Theoretical Economics Letters, 2024, 14, 617-631
https://www.scirp.org/journal/tel
ISSN Online: 2162-2086
ISSN Print: 2162-2078

Equivalences of Freight Trip Generation Functional Forms: Formulations, Approximations and Error Formalization Issues

Jesus Gonzalez-Feliu

Center for Research in Innovation and Intelligence in Management (CERIIM), Excelia Business School, La Rochelle, France
Email: gonzalezfeliuj@excelia-group.com

How to cite this paper: Gonzalez-Feliu, J. (2024). Equivalences of Freight Trip Generation Functional Forms: Formulations, Approximations and Error Formalization Issues. *Theoretical Economics Letters, 14,* 617-631.
https://doi.org/10.4236/tel.2024.142033

Received: January 30, 2024
Accepted: April 22, 2024
Published: April 25, 2024

Abstract

The definition of freight intensity patterns, mainly on the form of freight trip or commodity generation, is essential in urban economics. Those patters are in general defined in a set of models that relate trips or commodity quantities to each individual activity size, in the form of employment mainly. Those models, which can have different functional forms, are defined at the level of each single establishment, but in some cases only aggregated zonal data is available, making it possible to define constant and linear models (since their formulations have a transitivity property) but not non-linear models directly, those last requiring the definition of individual employment for each establishment. This paper aims to overcome this limitation by proposing a forma aggregated formulation of four functional forms (constant, linear, logarithmic and potential) and defining, via mathematical transformation, equivalences based on quasi-arithmetic means, which are then approximated by the use of the arithmetic mean instead (which is calculable using aggregated data, where the total number of establishments and the total employment are known). The paper analyses those approximations and proposes a theoretical calculation of the maximum error those approximations can have, via the definition of the statistical limiting error as the limit of a percentage error calculated on those equivalences when the variability of data (then the standard deviation) is very important, covering 99.7% of the statistical distribution of this error. Results show that those errors are contained and using non-linear models, even with an approximated number of establishments, results on more accurate models than using linear forms when the most suitable model is non-linear.

Keywords

Urban Economics, City Logistics, Freight Trip Generation, Functional Form, Data Estimation, Error Analysis, Statistical Limiting Error

1. Introduction

Urban economics and development face many challenges, most of which are related to the production and consumption of goods and services within the city area (Holguín-Veras et al., 2019; Malhotra & Mishra, 2019). The characterization and the identification of freight transport demand (Ramfou & Sambracos, 2023), essential to address those challenges, is mainly done via freight trip generation (FTG) and freight commodity generation (FG) models (Holguín-Veras et al., 2011).

A functional form can be defined as the definition of the algebraic form of the function $f(X;a)$ that explains a phenomenon (or explained variable) Y (Lau, 1986), where $Y= f(X;a)$. In the current context, that of FTG, given an establishment I belonging to category a, the set of establishments of category a being noted as V^a, the number of trips T_i is in general related to the size of the establishment, approached by either the Employment or the Area (Sánchez-Díaz et al., 2016). In general, five following functional forms can be defined (Lagorio et al., 2018): constant, linear, logarithmic, potential and exponential.

Traditionally, FTG models followed constant estimation rates, mainly in the 1960's-1990's when only count data was available (Institute of Transportation Engineers, 2008). Then, at the beginning of the 21st Century, linear models related to categorial classifications of activities (Bastida & Holguin-Veras, 2009) were proposed in a goal of higher accuracy with respect to both constant rates per establishment and constant rates per employee (Holguín-Veras et al., 2014). Then, non-linear models, like logarithmic, potential and exponential ones, have been proposed and validated (Holguín-Veras et al., 2021). However, all those models are supposed to have a dataset of urban establishments with individual data concerning category of activity and employment. However, in some cities, only aggregated zonal data is available (i.e., total number of establishments and corresponding total workforce for a category of activity within a zone) without the possibility of disaggregating it at the individual level (Regal et al., 2023). In such cases, when needing to define freight trip generation intensity, only constant estimations and linear models, able to be directly aggregated to apply the available data, have been used, resulting in an important reduction of accuracy of the estimations (see Regal et al., 2023 for a comparison of constant and linear models using aggregate data without the possibility to disaggregate them to use non-linear models, and the analysis of the disparities in resulting estimations).

Deploying linear models for a non-linear phenomenon result in theselection a less suitable functional form, introducing a higher error with respect to reality than using the most suitable relationship. On another hand, using non-linear models in those contexts needs the proposal of an estimation of employment instead of the real disaggregated employment, resulting also in an error raise. To assess which of the two approximation approaches is more suitable, it is possible to calculate, formally (i.e. via mathematical decompositions and equivalences), those two errors and compare them on a function analysis basis.

This paper aims to formally analyze the main functional forms of freight trip generation, examining the aggregation patterns and possibilities, and propose a set of approximations to be used when individual data is not available but an aggregate (total values) appreciations is possible. The proposed approximations will be developed using a mathematical decomposition on the basis of quasi-arithmetic means, then related to the arithmetic mean as an approximation of the real quasi-arithmetic mean related respectively to the logarithmic and to the potential models. First, the main theoretical issues of freight trip generation main functional forms (constant, linear, logarithmic and potential) are defined, then a set of approximations to deal with the lack of disaggregated data are presented. After that, a formal analysis of the estimation error of those approximations with respect to the complete functional forms (using disaggregated data) is presented and commented.

