The influence of aggregation level and category construction on estimation quality for freight trip generation models

Jesus Gonzalez-Feliu, Iván Sánchez-Díaz

The influence of aggregation level and category construction on estimation quality for freight trip generation models

Jesus Gonzalez-Feliu (Corresponding Author)
Mines Saint-Etienne, UMR 5600 Environnement, Ville et Société, CNRS, Institut Henri Fayol, 158 cours Fauriel, 42023, Saint-Etienne cedex 2, France, e-mail: jesus.gonzalez-feliu@emse.fr

Iván Sánchez-Díaz
Service Management and Logistics, Technology Management and Economics, Chalmers University of Technology, Vera Sandbergs Allé 8, 412 96 Gothenburg, Sweden, Phone: +46 0721 585 593, e-mail: ivan.sanchez@chalmers.se

Abstract

This paper analyzes the impacts of aggregation level and category construction on the relevance and quality of freight trip generation (FTG) models. More precisely, constant generations and functional form models are compared, as well as activity and activity-workforce categories. The paper proposes a method to compare constant generation and functional form models on different category classifications based on MAPE estimations. Functional forms are assessed via linear regression and compared using Pearson coefficient. Results show that the aggregation level has not always a positive impact on the model's accuracy and the choice of suitable functional form leads to more accurate models.

Keywords: freight trip generation, aggregation level, comparative analysis, error assessment.

The influence of aggregation level and category construction on estimation quality for freight trip generation models

Abstract

This paper analyzes the impacts of aggregation level and category construction on the relevance and quality of freight trip generation (FTG) models. More precisely, constant generations and functional form models are compared, as well as activity and activity-workforce categories. The paper proposes a method to compare constant generation and functional form models on different category classifications based on MAPE estimations. Functional forms are assessed via linear regression and compared using Pearson coefficient. Results show that the aggregation level has not always a positive impact on the model's accuracy and the choice of suitable functional form leads to more accurate models.

Keywords: freight trip generation, aggregation level, comparative analysis, error assessment.

1. Introduction

Urban freight demand modeling is a key subject in urban freight transport research. The importance of this subject has led to multiple applications, including diverse methods, software tools and practical studies. In general, urban freight demand models are divided into multiple steps to cover demand generation, flow distribution, routing, vehicle class selection, and route assignment, among others (Gonzalez-Feliu, 2018).

The first step when modeling freight demand is to identify and estimate the demand generation. This step consists on assessing the amount of freight and the number of freight trips necessary to satisfy a city's need for goods. However, it is addressed in different ways throughout the literature (Ambrosini et al., 2008; Russo and Comi, 2010; Comi et al., 2012; Gonzalez-Feliu and Routhier, 2012; Browne and Goodchild, 2013; Anand et al., 2015; Sánchez-Diaz et al., 2016b). Holguin-Veras et al. (2011) proposes two explained variables to quantify freight demand: freight trip generation (FTG) and freight generation (FG). This paper focuses on FTG, defined as the number of freight trips that are attracted and produced at an establishment without making any differentiation between attraction and production. FTG models were first used in the United States (Demetsky, 1974; Meyburg and Stopher, 1974; Loebl and Crowley, 1976;) but are now employed worldwide (Holguin-Veras et al., 2011, 2013; Alho and de Abreu e Silva, 2014; Jaller et al., 2015; Aditjandra et al., 2016; Ducret and Gonzalez-Feliu, 2016; Oliveira et al., 2017; Sanchez-Diaz, 2017). At this stage, the choice of the data granularity (Gonzalez-Feliu and Routhier, 2012) and the functional form defining the model (Sanchez-Diaz et al., 2016b) are major issues. Taking into account the model's aim and scope (Gonzalez-Feliu and Routhier, 2012), the needs of data aggregation and of model accuracy can differ from one model to another. Moreover, data quality and availability depend on the context and the application, reposing mainly on establishment surveys (Holguin-Veras and Jaller, 2014), which are directly related to the resources given to collect data but can additionally rely on confidentiality reasons that sometimes make data unavailable. For all those reasons, it seems important to study the relationships between data aggregation and the quality of the resulting models to investigate the interest and impact of the categorization choices of data, as well as the suitability of the main modeling frameworks for FTG.

The literature on FTG (as will be shown below) presents different modeling approaches, most of which are based on activity classifications. Those classifications, mainly related to hierarchical divisions of economic activities (NAICS or NACE codes, respectively, in the USA and Europe), can have different degrees of disaggregation. Moreover, the most recurrent FTG determinant at the single establishment level is the number of employees. Although other variables are used in some models, such as income or establishment's surface, related input data are not always available for modelling purposes (Gonzalez-Feliu, 2018). Therefore, employment is mainly used either as an explanatory variable (Holguin-Veras et al., 2011, 2013) or to build classes of workforce for constant generation (Aubert and Routhier, 1999). Most category classifications present hierarchic structures (i.e. macro-classes, classes and sub-classes). However, most of those categories are chosen with respect to available data or for practical reasons, and few works show the links between the choice of a category and the quality of a model. Therefore, it seems important to identify when to use what type of categorization level and how it affects forecasting accuracy. To the best of our knowledge, existing works focus on producing the model itself but not on the impacts of the way the reality is represented (by the means of a classification, functional form or other modeling hypotheses) and its accuracy within the framework of a comparative analysis (i.e., by comparing different categorization and functional form relationships).

The aim of this paper is to analyze the implications of using different category-construction frameworks and aggregation levels on the classification systems and its impacts on the quality of the resulting FTG model. More specifically, it aims to address the following research question: does the aggregation level of establishment categories have an influence on FTG modeling patterns and accuracy? To address that question, three assessments and the resulting comparative analyses are proposed based on a comprehensive dataset of more than 2900 establishments: the first assessment addresses the comparison of constant FTG estimations for three different aggregation levels of activity-based categories; the second derives from the first, such that each activity-based category classification is completed through workforce classes (to include a first vision of the impact of employment on FTG) but continues to deploy constant estimations; the third aims to assess the impact of searching the most suitable functional form for each category (extending the works of Holguin-Veras et al., 2011 and Sanchez-Diaz et al., 2016b), starting from the same three category classifications as in the first assessment but testing four types of models, namely, a constant generation, a linear model, a lin-log relation and a log-log relation. This paper is organized as follows. Section 2 presents an overview of the literature on freight demand generation modeling, discusses the main issues concerning the use of classifications in FTG and FG modeling and the aims of this paper. Section 3 presents the methodology used for model assessment and comparative analyses, as well as an overview on data used for model construction and validation. Section 4 presents the main results of such analyses, and Section 5 a synthesis of the three analyses, the discussion of the results and the main practical and research implications. Finally, Section 6 presents the conclusions of the proposed work as well as its main limits and further developments..

2. Literature review and research motivation

The literature of freight modeling presents a plethora of demand generation models (Holguin-Veras et al., 2012). Although several works try to classify freight demand models (which include demand generation but further assess the physical freight transport flows), those classifications do not focus on the generation methods but instead on the distribution and route construction ones. Generation models can be classified according to the following elements:

- **Modeling unit.** Although the literature proposes different modeling units for freight transport models (shipment, movement, trip, vehicle, route, commodity quantity), in terms of demand generation, all those units are mainly related to two frameworks (Holguin-Veras et al., 2011): commodity quantities, including the Freight Generation (FG) framework, and the loading/unloading operation, which is related to the shipment and the trip and is known as the Freight Trip Generation (FTG) framework. Models using complex units (as vehicles or routes) relate at one point to one of those two units: for example, Sonntag's (1985) model is based on routes, but to generate them, an FTG model is used to generate the origins and destinations composing those routes. Gentile and Vigo (2013) propose an OD model that can feed vehicle route construction algorithms but which starts by generating the number of deliveries per destination before converting them into commodity quantities; Russo and Comi (2010) and Nuzzolo and Comi (2014) generate first commodity quantities at retailers to then convert them into trips to construct routes using a choice model on a categorization of route types.
- **Spatial scale of the model.** Models can be divided, according to their scale, into three categories: models that estimate FTG at a macroscopic (an urban area or a city for overall FTG estimations), mesoscopic (a neighborhood or a street in an aggregate way) or microscopic level (at a single establishment of a small area for microsimulation purposes). Most models in the literature are macroscopic or mesoscopic (Comi et al., 2012; Gonzalez-Feliu and Routhier, 2012; Holguin-Veras and Jaller, 2014). Indeed, even those that generate FTG rates for each single establishment are based on average values and data are finely aggregated by category types (Holguin-Veras et al., 2011, 2013) or zones (Aubert and Routhier, 1999; Gentile and Vigo, 2013; Alho and de Abreu e Silva, 2014; Bonnafous et al., 2014) to propose macroscopic or mesoscopic estimations. Additionally, several microscopic approaches are found and are mainly related to the needs of traffic micro-simulation (Wisetdjindawaat and Sano, 2003; Lopez et al., 2016) or vehicle routing optimization (Gonzalez-Feliu et al., 2014b); these approaches use probabilistic random-based generations.
- **Functional form.** The functional form is defined as the mathematical formulation that can represent the causal relations between an explained variable (here the freight transport demand) and one or more explanatory variables. According to Holguin-Veras et al. (2011), FTG and FG models have been traditionally developed by associating constant generation rates per establishment or per employee (Demetsky, 1974; Institut de Recherches sur les Transports, 1977; Sonntag, 1985; Eriksson, 1996; Aubert and Routhier, 1999, among others). Another common functional form is to associate to FTG and FG linear functions, either related to only employment (Holguin-Veras et al., 2011, 2013; Gonzalez-Feliu et al., 2014a) or to other variables such as the

establishment's area (Jaller et al., 2015), the sales (Sanchez-Diaz et al., 2013) or a combination of them (Alho and de Abreu e Silva, 2014). Those relations can also be non-linear, including exponential, logarithmic and potential relations (Sanchez-Diaz et al., 2015). An alternative to those functions is that of utility functions, which are generally used for distribution and choice purposes but have also been applied at the generation level (Russo and Comi, 2010; Nuzzolo and Comi, 2014). Finally, and mainly for dynamic simulation requirements, probability-based random generation models are defined by assigning to each generator a random demand rate that follows a determined probability distribution instead of an average value (Gonzalez-Feliu et al., 2014b; Lopez et al., 2016).

