

# ビッグデータを用いた特許の出願数と会社業績の関係<sup>1)</sup>

藤本 祥二<sup>2)</sup>・石川 温<sup>2)</sup>・水野 貴之<sup>3)</sup>・渡辺 努<sup>4)</sup>

## Relation between Firms Performance and Number of Patent Applications using Big Data<sup>1)</sup>

Shouji FUJIMOTO<sup>2)</sup>, Atushi ISHIKAWA<sup>2)</sup>, Takayuki MIZUNO<sup>3)</sup> and Tsutomu WATANABE<sup>4)</sup>

### 要 旨

全世界の企業別に集計した特許の出願件数と被引用件数に関する分布調べた。企業単位で集計した特許と、各企業の売上などの業績データとの関係を調べた。関係を調べる際には企業の規模に関して条件を揃えることで偽相関などの可能性を排除することができる。

キーワード：特許、経済物理、幂分布（パレート則）、生産関数

### 1. はじめに

「イノベーション（革新、刷新）」というカタカナ語は誰もが一度は聞いたことがある言葉であろう。イノベーションはシュペンター著の『経済発展の理論』<sup>[1]</sup>で展開される理論の中心概念である。『経済発展の理論』ではイノベーションを次の5つに分類している。

1. 新しい生産物または生産物の新しい品質の創出と実現
2. 新しい生産方法の導入
3. 産業の新しい組織の創出
4. 新しい販売市場の開拓
5. 新しい買い付け先の開拓

全ての項目が経済発展に重要であろうことは理解できるが、新しいというイメージを強調するためだけにイノベーションという言葉を使うのは本末転倒である。イノベーションを確実に経済発展を繋げるには様々な種類のイノベーションと経済発展の関係の定量的な検証が必要である。特許は上記分類の1、2番に関係している。本論文ではイノベーションの一つの側面を定量的に調べるために企業の持つ特許と売上の関係を調べる。

この2、3年の間に「ビッグデータ」という言葉も市民権を得てきた。コンピュータの処理速度の向上やインターネットの普及に伴い大量のデジタルデータが記録・利用しやすい状況で管理されている。『経済物理学』は、近年盛んに新しい研究がなされている学問の分野である。具体的には経済現象を大量の実データを元に物理学の手法を取り入れて実証分析を行う分野であり、ビッグデータを用いた分析を牽引してきた分野である。本論文では世界中の特許を網羅したビッグデータと世界中の企業の財務データを網羅したビッグデータを用いて分析を行う。

1)：平成25年10月10日受付；平成25年10月31日受理。

Received Oct. 10, 2013 ; Accepted Oct. 31, 2013.

2)：金沢学院大学 経営情報学部；Faculty of Business Administration and Information Science, Kanazawa Gakuin University.

3)：国立情報学研究所，総研大；National Institute of Informatics, The Graduate University for Advanced Studies

4)：東京大学 経済学部；Faculty of Economics, The University of Tokyo.

特許についてはこれまでにいくつかの研究が報告されているが（例えば<sup>[2]</sup>を参照）、世界中の企業特許を網羅した研究は全く報告されておらず、本論文は最初のステップになるものである。

## 2. データの詳細

特許のデータは経済開発協力機構（OECD）が日本国特許庁、欧州特許庁、アメリカ合衆国特許商標庁、世界知的所有権機関、欧州委員会、全米科学財団からなる特許統計タスクフォース）の協力のもと、特許統計データベース、PATSTAT（Worldwide Patent Statistical Database）の開発を進めている。PATSTAT データベースは欧州特許庁（EPO）のサイトで閲覧できる<sup>[3]</sup>。

Bureau van Dijk 社の ORBIS データベース<sup>[4]</sup>は全世界（欧州、アジア、日本、ロシア、北米、ヨーロッパ）の包括的なデータを提供する世界最大の企業財務データベースである。我々は PATSTAT データベースに Bureau van Dijk 社が企業コードを付加したものを入手し、それによって企業が出願している特許と企業の財務データを紐付したデータを扱えるようになった。その結果全世界を網羅した企業特許のデータを詳しく調べることができるようになったわけである。

