

Beata Jackowska*

Analiza kohortowa czasu istnienia mikroprzedsiębiorstw w Gdańsku

Wstęp

Kondycja przedsiębiorstw, a w szczególności ich czas istnienia na rynku, jest pochodną funkcjonowania całej gospodarki narodowej. Od 2002 r. GUS prowadzi w cyklach 5-letnich reprezentacyjne badania panelowe przedsiębiorstw w Polsce. Każdy cykl oparty jest na próbie reprezentatywnej ok. 3 tys. przedsiębiorstw zarejestrowanych w danym roku (1,1% populacji). Początkowo badania obejmowały dominujące w polskiej gospodarce mikroprzedsiębiorstwa (zatrudniające mniej niż 10 osób), a od 2005 r. badania rozszerzono do małych przedsiębiorstw (zatrudniających mniej niż 50 osób)¹. Każdego roku, jesienią, wylosowane przedsiębiorstwa, które są nadal aktywne, poddawane są badaniom ankietowym². Przyjmuje się, że „jednostki aktywne to takie, które w chwili przeprowadzania badania prowadzą działalność stałą lub sezonową” [Warunki powstania ..., 2015, s. 14]. GUS co rok publikuje wyniki badań uogólnione na całą populację w różnych przekrojach – niestety wyniki badań nie są opracowywane w przekrojach regionalnych. Do analizy czasu trwania wykorzystywane są jedynie roczne indeksy dynamiki, nazywane w opracowaniu „wskaźnikami przeżycia” przedsiębiorstw. Jednak w odniesieniu do przedsiębiorstw należałoby raczej użyć określenia „wskaźnik przetrwania”³. W tabelicy 1 zaprezentowano wyniki ostatniego pełnego 5-letniego cyklu badania dla małych przedsiębiorstw powstałych w 2009 r. i obserwowanych w latach 2010–2014.

* Dr hab., Katedra Statystyki, Wydział Zarządzania, Uniwersytet Gdański, ul. Armii Krajowej 101, 81-824 Sopot, beata.jackowska@ug.edu.pl

¹ Mikroprzedsiębiorstwa zawierają się w kategorii małych przedsiębiorstw.

² Danych nie udaje się pozyskać dla około 8–9% aktywnych przedsiębiorstw [Warunki powstania ..., 2015, s. 18].

³ Dyskusję nad stosowaniem terminologii właściwej dla populacji ludzi w stosunku do zbiorowości przedsiębiorstw można znaleźć w artykułach [Domański, Szreder, 2010; Ptak-Chmielewska, 2012].

Tablica 1. Przedsiębiorstwa zatrudniające poniżej 50 pracowników zarejestrowane w Polsce w 2009 r.

Rok	Osoby fizyczne**			Osoby prawne***		
	liczba aktywnych*	wskaźnik przetrwania		Liczba aktywnych*	wskaźnik przetrwania	
		rok poprz. = 100	rok 2009 = 100		rok poprz. = 100	rok 2009 = 100
2009	259 907	x	100,00	15 400	x	100,00
2010	200 603	77,18	77,18	11 407	74,07	74,07
2011	153 534	76,54	59,07	10 889	95,46	70,70
2012	111 362	72,53	42,85	9 662	88,74	62,74
2013	89 403	80,28	34,40	8 630	89,32	56,04
2014	78 612	87,93	30,25	7 723	89,49	50,15

* W 2009 r. liczba zarejestrowanych.

** Osoby fizyczne oraz spółki cywilne.

*** Zaliczono tu wszystkie pozostałe podmioty.

Źródło: Opracowanie własne na podstawie [Warunki powstania ..., 2015, s. 94–95].

Wskaźniki przetrwania z tablicy 1 wskazują, że czas istnienia przedsiębiorstw zależy od formy prawnej. Spośród małych przedsiębiorstw powstałych w 2009 r. 5 lat przetrwało tylko 30% osób fizycznych, ale już 50% osób prawnych⁴. Dodatkowo na czas przetrwania wpływa wielkość przedsiębiorstw⁵, lecz niestety od 2005 r. sposób agregacji wyników przez GUS nie pozwala na wyodrębnienie mikroprzedsiębiorstw spośród małych przedsiębiorstw⁶. Charakterystyczną cechą polskiej gospodarki jest dominacja mikroprzedsiębiorstw. Sposób funkcjonowania tej grupy przedsiębiorstw różni się od pozostałych, czego widocznym przejawem jest długość okresu istnienia. W celu zwiększenia jednorodności badanej populacji w artykule skupiono się na prześledzeniu długości okresu istnienia mikroprzedsiębiorstw osób fizycznych.

⁴ Wskaźnik przetrwania dla powstałych w 2009 r. małych przedsiębiorstw ogółem wyniósł 31,4% [Warunki powstania ..., 2015, s. 95].

⁵ Świadczą o tym pośrednio m.in. wskaźniki przetrwania dla przedsiębiorstw w podziale na zatrudniających i niezatrudniających pracowników najemnych [Warunki powstania..., 2015, s. 95].

⁶ Wśród małych przedsiębiorstw dominują mikroprzedsiębiorstwa, więc podawane przez GUS zagregowane wyniki w przybliżeniu opisują mikroprzedsiębiorstwa, natomiast nie charakteryzują przedsiębiorstw zatrudniających od 10 do 49 pracowników.

Analiza przeprowadzonych dotychczas w Polsce badań nad czasem istnienia przedsiębiorstw wskazuje na występowanie zróżnicowania regionalnego, a także wpływ wielkości miejscowości na rozkład czasu trwania firm. Badania I. Markowicz oparte na danych z rejestru REGON pokazały, że 5 lat przetrwało 69,67% firm powstałych w Szczecinie w 2001 r. oraz 63,82% firm powstałych w Szczecinie w 2002 r. [Markowicz, 2012, s. 96–97]. Natomiast badania przeprowadzone przez A. Ptak-Chmielewską, także na podstawie danych z rejestru REGON dla pewnego województwa (nie podano jakiego), wykazały, że 5 lat przetrwało 57,50% przedsiębiorstw powstałych w 2003 r. [Ptak-Chmielewska, 2010]⁷. Mimo że A. Ptak-Chmielewska wskazuje na nieaktualność części danych w bazie REGON⁸, uzyskane wyniki mogą służyć analizie porównawczej badań opartych na rejestrze REGON oraz analizie zróżnicowania czasu istnienia ze względu na cechy firm. Badania te potwierdziły, że czas istnienia jest zróżnicowany ze względu na wielkość przedsiębiorstwa (deklarowaną liczbę zatrudnionych), formę prawną i rodzaj działalności, a dodatkowo ze względu na to, czy firma została zarejestrowana w gminie miejskiej czy wiejskiej.

