

Article

Modelling Freight Trip Generation Based on Deliveries for Brazilian Municipalities

Leise Kelli de Oliveira ^{1,2,*} , Gracielle Gonçalves Ferreira de Araújo ², Bruno Vieira Bertoncini ³ , Carlos David Pedrosa ³ and Francisco Gildemir Ferreira da Silva ⁴ 

¹ Department of Transportation and Geotechnical Engineering, Federal University of Minas Gerais, Belo Horizonte 31270-901, Brazil

² Postgraduate Program in Civil Engineering, Federal University of Pernambuco, Recife 50670-901, Brazil

³ Postgraduate Program in Transport Engineering, Federal University of Ceará, Fortaleza 60455-760, Brazil

⁴ Economy Graduate Program, Universidade Federal do Ceará, Fortaleza 60020-060, Brazil

* Correspondence: leise@etg.ufmg.br

Abstract: Freight trip generation modelling is important for forecasting freight movements, and understanding freight movements is crucial to enabling sustainable freight transportation planning. The existing literature focuses on model development, and most of the previous models are estimated by ordinary least squares regression. However, few studies have carefully considered the OLS assumptions. The objective of this paper is to estimate freight trip generation models using deliveries to commercial establishments in Brazilian municipalities. A procedure is described to estimate models by ordinary least squares (OLS), and alternative techniques are considered to address the violations of the OLS assumptions. The analysis was conducted with data from 860 commercial establishments in nine Brazilian municipalities, and models were estimated for capital, non-capital, small, medium, and larger municipalities. The findings showed that alternative techniques to OLS regression can provide better-estimated parameters and more accurate results. Not evaluating the OLS assumptions could compromise the quality of the model and, consequently, planning using these models. Moreover, the results showed that the number of employees has a more significant influence in small cities and a lower influence in medium-sized municipalities. Finally, the findings demonstrated the importance of local models that include the municipalities' characteristics and that can support freight transportation planning. These models can also include sustainable strategies for freight transport.

Keywords: freight trip generation modelling; urban freight transport; ordinary least squares; robust regression



Citation: Oliveira, L.K.d.; Araújo, G.G.F.d.; Bertoncini, B.V.; Pedrosa, C.D.; Silva, F.G.F.d. Modelling Freight Trip Generation Based on Deliveries for Brazilian Municipalities. *Sustainability* **2022**, *14*, 10300. <https://doi.org/10.3390/su141610300>

Academic Editor: Guido Perboli

Received: 7 July 2022

Accepted: 17 August 2022

Published: 18 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Behavioural choices are influenced by needs, opportunities, and skills [1]. In addition, the demand for products and services to meet population needs results from behavioural decisions. Consumption relationships are closely related to the freight demand, which influences urban freight transport (UFT) planning. However, UFT planning is challenging, as demand is not always known, and surveys to obtain data are expensive. Freight demand modelling can then be used to support UFT planning.

Furthermore, cargo movements are much more complex than passenger transport given that cargo movements involve interactions between companies at different supply chain stages [2] and different stakeholders and decision makers throughout the process. However, those who should plan urban transport policies usually lack a detailed view of the complex factors that influence UFT [2]. Thus, several problems (e.g., congestion, noise, visual and atmospheric pollution, safety, and damage to the infrastructure) are reflections of the lack of sustainable planning and operation [3–5]. UFT directly affects the product cost.

In order to understand UFT as a complex phenomenon, Cassiano et al. [6] proposed a conceptual model to represent the operation of the last mile in urban areas. The model focused on impact measures that were capable of guiding decision-makers toward sustainable UFT. In addition, this model was based on (1) the relationship between urban activities' subsystems and the transport subsystem, and (2) on how this relationship impacts UFT operation and consequently the stakeholders. The model proposed by Cassiano et al. changes the classic idea of reducing the urban environment to the space where the freight trip generation process takes place. The framework in Cassiano et al. considers three major steps: UFT starts in the purchase process from the urban activities' subsystem, moves around the transportation system with the freight trip generation model, and reaches the destination. Therefore, freight transport models are important and can be used to plan freight policies [7]. Furthermore, understanding the movements of goods is essential to supporting sustainable planning.

However, freight transport modelling is still in the early stages of development in many developing countries [8]. For example, in Brazil, freight trip generation models (FTGM) were estimated for some economic sectors, such as pubs and restaurants [9,10], supermarkets [10], shopping centres [9], buildings under construction [11], and warehouses [12]. Moreover, FTGM were used to evaluate parking needs in historical city centres [13] and freight trip flows in urban areas [14]. As a consequence of the lack of knowledge related to freight movements, freight policies are flawed and do not address freight problems [15]. Therefore, modelling efforts are needed to support freight policies. However, the lack of data, political priorities, and the slow pace of the knowledge-building process are some of the reasons for these limited research efforts [8]. This context, which is described by authors in India, is similar in Brazil.

UFT planning in Brazil is neglected by transportation authorities and planning agencies. Furthermore, freight transport is considered a private activity with limited impacts on the local economy. However, freight transport is of great importance for economic development. Creating knowledge and showing the impact of cargo movements in Brazilian cities are important to include this activity in urban planning.

To the best of the authors' knowledge, no paper has addressed FTGM estimation using deliveries to commercial establishments in Brazilian municipalities, as proposed in this paper. Thus, this paper estimates FTGM considering retail deliveries in Brazilian municipalities, which are based on their administrative and populational characteristics. These differences in administrative and populational characteristics allow dissimilarities between the estimated models to be discussed, which is better for planning freight movements.

Additionally, FTGM are generally estimated using linear regression [8,16]. However, this technique has some assumptions that need to be validated to ensure the accuracy of the estimated models. However, few studies have considered these assumptions explicitly in the analysis, especially among the existing Brazilian studies. Thus, a procedure is used to estimate FTGM using ordinary least squares (OLS) regression, and alternative techniques are considered to address the violations of the OLS assumptions. This procedure is shown by estimating FTGM using deliveries to commercial establishments in Brazilian municipalities. Furthermore, this study shows that alternative techniques to linear regression can provide better-estimated parameters and more accurate results.

The contribution of this paper is twofold. First, FTGM are estimated using delivery data, which contributes to understanding freight movements in Brazilian cities. Second, this study shows the importance of evaluating the OLS assumptions in FTGM, which is not always addressed by other studies.

This paper is structured in five sections. Section 2 presents a review of the FTGM literature and discusses modelling issues for better model accuracy. Section 3 details the approach used, and an application is shown in Section 4. The conclusion is presented in Section 5.

2. Freight Trip Generation Models—The Literature

This paper conducted a systematic literature review to address the research question: “How does the literature address FTGM?” Figure 1 describes the literature selection procedure. Considering this research question, “freight trip generation” and “modelling” were selected as search keywords. The search was conducted in Scopus, Web of Science, and Google Scholar®. Google Scholar® was included since the authors knew beforehand that some classical papers would not be found in the other two databases. The search was limited to documents up to August 2021. Then, the search procedure yielded 708 documents: 160 in Scopus, 76 in Web of Science, and 476 in Google Scholar®. The inclusion criteria considered records (1) that were related to the research question and that estimated FTGM, (2) that were available online in full text, and (3) that were published in English or Portuguese. The exclusion criteria considered books, chapters, reports, and duplicated records. The inclusion and exclusion criteria resulted in 43 papers, whose abstracts were read. Another 21 papers were excluded, which led to 22 remaining papers for the analysis. Lastly, these 22 papers were read, categorised, and analysed.

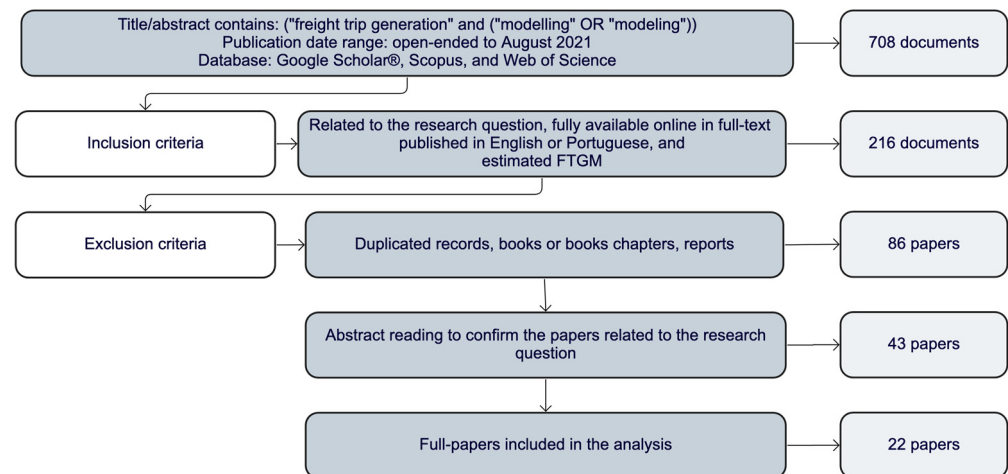


Figure 1. Literature selection procedure.

