

국내 특허의 생산함수 추정

이원경 · 김병수

기술혁신은 현대산업사회 발전에 중추적인 기능을 담당하고 있다. 본 연구에서는 이러한 기술혁신의 지표로 특허건수를 사용하였고 연구개발비를 설명변수로 사용하여 특허의 생산함수를 추정하였다. 특허건수는 기본적으로 포아송분포를 이루지만, 포아송모형을 수정하여야 할 과산포현상이 관찰되었다. 또한, 국내의 특허건수는 특허를 1건 이상 등록한 기업에 대해서만 자료수집이 이루어지고 있어 본 논의에서는 0에서 절삭된 포아송분포를 구성하여 이 분포하에서 과산포를 모형에 포함하여 특허의 생산함수를 추정해 보았다.

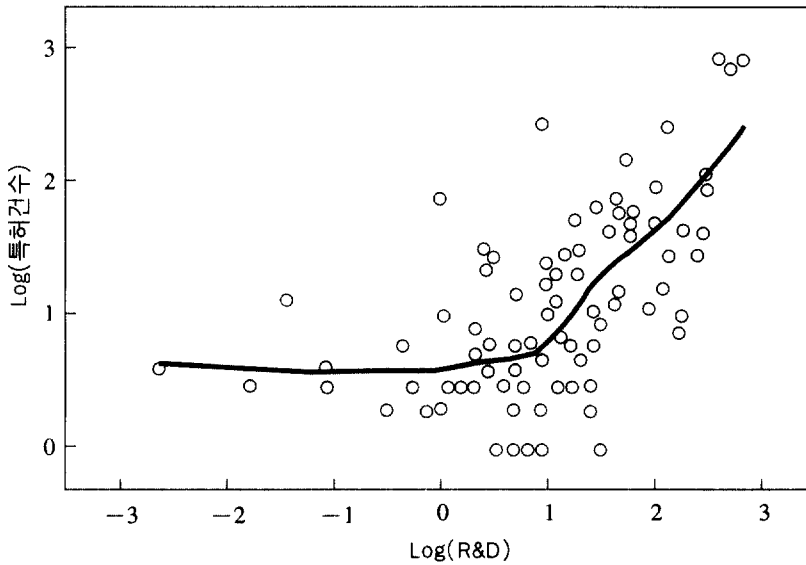
I. 서 론

연구개발과 혁신의 과정을 의미하는 기술혁신은 생산성과 효율성 증대를 통해 기업의 경쟁력과 수익성 확보의 원천이 되며, 결국 사회 전체의 후생을 증진시키는 요소라 할 수 있다. 특히 부존자원이 빈약한 우리 나라의 경우 제품의 고부가가치화를 위해 기술집약적인 첨단산업의 육성이 필요하며 이를 위해서는 지속적인 기술혁신이 이루어져야 한다(정갑영 [2]).

일반적으로 특정산업 또는 전체산업에서 기술혁신 정도를 비교적 정확히 가늠하는 지표로 특허건수, 연구개발투자의 집약도, 생산성의 향상 등이 활용되고 있다. 본 연구에서는 기술혁신의 지표로 특허건수를 사용하였다. <그림 1>은 특허건수와 연구개발비(R&D, 단위: 십억 원)에 대해서 각각 상용로그를 취한 후 $S+$ 를 이용하여 loess방법(Cleveland [4])으로 적합시켜 보았다. <그림 1>에서 볼 수 있듯이 $\log(\text{특허건수})$ 는 $\log(\text{R\&D})$ 의 2차 함수로 나타낼 수 있을 것 같다.

조흥은행 경제연구소 책임연구원, 서울시 중구 삼각동 34번지 조흥은행 백년관 16층, 100-200.
연세대학교 응용통계학과, 서울시 서대문구 신촌동 134번지, 120-749.

〈그림 1〉 국내특허건수와 R&D와의 loess 적합선



한편, 특허건수는 기본적으로 포아송분포를 가지는 자료로서 자료의 이질성 때문에 분산이 평균보다 커지는 과산포현상이 종종 나타난다(Hausman *et al.* [6], Wang *et al.* [7]). 이 점을 고려하지 않을 경우 모수추정에 대한 분산을 과소추정하게 되고 결국 낮은 검정력을 초래한다(김병수 등 [1]).