2. Functional forms and Aggregation Issues: A Theoretical Approach

2.1. A Recall on Main Functional Forms in FTG

Freight Trip Generation (FTG) models rely on the definition of the most suitable functional form that relates trip rates to other variables, such as employment or area (Sánchez-Díaz et al., 2016; Gonzalez-Feliu & Sánchez-Díaz, 2019). In the current context, that of FTG, given an establishment i or category a, the set of establishments of category a being noted as V^a, the number of trips T_i^a can be estimated as a function of Employment of establishment i, Emp_i, as follows:

$$T_i^a = f^a\left(Emp_i\right) \tag{1}$$

In the present work, we chose to develop the relationships and equivalences with respect to only one variable, since literature shows that most non-linear models are related to only one variable, for the most Employment, but this reasoning can be made for any exploratory variable. The four functional forms selected for the analysis are the constant, the linear, the logarithmic and the potential (or power) relationships.

Constant estimations are formally defined as follows:

$$T_i^a = K^a \tag{2}$$

where K^a is a constant, mainly estimated by arithmetic mean of the FTG rates of a set of establishments of category a surveyed.

Linear models are defined as follows:

$$T_i^a = K^a + a^a \cdot Emp_i; i \in V^a \tag{3}$$

where K^a and a^a are parameters defined by linear regression.

Logarithmic relationships see trips related to the logarithm of the employment. The mathematical relationship can be defined as follows:

$$T_i^a = K^a + a^a \cdot \log\left(Emp_i\right); i \in V^a \tag{4}$$

where K^a and a^a are parameters defined by linear regression.

Potential relationships are those relating trips to a potential expression where employment is multiplied by a first coefficient a^a and powered by a second coefficient b^a:

$$T_i^a = a^a \cdot Emp_i^{b^a}; i \in V^a \tag{5}$$

$$\log\left(T_i^a\right) = \log\left(a^a\right) + b^a \cdot \log\left(Emp_i\right); i \in V^a \tag{6}$$

By renaming the constant term, we obtain the following relationship:

$$\log\left(T_i^a\right) = c^a + b^a \cdot \log\left(Emp_i\right); i \in V^a \tag{7}$$

In order to linearize the relationship and then allow its calibration via linear regression.

All those models can then be estimated by linear regression (Gonzalez-Feliu & Sánchez-Díaz, 2019) and constant models are estimated by calculating arithmetic means.

Holguín-Veras et al. (2011, 2013) propose for the first time a comparative analysis of models (mainly comparing constant to linear ones) for FTG, showing that the choice of the functional form plays a crucial role in the model's accuracy. Sánchez-Díaz et al. (2016) show the potential of logarithmic models when comparing them to linear ones. A generalization of those approaches to choose both the best aggregation level and functional form is made in Gonzalez-Feliu and Sánchez-Díaz (2019), where 4 functional forms are envisaged (constant, linear, logarithmic and potential), showing that the potential relationships has, in some categories, a better representation capability than the other forms, but not in all, so the choice of the functional form (which can be made with the abductive method proposed by the authors) is crucial in FTG modelling, more than the aggregation level of the categories.

Those models are defined at the establishment level, i.e. it is necessary to have individual data (i.e. at the level of each single establishment) of employment and area. This is possible when establishment databases including geolocation or a zoning categorization with a detail on individual Employment is available and accessible (Holguín-Veras et al., 2019), but in some cases this information is difficult to be processed or obtained and, at most, the available information is given at a zonal level, with a total number of establishments, total number of employments and, in the best cases, also the total areas, per category of activity. When collected data on FTG is not available, the use of FTG models can be a suitable estimation for policy and planning issues (Ambrosini et al., 2008; Gentile & Vigo, 2013). However, in a context where the explanatory variables are aggregated at a zonal level and the FTG models defined at an individual (establishment) level, it is necessary to define equivalent models to pass from individual to zonal level and vice-versa (Gonzalez-Feliu & Peris-Pla, 2017). This is direct for constant and linear models (as shown in Regal et al., 2023) but not for non-linear models.

2.2. Considerations on Zonal Aggregation Issues: From Individual to Zonal FTG Models

Given a zone *z*, and a category of establishments a, the FTG of the zone can be defined as the sum of the individual FTG of each establishment of category *a*, so:

$$FTG^z = \sum_a FTG_a^z \tag{8}$$

where

$$FTG_a^z = \sum_{i \in V^{a,z}} T_i^a \tag{9}$$

For constant models, if we consider that the number of establishments of category *a* in zone *z* is n_a^z, this sum can be expressed as follows:

$$FTG_a^z = \sum_{i \in V^{a,z}} K^a = K^a n \tag{10}$$

In the case of linear functional forms, a similar aggregated formula can be defined:

$$FTG_a^z = \sum_{i \in V^{a,z}} \left(K^a + a^a \cdot Emp_i \right) = K^a \cdot n_a^z + a^a \cdot \overline{Emp_a^z} \cdot n_a^z \tag{11}$$

