Modeling and Analysis of Transportation Flows Created by E-commerce Transactions

Adam Ho, Erhan Kutanoglu, Michael Cole
Department of Industrial Engineering
University of Arkansas, Fayetteville, AR

Michael Bartolacci
Computer Science
Pennsylvania State University, Reading, PA

1. INTRODUCTION

1.1 Motivation

Many studies and surveys show the signs of an exponentially growing Internet-based economy. The recent growth of business and trade realized over the Internet has drawn a lot of attention to electronic business, whether it is business-to-business or business-to-consumer. The increasing availability of e-commerce solutions provides firms with new potential for reaching new customers and business partners. Traditionally, the two most formidable barriers for this type of extended business have been distance and the lack of access to key sales and marketing areas. With the potential removal of such barriers in the new economy, United States electronic sales are projected to be $380 billion in 2002 (U.S. Department of Commerce News, 2001). Ultimately, e-commerce will change and make an impact on the United States economy in terms of sales, jobs, and business opportunities. The goal of this study is to identify the effects of e-commerce-based transactions on transportation freight movements between regions. We believe that e-commerce has diminished and will continue to diminish the barrier effect of distance in the U.S. economy.

The question that arises is “How will growing e-commerce affect the physical transportation network?" Similar to traditional commerce transactions, an e-commerce transaction may result in transfer of goods. This physical exchange of goods relies heavily on the traditional transportation network. We hypothesize that the growing number of e-commerce transactions affects the distribution of loads on the transportation network, with potential changes on the usage of different modes such as air, rail, road, and inland water. In e-commerce, instead of shipping 100 computers in one truckload to a local store, 100 boxes, each with one computer, are shipped to a dispersed set of customers. For example, on a single Saturday in July 2000, 100 airplanes and 9,000 trucks delivered more than 250,000 copies of *Harry Potter and the Goblet of Fire* to Amazon.com customers all over the United State (Environmental News Network, 2000). Also, according to a senior fellow at Inform, an environmental research organization in New York City, “It's unlikely e-commerce will save the planet as some have claimed," says Bette Fishbein. "There might be some reductions in energy use, but a huge increase in packaging and shipping by air results in much more air pollution (Environmental News Network, 2000).

A quick analysis of the U.S. Census Bureau's commodity flow surveys (1993 and 1997) indicates an increase in the average distance of each ton of products shipped (www.census.gov). This implies that over the years, an average load is shipped to a destination that is farther away from the origin. Although not all of such changes are due to e-commerce, we believe that the growth of e-commerce results in a diminishing effect of distance on transportation flows between distant regions. Our goal is to model the

changes in the distribution of transportation flows given increasing amounts of e-commerce and the corresponding diminishing importance of distance.

For the purposes of planning by governmental agencies and transportation providers, surveys have already been undertaken through partnerships between the Census Bureau, the Department of Commerce and the Department of Transportation to collect data on the movement of goods (not necessarily e-commerce initiated). The data from this survey, referred as commodity flow survey (CFS), are used by public analysts and transportation providers to assess the demand for transportation facilities and services, energy use, and environmental concerns. We foresee that public analysts and transportation analysts can make use of the knowledge on changing pattern of flows due to e-commerce to allocate resources and plan for the future.

Currently, e-commerce represents approximately 1% of the total U.S. economy (U.S. Department of Commerce News, 2001). E-commerce generated flows are only a fraction of the total flows. However, due to its expected exponential growth, e-commerce will represent a significant part of the economy in the near future. At this point, the Census Bureau claims that producing a separate series of data at micro levels for e-commerce can be very difficult and expensive (Atrostic, Gates and Jarmin, 2000). Therefore, in this research, we seek to model the flow of goods for e-commerce by using the readily available commodity flow survey data from the Census Bureau. We intend to focus on a handful of selected SCTG (Standard Classification of Transported Goods) codes. Some of these products are more e-commerce related in the sense that their flows will be more likely be affected by e-commerce much earlier than others. A model developed for an e-commerce-oriented product will give us a better feel of the future

directional distribution of other products when e-commerce grows even further. We employ the use of gravity model in this research because it is capable of capturing the major components that contribute to the transportation flows system.

1.2. Report Outline

The rest of the report is organized as follows: In Chapter 2, we provide a rather extensive review of the relevant literature on different applications of gravity models, a model that is primarily used in estimating transportation flows between regions. In this research, this model has been further developed to estimate freight movements with the effect of e-commerce. We discuss data sources in Chapter 3. In Chapter 4, we discuss our modeling effort and how we calibrate the base flow data obtained from the 1997 commodity flow survey. Base flow is the historical flows of goods exchanging between regions. We also present the way we determine different parameters to estimate future flows, and the process in assigning the flows to different transportation modes. We provide our first preliminary analysis using SCTG code '35' (electronic and electrical products, and office equipments) as our base flow condition in Chapter 5. We also show our preliminary output of the gravity model and the way we use two other product flow data to eliminate the bias effect toward product code '35'. We show that we can use the results to validate the use of gravity modeling for quantifying the directional distribution of transportation flows under diminishing effects of distance in e-commerce. We also present the findings of the expected usage of different modes in year 2005. Finally, we summarize our main findings and provide some insights for future research in Chapter 6.

2. LITERATURE REVIEW

In this chapter, we present a review of literature on gravity modeling, which is the modeling tool that we use in this research. We first give an overview of gravity modeling applications in trip and freight distribution as well as other economic applications. While we review the relevant literature, we also highlight several differences between previous research and our implementation for e-commerce flows. To our knowledge, no previous research effort has used gravity models to capture the effect of e-commerce on the United States transportation network. In that sense, this research can be viewed as the first attempt toward such a goal.

2.1 Introduction to Gravity Models

Gravity modeling was first introduced into transportation modeling in the 1950's. Gravity models belong to a class of models called *synthetic models*, and they are generally used for the rough approximation of actual movements (Hamburg, Kaiser and Lathrop, 1983). Gravity models are often used for estimating trip distribution in a transportation context. These models have also been modified and used to estimate freight flows between a set of production and consumption regions. The gravity model is particularly useful when there are sizable distances and cost differences between each pair of production and consumption regions. Such characteristics are present in the world of electronic commerce.

Gravity models were originally developed from an analogy with Newton's gravitational law (Ortuzar and Willumsen, 1990). The simplest formulation of the gravity model is

$$T_{ij} = \frac{\alpha_{ij} O_i O_j}{d_{ij}^2} \tag{1}$$

where T_{ij} = the number of trips between origin i and destination j,

O_i = population of origin region i,

O_j = population of destination region j,

d_{ij} = distance between origin i and destination j,

α_{ij} = proportionality factor

The initial gravity model, Equation (1), was later modified by modeling the effect of distance with a more generic function $f(c_{ij})$, which represents the disincentive to travel as distance, time or cost increases. The modified model thus became

$$T_{ij} = \alpha_{ij} O_i O_j f(c_{ij}) \tag{2}$$

The *deterrence* function $f(c_{ij})$ (also called the *impedance*) is usually defined in terms of distance between region i and j. The potential problem with any gravity model application is that, in the flow matrix, the consumption of say Region 1, 2 and 3 may not be equal the production of Region A that has produced the flows. Since gravity model is a closed system, where all flows that are created are consumed within the system, the summation of consumption capacities should equal the production capacities. Therefore, an iterative process is employed to adjust T_{ij} to achieve equality between production and consumption (Hamburg, Lathrop, and Kaiser, 1983). This procedure will be further discussed in section 4.3.

2.2 Gravity Modeling in Trip Distribution Problems

The *trip distribution* problem deals with the assignment of traffic from given origin zones to given destination zones. This problem is built on the idea of accessibility of one region from another, thus creating the inter-activity between regions. In reference to the traditional form of gravity mode shown in Equation (1), the population of origin region O_i is substituted with P_i, which represents the production capacities, and the population of destination region O_j is substituted with C_j, which represents the consumption capacities. The relative number of opportunities such as work opportunities can be used as an accessibility measure for a zone. In this research, this can be viewed as the opportunity for online businesses to reach additional sellers/customers in further reaching regions, which then creates additional transportation flows.

The types of *marginal constraints* with which we shall be primarily concerned are of the forms

$$\sum_j T_{ij} = P_i \tag{3}$$

$$\sum_i T_{ij} = C_j \tag{4}$$

T_{ij} = Number of trips flowing from region i to region j

P_i = Number of trips originated from region i

C_j = Number of trips consumed by region j

These marginal constraints eliminate the gravity model problem discussed in Section 2.1 where all flows from region i to region j within a system should equal the production and consumption capacities. These constraints can also be represented as shown in Equation 5 and 6.

$$(\sum_i P_i = \sum_j C_j = 1) \qquad (5)\ (6)$$

Trip distribution models involving these types of marginal constraints are referred to as *doubly-constrained* distribution models (Erlander and Stewart, 1990). The gravity model that we develop in this research is also a doubly-constrained gravity model. A doubly-constrained gravity model could come in different forms, and such forms are governed by *impedance* values. *Impedance* values are determined from its functional form called *deterrence* function. *Impedance* values set the level of inter-activity between two regions. Erlander and Stewart (1988) present several basic forms, which we review briefly in Section 2.2.1.

The Bureau of Public Roads (Connor and Whitton, 1965) for urban area planning suggests that the most effective representation for *impedance* value is travel times. The total travel time is usually the minimum total driving time over a path between zones (or regions) plus the terminal times at both ends of the trip. Travel times provide a realistic measure of the actual spatial separation between regions, as it is likely to influence automobile drivers in their decisions as to places to work, shop, etc. In effect, the travel time factor measures the probability of making a trip during each time unit. Distance, travel cost, and many other spatial separation inter-relations have been used in the past as the factor to determine the *impedance* value.

Different Forms of Gravity Models:

a) Doubly Constrained Gravity Model with Given Inter-Zonal Weights

This type of gravity model assigns a set of inter-zonal weights for origin-destination pairs. These weights are usually viewed as constants, which can be interpreted as *a priori* weights. Erlander and Stewart (1988) define a gravity model with inter-zonal weights as follows: Given $W_{ij} \in (0,1)$, $(i,j) \in L$ (set of all possible origin-destination pairs or *Links),* T_{ij} is a solution of the doubly-constrained gravity model with given inter-zonal weights W_{ij},

$$T_{ij} = P_i C_j W_{ij} \qquad P_i > 0,\ C_j > 0,\ (i,j) \in L \tag{7}$$

where T_{ij} = Number of trips flowing from region i to region j

P_i = Number of trips originated from region i

C_j = Number of trips consumed in region j

W_{ij} = Inter-zonal weight between region i and region j

L = Set of origin-destination pairs

b) Doubly-Constrained Gravity Model with Exponential *Deterrence* Function

According to Erlander and Stewart (1988), the exponential *deterrence* function is the most widely used *deterrence* function in trip distribution modeling. The exponential *deterrence* function specifies the inter-zonal weights in terms of parameter $\gamma \geq 0$, and constants $c_{ij.}$ Given $\gamma \geq 0$, and $c_{ij} \geq 0$, $(i,j) \in L$, a doubly constrained gravity model with exponential *deterrence* function is as follows:

$$T_{ij} = P_i C_j e^{(-\gamma c_{ij})} \qquad P_i > 0,\ C_j > 0,\ (i,j) \in L \tag{8}$$

c) Doubly-Constrained Gravity Model with Exponential *Deterrence* Function and *Socio-economic* Factor

This new form is a modification of the previous one with additional constants K_{ij} that are interpreted as *socio-economic* factors. *Socio-economic* factors are included in trip distribution models in order to account for trip-making potentials of individuals, or the trip production potential of origins and the trip attraction potential of destinations (Kanafani, 1983). Given $\gamma \geq 0$, $C_j \geq 0$, *and* $K_{ij} \in (0,1)$, $(i,j) \in L$, a doubly-constrained gravity model with exponential *deterrence* function $\exp(-\gamma c_{ij})$ and socio economic factors K_{ij} is as follows:

$$T_{ij} = P_i\, C_j\, [K_{ij}\, \mathrm{e}^{(-\gamma c_{ij})}] \quad P_i > 0\,,\ C_j > 0\,,\ (i\,,j) \in L \tag{9}$$

2.3 Regression and Least Square Analysis

The third form of gravity model discussed in section 2.2 is as shown in equation (10)

$$T_{ij} = P_i C_j [K_{ij} \mathrm{e}^{(-\gamma c_{ij})}] \tag{10}$$

This model is linear by itself, and with a logarithmic transformation, we can calibrate it using simple linear regression to determine better γ values (Kanafani, 1983). The calibration process helps to better estimate the *impedance* values that will properly set the inter-activity between origin and destination pairs. Note that

$$\ln [\, T_{ij} / P_i\, C_j \,] = \ln(\mathrm{K_{ij}}) - \gamma c_{ij} \tag{11}$$

According to Kanafani, in order to avoid any possible distortion in the estimate of γ when there are large c_{ij} values, a least squares function can be used. That is, one can try to minimize the sum of squared errors to fine-tune the value of γ. The sum of squared

errors or the least squares function is defined as

$$S^2 = \sum_{(i,j)\in L} (T'_{ij} - T_{ij})^2 \tag{12}$$

where T'_{ij} = Observed origin-destination flows of the base flow condition

T_{ij} = Estimated origin-destination flows

The values used as base flow conditions are obtained from 1997 commodity flow surveys. They are historical values measured in tons, which represents the flows of products from region *i* to region *j*.

In this research, T'_{ij} is obtained from the U.S Census Bureau's commodity flow survey, and T_{ij} is estimated using the model that we have developed. The least squares estimation procedure attempts to seek the closest agreement between T_{ij} and T'_{ij} by minimizing the sum of squares. This is a method to improve and to evaluate the performance of the newly developed model and see how well the model is calibrated to base flow condition (Kanafani, 1983). We employ a similar procedure in our research.