1995년 12월 말 현재 국내의 5인 이상 제조업체 중에서 기술개발연구를 하는 기업은 9만 2,578개 회사이다. 이 중 특허 또는 실용 건수가 1개 이상인 회사는 2,380社로 전체 가운데 2.57%를 차지하고 있다(특허청, 개인적 교신, 1998). 주목할 것은 국내의 특허건수는 특허를 1건 이상 등록한 기업에 대하여 이루어지고 있다는 점이다. 반면, 외국에서는 특허건수가 0인 기업에 대해서도 R&D나 기타 특허관련 변수에 대한 조사가 이루어지고 있다. Wang *et al.* [7]의 논문에서는 특허건수가 0인 기업도 포함된 제약 및 생의학회사에 대한 자료를 분석한 반면, 본 연구에서는 특허건수가 0인 국내기업들을 조사할 수 없기 때문에 0에서 절삭된 포아송분포(positive poisson 분포; 이후 PP분포라 칭함)를 구성하고 이 분포에서 과산포를 모형에 반영하여 특허의 생산함수를 추정하고자 한다.

II. 특허의 생산함수모형

1. PP회귀모형과 TZNB회귀모형

포아송분포를 이루는 可算자료(count data)의 회귀모형으로 흔히 포아송회귀모형을 쓸 수 있는데 본 연구에서는 우리 나라 특허건수가 0에서 절삭되었기에 PP분포를 이용하여 PP회귀모형을 다음과 같이 구성하기로 한다.

i 번째 기업의 특허건수 Y_i 는 다음과 같은 PP분포를 갖는다.

$$f(y_i | \beta, x_i, y_i > 0) = \frac{f(y_i | \beta, x_i)}{1 - \exp(-\lambda_i)}, \quad y_i = 1, 2, \dots \quad (1)$$

$$\text{단, } f(y_i | \beta, x_i) = \frac{1}{y_i!} \lambda_i^{y_i} \exp(-\lambda_i), \quad y_i = 0, 1, 2, \dots$$

$$\lambda_i = \exp(x_i' \beta), \quad \beta = (\beta_0, \beta_1, \beta_2)'$$

$$x_i = (1, \log(R\&D_i), (\log(R\&D_i))^2)', \quad i = 1, \dots, n$$

PP분포는 포아송분포에 비해 평균이 크고 분산이 작은 특징을 갖는다. 즉, PP분포에서는 다음 식이 성립한다.

$$E(y_i | x_i, y_i > 0) > E(y_i | x_i) \quad (2)$$

$$Var(y_i | x_i, y_i > 0) < Var(y_i | x_i) \quad (3)$$

과산포가 있는 경우 PP분포를 사용하는 것은 적당하지 않으며, 대안으로서 0에서 절삭된 음이항분포(truncated at zero negative binomial분포; 이후 TZNB분포라 칭함)를 설정할 수 있고, 과산포검정의 문제를 PP분포에 대한 TZNB분포로의 이탈에 관한 문제로 놓을 수 있다.

TZNB분포는 다음과 같다.

$$f(y_i, \alpha | \beta, x_i, y_i > 0) = \frac{f(y_i, \alpha | \beta, x_i)}{\{1 - (1 + \alpha \lambda_i)^{-\alpha^{-1} \lambda_i^{j+1}}\}} \quad (4)$$

$$y_i = 1, 2, \dots, \quad j = 0, 1$$

$$\text{단, } f(y_i, \alpha | \beta, x_i) = \frac{\Gamma(y_i + \alpha^{-1} \lambda_i^{-j+1})}{\Gamma(\alpha^{-1} \lambda_i^{-j+1}) \Gamma(y_i + 1)} \times$$

$$(1 + \alpha \lambda_i^j)^{-\alpha^{-1} \lambda_i^{-j+1}} (1 + \alpha^{-1} \lambda_i^{-j})^{-y_i}, \quad y_i = 0, 1, 2, \dots, \quad j = 0, 1$$

$$\lambda_i = \exp(x_i' \beta), \quad \beta = (\beta_0, \beta_1, \beta_2)'$$

$$x_i = (1, \log(R\&D_i), (\log(R\&D_i))^2)', \quad i = 1, \dots, n$$

$\Gamma(\cdot)$ 는 감마함수, $0 < \alpha < \infty$

2. 과산포 검정

PP분포와 TZNB분포의 분산간에는 다음과 같은 관계가 성립한다.

$$\text{Var}_{TZNB}(y_i | x_i, y_i > 0) \geq \text{Var}_{PP}(y_i | x_i, y_i > 0) \quad (5)$$

식 (5)로부터 TZNB분포는 PP분포와 비교하여 과산포현상을 허용한다. 따라서, 다음과 같이 0에서 절삭된 음이항분포로의 이탈에 관한 가설을 설정할 수 있다.

$$H_0 : \alpha = 0 \quad \text{대} \quad H_a : \alpha > 0$$

여기서 $\alpha = 0$ 은 $\alpha \rightarrow 0$ 으로 해석한다.

