# EXPLORATORY RESEARCH ON THE PERFORMANCE OF EMPTY TRIP MODELS WITH PROBABILITIES THAT DEPEND ON TRIP CHARACTERISTICS


***José Holguín-Veras***
*Associate Professor, Ph.D., P.E.*
*Department of Civil and Environmental Engineering, Rensselaer Polytechnic Institute*
*JEC 4030, 110 8th Street, Troy, NY 12180, U.S.A*
*Tel: (518) 276- 6221 Fax: (518) 276-4833, Email: jhv@rpi.edu*

***Ellen Thorson***
*Senior Research Fellow, Ph.D.*
*University Transportation Research Center, The City College of New York*
*Marshak Hall Suite 910A 138th Street & Convent Avenue New York, NY 10031*
*Tel: 212-650-8144, Fax: 212-650-8374*
*Email: ethor@utrc2.org*

***Juan C. Zorrilla***
*Transportation Analyst*
*100 Cambridge Park Drive, Suite 400*
*Cambridge, MA 02140, Email: jzorrilla@camsys.com*



***Abstract***

This paper considers enhanced formulations to model commercial vehicle empty trips. These formulations relax a limitation of the original trip chain models, i.e., the assumption of a constant probability of a zero order trip chain for the entire range of trip types. The formulations considered in the paper assumed that $p$ was a function of the opposing commodity flow, or a function of the trip distance. The consideration of a variable $p$ improved the relative performance of the models in rather unexpected ways. As shown in the paper, in all three datasets considered, the simplest models, and the one with the smaller number of parameters, outperformed the most complex formulations.

## *1. Introduction*

One of the most important supporting activities in transportation planning is the estimation of future transportation demand which, together with network (supply) analyses, provide the basis for decisions pertaining to transportation investment needs. In order to make sound decisions, a comprehensive set of transportation demand models should be developed to represent the behavior of the different user segments (e.g., passengers, freight) as a function of social, economic and policy variables. Once such models have been estimated and calibrated to ensure they reflect the local conditions, they are used for forecasting purposes. Unfortunately, the lack of adequate freight transportation demand models remains a formidable obstacle to taking full advantage of the process described above.

There are several reasons that may help explain why freight transportation modeling is lagging behind passenger transportation modeling. A first reason may be related to the fact that transportation modeling started as an exercise that focused almost exclusively on passenger transportation, with freight being no more than an afterthought. This point of view may have had its origins in the fact that commercial truck traffic usually represents a relatively small proportion of total traffic (typically between 6-10%). However, in spite of the relatively small numerical significance of commercial truck traffic, its economic value and its contributions to network congestion are considerable. With travel time values that routinely reach $50-70/hour, compared to $10-15/hour for passenger demand, the economic value of commercial traffic could easily represent 20-35% of total economic value of the trips made in a region. (The same line of reasoning applies to network congestion.) A second set of reasons are the chronic lack of understanding and appreciation of the importance of freight activity to the economy, and the history of confrontations between the government and the freight industry (for discussions of the American case, see Holguín-Veras and Jara-Díaz, 1999; and Holguín-Veras et al. 2005) that have shaped the minds of countless transportation professionals and community leaders into believing that freight activity, and particularly truck traffic, is something to get rid of; not a topic that deserves to be understood and researched.

A third set of factors are associated with the staggering complexity of the underlying dynamics of freight activity. There are multiple players that dynamic interact in a highly confidential and commercially sensitive environment (e.g., shippers, freight forwarders, third party logistic systems, freight carriers, receivers, government regulatory agencies). There are literally hundreds of thousands of different commodities with a very wide range of opportunity costs and handling needs. More important to the purposes of this paper, freight demand can be measured and defined according to different dimensions: cargo value, weight and number of vehicle-trips.

The multidimensional nature of freight demand has given rise to two major modeling platforms: vehicle-trip based and commodity based (cargo value is only used in Input-Output models). Vehicle-based models focus on modeling the actual number of vehicle trips, which has several advantages. Among them are the relative ease and high-quality with which traffic data can be obtained; and, since the model focuses on vehicle trips, no distinction is made between empty and loaded trips. A key limitation of vehicle-trip modes is that they cannot be applied to multimodal systems because the vehicle-trip is already the result of a mode choice that already took place (Holguín-Veras and Thorson, 2003a). Furthermore since the models assume that the vehicle-trip is the unit of demand, as opposed to the commodity being transported, there is no way to consider the economic characteristics of the shipments. This is a rather serious limitation because the commodity type has been found to be a very important explanatory variable of a number of choice processes involving freight (Holguín-Veras, 2002).

Commodity based models, as the name points out, focus on modeling the flow of goods between zones (measured in a unit of weight). Since the cargo's weight is the unit of demand, the consideration of cargoes' attributes (e.g., value, weight, type) is straightforward. In this platform, the loaded trips are estimated by dividing the total flow from one region to the other by a suitable payload for all loaded trucks. The problem with commodity-based models is that they are unable to model empty trips, which can make up about 30 to 40 percent of the total trips in a region (Holguín-Veras and Thorson, 2003a). This occurs because the commodity flow in one direction determines the corresponding loaded trips, but does not bear a direct relationship with the empty trips. To resolve this, complementary empty trip models have been developed, such as Noortman and van Es' (in Hautzinger 1984), Hautzinger's (1984), and Holguín-Veras and Thorson (2003a).

