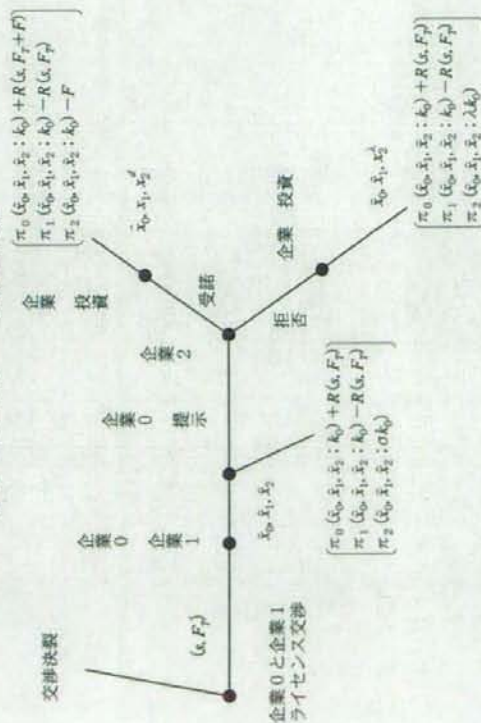


図 11 インサイダー 企業秘密交渉



は以下のとおりになる。式(4)から企業0の利益は、投資費用が k_0 である企業の数 (自分とライセンスの2企業) にも依存し、どの企業がライセンス契約を締結しているかに関係ないことを指摘しておく。

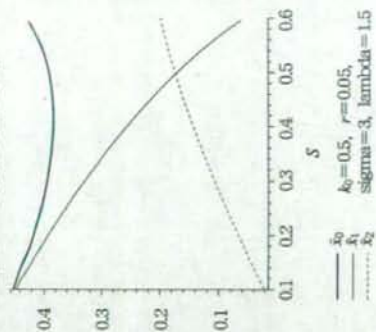
$$U_0 = m(h_0, h_0, \lambda h_0) + F, \quad U_1 = U_2 = m(h_0, \lambda h_0, h_0). \quad (6)$$

3.2 企業秘密による場合

ライセンス交渉のタイミングはアウトサイダーの場合と同じであるが、企業0も投資を行うので、ライセンス収入の他に自分の投資による利益があることが異なる。ライセンス交渉のタイミングと利得は図11にまとめてある。

ライセンスが承諾され、ロイヤルティが s と決まったら、全企業が開発投資を同時に決定する。企業0はライセンス料と自社の投資利益の和を最大にするように投資を決め、以下の期待利益が利得である非協力ゲームになっている。企業1が承諾された場合は特許化される。企業1が企業秘密ライセンスを結んだ場合は、費用係数が $k_i = h_0, k_i = \sigma h_0$ となり、企業は投資 $x_i, i=0, 1, 2$ を以下の利得が最大になるように同時に決める。

図 12 均衡投資 ($\bar{x}_0, \bar{x}_1, \bar{x}_2$)



$$\bar{x}_0 = \frac{s x_0}{x_0 + x_1 + x_2 + F} + \frac{x_0}{x_0 + x_1 + x_2 + F} - \lambda h_0 s,$$

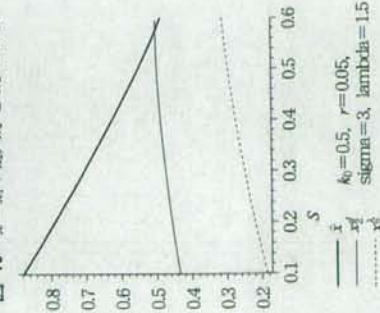
$$\bar{x}_1 = \frac{(1-s)x_1}{x_0 + x_1 + x_2 + F} - \lambda h_0 s, \quad \bar{x}_2 = \frac{x_2}{x_0 + x_1 + x_2 + F} - \sigma \lambda h_0 s$$

