitly designed to answer some of these questions, providing some guidance to practitioners on how best to allocate resources during a drug-crackdown operation.

The Sequential Crackdown Model

The Sequential Crackdown Model (SCM) restricts attention to crackdown on street drug markets targeting “street-corner” drug dealers. To develop the model, in-depth interviews with enforcement officials in Camden and Philadelphia were conducted. The interviews were organized in a semi-structured format to cover all the issues but allowed flexibility for the enforcement officials to provide additional inputs and factors relevant to the model, not considered beforehand.

Multiple officials were involved in the interviews to reduce misperceptions, biases, and loss of information. This enhances the creative potential of the study; their convergent insights boost confidence in findings while their divergent insights increase the chance of surprising findings and delay premature closure

To capture the insights of the enforcement officials, questions were often posed as comparisons. As the model was being mapped out, the officials were consulted at every stage checking for validity. This validation process often involved seeking feedback from officials on questions posed as “what-if” scenarios.

Based on these interviews, the following model assumptions were made:

1. The process of decision making for each individual dealer is the same, i.e., we do not distinguish drug dealers from each other.
2. On any given day, the drug dealer deals in the market with a probability.
3. The probability of dealing on any day depends on the dealer’s expected enforcement on that day, and a financial hardship factor that quantifies the financial desperation or need for the drug dealer to deal that day.
4. We assume that enough resources are allocated imposing a significant incarceration sentence to an arrested dealer. In other words, incarceration of drug dealers during a crackdown does impose a threat to their lifestyle.
5. The police officials make sequential decisions and decide on a daily basis the enforcement allocated for that day.
6. Based on the enforcement allocated on any given day, and the number of dealers actually dealing on that day, there is an associated probability of incarceration for a drug dealer.
7. Displacement of drug dealers is accounted for through the probability of not dealing. Therefore the probability of not dealing includes the possibility of dealers moving to other markets on a temporary basis.
8.

Notation used:

E_i : Crackdown enforcement imposed on a drug market on day *i* of the crackdown. E_0 is the regular base-level enforcement in the market before and during the crackdown. E_{max} is the maximum permissible enforcement on a given day,

$E_c(E_i)$: Expected Enforcement on day *i* by a dealer,

D_i : Financial hardship of a drug dealer on day *i*,

$E(M_i)$: Expected number of drug-dealing days missed by a dealer until day *i*. The phrase ‘until day *i*’ will be used throughout to imply days 1 through *i*-1 but not including day *i*,

τ_i : Probability of a drug dealer dealing on day *i*,

P_i : Probability of incarceration of a drug dealer on day *i*, and

N_i : Number of dealers available to deal on day *i*.

The Sequential Crackdown Model (SCM) considers the probability of a drug dealer dealing on day *i*, τ_i, as a function of the financial hardship D_i and the expected enforcement $E_c(E_i)$. The financial hardship factor itself depends on the economic conditions of the neighborhood and the expected number of days missed until day *i*. Further, the dealer's expected enforcement depends on the history of crackdown enforcement until day *i*.

Based on the probability of dealing and the pool of dealers available to deal, the SCM estimates the expected number of dealers dealing on day *i*. This in turn is then used to calculate the optimal enforcement which maximizes the number of dealers incarcerated per unit resource spent. This is visually depicted in Figure 4.1. Each of the individual components of the model are discussed next.

<u>Probability of Incarceration</u>

In this model we consider that an arrest results in an incarceration of a drug dealer for a significant length of time. The SCM models the probability of incarceration of a drug dealer on day *i*, P_i, as a function of the total enforcement per dealer. The expected number of dealers dealing on day *i* is estimated based on the pool of available dealers, N_i, and the probability of dealing τ_i. Mathematically we are looking for the behavior of P_i as a function of e_i (= $\frac{E_0+E_i}{\tau_i N_i}$), where e_i is the enforcement per dealer *actually* dealing in the market. For small values of e_i, we expect P_i to increase rapidly initially. Thereafter, beyond a certain value of e_i, P_i will increase only with diminishing returns. Therefore P_i will approach asymptotically a theoretical maximum value of P_{max}. This maximum value of probability of incarceration, P_{max}, would be low for markets where arrests are difficult, and close to one for markets where drug dealing is easy to spot and dealers have little opportunity for escaping an arrest.

Mathematically we are looking for a function with $\frac{\mathrm{d}\,P_i}{\mathrm{d}\,e_i} \geq 0$, $\left.\frac{\mathrm{d}P_i}{\mathrm{d}e_i}\right|_{e_i \to \infty} = 0$, $\left.\frac{\mathrm{d}^2 P_i}{\mathrm{d}e_i^2}\right|_{low\ e_i} > 0$ and $\left.\frac{\mathrm{d}^2 P_i}{\mathrm{d}e_i^2}\right|_{high\ e_i} < 0$. We use a logistic growth function that satisfies these conditions and is mathematically tractable.

Therefore we can write,

$$P_i = \frac{A}{1+Be^{-\frac{E_0+E_i}{\tau_i N_i}}}, \qquad \text{where A and B are positive constants.} \tag{1}$$

Incorporating end conditions we get,

$$A = P_{max}\,. \tag{2}$$

Before the start of the crackdown when baseline enforcement E_0 alone is used, the associated probability of incarceration is P_0. Using this in (1), we get

$$B = (\phi\text{-}1)\, e^{\frac{E_0}{\tau_i N_i}}, \qquad \text{where } \phi = \frac{P_{max}}{P_0}. \tag{3}$$

Using (2) & (3) in (1), we get

$$P_i = \frac{P_0\phi}{1+(\phi-1)e^{-\frac{E_i}{\tau_i N_i}}} \tag{4}$$

The incarceration probability improvement index, $\phi = \frac{P_{max}}{P_0}$, represents the ratio of probability of incarceration with "infinite" enforcement resources to the probability of incarceration under the pre-crackdown enforcement E_0. Therefore ϕ is a market attribute that quantifies the impact of enforcement on the probability of incarceration.

Financial Hardship Factor

One of the arguments put forward by experts is that the drug dealing is a source of employment for the dealers, albeit an illegal one. A period of no drug dealing imposes a financial hardship that may ultimately increase their chance of dealing on the street due to monetary desperation[1].

Here we postulate that the financial hardship factor on a given day *i*, D_i, depends on the expected number of days of drug dealing missed by an individual until day *i*, $E(M_i)$, and the alternate economic opportunity available for a drug dealer in the market. In interviews with the local police departments, it was evident that poor neighborhoods consistently had a high level of financial hardship among dealers, forcing them to deal. On the other hand, in the "not so poor neighborhoods", the financial hardship factor could be low but would change significantly with an increase in the expected number of drug-dealing days missed. Figure 4.2 visually depicts this mental model presented by the narcotics officials.

One way to model this behavior is,

$$D_i = \frac{\alpha}{1+\beta e^{-E(M_i)}} \quad \text{where } \alpha, \beta \text{ are positive constants.} \tag{5}$$

If $E(M_i) = 0$, financial hardship level is at its minimum, D_{min}. On the other hand, when $E(M_i)$ is very large, the hardship level reaches its theoretical maximum value, D_{max}.

Using this we get $\alpha = D_{max}$ and $\beta = \frac{D_{max}}{D_{min}} - 1$

Rewriting D_i in terms of D_{max} and D_{min}, we get

$$D_i = \frac{D_{max}}{1+(\psi-1)e^{-E(M_i)}}, \quad \text{where } \psi = \frac{D_{max}}{D_{min}}. \tag{6}$$

The hardship increment index, ψ, quantifies the increase in financial desperation of a drug dealer with an increase in the expected number of missed drug-dealing days.

The inflection point of the hardship curve is given by $\ln(\psi-1)$. For a poor neighborhood where the curve is strictly concave, $\ln(\psi-1)$ is strictly less than zero. This implies that $1 < \psi < 2$, i.e., in a relatively poor neighborhood we do not expect an increase of financial hardship by more than a factor of 2. On the other hand, not so poor neighborhoods will have a $\psi > 2$. This criterion can be used in deciding the value of D_{min} and D_{max} as well as in categorizing the neighborhoods.

Expected number of days missed

Recall that the number of days missed by a drug dealer until day *i* is $E[M_i]$

Let

---

[1] In discussions with the Philadelphia, PA and Camden, NJ Police Departments, this factor was emphasized by the police officers as the single most important factor that makes drug dealers come back to deal during a crackdown operation.

$$X_j = \begin{cases} 1, & \text{when dealer is not dealing on day } i \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Then } E[M_i] = E[\sum_{j=1}^{i-1} X_j] = \sum_{j=1}^{i-1} E[X_j] = \sum_{j=1}^{i-1}(1-\tau_j)$$

$$\therefore E(M_i) = \sum_{j=1}^{i-1}(1-\tau_j) \quad (7)$$

In words, the expected number of days missed until day *i* is the sum of the probability of not dealing on each of the previous days. This is a convenient expression, readily calculated for day *i*, since the probability of dealing for previous days, τ_j, is available.

Expected Enforcement

The other factor that affects the probability of dealing is the expected enforcement on day *i* by a dealer, $E_c(E_i)$. We postulate that drug dealers estimate the level of enforcement on day *i* based on their past experience of previous (*i*-1) days of the crackdown. Mathematically, the expected enforcement on day *i*, $E_c(E_i)$, is a function of $E_1, E_2, ..., E_{i-1}$. This expected enforcement is then formulated to be a linear weighted function of the past enforcement, i.e.,

$$E_c(E_i) = \sum_{j=1}^{i-1} w_j E_j \text{, where } w_j \text{ are non-negative weights such that } \sum_{j=1}^{i-1} w_j = 1 \quad (8)$$

It is not unreasonable to assume that a dealer assigns more weight to the more recent enforcement which implies that $0 \le w_1 \le w_2 \ldots\ldots \le w_{i-1} \le 1$. One simple way to assign weights that satisfy the above conditions is as follows:

$$w_j = a\rho^{i-j-1},\ j = 1, 2, ..., i-1, \quad \text{where } a > 0, 0 \le \rho < 1. \quad (9)$$

Using the fact that $\sum_{j=1}^{i-1} w_j = 1$, we get,

$$w_j = \frac{(1-\rho)\cdot\rho^{i-j-1}}{1-\rho^{i-1}},\ j = 1, 2, ..., i-1. \quad (10)$$

Using (8) and (10), we get,

$$E_c(E_i) = \sum_{j=1}^{i-1} \frac{(1-\rho)\rho^{i-j-1}}{(1-\rho^{i-1})}\cdot E_j\ . \quad (11)$$

ρ can be considered as dealers' memory parameter. If ρ is close to 1, the dealers evenly weigh recent and past enforcement in their mind. On the other hand, $\rho = 0$ implies dealers having short-term memory, basing their judgement primarily on the most recent enforcement level. This method can be very useful for the modeler in quantifying the value of ρ.

Probability of dealing

We are now ready to derive an expression for the probability of dealing for a dealer on day *i*, τ_i, which is a function of $E_c(E_i)$, the conditional expected enforcement, and D_i, the financial hardship factor on day *i*. Recall that D_i is between D_{min} and D_{max} and $E_c(E_i)$ is between E_0 and E_{max}.

To understand the behavior of τ_i as a function of $E_c(E_i)$ and D_i, we first look at the end conditions.

<u>Case I</u>:

Consider $E_c(E_i) = E_0$, the case when a dealer expects no additional enforcement due to the crackdown. If $D_i = D_{max,}$ i.e., a dealer is very desperate to deal on that day due to financial hardship, we would expect the probability of dealing on day *i* to be maximum, i.e., $\tau_i = \tau_{max}$. These conditions would be the most conducive for a dealer to deal since the enforcement is expected to be at a minimum and the financial hardship level is at the highest.

Mathematically,

$$E_c(E_i) = E_0,\ D_i = D_{max} \qquad \tau_i = \tau_{max} \tag{12}$$

<u>Case II</u>:

On the other hand, if $E_c(E_i) = E_{max}$, the maximum possible enforcement, and $D_i = D_{min}$, the least possible hardship, we would expect the probability of dealing to be at the lowest, i.e., $\tau_i = \tau_{min}$.

Mathematically,

$$E_c(E_i) = E_{max},\ D_i = D_{min} \qquad \tau_i = \tau_{min} \tag{13}$$

All the other cases will fall between these two extreme cases. We define the pre-crackdown probability of dealing to be τ_0 that is between τ_{min} and τ_{max}. Since little else is known about the relative effect of D_i and $E_c(E_i)$ on the parameter τ_i, for this work we assume a linear weighted relationship between the two extreme cases presented above.

Mathematically,

$$\tau_i = \tau_0 + (\tau_{max}-\tau_0)\left|\frac{D_i - D_{min}}{D_{max} - D_{min}}\right| - (\tau_0 - \tau_{min})\left(\frac{E_c(E_i) - E_0}{E_{max} - E_0}\right| \tag{14}$$

One can easily verify that conditions (12) and (13) are satisfied in (14).

The functional relationship used above is restrictive in that no interaction terms between D_i and $E_c(E_i)$ have been incorporated. Given the primitive understanding of drug dealing and illicit drug markets, this linear approximation may still capture the dynamics without making it too complicated mathematically. Later we develop a more generalized function for τ_i that *does* consider interaction terms between D_i and $E_c(E_i)$.

<u>The Objective Function</u>

Typically crackdowns have multiple objectives like minimizing the violence associated with drug dealing, improving the quality of life in the neighborhood, increasing the perception of safety for residents, etc. Since these objectives are difficult to quantify, we use the expected number of arrests as a surrogate measure of reduction in “negativities” associated with drug dealing. Therefore, we consider an objective function that maximizes the expected number of dealers incarcerated on a given day per unit enforcement spent on that day. This is the typical output/input objective used in many applications. Using such an objective function however is essentially ignoring the cost of incarceration. Some public policy experts even argue that minimizing number of arrests subject to a constraint/goal of crime reduction would be a more effective formulation of the problem. Despite this limitation, it is important to note that the objective function used by SCM (that emphasizes operational efficiency over effectiveness) continues to be used across the country by enforcement officials to measure the success of a crackdown. While this *does not justify* the choice of our objective function, it does however offer some validity for its choice.

In words, the objective is therefore to find that enforcement level (above the baseline level, E_0) on day *i* of a crackdown operation, that maximizes the expected number of dealers incarcerated per unit enforcement resources spent.

Mathematically,

$$\underset{E_i}{\text{Maximize}}\ Z_i = \frac{\text{Expected number of dealer incarcerated on day } i}{\text{Enforcement spent on day } i}$$

The expected number of dealers incarcerated on day *i* is the product of expected number of dealers actually dealing on day *i* and the probability of incarceration on day *i*. Further, the expected number of dealers dealing on day *i* is the product of the number of dealers in the market, N_i and the probability of dealing, τ_i. The enforcement used will be the sum of the baseline enforcement level E_0 and the crackdown enforcement E_i. Since the expected number of dealers dealing is constant on a given day in a sequential model, effectively we are maximizing the probability of incarceration per unit enforcement used.

Mathematically,

$$\underset{E_i}{\text{Max}}\ \ Z_i = \frac{N_i\,\tau_i\,P_i}{E_0 + E_i} = \frac{N_i \tau_i P_0 \phi}{[1+(\phi-1)e^{-\frac{E_i}{\tau_i N_i}}](E_0+E_i)}$$

Since the decision is being made on a daily basis; for simplicity, we drop the subscript *i* from all the variables.

$$\underset{E}{\text{Max}}\ Z = \frac{N\,\tau P_0\,\phi}{[1+(\phi-1)e^{-\frac{E}{\tau N}}](E_0+E)} \qquad (15)$$

It is important to note that a day-to-day model captures the prevalent myopic outlook (vs. a more longtime one) of both the dealers and the police. Obviously this short-term outlook is limited and may even be at odds with the effectiveness objective. A summary of the model developed so far:

- Probability of incarceration on day *i*, $P_i = \dfrac{P_0\phi}{1+(\phi-1)e^{-\frac{E_i}{\tau_i N_i}}}$ where $\phi = \frac{P_{\max}}{P_0}$,
- Expected enforcement on day *i*, $E_c(E_i) = \sum_{j=1}^{i-1} \dfrac{(1-\rho)\rho^{i-j-1}}{(1-\rho^{i-1})} \cdot E_j$,
- Expected number of days missed until day *i*, $E(M_i) = \sum_{j}^{i-1} (1-\tau_j)$,
- Financial hardship factor on day *i*, $D_i = \dfrac{D_{\max}}{1+(\phi-1)e^{-E(M_i)}}$, where $\psi = \frac{D_{\max}}{D_{\min}}$,
- Probability of dealing of a dealer on day *i*,

$\tau_i = \tau_0 + (\tau_{max} - \tau_0) \left| \frac{D_i - D_{min}}{D_{max} - D_{min}} \right| - (\tau_0 - \tau_{min}) \left(\frac{E_c(E_i) - E_0}{E_{max} - E_0} \right)$, and

- Objective function, $\underset{E_i}{\text{Max}} \; Z_i = \dfrac{N_i \tau_i P_0 \phi}{[1 + (\phi - 1)e^{-\frac{E_i}{\tau_i N_i}}](E_0 + E_i)}$.