The literature review was then systematised concerning objectives, techniques, goodness-of-fit measures, and OLS assumptions. Table 1 summarises the literature.

The OLS is the most used regression technique to estimate FTGM. Moreover, alternative regression models have been used to solve the limitations of the OLS assumptions. Examples include generalised linear regression [11,16–18], logit or probit models [17,19,20], non-linear regression [13,21], negative binomial regression [22–24], and spatial models [22,23,25–27].

The number of employees and the establishment area are generally the independent variables used to explain FTGM [11]. In addition, the usual measures for evaluating the models’ goodness-of-fit are the *t*-test, the F-test, and the R-squared. Furthermore, the Akaike information criteria (AIC), the Bayesian information criterion (BIC), the root mean squared error (RMSE), and the mean absolute percentage error (MAPE) are used to compare and measure the accuracy of the models.

The reliability of the OLS estimators is verified by evaluating the OLS assumptions: endogeneity, multicollinearity, linearity of parameters, autocorrelated errors, homoscedasticity, and the normality of the error distribution [18]. Econometric-related studies have carefully addressed these assumptions, as reported by [28,29], and recommended such considerations for transportation data analysis [30]. However, few studies have evaluated all the OLS assumptions [16,31]. Therefore, not evaluating the OLS assumptions compromises the accuracy of the models, which could lead to errors in prediction and thus negatively impact the planning process.

Some papers focused on the role of the urban environment in freight trip movements [25,32]. Other studies considered freight trip attraction models using the business characteristics of commercial establishments [8,10,17,33,34]. Finally, several studies evaluated parking needs [13,16,35], the relationship between freight trip movements and accessibility [34], and the role of aggregating commercial establishments by categories [25,36]. FTGM can support both planning and public policies, as they can help regulate freight activities and forecast future scenarios [8,25,26,34,37].

The literature shows that considerable effort has been made to find the best models to explain FTGM. Moreover, the OLS regression is usually considered to estimate FTGM. However, many researchers have not used suitable techniques to support the associated methodological effort. For example, few studies have evaluated the OLS assumptions when estimating OLS models. Therefore, this paper intends to contribute to this issue by describing the fundamental steps for a reliable trip-generation estimation. This procedure is presented in the next section.

3. Procedure for Estimating FTGM

This section describes the procedure used to estimate FTGM. The OLS regression is a classical technique to estimate the coefficients of linear regression models. Such regression is well detailed in the econometric literature since it is a common method to develop theories or to test existing hypotheses. Nonetheless, in the transportation literature, OLS regression is a common procedure that presents some flaws. Thus, this work shows, in a simple way, how to use the classical linear regression theory to estimate FTGM. Figure 2 presents the steps proposed in this procedure, which are detailed in the next subsections.

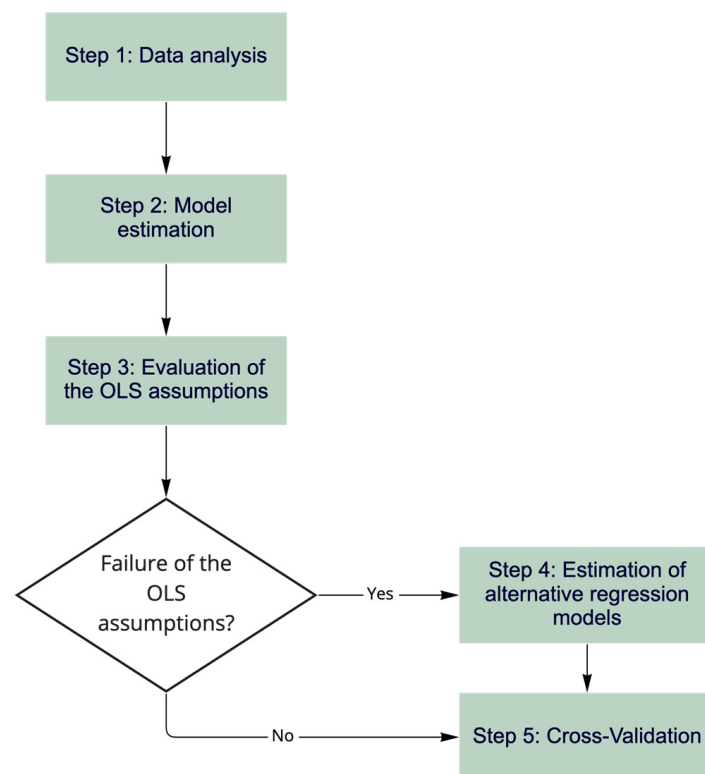


Figure 2. Procedure to estimate FTGM.

This study considered the usual variables in FTGM: The dependent variable (Y) is the number of deliveries per week, and the independent variables are the number of employees (X1) and the establishment areas (X2). However, other variables, if collected, can be used for FTGM. These data were obtained from an establishment-based survey that was conducted in Brazilian cities. Oliveira et al. [38] presented another study using the same data.

3.1. Step 1: Data Analysis

Figure 3 shows the sub-steps from the data analysis. Step 1.1 is related to the functional form of the linear models (i.e., model specification). A functional form refers to the algebraic formula that establishes a relationship between a dependent variable and the explanatory variables. However, linear regression models may suffer from functional form misspecification [28]. The Ramsey Regression Equation Specification Error Test (RESET) proposed by Ramsey [39] can be used to evaluate the general functional form misspecification. The null hypothesis considers that the functional form is correctly specified, and the rejection of the null hypothesis indicates that the model is misspecified [28]. The null hypothesis is rejected at a 95% significance level when the *p*-value is lower than 0.05. It should be noted that the RESET test is a functional form test; however, a model can also be misspecified due to omitted variables, which might not be identified by the RESET test [28]. Therefore, it is essential to ensure that the model variables explain the phenomenon well. The *lmtest* package [40] in the R environment was used to perform the RESET test.

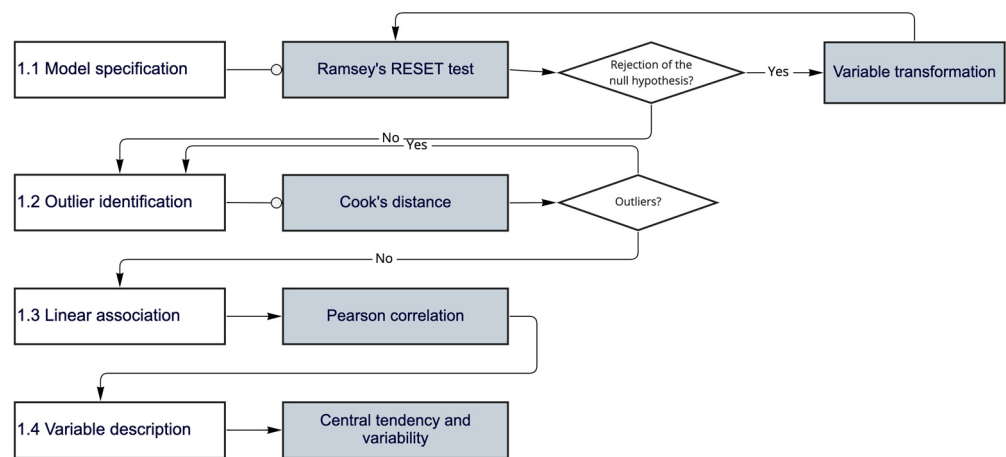


Figure 3. Step 1 flowchart: data analysis.

In a misspecified model, variable transformation is required to identify linear patterns. The most usual variable transformations are presented in Table 2. After the variable transformation, the RESET test can be used to verify the linear pattern of the functional forms.

Table 2. Variable transformation.

