**Flash Estimates and increasing data quality as a result of modern IT-applications and job satisfaction**

Sabine Schuster, Statistics Austria, sabine.schuster@statistik.gv.at

Thomas Karner, Statistics Austria, thomas.karner@statistik.gv.at

**Abstract**

*From 2017 onwards, Statistics Austria publishes flash estimates for road freight transport statistics. These flash estimates enable data users to get first key figures of freight transport of Austrian road freight vehicles already one month after the reference quarter (t+1) instead of five months (t+5).*

*Road freight transport statistics is – with reference to Council Regulation (EC) No 70/2012 on statistical returns in respect of the carriage of goods by road (recast) – performed as a sample survey. Weekly, questionnaires are sent to the respondents, who have to report all journeys of the selected trucks and return the web- or paper-based form to Statistics Austria two weeks later.*

*To check the plausibility of the filled-in forms a Java-application is used since 2014. Prior to that, the reprocessing was very complicated and included different steps and different people in charge. Now, each questionnaire is attributed to one staff member, who is registering its completeness, checking its plausibility and - if necessary - is getting in contact with the respondent to clarify certain parameters. Once finalised, each questionnaire is directly added to the authentic database. Following the two-factor theory of Herzberg about job satisfaction, it was noticed, that the motivation of the staff members improved due to intrinsic factors such as increased responsibility and decision autonomy. The work tasks have become more demanding and therefore interesting with a better insight and understanding of the whole process. Thereby the quality and availability of the produced data enhanced.*

*It was analysed, that one month after the reference quarter already more than 70% of the questionnaires could be correctly finalised. Calculated flash estimates showed similar outcomes to the final key results with a slightly higher sampling error which is still in the frame of precision requirements.*

**Keywords:** Motivation, Quality, Flash Estimates, IT-Applications

1. Preface

Transport statistics includes surveys of different modes of transport (road, rail, inland waterway, air and pipeline), which differ in terms of concepts, legal bases and survey techniques.

Road freight transport statistics is performed as a sample survey. Based on the nationality principle only trucks or semitrailer tractors registered in Austria and above 2 tonnes load capacity are included. The Austrian sample population consists of approximately 66 000 vehicles, whereof a stratified sample of 26 000 vehicles is drawn. For these vehicles all journeys within one reference week have to be reported to Statistics Austria.

After validating and if necessary correcting the received questionnaires the data will be extrapolated to the sample population. The results are published with a time lag of five months after the respective reference quarter (t+5).

The first part of this article reviews the improvements in the validation process due to the development of a Java-application and the resulting impact on the work of the employees and their motivation. The second part describes the analysis before publishing flash estimates. Flash Estimates provide data on transport volume and transport performance already one month after the end of the reporting period instead of five months for regular estimates.

1. Process improvement and its impact on data availability and quality
	1. Implementation of a modern, IT-supported application

IT-supported applications contribute to the statistical process to an increasing degree and enable the integration of several validation steps into one process. In 2014, such a Java-application was developed for road freight transport statistics. The survey for road freight transport statistics is a complex questionnaire which is available in three versions (web questionnaire, personalized Excel and paper-based form). Depending on the use of the road freight vehicle during the reporting week, the questionnaire can be easy to fill in (e.g. only a few trips over long distances) or very complicated and detailed (e.g. many laden journeys within a day).

Prior to the implementation of the application, the validation of one questionnaire involved several independent sub-steps with different people in charge. As some companies have to report several vehicles per reference week, in a first step, the completeness of all received questionnaires was checked by an employee and where required the company was contacted.

Once the completeness was given, the questionnaire was transferred to the host system with other employees in charge. In case of a paper-based questionnaire, the data had to be entered manually while the transfer of the web questionnaire happened partially automated. The plausibility checks at this stage applied to the input and the correctness of valid variable values. The consistence of specific variables in relation to each other (e.g. maximum permissible total weight in relation to loading) was only partially tested. Occasionally, journeys were reconstructed using street maps or route planners in order to understand the logic and sequence of loading and unloading locations.

At a later stage, more complex plausibility checks or the automated imputation of missing empty journeys were performed across the entire database and lead to further manual checks and editing. Completely examining and verifying a questionnaire was a complex matter, required special skills (e.g. very good geographic knowledge) and took several months.