Celem niniejszego badania jest określenie rozkładu czasu istnienia dla grupy mikroprzedsiębiorstw osób fizycznych powstałych w Gdańsku oraz znalezienie prawidłowości w długości okresu istnienia tych firm na gdańskim rynku. Badanie pozwoli m.in. odpowiedzieć na pytania:

1. W jakim tempie zmniejsza się liczebność powstałych w tym samym roku przedsiębiorstw?
2. W którym roku działalności ryzyko likwidacji jest największe?

Wyniki badania mogą być wykorzystane przez Urząd Miejski w Gdańsku w kreowaniu polityki gospodarczej miasta oraz mogą być przydatne do porównań z wynikami analogicznych badań przeprowadzanych w innych miastach czy województwach.

1. Opis badania

W niniejszym badaniu własnym prześledzono 5-letnią historię kohorty mikroprzedsiębiorstw powstałych w Gdańsku w 2009 r. Dane dotyczące przedsiębiorstw w zakresie: formy prawnej, przewidywanej

⁷ Dla porównania, zgodnie z wynikami badań panelowych GUS, spośród mikroprzedsiębiorstw powstałych w Polsce w 2001, 2002 i 2003 roku 5 lat przetrwało odpowiednio: 28,1%, 24,6% i 29,8% jednostek.

⁸ Problem ten opisano szerzej w artykule [Ptak-Chmielewska, 2012].

(w momencie rejestracji) liczby pracujących oraz momentu rozpoczęcia i zakończenia działalności pochodzą z rejestru REGON. Badania ograniczono do przedsiębiorstw osób fizycznych, aby z jednej strony zwiększyć jednorodność badanej zbiorowości (por. tablica 1), a z drugiej strony zapewnić lepszą jakość danych, gdyż założono, że wprowadzenie dla osób fizycznych od 31 marca 2009 r. jednego formularza rejestracji działalności gospodarczej w REGON, ZUS/KRUS oraz urządzie skarbowym a także jednego formularza wyrejestrowania, poprawiło jakość bazy REGON. Jakość ta w przeszłości uzależniona była od realizacji przez podmioty ustawowego obowiązku składania oddzielnych wniosków o wpis i skreślenie w rejestrze REGON⁹. Poprawa aktualności bazy REGON dotyczy zdarzeń od 2009 r., z tego powodu rejestr nadal nie powinien być wykorzystywany do badań przekrojowych, natomiast otwiera to pole do badań kohortowych¹⁰.

W Gdańsku w 2009 r. powstało 4786 przedsiębiorstw, z czego 98,7% stanowiły mikroprzedsiębiorstwa, tzn. w momencie rejestracji zadeklarowano przewidywaną liczbę pracujących poniżej 10 osób. Wśród zarejestrowanych 88,1% przedsiębiorstw należało do osób fizycznych. Badanie objęło ostatecznie 4174 mikroprzedsiębiorstwa osób fizycznych. Jednostki te stanowiły 87,2% wszystkich przedsiębiorstw, które rozpoczęły działalność w Gdańsku w 2009 r. Tablica 2 przedstawia strukturę tych przedsiębiorstw według przewidywanej liczby pracujących oraz formy prawnej.

Dane pochodzące z bazy REGON posłużyły do zbudowania kohortowej tablicy czasu istnienia mikroprzedsiębiorstw w Gdańsku. Wykorzystano ideę badań kohortowych rozwiniętą w ramach dziedziny demografii oraz metody analizy przeżycia pozwalające na estymację rozkładu czasu upływającego od zdarzenia początkowego do zdarzenia

⁹ Por. [Notatka ..., 2015]. Szczególnie w początkowym okresie transformacji gospodarczej, z powodu braku sankcji prawnych często nie dopełniano formalności rejestracyjnych. A. Raczyk w celu weryfikacji kompletności i aktualności rejestru REGON przeprowadził badania ankietowe w gminie miejskiej Milicz (miasto średniej wielkości) w 2006 r. i 2007 r. Analiza przekrojowa podmiotów zarejestrowanych w REGON wykazała, że fakt czy przedsiębiorstwo rzeczywiście istnieje, zależy od roku rejestracji działalności [Raczyk, 2009]. Przyczyn tego zjawiska należy doszukiwać się w zmianach w otoczeniu prawnym i gospodarczym.

¹⁰ W przypadku badań kohortowych opartych na rejestrze REGON należy zachować ostrożność przy porównywaniu wyników badań dla kohort przedsiębiorstw powstałych przed 2009 r. oraz powstałych począwszy od 2009 r. Z tego powodu w niniejszym artykule nie dokonano szczegółowych porównań z przywołanymi we wstępie wynikami badań opartymi na bazie REGON.

końcowego¹¹. Jednostki były obserwowane od momentu rozpoczęcia działalności w 2009 r. do momentu zakończenia działalności lub zakończenia badania w dniu 31 grudnia 2014 r., o ile przedsiębiorstwo było aktywne do tego dnia. Znajomość dokładnych dat tych zdarzeń umożliwia dokładniejsze przybliżenie rozkładu czasu trwania niż za pomocą jedynie wskaźników przetrwania roku kalendarzowego, jak to ma miejsce w badaniach panelowych GUS. Przy odpowiedniej liczebności zbioru danych można oszacować prawdopodobieństwa przetrwania kolejnych lat (licząc od momentu powstania), jak też okresów krótszych niż rok. Metody analizy przeżycia pozwalają na oszacowanie nie tylko prawdopodobieństwa przetrwania, ale także innych funkcji probabilistycznych określających rozkład czasu trwania, zarówno w przypadku danych kompletnych, jak też cenzurowanych.

Tablica 2. Struktura przedsiębiorstw, które rozpoczęły działalność w Gdańsku w 2009 r. według przewidywanej liczby pracujących oraz formy prawnej

Cecha	Liczba	w %
Przewidywana liczba pracujących		
do 9 osób	4723	98,68
10–49 osób	60	1,25
50–249 osób	3	0,06
Ogółem	4786	100,00
Forma prawna		
osoba fizyczna prowadząca działalność gospodarczą	4214	88,05
jednostka organizacyjna niemająca osobowości prawnej	336	7,02
osoba prawna	236	4,93
Ogółem	4786	100,00

Źródło: Badania własne na podstawie bazy REGON.

¹¹ Metody te są uniwersalne i znajdują szerokie zastosowania nie tylko do analizy czasu upływającego od urodzenia do śmierci, zob. np. [Balicki 2006; Kleinbaum, Klein, 2005; Rossa, 2005].