図1は1995以降の特許の全世界での出願数である。特許の出願数は右肩上がりで推移している。全データでは1995年の約150万件から2011年の約300万件に倍増している。企業の特許出願数は50万件から150万件と3倍になっている。本論文ではこのデータを元に企業特許の分布を調べ、財務データとの関連をみる。

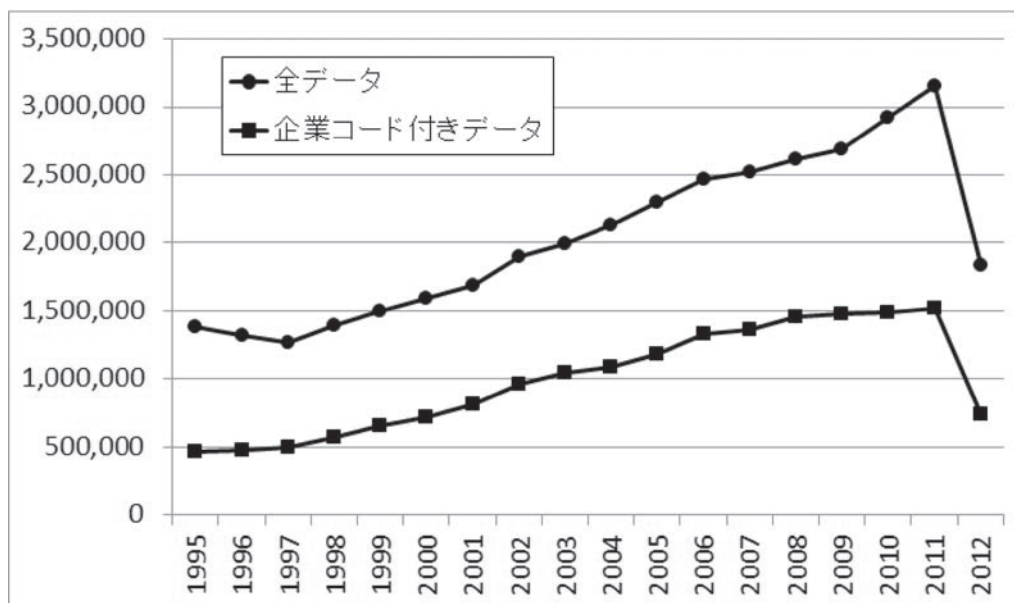


図1：1995年以降の特許の全世界出願数。丸印は企業個人全てのデータを用いた集計。四角印は企業が出した特許のみの集計。2012年は現段階で集計が終わっていない。

## 3. 出願件数の分布

特許のデータで最も基本的なものは出願件数である。出願件数には企業の特許に対する取り組み姿勢が表れると考えられる。図2は2011年の企業特許出願件数の累積分布を両対数プロットしたものである。分布は経済物理学で早くから注目されよく知られているベキ分布<sup>[5]</sup>になる。ベキ分布の特徴は図2の傾きで評価できるベキ指数で表される。このベキ指数はデータ間の格差を表す指標としても考えることができ、ベキ指数が小さいほど格差が大きくなることが知られている。

表3には2003年から2011年の様々なベキ指数をまとめているが、企業別の出願件数の分布のベキ指数は1.13から1.17の間で年が変わってデータ数が変わっても変化が殆どないことが確認された。

表1は2010年と2011年の特許出願件数のトップ10企業名である。トップ10には Samsung, IBM, Panasonic, などの大手製造業や、Qualcomm, ZTE, Bosch, などの通信業がランクインしている。