Simplifying,

$$FTG_a^z = n_a^z \cdot \left(K^a + a^a \cdot \overline{Emp_a^z} \right) \tag{12}$$

where $\overline{Emp_a^z}$ is the average number of employees for all establishments of category *a* in zone *z*, and n_a^z the number of establishments of category *a* in zone *z*. This average value is an arithmetic mean calculated as follows:

$$\overline{Emp_a^z} = \frac{1}{n_a^z} \cdot \sum_{i \in V^{a,z}} Emp_i \tag{13}$$

For logarithmic and potential functional forms, an analogous development can be made. However, and since neither the logarithmic nor the potential functions have a transitivity property on sums, those types of aggregations are not direct. In other words, the sums do not lead to a direct aggregation that can be obtained knowing only the total workforce in the zone without having the detail of the individual workforce of each establishment. For the logarithmic relationship, the resulting aggregation is the following:

$$FTG_a^z = \sum_{i \in V^{a,z}} \left(a^a + b^a \cdot \log\left(Emp_i \right) \right) = n_a^z \cdot a^a + b^a \cdot \sum_{i \in V^{a,z}} \log\left(Emp_i \right) \tag{14}$$

Analogously, the potential form, or potential function, leads to the following mathematical formulation for FTG at the zonal level:

$$FTG_a^z = \sum_{i \in V^{a,z}} \left(a^a \cdot Emp_i^{b^a} \right) = a^a \cdot \sum_{i \in V^{a,z}} \left(Emp_i^{b^a} \right) \tag{15}$$

We can conclude that, in both cases, it is not possible to relate the aggregated FTG formulation proportionally to the total number of establishments and/or employees in zone *z*, since the functions are non-linear. However, it is possible to relate them to other means, and then examine the possibilities of approximation by substituting those means by the arithmetic one. This will be analyzed in the following sub-section.

2.3. Approximations for Logarithmic and Potential Models at a Zonal Level

Let us start by considering the logarithmic functional form, which was written above as follows:

$$FTG_a^z = n_a^z \cdot a^a + b^a \cdot \sum_{i \in V^{a,z}} \log\left(Emp_i\right) \tag{16}$$

Taking into account the properties of the logarithmic function we can obtain the following equivalence:

$$\sum_{i \in V^{a,z}} \log\left(Emp_i\right) = \log\left[\prod_{i \in V^{a,z}} \left(Emp_i\right)\right] = n_a^z \cdot \log\left(\overline{\overline{Emp_a^z}}\right) \tag{17}$$

Where $\overline{\overline{Emp_a^z}}$ is the geometric mean, calculated as follows:

$$\overline{\overline{Emp_a^z}} = \left[\prod_{i \in V^{a,z}} \left(Emp_i\right)\right]^{\frac{1}{n_a^z}} \tag{18}$$

So

$$\overline{\overline{Emp_a^z}} = \left[\prod_{i \in V^{a,z}} \left(Emp_i\right)\right]^{\frac{1}{n_a^z}} \tag{19}$$

$$\log\left(\overline{\overline{Emp_a^z}}\right) = \frac{1}{n_a^z} \cdot \log\left(\prod_{i \in V^{a,z}} \left(Emp_i\right)\right) \tag{20}$$

And finally

$$n_a^z \cdot \log\left(\overline{\overline{Emp_a^z}}\right) = \log\left(\prod_{i \in V^{a,z}} \left(Emp_i\right)\right) \tag{21}$$

After that, we take the potential functional form, which can be mathematized as follows (as shown above):

$$FTG_a^z = a^a \cdot \sum_{i \in V^{a,z}} \left(Emp_i^{b^a} \right) \tag{22}$$

Taking into consideration the definition of a quasi-arithmetic mean of *p*-order also known as power mean of exponent *p*), defined for a general variable *x* as follows:

$$M_p(x) = \left(\frac{1}{n} \cdot \sum_i \left(x_i^p\right)\right)^{1/p} \tag{23}$$

We can write

$$\left(M_p(x)\right)^p = \frac{1}{n} \cdot \sum_i \left(x_i^p\right) \tag{24}$$

So

$$n \cdot \left(M_p(x)\right)^p = \sum_i \left(x_i^p\right) \tag{25}$$

We can re-write the potential aggregated functional form as follows:

$$FTG_a^z = a^a \cdot n_a^z \cdot \left(M_{b^a}\left(Emp_a^z\right)\right)^{b^a} \tag{26}$$

Recapitulating, we can define the three possible functional forms for FTG at zonal level, which can be approximated, when the non-arithmetic mean is not able to be obtained, by the arithmetic mean (see **Table 1**).

We observe that for constant or linear estimators, the transitivity properties allow a direct aggregation knowing the total number of establishments and the total employment. For the logarithmic and potential functional forms, if individual data is not available, we can approximate the quasi-arithmetic means by an arithmetic one, since it is easy to estimate it knowing the total number of establishments and the total employment of a category a of activities in a zone *z*. The relationships between arithmetic, p-order and geometric means are known. Therefore, we can state that the values of those means can be also defined considering the following relationship.