Zonal models are generally related to main socio-demographic variables, such as population, activity density or general income of families, and are typically applied to aggregated sets of data (Holguin-Veras et al., 2012; Lawson et al., 2012; Sanchez-Diaz et al., 2013). Establishment FTG models are non-spatial (i.e., they relate trip generation rates to each single establishment) and they aim to establish a link between the establishment characteristics and the trip generation rates (Holguin-Veras et al., 2011; Bonnafous et al., 2014). To do that, FTG models rely on the definition and use of categories that summarize the main characteristics of those establishments. Those categories use multiple categorization criteria, such as the activity type (Holguin-Veras et al., 2011, 2012), mainly related to national classifications that are made for censorial and descriptive statistics purposes, average workforce statistics, area (Gonzalez-Feliu et al., 2012b; Jaller et al., 2015), nature of premises (Aubert and Routhier, 1999), or the requirements of a land-use ordinance (Lawson et al., 2012; Sanchez-Diaz et al., 2013). Consequently, those classifications are not always adapted for logistics decision making and modeling (Liedtke and Schepperle, 2004). Moreover, the choice of the classification will depend on the objective of the model. The selection of categories has traditionally been linked to the purpose of data collection and to the type of classification systems available in each country.

Establishment-based FTG models rely mainly on categories constructed from existing activity-based classifications based on the nature of the activity and organized into a hierarchical structure (i.e., from a small set of macro-categories, different levels of disaggregation can be defined making able to define different classifications; see Holguin-Veras et al., 2011, 2013; Bonnafous et al., 2014). Although those categories differ for each country, in Europe, most codes derive from the Classification of Economic Activities in the European Community, NACE (European Commission, 2008), and US states use the North American Industry Classification System, NAICS (Holguin-Veras et al., 2012). In both cases, several levels of disaggregation can be observed, starting from the main economic activity sectors then defining different hierarchies of subcategories.

Those categories are not always adapted to FTG, mainly when dealing with urban logistics, since they have been conceived for economic classification and censorial aims, and not for specifically defining freight transport demand or behavior patterns (Liedtke and Schepperle, 2004). For those reasons, some models define their own categories. Aubert and Routhier (1999) and Bonnafous et al. (2014) use three criteria to define activity categories: activity type based on French declination of NACE codes; nature of the premise and workforce class, defining categories based on the needs of urban goods transport (i.e., by going on a high level of detail in retailing and stores subcategories but remaining at higher levels of aggregation for the remaining sectors, such as agriculture, industry or wholesaling).

A third group of works, which are focused on specific types of activities, do not use extensive category classifications but use a dataset containing one or two categories: Demetsky (1975) focuses on stores, whereas Maejima's (1978) work addresses construction. More recently, Alho and de Abreu e Silva (2014), de Oliveira et al. (2017) and Sanchez-Diaz (2018) propose models for the Hotels, Restaurants and Catering (Ho.Re.Ca.) sector, and Aditjandra et al. (2016) have collected data on the different establishments at a university campus site. In those cases, the category is defined by a type of activity that remains general (and does not include subcategories).

Zonal models, defined for city areas, use classifications; related to the type of urban zone, or urban form (Sanchez-Diaz et al., 2012; Ducret and Gonzalez-Feliu, 2016). Those types of urban zones define the residential/industrial/retailing predominance in the zones and can also take into account the mix of activities (Ducret and Gonzalez-Feliu, 2016). Other models define only macro-zonal categories (as center or periphery) but relate the FTG rates to socio-demographic variables such as employment, population or household density (Gonzalez-Feliu et al., 2012a; Gonzalez-Feliu, 2018).

Although in other types of models (e.g., activity-based flow estimation models; four-step models; behavioral models; land-use models; transport models) other types of categories can be deployed, those categories are not described in detail in this paper since FTG models are mainly related to four abovementioned types of categories:

- Activity-based categories (establishment models, most of them using employment as explanatory variable);
- Activity-workforce categories (constant-based establishment models);
- Type of urban zone (zonal models);
- Zone location, i.e., center or periphery, with eventually different degrees of centric or peripheral location (zonal models).

The choice of the category is then mainly related to the aim and the expected accuracy of the model. In FTG, not all context have the required data to define the different categories or assess each model, since the criteria that define some categories (mainly zonal) can differ from one country to another (see the differences between types of urban space in the USA and France in Sanchez-Diaz et al., 2013; Ducret et al., 2016), and the spatial and statistical distributions of the different premises are not the same for each country, making it difficult sometimes to go into detail for some categories since the number of premises can be very small. However, the activity sectors remain similar at a macroscopic level: agriculture, craftsmen, services, warehousing, retailing, stores and tertiary mainly (Holguin-Veras et al., 2012; Gonzalez-Feliu, 2018). For those reasons, it seems important to analyze the implications of disaggregation and categorization in FTG modeling accurately and to support the definition of the most suitable model for each use, using for that the most relevant category classification. Thus, constant estimations based on activity classifications or on mixed activity-workforce categorizations and functional form models will be assessed and compared.

3. Method and data description

As shown in the previous section, the quality and accuracy of an FTG model depends on many factors that can be grouped into two main types: the aggregation level and the relevance of the functional form. Therefore, the selection of the category classification and then of the data aggregation level needs to be made in relation to the selection of the functional forms. In this context, it is important to first analyze those relationships, as well as the impacts of functional aggregations on the related functional form. This section presents the methodological framework proposed to conduct those analyses as well as the data used and applied to this purpose.

3.1. Methodological framework

The overall method includes four sequential steps. The first two steps focus mainly on the aggregation levels using basic statistical methods, while the third step explores more expanded functional forms, and the fourth step assesses the tradeoffs between aggregation levels and functional forms using error metric analyses.

To do this, the following sequence is proposed:

1. Analysis of the impacts of data aggregation on the accuracy of activity-based constant estimations;
2. Analysis of the impacts of data aggregation on the accuracy of activity-workforce constant estimations, i.e., completing previous classifications by adding to activity categorization a second subdivision by workforce categories;
3. Assessment of functional forms from only activity categories and analysis of the accuracy of those models;
4. Comparison of the three types of modeling approaches.

The first and second steps seek to analyze the impacts of the aggregation level on the quality of the estimations in the case of using constant rates per establishment. This is used as a first assessment of the data aggregation implications in modeling, and the purpose is to isolate the impacts of the disaggregation level from other phenomena related to functional forms. In the literature, there are several approaches to measure the accuracy of FTG models, mainly the root mean square error (RMSE), the mean average percentage error (MAPE) and the Akaike Information Criterion (AIC), among others (Sanchez-Diaz et al., 2016b). The RMSE is scale-dependent, so it can be used only to compare models for one particular set of data; this metric is particularly sensitive to large errors and heavily penalizes outliers. The MAPE is a relative metric based on the absolute magnitude of the error; this metric is particularly sensitive to low values for the variable of interest and tends to favor models that estimate lower values. The AIC is useful for comparing models – this criterion does not focus on the absolute quality of a model but rather on the quality of one model relative to other models.

The proposed research aims to compare models that are defined under the same assumptions with respect to a set of collected data. From the three indicators, MAPE seems to be the easiest of interpretation for both researchers and practitioners. Moreover, since presented as the average of individual absolute percentage errors, its calculation formula is quite intuitive. For those reasons, the authors selected MAPE to measure the accuracy of the models in this paper.

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

$$MAPE_i = \frac{\sum_{j=1}^{n} \left| \frac{FTG_j^{predicted} - FTG_j^{observed}}{FTG_j^{observed}} \right|}{n} \quad (1)$$

where FTG_j is the FTG predicted/estimated for observation *j* belonging to category *i.*

The third step is to search for the most suitable functional forms. Three functional forms (linear, lin-log and log-log) are tested following Sanchez-Diaz et al.'s (2016b) framework. The two main determinants of FTG showed in literature are the presence or not of an activity (represented by the constant) and the employment (represented by the explanatory variable parameter). For each category, the relevance of those two determinants is different, and not all categories present both terms (Holguin-Veras et al., 2011, 2013; Gonzalez-Feliu, 2018). Thus, for each functional form (lin-lin, lin-log and log-log), a model is obtained by linear regressions as follows. First, the constant and a variable term are assessed statistically, i.e., for each functional form all three following possibilities are assessed via linear regression: (i) a constant, (ii) a function of employment without constant term and (iii) a function with both constant and employment-based coefficients. The parameters for the models are estimated using sandwich estimators that are robust to outliers (Freedman, 2006). These robust estimators depend on asymptotic properties, therefore the resulting estimators would not always fit the t-distribution which is a necessary assumption for the t-test (Imbens and Kolesar, 2016). As this can induce bias in estimations made with small datasets, the authors applied a degree-of-freedom correction on standard errors (Imbens and Kolesar, 2016).

