Title

Bivariate Quadrant Analysis of the Dispersion of Retail Establishments and Population in Urban Areas

Authors
Rogers, Andrei
Martin, Juan

Publication Date
1967-05-01

Peer reviewed

BIVARIATE QUADRAT ANALYSIS OF THE DISPERSION OF
RETAIL ESTABLISHMENTS AND POPULATION
IN URBAN AREAS

Andrei Rogers
and
Juan Martin

May 1, 1969

Working Paper No. 97

The research reported in this paper was partially supported by a grant from the Economic Development Administration.

Reprinted August, 1971

PREFACE

This paper is the fifth working paper of a group that have as their common concern the development of statistical methods for the analysis of the spatial structure of population and employment in urban areas and regions. Its principal concern is the application of bivariate quadrat methods to the analysis of urban dispersion.

This work is being conducted under the sponsorship of a grant from the Economic Development Administration of the U.S. Department of Commerce. A complete listing of the Working Papers under this grant appears on the back page of this paper. They deal with regional patterns, development, and policies.

Andrei Rogers
University of California, Berkeley
May 1, 1969

Working Papers on Spatial Pattern Analysis

#86 Andrei Rogers and Norbert G. Gomar, "Statistical Inference in Quadrat Analysis," October, 1968, to be published in Geographical Analysis.

*#91 Shlomo Angel, "Point Patterns in Urban Regions: Two Complementary Tests," January, 1969.

*#93 Andrei Rogers, "Quadrat Analysis of Urban Dispersion," March, 1969.

*#96 Juan Martin, Norbert Gomar and Andrei Rogers, "Some Computer Programs for Quadrat Analysis," May, 1969.

#97 Andrei Rogers and Juan Martin, "Bivariate Quadrat Analysis of the Dispersion of Retail Establishments and Population in Urban Areas," May, 1969

* Out of Print.

1. INTRODUCTION

A recent paper [Rogers (1965)], outlined a statistical procedure for analyzing the spatial clustering of retail establishments in urban areas. There it was concluded that competing shopping goods establishments tend to cluster together in order to share in the increased volume of sales that such aggregation produces. The empirical findings, in other words, supported the marketing principle that [Nelson (1958), p. 58] "a given number of stores dealing in the same merchandise will do more business if they are located adjacent or in proximity to each other than if they are widely scattered."

The empirical findings for convenience goods stores, however, were less clear. Marketing researchers have asserted that such establishments do not profit from cohesion and, therefore, typically do not group together [Nelson (1958), p. 179]. Yet analysis of the data revealed that convenience goods establishments tend to exhibit a more highly clustered pattern than shopping goods stores. This spatial clustering was attributed to the spatial clustering of the residential population. It was argued that because of the unspecialized and recurrent function that they perform, convenience goods stores tend to be drawn toward the consumers they serve and thus assume a distributional pattern which mirrors that of the residential population. Hence, it was concluded, both spatial patterns must be considered simultaneously.

In this paper, we extend the previous efforts in this direction by introducing bivariate clustering models which may be used to account for both the number of convenience goods establishments in a cell and the number of people residing in that same cell. The data used for empirical testing are the spatial patterns of retail establishments, retail employment, and residential population in Ljubljana, Yugoslavia during the year 1966.

2. BIVARIATE QUADRAT MODELS OF SPATIAL DISPERSION

We begin our analysis by considering an urban area that has been gridded into a network of squares and focus on a model which described the clustering of establishments within these cells by means of a stochastic process. In what follows it will be shown how different postulates for this stochastic process will lead to different models of urban dispersion.

2.1 Bivariate Correlated Poisson (BICP) Model

Postulates:

1) Two classes of establishments, type X establishments and type Y establishments, are randomly (i.e., Poisson) distributed in the study area.

2) The number of type X establishments, say x, and the number of type Y establishments, say y, in each cell are correlated and may be described by the probability law P(x,y), called the BICP distribution.

It will be shown that under the above postulates the univariate distributions of type X and type Y establishments are Poisson with parameters a and b, respectively. The derivation of the model proceeds as follows:

Let X and Y be two random variables with the joint BICP distribution (X,Y) such that

$$X = X' + U$$

$$Y = Y' + U$$

and X', Y', U are independent Poisson variables with parameters $a' = a - d$, $b' = b - d$, and d, respectively.

The probability density function of the BICP, $P(X = x, Y = y) = P(x,y)$, is given by:

$$P(x,y) = e^{-(a+b-d)} \sum_{k=0}^{z} \frac{(a-d)^{x-k}\,(b-d)^{y-k}\,d^{k}}{(x-k)!\,(y-k)!\,k!} \tag{1}$$

where a and b are marginal means, d is the covariance and $z = \min(x,y)$.