If $b^a < 1$

$$\overline{\overline{Emp_a^z}} < M_{b^a}\left(Emp_a^z\right) < \overline{Emp_a^z} \tag{27}$$

If $b^a > 1$

$$\overline{\overline{Emp_a^z}} < \overline{Emp_a^z} < M_{b^a}\left(Emp_a^z\right) \tag{28}$$

Table 1. Exact functional forms and approximated estimates for aggregated data situations (own elaboration from conclusions of analyses presented above).

Form type	Exact functional form	Approximation
Constant	$FTG_a^z = n_a^z \cdot K^a$	
Linear	$FTG_a^z = n_a^z \cdot \left(K^a + a^a \cdot \overline{Emp_a^z}\right)$	
Logarithmic	$FTG_a^z = n \cdot a^a + b^a \cdot n_a^z \cdot \log\left(\overline{\overline{Emp_a^z}}\right)$	$FTG_a^{z*} = n \cdot a^a + b^a \cdot n_a^z \cdot \log\left(\overline{Emp_a^z}\right)$
Potential	$FTG_a^z = a^a \cdot n_a^z \cdot \left(M_{b^a}\left(Emp_a^z\right)\right)^{b^a}$	$FTG_a^{z*} = a^a \cdot n_a^z \cdot \left(\overline{Emp_a^z}\right)^{b^a}$

Source: own elaboration by aggregation of developments presented above.

So we can state than the use of an arithmetic mean as an approximation on the equivalence of the logarithmic functional form will give higher results than using the exact equivalence, as for the potential functional form, if the exponent is higher than one, and the approximation will be lower for the potential functional form with an exponent lower than 1. In the current literature, no exponent lower than zero has been found (Sánchez-Díaz et al., 2016; Gonzalez-Feliu & Sánchez-Díaz, 2019). Moreover, although theoretically possible, a negative exponent would represent that FTG is inverse potentially related to the size of the establishment, so the smallest an activity is, the highest it would generate, which is incoherent physically (Gonzalez-Feliu & Sánchez-Díaz, 2019). Therefore, we assume the hypothesis that in FTG, the potential functional form has always positive exponents.

Remains then to analyze the average error of using those using of approximations, which will be done via a numerical analysis as proposed in next section.

3. Using Non-Linear Functional Form Approximations versus Linear Functional Form Formulations: A Formal Analysis of Error

We have seen in section 2 two equivalences for FTG generation that allow providing aggregate results respectively for logarithmic and potential functional forms. The main use of those approximations is to introduce a non-linear behavior when only aggregate data is available so only a total number of establishments and a global employment (by category) is available, without the possibility to identity each establishment's individual characteristics.

Those two equivalences rely on the use of quasi-arithmetic means, which are not able to be calculated when only the total number of establishments and total employment is known for each zone of application. For those reasons, an approximation of those equivalences has been proposed by substitution of the quasi-arithmetic mean by a pure arithmetic mean. We recall that the main use of those approximations is to introduce a non-linear behaviour when only aggregate data is available so only a total number of establishments and a global employment (by category) is available, without the possibility to identity each establishment's individual characteristics. To represent non-linear behaviours without having disaggregate data, using the approximations proposed in section 2 will introduce an error that needs to be characterized and analyzed. This section proposes a formal analysis of this error and its limits.

Different error measures can be used in such analyses (Gonzalez-Feliu & Sánchez-Díaz, 2019). We propose here a formal analysis based on Mean Average Percentage Errors (MAPE), which are linear (allowing different transformations and able to be analyzed using equivalences and approximations). An alternative, the Mean Absolute Error (MAE), is not related to the proportion of the difference (Hodson, 2022), so a percentage error is chosen instead. Moreover, MAPE is appointed as giving the same weight to each error, where another similar error, the Root Mean Square Error RMSE (which avoids calculating absolute values by the square root of the quadratic error) highlight bigger errors (Chai & Draxler, 2014). This leads to lower values of error for MAPE with respect to RMSE.

The MAPE error metric is computed on the basis of the following equation:

$$MAPE_i = \left[\sum_{j=1}^{n} \left| \frac{FTG_j^{predicted} - FTG_j^{observed}}{FTG_j^{observed}} \right| \right] \Big/ n \tag{29}$$

where FTG_j are the FTG rates predicted/estimated for observation j.

