APPLICATION OF DISCRETE NORMAL DISTRIBUTION
FOR DYNAMIC RURAL RETAIL SECTOR ANALYSIS:
PRELIMINARY RESULTS

Thomas R. Harris
J. Scott Shonkwiler

and

Yuanfang Lin

Thomas R. Harris is a Professor in the Department of Applied Economics and Statistics and Director of the University Center for Economic Development at the University of Nevada, Reno.

J. Scott Shonkwiler is a Professor in the Department of Applied Economics and Statistics at the University of Nevada, Reno.

Yuanfang Lin is a Former Research Graduate Student in the Department of Applied Economics and Statistics at the University of Nevada, Reno.

APPLICATION OF DISCRETE NORMAL DISTRIBUTION
FOR DYNAMIC RURAL RETAIL SECTOR ANALYSIS:
PRELIMINARY RESULTS

Abstract

This paper expands on previous count data models for estimating demand thresholds for rural retail sector counts. A discrete normal distribution is used for dynamic analysis of rural retail establishment counts.
APPLICATION OF DISCRETE NORMAL DISTRIBUTION FOR DYNAMIC RURAL RETAIL SECTOR ANALYSIS: PRELIMINARY RESULTS

INTRODUCTION

Traditionally, rural economic development has concentrated on the recruitment and attraction of export oriented goods-producing industries. Industrial recruitment yields industries that are primarily export-oriented and provide a base for existing local economic sectors while generating input demands for further economic development. Importantly for local economic development professionals, attraction of a goods-producing industry, such as manufacturing is highly visible. The direct employment and income effects of the relocated industry are measurable and the local community economic development team usually reaps abundant media coverage.

Nonetheless, industrial recruitment programs prove to be costly, risky and often yield little payoff. Rural communities are often unsuccessful at industrial recruitment because these communities have very limited resources (Hansen 1970). In order to attract goods-producing industries, rural communities with meager resources often grant tax concessions to new or relocating firms thereby eliminating opportunities for fiscal gain (Kieschnick 1981, Shaffer 1989). Usually the outcome of this type of industrial recruitment is that the local tax burden of the resident populace in the local community increases because increased community services for the new industry are incurred without an expanding tax base due to the tax moratoria (Tweeten and Brinkman 1976). Moreover, firms that are willing to relocate because of incentives and tax abatements are also likely to leave the community if other communities offer
better inducements. Results of recent surveys (Smith and Fox 1990; McNamara and Kriesel 1993) continue to show that planning commissions still emphasize the recruit of export oriented or goods-producing industries, while the pursuit of alternative economic development strategies, such as local services and retail sector development, are largely overlooked and often neglected.

Questions regarding the development and expansion of rural commercial sectors may be addressed by the economic development strategy of import substitution. Import substitution seeks to replace goods and services imported from outside the area with local sources of supply (Shaffer 1989). Import substitution strategies strengthen linkages within the local economy because expenditures remain inside the local economy instead of being lost to imports. Also, keeping earned surplus within the local economy enhances local employment and incomes (Smith 1994). For current and future time periods, local economic development strategies must give balanced emphasis to the formulation of import substitution strategies as well as relocation of goods-producing industries.

A commercial sector market analysis tool commonly used to estimate rural commercial sector activity is demand threshold analysis. The demand threshold is defined as the minimum market size required to support a particular good or service and still yield an acceptable rate of return for the business owner (Berry and Garrison 1958a, 1958b; Parr and Denike 1970; Salyards and Leitner 1981; King 1984). The concept is based on the internal economy of the firm and the characteristics of consumer demand. As dictated by central place theory, the foundation for threshold analysis, thresholds are not absolute but vary by good and service. Demand thresholds are usually measured in terms of population required to support one or more firms of a certain type.
Empirical estimates of market thresholds are numerous (Berry and Garrison 1958a, 1958b; Foust and Pickett 1974; Murray and Harris 1978; Salyards and Leitner 1981). However all of these past studies employed ordinary least squares procedures and truncated data sets to estimate threshold levels for rural retail establishments. Studies by Harris et al. (1996), Harris and Shonkwiler (1993) and Wensley and Stabler (1998) have introduced use of count-data techniques when data is truncated.

Usual count data models, such as the Poisson distribution can only cover zero and positive integer values. For dynamic analysis of count changes, counts can be positive, zero and negative. Using procedures outlined by Kemp (1997), a discrete normal distribution can be employed to incorporate zero, positive and negative counts.