2.4 Relevant Applications of Gravity Models

Carter (1993) states that gravity modeling is an accepted market analysis tool for determining the economic feasibility of retail stores. Retail gravity models were originally used to forecast the number of consumers shopping in a city. Carter (1993) uses them to evaluate the value of retail property depending on the demand for the products sold by stores. His research allocates the consumer dollars that will be spent for a type of product within a trade area based on a reasonable assumption about consumer behavior. The retail model assumes that, within a trade area, the probability that a consumer will shop at a particular store is directly proportional to some power of the size

of the store and is inversely proportional to some power of the distance between the consumer and the store. Distance is considered to be a dominating factor when it comes to trading, even if a large trade area is considered. However, in our view, this will change as e-commerce grows over time.

Retail gravity modeling is also used to quantify the economic viability of a proposed project. Bottum (1989) introduces additional parameters governing the retail gravity model. In the revised model, consumer behavior not only depends on the size of stores and distance, but also is a function of accessibility, physical barriers, driving time and income levels. This approach is feasible when a small trade area is considered.

Gravity modeling is also used in the travel industry to analyze the foreign tourist market. For example, Webster (1993) uses gravity modeling to predict the flow of tourists between a pair of countries as a direct function of each country's population and as an inverse function of the distance between them. Here distance serves as the main *impedance* contributor for tourism. However, later findings in Webster's research showed that there is a lack of significance displayed by the distance variable relative to the number of trips. Travel time turned out to be the best *impedance*.

2.5 Gravity Modeling for Freight Flow Distribution

Freight flow distribution can be defined as the movement of goods from several origins to several destinations. Modeling freight flows can be considered from multiple dimensions, such as volume, weight, and trips. Veras and Thorson (2000) consider the amount of freight measured in tons (or any comparable unit of weight) as a unit of measure for freight demand and supply. This allows commodity-based models such as

gravity models to more accurately capture the fundamental economic mechanisms driving freight movements, which largely are determined by the freight attributes such as tonnage.

In commodity flow surveys, data for both tonnage and dollar freight values are available. However, Veras and Thorson (2000) suggest avoiding using shipment dollar values since they believe that shipment values ($) exhibit more variability from one product to another. For example, freight values may be as low as $9/ton for products such as gypsum; and the value may very well exceed $500,000/ton for products such as computer chips. In addition, Veras and Thorson also discuss that using "trips traveled", may result in inaccurate results since empty trips may represent 15 to 50 percent of the total trips and the goal is to estimate actual freight being transported. Based on this, we use tonnage as the unit of measure of flow for our gravity model implementation.

2.6 Linear Programming for Freight Flow Distribution

Hamburg, Lathrop and Kaiser (1983) use linear programming (LP) for estimating freight distribution. Their LP formulation of freight flow distribution can be expressed mathematically as

Minimize $$\sum_i \sum_j c_{ij} T_{ij} \tag{13}$$

such that

$$\sum_j T_{ij} = P_i \qquad \forall i \tag{14}$$

$$\sum_i T_{ij} = C_j \qquad \forall j \tag{15}$$

$$T_{ij} \geq 0, \qquad \forall i, j \tag{16}$$

where T_{ij} = Shipment from production area i to consumption area j,

P_i = Production in Region i,

C_j = Consumption in Region j,

c_{ij} = *Impedance* value between Region i and Region j (normally distance or cost).

There are pros and cons in using LP to solve freight distribution problems. The major attraction of LP is its underlying basis of economic rationality, which is to minimize overall transportation cost. However, there is no rational central authority that could make all flow decisions between regions. In a way, each entity or region acts independently, which undermines the validity of LP approach. Moreover, the overall attractiveness is also damaged by inherent characteristics of LP, which have created some limitations in solving freight flow distribution problems Hamburg, Lathrop and Kaiser, 1983). First of all, for a system comprised of n regions, a normal solution to LP will produce no more than $(2n\text{-}1)$ of the $n(n\text{-}1)$ potential inter-regional flows, i.e., the optimal solution of the LP model will have only $2n\text{-}1$ positive flows. Secondly, LP does not allow freight flows in both directions along a link (from i to j and from j to i), which is called cross hauling. Very few commodity movements exist without some cross hauling. Lastly, in many cases, unit transport costs are not linear with distance or shipment size, as is assumed inherently in an LP formulation (Hamburg, Lathrop and Kaiser, 1983). Due to these limitations and the widespread use of gravity modeling for similar freight flow estimation problems, we use gravity modeling in this research.

3. DATA COLLECTION

The unavailability of good data is perhaps the greatest challenge we face in this research. Our goal is to model the directional distribution of flows generated by e-commerce, but there is currently no data source that has a direct measure of such flows. Estimated e-commerce sales volume in the United States (as a whole) is available, but it is not broken down into region-to-region basis. The Census Bureau has just begun to collect some survey data on the Internet economy. (Atrostic, Gates, Jarmin, 2000)

Since there is no readily available e-commerce data, we model the e-commerce flows based on the existing commodity flow survey data. Commodity flow surveys capture data on shipments originating from selected types of business establishments located in the fifty states and the District of Columbia. Businesses that participate in this program provide information on the total value of shipments, total weights, major commodity type, modes of transportation used, miles traveled, and the origin and destination of shipments. We estimate the flows due to e-commerce from the existing commodity flow survey.

Two sets of commodity flow survey data are available, 1993 and 1997. In 1993, there were virtually no significant e-commerce transactions. Therefore, we initially planned to compare the flows of a selected product code in 1993 to the flows in 1997. However, the 1993 survey data uses the detailed STCC (Standard Transportation Commodity Classification) coding system, whereas the 1997 commodity flow survey uses more aggregate SCTG (Standard Classification of Transported Goods) coding system. That is, goods are grouped within fewer product codes in 1997. Therefore, a

direct comparison between the 1993 and 1997 data is not possible. As a result, we used the 1997 commodity flow survey as our main data source.

Another set of data that we have looked at is the distribution of Internet domains registered in United States. As of June 2000, there were 13,260,000 active Web sites registered in the United States (U.S. Map New Stat, 2000). The data from this source indicates that more populous states top the list for largest number of domain name registrations. Though most web sites are inactive and do not conduct business online, surveys indicate that 80% of businesses that have registered a web address have done so to develop an on-line presence for an existing business (Network Solutions, 2000). In other words, these are companies with established business models and real products – the so-called "Click/Brick and Mortar" companies. These companies have nonetheless become the driving force behind the Internet economy, using the efficiencies and reach of the Internet to extend their traditional business models. Also note that many companies have distribution centers that would initialize shipment flows throughout the United States although they have registered their web site in another state. That is, a company's products may not come from the location where its domain is registered. The domain distribution data is not incorporated into the model, but it has provided us with better insights on the intensity of e-commerce in all the states in United States.

4. METHODOLOGY

4.1 Introduction

In this section, we first provide a brief overview of the formulation process of our gravity model that captures the directional distribution of flows. We explain the formulation procedure in a step-by-step manner. In Section 4.2, we describe the reverse derivation procedure. We use reverse derivation to determine the historical impedance of the base flow condition that leads to the flow distribution of the base flow condition. The *deterrence* function formulation will also be discussed in this section. In section 4.3, we describe the iterative procedure we use to adjust the calculated commodity flows to within 10 percent of the originally specified values. In section 4.4, we present the concept of an '*extreme*' case in *impedance* values, and show how the growing e-commerce economy is moving the *impedance* values to this *extreme*. In section 4.5, we describe how to calculate the average mile statistic. The average mile is the average distance traveled by each ton of product. We project the increase in average mile due to e-commerce such that the average exponent n can be estimated. In Section 4.6, we describe the process of determining the appropriate smoothing constant λ (a value to set the intermediate condition between current and future estimated condition). Finally in Section 4.7, we describe how the distributed flows are assigned to different modes of transportation.

We explain the steps of the procedure in more detail below:

1) Determine the geographic regions for the model. We use the 48 contiguous states. Hawaii and Alaska are eliminated due to the lack of flows and many missing data.

2) Pick product Code '35' (electronic and electrical products, and office equipment) as the representative e-commerce product. The base flow condition of our model will be based on product flows of Code '35'.

3) The *impedance* of the base flow is determined by doing a reverse derivation of

$$\left(\frac{T'_{ij}}{P'_{ij}\, C'_{j}} \right)$$

of each state-to-state pair, where T'_{ij} are actual flows of base condition obtained from the commodity flow survey of product code '35'. Note that all base (actual) conditions are differentiated with a prime (') sign.

4) Develop a distance and population based *deterrence* function that will represent the *impedance* of the new *deterrence* function. This function takes the form

$$F_{ij} = \frac{1}{d_{ij}^{n}} O_i O_j \left[\frac{1}{R_{ij}} \right] \tag{17}$$

where O_i and O_j are populations of state i and state j, respectively, and R_{ij}'s are *proportionality factor*s of the total commodity flow (ranging from 1 to 5) that we will define below. We use population as a part of the *deterrence* function since it represents the extent of demand and economic activity. We also use distance, as it is still a major contributor to the movement of goods. This function will be further calibrated in step 6.

5) An iterative procedure is performed to ensure that the production and consumption capacities that were initially specified are satisfied within 10 percent.

6) Fine-tune the *deterrence* function such that the sum of squared errors are minimized in order to determine a better n value.

$$Min\ S^2 = \sum_{ij} (T'_{ij} - T_{ij})^2 \tag{18}$$

where T'_{ij} is the base flow for each origin-destination pair directly obtained from the commodity flow survey for product code '35', and T_{ij} is the estimated flow distributed by the new *deterrence* function at the last iteration for product code '35'. We determine the n value that gives us the lowest summation of squared errors.

7) Project the average miles for product code '35' in year 2005 from the base flow. This value is used as our benchmark to determine smoothing constant λ.

8) Repeat the whole process for product codes '30' and '6' to eliminate the bias toward product code '35'. The average of n and λ values determined from the three product types is used in the model to distribute the total projected flows for year 2005.

9) Assign the flows to different transportation modes to estimate the impact of e-commerce on different transportation modes.

10) Report the increase in total ton-miles and the percent share of different mode usage in 2005.

4.2 *Deterrence* Function Formulation

4.2.1 Geographic Regions Determination

We first determine the boundaries of our study area. Our initial idea was to formulate the gravity model based on the 9 geographic divisions used by the Census Bureau because data is available for these regions. However, some of these regions are too big. The gravity model is primarily 'distance sensitive', and large regional sizes do not accurately represent where products originate and arrive. We think the model would not perform well if such issues were not carefully considered. Therefore, we use a more granular regional structure and we have decided to use the 48 contiguous state boundaries. This gives us a 48 by 48 matrix with 2,304 origin-destination pairs.

4.2.2 Base Flow *Impedance*

With the availability of base year condition (flow data for code '35'), a reverse derivation procedure is used to determine the (empirical) *impedance* values. We want to calibrate the *deterrence* function that we develop such that the sum of squared errors between the flows determined from the base and the newly developed *deterrence*

function is minimized. This procedure will be discussed in Section 4.2.4. We determine the base *impedance* using

$$F'_{ij} = \frac{T'_{ij}}{P'_i C'_j} \tag{19}$$

where F'_{ij} = *Impedance* value between origin i and destination j for base flow condition

T'_{ij} = Observed flows in tons from region i to region j for base flow condition

P'_i = Production from region i for base flow condition

C'_j = Consumption for region j for base flow condition

4.2.3 Main Components of *Deterrence* Function

The *deterrence* function F_{ij} is a function that reflects the *impedance* of product flow. *Deterrence* functions are typically assumed to be either a linearly or exponentially decreasing function of distance. However, such thought primarily applies to trip distribution. In the inter-regional commodity flows in the United States, the observed distribution is not only the result of the *impedance* function due to distance, but also of the economic activity level. For example, we observe significant flows between large western states, such as California, and large eastern states, such as New York, although they are far away from each other. Therefore, there are other factors involved than just distance.

To achieve better modeling of this pattern, we assume that such strong economic trade level is primarily based on the population of those regions. Also, similar to the *socio-economic* factors in the literature (see section 2.2.1(b)), we introduce a

multiplication factor K_{ij} as a representation for such activity (We discuss the details of determining K_{ij} factors in the following section). The resulting estimated flow T_{ij} between region i and j takes the form:

$$T_{ij} = \alpha P_i C_j F_{ij} = \alpha P_i C_j f(d_{ij}) K_{ij} = \alpha P_i C_j \frac{1}{d_{ij}^n} O_i O_j \left[\frac{1}{R_{ij}} \right] \tag{20}$$

where $f(d_{ij})$ is a function of distance.

4.2.4 Determination of the *deterrence* function, F_{ij}

The first segment of our *deterrence* function $f(d_{ij})$ depends on distance. Specifically, we represent $f(d_{ij})$ as the inverse of distance raised to power of n, which is a parameter that we can modify for better accuracy of the model. Assuming K_{ij} between each origin-destination pair is a constant, we follow a trial and error approach to find the exponent n. The following is the *deterrence* function for origin i and destination j:

$$F_{ij} = f(d_{ij}) K_{ij} = \frac{1}{d_{ij}^n} O_i O_j \left[\frac{1}{R_{ij}} \right] \tag{21}$$

The factor K_{ij} is based on the 1997 statewide population estimates from the U.S. Census Bureau. The estimated population data is the computed number of persons living in each state. It is calculated from a demographic component of change model that incorporates information on natural change (births and deaths) and net migration (net domestic migration and net movement from abroad) that has occurred in each state since the reference data of the 1990 Census.

For each origin-destination pair (i,j), K_{ij} takes the following form

$$K_{ij} = O_i O_j \left(\frac{1}{R_{ij}} \right) \qquad \forall \ i, j \tag{22}$$

where O_i = Population of origin state i,

O_j = Population of destination state j,

R_{ij} = *Proportionality factor* of each origin-destination pair.

The inverse of the *proportionality factor* of total commodity flow for every origin-destination pair is introduced to differentiate the K_{ij} factor between two origin-destination pairs depending on the magnitude of the existing overall commodity flow between the states. X'_{ij} is the total commodity flow between origin state i to destination state j of the base condition, which includes all product types listed in the SCTG codes.

Table 1 shows the breakpoints in the values of R_{ij}.