<u>Properties of the Sequential Crackdown Model</u>

The objective function is clearly non-linear in the decision variable, E, and cannot be solved explicitly. Even though numerical solutions can be readily obtained, we first analytically explore the behavior of the objective function vis-à-vis different parameters.

Taking the derivative of Z with respect to E, we get

$$\frac{\mathrm{d}Z}{\mathrm{d}E} = \frac{-N\tau P_0 e^{-\frac{E}{\tau N}}[e^{\frac{E}{\tau N}} - \{(\phi-1)\frac{E}{\tau N} + (\frac{E_0}{\tau N} - 1)(\phi - 1)\}]}{[1 + (\phi-1)e^{-\frac{E}{\tau N}}]^2 [E_0 + E]^2} \; . \qquad (16)$$

This expression is nonlinear in E and explicit solutions of $\frac{\mathrm{d}Z}{\mathrm{d}E} = 0$ (which we refer as stationary points) are not possible. However,

$$Sign(\tfrac{\mathrm{d}Z}{\mathrm{d}E}) = -Sign[e^{\frac{E}{\tau N}} - \{(\phi-1)\tfrac{E}{\tau N} + (\tfrac{E_0}{\tau N} - 1)(\phi - 1)\}] \, . \qquad (17)$$

Next we establish the necessary and sufficient conditions for a stationary point to exist via the following Lemmas.

<u>Lemma 1:</u>

The objective function has a unique global maximum if $(\frac{E_0}{\tau N} - 1)(\phi - 1) > 1$.

<u>Proof</u>:

In (17) we represent the right hand side as a difference of two functions:

$f_1(E) = e^{\frac{E}{\tau N}}$, and $f_2(E) = (\phi-1)\frac{E}{\tau N} + (\frac{E_0}{\tau N} - 1)(\phi - 1)$.

The function $f_1(E)$, is non-linear in E with slope, $m_1 = \frac{1}{\tau N} e^{\frac{E}{\tau N}}$, and intercept $c_1 = 1$. The function $f_2(E)$ is linear in E having a slope of $m_2 = \frac{(\phi - 1)}{\tau N}$ and intercept $c_2 = (\frac{E_0}{\tau N} - 1)(\phi - 1)$. The graphs of f_1 and f_2 are shown in Figure 4.3(a).

From Figure 4.3(a), it is easily verified that since f_1 is convex and f_2 is linear, f_1 and f_2 will meet uniquely at one point if intercept $c_2 > c_1$. This implies that if $(\frac{E_0}{\tau N} - 1)(\phi - 1) > 1$, there exists a unique point E* where $\frac{\mathrm{d}Z}{\mathrm{d}E} = 0$. From (17) we can also see that sign of $\frac{\mathrm{d}Z}{\mathrm{d}E}$ is the sign of $(f_2 - f_1)$. Therefore from Figure 4.3(a) it is clear that $\frac{\mathrm{d}Z}{\mathrm{d}E} > 0$ for $E < E^*$ and $\frac{\mathrm{d}Z}{\mathrm{d}E} < 0$ for $E > E^*$ implying E* is a global maximum.

<u>Lemma 2:</u>

There exists at least one non-negative stationary point for the objective function if the following necessary condition holds:

$$\begin{cases} (\frac{E_0}{\tau N} - 1)(\phi - 1) > 1 & \text{for } 1 < \phi < 2 \\ (\frac{E_0}{\tau N} - 1) + \ln(\phi - 1) \geq 1 & \text{for } \phi \geq 2 \end{cases}$$

Proof:
Consider a point E′ where the slopes m_1 and m_2 are equal (Figure 4.3(b)). If E′ ≥ 0 then at least one positive solution to $\frac{dZ}{dE}$ = 0 will exist if the value of f_2 at E′ is greater than or equal to the value of f_1 at E′. However, if E′ < 0, at least one non-negative solution to $\frac{dZ}{dE}$ = 0 will exist iff $f_2(0) \geq f_1(0)$.
Mathematically $m_1\big|_{E'} = m_2\big|_{E'}$ yields,

$$\frac{1}{\tau N} e^{\frac{E'}{\tau N}} = \frac{(\phi - 1)}{\tau N},$$

$$E' = \tau N \ln(\phi - 1). \qquad (18)$$

CASE I: Consider E′ ≥ 0 which implies from (18) that $\phi \geq 2$. As argued above the necessary condition for at least one positive solution to $\frac{dZ}{dE}$ = 0 is $f_2(E') \geq f_1(E')$. This yields,

$$\ln(\phi-1) + \left(\frac{E_0}{\tau N} - 1\right) \geq 1 \quad \text{for } \phi \geq 2. \qquad (19)$$

CASE II: In the case E′ < 0 or alternatively $1 < \phi < 2$, the necessary condition for at least one positive solution to $\frac{dZ}{dE}$ = 0 is $f_2(0) \geq f_1(0)$. This yields,

$$1 < \left(\frac{E_0}{\tau N} - 1\right)(\phi - 1) \qquad \text{for } 1 < \phi < 2. \qquad (20)$$

Combining the two cases given in [19] and [20] we get the Lemma.

Lemma 3:
The objective function, $\underset{E}{\text{Max}}\ Z$, can have at most two stationary points.

Proof:
Recall from Lemma 2 that E′ is the point where the slopes m_1 and m_2 are equal. It is easy to show that the difference $(f_1 - f_2)$ is strictly decreasing for E < E′ and strictly increasing for E > E′. If the difference $(f_1 - f_2)$ at E′ is positive then there exists no stationary point. On the other hand, if this difference is negative at E′ *and* the difference at E = 0 is positive then we have two positive stationary point solutions. The two solutions are on either side of E′, i.e., $E^{s_1} < E' < E^{s_2}$ as shown in Figure 4.3(b). It is easy to verify that E^{s_2} will be the point of local maximum and of interest to us.

Finally, if $(f_1 - f_2)$ is negative at E = E′ *and* E = 0 then only one point exists with $\frac{dZ}{dE}$ = 0. These are the only possibilities and the Lemma is established.

Theorem 1:
The equation $\frac{dZ}{dE}$ = 0 has either zero, one or two positive solutions. The conditions for each possibility are given graphically in Figure 4.4.

Proof:
The condition for a unique solution from Lemma 1generates the region I in Figure 4.4. The condition from Lemma 2 yields the lower region in Figure 4.4. Notice that the two curves meet when

$(\frac{E_0}{\tau N}$-1)(ϕ-1) = ($\frac{E_0}{\tau N}$-1) + ln(ϕ-1)

($\frac{E_0}{\tau N}$-1) = 1 and (ϕ-1) = 1

The second part of the condition in Lemma 2 for 1 < ϕ < 2 is the same as the condition for unique solution in Lemma 1. i.e., the upper and lower curves merge for 1 < ϕ < 2 and ($\frac{E_0}{\tau N}$-1) > 1.

Region II in Figure 4.4 bounded between lower and upper curves has at least one solution but since it is below the upper curve it does not have a unique solution. From Lemma 3 it is clear that the bounded region II will have exactly two solutions. Region III does not satisfy the necessary condition for at least one solution. That is, $\frac{dZ}{dE}$ is never equal to zero in that region.

Theorem 1 will be useful later in applying the SCM to examples from the city of Philadelphia.

Having established the conditions for solutions to $\frac{dZ}{dE}$ = 0, next we explore the properties of the optimal solution.

Effect of Size of the Market on the Optimal Enforcement

The following Lemma and Theorems look at the effect of the size of the market, Nτ, on the optimal enforcement value, E*.

Lemma 4:

If an optimal solution E* > 0 exists then it must satisfy the condition $(\phi-1)e^{-E^*/\tau N} < 1$.

Proof:

In Lemma 3, E* = E^{s_2} > E′ is a point of local maximum. This yields:

E′ < E*.

Using (20) we get,

τNln(ϕ -1) < E*,

ln(ϕ -1) < $\frac{E^*}{\tau N}$,

$$(\phi -1)e^{-\frac{E^*}{\tau N}} < 1. \qquad (21)$$

Theorem 2:

If an optimal enforcement E* > 0 exists, then the optimal enforcement per dealer is a decreasing function of the number of dealers.

Proof:

From (16) it is clear that E* satisfies the condition (where w is defined as $\frac{1}{\tau N}$ for simplicity):

$$e^{E^*w} = (\phi-1)[E^*w + E_0 w - 1].$$

Taking the first derivative with respect to w we get,

$$e^{E^*w}[\frac{d(E^*w)}{dw}] = (\phi-1)[\frac{d(E^*w)}{dw} + E_0],$$

$$\frac{\mathrm{d}(E^*w)}{\mathrm{d}w} = \frac{E_0}{[\frac{e^{E^*w}}{(\phi-1)} - 1]}.$$

From Lemma 4, the denominator is positive, which shows that $\frac{\mathrm{d}(E^*w)}{\mathrm{d}w} > 0$. This implies that as *w* increases (or size of the market shrinks), the optimal enforcement per dealer dealing increases.

This result suggests that as the market shrinks in size (during a crackdown), the optimal enforcement per dealer continues to increase further increasing the probability of incarceration. The optimal strategy exhibits a positive feedback characteristic.

Theorem 3:

The optimal objective function, Z*, is an increasing function of the number of dealers.

Proof:

From (15) and (16) we get,

$$Z^* = \frac{P_0\phi}{(1+(\phi-1)e^{-E^*w})(1+\frac{e^{E^*w}}{\phi-1})}, \qquad \text{where } w = \frac{1}{\tau N}.$$

Let $\lambda(w) = \frac{e^{E^*w}}{\phi-1}$. From Lemma 4, $\lambda(w) > 1$. Denominator can be written as,

$$\left(1+\frac{1}{\lambda(w)}\right)(1+\lambda(w)) = 4 + \left(\sqrt{\lambda(w)} - \frac{1}{\sqrt{\lambda(w)}}\right)^2.$$

It is easy to see that the denominator attains a minimum value at $\lambda(w) = 1$, and is increasing in $\lambda(w)$. Furthermore from Theorem 2, $\lambda(w)$ is an increasing function in *w* implying that denominator is an increasing function of *w*. Therefore Z* is a decreasing function of *w or* an increasing function of the number of dealers dealing in the market.

Conditions for a Global Maximum

In Theorem 1 (Figure 4.4) it was established that for region III E* = 0 is optimal, suggesting not undertaking a crackdown operation for markets lying in that region. On the other hand, region I has a unique global maximum, while region II has two stationary points, E^{s_1} and E^{s_2}, which are local minimum and maximum respectively. The point E^{s_2}, need not be a global maximum in which case E* = 0 may still be optimal. Therefore necessary and sufficient conditions for a global maximum still need to be established. The following Lemmas present these conditions.

Lemma 5:

A *sufficient* condition for E* to a global maximum is,

$$\frac{E^*}{E_0} < \frac{\phi}{2} - 1.$$

Proof:

Since $E^{s_1} < E^{s_2}$ (where E^{s_1} is a local minimum and E^{s_2} is a local maximum) it is not difficult to see that E* = E^{s_2} will be a global maximum if Z(E*) > Z(0).

From (15) we get

$$\frac{N\tau P_0\phi}{[1+(\phi-1)e^{-E^*/\tau N}][E_0+E^*]} > \frac{N\tau P_0}{E_0},$$
$$\frac{\phi}{[1+(\phi-1)e^{-E^*/\tau N}][1+\frac{E^*}{E_0}]} > 1.$$

From Lemma 4, $(\phi-1)e^{-E^*/\tau N} < 1$. Using this bound, we get,

$$\frac{\phi}{[1+1][1+\frac{E^*}{E_0}]} > 1,$$

or $$\frac{E^*}{E_0} < \frac{\phi}{2} - 1.$$

We will see later, this condition is almost always satisfied since ϕ is fairly high for most realistic situations.

<u>Lemma 6:</u>

Necessary condition for E* to be a global maximum is,

$$\frac{E_0\phi}{N\tau} > 4.$$

<u>Proof:</u>

From Lemma 5,

$$\frac{N\tau P_0\phi}{[1+(\phi-1)e^{-E^*/\tau N}][E_0+E^*]} > \frac{N\tau P_0}{E_0}.$$

Conditions for E* in (16) imply that,

$$1+\frac{e^{-\frac{E^*}{\tau N}}}{(\phi\text{-}1)} = \frac{E^*+E_o}{\tau \mathrm{N}}. \qquad (22)$$

Using (22) we get,

$$\frac{P_0\phi}{[1+(\phi-1)e^{-E^*/\tau N}][1+(\phi-1)^{-1}e^{-E^*/\tau N}]} > \frac{N\tau P_0}{E_0},$$

or $$\frac{E_0\phi}{N\tau} > 2+(\phi-1)e^{-E^*/\tau N} + \frac{e^{E^*/\tau N}}{\phi-1},$$

$$\frac{E_0\phi}{N\tau} > 2+2+\underbrace{\left[\sqrt{(\phi-1)e^{-E^*/\tau N}} - \sqrt{\frac{e^{E^*/\tau N}}{\phi-1}}\right]^2}_{\text{Always positive}}.$$

Therefore $\frac{E_0\phi}{N\tau} > 4$.

In words, if $\frac{E_0\phi}{N\tau}$ is less than 4, E* = 0 will be optimal, a condition where the model suggests not undertaking a crackdown operation.

<u>Effect of the Incarceration Probability Improvement Index on the Optimal Enforcement</u>

Here we discuss the behavior of E* as a function of the parameter ϕ. Since there is no explicit expression for E*, we establish the result via the following Theorem.

<u>Theorem 4:</u>

If there exists an optimal solution E* > 0, then E* must be an increasing function of the incarceration probability improvement index, ϕ.

Proof:

From (16) we know that E* satisfies the condition:

$$e^{\frac{E^*}{\tau N}} = (\phi - 1)[\tfrac{E^*}{\tau N} + \tfrac{E_0}{\tau N} - 1]$$

Taking derivative with respect to ϕ and rearranging terms we get,

$$\frac{\mathrm{d}E^*}{\mathrm{d}\phi} = \frac{e^{\frac{E^*}{\tau N}} / (\phi - 1)}{\frac{(\phi-1)}{\tau N}[\frac{e^{E^*/\tau N}}{(\phi-1)} - 1]} \tag{23}$$

Using Lemma 4, it is clear that $\frac{\mathrm{d}E^*}{\mathrm{d}\phi} > 0$ which proves the Theorem.

Intuitively this means that it is optimal to spend more enforcement resources in markets where there is greater possibility of improvement in the probability of incarceration.

Application of the Sequential Crackdown Model

The model developed here was motivated by law enforcement operations on illicit drug markets in Philadelphia, PA and Camden, NJ. Clearly, the advantage of the sequential model developed is in its ability to capture the day-to-day changes in dealer's perception of enforcement and financial hardship. In practice, the optimal enforcement level on day *i* can be calculated based on the number of prior arrests *actually* made. However, to forecast the behavior of a drug market according to the strategies suggested by the SCM, the model needs to update the factors sequentially based on an estimate of arrests. Presentation of such projected scenarios to decision makers is important because (i) it helps in understanding and validating the interactive dynamics of different factors in the decision making process, and (ii) it can be used to find the implication of availability of additional resources. In fact the second reason is considered very critical by enforcement officials in their ability to plan and seek additional resources from local, state and federal funding agencies.

The first step in using the model is the estimation of parameters. The minimum, maximum and pre-crackdown probabilities of dealing need to be estimated. The pre-crackdown level of enforcement, number of dealers, maximum available enforcement on a day, the extremes of the financial hardship factor, and the incarceration improvement index are all inputs to the model.