Therefore, the primary objective of this paper is to develop an empirical analysis of the dynamic changes in retail sector counts for the Mountain States. Specific objectives are to review past demand threshold studies, discuss the Mountain States study area, present the discrete normal distribution for dynamic count data analysis and discuss preliminary results of the dynamic commercial sector threshold analysis for the Mountain States.

A REVIEW OF MARKET THRESHOLD ANALYSIS

Threshold analysis is rooted in central place theory (CPT) in two ways. First, CPT predicts that there is a direct and positive relationship between the population of the central place and the number of firms. Here, number of functions can be proxied by the number of firms within the central place. In other words, as the population of the central place increases, so do the number of firms within the place.
Second, and perhaps more fundamental, CPT predicts that goods will have a specific limitation to the size of their market in a spatial sense. The radius of this market determines the range of the good. The larger the range of the good, the larger the spatial size of the market supporting that good. The key determinants of a good’s range are the demand for the good and the cost of supplying the good. Specifically, the interaction of the Losch demand cone and the firm’s average cost curve determines the range or market size of the good. Given that the cost structure facing the firm is determined exogenously from CPT (i.e., factor prices and good’s production technology) the primary determinant of a good’s range, or spatial market, will be the characteristics of the good’s aggregate demand structure (i.e., demand cone). A spatial equilibrium is achieved when the dollar volume under the demand structure is just sufficient to cover operating costs and allow an acceptable rate of return.

Threshold analysis attempts to proxy the demand structure for a good by relating population to the number of functions (i.e., number of businesses) within a particular central place. Berry and Garrison (1958a, 1958b) suggested that this relationship can be expressed as

$$P = \alpha B^\beta$$

(1)

where $P$ is the place’s population, $B$ is the number of businesses of a particular type within the place and $\alpha$ and $\beta$ are parameters to be estimated. The nonlinear specification follows from CPT. In practice, the estimated equation is a double-log model. Given estimates of $\alpha$ and $\beta$, one may substitute $B = 1$ and solve for the population required to support one firm. Hence, a proxy measure for the size of the supporting demand structure for the good is provided.

The use of this specification for estimating market thresholds raises several problems. First, the use of a logarithmic transformation affects the nature of the estimates produced. The regression procedures estimate the logarithm of the number of businesses, not the number of
businesses themselves. The antilog of these estimates are biased estimates of the number of businesses (Haworth and Vincent 1979).

A second difficulty arises by the use of the logarithmic transformation when a place’s number of businesses for a particular type is zero. Since the logarithm of zero is negative infinity, a small positive number is usually added to all observations or zero observations are removed from the sample. In rural areas where there are numerous places with no retail activity in some sectors, this difficulty can lead to serious problems. Adding a small positive number will result in upward, nonparallel shift of the relationship and biased estimates of threshold populations.

A third problem many past researchers seemed to share was a reversal of the logical cause-effect relationship between population and number of businesses (Chrisman 1985). Berry and Garrison (1958a) for example, regress number of businesses onto population. Because the number of businesses is the random variable within the problem, placing it on the right-hand-side of the equation results in both biased and inconsistent estimates. Not all threshold studies, however, are subject to this shortcoming (Foust and de Souza 1977; Foust and Pickett 1974).

A fourth shortcoming of the bulk of the empirical threshold literature is the sparseness of the specification of the estimated equation. Numerous studies use population as the sole determinant of market demand. As argued by Murray and Harris (1978) the number of businesses supported by a given population is influenced by many factors. Other studies or retail activity have determined that socioeconomic factors, such as income levels and distribution, population density and spatial competition can dramatically affect the size and shape of the market demand cone (Deller and Chicoine 1989; Henderson 1990). By omitting relevant variables, the parameter estimates will be biased.
A final problem concerns the use of OLS procedures to estimate numbers of businesses. Ordinary least squares assume that the number of businesses are normally distributed which implies that the possible values which can be taken by the random variable are normally distributed around the estimate. There is little reason to suppose the values are normal. In fact, the number of firms are non-negative and integer which would suggest count data procedures.

Harris et al. (1996) and Harris and Shonkwiler (1993) applied count data procedures to estimate minimum demand thresholds at the county level. Wensley and Stabler (1998) employed count data procedures to estimate demand thresholds for rural Saskatchewan at the local or community level. In so doing, they highlighted a common observation that rural areas are characterized by lower demand thresholds and, therefore, higher frequency of business establishments relative to areas that are more proximate to urban centers, other things being equal.

However, employing count data models such as the Poisson or negative binomial only involve non-negative integers. Quantifying those factors that cause commercial sector counts to decrease may be as important as quantifying factors relating to increases in commercial sector counts. In order to derive factors that cause change in rural Mountain States’ commercial sector counts, the discrete normal distribution is employed.