均衡投資 ($\bar{x}_0, \bar{x}_1, \bar{x}_2$) はナッシュ均衡である。式は補遺にあるが、ここではプロット (図12) を示す。ロイヤルティー支払が増加すると、企業1の限界収入が減るので、投資は減少する。¹¹⁾ 企業0の投資はいくつかの効果に依存する。まず、 s の増加の効果は2つある。一方で企業1の収入の取り分が増加するが、 x_1 が減少するので企業1の収入自体は減少し、 m の第1項は増加も減少もしうる。また、企業0の投資は2つの効果がある。投資を増やすと、自分の勝利の確率 (第2項) は増加するが、企業1の収入 (第1項) は減少する。2つの効果のどちらかが、また企業0の投資の効果のどちらが勝るかによって \bar{x}_0 は決まり、U字型になっている。企業2の投資は s とともに増加しているが、これは図13で見ると、 \bar{x}_0 と \bar{x}_1 の和が減少しているからである。

次にロイヤルティー水準を検討する。アウトサイダーの場合と同様に企業0

11) もちろん厳密には直接効果 (direct effect) の他に戦略効果 (strategic effect) がある。

図13 $\bar{x} = \bar{x}_1 + \bar{x}_2$, x_1^d と x_2^d の比較



$k_0=0.5$, $r=0.05$,
 $\sigma=3$, $\lambda=1.5$

が企業1にライセンスした後に企業2とも契約することを防ぐライセンスを企業1は結ぶ必要がある。企業2がライセンスを受諾した場合と、拒否した場合のそれぞれ企業2の投資を x_2^d と x_2^r とする。アウトサイダーの場合と同様に、ライセンス（企業1）と企業0はすでに投資をした後なので、企業2にとって企業1と企業0の投資は所与である。 $\bar{x} = \bar{x}_0 + \bar{x}_1$ とすると、最適投資は以下のようになる。

$$x_2^d = \arg \max_{x_2} \frac{x_2}{\bar{x} + x_0 + r} - k_0 x_2 = \sqrt{\frac{\bar{x} + r}{k_0}} - (\bar{x} + r)$$

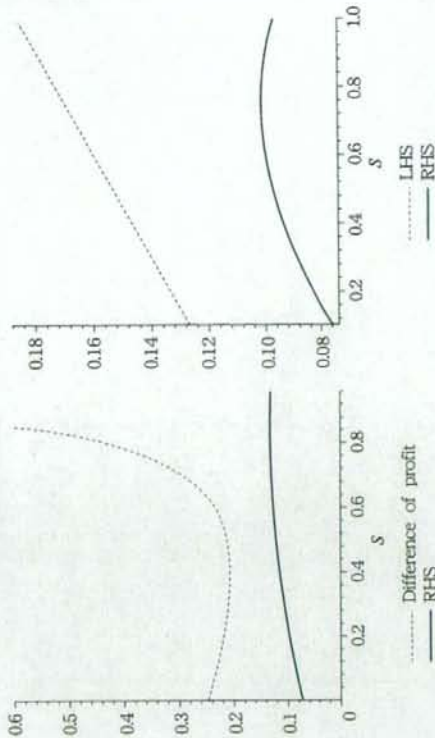
$$x_2^r = \arg \max_{x_2} \frac{x_2}{\bar{x} + x_0 + r} - \lambda k_0 x_2 = \sqrt{\frac{\bar{x} + r}{\lambda k_0}} - (\bar{x} + r)$$

図13にあるように、 s が増加すると double marginalization がひどくなり、企業0と企業1の合計投資 $\bar{x} = \bar{x}_0 + \bar{x}_1$ が減少する。具体的な \bar{x} の式は補遺参照。アウトサイダーの場合と同様に交渉過程のスピルオーバーにより企業2の費用が低下し投資が増える ($x_2^d > x_2^r$)。そのときの企業2の利得は以下のとおりで、明らかに $\pi_2^d > \pi_2^r$ であり、これらの差が企業2から企業0に支払われるライセンス料になる。

$$\pi_2^d = (1 - \sqrt{k_0(\bar{x} + r)})^2, \quad \pi_2^r = (1 - \sqrt{\lambda k_0(\bar{x} + r)})^2$$

