

フィンランド統計局のビジネスレジスターを用いた プロファイリング対象の選定の実験¹

菅 幹雄（法政大学経済学部）

1. はじめに

筆者が代表者を務める研究グループ（平成 24～25 年度科学研究費補助金基盤研究(C)、研究課題名「ビジネスレジスターによる企業動態統計の開発」）ではフィンランド統計局より、2002 年についてのビジネスレジスターのマイクロデータの提供を受けた。

ビジネスレジスターの作成・メンテナンスには、税務記録情報など行政記録情報が使用されているため、その情報の秘匿は法律で厳しく規定されており、ビジネスレジスターのマイクロデータを一般に提供している国はほとんどない²。その中であってフィンランドはビジネスレジスターのマイクロデータを一般に有償で提供している希有な事例である。

本稿はフィンランド統計局が提供しているビジネスレジスターのマイクロデータの構造と分布を観察した上で、このマイクロデータを用いたプロビット・モデルによるプロファイリング対象の選定の実験について報告する。

2. フィンランドのビジネスレジスターのマイクロデータの構造と分布

ビジネスレジスターのマイクロデータは企業データ（Data on enterprises）と事業所データ（Data on establishments）の二つのファイルから構成されている。企業データに収録されているのは表 1 の一番左の列、事業所データに収録されているのは表 1 の真ん中の列に示された項目である。

事業者識別番号が企業データと事業所データの両方にあるので、二つのファイルはマッチング可能である。このほかに両方のデータにある項目としては、市町村コード、産業分類、従業者数の規模階級、開設年月日がある。企業データにしかない項目には、従業者数、売上高の規模階級、法的形態（経営組織）、所有者の区分、雇用主／非雇用主の別、付加価値税納税者／非納税者の別、輸入／輸出の有無である。一方、事業所データにしかない項目には、事業所コード、郵便番号、市町村の従業者数、経度・緯度がある。

雇用主／非雇用主の別、付加価値税納税者／非納税者の別、輸入／輸出の有無、法的形態（経営組織）が企業データにしかないのは、企業が法的単位（legal unit）であるからである。すなわち、雇用者の税金・社会保険等の源泉徴収、付加価値税の納税、EU域内取引の輸入に関する付加価値税の仕入税額控除は法的単位で申告されるからである。

¹ 本稿の内容については、フィンランド統計局のユッカ・パコラ（Jukka Pakola）氏、慶應義塾大学の宮内環氏、総務省統計局基本構造統計課の永島勝利氏、高部勲氏からコメントをいただいた。ここに記して感謝したい。

² 米国では研究者からビジネスレジスターのマイクロデータの提供の要望が強くあり、合成データ（synthetic data、擬似マイクロデータのこと）による提供を模索している Kinney, Reiter[2013]。

売上高の規模階級が企業データにしかないのは、フィンランド統計局のビジネスレジスターでは、売上高を付加価値税の情報に基づいて把握しているが、付加価値税は企業単位で申告されるのであり、事業所単位ではないからである。このように行政記録の単位によって、存在する情報の種類が異なる。

表 1 フィンランド統計局のビジネスレジスターのマイクロデータに収録された項目

企業データ Data on enterprises	事業所データ Data on establishments	内容
Business ID	Business ID	事業者識別番号(8桁)
	Establishment code	事業所コード(9桁)
	Postal code	郵便番号(5桁)
Municipality	Municipality	市町村コード(3桁)
Industry, TOL 2002	Industry, TOL 2002	産業分類(5桁)
Personnel		従業者数
Size category of personnel	Size category of personnel	従業者数規模階級
Size category of turnover		売上高規模階級
Start of operation	Start of operation	開業・開設年月日(企業は年月まで)
Legal form		法的形態(経営組織)
Type of owner		所有者の区分
Employer activity		雇用主/非雇用主の別
VAT activity		付加価値税納税者/非納税者の別
Importer/Exporter		輸入/輸出の有無
	Number of personnel in the municipality	市町村の従業者数
	Co-ordinates	経度・緯度

