An empirical analysis of endogenous growth without scale effect*

Francesco Schettino†

May 31, 2005

Abstract

The aim of this paper is to empirically examine the endogenous growth without scale effect as in Segestrom (1998). We firstly build a concordance table between the 1972 Standard Industrial Classification System (SIC) and USPCS (as of December, 31 2002) industrial codes using the USPTO concordance new file. Then we approach the analysis of the difficulty index of each industry starting from the NBER patent data set (63_99) using the mean forward citations lag data as the fundamental variable of our analysis. In fact we follow the idea that each patent number of citations grows untill a point from which it starts to decade. We explain that point as the moment that draws the obsolescence of the invention. Once we have eliminated by that way the lag truncations problem we obtain the difficult index serie of each industrial sector. Finally we describe the relationship between the growth of the difficulty index and the number of scientists and engeneers. Our investigation concludes showing a difference between traditional and new sectors where in the latters the absence of the scale effect is more clear than in the others.

J.E.L. Classification: O340, O140, O410
Keywords: Patents, Endogenous Growth, Scale effect

*I thank G.Cozzi, M.Giammatteo and M.Tancioni for helpful comments.

†Department of Public Economics, University of Rome, 'La Sapienza', via del Castro Lurenziano, 9 - 00161 - Rome. Email: Francesco.Schettino@uniroma1.it; francesco.schettino@gmail.com - tel: +39.06.49766843

1 Introduction

The idea of the endogenous growth with scale effect was born in the first articles of the neoschumpeterian framework as in Romer (1990), Grossman-Helpman (1991) and Aghion-Howitt(1992). They underline that an increasing number of scientists and engeneers (from now on S&E) could determine a proportional increase of the whole economy growth. Segestrom (1998) has gone against this ideas and shows that empirically never has been this direct corrispondence. In fact in the U.S. the S&E number growth has never corrisponded to the rate of growth of the U.S. economy. That is why he introduces a new element, the *difficulty index*, that determines a less than proportional relation between the number of S&E and the growth of the economy. No empirical evaluations of these fundamental intuition have been carried out so far. That is why we first need to develop a new concordance table between SIC and USPCS industry codes using the NBER concordance software. Then we have been able to proceed and we define empirically the *difficulty index* that finally permits us to evaluate the relation between the rate of growth of the difficulty index and the probability of innovative success for each industry.

The article is organized as follows: in the first section we use the USPTO file obtaining a concordance table between USPCS (as December, 31 2002) and 1972 Standard Industrial Classification System (SIC) codes as the necessary methodological tool used to work on the original patent data file (NBER). In the second section we develop the idea on how to evaluate the difficulty index using the "NBER 63_99" data file, giving to the mean forward lag citations the principal importance. Thus, we obtain the historical serie of the difficulty index by industry. In the last section, finally, we find the stable relation between the difficulty index of each industry and the corrispondent probability of innovative success.

2 Methodological Tools: the SIC-USPCS Concordance

Hirabayashi (2003) shows the historical great difficulty which any researcher gots into when he has to compare the available economic data of patents with the other economic variables cause the different industrial codes they are classified. Our goal is to show the relation between the evolution of the difficulty index (that we derive from the NBER data file as we observe in section 3), classified in USPCS, and the number of the S&E of each sector, classified by NSF in SIC (1972); that is why we have reputed fundamental build a concordance table between the two codes starting from the USPTO software[1]. In table 1 we present the results.

3 Evaluating the Difficulty Index

Once we have found a concordance table between SIC and USPCS codes we are able to proceed further. In this section we want to emprically evaluate the R&D races as in Segerstrom (1998). In this part of the model he defines the innovative success probability at time t as:

$$I(\omega, t) = \frac{AL_I(\omega, t)}{X(\omega, t)}, \tag{1}$$

where L_I is the number of S&E in the whole economy A is a given exogenous technological parameter and X is the *difficulty index* . Moreover he sustains that as industries do more R&D the difficulty index increase as:

$$\frac{\dot{X}(\omega, t)}{X(\omega, t)} = \mu I(\omega, t). \tag{2}$$

Substituing (1) in (2), and considering $A = 1$ without loosing genrality, we obtain the following relation that we want to test:

$$\dot{X}(\omega, t) = \mu L_I(\omega, t). \tag{3}$$

At this point of our analysis, to go further we need to define more than conceptually $X(\omega, t)$. In fact Segerstrom (1998) describes it as function of time and industry (ω). Hall, Jaffe and Traitenberg (2003) underline that the only instrument we have to quantify to innovation trend is the patent data because actually it is impossible to evaluate the number of the not patented innovation. That is why we use the NBER data file (PAT63_99) to observe the difficulty index evolution in the last 30 years. The data on the mean forward citation lag has been fundamental in our work. In fact our idea is based on how each patent is citated for the x periods of its "life". More

[1]Available on line at ftp://ftp.uspto.gov/pub/taf/sic_conc/;

Table 1: Concordance table SIC to USPCS

SECTOR	SIC	USPCS
Food, kindred and tobacco products	**20,21**	83 127 205 426 428
Textiles and apparel	**22,23**	8 16 24 57 66 87 114 119 139 152 181 182 205 241 242 264 405 427 428 435 442 474 492
Chemicals and allied products	**28**	2 4 5 8 12 15 16 23 24 29 34 36 40 44 47 48 49 51 52 59 62 65 71 75 81 95 102 104 106 108 110 114 116 119 126 127 128 131 134 135 137 138 148 149 150 152 156 160 164 165 166 168 174 175 181 188 201 203 204 205 206 208 215 220 221 222 223 224 228 229 238 239 242 246 248 249 251 252 256 260 264 267 277 280 285 294 295 301 335 340 349 359 376 383 384 385 403 405 411 416 420 422 423 424 427 428 429 431 435 436 441 451 454 464 474 482 492 501 502 504 507 508 510 512 514516 518 520 521 522 523 524 525 526 527 528 530 532 534 536 540 544 546 548 549 552 554 556 558 560 562 564 568 570 585 588 604 607 800 930 968 976 987
Petroleum refining and extraction	**13, 29**	44 48 166 175 204 205 208 340 428 435 508
Rubber products	**30**	2 4 5 8 12 15 16 24 29 36 40 47 49 52 59 62 81 106 108 114 116 119 126 128 135 137 138 150 152 156 160 165 168 181 188 204 205 206 215 220 221 222 223 224 229 238 239 242 248 251 256 264 267 280 285 294 301 383 384 403 411 416 422 427 428 429 441 474 482 492 521 523 524 525 527 604 968
Stone, clay, and glass products	**32**	4 8 15 29 40 47 51 52 65 106 110 119 126 131 138 156 166 174 181 188 205 215 220 222 238 239 242 251 256 264
Primary metals	**33**	29 59 75 104 138 148 156 164 166 174 188 205 228 238 246 249 264 295 385 411 416 420 427 428 464
Fabricated metal products	**34**	2 4 5 7 14 15 16 24 28 29 30 37 38 40 43 47 49 52 54 56 59 62 69 70 72 75 76 79 81 99 104 105 109 110 111 114 116 119 122 125 126 131 134 135 137 138 140 141 144 148 156 160 165 166 168 172 180 181 182 185 186 188 193 204 205 206 211 215 220 221 222 223 224 232 237 238 239 242 244 245 246 248 249 250 251 254 256 258 261 267 269 280 285 289 292 293 294 295 296 297 300 301 310 312 359 376 403 404 405 407 410 411 413 414 416 419 427 428 431 441 454 464 474 482 492 968 976
Machinery	**35**	4 12 15 19 26 27 28 29 30 34 37 38 40 43 47 48 52 53 55 56 57 59 60 62 65 66 68 69 72 73 74 76 79 81 82 83 87 91 92 96 99 100 101 104 105 108 110 111 112 114 116 117 118 119 122 123 125 126 127 131 134 137 138 139 140 141 142 144 147 156 157 159 162 163 164 165 166 169 171 172 173 174 175 177 180 181 182 184 185 186 187 188 192 193 194 196 198 199 202 204 205 206 209 210 211 212 213 219 221 222 223 225 226 227 228 231 234 235 237 239 241 242 244 246 248 249 250 251 254 261 264 266 267 269 270 271 276 277 278 279 280 289 290 293 294 298 299 300 303 305 307 312 335 341 345 346 347 349 356 358 360 366 369 376 380 382 384 386 388 392 395 400 403 404 405 406 407 408 409 412 413 414 415 416 417 418 422 425 427 428 432 435 440 445 451 452 453 454 460 464 470 474 475 476 477 483 492 493 494 505 526 700 701 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 901 902 968 976
Electrical equipment	**36**	15 29 30 34 38 49 52 62 68 73 83 91 96 99 104 112 116 117 118 119 123 125 126 128 134 136 140 148 156 160 165 174 178 181 187 188 191 192 200 204 205 206 211 216 218 219 220 221 222 228 236 238 241 242 244 246 250 256 257 261 264 267 290 307 310 312 313 314 315 318 320 322 323 324 326 327 329 330 331 332 333 334 335 336 338 340 341 342 343 345 346 348 349 356 358 359 360 361 362 363 365 366 367 369 370 372 373 375 377 378 379 380 381 385 386 388 392 414 416 422 427 428 429 431 434 438 439 445 451 455 477 502 505 600 601 607 700 704 706 714 725 976
Tranportation equipment	**37**	14 15 29 37 42 49 52 60 62 73 74 86 89 91 102 104 105 110 114 116 119 123 124 126 135 137 152 157 160 165 166 169 175 180 181 182 184 186 188 191 192 213 219 220 237 239 242 244 246 254 258 264 267 278 280 291 293 295 296 297 298 301 303 305 307 315 359 384 405 406 414 416 428 431 434 439 440 441 454 464 474 475 476 477 505 701 706 976