After assessing for each functional form the three variants (constant, function of employment without intercept and function of employment with intercept), a Fisher's test is made on each resulting model to state on the model's relevance, and a Student's test on each parameter to state on its statistical significance. Only relevant models with significant parameters are retained.

Once the relevant models for each functional form are assessed, the most suitable relation is selected based on the highest Pearson correlation coefficient between FTG and number of employees (lin), FTG and ln(employees) (lin-log), and ln(FTG) and ln(employees) (log-log). The Pearson coefficient is the standard measure of a linear correlation between two variables and represents the R^2 metric in a bivariate linear regression. The equation to compute the Pearson correlation is as follows:

$$\rho_{y,x} = \frac{\mathrm{cov}(Y,X)}{\sigma_y \sigma_x} \qquad (2)$$

where σ_y is the standard deviation of the dependent variable, i.e., FTG or ln(FTG), and σ_x is the standard deviation of the independent variable, i.e., employees or ln(employees).

Once functional form models are defined for each category classification, they are assessed and resulting models[1] compared on the basis of MAPE, to allow a comparison among both constant generation approaches and functional form models.

3.2. Data, category description and statistical relevance for average and functional form modeling

To conduct the proposed analyses, this paper uses data issued from the French Urban Goods Surveys of 1995-1998. For a detailed description of the surveys, readers can refer to Ambrosini et al. (2010). This dataset was selected because of the sample size and the variety of sectors covered, which is rare in FTG studies. The surveys used for the data collection comprise three parts: the first one describes the establishments' attributes, the second studies

[1] Resulting models will be then of form $y=a+b.x$; $y=a+log(x)$ or $y=a.x^b$

the pickup and delivery operations, and the third presents the truck driver's patterns, vehicle rounds and paths. Those surveys have been used by different authors to develop FTG models (Aubert and Routhier, 1999; Patier, 2001; Deprez and Bourcier, 2002; Gonzalez-Feliu et al., 2014a,b ; Battaia et al., 2016; Ducret and Gonzalez-Feliu, 2016; Sanchez-Diaz et al., 2016a; Gonzalez-Feliu and Peris-Pla, 2017) and constitute, to the best of our knowledge, the most extensive dataset available[2]. For FTG, the authors use a set of data extracted from the first two parts of the survey. The data were collected from three cities with different characteristics (Bordeaux, Dijon and Marseille) and include 2,970 establishments and 11,588 freight trips, including both attraction and production. Although this dataset is outdated, it provides a unique opportunity to assess the modeling implications of different aggregations of classification systems because of the number of observations within each sector: 2,970 establishments with at least one operation of any type, 2,613 establishments with at least one reception, and 1,500 establishments with at least one expedition. Moreover, the FTG rates per establishment within each category were considered homogeneous independent of the city when designing the survey; thus, the three sets can be used jointly for modeling purposes (Ambrosini et al., 2010). This study focuses on FTG, adding both attraction and production. FTG is the focus for two reasons: first, FTG, without specifying the nature of the operation, can be useful for some planning purposes (identification of truck types and mileages for congestion estimation, definition and dimensioning of parking and delivery bay facilities, etc.), and making no distinction between the type of operations provides a larger dataset to enhance the statistical significance of the results.

The classification system used to group the data within categories is an adaptation of the classification used in the French urban goods surveys (and resulting in both the data collection and descriptive analyses). This classification results from an aggregation of NAF codes (the French declination of NACE codes). The main elements determining the categories are:

- The activity of the firm, eventually coupled with the main function of the establishment: 8, 27 or 43 activities distinguished;
- The class of workforce which allows for introducing the employment as a determinant for trip generation and divides respectively the initial 8, 27 and 43 categories into 17, 72 and 105 categories[3].

The first level of disaggregation distinguishes 8 categories of establishments (ST8). Then, each ST8 category is divided into one or more categories, leading to a second level of 27 categories. The third aggregation comprises 43 categories (ST43), and its notation involves repeating the ST27 codes and adding sub-indexes only when the category is subdivided. In all three classifications, the only discriminating element is the economic activity of the establishment. From those categories, it is possible to define activity-workforce categories, to add information on workforce: instead of defining employment by a numerical variable, classes of employment are assigned to each category (Ambrosini et al., 2010).

[2] The surveys have been conducted on the basis of the Freturb model (Aubert and Routhier, 1999 ; Bonnafous et al., 2014), but they have been also used to propose alternative FTG (Patier, 1999, 2001; Deprez and Bourcier, 2002; CERTU, 2013; Gonzalez-Feliu et al., 2014a; Ducret and Gonzalez-Feliu, 2016) and FG models (Henriot and Routhier, 2010; Abdelhai, 2013; Gonzalez-Feliu, 2018). Thus, this work uses the establishment-based data of the survey and the category classifications used in the data collection and analysis process (which do not exactly coincide with the Freturb categorization) to conduct the aggregation analysis, independently of the Freturb modeling choice and hypotheses. Indeed, in this work, raw data in terms of freight trips and frequencies issued directly from the surveys source databases have been used instead of the Freturb-calibrated data.

[3] Categories have been updated with respect to the original work (Ambrosini et al., 1996, 1999a,b) to ensure that each category contains a minimum number of individuals for statistical relevance purposes.

One of the first actions to take when modeling is to verify that the data are relevant for the given purposes. More precisely, there needs to be a sufficient amount of data to be able to obtain a model that represents the FTG patterns of the establishments belonging to a given category. The main criteria to determine whether a sample of data included in one category is representative of other establishments in this category are the random process to select the data, the response rate, and the sample size. The process to select the data and ensure a high response rate is part of the data collection design (Lyberg and Kasprzyk, 1991), while the sample size per category depends on the aggregation level selected by the modeler. It is noteworthy that data collection designs should be in line with the level of aggregation that will then be selected for modeling. As the aggregation level is the central theme of this paper, the number of observations in each category is evaluated to check whether they meet the necessary threshold to be considered relevant for defining FTG patterns and those for each category.

This threshold is important when addressing the mean estimation and regression-based analysis, as the normal distribution hypothesis is often made when conducting statistical tests (such as Student's and Fisher's tests) and ordinary-least-squares regressions. This hypothesis can be assumed for a number of observations equal to or higher than 30, thus obeying the Law of Big Numbers (Wonnacott and Wonnacott, 2001). However, for some categories, it is necessary to conduct statistics with a lower number of observations. In works with small sample sizes (Brown and Forsythe, 1974; Lovric, 2011), the means can be defined for sets having at least 5 observations (Lewis, 1993), although classical dispersion analyses are robust for sets having at least 9 individuals, and range-based dispersion estimators can be applied to sets having between 6 and 8 individuals (Dean and Dixon, 1951).

That choice leads to the definition of the following activity-workforce category classifications:

- The ST8 categorization is converted into an ST17 activity-workforce category classification.
- The ST27 categorization is converted into an ST67 activity-workforce category classification.
- The ST43 categorization is converted into an ST105 activity-workforce category classification.

Those categories are defined from the initial categories defined in the French surveys (Ambrosini et al., 2010), grouping then those that did not meet the defined requirements (i.e., having less than 6 observations) into the closest category in order to have the minimum number of observations for modeling purposes.

A synthesis of the three activity-based levels of disaggregation is presented in Table 1.

Table 1: Synthesis of the three first levels of disaggregation (ST8-ST27-ST43)

ST8 code	Description	ST27 code	Description	ST43 code	Description
1	Agriculture	1	Agriculture	1	Agriculture
2	Craftsmen/ services	2	Craftsmen	2-2	Repair activities
				2-3	Manufacturing or installation
		26	Services	26Ha	Tertiary services: high flows
				26Mi	Tertiary services: mixed flows
				26Mo	Tertiary services: average flows
3	Industry	3	Chemical industry	3	Chemical industry
		34	Construction industry	34-2	Repair industry
				34-3	Construction – Manufacturing or installation
		4	Production and intermediate	4-2	Production and intermediate – basic bulk
				4-6	Production and intermediate – small objects
				4-7	Production and intermediate – bulk
		5	Consumption goods	5-2	Fragile foodstuffs
				5-4	Non-fragile foodstuffs
				5-5	Non-fragile consumer goods, equipment of the house and the individual
4	Wholesalers	7	Intermediary products	7-2	Fragile intermediate products
				7-3	Other intermediate products
		8	Non-food	8-2	Non-food fragile consumer goods
				8-3	Non-food non-fragile consumer goods
		9	Food products	9-2	Fragile food consumer goods
				9-3	Other food consumer goods
5	Department stores	10	Department stores	10	Hypermarkets and big department stores
				11	Supermarkets
				12	Specialized department stores
6	Retailers	13	Minimarkets	13	Minimarkets
		14	Clothing, shoes, leather	14	Retail trades, clothing, shoes, leather
		15	Butcher's shops	15	Butcher's shops
		16	Small groceries	16	Grocer's shops
		17	Bakery retailers	17	Bakeries – Cake shops
		18	Ho.,Re.,Ca.	18	Ho.,Re.,Ca.: Hotels, Restaurants, Cafés
		19	Pharmacies	19	Pharmacies
		20	Hardware stores	20	Hardware stores
		21	Furnishing shops	21	Furnishing shops
		22	Bookshops	22	Bookshops
		23	Other retail shops	23	Other retail shops
		29	Street trading	29	Street trading (outdoor trading centers and marketplaces)
7	Tertiary/ offices	6	Pure transport	6	Transport (except storage)
		24	Other tertiary	24	Other tertiary activities with low flows
		25	Offices	25	Pure tertiary sector (offices)
				27-2	Not tertiary offices (agriculture, wholesales)
				27-3	Not tertiary offices (retail trade, industry, transport, administration)
8	Warehousing	28	Warehousing	28-2	Warehouses (bulk)
				28-3	Warehouses (of which transport)

Table 2 shows the percentage of categories meeting the threshold of observations to be considered for inferential statistical analyses for each aggregation level. Table 2 also shows the number of categories meeting two thresholds corresponding to minima of 30 and 6 observations for each aggregation level. The activity-workforce categories have been adapted from the original ones to ensure that the threshold of 6 observations is verified for all categories.