이 가설에 대한 검정통계량은 다음과 같다(Gurmu [5]).

$$T = \frac{\sum_i \hat{\lambda}_i^{j-1} \{ \hat{\epsilon}_i^2 - y_i + (\hat{\epsilon}_i + y_i) \hat{\delta}_i \}}{2 \{ I_{\alpha\alpha}(\hat{\rho}) - I_{\alpha\beta}(\hat{\rho}) I_{\beta\beta}^{-1}(\hat{\rho}) I_{\beta\alpha}(\hat{\rho}) \}^{1/2}} \underset{d}{\sim} N(0, 1) \quad (6)$$

$$\text{단, } I(\hat{\rho}) = \begin{bmatrix} I_{\beta\beta}(\hat{\rho}) & I_{\beta\alpha}(\hat{\rho}) \\ I_{\alpha\beta}(\hat{\rho}) & I_{\alpha\alpha}(\hat{\rho}) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_i \hat{\mu}_i (1 - \hat{\delta}_i) x_i x_i' & \frac{1}{2} \sum_i \hat{\theta}_i^j \hat{\mu}_i \hat{\delta}_i x_i \\ \frac{1}{2} \sum_i \hat{\theta}_i^j \hat{\mu}_i \hat{\delta}_i x_i' & \frac{1}{2} \sum_i \hat{\theta}_i^{2j-1} \hat{\mu}_i (1 - \frac{1}{2} \hat{\theta}_i \hat{\delta}_i) \end{bmatrix}$$

$$\hat{\lambda}_i = \exp(x_i' \hat{\beta}), \quad \hat{\delta}_i = \frac{\hat{\lambda}_i}{\exp(\hat{\lambda}_i) - 1}, \quad \hat{\mu}_i = \hat{\lambda}_i + \hat{\delta}_i, \quad \hat{\varepsilon}_i = y_i - \hat{\lambda}_i$$

$$x_i = (1, \log(R\&D_i), (\log(R\&D_i))^2)', \quad i=1, \dots, n$$

‘a’는 ~를 중심으로 좌측의 통계량이 우측의 분포를 귀무가설하에서 점근적으로 접근함을 의미한다.

한편, 포아송모형에서 과산포에 대한 스코어검정은 소표본(small sample) 하에서 검정의 실제 크기(actual size)는 명목크기(nominal size)보다 훨씬 작다. Gurmu [5]는 이러한 단점을 보완하기 위한, 크기를 보정한 검정통계량을 식 (7)과 같이 제안하였다.

$$T_a = \frac{\sum_i \hat{\lambda}_i^{j-1} [\hat{\varepsilon}_i^2 - y_i + (\hat{\varepsilon}_i + y_i) \hat{\delta}_i + \hat{H}_i \{ \hat{\mu}_i + \hat{\delta}_i (\hat{\mu}_i - 2) \}]}{2 \{ I_{\alpha\alpha}(\hat{\rho}) - I_{\alpha\beta}(\hat{\rho}) I_{\beta\beta}^{-1}(\hat{\rho}) I_{\beta\alpha}(\hat{\rho}) \}^{1/2}} \quad \text{a} \quad N(0, 1) \quad (7)$$

$$\text{단, } \hat{H}_i = \hat{\mu}_i (1 - \hat{\delta}_i) x_i' \{ \sum_j \hat{\mu}_j (1 - \hat{\delta}_j) x_j x_j' \}^{-1} x_i$$

$$x_i = (1, \log(R\&D_i), (\log(R\&D_i))^2)', \quad i=1, \dots, n$$

III. 국내 특허자료

본 연구에서 사용되는 특허자료는 1996년에 등록된 것으로, 특허건수는 특허청으로부터 자료를 수집하였으며, R&D는 한국신용평가정보주식회사의 상장법인 재무자료에서 수집하였다(한국신용평가정보주식회사 [3]). 여기서 R&D는 경상연구개발비, 개발비, 연구개발비 상각, 연구개발비의 네 항목을 합한 것으로 사용하였다. 이 같은 정의는 일시적인 연구개발투자로 인한 과대계상을 피하고 해당기업이 한 해 동안 연구개발에 투자한 규모에 중점을 두었기 때문이다. <표 1>에서 1996년 기업들의 특허건수의 평균은 55건이고 평균 R&D는 633억 원 정도이다.