Study Area

The previous literature suggests several paths for research and there are various market areas that can be investigated. Our study will still be at the county level, because “in general, counties are the smallest geographical areas for which significant amounts of data are available, thus permitting estimation of effects of a rich set of predetermined and exogenous variables” (Carlino and Mills, 1987).
There are 280 counties in the eight U.S. mountain states - Arizona, Colorado, Idaho, Montana, New Mexico, Nevada, Utah and Wyoming. Among them, 70 rural counties which are not adjacent to a metropolitan area, and have population of less than 7,000 (in the year 1988) were selected as the study area. Sparsely populated counties were chosen for this study because of the concern that rural areas were not benefiting from national economic growth and because of their peculiar market and structure. For the dynamic commercial sector analysis, retail and service sector establishment numbers were gathered for 1988 and 1997.

The change in commercial sector establishment numbers was derived by finding the difference between the 1997 and 1998 values. By employing the discrete normal distribution, which can consider positive, negative and zero changes, this study will expect the following results. First, we will estimate variables that affect business entry and exit in rural areas. Second, whether the model applied in our study has predictive power for the future variation of establishment changes. Last, what is the marginal impact of specific factors to the change in commercial sector counts.

**Statistical Methods:**

To model firm dynamics, the change in number of firms in an industry between two successive time periods is analyzed. If this change is negative, more firms have exited the industry than have entered; if it is zero then firm entry and exit have balanced out; and if it is positive there has been a net influx of firms. Given the focus on isolated rural communities, it is expected that these counts would be comprised of small (positive and negative) integer values. These discrete outcomes are generated by a dynamic process. This framework requires a stochastic model of the counts of firms. Changes in this stochastic process between the two time periods then should generate the model required for the firm change data. For this reason, the
discrete normal distribution of Kemp (1997) is adopted, because it can be related to the
difference of two related Heine distributions. The discrete normal distribution is outlined below
as an approach for estimating the differenced data, and then the Heine distribution as an adjunct
model to represent the firm levels data on the total counts of firms is discussed.

Kemp (1997) presented the discrete normal as a maximum entropy distribution by
characterizing its probability mass function (pmf) as

\[ P(Y=y) = \frac{\lambda^y q^{1/2(y^2-y)}}{\sum_y \lambda^y q^{1/2(Y^2-Y)}} \]  

(2)

\[ \lambda > 0; \ 0 < q < 1; \text{ and } \ Y = \ldots, -3, -2, -1, 0, 1, 2, \ldots \]

Note that the numerator of the pmf has as its argument the realization of the random
variable (y) while the denominator is a normalizing factor that is summed over the support of the
distribution (all possible values of Y). Further this characterization does not permit closed form
expressions of the mean and variance of Y.

Although not recognized by Kemp, the parameters \( \lambda \) and q may be transformed to permit
a representation of the pmf in terms of parameters associated with the mean and variance. Begin
by letting \( \lambda^2 = q\beta^{-2}\mu \) with \( 0<\beta<1 \) and \( -\infty<\mu<\infty \). Without loss of generality \( q = \beta \) can be set.

Then this implies

\[ P(Y = y) = \frac{\beta^{5y^2-\mu y}}{\sum_y \beta^{5Y^2-\mu Y}} \]  

(3)

so that \( E(Y)=\mu \) and \( V(y)=-1/\ln\beta \). It is also concluded that \( \beta = e^{-1/\sigma^2} \) from this latter relationship.

After some algebra the discrete normal is derived as
\[ P(Y = y) = \frac{e^{-\frac{5(y-\mu)^2}{\sigma^2}}}{\sum_{Y} e^{-\frac{5(Y-\mu)^2}{\sigma^2}}} \]

with \( E(Y) = \mu \) and \( V(Y) = \sigma^2 \). To recover the original parameters of Kemp, the relationships:

\[ \lambda = e^{\frac{(\mu - 5)}{\sigma^2}} \text{ and } q = e^{-1/\sigma^2} \]

are employed. Another advantage of this parameterization of the discrete normal is that it permits a generalization to the multivariate case. This means that compromise sectors can be jointly modeled and tests of interdependence can be constructed.

As mentioned, Kemp showed that the discrete normal can be obtained as the difference between two related Heine distributions which depend on the same parameters \( \lambda \) and \( q \). This is an especially important result because in a previous article Kemp (1992a) discussed a class of discrete distributions of which the Heine is a member. The Heine and associated Euler distribution were shown by Kemp to represent certain Markov processes. From these distributions it is possible to derive the transition probabilities which describe the likelihood of an industry making marginal changes in the number of firms given an initial state. This will be demonstrated shortly, but first Heine and Euler probability mass functions are presented.