Based on these inputs, the optimal enforcement for day *i* can be calculated. Using this optimal enforcement level, the probability of incarceration on day *i* is then estimated. The dealer's expected enforcement and the hardship level are subsequently updated for the $(i+1)^{st}$ day which then is used to find the probability of dealing for the $(i+1)^{st}$ day. The expected number of dealers dealing on day (i+1) is finally estimated and used to calculate the optimal enforcement level for day (i+1). These above outlined steps have been summarized via a flowchart in Figure 4.5.

For the purpose of the analysis here, it is assumed that the start of a crackdown operation is unknown to the dealers and comes as a surprise. This would imply that the probability of dealing on day 1 is the same as the pre-crackdown probability of dealing, i.e., $\tau_1 = \tau_0$.

This assumption may not be valid if the dealers are aware of the crackdown beforehand. In the latter case τ_1 can be calculated based on the dealer's expected enforcement level at the start of crackdown.

Note that the procedure can be easily used for different E_{max} levels providing insight into the impact of additional resources. Next, we consider two examples of this model applied to illicit drug markets in the city of Philadelphia.

Examples:

The two illicit street drug markets, A and B, considered in the example here, are similar in size. To estimate the number of dealers, number of "drug dealing slots" were estimated. This idea of drug-dealing slots was easily understood by the police officers and served as a useful mechanism of estimating the number of drug dealers. The use of these drug-dealing slots is in agreement with street-level studies where dealers were found to work specific "spots" on either an hourly or permanent basis. It should also be noted that in terms of drug dealing organizational hierarchies, these slot-level dealers would be on one of the two bottom levels, while the entrepreneurs, whom they work for, retaining a greater share of the profit.

After a series of independent interviews with narcotic police officers, it was estimated that at a given time both the markets had an average of ten slots (the number was higher at certain times of the day). Since drug dealing is typically carried out round the clock, a number of N=30 slots seemed reasonable.

So far we have not been specific in defining the enforcement level E, allowing multiple definitions within the context of the model. However, from discussions with the police officers it was clear that the risk imposed to the dealers was a function of the *hours* of police patrolling on the street. Using this as a criterion for estimating enforcement level we estimated the hours of police patrolling for a 24-hour period in a drug market. It was estimated that prior to the crackdown, 10 police officers working 8-hour shifts covered about 2 districts (or 10 drug markets) in a 24-hour period. This translated to E_0 = 8 hours.

The baseline pre-crackdown probability of incarceration, P_0, was calculated based on average weekly arrests. The narcotics police officers indicated that an average of one arrest was made per week in the two markets before the crackdown. This means $P_0 = 1/(30*7) \cong 0.005$.

The maximum probability of incarceration was clearly high; the officers felt that if the enforcement level were increased sufficiently, the probability of arresting a dealer dealing would be between 80 to 90%. Based on this information we used an incarceration improvement index $\phi = 0.85/0.005 = 170$.

Despite these similarities, the two markets were distinct and different from each other in several ways. Market B was in an extremely poor neighborhood with severe financial hardship and lack of opportunities for legal employment. The probability of dealing was consistently high, and τ_0 of 0.85 was estimated. Furthermore, the values of τ_{min} and τ_{max} were conjectured to be 0.50 and 0.95, respectively. The hardship level was consistently high and ranged between 0.7 and 0.95 which was consistent with a value of $1 < \psi < 2$ stated earlier.

For market A where the dealers were more integrated with the society, the estimates were $\tau_0 = 0.80$, $\tau_{min} = 0.1$, $\tau_{max} = 0.9$, $D_{min} = 0.1$, $D_{max} = 0.9$, clearly with $\psi > 2$.

At this point, it is important to emphasize that the estimates of the probability of dealing and the financial hardship factors were based on comparative questions. Therefore, the relative levels are far more reliable than the actual figures.

Using these estimates scenarios were generated based on the procedure outlined by the flowchart in Figure 4.5. Scenarios 1 and 2 are based on a E_{max} level of 100 and 200, respectively. The results are summarized in Figure 4.6 (a), (b) and Figure 4.7 (a), (b), respectively.

The solutions to the SCM displayed in Figures 4.6 and 4.7 were obtained using IMSL subroutines run on a UNIX mainframe computer. Bounds for the non-linear solver were estimated based on the properties discussed earlier. The results are revealing and several observations can be made.

(a) Towards the beginning of the crackdown, the optimal E* values are much higher than the baseline level of enforcement. For Market A the optimal value on day one was 167.31 ($\cong 21E_0$) and for market B it was 177.67 ($\cong 22E_0$).

(b) From Figure 4.6 (a) and Figure 4.6 (b) it is clear that $E_{max} = 100$ is a binding constraint on the first day of the crackdown imposing serious limitation on the possibility of reducing drug dealing. Only 20% market was collapsed on the first day as against a 60% reduction (Figure 4.7 (a) and Figure 4.7(b)) with $E_{max} = 200$. Clearly additional resources could result in a significant reduction in the size of the market on the first day of the crackdown.

(c) In the case of the more severe constraint with $E_{max} = 100$ the cumulative resources used during the crackdown were actually more than with $E_{max} = 200$. For market A, the total hours of enforcement resources were 327.23 with $E_{max} = 100$ as against 281.77 with $E_{max} = 200$. Similarly for market B the numbers were 331.56 and 282.62, respectively.

(d) Market A (where financial hardship was less severe) exhibited a cyclic optimal enforcement initially. Based on a low probability of dealing, the optimal enforcement was lower on the second day of the crackdown as compared to the first and the third day. This strategy is similar to those suggested by some researchers based on a qualitative analysis where they argued for a crackdown-backoff strategy based on residual deterrence.

(e) Market B, where the financial hardship level was quite high, did not display this cyclic behavior. The optimal enforcement decreased steadily since the probability of dealing continued to be high. The enforcement officials confirmed that in such poor neighborhood the dealers continue to deal despite the high risk of arrest during a crackdown operation.

While the market profile was easily obtained as a function of days of crackdown using the SCM, a serious concern raised by the enforcement officials was the possibility of "substitute" dealers replacing the regular drug dealers..

Generalizations of the SCM

The pool of dealers in a drug market can have a large turnover. In the model developed thus far we considered the situation where the number of "slots" for dealing were the same as the number of dealers wanting to deal. However, in neighborhoods with a severe drug problem there are "substitute" or "replacement" drug dealers filling in any available dealing slots. To incorporate this phenomenon into the model we distinguish two types of dealers — Regular and Substitute. As before, the probability of dealing for regular dealers can be calculated based on the financial hardship level and the dealer's expectation of enforcement. The probability of dealing for the substitute dealers was postulated to depend on the availability of slots instead of the financial hardship, since these dealers don't *expect* to deal by definition unless an opportunity arises.

The flowchart for the implementation of this new model is presented in Figure 4.8. The Philadelphia police officials felt that Market B experienced this situation of substitute dealers. Therefore this new model was tested for Market B with E_{max} = 100 and E_{max} = 200. The results are summarized in Figures 4.9(a) and 4.9(b). It is not surprising that the resources needed to collapse the market were significantly higher due to the existence of a "parallel" or "replacement" market. The constraint of E_{max} has a severe impact on the total resources used during the crackdown. Over the duration of the entire crackdown, the cumulative enforcement resources used with E_{max} = 200 is about half of those needed with E_{max} = 100. Even though the model developed for this situation requires more calculations, indeed, the core of the model remains unchanged because the computations are essentially sequentially performed as before.

The SCM considers sequential decision-making where every stage i is a *day*. In reality this time-window may be too large and the enforcement level during the morning may affect the dealer's strategy in the afternoon. To incorporate a smaller time duration in the model, SCM could be modified to consider each stage i as an eight-hour shift with three times as many iterations as the number of days. This means that the parameters need to be adjusted so that i represents an eight-hour period. Using this altered model, the daily variability can be modeled and observed.

Using other objective functions in SCM is possible and computationally poses little difficulty. Similarly incorporating more realistic functions to model the hardship factor and the probability of dealing is possible if better information becomes available. For example, in the model formulation thus far, the functional relationship of τ_i (equation 14) did not involve any interaction terms between D_i and $E_c(E_i)$. This restrictive assumption can be relaxed by using a function for τ_i such as:

$$\tau_i = a\,(D_i - D_{min})\,(E_{max} - E_c(E_i)) + bD_i - cE_c(E_i) + d$$

Where a, b, c, d are positive constants. We can easily verify that this function satisfies the necessary conditions that τ_i increases with an increase in D_i and decreases with an increase in $E_c(E_i)$. However, the function now needs four end-conditions (recall that the earlier relationship in (14) required only two) to find the value of the four constants a, b, c and d.

Answers to the Questions Raised

One of the goals of this research was to answer questions raised by enforcement officials in the cities of Philadelphia and Camden. Below we discuss the insights gained from the model vis-à-vis the questions raised.

1. *Which markets are conducive to crackdowns?*

This question was answered in the context of the model's objective function via Figure 4.4. Region III in the Figure represents markets where E* = 0 is optimal. Additionally Lemma 6 showed that for the two-solution case E* = 0 is optimal again, if $\frac{E_0\phi}{N\tau} < 4$. Markets with low incarceration improvement index, ϕ, and/or low pre-crackdown enforcement per dealer are likely to be least conducive to a crackdown operation. Intuitively, markets where enforcement has little impact on improving the incarceration probability are not likely to offer the desired benefits. Incarceration probability can often be improved by the use of tactics that barricade possibilities of escape for drug dealers. Therefore, the city officials may have to assist enforcement officials by "boarding-up" empty houses, improving lighting, etc *prior* to the crackdown operation. Similarly the pre-crackdown enforcement per dealer can be improved by reducing the probability of dealing by offering alternate avenues of legal employment in the neighborhood. In effect the model states that markets with low ϕ and/or low $\frac{E_0}{\tau N}$ should first focus on improving these two parameters before a crackdown could yield desired results.

2. *Does maximum enforcement imply maximum reduction in violence per unit money spent?*

The SCM objective function did not model violence directly. The expected number of dealers arrested was used as a surrogate measure for reduction in violence. Based on this assumption, we established that maximization of number of arrests per unit resource *need not* be achieved by using maximum enforcement. In fact, some markets may have a cyclic optimal enforcement strategy.

Nevertheless, using maximum enforcement could still be optimal or for an extended period as determined by the size of the markets and the availability of resources on any day.

3. *Effect of size of the market*

This question was discussed where we established that the optimal enforcement per dealer is an increasing function of the number of dealers. This implies that the probability of arresting improves as the market size shrinks. Therefore, it is easier to collapse smaller markets than larger ones. Additionally, from Theorem 3, it was shown that the objective function increases as the size of market increases. Even though the probability of arresting a dealer is low on the first day due to the size of the market, the number of dealers captured will be still high. This results in a high value of the objective function in a large-sized market. Indeed, the impact of a crackdown will be evident greatest in the initial days of the crackdown.

4. *Effect of Arrest probability*

This was partly addressed in question 1 above. Effect of ϕ on E was discussed earlier suggesting undertaking a crackdown operation with more resources in markets where the probability of incarceration has a high potential for increasing.

5. *Effect of Financial Hardship*

Financial hardship was explicitly incorporated as a parameter in the model impacting dealer's probability of dealing directly. A higher hardship level led to an increased probability of dealing thereby increasing the number of dealers actually dealing. Thus, this question is indirectly addressed in question 3 above.

<u>Conclusions</u>

The model developed here incorporated the sequential (and myopic) decision making of drug dealers and the enforcement officials. A probabilistic framework was developed which reflected the underlying dynamics at work (e.g., the probability of incarceration and the probability of dealing). Parameters were estimated and the model was tested for two drug markets in the city of Philadelphia. Results showed that using maximum enforcement for a significant number of days in the crackdown may be optimal if there is a severe drug problem in the neighborhood. A cyclic enforcement strategy was shown to be optimal for markets where residual deterrence dominates the financial hardship. The model also provides guidelines for identifying markets where crackdowns may not be appropriate as well as addresses some other questions that relate the market characteristics with the success of a crackdown operation. Before closing, let us discuss some of the limitations of the model and its results.

The underlying model is a simplification of reality and therefore the usefulness of its results is closely tied with the validity of the embedded assumptions. Even though a careful process of interviews ensured development of an accurate model, data limitations in an illicit activity such as drug dealing constrain a thorough validation. While the work does make a contribution toward quantifying and modeling the dynamics of illicit drug activity on street corners, model assumptions need to be looked at carefully by practitioners in the context of the application before adopting the results.

The Sequential Crackdown Model also raises several unanswered questions. The model suggests possibilities of coupling crackdowns in markets where residual deterrence can lead to cyclic use (and freeing) of resources. However, a formal analysis is needed for sequencing crackdowns on drug markets across the city. The number of drug markets can be large (e.g., city of Camden is estimated to have 130 drug markets) and significant savings can be accrued by using optimal scheduling strategies, underscoring the need for researchers and practitioners to study this problem.

There exists an opportunity for researchers and practitioners to better quantify some of the model parameters such as the financial hardship level, the incarceration probability improvement index, the probability of incarceration, etc. The estimates provided here were based on interviews and can be further authenticated by collecting information from arrested dealers directly.

Some of the interview techniques used by researchers for the development of the model can also be used to build theories from case studies. This technique, often used to empirically develop hypothesis in the area of strategic management, can also be applied for this application where expertise and experience of individuals exists more readily than raw data.

The model developed here indirectly addressed the issue of dealer displacement to other neighborhoods as a result of the crackdown operation. One way to extend the model would be by incorporating a displacement probability factor based on (i) mobility of the dealers and, (ii) the availability of alternate locations for drug dealing. The model's objective function would then have to be altered appropriately incorporating the cost of the displacement effect.

For the purpose of the model we assumed that arrests implied a significant time of incarceration for the dealers. Given the severity of the constraints on the judicial system and prisons across the country, this assumption may not be valid in several situations and the model needs to be altered accordingly.

Finally, the intent of this work was not to argue for or against the use of crackdowns vis-à-vis other strategies, but only to provide some guidance to practitioners for the most efficient way of cracking down on street markets for illicit drugs. The model and the results are presented with a view to encourage others to look into better ways of managing and controlling this problem that continues to affect millions of citizens across the United States.