또한, 특허자료에 대한 패널자료를 수집하였다. 패널자료는 1995년부터 1997년까지 3년 동안 68개 기업체를 대상으로 수집하였으며 특허건수와 연구개발비를 변수로 구성하였다. <표 2>에서 알 수 있듯이 R&D는 1996년에 크게 증가한 반면 특허건수는 1997년에 크게 증가하였다.

〈표 1〉 1996년 특허자료의 기술통계량

변 수	평 균	분 산	최소값	최대값	표본수
특허등록건수	55.38	23616.33	1	868	90
R&D	63.269	17435.454	0.002286	792.387	90

주 : R&D의 단위는 십억 원.

〈표 2〉 1995년부터 1997년까지의 특허 패널자료의 기술통계량

연 도	변 수	평 균	분 산	최소값	최대값	표본수
1995	특허등록건수	60.53	32242.61	1	1164	68
	R&D	63.896	14783.006	0.012	724.504	68
1996	특허등록건수	72.25	30187.35	1	868	68
	R&D	81.885	21707.500	0.035549	792.387	68
1997	특허등록건수	132.51	125459.448	2	2210	68
	R&D	84.440	22508.750	0.044517	811.533	68

주 : R&D의 단위는 십억 원.

IV. 실증분석

1. 과산포 검정

국내 특허건수에 대한 과산포현상을 두 가지 방법으로 분석하였다. 첫째, PP분포의 평균이 분산보다 크다는 성질을 이용하여 표본의 평균과 분산을 계산하여 표본분산이 표본평균보다 크다는 것을 보임으로써 과산포를 나타낼 수 있다. 둘째, 과산포 검정통계량인 식 (6)의 T 와 식 (7)의 T_0 를 사용하여 검정을 실시하였다. TZNB모형에서 $j=0$ 인 경우 모수가 수렴하지 않는 문제가 발생하였기 때문에 본 연구에서는 $j=1$ 인 TZNB모형만을 다루었다.

1) 표본의 평균과 분산의 비교

연구개발비의 규모가 비슷한 기업끼리 군을 이루어 모두 15군을 구성하여 각각에 대한 표본의 평균과 분산을 계산하였다. 계산결과 표본 전체에서는 표본분산이 표본평균보다 426배 정도가 크다. 또한, 15개 그룹에 대해서도 그룹 13의 표

본평균이 표본분산보다 약간 작은 것을 제외하고는 표본분산이 표본평균보다 작게는 2배, 크게는 370배 정도 큼을 알 수 있다. 즉, 표본 전체뿐만 아니라 15개 그룹에 대하여 표본분산이 표본평균보다 현격하게 큼을 알 수 있다. 따라서, 국내 특허건수는 연구개발비를 기준으로 '비교적' 동질적인 15개 그룹의 표본분산이 표본평균보다 현격하게 크므로 과산포현상이 존재함을 알 수 있다.

2) 검정통계량 T 와 T_a

PP분포로부터 TZNB분포로의 이탈에 관한 검정, 즉 과산포현상에 대한 검정통계량 T 를 추정하기 위해 먼저 귀무가설하에서, 즉 PP모형하에서 β 의 최우추정치 $\hat{\beta}$ 를 추정한다. 다음으로 추정된 $\hat{\beta}$ 를 사용하여 식 (6)에서의 $\hat{\lambda}, \hat{\delta}_i, \hat{\mu}_i, \hat{\varepsilon}_i (i = 1, \dots, n)$ 을 계산한다. 이 값을 가지고 식 (6)에서의 정보행렬 $I(\hat{\rho})$ 을 추정한다. 또한, 크기를 보정한 검정통계량 T_a 를 추정하기 위해 식 (6)에서 구한 $\hat{\delta}_i, \hat{\mu}_i (i = 1, \dots, n)$ 을 사용하여 $H_i (i = 1, \dots, n)$ 를 추정한다. 이러한 절차를 걸쳐 검정통계량 T 와 크기를 보정한 검정통계량 T_a 의 값이 각각 400.59와 400.92로 추정되었다. 이 값은 국내 특허건수에 과산포현상이 있음을 강력하게 증거하고 있다.