Except for the fact that the origin $a = b = d = 0$ is excluded, the parameter space is defined by the inequality

$$0 \leq d \leq \min(a,b). \tag{2}$$

The correlation coefficient is

$$\rho = \frac{d}{\sqrt{ab}},$$

and the inequality (2) implies a restriction on the correlation coefficient. First, suppose $\min(a,b)$ is a then

$$0 \leq d \leq a,$$

squaring and dividing by b yields

$$0 \leq \frac{d^2}{ab} \leq \frac{a}{b},$$

or

$$0 \leq \frac{d}{\sqrt{ab}} \leq \sqrt{\frac{a}{b}}.$$

Similarly, if min(a,b) is b then

$$0 \le \frac{d}{\sqrt{ab}} \le \sqrt{\frac{b}{a}} .$$

Hence

$$0 \le \rho \le \min\left[\sqrt{\frac{a}{b}} , \sqrt{\frac{b}{a}}\right] .$$

The probability generating function of the BICP is given by

$$p_{xy}(s,t) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P(x,y)s^i t^j$$

$$= e^{-(a+b-d)} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \sum_{k=0}^{\min(x,y)} \frac{(a-d)^{i-k}(b-d)^{j-k}d^k}{(i-k)!\,(j-k)!\,k!} s^i t^j$$

$$= \exp\{a(s-1)+b(t-1)+d(s-1)(t-1)\}$$

$$= \exp\{(a-d)(s-1)+(b-d)(t-1)+d(st-1)\} \quad (3)$$

The marginal distribution of X and Y are Poisson with parameters a and b respectively, since

$$g_x(s) = p_{xy}(s,1) = e^{-a(1-s)} ,$$

with mean and variance

$$E(x) = VAR(x) = a ,$$

and

$$h_y(t) = p_{xy}(1,t) = e^{-b(1-t)} ,$$

with mean and variance

$$E(Y) = VAR(Y) = b .$$

Since the parameter d is the covariance it follows that two uncorrelated Poisson variables distributed according to (1) are independent. The probability generating function, in this case, is

$$p_{xy}(s,t) \quad \exp a(s-1) + b(t-1) \quad (4)$$

2.2 Generalized Processes

Postulates:

1) The spatial distribution of type X establishments is a clustered one.

2) The number of type X establishments in an X cluster, say x, is a random variable with a probability function G(x).

In order to include the spatial distribution of type B establishments, we also adopt the following postulates:

3) The spatial distribution of type Y establishments is a clustered one.

4) The number of type Y establishments in a Y cluster, say y, is a random variable with a probability function H(y).

To consider the joint spatial pattern of X and Y establishments we include the additional postulate:

5) The number of type X establishments, say x, and the number of type Y establishments, say y, in each cell are correlated and may be described by the probability law P(x,y).

It is easily established that a generalized bivariate distribution has a probability generating function of the form

$$f(s,t) = p[g(s),h(t)] \ . \tag{5}$$

where $p[\ \circ\ ]$, $g(s)$, and $h(t)$ are the probability generating functions of $P(x,y)$, $G(x)$ and $H(y)$, respectively. A series of bivariate models may be derived by assuming arbitrary probability density functions for $P(x,y)$, the distribution defining the number of type X and type Y clusters in each cell, or $G(x)$ and $H(y)$, the distributions defining the number of type X and type Y establishments, respectively, per cluster.

Alternatively, we may reformulate the fundamental model by postulating a clustering process which is univariate in the distribution of clusters, but is bivariate in the number of establishments in each cluster. Such a formulation leads to slightly different results [Holgate (1966)].

2.2.1 The Generalized Bivariate Correlated Poisson (GBCP) Model

For the bivariate clustering model defined above, let: $P(x,y)$ be the Bivariate Correlated Poisson, with probability generating function from (3)

$$p_{xy}(s,t) = \exp\{(a-d)(s-1)+(b-d)(t-1)+d(st-1)\}.$$

Let $G(x)$ be an univariate Poisson distribution with parameter m and probability generating function

$$g_x(s) = e^{m(s-1)} ,$$

and let $H(y)$ also be a univariate Poisson distribution with parameter n and probability generating function

$$h_y(t) = e^{n(t-1)} .$$

Then the probability generating function for the generalized process in (5) becomes

$$f_{xy}(s,t) = \exp\{(a-d)(e^{m(s-1)}-1)+(b-d)(e^{n(t-1)}-1)+d(e^{m(s-1)+n(t-1)}-1\} \qquad (6)$$

This is the probability generating function of a bivariate generalization of Neyman's Univariate Type A distribution. Differentiation of (6) x+1 times with respect to a and y times with respect to b leads, on setting a = b = 0, to the recurrence relation:

$$F(x+1,y) = \frac{dme^{-(m+n)}}{x+1} \sum_{i=0}^{x} \sum_{j=0}^{y} \frac{m^i n^j}{i!j!} F(x-i,y-j)$$

$$+ \frac{(a-d)me^{-m}}{x+1} \sum_{i=0}^{x} \frac{m^i}{i!} F(x-i,y)$$

with the initial condition

$$F(0,0) = f_{xy}(0,0) = \exp\{(a-d)(e^{-m}-1)+(b-d)(e^{-n}-1)+d(e^{-(m+n)}-1)\} ,$$

where a and b are the means of the G(x) and H(y) distributions and d its covariance

$$0 \le d \le \min[a,b] .$$

The correlation between the number of type X clusters and the number of type Y clusters in a cell, is given by

$$\rho_1 = \frac{d}{\sqrt{ab}} .$$

The marginal distributions are univariate Neyman Type A distributions

$$g_x(s) = f_{xy}(s,1) = e^{a(e^{m(s-1)}-1)} ,$$

with mean and variance

$$E(x) = am$$

$$VAR(x) = am(m+1)$$

and

$$h_y(t) = f_{xy}(1,t) = e^{b(e^{n(t-1)}-1)}$$

with mean and variance

$$E(x) = bn$$

$$VAR(y) = bn(n+1)$$

and covariance

$$COV(x,y) = dmn .$$

Finally, the correlation between the number of establishments of type X and the number of type Y is given by

$$\rho = \frac{wmn}{\sqrt{am(m+1)bn(n+1)}} = \frac{mn}{\sqrt{mn(m+1)(n+1)}}\,\rho_1$$

2.3 Compound Process

Postulates:

1) The spatial distribution of type X establishments is a clustered one.