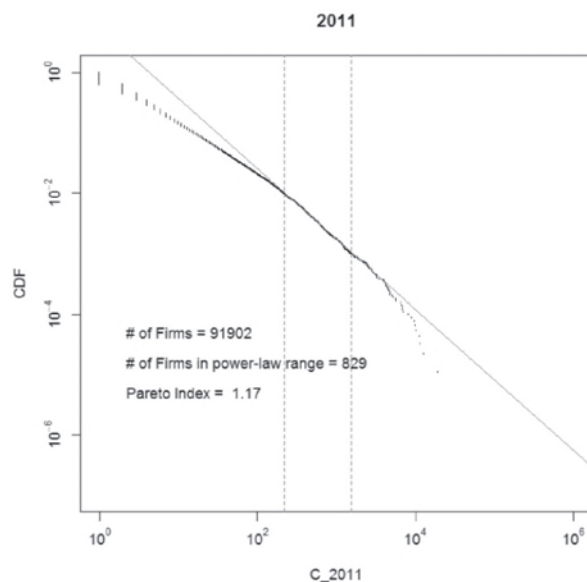


図2：2011年の企業特許出願数の累積分布。企業数は91,902社、ベキ分布の領域には829社、ベキ指数は1.17である。

表1：特許出願件数トップ10企業名。

No.	2010	2011
1	Samsung	Samsung
2	IBM	IBM
3	LGE	ZTE
4	Panasonic	LGE
5	Qualcomm	Panasonic
6	ZTE	Bosch
7	Bosch	Qualcomm
8	Sony	Canon
9	Canon	Sony
10	Microsoft	Microsoft

#### 4. 被引用件数の分布

特許の価値を表す指標の一つに被引用件数がある。特許は他の特許を元にしたたり参考にしたたりしているが、その情報は引用参照として扱われている。被引用件数とは対象となる特許が他の特許から何回引用されているかの件数である。引用回数の多い特許は他の特許に与える影響も多いのでそれだけ価値が高いと考えられる。また、件数を扱うことで定量的な評価ができるのでこの量に注目する意義も高い。

図3は2011年の特許の被引用件数の企業で集計する前の累積分布である。直線部分があるのでその傾きからベキ

指数を見積もると、ベキ指数は約3になる。企業別に集計する前の1つ1つの特許では引用されてもせいぜい百件単位であるため分布の裾の部分がそれほど伸びない。そのためベキ指数は3と大きな値になるわけである。

企業別に集計した被引用数には大きな格差があることが確認できる。図4は2011年の特許の被引用件数を企業別に集計したものの分布である。表3には2003年から2011年のベキ指数をまとめたが、ベキ指数は1.05から1.07の間で年が変わっても安定して殆ど同じ値になる事が確認できる。

表2は特許の被引用件数の企業別トップ10である。出願件数でトップ10に入っていた通信業がなくなり製造業ばかりになっている。出願数と被引用件数で業種間の違いが垣間見える。