従業者数規模階級は企業と事業所でやや異なっており、事業所では 200 人以上がまとめられている(表 2)。なお企業については従業者数の情報があるので、これとは異なる規模階級を設定することが可能である。

表 2 従業者数規模階級

企業 (単位: 人)	事業所 (単位: 人)
1 = 0-4	1 = 0-4
2 = 5-9	2 = 5-9
3 = 10-19	3 = 10-19
4 = 20-49	4 = 20-49
5 = 50-99	5 = 50-99
6 = 100-199	6 = 100-199
7 = 200-499	7 = 200-
8 = 500-999	
9 = 1000-	

売上高規模階級情報は事業所データにはなく、企業データだけにある。表 3 は企業デー

タの売上高規模階級を示している。売上高の一番小さい規模階級1は0～999ユーロであり、1ユーロ約120円で換算すると0～約12万円しかない。表4の法的形態別売上高規模階級別企業数を見るとその多くは有限責任会社(Limited company)、自然人(Natural person)、住宅営団(Housing corporation)である。自然人は我が国の個人企業に相当する。このように売上高が少ない企業は、おそらく有限責任会社の場合は活動実態がほとんどなく、自然人(個人企業)の場合は趣味であろう。その次の規模階級2(1,000～39,999ユーロ)は約12～約480万円、規模階級3(40,000～99,999ユーロ)は約480～1200万円である。表4を見るとその多くは自然人(個人企業)、有限責任会社、合名会社であり、零細な自営業者を多く含んでいると考えられる。ちなみにフィンランド統計局のビジネスレジスターにおいて「企業」(Enterprise)と呼ぶ場合は、このように会社企業以外も含んでいる。

表3 売上高規模階級

企業 (単位:ユーロ)
1 = 0 - 999
2 = 1 000 - 39 999
3 = 40 000 - 99 999
4 = 100 000 - 399 999
5 = 400 000 - 1 999 999
6 = 2 000 000 - 9 999 999
7 = 10 000 000 - 39 999 999
8 = 40 000 000 - 199 999 999
9 = 200 000 000 -

表4 法的形態別売上高規模階級別企業数

法的形態	売上高規模階級(単位:ユーロ)									合計
	0-999	1,000-39,999	40,000-99,999	100,000-399,999	400,000-1,999,999	2,000,000-9,999,999	10,000,000-39,999,999	40,000,000-199,999,999	200,000,000-	
11 自然人	1579	43888	30281	13130	1400	476	2			90756
12 遺産	39	71	33	16	8	4				171
13 納税義務がある法人	119	277	151	136	34	3	1	1		722
14 合名会社	78	1602	2275	3150	729	88	4			7926
15 破産財団	9	3	8	7	5	1				33
21 合資会社	181	4649	6643	9298	2905	302	8	1		23987
22 共同所有輸送会社						3				3
31 有限責任会社	3770	14589	17797	31499	21360	6868	1706	610	144	98343
32 相互保険会社	19									19
33 貯蓄銀行	40									40
34 年金財団	35									35
35 住宅営団	1165									1165
41 協同組合	336	428	202	209	87	54	31	24	8	1379
51 財団	40									40
52 任意団体	48									48
53 相互損害賠償保険組合	115									115
54 経済団体	708	65	33	37	15	6	1		1	866
62 国営企業	1					1		1	4	7
63 公共団体、公社、公団	1									1
72 その他の宗教団体	1									1
90 その他の法的形態	138	37	46	134	199	142	53	18	2	769
合計	8422	65609	57469	57616	26742	7948	1806	655	159	226426

注) 濃い網掛けは1万社以上、薄い網掛けは1千社以上。

表5は事業所数規模階級別従業者数規模階級別企業数を示している。わが国では統計に

における事業所の要件に「財又はサービスの生産と供給が、人及び設備を有して、継続的に行われていること」があるので、必ず事業所数 \leq 従業者数の関係が成り立つはずである。フィンランドでも、一部例外があるようであるが、ほぼ成り立っているようである。相対的に高い規模階級を見ると、同じ従業者数規模階級でも事業所数がかなりばらついている。カテゴリーデータである2項目間の関連性の強さを表わすクラメールの連関係数³が0.244と低いことから、従業者数規模が大きい企業が事業所数規模も大きいとは限らないことがわかる。

