THE FUNCTIONAL FORM OF THE GRAVITY MODEL: A NEW TECHNIQUE WITH EMPIRICAL RESULTS

James B. Kau and C. F. Sirmans

#359

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign
THE FUNCTIONAL FORM OF THE GRAVITY MODEL:
A NEW TECHNIQUE WITH EMPIRICAL RESULTS

James B. Kau and C. F. Sirmans

#359
THE FUNCTIONAL FORM OF THE GRAVITY MODEL: A NEW TECHNIQUE WITH EMPirical RESULTS

by

James B. Kau
Visiting Associate Professor
Department of Finance
University of Illinois
Urbana, Illinois 61801

C. F. Sirmans
Assistant Professor
Department of Finance
University of Illinois
Urbana, Illinois 61801
THE FUNCTIONAL FORM OF THE GRAVITY MODEL:
A NEW TECHNIQUE WITH EMPIRICAL RESULTS

1. Introduction

Gravity models have probably been used more in regional planning and transportation studies to examine spatial interaction than any other form of mathematical model. Regional economists generally agree that spatial interaction is positively related to mass and inversely related to distance. There is much less agreement, however, on the theoretical foundations of the relationship and its precise functional form. The purpose of this paper is to examine the latter, i.e., the functional form of the gravity model. The examination of the precise functional form is important since misspecified equations may lead to incorrect conclusions concerning many types of policy questions. For example, gravity model formulations are used in many modeling efforts to measure the impact of changing a transportation system on spatial interaction. Incorrect specification of the gravity model will lead to unreliable predictions.

One of the criticisms of using the gravity model has been the lack of a technique for choosing among alternative functional forms. This paper demonstrates a functional form technique developed by Box and Cox [7] which not only allows for discrimination between the linear, the double-log, and exponential specifications but also provides a generalized functional form approach to reduce specification error for
use in future research. It is suggested that this transformation of variables procedure is a powerful technique to handle the general problem of choice of functional form of the gravity model in regional planning and transportation studies.

2. The Traditional Gravity Model

The standard functional form used in regional economic theory to explain spatial interaction phenomena is based on a potential energy definition and thus referred to as a gravity model. The traditional equation is:

\[ I_{ij} = G M_i^{\beta_1} M_j^{\beta_2} C_{ij}^{\beta_3} \]

where

- \( I_{ij} \) = any transportation or communication flow between origin (i) and destination (j), such as journey-to-work trips;
- \( G \) = a constant term;
- \( M_i \) and \( M_j \) = some measure of mass at i and j;
- \( C_{ij} \) = cost of interaction between i and j; and
- \( \beta_1, \beta_2, \beta_3 \) = exponents of \( M_i, M_j \) and \( C_{ij} \), respectively.

In many studies (for example, G. K. Zipf [37, 38] and J. D. Carroll and H. W. Bevis [9]), \( \beta_1 = \beta_2 = 1 \) and \( \beta_3 \) is determined by regression analysis on the log functional form of equation (1). Equation (1) can be modified to consider other possible factors such as income and education. This could be accomplished by weighting the mass variable with an appropriate index of income and education (Isard and Freutal [21]). In most studies the ratio of actual to expected total person trips is
assumed to be a function of distance. Equation (1) becomes

\[ \frac{T_{ij}}{T_{ij}} = \frac{1}{d_{ij}^\beta} \]  

where

\[ T_{ij} = kM_i M_j / M; \]
\[ k = \text{average number of trips per capita}; \]
\[ d_{ij} = \text{distance between } i \text{ and } j \text{ and is used as a proxy for } C_{ij}; \]
\[ M = \text{total mass for area}. \]

The issue of this paper concerns the general form of the function, in particular the influence of distance. As discussed by Isard [20, p. 510], the Carroll and Bevis [9] empirical study suggests a quadratic rather than a double-log function. Other empirical studies also question the applicability of the double-log function. For example, Anderson [3] tests the hypothesis that the exponent of \( d_{ij} (\beta_i) \) is a function of population. Carroll and Bevis [9] have also noted that for the distance variable, different exponents apply to different trip purposes.