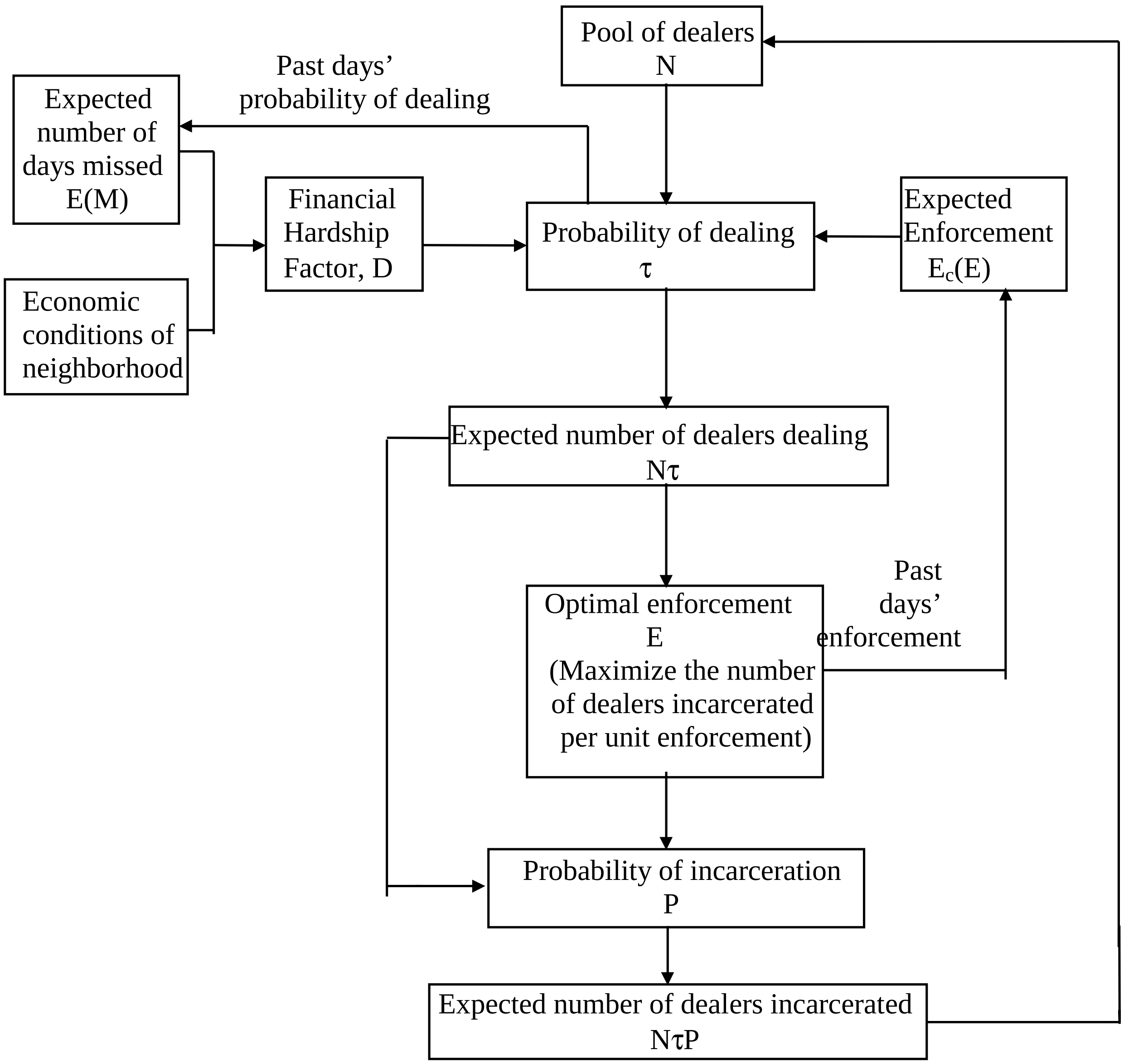


Figure 4.1: The Sequential Crackdown Model (SCM)

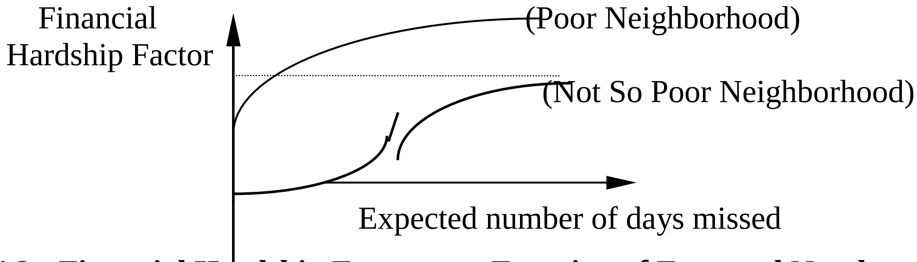


Figure 4.2: Financial Hardship Factor as a Function of Expected Number of Days Missed

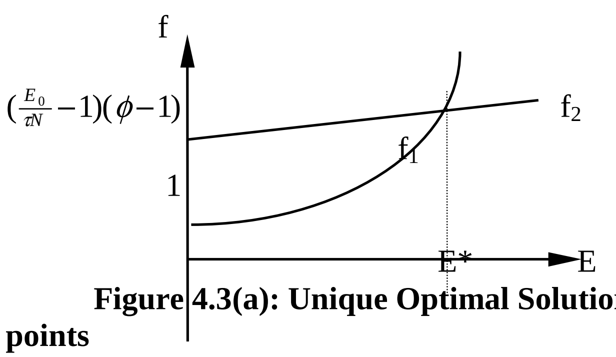


Figure 4.3(a): Unique Optimal Solution

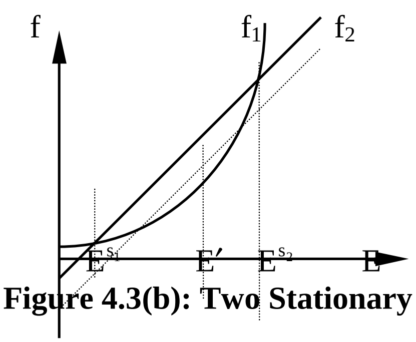


Figure 4.3(b): Two Stationary points

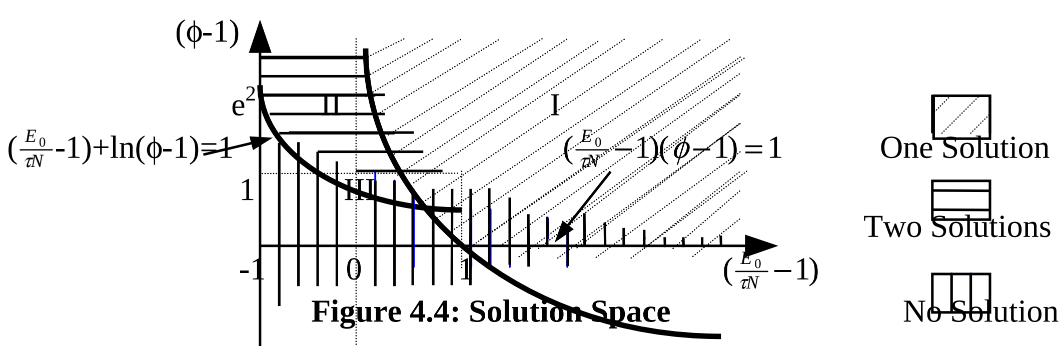


Figure 4.4: Solution Space

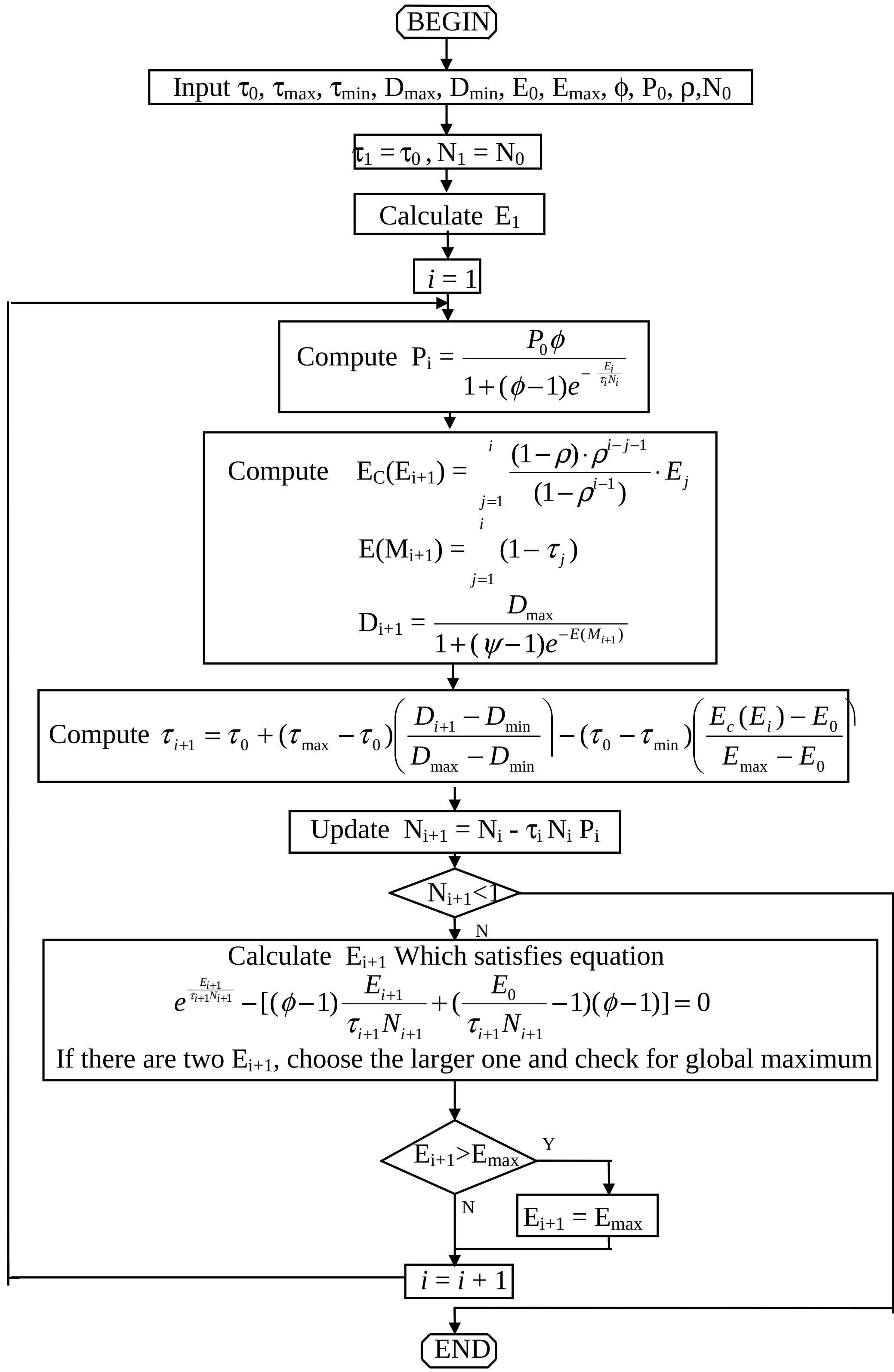


Figure 4.5: Estimating Crackdown Scenarios Using the SCM

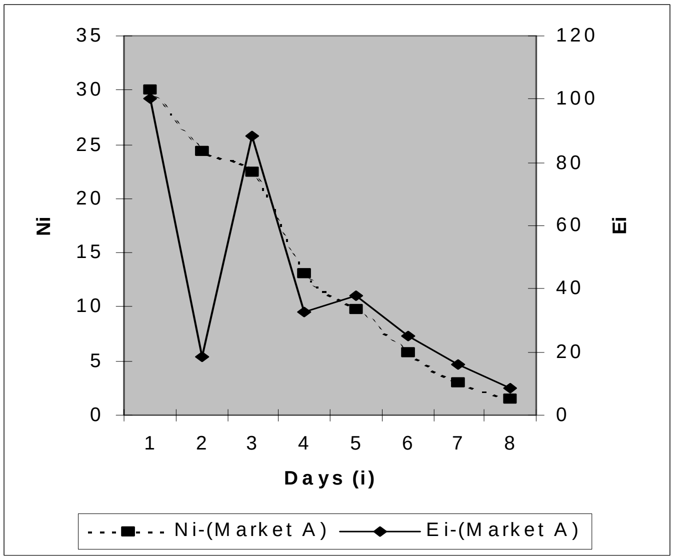


Figure 4.6(a): Market A with E_{max} = 100

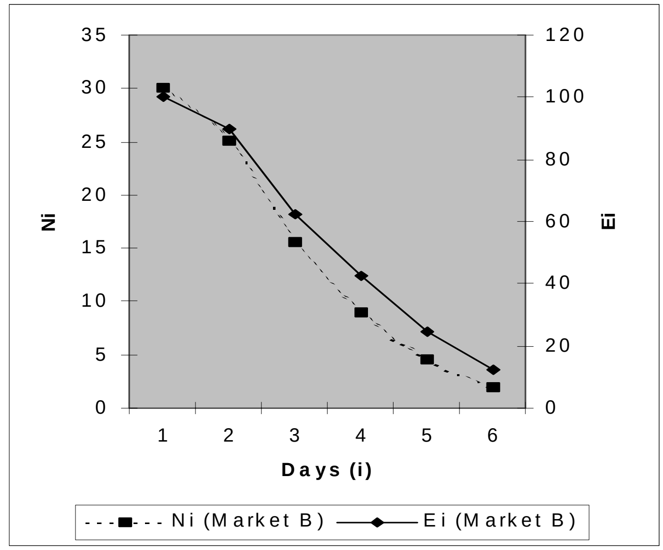


Figure 4.6(b): Market B with E_{max} = 100

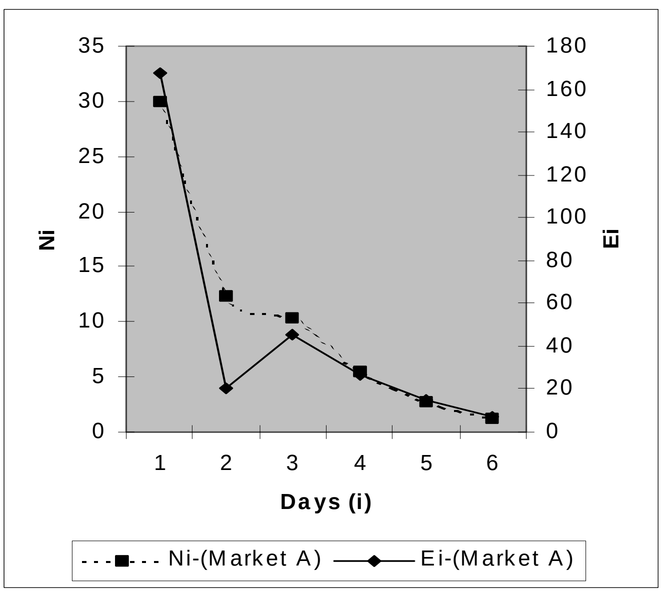


Figure 4.7(a): Market A with E_{max} = 200

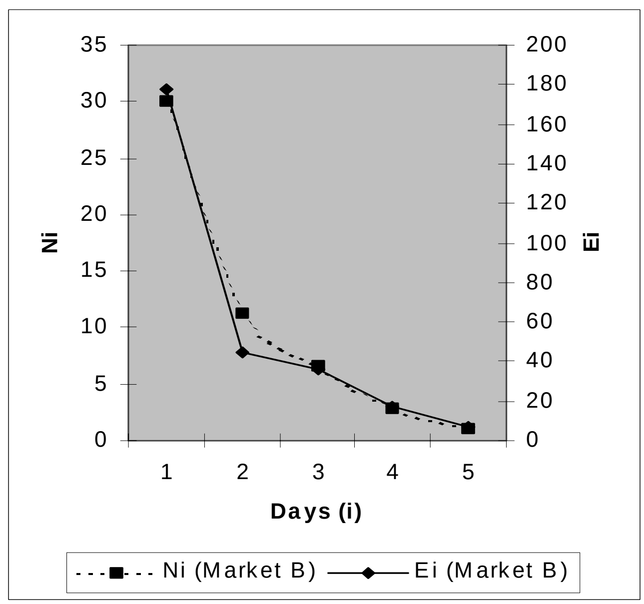


Figure 4.7(b): Market B with E_{max} = 200

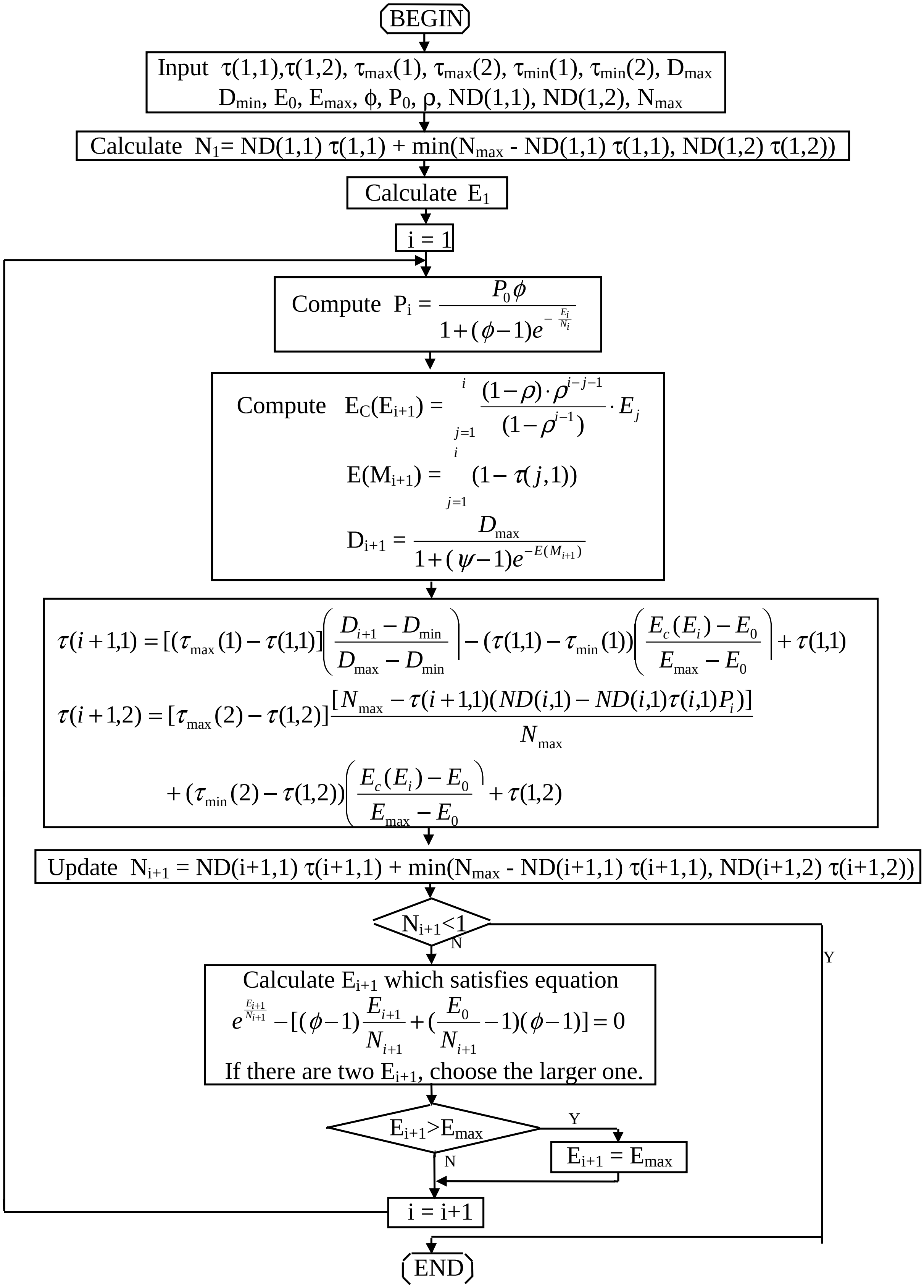


Figure 4.8: Incorporating "Substitute" or "Replacement" Dealers in the SCM

NOTATION:

$\tau(i,1)$: Probability of dealing on day *i* of a crackdown for a regular dealer.
$\tau(i,2)$: Probability of dealing on day *i* of a crackdown for a substitute dealer.
$\tau_{max}(1)$: The maximum probability of dealing for a regular dealer.
$\tau_{max}(2)$: The maximum probability of dealing for a substitute dealer.
$\tau_{min}(1)$: The minimum probability of dealing for a regular dealer.
$\tau_{min}(2)$: The minimum probability of dealing for a substitute dealer.
$ND(i,1)$: Pool of regular dealers on day *i*.
$ND(i,2)$: Pool of substitute dealers on day *i*.
$N(i)$: Expected number of dealers dealing (including regular and substitute) on day *i* during the crackdown.
N_{max}: The total number of "dealing slots" available on any day.