表5 事業所数規模階級別、従業者数規模階級別企業数

	従業者数規模階級										合計
	0-4	5-9	10-19	20-49	50-99	100-199	200-499	500-999	1000-		
0	29	6	3	2	1						41
1-9	193835	16448	8397	4800	1390	616	334	89	33		225942
10-29			3	28	52	56	70	19	13		241
30-39			1	2	4	14	36	12	12		81
40-49					2	4	14	6	12		38
50-59						1	6	8	3		18
60-69					1		3	3	4		11
70-79						3	3	8	2		16
80-89							1	1	2		4
90-99								1	3		4
100-199					1			2	2		5
200-299								8	8		16
300-399						1		1			2
400-499									3		3
500-599									1		1
600-699									1		1
700-799									0		0
800-899									2		2
合計	193864	16454	8404	4832	1451	695	467	158	101		226426

表6は事業所数規模階級別、売上高規模階級別企業数を示している。売上高規模階級が一番低い階級(0~999ユーロ)に事業所数が多い企業がある。参考までに売上高規模階級1(0~999ユーロ)について事業所数規模階級別法的階級別企業数を見ると、事業所数が多い企業の法的形態は主に有限責任会社である。同階級について事業所数規模階級別産業分類5桁別企業数を見ると、事業所数が多い5桁産業は「貯蓄銀行による金融仲介業」(65120 Monetary intermediation by deposit banks)、「非生命保険会社」(66031 Non-life insurance companies)である。したがって、売上高が非常に小さいにもかかわらず、事業所数が多い企業は金融・保険業であること、この産業の売上高にはアウトプットの一部しか計上されていないと推察される。

表6に戻ると、売上高規模階級2(1,000~39,999ユーロ)から5(400,000~1,999,999

³ クラメールの連関係数は $\sqrt{\frac{\chi^2}{n(\min[k,l]-1)}}$ で計算した。 χ^2 はカイ自乗値。 n はデータの個数、 k,l はクロス集計表の列数と行数である。

ユーロ) までは傘下事業所数はほとんどの企業で 9 カ所以下である。それより高い規模階級を見ると、同じ売上高規模階級でも事業所数はかなりばらついている。クラメールの連関係数は表 5 よりさらに低く 0.182 である。このことから売上高規模と事業所数規模の関係性は弱いことがわかる。

表 6 事業所数規模階級別、売上高規模階級別企業数

	売上高規模階級									合計
	0-999	1,000-39,999	40,000-99,999	100,000-399,999	400,000-1,999,999	2,000,000-9,999,999	10,000,000-39,999,999	40,000,000-199,999,999	200,000,000-	
0	2	13	6	7	5	8				41
1-9	8377	65596	57463	57609	26732	7878	1712	503	72	225942
10-29	19				5	53	59	76	29	241
30-39	15					6	14	33	13	81
40-49	1					1	10	15	11	38
50-59	1						3	10	4	18
60-69						1	1	4	5	11
70-79	2						5	4	5	16
80-89	1							2	1	4
90-99									4	4
100-199	1					1			3	5
200-299	2						1	7	6	16
300-399							1		1	2
400-499	1							1	1	3
500-599									1	1
600-699									1	1
700-799									0	0
800-899									2	2
合計	8422	65609	57469	57616	26742	7948	1806	655	159	226426

3. 規模別開設率

ビジネスレジスターの精度向上にはプロファイリングが欠かせない。プロファイリングとは行政記録情報等から入手した開業・開設、廃業・閉鎖、規模拡大・縮小などの企業組織構造の変化に関する情報を確認することである。その確認方法は、米国センサス局のように紙媒体の調査票によるものや、カナダ統計局のように電話やインターネットによるものや、英国統計局のように両方を実施するなどさまざまである。共通しているのは、全部の事業者について毎年確認することは不可能であるので、対象を絞って行っていることである。そして現在の我が国で問題になっているのは、その対象をいかに絞るかである。