In this context, the empty trip models use the commodity flows estimated by a freight demand model as an input for the estimation of the corresponding empty trips. Having done that, the empty trips are added to the loaded trips to obtain the total vehicle trips that would be used in the traffic assignment process. The main advantage of this sequential estimation is that it enables the researcher to take advantage of existing commodity flow models. It is also clear that if comprehensive freight models—which are able to estimate both commodity flows and vehicle trips—are used, there is no need to use separate empty trip models (e.g., Holguín-Veras, 2000). Far from being of purely academic interest, the correct estimation of commercial vehicle empty trips is very important for transportation planning purposes because not doing correctly will lead to severe directional errors in the estimation of commercial vehicle traffic, as shown in Holguín-Veras and Thorson (2003b). This, in turn, may have important implications in terms of determining road capacity improvement needs.

The main objective of this paper is to contribute to freight transportation modeling by enhancing the methodologies used to estimate empty trips from previously estimated commodity flow matrices. The paper builds on the developments of Noortman and van Es (1978), Hautzinger (1984) and Holguín-Veras and Thorson (2003a). The paper considers enhanced formulations of Holguín-Veras and Thorson's that relax a key assumption of their original formulations, i.e., that the probability of a trip chain with only one stop (zero order trip chain) is constant. The formulations considered in this paper, originally suggested in Holguín-Veras et al., (2005), outperform all others.

The paper has four major sections. *Previous Developments* provides the reader with an idea about the alternative approaches to model (or to avoid modeling) commercial vehicle empty trips. *Further Improvements* discusses the new formulations developed in this paper. *Brief description of the data*, as the name implies, provides the reader with a succinct idea about the data used in the paper. *Results* presents the findings from a case study. *Conclusions* summarizes the key findings from this research.

## *2. Previous Developments*

This section discusses the empty trip models reported in the literature. In order to ensure consistency with these works, the authors follow the notation used by Hautzinger (1984) and Holguín-Veras and Thorson (2003a). Because of the obvious linkages between the Holguín-Veras and Thorson (2003a) and the models proposed in this paper, the latter models are discussed in more detail than the other previous developments. This would enable the reader to fully understand the differences between these models and the models introduced in this paper.

Notation:

$m_{ij}$ = commodity flow between origin *i* and destination *j*

$d_{ij}$ = distance (or any measure of trip impedance) between origin *i* and destination *j*

$a_{ij}$ = average payload (tons/trip) for loaded trips between *i* and *j*

$x_{ij} = \frac{m_{ij}}{a_{ij}}$ = estimated number of loaded trips between *i* and *j*

$y_{ij}$ = estimated number of empty trips between *i* and *j*

$z_{ij} = x_{ij} + y_{ij}$ = estimated total number of trips (loaded + empty) between *i* and *j*

### *2.1 Naïve Proportionality Model*

This model attempts to approximate the total number of trips from *i* to *j* as a function of the commodity flow in the same direction, as shown in Equation (1). The constant *M*, to be determined empirically, is multiplied by the average payload in order to obtain a "load factor" that takes into account empty trips as well. Although the estimates produced by this model could be made to match the total number of trips in a region, it leads to significant errors in the directional flows. This is because of its mathematical construction that produces overestimation of the vehicle-trips flows in the predominant direction of the commodity flows, and underestimation of the vehicle-trips in the opposite direction (see Holguín-Veras and Thorson, 2003b). For example, if the loaded trips from *j* to *i* increase, it is likely that the number of trips from *i* to *j* will also increase (because of the return trips), but this model will not account for it since the commodity from *i* to *j* remains unchanged (assuming return trips are empty). Other variations of naïve formulations exist, however they all have the same limitations, as shown in Holguín-Veras and Thorson (2003b).

$$z_{ij} = \frac{m_{ij}}{Ma_{ij}} \qquad (1)$$

### *2.2 Modeling empty trips as a commodity*

Another approach that has been used entails modeling the empty trips as a (rather unique) commodity (e.g., Tamin and Willumsen, 1984; Fernández et al. 2003). In this type of application, distribution models of commercial vehicle empty trips are calibrated as a function of socio-economic variables, typically using an observed matrix of empty trips as the input. There are, at least, two major problems with this approach. First, there is no way to ensure consistency between the forecasts of loaded and empty trips. This is important because the percentage of empty trips in a given area has been found to be very stable between 30-50% of total trips and it is not likely that independently estimated models maintain such consistency. Second, since the empirical evidence demonstrates the existence of a significant correlation between the empty trips and the opposing commodity flows, the distribution models of empty trips are likely to end up using variables traditionally associated with trip production as trip attraction variables, and vice versa. Needless to say, this is of questionable conceptual validity.

### *2.3 Noortman and van Es*

Noortman and van Es (1978) modeled the empty trips explicitly as a function of the commodity flow in the opposite direction, which undoubtedly represents a step forward. A fraction, $p$, of the loaded trips in the opposite direction are expected to return empty to their origin. Therefore the number of empty trips in direction $i$-$j$ can be estimated as the total number of loaded trips in the opposite direction multiplied by this constant $p$, see Equation (2) below. Note that this model assumes that the total number of trips from $i$ to $j$ is given only by the commodity flow between the two areas therefore not considering trip chains, which occur more often than not. Nevertheless, it still represents an improvement over the naïve proportionality model.

$$z_{ij} = x_{ij} + y_{ij} = \frac{m_{ij}}{a_{ij}} + p\frac{m_{ji}}{a_{ji}} \qquad (2)$$

