



## Sampling for freight trip generation studies

Downloaded from: <https://research.chalmers.se>, 2026-10-12 20:38 UTC

Citation for the original published paper (version of record):

Sanchez-Diaz, I., Hurtado, K. (2026). Sampling for freight trip generation studies. Transportation Research Procedia, 98: 36-43. <http://dx.doi.org/10.1016/j.trpro.2026.04.102>

N.B. When citing this work, cite the original published paper.

The 13th International Conference on City Logistics (City Logistics 2025)

## Sampling for freight trip generation studies

Ivan Sanchez-Diaz<sup>a\*</sup>, Kevin Jiménez Hurtado<sup>b</sup>

<sup>a</sup>*Chalmers University of Technology, Gothenburg, 412 17, Sweden*

<sup>b</sup>*École des Mines de Saint-Étienne, Saint-Étienne, 420 23, France*

---

### Abstract

Freight Trip Generation (FTG) models link establishment characteristics to delivery activity and are widely used in urban freight planning. While their reliability depends on establishment surveys, the implications of sampling design remain poorly understood. This paper examines how alternative sampling strategies affect FTG model performance and cost, using population data to sample from establishments in Stockholm. The authors compare simple random, stratified, and hybrid sampling across proportional, Neyman, and optimal allocation rules, as well as three levels of sector aggregation. Results show that simple random sampling overrepresents large, low-impact sectors and underrepresents freight-intensive ones. Stratified approaches improve accuracy, but only when allocation accounts for sectoral variance and survey costs. Aggregation reduces required sample sizes, with moderate aggregation offering a practical compromise and high aggregation suited for strategic planning.

© 2026 The Authors. Published by ELSEVIER B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

The peer review will be performed under the responsibility of the 13th International Conference on City Logistics (City Logistics 2025).

**Keywords:** Freight Trip Generation; Sampling Techniques; Simple Random Sampling; Stratified Random Sampling; Performance Assessment

---

### 1. Introduction

Freight Trip Generation (FTG) models are a cornerstone for urban freight research and practice. Their primary role is to estimate the number of delivery trips generated by commercial establishments, which is a critical input for both short-term management measures—such as curbside regulations and loading zone allocation—and long-term planning decisions, including infrastructure development and land use zoning. By linking establishment attributes (e.g., economic activity, number of employees or floor area, or location) with freight deliveries, FTG models provide a structured way to forecast freight traffic and assess policy interventions.

The quality of FTG models, however, depends directly on the quality of the data from which they are derived. Most FTG studies rely on establishment surveys (Sanchez-Diaz 2016), yet the way these surveys are sampled is often under-

\*Corresponding author. Email address: [ivan.sanchez@chalmers.se](mailto:ivan.sanchez@chalmers.se)

documented, under-theorized, and treated as a secondary concern. In the literature, details on sample design are scarce, even though sampling choices strongly influence model performance, estimation costs, and ultimately the reliability of policy analysis. A sample designed for evaluating the long-term effects of infrastructure investment may not be suitable for assessing the operational impacts of consolidation policies in a specific district. This mismatch risks undermining the policy relevance and transferability of FTG models.

This paper addresses that gap by explicitly examining the role of sampling in FTG studies. The paper evaluates how different sampling strategies—simple random sampling (SRS), stratified random sampling (STR), and hybrid designs—affect model form, estimation accuracy, and data collection costs. It also assesses how these trade-offs shape the usefulness of FTG models for various policy purposes. Accordingly, this study is guided by two research questions, (1) which sample designs are most appropriate for different types of FTG studies? and, (2) how do sampling design choices affect the reliability and cost of FTG models?

This paper uses data from Stockholm to provide a systematic evaluation of sampling strategies in FTG research, offering guidance for both modeling practice and policy-oriented data collection. The remainder of the paper is organized as follows: Section 2 outlines the key principles of sampling theory. Section 3 describes the conceptual framework of FTG models. Section 4 presents the methodology adopted in this study. Section 5 describes the main results of this research, while the final section summarizes the key findings and policy implications.