Functional Form	Original Form	Linearised Form	Notes
Linear	$Y = b_0 + b_1X$	$Y = b_0 + b_1X$	
Log-log	$Y = b_0X^{b_1}$	$\ln(Y) = B_0 + b_1\ln(X)$	$B_0 = \ln(b_0)$
Log-linear exponential	$Y = b_0b_1^X$	$\ln(Y) = B_0 + B_1X$	$B_0 = \ln(b_0)$ and $B_1 = \ln(b_1)$
Linear-log or logarithm-X		$Y = b_0 + b_1\ln(X)$	
Inverse	$Y = b_0 + b_1X$	$Y = b_0 + b_1/X$	
Log-inverse		$\ln(Y) = b_0 + b_1/X$	
Reciprocal-Y	$1/Y = b_0 + b_1X$	$Z = b_0 + b_1X$	$Z = 1/Y$
Double reciprocal	$1/Y = b_0 + b_1/X$	$Z = b_0 + b_1W$	$W = 1/X$ and $Z = 1/Y$
Quadratic	$Y = b_0 + b_1X_1 + b_2X_2$	$Y = b_0 + b_1X_1 + b_2W$	$W = X_2$

Assuming the functional form has been verified, step 1.2 concerns outlier identification, i.e., observations that lie at an abnormal distance from other observations in a random sample. Outliers are generally observed when modelling systems use real data. However, outliers can increase the variation of the explanatory variables and lead to models with lower predictive power. Thus, the decision to keep such observations in a regression analysis can be challenging [28]. In this study, outliers are removed. Since the data come

from information declared by interviewees, the questions may be misinterpreted. In addition, interviewees can answer the questions with inaccurate estimations. For example, respondents presented difficulties getting to know the establishment area. In addition, deliveries are known to fluctuate over time due to bank holiday dates. For this question, the number of deliveries was also an estimate given by the respondents, which was based on their experiences. Considering the size of the available database, removing the outliers would be beneficial to improve model estimation.

Thus, influential outliers were identified by the Cook's distance. First, the regression model was estimated and the Cook's distance was calculated to identify the influential observations, i.e., observations with a Cook's distance greater than four times the average. The influential observations were removed from the sample. The process was repeated as long as there were influential points in the sample. Cook [41] provides additional details about the mathematical procedure to identify influential measures. The outlier removal process provides the final database for estimating FTGM.

Step 1.3 requires measuring the level of association between the variables by the Pearson correlation coefficient. The correlation level varies from -1 (negative correlation) to $+1$ (positive correlation). A value below ± 0.38 indicates a weak correlation, a value between ± 0.4 and ± 0.69 indicates a moderate correlation, a value between ± 0.7 and ± 0.89 indicates a strong correlation, and a value above ± 0.9 indicates a very strong correlation. A strong correlation between the independent and the dependent variables is desirable.

Finally, step 1.4 summarises the data by calculating their central tendency and the associated variability.

3.2. Step 2: Estimation of the Linear Model

Figure 4 shows the procedures in step 2. The OLS technique was used to estimate the linear models. The functional form of the OLS model is represented by Equation (1), where Y is the dependent variable, X_1, \dots, X_n are the explanatory variables, β_1, \dots, β_n are the estimators, ϵ_n is the error, and n is the number of observations. The OLS regression minimises the difference between observed and predicted values (i.e., the sum of squared errors).

$$Y_n = \beta_1 + \beta_2 X_1 + \dots + \beta_n X_n + \epsilon_n \quad (1)$$

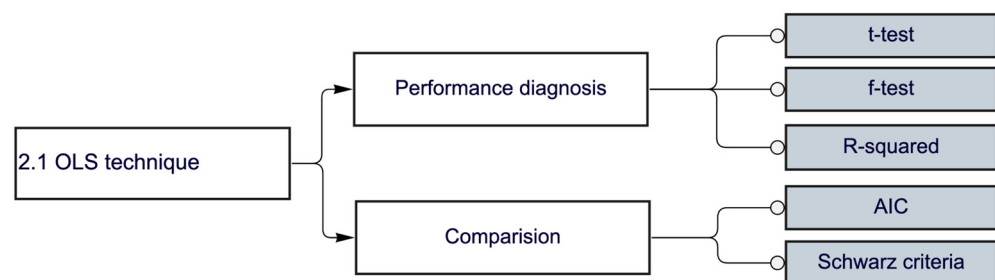


Figure 4. Step 2 flowchart: estimation of the linear models.

The goodness-of-fit tests indicate the models' performance. The usual goodness-of-fit measures are the hypothesis tests of estimators (T -test) and of the model (F -test), and the measure of the model fit (R -squared). The hypothesis tests of the estimators are conducted by the T -test, which identifies the linear association between the estimator and the dependent variable. The F -statistic tests the statistical significance of the independent variables in the linear regression. Moreover, the R -squared measure evaluates the prediction power of the OLS model. These statistics are usually verified in FTGM, as pointed out by the existing literature.

The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are used for comparing models. The AIC is a measure for scoring, comparing, and selecting the estimated models by different techniques. The model with the smallest AIC

is recommended [42]. The BIC, also called the Schwarz information criterion, is used for selecting econometric models. The model with the lowest BIC is preferred [43].

3.3. Step 3: Evaluation of the OLS Assumptions

Verifying OLS assumptions avoids incorrect parameter estimation [18]. Violation of the OLS assumptions results in misuse and inaccurate models. The main classical OLS assumptions are [29,30] (i) linearity of the parameters, (ii) uncorrelated regressors and errors, (iii) homoscedasticity, (iv) no autocorrelation between the errors, (v) multicollinearity, and (vi) normally distributed errors. Figure 5 details the required test to evaluate each of these assumptions.

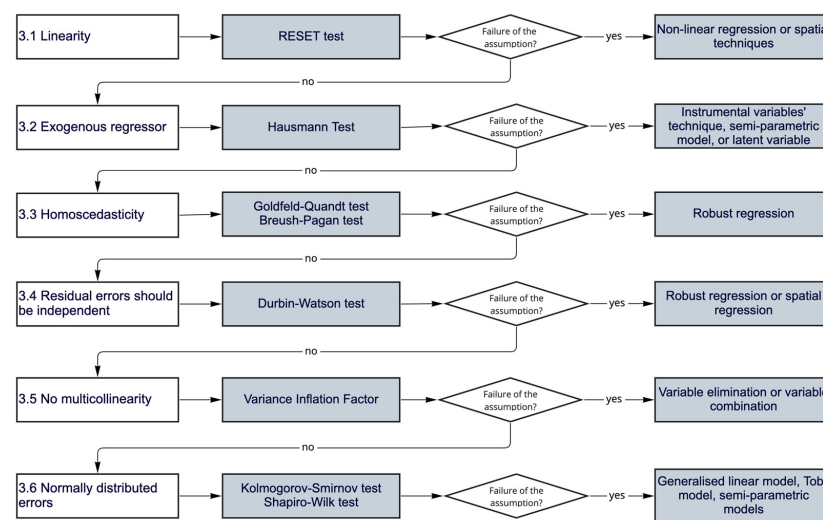


Figure 5. Step 3 flowchart: evaluation of the OLS assumptions.

The linearity assumption (step 3.1) considers a linear pattern of the parameters. The RESET test evaluates the models' functional form, as explained earlier. Additionally, the response variable Y must be continuous [30] and positive [29]. If a linear pattern is not identified, alternative techniques (e.g., non-linear regression model, generalised least-squares, maximum likelihood estimation, Bayesian regression, kernel regression, or gaussian process regression) must be used to estimate the parameters.

The assumption of uncorrelated regressors and errors (Step 3.2) is known as regressor exogeneity. An exogenous regressor is uncorrelated with the random error term, i.e., the regressor has zero variance with the error term. An endogenous regressor is the opposite of an exogenous regressor and is unsuitable for OLS estimation. The Durbin–Wu–Hausman (DWH) test (also called the Hausman specification test) checks exogenous regressors. The null hypothesis assumes an exogenous regressor, and the rejection of the null hypothesis indicates an endogenous regressor. The instrumental variable technique is an alternative to endogenous regressors.

Homoscedasticity or constant error variance is evaluated in step 3.3. Homoscedasticity means constant variance, whereas heteroskedasticity means variance in the errors. The Goldfeld–Quandt test and the Breush–Pagan test check for homoscedasticity. The null hypothesis of the Goldfeld–Quandt test is homoscedasticity, whereas the null hypothesis of the Breush–Pagan test is heteroskedasticity. The rejection of the null hypotheses indicates heteroskedasticity and homoscedasticity, respectively. Thus, it is expected that the null hypothesis of the Goldfeld–Quandt test will not be rejected and that the null hypothesis of the Breush–Pagan test will be rejected. Robust regression can be used to solve the heteroskedasticity problem.