2) The number of type X establishments in a cell, say x, is a random variable with probability function $G(x|\lambda)$. It is assumed that for any randomly selected cell the parameter λ is a particular value of a positive random variable with probability function $L(\lambda)$.

3) The spatial distribution of type Y establishments is a clustered one.

4) The number of type Y establishments in a cell, say y, is a random variable with probability function $H(y|\lambda)$.

5) The number of type X and type Y establishments in each cell may or may not be correlated, and may be described by the conditional probability law $P(x,y|\lambda)$.

It can readily be shown that a compound bivariate function has an unconditional probability density function of the form:

$$F(x,y) = \int_0^\infty P(x,y|\lambda)\, L(\lambda)\, d\lambda \qquad (7)$$

and once again several models may be derived by assuming arbitrary probability functions for $L(\lambda)$, $G(x)$, $H(y)$ and the presence or absence of correlation between both types of establishments.

2.3.1 The Compound Bivariate Uncorrelated Poisson (CBUP) Model

For the compound process defined above, let $L(\lambda)$ be the gamma distribution:

$$L(\lambda) = \frac{\beta^{\alpha}\lambda^{\alpha-1}e^{-\beta\lambda}}{\Gamma(\alpha)} \quad ; \lambda>0, \beta>0, \alpha>0$$

$$= 0 \qquad\qquad , \lambda \leq 0$$

let $G(x|\lambda)$ be the Poisson function with parameter $a\lambda$

$$G(x|\lambda) = \frac{e^{-\lambda a}(\lambda a)^{x}}{x!} \quad ,$$

let $H(y|\lambda)$ be the Poisson function with parameter $b\lambda$

$$H(y|\lambda) = \frac{e^{-\lambda b}(\lambda b)^{y}}{y!} \quad ,$$

and assume that $G(x|\lambda)$ and $H(y|\lambda)$ are stochastically independent. Then, given $\lambda = \lambda_o$, their conditional joint probability generating function is:

$$f_{xy}(s,t|\lambda_o) = \exp\{\lambda_o[a(s-1)+b(t-1)]\} \ .$$

Neyman and Bates (1952) have shown that the unconditional probability generating function is of the form

$$f_{xy}(s,t) = [1+ \tfrac{a}{\beta}(1-s) + \tfrac{b}{\beta}(1-t)]^{-\alpha} \quad , \qquad (8)$$

and they have termed it, in its most general form, the multivariate negative binomial distribution. Measuring b in units of a

$$f_{xy}(s,t) = \frac{\beta^{\alpha}}{[\beta + (1-s) + A(1-t)]^{\alpha}} \quad , \qquad (9)$$

where

$$A = \frac{b}{a} = \frac{b\lambda}{a\lambda} = \frac{E(y)}{E(x)} \quad .$$

By expanding (9) in powers of s and t, we obtain as the coefficient of $s^x t^y$

$$F(x,y) = \frac{\Gamma(\alpha+x+y)}{x!\,y!\,\Gamma(\alpha)} \beta^{\alpha} A^{y} (\beta+A+1)^{-(\alpha+x+y)} \quad .$$

Neyman and Bates have shown that whatever the distribution function $L(\lambda)$, provided it possesses the first two moments, the correlation coefficient (ρ) between x and y is given by

$$\rho = \{(1 + \frac{E(\lambda)}{a\ Var(\lambda)})^{-1} \ (1 + \frac{E(\lambda)}{b\ Var(\lambda)})^{-1}\}^{1/2}$$

In particular if $L(\lambda)$ is the Gamma distribution

$$E(\lambda) = \alpha$$

$$VAR(\lambda) = \alpha^2/k \ ,$$

and if follows that

$$\rho = [(1 + \frac{k}{a\alpha}) \ (1 + \frac{k}{\beta\alpha})^{-1/2} \quad .$$

The marginal distributions of x and y and the distribution of the sum $z = x+y$ are all univariate negative binomial. Thus from (9)

$$g_x(s) = f_{xy}(s,1) = [1 + \frac{1}{\beta} - \frac{1}{\beta} s]^{-\alpha} \ ,$$

with parameters

$$k = \alpha$$

$$m = \frac{\alpha}{\beta} \ ,$$

and mean and variance

$$E(x) = \frac{\alpha}{\beta}$$

$$VAR(x) = \frac{\alpha}{\beta} (1 + \frac{1}{\beta})$$

Similarly

$$h_y(t) = f_{xy}(1,t) = [1 + \frac{A}{\beta} - \frac{A}{\beta} t]^{-\alpha}$$

with parameters

$$k = \alpha$$

$$m = A \frac{\alpha}{\beta}$$

and mean and variance

$$E(Y) = A \frac{\alpha}{\beta}$$

$$VAR(Y) = A \frac{\alpha}{\beta} [1 + \frac{A}{\beta}] .$$

Finally

$$p_z(\gamma) = f_{xy}(\gamma,\gamma) = [1 + \frac{1+A}{\beta} - \frac{1+A}{\beta} \gamma]^{-\alpha} ,$$

with parameters

$$k = \alpha$$

$$m = (1+A) \frac{\alpha}{\beta} ,$$

and mean and variance

$$E(z) = (1+A) \frac{\alpha}{\beta}$$

$$VAR(z) = (1+A) \frac{\alpha}{\beta} [1 + \frac{1+A}{\beta}] .$$

2.3.2 The Compound Bivariate Correlated Poisson (CBCP) Model

For the compound stochastic process, let $L(\lambda)$ be the Gamma distribution:

$$L(\lambda) = \frac{\beta^{\alpha}\lambda^{\alpha-1}e^{-\beta\lambda}}{\Gamma(\lambda)} \qquad \lambda>0,\beta>0,\alpha>0$$

$$= 0 \qquad , \lambda\leq 0$$

let $G(x|\lambda)$ be the Poisson function with parameter $a\lambda$

$$G(x|\lambda) = \frac{e^{-\lambda a}(\lambda a)^x}{x!} ,$$

let $H(y|\lambda)$ be the Poisson function with parameter $b\lambda$

$$H(y|\lambda) = \frac{e^{-\lambda b}(\lambda b)^y}{y!}$$

as in the CBUP model. Now assume that there exists a positive correlation between the number of type X establishments given by $G(x|\lambda)$ and the number of type Y establishments given by $H(y|\lambda)$, in each cell. Then, given $\lambda = \lambda_0$, their conditional joint probability generating function is

$$f_{xy}(s,t|\lambda_0) = \exp\{\lambda_0[a(s-1) + b(t-1) + d(s-1)(t-1)]\}$$

$$\lambda_0 > 0$$

$$0 \leq d \leq \min(a,b) \quad ,$$

where $\lambda_0 a$ and $\lambda_0 b$ are the means of the marginal Poisson distributions and $\lambda_0 d$ is the covariance. Then, as Edwards and Gurland (1961) have shown, the unconditional distribution of type X and type Y establishments in a cell has probability generating function:

$$f_{xy}(s,t) = [1+\frac{a-d}{\beta}(1-s) + \frac{(b-d)}{\beta}(1-t) + \frac{d}{\beta}(1-st)]^{-\alpha} \qquad (10)$$

$$= (A + B_1 s + B_2 t + B_{12} st)^{-\alpha} \qquad (11)$$

where

$$B_1 = -\frac{1}{\beta}(a-d) < 0$$

$$B_2 = -\frac{1}{\beta}(b-d) < 0$$

$$B_{12} = -\frac{d}{\beta} < 0$$

and

$$A = 1 + \frac{a}{\beta} + \frac{b}{\beta} - \frac{d}{\beta} = (1 - B_1 - B_2 - B_{12}) > 0 \ .$$

Edwards and Gurland have called this distribution the Compound Bivariate Correlated Poisson and they have shown that the CBUP is a special case of it when the correlation term d is equal to zero. Notice that when d = 0 then (10) reduces to the probability generating function of the CBUP in (8).

By expanding (11) in powers of s and t, we obtain as the coefficient of $s^x t^y$

$$F(x,y) = A^{-(\alpha+x)} B_1^{(x-y)} B_{12}^{y} \sum_{i=0}^{y} \frac{(-1)^{x+i}\Gamma(\alpha+x+i)}{(x-y+i)!(y-i)!i!\Gamma(\alpha)} \left[\frac{B_1B_2}{AB_{12}}\right]^i$$

$x \geq y$ and $x, y, i = i, 2, 3, \ldots$,

and when $y>x$, $F(x,y)$ may be obtained by reversing the positions of x and y, B_1 and B_2.

Several recurrence relations can be obtained for the joint probabilities of the CBCP distribution. Two of the more useful are

$$F(x,y) = (-A)^{-x} B_1^{x-y} B_{12}^{y} \sum_{i=0}^{y} \frac{\Gamma(\alpha+x+i)}{(x-y+i)!(y-i)!\Gamma(\alpha+i)} \left[\frac{B_1}{B_{12}}\right]^i F(0,i)$$

$$F(r+t,r) = \left[\frac{B_1}{B_2}\right]^t F(r,r+t) \qquad r \geq s \text{ and } x, y, r, t, i = 1, 2, 3, \ldots$$

with the initial condition from the density function above

$F(0,0) = A^{-\alpha}$.

The correlation between the number of type X and type Y establishments in a cell is given by

$$\rho = \frac{B_{12}\,(B_1+B_2+B_{12}-1) + B_1B_2}{\sqrt{(B_1+B_{12})(1-(B_1+B_{12}))(B_2+B_{12})(1-(B_2+B_{12}))}} \ .$$

As in the CBUP, the marginal distributions of x and y are negative binomial. However, the distribution of the sum, $z = x+y$, is not a univariate negative binomial distribution unless d is equal to zero. From (11)

$$g_x(s) = f_{xy}(s,1) = [(A+B_2) + (B_1+B_{12})s]^{-\alpha}$$
$$= [1-(B_1+B_{12}) + (B_1+B_{12})s]^{-\alpha} \quad ,$$

with parameters

$$k = \alpha$$

$$m = -\alpha(B_1+B_{12}) \quad ,$$

and mean and variance

$$E(x) = -\alpha(B_1+B_{12})$$

$$VAR(x) = -\alpha(B_1+B_{12})(1-(B_1+B_{12})) \quad .$$

Similarly

$$h_y(t) = f_{xy}(1,t) = [1-(B_2+B_{12}) + (B_2+B_{12})t]^{-\alpha} \quad ,$$

with parameters

$$k = \alpha$$

$$m = -\alpha(B_2+B_{12}) \quad ,$$

and mean and variance

$$E(x) = -\alpha(B_2+B_{12})$$

$$VAR(x) = -\alpha(B_2+B_{12})(1-(B_2+B_{12})) \quad .$$

2.4 Contagious Processes

Postulates:

1) Initially (i.e., at time t = 0), none of the cells contains any type X establishments. The number of type X establishments in a given cell, say x, at time t is given by the probability law $G(x,t)$.