Table 1. Breakpoints in determining the values of R_{ij}

Total Base Flow of Each Origin-Destination Pair (X'_{ij})	R_{ij}
$X'_{ij} < 500$K tons	5
$500\text{K} \le X'_{ij} < 1500$K tons	4
$1500\text{K} \le X'_{ij} < 2500$K tons	3
$2500\text{K} \le X'_{ij} < 4000$K tons	2
$X'_{ij} \ge 4000$K tons	1

We currently set R_{ij} factor to a value between 1 and 5, depending on the overall base flow. Due to the *proportionality factor*, K_{ij} may not necessarily be the same as K_{ji}. Such differences can add more validity to the model as they take into account the economic interaction between states. Preliminary results show that using this factor decreases the sum of squared errors between the base flows and the calculated flows at the 9th iteration. The least squares procedure is discussed later in this chapter. Finally, the *deterrence* function that we have developed, F_{ij} is as follows:

$$F_{ij} = f(d_{ij})K_{ij} = \frac{1}{d_{ij}^{n}} O_i O_j \left(\frac{1}{R_{ij}} \right) \tag{23}$$

The next step is to calibrate our model by fine-tuning the exponent n in the *deterrence* function such that the *deterrence* function can be refined using the least

square analysis. We search for a value of n that gives us small error. We employ Kanafani's method (1983), which minimizes the sum of squared errors between base flow and calculated flows. The following is the sum of squared errors that we try to minimize by changing n:

$$Min\ S^2 = \sum_{ij} (T'_{ij} - T_{ij})^2 \tag{24}$$

where T'_{ij} = Base flow directly obtained from commodity flow survey for product code '35'

T_{ij} = Estimated flows distributed by the new *deterrence* function at 9th iteration for product code '35'

This calibration process helps us to determine the proper n-value for our *deterrence* function.

4.3 Iterative Procedure in Gravity Modeling

Hamburg, Kaiser and Lathrop (1983) introduced three concepts as the basis of the iterative procedure in gravity modeling: attraction factors (A_{ij}), accessibility indices (I_i), and production indices (U_i). An iterative procedure is employed in gravity model to ensure that the production and consumption capacity is satisfied to within 5 to 10 percent of the estimated value. This iterative procedure will be undertaken for all rows and columns of the 48 by 48 matrix of the gravity model.

4.3.1 Attraction Factors, Accessibility Indices and Production Indices

Knowing the *deterrence* function F_{ij}, production capacities P_i, and consumption capacities C_j we determine the attraction factor (A_{ij}) for every pair of regions i and j:

$$A_{ij} = C_j F_{ij} \tag{25}$$

The accessibility index (I_i) for production region i is

$$I_i = \sum_{j=1}^{n} C_j F_{ij} \tag{26}$$

The production index (U_i) for region i is

$$U_i = \frac{P_i}{I_i} \tag{27}$$

The matrix of the model up to 1st iteration takes the form of Table 2.

Table 2. A representative table for the first iteration of the gravity model

	Consumption Capacity, C_j			Accessibility Index, I_i $I_i = \sum_{j=1}^{n} C_j F_{ij}$	Production Index, U_i $U_i = \frac{P_i}{I_i}$
	j_1	$j2$	……..jn^{th}		
i_1	$A_{ij}=C_{j1}F_{i1,j1}$	$C_{j2}F_{i1,j2}$			
i_2	$C_{j1}F_{i2,j1}$	$C_{j2}F_{i1,j2}$			
Production Capacity, P_i : : in^{th}					
Total Commodity to Region j ; ΣT_{ij}	ΣT_{ij}				
% Deviation for Iteration 0	$1-(\Sigma Tij / C_{j1})$				
Adjusting Factor for Iteration 1	$C_{j1}/\Sigma Tij$				
Total Commodity to Region j ; ΣT_{ij}					
% Deviation for Iteration 1					
Adjusting Factor for Iteration 2					

Table 3 illustrates the implementation of the formulation process in Table 2. We perform the iterative procedure up to the 2nd iteration. We present an example problem that involves inter-regional flows between 3 regions. The production, consumption and *impedance* values between the regions are given.

Production Capacity	**Production**	**Consumption**
Region A	10	15
Region B	20	20
Region C	30	25

Impedance* Values Between Origin Region *i* and Destination Region *j

	Region A	**Region B**	**Region C**
Region A	6	8	9
Region B	5	6	8
Region C	6	8	10

Table 3. A sample problem of inter-regional flows between 3 regions.

$$I_i = \sum_{j=1}^{n} C_j F_{ij}$$

$$U_i = \frac{P_i}{I_i}$$

			Consumption				
			Region A	**Region B**	**Region C**	**Accessibility Index, I_i**	**Production Index, U_i**
	Capacity		**15**	**20**	**25**		
Region A	**10**	***Impedance***	6	8	9		
		Attraction Factor; A_{ij}	6x15=**90**	8x20=**160**	9x25=**225**	**475.00**	**0.02**
		Flow @ Iteration 1	0.02(90)=**1.8**	0.02(160)=**3.2**	0.02(225)=**4.5**		
		Attraction Factor; A_{ij}	1.19x1.8=**2.14**	1.18x3.2=**3.78**	0.74x4.5=**3.33**	**9.25**	**1.08**
		Flow @ Iteration 2	1.08x2.14=**2.31**	1.08x3.78=**4.08**	1.08x3.33=**3.60**		
Region B	**20**	***Impedance***	5	6	8		
		Attraction Factor; A_{ij}	5x15=**75**	6x15=**90**	8x25=**200**	**365.00**	**0.06**
		Flow @ Iteration 1	0.06(75)=**4.5**	0.06(90)=**5.4**	0.06(200)=**12**		
		Attraction Factor; A_{ij}	1.19x4.5=**5.36**	1.18x5.4=**6.37**	0.74x12=**8.88**	**20.61**	**0.97**
		Flow @ Iteration 2	0.97x5.36=**5.20**	0.97x6.37=**6.18**	0.97x8.88=**8.61**		
Region C	**30**	***Impedance***	6	8	10		
		Attraction Factor; A_{ij}	6x15=**90**	8x15=**120**	10x25=**250**	**460.00**	**0.07**
		Flow @ Iteration 1	0.07(90)=**6.3**	0.07(120)=**8.4**	0.07(1575)=**17.5**		
		Attraction Factor; A_{ij}	1.19x6.3=**7.50**	1.18x8.4=**9.91**	0.74x17.5=**12.95**	**30.36**	**0.99**
		Flow @ Iteration 2	0.99x7.5=**7.43**	0.99x9.91=**9.81**	0.99x12.95=**12.82**		
		Total Commodity to Region j ; ΣTij	**12.6**	**17**	**34**		
		% Deviation for Iteration 1	1-(12.6/15)=**0.16**	1-(17/20)=**0.15**	1-(34/25)= **-0.36**		
		Adjusting Factor for Iteration 2	(15/12.6)=**1.19**	(20/17)=**1.18**	(25/34)=**0.74**		
		Total Commodity to Region j ; ΣTij	**14.94**	**20.07**	**25.03**		
		% Deviation for Iteration 2	1-(14.94/15)=**0.004**	1-(20.07/20)= **-0.0035**	1-(25.03/25)= **-0.0012**		
		Adjusting Factor for Iteration 3	(15/14.94)=**1.00**	(20/20.07)=**1.00**	(25/25.03)=**1.00**		

In the first iteration, the commodity flow from region i to region j is

$$T_{ij} = U_i C_j F_{ij} \tag{28}$$

This process automatically generates flows that satisfy the production capacity constraints of our doubly-constrained gravity model. The next step is to correct flows for the consumption capacity of each region. The adjusting factor for consumption in each region j is

$$\frac{C_j}{\sum_{i=1}^{n} T_{ij}} \tag{29}$$

The adjusting factor of each consumption region is multiplied by the attraction factor of the corresponding origin-destination pair to determine the new attraction factor for the next iteration of the procedure. We repeat the same procedure of calculating attraction factors, accessibility indices, production indices, and finally new adjusting factors until the percent deviation from the actual consumption value falls within 10%. The resulting T_{ij} for each individual member of the matrix is the flow for each origin-destination pair.

We provide a summary of the iterative procedure as follows:

1) Calculate the attraction factor of each origin-destination pair.
2) Calculate the accessibility index for each origin/production region i (summation of all attraction factors).
3) Calculate the production index for each production region i (divide the production capacity of each region by its accessibility index).

4) Compute the initial flow for each origin-destination (production-consumption) pair by multiplying the production index of the origin with the attraction factor of the corresponding destination.
5) Calculate the adjusting factor for the consumption capacity by dividing the consumption capacity of each destination region *j* by the summation of all flows coming into region *j*. If the adjusting factor is small (close to 1), stop and report the latest calculated flows. Otherwise, multiply the adjusting factor for each consumption region with the corresponding attraction factor to determine the new attraction factor for the next iteration and return to Step 2.

4.4 *Extreme Impedance*

E-commerce tools and technologies are increasingly bringing buyers and sellers, and suppliers and customers, closer. The three basic spatial separation constraints (distance, time, and shipping cost) are beginning to impact buyers less. Consumers are no longer restricted to buy things from local stores. Products can be delivered overnight or in a very short amount of time. The difference in shipping costs among the regions is getting smaller. For example, according to the shipping rate provided by UPS, shipping a 100 pound parcel from Scranton, Pennsylvania to Conway, New Hampshire costs approximately $46, whereas shipping the same parcel to San Francisco, California (five times the distance), costs only $54. (www.ups.com)

Due to these observations, we introduce the following *extreme impedance* function to generate multiple scenarios for simulating the effect of diminishing effects of distance on the flows of goods. The *extreme impedance* function is a constant value, i.e.,

every origin-destination pair has the same value between the regions. We then have two *deterrence* functions, the *extreme* function (constant), and the original *deterrence* function in equation (23). We believe that as e-commerce continues to grow, the *deterrence* function will fall somewhere between these two functions, and it will move closer to the extreme case over time. Therefore, we employ the use of a smoothing method to develop an intermediate *deterrence* function. We estimate the intermediate *deterrence* function that lies between the current *deterrence* and the *extreme deterrence* using

$$f=(1-\lambda)(\text{distance based-deterrence function})+\lambda(\text{extreme impedance function}) \qquad (30)$$

where λ is the a smoothing constant between 0 and 1. λ is equal to 1 when we are at the extreme condition, and λ is equal to 0 when we are at the traditional base condition.

4.5 Determining the Average Miles

One way to illustrate the future impact of e-commerce on inter-regional flows is to compare the average miles of the base flow condition to that of the inter-regional flows determined from the projected future production and consumption capacity. Average miles traveled by each ton of the current and future flows can be determined by using the following function.

$$\text{Average Miles} = (\sum\text{ton-miles} / \sum\text{ton}) \qquad (31)$$

According to the U.S. Census Bureau, ton-mile is simply the shipment weight times the mileage for a shipment. Respondents of the commodity flow survey reported shipment weight in pounds; mileage was calculated as the distance between the shipment origin zip code and destination zip code. Aggregated pound-miles were converted to ton-

miles. The summation of ton-miles for every origin-destination pair divided by the total tonnage shipped results in average distance traveled by one ton of products for all the 48 contiguous states. The significant increase in average distance traveled by products for future projected flows is one indicator that shows the diminishing effect of distance due to e-commerce on inter-regional transportation flows.

4.6 Determining the Appropriate λ Value for Smoothing

We now determine the appropriate λ value for smoothing between our distance based deterrence function and the extreme impedance function. This process involves a benchmarking procedure where an assumption is made on the average expected increase in average miles with the effect of e-commerce on future flows. We seek to determine the λ value for the year 2005.

The U.S. Census Bureau reports that online purchases accounted for 11 percent of all cost of materials at manufacturing plants in 1999. Also, 12 percent of all manufacturing shipments were for orders accepted online. E-commerce transactions were significant in the machinery sector (12 percent), Computer and Electronic Products sector (12 percent), and Electrical Equipment sector (10 percent). (U.S. Census Bureau, 2001). A survey conducted by U.S. Census Bureau of 38,985 manufacturing plants shows that 16 percent of reporting plants have engaged in both e-shipments (shipping products to customers that have made their purchases online) and e-purchases (making online purchases). The result of the survey is presented in Table 4.

Table 4. Status of E-commerce Engagement for Manufacturing Plants (U.S. Census Bureau, 2001)

	Status of E-Shipments			
Status of E-purchases	**All plants**	**Make E-shipments**	**Do Not Make E-shipments**	**Unknown**
All Plants	38,985	12,069	26,462	454
Make E-Purchases	13,233	6,063	7,061	109
Do Not Make E-Purchases	25,237	5,901	19,203	133
Unknown	515	105	198	212

From this latest report by U.S. Census Bureau, we know that e-commerce is beginning to play a major role in the United States economy. Though this report does not cover the entire U.S. economy, it surveys the North American Industry Classification System (NAICS) industries that accounted for approximately 70 percent of economic activity measured in the 1997 Economic Census. We are therefore looking at a major portion of the U.S. economy that has gone online.

With this new information released by U.S. Census Bureau, we will conservatively assume the share of e-commerce flows for 2005 will be 15%. For the remaining 85% of the products, we assume the flows were created under the traditional economy.

In estimating the average miles of the product in year 2005, we evaluate the 1993 and 1997 commodity flow survey. Most products have undergone an average increase of 5% in average miles within the 4 years period (see Figure 1). Therefore, for year 2005, we assume the average mile will increase by 10% since 1997. We apply this percent increase to the 85% of the products, which flow under traditional economic methods. As for the 15% e-commerce products, we assume the average mile will increase by 20%

from 1997 to 2005. The resulting average mile computed from the two shares of economy is reported in the next chapter.

Based on the estimated average mileage transported by each ton of products for 2005, the λ value is determined. A 20% increase in production and consumption capacities is assumed for all states. Such assumption is based on the average increase of production capacity of 14% every 4 years. Since the projection is made from 1997 to 2005, we conservatively assume the production and consumption capacities will grow by 20% in those 8 years. An interpolation process is employed to determine the exact λ value that gives us the estimated average miles of 2005. Since the λ values may vary across product lines, the process of determining λ value will be performed each time a different product is considered.