Step 3.4 evaluates the assumption that the residual errors should be independent (i.e., without autocorrelation). The autocorrelation of the error terms is a typical problem in time-series analysis. The Durbin–Watson test is used to evaluate this assumption. The

null hypothesis assumes that the residual errors are independent, and the rejection of the null hypothesis indicates autocorrelated errors. Robust regression can be implemented to account for this violation.

Step 3.5 evaluates the multicollinearity assumption between the regressors. Multicollinearity is observed when the variables are correlated, and it implies inaccurate estimation. If multicollinearity is observed, the OLS estimators have no statistical significance. Multicollinearity can be detected by measuring the variance inflation factor (VIF). Alternatively, the Farrar–Glauber test can be used to measure the orthogonality of the variables, and the condition index can be calculated to diagnose this problem. Eliminating variables with $VIF > 10$ might solve the multicollinearity problem. In addition, combining variables may minimise the multicollinearity problem.

Finally, the assumption of normally distributed errors (step 3.5) is considered. This assumption is not a requirement for estimating linear regression models, but having normally distributed errors is required to make inferences about the model parameters [30]. The Kolmogorov–Smirnov and the Shapiro–Wilk tests check whether the errors are normally distributed (i.e., null hypothesis). Generalised linear models (GLM) are an alternative for count data, whereas tobit regression can be used for censored data. Moreover, a semi-parametric regression is an alternative for other cases [44]. More information about the OLS assumptions is provided in [29].

3.4. Step 4: Estimation of Alternative Regression Models

Violating OLS assumptions requires estimating models by alternative techniques. This paper used two methods due to the violation of OLS assumptions: robust regression and tobit regression.

Robust regression is an alternative to OLS regression with less restrictive assumptions. The residual standard error (RSE) measures the standard deviation of the residuals. A low RSE value indicates that the model fits the data. More information about robust regression can be found in [45]. Robust regression models were estimated using the MAAS package [45] in the R environment. Tobit regression is an alternative for censored data and for errors that are not normally distributed. For example, Figure 6 shows that the explanatory variable delivery is right-censored. Therefore, tobit regression can be used. Wooldridge [28] details the formulation of the tobit model. Tobit models were estimated using the VGAM package [46] in the R environment. The statistical significance of the estimated coefficients in both cases was evaluated by the z-value. Finally, the estimated models were compared by AIC and BIC.

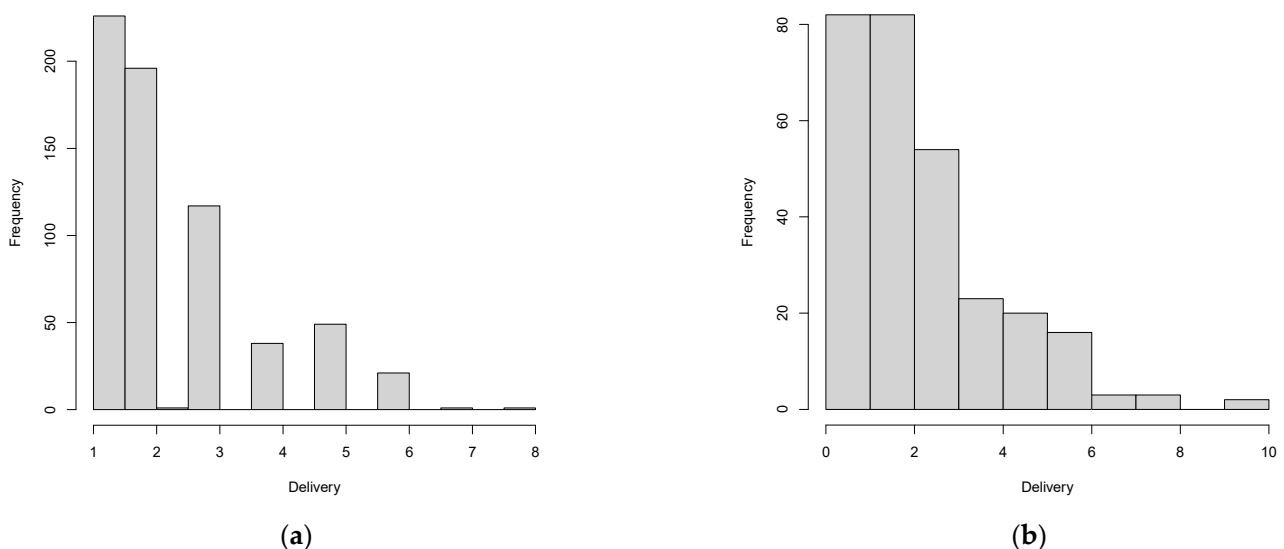


Figure 6. Cont.

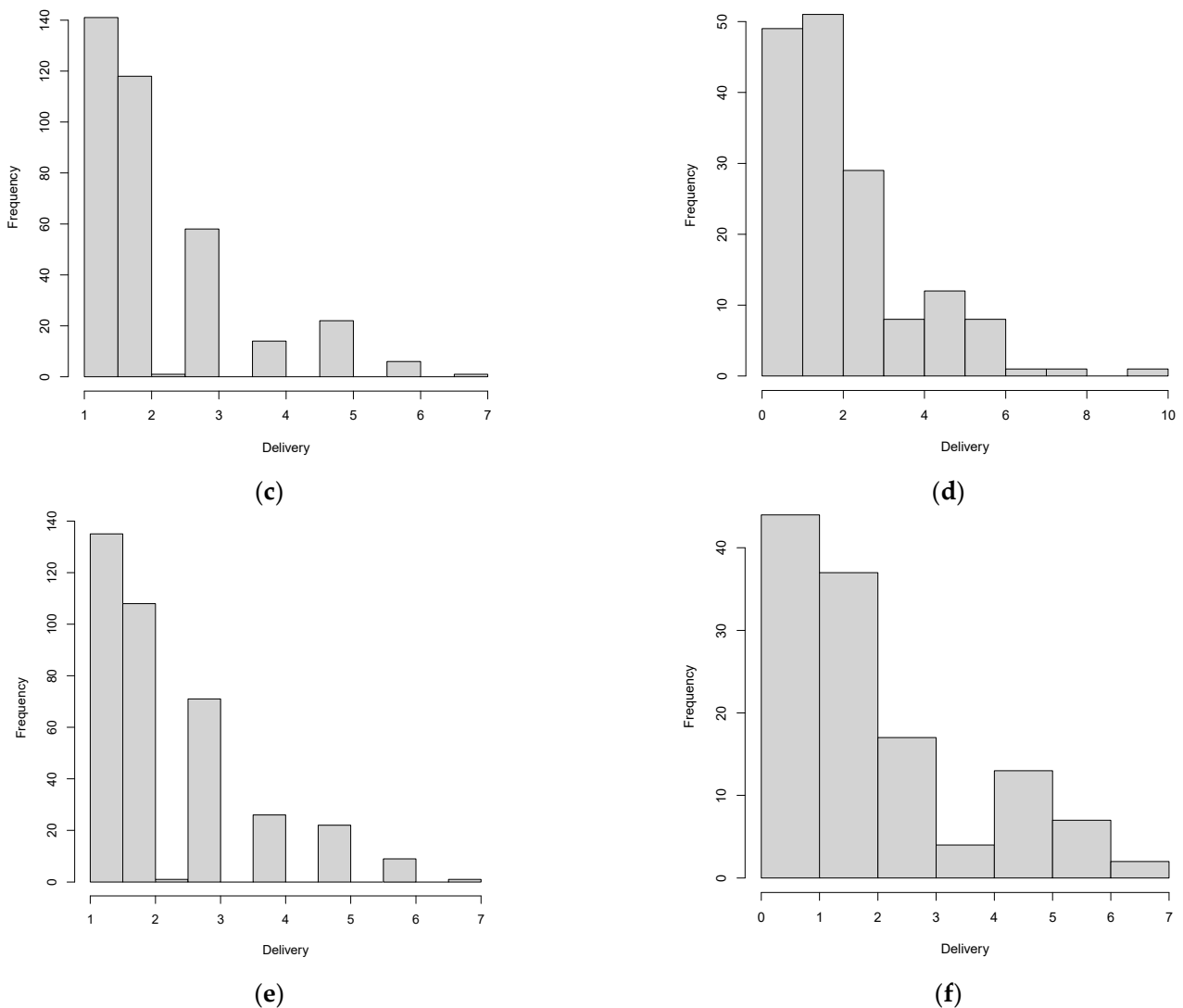


Figure 6. Histogram of the explanatory variable, by subset. (a) Brazil data, (b) capital cities, (c) non-capital cities, (d) larger cities, (e) medium cities, (f) small cities.

