BURDEN ANALYSIS FOR SHORT-TERM BUSINESS SURVEYS RESPONDENTS[1]

Silvia Binci, Katia Bontempi, Claudio Ceccarelli, Francesca Monetti, Francesca Rossetti

Abstract. Short-term business statistics are produced by the Italian National Institute of Statistics (ISTAT) and governed by specific European regulations which outline the structure of questionnaires and surveys, the scope of observation, the reference population, the sampling methodology and the precision of the estimates.
The Italian productive sector, characterized by a fragmentation of enterprises in terms of size, along with the sampling strategies mandated by European regulations, results in the participation of specific business types, notably medium and large enterprises, in most of the surveys conducted by ISTAT.
This work analyses the burden on respondents and their behaviour over time by assessing the profiles of respondents and non-respondents. The objective, using survival analysis methods, is to identify the factors contributing to the permanent exit of respondents from surveys, and particularly to measure the extent of this withdrawal. This involves analysing the duration of cooperation between respondent units and ISTAT.
The data analysed come from the *The Business Statistical Portal*, i.e. the system that collectively manages all short-term business surveys. The information used in the study covers all the survey units involved in short term business surveys in the period 2016 to 2023.

1. Introduction

Short-term business statistics (STS) are subject to specific European regulations that outline the structure of questionnaires and surveys, the scope of observation, the reference population, the sampling methodology and the precision of the estimates.

In the context of enterprises, particularly within the Italian production sector characterized by the prevalence of small businesses, regulatory criteria and constraints necessitate the consistent inclusion of large enterprises in samples across multiple surveys over time.

The aim of this work is to analyse the burden on respondents and their behaviour over time by assessing the profiles of respondents and non-respondents.

[1] The article was only possible thanks to the joint work of the authors. In particular, Claudio Ceccarelli wrote section 1 and 6, Silvia Binci wrote section 2, Francesca Monetti wrote section 3, Francesca Rossetti wrote section 4 and Katia Bontempi wrote sections 5.

The information used in the study, appropriately normalised and standardised, covers all the survey units involved in short term business surveys in the period 2016 to 2023 and is organised in a single database.

Our purpose, using survival analysis methods, is to identify the factors contributing to the permanent exit of respondents from surveys and to analyse the extent and the duration of cooperation between respondent units and Italian National Institute of Statistics (ISTAT).

Section 2 gives a description of the purpose of "The Business Statistical Portal" and outlines the information collected. Section 3 describes the characteristics of the short-term business surveys on which the survival analysis is applied. Section 4 describes the method on which the work is based. Section 5 contains the main results and Section 6 presents the main conclusions.

2. 'The Business Statistical Portal': the system that manages all short-term business surveys collectively

In October 2010, following a Memorandum of Understanding signed in May 2010 between ISTAT, the National Ministry for Public Administration and Innovation and the Italian Union of Chambers of Commerce (Unioncamere), ISTAT started a project for the development of a platform dedicated to the acquisition of statistical information from enterprises. It resulted in the implementation of a comprehensive innovative tool, aimed at the integration and standardization of data collection processes both internally and externally (companies that provide statistical information): the Business Statistical Portal (BSP). It is an attempt to streamline the organization and management of business surveys as a whole in order to reduce the respondent burden, increase both ordinary and extraordinary communications on survey events and activities, and standardize and harmonize data collection procedures to increase process efficiency (Fazio et al. 2013).

The BSP is composed of a front-end and a back-office. The front-end is "enterprise-centred" with its main features being, for example, a single sign-on and single point of access to all the questionnaires with unique credentials; a delegation system to manage in a flexible way the assignment of the questionnaires to the appropriate contacts; a specific section dedicated to news about the conduct of the surveys (start, closure, extraordinary events, extensions and technical problems); and a status of obligations, a complete and updated framework of all the statistical activities required by each company and of the status of fulfilment of each questionnaire. The back-office supports all the users participating in the survey network: survey managers, interviewers, data analysts, representatives of administrative registers, supervisors, data reviewers and help desk operators. It

enables an integrated, harmonized and centralized management of all survey stages for all areas of production, allowing a shared access to information and data, acting on a role-based user profile system (Bellini et al. 2019).