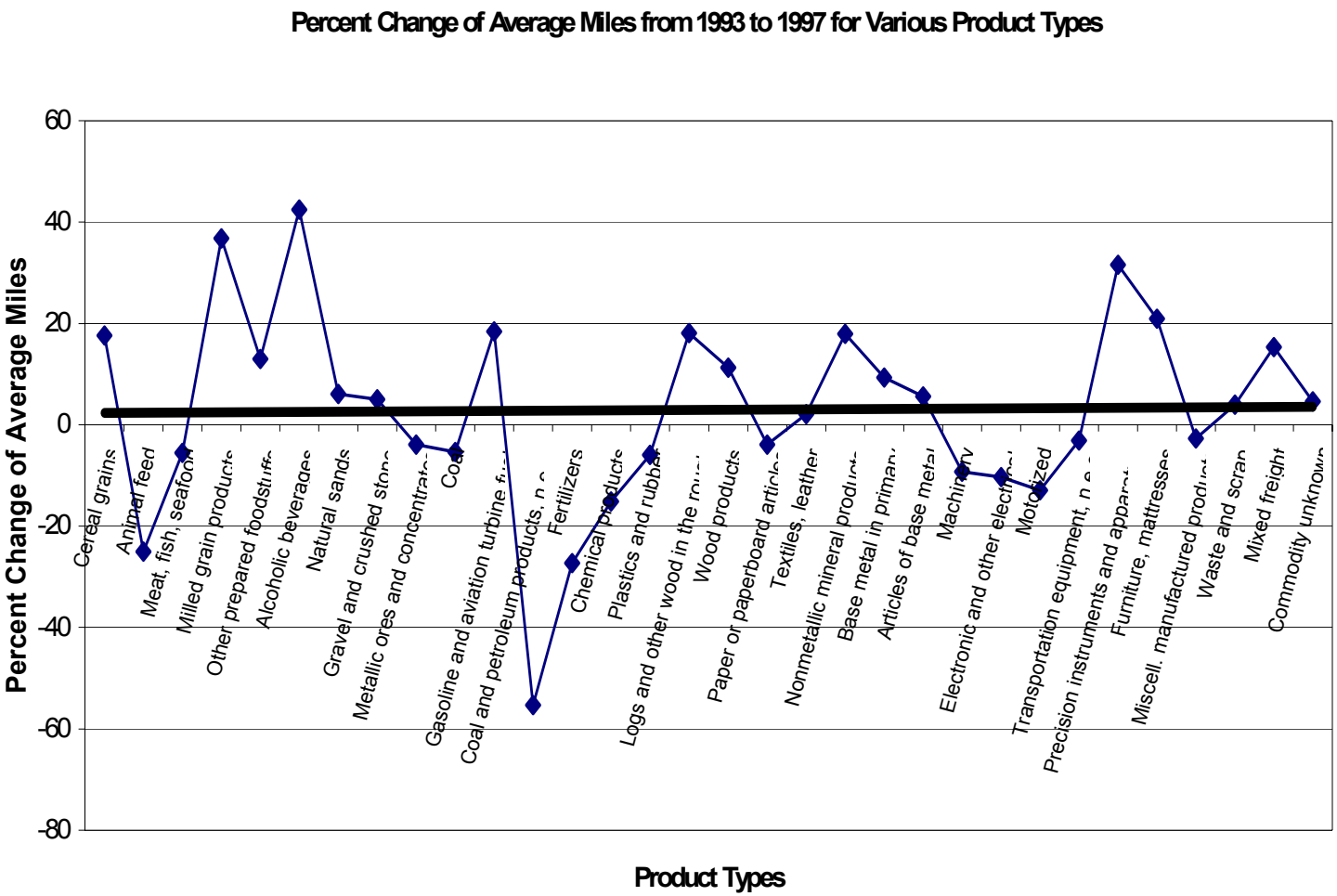


Figure 1. Graph of Percent Change of Average Miles for Various Product Types from 1993 to 1997

4.7 Mode Assignment

We perform mode assignment on the total (all SCTG codes) production and consumption capacities for every state. Mode assignment mainly assigns the distributed flows to different modes of transportation. We assume the production and consumption capacities have increased by 20% from 1997 to 2005. Also, the assumption on the percent share of e-commerce (15%) remains the same.

The result of mode assignment that we will obtain in this procedure will not ultimately represent the percent share of different mode usage in 2005 for United States. However, we perform this step to illustrate the impact of diminishing effect of distance on the choice of mode used. Though we realize that, other than distance, shippers make their selections on their choice of modes based upon variety of reasons. In fact, Gray (1980), is supported by other authors in concluding that there is no specific need to examine organizational mode selection in freight transport. Lambert (1993) found that "lowest rates" came only 40th after testing the importance of over hundred and fifty selection factors. In fact, various subjective factors that affect mode choice, such as reliability, trust and service level play a much important role (Pisharodi, 1991). Many studies have come up with different contrasting opinions, which make mode assignment a much greater challenge (Gray, 1982). Therefore, the output of mode assignment in this research is not intended for forecasting, but just to illustrate the mode usage distribution if distance is a significant influence of mode usage.

Before distributing the flows, we determine the proper n and λ values to be used. As discussed before, n and λ may vary across product lines. Therefore, the average n and λ values determined from the individual n and λ of product codes '35', '30' and '6' will

be used to distribute the flows. These three groups of products represent three distinct types of products in the commodity flow survey. Therefore, taking the average of the individual n and λ values of these three product types will be a good approach.

Having determined the n and λ values, we then distribute the projected production capacity to every origin-destination pair. Every origin-destination flow is considered separately for mode assignment. We perform the mode assignment at a very high level in this research. We assign the distributed flows to different transportation modes based on the average distance between the origin-destination pair. Table 5 gives us the percent share of mode choice under different distance categories. The values shown in these tables are average percent shares across all states for all shipment types. A graph of flows versus modes used in 1997 and 2005 (Figure 13) will be plotted to observe the change in mode usage due to e-commerce. This graph will be shown in Chapter 5. The next chapter validates the approach we have taken and presents the results from this research.

Table 5. Percent Usage of Different Transportation Modes Relative to Distance in 1997 U.S. Commodity Flow

	Less than 50 miles		**50 to 99 miles**		**100 to 249 miles**	
Mode	**Tons(000)**	**Percentage**	**Tons(000)**	**Percentage**	**Tons(000)**	**Percentage**
All modes	***6444454***		***1079841***		***1311278***	
Single modes	***6086713***		***1053694***		***1238913***	
Truck	5212913	80.90%	866735	80.75%	770562	58.76%
Rail	254985	3.96%	107608	10.03%	285232	21.75%
Water	179449	2.79%	53806	5.01%	109425	8.34%
Air (includes truck and air)	0	0.00%	202	0.02%	683	0.05%
Pipeline	439350	6.82%	25342	2.36%	73011	5.57%
Parcel, US Postal Service or courier	4307	0.07%	1704	0.16%	3546	0.27%
Multiple Modes(truck&rail,truck& water,rail&water,other multiple modes	352315	5.47%	17981	1.68%	68818	5.25%
Total	6443319	100%	1073378	100%	1311277	100%

Table 5. Continues.

	250 to 499 miles Tons(000)	Percentage	500 to 749 miles Tons(000)	Percentage	750 to 999 miles Tons(000)	Percentage
All modes	***905504***		***541782***		***383327***	
Single modes	***844909***		***501869***		***343183***	
Truck	415852	46.05%	191915	35.79%	103369	27.03%
Rail	322529	35.71%	213720	39.86%	173661	45.41%
Water	65618	7.27%	84391	15.74%	51822	13.55%
Air (includes truck and air)	622	0.07%	485	0.09%	455	0.12%
Pipeline	40288	4.46%	11359	2.12%	13876	3.63%
Parcel, US Postal Service or courier	3611	0.40%	2869	0.54%	2257	0.59%
Multiple Modes(truck&rail,truck& water,rail&water,other multiple modes	54566	6.04%	31412	5.86%	36999	9.67%
Total	903086	100.00%	536151	100.00%	382439	100.00%

Table 5. Continues.

	1,000 to 1,499 miles		1,500 to 1,999 miles		2,000 miles or more	
	Tons(000)	**Percentage**	**Tons(000)**	**Percentage**	**Tons(000)**	**Percentage**
All modes	***302918***		***77130***		***43499***	
Single modes	***269791***		***65872***		***31594***	
Truck	79277	26.49%	37500	50.51%	22552	51.88%
Rail	159207	53.20%	25144	33.87%	7731	17.79%
Water	15661	5.23%	0	0.00%	358	0.82%
Air (includes truck and air)	684	0.23%	374	0.50%	954	2.19%
Pipeline	14961	5.00%	0	0.00%	0	0.00%
Parcel, US Postal Service or courier	2270	0.76%	1542	2.08%	1585	3.65%
Multiple Modes(truck&rail,truck& water,rail&water,other multiple modes	27196	9.09%	9680	13.04%	10289	23.67%
Total	299256	100.00%	74240	100.00%	43469	100.00%

5. RESULTS

5.1 Initial Observation and Validation Process

We provide some preliminary results and summarize our initial observations to support the assumptions and conjectures of our research.

One of the initial assumptions of this research is to model the flow based on SCTG Code '35' for electronic and other electrical equipment, and components, and office equipment. Latest data released by U.S. Census Bureau shows that electrical and electronic equipment is one of the leading manufacturing products in e-commerce. (U.S. Census Bureau, 2001)

Using the flow data for product Code '35', we hereby validate our steps in determining the *deterrence* function. The development of the *deterrence* function is determined by taking the least square of the base flow and calculated flows distributed at 9th iteration using the formulated function. The initial form of the *deterrence* function is $1/d^2$. The *impedance* values of base flow condition that contribute to the distributed flows are plotted along with the initial form of deterrence function $1/d^2$ in Figure 2. To improve the *deterrence* function, population factors are being introduced. The results show that population values of production and consumption states as well as the *proportionality factor*s improve the *deterrence* function even better as it gives a lower summation of least squared errors (Figure 3). These observations prove that both population and *proportionality factor*s have a significant impact on the performance of the *deterrence* function. Population factor is expected to be a useful parameter because of its influence on the distribution of commercial (.com) web sites in the United States. In addition, the *proportionality factor*s capture the base effect, and help tune the *deterrence* function to

the base condition. Such effect will add more realism to the *deterrence* function, and disallow the calculated flow to deviate far from the base condition.

Based on our initial assumptions, as e-commerce continues to grow, the diminishing importance of distance will increase the proportion of flows between more and more distant regions. This creates an opposite effect on flows that travel on shorter distance because increasing long distance flows will "steal" part of the flows from short-distance flows. To show this behavior, we determine several versions of the proposed smoothing-based *deterrence* (using different λ values). As λ increases (from 0 to 1), the effect of distance diminishes and the *deterrence* function becomes closer to the *extreme* case. The observed *impedance* values are being inserted into the gravity model to obtain the distributed flows at 9th iterations. Based on our analysis, flows from 9th iteration are used because at this stage, the deviation of all states' calculated consumption value from its desired consumption value is at an average of 10%. Therefore, the distributed flows are satisfying the terminating condition. Figure 4 shows the graph of flows from 4 *deterrence* functions with different λ values plotted over distance.

Next, we provide the observation on the distributed flows by taking the summations of flows for two ranges of distances calculated for different λ values. The first range with 500 miles or less represents short distance, whereas the range of greater than 2000 miles represents the long distance. Their proportions from total flows are collected and compared over different values of λ. Table 6 is a summary table of the values obtained.

Table 6. Summary Table for Two Ranges of Distance

λ	Total Flow	Total Flow for distance <500	Proportion from Total	Total Flow for distance >2000	Proportion from Total
0.125	39435.00	11335.90	0.287	6179.59	0.157
0.250	39435.00	9964.96	0.253	7053.85	0.179
0.375	39435.00	9343.06	0.237	7517.24	0.191
0.500	39435.00	8974.48	0.228	8974.48	0.228

Based on the summary presented in Table 6, we see that as λ increases the effect of distance diminishes and more and more long-distance flows take place while the opposite is true for short-distance flows. Assuming that the growing e-commerce will change the inherent *deterrence* function in such a way that it will move from current to the *extreme* case, we can see that long-distance regions will "steal business from neighbor regions." Consumers and businesses will soon begin to make purchases over the Internet without any concerns for the locations of the sellers.

All the preliminary results presented here help support the validity of the approach we took up to this point. It is very important that the *deterrence* function be calibrated to the base condition, as future *deterrence* functions will primarily be based on this. Though changing the value of λ in exponential smoothing will adjust the diminishing importance of distance, we will perform a sensitivity analysis on the flows distributed using different λ values to come up with a λ value that will more likely represent the conditions of e-commerce in the near future. Also, the newly developed *deterrence* function will be inserted into the gravity model for projected production and

consumption values (for 2005) in order to estimate the flows of each origin-destination pair. The increase in ton-miles will be estimated.