In a companion article, Kemp (1992b) described the Heine and Euler distributions and remarked on a number of their properties. For the Heine distribution the pmf is given as:

\[ P(Y = y) = P(Y = 0)\lambda^y q^{y(y-1)/2} \{(1-q)(1-q^2)...(1-q^y)\} \]

such that \( \lambda > 0; \ 0 < q < 1; \) and \( y = 1, 2, 3... \) For the Euler distribution the pmf is given as:

\[ P(Y = y) = P(Y = 0)\alpha^y \{(1-q)(1-q^2)...(1-q^y)\} \]

such that \( 0 < \alpha < 1 \) and \( 0 < q < 1 \).
Note that evaluation of $P(Y = 0)$ requires an infinite recursion for both distributions and as a consequence there are no closed form expressions for the means and variances of the distributions.

As mentioned, Kemp (1992a) related these distributions to certain queue-lengths with transition probabilities that depend on their parameters. In the case of the Heine distribution, define $\pi = \lambda/(1+\lambda)$ and for the Euler distribution define $\pi = \alpha/(1+\alpha)$ then following Kemp we have

**Heine Transition Probabilities**

\[
P_{y,y+1} = \pi q^y \\
P_{y,y} = \pi(1-q^y) + (1-\pi)q^y \\
P_{y,y-1} = (1-\pi)(1-q^y)
\]

**Euler Transition Probabilities**

\[
P_{y,y+1} = \pi \\
P_{y,y} = (1-\pi)q^y \\
P_{y,y-1} = (1-\pi)(1-q^y)
\]

Clearly these transition probabilities have different implications about the dynamics of firm entry and exit. Additionally they can be specialized to individual communities so that probabilities can be calculated for the events that i) there is a net gain in firms ii) there is no change in firm numbers and iii) there is a net loss in firms--conditional on the total number of firms in the community.

**Statistical Model**

For the analysis for commercial sector establishment count changes in the Mountain States, the following equation was used:

\[
\Delta E_{ij} = \alpha + b_1 \Delta POP_j + b_2 \Delta Pd_j + b_3 \Delta PCI_j + b_4 \Delta EMP_j + b_5 \Delta HW_j + \varepsilon_i
\]

Where $\Delta E_{ij}$ is the change in establishment count for commercial sector in county $j$ from 1988 to 1997;

$\Delta POP_j$ is change in County population in county $j$ from 1988 to 1997;
\( \Delta P_d_j \) is change in county population density in county \( j \) from 1988 to 1997;

\( \Delta PCI_j \) is change in county real per capita income in county \( j \) from 1988 to 1997;

\( \Delta EMP_j \) is the change in employment in county \( j \) from 1988 to 1997; and

\( HW_j \) is a 0/1 dummy variable denoting if an interstate highway existed in county \( j \) in 1997.

Univariate discrete normal models were run for seven retail sectors. The maximum likelihood results are shown in Table 1. The change in population was statistically significant for all retail sectors except for the Miscellaneous Retail Sector. For all statistically significant retail sectors, change in population will increase changes in retail sector counts.

A similar result is found for change in employment. From Table 1, the change in employment was statistically significant for all retail sectors, except for the General Merchandise Store. For all statistically significant retail sectors, change in employment will yield a positive increase in retail sector counts.

From Table 1, a positive change in per capita income will decrease retail sector counts for the Building Materials and Garden Supplies Sector; the General Merchandise Sector; the Food Sector; the Apparel and Accessory Store Sector; the Furniture and Home Furnishings Sector; and the Miscellaneous Retail Sector. One explanation may be that as per capita incomes increase, shoppers in these rural areas will desire better quality items that are available in metropolitan areas.

Regarding the interstate highway variable, only the Furniture and Home Furnishing Sector and the Miscellaneous Retail Sector yielded statistically significant values. The coefficients for both of these retail sectors is negative, which means the existence of an interstate highway leads to decrease retail sector counts for these two retail sectors.
**Table 1. Results of Discrete Normal Change in Establishment Numbers for Retail Sector in the Mountain States, 1988 to 1997.**