## 2. Sampling Theory and Background

The reliability of establishment-based FTG models depends fundamentally on the sampling strategies used to collect data. Since it is impractical to survey all establishments in a city, researchers must rely on a sample of units, and the quality of this sample conditions both model performance and policy relevance. Despite this, several problems are recurrent in FTG sampling practice. First, non-random or convenience samples introduce bias by over-representing certain types of establishments. Second, the omission of relevant establishment attributes limits explanatory power and weakens the statistical robustness of models. Third, some freight-intensive sectors are frequently under-represented, which leads to systematic underestimation of freight trips. Finally, poor documentation of sampling procedures makes it difficult to replicate results or assess external validity. Together, these issues produce some negative consequences: mis-estimated trip volumes, inaccurate forecasts of policy impacts, limited transferability of results across contexts, and potential misalignment between model outputs and urban freight interventions.

From a statistical perspective, a well-designed sample aims to balance representativeness and cost. Let the total population of establishments be  $N$ , divided into  $H$  strata of size  $N_h$ . A sample of size  $n$  can be drawn using different strategies. In Simple Random Sampling (SRS), each unit has equal probability of selection (Cochran 1977), with expected sample variance given by Equation 1, where  $s_1$  and  $\bar{y}_1$  are the estimated population standard deviation and mean,  $t_{\alpha/2}$  the critical value of the t-distribution at  $\alpha/2$ ,  $r$  the desired relative error and  $n_1$  the pre-sample size. SRS is straightforward to implement but assumes relative homogeneity across establishments, which rarely holds in urban freight. Stratified Random Sampling (STR) addresses this limitation by dividing the population into strata according to sectoral or other relevant attributes (Cochran 1977). The overall sample mean, i.e.,  $\bar{y}_{st}$ , is then a weighted average of stratum-specific means (see Equation 2), where the weights are defined as  $W_h = N_h/N$ .

$$n \geq \left( \frac{s_1 t_{\alpha/2}}{r \bar{y}_1} \right)^2 \left( 1 + \frac{2}{n_1} \right) \quad (1)$$

$$\bar{y}_{st} = \sum_{h=1}^H W_h \bar{y}_h \quad (2)$$

Sample size allocation across strata is central to STR. Under Proportional Allocation (PA),  $n_h = n W_h$ , which reflects population structure without accounting for heterogeneity, while Neyman Allocation (NA) adjusts for stratum variance, giving more weight to high-variance strata (Equation 3). Optimal Allocation (OA) further incorporates data collection costs (see Equation 4), allocating observations based on both variance and per-unit cost ( $c_h$ ).

$$n_h = n \frac{N_h s_h}{\sum_{h=1}^H N_h s_h} \quad (3)$$

$$n_h \geq n \frac{N_h s_h / \sqrt{c_h}}{\sum_{h=1}^H N_h s_h / \sqrt{c_h}} \quad (4)$$

In practice, hybrid designs that combine elements of SRS and STR are often employed to balance simplicity and representativeness. Furthermore, aggregation choices—such as grouping sectors into broader categories—can reduce sample sizes and costs, but at the expense of detail. These methodological choices must therefore be aligned with the intended purpose of the FTG model: while long-term infrastructure planning may tolerate aggregated sampling, operational studies for curbside management or consolidation require more detailed and representative designs.

### 3. Freight Trip Generation (FTG) Models

FTG models establish the relationship between establishment characteristics and the number of freight trips they generate over a given time period. Their main function is to serve as input for estimating freight traffic volumes, which are essential for both operational policies (e.g., allocation of curb space, consolidation schemes) and long-term planning (e.g., infrastructure and land-use development) (Holguín-Veras & Sánchez-Díaz, 2016).