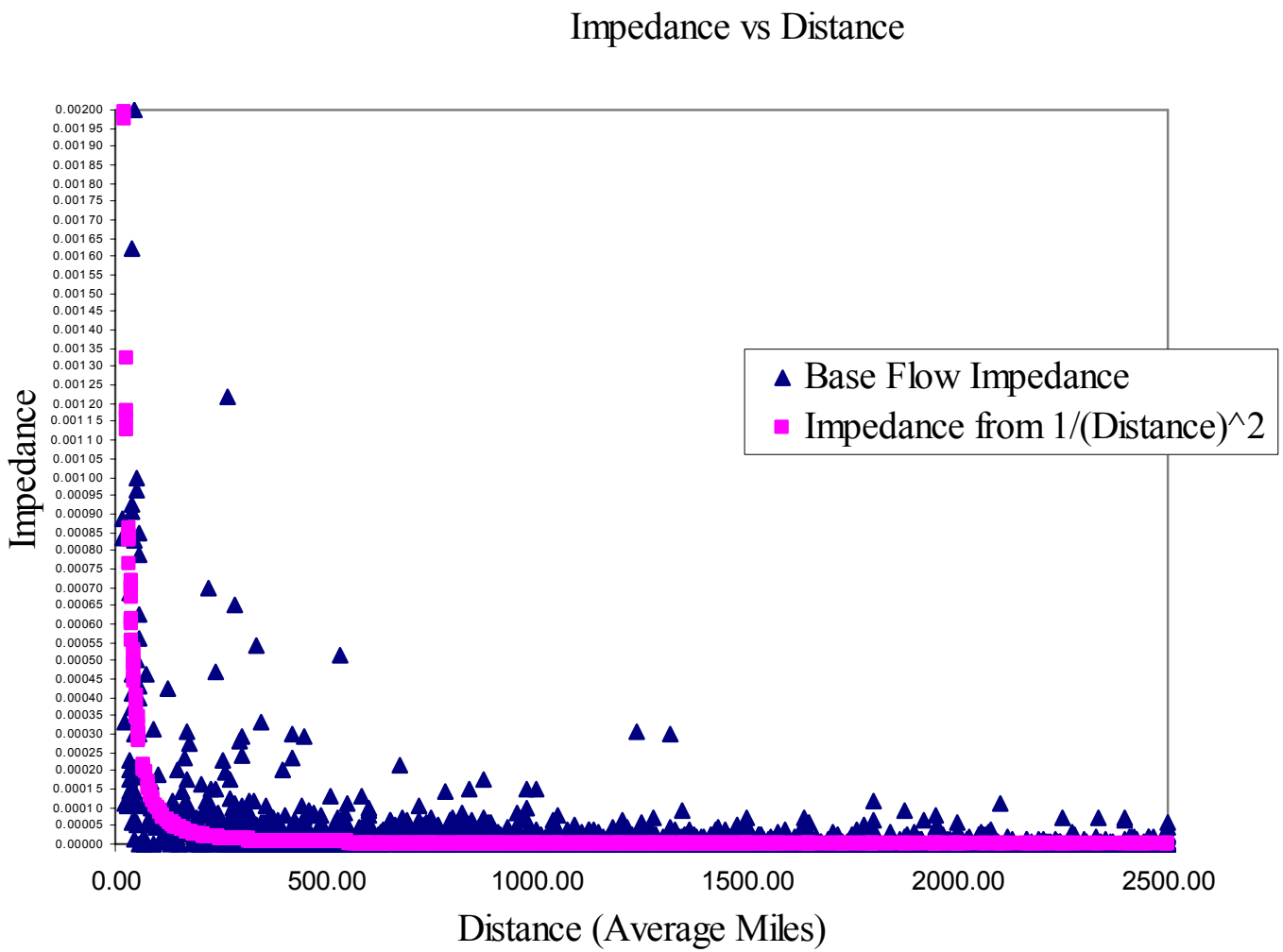


Figure 2. Graph of Base Flow *Impedance* and *Impedance* from (1/(Distance)^2) versus Distance

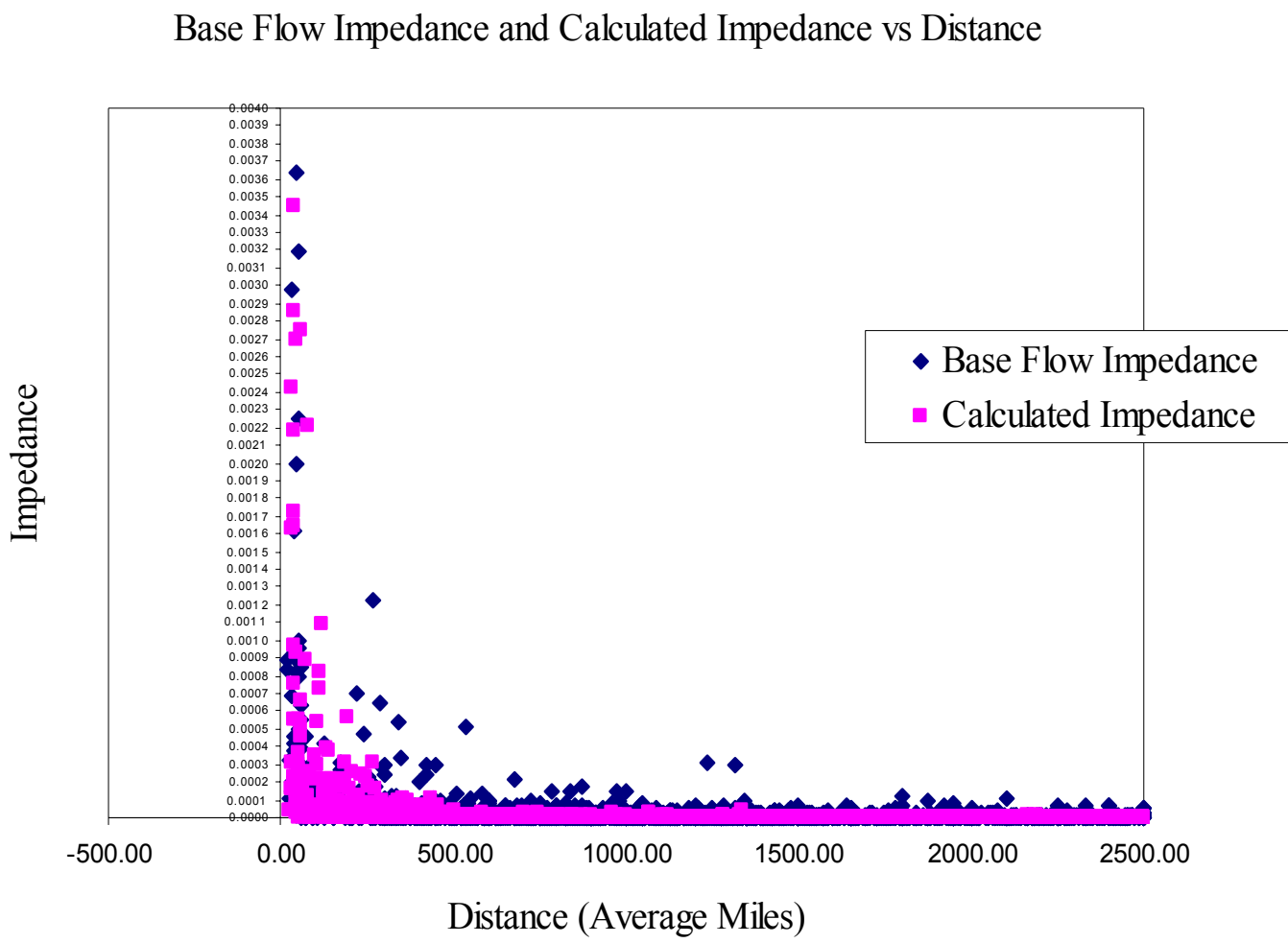


Figure 3. Graph of Base Flow *Impedance* and Calculated *Impedance* (from the developed deterrence functio

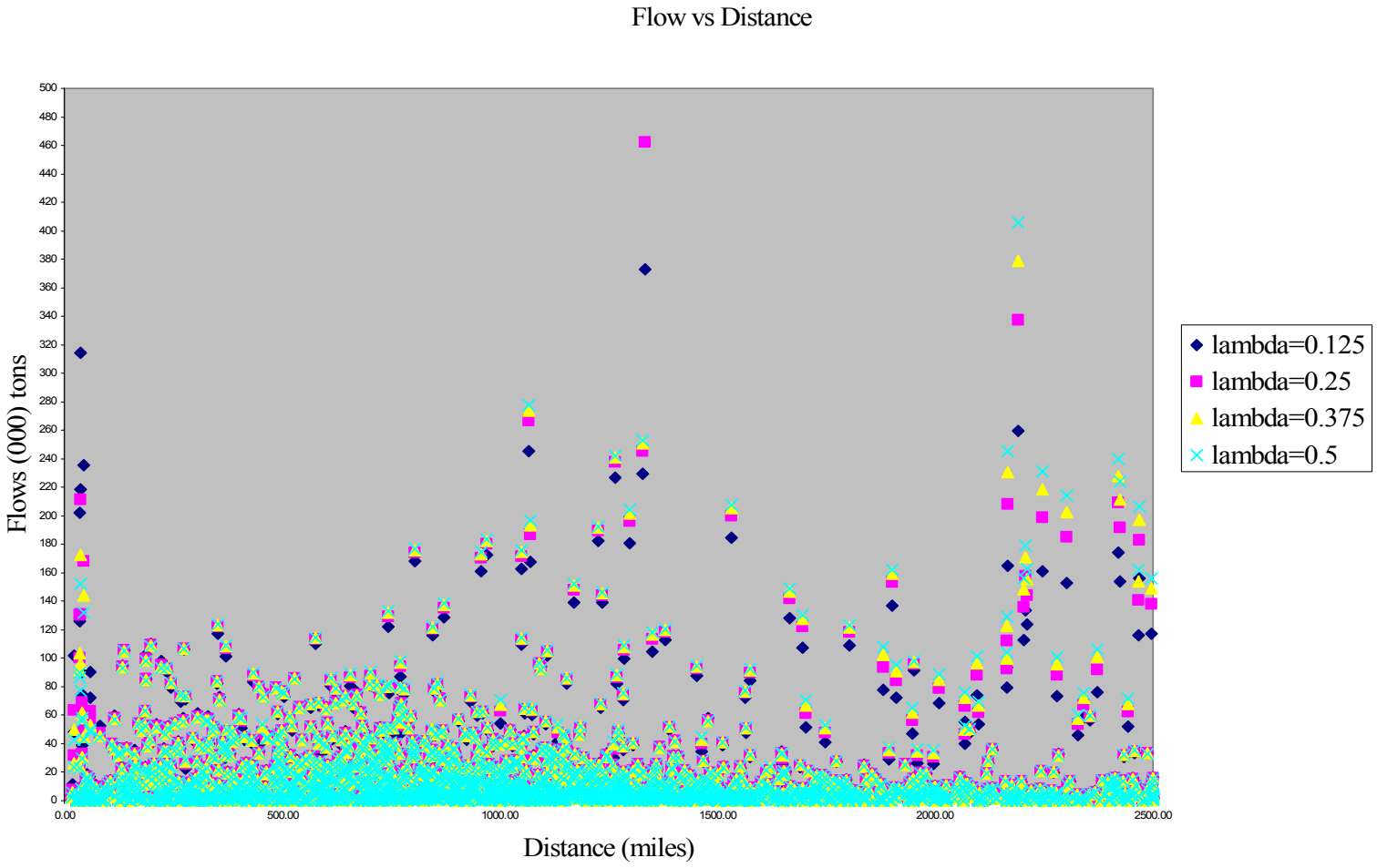


Figure 4. Graph of Estimated Flows for Different Values Plotted Over Distance

To ensure that the developed model is not biased toward product code ‘35’ and can also be easily calibrated and applied to other product types, we test the model with two other sets of product flow data. The *deterrence* function of the model is evaluated to see if it can be calibrated to the base flow condition of these two-product types. Flow data for product code SCTG 6 for milled grain products and product code SCTG 30 for textiles and leather have been selected for this evaluation. These two products are very distinct product types as compared to product code '35', and they may very well be affected differently by e-commerce.

During this evaluation process, as expected, the *impedance* values vary across the product types. However, the variability is not significant (see Figure 5). Note that the *impedance* values of these three product types generally follow the same pattern. Graphs of *impedance* of product code '6' and *impedance* of product code '30' are plotted over distance in Figure 6 and Figure 7, together with impedance values from the developed deterrence function, respectively. Figure 6 and Figure 7 are plotted prior to calibration process. Observations from the graphs show that our newly developed *deterrence* function can be applied in both sets of data. The *impedance* obtained from the new *deterrence* function is seen to be capable of capturing the decreasing effect for product code '6' and product code '30'. Calibration process will be required to reduce the sum of squared errors and better improve the newly developed *deterrence* function to fit the two product types.

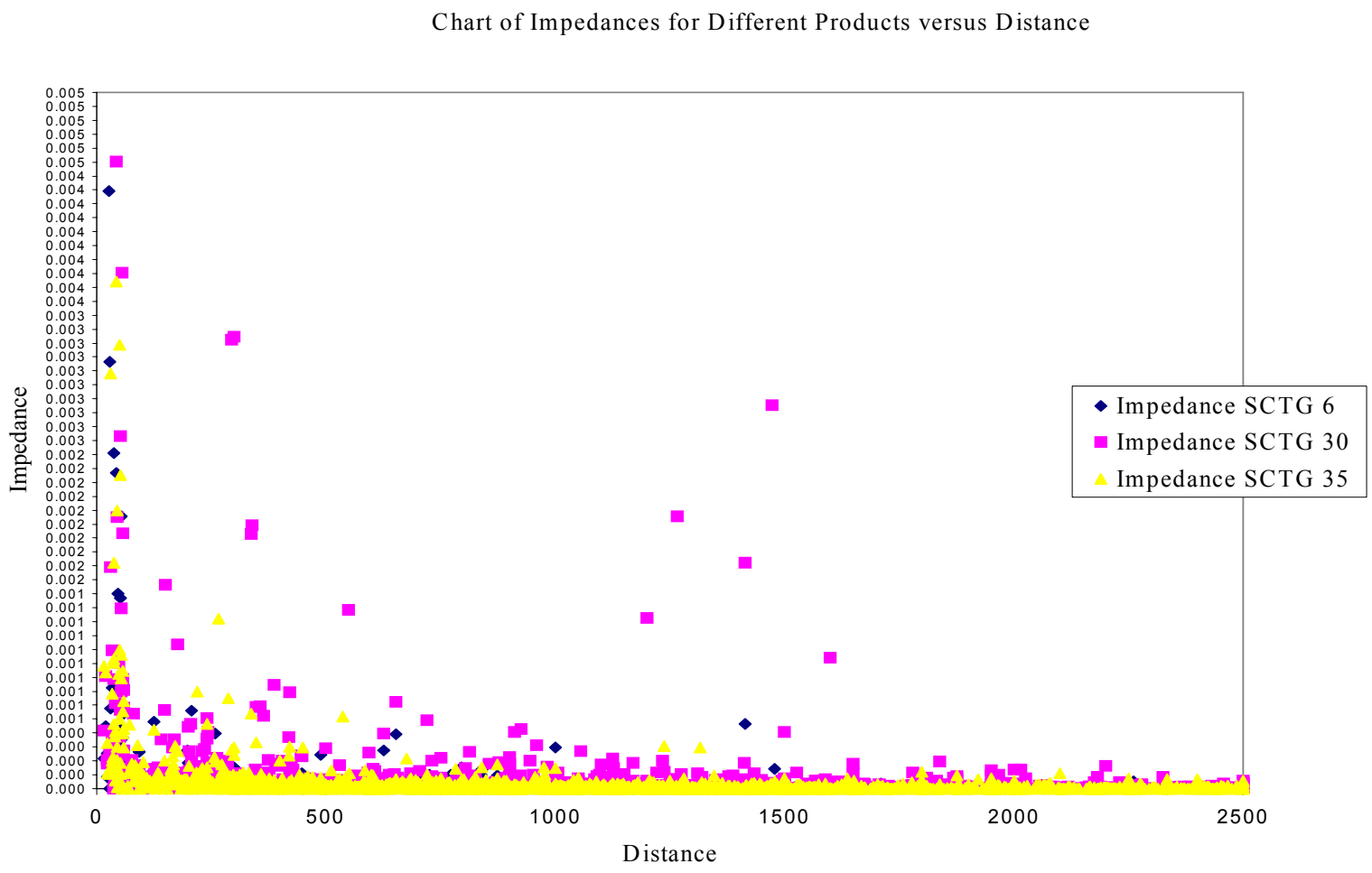


Figure 5. Graph of *Impedance* versus Distance for Product Code '6', '30' and '35