simply, as we show in figure 1, generally the distributional form of the mean forward citation lag in each sector and in each year is firstly increasing and after decreases.

In that way it shows constantly a peak (the mode of the distribution) that we want to analyze. We think that in the increasing citation side of the distribution graph the patent hasn't reached the time of its obsolescence: the side after the peak (mode), shows that the patent has been substituded by another innovation because its citations decline. Thus, we cumulate the frequencies of the citation (weighted) untill the peak and we define them as the denominator of the difficulty index (from now on S). The idea is, in fact, that greater is the percentual of the scientists that work using that patent untill it reaches its peak, more difficult is its substitution.

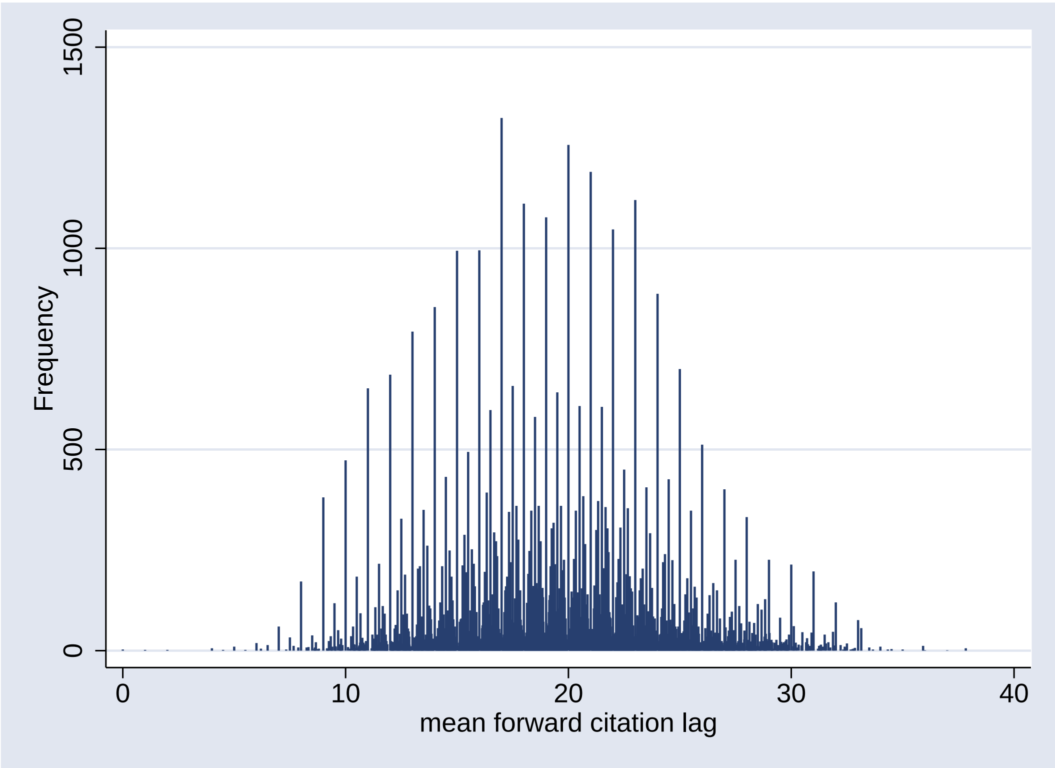

Figure 1: Distribution of the mean lag citation forward

It's clear that here lives the problem of the *truncation lag* as in Hall, Jaffe and Traitenberg (2003) as we observe in figure 2 where we report the different mean forward citations lag weighted distribution modes. In fact the mean forward citation lag become lower and lower when we proceed through the last year of the data base (1999). We solve this problem in two ways. We firstly esclude the last 5 years from the statistics and then we normalize S dividing it with the square of the number of year lags:

$$X(\omega, t) = \frac{S}{y^2}. \tag{4}$$

Thus we obtain the historical serie (see table 2) of the difficulty index for each industry.