2. Zasady konstrukcji kohortowych tablic czasu trwania

Obserwacji podlega kohorta przedsiębiorstw, które rozpoczęły działalność w 2009 r. Czas w badaniu można określić na dwa sposoby:

- 1) jako czas kalendarzowy (historyczny) – odnotowując daty zakończenia działalności,
- 2) jako czas własny kohorty – czas upływający od momentu rozpoczęcia do momentu zakończenia działalności.

Ideę badań kohortowych, nazywanych inaczej badaniami wzdłużnymi. ilustruje dobrze wykorzystywany w demografii diagram Lexisa (rysunek 1) pozwalający odnotowywać badane zdarzenia według czasu kalendarzowego i własnego kohorty [Pressat, 1966, s. 32–49]. Ze względu na określenie czasu na dwa sposoby w demografii wyróżnia się dwa rodzaje liczby dożywających (rysunek 1):

- 1) liczba dożywających I rodzaju – w tym badaniu odpowiada to liczbie przedsiębiorstw, które dotrwały do pełnej rocznicy, licząc od momentu swojego powstania, i będzie nazywana liczbą istniejących I rodzaju;
- 2) liczba dożywających II rodzaju – w tym badaniu odpowiada to liczbie przedsiębiorstw, które dotrwały do końca roku kalendarzowego, i będzie nazywana liczbą istniejących II rodzaju¹².

Ponieważ zakończenie badania było z góry zaplanowane na dzień 31 grudnia 2014, więc nie dla wszystkich przedsiębiorstw odnotowano datę zakończenia działalności. W analizie przeżycia mówi się, że w badaniu występuje cenzurowanie prawostronne, gdy nie jest możliwe odnotowanie zdarzenia końcowego dla wszystkich obserwowanych jednostek. Taka sytuacja ma miejsce najczęściej, gdy:

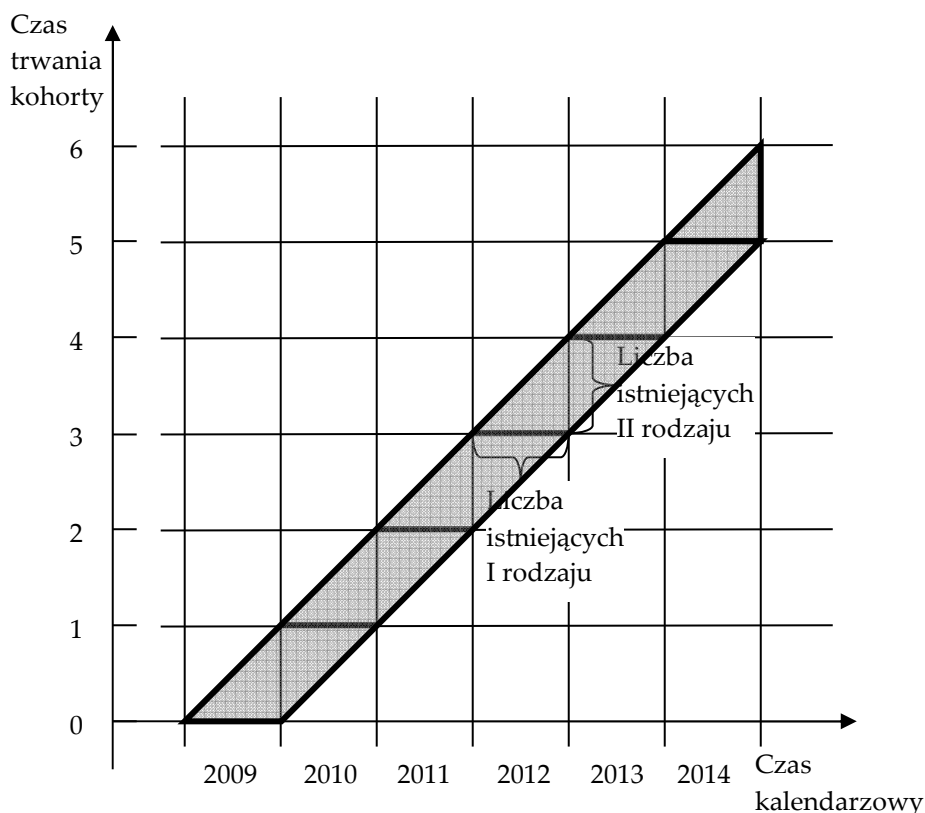
- zakończono badanie, zanim wszystkie jednostki doznały zdarzenia,
- część jednostek została utracona z pola obserwacji,
- czas trwania jednostek skończył się z innej przyczyny niż badane zdarzenie końcowe (nastąpiła realizacja ryzyka konkurencyjnego).

Jeżeli prawostronne cenzurowanie jest z góry zaplanowane w badaniu, to nazywane jest cenzurowaniem prawostronnym arbitralnym

¹² Podawane w opracowaniu GUS [Warunki powstania ..., 2015] liczby zarejestrowanych przedsiębiorstw są liczbami istniejących I rodzaju, natomiast odnotowane jesienią kolejnych lat liczby aktywnych przedsiębiorstw są liczbami istniejących II rodzaju (por. tablica 1 i rysunek 1). Prezentowane w opracowaniu GUS „wskaźniki przeżycia” są ilorazem liczb istniejących II rodzaju dla sąsiednich lat, z wyjątkiem wskaźnika dla pierwszego roku będącego relacją liczby istniejących II rodzaju do liczby istniejących I rodzaju.

i z taką sytuacją mamy do czynienia w niniejszym badaniu [Rossa, 2005, s. 13–15]. Wszystkie czasy cenzurowania znajdują się w ostatnim szóstym rocznym przedziale według czasu własnego kohorty i co więcej – wszystkie zdarzenia z ostatniego przedziału zawarte są w ostatnim trójkącie dolnym¹³ na siatce Lexisa (rysunek 1). Ponieważ momenty rozpoczęcia działalności gospodarczej w roku 2009 są losowe, więc czasy cenzurowania wewnątrz ostatniego trójkąta dolnego są także losowe.

Rysunek 1. Diagram Lexisa dla kohorty przedsiębiorstw powstałych w 2009 r.



Źródło: Opracowanie własne.

Nieparametryczny model rozkładu czasu trwania w demografii najczęściej prezentuje się w postaci tablic wymieralności, nazywanych inaczej tablicami trwania życia. Tablice wymieralności można określić jako formalną konstrukcję przyporządkowującą przedziałom czasu

¹³ W siatce Lexisa występują trójkąty prostokątne: trójkąty dolne i trójkąty górne, których nazwa pochodzi od położenia kąta prostego.

wartości wybranych funkcji oraz wartości parametrów opisujących rozkład czasu trwania (przy czym czas trwania jest tu czasem własnym kohorty) [Chiang, 1984]. Metody konstrukcji tablic wymieralności można uogólnić i wykorzystać do badania rozkładów czasu trwania nie tylko w populacjach ludzi [Balicki, 2006].