2) The probability that a type X establishment locates in a cell, $P(x,t)$, is a function of the number of type X establishments already in the cell.

3) The number of type Y establishments, say y, at time t, for a given number of type X establishments, say x, is given by the conditional probability function $H(y,t|x)$.

4) The probability that a type Y establishment locates in a cell, $Q(x,y,t)$, is a function of both the number of type X and type Y establishments already in the cell.

In another paper [Rogers (1969)], it has been shown that in the univariate case the above postulates lead to the following single differential equation

$$dg_x(u) = (u-1)R_x(u)dt \tag{12}$$

where

$$R_x(u) = \sum_{x=0}^{\infty} u^x P(x,t)G(x,t) \quad .$$

2.4.1 The Contagious Bivariate Negative Binomial (CBNB) Model

Let the probability $P(x,t)$ that a type X establishment locates in a cell be given by the linear function

$$P(x,t_1) = \beta + \gamma x \qquad \beta>0, \gamma>0 \ ,$$

where x is the number of type X establishments already in the cell. Then (12) becomes

$$\frac{\partial}{\partial t} \ g_x(u) = (u-1) \ [\beta g_x(u) + \gamma u \frac{\partial}{\partial u} \ g_x(u)] \quad ,$$

the solution of which is

$$g_x(u) = [e^{\gamma t}1 - u(e^{\gamma t}1 - 1)]^{-\beta/\gamma}$$

Thus the distribution of type X establishments is negative binomial with probability density function

$$G(x,t_1) = e^{-\beta t_1} \frac{\Gamma(\beta/\gamma+x)}{x!\Gamma(\beta/\gamma)} (1-e^{-\gamma t_1})^x$$

Now let the probability that a type Y establishment locates in a cell, $Q(y,t_2|x)$ be given by the linear function

$$Q(y,t_2|x) = \beta + \gamma(x+y) \quad \beta>0,\gamma>0 \quad ,$$

where x and y are the number of type X and type Y establishments, respectively, already in the cell. In this case the solution of (12) becomes

$$h_y(v|x) = [e^{\gamma t_2} - v(e^{\gamma t_2} - 1)]^{\beta/\gamma-x} \quad ,$$

whence

$$H(y,t_2|x) = e^{-(\beta+\gamma x)t_2} \frac{\Gamma(\beta/\gamma+x+y)}{\Gamma(\beta/\gamma+x)y!} (1-e^{-\gamma t_2})^y \quad ,$$

and the joint distribution of type X and type Y establishments, measuring t_2 in units of t_1, as before, is given by

$$F(x,y) = G(x) \cdot H(y|x)$$

$$= e^{-\beta(1+A)} \frac{\Gamma(\beta/\gamma+x+y)}{\Gamma(\beta/\gamma)x!y!} (e^{-\gamma A}-e^{-\gamma(1+A)})^x(1-e^{-\gamma A})^y$$

where $t_1 = 1$ and $t_2 = A = t_2/t_1 = \frac{E(Y)}{E(X)}$,

with probability generating function

$$f_{xy}(u,v) = [e^{\gamma(1+A)}-u(e^{\gamma}-1) -ve^{\gamma} (e^{\gamma A}-1)]^{-\beta/\gamma}$$

The marginal distribution of x is negative binomial with probability generating function

$$g_x(u) = f_{xy}(u,1) = [e^{\gamma} -u(e^{\gamma} -1)]^{-\beta/\gamma}$$

Setting

$$p = e^{\gamma}-1$$

$$k = \beta/\gamma$$

$$g_x(u) = [1 + p - pu]^{-k}$$

we have the parameters

$$k = \beta/\gamma$$

$$m = \frac{\beta}{\gamma}\,(e^{\gamma}-1)$$

and the mean and variance

$$E(x) = \frac{\beta}{\gamma}\,(e^{\gamma}-1)$$

$$VAR(x) - \frac{\beta}{\gamma}\,(e^{\gamma}-1)e^{\gamma} \quad .$$

The marginal distribution of y has probability generating function

$$h_y(v) = f_{xy}(1,v) = [e^{-\gamma} + (1-v)(e^{\gamma A}-1)]^{-\beta/\gamma} \quad .$$

thus setting

$$p = e^{\gamma A}-1$$

$$k = -\beta/\gamma$$

$$h_y(v) = [e^{-\gamma}+p - pv]^{-k}$$

we obtain the marginal conditional mean and variance of y:

$$E(y|x) = (\frac{\beta}{\gamma} + x)\ (e^{\gamma(1+A)}-1)$$

$$VAR(y|x) = (\frac{\beta}{\gamma} + x)\ (e^{\gamma(1+A)}-1)\ e^{\gamma} \quad .$$

Finally, the correlation coefficient between the number of type X and type Y establishments is

$$\rho = [(1 + \frac{1}{e^{\gamma}-1})\ (1 + \frac{1}{e^{\gamma}(e^{\gamma A}-1)})]^{-1/2}$$

3. PARAMETER ESTIMATION

In order to carry out statistical tests on the goodness-of-fit of an empirical distribution to a hypothesized theoretical model, it is first necessary to obtain estimates of that theoretical model's parameters. In this paper, the method of moments is utilized to derive estimates of the parameters for each of the above described theoretical models. Maximum likelihood estimation is only carried out for the CBUP model.