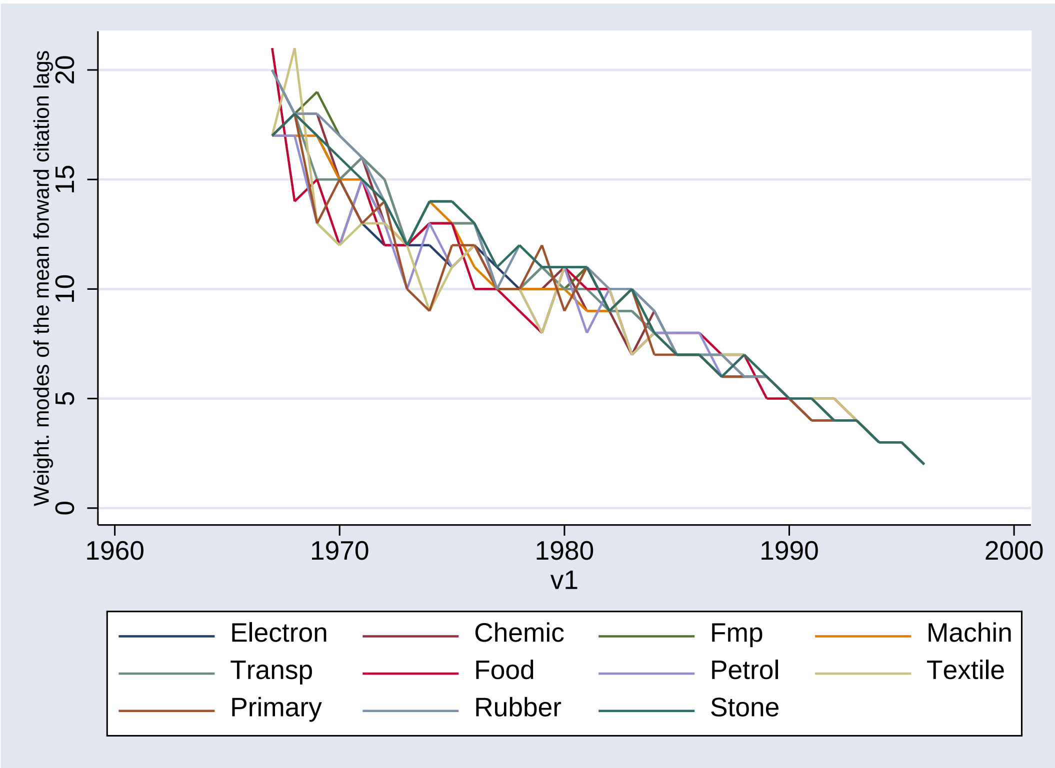

Figure 2: Weighted modes of the mean forward lag citation

Table 2: Difficulty index serie

	Electron	Chemic	Fmp	Machinery	Transport	Food	Petroleum	Textile	Primary_met	Rubber	Stone_clay
1967	0.10	0.13	0.12	0.10	0.13	0.15	0.11	0.11	0.10	0.11	0.09
1968	0.15	0.13	0.13	0.13	0.14	0.13	0.14	0.16	0.14	0.12	0.13
1969	0.16	0.16	0.16	0.17	0.13	0.18	0.12	0.12	0.10	0.15	0.15
1970	0.18	0.16	0.18	0.17	0.20	0.17	0.13	0.12	0.17	0.17	0.17
1971	0.19	0.20	0.19	0.21	0.21	0.22	0.23	0.18	0.17	0.18	0.19
1972	0.23	0.24	0.24	0.25	0.25	0.27	0.26	0.25	0.26	0.22	0.24
1973	0.31	0.28	0.25	0.30	0.28	0.33	0.29	0.31	0.26	0.25	0.28
1974	0.34	0.32	0.31	0.34	0.33	0.35	0.36	0.29	0.28	0.30	0.32
1975	0.35	0.33	0.33	0.35	0.40	0.35	0.35	0.32	0.32	0.31	0.33
1976	0.39	0.37	0.36	0.36	0.38	0.37	0.40	0.35	0.37	0.35	0.37
1977	0.43	0.38	0.36	0.39	0.40	0.43	0.43	0.36	0.38	0.33	0.38
1978	0.45	0.40	0.37	0.41	0.43	0.45	0.44	0.37	0.40	0.40	0.43
1979	0.49	0.45	0.45	0.46	0.50	0.48	0.46	0.37	0.50	0.43	0.46
1980	0.51	0.49	0.44	0.48	0.49	0.54	0.54	0.49	0.43	0.46	0.49
1981	0.59	0.51	0.55	0.51	0.55	0.59	0.54	0.55	0.58	0.53	0.55
1982	0.63	0.58	0.55	0.60	0.60	0.65	0.65	0.61	0.60	0.58	0.57
1983	0.70	0.50	0.67	0.70	0.67	0.72	0.58	0.49	0.70	0.65	0.67
1984	0.74	0.77	0.76	0.70	0.73	0.85	0.82	0.71	0.65	0.73	0.71