Most FTG models are specified as regression equations of the form (Sanchez-Diaz 2016):

$$Y_i = \alpha + \sum_{k=1}^K \beta_k X_{ik} + \varepsilon_i \quad (5)$$

Where  $Y_i$  is the number of trips generated by establishment  $i$ ,  $X_{ik}$  represents the  $k^{th}$  explanatory variable describing the size of the establishment,  $\beta_k$  are the estimated coefficients, and  $\varepsilon_i$  is the error term. The interpretation of these coefficients depends on the sample: over- or under-representation of sectors may bias  $\beta_k$  or limit external validity. The choice of explanatory variables and the aggregation level of economic activities is crucial. Highly disaggregated models capture heterogeneity better but require larger samples, while aggregated models reduce sampling requirements and estimation variance but obscure sector-specific effects. This study tests different aggregation levels.

A literature review shows that the two most common sampling techniques in establishment-based FTG studies are SRS and STR. In SRS, sample sizes are often estimated using widely known formulas from the survey sampling field (Dhonde and Patel 2021, Pani, Sahu et al. 2021, Pani, Sahu et al. 2023). While SRS is straightforward and statistically robust, it assumes homogeneity across sectors, which is limiting since different sectors generate distinct trip patterns. Alternatively, some studies have used STR to better capture heterogeneity across sectors (Lawson, Holguín-Veras et al. 2012, Alho and de Abreu e Silva 2015, Sánchez-Díaz, Holguín-Veras et al. 2016, Gonzalez-Calderon, Sanchez-Diaz et al. 2018, Holguín-Veras, Ramirez-Rios et al. 2019, Cheah, Mepparambath et al. 2021), though reporting of procedures is often limited. In many cases, sectoral allocation is described only in general terms, and sample sizes are justified by resource constraints or by the minimum number of observations required to run a statistically valid model, rather than by explicit allocation rules. Samples are frequently defined as a population percentage or drawn from earlier surveys, leaving open the question of how to balance representativeness, sectoral variability, and cost.

Model performance is evaluated through both estimation accuracy (root mean squared error, RMSE; standard errors of coefficients) and predictive validity (comparing total FTG estimates to benchmarks). This dual approach ensures that models are not only statistically robust but also meaningful for policy applications.

It is noteworthy that the relationship between FTG model quality and sampling design lies in how well the sample captures sectoral heterogeneity. Stratified and hybrid designs can reduce the variance of coefficient estimates by ensuring representation of high-variance sectors, while simple random samples risk imbalances that inflate errors. At the same time, data collection costs increase with disaggregation, creating a trade-off between detail and feasibility.

### 4. Methodology

To empirically assess the implications of sampling design for FTG modelling, the authors used establishment data from Stockholm. The total number of establishments includes 11,279 establishments across seven economic sectors. The geographical area of the study was selected based on the target of the urban freight strategies formulated in the Stockholm Freight Plan. From this population, the authors designed and drew samples under different strategies and evaluate the resulting FTG models in terms of statistical performance, predictive accuracy, and data collection costs.

#### 4.1. Sampling Strategy: Design and Assessment

The method for this paper follows five steps:

1. **Pre-sample model specification.** A reference FTG model is estimated to serve as a benchmark (Sanchez-Diaz 2016). This model links freight trips per establishment to explanatory variables such as employment, floor area, and sectoral classification.
2. **Computation of sample sizes.** For each sampling strategy, sample sizes are calculated using standard formulas (see Section 2), ensuring comparability in confidence and precision levels. Both simple random sampling (SRS) and stratified random sampling (STR) designs are tested, along with hybrid approaches that combine proportional SRS with stratification in selected sectors.
3. **Sample selection and aggregation levels.** Establishments are drawn according to each design. In the stratified case, the authors also test alternative levels of sectoral aggregation, since the degree of disaggregation influences both cost and accuracy:
  - **AL1:** All seven sectors modeled separately (maximum detail).
  - **AL2:** Retail sectors and food services are combined into broader categories, reducing the number of strata.
  - **AL3:** In addition to AL2, wholesale and manufacturing are aggregated, further reducing complexity.

These aggregation levels are not based on generic sampling theory, but design choices specific to this study, reflecting the balance between policy detail and resource constraints. For each aggregation level, proportional allocation (PA), Neyman allocation (NA), and optimal allocation (OA) are applied.