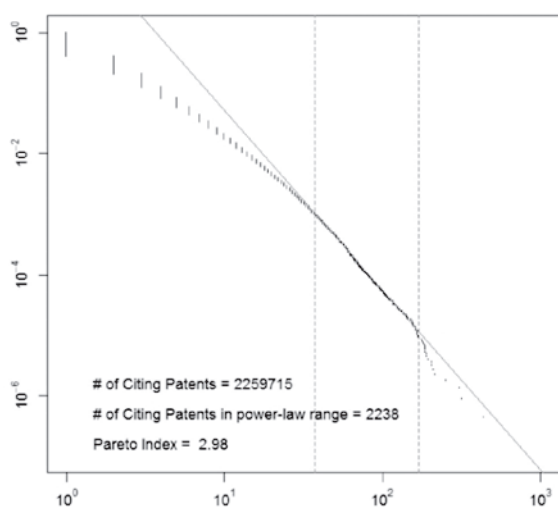


図3：2011年の特許の被引用件数の累積分布。特許数は2,259,715個、ベキ分布の領域には2,238個、ベキ指数は2.98である。

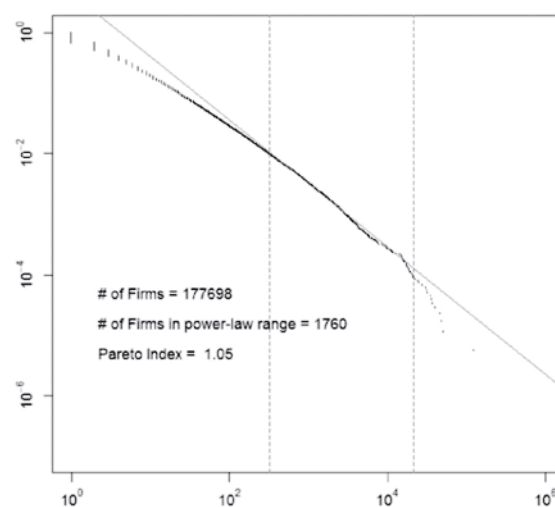


図4：2011年の特許の企業別に集計した被引用件数の累積分布。企業数は177,698社、ベキ分布の領域には1,760社、ベキ指数は1.05である。

表2：被引用件数トップ10企業名。

No.	2010	2011
1	IBM	IBM
2	Microsoft	Samsung
3	Samsung	Microsoft
4	Canon	Canon
5	Panasonic	Panasonic
6	Hitachi	Hitachi
7	Toshiba	Toshiba
8	Intel	Intel
9	Sony	Motorola
10	Motorola	Sony

表 3：年別各変数のベキ指数表。

年	企業別出願数	被引用数	企業別被引用数
2003	1.13	3.44	1.07
2004	1.13	3.66	1.06
2005	1.12	4.05	1.07
2006	1.13	3.75	1.07
2007	1.15	3.73	1.07
2008	1.14	3.10	1.07
2009	1.13	3.16	1.07
2010	1.17	3.25	1.05
2011	1.17	2.98	1.05

## 5. 売上と出願数の相関

これまでの我々の研究<sup>[6]</sup>で、企業の売上  $Y$ 、有形固定資産  $K$ 、従業員数  $L$ 、の間には

$$Y = AK^{\alpha}L^{\beta} \quad (1)$$

の関係があることが分かっている。これは生産に関わる投入量  $K, L$  と生産量  $Y$  との関係である生産関数として解釈でき、上式の形の生産関数はコブ・ダグラス型生産関数<sup>[7]</sup>と呼ばれる。

生産関数には様々な形の関数が考えられるがコブ・ダグラス型が実際のデータと良く合う。その理由1つとして  $Y, K, L$  がベキ分布をすることが上げられる。上式の両辺で対数を取ると各変数の間には線形の関係が成り立ち、ベキ分布と非常に相性が良いというわけである。各企業の添え字を付けて上式の対数を取ったものを書くと

$$\log Y_i = \alpha \log K_i + \beta \log L_i + \log A_i \quad (2)$$

になる。この式でデータを重回帰分析すると、弾性値を表す定数  $\alpha, \beta$  が決まる。 $A_i$  は定数切片と残差を合わせたものである。この残差の部分には  $K, L$  のみでは説明しきれない  $Y$  に関する情報が入っていると考えられる。 $A$  は全要素生産性 (TFP) と呼ばれる生産効率と関係する値である。生産効率には企業の技術力も含まれるため特許による技術力とも何か関係があるのではないかと思われる。

特許出願件数 (又は被引用件数) を  $C$  とすると、 $C$  も  $Y, K, L$  と同様ベキ分布に従うので生産関数に

$$Y = BC^{\gamma}K^{\alpha}L^{\beta} \quad (3)$$

の形で組み込めるのではないかと考えられる。この式から

$$A = BC^{\gamma} \quad (4)$$

の関係式が成り立ち、 $A$  と  $C$  に対数相関があることになる。これを確認するために  $A$  と特許の出願件数や被引用件数に対数相関があるのか調べてみた。結果の図は掲載しないが相関係数は殆ど0であった。おそらく何らかの相関はあると思われるのだが、 $A$  を決める際にデータを均してしまう効果が凄まじく大きいので、手法としてあまり上手い方法ではないと考えられる。