そこでフィンランド統計局のマイクロデータを用いて、対象企業を選定する方法を検討するための実験を行う。フィンランド統計局のマイクロデータには事業所の開設年月の情報があるので、開設事業所を特定することはできる。ただし、閉鎖した事業所のデータはない。そこで企業組織構造の変化の有無ではなく、開設事業所の有無について分析する。このとき開設事業所とは過去 2 年以内（2000 年 1 月以降）に開設した事業所とする。

まず規模別開設率を観察する。ここでは開設率とは、開設事業所が有った企業の割合と定義する。表 7 は事業所数規模階級別、従業者数規模階級別開設率である。事業所数規模階級で見ると規模が大きい方が開設率も高い傾向が観察される。従業者数規模階級別に見ると、一番小さい階級（0～4 人）を除けば、同様に規模が大きい方が開設率も高い傾向が

観察される。表 8 は事業所数規模階級別、売上高規模階級別開設率である。売上高規模階級別に見ると、下から 5 番目の階級（400,000～1,999,999 ユーロ）以降は、規模が大きい方が開設率も高い傾向が観察されるが、それより下の階級ではむしろ逆に傾向が観察される。

表 7 事業所数規模階級別、従業者数規模階級別開設率

	従業者数規模階級									合計
	0-4	5-9	10-19	20-49	50-99	100-199	200-499	500-999	1000-	
0	0%	0%	0%	0%	0%					0%
1-9	18%	11%	12%	16%	20%	26%	28%	27%	27%	17%
10-29			67%	75%	79%	57%	66%	63%	46%	66%
30-39			100%	100%	100%	79%	78%	83%	75%	80%
40-49					50%	100%	86%	100%	75%	84%
50-59						100%	83%	100%	67%	89%
60-69					100%		100%	100%	75%	91%
70-79						100%	100%	88%	50%	88%
80-89							100%	100%	100%	100%
90-99								100%	100%	100%
100-199					100%			100%	100%	100%
200-299								100%	88%	94%
300-399						100%		100%		100%
400-499									100%	100%
500-599									100%	100%
600-699									100%	100%
700-799									100%	100%
800-899									100%	100%
合計	18%	11%	12%	16%	23%	31%	41%	53%	59%	17%

表 8 事業所数規模階級別、売上高規模階級別開設率

	売上高規模階級									合計
	0-999	1,000-39,999	40,000-99,999	100,000-399,999	400,000-1,999,999	2,000,000-9,999,999	10,000,000-39,999,999	40,000,000-199,999,999	200,000,000-	
0	0%	0%	0%	0%	0%	0%				0%
1-9	21%	24%	17%	13%	11%	14%	21%	26%	26%	17%
10-29	53%				60%	85%	68%	63%	48%	66%
30-39	73%					100%	93%	76%	77%	80%
40-49	100%					100%	80%	87%	82%	84%
50-59	100%						100%	80%	100%	89%
60-69						100%	100%	100%	80%	91%
70-79	100%						80%	100%	80%	88%
80-89	100%							100%	100%	100%
90-99									100%	100%
100-199	100%					100%			100%	100%
200-299	100%						100%	100%	83%	94%
300-399							100%		100%	100%
400-499	100%							100%	100%	100%
500-599									100%	100%
600-699									100%	100%
700-799									100%	100%
800-899									100%	100%
合計	21%	24%	17%	13%	11%	15%	24%	37%	52%	17%

4. プロビット・モデルによるプロファイリング対象の選定

規模別開設率の観察からは、事業所数規模階級と開設率との関係性は明確であった。そこで、開設事業所の有無を被説明変数とし、事業所数規模階級と従業者規模階級を説明変数とする。また、従業者数規模階級については、一番下の階級（0～4人）を除けば、開設率との関係性は明確であった。このように従業者数規模が非常に小さい企業のプロファイリングは費用の面で割り合わないので、行政記録から得られた企業組織構造の変化に関する情報をそのままビジネスレジスターに反映することが国際的になされている。そこで、従業者数規模階級が一番下の階級は分析の対象外とする。