Table 2: Aggregation level and data quantity requirements for statistical modeling purposes

Categories of establishments		Number of categories ≥ 30 observations	Number of categories ≥ 6 observations
ST8	8 categories	8/8 (100%)	8/8 (100%)
ST27	27 categories	26/27 (96%)	27/27 (100%)
ST43	43 categories	34/43 (79%)	43/43 (100%)
ST17	17 categories obtained from the initial 8 classes	60/17 (54%)	105/105 (100%)
ST67	72 categories obtained from the initial 43 classes	47/67 (70%)	105/105 (100%)
ST111	105 categories obtained from the initial 43 classes	60/105 (57%)	105/105 (100%)

In the current context, it is difficult to split the dataset in two representative subsets to provide a construction-prediction test for all models (several categories, mainly for activity-workforce classifications, present less than 30 individuals, which makes difficult to split the complete dataset into two representative subsets). For that reason, the data used to obtain the constant FTG rates and define the most suitable functional forms will be the same as the data used for comparing the resulting models.

4. Results and discussion

The results of the proposed analyses are presented below. The results are organized in three groups: first, FTG constant rates for the three activity-based aggregation levels; second, an analogous analysis for constant FTG estimations on activity-workforce categories; and third, the assessment of functional forms, the consequent comparison and a reproducibility assessment. As explained in Section 2, MAPE is selected as the error metric.

4.1. Impacts of the Aggregation Level on the Quality of the Estimations - Constant Rates per Establishment

The first set of results presents the comparison of the three activity-based category levels on the basis of accuracy of resulting models.

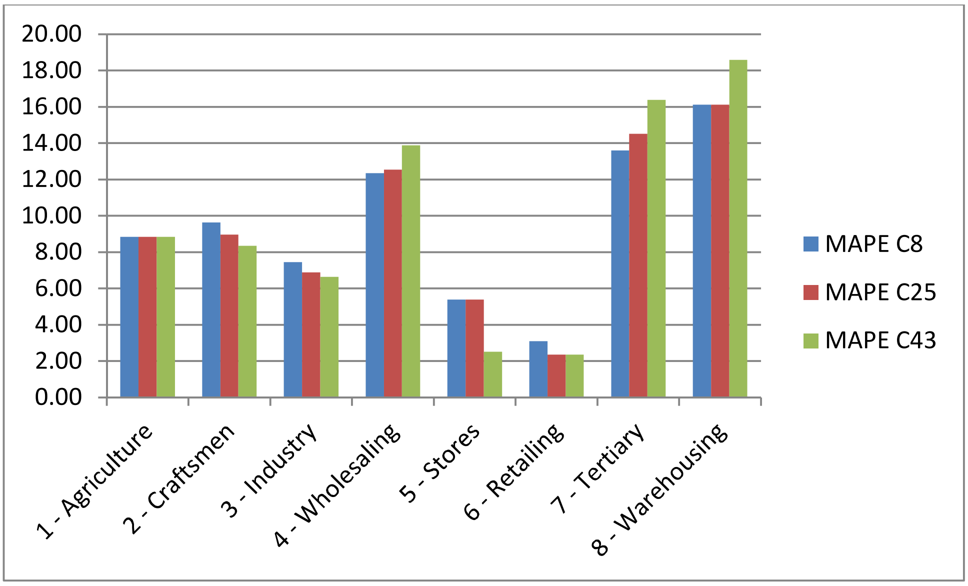

Figure 1: MAPE estimation for constant generation on activity categorizations at the sector level

Figure 1 shows a graphical representation of the MAPE for the different aggregation levels. Although the disaggregation of data leads to more accurate estimations, this trend is not identical for all the categories. For four sectors, i.e., craftsmen, industry, department stores, and retailers (categories 2, 3, 5 and 6 respectively), disaggregating data results in a gain of accuracy, which is small except for sector 5-stores, where the estimation with ST45 category classification results in a MAPE approximately half of that for the other two category classifications. For sector 1-agriculture, the models are the same (i.e., the data available did not allow subdividing it), consequently the MAPE values were also identical. For sectors 4-wholesaling, 7-tertiary and 9-warehousing, the disaggregation results in a decrease in accuracy, including a small decrease between ST8 and ST27 and a higher decrease between those two category classifications and ST44. This difference is due to sectors 4, 7 and 8 being more heterogeneous than the others. Although industry can be heterogeneous, the proposed categorization remains very aggregated by three macro-types of industries for all three aggregation levels. Thus, the assumptions made for subdividing those categories for ST27 and ST43 do not capture the heterogeneity of the sectors in a more accurate way than for ST8.

In conclusion, applying constant rates to more detailed categories does not always tend to result in more accurate FTG estimates. Moreover, having more detailed categories also requires a larger amount of data, and the high disaggregation can lead to representativeness risks. A potential solution to find a balance between aggregation and performance is through more elaborated functional forms that capture the heterogeneity observed within the aggregated categories.

4.2. Use of activity-workforce categories

To analyze the influence of using activity-workforce categories in FTG estimation, this paper proposes a second analysis based on constant FTG estimations on categories which, starting from the previous activity-based classifications, add information on workforce through the addition of a workforce class subdivision. Figure 2 shows the results of that assessment. As shown, when combining employment and activity types in the definition of categories, the impact of data disaggregation is small, except in sector 5-stores, for which the number of observations remains small, and sector 8-warehousing, which is already seen as a heterogeneous one.

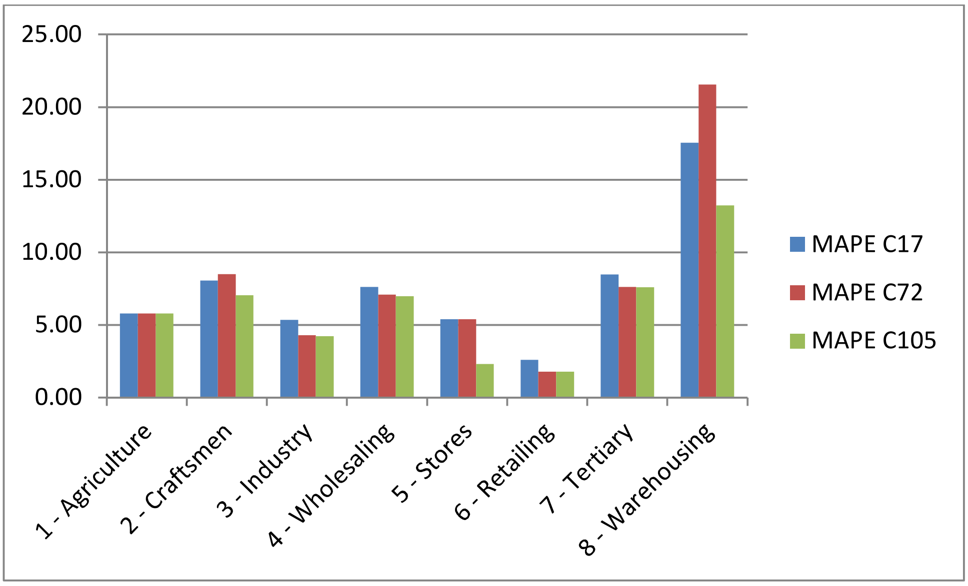

Figure 2: MAPE estimations for constant estimations in activity-workforce categorizations at the sector level

The impact of combining employment and activity type on sector 1-agriculture is almost negligible. For sector 3-industry, 4-wholesaling, 6-retailing and 7-tertiary, a decrease of MAPE is observed when data is more disaggregated but remains small. For sector 2-craftsmen and services, a similar trend with respect to sector 8-warehousing is observed. However, the differences of MAPE among models are smaller for sector 2-craftsmen and services. In both cases, as well as for sector 7-tertiary, the difficulty of assessing models can be related to the heterogeneity of the subcategories (in ST27 and ST43) and the consequent difficulty of defining coherent classes for those sectors in disaggregated categorizations.

4.3. Assessing functional forms

To study the potential of using more elaborated functional forms to capture the heterogeneity within aggregated categories, the authors assess the different functional relationships between the dependent (FTG) and the independent variables (i.e., constant per establishment, number of employees). The alternatives for these functional relationships are linear, lin-log and log-log evaluated for ST8, ST27 and ST43. As explained in Section 2, the decision for the functional relationship is based on a Pearson correlation analysis, and the inclusion of variables that determine the final functional form is based on the statistical significance of the variables assessed using regression analysis with sandwich estimators. Only variables significant at the 5% level are kept in the models.

Table 3 shows a short description, the number of observations and the functional form selected for each category. More details, such as the model parameters and error assessments, are presented in the appendix.