* 1. Reorganization of the workflows

With the use of the new application, the workflows of the validation process changed completely. Before different employees were in charge of different, independent work steps. Now each employee is responsible for his assigned questionnaire and its correctness – with the option to include a supervisor if necessary. Instead of only manually entering the questionnaire into the system or verifying the correctness of individual variables, the employees can now identify relationships between variables. With a route planner included in the application, the reported journeys are displayed automatically and can be easily checked whether the sequence of journeys or individual deliveries is plausible. To check the correctness of the types of goods reported, the economic activity of a company can be easily verified e.g. via the NACE assignment in the business register or internet research.

The new application allows an automatic or manual imputation of empty journeys which can be done by the person who is responsible for the case. This is important, as empty journeys are often missing, even though the respondents should report all trips during a reference week. As another result of the new application every employee is now linked to specific companies. This improves the communication between Statistics Austria and the respondents.

Beside the presented technical improvements, the reorganization of the workflows had a positive impact on the daily routine of the employees, which became more diverse and demanding, leading to higher motivated employees. Referring to the classical two-factor theory of Herzberg (2010), which deals with motivation and in particular work motivation, some connections shall be described in the following section.

* 1. The 2-factor theory of motivation by Herzberg

The theory of motivation, introduced by Herzberg 1958, differentiates between factors related to the content of work (motivators) and those related to the context of work (hygiene factors). According to Herzberg, satisfaction and dissatisfaction are not polar opposites. While motivators represent a spectrum of "not satisfied - satisfied", the spectrum of hygiene factors applies to "dissatisfied - not dissatisfied".

Typical work motivators are sense of achievement, meaningfulness of work, recognition, independent work or responsibility. Improvements can be achieved through e.g. additional work tasks at the same level (job enlargement) or at a higher level (job enrichment). Hygiene factors include remuneration, leadership skills of supervisors, working conditions, interpersonal relationships with employees and supervisors or job security. Hygiene factors with a positive connotation do not lead to motivation but rather to no being dissatisfied with the work situation. Satisfaction and motivation are only formed when motivators exist.

Relating to this model, it can be seen, that many aspects of the factors described – in particular motivators – arose by reorganizing the workflows due to the new application. Independent work and personal responsibility is positively observed by the employees. By solving and clarifying plausibility errors on their own, the employees experience a sense of achievement as result of their own work performance. Since work steps of different levels have now been combined, job enlargement and job enrichment was noticed.

As the reorganization of the workflows does not directly influence hygiene factors, it is necessary to ensure that dissatisfaction does not arise – through general conditions and appropriate behaviour of the supervisors.

It is evident that the use of IT-based applications for statistical processes is not just a technical adjustment, but can improve work to a new, more sophisticated level. Many activities which could be perceived as boring routine can be facilitated and eliminated and reported questionnaires can be checked in a complete and responsible manner. This increases work performance and motivation and thus enhances the quality and availability of the produced data, the basic requirement for the calculation and publication of flash estimates.

It should not go unmentioned that IT-based applications require constant maintenance. Since the technical environment and standards are constantly changing permanent adaptations are necessary.

1. Statistical analyses before publishing flash estimates

Before publishing flash estimates, it is important to clarify the timing when enough questionnaires are reported and validated in order to calculate results that permit valid and accurate results for the reference period. Furthermore it is essential to ensure that there are no systematic differences between the available data at the time of estimating flash estimates and final results. Statistics Austria decided to fix the release dates of the flash estimates instead of determine the time of publication depending on the degree of data availability; providing the data users with a regular publication of results.

* 1. Time lack of flash estimates and sampling error

Statistics Austria receives the majority of the questionnaires usually immediately after the respective reference week. As the new validation workflow described above proved to be much more time efficient than the old one, it was examined whether it was possible to publish flash estimates with acceptable sampling errors one month after the respective quarter.

Analysing the reference year 2016, the degree of received and validated data of the individual quarters at the time of producing flash estimates were 58% (Q1), 79% (Q2), 71% (Q3) and 77% (Q4). At the time t+5 – when the regular results were estimated – the degree was 96%.

It can be seen, that the validation levels differed between the individual quarters. Among other things, this occurred due to the quality of the individual questionnaires and the number of the employees (e.g. staff away sick or on vacation in winter months or during holiday season).