また、欧米諸国で行われている実際のプロファイリングでは、従業者数あるいは売上高の規模の大きい企業を、毎年、プロファイリングするのが一般的である。そこで従業者数規模あるいは売上高規模において一番上の階級に区分される企業は、優先的にプロファイリングの対象となると想定する。ちなみにこれに該当する企業は191社である。

以上より、開設事業所の有無を被説明変数とし、事業所数規模階級と従業者規模階級を説明変数とし、従業者数規模階級の一番下の階級と一番上の階級を除き（10人以上999人以下）かつ売上高規模階級の一番上の階級を除いた（1,999,999ユーロ以下）の企業を分析対象とする。事業所数規模階級を説明変数とするのは、事業所数が多い企業は、開設に関するノウハウがあるので、開設しやすいと考えるからである。従業者数規模階級を説明変数とするのは、従業者数が多い企業は、開設に必要な人員に余裕があると考えられるからである。

開設事業所の変化の有無は二項選択であるので、プロビット・モデルを適用する。なおプロビット・モデルの適用についてはColin & Triverdi[1998]を参照した。モデルは次の通りである。

$$Y_i^* = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$
$$Y_i = \begin{cases} 1 & Y_i^* > 0 \\ 0 & Y_i^* \leq 0 \end{cases}$$

ここで Y_i は開設事業所の変化の有無、 X_{2i} は従業者数規模階級、 X_{3i} は事業所数規模階級、 ε_i は誤差項である。このモデルに誤差項の分布として標準正規分布を仮定したのがプロビット・モデルである。表7はプロビット・モデルの推定結果である。サンプルの大きさが32,374社と大きいので、z統計量のp値はすべて0.000%である。マクファーデンの決定係数(McFadden R-squared)が0.035677とかなり低い。従業者数規模階級(PERSONNEL)、事業所数規模階級(ESTAB)の係数のいずれも正值であり、これは従業者数が多い、あるいは事業所数が多い企業の開設率が高いことを示しており、妥当である。

モデルの当てはまり(表8)を見ると、正しく判定した割合は高い(0.6%+86.3%=86.9%)。開設事業所が有ると判定した企業253社のうち、74.7%(189社)は実際に開設事業所があったことになり、これに従ってプロファイリングした場合の効率は良い。だが、実際には開設事業所が有った企業4,378社のうち4.3%(189社)しか、「開設事業所有り」と判定されておらず、大半の企業が漏れてしまう。

表7 プロビット・モデルの推定結果(モデル1)

Dependent Variable: NEW_ESTAB				
Method: ML - Binary Probit (Quadratic hill climbing)				
Date: 06/02/14 Time: 15:36				
Sample: 1 32374				
Included observations: 32374				
Convergence achieved after 5 iterations				
Covariance matrix computed using second derivatives				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-3.104278	0.107489	-28.88002	0.0000
ESTAB	0.798036	0.054970	14.51773	0.0000
PERSONNEL	0.129989	0.007461	17.42184	0.0000
McFadden R-squared	0.035677	Mean dependent var	0.135232	
S.D. dependent var	0.341976	S.E. of regression	0.334677	
Akaike info criterion	0.764339	Sum squared resid	3625.823	
Schwarz criterion	0.765116	Log likelihood	-12369.36	
Hannan-Quinn criter.	0.764588	Deviance	24738.72	
Restr. deviance	25653.99	Restr. log likelihood	-12826.99	
LR statistic	915.2645	Avg. log likelihood	-0.382077	
Prob(LR statistic)	0.000000			
Obs with Dep=0	27996	Total obs	32374	
Obs with Dep=1	4378			

表8 プロビット・モデルの当てはまり(モデル1)