W celu zbudowania tablicy dla kohorty rzeczywistej czas trwania kohorty dzieli się arbitralnie na przedziały, wybierając granice i szerokości przedziałów (tablica 3). W każdym przedziale czasu oddzielnie dokonuje się estymacji funkcji tablicowych, wychodząc od określenia:

- liczby istniejących jednostek – w tym badaniu liczby przedsiębiorstw istniejących (I rodzaju) na początku przedziału czasu,
- liczby zdarzeń końcowych – w tym przypadku liczby przedsiębiorstw, które zakończyły działalność w przedziale czasu,
- liczby jednostek cenzurowanych – liczby przedsiębiorstw, których obserwacja zakończyła się w danym przedziale czasu przed odnotowaniem zakończenia działalności.

Tablica 3. Schemat empirycznej tablicy czasu trwania kohorty rzeczywistej w przypadku danych cenzurowanych losowo

Lp.	Czas trwania	Szerokość przedziału czasu trwania	Liczba jednostek istniejących	Liczba zdarzeń końcowych	Liczba jednostek cenzurowanych
i	$[x_i, x_{i+1})$	c_i	N_i	D_i	R_i
0	$x_0 - x_1$	c_0	N_0	D_0	R_0
1	$x_1 - x_2$	c_1	N_1	D_1	R_1
...
ω	x_ω i więcej	c_ω	N_ω	D_ω	R_ω

Źródło: Opracowanie własne na podstawie [Balicki, 2006, s. 79].

3. Estymatory funkcji tablicowych

W przypadku występowania danych cenzurowanych do budowy tablicy można zastosować metodę aktuarialną, nazywaną inaczej metodą tablicy wymieralności. W metodzie tej przyjmuje się założenie, że liczba jednostek narażonych na wystąpienie zdarzenia wynosi zgodnie z oznaczeniami z tablicy 3:

$$N'_i = \frac{1}{2}((N_i - R_i) + N_i) = N_i - \frac{1}{2}R_i, \quad (1)$$

co oznacza, że jednostki cenzurowane w danym przedziale czasu narażone były na zdarzenie średnio przez połowę długości tego przedziału [Elandt-Johnson, Johnson, 1980, s. 157]. Jest to równoważne z założeniem równomiernego rozkładu czasów cenzurowanych w każdym z przedziałów. W tabelicy 4, przy założeniu aktuarialnym, przedstawiono postać estymatorów najważniejszych funkcji określających rozkład czasu trwania.

Tabela 4. Estymatory aktuarialne funkcji tablicowych w przypadku występowania danych cenzurowanych

Nazwa funkcji	Estymator	Standardowy błąd szacunku
Warunkowe prawdopodobieństwo wystąpienia zdarzenia	$\hat{q}_i = \frac{D_i}{N'_i}$	$S(\hat{q}_i) = \sqrt{\frac{\hat{p}_i \hat{q}_i}{N'_i}}$
Warunkowe prawdopodobieństwo przetrwania	$\hat{p}_i = 1 - \hat{q}_i$	$S(\hat{p}_i) = \sqrt{\frac{\hat{p}_i \hat{q}_i}{N'_i}}$
Funkcja trwania (przetrwania)	$\hat{S}_0 = 1$ $\hat{S}_i = \prod_{k=0}^{i-1} \hat{p}_k$	$S(\hat{S}_0) = 0$ $S(\hat{S}_i) = \hat{S}_i \sqrt{\sum_{k=0}^{i-1} \frac{\hat{q}_k}{N'_k \hat{p}_k}}$
Funkcja gęstości	$\hat{f}_i = \frac{\hat{S}_i - \hat{S}_{i+1}}{c_i}$	$S(\hat{f}_0) = \frac{\hat{S}_0 \hat{q}_0}{c_0} \sqrt{\frac{\hat{p}_0}{N'_0 \hat{q}_0}}$ $S(\hat{f}_i) = \frac{\hat{S}_i \hat{q}_i}{c_i} \sqrt{\sum_{k=0}^{i-1} \frac{\hat{q}_k}{N'_k \hat{p}_k} + \frac{\hat{p}_i}{N'_i \hat{q}_i}}$ Jeżeli $\hat{q}_i = 0$, to $S(\hat{f}_i) = 0$
Funkcja intensywności zdarzeń	$\hat{\lambda}_i = \frac{2\hat{q}_i}{c_i(1 + \hat{p}_i)}$	$S(\hat{\lambda}_i) = \sqrt{\frac{\hat{\lambda}_i^2}{N'_i q_i} \left[1 - \left(\frac{\hat{\lambda}_i c_i}{2} \right)^2 \right]}$ Jeżeli $\hat{q}_i = 0$, to $S(\hat{\lambda}_i) = 0$
Mediana dalszego czasu trwania	Jeżeli $\hat{S}_{i+1} \leq 0,5\hat{S}_j \leq \hat{S}_i$, to $me_j = t_i - t_j + \frac{\hat{S}_i - 0,5\hat{S}_j}{\hat{S}_i - \hat{S}_{i+1}} c_i$	$S(me_j) = \sqrt{\frac{\hat{S}_j^2}{4N'_j \hat{f}_j^2}}$

Źródło: Opracowanie własne na podstawie [Jackowska, 2013].

W badaniach kohortowych, zwykle w ostatnim przedziale tablicy, występują jednostki cenzurowane. W takich przypadkach nie jest możliwe oszacowanie rozkładu w całości, tzn. do maksymalnego możliwego czasu trwania. Nie można także obliczyć przeciętnego czasu trwania, lecz można obliczyć medianę czasu trwania, o ile co najmniej połowa jednostek doznała zdarzenia końcowego w okresie badania. W przedziałach czasu, w których nie ma obserwacji cenzurowanych, estymatory aktuarialne są równoważne estymatorom największej wiarygodności zaprezentowanym w tablicy 5.