重要なのは売上 (生産量) と特許の関係なので  $Y$  と  $C$  の関係を詳しくみてみよう。図5に  $C$  と  $Y$  の散布図を両対数でプロットしたものを示す。見た目で相関があることが分かるが、 $\log C$  と  $\log Y$  の相関係数である対数相関係数は0.48なので相関はある。ただしこの値は  $\log K$  と  $\log Y$  の相関や  $\log L$  と  $\log Y$  の相関の約0.9に比べて弱い値である。図6に  $C$  と  $K$  の散布図を示す。 $\log C$  と  $\log K$  の相関係数は0.45である。また、 $\log C$  と  $\log L$  の相関は0.51である。ここで注意しなければならないのは  $C$  と  $Y$  の相関は直接の相関ではなく企業の規模を表す変数  $K$  や  $L$  を通じて現れる相関、つまり偽相関なのではないかということである。特許が売上と直接関係しているの

ではなく、「規模の大きい企業は特許も多く売り上げも多いので結果として相関があるように見える」という可能性である。

企業の規模の影響を取り除くためには規模に関わる変数  $K, L$  を固定してその条件の下で  $C$  と  $Y$  の関係を調べる必要がある。(3)式で  $K, L$  の部分を定数とみなして

$$Y = BC^r \times (\text{定数}) \quad (3)$$

の関係を調べるわけである。ただし、完全に固定してしまうとその条件に当てはまるデータは1個もなくなってしまいうため、ある程度幅を持たせて制限をつけることになる。

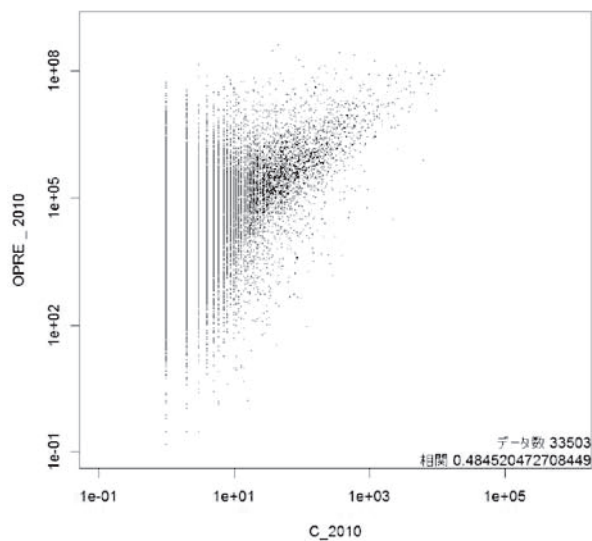