Referring to the completion status of the questionnaires, data are stored in an Oracle table. The table contains a record for each responding unit, for each survey and for each reference period in which it is involved; information on the fulfilment status or any reasons for exclusion, and the relative date of completion or exclusion, are saved in real time. In particular, for short-term statistics, the response date registered is that corresponding to the first data transmission, so as to prevent the enterprise from incurring sanctions in the event that it overwrites updates after the deadlines established in the information letter. Storing data in an Oracle table allows for great flexibility in analysing the performance of statistical surveys and the behaviour of each responding unit.

The portal entered a testing phase in June 2013, involving a purposive sample of about 70 enterprises across a subset of representative surveys of economic statistics. It became fully operational in early 2014 and then progressively hosted nearly all ISTAT economic surveys and their related survey units.

Table 1 – *Surveys and Authorized Users of the Business Statistical Portal - May 2024.*

Surveys included in the Portal	Short-term surveys included in the Portal	Number of companies authorized to access	Number of registered NSI external users	Number of registered NSI internal users
117	55	1,032,910	1,125,363	765

Elaboration of data from "Business Statistical Portal"

3. Short-term business statistics overview

Short-term business statistics are a set of indicators, with a monthly or quarterly frequency, used for tracking closely the business cycle of an economy. STS indicators are important tools for the formulating and monitoring of economic and monetary policies. The legal authority for these indicators is Regulation EC No 2152/20219 amending Regulation EC No 1158/2005 and further implementing and amending all appropriate regulations; they are also included in the National Statistical Program that contains, in a specific list, instances of non-compliance with the obligation to provide the required data, to which the law equates knowingly incorrect or incomplete data, which are punished with a financial administrative penalty. Administrative penalties are not provided for all the units involved in

statistical surveys, but only for a subset. For the majority of short-term surveys, the threshold for penalties is set at a staff of 100 employees or a turnover of 500 million euros (the turnover criterion has been introduced from 2022). The introduction of Centralised Data Collection has required a complete revision of the procedures adopted in order to standardize and generalize all the activities performed. In the following paragraphs, the main innovative issues regarding the STS are presented (Bellini et al. 2019).

In 2018 a new penalty system was introduced: the new criteria concern the organization of penalties on an annual basis, the provision of the administrative penalties in the event of non-compliance within a single period (month or quarter) and the possibility of delayed delivery, with two types of flexibility: the number of days of tolerance with respect to each monthly or quarterly deadline, varying from survey to survey, and the annual cumulative delay tolerated, lower than the sum of the delays tolerated in the single periods. For legal reasons, the amount of the penalty applied is fixed and independent of the number of periods of the year omitted (months or quarters). Only since 2020 has the new penalty scheme been fully operational.

The information letter is the official document that ISTAT uses to communicate to enterprises their involvement in the surveys; it also illustrates the objectives and the regulations in force at the time of each survey and the method of communication of any penalties. Since 2018, the layout of the information letter has been revised, paying particular attention to the deadlines for the provision of data. Strict adherence to deadlines is an essential prerequisite for the data processing, subsequent transmission to Eurostat and national data dissemination to take place by the due dates.

The introduction of the new sanctioning system has led to a strict scheduling of formal and informal communications by certified email (PEC) and ordinary email, jointly with telephone reminders addressed to the units involved in the surveys. The deadlines indicated in the information letter involve the definition of a strict timetable in order to remind the respondents about the required collaboration. The texts of the communications have also been revised recently in order make them more effective with respect to the units.

The following standards have been adopted: i) a single centralized access point both for the data capturing systems (Business Statistical Portal see par. 2) and for the incoming contacts (free of charge inbound Contact Centre - CC) by telephone or by email; ii) centralized outbound CC services; and iii) a system of harmonized standard answers to provide timely assistance to enterprises on non-thematic and recurring thematic requests. An analysis of the response rates of selected short-term surveys in the period 2016-2023 highlighted that the introduction of the current penalty rules,

jointly with the other main innovative fields, had an average impact of 29 percentage points (pp), starting from 59 in 2016.