3.5. Step 5: Cross-Validation Analysis

Cross-validation measures the model's predictive performance using test data. Two cross-validation methods were used: leave-one-out cross-validation (LOOCV) and K-fold. LOOCV uses all observations and reduces potential bias. In this case, one data point is left out, and the model is estimated with the rest of the data set. Thus, the model's predictive power is evaluated using the data point left out and the test error associated with the prediction is recorded. The process is repeated for all data points, which enables the root mean squared error (RMSE) and the mean absolute error (MAE) to be computed. The RMSE is used to measure the difference between observed and predicted values. The MAE is the average of the absolute error values.

In the K-fold procedure, the dataset is split into k-subsets. Thus, one subset (testing data) is selected for testing and the model is estimated with the other subsets (training data). The model is predicted using the testing data, and the prediction error is calculated. This process is repeated until k subsets are used as a testing set. Finally, the RMSE and the MAE are computed.

4. Results

A questionnaire was designed to collect data related to the area of the commercial establishment, the number of employees, and the number of deliveries per week. The researchers visited several commercial establishments and invited the employers to answer the questionnaire. This process resulted in data from 860 commercial establishments from nine Brazilian cities: Belo Horizonte, Betim, Caruaru, Contagem, Divinópolis, Itabira, Nova Lima, Palmas, and Quixada. In the analysis, no distinction was made between the various economic sectors of the commercial establishments. However, we acknowledge the importance of the economic sector for FTGM and plan to address this issue in the future.

Data grouping was used to analyse differences in the models generated for cities with similar characteristics. The administrative and population characteristics were then considered, and the models were estimated for the following subsets: (i) **all data**; (ii) state **capital** cities (Belo Horizonte and Palmas); (iii) **non-capital** cities (Betim, Caruaru, Contagem, Divinópolis, Itabira, Nova Lima, and Quixada); (iv) **larger cities**, i.e., cities with a population greater than 50,000 inhabitants (Belo Horizonte, Betim, and Contagem); (v) **medium cities**, i.e., cities with a population between 100,000 and 500,000 inhabitants (Caruaru, Divinópolis, Itabira, and Palmas); and (vi) **small cities**, i.e., cities with a population lower than 100,000 inhabitants (Nova Lima and Quixada).

Table 3 shows the Ramsey RESET test results to evaluate the functional form. The linear functional form was unsuitable for all subsets, i.e., the null hypothesis of the Ramsey RESET test was rejected, and the functional form was not correctly specified. The log–log functional transformation was suitable for all subsets. Moreover, the linear–log, the inverse, and the log–inverse transformations were suitable for some groups: linear–log functional form for the subsets “all data,” “capital,” “non-capital cities,” and “larger cities”; inverse functional form for the subsets “larger cities,” “medium cities,” and “small cities”; and the log–inverse functional form for the subset “non-capital cities.” The results showed the importance of evaluating the functional form to obtain an accurate linear regression model. Therefore, unfamiliarity can produce biased estimated models by simply using inappropriate functional forms.

Table 3. Ramsey RESET results.

Variable Transformation	All Data	Capital	Non-Capital Cities	Larger Cities	Medium Cities	Small Cities
Linear	4.90 (0.01) ^{NS}	13.76 (0.00) ^{NS}	4.016 (0.02) ^{NS}	3.20 (0.04) ^{NS}	21.64 (0.00) ^{NS}	7.04 (0.00) ^{NS}
Log–log	0.234 (0.79)	0.11 (0.89)	1.06 (0.35)	1.13 (0.33)	0.24 (0.79)	3.15 (0.05)
Log–linear	17.34 (0.00) ^{NS}	3.21 (0.04) ^{NS}	13.83 (0.00) ^{NS}	3.76 (0.02) ^{NS}	3.94 (0.02) ^{NS}	17.29 (0.00) ^{NS}
Linear–log	2.133 (0.12)	0.93 (0.40)	2.37 (0.09)	0.88 (0.42)	3.86 (0.02) ^{NS}	7.58 (0.00) ^{NS}
Inverse	4.62 (0.01) ^{NS}	17.68 (0.00) ^{NS}	5.97 (0.00) ^{NS}	1.30 (0.28)	1.06 (0.35)	20.85 (0.00)
Log–inverse	28.03 (0.00) ^{NS}	45.48 (0.00) ^{NS}	0.74 (0.07)	41.50 (0.00) ^{NS}	3.45 (0.03) ^{NS}	4.46 (0.01) ^{NS}

^{NS} Without statistical significance.

The well-specified model (i.e., the log–log functional form) was considered to identify the outliers with the Cook’s distance. The outliers were then removed from the sample. Table 4 shows the variables’ descriptive statistics with and without outliers. The data with outliers presented a high variation, which showed the heterogeneity of commercial establishment characteristics in urban areas. For example, in the same downtown area, there are big stores and nano stores, which were all considered in the sample. After removing the outliers, the dataset was considered suitable for modelling since influential points were not considered. Moreover, the sample variation without outliers was reduced. In all the analysed cases, the sample was greater than 20 observations, which is the minimal sample size recommended for estimating OLS models (Hair et al., 2019). The data without outliers showed that the number of deliveries per week varied from 2.11 (non-capital cities) to 2.67 (capital cities). The area varied from 112.72 (medium cities) to 132.87 (small cities), and the number of employees varied from 5.44 (medium cities) to 7.93 (small cities).

Table 4. Descriptive statistics of the variables.

Outliers	Model	Sample	Deliveries Per Week		Area		Employees	
			Ave.	Sd.	Ave.	Sd.	Ave.	Sd.
With outliers	All data	860	3.23	4.57	150.35	223.01	7.08	7.62
	Capital	374	4.12	5.81	170.40	146.74	7.00	7.38
	Non-capital cities	486	2.53	3.19	134.95	201.79	7.15	7.80
	Larger cities	221	7.96	5.07	150.50	192.52	4.14	8.90
	Medium cities	468	3.18	4.95	154.00	230.00	6.55	6.40
	Small cities	171	2.21	1.75	140.10	239.41	7.42	8.73
Without outliers	All data	650	2.32	1.39	122.81	174.46	6.48	7.09
	Capital	285	2.67	1.73	130.85	155.81	6.03	6.32
	Non-capital cities	361	2.11	1.26	119.79	191.79	6.98	8.02
	Larger cities	160	2.51	1.64	131.03	163.29	7.16	7.29
	Medium cities	373	2.26	1.32	112.72	142.24	5.44	5.12
	Small cities	124	2.44	1.66	132.87	195.72	7.93	8.62

The results from the Pearson correlation show no patterns between the variables. However, a strong correlation was observed between the independent variables in the “capital” and “small cities” dataset, and a moderate correlation was observed in the other cases. In addition, weak or moderate correlations were observed in the datasets between the independent variables and the dependent variables. This preliminary result indicates that the OLS technique was not suitable for the datasets, as presented forward.

Table 5 shows the models’ estimated parameters. The estimated parameters had no statistical significance in the “capital” and “non-capital cities” models, which indicates that these variables did not influence FTGM. However, other models presented statistical significance in the F-test, which suggests that they could explain FTGM. The explanatory power of the model (R-squared) varied from 0.40 (non-capital cities) to 0.52 (capitals). The performance diagnosis parameters show that the models could be used for FTGM generation, except for the capital and non-capital models. These two models did not present statistical significance for the parameters. However, performance diagnosis is insufficient to ensure the non-violation of the OLS assumptions. Other statistical tests should be run for it.

Table 5. Estimated parameters using OLS.

Model	Variables	Estimated Parameters	T-Value	F-Statistics	R ²	AIC	BIC
All data	Intercept	−0.22	−2.53 **	182.6 ***	0.46	844.15	862.06
	Area	0.07	2.70 **				
	Employees	0.39	12.88 ***				
Capital	Intercept	−0.18	−1.02	61.81 ***	0.52	441.96	456.57
	Area	0.08	1.78				
	Employees	0.41	8.53 ***				
Non-capital cities	Intercept	−0.20	−2.18 *	156.1 ***	0.40	369.47	385.02
	Area	0.02	0.80				
	Employees	0.44	11.39 ***				
Larger cities	Intercept	0.69	2.88 ***	35.59 ***	0.51	243.25	255.55
	Area	−0.19	−3.03 ***				
	Employees	0.56	8.03 ***				
Medium cities	Intercept	−0.42	−3.96 ***	90.79 ***	0.47	492.57	508.26
	Area	0.17	5.50 ***				
	Employees	0.26	6.39 ***				
Small cities	Intercept	0.59	2.69 ***	116.4 ***	0.41	133.39	144.68
	Area	−0.40	−5.46 ***				
	Employees	1.06	12.40 ***				

Significance codes: *** p -value < 0.001; ** p -value < 0.01; * p -value < 0.05.