Table 3: Selected Freight Trip Generation Models

ST8 code	ST27 code	ST43 code	Description	Obs.	Functional relationship
1	1	1	Agriculture	41	log-log
2	All	All	Craftsmen and services	375	log-log
	2	All	Craftsmen	188	lin-log
		2-2	Repair activities	52	log-log
		2-3	Manufacturing/ installation	135	lin
	26	All	Tertiary services: craftsmen	187	log-log
		26Ha	Tertiary services: high flows	32	log-log
		26Mi	Tertiary services: mixed flows	140	log-log
		26Mo	Tertiary services: average flows	15	log-log
3	All	All	Industry	623	log-log
	3	3	Chemical industry	46	log-log
	34	All	Construction industry	125	log-log
		34-2	Construction repairs	28	lin-log
		34-3	Construction manufacturing/installation	97	lin
	4	All	Primary and intermediate products	256	lin
		4-2	Basic bulk	144	lin
		4-6	Small objects	92	log-log
		4-7	Bulk	112	log-log
	5	All	Food and non-fragile consumer goods	196	log-log
		5-2	Consumer goods (fragile foodstuffs)	8	lin
		5-4	Non-consumer goods (fragile foods)	105	lin
		5-5	Non-fragile consumer goods (house)	82	lin
4	All	All	Wholesale	414	log-log
	7	All	Intermediary products	200	log-log
		7-2	Fragile intermediate products	80	log-log
		7-3	Other intermediate products	120	log-log
	8	All	Nonfood consumer goods	118	log-log
		8-2	Non-food fragile consumer goods	92	log-log
		8-3	Non-food non-fragile consumer goods	33	lin-log
	9	All	Food	96	log-log
		9-2	Fragile food consumer goods	32	log-log
		9-3	Other food consumer goods	64	log-log

ST8 code	ST27 code	ST43 code	Description	Obs.	Functional relationship
5	10	All	Department stores	47	log-log
		10	Hypermarkets and department stores	13	lin
		11	Supermarkets	20	lin
		12	Specialized department stores	14	log-log
6	All	All	Retailers	1080	log-log
	13	13	Minimarkets	13	lin
	14	14	Clothing, shoes, leather	109	lin
	15	15	Butcher's shops	70	lin
	16	16	Small groceries	97	lin
	17	17	Bakery retailers	112	lin
	18	18	Hotels, restaurants, cafés	145	lin
	19	19	Pharmacies	61	lin
	20	20	Hardware stores	43	lin
	21	21	Furnishing shops	49	lin-log
	22	22	Bookshops	90	log-log
	23	23	Other retail shops	261	lin-log
	29	29	Street trading (marketplaces)	30	lin
7	All	All	Tertiary/ offices	322	log-log
	6	6	Transport except storage	34	log-log
	24	24	Other tertiary activities with low flows	32	log-log
	25	All	Offices	256	log-log
		25	Pure tertiary sector (offices)	178	log-log
		27-2	Non-tertiary offices (agriculture. wholesale)	48	log-log
		27-3	Non-tertiary offices (other activities)	30	log-log
8	28	All	Warehouses/ transport	68	log-log
		28-2	Warehouses (bulk)	21	log-log
		28-3	Warehouses (with transport)	68	lin-log

As shown in the table, a substantial majority of the models display a functional form of type log-log. For ST8, all the categories show a log-log model as the best relationship, three of the eight models have only an employment term and no constant (1–Agriculture, 2–Craftsmen and 5–Stores), and the other five have both a constant and an employment term, i.e., having the form “y=a.x^b.” For retailing, seven of the twelve categories result in a linear relationship, six of type “y=a+ b.x” and one of type “y=c.x”; two are of type lin-log, i.e., y=log(bx); one is of type log-log; and for two categories, no elaborated functional form was found thus, a constant rate per establishment was proposed. For the remaining 15 categories of ST27, 13 are of the type log-log, one is lin-log, and one is linear. Finally, for the remaining 31 categories of ST43, 20 are log-log, eight are linear, two are lin-log, and one is a constant rate.

4.4. Assessment of Aggregation Levels and Functional Form

The results of the previous section show that there is no functional form that works for all the categories and that capturing the dispersion observed within categories requires a large effort in modeling terms. The remaining question is whether these more elaborated functional forms can offset a lack of disaggregation between categories.

Figure 3 summarizes the results from the MAPE analysis in a graphical way.

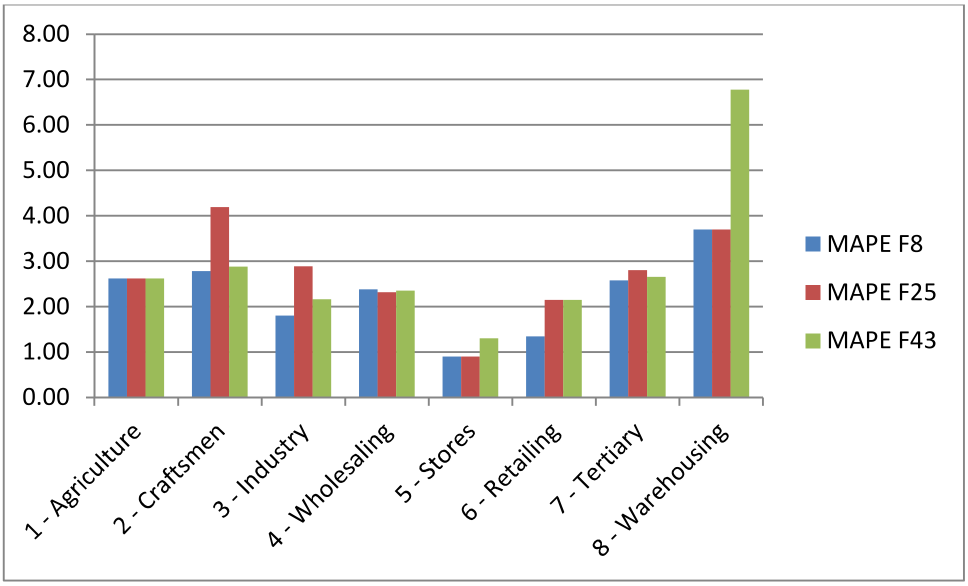

Figure 3: MAPE estimations for functional form assessments in activity categorizations at the sector level

The results show that models with elaborated functional forms present lower MAPE values than those applying the constant rate for the most disaggregated categories, but the data disaggregation does not have a significant impact on model quality. Indeed, for three sectors (agriculture, wholesaling, and services), MAPE remains very close for the three levels of disaggregation. In four other sectors, two of the aggregation levels present similar results, and the third is in general less accurate. Finally, retailing shows ST27 and ST44 having similar accuracy results but ST8 presents lower MAPE. In all cases, ST8 has MAPE of the same order of magnitude than more disaggregated results, and sometimes, models for that aggregation level present a lower MAPE than the others. In that context, we can state that capturing the most suitable functional form can overcome the issues of heterogeneity representation. In other words, sets with representative and significant data, even if aggregated, can lead to accurate models if the functional forms are examined in detail and chosen in a systematic way.

In addition to the previous analysis, a bootstrap procedure was performed to assess the impact of using different sample sizes on the quality of ST8 nonlinear models. The bootstrap procedure uses Monte Carlo simulations to draw samples of 5, 30 and 100 observations from each category in the ST8 level, the best model is then calibrated with the new sample and applied to all the observations in each category to compute the MAPE. For each number of observations, ten different runs of the simulation are performed. The results are summarized in Table 7.

Table 7 MAPE for ST8 models using bootstrapping

ST8	Description	Average MAPE-5 obs	Average MAPE-30 obs	Average MAPE-100 obs
1	Agriculture	2.78	2.60	2.45
2	Craftsmen and services	34.33	5.97	2.59
3	Industry	2.74	1.97	2.19
4	Wholesale	>1000	2.46	3.04
5	Department stores	1.20	1.04	0.92
6	Retailers	1.84	1.44	1.30
7	Tertiary/ offices	5.34	3.33	2.64
8	Warehouses/ transport	6.64	3.50	0.87

As shown, the results from the bootstrap using 30 and 100 observations are fairly robust. Although models calibrated with 100 observations display lower MAPE than models using 30 observations, this difference is not very large for most categories; only categories 2 and 8 display important differences. For models calibrated using 5 observations, the results are mixed. For categories 1, 3, 5 and 6, the MAPEs have a similar magnitude as the ones for models calibrated using 30 observations. For categories 7 and 8, the MAPEs are almost double those for models using 30 or 100 observations. However, for categories 2 and 4, the MAPEs are extremely large, thus making this sampling size inappropriate. The most interesting results from this bootstrap analysis is that except for categories 2 and 4 estimated using 5 observations, all the ST8 nonlinear models produce lower MAPE than the ST105 constants, independently from the sample size.

In essence, finding the most suitable functional form results in more accurate models than making very disaggregated constant estimations, and this is true even when the sample for estimations is relatively small. Moreover, having gains in accuracy by developing more elaborated functional forms instead of using more disaggregated categories has major benefits in terms of data collection costs.

5. Discussion and practical implications

The results show that for a given classification and FTG estimation methodology, the impact of aggregation is not systematic and remains small in most cases. However, it is still necessary to compare the three assessments proposed above in order to address the potential of each, as well as the relevance of functional form modeling with respect to category choice. Figure 4 shows the MAPE values for all models and categorization choice. As shown, for three sectors (1-agriculture, 4-wholesaling and 7-tertiary), there is a positive impact on accuracy when introducing information about employment: activity-workforce models are clearly more accurate than constant activity-based estimations, but functional form models remain the most accurate ones. The impact of employment is less clear for stores where only functional form models result more accurate than constant ones. and for retailing where MAPE remain very close for all models. For the remaining three activity sectors (craftsmen, industry and warehousing), functional form models are more accurate than other models, but employment-based categories do not reflect clearly the impact of employment on FTG since those models remain close to constant-based ones in terms of accuracy.