Table 2 – *Selected short-term business statistics: main characteristics.*

Survey	PI	Observation field	Sampling design	Survey unit
Employment in large enterprises (OCC)	M	Enterprises with at least 500 employees	Census survey for enterprises with at least 500 employees	Functional unit
Retail trade (DETT)	M	Enterprises with main economic activity in sec G of the Nace Rev. 2 classification	Stratified random sampling for enterprises with less than 50 employees - Census survey for enterprises with at least 50 employees	Enterprises
Service turnover (FAS)	M	Enterprises with main economic activity in secs G, H, I, J, L, M and N of the Nace Rev. 2 classification	Stratified random sampling in sectors featuring a substantial number of enterprises - Random sampling for other sectors	Unit of Economic Activity
Industrial producer prices (PPI)	M	Products included in Sections from B to E36 of the CPA classification - Enterprises with main economic activity in secs B, C, D and E36 of the Nace Rev. 2 classification	Sample annually updated- Reasoned sample and cut-off criterion	Enterprises
Industrial production (IPI)	M	Enterprises with main economic activity in secs B, C and D of the Nace Rev. 2 classification	Reasoned sample – Cut-off criterion	Local unit
Industry turnover (FATT)	M	Enterprises with main economic activity in secs B and C of the Nace Rev. 2 classification	Cut-off criterion for enterprises with less than 20 employees - Reasoned sample for enterprises with at least 20 employees	Unit of Economic Activity
Import prices (PREIMP)	M	Products listed in sections B, C and D of the CPA classification -Enterprises with main economic activity in secs B, C, D, E and G of the Nace Rev. 2 classification	Sample annually updated- Reasoned sample and cut-off criterion	Enterprises
Job vacancies and hours worked (VELA)	Q	Enterprises and private institutions with employees operating in secs B-S, excluding O of the Nace Rev. 2 classification	Stratified and random sampling for enterprises with less than 500 employees	Enterprises and private institutions with employees

Istat

4. Data and indicators in the analysis: survival analysis methods

The standardization of the data collection processes and tools represented by the BSP, along with the uniform methodological rules governing the STS, allows for the observation of enterprise response behaviour across all surveys as a whole and over the entire period under consideration. In this context, the use of survival curves is intended to analyse the burden on respondents and their behaviour over time by analysing the profiles of respondents and non-respondents.

Kaplan-Meier survival curves are a statistical analysis tool used to plot the relationship between the probability of survival and the time observed and to measure the risk of a particular event occurring (Cox et al. 2018). It is crucial for survival analysis to set the 'event' variable to indicate whether an event has occurred and the 'survival' or 'time' variable to measure the duration of observation.

The concept behind Kaplan-Meier is conditional probability, i.e. the probability of surviving to a given point in time is conditional on the probability of being alive in earlier periods. Letting *S(t)* be the survival function at time *t*, the Kaplan-Meier estimate of the survival function at time t_i is the cumulative product

$$\hat{S}(t_i) = P(T > t_i) = \prod_{j=1}^{i}(1 - \frac{d_j}{Y_j}) \quad (1)$$

where $t_1 < t_2 < \ldots < t_D$ represents the distinct event times, for each $i=1,\ldots,D$, Yi is the number of surviving units (the size of the risk set) just prior to t_i and d_i is the number of units that fail at t_i.

The method distinguishes between events and censoring: an event indicates that the unit experienced the event being studied during the observation period, while censoring refers to units that leave the study for reasons other than the event.

The versatility of survival analysis allows it to be used not only in medical research but also in industrial durability testing and in fields ranging from physics to econometrics. By examining how different variables influence survival outcomes, researchers can better understand the effectiveness of treatments and identify important prognostic factors. An important aspect of survival data analysis is the comparison of survival curves. It is crucial to determine whether the underlying population groups have identical survival functions. Several tests can be used for this purpose, including the log-rank test, the Wilcoxon test, the Tarone-Ware test and the Peto-Peto test. In this application, the log-rank test was used.