The different degrees of validated data of the individual quarters affect the magnitude of the sampling errors. Therefore, it was analysed whether at least the flash estimates of the main results (transport volume and transport performance) could be published with a reasonable statistical uncertainty at the planned release time (t+1).

The measure used to assess the quality of the extrapolation was the margin of error at 95 % confidence, which is in line with the precision requirements set by the EU for the European road freight transport survey.

The estimation of the sampling error was done by bootstrapping based on 1 000 simulated resamples. Unlike analytical methods for estimating the sampling error, bootstrapping does not make any assumptions about the distribution of the population, making it more suitable for complex sampling designs.

Table 1 shows the sampling errors of flash estimates and final results. It can be seen, that - except for the first quarter, where the degree of received and validated questionnaires was only 58% - the sampling errors for the flash estimates are within an acceptable range. The flash estimates compared with the final results show only a slight deviation from each other. For the transport volume of the individual quarters of 2016, the differences consisted in: Q1: 1.80%, Q2: -1.71%, Q3: -2.09% and Q4: -0.24%. Figure 1 illustrates this comparison and their sampling errors.

**Table 1. Transport volume and performance 2016 – flash estimates and final results**

|  |
| --- |
| **Transport volume in tonnes** |
| **Period** | **Flash Estimates** | **Final result** | **Difference in %** |
|  | **Tonnes** | **95% CI (relative)** | **Tonnes** | **95% CI (relative)** |  |
| 2016-Q1 | 71585233 | 7.6 | 72873201 | 4.2 | 1.80 |
| 2016-Q2 | 103228986 | 4.2 | 101461595 | 3.9 | -1.71 |
| 2016-Q3 | 111860854 | 5.0 | 109518931 | 4.3 | -2.09 |
| 2016-Q4 | 94056358 | 5.3 | 93833462 | 5.4 | -0.24 |
| **Transport performance in 1000 tonnes-kilometres domestic** |
| **Period** | **Flash Estimates** | **Final result** | **Difference in %** |
|  | **Domestic-Tkm** | **95% CI (relative)** | **Tkm** | **95% CI (relative)** |  |
| 2016-Q1 | 4062471 | 6.2 | 4002077 | 3.5 | -1.49 |
| 2016-Q2 | 4713406 | 3.6 | 4665667 | 3.4 | -1.01 |
| 2016-Q3 | 5086700 | 4.3 | 4956855 | 3.3 | -2.55 |
| 2016-Q4 | 4467127 | 3.7 | 4452743 | 3.6 | -0.32 |

**Figure 1. Comparison sampling errors final results and flash estimates 2016**



Both the sampling errors as well as the absolute differences of the flash estimates were consistent with the final results. Hence, the statistical uncertainty for the analysed period is within a reasonable range and publishing flash estimates of the main results with a time lag of one month is possible.

* 1. Percentage of vehicles in the sampling stratum

Beside the analysis of sampling errors, it must be ensured that the questionnaires received at the time of the estimation of the flash estimates are structurally indistinguishable from those at the time of the final results. Systematic differences – e.g. "Vehicles from Vienna are underrepresented at t+1” – can lead to a bias in the results.

The survey sample for road freight transport statistics is performed as a disproportionate sample stratified by size of company, type of transport, region and weight class/vehicle type. In total the sample population is split into 25 different strata.

The distribution of the vehicles across the strata showed no structural differences between flash estimates and final results. The percentage of reported vehicles regarding the different groups differed only marginal from the four quarters of 2016. Table 2 shows the number of vehicles at the time of sampling, at the time of the flash estimates and the final results for the first quarter of 2016. For all three groups, the distribution of the vehicles in their individual stratum is included. Although the degree of received and verified questionnaires of the first quarter was only 58%, it can be seen that the share of vehicles in each stratum is nearly the same. This was also examined for the other three quarters. Thus, structural biases of the results of the flash estimates can be excluded.