Recent theoretical studies have suggested several possible structural forms. Beckmann and Wallace [6] and Golob and Beckmann [15] in modeling individual trip preferences have incorporated the interrelationship between opportunity interactions and their trips. An individual's net utility is derived from the utility of interaction for each spatial opportunity \( i \) minus the reduction in utility from traveling to \( i \). Smith [31], following the same line of development, presents a theory of travel preferences which leads to distance-dependent utility functions. Trip-makers
are assumed to discount their anticipated opportunity interactions in terms of the distance to the opportunities. Smith [31] demonstrates the possibility of exponential spatial discounting behavior within the axiomatic theory of spatial discounting. Isard [18] drawing on the contributions made by Niedercon and Bechdolt [24], Allen [1], Golob, Gustafson, and Beckmann [16], Smith [32], and Wilson [33] develops a rationale for travel behavior which is consistent with both the gravity model trip pattern and exponential spatial discounting.

This study accepts the theoretical argument that spatial interaction is related to distance and presents a technique that allows the data to give information on the correct functional form. Using methods suggested by Cesario [12] and devised by Box and Cox [7] and Box and Tidwell [8], journey-to-work data from specified SMSAs are used as an example of its application.

3. The Generalized Functional Form

The relationship between the interaction variable and distance can be generalized by power transformations on the dependent and independent variables of equation (2). The general form is

\[(I_{ij}/M_i M_j)^{\lambda_1} = Y_0 + \gamma_1 d_{ij}^{\lambda_2}\]

where \(\lambda\) is a functional form parameter and \(d\) represents the cost of interaction, i.e., distance or time. If \(\lambda_1, \lambda_2 = 1\) then equation (3) represents the linear case. If \(\lambda_1, \lambda_2 \to 0\)
then equation (3) represents the double-log case. If \( \lambda_1, \lambda_2 \neq 1 \) or 0, then equation (3) will represent some form of equation (2) in which the variables are expressed in a non-log power transformation relationship. If \( \lambda_1 \) is allowed to vary and \( \lambda_2 \) is assumed to approach one, then equation (3) is used to test for the existence of an exponential distance variable.

The generalized functional form [equation (3)] allows the elasticity of trip demands with respect to distance to vary with changes in distance. The distance elasticity for trip demand from equation (3) can be shown to be (assuming \( \lambda_1 = \lambda_2 = \lambda \))

\[
Nd = \frac{\partial \frac{I_{ij}}{M_{ij}}}{\partial d_{ij}} \cdot \frac{d_{ij}^{\lambda_1}}{I_{ij}^{\lambda_1}} = \gamma_1 \left[ \frac{d_{ij}}{I_{ij}} \right]^{\lambda_1}.
\]

If \( \lambda = 1 \), the linear case, then \( Nd = \gamma_1 \left( \frac{d_{ij}}{I_{ij}} \right) \). In the log case as \( \lambda \to 0 \), \( Nd = \gamma_1 \), thus if the distance elasticity is initially one then as the distance of the trip increases the elasticity remains one. The distance elasticity increases as distance grows if \( \lambda > 0 \); it decreases if \( \lambda < 0 \), and \( d_{ij} > 1 \).

4. The Estimation of \( \lambda \)

Based upon Box and Cox [7], Box and Tidwell [8], and Zarembka [36], the generalized functional form [equation (3)] for describing the relationship between the interaction variable (trips in this case) and distance can be defined as

\[
\frac{(T_{ij}^\lambda - 1)}{\lambda} = \gamma_0 - \gamma_1 \left( \frac{(d_{ij}^\lambda - 1)}{\lambda} \right) + \epsilon_i
\]
where equation (5) represents the general functional form of equation (3) with $T_{ij} = \frac{I_{ij}}{M_i M_j}$ and $\gamma_0$ and $\gamma_1$ being regression parameters and where the disturbance term $(\varepsilon_i)$ is normally distributed with zero means and variance of $\sigma^2$. Using maximum likelihood techniques developed by Box and Cox [7], $\lambda$ and the regression coefficients are estimated from the data.

Under the assumption of normality, the probability density function for $\varepsilon_i$ in equation (5) is written as

$$f(\varepsilon_i) = (2\pi \sigma^2)^{-1/2} \exp(-1/2[\varepsilon_i^2/\sigma^2])$$.

If the $\varepsilon_i$'s are identically and independently distributed, then the log likelihood function for equation (6) can be written as

$$\log L = \frac{-n \log 2\pi^2}{2} + (\lambda - 1) \sum_{i=1}^{n} \log T_{ij} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left\{ \frac{T_{ij}^\lambda - 1}{\lambda} \right\}^2 - \gamma_0 + \gamma_1 \left[ \frac{d_{ij}^\lambda - 1}{\lambda} \right]^2$$.