In this study, the reference population consists of the units participating in short-term business surveys through the BSP. Each unit was thoroughly observed from 2016 to 2023, covering all STS and related survey occasions as a cohesive lifecycle.

In order to allow the contextual use of data from several surveys with different reference units (e.g. enterprises, local units, functional units, etc.), it was decided to use the enterprise as the statistical unit for all the surveys. For this purpose, in surveys where the reference units are parts of the enterprise, the enterprise is considered to have fallen if no reference unit responded on the occasion of the survey. The time is represented by the number of survey occasions in which each enterprise participated. Time T1 is therefore defined as the first opportunity for each enterprise to be interviewed: it is not necessarily the same for all units. We considered two possible events: 'first fall', defined as the first survey occasion when the unit does not respond, and "definitive fall', defined as the survey occasion when the enterprise stops cooperating. The censored units are those enterprises that are no longer included in any sample of a short-term survey and are therefore no longer required to cooperate with ISTAT.

The analysis included 163,097 enterprises. Table 3 shows the extent to which the enterprises considered were involved in STS during the observation period in terms of the number of surveys in which they were asked to participate. The number of times an enterprise is asked to respond depends not only on its duration in the samples over time but also on the type of survey in which it is involved, whether monthly (12 times a year) or quarterly (4 times a year). During the observation period, the highest number of survey occasions on which the enterprises were asked to participate was 530. It can be observed that a small number of enterprises were involved in more than five surveys, probably due to the sampling design used for short-term business surveys. These enterprises are typically large (more than 250 employees).

Table 3 – *Distribution of enterprises registered in the Business Statistical Portal by number of short-term surveys in the period 2016-2023.*

N. of short term surveys (absolute value)	N. of enterprises (absolute value)	Average number of interview occasions each company participated in (absolute value)	Average response rate (percentage value)	Percentage of enterprises with at least one fall event (percentage value)	Percentage of enterprises with a definitive fall event (percentage value)
1	127,529	15.1	62.8	48.8	40.3
2	25,065	49.3	86.3	41.8	16.8
3	6,785	100.0	92.0	50.1	9.9
4	1,940	187.5	95.1	57.7	5.5
5	981	271.8	96.5	63.3	2.3
6	586	351.8	97.8	67.9	1.5
7	198	432.3	98.3	67.7	2.0
8	12	392.5	98.6	66.7	-
9	1	184.0	98.9	100.0	-

Elaboration of data from "Business Statistical Portal"

It can also be observed that cooperation increases with the number of surveys in which the enterprises participate: those participating in only one survey have an average response rate of 62.8%, those participating in two surveys have an average response rate of 86.3% and those participating in three or more surveys have an average response rate of over 92%.

5. Results

Figure 1 shows the Kaplan-Meier survival curve for the first fall event: the steep decline corresponding to the first survey occasion is evident. In fact, the probability of survival at time T1 is about 70%, while the probability of maintaining the collaboration until the end of the observation period is about 18%. These results suggest that a large proportion of enterprises interrupt their cooperation by not responding to one of the surveys in which they are involved at least once.

