

25th Meeting of the Wiesbaden Group on Business Registers
- International Roundtable on Business Survey Frames

Tokyo, 8 – 11 November 2016

*Arturo Blancas,
Susana Pérez,
Hugo Hernández
INEGI
Session No. 6*

Business Demography and Data Products from the Business Registers

***Punctual monitoring of establishments in the period 2009-2014 and its comparison against results
from a study on the demography of establishments***

1 Introduction.

Since 1930, México has conducted Economic Censuses every five years; this has allowed INEGI to provide information with diverse geographic and economic disaggregation levels, up to the maximum level permitted by the principle of confidentiality, which must be respected by law. Results from Economic Censuses support decision-making in all government levels and offer various products to the business sector, researchers, academia and to the public in general.

In 2010 INEGI built its Statistical Business Register (SBR) as the only updating source for its various statistical projects. Right now, INEGI's SBR serves as backbone for the extraction of frozen frames and directories and regular maintenance of institutional products, including the National Statistical Directory of Economic Units (DENUE), which has the advantage of possessing a statistical and geographic infrastructure that allows displaying establishments on maps together with the location where their economic activity takes place.

In 2014, during the Economic Census' data collection, a punctual follow-up was given to every establishment through the Statistical Business Code (CLEE), detecting for each surviving establishment changes presented in the period between 2009 and 2014; furthermore it was possible to know the deaths and births of establishments in the same period on an individual basis; thus demographic events were recorded in the SBR.

On the other hand, INEGI developed a model on the demography of establishments in Mexico, producing survival tables built by monitoring 16 generations with information from the 1989-2009 Economic Censuses.

Now, with the purpose of evaluating the estimations derived from the model for survivals and deaths in 2014, an exercise was performed using the survival tables for the establishments recorded in the SBR (those collected by the 2009 Economic Censuses). The results of this exercise were compared against the number of survivals and deaths, which are individually recorded in the SBR. This paper presents the results from this comparison.

The paper is comprised of five sections, including this Introduction. Section two describes the follow-up performed in 2014 on the establishments that were collected by the 2009 Economic Censuses. Section three includes a brief description of the survival model developed by INEGI for estimating the number of surviving establishments in the years following 2009, and displays results produced when applying the model to the 2009 Economic Censuses data. Section four includes a comparison between

recorded demographic data for every establishment in the SBR and those obtained through the model. Section five presents conclusions and future work.

2 Follow-up on every establishment in the period 2009-2014.

2.1 The SBR of Mexico as the basis for longitudinal and demographic studies.

The first SBR version was comprised of results from the 2009 Economic Censuses, assigning to each establishment a unique identifier, called Statistical Business Code (CLEE), through which follow-up is given to each and every one of the establishments as long as they remain active, thus opening up the possibility of keeping the same identifier throughout time and by all institutional projects.

In this way, for the 2014 census exercise, the initial directory was extracted from the SBR and it was stored into mobile computing devices together with the associated digital cartography and satellite images and all establishments were located on digital maps through geographical coordinates corresponding to the places in which they developed their economic activity. All this was possible thanks to advancements in Information and Communication Technology and the staff's experience gained through past Economic Censuses exercises.

During the 2014 on-field operation, during the exhaustive on-site visit to all the urban areas in the country, punctual monitoring was given to each establishment through the CLEE, enabling precise recording of demographic events, such as updating data on survivors and detecting establishments that definitely ceased operations (deaths), as well as new establishments (births).

SBR's update was based upon the results from this 2014 census exercise; this helped increasing SBR's statistical potential. Produced results serve two important purposes:

- a) To serve as the basis for conducting longitudinal studies for period 2009-2014, since it is possible to link information captured on surviving establishments via CLEE during the 2009 and 2014 Economic Censuses.
- b) To know, on an individual basis, survivals, births and deaths of establishments for period 2009-2014.

Results related to these two purposes are analyzed in this paper.

2.2 Identification of on-site demographic events for each establishment.

The following events were defined taking into account SBR recorded updates for period 2009-2014:

Survivors: Existing establishments detected in 2009 that were collected again during the 2014 Economic Censuses, applying the continuity rules recommended by Eurostat- OECD.

Births: Establishments that initiated activities after 2009 or were declared as new establishments, according to the continuity rules.

Deaths: Establishments that were active in 2009 but subsequently ceased operations definitely; according to the continuity rules, it can also be determined that an establishment collected in 2009 is no longer the same entity, giving rise to a new establishment.

Continuity Rules: It is considered that a new establishment is born from an existing one when at least two out of its three following characteristics have changed:

- Company name
- Address
- Economic activity

2.3 Results from demographic events recorded in the SBR.

The results obtained from counting survivals, births and deaths of establishments one by one in the 2009-2014 period are shown in Table 1.

Table 1. Recount of survivors, births and deaths for period 2009-2014.

Demographic event	SBR recount
Survivors	2'461,851
Births	2'437,620
Deaths	1'796,420

These results were used to compare them with the estimations from the survival model in order to evaluate the precision of the model, as will be seen below in this paper.