Figure 4.9(a)

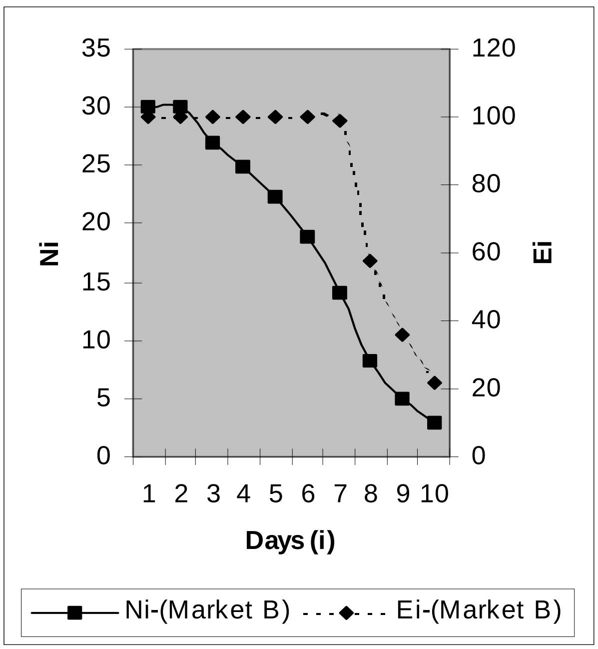


Figure 9(b)

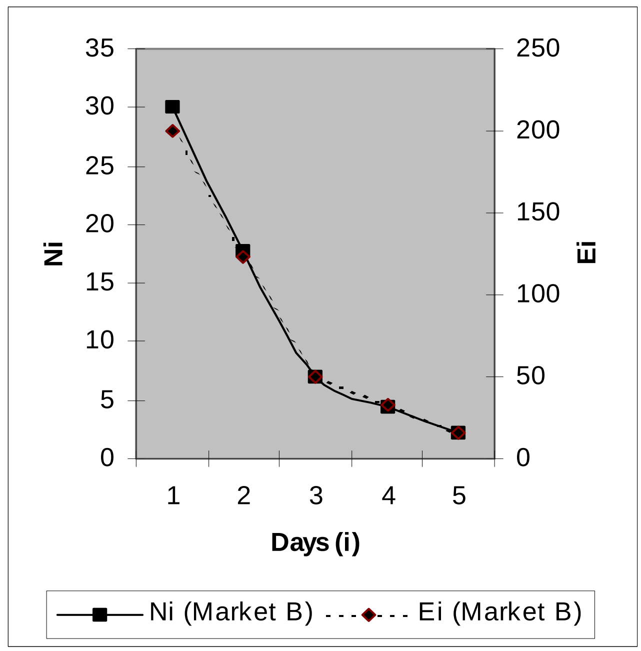


Figure 4.9(a): Market B with "Substitute Dealers with E_{max} = 100
Figure 4.9(b): Market B with "Substitute Dealers" and E_{max} = 200

4.3 Data-Driven Strategic Policing Decision-Support System

One of the greatest technical limitations on the criminal justice system is the lack of useful information to help in the decision making. There has been some progress in recent times with a growing presence of information technology in law enforcement applications. Now police managers regularly retrieve and use information on crime patterns, responses to calls for service, vehicle locations, personnel, finances, and various aspects of departmental performance. Emerging technologies in data collection and usage have opened possibilities for police departments to develop and test new problem-solving techniques. However many of these policing trends, such as rapid-response to 911-calls, are *reactive* instead of being *proactive* in problem solving.

Recent emphasis on community policing and problem-oriented policing shows a definite shift in policing philosophy toward better resource management. Scarcity of resources and increased public attention to crime reduction has put additional pressure on police departments to improve their performance. Police departments need to be able to sort through the clutter of approaches and controversies to find tactics and initiatives that really work. Therefore, there is a growing effort among law enforcement departments, local government agencies and academics toward institutionalizing the process of 'learning' in police departments so that they can better serve and strengthen their communities. Recent research has argued that "the informal network among police planners appears to be a critical element in the research/planning process". This component of the project is an effort in this direction – providing support for the expansion of the informal network of police professionals enabling better police management practices.

This project furnishes a technical platform for police departments to make more informed decisions via dialog with cities having similar conditions and problems. The application software developed here utilizes case-based reasoning – CBR, an artificial intelligence technique, and computing power to compare several cities across multiple dimensions - demographic, crime and enforcement. The software retrieves 'matches' that would have been otherwise difficult to generate due to limits in human cognitive abilities particularly in short term memory. The system also addresses the issue of information availability, helping departments share knowledge. This sharing of knowledge will allow police departments to benefit from the cumulative experiences of others.

For years, analytical techniques and computers have been used for law enforcement applications. For example, analytical and mathematical models have been recently used to manage and control illicit drug activity. On the other hand, use of computer systems included those that keep track of arrests, crimes and their types, criminal history, and missing persons. Recently, more advanced systems have helped police respond to crime - Automated Fingerprint Identification and Computer-aided Dispatch (CAD) systems, particularly tied to 9-1-1 systems. While these systems can also help an officer by providing much information in preparing to respond to a call, they tend to be geared to fast response.

Other systems include those that allow sharing of information among agencies (e.g. FBI's National Crime Information Center, NCIC). They provide cross-checking capability for firearms and child care checks, speed dispatch through use of global positioning systems (GPS) for automated vehicle location (AVL), aid identification of suspect vehicles and enhance communication by putting terminals in police cars. Still these are geared toward day-to-day, and minute-to-minute tactical needs of police. These innovations are ways of primarily "managing the environment, rather than changing it".

A recent emphasis on Geographic Information Systems (GIS) allows mapping of information to help crime analysis and problem detection. These systems can be used to map crime patterns within a city where various kinds of data can be overlaid over the maps to facilitate visualization of "hot spots" of crime or trends. This kind of application can aid in planning and problem-solving.

The software developed in this project, however, provides an additional dimension for policing. It promotes a wider view, a longer-range perspective, encouraging greater communication among various agencies and community groups, creating opportunities for cooperation. It facilitates the identification, dissemination and use of the more useful and successful systems and strategies.

Technologically, the project has some commonality with some previously developed software. The FBI developed an automated crime profiling expert system which found previous incidents for solving current crime incidents. The similarities were judged using "rules" that were developed by experts in criminal justice profiling. The differences from the current project are both in technology and in use. Rules were used instead of general similarity measurement techniques, and crime investigation was the focus instead of communication and strategic planning. Constructing rules is a painstaking process, and is only possible if genuine experts exist. Even *with* experts, it may be hard to formulate rules for problem-solving. In the current project, using rule-based AI technology would not be desirable due to a scarcity of experts for an intrinsically difficult task.

The technique used here, CBR, has been an active research area in computer science for over 15 years and has been successfully employed for numerous applications. The range of applications of CBR includes automobile diagnosis, medical diagnosis, banquet planning, identification of cloud types, range-land management, explanation generation, route planning, device design and architectural design. More recently, this technique has received much attention due to its use for commercial purposes. For instance, Dell Computers reports great success using CBR for its customer support services. Other commercial uses include matching the colors of plastics by General Electric and detecting rail defects by Dutch Railways. While research prototypes have addressed legal issues, such as trade secrets law and criminal sentencing by judges, the knowledge CBR has not been used for many law enforcement tasks. Thus, the project is both a new use of the technology and a new capability for those involved in law enforcement.

Modeling Framework

Conceptually, this model seeks to generate a list of matching cities which will enable the 'cue' city to satisfy one of the following broad goals - (a) crime reduction, (b) reducing policing expenditure while maintaining same level of service, (c) making a case for getting additional funding, and (d) increasing cooperation among departments (for instance, departments facing similar problems may request joint funding from government agencies). The guiding principle in this project is to let the computer do what it does best – handle a large volume of data and calculations - and let humans do what the computer does not do best, communicating with people, and synthesizing lessons from people's experience.

To achieve the above goals, the software requires a process of appropriately 'matching' communities based on input data. The approach used here considers each community or city a "case" with the socioeconomic, law enforcement, and crime factors as part of the case problem description; these are inputs to the model. Based on the problem description 'Crime Similarity System' (CSS), retrieves communities that have relevant similarity, allowing the law enforcement user to contact representatives of the retrieved communities for discussions enabling dialog, discussion and possible learning.

The steps in the execution of this project involved identification of the data, development of a method for determining similarity and using similarity matching to meet different user goals. Therefore, three major aspects defining the modeling framework for CSS are (i) the input factors, (ii) the matching process that forms the intelligence core of the software, and (iii) the user goals. Below we discuss each of these components.

Input Factors

A *factor* is a piece of data describing a collection of related instances. The terms "attribute" in database systems, "feature" in CBR, "field" in data processing, or "variable" in many areas of social science capture the same idea as a factor. To arrive at the input factors to the model, extensive discussions were held with law enforcement officials from the partner police departments in the cities of Camden, NJ and Philadelphia, PA. The primary Camden police department contact was the deputy chief of police assisted by two planning officers and two special task force officers. The primary Philadelphia police department contact was a headquarters planner/investigator with specialized training in use of technology and analysis. In addition, discussions were held with academics from relevant disciplines of criminal justice, geography and public policy. Multiple interviews, including brainstorming sessions were held to identify and prioritize the inputs. Questions were posed in a semi-formal format, ensuring focus, yet allowing flexibility for the end-users to raise relevant issues not considered beforehand[2]. Sample printouts of data were provided to the officials to assist them in the discussions. Subsequently a tentative list of relevant factors was developed, which was then fine-tuned via follow-up interviews.

While the software could easily deal with hundreds of factors, it was our goal to limit the number of factors to a manageable number. In particular, when a similarity match is returned, a user may want/like to see how the community matches up. Presenting the user with *numerous* pieces of data for their community and the retrieved communities was thought to be more overwhelming than useful. Therefore, it was a major goal of the interviews to come up with the *best small set of factors*.

[2] An example question posed was: "How important do you think population is in judging similarity between communities?" This question was then followed up with an open-forum discussion allowing the experts to identify other related factors and/or a combination of factors that, in their view, had an important bearing in judging similarity among communities.

Table 1 - Factors Used in Environment, Enforcement and Crime Dimensions

ENVIRONMENT	Population density
	Median household income
	percent of households receiving public assistance
	percent of population between age 16 and 24
	percent of adults who haven't completed high school
	percent of households that are owner occupied

ENFORCEMENT	Number of police officers
	Police officers per 100,000 population
	Number of requests for police service per officer
	Police operating budget per 100,000 population
	Percent of officers on patrol
	Whether the police had a special gang unit
	The percent of police officers assigned to special drug units
	The racial match of the police force to the community

CRIME	Total violent crime rate
	Murder rate
	Drug arrest rate
	Total non-violent crime rate

After the identification phase, these factors were aggregated into three *dimensions*: environment, enforcement and crime, where a dimension is an aggregation of factors describing one aspect of a community. The factors incorporated in each dimension are shown in Table 1. The environment dimension deals mainly with the socioeconomic conditions of the community. These factors were taken directly from the U.S. Census data where available, while others were generated from multiple raw factors. For example, population density was calculated from population and land area, and percent of adults over 25 without a high school diploma was calculated from numbers of adults in each educational sub-category.

The enforcement dimension measures resources available for law enforcement, demand for service on those resources, and deployment of these resources toward fighting crime. The enforcement factors were either obtained directly from the Law Enforcement Management and Administrative Statistics (LEMAS) survey or were calculated via a combination with the Census data. The racial match factor – developed to measure the sensitivity of deployment to ethnic and minority groups (a key factor for garnering community help in community policing) – involved a slightly more complex calculation. It quantitatively measures whether the ethnic composition of the police force matches that of the population it serves (see Appendix).

The crime dimension quantifies the prevalence of crime in a community, and each factor incorporated in this dimension has been normalized per 100,000 population to ease comparison. The factors chosen (Table 1) were obtained from the Federal Bureau of Investigation's (FBI)

Uniform Crime Reports.

The Case-Based Reasoning Model

As mentioned above, a technique from the field of Artificial intelligence, Case-Based Reasoning (CBR) was used. CBR's intelligence lies in its 'memory' of a successful case of problem solving, used to solve new instances of similar problem, without complete model or knowledge of the task. A central step in any Case-Based Reasoning system is the *retrieval* of appropriately similar previous cases. CBR's inherent strength in targeted retrievals makes it desirable for use in the application. One popular form of case-based retrieval is nearest neighbor in which all factors are combined to derive similarity between cases. The core of CSS uses three concurrent nearest neighbor retrievals, one for each of the three dimensions.

In a nearest-neighbor retrieval, cases in 'memory' are compared to the given (or "cue") case on all relevant problem-description factors. For each factor, the difference between the cue and the stored cases is calculated. The resulting differences are combined using one of several potential metrics. The 'Manhattan' metric used by CSS (see Equation 1) adds the absolute value of the weighted differences together. Regardless of the metric, the case in memory with the smallest result is retrieved as the 'closest' case i.e., it is the "least distance away". As typical of most uses of nearest neighbor approach, each factor is weighted to capture their relative impact.

$$\text{Distance}_j = w_1|c_1 - m_{1j}| + w_2|c_2 - m_{2j}| + \ldots\ w_n\ |c_n - m_{nj}| \qquad (24)$$

Where, c_n is the value for the nth attribute for the cue community, m_{ij} is the value for the ith attribute for the jth community in memory and w_n is the weight for the nth attribute.

This nearest-neighbor approach takes advantage of the computer's capability to consider all factors simultaneously, whereas a person asked to judge similarity, would have to focus on only a few relevant factors due to short-term memory limitations. The software program has the additional advantage of being able to keep in its memory hundreds or thousands of cases efficiently.

Users from partner police departments indicated that cities of drastically different populations would not be useful matches even if many attributes were similar, since law enforcement issues are significantly different qualitatively. Therefore, in CSS, the nearest-neighbor retrievals are limited by city population. Accordingly we chose to perform distance measurement only on comparable cities - those with a similar population to the cue-city. Utilizing FBI's standard categorization of cities into groups by population, CSS's nearest-neighbor retrieval only considers cities in the same or adjoining population categories as the cue-city.