Table 6 shows the analysis of the OLS assumptions. The results show that no model met all the OLS assumptions. The assumptions of multicollinearity, linearity, and homoscedasticity were met in all models. However, the endogeneity assumption failed for the employees' estimated parameters in capital and non-capital cities. In addition, all models had autocorrelated errors, and the errors were not normally distributed. Therefore, since the OLS assumptions were not entirely met, the linear regression models estimated by OLS should not be used to forecast freight trip demand. This is different from the conclusions if solely analysing performance diagnosis parameters.

Table 6. Analysis of the OLS assumption (DW = Durbin–Watson; GQ = Goldfeld–Quandt; BP = Breusch–Pagan; KS = Kolmogorov–Smirnov; SW = Shapiro–Wilk).

Model	Variables	Endogeneity	Multicol.	Linearity	Autocorr. Errors	Homoscedasticity	Normality of Errors
		Hauss. Test	VIF	RESET	DW Test	GQ Test	BP Test
All data	Area	166 (0.00)	1.68	0.11 (0.89)	1.62 (0.00)	1.09 (0.22)	0.28 (0.87)
	Employees	7.28 (0.01)	1.68				
Capital	Area	72.76 (0.00)	1.35	0.33 (0.72)	0.99 (0.00)	1.07 (0.34)	1.10 (0.58)
	Employees	3.15 (0.08) ^{NS}	1.35				
Non-capital cities	Area	129.7 (0.00)	2.17	0.58 (0.56)	1.80 (0.03)	1.20 (0.10)	0.38 (0.82)
	Employees	0.64 (0.42) ^{NS}	2.17				
Larger cities	Area	64.53 (0.00)	1.65	0.06 (0.94)	0.84 (0.00)	1.04 (0.44)	0.97 (0.61)
	Employees	9.16 (0.00)	1.65				
Medium cities	Area	40.81 (0.00)	1.59	1.12 (0.33)	1.55 (0.00)	1.10 (0.25)	1.26 (0.53)
	Employees	30.24 (0.00)	1.59				
Small cities	Area	153.8 (0.00)	3.30	0.14 (0.87)	1.73 (0.06)	0.67 (0.94)	3.62 (0.16)
	Employees	166 (0.00)	1.68				

^{NS} Without statistical significance.

Robust regression can be used when the error terms are not autocorrelated or not normally distributed. Using this technique, the statistics are well measured even with bias, which prevents false positive decisions. Except for the “small cities” model, the other models violated the assumptions of autocorrelated errors and that the errors were normally distributed. Thus, robust regression models were estimated, and the estimated coefficients are shown in Table 7. The estimated parameters had no statistical significance in the “capital” and “non-capital cities” models. The robust regression models had slightly higher AIC and BIC values when compared to the previous models. However, this technique provided more reliable results since the OLS assumptions were violated.

Tobit regression can be used for censored data and for errors that are not normally distributed. Table 8 shows the estimated parameters using this technique. The estimated parameters had no statistical significance in the “non-capital” and “larger cities” models. Tobit regression could be an alternative to OLS linear regression models. However, the robust regression models provided lower AIC and BIC values when compared to tobit regression. Thus, the estimated model using robust regression provided more accurate parameters. However, the estimated parameters were not significant for the “non-capital” and “larger cities” models.

Cross-validation was conducted only for the subsets “all data,” “larger cities,” “medium cities,” and “small cities.” These models were estimated by robust regression. For these subsets, the estimated models using the log–log functional form and the robust regression had statistical significance. Thus, the cross-validation technique was performed to verify the accuracy of the estimated models. Table 9 shows the RMSE and the MAE values using LOOCV and K-fold. In all cases, low error values were obtained, which shows the accuracy of the estimated robust regression models.

Table 7. Estimated parameters using robust regression.

Model	Variables	Estimated Parameters	z-Value	Residual Standard Error	AIC	BIC
All data	Intercept	−0.24	−2.53 **	0.51	844.47	862.38
	Area	0.07	2.49 **			
	Employees	0.40	12.07 ***			
Capital	Intercept	−0.21	−1.09	0.61	441.99	456.60
	Area	0.08	1.76 ′			
	Employees	0.42	8.17 ***			
Non-capital cities	Intercept	−0.20	−2.11 *	0.46	369.66	385.22
	Area	0.02	0.63			
	Employees	0.45	11.26 ***			
Larger cities	Intercept	0.69	2.75 **	0.58	243.26	255.56
	Area	−0.20	−2.92 ***			
	Employees	0.56	7.69 ***			
Medium cities	Intercept	−0.43	−3.79 ***	0.54	492.82	508.51
	Area	0.17	5.12 ***			
	Employees	0.27	6.19 ***			
Small cities	Intercept	0.57	2.43 *	0.38	492.82	508.51
	Area	−0.40	−5.10 ***			
	Employees	1.06	11.67 ***			

Significance codes: *** p -value < 0.001; ** p -value < 0.01; * p -value < 0.05; ′ p -value < 0.1.

Table 8. Estimated parameters using tobit regression.

Model	Variables	Estimated Parameters	z-Value	Log-Likelihood	AIC	BIC
All data	Intercept	−0.79	−6.04 ***	−596.36	1200.7	1218.6
	Area	0.11	2.89 ***			
	Employees	0.55	12.11 ***			
Capital	Intercept	−0.59	−2.42 **	−283.12	574.24	588.85
	Area	0.10	1.67 ′			
	Employees	0.54	8.23 ***			
Non-capital cities	Intercept	−0.79	−5.34 ***	−292.91	593.81	609.37
	Area	0.05	1.05 (0.29)			
	Employees	0.63	10.27 ***			
Larger cities	Intercept	0.52	1.62	−155.27	318.54	330.84
	Area	−0.26	−2.91 ***			
	Employees	0.74	7.74 ***			
Medium cities	Intercept	−1.18	−6.93 ***	−342.81	693.62	709.31
	Area	0.26	5.58 ***			
	Employees	0.39	6.33 ***			
Small cities	Intercept	0.52	1.72 ′	−91.33	190.66	201.94
	Area	−0.53	−5.00 ***			
	Employees	1.34	10.23 ***			

Significance codes: *** p -value < 0.001; ** p -value < 0.01; ′ p -value < 0.1.

Table 9. Cross-validation results from robust regression.

Technique	Statistics	All Data	Larger Cities	Medium Cities	Small Cities
LOOCV	RMSE	0.343	0.397	0.307	0.322
	R-squared	0.178	0.093	0.18	0.346
	MAE	0.262	0.307	0.238	0.241
K-fold	RMSE	0.343	0.39	0.305	0.316
	R-squared	0.188	0.147	0.214	0.408
	MAE	0.262	0.305	0.238	0.241

5. Discussion of Results

FTGM are critical to understanding freight transport flows. However, as explored in the literature review, some models reported in the literature have no statistical validation, which motivated the development of this study. Moreover, verifying the OLS assumptions is critical for OLS linear regression.

First, this study showed the importance of analysing the functional form of linear regression models. The Ramsey RESET test indicated that the linear functional form should not be used, although this form is common in the transportation literature to estimate linear regression models. Thus, predicting the linear pattern of a phenomenon should be confirmed by statistical tests.

Although OLS models have statistical validity, not all the assumptions of OLS regression were met. Thus, estimating OLS models without evaluating their assumptions would generate biased estimates, and the use of such models would generate distorted results.

However, this problem can be addressed by estimating FTGM using alternative regression techniques. There are specific regression techniques for each violation of the OLS assumptions, which enable FTGM to be accurately estimated. The results in this study indicated that the robust regression provided models for “all data”, “larger cities”, “medium cities”, and “small cities”. Therefore, finding the modelling technique that best fits the dataset is important for proper model estimation.

The estimated models presented different characteristics considering the several city groups. However, the coefficient signs show that the estimated parameters of the “all data” model had the same sign as the “medium cities” model: The intercepts had a negative sign, and the independent variables had a positive sign. Thus, the size of the “medium cities” subset could influence the results of the “all data” model. Looking at and understanding the data is a crucial step in any modelling framework.