### *2.4 Hautzinger*

Hautzinger pointed out that Noortman and van Es' model implies that the difference in vehicle flows (between two zones) is directly proportional to the difference in commodity flows. He claims that empirical evidence shows that even in the cases of extreme differences between commodity flows, $z_{ij}$ and $z_{ji}$ tend to be approximately equal. To correct this he developed the model shown in Equation (3). Here $p_i$ represents the probability of a trip going from *i* to *j* returning empty to *i*. It can clearly be seen that with this formulation $z_{ij}$ is equal to $z_{ji}$. Hautzinger also specified a formulation for the probabilities, assuming that $p_i = \exp(-\lambda(m_{ji} / m_{ij})^2)$.

$$z_{ij} = \frac{(p_i m_{ij} + p_j m_{ji})}{a(1-(1-p_i)(1-p_j))} \quad (3)$$

However, the analyses of the dataset used in this paper highlighed some limitations of the Hautzinger model. First, attempts to compute $p_i$ proved unsuccessful because of the numerical impossibility of computing exponentials of large ratios. Second, the authors also found that the difference in vehicle flows is in fact directly proportional to the difference in commodity flows, as originally assumed by Noortman and van Es (the Pearson correlation coefficient between the two variables for the dataset in study was relatively high, i.e., 0.78). This finding was corroborated by the analyses of two different data sets (see Holguín-Veras et al. 2005). There was also an overall considerable difference in directional vehicle flows in the dataset, reaching up to 32% of the flows.

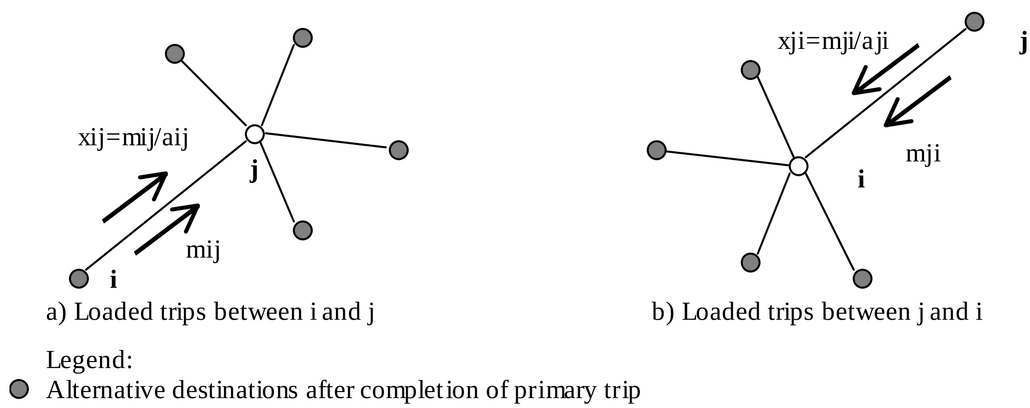
In any case, Hautzinger's model still presents the same limitation as Noortman and van Es', i.e., it assumes that the vehicle flow between two zones is only a function of the commodity flow between these two zones. Although it is true that a large portion of the commercial traffic is a function of the commodity flow between the two zones in consideration, trip chains are becoming increasingly common. As a result, there is an increased need for formulations that explicitly consider trip chains. In order to overcome this limitation, Holguín-Veras and Thorson (2003a) introduced a formulation based on trip chains to model commercial vehicle empty trips.

### *2.5 Holguín-Veras and Thorson*

The next stage in the development of empty trip models involved the use of probability and spatial interaction concepts to model empty trips as trip chains (Holguín-Veras and Thorson, 2003a). These formulations attempt to model the trip chain formation process, as opposed to Noortman and van Es' and Hautzinger's that only focus on modeling empty trips back to the home base. Holguín-Veras and Thorson (2003a) defined order of a trip chain as the number of destinations –in addition to the primary trip– that a vehicle visits in a tour. In this context, the models proposed by Noortman and van Es, and Hautzinger effectively model zero order trip chains.

The models developed by Holguín-Veras and Thorson (2003a) are based on a simplified depiction of commercial vehicle trip chains. These formulations are able to consider the flow of commercial vehicles as part of a trip chain of first order, i.e., in addition to modeling the vehicle flow from *i* to *j*, the model considers a third zone. Although in real life trip chains are far more complicated than presented in this model, it still represents a step forward. An example of a 1st order trip chain is shown in Figure 1 (after Holguín-Veras and Thorson, 2003a). Figure 1a presents the flow of vehicles from *i* to *j*, and a set of possible destinations after *j*. Figure 1b presents the flow in the opposite direction.

**Figure 1: Schematic of commercial vehicle trip chains**




$$E(z_{ij}) = E(x_{ij}) + E(y_{ij}) \quad (4)$$

Using the model proposed by Noortman and van Es as a basis, Holguín-Veras and Thorson defined the vehicle flow from zone *i* to zone *j* as the summation of the expected number of loaded trips and the expected number of empty trips. As defined earlier, the expected number of loaded trips is equal to the commodity flow, $m_{ij}$, divided by the average payload, $a_{ij}$. Therefore, after substitution, equation (4) reduces to the model below.

$$E(z_{ij}) = \frac{m_{ij}}{a_{ij}} + E(y_{ij}) \quad (5)$$

Holguín-Veras and Thorson (2003a) decomposed the second term into a summation of the vehicle-trips associated with the trip chains of different orders. They defined $x_{ij}^{n}$ as the number of loaded trips from *i* to *j* as part of an nth order trip chain. Similarly, $y_{ij}^{n}$ is the number of empty trips from *i* to *j* as part of an n[th] order trip chain. Then, Equation (5) can be expressed as the summation of the loaded and empty trips from *i* to *j* associated with all the trip chains of different orders. This is shown in Equation 6 below.