The logarithmic likelihood is maximized with respect to $\sigma^2$, $\gamma_0$, and $\gamma_1$ given $\lambda$. The estimate of $\sigma^2$ for a given $\lambda$, $\hat{\sigma}^2(\lambda)$ is then the estimated variance of the disturbances of regressing $(T_{ij}^\lambda - 1)/\lambda$ on the explanatory variable $d_{ij}^\lambda$. Replacing $\sigma^2$ by $\hat{\sigma}^2(\lambda)$, the maximum log likelihood $[\log L]\text{max}(\lambda)$ for equation (3) is

$$\log L\text{max}(\lambda) = \frac{-n}{2} \log 2\pi \hat{\sigma}^2(\lambda) - \frac{n}{2} + (\lambda - 1) \sum_{i=1}^{n} \log T_{ij}.$$
Box and Cox [7, p. 216] indicate that an approximate 95% confidence region for \( \lambda \) is obtained from

\[
(9) \quad L_{\text{max}}(\hat{\lambda}) - L_{\text{max}}(\lambda) < 1/2\chi^2(.05) = 1.92.
\]

Thus the general-functional-form [equation (3)] is used to test whether the functional form parameters \( \lambda_1 \) and \( \lambda_2 \) are significantly different from zero and/or one.

5. The Empirical Results

The basic assumption underlying the Box and Cox method is that some transformation of trips \( T_{ij} \) exists which is normally distributed and linearly related to distance. The data used in this paper consist of 489 intercounty trip flows between place of residence and place of employment in four different SMSAs in the State of Georgia. A detailed time and distance matrix with approximately 700 different travel nodes for 1970 was used. Population and employment data are used as a measure of mass and were collected for each destination and origin of a trip.

The functional form parameter is determined by transforming \( T_{ij}/M_iM_j \) and \( d_{ij} \) in accordance with equation (3) using \( \lambda \)'s between -.50 and 1.50 at intervals of 0.1. Twenty-one different regressions are estimated for each equation set. The \( L_{\text{max}}(\lambda) \)'s of the equation sets are calculated by equation (8) and listed in the Appendix. There are eight separate equation sets, four of which concentrate on the appropriate power transformation for the dependent variable and four of which have power transformations on both
the dependent and independent variable as specified by equation (3). The separate equations are derived by substituting population \((P_iP_j)\) or employment \((E_iE_j)\) for \(M_iM_j\) and using travel time \((t_{ij})\) and distance \((d_{ij})\) as measures of the costs. Population is measured as the total population living in origin \((P_i)\) and the destination \((P_j)\). When employment is used as the measure of mass, \(E_i\) is the number of trip origins at area \((i)\) (the number of employed persons living in area \(i\)) and \(E_j\) is the total number of trip-destinations in area \((j)\) (i.e., the number of persons working in area \(j\)). Thus this paper also determines whether using employment or population as the base and whether using travel time or distance makes significant changes in the functional form and the corresponding estimated coefficients.

Initially the empirical tests concentrate on the issue of the appropriate power transformation for the dependent variable \((I_{ij}/M_iM_j)\) thus iterating \(\lambda_1\) and fixing \(\lambda_2\) at unity so that distance enters linearly. The sensitivity of the maximum likelihood estimates of \(\lambda_1\) to introducing spatial interaction of employment centers as compared with population is also determined. The results of these iterations for the first equation set are summarized in Figure 1. The likelihood maximizing value of \(\lambda_1\) is 0.0. The optimal value of the power transformation is not affected by introducing travel time as the independent variable rather than distance.

Using the \(\chi^2\) test with one degree of freedom, it is possible to reject the linear dependent variable. The optimal
Figure 1
LOG MAXIMUM VALUES FOR ALTERNATIVE λ'S FOR THE SEMI-LOG CASE WITH POPULATION AND EMPLOYMENT*

Figure 2
LOG MAXIMUM VALUES FOR ALTERNATIVE λ'S FOR THE DOUBLE-LOG CASE WITH POPULATION AND EMPLOYMENT*

*When travel time is introduced in the place of d_{ij} there is no significant changes in the optimal λ or the log likelihood values.
A value of zero suggests that the natural logarithmic transformation dominates all other forms. This initial result thus provides a case for exponential spatial distribution behavior. While the population spatial model appears to dominate the employment base model in all cases, it is not possible to interpret the empirical results in this manner.

The second equation set allows for iterations on both the dependent and independent variables. The results of these iterations are summarized in Figure 2. The likelihood maximizing value of $\lambda_1$ and $\lambda_2$ is 0.0 and as before the optimal $\lambda$'s are not affected by introducing travel time instead of distance and again the employment model dominates. The logarithmic likelihood values for the semi-log case are similar to the double-log functional form.