4. **Model estimation.** For each sample, FTG models are estimated using ordinary least squares (OLS). Estimated coefficients, standard errors, and overall fit are compared against the population benchmark to assess the impact of sampling design on model form and parameter stability.
5. **Validation and performance evaluation.** Model outcomes are evaluated across three dimensions:
  - **Estimation accuracy**, measured by the RMSE of predictions relative to observed trips in the sample.
  - **Predictive validity**, assessed by comparing total FTG estimated from each sample-based model against a random sample from the FTG benchmark dataset using Leave-One-Out-Cross-Validation.
  - **Cost efficiency**, calculated by multiplying the number of observations required under each design by the estimated per-survey cost in euros.

#### 4.2. Case Study Context

The Stockholm case is well-suited for this analysis, with detailed establishment records stratified by sector and location and an active freight plan linking establishment activity to policy measures. Comparing models across sampling designs, aggregation levels, and allocation rules, this case illustrates trade-offs between accuracy, representativeness, and cost, providing policy-relevant guidance for selecting FTG sampling strategies.

Although an ideal dataset would contain FTG for the entire population to benchmark sample-based FTG models against ground truth, this is rarely possible in practice. Alternatively, the authors use the data from studies conducted in Sweden to test the different sampling strategies: the FTG study in Gothenburg (Sanchez-Diaz 2016) is used for pre-sample and the FTG study in Stockholm (Sánchez-Díaz 2018, Sanchez-Diaz 2020) for assessment.

The sample for the FTG study in Stockholm ( $n=388$ ) was selected using a random drawing from a database containing all the establishments in the study area. The sample size was determined based on the FTG mean and variance from previous studies conducted in Gothenburg, Sweden (Sanchez-Diaz 2016). The data were segmented into eight different strata using commercial sectors as designated by the Swedish National Industry (SNI) codes. This paper focuses on the data collected for the following sectors: accommodation (A), food services (FS), retail perishable (RP), retail non-perishable (RNP), wholesale (W), manufacturing (M) and offices (O) (see Table 1).

## 5. Results

### 5.1. Sample Sizes Across Designs

Table 2 shows the required sample sizes under different designs using Table 1 inputs (5% significance level, 4% target relative error). Sample sizes were adjusted when exceeding available data (set to maximum) or too small for FTG estimation (set to 12). For simple random sampling (SRS) with all seven sectors (AL1), the estimated minimum

sample size is 296 establishments. In contrast, stratified random sampling (STR) with greater aggregation and optimal allocation (AL3, OA) needs only 73 establishments, reflecting the efficiency gains from targeted allocation. Intermediate designs, such as hybrid SRS–STR, yield sample sizes between these extremes. These results illustrate the strong influence of both aggregation level and allocation rule on the number of establishments to be surveyed.

Since the OA method accounts for cost, and the W sector has the highest data collection cost (along with O), its sample size under OA is lower than under NA. In contrast, although the RP sector also has high variability, its small population size requires adjusting sample sizes across all allocation methods. Notably, total sample sizes with PA method are always larger than with NA or OA—even when no adjustments are made (e.g., in SRS–STR). This is because PA ignores sectoral variance, thus requiring equal or larger sample sizes to meet the same relative error threshold.

Table 1. Input data to estimate the sample sizes in Stockholm.

| Sector                                 | FTG Reg Mean<br>(trips/week) | FTG Reg Std<br>Dev (trips/week) | Population<br>Size | Pre-sample<br>Size | Cost<br>(€/unit) | Max<br>Sample Size |
|----------------------------------------|------------------------------|---------------------------------|--------------------|--------------------|------------------|--------------------|
| Accommodation (A)                      | 6.12                         | 0.41                            | 190                | 95                 | 57.72            | 47                 |
| Food Services (FS)                     | 5.39                         | 0.67                            | 1896               | 41                 | 57.72            | 44                 |
| Manufacturing (M)                      | 9.32                         | 1.09                            | 245                | 125                | 57.72            | 22                 |
| Offices (O)                            | 3.74                         | 0.48                            | 8537               | 256                | 115.44           | 117                |
| Retail non-perishable (RNP)            | 4.33                         | 1.15                            | 2514               | 65                 | 57.72            | 63                 |
| Retail Perishable (RP)                 | 9.30                         | 1.91                            | 519                | 53                 | 57.72            | 65                 |
| Wholesale (W)                          | 18.78                        | 3.23                            | 932                | 414                | 115.44           | 30                 |
| Accommodation and Food services (A&FS) | 5.61                         | 0.43                            | 2086               | 136                | 57.72            | 91                 |
| Retail (R)                             | 5.98                         | 1.10                            | 3033               | 118                | 57.72            | 128                |
| Wholesale and Manufacturing (W&M)      | 17.26                        | 2.49                            | 1177               | 539                | 103.43           | 52                 |