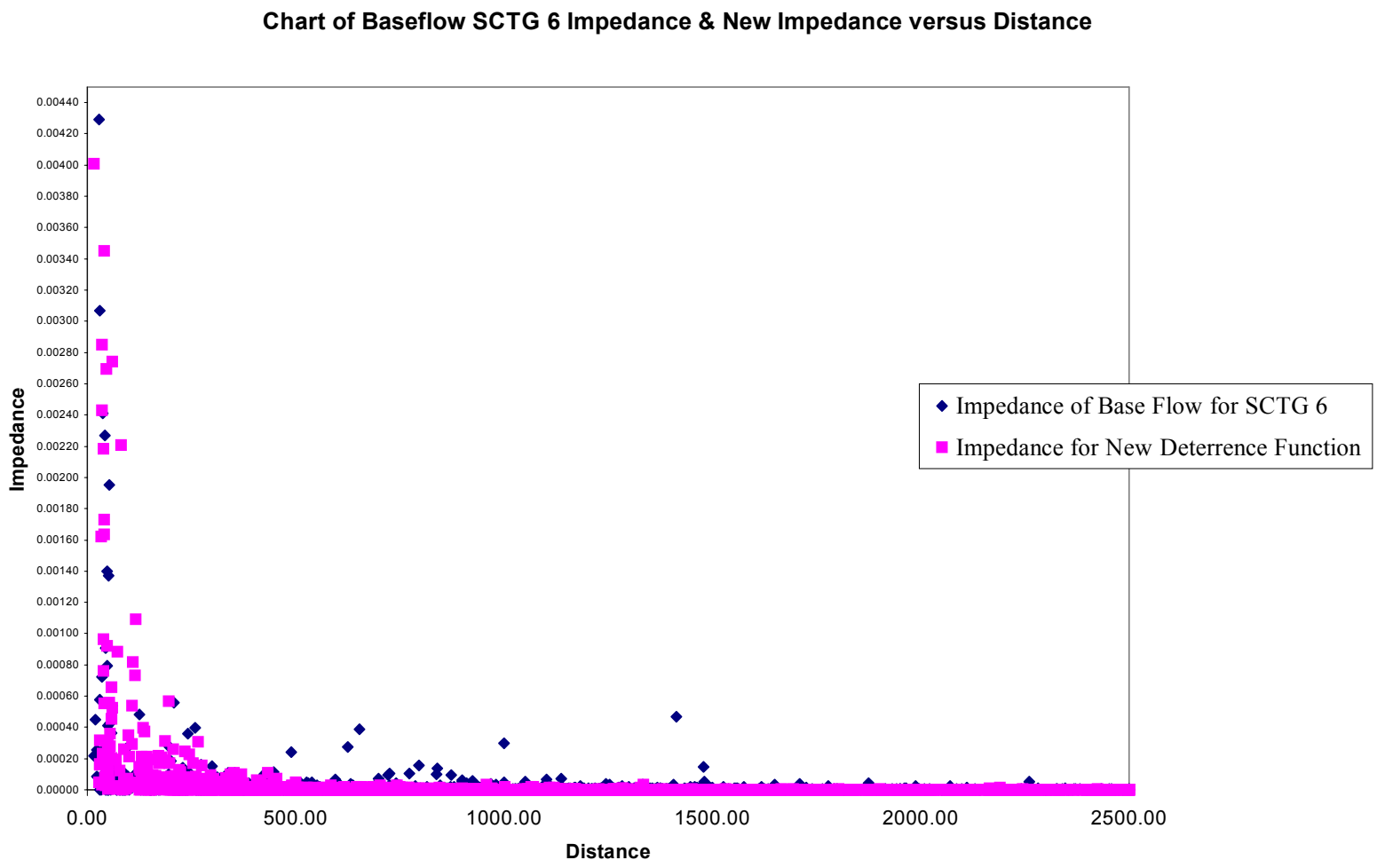


Figure 6. Graph of *Impedance* for Product Code '6' and *Impedance* from New *Deterrence* Function versus Distance

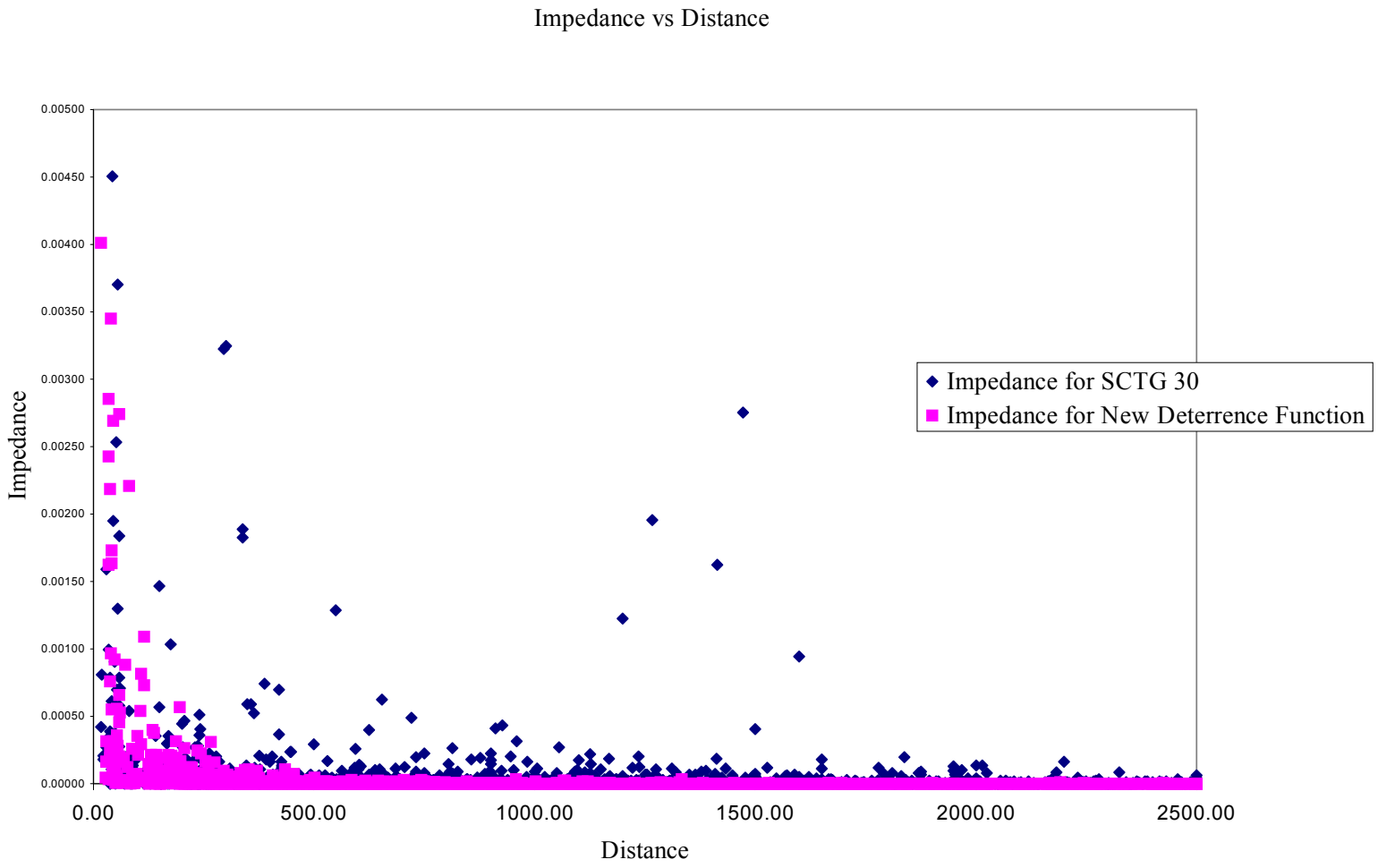


Figure 7. Graph of *Impedance* for Product Code '30' and *Impedance* from New *Deterrence* Function versus Distance

5.2 Selecting the Appropriate *n* and λ Values

Finding the appropriate *n* value is primarily based on the least square method discussed in Section 4.2.4. Also, we estimate the λ value for exponential smoothing method in order to determine the future form of *deterrence* function, which takes into consideration the effect of e-commerce.

As discussed before, testing the model with different sets of commodity flow survey data (Codes '35', '30' and '6') will more likely give us *n* and λ values that will differ from one product to another. Since this model is primarily developed for future modeling purposes, which will represent flows of various product types, we will be using the average *n* and λ values in the model for all the three products for forecasting purposes. For every product type, we select the *n* value that gives us the least squared error, and we select the λ value that gives us the year 2005 estimated average miles that takes into consideration 20% growth in average miles for 15% of the products, and 10% growth in average miles for the remaining 85% of the products. The estimated resulting intermediate average mile is the average mile that will be used as the benchmark to determine *n* and λ values. Values for the intermediate average miles are shown in Table 7. A graphical plot of the data in Table 7 for different product types is shown in Figures 8-10 for product code '35','30' and '6', respectively. The n and λ values that are determined from these benchmark intermediate average miles are shown in Table 8. Note that the final average values for n and λ are as follows.

Average n = (0.5 + 0.75 + 0.75) / 3 = 0.6667

Average λ = (0.525 + 0.995 + 0.067) / 3 = 0.528

Table 7. The Estimated Average Miles for 2005 Used in Determining n and λ Values.

SCTG 35

Year	Average Miles (Traditional)	Average Miles (E-Commerce)	Intermediate Average Miles
1997	640	640	640
2001	660	792	679.8
2005	680	974.4	724.16

SCTG 30

Year	Average Miles (Traditional)	Average Miles (E-Commerce)	Intermediate Average Miles
1997	912	912	912
2001	932	1118.4	959.96
2005	952	1366.08	1014.112

SCTG 6

Year	Average Miles (Traditional)	Average Miles (E-Commerce)	Intermediate Average Miles
1997	452	452	452
2001	472	566.4	486.16
2005	492	703.68	523.75

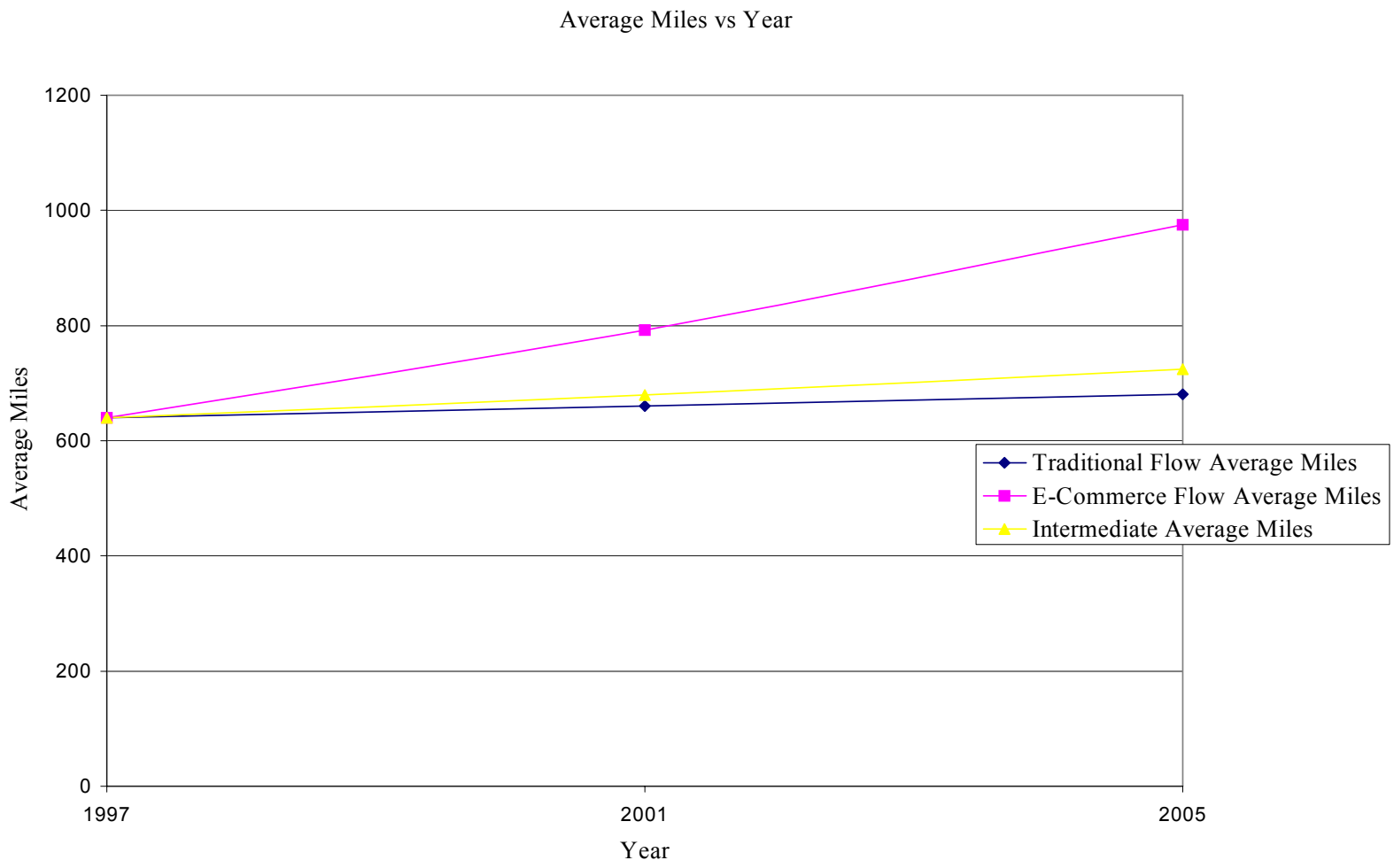


Figure 8. Graph of Average Miles versus Year for Product Code '35'