**Table 2. Breakdown vehicles/stratum, final results and flash estimates – 1Q2016**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Stratum** | **Vehicles in stratum** | **Vehicles – flash estimates** | **Vehicles – final results** | **Vehicles in stratum %** | **Vehicles – flash estimates %** | **Vehicles – final results %** |
| 1 | 500 | 290 | 434 | 7,74 | 7,74 | 7,67 |
| 2 | 77 | 45 | 69 | 1,19 | 1,20 | 1,22 |
| 3 | 16 | 10 | 13 | 0,25 | 0,27 | 0,23 |
| 4 | 494 | 257 | 419 | 7,65 | 6,86 | 7,40 |
| 5 | 43 | 20 | 38 | 0,67 | 0,53 | 0,67 |
| 6 | 352 | 171 | 305 | 5,45 | 4,57 | 5,39 |
| 7 | 59 | 39 | 56 | 0,91 | 1,04 | 0,99 |
| 8 | 10 | 7 | 9 | 0,15 | 0,19 | 0,16 |
| 9 | 87 | 51 | 72 | 1,35 | 1,36 | 1,27 |
| 10 | 10 | 5 | 10 | 0,15 | 0,13 | 0,18 |
| 11 | 884 | 521 | 774 | 13,69 | 13,91 | 13,68 |
| 12 | 49 | 26 | 42 | 0,76 | 0,69 | 0,74 |
| 13 | 343 | 190 | 294 | 5,31 | 5,07 | 5,20 |
| 14 | 38 | 23 | 36 | 0,59 | 0,61 | 0,64 |
| 15 | 10 | 3 | 9 | 0,15 | 0,08 | 0,16 |
| 16 | 887 | 496 | 744 | 13,74 | 13,24 | 13,15 |
| 17 | 46 | 24 | 41 | 0,71 | 0,64 | 0,72 |
| 18 | 1155 | 745 | 1062 | 17,89 | 19,89 | 18,77 |
| 19 | 15 | 13 | 15 | 0,23 | 0,35 | 0,27 |
| 20 | 10 | 3 | 7 | 0,15 | 0,08 | 0,12 |
| 21 | 41 | 31 | 37 | 0,64 | 0,83 | 0,65 |
| 22 | 10 | 6 | 8 | 0,15 | 0,16 | 0,14 |
| 23 | 810 | 499 | 722 | 12,55 | 13,32 | 12,76 |
| 24 | 61 | 11 | 56 | 0,94 | 0,29 | 0,99 |
| 25 | 449 | 259 | 387 | 6,95 | 6,92 | 6,84 |
| **Total** | **6456** | **3745** | **5659** | **100,00** | **100,00** | **100,00** |

1. Conclusion

This paper describes the development of flash estimates for road freight transport statistics in Austria. In 2014, a new IT-based application was implemented in the statistical process. This application had a mayor impact both on the validation process and the work of the employees. Following the theory of motivation by Herzberg (2010) it was examined how job satisfaction and motivation of employees improved and data availability and quality enhanced. An early availability of high quality data is essential for developing flash estimates.

On the basis of the road freight transport statistics an early publication of estimates for the main results transport volume und transport performance was aimed. Analyses regarding sampling error and distribution within the sampling strata were done and lead to satisfying results. At the time of estimating flash estimates a sufficient large number of validated and uniformly distributed questionnaires is available. This allows the estimation and publication of the main results already one month after the reference period – compared to five months for regular estimates.

**5. References**

Council of the European Union (2006). “Regulation (EC) No 1893/2006 of the European Parliament and of the Council of 20 December 2006 Establishing the Statistical Classifi- cation of Economic Activities Nace Revision 2 and Amending Council Regulation (EEC) No 3037/90 As Well As Certain EC Regulations on Specific Statistical Domains.” URL <http://data.europa.eu/eli/reg/2006/1893/oj>.

Council of the European Union (2012). “Regulation (EU) No 70/2012 of the European Parliament and of the Council of 18 January 2012 on Statistical Returns in Respect of the Carriage of Goods by Road.” URL <http://data.europa.eu/eli/reg/2012/70/oj>.

European Commission (2003a). “Commission Regulation (EC) No 6/2003 of 30 December 2002 Concerning the Dissemination of Statistics on the Carriage of Goods by Road.” URL <http://data.europa.eu/eli/reg/2003/6/oj>.

European Commission (2003b). “Commission Regulation (EC) No 642/2004 of 6 April 2004 on Precision Requirements for Data Collected in Accordance With Council Regulation (EC) No 1172/98 on Statistical Returns in Respect of the Carriage of Goods by Road.” URL <http://data.europa.eu/eli/reg/2004/642/oj>.

Herzberg, F., Mausner, B. & Bloch Snyderman, B. (2010). The motivation to work, New Brunswick (U.S.A.) and London (U.K.): Transaction published (twelfth printing).