〈표 3〉 15개 群別 표본평균 및 표본분산

	표본평균	표본분산	표본분산÷표본평균	표본수
표본 전체	55.38	23616.33	426.4	90
그룹 1	461.00	170586.40	370.0	6
그룹 2	23.50	323.67	13.8	4
그룹 3	78.33	9528.67	121.6	6
그룹 4	83.50	2099.50	25.1	6
그룹 5	27.00	860.40	31.9	6
그룹 6	17.60	744.30	42.3	5
그룹 7	20.50	386.86	18.9	8
그룹 8	16.80	169.20	10.1	5
그룹 9	19.00	36.67	1.9	4
그룹 10	71.75	19092.92	266.1	4
그룹 11	4.09	16.29	4.0	11
그룹 12	15.33	171.07	11.2	6
그룹 13	4.40	4.80	1.1	5
그룹 14	15.83	908.17	57.4	6
그룹 15	4.75	12.50	61.0	8

2. 특허의 생산함수 추정

생산함수는 GAUSS Ver. 3.2(Aptech Systems, Inc. [8])와 MAXLIK 모듈(Aptech Systems, Inc. [9])을 사용하여 추정하였다. 국내 특허건수에 대한 PP모형과 TZNB모형의 추정방정식은 다음과 같다. 괄호 안의 숫자는 t 값을 나타낸다.

PP모형

$$\log(\hat{\lambda}_i) = 1.7733 + 0.4547 \log(R\&D_i) + 0.4143 \{\log(R\&D_i)\}^2$$

(43.395) (15.861) (43.469)

$$\text{대수우도값} = -2630.33$$

TZNB모형

$$\log(\hat{\lambda}_i) = 1.7392 + 0.7477 \log(R\&D_i) + 0.2471 \{\log(R\&D_i)\}^2, \hat{\alpha} = 2.2956$$

(6.953) (3.562) (2.622) (8.2457)

$$\text{대수우도값} = -372.34$$

V. 결 론

우리는 두 가지 방법을 통하여 국내 특허건수에 과산포현상이 존재함을 알았다. 첫째, 표본을 15개 群으로 나누어 계산한 표본분산이 표본평균보다 모두 크므로 PP분포의 평균이 분산보다 크다는 조건을 만족하지 않는다. 둘째, 대립가설 TZNB모형에 기초한 검정통계량 T 와 T_0 값에 대한 p -value가 모두 0.00으로 국내 특허건수에 과산포현상이 존재함을 알 수 있다. 또한, TZNB모형에서 추정된 α 값과 이에 대한 t 값이 각각 2.2956과 8.2457로써 국내 특허건수의 생산함수로 PP모형이 적합하지 않음을 알 수 있다. 따라서, 국내 특허의 생산함수로서 PP모형보다는 TZNB모형이 더 타당한 모형임을 알 수 있다.

VI. 향후 연구과제

특허자료에 대해서 앞으로 연구할 과제는 다음과 같다. 첫째, PP모형의 유한혼합분포를 사용해 연구개발투자의 집약도(=연구개발비 ÷ 매출액)를 혼합확률의 설명변수로 사용하여 기업활동을 기업들의 질적인 특성에 따라 몇 개 군으로 분류해 볼 수 있다. 이렇게 분류된 군마다 생산함수모형을 추정한다(Wang *et al.* [7]). 둘째, 특허의 패널자료를 사용하여 포아송회귀모형을 PP회귀모형으로 확장하여 과산포현상에 대해 검정하고, 특허에 대한 기업효과(firm effect)에 대해 살펴본다(Hausman *et al.* [6]).

◆ 참고 문헌 ◆

1. 김병수 · 오경주 · 박철용, “가산자료(count data)의 과산포 검색: 일반화 과정”, 「응용통계연구」, 8, 한국통계학회, 1995, pp. 147~161.
2. 정갑영, 『한국의 산업조직: 구조적 접근』, 박영사, 1998.
3. 한국신용평가정보주식회사, 『한국기업총람』, 한국신용평가정보주식회사, 1996, 1997, 1998.
4. Cleveland, W. S., “Robust Locally Weighted Regression and Smoothing Scatter Plots,” *Journal of the American Statistical Association*, 74, 1979, pp. 829~836.
5. Gurmu, S., “Tests for Detecting Overdispersion in the Positive Poisson Regression Model,” *Journal of Business & Economic Statistics*, 9, 1991, pp. 215~222.
6. Hausman, J. A., Hall, B. H. and Z. Griliches, “Econometric Models for Count Data with an Application to the Patents-R&D Relationship,” *Econometrica*, 52, 1984, pp. 909~938.
7. Wang, P., Cockburn, I. M. and M. L. Puterman, “Analysis of Patent Data - a Mixed - Poisson - Regression - Model Approach,” *Journal of Business & Economic Statistics*, 16, 1998, pp. 27~41.
8. Aptech Systems, Inc., *GAUSS Command Reference*, Aptech Systems, Inc., 1994.
9. _____, *Maximum Likelihood 4*, Aptech Systems, Inc., 1995.