1985	0.80	0.76	0.74	0.77	0.83	0.91	0.90	0.79	0.82	0.69	0.77
1986	0.95	0.91	0.94	0.93	1.02	1.13	1.06	0.96	1.02	0.86	0.95
1987	1.00	0.93	1.19	0.95	1.09	1.23	0.84	1.04	1.02	1.12	0.97
1988	1.29	1.23	1.33	1.23	1.40	1.41	1.22	1.25	1.25	1.23	1.38
1989	1.64	1.52	1.64	1.53	1.72	1.42	1.37	1.43	1.62	1.53	1.58
1990	1.85	1.69	1.89	1.73	2.02	1.89	1.47	1.53	1.82	1.70	1.75
1991	1.80	2.25	2.47	2.28	2.02	2.30	2.00	2.05	1.85	2.31	2.31
1992	2.64	2.52	2.72	2.43	2.93	2.80	1.97	2.59	2.60	2.51	2.46
1993	3.74	3.69	3.88	3.46	3.96	3.60	3.06	3.29	3.68	3.74	3.61
1994	4.52	4.79	4.96	4.33	5.01	4.86	3.97	4.34	4.69	4.85	4.66
1995	7.46	7.18	7.43	7.06	7.43	7.19	6.86	6.88	7.40	7.34	7.13
1996	12.00	11.48	11.55	11.72	11.94	12.98	11.24	11.43	12.22	11.67	11.30

4 Relationships Between Difficulty Index Growth and the Probability of Innovation

Once we have defined the difficulty index as in table 2 we are interested to analyze the evolution year after year. As we sustain in the Introduction and in section 3, we prefer use the data 67_93. Thus we regress $X_{t+1} - X_t$ on $L_I(\omega, t)$ in each manufacturing sector. We use the $L_I(\omega, t)$ serie created in SIC code of the NSF. We use the simply ordinary least squared to estimate the parameter μ. The results are summarized in table 3:

Table 3: Each sector's μ

SIC Sectors	μ
Electrical Equipments	**0.011421**
Chemical and Allied Products	**0.013431**
Machinery	**0.016793**
Rubber Products	**0.039677**
Petroleum refining and extraction	**0.073715**
Fabricated metal products	**0.118065**
Food, kindred an tobacco products	**0.170971**
Primary Metals	**0.131779**
Stone, Clay and Glass products	**0.208046**
Textile and Apparel	**0.555105**

Clearly, the estimation outputs are not exceptional but we are regressing on an indipendent variable only with one dipendent variable, without any constant. The output tables are the following:

Table 4: Chemical and Allied Products

Dependent Variable: Xdot				
Method: Least Squares				
Date: 05/18/05 Time: 16:37				
Sample: 1969 1991				
Included observations: 23				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
L	0.013431	0.003139	4.27921	0.0003
R-squared	0.185454	Mean dependent var	0.67887	
Adjusted R-squared	0.185454	S.D. dependent var	0.989059	
S.E. of regression	0.892647	Akaike info criterion	2.653255	
Sum squared resid	17.53003	Schwarz criterion	2.702624	
Log likelihood	-29.51243	Durbin-Watson stat	2.074821	

Table 5: Petroleum refining and extraction

Dependent Variable: X				
Method: Least Squares				
Date: 05/18/05 Time: 16:32				
Sample: 1969 1991				
Included observations: 23				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
L	0.073715	0.029036	2.538797	0.0187
R-squared	0.019342	Mean dependent var	0.764135	
Adjusted R-squared	0.019342	S.D. dependent var	1.50932	
S.E. of regression	1.494652	Akaike info criterion	3.684169	
Sum squared resid	49.14766	Schwarz criterion	3.733538	
Log likelihood	-41.36794	Durbin-Watson stat	2.554235	

Table 6: Primary Metals

Dependent Variable: X				
Method: Least Squares				
Date: 05/18/05 Time: 16:33				
Sample: 1969 1991				
Included observations: 23				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
L	0.131779	0.054782	2.405505	0.025
R-squared	-0.098449	Mean dependent var	1.056612	
Adjusted R-squared	-0.098449	S.D. dependent var	1.735837	
S.E. of regression	1.819277	Akaike info criterion	4.07726	
Sum squared resid	72.81488	Schwarz criterion	4.126629	
Log likelihood	-45.88849	Durbin-Watson stat	1.201854	

Table 7: Electronical equipment

Dependent Variable: X				
Method: Least Squares				
Date: 05/18/05 Time: 16:39				
Sample: 1969 1991				
Included observations: 23				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
L	0.011421	0.003674	3.108616	0.0051
R-squared	0.049898	Mean dependent var	1.089929	
Adjusted R-squared	0.049898	S.D. dependent var	1.838488	
S.E. of regression	1.792033	Akaike info criterion	4.047084	
Sum squared resid	70.65044	Schwarz criterion	4.096453	
Log likelihood	-45.54146	Durbin-Watson stat	1.388479	

Table 8: Fabricated metal products

Dependent Variable: X				
Method: Least Squares				
Date: 05/18/05 Time: 16:29				
Sample(adjusted): 1969 1990				
Included observations: 19				
Excluded observations:				
3 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
L	0.118065	0.03812	3.097189	0.0062
R-squared	0.074661	Mean dependent var	0.924083	
Adjusted R-squared	0.074661	S.D. dependent var	1.467636	
S.E. of regression	1.411786	Akaike info criterion	3.578784	
Sum squared resid	35.87652	Schwarz criterion	3.628491	

Table 9: Machinery

Dependent Variable: X				
Method: Least Squares				
Date: 05/18/05 Time: 16:31				
Sample: 1969 1991				
Included observations: 23				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
L	0.016793	0.003033	5.537161	0
R-squared	0.328133	Mean dependent var	0.991152	
Adjusted R-squared	0.328133	S.D. dependent var	1.299469	
S.E. of regression	1.065142	Akaike info criterion	3.006598	
Sum squared resid	24.95961	Schwarz criterion	3.055968	
Log likelihood	-33.57588	Durbin-Watson stat	1.620574	