Table 2. Sample sizes obtained for each sample design.

| Sector | AL1 |     |     |     | AL2 |         |     |     |     |    | AL3 |     |         |     |     |     |    |     |
|--------|-----|-----|-----|-----|-----|---------|-----|-----|-----|----|-----|-----|---------|-----|-----|-----|----|-----|
|        | SRS | STR |     |     | SRS | SRS-STR |     |     | STR |    |     | SRS | SRS-STR |     |     | STR |    |     |
|        | SRS | NA  | OA  | PA  | SRS | NA      | OA  | PA  | NA  | OA | PA  | SRS | NA      | OA  | PA  | NA  | OA | PA  |
| A      | 16  | 12  | 12  | 12  |     | 12      | 12  | 12  |     |    |     |     | 12      | 12  | 12  |     |    |     |
| FS     | 44  | 12  | 12  | 19  |     | 44      | 44  | 43  |     |    |     |     | 44      | 44  | 43  |     |    |     |
| M      | 22  | 12  | 12  | 12  | 22  | 22      | 22  | 22  | 12  | 12 | 12  |     | 12      | 12  | 20  |     |    |     |
| O      | 56  | 28  | 24  | 83  | 56  | 56      | 56  | 56  | 24  | 21 | 69  | 56  | 56      | 56  | 56  | 23  | 21 | 57  |
| RNP    | 63  | 20  | 24  | 25  |     | 63      | 63  | 63  |     |    |     |     | 63      | 63  | 63  |     |    |     |
| RP     | 65  | 12  | 12  | 12  |     | 50      | 50  | 36  |     |    |     |     | 50      | 50  | 36  |     |    |     |
| W      | 30  | 21  | 18  | 12  | 30  | 30      | 30  | 30  | 18  | 16 | 12  |     | 30      | 30  | 30  |     |    |     |
| A&FS   |     |     |     |     | 21  |         |     |     | 12  | 12 | 17  | 21  |         |     |     | 12  | 12 | 14  |
| R      |     |     |     |     | 116 |         |     |     | 20  | 25 | 25  | 116 |         |     |     | 19  | 24 | 21  |
| W&M    |     |     |     |     |     |         |     |     |     |    |     | 52  |         |     |     | 17  | 16 | 12  |
| Total  | 296 | 117 | 114 | 175 | 245 | 277     | 277 | 262 | 86  | 86 | 135 | 245 | 267     | 267 | 260 | 71  | 73 | 104 |

Note: \*In shadow the sample sizes that had to be adjusted.

## 5.2. Model Performance

Estimation accuracy is first assessed using the root mean squared error (RMSE). Models estimated with SRS at AL1 and hybrid SRS–STR designs at AL2 achieve the lowest RMSE values, indicating superior predictive performance. In contrast, pure STR designs at AL2 perform poorly, particularly in sectors with small population weights but high variance, where under-sampling leads to unstable coefficients. Overall, hybrid designs balance the benefits of randomization with the representativeness of stratification, resulting in models that approximate the population benchmark most closely.

## 5.3. Estimation Validity

Estimation errors are further analyzed by comparing model predictions with observed trips in the population. The lowest relative errors are observed for SRS (AL1) and SRS–STR (AL2), which remain within 5% of the benchmark. STR-based models exhibit higher errors in low-weight, high-variance sectors, confirming the sensitivity of regression

coefficients to uneven representation. These results suggest that while STR can reduce variance theoretically, in practice it may underestimate freight activity in certain sectors if allocation is not carefully calibrated.