3.1. Calculation of the Error for the Logarithmic Functional Form

Let us consider the logarithmic functional form $FTG_{a\,i}^{z}$ for an establishment i belonging to an activity category a in a zone z, and its approximation using the arithmetic mean $FTG_{a\,i}^{z\,*}$. If we aim to calculate the MAPE of this estimation with respect to the formal functional form, we obtain the following development:

$$MAPE_a^z = \left[\sum_{i=1}^{n} \left| \frac{FTG_{a\,i}^{z\,*} - FTG_{a\,i}^{z}}{FTG_{a\,i}^{z}} \right| \right] \Big/ n \tag{30}$$

For each single establishment i, we can write the percentage error $PE_{z\,i}^{a}$ as follows:

$$PE_{a\,i}^{z} = \frac{FTG_{a\,i}^{z\,*} - FTG_{a\,i}^{z}}{FTG_{a\,i}^{z}} \tag{31}$$

$$PE_{a\,i}^{z} = \frac{\left[a^a + b^a \cdot \log\left(\overline{Emp_a^z}\right) \right] - \left[a^a + b^a \cdot \log\left(Emp_{ai}^z\right) \right]}{\left[a^a + b^a \cdot \log\left(Emp_{ai}^z\right) \right]} \tag{32}$$

$$PE_{a\,i}^{z} = \frac{a^a - a^a + b^a \cdot \log\left(\overline{Emp_a^z}\right) - b^a \cdot \log\left(Emp_{ai}^z\right)}{a^a + b^a \cdot \log\left(Emp_{ai}^z\right)} \tag{33}$$

$$PE_{a\,i}^{z} = \frac{b^a \cdot \left[\log\left(\overline{Emp_a^z}\right) - \log\left(Emp_{ai}^z\right) \right]}{a^a + b^a \cdot \log\left(Emp_{ai}^z\right)} \tag{34}$$

$$PE_{a\,i}^{z} = \frac{b^a \cdot \log\left[\frac{\overline{Emp_a^z}}{Emp_{ai}^z} \right]}{a^a + b^a \cdot \log\left(Emp_{ai}^z\right)} \tag{35}$$

Emp_{ai}^z is a real variable, which is positive non-null (since it is defined at the workforce of establishment i, including the owner of the activity), as defined by Aubert and Routhier (1999).

Knowing the statistical distribution of the explanatory variable (the employment), it is possible to estimate the maximum value of this error as the statistical limiting error. This value is the limit of the MAPE function (which is a function of the Employment) for higher and lower values of the explanatory variable (the Employment). Assuming that the Employment follow a normal distribution of average $\overline{Emp_a^z}$ and standard error σ_z^a (which is a valid assumption on the basis of the Law of Large Numbers (Wonnacott & Wonnacott, 2015). In this case, we can use this Law to estimate a possible maximum gap value of the statistical distribution. Let us assume a target of having 95% of the statistical distribution. In this case, Emp_{ai}^z will be, in 99.7% of the cases, lower than $\overline{Emp_a^z} \pm 3\sigma_z^a$ (Wonnacott & Wonnacott, 2015). So the relationship can be written as follows:

$$\log\left[\frac{\overline{Emp_a^z}}{Emp_{ai}^z}\right] = \log\left[\frac{\overline{Emp_a^z}}{\overline{Emp_a^z} \pm 3\sigma_a^z}\right] = \left(\log\left[\frac{\overline{Emp_a^z} \pm 3\sigma_a^z}{\overline{Emp_a^z}}\right]\right)^{-1} \tag{36}$$

By the properties of the logarithm function, we obtain:

$$\log\left[\frac{\overline{Emp_a^z}}{Emp_{ai}^z}\right] = -\log\left[\frac{\overline{Emp_a^z} \pm 3\sigma_a^z}{\overline{Emp_a^z}}\right] = -\log\left[1 \pm 3CV_a^z\right] \tag{37}$$

Where CV_a^z is the coefficient of variation calculated as follows: $CV_a^z = \frac{\overline{Emp_a^z}}{\sigma_z^a}$.

Since in MAPE we consider only absolute value, we do not take into account the sign of the result. And aiming to take the maximum value of the absolute percentage error APE_{ai}^z, we can formulate the following relation:

$$APE_{ai}^z = \left|PE_{ai}^z\right| = \left|\log\left[\frac{\overline{Emp_a^z}}{Emp_{ai}^z}\right]\right| \le \log\left[1 + 3CV_a^z\right] \tag{38}$$

So

$$APE_{ai}^z \le \frac{b^a \cdot \log\left[1 + 3CV_a^z\right]}{a^a + b^a \cdot \log\left(\overline{Emp_a^z} + 3\sigma_a^z\right)} \tag{39}$$

And also

$$APE_{ai}^z \le \frac{b^a \cdot \log\left[1 + 3CV_a^z\right]}{b^a \cdot \log\left(\overline{Emp_a^z} + 3\sigma_a^z\right)} = \frac{\log\left[1 + 3CV_a^z\right]}{\log\left(\overline{Emp_a^z} + 3\sigma_a^z\right)} \tag{40}$$

If we calculate the limit for the highest levels of σ_a^z we obtain:

$$\lim_{\sigma_a^z \to \infty} \frac{\log\left[1 + 3CV_a^z\right]}{\log\left(\overline{Emp_a^z} + 3\sigma_a^z\right)} = \lim_{\sigma_a^z \to \infty} \frac{\log\left[1 + 3 \cdot \frac{\sigma_a^z}{\overline{Emp_a^z}}\right]}{\log\left(\overline{Emp_a^z} + 3\sigma_a^z\right)} \tag{41}$$

$$\cong \lim_{\sigma_a^z \to \infty} \frac{\log\left[3 \cdot \frac{\sigma_a^z}{\overline{Emp_a^z}}\right]}{\log\left(3\sigma_a^z\right)} \cong \lim_{\sigma_a^z \to \infty} \frac{\log\left(3\sigma_a^z\right)}{\log\left(3\sigma_a^z\right)} \cong 1 \tag{42}$$