		判定		
		有り	無し	計
実際	有り	189	4189	4378
	無し	64	27932	27996
	計	253	32121	32374

		判定		
		有り	無し	計
実際	有り	4.3%	95.7%	100.0%
	無し	0.2%	99.8%	100.0%
	計	0.8%	99.2%	100.0%

		判定		
		有り	無し	計
実際	有り	74.7%	13.0%	13.5%
	無し	25.3%	87.0%	86.5%
	計	100.0%	100.0%	100.0%

		判定		
		有り	無し	計
実際	有り	0.6%	12.9%	13.5%
	無し	0.2%	86.3%	86.5%
	計	0.8%	99.2%	100.0%

最初のモデルでは、実際のプロファイリングに適用するには不十分であるので、追加すべき説明変数を検討する。企業の開業年と、傘下の事業所の開設率を見ると、2000年以降に開業した企業の開設率が70%と他の年次と比較して非常に高い（表9）。1999年以前に開業した企業について見ると、新しい企業の開設率が高いとは必ずしも言えない。したがって、2000年以降に開業した企業に着目して説明変数を追加することが考えられる。

表9 企業の開業年と傘下の事業所の開設率

開業年	分析対象企業数	内開設事業所が 有った企業数	開設率
1900-1909	12	3	25%
1910-1919	33	6	18%
1920-1929	35	1	3%
1930-1939	33	5	15%
1940-1949	118	23	19%
1950-1959	215	26	12%
1960-1969	760	64	8%
1970-1979	3870	224	6%
1980-1989	9420	625	7%
1990-1999	15041	1405	9%
2000-	2837	1996	70%
計	32374	4378	14%

以上を踏まえて2000年以降に開業した企業を1、1999年以前に開業した企業を0とするダミー変数を説明変数として追加してモデルを修正した。

$$Y_i^* = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 D_i + \varepsilon_i$$

$$Y_i = \begin{cases} 1 & Y_i^* > 0 \\ 0 & Y_i^* \leq 0 \end{cases}$$

ただし、 D_i はダミー変数である。このモデルの推定結果（表10）を見ると、上のモデルと同様にサンプルの大きさが32,374社と大きいため、z統計量のp値はすべて0.000%である。マクファーデンの決定係数は0.269020とかなり改善している。従業員数規模階級（PERSONNEL）、事業所数規模階級（ESTAB）、ダミー変数（START_2000S）の係数のどれも正值であるが、これは従業員数が多い、あるいは事業所数が多い、あるいは最近開業した企業の開設率が高いことを示している。

モデルの当てはまり（表11）を見ると、正しく判定した割合は高い（6.6%+83.8%=90.4%）。開設事業所があると判定した企業3,005社のうち、70.9%（2,130社）は実際に開設事業所があったことになり、これに従ってプロファイリングした場合の効率は良い。また、実際には開設事業所があった企業4,378社のうち48.7%（2,130社）が、「開設事業所有り」と判定されており、約半分がカバーされている。

表 10 プロビット・モデルの推定結果 (モデル2)

Dependent Variable: NEW_ESTAB				
Method: ML - Binary Probit (Quadratic hill climbing)				
Date: 06/02/14 Time: 22:31				
Sample: 1 32374				
Included observations: 32374				
Convergence achieved after 5 iterations				
Covariance matrix computed using second derivatives				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-3.657472	0.110877	-32.98684	0.0000
ESTAB	0.820630	0.056479	14.52975	0.0000
PERSONNEL	0.189582	0.008119	23.35142	0.0000
START_2000S	2.017117	0.027350	73.75304	0.0000
McFadden R-squared	0.269020	Mean dependent var	0.135232	
S.D. dependent var	0.341976	S.E. of regression	0.285233	
Akaike info criterion	0.579494	Sum squared resid	2633.545	
Schwarz criterion	0.580530	Log likelihood	-9376.272	
Hannan-Quinn criter.	0.579825	Deviance	18752.54	
Restr. deviance	25653.99	Restr. log likelihood	-12826.99	
LR statistic	6901.444	Avg. log likelihood	-0.289624	
Prob(LR statistic)	0.000000			
Obs with Dep=0	27996	Total obs	32374	
Obs with Dep=1	4378			