Besides its obvious advantage over manual searches, the nearest-neighbor approach also provides several advantages over human-assisted database retrieval. A person performing database retrievals still must choose a criterion to use, which requires picking a small set of factors and determining cut-off(s) for each factor. In such a database retrieval, using a very small set of factors could lead to inaccuracies if some important factors are left out. On the other hand, using a large number of factors would be tedious and impractical for the user. Additionally, the user may have to do experimentation for selecting most appropriate cut-offs, and wrong choices may lead to too few or too many retrievals. In contrast the nearest-neighbor approach weighs all the factors appropriately in finding the best 'partial' match where no perfect match for a particular factor is required.

Assigning Weights

The process of finding the appropriate factor weights required multiple interviews with the experts from partner police departments. As discussed above, the experts were first asked to suggest possible factors. Next they were asked to place the factors for a dimension into 2 lists – list A consisting of the factors perceived more important and list B consisting of the relatively less important ones. They were then asked to assign a cumulative percent weight on the factors in list A. After assigning a cumulative weight on each list, they ranked the factors in each list. The experts were then asked to further split each list into two smaller sub-lists and repeat the above procedure if necessary until they felt "comfortable" to assign a numerical weight on *each* factor.

In the second stage the experts were shown the impacts of the weights via results and were given the opportunity to revise their ratings. Even though this stage was tedious (for it required looking at a myriad of data), its impact on the quality of the model solution was invaluable. Fine-tuning of the weights was an on-going process through all the stages of this project. It is important to note that this technique avoids the problem of prohibitive number of comparisons (the task at hand involved over 300 cities and 20 factors) that methods such as a conventional Analytic Hierarchy Process approach would encounter.

User Goals

Initial discussions with the partners, Philadelphia and Camden police departments, focused on the question, “How can the software aid strategic planning and decision making?” While many possible uses were considered, we finally converged to the following four: (a) “*find very similar*” cities, (b) “*find more efficient*” cities, (c) “*find more effective*” cities, and (d) “*find funding argument*” city matches. For each of these user goals, CSS performs three nearest neighbor retrievals, one for each of the dimensions. Depending on the user’s goal, these individual retrievals are then combined to produce a final retrieval result.

The “*find very similar*” goal is targeted to yield communities that are similar to the cue-city on all three dimensions - enforcement, environment and crime. If the software outputs a “very similar” match, it implies that the match was among the top 15% in similarity on all three dimensions. A top 15% cut-off in similarity was implemented since it yielded reasonable results. Unlike typical nearest-neighbor retrievals where all factors would contribute to *one* similarity measurement, our approach combined three nearest-neighbor retrievals, forcing retrieved communities to be similar on *all* three dimensions. This ensures that a close similarity on one dimension (e.g. crime) cannot compensate for a lower similarity on another dimension (e.g. enforcement). For instance, Compton is the most similar city to Camden on socioeconomic environment, and 10th most similar on crime, but is only 187th most similar on enforcement out of 325 communities. Using our approach, Compton will not be considered a very similar community to Camden while in a single-weighted nearest-neighbor retrieval it would be 2nd most similar (using same weights as CSS). While Compton may be relevantly similar to Camden in terms of one of the *other* goals available in CSS, the difference in enforcement however suggests that it shouldn’t be considered “very similar”. In contrast, consider Hartford, one of the cities retrieved by CSS’s “very similar” search for Camden. Hartford rates 2nd most similar to Camden on socioeconomic environment, 17th most similar on crime, and 13th most similar on enforcement, a good all around match.

In identifying these “very similar” communities, the hope is that it will help develop channels of communication for sharing experiences and learning from each other. Such a dialog could lead to the identification of strategies that have worked in similar situations elsewhere or warn against pitfalls of failed strategies. Based on the feedback from the partners, it was felt that this information-sharing holds exciting possibilities in making the policing efforts to be proactive and strategic.

For the “*find more efficient*” goal, the application displays communities that are similar to the cue community (top 15%) on environment and crime, but significantly lower (at least 20% lower, see Figure 4.10) in enforcement resources. A possible implication is that the matching communities could be using their resources more efficiently, offering a possible opportunity for learning on how to reduce spending or re-deploy officers. Note that we are not arguing that such a match for any of the user goals is *necessarily* going to be fruitful, but that the potential is worthy of further exploration. The efficiency measure described here is, of course, limited by the data factors incorporated into CSS. Nevertheless, this measure does provide a starting point for investigating the efficient use of resources by police departments.

The third goal of “*find more effective*” outputs communities that are similar to the cue community in the environment and enforcement dimension (top 15%), but are significantly lower in crime (at least 20% lower). A link with such communities could result in initiation of similar programs, which utilize resources more effectively. It may be argued that since enforcement and crime are correlated, a community with similar enforcement and lower crime than the cue-city is an exception or an “outlier” and therefore should be ignored. For this application however, these exceptions are potentially *very interesting* and could be a function of unusual efforts of the community in reducing crime, worthy of serving as benchmark.

The “*find funding argument*” search yields communities that are similar to the cue community on environment (top 15%), but significantly higher in enforcement and significantly lower on crime. Benchmarking such communities, the user police department could make an argument that additional resources (from state and federal agencies) could possibly help bring down crime (similar to the levels of the matching community), given that both have a similar environment.

While these were the four goals implemented in the current version of the CSS application, the modeling framework allows flexibility for additional goals.

Prototype Development

Developing the CSS software was an elaborate and time-intensive process. Multiple data sources with differing formats had to be combined, resulting in additional difficulty. Next we discuss the various sources of data, extraction of items from the parent sources, coalescing of data from different sources into one dataset, normalization of the data, development of a graphical user interface, and system evaluation. Figure 4.11 provides an overview of the process of turning raw data into useful retrieval.

Data Sources

Data for the various socioeconomic factors was obtained from the U.S. Census data, available through CD-ROMs (U.S. Department of Commerce, 1992). Crime data was obtained from two sources: (a) the non drug-offenses data for the “index crime” obtained from the Uniform Crime Report on the FTP server at the University of Alaska’s Criminal Justice Center, and (b) The drug arrest data on tapes purchased directly from the FBI. The enforcement profile data was obtained from the Law Enforcement Management and Administrative Statistics (LEMAS) survey of police departments nationwide (U.S. Department of Justice, 1992), available through the Inter-University Consortium for Political and Social Research's (ICPSR) WWW site at the University of Michigan-Ann Arbor.

Data Extraction

Since the data sets were not in ready-to-use format for this application, extensive computer programming had to be performed for extracting relevant information. For example, drug arrest data from the FBI dataset provided, for every city, monthly arrest data for each offense further branched by each age/race/sex combination. Since the model did not require breakdown by time, age, race, or gender, aggregation was performed to obtain total annual drug arrests for each city. Using each data source required pulling out relevant fields from among many fields, sometimes requiring use of statistical programs like SAS.

Data Coalescence

Following extraction, datasets from all the sources were matched up. The complications included (i) the city names were not unique, and (ii) different sources used different formats for the names (e.g. all capital letters versus mixed case etc.). The data coalescence was achieved using a Pascal program that converted the most “readable” community names, from the census, to each of the other forms (e.g. removing all blank spaces, converting all lower case letters to uppercase, stripping words like “borough”, etc.) and then matched across datasets. Additionally, since city names were not unique, the matching involved checking the state and the county as well. At this stage, in addition to data coalescence, pre-processing of data was performed. For example, the Census data provided raw data on several educational levels of residents, which were converted to percentages.

Data Normalization

The successful use of a nearest neighbor approach depends on the "normalization" of data - putting all data into the same relative scale. This becomes necessary to ensure that a small difference on one factor (e.g. $10,000 difference in police budget) does not override a very important difference on another factor (e.g. difference of 500 in number of police officers).

Normalization of data was done by first calculating the standard deviation for each factor. All values beyond three standard deviations above or below the mean were set aside temporarily, and among the rest the largest and smallest values were found. The largest value (and anything above it – i.e. those that were temporarily set aside) was set to a normalized value 10, while the smallest number (and anything below it– i.e. those that were temporarily set aside) was set to a normalized value of 0. The range between the smallest and the largest values was then divided into 9 equal size sub-ranges numbered 1-9. Raw data falling in each sub-range was assigned the corresponding normalized value. For any factor in which a high number was considered ‘undesirable’ for a community, the normalized values were then reversed ($0 \rightarrow 10$, $1 \rightarrow 9$, $2 \rightarrow 8$, etc.), ensuring consistency and ease in interpretation.

The three standard deviation cutoff makes sure that 'unusual' data does not force all other values to fall on one side of the scale. The extreme values for a factor become 0 and 10, and values in between are spread into the range 1-9. Intuitively, the 'equally spaced' interval normalization process captures the mental model that magnitudes of difference between data is the key indicator of similarity, and should be preserved. Thus, for instance, it was not desirable to use z-scores, which tends to provide fine distinctions among data points that are close together in raw score but are separated by many other data points. The normalization scheme used here, on the other hand, preserves the original clustering of data.

Software Application and the Graphical User Interface

The software code for this component was written in the "C" programming language (see Appendix) and is portable to multiple computer platforms. For the CSS application to be used easily and effectively by police officials, development of a good graphical user interface (GUI) was imperative. This front-end of the software was developed using an application development toolkit. The GUI (see Figure 4.12) allows the user to specify their desired goal, using user-friendly point-and-click technology. To specify the goal, or choose the cue community, the user clicks on the appropriate choice in the list box. The appropriate core algorithm is invoked once the user makes a goal selection. Based on a follow-up suggestion from the partner police departments, the current version of CSS now allows users to view the data comparison between the cue-city and any city (of interest to the user) in the retrieval list. The layout and functionality of the software have been continually updated based on ongoing interactions with the police officials.

System Evaluation

After routine unit and system testing, police users checked the reasonableness of the retrievals using their knowledge and experience. Detailed factor-by-factor calculations of similarity were then presented to officials to show them the underlying process. Weights for different factors in the nearest-neighbor algorithm were fine-tuned based on their feedback and comments. Next, the users were asked for feedback on the usability of the program. It was at this stage of the evaluation that the need for displaying the underlying data was identified by the users. Displaying data comparisons enables on-line exploration and analysis of the similarity measurement. Besides helping the users in evaluating the "goodness" of the match, this capability generates user confidence in the reliability of the software.

Ideally, we would like to evaluate the quality of strategic decisions made by police departments supported by CSS vs. the quality of those without it. However, such a first order test is infeasible due to several reasons. First, gathering a sufficiently large sample of strategic decision-makers from police departments is impractical in itself. Second, the strategic decisions being made by different police departments in a given time frame would tend to be intrinsically different and therefore hard to compare. Finally, the success of strategic decisions is best judged after observing the effects over a period of years – and even then, the effects are confounded with many influences making it a complex task. Hence, a second-order test evaluating the system's support for retrieving relevantly similar communities was performed. The natural standard for comparison for the retrieval task seemed to be a database. This fact was further substantiated when, during the course of this project, we would often be confronted with the question both from end users and other participants “why not just use a database?” A web search engine was clearly not a good standard for comparison, because the task is not well suited to keyword search. Hence, we evaluated the support provided for retrieval by CSS in comparison to that provided by a database. The experiment used 20 MBA students as human subjects.

The experiment required the subject to retrieve cities similar to two cue cities, Camden, NJ and Cincinnati, OH, using the CSS software and Microsoft Access database software. While the subjects had no prior experience in using either software, they were computer proficient with strong analytical skills, and therefore easily trained. A ten-minute training was given on Access followed by a practice exercise on a similar, scaled-down task that lasted approximately 5 minutes. Since the training focussed exclusively on the query capability of Access – the only part needed for the experimental task – ten minutes of training time was adequate, which was verified by pilot testing. The practice exercise directly paralleled the experimental task; it was just a smaller scale – fewer factors and fewer records in the database. In addition to the ten-minute training on Access, each subject was trained on CSS for about 1 min (with no practice exercise). For all subjects, the time spent and their retrievals for both tasks were recorded. In addition, a follow up questionnaire was given to help ascertain their level of satisfaction in using each software.

While time was used as a performance measure directly, the actual retrievals had to be translated into a measure that represented quality. It may seem logical to have an expert compare the quality of retrievals by CSS and the subjects, in reality, the myriad of factors and dimensions makes such a task practically impossible (such an evaluation was indeed attempted but the expert was overwhelmed by the magnitude of the task). Therefore, the comparison was done using an independent quantitative measure. Closeness to the cue-city represents desirable retrieval for this task and is an indicator of quality. This closeness of retrievals for each subject was calculated by first ranking all the communities in the database on each of the input factors. The difference of ranks between the cue and retrieved city on all input factors was then summed up to get a closeness measure for that particular retrieval for each subject. Finally all of these closeness measures were summed over all the retrievals obtained by the subject. Equation (2) summarizes this calculation. In summary, a lower difference rank sum represents a closer match to the cue-city and in effect represents good quality. To maintain fairness to both methods, this quality measurement is different and independent from the similarity measurement used by CSS.

$$Q_{ij} = \sum_{s\in S_i}\sum_{k=1}^{18} | R_{sk} - R_{jk} | \quad \text{for } i = 1...20,\ j = 1,2 \qquad (25)$$

where

i = subject i

j = cue cities; 1 = Camden, 2 = Cincinnati

k = input factors

S_i = set of retrieved cities for subject i

R_{lm} = Rank of city l on input factor m

Q_{ij} = Quality of retrievals obtained by subject i for cue city j.

In the initial pilot run we observed some learning effect i.e., time taken by subjects for second city was less than the first one, for both CSS and Access, thus the experiment was designed to counter-balance the ordering of software and city searches. Table 2 summarizes the average time taken for completing the task for each software and city.

Table 2 : Average Retrieval Time (in seconds)

Software	**CSS**		**ACCESS**	
City	Camden	Cincinnati	Camden	Cincinnati
1st City	27.7	27.3	1003.7	1200.2
2nd City	16.6	12.6	681.1	494.8

For example, using Access, it took an average of 1003.7 seconds for subjects to find matching cities for Camden when done first and 681.1 seconds when they did the retrieval for Cincinnati *before* doing Camden. Since second row values are less than the corresponding values in the first row, the learning effect is clear. To accommodate this learning effect, a crossover analysis of variance design was employed using a nested, three factor model. The statistical analysis showed that CSS did significantly better than Access ($p<0.001$) on the time performance measure.

Table 3 summarizes the calculated, average quality measurement for each software and cue-city. Since crossover learning effect was not found to be significant for the quality performance measure, aggregate measurements are provided. An analysis of variance showed CSS better than Access on the quality measurement as well ($p< 0.01$).

Table 3: Average Quality Measurement (Low is good)

Software	**CSS**		**ACCESS**	
City	Camden	Cincinnati	Camden	Cincinnati
Quality of Retrievals	1283.17	1393.13	1530.3	1511.7

Table 4: Post – Experiment Questionnaire Results

(on scale 1-5 where 5 is good and 1 is bad)

Software	**CSS**	**ACCESS**
Searching for Matches	4.95	2.63
Ease of use	5.00	3.42

Finally, the post-experiment questionnaire showed that the users clearly favored CSS over Access on both (1) efficiency and quickness of search, and (2) ease of use (See Table 4).

In looking at the results, it is not surprising that search performed using CSS was much faster than search by using Access. CSS was developed to specifically support the task, while a database is a general-purpose program supporting many different searches. It may be argued that given CSS is specialized software, a comparison with a general-purpose database is not fair. However, as discussed earlier Access is the only available alternative to CSS, and therefore a valid choice for comparison. Further, the fact that the task is difficult and time consuming using a database only goes on to emphasize the need for a specialized software such as CSS.

The experiment also demonstrates the better performance of CSS than the human-assisted database retrievals on the quality dimension. Perhaps, this can be attributed to CSS' ability to capture all the factors to generate a best partial match.

As discussed above, ideally we would have liked police administrators and decision-makers as subjects in this evaluation experiment. However, that was not possible due to the time commitment required from a large group of senior officials. Despite this limitation, the experiment does demonstrate the effectiveness of the CSS software in generating quality retrievals expeditiously within a subject population group.