3 Estimating the number of surviving establishments from the survival model.

In 2014 INEGI undertook a demographic study on life expectancy of businesses in Mexico¹, which produced several survival tables² built through the monitoring of 16 generations of establishments.

This study provides, among other things, survival probabilities for establishments in Mexico through models based upon demographic and linear regression concepts; such models were built from data collected in the Economic Censuses conducted by INEGI in the years 1989, 1994, 1999, 2004 and 2009. In order to develop the exercise shown below, the survival probabilities were extrapolated towards 2014.

In this exercise the survival probabilities from the demographic model were interpreted as the proportion of establishments which survive $x+1$ years after their birth year; this proportion is denoted as $p(x)$; in other words, $p(x)$ is the proportion of establishments from the same generation that make it to $x+1$ years.

Survival table A1, found in the annex of this paper, shows proportions $p(x)$ for all possible values of x . The survival model predicts that after an establishment reaches 20 years of age, the probability of it dying is practically zero.

In order to estimate the proportion of establishments collected during the 2009 Economic Censuses that survived up to the year 2014, the recounts of establishments collected by those censuses, and

¹ http://internet.contenidos.inegi.org.mx/contenidos/productos//prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825068431.pdf (methodological paper in Spanish).

² <http://www.inegi.org.mx/inegi/contenidos/investigacion/Experimentales/Esperanza/doc/anexos.pdf>; see table "Total population of establishments", annex 4 page 115.

broken down according to their birth year, served as a starting point. Table A2 (see annex) shows such recounts.

In order to do the actual calculations for estimating the number of surviving establishments in the years following 2009, the procedure schematized in table 2 was used.

Table 2. Procedure for estimating the number of surviving establishments for the years following 2009.

Recount of establishments in 2009	Estimated number of survivals				
	2010	2011	2012	2013	2014
N_{2009}	$[p(0) * N_{2009}]$	$[p(1) * N_{2009}]$	$[p(2) * N_{2009}]$	$[p(3) * N_{2009}]$	$[p(4) * N_{2009}]$
$\left[p(0) * \frac{N_{2008}}{p(0)} \right]$	$\left[p(1) * \frac{N_{2008}}{p(0)} \right]$	$\left[p(2) * \frac{N_{2008}}{p(0)} \right]$	$\left[p(3) * \frac{N_{2008}}{p(0)} \right]$	$\left[p(4) * \frac{N_{2008}}{p(0)} \right]$	$\left[p(5) * \frac{N_{2008}}{p(0)} \right]$
$\left[p(1) * \frac{N_{2007}}{p(1)} \right]$	$\left[p(2) * \frac{N_{2007}}{p(1)} \right]$	$\left[p(3) * \frac{N_{2007}}{p(1)} \right]$	$\left[p(4) * \frac{N_{2007}}{p(1)} \right]$	$\left[p(5) * \frac{N_{2007}}{p(1)} \right]$	$\left[p(6) * \frac{N_{2007}}{p(1)} \right]$
$\left[p(2) * \frac{N_{2006}}{p(2)} \right]$	$\left[p(3) * \frac{N_{2006}}{p(2)} \right]$	$\left[p(4) * \frac{N_{2006}}{p(2)} \right]$	$\left[p(5) * \frac{N_{2006}}{p(2)} \right]$	$\left[p(6) * \frac{N_{2006}}{p(2)} \right]$	$\left[p(7) * \frac{N_{2006}}{p(2)} \right]$
.
.
.

In table 2, N_{YYYY} represents the number of establishments collected during the 2009 Economic Censuses, which initiated operations in the year $YYYY$ (these values are shown in table A2); $p(x)$ is a survival proportion (see table A1) and operator $[]$ indicates that decimals are truncated off from the result obtained after computing the specified operation inside the brackets. Table 2 shows only the formulas used in the first four rows; for computing successive rows, the logical indicated sequence is followed. Thereby, N_{2009} refers to the number of establishments detected in 2009 and born in that same year; N_{2008} is the number of establishments born in 2008, which survived up to 2009, so $\frac{N_{2008}}{p(0)}$ is the original number of establishments that were born in 2008, according to the survival model; similarly, $\frac{N_{2007}}{p(1)}$ is the original number of establishments from generation 2007 according to the survival model, from which N_{2007} survived up to 2009, and so on.

By combining the formulas in Table 2 with the data from tables A1 and A2, estimates for the number of surviving establishments are obtained for every year in the 2010-2014 period; these estimates are shown in Table 3.

Table 3. Estimations for the number of surviving establishments, according to the indicated operations in table 2.

Birth year	Recount of establishments in 2009	Estimated number of survivals				
		2010	2011	2012	2013	2014
2009	343,247	219,337	124,373	116,681	109,465	102,695
2008	791,103	448,588	420,845	394,818	370,400	347,492
2007	393,089	368,778	345,971	324,574	304,500	285,668
2006	294,040	275,854	258,794	242,789	227,773	213,687
.
.
.
Total	4'291,268	3'666,034	3'398,272	3,233,780	3081378	2'940,850

From the totals at the bottom of table 3, we can see that out of the 4'291,268 establishments accounted during the 2009 Economic Censuses, 2'940,850 would survive up to 2014, according to the survival model.