表 11 プロビット・モデルの当てはまり (モデル2)

		判定		
		有り	無し	計
実際	有り	2130	2248	4378
	無し	875	27121	27996
	計	3005	29369	32374

ヨコ構成比		判定		
		有り	無し	計
実際	有り	48.7%	51.3%	100.0%
	無し	3.1%	96.9%	100.0%
	計	9.3%	90.7%	100.0%

タテ構成比		判定		
		有り	無し	計
実際	有り	70.9%	7.7%	13.5%
	無し	29.1%	92.3%	86.5%
	計	100.0%	100.0%	100.0%

総計に対する構成比		判定		
		有り	無し	計
実際	有り	6.6%	6.9%	13.5%
	無し	2.7%	83.8%	86.5%
	計	9.3%	90.7%	100.0%

プロビット・モデルでは、開設確率を

$$P = F(Y^*) = F(\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 D_i)$$

と推定することができる。Fは正規分布の累積分布関数である。表12はプロビット・モデルによって推定した開設確率である。開設確率が高い企業から順番にプロファイリングすることにより、少ない費用で企業組織構造の変化を把握できると考えられる。

表 12 プロビット・モデルによる開設確率

企業の開業年が1999年以前の場合		従業者数規模階級						
		5-9	10-19	20-49	50-99	100-199	200-499	500-999
事業 所 数 規 模 階 級	0	1%	1%	2%	3%	4%	7%	9%
	1-9	5%	7%	10%	14%	19%	25%	31%
	10-29	21%	27%	33%	40%	48%	55%	63%
	30-39	50%	58%	65%	72%	78%	83%	87%
	40-49	80%	84%	89%	92%	94%	96%	98%
	50-59	95%	97%	98%	99%	99%	100%	100%
	60-69	99%	100%	100%	100%	100%	100%	100%
	70-79	100%	100%	100%	100%	100%	100%	100%
	80-89	100%	100%	100%	100%	100%	100%	100%
	90-99	100%	100%	100%	100%	100%	100%	100%
	100-199	100%	100%	100%	100%	100%	100%	100%
200-299	100%	100%	100%	100%	100%	100%	100%	

企業の開業年が2000年以降の場合		従業者数規模階級						
		5-9	10-19	20-49	50-99	100-199	200-499	500-999
事業 所 数 規 模 階 級	0	33%	40%	48%	55%	62%	69%	76%
	1-9	65%	72%	78%	83%	87%	91%	94%
	10-29	89%	92%	94%	96%	97%	98%	99%
	30-39	98%	99%	99%	100%	100%	100%	100%
	40-49	100%	100%	100%	100%	100%	100%	100%
	50-59	100%	100%	100%	100%	100%	100%	100%
	60-69	100%	100%	100%	100%	100%	100%	100%
	70-79	100%	100%	100%	100%	100%	100%	100%
	80-89	100%	100%	100%	100%	100%	100%	100%
	90-99	100%	100%	100%	100%	100%	100%	100%
	100-199	100%	100%	100%	100%	100%	100%	100%
200-299	100%	100%	100%	100%	100%	100%	100%	

5. おわりに

ビジネスレジスターのマイクロデータを用いた分析を応用することにより、企業組織構造に変化があった企業を効率的に見つけることができる可能性がある。実施のプロファイリングでは、新聞記事等の情報も活用できるわけであり、もっと正確に組織構造に変化があった企業を見つけることができるであろう。

参考文献

- Cameron, A. Colin, Pravin K. Trivedi[1998] *Regression Analysis of Count Data*, Econometric Society Monograph No.30, Cambridge University Press
- Kinney, Satkartar K., Jerome P. Reiter[2013] *SynLBD: Providing firm characteristics on synthetic establishment data*, Proceedings of the 59th World Statistics Congress of the International Statistical Institute, 2013