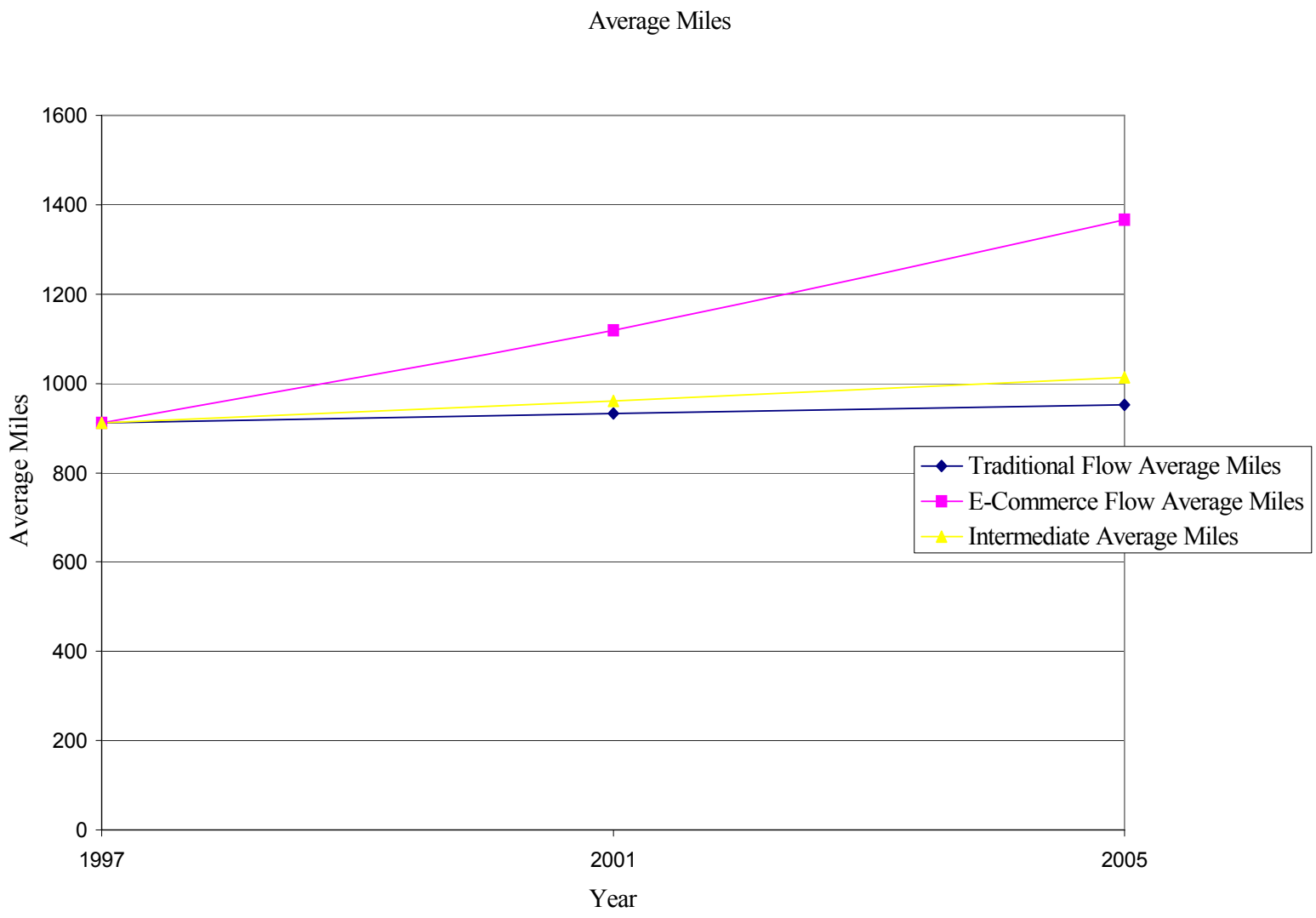


Figure 9. Graph of Average Miles versus Year for Product Code '30'

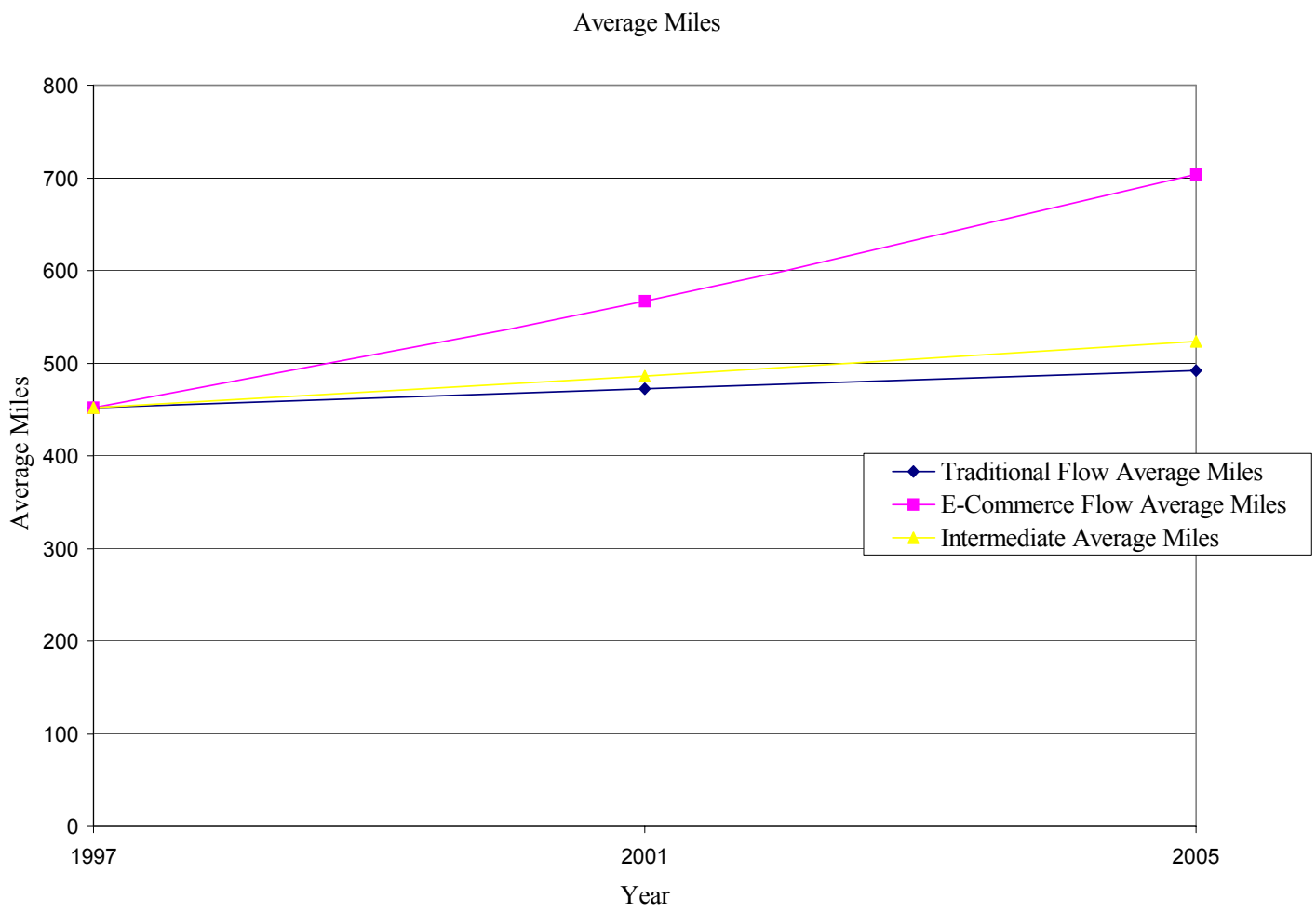


Figure 10. Graph of Average Miles versus Year for Product Code '6'

Table 8. Values of n and λ Obtained for Different Product Types.

Product Code '35'

n	Flow Least Square
0.25	3,443,232.54
0.50	***1,763,652.53***
0.75	2,799,210.08
1.00	5,834,709.66
1.25	9,588,275.87
1.50	13,134,223.12
1.75	16,219,200.91
2.00	18,782,414.71

Product Code '35'

λ	Average Miles
0.500	715
0.525	***724***
0.550	735
0.575	745
0.600	756

Product Code '30'

n	Flow Least Square
0.25	12,037,682.84
0.50	4,406,906.77
0.75	***1,970,933.80***
1.00	3,796,608.50
1.25	7,930,851.80
1.50	13,005,731.37
1.75	18,323,976.54
2.00	23,658,138.75

Product Code '30'

λ	Average Miles
0.9940	1,012
0.9945	***1,014***
0.9950	1,016

Table 8. "Continues"

Product Code '6'

n	Flow Least Square
0.25	70,161,378.57
0.50	52,440,376.90
0.75	***44,882,946.09***
1.00	45,923,596.87
1.25	52,073,066.11
1.50	60,305,219.91
1.75	68,779,832.10
2.00	76,813,538.13

Product Code '6'

λ	Average Miles
0.025	487
0.050	510
0.067	***523***
0.075	529
0.100	546

Knowing the *n* and λ values, the *deterrence* function now takes the following form

$$F_{ij} = \frac{1}{d_{ij}^{(0.6667)}} O_i O_j \left(\frac{1}{R_{ij}} \right) \tag{32}$$

and the smoothed function takes the following form

$$f = (1\text{-}0.5288)(F_{ij}) + 0.5288(\textit{extreme impedance}\ \text{function}) \tag{33}$$

where the constant used to represent the *extreme impedance* function for every pair is 0.037.

Equation 33 is the *deterrence* function for the projected production and consumption capacities for the year 2005. The *deterrence* function, which takes into consideration the effects of e-commerce, yields an estimated summation of **10,668,400**

million ton-miles in 2005. As compared to the value of 2,511,728 million ton-miles in 1997, such enormous increase in ton-miles from 1997 to 2005 is due to the opportunities provided by e-commerce, which allows businesses to reach farther and newer customers. Even though we have taken a rather conservative approach so as to not overestimate the effect of e-commerce, the result still sees a tremendous exponential growth to the estimated ton-miles placed on the transportation network. The reported state-to-state ton-miles can be used by planners and governmental agencies to properly allocate their resources for expanding and improving the transportation network throughout United States.

Table 9 is the list of all assumptions used to obtain the results of this research. The table includes the supporting ideas on why certain assumptions are made.

Table 9. List of Assumptions Made in this Research

Items	Assumption	Supporting Ideas
Percent share of E-Commerce in 2005	**15%**	Latest data released by U.S. Census Bureau claims that in 1999, approximately 10% of manufacturing shipments are due to e-commerce. 15% is therefore a projected increase for year 2005.
Rate of increase used to project year 2005 production and consumption capacities	**20%**	Capacities increased by 14% from 1993 to 1997. To avoid over estimating the capacities, we assume the capacities will increase by 20% from 1997 to 2005.
Percent of miles beyond the linear increase in average miles every 4 years that is used on e-commerce products. This applies to 15% of the product.	**20%**	Some selected e-commerce products have undergone increase in average miles up to 15% every 4 years. Therefore, with the effect of e-commerce, which creates flows to even further regions, 20% beyond the linear increase is a reasonable assumption.
Percent of miles beyond the linear increase in average miles every 4 years that is used on non e-commerce product, the assumption is as follows.	**10%**	For non e-commerce product, we make assumption based upon the traditional condition of the commodity flow survey from 1993 to 1997.
Miles between each origin-destination pair	**(Σton-miles) / (Σton)** for all products flowing between each origin-destination pair	This assumption gives us the average distance between all the shipment locations between each origin-destination pair.

5.3 Estimated Percentage of Different Modes Usage

A side usage of the flows calculated using the proposed method is to estimate the mode usage. We assign the flows between the states to different types of transportation modes using the 1997 commodity flow survey data. Each origin-destination pair was evaluated separately for this analysis.

The resulting conclusions from this analysis indicate that parcel, air, and multi-mode shipments will undergo a significant growth. This growth will be driven by the impact of e-commerce and resultant in farther distance shipments. As for truck shipments, its percent share in mode usage will decrease quite significantly, whereas the rail is becoming a more dominant mode. Bob Davidson, an expert from ABF trucking company claims that such observation is very unlikely to happen, provided that fuel costs increases tremendously in the next few years. On top of that, the limitation of railroad network also prohibits shippers from using rail shipments. Most rail shipments involve large shipment size, and longer hauls. The long haul condition is satisfied due to e-commerce, but e-commerce involves greater amount of smaller shipments. Therefore it is not possible for shippers to use rail to move the shipments.

Corresponding ton-miles placed on the transportation network will still continue to grow due to the increase in average miles shipped, and the growth in production capacities. The percent shares of different mode usage in 1997 and 2005 are shown in Table 10. Values from this table are plotted in Figure 11.