For the independent variables, the estimated parameter for the area presented more influence on medium cities and lower influence on small cities. Conversely, the number of employees presented more influence on small cities and lower influence on medium-sized cities. Thus, FTGM are different in Brazilian cities. Therefore, comparing the AIC and BIC values, FTGM can be used to predict freight trip generation considering the estimated parameters for larger, medium, and small cities.

Not properly considering autocorrelation, endogeneity, multicollinearity, and heteroscedasticity implies calibrating biased parameters, which can result in approving intervention that will result in adverse effects (i.e., different from the intended effects). This occurs because violating these assumptions results in reducing or amplifying the magnitude of the parameters. For example, a variable might indicate that reducing fares of a transportation mode implies an increase in demand. However, this parameter could have an opposite sign if the methodological procedures in this paper (i.e., analysing the method assumptions) were followed. Then, if the wrong model is used, decision-makers could make the operation of this mode of transport unfeasible. By reducing the fares, the profit margin of the operation is also reduced, which could result in frequent damage to the operator. Consequently, this situation could result in bankruptcy or abandoning the operation.

Additionally, the results indicate similar signs in all models, which shows that, regardless of the technique, the estimates suggest the correct direction of the effect of the explanatory variables on the dependent variable. However, violating the OLS assumptions compromises the metrics for evaluating the statistical significance and the magnitude of the effects. This requires an alternative estimation, such as using robust regression. For example, results indicate that the area parameter in the capital model is significant at the 10% level. If the data have enough quality to consider this level of significance, then the OLS model would induce the policymaker to disregard the area as a policy variable. Therefore, inefficient decisions would be made for capital cities. Conversely, if the model aims to predict freight trips, the robust regression would be more suitable since this tech-

nique presented lower AIC values. This also occurs in small cities, where the OLS is more appropriate than the robust regression.

In addition to the modelling process, the results influence freight transportation planning. The FTGM-estimated coefficients vary depending on the cities' characteristics. The models obtained in this research estimated freight movements according to the population characteristics of the cities (small, medium, and larger cities). As the models can be used to estimate freight flows in different cities, the models can also be considered to elaborate public policies to improve freight transport. This occurs because commercial establishments have different characteristics. For example, Cheah et al. [47] suggested using FTGM for evaluating building-level urban logistic management initiatives. Moreover, Silva et al. [13] evaluated the usage of on-street parking using FTGM. Suitable strategies can be set to reduce the externalities associated with freight transport, such as congestion and emissions. In addition, sustainable solutions can be evaluated and implemented by public managers to improve the quality of life in urban centres, which contributes to economic and social development.

6. Conclusions

This paper estimated models using data on deliveries to commercial establishments. The models were obtained for Brazil and considered the cities' populational characteristics, i.e., for larger cities, medium cities, and small cities.

Data were obtained from field research, and the influential observations were removed from the sample. Therefore, the influential observations were anomalies due to the incorrect answers reported by the respondents. Identifying data collection strategies to reduce inconsistencies might be interesting for future works.

The first contribution of this paper concerns the estimated models using delivery data, which are easy to collect in a scenario with budget restrictions for data collection. The second contribution is related to the methodological procedure. Although evaluating the OLS assumptions is common in econometrics, few studies have conducted this procedure for estimating FTGM. However, the use of different techniques is crucial for accurately estimating the models.

Finally, results in this study provided insights for policymakers since accurate models were obtained for freight transport in Brazilian cities. The results could support freight policies to improve freight operations in urban areas. In accordance with good scientific practice, the estimated models support forecasting and proposing public policies. Public managers can use these models for feasibility studies and for solutions with light interventions and minimal side effects. Coherent measures require following the precepts of the adopted model. Therefore, the assumptions of the estimation methods are essential to support policies or feasibility applications.

The models could be used by practitioners to estimate freight movements in urban areas. The results allow the city impacts and the regions with high freight flow levels to be identified. Thus, urban planners can identify strategies to accommodate cargo flows, which are essential for economic development. In addition, planning based on understanding the problems and based on cargo flows allows suitable alternatives to be identified to reduce and/or organise freight trips, reduce environmental impacts, and contribute to minimising the externalities perceived by society. In this way, a sustainable UFT is supported.

Although not evaluated in this paper, the economic sector also influences cargo flows. Thus, for future works, we suggest estimating FTGM by economic sector. Moreover, many studies have explored the usage of spatial analysis to understand freight trip movements. We recommend incorporating spatial and temporal factors. In addition, we recommend the FTGM estimation by using spatial techniques.

Author Contributions: Conceptualisation, L.K.d.O. and B.V.B.; methodology, L.K.d.O., B.V.B. and F.G.F.d.S.; validation, L.K.d.O., B.V.B. and F.G.F.d.S.; formal analysis, G.G.F.d.A. and C.D.P.; investigation, L.K.d.O., G.G.F.d.A., B.V.B., C.D.P. and F.G.F.d.S.; data curation, L.K.d.O. and G.G.F.d.A.; writing—original draft preparation, L.K.d.O., G.G.F.d.A., B.V.B., C.D.P. and F.G.F.d.S.; writing—review and editing, L.K.d.O., G.G.F.d.A., B.V.B., C.D.P. and F.G.F.d.S.; funding acquisition, L.K.d.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Council for Scientific and Technological Development (CNPq), grant numbers 406873/2018-6 and 303171/2020-0.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

- van Wee, B.; Annema, J.A.; Banister, D. Accessibility: Perspectives, Measures and Applications. In *The Transport System and Transport Policy*; Edward Elgar: Cheltenham, UK, 2013.
- McLeod, S.; Schapper, J.H.M.; Curtis, C.; Graham, G. Conceptualizing Freight Generation for Transport and Land Use Planning: A Review and Synthesis of the Literature. *Transp. Policy* **2019**, *74*, 24–34. [[CrossRef](#)]
- Lindholm, M.; Behrends, S. Challenges in Urban Freight Transport Planning—A Review in the Baltic Sea Region. *J. Transp. Geogr.* **2012**, *22*, 129–136. [[CrossRef](#)]
- Holguín-Veras, J.; Amaya Leal, J.; Sanchez-Diaz, I.; Browne, M.; Wojtowicz, J. State of the Art and Practice of Urban Freight Management Part II: Financial Approaches, Logistics, and Demand Management. *Transp. Res. Part A Policy Pract.* **2020**, *137*, 383–410. [[CrossRef](#)]
- Holguín-Veras, J.; Amaya Leal, J.; Sánchez-Diaz, I.; Browne, M.; Wojtowicz, J. State of the Art and Practice of Urban Freight Management: Part I: Infrastructure, Vehicle-Related, and Traffic Operations. *Transp. Res. Part A Policy Pract.* **2020**, *137*, 360–382. [[CrossRef](#)]
- Cassiano, D.R.; Bertoncini, B.V.; de Oliveira, L.K. A Conceptual Model Based on the Activity System and Transportation System for Sustainable Urban Freight Transport. *Sustainability* **2021**, *13*, 5642. [[CrossRef](#)]
- Comi, A.; Site, P.D.; Filippi, F.; Nuzzolo, A. Urban Freight Transport Demand Modelling: A State of the Art. *Eur. Transp. Trasp.* **2012**, *51*, 1–17.
- Sahu, P.K.; Pani, A. Freight Generation and Geographical Effects: Modelling Freight Needs of Establishments in Developing Economies and Analyzing Their Geographical Disparities. *Transportation* **2020**, *47*, 2873–2902. [[CrossRef](#)]
- Gasparini, A.; Campos, V.B.G.; D’agosto, M.D.A. Modelos Para Estimativa Da Demanda de Viagens de Veículos de Carga Para Supermercados e Shopping-Centers. *Transportes* **2010**, *18*, 58–65. [[CrossRef](#)]
- Oliveira, L.K.; Oliveira, R.L.M.; Ramos, C.M.F.; Ebias, D.G. Modelo de Geração de Viagens de Carga Em Áreas Urbanas: Um Estudo Para Bares, Restaurantes e Supermercados. *Transportes* **2016**, *24*, 53–67. [[CrossRef](#)]
- Oliveira, L.K.; Herédia, R.T.; Bertoncini, B.V.; Oliveira, R.L.M. Freight Trip Generation to Buildings under Construction: A Comparative Analysis with Linear Regression and Generalised Linear Regression. *Transportes* **2020**, *28*, 28–42. [[CrossRef](#)]
- de Oliveira, L.K.; Lopes, G.P.; de Oliveira, R.L.M.; Bracarense, L.; dos, S.F.P.; Pitombo, C.S. An Investigation of Contributing Factors for Warehouse Location and the Relationship between Local Attributes and Explanatory Variables of Warehouse Freight Trip Generation Model. *Transp. Res. Part A Policy Pract.* **2022**, *162*, 206–219. [[CrossRef](#)]
- Silva, K.; da Silva Lima, R.; Alves, R.; Yushimito, W.F.; Holguín-Veras, J. Freight and Service Parking Needs in Historical Centers: A Case Study in São João Del Rei, Brazil. *Transp. Res. Rec.* **2020**, *2674*, 352–366. [[CrossRef](#)]
- Ferreira, B.L.G.; Silva, M.A.V. Truck Trips in Urban Areas and Its Relation to Socioeconomic Variables. *Rev. Gestão Da Produção Operações E Sist.* **2016**, *11*, 197–212. [[CrossRef](#)]
- Barbosa, M.W.; de Sousa, P.R.; de Oliveira, L.K. The Effects of Barriers and Freight Vehicle Restrictions on Logistics Costs: A Comparison before and during the COVID-19 Pandemic in Brazil. *Sustainability* **2022**, *14*, 8650. [[CrossRef](#)]
- Alho, A.R.; Silva, J.D.A.E. Freight-Trip Generation Model: Predicting Urban Freight Weekly Parking Demand from Retail Establishment Characteristics. *Transp. Res. Rec.* **2014**, *2411*, 45–54. [[CrossRef](#)]
- Alho, A.R.; Silva, J.A. Modeling Retail Establishments’ Freight Trip Generation: A Comparison of Methodologies to Predict Total Weekly Deliveries. *Transportation* **2015**. [[CrossRef](#)]
- Mommens, K.; van Lier, T.; Macharis, C. Loading Unit in Freight Transport Modelling. *Procedia Comput. Sci.* **2016**, *83*, 921–927. [[CrossRef](#)]
- Sánchez-Díaz, I. Modeling Urban Freight Generation: A Study of Commercial Establishments’ Freight Needs. *Transp. Res. Part A Policy Pract.* **2017**, *102*, 3–17. [[CrossRef](#)]