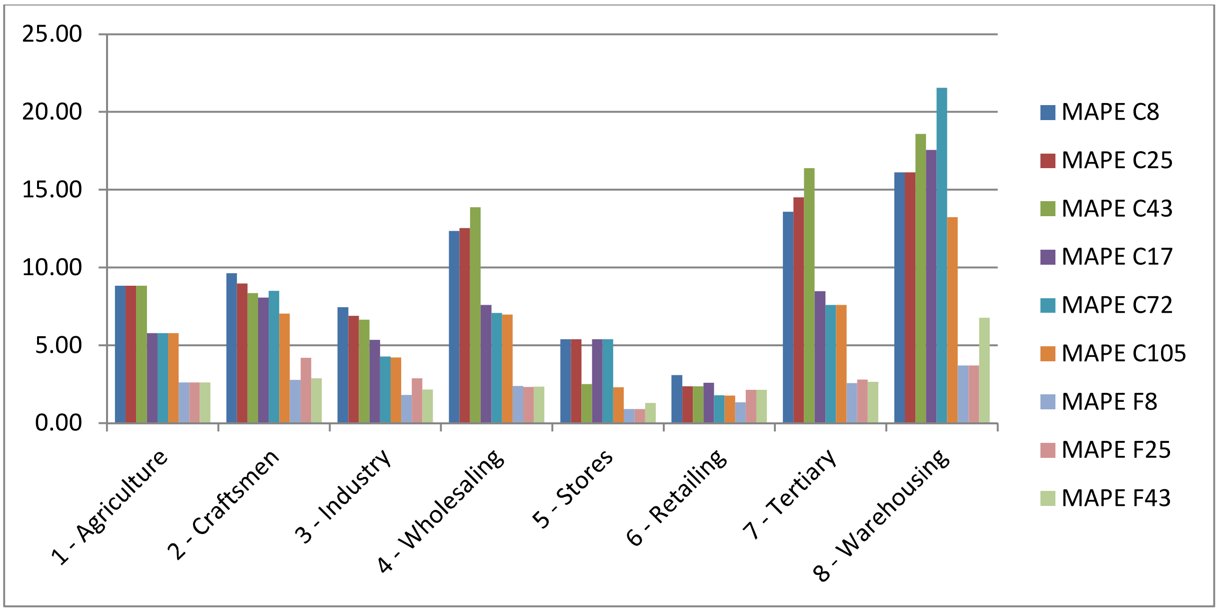

Figure 4: MAPE for all assessed models and estimators

In conclusion, functional form models remain more accurate than pure categorical ones even when introducing employment into categories, except for retailing where the accuracy is similar. The proposed analysis shows the importance of selecting the most suitable functional forms since the data aggregation plays a secondary role once those forms are identified. However, the use of the different models will also depend on the aims of the estimation and its scope.

This research has a number of practical implications. The first is of the operational and financial order. As shown in Figure 3, when choosing suitable functional forms, the level of disaggregation plays a minor role in the estimation quality. In the context of survey-based declarative data, it is essential to calibrate FTG models; the cost of the data collection process is strongly dependent on both the disaggregation level, the need of collecting a minimum number of observations per category to ensure representativeness, and the number of questions asked to each individual. Moreover, the data collection cost is not linear, and higher levels of detail in categorization often imply costs that make freight surveys prohibitive (Holguin-Veras and Jaller, 2014). Since models with more elaborated functional forms display a degree of accuracy equivalent or superior to estimating simple rates at a more disaggregated category level, more aggregated models will imply a lower need for large sample sizes, thus lowering costs. Moreover, the results show that functional forms that relate FTG to employment can be found, and their quality remains similar for all aggregation levels, so the choice of disaggregating data is related to the use of the model and the resources used to develop it.

A second implication is related to the use of the proposed results by practitioners, mainly related to the data inputs available to them for application. As shown above, functional forms depend on different explanatory variables. In this paper, only employment was used, since for French surveys, data on commercial areas or income are not publicly available. Data on employment were collected in the surveys, and thus, it was possible to identify the proposed functional forms. However, employment is often publicly available but only as a range (in France, in the USA, and in Sweden, among other countries), which in several cases refers to the entire company and not to the single location considered for FTG analysis. The results in this paper reveal that depending on availability, there are two ways to overcome this challenge: (i) converting range data into an average employment rate for each establishment; or (ii) using paid datasets that disclose estimated or real numbers of employees. For (i), the accuracy of the FTG models will be reduced, not because of the models' accuracy but of the data input bias related to the need of estimating average employment rates, while for (ii), the cost will be increase because of acquiring the application data but the estimations would be more accurate.

Finally, it is important to address operational and data collection implications. As shown in the analysis, for some complex categories for which it is difficult to collect data, the influence of data aggregation on the model quality is low, thus there is no need for much disaggregated data to produce FTG estimates. This is interesting in categories 2-craftsmen), 4- wholesalers) and 7-tertiary), where it is very difficult to have a representative sample of all subcategories, but it is possible to have a good sample for the overall category. For some sectors, such as, sector 5-department stores), the main problem is finding enough establishments in a city of a single sub-category. In those cases it is valuable to know that an aggregated categorization can lead to models with similar accuracy than a disaggregated one. Therefore, a less detailed data collection can however lead to a good assessment model if the most suitable functional form is identified and assessed. Moreover, since the availability of databases and the budget to conduct extensive surveys vary across cities (Holguin-Veras and Jaller, 2014); the results presented in this paper can help practitioners to design samples requiring less resources for data collection in certain sectors while ensuring robust and statistically significant samples. For sector 6-retailers, there is often a need for more mode detailed information, such as about the type of shop and the type of goods handled,. In those cases, the sampling strategy cannot depend only on the potential for accurate estimates. For that reason, it will be important to define the objectives of a survey prior to set data aggregation and granularity levels.

6. Conclusion

This paper is a first step in understanding how the classification of data affects the quality of the estimators for urban freight trip generation (FTG). To that end, the paper proposed two sets of analysis. The first combined a dispersion-based analysis, as well as the assessment of constant-based generation models on four levels of disaggregation to address the implications of category disaggregation on both dispersion and constant estimation accuracy. The second assessed the relevance of defining suitable functional forms and combined the assessment of three functional forms (i.e., linear, lin-log and log-log) and the subsequent choice of the best functional form for each category and classification level, followed by a comparison of the best functional forms for each category aggregation. The results show that dispersion is not negligible, so invariances and, thus, a systematic constant generation cannot be proven; however, for some aggregated uses (i.e., grouping individual FTG rates into zones), the latter can be suitable. In that context, the disaggregation level can play a role in the quality of constant-based FTG rates. Although not in a proportional way, more disaggregated estimations generally result in more suitable results. However, the analysis in this paper shows that the aggregation level plays a less important role than the identification of the suitable functional form. Moreover, the functional form models result in lower MAPE than the best constant-generation models, and that independently of the aggregation level. In that context, it is possible to conclude that finding suitable functional forms can reduce the needs for more disaggregated data, thus resulting in less costly data collection procedures.

The limitations of the study are mainly related to the incorporation of additional variables that capture logistics decisions that differ by type of activities and to the transferability of the results. Results have also their limits due to data availability that did not allow the application of a prediction dataset different from the construction dataset. For the former, the introduction of nonlinear functional relationships is an important step in adapting the model to the effects of shipment size and mode changes on FTG, but further research is necessary in identifying relevant and practical variables that can explain FTG generation patterns, mainly in urban areas (where area, frontline and other variables are starting to be used but its use remains less generalized than employment). In terms of transferability of the findings, a direction for further research is to compare data from different countries, e.g., France, Sweden and the USA, and to conduct in-depth statistical analyses to analyze data dispersion within categories as well as the role of better functional forms.

Acknowledgements

Part of this work has been conducted in the context of the French National Agency project ANNONA, and the first ideas and developments were started during the first author's former position as Research Engineer in data production and analysis at the French National Center of Research and Laboratory of Transport Economics. The authors also thank Christian Ambrosini, former senior lecturer at Lumière University Lyon 2 and Laboratory of Transport Economics, who helped the authors with his valuable comments at an early stage of the research. Moreover, the data used for the analysis are extracted from the French Urban Goods Transport Surveys steered by the French Ministry in charge of Transport (Direction of Research and Innovation, DRI) and co-financed with the French Agency of the Energy and the local authorities of Bordeaux, Marseille and Dijon. The Laboratory of Transport Economics was the scientific lead and methodology developer. Various exchanges with the data collection team during the first author's stay at the Laboratory of Transport Economics were necessary to understand the hypotheses and methods behind the dataset. The authors thank Michel Julien from DRI, who was in charge of maintaining and disseminating those surveys and provided pertinent advice and comments to the authors.

The authors would also like to acknowledge the support from the Volvo Research and Educational Foundations (VREF), which fund the Gothenburg Urban Freight Platform (UFP) through its Future Urban Transport research program. Last but not least, the authors would like to acknowledge the valuable comments and suggestions from the reviewers that helped to substantially improve the paper.

References

[1] Abdelhai, L. (2013). Analyse des impacts de la mutualisation sur les organisations logistiques existantes. Mémoire de Master, Université Lyon 2.