Figure 1 – *Kaplan-Meier survival curve for the first fall event.*

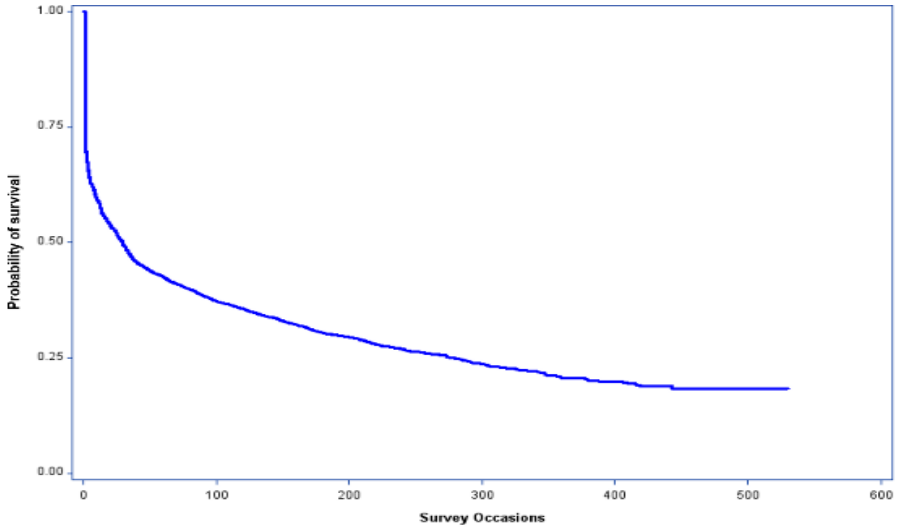


Figure 2 shows the Kaplan-Meier survival curve for the definitive fall event: again, a sharp drop at time T1 can be observed, but in this case the probability of survival at time T1 is about 77%. However, the most important finding is that the probability of continuing to cooperate until the end of the period is about 50%. This leads us to conclude that the fact that an enterprise does not continuously participate in the surveys does not necessarily imply a decision to stop cooperating in the long

run. For this reason, we focused the survival analysis on the definitive fall event and examined some factors that could influence the decision of enterprises to stop participating in ISTAT surveys. We used the Kaplan-Meier survival curve method stratifying by the selected factors; the comparison of the curves was consistently supported by a significant log-rank test ($p < 0.05$).

Figure 2 – *Kaplan-Meier survival curve for the definitive fall event.*

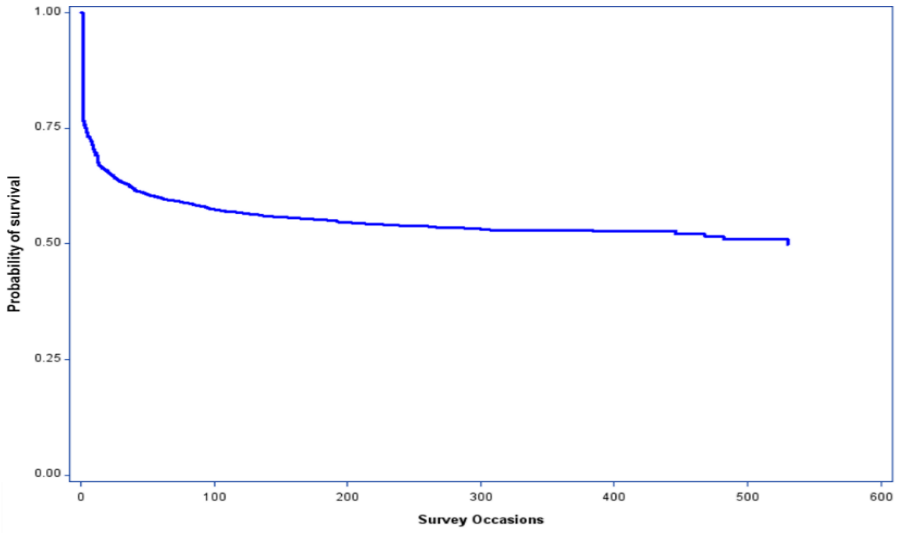


Figure 3 shows the Kaplan-Meier survival curves for the definitive fall event by employee class: it is noteworthy that enterprises in the 0-2 employee class have a much lower probability of survival at time T1 than the other categories, around 65%. Furthermore, although these enterprises were asked to participate in fewer survey occasions than the others (at most 317 occasions), they are much less likely to continue to cooperate until the end of the observation period (only about 12%). On the other hand, enterprises in the 100-249 and 250+ employee classes have a higher probability of survival at time T1 (97% and 99% respectively) and a similar probability of survival at the end of the observation period (around 86%).