$$E(z_{ij}) = \sum_{n=0}^{N} E(x_{ij}^{n}) + \sum_{n=0}^{N} E(y_{ij}^{n}) \tag{6}$$

$$E(z_{ij}) = \frac{m_{ij}}{a_{ij}} + E(y_{ij}^{0}) + E(y_{ij}^{1}) + \sum_{n>1} E(y_{ij}^{n}) = \frac{m_{ij}}{a_{ij}} + E(y_{ij}^{0}) + E(y_{ij}^{1})\left[1 + \frac{\sum_{n>1} E(y_{ij}^{n})}{E(y_{ij}^{1})}\right] \tag{7}$$

Then, by assuming that the term within brackets is equal to a constant $\gamma^{*}$, Holguín-Veras and Thorson arrived to the following model:

$$E(z_{ij}) = \frac{m_{ij}}{a_{ij}} + E(y_{ij}^{0}) + \gamma^{*} E(y_{ij}^{1}) \tag{8}$$

As shown in Equation (9), Holguín-Veras and Thorson assumed (in a manner similar to the model of Noortman and van Es) that zero order empty trips are a function of the opposing commodity flow and the probability, *p*, of a of a zero order trip chain. Empty trips of first order were modeled using the total number of vehicles arriving at zone *i* multiplied by the probability that these will not be involved in a zero order chain (*1-p*). Furthermore these vehicles would then have to choose zone *j* as the next destination and be empty, i.e., $P(j)P(E/j)$.

$$E(z_{ij}) = \frac{m_{ij}}{a_{ij}} + p\frac{m_{ji}}{a_{ji}} + (1-p)\gamma^{*} \sum_{h \neq j} \frac{m_{hi}}{a_{hi}} \left(P^{h}(j) P^{h}(E/j)\right) \tag{9}$$

If the probability *p* is constant, equation (9) can be rewritten as:

$$E(z_{ij}) = \frac{m_{ij}}{a_{ij}} + p\frac{m_{ji}}{a_{ji}} + \gamma \sum_{h \neq j} x_{hi} \left(P^{h}(j) P^{h}(E/j)\right) \tag{10}$$

Where:

*p* = probability of a zero order trip chain

$\gamma = (1 - p)\gamma^*$ = parameter to be determined empirically

$x_{hi}$ = number of loaded trips from *h* to *i*

$P^h(j)$= probability that a vehicle that came from *h* to *i* chooses *j* as the next destination

$P^h(E/j)$ = probability that a vehicle following the tour *h-i-j* does not get cargo to *j*

Holguín-Veras and Thorson proposed a number of different specifications of the destination choice probability function. In general terms, they assumed that the probability that a vehicle that came from *h* to *i* chooses *j* as the destination, $P^h(j)$, is a function of the commodity flow in the same direction, $m_{ij}$, and the trip impedance between *i* and *j*. Some formulations expressed the choice probabilities as a function of the previous origin *h*, while others only consider the interaction between zones *i* and *j*. These formulations are shown in Table 1. The resulting empty trip models are shown in Table 2.

**Table 1: Destination choice probability functions**

| | |
|---|---|
| $P(j) = \dfrac{m_{ij}}{\sum_l m_{il}}$ | (11) |
| $P(j) = \dfrac{m_{ij} e^{-\beta d_{ij}}}{\sum_l m_{il} e^{-\beta d_{il}}}$ | (12) |
| $P(j) = \dfrac{m_{ij} (d_{ij})^{-\beta}}{\sum_l m_{il} (d_{il})^{-\beta}}$ | (13) |
| $P^h(j) = \dfrac{m_{ij} (d_{ij} + d_{hi})^{-\beta}}{\sum_l m_{il} (d_{il} + d_{hi})^{-\beta}}$ | (14) |

The first equation (11) only considers the commodity flow between the two zones, while equations 12 and 13 also take into account the impedance between zones *i* and *j*. These formulations are more effective given that they consider the fact that trip chains with short trips are more likely than with long trips. Finally, equation 14 also takes into account the amount of travel already done by the vehicle starting at node *h*.

**Table 2: Holguín-Veras and Thorson's Trip Chain Formulations**

HV-T Model 1: $$E(z_{ij}) = \frac{m_{ij}}{a_{ij}} + p\frac{m_{ji}}{a_{ji}} + \gamma\sum_{h\neq j} x_{hi}\frac{m_{ij}}{\sum_l m_{il}} P(E/j) \quad (15)$$

HV-T Model 2: $$E(z_{ij}) = \frac{m_{ij}}{a_{ij}} + p\frac{m_{ji}}{a_{ji}} + \gamma\sum_{h\neq j} x_{hi}\frac{m_{ij}e^{-\beta(d_{ij})}}{\sum_l m_{il}e^{-\beta(d_{il})}} P(E/j) \quad (16)$$

HV-T Model 3: $$E(z_{ij}) = \frac{m_{ij}}{a_{ij}} + p\frac{m_{ji}}{a_{ji}} + \gamma\sum_{h\neq j} x_{hi}\frac{m_{ij}(d_{ij})^{-\beta}}{\sum_l m_{il}(d_{il})^{-\beta}} P(E/j) \quad (17)$$

HV-T Model 4: $$E(z_{ij}) = \frac{m_{ij}}{a_{ij}} + p\frac{m_{ji}}{a_{ji}} + \gamma\sum_{h\neq j} x_{hi}\frac{m_{ij}(d_{ij}+d_{hi})^{-\beta}}{\sum_l m_{il}(d_{il}+d_{hi})^{-\beta}} P(E/j) \quad (18)$$

Note: The initials HV-T are used when discussing the models presented by Holguín-Veras and Thorson

***3. Further improvements***

As shown in equation (9), Holguín-Veras and Thorson originally assumed that the probability of a zero order trip chain, *p*, is constant which enables equation (10) to be obtained. However, there is compelling evidence and theoretical support to argue that *p* should be a function of a number of variables including trip distance and the magnitude of the opposing commodity flow $m_{ji}$. In this section, a number of alternative formulations for *p* are hypothesized and tested with real life data available to the authors. Two cases are considered in this section, which are described next.