<table>
<thead>
<tr>
<th>Sector</th>
<th>Intercept</th>
<th>ΔPOP</th>
<th>ΔPₜ</th>
<th>ΔPCI</th>
<th>ΔEMP</th>
<th>HW</th>
<th>AO</th>
<th>Log Likelyhood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building Materials and Garden Supplies (SIC 52)</td>
<td>-0.2689 (2.1018*)</td>
<td>0.0137</td>
<td>0.2379</td>
<td>-0.0247</td>
<td>0.0470</td>
<td>0.0256</td>
<td>0.1696</td>
<td>-114.851</td>
</tr>
<tr>
<td>General Merchandise Stores (SIC 53)</td>
<td>-0.2691 (2.0357*)</td>
<td>0.0057</td>
<td>-0.0019</td>
<td>-0.0326</td>
<td>-0.0059</td>
<td>-0.0954</td>
<td>0.1560</td>
<td>-114.498</td>
</tr>
<tr>
<td>Food Stores (SIC 54)</td>
<td>-0.4327 (3.9803*)</td>
<td>0.0049</td>
<td>-0.2973</td>
<td>-0.0009</td>
<td>0.0597</td>
<td>-0.1094</td>
<td>0.6219</td>
<td>-127.041</td>
</tr>
<tr>
<td>Auto Dealers and Service Station (SIC 55)</td>
<td>-0.1747 (2.1668*)</td>
<td>0.0187</td>
<td>0.0133</td>
<td>0.0057</td>
<td>0.0176</td>
<td>-0.1538</td>
<td>1.5092</td>
<td>-153.476</td>
</tr>
<tr>
<td>Apparel and Accessory Stores (SIC 56)</td>
<td>-0.7257 (4.5631*)</td>
<td>0.0188</td>
<td>0.1906</td>
<td>-0.0095</td>
<td>0.0673</td>
<td>-0.2649</td>
<td>-0.0643</td>
<td>-108.962</td>
</tr>
<tr>
<td>Furniture and Home Furnishings (SIC 57)</td>
<td>-0.0350 (2.5080*)</td>
<td>0.0603</td>
<td>-0.3516</td>
<td>-0.0045</td>
<td>0.0110</td>
<td>-0.3266</td>
<td>0.4351</td>
<td>-121.884</td>
</tr>
<tr>
<td>Eating and Drinking Places (SIC 58)</td>
<td>0.1069 (1.8232*)</td>
<td>0.0469</td>
<td>-0.0251</td>
<td>0.0135</td>
<td>0.0230</td>
<td>-0.1396</td>
<td>2.2773</td>
<td>-178.17</td>
</tr>
<tr>
<td>Miscellaneous Retail (SIC 59)</td>
<td>-0.0531 (0.7377)</td>
<td>-0.0009</td>
<td>0.3817</td>
<td>-0.0017</td>
<td>0.0568</td>
<td>-0.3266</td>
<td>1.7812</td>
<td>-162.077</td>
</tr>
</tbody>
</table>

*denotes the White’s (W) t-ratio indicates significance at the 10% level.
An important attribute of the discrete normal distribution is that marginal impacts to retail sector counts can be derived. Estimation of marginal impacts requires application of numeric derivatives. For the Eating and Drinking Sector, the estimated marginal effect from a change in county population was estimated to be 0.48. This value means that a 100 person increase in a county would yield a 0.48 increase in establishment numbers for the Eating and Drinking Sector. In other words, an increase of approximately 200 persons in county population would yield a one establishment increase in the Eating and Drinking Sector.

Marginal effects also can be derived for changes in county employment. The estimated marginal impact for county population was about six times greater than for county employment for the Furniture and Home Furnishings Sector. This indicates that increases in establishment numbers for the Furniture and Home Furnishings would more likely occur where people live than where they work.

As for per capita income, the marginal impact for the Eating and Drinking Sector was the only positive value for all the retail sectors. The value of 0.1382 means that a $1,382 increase in county per capita income is necessary for a one establishment increase in the county Eating and Drinking Sector. However, for the Apparel and Accessory Sector, the calculated marginal impact value was -0.0131. This means that for a $131 increase in per capita income, county Apparel and Accessory Sector establishment numbers will decrease by one firm.

**Conclusions**

Expansion of retail sectors has recently become an issue for rural economies. Past demand threshold studies have only investigated non-negative establishment count data. However, factors that significantly influence the decline in retail establishment numbers may be as needed as well as those that increase retail sector establishment counts.
This study investigated the use of the discrete normal distribution for estimation of changes in retail sector establishment counts. This paper has shown how results of the discrete normal distribution can be used for estimating significant factors that influence retail sector numbers in the Mountain States of the nation. Procedures that estimate the marginal effects of county factors that influence county retail sector business counts were also derived.
REFERENCES

Berry, B. and W. Garrison. “A Note on Central Place Theory and the Range of Good.”


Chrisman, J.J. “Population Change and Its Effect on the Retail Sector: An Exploratory Study,”


   Unpublished manuscript, Department of Geography, University of Wisconsin, Eau Claire, 1974.