Tablica 5. Estymatory największej wiarygodności funkcji tablicowych w przypadku danych pełnych

Nazwa funkcji	Estymator	Standardowy błąd szacunku
Warunkowe prawdopodobieństwo wystąpienia zdarzenia	$\hat{q}_i = \frac{D_i}{N_i}$	$S(\hat{q}_i) = \sqrt{\frac{\hat{p}_i \hat{q}_i}{N_i}}$
Warunkowe prawdopodobieństwo przetrwania	$\hat{p}_i = 1 - \hat{q}_i$	$S(\hat{p}_i) = \sqrt{\frac{\hat{p}_i \hat{q}_i}{N_i}}$
Funkcja trwania (przetrwania)	$\hat{S}_i = \frac{N_i}{N_0}$	$S(\hat{S}_i) = \sqrt{\frac{\hat{S}_i(1 - \hat{S}_i)}{N_0}}$
Funkcja gęstości	$\hat{f}_i = \frac{D_i}{c_i N_0}$	$S(\hat{f}_i) = \sqrt{\frac{\hat{f}_i(1 - c_i \hat{f}_i)}{c_i N_0}}$
Funkcja intensywności zdarzeń	$\hat{\lambda}_i = \frac{D_i}{N_{i+1}c_i + \frac{1}{2}D_i c_i}$	$S(\hat{\lambda}_i) = \frac{\sqrt{D_i}}{(N_{i+1}c_i + \frac{1}{2}D_i c_i)}$
Mediana dalszego czasu trwania	Jeżeli $\hat{S}_{i+1} \leq 0,5\hat{S}_j \leq \hat{S}_i$, to $me_j = t_i - t_j + \frac{\hat{S}_i - 0,5\hat{S}_j}{\hat{S}_i - \hat{S}_{i+1}} c_i$	$S(me_j) = \sqrt{\frac{\hat{S}_j^2}{4N_j \hat{f}_i^2}}$
Przeciętny dalszy czas trwania	$\hat{e}_i = \frac{1}{2N_i} \sum_{k=i}^{\omega} (N_k + N_{k+1})c_k$	$S(\hat{e}_i) = \sqrt{\sum_{k=i}^{\omega-1} \left(\frac{N_k}{N_i}\right)^2 \left(\hat{e}_{k+1} + \frac{c_k}{2}\right)^2 S^2(\hat{p}_k)}$

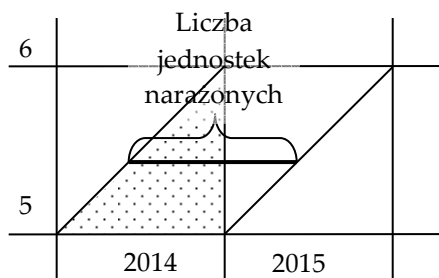
Źródło: Opracowanie własne na podstawie [Jackowska, 2013].

4. Wyniki estymacji kohortowej tablicy czasu istnienia mikroprzedsiębiorstw

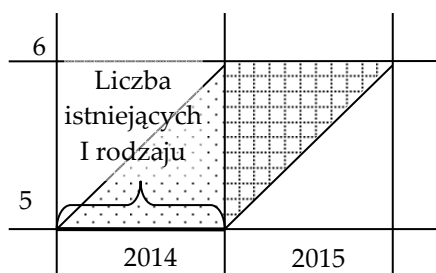
W niniejszym badaniu wszystkie obserwacje cenzurowane pojawiają się w szóstym roku istnienia kohorty przedsiębiorstw, w ostatnim trójkącie dolnym na siatce Lexisa (rysunek 1). Dla pierwszych 5 lat istnienia przedsiębiorstw można więc wykorzystać estymatory największej wiarygodności. Ponieważ po zakończeniu badania istnieją jeszcze aktywne przedsiębiorstwa, więc rozkładu nie można określić w całości. Można jednak oszacować rozkład w szóstym roku przy założeniu aktuarialnym, w którym zdarzenia z trójkąta dolnego odnoszone są do liczby jednostek narażonych (rysunek 2a). Założenie aktuarialne można zastąpić założeniem równomiernego rozkładu liczby zdarzeń wewnątrz rocznego przedziału czasu trwania (według kalendarza własnego kohorty). Założenie to pozwoli na oszacowanie brakującej liczebności zdarzeń w trójkącie górnym (rysunek 2b).

Rysunek 2. Fragment diagramu Lexisa – ostatni rok badania kohorty przedsiębiorstw powstałych w 2009 r.

a) Założenie aktuarialne



b) Założenie równomiernego rozkładu zdarzeń wewnątrz rocznego przedziału



Źródło: Opracowanie własne.

Przyjmując roczne przedziały czasu trwania (według czasu własnego kohorty), otrzymano kohortową tablicę czasu trwania mikroprzedsiębiorstw osób fizycznych w Gdańsku (tablica 6). Spośród 4174 mikroprzedsiębiorstw rozpoczynających działalność w 2009 r. 5 lat przetrwało 2009 jednostek (liczba istniejących I rodzaju). Prawdopodobieństwo przetrwania 5 lat dla tych przedsiębiorstw wyniosło 48,1%. Największa intensywność likwidacji wystąpiła w trzecim roku, a w następnej kolejności w piątym roku istnienia przedsiębiorstw. W rozkładzie czasu

istnienia przedsiębiorstw nie można wyznaczyć wartości przeciętnej, lecz można wyznaczyć medianę, która wyniosła około 4 lata 9,5 miesiąca (4,791 roku). Z powodu prawostronnego cenzurowania arbitralnego w ostatnim przedziale znalazło się 1879 jednostek cenzurowanych (liczba istniejących II rodzaju). W szóstym roku istnienia kohorty rozkład został oszacowany przy założeniu równomiernego rozkładu liczby zdarzeń (por. rysunek 2b).

Tablica 6. Kohortowa tablica czasu istnienia mikroprzedsiębiorstw osób fizycznych powstałych w Gdańsku w 2009 r. – estymacja w przedziałach rocznych

Rok kalendarzowy	Przedział czasu trwania w latach	Prawdop. zakończenia działalności	Funkcja przetrwania	Gęstość rozkładu	Intensywność zdarzeń	Mediana dalszego trwania w latach
$t/(t+1)$	$[x, x+1)$	q_x	$S(x)$	$f(x)$	$\mu(x)$	$me(x)$
2009/2010	[0, 1)	0,1150	1,0000	0,1150	0,1220	4,791
2010/2011	[1, 2)	0,1326	0,8850	0,1174	0,1421	4,648
2011/2012	[2, 3)	0,1601	0,7676	0,1229	0,1740	–
2012/2013	[3, 4)	0,1145	0,6447	0,0738	0,1214	–
2013/2014	[4, 5)	0,1569	0,5709	0,0896	0,1703	–
2014/2015	[5, 6)*	0,1294	0,4813	0,0623	0,1384	–

* Brakującą liczebność zdarzeń w trójkącie górnym oszacowano przy założeniu równomiernego rozkładu zdarzeń wewnątrz rocznego przedziału czasu trwania (według kalendarza własnego kohorty).

Źródło: Badania własne na podstawie bazy REGON.