So

$$MAPE_{a\,i}^{z} \le \left[\sum_{i} APE_{a\,i}^{z} \right] \Big/ n = \frac{n}{n} = 1 \tag{43}$$

We observe a limit for MAPE which is lower of equal to 1. Remains then to study the variation of the function. Since the function is a fraction in which both terms contain the standard error (measure on which we study the limits), and both terms inside the logarithmic functions are higher than 1, the result of each term is positive. When observing them, the numerator will be, for Emp higher or equal to 1, lower than the denominator. So the function is decreasing. In this case, for small standard errors, the value of APE in the numerator is close to log(1), which is 0, so each individual error will vary between 0 and 1 and the MAPE will also be on this internal, i.e. [0, 1].

3.2. Calculation of the Error for the Potential Functional Form

Let us then consider the potential functional form $FTG_{a\,i}^{z}$ for an establishment *i* belonging to an activity category *a* in a zone *z*, and its approximation using the arithmetic mean $FTG_{a\,i}^{z\,*}$. If we aim to calculate the MAPE of this estimation with respect to the formal functional form, we obtain the following development:

$$MAPE_{a}^{z} = \left[\sum_{i=1}^{n} \left| \frac{FTG_{a\,i}^{z\,*} - FTG_{a\,i}^{z}}{FTG_{a\,i}^{z}} \right| \right] \Big/ n \tag{44}$$

For each single establishment i, we can write the percentage error $PE_{z\,i}^{a}$ as follows:

$$PE_{a\,i}^{z} = \frac{FTG_{a\,i}^{z\,*} - FTG_{a\,i}^{z}}{FTG_{a\,i}^{z}} \tag{45}$$

$$PE_{z}^{a} = \frac{a^{a}\left(Emp_{z\,i}^{a}\right)^{b^{a}} - a^{a}\left(\overline{Emp_{z}^{a}}\right)^{b^{a}}}{a^{a}\left(Emp_{z\,i}^{a}\right)^{b^{a}}} \tag{46}$$

$$PE_{z}^{a} = \frac{a^{a}\left[\left(Emp_{z\,i}^{a}\right)^{b^{a}} - \left(\overline{Emp_{z}^{a}}\right)^{b^{a}}\right]}{a^{a}\left(Emp_{z\,i}^{a}\right)^{b^{a}}} = \frac{\left(Emp_{z\,i}^{a}\right)^{b^{a}} - \left(\overline{Emp_{z}^{a}}\right)^{b^{a}}}{\left(Emp_{z\,i}^{a}\right)^{b^{a}}} \tag{47}$$

$$PE_{z}^{a} = 1 - \frac{\left(\overline{Emp_{z}^{a}}\right)^{b^{a}}}{\left(Emp_{z\,i}^{a}\right)^{b^{a}}} \tag{48}$$

If we take the fraction and develop it following the same assumptions than for (37) to (44), we obtain

$$\frac{\left(\overline{Emp_{z}^{a}}\right)^{b^{a}}}{\left(Emp_{z\,i}^{a}\right)^{b^{a}}} = \left(\frac{\overline{Emp_{z}^{a}}}{Emp_{z\,i}^{a}}\right)^{b^{a}} \le \left(\frac{\overline{Emp_{z}^{a}}}{\overline{Emp_{z}^{a}} - 3\sigma_{z}^{a}}\right)^{b^{a}} \tag{49}$$

$$\left(\frac{\overline{Emp_z^a}}{\overline{Emp_z^a}-3\sigma_z^a}\right)^{b^a}=\left(\frac{\overline{Emp_z^a}-3\sigma_z^a}{\overline{Emp_z^a}}\right)^{-b^a}=\left(1-\frac{3\sigma_z^a}{\overline{Emp_z^a}}\right)^{-b^a}=\left(1-3\cdot CV_z^a\right)^{-b^a} \quad (50)$$

So

$$PE_z^a=1-\left(1-3\cdot CV_z^a\right)^{-b^a} \quad (51)$$

If we calculate the limit for the highest levels of σ_a^z we obtain:

$$\lim_{\sigma_z^a\to\infty}\left[1-\left(1-\frac{3\sigma_z^a}{\overline{Emp_z^a}}\right)^{-b^a}\right]\cong\lim_{\sigma_z^a\to\infty}\left[1-\left(\frac{3\sigma_z^a}{\overline{Emp_z^a}}\right)^{-b^a}\right]\cong\lim_{\sigma_z^a\to\infty}\left[1-\left(3\sigma_z^a\right)^{-b^a}\right] \quad (52)$$

$$\lim_{\sigma_z^a\to\infty}\left[1-\left(1-\frac{3\sigma_z^a}{\overline{Emp_z^a}}\right)^{-b^a}\right]\cong 1-0\cong 1 \quad (53)$$

So

$$MAPE_{ai}^z\le\left[\sum_i APE_{ai}^z\right]\Big/n=\frac{n}{n}=1 \quad (54)$$

The value for small values of sigma is close to 0 since the difference between the FTG rates obtained with the approximation and those with the exact equivalence will be near the same. Indeed:

$$\lim_{\sigma_z^a\to 0}\left[\frac{FTG_{ai}^{z\,*}-FTG_{ai}^z}{FTG_{ai}^z}\right]=\lim_{\sigma_z^a\to 0}\left[1-\left(\frac{\overline{Emp_z^a}}{\overline{Emp_z^a}-3\sigma_z^a}\right)^{b^a}\right]\cong 1-1=0 \quad (55)$$

So, MAPE varies in this case also inside the interval [0,1].