#### 5.4. Total FTG Estimates

The total reported freight trips per week in the benchmark dataset in Stockholm is 160,904. The SRS (AL1) design yields an almost identical estimate (160,904), while SRS–STR (AL2) produces a total of 160,243, a difference of only 0.4%. By contrast, STR designs systematically underestimate total trips, with deviations that can exceed 5%. This finding underlines the importance of matching sampling design to policy application: while small deviations may be acceptable for exploratory analysis, underestimation of demand could seriously misinform infrastructure planning.

#### 5.5. Data Collection Costs

Data collection costs were calculated by multiplying required sample sizes by an estimated cost per survey based on previous experience. Under this assumption, SRS (AL1) would cost approximately €22,000, compared to only €6,200 for STR with OA at AL3. Hybrid designs fall in between, with costs around €21,000. This highlights the financial advantages of aggregated STR designs, although these come at the expense of predictive accuracy.

#### 5.6. Trade-offs Between Cost and Accuracy

Taken together, the results reveal a clear trade-off. Disaggregated samples (SRS, AL1) maximize model accuracy but are costly. Aggregated stratified samples (STR, AL3, OA) minimize cost but systematically underestimate trips. Hybrid approaches provide the most balanced outcome, delivering acceptable accuracy at moderate costs. From a policy perspective, the choice of design should therefore be aligned with the intended application: short-term urban management requires higher accuracy and thus larger samples, while long-term strategic planning may tolerate aggregated designs at lower cost.

### 6. Discussion and Policy Implications

The results show that the performance of FTG models depends not only on statistical efficiency but also on sector heterogeneity and practical constraints in data collection. By linking sector characteristics, allocation rules, and aggregation levels, it can be better understood how sampling design decisions translate into policy-relevant outcomes.

#### 6.1. Sectoral Priorities in Sampling

As summarized in Table 4, sectors differ sharply in terms of population size, freight intensity, and variance. High-variance, high-intensity sectors such as retail perishable and wholesale are critical for reliable FTG estimation, yet they are often under-sampled in simple random designs. Conversely, large but low-intensity sectors such as offices absorb a disproportionate share of the sample in proportional allocation, yielding little marginal information for freight modeling. This implies that purely random or proportional strategies are misaligned with the structure of freight activity, and careful prioritization of sectors is needed to avoid systematic bias in trip estimation.

Table 3: Implications of sectors FTG characteristics for the sampling strategy.

| FTG    |            |           |           |      |                                        |                                |
|--------|------------|-----------|-----------|------|----------------------------------------|--------------------------------|
| Sector | Obs.       | Mean      | Variance  | Cost | Allocation Strategy                    | Sampling Method                |
| A      | Small      | Moderate  | Very low  | Low  | Low priority                           | Undersampled in SRS            |
| FS     | Large      | Moderate  | Low       | Low  | Medium priority                        | Well covered in SRS            |
| M      | Small      | High      | Moderate  | Low  | Medium priority; downsized under OA    | Undersampled in SRS            |
| O      | Very large | Very low  | Very low  | High | Overallocated in PA; down-weight in OA | Overrepresented in SRS         |
| RNP    | Large      | Moderate  | High      | Low  | High priority under NA/OA              | Adequately covered             |
| RP     | Moderate   | High      | Very high | Low  | Very high priority                     | At risk of being missed in SRS |
| W      | Moderate   | Very high | Very high | High | Adjust for cost in OA                  | Undersampled in SRS            |

## 6.2. Implications of Allocation Strategies

The choice of allocation strategy has significant implications for both accuracy and cost (Table 5). Proportional allocation is simple to implement but tends to over-sample sectors like offices, resulting in inefficiencies. Neyman allocation improves statistical efficiency by directing observations toward sectors with high variance, such as wholesale or retail perishable, but may become cost-prohibitive. Optimal allocation provides a middle ground by incorporating cost considerations, ensuring that high-variance but costly sectors such as manufacturing and wholesale are not oversampled. From a policy standpoint, this makes OA more suitable for large-scale or resource-constrained surveys, while NA may be preferable in research contexts where cost is less of a concern.