Liczebność kohorty pozwoliła na skonstruowanie tablicy bardziej szczegółowej o przedziałach szerokości jednego kwartału (tablica 7). Dzięki temu, że funkcje określające rozkład zostały estymowane w przedziałach czterokrotnie krótszych, można stwierdzić, że wysoka intensywność likwidacji firm miała miejsce pod koniec drugiego roku i w pierwszej połowie trzeciego roku, a także w drugiej połowie piątego roku i na początku szóstego roku. Dokładniejsze jest też oszacowanie mediany, która wyniosła około 4 lata i 10 miesięcy (19,28 kwartału \approx 4,82 roku). W przypadku ostatnich czterech przedziałów kwartalnych, w których występują dane cenzurowane, założenie równomiernego rozkładu liczby zdarzeń wewnątrz rocznego przedziału czasu trwania zostało zastąpione założeniem równomiernego rozkładu zdarzeń we-

wnątrz kwartalnych przedziałów czasu trwania. Pozwoliło to na dokładniejsze oszacowanie rozkładu w szóstym roku.

Tablica 7. Kohortowa tablica czasu istnienia mikroprzedsiębiorstw osób fizycznych powstałych w Gdańsku w 2009 r. – estymacja w przedziałach kwartalnych

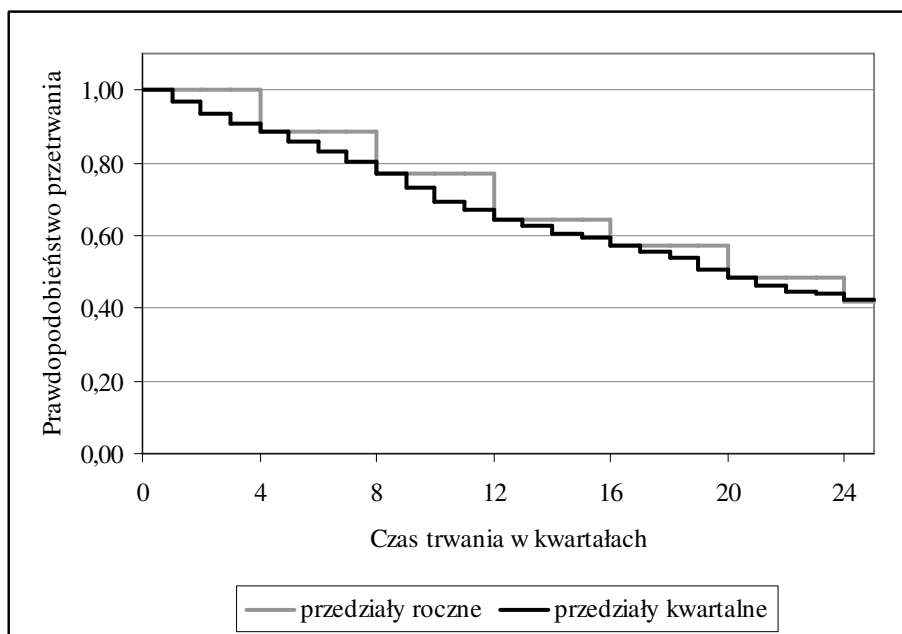
Rok	Kwartał	Przedział czasu trwania w kwartałach	Prawdop. zakończenia działalności	Funkcja przetrwania	Gęstość rozkładu	Intensywność zdarzeń	Mediana dalszego trwania w kwartałach
		$[x, x+1)$	q_x	$S(x)$	$f(x)$	$\mu(x)$	$me(x)$
1	1	[0, 1)	0,0340	1,0000	0,0340	0,0346	19,28
	2	[1, 2)	0,0322	0,9660	0,0311	0,0328	18,94
	3	[2, 3)	0,0269	0,9348	0,0252	0,0273	18,70
	4	[3, 4)	0,0271	0,9097	0,0247	0,0275	18,49
2	5	[4, 5)	0,0317	0,8850	0,0280	0,0322	18,52
	6	[5, 6)	0,0282	0,8570	0,0242	0,0286	18,79
	7	[6, 7)	0,0328	0,8328	0,0273	0,0333	–
	8	[7, 8)	0,0470	0,8055	0,0379	0,0481	–
3	9	[8, 9)	0,0506	0,7676	0,0388	0,0519	–
	10	[9, 10)	0,0470	0,7288	0,0343	0,0481	–
	11	[10, 11)	0,0359	0,6945	0,0249	0,0365	–
	12	[11, 12)	0,0372	0,6696	0,0249	0,0379	–
4	13	[12, 13)	0,0286	0,6447	0,0184	0,0290	–
	14	[13, 14)	0,0333	0,6263	0,0208	0,0338	–
	15	[14, 15)	0,0222	0,6054	0,0134	0,0224	–
	16	[15, 16)	0,0356	0,5920	0,0211	0,0363	–
5	17	[16, 17)	0,0260	0,5709	0,0149	0,0264	–
	18	[17, 18)	0,0349	0,5561	0,0194	0,0355	–
	19	[18, 19)	0,0549	0,5367	0,0295	0,0565	–
	20	[19, 20)	0,0510	0,5072	0,0259	0,0524	–
6	21	[20, 21)*	0,0410	0,4813	0,0197	0,0418	–
	22	[21, 22)*	0,0299	0,4616	0,0138	0,0303	–
	23	[22, 23)*	0,0228	0,4478	0,0102	0,0231	–
	24	[23, 24)*	0,0263	0,4376			

* Brakującą liczebność zdarzeń w trójkącie górnym oszacowano przy założeniu równomiernego rozkładu zdarzeń wewnątrz kwartalnego przedziału czasu trwania (według kalendarza własnego kohorty).

Źródło: Badania własne na podstawie bazy REGON.

Na rysunku 3 porównano wykresy funkcji przetrwania oszacowanej w rocznych przedziałach czasu (tablica 6) oraz w kwartalnych przedziałach czasu (tablica 7). W przypadku modeli nieparametrycznych funkcja ta jest stała w przyjętych przedziałach. Im węższe przedziały, tym model bardziej przybliża się do rzeczywistej funkcji opisującej rozkład czasu istnienia firm. Z drugiej strony zbyt rozbudowane tablice mogą być nieczytelne dla użytkowników i utrudniać zastosowanie w praktyce.

Rysunek 3. Funkcja przetrwania dla mikroprzedsiębiorstw osób fizycznych powstałych w Gdańsku w 2009 r. – estymacja w przedziałach rocznych i kwartalnych



Źródło: Badania własne na podstawie bazy REGON.