6. Variations in the Estimated Coefficients

Since this study has somewhat accurate information on travel time and distance between residence and work place, it is of interest to note the coefficients for the independent variable under alternative specifications and definitions.

Even though using travel time or distance did not alter the functional form of the gravity model, there were significant differences in the coefficients in all cases. As Table I indicates, the coefficients for travel time are significantly less relative to distance in both the employment and population gravity models. For example, in the employment double-log-gravity models a 1% change in time or distance led to a 2.5% and a 5.4% decrease in spatial interaction, respectively.
TABLE I

Regression Coefficients for the Alternative
Gravity Model Formulations*

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Constant Term</th>
<th>Independent Variable</th>
<th>$R^2$</th>
<th>d.f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{i,j}$/$P_{i,j}$</td>
<td>-2.440</td>
<td>-8.140 $d_{i,j}$</td>
<td>.44</td>
<td>487</td>
</tr>
<tr>
<td>$P_{i,j}$/$P_{i,j}$</td>
<td>-8.090</td>
<td>-2.345 $t_{i,j}$</td>
<td>.45</td>
<td>487</td>
</tr>
<tr>
<td>$P_{i,j}$/$E_{i,j}$</td>
<td>-2.554</td>
<td>-5.418 $d_{i,j}$</td>
<td>.45</td>
<td>487</td>
</tr>
<tr>
<td>$P_{i,j}$/$E_{i,j}$</td>
<td>-5.350</td>
<td>-2.456 $t_{i,j}$</td>
<td>.47</td>
<td>487</td>
</tr>
<tr>
<td>$P_{i,j}$/$E_{i,j}$</td>
<td>-14.253</td>
<td>-0.065 $d_{i,j}$</td>
<td>.43</td>
<td>487</td>
</tr>
<tr>
<td>$P_{i,j}$/$E_{i,j}$</td>
<td>-14.441</td>
<td>-0.051 $t_{i,j}$</td>
<td>.43</td>
<td>487</td>
</tr>
<tr>
<td>$E_{i,j}$/$E_{i,j}$</td>
<td>-11.827</td>
<td>-0.0678 $d_{i,j}$</td>
<td>.44</td>
<td>487</td>
</tr>
<tr>
<td>$E_{i,j}$/$E_{i,j}$</td>
<td>-11.989</td>
<td>-0.053 $t_{i,j}$</td>
<td>.45</td>
<td>487</td>
</tr>
</tbody>
</table>

* t-values are in parentheses.
For the population base models the change was 2.3% and 8.14%, respectively. This result suggests that studies using distance as a proxy for the time cost of travel will overstate the impact of transportation cost on spatial interaction.

The employment and income based gravity models were separately weighted with education and then income for each origin and destination as suggested by Isard. In all cases, the functional form was not altered by the weights. An alternative treatment would be to allow these variables and others to be treated as independent variables since they probably influence travel behavior.

7. Implications for Future Research

This paper has employed the log likelihood value to determine the correct specification of the functional form of the relationship between spatial interaction and transportation cost. Using the critical value of 1.92 from equation (8), the results indicate that the logarithmic specifications dominate the linear specifications. But the log likelihood values for the semi-log (exponential case) are similar to the double-log functional form. Thus for the semi-log and the double-log functional forms there is no statistical justification for preferring a priori one specification over another. These results are for one particular test area; thus as yet no generalized statement concerning the functional form of a gravity model should be made. The major purpose of this paper has been to provide a generalized functional form [equation (3)] and a transformation technique that can be
used for different sets of data to derive the optimal specification.

Wilson [35] and others have suggested that using employment data to measure spatial interactions is superior to population data. The empirical results of this study do not necessarily provide any insight into this and hence it remains an open question. Some insight was also provided on the debate of whether distance is a suitable proxy for travel time. In every case, distance overstated, relative to travel time, the effect of transportation cost. This means that the many models using distance as a proxy for the time cost of travel may overestimate the impact on spatial interaction of various transportation facilities, such as new subways or roads.

The gravity model has been widely used to explain the relationship between spatial interaction and transportation cost. Many land-use models have used various formulations of the gravity model. (See BASS [4], Lowry [23], Alonso [2], Goldner [14], and Putman [26].) These models have been applied to different areas under different circumstances; the possibility of measurement errors, specification bias, and corresponding inaccurate estimates exist in all cases. The Box and Cox technique used in this paper could be incorporated in all land-use models to reduce specification bias and improve predictive power.