3.3. Discussion

The proposed analyses show that in both cases, the error between the approximation and the use of the real functional form is lower than or equal to 1. If we observe Gonzalez-Feliu and Sánchez-Díaz's (2019) results, which calculate MAPE and RMSE for different functional forms and different aggregations of categories, we can state that the error between the most suitable functional form (estimate) and the real data (observed), vary between 1 and 6, and the constant ones between 2 and 20. So adding 1 to the estimation error (which is in general between 1 and 20) does neither alter its nature nor its magnitude (the error remains of the same type and the order of magnitude remains similar). Authors did not assess MAPE for only linear functions but present the MAPE of the most suitable functional form for each category.

In this context, a more detailed analysis could be carried out to compare the suitability of the different models of approximations on a practical basis, but as an exploratory and theoretical analysis we can state that the proposed approximations remain of the same orders of magnitude than current errors using disaggregated functional forms (those disaggregated models can then be aggregated zonally as on Gonzalez-Feliu and Peris-Pla, 2017 or Pani et al., 2019). In any case, the proposed work shows that using those approximations allow to have very similar results, in terms of error, that using the real functional forms, and it appears that those approximations remain more suitable than using a linear function when the generation phenomenon is proven to not be linear.

This goes on the same direction than the results of Gonzalez-Feliu and Sánchez-Díaz (2019) and Gonzalez-Feliu et al. (2024): the definition of the most suitable functional form leads to a higher approximation than other transformations of data to allow using constant or linear functions. This work adds to that statement a formal proof that defining a suitable aggregated function (related to a quasi-arithmetic mean) and approximating it via the use of the arithmetic mean results on contained errors that seem to be lower than those of using linear functions to represent a non-linear phenomenon.

4. Conclusion

This paper proposed a formal and theoretical analysis of the possibilities of approximating non-linear functional forms for freight trip generation using an aggregate estimation (on the non-linear form) based on the definition of quasi-arithmetic means (geometric mean and power mean of order p respectively for logarithmic and potential models) and their exponent via the arithmetic mean. To validate those approximations, the definition of a Mean Absolute Percentage Error (MAPE) in a formal way and its development led to a formal analysis and the calculation of the limit (maximum error), which was in both cases of 1. Taking into account the values of MAPE observed in previous works (from 1 to 20 in purely constant models and from 1 to 6 when mixing functional forms, linear and non-linear), the error introduced by the approximation, although non-negligible, is of the same order, and appears being lower than the approximation via a linear generation (which MAPE is set between the two ranges presented above).

Those results show the interests of the proposed analysis on both theoretical and practical viewpoints. On a theoretical viewpoint, the determination of approximations via mathematical deduction justifies the proposed estimators and their internal coherence, then the error estimation analysis allows validating them. On a practical viewpoint, those estimators can be used when aggregated data at a zonal level is available but not individual data, giving the possibility of accounting of non-linear behaviours when they can be proven even if non-disaggregated data is available. The work also confirms and completes Gonzalez-Feliu and Sánchez-Díaz's (2019) conclusions showing that the choice of the most suitable functional form allows in all cases a more accurate estimation of FTG, even when disaggregated data is non-available, confirming that constant and linear estimations are not suitable for representing non-linear behaviours and a mean-based approximation remains more accurate than a linearization of the phenomena.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Ambrosini, C., Meimbresse, B., Routhier, J. L., & Sonntag, H. (2008). Urban Freight Modeling: A Review. In Taniguchi, E., & Thompson, R. G. (Eds.), *Innovations for City Logistics* (pp. 197-212). Nova Science.

Aubert, P. L., & Routhier, J. L. (1999). Freturb: Un modèle de simulation des transports de merchandises en ville. In *Transport Modes and Systems: Proceedings of the 8th World Conference on Transport Research* (pp. 531-544). Pergamon.

Bastida, C., & Holguín-Veras, J. (2009). Freight Generation Models: Comparative Analysis of Regression Models and Multiple Classification Analysis. *Transportation Research Record, 2097,* 51-61. https://doi.org/10.3141/2097-07

Chai, T., & Draxler, R. R. (2014). Root Mean Square Error (RMSE) or Mean Absolute Error (MAE). *Geoscientific Model Development Discussions, 7,* 1525-1534. https://doi.org/10.5194/gmdd-7-1525-2014

Gentile, G., & Vigo, D. (2013). Movement Generation and Trip Distribution for Freight Demand Modelling Applied to City Logistics. *European Transport/Trasporti Europei, 54,* 1-27.