5. Analiza Kaplana-Meiera czasu istnienia mikroprzedsiębiorstw

Za pomocą estymatora Kaplana-Meiera można otrzymać model nieparametryczny, niewymagający arbitralnego podziału czasu na przedziały, jak to ma miejsce w tablicach czasu trwania. W tym przypadku czas trwania dzielony jest losowo na przedziały w momentach wystąpienia zdarzeń dla poszczególnych jednostek. W każdym przedziale

czasu oddzielnie dokonuje się estymacji wartości funkcji przetrwania według wzoru:

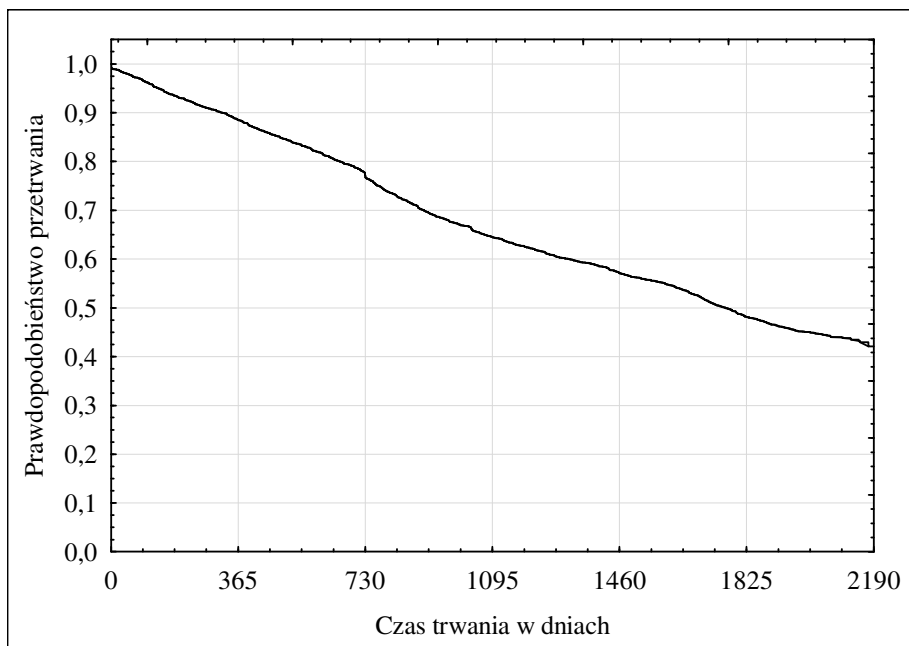
$$\hat{S}_i = \begin{cases} 1 & \text{dla } i = 0 \\ \prod_{k=1}^i \left(\frac{N_k - D_k}{N_k} \right) & \text{dla } i = 1, 2, \dots, \omega - 1 \\ 0 & \text{dla } i = \omega, \text{ gdy w ostatnim przedziale,} \\ & \text{nie ma czasow cenzurowanych} \\ \text{nieokreslony} & \text{dla } i = \omega, \text{ gdy w ostatnim przedziale} \\ & \text{sa czasy cenzurowane} \end{cases} \quad (2)$$

gdzie oznaczenia są zgodne z tabelą 3. Estymator wyrażony powyższym wzorem nazywany jest estymatorem Kaplana-Meiera [Rossa, 2005, s. 34; Kleinbaum, Klein, 2005, s. 48–57]. W wyniku estymacji otrzymuje się funkcję także przedziałami stałą, lecz o skokach w losowych punktach wyznaczonych przez obserwacje pełne (niecenzurowane). Im większa próba, tym krzywa Kaplana-Meiera ma mniejsze skoki i dokładniej przybliża nieznaną rozkład zmiennej ciągłej będącej czasem trwania.

Na rysunku 4 przedstawiono krzywą Kaplana-Meiera dla mikroprzedsiębiorstw osób fizycznych. Duża liczebność kohorty spowodowała, że skoki funkcji przetrwania są prawie niewidoczne na wykresie. W kohorcie mikroprzedsiębiorstw powstałych w 2009 r. prawdopodobieństwo przetrwania stopniowo spadało do poziomu 48,1% po pięciu latach, a następnie do poziomu 42,1% na końcu 2014 r. Z powodu istnienia obserwacji cenzurowanych w momencie zakończenia badania nie jest możliwe określenie rozkładu czasu trwania w całości, tzn. do końca istnienia kohorty. Na podstawie krzywej Kaplana-Meiera można wyznaczyć dokładną wartość mediany, o ile w okresie badania co najmniej 50% jednostek doświadczyło zdarzenia końcowego. Mediana czasu istnienia mikroprzedsiębiorstw wyniosła około 4 lata 10 miesięcy (1764 dni \approx 4,83 roku). Okazało się więc, że oszacowanie mediany na podstawie tablic z przedziałami kwartalnymi jest zbliżone do dokładnej wartości empirycznej.

Model nieparametryczny określony krzywą Kaplana-Meiera jest dokładniejszy od modeli przedstawionych w postaci tablic czasu trwania, dlatego jest ważny z teoretycznego punktu widzenia. Jednakże tablice dają syntetyczny obraz badanego zjawiska i dla praktyków mogą stanowić podstawowe źródło wiedzy.

Rysunek 4. Funkcja przetrwania dla mikroprzedsiębiorstw osób fizycznych powstałych w Gdańsku w 2009 r. – estymator Kaplana-Meiera



Źródło: Badania własne na podstawie bazy REGON.

Zakończenie

Niniejsze badanie pozwoliło na oszacowanie rozkładu czasu istnienia mikroprzedsiębiorstw osób fizycznych powstałych w Gdańsku w 2009 r. 5 lat przetrwało 48% jednostek, można spodziewać się, że 6 lat przetrwa 42% jednostek. Połowa mikroprzedsiębiorstw istniała nie dłużej niż 4 lata i 10 miesięcy, a druga połowa nie krócej niż 4 lata i 10 miesięcy. Wysoka intensywność likwidacji została odnotowana na przełomie drugiego i trzeciego roku oraz na przełomie piątego i szóstego roku istnienia firm. Pierwszy z zaobserwowanych krytycznych okresów w działalności gospodarczej zbiega się z zakończeniem 24-miesięcznego okresu ulgi w opłacaniu składek na ubezpieczenie społeczne (z ulgi mogą skorzystać rozpoczynający działalność i spełniający określone warunki) [ustawa, 2005, art. 1]. Przedsiębiorstwa, które były w stanie funkcjonować po zakończeniu okresu preferencyjnego, były szczególnie narażone w drugim krytycznym okresie. Potrzebne są dalsze badania, aby sprawdzić czy przedsiębiorstwa, które przetrwały 5 lat są już w miarę „bezpieczne” – dobrze dostosowały się do rynku.