20. Holguín-Veras, J.; Sánchez-Díaz, I. Freight Demand Management and the Potential of Receiver-Led Consolidation Programs. *Transp. Res. Part A Policy Pract.* **2016**, *84*, 109–130. [[CrossRef](#)]
21. Venkadavaran, M.; Raj, C.T.; Marisamynathan, S. Development of Freight Travel Demand Model with Characteristics of Vehicle Tour Activities. *Transp. Res. Interdiscip. Perspect.* **2020**, *8*, 100241. [[CrossRef](#)]
22. Middela, M.S.; Ramadurai, G. Incorporating Spatial Interactions in Zero-Inflated Negative Binomial Models for Freight Trip Generation. *Transportation* **2021**, *48*, 2335–2356. [[CrossRef](#)]
23. Sanchez-Diaz, I. Assessing the Magnitude of Freight Traffic Generated by Office Deliveries. *Transp. Res. Part A Policy Pract.* **2020**, *142*, 279–289. [[CrossRef](#)]
24. Wang, X.C.; Zhou, Y. Deliveries to Residential Units: A Rising Form of Freight Transportation in the U.S. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 46–55. [[CrossRef](#)]
25. Sánchez-Díaz, I.; Holguín-Veras, J.; Wang, X. An Exploratory Analysis of Spatial Effects on Freight Trip Attraction. *Transportation* **2016**, *43*, 177–196. [[CrossRef](#)]
26. Ducret, R.; Gonzalez-Feliu, J. Connecting Demand Estimation and Spatial Category Models for Urban Freight: First Attempt and Research Implications. *Transp. Res. Procedia* **2016**, *12*, 142–156. [[CrossRef](#)]
27. Pani, A.; Sahu, P.K.; Chandra, A.; Sarkar, A.K. Assessing the Extent of Modifiable Areal Unit Problem in Modelling Freight (Trip) Generation: Relationship between Zone Design and Model Estimation Results. *J. Transp. Geogr.* **2019**, *80*, 102524. [[CrossRef](#)]
28. Wooldridge, J.M. *Introductory Econometrics: A Modern Approach*; Cengage Learning: Boston, MA, USA, 2012.
29. Gujarati, D.N.; Damodar, G.; Porter, D. *Basic Econometrics*; Irwin/McGraw-Hill: New York, NY, USA, 2008.
30. Washington, S.P.; Karlaftis, M.G.; Mannering, F.; Anastasopoulos, P. *Statistical and Econometric Methods for Transportation Data Analysis*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2020.
31. Motuba, D.; Tolliver, D. Truck Trip Generation in Small-And Medium-Sized Urban Areas. *Transp. Plan. Technol.* **2017**, *40*, 327–339. [[CrossRef](#)]
32. Lawson, C.T.; Holguín-Veras, J.; Sánchez-Díaz, I.; Jaller, M.; Campbell, S.; Powers, E.L. Estimated Generation of Freight Trips Based on Land Use. *Transp. Res. Rec.* **2012**, *2269*, 65–72. [[CrossRef](#)]
33. Dhonde, B.; Patel, C.R. Implementing Circular Economy Concepts for Sustainable Urban Freight Transport: Case of Textile Manufacturing Supply Chain. *Acta Logist.* **2020**, *7*, 131–143. [[CrossRef](#)]
34. Gonzalez-Feliu, J.; Peris-Pla, C. Impacts of Retailing Land Use on Both Retailing Deliveries and Shopping Trips: Modelling Framework and Decision Support System. *IFAC-Pap.* **2018**, *51*, 606–611. [[CrossRef](#)]
35. Campbell, S.; Holguín-Veras, J.; Ramirez-Rios, D.G.; González-Calderón, C.; Kalahasthi, L.; Wojtowicz, J. Freight and Service Parking Needs and the Role of Demand Management. *Eur. Transp. Res. Rev.* **2018**, *10*, 47. [[CrossRef](#)]
36. Gonzalez-Feliu, J.; Sánchez-Díaz, I. The Influence of Aggregation Level and Category Construction on Estimation Quality for Freight Trip Generation Models. *Transp. Res. Part E Logist. Transp. Rev.* **2019**, *121*, 134–148. [[CrossRef](#)]
37. Puente-Mejía, B.; Palacios-Argüello, L.; Suárez-Núñez, C.; Gonzalez-Feliu, J. Freight Trip Generation Modeling and Data Collection Processes in Latin American Cities. Modeling Framework for Quito and Generalization Issues. *Transp. Res. Part A Policy Pract.* **2020**, *132*, 226–241. [[CrossRef](#)]
38. Oliveira, L.K.; Barraza, B.; Bertocini, B.V.; Isler, C.A.; Pires, D.R.; Madalon, E.C.N.; Lima, J.; Vieira, J.G.V.; Meira, L.H.; Bracarense, L.S.F.P.; et al. An Overview of Problems and Solutions for Urban Freight Transport in Brazilian Cities. *Sustainability* **2018**, *10*, 1233. [[CrossRef](#)]
39. Ramsey, J.B. Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *J. R. Stat. Society. Ser. B (Methodol.)* **1969**, *31*, 350–371. [[CrossRef](#)]
40. Zeileis, A.; Hothorn, T. Diagnostic Checking in Regression Relationship. *R News* **2002**, *2*, 7–10.
41. Cook, R.D. Detection of Influential Observation in Linear Regression. *Technometrics* **1977**, *19*, 15–18. [[CrossRef](#)]
42. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: Berlin/Heidelberg, Germany, 2016.
43. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
44. Zhu, L.; Cui, H. A Semi-Parametric Regression Model with Errors in Variables. *Scand. J. Stat.* **2003**, *30*, 429–442. [[CrossRef](#)]
45. Venables, V.M.; Ripley, B.D. *Modern Applied Statistics with S-PLUS*; Springer: Berlin/Heidelberg, Germany, 2002.
46. Yee, T.W. *Vector Generalized Linear and Additive Models: With an Implementation in R*; Springer: Berlin/Heidelberg, Germany, 2015.
47. Cheah, L.; Mepparambath, R.M.; Ricart Surribas, G.M. Freight Trips Generated at Retail Malls in Dense Urban Areas. *Transp. Res. Part A Policy Pract.* **2021**, *145*, 118–131. [[CrossRef](#)]