受諾されるライセンス契約 (s, F_T) を特定するためには、交渉を開始しな

図14 利益の差と機会費用の差の比較



$\sigma=3$, $\lambda=1.7$, $k_0=0.8$

った場合と、交渉が決裂した場合の企業2の投資を考慮した企業0の利益（ライセンス収入以外のもので、開発に成功した場合の期待利益）を考える必要がある。企業2とのライセンスが成立と交渉決裂のそれぞれの場合の企業0利得は以下である。

$$\pi_0^d = \frac{\bar{x}_0}{\bar{x}_0 + \bar{x}_1 + x_2^d + r} - k_0 \bar{x}_0, \quad \pi_0^r = \frac{\bar{x}_0}{\bar{x}_0 + \bar{x}_1 + \bar{x}_0 + r} - k_0 \bar{x}_0$$

ライセンスのロイヤルティーは企業0の機会主義的行動を阻止するようにならなければならない。アウトサイダーの場合と同様に右辺は企業2から要求できるライセンス料で、左辺は企業0の利益の差であるが、インサイダーなので自社の投資利益も含まなければならない。

$$\left(\frac{s \bar{x}_0}{\bar{x}_0 + \bar{x}_1 + \bar{x}_0 + r} + \bar{\pi}_0 \right) - \left(\frac{s \bar{x}_0}{\bar{x}_0 + \bar{x}_1 + x_2^d + r} + \pi_0^d \right) \geq \pi_2^d - \pi_2^r \quad (7)$$

シミュレーションをすると、不等式が成立しないパラメーターは見つからなかった。(図14右) $s=0$ に対しても成立するので、自社の投資利益の差 $\bar{\pi}_0 - \pi_0^d$ が右辺より大きくなければならない。企業0の利益は企業2の利益よりも大き

いは明らかであるが、投資が複雑に s に依存しているの、差の関係を証明するのは困難であるので、図14の左側のよりに、 $\pi_0 - \pi_0^* \geq \pi_1^* - \pi_1^*$ の右辺をプロットした。これは常に、

$$\pi_0 - \pi_0^* \geq \pi_1^* - \pi_1^*$$

が成立することを示唆している。完全開示（スピルオーバーなしから完全開示）による企業0の損失のほうが、企業2の限界の開示（スピルオーバーから完全開示）による利益の増分よりも大きいことである。単に企業2の費用の変化を考えると、 σh_0 から h_0 の効果のほうが λh_0 から h_0 の効果より大きいと考えられるが、さらに企業の費用が均一化（企業0と企業1の費用はともに h_0 である）することによって競争が激化する効果もあるはずである。

事後的ロイヤルティーを使う必要がないので、ライセンス料は企業1の残りの利益とライセンスが成立しなかった場合の利益（特許化された場合）の差を固定料金として課せばよい。企業0のライセンス収入 T は、以下のようになる。

$$\begin{aligned} T &= F_T + \frac{s\lambda}{\sigma_0 + \lambda_1 + \lambda_2 + \tau} = \pi(h_0, h_0, \sigma h_0) - \pi(h_0, h_0, \lambda h_0, h_0) \\ &= \pi(h_0, h_0, \sigma h_0) - \pi(h_0, h_0, \lambda h_0). \end{aligned}$$

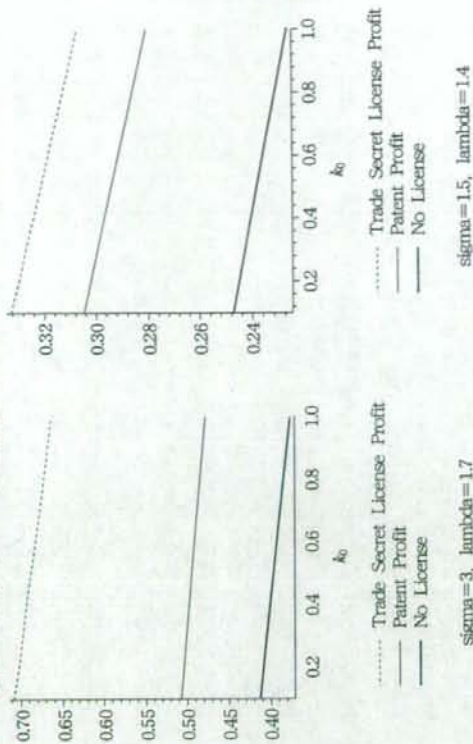
2つ目の等式は $\pi(h_0, h_0, h_0)$ の定義(4)による。明らかに、 $F_T > F_P$ である。企業秘密のライセンスをするのがいつでも可能なだけでなく必ず特許化より利益が大きい。