図5：2011年の企業別特許出願数  $C$  と売上  $Y$  の散布図。  
対数相関係数は0.485。

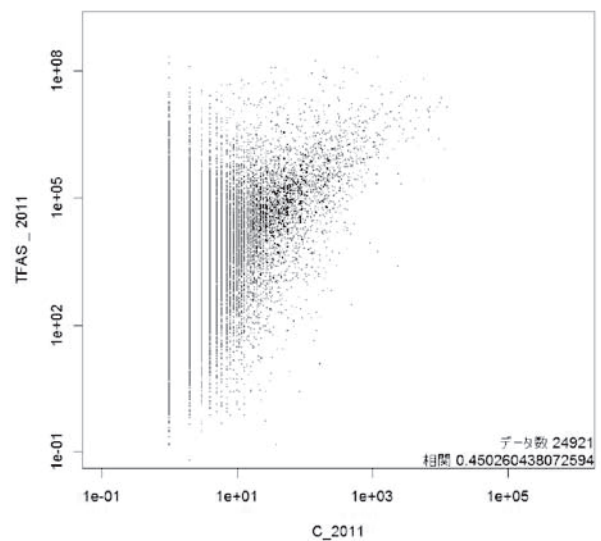


図6：2011年の企業別特許出願数  $C$  と有形固定資産  $K$  の散布図。対数相関係数は0.450。

図7は図5のデータの中で  $10^2 \leq K \leq 10^3$ ,  $10^0 \leq L \leq 10^1$  のデータを抽出してプロットしたものである。この領域の  $C$  と  $Y$  の対数相関は0.02となり殆どない。図8は図5のデータの中で  $10^4 \leq K \leq 10^5$ ,  $10^2 \leq L \leq 10^3$  のデータを抽出してプロットしたものである。この領域の  $C$  と  $Y$  の対数相関は0.22となり小さいながらも相関はある。表3は  $K, L$  を1桁ごとの領域に区切って  $C, Y$  の相関を調べた結果を表にしたものである。特許の出願件数と売り上げ間には従業員数や有形固定資産が大きな企業では小さいながらも相関があると考えられる。



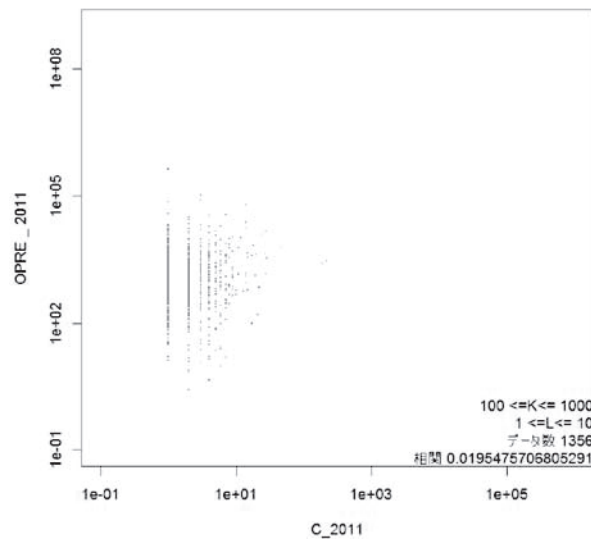


図7： $10^2 \leq K \leq 10^3$ [\$],  $10^0 \leq L \leq 10^1$ [人]の条件でのCとYの散布図。データ数は1356、対数相関は0.02。

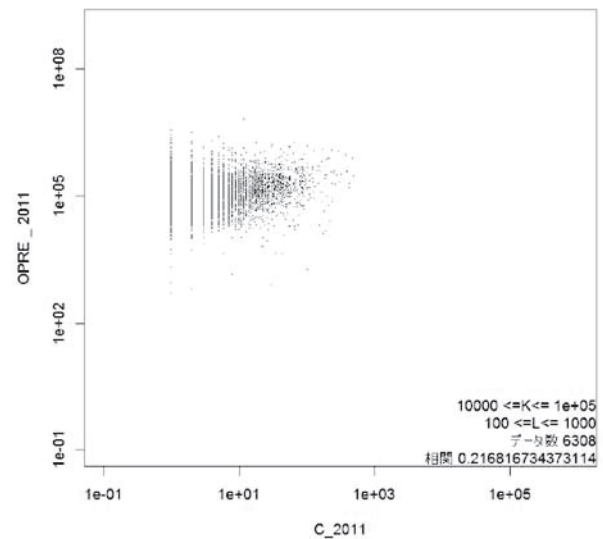


図8： $10^4 \leq K \leq 10^5$ [\$],  $10^2 \leq L \leq 10^3$ [人]の条件でのCとYの散布図。データ数は6308、対数相関は0.22。