Figure 3 – *Kaplan-Meier survival curves for the definitive fall event by employee class.*

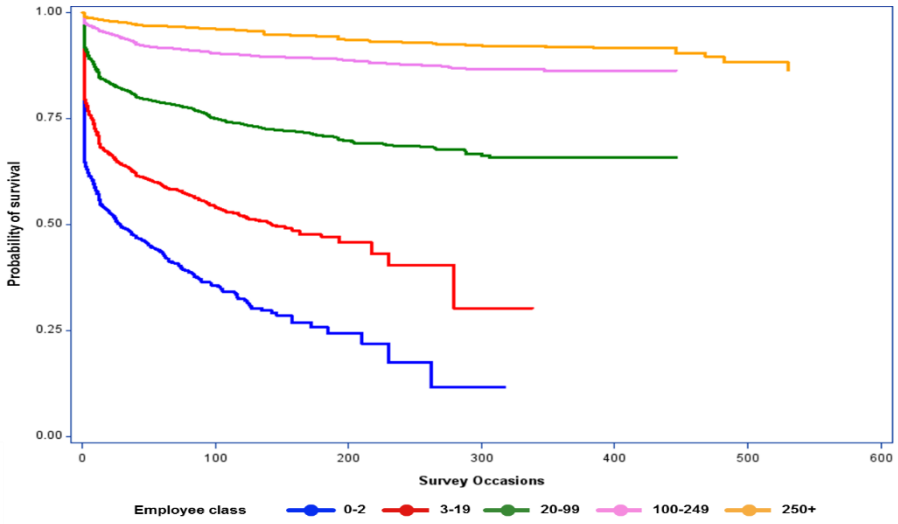


Figure 4 – *Kaplan-Meier survival curves for the definitive fall event by geographical area*.

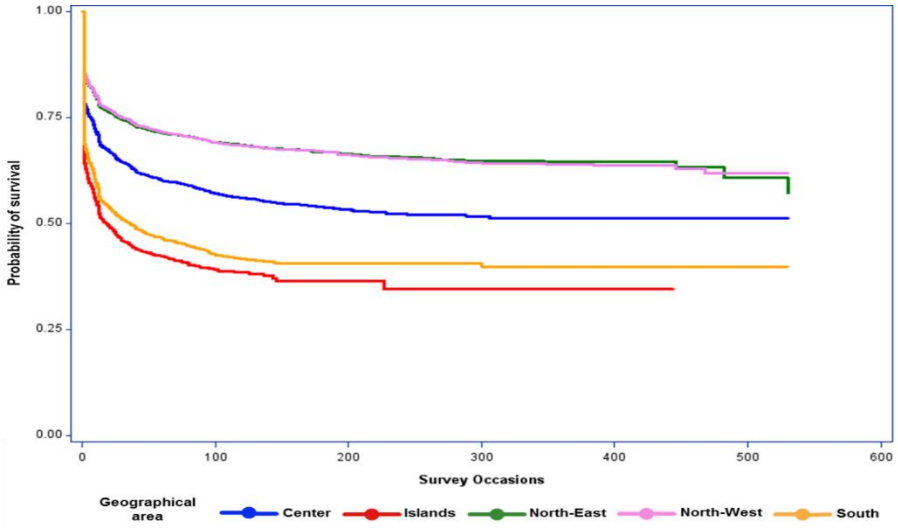
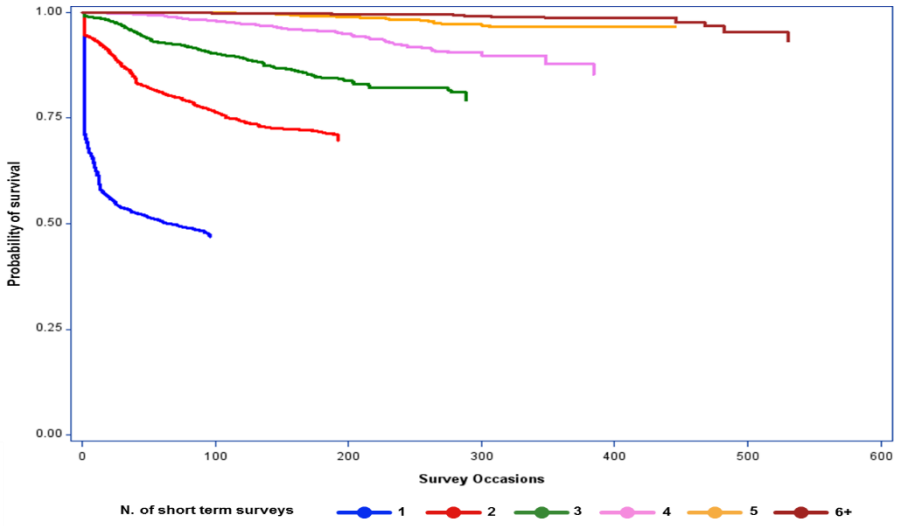