Examples

Here we illustrate how the CSS software can be used to achieve two of the supported user goals – "Find More Effective" and "Find Funding Argument". The "Find More Effective" feature of the software helps the user in finding cities that are similar to the cue community in the environment and enforcement dimensions, but significantly lower in crime. For example, a run of the software for finding a more effective city using Harrisburg, PA as the cue-city yields the city of New Bedford, MA as a match. The weighted differences of 51 and 60 shown in tables 5 and 6 show that the two cities are quite similar in the environment and enforcement dimensions respectively (differences can range from 0 on the low end to 1000 on the high end). The CSS software then investigates the crime dimension (Table 7) indicating a significant difference of 314 between the two cities (recall, a higher normalized score implies a lower crime level). In fact, New Bedford's crime level of 801 is 61% better than Harrisburg's 487 crime level. Given the same resources and environment but a lower crime level suggests that New Bedford seems be more effective than Harrisburg in its law enforcement efforts. Harrisburg may be able to gain valuable information and learn new strategies from New Bedford police department that could help them in reducing the crime level. The software can even help identify specific opportunities for learning. In this example, the difference in drug related arrests in both communities seems to be significant and perhaps a point worthy of further exploration. If New Bedford has some specialized drug-related tactics that are contributing to a lower crime level, Harrisburg can identify and implement similar programs.

Table 5 - Calculations for the Environment Dimension (Find More Effective)

ENVIRONMENT	**% People Receiving Public Assistance**	**% age 16-24**	**Population Density**	**% People Less High School Education**	**Median Household Income**	**% House Owner Occupied**
Harrisburg	4	6	8	5	3	4
New Bedford	4	6	8	2	3	4
Difference	0	0	0	3	0	0
Weight	28	22	17	17	11	6
Weighted Difference	0	0	0	51	0	0

Total Weighted Difference: 51

Table 6 - Calculations for the Enforcement Dimension (Find More Effective)

ENFORCE MENT	Police officers	Police officers /100K population	Police operating budget $/1000	Police requests per officer	% officers sworn on patrol	Racial match factor	Gang units	% police officers assigned to drug units
Harrisburg	3	1	2	7	7	5	10	3
New Bedford	3	1	2	6	8	8	0	4
Difference	0	0	0	1	1	3	10	1
Weight	29	29	21	9	9	3	3	3
Weighted Difference	0	0	0	9	9	9	30	3

Total Weighted Difference: 60

Table 7 - Calculations for Crime Dimension (Find More Effective)

CRIME	Murders/100K population	Violent Crimes/100K population	Drug Arrests/100K population	Non Violent Crime/100K population
Harrisburg	6	5	2	6
New Bedford	10	7	6	8
Difference	4	2	4	2
Weight	36	29	21	14
Weighted Difference	144	58	84	28

Total Weighted Difference: 314

Weighted Level for Harrisburg: 487

Weighted Level for New Bedford: 801

Next let us consider the "Find Funding Argument" user goal. Often, State/Federal criteria for funding puts police departments in the position of meticulously justifying requests for additional funding. The "*funding match*" feature identifies communities with similar environments, but which enjoy access to greater enforcement resources and lower crime levels. The user police department can cite the examples of such communities to argue for additional resources, "Since Community X is similar to ours, we can reduce our crime level to match theirs, if we had access to comparable enforcement resources." For instance, if the Hawthorne, CA, police department used the funding match features of CSS, it will identify Yonkers, NY, as a potential community to use as justification for additional enforcement resources. Yonkers has almost twice the number of police officers and per capital operating budget than Hawthorne, even though their environments are similar. Perhaps, this is the reason why Yonkers enjoys a 50% lower murder rate and 60% lower violent crime rate than that of Hawthorne. Hawthorne can make a compelling case for additional resources citing the example of Yonkers.

Concluding Remarks

"An African proverb goes, 'No one tests the depth of a river with both feet'. Yet, thoughtful police sometimes wonder if their department is an exception to this rule. They watch bewildered and despairing as their organization leaps from one tactic and program to another, rarely bothering to conduct a meaningful feasibility study or figure out what did not work and under what conditions the last time a similar problem was tackled". The CSS software developed here is geared toward helping departments go beyond learning from their own experiences alone, benefiting from the cumulative experiences of other departments. The richness of shared relevant experiences holds immense possibilities for cooperation and innovation.

A recent National Institute of Justice report emphasizes the significance of an informal network among police planners and suggests its enhancement by: (1) acknowledging and encouraging the network of communication among police organizations, (2) providing resources to key organizations in this network to support their dissemination activities, (3) continuing efforts to enhance research capacity of police organizations, (4) choosing sites for research and demonstration projects on the basis of an agency's prominence in the communication network, and (5) continuing efforts to make research available via electronic media. This project is in consonance with the above recommendations because it (1) specifically encourages communications among police organizations, (2) will help successful organizations disseminate their success stories by leading other departments to them, via the *find more efficient* and *find more effective* capabilities, and (3) helps police departments do their research by providing a quick link to relevant departments. Further, with the possibility of collaborative funding, demonstration and evaluation of policing strategies could be an indirect outcome of this project.

Besides the practical usefulness of this project in enhancing the capability of police department to make more informed decisions, the project makes three significant contributions. First, it provides a unique modeling framework classifying and aggregating input data into three dimensions – environment, enforcement and crime. Relevant factors are identified and new measures (e.g. racial match index) developed that help define these dimensions. Second, this work identifies four important strategic goals that assist police departments in moving toward a direction of proactive management. Third, this work develops a unique combination of three similarity measurements that help in matching user goals.

The CSS software developed here is a prototype and can be improved in several ways. A possible enhancement, suggested by the Philadelphia Police representatives, would be a feature in which a city can target to reach a certain reduced crime level. The software would analyze the goal's feasibility displaying cities that have similar environment to the cue-city and are closest in crime rate to the target. Using this potential functionality of the software, police managers could make future plans to reach the specified crime level reduction target *utilizing* the experiences of the retrieved cities. This feature could be especially useful in helping communities to revitalize by emulating other successful communities.

While an extensive effort went into development of the modeling framework and software application, CSS is certainly not a "foolproof" tool. Since its inception, CSS was meant to be an Artificial Intelligence tool that learns from experience enhancing its capability to provide better matches. A future improvement would add a feature allowing the software to incorporate feedback for better matching. On coming across a non-useful match, a user can provide useful input allowing the software to 'learn' and adjust the weights used in the nearest neighbor retrieval.

Another improvement would be to make the program run on the World-Wide-Web by converting it to Java, so that it would be more accessible and amenable to knowledge sharing. Knowledge management systems are gaining popularity, especially since distilling and imparting lessons learned has become more critical in today's rapidly changing environment. A knowledge-capture feature could be added in CSS to document actual contacts made between police departments using this tool. The departments could annotate the data with comments about the cities contacted and useful information obtained. This would then advance the current application into a "lessons learned" system.

In addition to this current application, the modeling framework developed here can support other purposes beyond law enforcement. It could be a useful tool for community revitalization and business relocation efforts. For instance, using a time-series feature in the software, communities can lure new businesses by showing lowering trends in crime in their neighborhood vis-à-vis other competing communities. Further, a community can attempt to attract businesses from other targeted communities - communities that are similar, but whose environment or crime situations have declined over the preceding years.

In contrast to the AI approach used here of finding similar cases and generating matches, alternate paradigms such as Linear Programming based Data Envelopment Analysis (DEA) could be utilized. A DEA analysis could generate relative efficiency lists and find weight assignments enabling comparison. However the CBR approach has two significant advantages over the DEA approach. First, the application provides only those efficient communities that are similar to the cue community. In other words, CSS will not give a retrieval of an efficient community that is significantly different from the cue community in crime, thereby increasing the probability of usefulness of CSS's matches. Second, unlike the DEA methodology, the approach used here is able to distinguish two separate kind of inefficiencies – (a) higher inputs (via the find more efficient goal), and (b) not as good outputs (via the find more effective goal).

While the software was intended for use by police departments and law enforcement agencies, the concept is potentially useful to a wide range of managers in diverse industries. The framework of finding similarities and differences across multiple dimensions is really a novel framework for data mining. This type of framework can be applied to applications involving use and analysis of multi-dimensional data. For instance, in credit risk analysis, instead of combining dissimilar factors into one measure of credit-worthiness, the separate dimensions of "economic resources" factors, "personal stability" factors (such as how long in their current job), and collateral factors (e.g. appraisal of a house) could be preserved separately and then compared in a meaningful and effective manner. Indeed, as data mining gains more importance, the current model can offer a framework useful in many applications. The modeling approach taken here could provide a new framework for data mining, applicable to a variety of tasks.

CSS is an effort toward bringing a strategic, co-operative, learning and proactive viewpoint among police departments. The current version is a completed prototype that illustrates the value of this approach. The software's usefulness lies in its easy usage, displaying only relevant information while saving the user from the tedious task of calculations and uncovering relevant factors. However, it is a decision-making tool that is targeted to *assist* and not *replace* the decision-maker. With an Internet application interface effort under exploration, its usage could become nationwide, offering possibilities and innovations beyond the ones presented here. Indeed, these exciting possibilities of learning and sharing among communities are valuable in and of themselves.

Figure 10: Efficiency Goal Flowchart

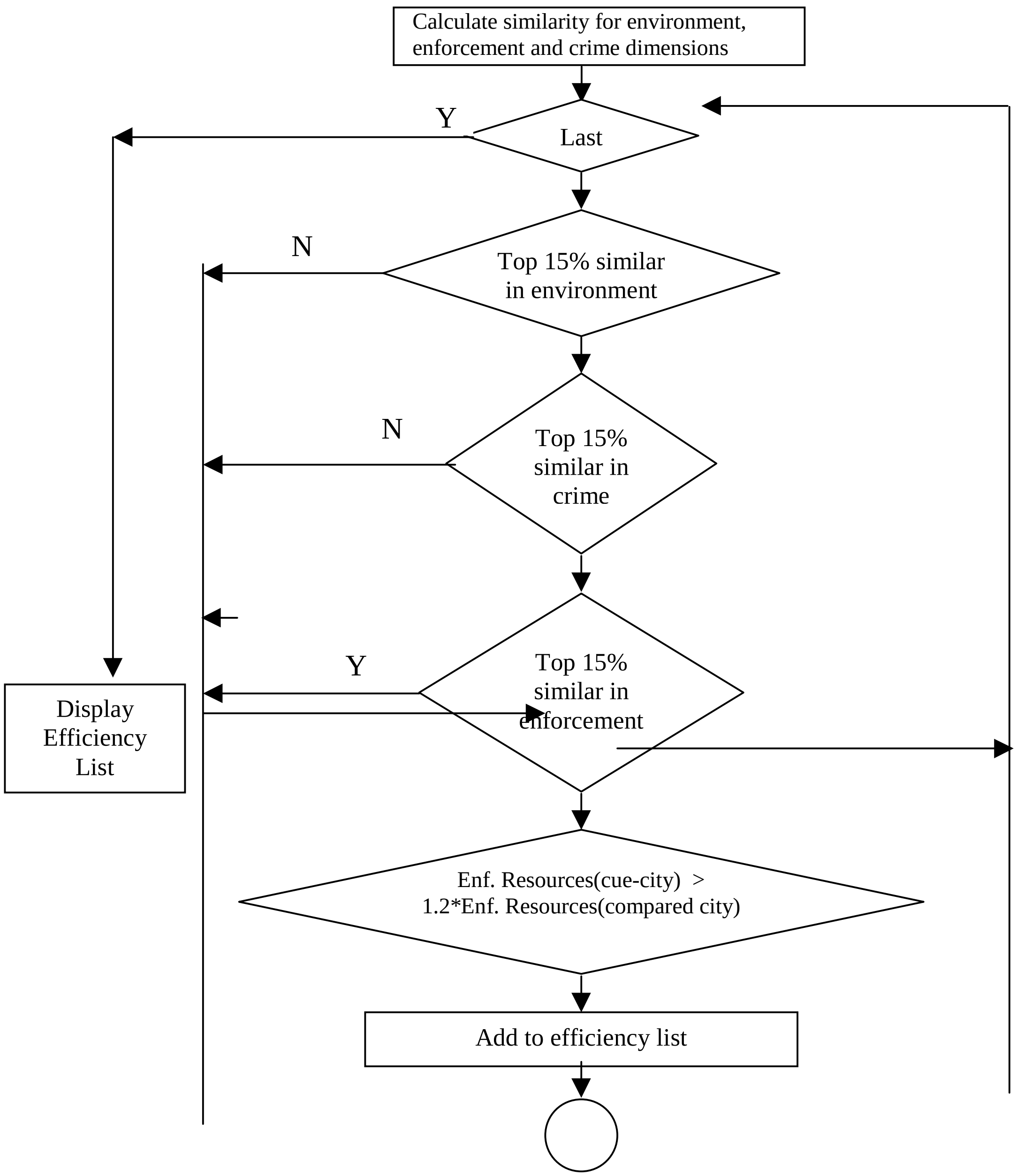


Figure 11: Process Flow Diagram for CSS

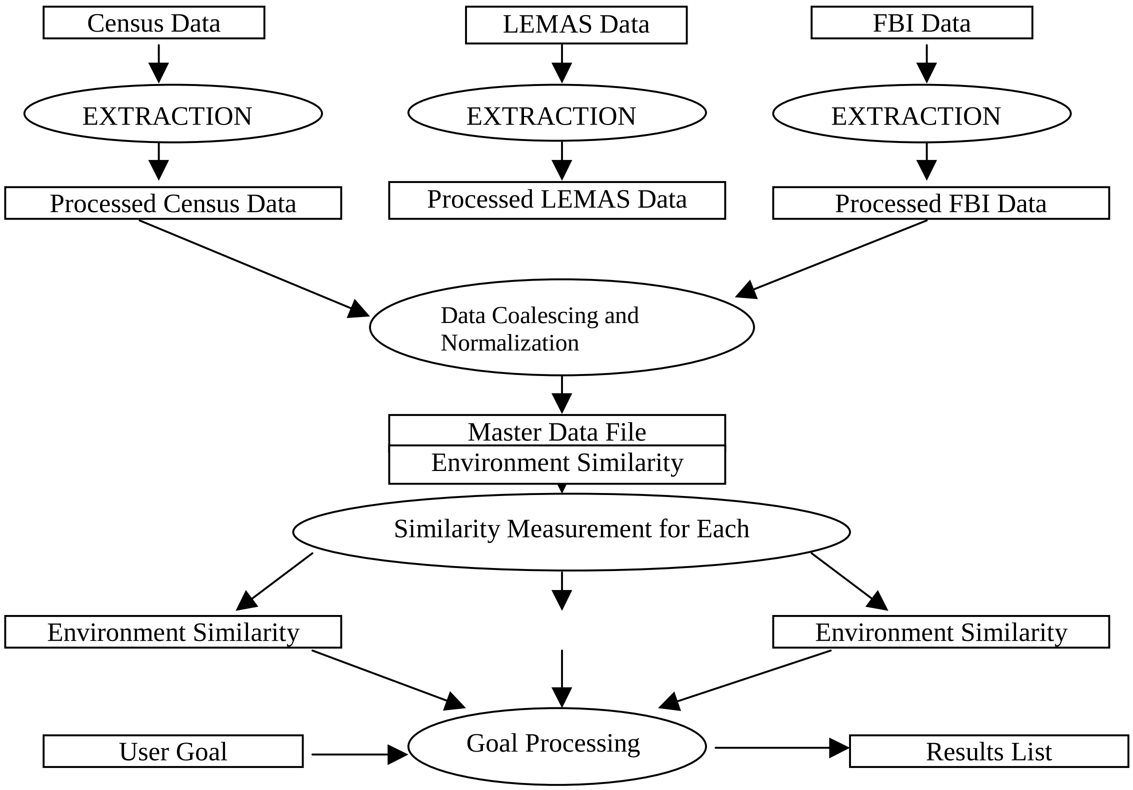
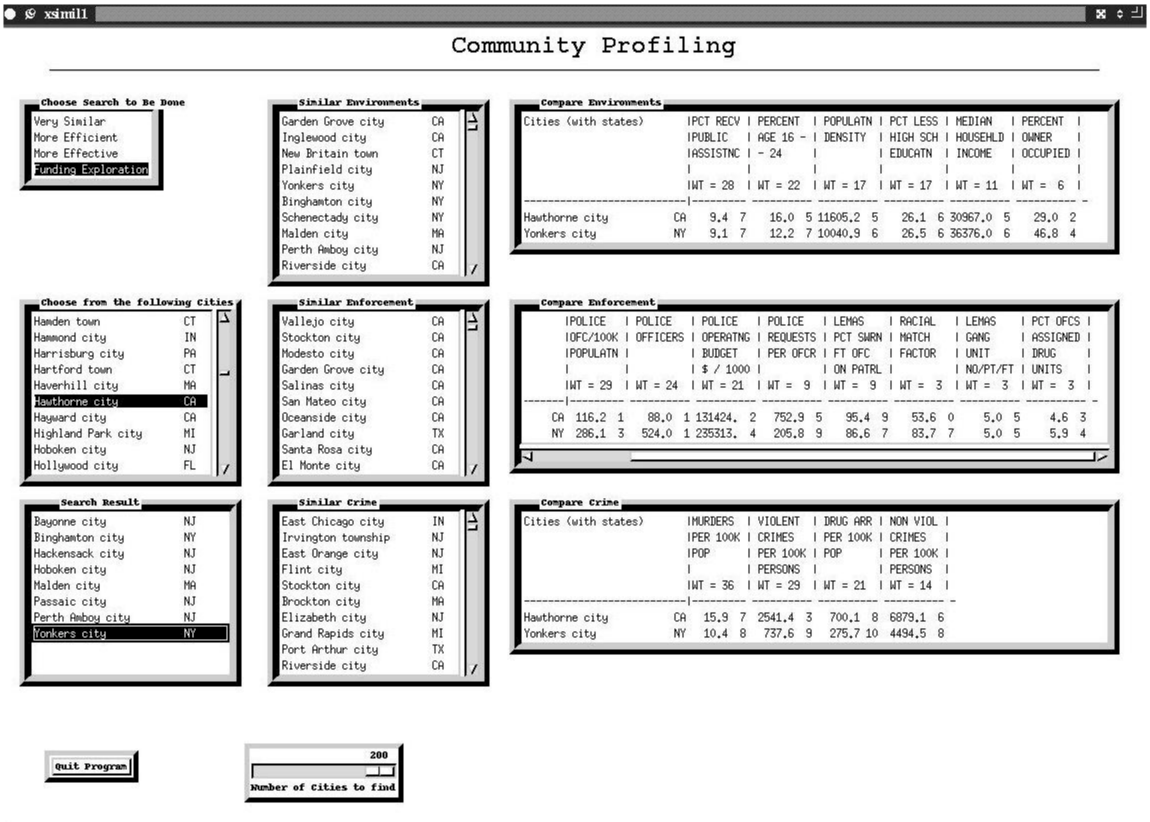