3.1 The Bivariate Correlated Poisson

For this model we have to estimate three independent parameters: a, b, and d. The first two parameters can be estimated from the marginal distributions of x and y. Thus

$$P_x(s) = e^{-a(1-s)} ,$$

hence

$$m_1 = a \qquad ,$$

and we obtain the moment estimator:

$$\hat{a} = E(X)$$

Similarly

$$P_y(t) = e^{-b(1-t)} .$$

hence

$$m_1 = b$$

and we have the moment estimator:

$$\hat{b} = E(Y) \quad .$$

The sample covariance cov(x,y) is a consistent estimator of d. Thus

$$\hat{d} = \text{cov}(x,y)$$

Numerical computations of the efficiency of moments [Holgate (1964)] show that although the method of moments is fully efficient for $d = 0$, its efficiency falls off with increasing correlation. Maximum likelihood estimation is recommended for such cases. Estimation of the parameters by maximum likelihood yields $\hat{a} = E(x)$, $\hat{b} = E(y)$, but only a polynomial of random degree in d [Teicher (1954)].

3.2 The Generalized Bivariate Correlated Poisson

For this distribution we have to estimate five independent parameters: a, m, b, n, and d. Moment estimators are

$$\hat{m} = \frac{VAR(x)}{E(x)} - 1 \qquad \hat{n} = \frac{VAR(y)}{E(y)} - 1$$

$$\hat{a} = \frac{E(x)}{\hat{m}} \qquad \hat{b} = \frac{E(x)}{\hat{n}}$$

and

$$\hat{d} = \frac{\text{cov}(x,y)}{\hat{m}\ \hat{n}}$$

The only restriction is that given by (2), namely that

$$0 \leq \hat{d} \leq \min(\hat{a},\hat{b}) \quad .$$

3.3 The Compound Bivariate Uncorrelated Poisson

Three independent parameters need to be estimated: A, β and α.

3.3.1 The Moment Estimator

Moment estimators are derived simply as

$$A = \frac{E(y)}{E(x)}$$

$$\hat{\alpha} = \frac{[E(z)]^2}{VAR(z) - E(z)}$$

$$\hat{\beta} = \frac{(1 + A)\ E(z)}{VAR(z) - E(z)}$$

where E(z) and VAR(z) are the mean and variance of the sum of X and Y.

3.3.2 The Maximum Likelihood Estimator

Let n be the number of independent observations on the pair of values (X,Y). The letter $n_{x,y}$ will then denote the random variable representing the number of pairs [(X = x), (Y = y)]. The joint frequency function of all the $n_{x,y}$ is represented by the product

$$J = C \prod_{x=0}^{\infty} \prod_{y=0}^{\infty} F(x,y)^{n_{x,y}} \quad , \tag{13}$$

where C is a factor independent of the parameters and

$$\sum_{x=0}^{\infty} \sum_{y=0}^{\infty} n_{x,y} = n \quad .$$

Recalling that

$$F(x,y) = \frac{\Gamma(\alpha+y+x)}{x!\,y!\,\Gamma(\alpha)}\ \beta^{\alpha}\ A^{y}\,(\beta+A+1)^{-(\alpha+x+y)} \tag{14}$$

and substituting it into (13), taking Neperian logarithms and dividing by n yields

$$\frac{1}{n}\log J = C_1 + \alpha\log\beta - (\alpha+\overline{X}+\overline{Y})\log(\beta+A+1) + \overline{Y}\log A + \sum_{t=0}^{\infty}\left(1-\sum_{r=0}^{t} q_r\right)\log(\alpha+t) \tag{15}$$

where $\overline{X}$ and $\overline{Y}$ are marginal means and q_r represents the relative frequency of pairs (X,Y) which have their sum X + Y = r. The maximum likelihood equations are obtained by differentiating (15) with respect to α, β, and A, and by equating the resulting derivatives to zero. Thus

$$\log \frac{\beta}{\hat{\beta}+\hat{A}+1} + \sum_{t=0}^{\infty} \frac{1-\sum_{r=0}^{t} q_r}{\hat{\alpha}+t} = 0 \tag{16}$$

$$\frac{\hat{\alpha}}{\beta} - \frac{\hat{\alpha}+\overline{X}+\overline{Y}}{\hat{\beta}+\hat{A}+1} = 0 \tag{17}$$

$$\frac{\overline{Y}}{\hat{A}} - \frac{\hat{\alpha}+\overline{X}+\overline{Y}}{\hat{\beta}+\hat{A}+1} = 0 \tag{18}$$

Equations (17) and (18) imply that

$$\hat{\alpha} = \overline{X}\hat{\beta} \tag{19}$$

$$\hat{A} = \frac{\overline{Y}}{\overline{X}} \tag{20}$$

Then (16) becomes

$$Q(\alpha) = \log\left(1 + \frac{\overline{X}+\overline{Y}}{\hat{\alpha}}\right) - \sum_{t=0}^{\infty} \frac{1-\sum_{r=0}^{t} q_r}{\hat{\alpha}+t} = 0. \tag{21}$$

Equation (21) must be solved iteratively using some standard method such as the Newton-Raphson which begins with an initial approximation of α, say α_0, and obtains an improved approximation, say α_1, by means of the expression $\alpha_1 = \alpha_0 - \frac{Q(\alpha_0)}{Q'(\alpha_0)}$,

where $Q'(\alpha)$ is the derivative of $Q(\alpha)$.