[2] Aditjandra, P. T., Galatioto, F., Bell, M. C., & Zunder, T. H. (2016). Evaluating the impacts of urban freight traffic: application of micro-simulation at a large establishment. *European Journal of Transport and Infrastructure Research*, 16(1), 4-22.

[3] Alho, A., & de Abreu e Silva, J. (2014). Freight-Trip Generation Model: Predicting Urban Freight Weekly Parking Demand from Retail Establishment Characteristics. *Transportation Research Record: Journal of the Transportation Research Board*, (2411), 45-54.

[4] Ambrosini, C., Bossin, P., Durand, S., Gelas, P., Le Nir, M., Patier, D., Routhier, J.L., (1996), Transport de marchandises en ville : Enquête quantitative réalisée à Bordeaux. Rapport final, MELT-DRAST-LET, Paris, France.

[5] Ambrosini, C., Meimbresse, B., Routhier, J.L., & Sonntag, H. (2008). Urban freight policy-oriented modelling in Europe. In Taniguchi, E., Thompson, R.G. (eds.) *Innovations in City Logistics* (pp. 197-212). Nova Science Publishing, New York,.

[6] Ambrosini, C., Patier D., Routhier, J.L, (1999a), Transport de marchandises en ville - Résultats de l'enquête quantitative réalisée à Dijon. Rapport final, MELT-DRAST-LET, Paris, France.

[7] Ambrosini, C., Patier D., Routhier, J.L, (1999b), Transport de marchandises en ville - Résultats de l'enquête quantitative réalisée à Dijon. Rapport final, MELT-DRAST-LET, Paris, France.

[8] Ambrosini, C., Patier, D., Routhier, J.L. (2010). Urban Freight Establishment and Tour Based Surveys for Policy Oriented Modelling. *Procedia-Social and Behavioral Sciences*, Vol. 2, No. 3, pp. 6013-6026.

[9] Anand, N., Van Duin, R., Quak, H., & Tavasszy, L. (2015). Relevance of city logistics modelling efforts: a review. *Transport Reviews*, *35*(6), 701-719.

[10] Aubert, P.L., & Routhier, J. (1999). FRETURB, un modèle de simulation des transports de marchandises en ville. *8th WCTR Antwerp proceedings*, Vol. 1, pp. 531-544.

[11] Battaïa, G., Gardrat, M., Toilier, F., Le Van, E., et al. (2016). Simulating logistic innovation in a growing urban environment. *Transportation Research Procedia*, *12*, 489-499.

[12] Bonnafous, A., Gonzalez-Feliu, J., Routhier, J.L. (2014). An alternative UGM paradigm to OD matrices: the FRETURB model. In *WCTR 2013 Selected Papers, Rio de Janerio, Brazil*.

[13] Brown, M. B., & Forsythe, A. B. (1974). The small sample behavior of some statistics which test the equality of several means. *Technometrics*, *16*(1), 129-132.

[14] Browne, M., & Goodchild, A. V. (2013). Modeling Approaches to Address Urban Freight's Challenges: A Comparison of the United States and Europe. In *Transportation Research Board Conference Proceedings* 50, 5-7.

[15] CERTU (2013). *Aménagement des aires de livraison : guide pour leur quantification, leur localisation et leur dimensionnement*, CERTU, Lyon.

[16] Comi, A., Delle Site, P., Filippi, F., & Nuzzolo, A. (2012). Urban freight transport demand modelling: A state of the art. *European Transport/Trasporti Europei*, 51 (7), pp. 1-17.

[17] Dean, R. B., & Dixon, W. J. (1951). Simplified statistics for small numbers of observations. *Analytical Chemistry*, *23*(4), 636-638.

[18] Demetsky, M.J. (1974). Measurement of Urban Commodity Movements. *Transportation Research Record*, 496, 57-67.

[19] Deprez, S., & Bourcier, A. (2002). Transport de marchandises dans l'agglomération Rouennaise: Pour une approche géographique. *TEC*, 173, 8-13.

[20] de Oliveira, L. K., de Albuquerque Nóbrega, R. A., & Ebias, D. G. (2017). Analysis of freight trip generation model for food and beverage in Belo Horizonte (Brazil). *REGION*, *4*(1), 17-30.

[21] Ducret, R., & Gonzalez-Feliu, J. (2016). Connecting demand estimation and spatial category models for urban freight: First attempt and research implications. *Transportation Research Procedia*, *12*, 142-156.

[22] Eriksson, J.R. (1997). Urban freight Transport forecasting – an empirical approach. *Urban Transportation and the Environment II*. Computational Mechanism Publications, Ashurst, 359-369.

[23] European Commission (2008). *NACE Rev. 2 – Statistical classification of economic activities in the European Community*. Office for Official Publications of the European Communities, Luxembourg.

[24] Freedman, D. A. (2006). On the so-called "Huber sandwich estimator" and "robust standard errors". *The American Statistician*, Vol. 60, No. 4,

[25] Gentile, G., & Vigo D. (2013). Movement generation and trip distribution for freight demand modelling applied to city logistics, *European Transport/Trasporti Europei*, 54(6), 1-27.

[26] Gonzalez-Feliu, J. (2018). *Sustainable Urban Logistics: Planning and Evaluation*. John Wiley & Sons.

[27] Gonzalez-Feliu, J., Ambrosini, C., Pluvinet, P., Toilier, F., & Routhier, J.L. (2012a). A simulation framework for evaluating the impacts of urban goods transport in terms of road occupancy. *Journal of Computational Science*, 6 (4), 206-215

[28] Gonzalez-Feliu, J., Ambrosini, C., & Routhier, J.L. (2012b). New trends on urban goods movement: modelling and simulation of e-commerce distribution. *European Transport/Trasporti Europei*, 50 (6), 1-23.

[29] Gonzalez-Feliu, J., Cedillo-Campos, M.G., Garcia-Alcaraz, J.L. (2014a). An emission model as an alternative to OD matrix in urban goods transport modelling. *Dyna*, Vol. 81(187), 249-256.

[30] 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.

[31] Gonzalez-Feliu, J., & Routhier, J.L. (2012). Modeling urban goods movement: How to be oriented with so many approaches? *Procedia-Social and Behavioral Sciences*, 39, 89-100.

[32] Gonzalez-Feliu, J., Salanova Grau, J. M., & Beziat, A. (2014b). A location-based accessibility analysis to estimate the suitability of urban consolidation facilities. *International Journal of Urban Sciences*, *18*(2), 166-185.

[33] Henriot, F., & Routhier, J. L. (2010). Scenarios of commercial zoning for reducing impacts of freight movement in the city, 12th WCTR, July 11-15, 2010. Lisbon, Portugal.

[34] Holguín-Veras, J., & Jaller, M. (2014). Comprehensive freight demand data collection framework for large urban areas. In Gonzalez-Feliu, J., Semet, F., Routhier, J.L. (eds.)

Sustainable Urban Logistics: Concepts, Methods and Information Systems. Springer, Heidelberg, pp. 91-112.

[35] Holguín-Veras, J., Jaller, M., Destro, L., Ban, X. Lawson, C., Levinson, H.S. (2011). Freight Generation, Freight Trip Generation, and the Perils of Using Constant Trip Rates. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2224, pp. 68-81.

[36] Holguín-Veras, J., Jaller, M., Sanchez-Diaz, I., Wojtowicz, J., Campbell, S., Levinson, H., et al. (2012). *Freight trip generation and land use* (No. NCHRP Project 08-80).

[37] Holguín-Veras, J., Sánchez-Díaz, I., Lawson, C., Jaller, M., Campbell, S., Levinson, H.S., Shin, H.S. (2013). Transferability of Freight Trip Generation Models. *Transport Research Record*, Vol. 2379, pp. 1-8.

[38] Imbens, G. W., & Kolesar, M. (2016). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics, 98(4)*, 701-712.

[39] Institut de Recherches sur les Transports (1977). Modèle de répartition modale pour les marchandises - Classification d'un ensemble d'expéditions pour l'année 1974. IRT, Paris.

[40] Jaller, M., Sánchez-Díaz, I., Holguín-Veras, J. (2015). Identifying Freight Intermediaries: Implications for Freight Trip Generation Modeling. *Transportation Research Record: Journal of Transportation Research Board*, Vol. 2478,, pp. 48-56.

[41] Lawson, C., Holguín-Veras, J., Sánchez-Díaz, I., Jaller, M., Campbell, S., Powers, E. (2012). Estimated Generation of Freight Trips Based on Land Use. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2269, pp. 65-72.

[42] Liedtke, G., & Schepperle, H. (2004). Segmentation of the transportation market with regard to activity-based freight transport modelling. *International Journal of Logistics Research and Applications*, *7*(3), 199-218.

[43] Lewis, J.R. (1993). Multipoint scales: Mean and median differences and observed significance levels. *International Journal of Human-Computer Interaction*, 5(4), 383-392.

[44] Loebl, S.A. & Crowley, K.W. (1976). Aspects of demand for urban goods movement in city centers. *Transportation Research Record: Journal of the Transportation Research Board*, 591, 38-40.

[45] Lopez, C., Gonzalez-Feliu, J., Chiabaut, N., & Leclercq, L. Assessing the impacts of goods deliveries' double line parking on the overall traffic under realistic conditions. In *Proceedings of the 6th International Conference in Information Systems, Logistics and Supply Chain, Bordeaux, France, June 1-4, 2016*, Kedge Business School, Bordeaux.

[46] Lovric, M., ed. (2011). *International Encyclopedia of Statistical Science*. Springer, Heidelberg.