Figure 12: A Sample Screen Capture from the Interface of CSS Software



APPENDIX

1. Racial Match Factor Calculation

Here we show calculation of the racial-match factor that is one of the factors in the enforcement dimension of the CSS. Consider the data shown in Table 8. First the percent of the police force in each racial group is calculated along with the percent of community population in each group. Next, for each racial group, the smaller of the two percentages (community and police) for each racial group is determined. Then, these minimum values are added together to get the racial match index. If there is a perfect match between the percentages in the community and the police force, the index will be 100. Any under-represented racial group leads to a lower index value. In the above example, a mismatch in percent of African-Americans in the community (41.3%) and percent of African-Americans in the police force (10.5%) lowered the racial match index. An index value significantly lower than 100 would indicate a poor racial match

Table 8 – Illustrating the Racial Match Index Calculation

Racial Group	Column 1 % Community	Column 2 % Police Force	Minimum (of Column 1 and 2)
White	55.0	87.7	55.0
African-American	41.3	10.5	10.5
Asian	1.7	0	0
		Racial Match Index =	65.5

Appendix. Data-Driven Framework for Understanding Criminal Activity

Consultant: Alok Baveja

Progress Report – January 31, 2001

To achieve the goal of this project – viz. understanding criminal activity, a set of appropriate inputs to the model needed to be identified. These inputs or *factors* are pieces of data that holistically describe criminal activity within a city of concern.

Input Factors Identification (Phase I of the project):

To arrive at the input factors to the model, extensive discussions were held with law enforcement officials from the partner police departments in the cities of Camden, NJ and Philadelphia, PA. Our primary Camden police department contact was the deputy chief of police assisted by two planning officers and two special task force officers. Our primary Philadelphia police department contact was a headquarters planner/investigator with specialized training in use of technology and analysis. In addition, discussions were held with academics from relevant disciplines of criminal justice, geography and public policy. Multiple interviews, including brainstorming sessions were held to identify and prioritize the inputs. Questions were posed in a semi-formal format, ensuring focus, yet allowing flexibility for the interviewees to raise relevant issues. Subsequently a tentative list of relevant factors was developed, which was then fine-tuned via follow-up interviews.

After the identification phase, these factors were aggregated into three *dimensions*: environment, enforcement and crime, where a dimension is an aggregation of factors describing one aspect to criminal activity within a city. The factors incorporated in each dimension are shown in Table below.

The environment dimension deals mainly with the socioeconomic conditions of the community. The enforcement dimension measures resources available for law enforcement, demand for service on those resources, and deployment of these resources toward fighting crime. The crime dimension quantifies the prevalence of crime in a community.

While the model could easily deal with hundreds of factors, it is our goal to limit the number of factors to a manageable number. Presenting *numerous* pieces of data will be more overwhelming than useful. Therefore, it is a major goal of this phase of the project to come up with the *best small set of factors*.

Table - Input factors aggregated by dimensions

ENVIRONMENT	Population density Median household income percent of households receiving public assistance percent of population between age 16 and 24 percent of adults who haven't completed high school percent of households that are owner occupied

ENFORCEMENT	Number of police officers Police officers per 100,000 population Number of requests for police service per officer Police operating budget per 100,000 population Percent of officers on patrol Whether the police had a special gang unit The percent of police officers assigned to special drug units The racial match of the police force to the community

CRIME	Total violent crime rate Murder rate Drug arrest rate Total non-violent crime rate

Progress Report – February 28, 2001

After identifying the appropriate data-pieces, the next step was to find reliable and accurate data sources for input. Multiple data sources with differing formats were identified. They had to be combined, resulting in additional difficulty. At this stage of the project, we identified the various sources of data, extracted items from the parent sources, coalesced different sources into one data-set and normalized the data.

Data Sources

Data for the various socioeconomic factors was obtained from the U.S. Census, available through CD-ROMs (U.S. Department of Commerce). Crime data was obtained from two sources: (a) the non drug-offenses data for the "index crime" obtained from the Uniform Crime Report on the FTP server at the University of Alaska's Criminal Justice Center, and (b) The drug arrest data on tapes obtained directly from the FBI. The enforcement profile data was obtained from the Law Enforcement Management and Administrative Statistics (LEMAS) survey of police departments nationwide (U.S. Department of Justice), available through the Inter-University Consortium for Political and Social Research's (ICPSR) WWW site at the University of Michigan-Ann Arbor.

Data Extraction

Since the data sets were not in ready-to-use format for our project, extensive computer programming had to be performed for extracting relevant information. For example, drug arrest information from the FBI data-set provided, for every city, monthly arrests for each offense further branched by each age/race/sex combination. Since our project does not require breakdown by time, age, race, or gender, an aggregation was performed to obtain total annual drug arrests for each city. Using each data source required pulling out relevant fields from among many fields, requiring extensive use of statistical programs like the SAS, and general-purpose programs written in C.

Data Coalescence

After data extraction, data-sets from all the sources were matched up. The complications included (i) the city names were not unique, and (ii) different sources used different formats for the names (e.g. all capital letters versus mixed case or hyphenated versus underscore etc.). The data coalescence was achieved using a Pascal program that converted the most "readable" community names, from the census, to each of the other forms (e.g. removing all blank spaces, converting all lower case letters to uppercase, stripping words like "borough", etc.) and then matched across data-sets. Additionally, since city names were not unique, the matching involved checking the state and the county as well. At this stage, in addition to data coalescence, pre-processing of data was performed. For example, the Census data provided raw data on several educational levels of residents, which were converted to percentages.

The result from the above three steps - data collection, extraction and coalescence - was a "master" file to be used as input for the project.

Data Normalization

To use all the different pieces of information on a common platform, "normalization" of data becomes necessary – i.e. putting all data into the same relative scale. This becomes necessary to ensure that a small difference on one factor (e.g. $10,000 difference in police budget) does not override a very important difference on another factor (e.g. difference of 500 in number of police officers).

Normalization of data was done by first calculating the standard deviation for each input factor. All values beyond three standard deviations above or below the mean were set aside temporarily, and among the rest the largest and smallest values were found. The largest value (and anything above it – i.e. those that were temporarily set aside) was set to a normalized value 10, while the smallest number (and anything below it– i.e. those that were temporarily set aside) was set to a normalized value of 0. The range between the smallest and the largest values was then divided into 9 equal size sub-ranges numbered 1-9. Raw data falling in each sub-range was assigned the corresponding normalized value. For any factor in which a high number was considered 'undesirable' for a community, the normalized values were then reversed (0 10, , , etc.) ensuring consistency and ease in interpretation.

The three standard deviation cut-off makes sure that 'unusual' data does not force all other values to fall on one side of the scale. The extreme values for a factor become 0 and 10, and values in between are spread into the range 1-9. Intuitively, the 'equally spaced' interval normalization process captures the mental model that magnitudes of difference between data is the key indicator of similarity, and should be preserved ensuring accurate comparisons later. Thus, for instance, it was not desirable to use z-scores, which tends to provide fine distinctions among data points that are close together in raw score but are separated by many other data points.

Progress Report – March 31, 2001

Following the extraction, coalescence and normalization of data, the next step in the project was to combine the input factors into single measures for each of the three dimensions.

This stage required multiple interviews with the experts from partner police departments. The experts were first asked to suggest possible factors and dimensions. The three dimensions and factors were identified as:

1. **Environment** dimension with input factors of population density, median household income, percent of households receiving public assistance, percent of population between age 16 and 24, percent of adults who have not completed high-school education, percentage of households that are owner occupied.

2. **Enforcement** dimension consisting of number of police officers, police officers per 100,000 population, number of requests for police service per officer, police operating budget per 100,000 population, percentage of officers on patrol, existence/absence of a special gang unit, the percent of police officers assigned to special drug units, the racial match of the police with the community they serve.

3. **Crime** dimension consisting of total violent crime, murder rate, drug arrest rate and total non-violent crime rate.

Next, extensive independent one-on-one interviews were held with the experts for finding the appropriate way of aggregating the input factors into a single measure for each of the three dimensions. The methodology involved a sequential process. Specifically the experts were asked to place the factors for a *particular* dimension into 2 lists **– list A** consisting of the factors perceived more important and **list B** consisting of the relatively less important ones. The experts were then asked to assign a cumulative percent weight on the factors in list A. After assigning a cumulative weight on each list, they ranked the factors in each list in the following way. They were required to further divide each list into two smaller sub-lists and repeat the above procedure if necessary until they felt "comfortable" to assign a numerical weight on *each* factor. This process of hierarchical and sequential weight assignment was developed after trying a few pilot interviews in which we found that the experts found it difficult to quantify the importance of a factor directly. This step-by-step procedure worked extremely well in trials and was therefore adopted.

In the second stage the experts were shown the impacts of the weights via preliminary results and were given the opportunity to revise their weighting. For example, they would be shown the cumulative crime dimension score for two cities they were familiar with to see if the score was representing reality accurately. Even though this stage was tedious (for it required looking at a myriad of data and comparisons), its impact on the quality of the model was invaluable. Fine-tuning of the weights is envisaged as an on-going process through all the subsequent stages of this project as well.

Progress Report – April 30, 2001

After routine unit and system testing, police users checked the reasonableness of the model using their knowledge and experience. Detailed factor-by-factor calculations were then presented to officials to show them the underlying process. Weights for different factors in the algorithms were fine-tuned based on their feedback and comments. Next, the users were asked for feedback on the usability of the program. It was at this stage of the evaluation that the need for displaying the underlying data was identified by the users. Adding a display of data comparisons enabled on-line exploration and analysis of the measurement. Besides helping the users in evaluating the quality of the model, this capability generated user confidence in the reliability of the system.

For this project, ideally we would have liked to evaluate the quality of decisions made by police departments supported by our model vs. the quality of those without it. However, such a first order test is infeasible due to several reasons. First, gathering a sufficiently large sample of high-level decision-makers from police departments is impractical in itself. Second, the strategic decisions being made by different police departments in a given time frame would tend to be intrinsically different and therefore hard to compare. Finally, the success of strategic decisions is best judged after observing the effects over a period of years – and even then, the effects are confounded with many influences making it a complex task. Hence, we performed a second-order test evaluating the system's support for retrieving relevantly similar communities. The natural standard for comparison for the task seemed to be a database. This fact was further substantiated when, during the course of this project, we would often be confronted with the question both from end users and other participants "why not just use a database?" Therefore the system evaluation was done using a lab experiment using student subjects.

The experiment required the subjects to retrieve cities similar to two cue cities, using our model and Microsoft Access database software. While the subjects had no prior experience in using either software, they were computer proficient with strong analytical skills, and therefore easily trained.

While time was used as a performance measure directly, the actual outcome had to be translated into a measure that represented quality. It may seem logical to have an expert compare the quality of results from the model with those from the database software. In reality, the myriad of factors and dimensions makes such a task practically impossible (we did attempt such an evaluation but the expert was overwhelmed by the magnitude of the task).

To accommodate a learning effect among subjects, a crossover analysis of variance design was employed using a nested, three factor model. The statistical analysis showed that our model did significantly better than Access ($p<0.01$) on the time and quality performance measures. Also the post-experiment questionnaire showed that the users clearly favored our model over MS Access on both 1) efficiency and quickness of search, and 2) ease of use

Despite the limitation of not using the actual end-users, the experiment did demonstrate the effectiveness of the model in providing quality and timely help to decision makers within a subject population group.

Progress Report – May 31, 2001

The final development and testing of the model/software has been accomplished. The partner police departments have validated the software's usefulness in:
(1) developing a comprehensive database that incorporates environmental, enforcement factors along with crime statistics, (2) understanding and measuring criminal activity based on a comparative, data-driven modeling framework, (3) encouraging meaningful communication among similar police departments. Further, with the possibility of collaborative funding among police departments, demonstration and evaluation of policing strategies could be an indirect outcome of this project. Besides the practical usefulness of this project in enhancing the capability of police department to make more informed decisions, the project makes two theoretical contributions. First, it provides a unique modeling framework classifying and aggregating input data into three dimensions – environment, enforcement and crime. Relevant factors are identified and new measures (e.g. racial match index) developed that help define these dimensions. Second, this project identifies four important strategic model goals that assist police departments in moving toward a direction of proactive management.

Dissemination of the project results is being done through: (1) A scholarly research article which is under final stages of preparation. This article is targeted for the European Journal of Operational Research; (2) A short public-policy piece is ready for submission to a practitioner outlet –targeting, Police Chief, the official publication of the International Association of Chiefs of Police (IACP). The International Association of Chiefs of Police is the world's oldest and largest nonprofit membership organization of police executives, with over 19,000 members in over 100 different countries. IACP's leadership consists of the operating chief executives of international, federal, state and local agencies of all sizes; (3) During the course of discussions with the partner police departments, an opportunity to model enforcement strategies was recognized. Based on these insights, a scholarly publication targeted for a special issue of Mathematical and Computer Modeling is also in final stages of preparation; (4) In addition, a screen display of the model and software will be put up on the world-wide web for public viewing.

Proj1.apr: An ArcView Project folder allowing one to view 1998 crime data from Buffalo Police Department (BPD).

Each one of the files below is actually a set of three or four files with the following extensions: .dbf, .shp., .shx, and sometimes, .spn

.dbf files contain the date, location of the crime (street address), *x*- and *y*- coordinates

In the case of robberies, the nature of the stolen property is described.

98_ robbery	
98_ robbery_ purse	Cases where purses were stolen
98_ robbery_ s_ arm	Strong-arm robberies
98_ robbery_ w_ cutting	Robberies with knives
98_ robbery_ wea_ stolen	Robberies with stolen weapons
98_ robbery_ bike	Robberies of bicycles
98_ robbery_ gun	Robberies using guns

96_ crime_ burglary	Burglaries
96_ crime_ drug	Drug crimes
96_ crime_ arson	Arsons
96_ crime_ arsonsub	Arsons for a small, high-crime area within Buffalo

The following files work with ArcView's Tracking Analyst, to animate the dot map of crimes as it evolves over time:

trackpurse
tracki _ 98_ bike
track_ 98_ s_ arm
track_ 98_ gun
track_ 98_ cutting

Each of these files is actually a set of three files containing extensions .trx, .shx, and .shp There is also a .dbf with the same name, though there is no file extension.

Program Description

cuwinpu.prg is a program run using the GAUSS programming language. It signals when changes in geographic patterns occur. The program is commented, with indications where parameters are set.