Based on the estimated number of survivors, it is also possible to estimate the number of deaths, since:

$$\text{Number of deaths} = \text{number of establishments in 2009} - \text{number of survivors.}$$

Table 4 summarizes the estimations obtained through the survival model for the number of establishments that were captured in 2009 and survived up to 2014, as well as the number of estimated deaths for this period.

Table 4. Estimations for the number of establishments that survived and died according to the survival model, considering exclusively the birth year (without applying the continuity rules).

Number of establishments		
In 2009	Survivals in 2014	Deaths up to 2014
	(according to the model)	
4'291,268	2'940,850	1'350,418

However, since the model does not consider the continuity rules (see Section 2.2), calculations (shown in Table 5) were performed for the sake of comparison against the SBR, in which the

continuity rules was indeed applied. For the case of the surviving establishments, the number of deaths obtained through the continuity rules was subtracted from the estimation of the model, and for the case of deaths, that same number was added up to the corresponding estimation of the model.

Table 5. Estimated number of surviving and dead establishments after including the results obtained by the application of the continuity rules.

Estimated number of survivors (S)	Number of dead establishments according to the continuity rule (DCR)	Estimated number of survivors after applying continuity rules results (S – DCR)
2'940,850	373,736	2'567,114
Estimated number of deaths (D)		Estimated number of deaths after applying continuity rules results (D + DCR)
1'350,418		1'724,154

4 Comparisons between real data and estimations.

Table 6 shows a comparison between the number of survivals and deaths recorded in the SBR, and the corresponding estimations shown in table 5.

Table 6. Comparative table between aggregated demographic events obtained from the SBR for each establishment and estimations computed through the use of the survival tables (period 2009-2014).

Demographic event	SBR	Estimated number after applying the continuity rules	Absolute difference	Absolute relative difference (%)
Survivals	2'461,851	2'567,114	105,263	4.3%
Deaths	1'796,420	1'724,154	72,266	4.0%

5 Conclusions and future work.

The survival model produces similar results when compared against aggregated demographic events registered in the SBR; thus, it is important to continue updating data on every establishment across the diverse institutional projects conducted by INEGI, including of course the Economic Censuses, since it is through them that the whole SBR universe is updated and thus the demographic events' records are maintained.

Inasmuch as the SBR is consolidated, it will facilitate the development of studies that comprehensively model the different demographic events on establishments, in an ever more accurate way.

Annex

Table A1. Survival proportions, where $p(x)$ is the proportion of establishments, coming from the same generation, which survive to the age of $x+1$ years.

x	$p(x)$	x	$p(x)$	x	$p(x)$	x	$p(x)$
0	0.63900701	6	0.26332547	12	0.17953098	18	0.12240127
1	0.36234375	7	0.24704001	13	0.16842781	19	0.11483131
2	0.33993446	8	0.23176172	14	0.15801133	≥ 20	0.10772953
3	0.31891109	9	0.21742833	15	0.14823906		
4	0.29918791	10	0.20398139	16	0.13907116		
5	0.28068452	11	0.19136609	17	0.13047025		

Table A2. Recount of establishments observed in the 2009 Economic Censuses, according to their birth year.

YEAR	COUNT	YEAR	COUNT	YEAR	COUNT	YEAR	COUNT	YEAR	COUNT	YEAR	COUNT
2009	343247	1989	67593	1969	10615	1949	1946	1929	316	1909	211
2008	791103	1988	31768	1968	8245	1948	1395	1928	199	1908	125
2007	393089	1987	24287	1967	3495	1947	902	1927	168	1907	74
2006	294040	1986	25813	1966	3201	1946	825	1926	166	1906	75
2005	252528	1985	45523	1965	6938	1945	1687	1925	245	1905	95
2004	221834	1984	28144	1964	5306	1944	601	1924	146	1904	79
2003	141738	1983	18472	1963	4543	1943	531	1923	140	1903	75
2002	128956	1982	20280	1962	3650	1942	687	1922	116	1902	76
2001	108786	1981	13948	1961	2092	1941	384	1921	97	1901	82
2000	183781	1980	47882	1960	11334	1940	2480	1920	581	1900 and earlier	88630
1999	162076	1979	33732	1959	5238	1939	790	1919	90		
1998	94306	1978	14670	1958	3589	1938	647	1918	71		
1997	67129	1977	7808	1957	4311	1937	455	1917	118		
1996	56086	1976	8733	1956	1899	1936	485	1916	49		
1995	83193	1975	15974	1955	2656	1935	676	1915	85		
1994	79136	1974	10385	1954	1735	1934	420	1914	65		
1993	47175	1973	6962	1953	1367	1933	253	1913	60		
1992	51086	1972	8534	1952	1619	1932	246	1912	66		
1991	37550	1971	4556	1951	974	1931	140	1911	56		
1990	97590	1970	22756	1950	6430	1930	1147	1910	729		