Figure 4 shows the Kaplan-Meier survival curves for the definitive fall event by geographical area. Enterprises located in the North-West and North-East areas have a very similar response behaviour: in fact, their curves partially overlap, although

the North-East enterprises have a lower probability of survival at the end of the observation period compared to the North-West enterprises (57% and 62% respectively). Enterprises located in the South of Italy, especially on the Islands, have the worst performance curves, while those in the Center area remain in an intermediate position, with a survival of more than 50% at the end of the observation period.

Figure 5 shows the Kaplan-Meier survival curves for the definitive fall event by number of short-term surveys in which the enterprises were asked to participate. As shown in Table 1, the willingness to cooperate seems to increase with the number of surveys in which a company is involved. In particular, enterprises involved in only one survey have a probability of survival at time T1 of around 71%, compared to around 95% for enterprises involved in two surveys. Furthermore, for enterprises involved in 6 or more surveys, the first definitive fall events occur at the 97th survey occasion.

Figure 5 – *Kaplan-Meier survival curves for the definitive fall event by number of short-term surveys in which the enterprises were asked to participate*.



6. Conclusion

The lack of cooperation observed during the initial survey occasions should prompt us to consider enhancing our communication methods and tools to improve company participation, with a particular focus on small and medium-sized enterprises.

The medium to long-term decline in the number of enterprises initially willing to collaborate should prompt us to reconsider the statistical burden on businesses.

This work, in addition to the specific improvements that we can make in relation to each individual survey, gives us the opportunity to provide some general considerations that are valid for all surveys.

Undoubtedly, the dimensions of the phenomenon are manifold and work must certainly be carried out in relation to communication on: the reduction of the statistical burden on respondents, simplifying both the specific questionnaire and general questionnaires; the integration of the surveys with other statistical sources; and, no less important, the return of information that can be useful to the respondents.

References

FAZIO N.R., MURGIA .M, NUNNARI A. 2013. The Business statistical Portal: a new way of organizing and managing data collection processes for business surveys in Istat. *Working Papers 14*, Seminar on Statistical Data Collection. Unece - Conference of European statisticians, 25-27 September 2013, Geneva, Switzerland.

BELLINI G., BINCI S., DE GAETANO L., MONETTI F., PAPA P. 2019. Effects of Istat CDC (Centralised Data Collection) approach on the reduction of the Total Survey Error: experiences in data collection implementation field. In *Workshop on Statistical Data Collection 'New Sources and New Technologies'*. Unece - Conference of European statisticians, 14-16 October 2019, Geneva, Switzerland.

COX, DAVID ROXBEE 2018. *Analysis of survival data*. Chapman and Hall/CRC.

Silvia BINCI, Italian National Institute of Statistics, binci@istat.it
Katia BONTEMPI, Italian National Institute of Statistics, katia.bontempi@istat.it
Claudio CECCARELLI, Italian National Institute of Statistics, clceccar@istat.it
Francesca MONETTI, Italian National Institute of Statistics, francesca.monetti@istat.it
Francesca ROSSETTI, Italian National Institute of Statistics, frrosset@istat.it