Table 10: Rubber Products

Dependent Variable: X				
Method: Least Squares				
Date: 05/18/05 Time: 16:35				
Sample(adjusted): 1969 1981				
Included observations: 12				
Excluded observations:				
1 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
L	0.039677	0.007372	5.38232	2E-04
R-squared	0.130847	Mean dependent var	0.30205	
Adjusted R-squared	0.130847	S.D. dependent var	0.21475	
S.E. of regression	0.200211	Akaike info criterion	-0.29924	
Sum squared resid	0.440928	Schwarz criterion	-0.25883	

Table 11: Stone, Clay and Glass products

Dependent Variable: X				
Method: Least Squares				
Date: 05/18/05 Time: 16:35				
Sample: 1969 1991				
Included observations: 23				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
L	0.208046	0.042701	4.872095	0.0001
R-squared	0.233935	Mean dependent var	1.013876	
Adjusted R-squared	0.233935	S.D. dependent var	1.346627	
S.E. of regression	1.178637	Akaike info criterion	3.2091	
Sum squared resid	30.56209	Schwarz criterion	3.258469	
Log likelihood	-35.90465	Durbin-Watson stat	1.760902	

Table 12: Textile and apparel

Dependent Variable: X				
Method: Least Squares				
Date: 05/18/05 Time: 16:36				
Sample: 1969 1991				
Included observations: 23				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
L	0.555105	0.137888	4.025783	0.0006
R-squared	0.151834	Mean dependent var	1.05059	
Adjusted R-squared	0.151834	S.D. dependent var	1.56192	
S.E. of regression	1.438465	Akaike info criterion	3.607535	
Sum squared resid	45.52201	Schwarz criterion	3.656905	
Log likelihood	-40.48666	Durbin-Watson stat	0.961233	

Table 13: Food Kindred and Tobacco products

Dependent Variable: X				
Method: Least Squares				
Date: 05/18/05 Time: 16:30				
Sample: 1969 1991				
Included observations: 23				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
FOOD	0.170971	0.040366	4.235503	0.0003
R-squared	0.112355	Mean dependent var	1.135912	
Adjusted R-squared	0.112355	S.D. dependent var	1.485298	
S.E. of regression	1.399373	Akaike info criterion	3.55243	
Sum squared resid	43.08136	Schwarz criterion	3.601799	
Log likelihood	-39.85294	Durbin-Watson stat	0.724139	

5 Conclusions

Empirically evaluating the Segerstrom (1998) model, we have found that the only way to estimate the difficulty index is using the mean forward citation lag of the U.S. patents. By that way we have been able to find the relationship between the growth of the difficulty index and the probability of innovative success. Our results confirm Segerstrom (1998) and, moreover, show that there are differencies between manufacturing sector. In fact in the traditional sectors the difficulty index series are greater then in new sectors; consequently the absence of scale effect is more visible in that sectors than in the others as it has confirmed by the stable estimate relation between the growth of the difficulty index and the probability of innovative success.

References

[1] AGHION, P. and HOWITT, P. (1992), A Model of Growth Through Creative Distruction, Econometrica 60, 323-351;

[2] GROSSMAN, M. and HELPMAN, E. (1991), Innovation and Growth in the Global economy, Cambridge, MA: MIT press;

[3] HALL B.H., JAFFE A. B., TRATJTENBERG M. (2001), The NBER Patent Citations Data File: Lessons, Insights And Methodological Tools, NBER Working Paper 8498;

[4] HIRABAYASHI, J. (2003), Presentation in WIPO-OECD Workshop on Statistics in the Patent Field, Geneva, September 2003;

[5] NATIONAL SCIENCE FOUNDATION (2002), Science and Engineering indicators 2002 vol.1 and 2, National Science Board;

[6] ROMER, P. (1990), Human Capital And Growth: Theory and Evidence, Carnegie-Rochester Conference Series on Public Policy: Unit Roots, Investment Measures and Other Essays, Vol. 32, pp. 251-286, Spring 1990;

[7] SEGERSTROM, P. (1998), Endogenous Growth without Scale Effect, AER, Vol. 98, No. 5 (december 1998), 1290-1310.