Table 10. Percent Share of Different Mode Usage in 1997 and 2005 for All Product Types

	1997 Proportion of Modes	2005 Proportion of Modes
Truck	70.04%	48.91%
Rail	15.04%	31.09%
Water	5.13%	6.91%
Air	0.06%	0.32%
Pipeline	4.59%	3.52%
Parcel, Courier, US Postal	0.26%	0.84%
Multiple and Unknown Modes	4.88%	8.42%

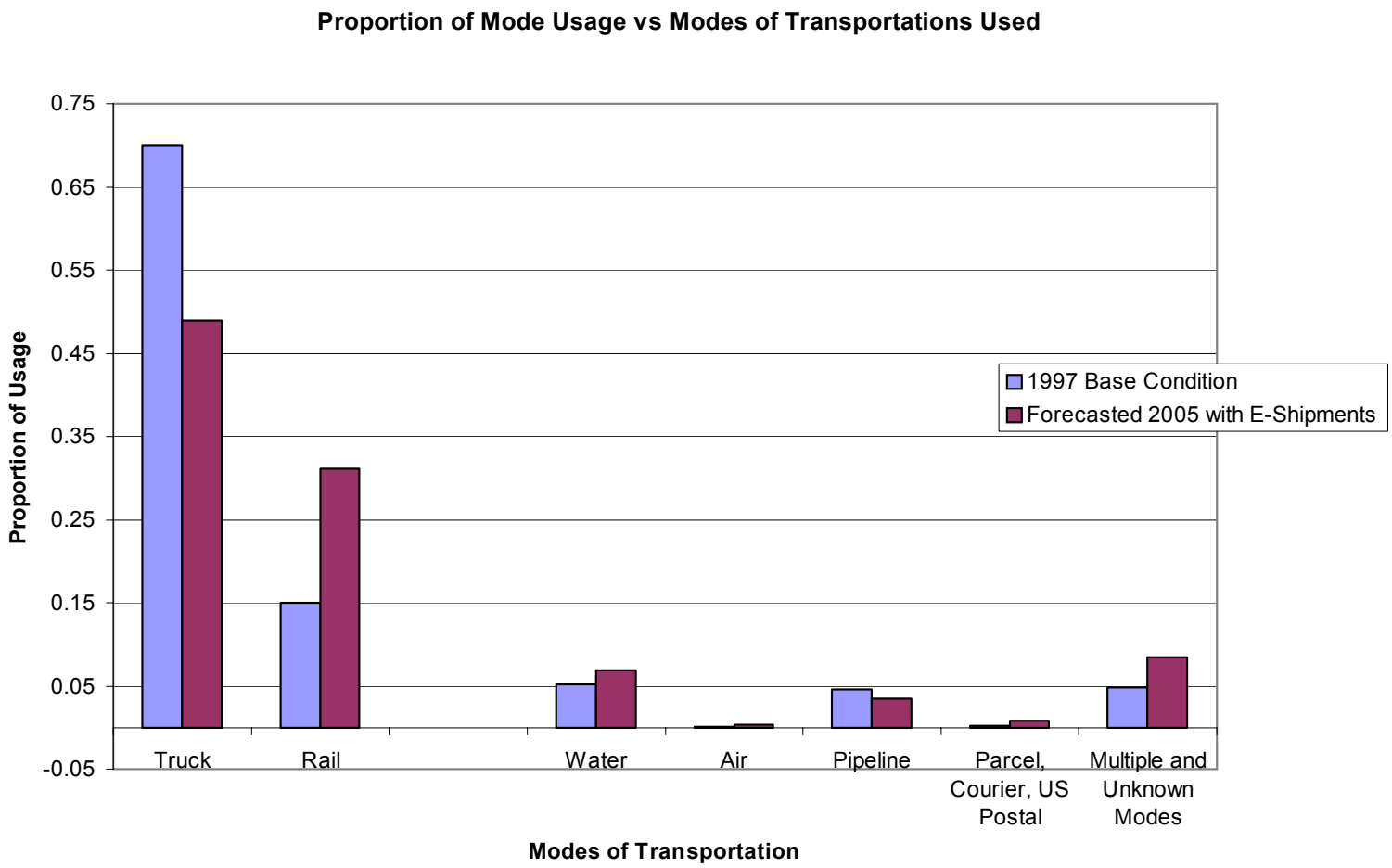


Figure 11. Proportion of Different Mode Usage versus Modes of Transportation Used in 1997 and 2005

As discussed before, the variability in factors that drive mode selection has placed a great limitation in better estimate the mode assignment. The method used in this research assumes that distance has a significant effect on mode choice. Therefore, estimated results should be carefully reviewed for any future estimation and forecasting purposes.

6. CONCLUSION

Understanding all the elements of sweeping changes drawn by the Internet economy is a great challenge. In government agencies, research institutions, and private companies around the world, analysts are trying to meet this challenge (Buckley, Henry, Gill, Laporte, Cooke, Pastore, Price, Shapiro, Mayer, 2000). The dynamism of this new economy is creating higher challenge for us in measuring its growth and impact on our society. The Internet has not only partially leveled the playing field for large and small firms, but it has also taken away the spatial barriers among themselves, and between them and the consumers. The increasing electronic connectivity has linked millions of people together in one small marketplace (Buckley, Henry, Gill, Laporte, Cooke, Pastore, Price, Shapiro, Mayer, 2000).

Looking at the Internet economy from a different perspective, the share of consumption, particularly from longer-distance regions, is increasing. This share will continue to rise as more and more businesses and consumers that are further away are getting connected and finding themselves closer and closer to the sellers. This trend of diminishing effect of distance caused by the growth in information technology will have significant effect on the transportation network, as more and more products will travel longer distance to reach its buyers.

Gravity modeling, the primary modeling approach used in this research, has the capability of representing the entities of e-commerce. Mainly, the entities such as the production and consumption capacities can be represented by tonnage values that are created and consumed within a region, and the relationship between the regions can be represented by *impedance* values derived from the *deterrence* function. The diminishing

effect of distance on the flows is captured by *impedance* values derived from the exponential smoothing method.

Recent e-commerce data released by Census Bureau provide us with a better insight on the percent of e-commerce from the total industry. This information helps us to better determine the λ value that can be used in the exponential smoothing method in order to determine the estimated directional distribution of flows in year 2005.

Observations and results from this research have shown positive capability of the gravity model to capture the diminishing effect of distance in e-commerce. We believe that this model can serve as a tool to estimate the future directional distribution of product flows in the United States due to e-commerce.

The growth of e-commerce also changes the demand for different transportation modes significantly. Since products begin to flow on longer distances and the expectation from e-consumers will continue to rise, companies will continually seek for better and cheaper transportation solutions. This will therefore placed a greater impact on transportation providers and open up lots of opportunities for researches to better improve and integrate different transportation mode usage and their networks. We believe that the U.S. Census Bureau should begin to take a bigger step in tracking the Internet economy. Better monitoring of the product flows due to e-commerce can help parties such as the Transportation Department to identify the point of need to allocate their resources to improve the transportation network. In addition, better understanding of e-commerce flows will also help transportation industry such as a trucking company to expand their truck travel network in providing better services to their customers. Understanding the e-

commerce will also help businesses from manufacturers to retailers to better manage their supply chain and provide better customer service level.

7. REFERENCES

Atrostic, B.K.; Gates, J.; Jarmin, R., "Measuring the Electronic Economy: Current Status and Next Steps", U.S. Census Bureau; available from http://www.census.gov/eos/www/ebusiness614.htm/; accessed April, 2000.

Barua, A.; Pinnell, J.; Shutter, J.; Whinston, A.B., "Measuring the Internet Economy: An Exploratory Study", University of Texas at Austin, 1999.

Bartolacci, M.R., Estimating E-Commerce Flows for Network Planning, Penn State Berks/ Lehigh Valley College, 1999. Unpublished.

Bottum, M.S., "Retail Gravity Model", *Appraisal Journal*, Vol. 557, pp. 166-172, 1989.

Carter, C., "Assumptions Underlying the Retail Gravity Model", *Appraisal Journal*, Vol. 61, pp. 509-518, 1993.

Connor, J.; Whitton, R., "Calibrating and Testing a Gravity Model for Any Urban Size Area, Bureau of Public Roads", U.S. Department of Commerce, 1965.

Buckley, P., Henry, D., Gill, G., Laporte, S., Cooke, S., Pastore, D., Price, L., Shapiro, R., Mayer, J., "Digital Economy 2000", Economic and Statistics Administration, U.S. Department of Commerce, available from http://www.esa.doc.gov/de2000.pdf/; accessed December, 2000.

Edwards, J.D., Transportation Planning Handbook, Institute of Transportation Engineers, Prentice Hall, 1992.

Environmental News Network, "E-Commerce: Friends or Foe of the Environment," available from http://cnn.com/; accessed December 12, 2000.

Erlander, S., Optimal Spatial Interaction and the Gravity Model, Springer-Verlag, pp. 1-13, 1980.

Erlander, S.; Stewart, N., The Gravity Model in Transportation Analysis, VSP, pp. 1-36, 1990.

Hamburg, J.R.; Kaiser, E.; Lathrop, G., Forecasting Inputs to Transportation Planning, *National Cooperative Highway Research Program Report 266*, pp. 19-63, 1983.

Kanafani, A.K, Transportation Demand Analysis, McGraw-Hill, Inc, pp. 165-185, 1983.

Mesenbourg, T.L., "Measuring Electronic Business: Definitions, Underlying Concepts, and Measurement Plans", U.S. Census Bureau; available from http://www.census.gov/eos/www/ebusiness614.htm/; accessed March, 2001.

Ortuzar, J.D., Modeling Transport, John Wiley and Sons, 1990.

Pisharodi R., The Transport Choice Decision Process, *International Journal of Physical Distribution and Logistics Management*, Vol. 21, No.5, 1991, pp. 13-22

"United States Department of Commerce News", U.S. Census Bureau; available from http://www.census.gov/mrts/www/current.html/; accessed April, 2001.

"U.S. Map New Stat", dotcom.com Network Solution; available from http://www.dotcom.com/; accessed December 10, 2000.

Veras, J.H.; Thorson, E., "Trip Length Distributions in Commodity Based and Trip Based Freight Demand Modeling", *Journal of Transportation Research Board*, Vol. 1707, pp. 37-55, 2000.

Werner, C., Spatial Transportation Modeling, University of California, Sage Publication, 1985.

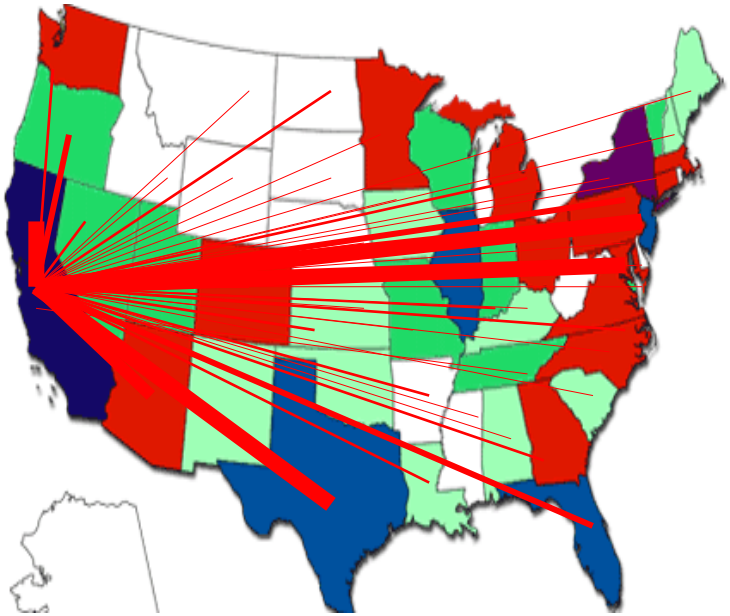
Okoruwa, A.; Nourse, H.; Terza, J., "Estimating Sales for Retail Centers: An Application of the Poisson Gravity Model", *Journal of Real Estate Research*, pp. 85-97, 1994.

Network Solution , "Web Site Usage Among Dotcoms", available from http://www.dotcom.com/; accessed December 10, 2000.

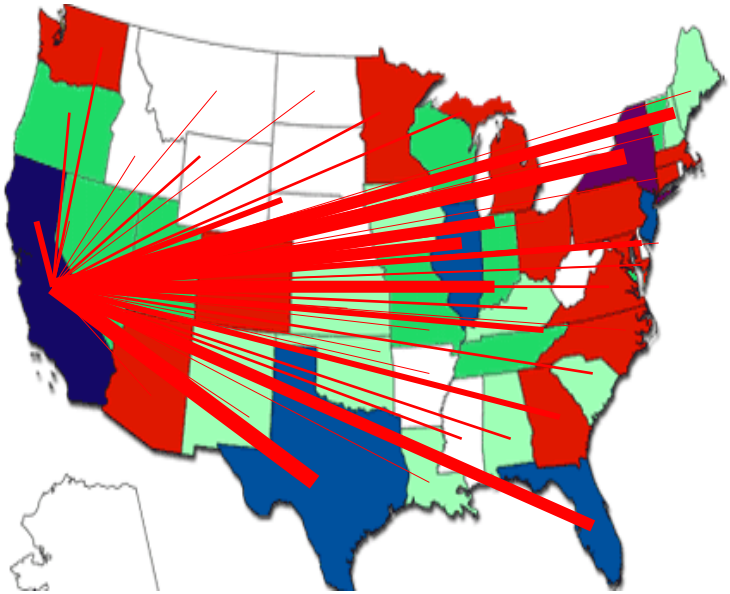
Webster, E., "A Gravity Model Analysis of the Effect of Regional Policies to Attract Foreign Tourists", *Journal of Applied Business Research*, Vol. 9, pp. 19-24, 1993.

Internet Statistics, Cyber Atlas; available from http://www.cyberatlas.com/; accessed April 10, 2001.

Regional Retail Sales Statistic, National Retail Foundation; available from http://www.nrf.com/; accessed December 15, 2000.



Traditional Base Flow Condition



Flow Condition with Strong E-Commerce Effect (n=2, λ=0.5)

>250K tons
151-200K tons
101-150K tons
51-100K tons
0-50K tons