スピルオーバーがひどくなると (λ が大きくなると)、特許、企業秘密に関係なくライセンス料が低くなる。特許ライセンスのライセンス料はライセンスがある利益とない利益の差であるので、スピルオーバーが大きいかほど差が縮まってしまふからである。これはライバルの企業の費用が上がるのは企業にとつて有利であり、既存の技術が悪いほど (σ が大きい) 保護方法に関係なく企業0の利益は増加することをシミュレーションが示している。

3.3 ライセンスをしない場合

すでにインサイダーの場合は自社の利益がライセンス行動の決定に大きく影響をしていることを見たが、インサイダーの場合はそもそもライセンスをせず、

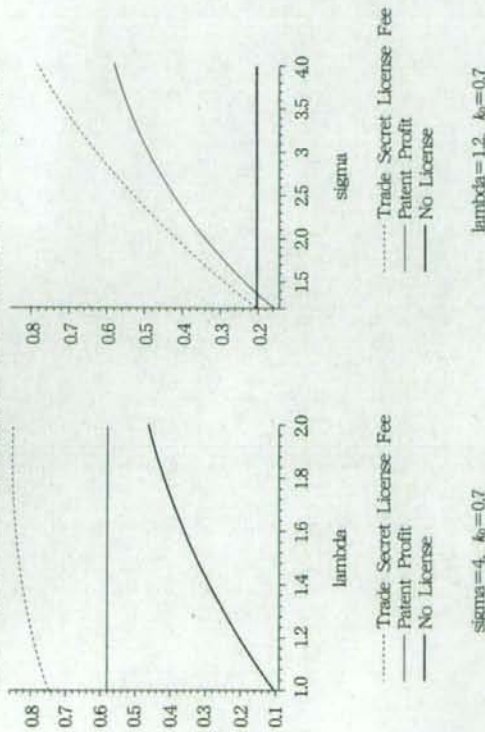
図15 特許ライセンス、企業秘密ライセンス、無ライセンスの比較 (1)



自社のみ中間技術を使うことが可能である。ライセンスをすることによってライセンス収入が入るからであるが、競争相手の費用を減らして、自分の立場を不利にしているからである。本項ではそもそもライセンスをするかという判断を分析する。特許化せずに、企業秘密にしたままライセンスもなかった場合は企業1と2は σh_0 の費用で技術開発投資をするので、企業 i の利得は(4)において、 $\pi(h_0, \sigma h_0, \sigma h_0)$ である。

まず、特許化した場合は必ず図9に示されているようにライセンスをして、特許料として利得(5)と(6)を得る。特許化せず、企業秘密の保護を選ぶと、ライセンスするかしないか選ぶ必要がある。ライセンスした場合の展開は図11と同じである。ライセンスをしないと、企業秘密のまま、費用係数は $h_0 = h_0 = \sigma h_0$ で、同時に投資を行う。均衡利得は(4)で計算できる。特許ライセンス、企業秘密ライセンスと企業秘密のままの利得をシミュレーションしたのが図15と図16である。 σ が小さい場合、つまり既存技術の性能がよい場合は特許ライセンスをしたほうがよい。競争相手のコストがともとも低いので、特許ライセンスをしても競争は前よりそれほど激化せず、ライセンス料が入るからである。

図16 特許ライセンス、企業秘密ライセンス、無ライセンスの比較 (2)



逆にそれ以外の場合は特許化よりも企業秘密がよい。しかし、それよりも企業秘密をライセンスすることのほうがさらに利益が大きい。競争相手のコストを下げても、インサイダーの場合は事後的ロイヤルティが必要ないのでライセンス収入が大きくなるからである。