Finally the generalized functional form of the gravity model [equation (3)] derived in this paper can be used to test the validity of various spatial distribution theories.
Empirically derived optimal $\lambda$ values determine and alter the functional form of the gravity model. Hence, the Box and Cox technique can be used to reduce specification bias in each test area and at the same time provide insight into the correct theoretical functional form.

In conclusion, Box and Cox [7] have ascertained that after suitable transformation of the variables, (1) the expected values of the transformed observations are described by a model of simple structure, (2) the error variance is constant, and (3) the observations are normally distributed. In addition, Box and Tidwell [8] have shown that the main purpose of transforming the independent variable is to reduce the function in these transformed variables to as simple a form as possible. From the implications of these transformations and the empirical results of this paper, it is concluded that both semi-log (exponential case) and double-log relationships are efficient functional forms to be used to describe the relationship between spatial interaction and transportation cost.

This study leaves unexplained the correct functional form for the more abstract models of travel demand. Further research is necessary to generalize such models as presented by Quandt and Baumol [28, 29], Quandt and Young [30], and Howrey [17]. The abstract mode models are cross-sectional demand models for passenger transportation. The functional form of these demand models has varied; the correct specification can be determined with the techniques
described in this paper. Further, as Lee [22] has recently noted, disaggregation of data to take into account differentials in socio-economic characteristics and trip purpose would probably result in substantial improvements in the descriptive and forecasting ability of gravity models. The technique suggested in this paper would lend itself easily to testing the functional form of travel behavior, for example, of whites and non-whites as well as any other type of disaggregation which theory would possibly suggest.
FOOTNOTES

1 For recent discussions of the theoretical foundations of the gravity model, see, for example, Isard [18] and Smith [31]. See Lee [22] for a discussion of the various problems and limitations of the gravity model.

2 The literature on gravity models is extensive and this paper will not attempt to review it. For a historical review, see Carrothers [10], Wilson [34], and Olsson [25].

3 For a review of gravity model derivations, see Isard [20, pp. 493-568] and Isard [19].

4 As $\lambda_1$ and $\lambda_2 \rightarrow 0$ the equation reduces to

$$
\left(\frac{I_{ij}}{M_i M_j}\right)^{\lambda} - 1 \rightarrow \frac{\gamma_0^{\lambda} - 1}{\lambda} + \frac{\gamma_1 (d_{ij}^{\lambda} - 1)}{\lambda}
$$

for some $\gamma_0$

$$
\lim_{\lambda \rightarrow 0} \left(\frac{I_{ij}}{M_i M_j}\right)^{\lambda} - 1 = \lim_{\lambda \rightarrow 0} \frac{\gamma_0^{\lambda} - 1}{\lambda} + \lim_{\lambda \rightarrow 0} \frac{\gamma_1 (d_{ij}^{\lambda} - 1)}{\lambda}
$$

$$
\log \left(\frac{I_{ij}}{M_i M_j}\right) = \log(\gamma_0) + \gamma_1 \log(d_{ij})
$$

through the use of l'Hospital's rule.

5 The Box-Cox technique can be handled with any regression program and solving for the log maximum (equation 8) by hand. A Box-Cox computer program is available from the authors.

6 Data on trips and employment are from the Census of Population 1970, Special Tabulation from Census Tapes. Data for distance and time are from a special tabulation by the Department of Transportation, State of Georgia.

7 The calculated log maximum likelihood values for alternative $\chi$'s are available in a data appendix. In addition to the generalized form of equation (3), several alternative gravity model forms were tested. These equations were:
\[ I_{ij}^{\lambda_1} = \gamma_0 + \gamma_1 (M_i M_j)^{\lambda_2} - \gamma_2 (C_{ij})^{\lambda_3} \]

and

\[ I_{ij}^{\lambda_1} = \gamma_0 + \gamma_1 (M_i)^{\lambda_2} + \gamma_2 (M_j)^{\lambda_3} - \gamma_3 (C_{ij})^{\lambda_4} \]

Where \( I_{ij} \) is the number of journey trips, \( M_i \) and \( M_j \) are the mass variables in \( i \) and \( j \) represented by employment and population and \( C_{ij} \) is the travel costs represented by distance or travel time between \( i \) and \( j \). Thus there are eight possible equation forms represented by these equations. The results indicated that the log-linear form was the optimal form of the various equations.
REFERENCES