Table 4: Implications of allocation strategy for FTG studies.

| Allocation Strategy                 | Why?                                                        | Implication                                                                 | Example                                                               |
|-------------------------------------|-------------------------------------------------------------|-----------------------------------------------------------------------------|-----------------------------------------------------------------------|
| <b>Proportional Allocation (PA)</b> | Allocates samples by sector's population share.             | Simple, but may oversample large low-impact sectors and miss critical ones. | Offices (many sites, low FTG) receive large samples.                  |
| <b>Neyman Allocation (NA)</b>       | Allocates more samples to sectors with higher FTG variance. | Statistically efficient; prioritizes heterogeneous sectors.                 | Wholesale (moderate size, high FTG variance) receives larger samples. |
| <b>Optimal Allocation (OA)</b>      | Balances variance and data collection cost.                 | Cost-efficient; optimizes sample allocation under budget limits.            | Wholesale (high variance & cost) gets fewer samples than under NA.    |

## 6.3. Aggregation and Policy Applications

Sector aggregation (Table 5) is another critical dimension. At the most detailed level (AL1), sectors are kept separate, allowing policymakers to capture fine-grained differences in FTG patterns. This is particularly valuable for operational measures, such as curbside regulations or consolidation schemes, where the spatial and temporal behavior of freight differs strongly across sectors. Moderate aggregation (AL2) reduces complexity while preserving the main differences across freight-intensive activities, offering a pragmatic balance between detail and feasibility. High aggregation (AL3) is cost-effective and easier to implement but masks sectoral distinctions, making it more suitable for long-term strategic planning where broad trends are more important than sector-specific interventions.

Table 5: Implications of aggregation level for FTG sampling.

| Aggregation Level                 | Sector Groupings                                                                                        | Description                                                                                             |
|-----------------------------------|---------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| <b>AL1 – Full Disaggregation</b>  | No grouping applied                                                                                     | Each sector treated as a separate stratum; maximum resolution in sectoral analysis.                     |
| <b>AL2 – Moderate Aggregation</b> | -Combined Accommodation & Food Services (A&FS) - Combined into Retail (R) -W, M, O separate             | Sectors with similar service characteristics grouped; reduces number of strata while preserving detail. |
| <b>AL3 – High Aggregation</b>     | -A&FS (as above) -R (as above) -W & M combined into Wholesale & Manufacturing (W&M) -O remains separate | Most consolidated structure; reduces complexity and balances strata sizes for cost-effective sampling.  |

## 6.4. Policy Relevance of Sampling Choices

Taken together, these findings highlight that sampling choices must be explicitly aligned with the intended use of FTG models. For short-term management (e.g., loading zones, delivery windows, consolidation pilots), disaggregated samples with careful allocation (AL1/AL2) are essential to capture sectoral detail. For long-term planning (e.g., freight corridors, zoning, or infrastructure investment), aggregated designs (AL3 with OA) may provide a sufficient level of accuracy at a fraction of the cost. Importantly, the results demonstrate that no design is universally superior; instead, sampling must be treated as a design variable itself, with transparent documentation of its rationale and limitations.

## 7. Conclusions

This paper has examined the implications of sampling design for FTG models through an empirical case study of Stockholm. By systematically comparing simple random sampling, stratified sampling, and hybrid designs across different allocation rules and aggregation levels, the paper demonstrates that sampling decisions exert a decisive influence on model accuracy, predictive validity, and data collection costs.

The results confirm that simple random sampling, while easy to implement, tends to misrepresent high-variance freight-intensive sectors and over-sample large but low-impact ones. Stratified and hybrid approaches improve accuracy by ensuring representation of critical sectors, but their efficiency depends strongly on allocation and aggregation choices. Neyman allocation directs observations toward high-variance sectors, enhancing statistical efficiency but at high cost, while optimal allocation balances accuracy and budget constraints. Similarly, aggregation reduces costs but at the expense of sectoral detail, with AL2 emerging as a pragmatic compromise and AL3 more suitable for strategic planning.