4 おわりに

本論文ではインサイダーとアウトサイダーの特許と企業秘密のライセンス行動を、両者の区別はスピルオーバーの起こり方によって交渉過程と実施できるライセンス契約が異なるという解釈に基づいて分析した。特許の場合は権利取得の段階で情報はすでに開示されているが、そのような開示が起きていない企業秘密の場合は、1つの企業にライセンスした後も、まだ他の企業にとつて情報は価値がある。しかし、ライセンス交渉すること自体が情報が漏れる原因なることを利用して正しいインセンティブを与える契約を結ぶことによつてさ

らなる情報の開示を防ぐことができる。この余分の制約が企業秘密ライセンスのコストであるが、スピルオーバーが大きいほど別の企業に情報を売るといふ機会的行為が防ぎやすい。よつて、アウトサイダーはスピルオーバーが十分に大きくなければ企業秘密のライセンスができない。

さらに、独自で開発投資ができないアウトサイダーにとつては企業秘密のライセンスが可能なのは、企業秘密ライセンスのほうが特許ライセンスよりも利益が大きいたことがわかつた。企業秘密の場合は維持のためにロイヤルティによるディインセンティブを与えコストがあるものの、ロイヤルティが既存の技術という高い投資費用のままであるのが非常に有利である。これに対して特許の場合は開示があるので、ライセンスのない企業のコストが低くなつている。

中間技術所有者が独自でも開発投資を行っているインサイダーの場合は、まずライセンスを増やすことは競争相手の投資費用を下げて、自分が技術開発に成功する確率を下げる。そのため、情報所有者が第2の企業に企業秘密情報を提供するインセンティブは制約とはならない。さらなるライセンスを防ぐ必要がないので、スピルオーバーが非ライセンス企業にならない企業秘密のほうが当然利益が高い。企業秘密はまったくライセンスをしなければ、すべてのライバルが既存技術のままであるが、それよりは企業にライセンスをしたほうがよいことがわかつたが、特許ライセンスは既存技術がかなりよい場合(投資費用が低い場合)はしないほうがよい。この場合は投資による利益は開示がない利益が低く、開示してもあまり低下せず、その損失以上に特許収入が入るからである。

ライセンス交渉過程をモデル化することによつて、特許ライセンスが企業秘密ライセンスより勝る場合や、ロイヤルティの設定のされ方などを分析することができた。残念ながらシミュレーションにたよつており、また、パラメータ範囲の特定化が不十分など、未完成的側面があるのが残念である。序論で述べたように、技術ライセンスについてはデータを既存の理論を使って説明・理解する必要がある。このようなアプローチの紹介が今後の実証分析を含む知識財と中間技術ライセンス契約の研究の刺激になれば幸いである。

補 遺

1. x_i の 導 出

企業 i の投資水準が $x_i, i=1, \dots, n$ とする。同企業の利益は、以下のようになる。

$$\pi_i = \frac{x_i}{\sum_{j=1}^n x_j + r} - k_i x_i.$$

最大化の一階条件は、 $x = \sum_{j=1}^n x_j$ とすると、

$$\frac{\partial \pi_i}{\partial x_i} = \frac{x+r-x_i}{(x+r)^2} - k_i = 0, \quad i=1, \dots, n \quad (8)$$

となる。 i について加算すると、

$$n(x+r) - x = \sum_{j=1}^n k_j (x+r)^2$$

となる。 $k = \sum_{j=1}^n k_j$ を代入して、 x の方程式として解いた解は以下のようになる。

$$x = x(k_1, \dots, k_n) = \frac{(n-1) + \sqrt{4kr + (n-1)^2}}{2k} - r. \quad (9)$$

これを(8)に代入すると、均衡投資がわかる。

$$x^* = x(k_1, \dots, k_n) + r = k_1(x(k_1, \dots, k_n) + r)^2.$$

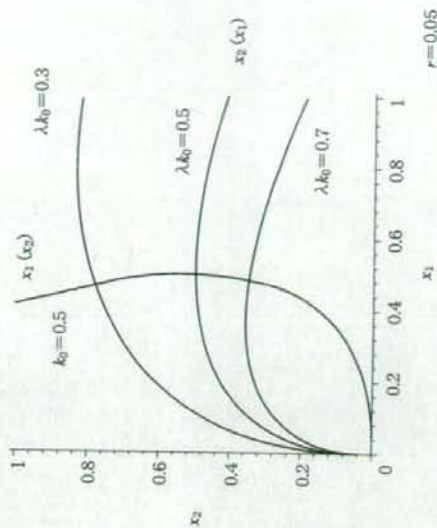
本文でも指摘したように、投資が変わったときの均衡投資の変化は大きさによる。(8)が最適反応関数を定義している。 $n=2$ の場合の最適反応関数を描いたのが、図17である。補足的、代替的部分がわかる。均衡における利益は、以下のようになる。