The first trial value, say α_0, may be conveniently obtained as follows. Substituting $\hat{\alpha}$, $\hat{\beta}$, and $\hat{A}$ in (14) for X = 0 and Y = 0, should give a result comparable to q_0

$$F(0,0) = \left(\frac{\hat{\beta}}{\hat{\beta}+\hat{A}+1}\right)^{\hat{\alpha}} = q_0 \quad ,$$

and using (19) and (20), we find that

$$\left(\frac{\alpha_0}{\alpha_0+\bar{X}+\bar{Y}}\right)^{\alpha_0} = q_0 \quad ,$$

which is equivalent to

$$\log(1+z) + \frac{\log q_0}{\bar{X} + \bar{Y}} z = 0 \ , \qquad (22)$$

where

$$z = \frac{\bar{X} + \bar{Y}}{\alpha_0} \quad .$$

In order to obtain α_0 we must solve (22) iteratively. Once α_0 is obtained, we may use it as a first approximation to α in (21) to get $\hat{\alpha}$, and then using (19) and (20) obtain A and $\hat{\beta}$.

3.4 The Compound Bivariate Correlated Poisson

In this model there are only three independent parameters to be estimated: B_2, B_{12}, and α, since

$$\hat{B}_1 = \frac{E(x)}{E(y)} \hat{B}_2 + \left(\frac{E(x)}{E(y)} - 1\right) \hat{B}_{12} \ .$$

One combination of moments that might be used to provide these estimates is E(x), E(y), VAR(y), and COV(x,y). Estimates of the parameters may then be obtained from the relations

$$\hat{\alpha} = \frac{E(x)}{k}$$

$$\hat{B}_{12} = \left[\frac{E(y)}{E(x)}\right] k^2 - \frac{COV(x,y)}{\hat{\alpha}}$$

$$\hat{B}_2 = -k - \hat{B}_{12} \quad ,$$

where

$$k = \frac{VAR(y) - E(x)\left[\frac{E(y)}{E(x)}\right]}{E(x)\left[\frac{E(y)}{E(x)}\right]^2}$$

3.5 The Contagious Bivariate Correlated Poisson

Three parameters have to be estimated in this model: A, β, and γ. The mean and variance of the sum $Z = X+Y$ may be used to derive these estimates, as well as the marginal means of X and Y. Thus

$$A = \frac{E(y)}{E(x)}$$

$$\gamma = \frac{\log VAR(Z) - \log E(Z)}{1 + A}$$

$$\beta = \frac{E(Z)\gamma}{e^{\gamma(1+A)}-1} \quad .$$

4. SOME EXPERIMENTAL RESULTS

The above models were fitted to the observed bivariate distribution of food stores and population in Ljubljana, Yugoslavia presented in Table 1. Table 2 presents the resulting sample statistics, parameter estimates and chi square statistics. Tables 3, 4, 5, and 6 present a few of the expected distributions that are generated by these sample statistics and parameter estimates.

TABLE 1 -- OBSERVED BIVARIATE DISTRIBUTION OF FOOD STORES AND POPULATION IN LJUBLJANA, YUGOSLAVIA

Population y 500's \ Food Stores x	0	1	2	3	4	5	6	7	8	9+	Marginal Total y
0	35	4	5								44
1	10	6	3	2	3	1				1	26
2	3	1	1	3	1	1	2			2	14
3		2	1	1	2	1				3	10
4	1		1		1	2				1	6
Marginal Total x	49	13	11	6	7	5	2	0	0	7	100

TABLE 21-- SAMPLE STATISTICS, PARAMETER ESTIMATES, AND CHI SQUARE STATISTICS

I. Sample Statistics:

E(x) =	2.0100
E(y) =	1.1100
Var(x) =	11.5454
Var(y) =	1.7353
Cov(xy) =	2.4189
R=Corr(xy) =	.5404

II. Parameter Estimates and Chi Square Statistics:

A. Bivariate Correlated Poisson:

Estimate		Chi square		d.f.	P.05
A =	2.0100	Along rows:	72.53	10d.f.	18.30
B =	1.1100	cols:	83.98	10d.f.	18.30
D =	1.1000	diag:	82.52	9d.f.	16.90

B. Generalized Bivariate Correlated Poisson (Heyman's Type A):

Estimate		Chi square		d.f.	P.05
EME =	4.7440	Along rows:	81.66	11d.f.	19.70
ENE =	.5633	cols:	144.29	13d.f.	22.40
U =	.4237	diag:	22.79	12d.f.	21.00
V =	1.9706				
W =	.4237				

C. Compound Bivariate Correlated Poisson (Negative Binomial):

Estimate		Chi square		d.f.	P.05
K =	.4237	Along rows:	52.49	10d.f.	18.30
B_1 =	1.4368	cols:	30.68	10d.f.	18.30
B_2 =	2.6098	diag:	33.91	9d.f.	16.90
B_{12} =	-.0100				
A =	5.0566				

D. Compound Bivariate Uncorrelated Poisson (Negative Binomial):

(i) Moments:

Estimate		Chi square		d.f.	P.05
ALFA =	.6469	Along rows:	29.23	16d.f.	26.30
BETA =	.3219	* cols:	12.97	13d.f.	22.40
A =	.5522	* diag:	21.33	14d.f.	23.70

(ii) Max. Likelihood:

Estimate		Chi square		d.f.	P.05
ALFA =	.5864	Along rows:	30.05	15	25.00
BETA =	.2917	* cols:	12.50	13	22.40
A =	.5522	* diag:	18.64	14	23.70

E. Contagious Bivariate Negative Binomial

Estimate		Chi square		d.f.	P.05
GAMMA =	1.1350	Along rows:	27.10	15	25.00
BETA =	.7343	* cols:	14.48	14	23.70
A =	.5522	* diag:	18.15	14	23.70

---

* Null hypothesis not rejected at the 5 percent level of significance.