A situation that may lead to a variable *p* involves taking into account that large values of the opposing commodity flow $m_{ji}$ may increase the probability of undertaking a zero order trip chain. In this context, a large value of $m_{ji}$ could make it more attractive to undertake a zero order trip chain because it increases the likelihood of getting a load for the return trip. A formulation that could be used to test this hypothesis is the one in equation (19):

$$p = p_0 e^{p_1 m_{ji}} \quad (19)$$

As shown, it has two parameters $p_0$ and $p_1$ and is a function of the opposing commodity flow $m_{ji}$. Substituting equation (19) in equation (9) leads to:

$$E(z_{ij}) = \frac{m_{ij}}{a_{ij}} + p_0 e^{p_1 m_{ji}} \frac{m_{ji}}{a_{ji}} + (1 - p_0 e^{p_1 m_{ji}})\gamma^* \sum_{h \neq j} \frac{m_{hi}}{a_{hi}} \left(P^{hi}(j) P^{hi}(E/j)\right) \quad (20)$$

The second formulation considers *p* as a function of distance. (It seems that both Noortman and van Es (1978) and Hautzinger (1984) realized the role of distance and hypothesized that the parameters of their models are a function of the trip length, though neither suggested formulae.) This would imply that *p* should also be a function of trip length, as outlined in equation (21):

$$E(z_{ij}) = \frac{m_{ij}}{a_{ij}} + p_0 (1 + p_1^{'} d_{ij}) \frac{m_{ji}}{a_{ji}} + (1 - p_0 (1 + p_1^{'} d_{ij}))\gamma^* \sum_{h \neq j} \frac{m_{hi}}{a_{hi}} \left(P^{hi}(j) P^{hi}(E/j)\right) \quad (21)$$

These formulations were applied, together with the formulations with constant *p*, to three datasets corresponding to different vehicle types (i.e., small trucks and pick-ups, large (2 and 3 axle) trucks, and semi-trailers). The results shed some light into the different behaviors exhibited by these different types of vehicles.

**Parameter Search Procedure**

The estimation of the parameters of the models considered in this paper was conducted by a parameter optimization search that consists of finding the parameters that best fit a given data set. The parameters were determined by finding the values which minimized the error:

$$\varepsilon = \sum_{i,j} (z_{ij}^a - z_{ij})^2 \quad (22)$$

Where $z_{ij}^a$ is the actual number of trips from *i* to *j*, and $z_{ij}$ is the estimated value. The minimization procedure for the models with more than one parameter consists of a multidimensional downhill simplex method (Nelder and Mead, 1965, Press et al, 1992). A simplex is a geometrical figure in *N* dimensions with *N* + 1 vertices. The initial simplex is defined by a starting point $P_0$ which is a vector with *N* dimensions and a set of *N* points defined by:

$$P_i = P_0 + \lambda e_i \quad (23)$$

Where the $e_i$ are *N* unit vectors and $\lambda$ is a constant.

The method then takes a series of moves such as moving the point of the simplex where the function is largest through the opposite simplex face to a lower point until the function value at all points are within a specified tolerance of a minimum value. For the one dimensional model (NVE with constant *p*), a golden search was conducted (Press et al, 1992).

### *4. Brief description of the data*

The models were applied to an origin-destination (OD) sample corresponding to a major modeling project conducted by the first author in Guatemala City. In that project, roadside origin-destination interviews were conducted and complemented by classified traffic counts to expand the sample according to time of day and type of vehicle. The origin-destination questionnaire included questions about: time of the interview, vehicle type, origin, destination, commodity type, load factor, number of units, % of load, type of trucking company, shipment size, economic sectors at both origin and destination, and activities performed at both origin and destination. The sample, comprised of 5,276 observations, was expanded by time of day and type of vehicle, and processed to eliminate double counting of trips. The overall expansion factor was 6.476. There were 17 survey stations, five within the city itself and the remaining 12 located in the surrounding suburban areas. A full description of the sample can be found in Holguín-Veras and Thorson (2000). Although the survey indicated the type of commodity being transported, the models were applied to flows which were aggregated across commodity types. The number of observations and expanded trips are shown for each type of truck in Table 3.

**Table 3: Number of observations and expanded trips**

| Vehicle Set | Number of Observations | Expanded Total Trips |
|---|---|---|
| **Small** | 1138 | 21572 |
| **Large** | 3288 | 9769 |
| **Semitrailer** | 848 | 3645 |

### *5. Results*

Tables 4, 5, and 6 show a comparison of the different models for three data subsets taken from the Guatemala City origin-destination dataset. These data sets consisted of the following three groups of vehicle types:

(1) small trucks – pickups and small 2 axle trucks

(2) large trucks – large 2 axle and small 3 axle single unit trucks

(3) semi trailers.