$$\pi_i^* = (1 - k_i(x(k_1, \dots, k_n) + r))^2.$$

2. アウトサイダーの企業秘密の投資と利益の導出

企業1の最適化の一階条件は開発費用が k_0 の代わりに $k_0/(1-s)$ であるのと同じになる。最適化の一階条件

図 17 最適反応関数 $x_2(x_1)$ と $x_1(x_2)$
best response



$$\frac{\partial \pi_1}{\partial x_1} = \frac{(1-s)(x_1+r)}{(x_1+x_2+r)^2} - k_0 = 0, \quad \frac{\partial \pi_2}{\partial x_2} = \frac{x_1+r}{(x_1+x_2+r)^2} - \sigma k_0 = 0$$

において、 $X = x_1 + x_2$ とおくと、以下のようになり、さらに X の方程式が導かれる。

$$x_1 = \sigma k_0 (X+r)^2 - r, \quad x_2 = \frac{k_0}{1-s} (X+r)^2 - r$$

$$\left(\frac{1}{1-s} + \sigma \right) k_0 (X+r)^2 = (X+2r)$$

X についてまず解き

$$X = \frac{-2 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1 + \sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0}$$

一階条件に代入すると x_1 と x_2 が得られる。 ¹²⁾

$$\begin{aligned} \bar{x}_1 &= \left(\frac{1 + \sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0} \right) \left[1 - \frac{k_0}{1-s} \left(1 + \frac{\sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0} \right) \right], \\ \bar{x}_0 &= \left(\frac{1 + \sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0} \right) \left[1 - \sigma k_0 \left(\frac{1 + \sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0} \right) \right]. \end{aligned} \quad (10)$$

第1項から第2項は $\bar{x}_1 + \bar{x}_0 + r = \left(\frac{1 + \sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0} \right)$ の関係によって導かれる。企業1の利益は、

$$\begin{aligned} \bar{\pi}_1 &= \pi(\bar{x}_1, \bar{x}_0; k_0) \\ &= \left(\frac{1 + \sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0} \right) \left\{ 1 - s - \frac{k_0}{1-s} \left(\frac{1 + \sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0} \right) \right\} \end{aligned}$$

となる。

企業1はすでに投資 \bar{x}_1 を行ってしまうので、 \bar{x}_1 を所与とした企業2単独の最適化問題として、(10)の \bar{x}_1 を代入すればよい。

$$\begin{aligned} x_2^d &= \sqrt{\sigma} \left(\frac{1 + \sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0} \right) - k_0 \left(\frac{1 + \sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0} \right)^2 \\ x_2^s &= \sqrt{\frac{\sigma}{\lambda}} \left(\frac{1 + \sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0} \right) - k_0 \left(\frac{1 + \sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0} \right)^2 \end{aligned}$$

開示の有無に対応する企業2の期待利益は、以下のとおりである。

$$\pi_2^d = \left\{ 1 - \sqrt{\sigma} k_0 \left(\frac{1 + \sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0} \right) \right\},$$

$$\pi_2^s = \left\{ 1 - \sqrt{\lambda \sigma} k_0 \left(\frac{1 + \sqrt{4 \left(\frac{1}{1-s} + \sigma \right) k_0 r + 1}}{2 \left(\frac{1}{1-s} + \sigma \right) k_0} \right) \right\}.$$

3 インサイダーの企業秘密の投資と利益の導出 最適化の一階条件は

$$\frac{\partial \pi_0}{\partial x_0} = \frac{(1-s)x_1 + x_0 + r}{(x_0 + x_1 + x_0 + r)^2} - k_0 = 0, \quad \frac{\partial \pi_1}{\partial x_1} = \frac{(1-s)(x_0 + x_0 + r)}{(x_0 + x_1 + x_0 + r)^2} - k_0 = 0,$$

$$\frac{\partial \pi_0}{\partial x_0} = \frac{x_0 + x_1 + r}{(x_0 + x_1 + x_0 + r)^2} - \sigma k_0 = 0$$