TABLE 3 -- EXPECTED BIVARIATE DISTRIBUTION OF FOOD STORES AND POPULATION IN LJUBLJANA, YUGOSLAVIA: BIVARIATE CORRELATED POISSON MODEL

Population y 500's \ Food Stores x	0	1	2	3	4	5	6	7	8	9+	Marginal Total y
0	13.27	12.07	5.49	1.67	.38	.07	.01				32.96
1	.13	14.71	13.33	6.06	1.84	.42	.08	.01			36.58
2		.15	8.16	7.36	3.34	1.01	.23	.04	.01		20.30
3			.08	3.02	2.71	1.23	.37	.08	.02		7.51
4				.03	.84	.75	.34	.10	.02	.57	2.65
Marginal Total X	13.40	26.93	27.07	18.13	9.10	3.48	1.03	.24	.05	.57	Total 100.00

TABLE 4 -- EXPECTED BIVARIATE DISTRIBUTION OF FOOD STORES AND POPULATION IN LJUBLJANA, YUGOSLAVIA: GENERALIZED BIVARIATE CORRELATED POISSON MODEL

Population y 500's / Food Stores x	0	1	2	3	4	5	6	7	8	9+	Marginal Total y
0	33.70	.34	.80	1.27	1.52	1.46	1.19	.86	.57	1.09	42.79
1	16.76	.36	.85	1.35	1.62	1.58	1.31	.97	.68	1.57	27.02
2	8.89	.24	.56	.90	1.08	1.06	.89	.68	.50	1.34	16.13
3	3.92	.13	.30	.48	.58	.57	.49	.38	.29	.88	8.02
4 +	2.14	.08	.20	.31	.38	.38	.33	.27	.21	1.72	6.03
Marginal TOTAL X	65.41	1.14	2.70	4.30	5,18	5.05	4.22	3.16	2.24	6.60	Total 100.00

5. REFERENCES

Bates, G. E., and Neyman, J. (1952). Contributions to the theory of accident proneness. _University of California Publications in Statistics_, 1, 215-276.

Edwards, C. B., and Gurland, J. (1961). A class of distributions applicable to accidents. _Journal of the American Statistical Association_, 56, 503-517.

Holgate, P. (1966). Bivariate generalizations of Neyman's Type A distribution. _Biometrika_, 53, 241-244.

Holgate, P. (1964). Estimation for the bivariate Poisson distribution. _Biometrika_, 51, 241-245.

Nelson, R. L. (1958). _The Selection of Retail Locations_ (New York: F. W. Dodge Corp.).

Rogers, A. (1965). A stochastic analysis of the spatial clustering of retail establishments. _Journal of the American Statistical Association_, 60, 1094-1103.

Rogers, A. (1969). Quadrat analysis of urban dispersion. _Environment_, forthcoming.

Teicher, H. (1954). On the multivariate Poisson distribution. _Skandinavisk Aktuarietidskrift_, 37, 1-9.

Institute of Urban and Regional Development
Working Papers Sponsored by the
Economic Development Administration

*WP-65 Ronald Choy, "A Review of the Role of Demographic Variables in Macro-Econometric Models," August, 1967.

*WP-66 Andrei Rogers, "The Aggregation Problem in Demography," October, 1967.

*WP-67 Michael Teitz, "Toward a Theory of Urban Public Facility Location," November, 1967.

*WP-70 Michael Teitz, "Locational Strategies for Competitive Systems," December, 1967.

*WP-72 William Alonso, "The Quality of Data and the Choice and Design of Predictive Models," December, 1967.

WP-74 William Alonso, "Industrial Location and Regional Policy in Economic Development," February, 1968.

WP-78 William Alonso, "Equity and Its Relation to Efficiency in Urbanization," July, 1968.

*WP-81 Andrei Rogers and Susan McDougall, "An Analysis of Population Growth and Change in Slovenia and the Rest of Yugoslavia," June, 1968.

*WP-84 Andrei Rogers, "Activity Allocation Models in Transportation Planning," September, 1968.

WP-86 Andrei Rogers and Norbert G. Gomar, "Statistical Inference in Quadrat Analysis," October, 1968.

WP-87 William Alonso, "Aspects of Regional Planning and Theory in the United States," October, 1968.

*WP-89 David Darwent, "Growth Pole and Growth Center Concepts: A review, Evaluation and Bibliography," October, 1968.

WP-90 William Alonso, "Beyond the Inter-Disciplinary Approach to Planning," November, 1968.

*WP-91 Shlomo Angel, "Point Patterns in Urban Regions: Two Complementary Tests," January, 1969.

*WP-92 Andrei Rogers and Susan McDougall Choy, "Multisectoral Models of Regional Demographic and Economic Growth," February, 1969.

*WP-93 Andrei Rogers, "Quadrat Analysis of Urban Dispersion," March, 1969.

*WP-96 Juan Martin, Norbert Gomar, and Andrei Rogers, "Some Computer Programs for Quadrat Analysis," May, 1969.

WP-97 Andrei Rogers and Juan Martin, "Bivariate Quadrat Analysis of the Dispersion of Retail Establishments and Population in Urban Areas," May, 1969.

*Out of Print.