Five different models are considered, the model proposed by Noortman and van Es and models 1-4 by Holguín-Veras and Thorson (equations 15, 16, 17, and 18). The first row for each model shows the results obtained when maintaining the parameter *p* as a constant, the second row shows the results of assuming *p* is a function of the commodity flow (equation 20), and the third row shows the results of assuming *p* is a function of the trip length (equation 21). The parameter values are given in columns 4 through 7 followed by the Sum of Squared Differences (SSD), that is, the square of the difference between the actual number of trips and the estimated ones. Although in previous papers, the authors used the Root Mean Squared Error (RMSE) as a performance measure, it was found that the RMSE does not always produce a consistent evaluation of performance. This is because different models produce a different number of estimates, *n*, which may translate into an artificially low value of the RMSE. In this context, to avoid this problem, the authors decided to focus on a measure of total error (SSD) instead.

**Table 4: Results for Small Trucks**

| | **Model Characteristics** | | **Model Parameters** | | | | **Performance Measures** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Dest. Choice** | **p function** | **p0** | **p1** | **Gamma** | **Beta** | **Error** | **% Diff-W** | **% Diff-B** |
| **NVE** | None | Constant | 1.09551 | | | | 2218882 | 8.034 | 54.907 |
| | **None** | **Commodity** | **1.97314** | **-1.28216** | | | **2053864** | **0.000** | **43.386** |
| | None | Distance | 1.18280 | -0.51677 | | | 2208492 | 7.529 | 54.181 |
| **HVT 1** | mij | Constant | 1.00460 | | -951.46837 | | 1631792 | 13.920 | 13.920 |
| | **mij** | **Commodity** | **1.00888** | **-0.01177** | **-951.44097** | | **1432399** | **0.000** | **0.000** |
| | mij | Distance | 1.00459 | 0.00079 | -951.46017 | | 1631014 | 13.866 | 13.866 |
| **HVT 2** | mij*exp(b*dij) | Constant | 0.61446 | | 12.92470 | -14.28841 | 1452190 | 1.216 | 1.382 |
| | **mij*exp(b*dij)** | **Commodity** | **0.55597** | **0.33605** | **12.98480** | **-14.37069** | **1434742** | **0.000** | **0.164** |
| | mij*exp(b*dij) | Distance | 0.60340 | 0.25377 | 13.03043 | -14.31673 | 1449113 | 1.002 | 1.167 |
| **HVT 3** | mij*dij^b | Constant | 0.59742 | | 6.63583 | -13.10547 | 1724791 | 3.290 | 20.413 |
| | **mij*dij^b** | **Commodity** | **0.22224** | **-0.26220** | **4.86220** | **-2.26979** | **1669848** | **0.000** | **16.577** |
| | mij*dij^b | Distance | 0.33908 | -0.00319 | 4.62928 | -2.56423 | 1683314 | 0.806 | 17.517 |
| **HVT 4** | mij*(dij+dhi)^b | Constant | 0.61032 | | 11.01942 | -9.42906 | 1456988 | 0.450 | 1.717 |
| | mij*(dij+dhi)^b | Commodity | 1.00180 | -0.95155 | 10.77352 | -9.28352 | 1854337 | 27.845 | 29.457 |
| | **mij*(dij+dhi)^b** | **Distance** | **0.59676** | **0.27171** | **11.25744** | **-7.84738** | **1450456** | **0.000** | **1.261** |

**Table 5: Results for (Two/Three Axle) Large Trucks**

| | Model Characteristics | | Model Parameters | | | | Performance Measures | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Dest. Choice** | **p function** | **p0** | **p1** | **Gamma** | **Beta** | **Error** | **% Diff-W** | **% Diff-B** |
| **NVE** | None | Constant | 0.68019 | | | | 74391 | 11.309 | 30.917 |
| | **None** | Commodity | 1.04663 | -0.92285 | | | 67913 | 1.616 | 19.516 |
| | None | **Distance** | **0.86912** | **-0.90411** | | | **66833** | **0.000** | **17.615** |
| **HVT 1** | mij | Constant | 0.41639 | | 2.20531 | | 60733 | 6.882 | 6.882 |
| | **mij** | Commodity | 0.73023 | -1.72904 | 2.61330 | | 57416 | 1.044 | 1.044 |
| | mij | **Distance** | **0.65870** | **-1.33082** | **1.84539** | | **56823** | **0.000** | **0.000** |
| **HVT 2** | mij*exp(b*dij) | Constant | 0.41101 | | 2.19208 | -0.77134 | 60326 | 5.338 | 6.165 |
| | **mij*exp(b*dij)** | Commodity | 0.62493 | -1.70467 | 2.25841 | -0.74801 | 58227 | 1.671 | 2.470 |
| | mij*exp(b*dij) | **Distance** | **0.56351** | **-1.26476** | **2.14558** | **-0.83153** | **57269** | **0.000** | **0.785** |
| **HVT 3** | mij*dij^b | Constant | 0.46593 | | 1.15076 | -0.32851 | 65562 | 12.824 | 15.379 |
| | **mij*dij^b** | **Commodity** | **0.66950** | **-0.76874** | **2.08079** | **-0.63325** | **58110** | **0.000** | **2.264** |
| | mij*dij^b | Distance | 0.40984 | -0.10488 | 1.71489 | -0.52622 | 61231 | 5.371 | 7.757 |
| **HVT 4** | mij*(dij+dhi)^b | Constant | 0.51828 | | 1.19294 | 0.57955 | 67391 | 14.818 | 18.599 |
| | mij*(dij+dhi)^b | Commodity | 1.17563 | -1.09935 | 1.06289 | 0.58325 | 69126 | 17.773 | 21.651 |
| | **mij*(dij+dhi)^b** | **Distance** | **0.72341** | **-0.96690** | **1.41656** | **0.14420** | **58694** | **0.000** | **3.293** |