[47] Lyberg, L., & Kasprzyk, D. (1991). Data collection methods and measurement error: an overview. Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., & Sudman, S. (eds.) *Measurement errors in surveys*. John Wiley & Sons, Hoboken, NJ, pp. 235-257.

[48] Maejima, T. (1979). An application of continuous spatial models to freight movements in Greater London. *Transportation*, 8, 51-63.

[49] Meyburg A.H., & Stopher P.R. (1974). Towards Trip Attraction Models For Freight Vehicle Trips To Shopping Plazas. *Transportation Research Record*, 496, 68-79.

[50] Nuzzolo, A., & Comi, A. (2014). Urban freight demand forecasting: a mixed quantity/delivery/vehicle-based model. *Transportation Research Part E: Logistics and Transportation Review*, 65, 84-98.

[51] Patier, D. (1999). Spécificité et régularité des déterminants de la logistique urbaine ; vers une méthode globale. In *Transport Modes and Systems: Proc., 8th World Conf. on Transport Research*, vol. 1 (pp. 573-586). Pergamon, Amsterdam.

[52] Patier, D. (2001). Diagnostic sur les enquêtes françaises: tentative de globalisation. In Patier, D. (ed.) *L'intégration des marchandises dans le system des déplacements urbains* (pp. 125-144). Laboratoire d'Economie des Transports, coll. Etudes & Recherches vol. 15.

[53] Russo, F., & Comi, A. (2010). A modeling system to simulate goods movements at an urban scale. *Transportation*, vol. 37(6), 987-1009.

[54] Sánchez-Díaz, I. (2017). Modeling urban freight generation: A study of commercial establishments' freight needs. *Transportation Research Part A: Policy and Practice*, *102*, 3-17.

[55] Sanchez-Diaz (2018). The Potential of Implementing Urban Freight Strategies in the Accommodation and Food Services Sector. *Transportation Research Record: Journal of the Transportation Research Board*. (In print). ISSN 0361-1981.

[56] Sanchez-Diaz, I., Gonzalez-Feliu, J., & Ambrosini, C. (2016a). Assessing the implications of aggregating data by activity-based categories for urban freight trip generation modeling. In *Proceedings of the 6th International Conference in Information Systems, Logistics and Supply Chain, Bordeaux, France, June 1-4, 2016*, Kedge Business School, Bordeaux, ISBN 978-2-9539787-3-5.

[57] Sánchez-Díaz, I., Holguín-Veras, J., Wang, X. (2016b). An Exploratory Analysis of Spatial Effects on Freight Trip Attraction. *Transportation*, 43(1), 177-196.

[58] Sanchez-Diaz, I., Holguin-Veras, J., & Wang, C. (2013). Assessing the role of land-use, network characteristics, and spatial effects on freight trip attraction. Transportation Research Board 92nd Annual Meeting. Washington DC.

[59] Sonntag, H. (1985). A computer model for urban commercial traffic-analysis, basic concept and application. *Transport Policy and Decision Making*, 3(2).

[60] Wisetjindawat, W., & Sano, K. (2003). A behavioral modeling in micro-simulation for urban freight transportation. *Journal of the Eastern Asia Society for Transportation Studies*, *5*(3), 2193-208.

[61] Wonnacott, T.H., Wonnacott, R.J. (2001). *Introductory statistics.* Pearson Education.

Appendix: details of the assessment results

This appendix contains the whole set of results of the assessments proposed in the main paper, i.e. the comparison of constant-based estimations, the results of the model assessment and those of functional-form comparison. Table 4 shows the results of this assessment.

Table 4: MAPE for constant generation

	Category	ST8	ST27	ST43	ST17	ST84	ST105	ST8	ST27	ST43
1	Agriculture	8.84	8.84	8.84	5.78	5.78	5.78	2.62	2.62	2.62
2	Craftsmen	9.63	8.96	9.63	8.06	8.49	7.04	2.78	2.88	4.19
3	Industry	7.45	6.89	7.45	5.34	4.28	4.22	1.80	2.16	2.89
4	Wholesaling	12.34	12.54	12.34	7.60	7.08	6.97	2.38	2.35	2.32
5	Stores	5.39	5.39	5.39	5.40	5.40	2.31	0.90	1.30	0.90
6	Retailers	3.09	2.36	3.09	2.59	1.78	1.78	1.35	2.15	2.15
7	Tertiary	13.59	14.52	13.59	8.47	7.60	7.60	2.57	2.66	2.81
8	Warehousing	16.12	16.12	16.12	17.55	21.56	13.23	3.70	6.78	3.70

The summarizing results from the assessment and choice of functional form models are shown in Table 5. The table shows a short description of the groups by category, the number of observations for each group, the functional relationship, the parameters for the FTG models and MAPE.

Table 5: Selected Freight Trip Generation Models

ST8 code	ST27 code	ST43 code	Description	Obs.	Functional relationship	Model: Constant	Model: Emp.	MAPE
1	1	1	Agriculture	41	log-log	-	0.67	2.62
2	All	All	Craftsmen and services	375	log-log	-	0.58	2.78
	2	All	Craftsmen	188	lin-log	2.10	5.22	5.60
		2-2	Repair activities	52	log-log	1.14	0.61	2.66
		2-3	Manufacturing/ installation	135	lin	-	1.27	3.56
		All	Tertiary services: craftsmen	187	log-log	-0.39	0.59	2.78
	26	26Ha	Tertiary services: high flows	32	log-log	-	0.97	2.36
		26Mi	Tertiary services: mixed flows	140	log-log	-0.60	0.59	2.66
		26Mo	Tertiary services: average flows	15	log-log	-	0.66	0.72
3	All	All	Industry	623	log-log	0.42	0.61	1.80
	3	3	Chemical industry	46	log-log	-	0.88	1.24
	34	All	Construction industry	125	log-log	-	0.62	1.71
		34-2	Construction repairs	28	lin-log	-	11.78	3.09
		34-3	Construction manufacturing/installation	97	lin		0.24	1.29
	4	All	Primary and intermediate products	256	lin	6.52	0.25	4.67
		4-2	Basic bulk	144	lin	9.34	-	2.25
		4-6	Small objects	92	log-log	0.57	0.70	1.97
		4-7	Bulk	112	log-log	-	0.66	1.48
	5	All	Food and non-fragile consumer goods	196	log-log	0.44	0.65	1.71
		5-2	Consumer goods (fragile foodstuffs)	8	lin	24.69	-	4.03
		5-4	Non-consumer goods (fragile foods)	105	lin	-	1.72	1.73
		5-5	Non-fragile consumer goods (house)	82	lin	5.96	0.21	4.98

ST8 code	ST27 code	ST43 code	Description	Obs.	Functional form	Model		MAPE
						Constant	Emp.	
4	All	All	Wholesale	414	log-log	0.91	0.66	2.38
	7	All	Intermediary products	200	log-log	0.83	0.65	2.50
		7-2	Fragile intermediate products	80	log-log	1.11	0.61	3.50
		7-3	Other intermediate products	120	log-log	0.72	0.64	2.03
	8	All	Nonfood consumer goods	118	log-log	0.88	0.61	2.08
		8-2	Non-food fragile consumer goods	92	log-log	1.08	0.54	2.24
		8-3	Non-food non-fragile consumer goods	33	lin-log	-	6.12	2.79
	9	All	Food	96	log-log	1.25	0.67	2.22
		9-2	Fragile food consumer goods	32	log-log	1.90	0.61	0.95
		9-3	Other food consumer goods	64	log-log	0.88	0.71	2.22
5	10	All	Department stores	47	log-log	-	0.89	0.90
		10	Hypermarkets and department stores	13	lin	-	0.54	0.65
		11	Supermarkets	20	lin	11.49	0.60	2.09
		12	Specialized department stores	14	log-log	-	0.85	0.80
6	All	All	Retailers	1080	log-log	0.98	0.45	1.35
	13	13	Minimarkets	13	lin	15.44	-	1.44
	14	14	Clothing. shoes. leather	109	lin	2.01	1.70	4.88
	15	15	Butcher's shops	70	lin	3.55	1.76	1.22
	16	16	Small groceries	97	lin	4.34	1.02	1.35
	17	17	Bakery retailers	112	lin	6.49	0.18	1.57
	18	18	Hotels. restaurants. cafés	145	lin	2.63	0.61	1.38
	19	19	Pharmacies	61	lin	15.94	1.94	2.02
	20	20	Hardware stores	43	lin	4.12	-	1.56
	21	21	Furnishing shops	49	lin-log	-	4.67	4.24
	22	22	Bookshops	90	log-log	1.98	-	1.28
	23	23	Other retail shops	261	lin-log	-	5.68	1.89
	29	29	Street trading (marketplaces)	30	lin	-	5.77	5.59
7	All	All	Tertiary/ offices	322	log-log	-0.81	0.53	2.57
	6	6	Transport except storage	34	log-log	-	0.38	5.74
	24	24	Other tertiary activities with low flows	32	log-log	-0.89	0.47	1.55
	25	All	Offices	256	log-log	-0.83	0.53	2.57
		25	Pure tertiary sector (offices)	178	log-log	-1.00	0.53	1.89
		27-2	Non-tertiary offices (agriculture. wholesale)	48	log-log	-	0.64	3.02
		27-3	Non-tertiary offices (other activities)	30	log-log	-1.70	0.75	4.28
8	28	All	Warehouses/ transport	68	log-log	1.68	0.66	3.70
		28-2	Warehouses (bulk)	21	log-log	-	1.53	3.08
		28-3	Warehouses (with transport)	68	lin-log	-	27.87	11.96