Wprowadzenie w 2009 r. jednego formularza rejestracji działalności gospodarczej w REGON, ZUS/KRUS oraz urzędzie skarbowym powinno przyczynić się do znacznej poprawy jakości bazy REGON. Dalsza poprawa jakości bazy REGON związana jest z powstaniem Centralnej Ewidencji i Informacja o Działalności Gospodarczej (CEIDG). Od dnia 1 lipca 2011 r. rejestracja i zmiany we wpisach działalności gospodarczej osób fizycznych dokonywane są w CEIDG za pośrednictwem formularza elektronicznego lub w urzędzie gminy, lub listem poleconym¹⁴. Rejestr REGON jest na bieżąco aktualizowany na podstawie danych z CEIDG. Ponadto od 2011 r. dwa razy w roku do aktualizacji rejestru REGON są wykorzystywane informacje dotyczące zgonów osób fizycznych pochodzące z Powszechnego Elektronicznego Systemu Ewidencji Ludności (PESEL).

Dzięki tym zmianom otwiera się pole do szczegółowych badań kohortowych czasu istnienia przedsiębiorstw: w dowolnej jednostce administracyjnej, według rodzaju działalności, według deklarowanej przewidywanej liczby zatrudnionych. Wyniki badań tego typu mogą przyczynić się do prowadzenia skutecznej lokalnej polityki wspierającej rozwój przedsiębiorczości na danym terenie.

Literatura

1. Balicki A. (2006), *Analiza przeżycia i tablice wymieralności*, PWE, Warszawa.
2. Chiang C. L. (1984), *The Life Table and its Applications*, Robert E. Krieger Company Malabar, Florida.
3. Domański Cz., Szreder M. (2010), *Demografia oderwana od populacji ludzkiej?*, „Wiadomości Statystyczne”, nr 12.
4. Elandt-Johnson R. C., Johnson N. L. (1980), *Survival Models and Data Analysis*, John Wiley & Sons, New York.
5. Jackowska B. (2013), *Modele dalszego trwania życia oraz ich zastosowania w przypadku osób starszych*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk.
6. Kleinbaum D. G., Klein M. (2005), *Survival Analysis. A Self-Learning Text*, Springer-Verlag, New York.

¹⁴ Do końca 2011 roku przeniesione zostały dane z ewidencji gminnych do CEIDG dotyczące przedsiębiorstw funkcjonujących, tzn. aktywnych i zawieszonych.

7. Markowicz I. (2012), *Statystyczna analiza żywotności firm*, „Rozprawy i Studia” nr 835, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin.
8. *Notatka na temat różnic między informacjami o podmiotach gospodarki narodowej z rejestru urzędowego REGON oraz o przedsiębiorstwach z badań statystycznych GUS* (2015), GUS, Warszawa, maj 2015 r., http://stat.gov.pl/gfx/portalinformacyjny/userfiles/_public/podmioty_gospodarcze/rejestr_regon_oraz_badania_statystyczne_material_educacyjny.pdf, dostęp dnia 24.08.2015.
9. Pressat R. (1966), *Analiza demograficzna. Metody, wyniki, zastosowania*, PWN, Warszawa.
10. Ptak-Chmielewska A. (2010), *Analiza przeżycia przedsiębiorstw w Polsce na przykładzie wybranego województwa*, w: *Prognozowanie w zarządzaniu firmą*, Dittmann P. (red.), „Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu” nr 103.
11. Ptak-Chmielewska A. (2012), *Dostępność i przydatność danych do analizy przeżycia przedsiębiorstw*, „Wiadomości Statystyczne” nr 6.
12. Raczyk A. (2009), *Metody badania przedsiębiorczości oparte na rejestrze podmiotów gospodarki narodowej*, w: *Rola przedsiębiorczości w kształtowaniu społeczeństwa informacyjnego*, Ziolo T., Rachwał T. (red.), „Przedsiębiorczość – Edukacja” nr 5, Zakład Przedsiębiorczości i Gospodarki Przestrzennej Instytutu Geografii Uniwersytetu Pedagogicznego w Krakowie.
13. Rossa A. (2005), *Metody estymacji rozkładu czasu trwania zjawisk dla danych cenzurowanych oraz ich zastosowania*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
14. Ustawa z dnia 1 lipca 2005 r. o zmianie ustawy o systemie ubezpieczeń społecznych oraz niektórych innych ustaw, Dz.U. Nr 150, poz. 1248.
15. *Warunki powstania i działania oraz perspektywy rozwojowe polskich przedsiębiorstw powstałych w latach 2009–2013* (2015), „Informacje i Opracowania Statystyczne”, GUS, Warszawa.

Streszczenie

W artykule przedstawiono wyniki badania kohortowego czasu istnienia mikroprzedsiębiorstw powstałych w Gdańsku w 2009 r. Celem badania było określenie rozkładu czasu istnienia mikroprzedsiębiorstw w Gdańsku oraz znalezienie prawidłowości w długości okresu istnienia firm na gdańskim rynku. Wykorzystano ideę analizy kohortowej rozwiniętą w ramach dziedziny

demografii oraz metody analizy przeżycia pozwalające na estymację rozkładu czasu upływającego od zdarzenia początkowego (rozpoczęcie działalności gospodarczej) do zdarzenia końcowego (zakończenie działalności gospodarczej). Skonstruowano empiryczną tablicę kohortową czasu istnienia mikroprzedsiębiorstw, a także wykorzystano estymator Kaplana-Meiera do oszacowania funkcji przetrwania. Wyniki badania mogą być wykorzystane do prowadzenia skutecznej polityki gospodarczej w Gdańsku wspierającej rozwój przedsiębiorczości oraz mogą być przydatne do porównań z wynikami analogicznych badań przeprowadzanych w innych miastach czy województwach.

Słowa kluczowe

mikroprzedsiębiorstwa, analiza przeżycia, analiza kohortowa

Cohort analysis of survival time of micro enterprises in Gdansk (Summary)

This paper presents the results of a longitudinal study of survival time of micro enterprises set up in Gdansk in 2009. The aim of the research was to determine a distribution of survival time of these enterprises, and to derive a pattern for firms' existence time on the market in Gdansk. A cohort analysis approach, developed within the field of demography, was employed. Survival analysis methods, which allow to estimate distribution of survival time from the initial event (outset of business activity) to the final event (termination of business activity), were also applied. An empirical cohort life table of micro enterprises was constructed. In addition, a survival function was estimated using the Kaplan-Meier estimator. Results of the study can be used to implement an effective economic policy to support business development in Gdansk, and for comparison with results of similar research carried out in other cities or regions.

Keywords

micro enterprises, survival analysis, cohort analysis