以下のように書き換えられる。

$$\frac{x_1 + x_0 + r - s x_1}{(x_0 + x_1 + x_0 + r)^2} = k_0, \quad \frac{x_0 + x_0 + r}{(x_0 + x_1 + x_0 + r)^2} = \frac{k_0}{1-s},$$

$$\frac{x_0 + x_1 + r}{(x_0 + x_1 + x_0 + r)^2} = \sigma k_0$$

さらに、 $z = x_0 + x_1 + x_0 + r$ として代入し、それぞれ $x_0 + x_0 + r = z - x_1$ などを使って x_1 について解くと、

$$\bar{x}_0 = z - k_0 z^2 - s \left(z - \frac{k_0}{1-s} z^2 \right), \quad \bar{x}_1 = z - \frac{k_0}{1-s} z^2, \quad \bar{x}_0 = z - \sigma k_0 z^2 \quad (11)$$

が得られる。3つの式を一緒にして得られる方程式、

$$(2k_0 + \sigma k_0) z^2 + (s-2)z - r = 0$$

を解くと、 z が得られる。

$$z = \bar{x}_0 + \bar{x}_1 + \bar{x}_0 + r = \frac{2-s + \sqrt{(2-s)^2 + 4k_0(2+\sigma)r}}{2k_0(2+\sigma)}$$

企業2にとって企業1と企業0の投資は所与であるので定数である。 $\bar{x}_1 = \bar{x}_0 + \bar{x}_1$ とすると、最適投資は以下のようなになる。

$$\bar{x}_2^d = \arg \max_{x_2} \frac{x_0}{\bar{x}_1 + x_0 + r} - k_0 x_2 = \sqrt{\frac{\bar{x}_1 + r}{k_0}} - (\bar{x}_1 + r),$$

$$\bar{x}_2^s = \arg \max_{x_2} \frac{x_0}{\bar{x}_1 + x_0 + r} - \lambda k_0 x_2 = \sqrt{\frac{\bar{x}_1 + r}{\lambda k_0}} - (\bar{x}_1 + r).$$

(11) から,

$$\begin{aligned} \bar{x} &= \bar{x}_0 + \bar{x}_1 = (2-s-2k_0z)z \\ &= \left(\frac{(1+\sigma)((2-s)z + \sqrt{(2-s)^2 + 4k_0(2+\sigma)z})}{(2+\sigma)} \right) \\ &\quad \left(\frac{(2-s + \sqrt{(2-s)^2 + 4k_0(2+\sigma)z})}{2k_0(2+\sigma)} \right) \end{aligned}$$

がわかる。

【参考文献】

- Anton, James, and Dennis Yao [1994] "Expropriation and Inventions: Appropriate Rents in the Absence of Property Rights," *American Economic Review*, 84(1), 191-209.
- Arundel, Anthony, and Isabelle Kabla [1998] "What Percentage of Innovations are Patented?" *Research Policy*, 27, 611-624.
- Bhattacharya, Sudipto, and Segei Guriev [2006] "Patents vs. Trade Secrets: Knowledge Licensing and Spillover," *Journal of the European Economic Association*, 4(6), 1112-1147.
- Bolton, Patrick, and Michael Whinston [1993] "Incomplete Contracts, Vertical Integration, and Supply Assurance," *Review of Economic Studies*, 60, 121-148.
- Chabchoub, Norhene, and Jorge Niosi [2005] "Explaining the Propensity to Patent Computer Software," *Technovation* 25, 971-978.
- Cohen, Welsley, Richard Nelson, and John Walsh [2000] "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)," National Bureau of Economic Research, NBER Working Paper No. 7552.
- Goto, Akira, and Akiya Nagata [1997] "Appropriability and Technological Opportunity in Innovation," NISTEP Research Report No. 48.
- Jensen, Richard, and Richard Thursby [2001] "Proofs and Prototypes for Sale: The Licensing of University Inventions," *American Economic Review*, 91(1), 240-259.