Gonzalez-Feliu, J., & Peris-Pla, C. (2017). Impacts of Retailing Attractiveness on Freight and Shopping Trip Attraction Rates. *Research in Transportation Business & Management, 24,* 49-58. https://doi.org/10.1016/j.rtbm.2017.07.004

Gonzalez-Feliu, J., & Sánchez-Díaz, I. (2019). The Influence of Aggregation Level and Category Construction on Estimation Quality for Freight Trip Generation Models. *Transportation Research Part E: Logistics and Transportation Review, 121,* 134-148. https://doi.org/10.1016/j.tre.2018.07.007

Gonzalez-Feliu, J., Ramirez-Rios, D. G., Pani, A., & Bouraoua, S. (2024). A Categorization-Modelling Procedure to Estimate Urban Goods Transport Demand Minimizing the Number of Categories of Establishments. *Transportation Research Procedia.* (Preprint)

Hodson, T. O. (2022). Root Mean Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not. *Geoscientific Model Development Discussions, 15,* 5481-5487. https://doi.org/10.5194/gmd-15-5481-2022

Holguín-Veras, J., Jaller, M., Destro, L., Ban, X., Lawson, C., & Levinson, H. S. (2011). Freight Generation, Freight Trip Generation, and Perils of Using Constant Trip Rates. *Transportation Research Record, 2224,* 68-81. https://doi.org/10.3141/2224-09

Holguín-Veras, J., Jaller, M., Sánchez-Díaz, I., Campbell, S., & Lawson, C. T. (2014). Freight Generation and Freight Trip Generation Models. In L. Tavasszy, & G. de Jong (Eds.), *Modelling Freight Transport* (pp. 43-63). Elsevier. https://doi.org/10.1016/B978-0-12-410400-6.00003-3

Holguín-Veras, J., Ramírez-Ríos, D. G., Encarnación, T., Gonzalez-Feliu, J., Caspersen, E., Rivera-González, C., González-Calderón, C., & Da Silva Lima, R. (2019). Metropolitan Economies and the Generation of Freight and Service Activity: An International Perspective. In M. Browne, S. Behrends, J. Woxenius, G. Giuliano, & J. Holguín-Veras (Eds.), *Urban Logistics. Management, Policy and Innovation in a Rapidly Changing Environment* (pp. 19-51). Kogan Page.

Holguín-Veras, J., Ramirez-Rios, D., & Pérez-Guzmán, S. (2021). Time-Dependent Patterns in Freight Trip Generation. *Transportation Research Part A: Policy and Practice, 148,* 423-444. https://doi.org/10.1016/j.tra.2021.03.029

Holguín-Veras, J., Sánchez-Díaz, I., Lawson, C. T., Jaller, M., Campbell, S., Levinson, H. S., & Shin, H. S. (2013). Transferability of Freight Trip Generation Models. *Transportation Research Record, 2379,* 1-8. https://doi.org/10.3141/2379-01

Institute of Transportation Engineers (2008). *ITE Trip Generation: An ITE Informational.* Report, ITE.

Lagorio, A., Gonzalez-Feliu, J., & Pinto, R. (2018). Urban Freight Demand Estimation: A Probability Distribution Basedmethod. In *Proceedings of the 7th International Conference InInformation Systems, Logistics and Supply Chain* (pp. 415-422). INSA.

Lau, L. J. (1986). Functional Forms in Econometric Model Building. In *Handbook of Econometrics* (Vol. 3, pp. 1515-1566). Elsevier. https://doi.org/10.1016/S1573-4412(86)03006-4

Malhotra, G., & Mishra, S. (2019). Effect of Economic Growth on the Logistics Sector in India. *Theoretical Economics Letters, 9,* 210-222. https://doi.org/10.4236/tel.2019.91016

Pani, A., Sahu, P. K., Chandra, A., & Sarkar, A. K. (2019). Assessing the Extent of Modifiable Areal Unit Problem in Modelling Freight (Trip) Generation: Relationship between Zone Design and Model Estimation Results. *Journal of Transport Geography, 80,* Article 102524. https://doi.org/10.1016/j.jtrangeo.2019.102524

Ramfou, I., & Sambracos, E. (2023). Exploring the Use of Systems Dynamics for the Valuation of Freight Transport Time Changes. *Theoretical Economics Letters, 13,* 1418-1433. https://doi.org/10.4236/tel.2023.136080

Regal, A., Gonzalez-Feliu, J., & Rodriguez, M. (2023). A Spatio-Functional Logistics Profile Clustering Analysis Method for Metropolitan Areas. *Transportation Research Part E: Logistics and Transportation Review, 179,* Article 103312. https://doi.org/10.1016/j.tre.2023.103312

Sánchez-Díaz, I., Holguín-Veras, J., & Wang, X. (2016). An Exploratory Analysis of Spatial Effects on Freight Trip Attraction. *Transportation, 43,* 177-196. https://doi.org/10.1007/s11116-014-9570-1

Wonnacott, T. H., & Wonnacott, R. J. (2015). *Introductory Statistics for Business and Economics.* John Wiley & Sons.