From a policy perspective, these findings underscore that there is no single “best” sampling strategy. Instead, the design must be aligned with the purpose of the FTG model: disaggregated, carefully allocated samples are needed for short-term management measures, while aggregated and cost-efficient designs may suffice for long-term strategic planning. Across all cases, transparent documentation of sampling procedures is essential to ensure reproducibility, comparability across studies, and credibility in policy analysis.

By providing the first systematic empirical evaluation of FTG sampling strategies, this study contributes both to research and practice. For researchers, it highlights the methodological importance of treating sampling as a central design variable rather than a technical afterthought. For practitioners, it offers guidance on how to balance accuracy and cost when commissioning establishment surveys for freight planning. Future work should extend this analysis to other urban contexts and explore how sampling interacts with emerging data sources such as digital transaction records or sensor-based observations.

## References

- Alho, A. R. and J. de Abreu e Silva (2015). "Modeling retail establishments' freight trip generation: a comparison of methodologies to predict total weekly deliveries." *Transportation*: 1-18.
- Cheah, L., R. M. Mepparambath and G. M. R. Surribas (2021). "Freight trips generated at retail malls in dense urban areas." *Transportation Research Part A: Policy and Practice* **145**: 118-131.
- Cochran, W. G. (1977). *Sampling techniques*. John Wiley & sons.
- Dhonde, B. N. and C. R. Patel (2021). "Estimating urban freight trips using light commercial vehicles in the Indian textile industry." *Transportation Research Interdisciplinary Perspectives* **11**: 100411.
- Gonzalez-Calderon, C., I. Sanchez-Diaz, I. Sarmiento-Ordosgoitia and J. Holguín-Veras (2018). "Characterization and analysis of metropolitan freight patterns in Medellin, Colombia." *European Transport Research Review (in print)*.
- Holguín-Veras, J., D. Ramirez-Rios, T. Encarnacion, J. González-Feliu, E. Caspersen, C. Rivera-Gonzalez, C. González-Calderón and R. Da Silva Lima (2019). Metropolitan Economies and the Generation of Freight and Service Activity: An International Perspective. *Urban Logistics: Management, Policy and Innovation in a Rapidly Changing Environment*. M. Browne, S. Behrends, J. Woxenius, G. Giuliano and J. Holguín-Veras, Kogan Page Limited: 19-51.
- Lawson, C. T., J. Holguín-Veras, I. Sánchez-Díaz, M. Jaller, S. Campbell and E. L. Powers (2012). "Estimated generation of freight trips based on land use." *Transportation Research Record* **2269**(1): 65-72.
- Pani, A., P. K. Sahu and F. A. Bhat (2021). "Assessing the spatial transferability of freight (trip) generation models across and within states of India: empirical evidence and implications for benefit transfer." *Networks and Spatial Economics* **21**(2): 465-493.
- Pani, A., P. K. Sahu, L. Tavasszy and S. Mishra (2023). "Freight activity-travel pattern generation (FAPG) as an enhancement of freight (trip) generation modelling: Methodology and case study." *Transport Policy* **144**: 34-48.
- Sanchez-Díaz, I. (2016). "Modeling Urban Freight Generation: A Study of Receivers' Freight Needs " *Transportation Research Part A: Policy and Practice* **102**: 3-17.
- Sanchez-Díaz, I. (2020). "Assessing the magnitude of freight traffic generated by office deliveries." *Transportation research part A: policy and practice* **142**: 279-289.
- Sánchez-Díaz, I. (2018). "Potential of implementing urban freight strategies in the accommodation and food services sector." *Transportation Research Record* **2672**(9): 194-203.
- Sánchez-Díaz, I., J. Holguín-Veras and X. Wang (2016). "An exploratory analysis of spatial effects on freight trip attraction." *Transportation* **43**(1): 177-196.