**Table 6: Results for Semi-trailers**

| | Model Characteristics | | Model Parameters | | | | Performance Measures | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Dest. Choice** | **p function** | **p0** | **p1** | **Gamma** | **Beta** | **Error** | **% Diff-W** | **% Diff-B** |
| **NVE** | None | Constant | 0.45495 | | | | 28926 | 4.430 | 28.127 |
| | **None** | **Commodity** | **0.66712** | **-0.66496** | | | **27698** | **0.000** | **22.692** |
| | None | Distance | 0.30892 | 1.13926 | | | 28089 | 1.410 | 24.422 |
| **HVT 1** | mij | Constant | 0.33736 | | 1.44461 | | 24099 | 6.748 | 6.748 |
| | **mij** | Commodity | 0.56115 | -0.96224 | 1.72861 | | 23243 | 2.956 | 2.956 |
| | mij | **Distance** | **0.10627** | **5.21738** | **1.44557** | | **22576** | **0.000** | **0.000** |
| **HVT 2** | mij*exp(b*dij) | Constant | 0.32278 | | 1.53966 | 1.38604 | 23731 | 3.386 | 5.116 |
| | **mij*exp(b*dij)** | **Commodity** | **0.48069** | **-0.96771** | **1.72134** | **1.04226** | **22953** | **0.000** | **1.673** |
| | mij*exp(b*dij) | Distance | 0.26922 | 0.69757 | 1.76085 | 1.05938 | 23319 | 1.594 | 3.294 |
| **HVT 3** | mij*dij^b | Constant | 0.33760 | | 1.43568 | -0.02260 | 24097 | 3.612 | 6.741 |
| | **mij*dij^b** | Commodity | 0.54938 | -0.86677 | 1.53964 | -0.04918 | 23310 | 0.228 | 3.255 |
| | mij*dij^b | **Distance** | **0.21787** | **1.16401** | **1.66981** | **0.04506** | **23257** | **0.000** | **3.020** |
| **HVT 4** | mij*(dij+dhi)^b | Constant | 0.32641 | | 1.47806 | 0.37582 | 24016 | 2.726 | 6.379 |
| | mij*(dij+dhi)^b | Commodity | 0.81774 | -0.80499 | 1.50157 | 0.27441 | 27138 | 16.081 | 20.209 |
| | **mij*(dij+dhi)^b** | **Distance** | **0.27425** | **0.67900** | **1.60462** | **0.12722** | **23378** | **0.000** | **3.556** |

Note: In each of the tables above, the best performing model for each group is shown in bold. % Diff-W is the percentage difference in error calculated relative to the best model within each group and % Diff-B is the percentage difference in error calculated relative to the best overall model.

As shown in the tables, the consideration of a variable probability of a zero order trip chain produces some rather remarkable results. The most noteworthy being that the simplest model (HVT 1), with the smallest number of parameters, outperformed the more complex formulations. Another important finding is that the variable *p* enhances model performance across the board with the exception of HVT 4 (commodity). As shown in the column % Diff-W (within group), models with variable *p* outperform the models with the constant *p*, by an average of 5.22% excluding HVT 4 (commodity).

For each type of truck, the best performing model overall was the simplest HVT model (HVT 1). For small trucks, the version with *p* as a function of commodity flow was best, while for the larger trucks, the best model was that with *p* as a function of distance. This difference could be explained by the fact that small trucks tend to be used for shorter trips, while distance plays a more important role for larger trucks which tend to be used for longer trips. While HVT 1 (commodity) had the lowest error for small trucks, the parameter values for *p* (1.001) and *gamma* (-951.468) raise concerns about the conceptual validity of these models when applied to this data subset. The fact that $p_0$ exceeds 1 in the NVE model indicates that there are more empty than loaded trips which in turn suggests that the trips captured in this subset are not limited to freight trips, but include passenger trips as well. This is not surprising considering that most of the trips in this subset were by pickup trucks.

The next best models for small trucks were the next simplest models (HVT 2) with *p* as a function of commodity flow having the next lowest relative error, followed by *p* as a function of distance, and then constant *p*. HVT 4 (distance) and HVT 4 (constant) also had very small relative errors (less than 2%) for small trucks.

For large trucks, both HVT 2 (distance) and HVT 1 (commodity) also had relative errors less than 2%. For semi-trailers, the model with the next lowest relative error was HVT 2 (commodity), but the coefficient for *beta* was positive which is not conceptually valid as it indicates that a zone’s attractiveness increases with increasing distance. The next best model is HVT 1 (commodity), followed by HVT 3 (commodity). HVT 3 (distance) actually has a lower relative error value, however its *beta* coefficient is also positive.

One of the most interesting trends in these results is that the simpler models appear to perform the best. In fact, the most complex model (HVT 4) produced conceptually invalid results for both of the larger truck data sets as their *beta* coefficient values were positive.