表3：K, Lの領域に対するC, Yの対数相関。カッコ内は領域内のデータ数。データ数と相関の両方がある程度大きなところを太字で強調した。

	log L	0	1	2	3	4	5
log K		1	2	3	4	5	6
0	1	0.00(627)	-0.03(131)				
1	2	-0.03(1414)	-0.02(1114)	0.08(39)			
2	3	0.02(1356)	0.01(4043)	0.06(528)	-0.18(9)		
3	4	0.14(490)	0.05(6157)	<b>0.20(4054)</b>	0.05(91)		
4	5	<b>0.27(50)</b>	0.13(1150)	<b>0.22(6308)</b>	<b>0.17(1603)</b>	0.17(10)	
5	6		-0.14(24)	0.10(764)	<b>0.17(2317)</b>	-0.01(369)	
6	7			-0.13(26)	<b>0.23(308)</b>	<b>0.18(498)</b>	<b>0.19(37)</b>
7	8				<b>0.45(34)</b>	<b>0.20(94)</b>	<b>0.40(48)</b>
8	9						0.76(4)

## 6. まとめと今後の展望

今回の研究では、特許の企業別出願件数の分布を調べた。分布はベキ分布になりベキ指数の2003年から2011年の推移をみた。結果は図2と表3に示している。ベキ指数は年に依らず安定して1.13～1.17の値になる。

次に特許の被引用件数についての分布を企業別に集計する前のものと企業別に集計した後のものと調べた。結果は図3、図4と表3に示している。企業別に集計する前のベキ指数は3～4であったが、集計後のベキ指数は1.05～1.07で年に依らずに安定している。

企業の特許データがベキ分布になる事が判明したので他のベキ分布する企業の財務データと関係があると推測される。それを確認するために様々なデータとの相関係数を調べた。重要なのは特許が直接売上と相関があるかどうかである。特許と売上の相関を見るには企業規模との偽相関の可能性があるので企業規模の条件を固定した企業同

士で相関を確認した。結果は表3に示している。従業員数や資産の大きな大企業では大きくはないが相関はある。

特許と売上には様々な経緯を通じて関係があると考えられる。直接的なのは特許収入であるが、企業特許には他社に自社技術を使われないようにという防衛的な側面が出願される側面もある。防衛的な出願は間接的に売上に反映されると考えられる。業種や国によって直接収入と防衛的な側面との割合が違うと考えられる。業種をうまく分けて自己引用などを精査することでその割合と売上の関係などを調べていきたい。

また、特許の質をどのように定量化するかという問題がある。今回は定量化が簡単な出願件数や被引用件数のみを見ていたが、すぐに特許収入に直接つながる特許もあれば誰にも目にされない特許も存在する。それらを全て一律に扱うのではなく論文雑誌のインパクトファクターのような定量化を模索していきたい。

## 謝辞

本研究は JSPS 科研費24510212および24710156の助成を受けたものである。

## 参考・引用文献リスト

- [1] 塩野谷祐一・中山伊知郎・東畑精一訳 *経済発展の理論*, (岩波書店, 1977).
- [2] G. Silverberg and B. Verspagen, MERIT–Infonomics Research Memorandum Series 2004–021(2004); X. Li, H. Chen, Z. Huang, M. C. Roco, *Journal of Nanoparticle Research* 9 (2007) 337; D. R. J. O’Neale and S. C. Hendy, *PLOS ONE* 7 (2012) e 49501.
- [3] EPO – PATSTAT online (Beta) [http : //www.epo.org/searching/subscription/patstat-online.html](http://www.epo.org/searching/subscription/patstat-online.html)
- [4] Bureau van Dijk 日本法人 ORBIS [http : //www.bvdinfo.co.jp/service/db\\_orbis.html](http://www.bvdinfo.co.jp/service/db_orbis.html)
- [5] 藤本祥二・石川温, 金沢学院大学紀要, 第9号, (2011), 25; V. Pareto, *Cours d’Economie Politique* (Macmillan, London, 1897).
- [6] A. Ishikawa, S. Fujimoto, T. Mizuno, T. Watanabe, *Prog. Theor Phys. Suppl.* **194** (2012) 122.
- [7] C. W. Cobb and P. H. Douglas, *American Economics Review*, **18** (1928) 139.