The results shown above enable a conceptual discussion of the relation between the probability of a zero order trip chain, *p*, and trip distance. Although the analyses are hampered by the structural limitations of the linear model assumed that, among other things, is not constrained to remain in the range valid for probabilities, the parameters of the *p* function do provide an indication of the about the overall trend of the relationship between *p* and distance. The meaning of these parameters is best appreciated by analyzing equation (24), that corresponds to equation (9). As can be seen in equation (24), *p* values that increase with distance indicate that for long trips zero order trip chains would dominate, because the term *(1-p)* may vanish. On the other hand, *p* values that decrease with distance indicate the opposite, i.e., that for long trips higher order trip chains would be the norm.

$$E(z_{ij}) = \frac{m_{ij}}{a_{ij}} + p\frac{m_{ji}}{a_{ji}} + (1-p)\gamma^* \sum_{h \neq j} \frac{m_{hi}}{a_{hi}} \left( P^h(j) P^h(E/j) \right) \qquad (24)$$

In the case of large trucks, *p* decreases with distance in all formulations (as evidenced by the negative value of $p_1$); while, for semi-trailers, *p* increases with distance (the results for small trucks are not discussed further because they are impacted by the flow of non-freight related trips). These results suggest two different types of behaviors that depend on the vehicle type: for large trucks, the longer the trip, the more likely they are to undertake a higher order trip chain; while for semi-trailers, the exact opposite happens, i.e., the longer the trip, the more likely is that a zero order trip chain would take place.

Conceptually, this makes sense because in developing countries like Guatemala, the cargoes are concentrated among a relatively small number of economic poles, from which cargo flows to the other regions of the country usually by semi-trailers. Since large trucks are used for regional distribution purposes, they are found to undertake long trip chains at relatively long distances. Small trucks complete the operation, since they are typically used to pick up cargoes from warehouses to individual stores or users.

## *6. Conclusions*

This paper considers enhanced formulations to model commercial vehicle empty trips. These formulations relax a limitation of the original formulations developed by Holguín-Veras and Thorson (2003a), i.e., the assumption of a constant probability of a zero order trip chain for the entire range of trip types. The formulations considered in the paper assumed that *p* was a function of the opposing commodity flow, or a function of the trip distance.

The consideration of a variable *p* improved the relative performance of the models in rather unexpected ways. As shown in the paper, in all three datasets considered, the simplest models, and the one with the smaller number of parameters, outperformed the most complex formulations.

In both 2-3 axle trucks and semi-trailers, *p* as a function of distance outperformed the other type of functions. (The results for small trucks were deemed not conceptually correct because of the passenger traffic in pickups.) The analyses of the results show that different vehicles exhibited different behaviors. In the case of 2-3 axle large trucks, the probability *p* decreases with distance; while in the case of semi-trailers *p* increases with distance. This implies that the probability that a semi-trailer is used to transport a cargo and return empty increases with trip distance; while the opposite happens with 2-3 axle trucks. However, these are far from definitive conclusions because of the structural limitations of the exploratory formulations used in the paper that, among other things are not bounded in the (0,1) interval.

In this context, the formulations considered here should be considered as nothing proofs of concept of more comprehensive formulations. Future research by the authors is expected to consider more complete formulations based on choice theory. In spite of the acknowledged limitations of the work, it is clear that considering variable *p* functions hold the potential to significantly improve the performance of empty trips models, which would hopefully facilitate the development of new paradigms of freight transportation modeling.

***References***

Fernández, J.E., de Cea, J. and A. Soto O. (2003), “A Multi-Modal Supply–Demand Equilibrium Model for Predicting Intercity Freight Flows” Transportation Research Part B, 37, pp. 615-640.

Hautzinger, H. (1984), “The Prediction of Interregional Goods Vehicle Flows: Some New Modeling Concepts,” Ninth International Symposium on Transportation and Traffic Theory, VNU Science Press, pp. 375-396.

Holguín-Veras, J. (2000), “A Framework for an Integrative Freight Market Simulation” Published by the IEEE 3rd Annual Intelligent Transportation Systems Conference ITSC-2000, Dearborn Michigan, October 2000, pp. 476-481.

Holguín-Veras, J. and E. Thorson. (2000) “An Investigation of the Relationships Between the Trip Length Distributions in Commodity-based and Trip-based Freight Demand Modeling.”

Journal of the Transportation Research Record No.1707, Transportation Research Board of the National Academies, Washington, D.C., pp. 37-48.

Holguín-Veras, J., (2002), "Revealed Preference Analysis of the Commercial Vehicle Choice Process," Journal of Transportation Engineering, Vol. 128, No. 4, American Society of Civil Engineers, July/August pp. 336-346.

Holguín-Veras, J. and E. Thorson (2000) "An Investigation of the Relationships between the Trip Length Distributions in Commodity-based and Trip-based Freight Demand Modeling", Transportation Research Record #1707 pp. 37-48.

Holguín-Veras, J., and E. Thorson (2003a), "Modeling Commercial Vehicle Empty Trips with a First Order Trip Chain Model" Transportation Research Part B 37, 129-148.

Holguín-Veras, J. and E. Thorson (2003b), "Practical Implications of Modeling Commercial Vehicle Empty Trips" Transportation Research Record #1833, pp. 87-94.

Holguín-Veras, J., J. C. Zorrilla, and E. Thorson (2005), "Modeling Commercial Vehicle Empty Trips: Theory and Application" XVI International Symposium of Traffic and Transportation Theory (ISTTT), pp. 585-608, H. Mahmassani (Editor), Elsevier.

Nelder, J. and R. Mead (1965), "A simplex method for function minimization" Computer Journal vol. 7, pp. 308-313.

Press, W., S. Teukolsky, W. Vetterling, and B. Flannery (1992) "Numerical Recipes in FORTRAN: The Art of Scientific Computing" 2nd Edition, Cambridge University Press, Cambridge, U.K.

Tamin, O.Z. and L.G. Willumsen (1988), "Freight Demand Model Estimation from Traffic Counts" PTRC Annual Meeting, University of Bath, England. Reprinted in Simplified Transport Demand Modeling, Juan de Dios Ortúzar, Ed. PTRC, 1992.