Kamien, Morton, and Yair Tauman [2002] "Patent Licensing: The Insider Story," *Manchester School*, 70(1), 7-15.

Nakamura, Kenta, and Hiroyuki Odagiri [2005] "R&D Boundaries of the Firm: An Estimation of the Double-Hurdle Model on Commissioned R&D, Joint R&D, and Licensing in Japan," *Economics, Innovation and New Technology*, 14(7), 583-615.

Scotchmer, Suzanne [2004] *Innovation and Incentives*, MIT Press.

Spiegel, Yossi [2007] "Licensing Interim R&D Knowledge," Recamati Graduate School of Business Administration, Tel Aviv University.

Suzuki, Jun, Kiminori Gamba, Schumpeter Tamda, Yoshihito Yasaki, and Akira Goto [2006] "Analysis of Propensity to Patent and Science-Dependence of Large Japanese Manufacturers of Electrical Machinery," *Scientometrics*, 68(2), 265-288.

青木治子・矢崎敬人 [2007] 「特許・知財の法と経済学」『経済研究』58(3)263-283.

伊藤秀史 [2003] 『契約の経済理論』有斐閣.

長岡貞男・中村健太 [2006] 「医薬品上流発明のライセンス契約に関する分析」(大学における知的財産権研究プロジェクト研究成果報告書, 2006年3月).

ランクサイズ回帰の検定について

小西葉子・西山慶彦

多くの発展研究では都市サイズ、企業の資源や売上高の規模などの研究対象がパレート性を持つことを、ランクサイズ回帰で観察してきた。具体的には、順位対数値をその規模の対数値に回帰することにより、その係数が-1になるかを調べる。また、パレートの有無には、二次項の係数が0であることも条件となるので、本稿では、二次項を加えたものを回帰モデルとする。パレートの検証には、一次項、二次項それぞれ β の検定、一次項の係数が-1、二次項の係数が0という複合仮説が成立しているかを F 検定で調べる方法がある。しかし、分析対象がパレート分布に従う時、 F 係数が大きくなると、 β は発散してしまふため通常の β 検定を行えないことがわかっており、 F 検定も同様の問題が観察された。そこで本稿では、 F 係数の発散をシミュレーションによって検証し、ランクサイズ回帰の複合仮説を検証可能とし、パレートの検定の新たな手法として提案した。

JEL Classification: C12, C16, R12

1. はじめに

都市・地域経済学でランクサイズ回帰のあてはまりがよく、古くから応用されている分野に都市人口分布の分析がある。まず一國の都市の人口を大きい順に並べ替え、1位、2位、...と順位(ランク)をつける。ランクサイズ回帰とは、都市の人口規模の対数値を当該都市の順位(ランク)の対数値に回帰したものである。すると、多くの国において定数項がほぼサンプルサイズ n の対数値に等しく、傾きはほぼ-1に等しくなるという結果が得られる。つまり人口規模が1番大きい都市から順に2番目の都市は1/2の人口、3番目は1/3、...と減少していく。

$S_i, i=1, \dots, n$ をある國の都市の人口とし、 $S_{(i)}$ をそれを大きい順に並べ替えた順序統計量とする。つまり $S_{(1)} \geq S_{(2)} \geq \dots \geq S_{(n)}$ である。

$\log S_{(i)} \approx \alpha + \beta \log i, i=1, \dots, n$ (1)
 $\alpha > 0, \beta < 0$ のとき、(1)式はランクサイズ回帰モデルと呼ばれ、 $\alpha = -1$ のとき Zipf's law (ジプフの法則)が成り立っているといえる。またパレート分布のパラメータが1の場合も同様の関係が観察される。

このような現象は様々な分野やトピックで観察されており、経済学では所得や資産が上位層に集中することや、有能な少数の従業員が全体

の生産に大きく貢献することなどが事例として挙げられている。これは、パレートの80:20の法則とも呼ばれ、集中度や不平等度の指標の一つとして用いられてきたが、近年では、マーケティングや経営の分野でも広く用いられている。この方法には、自然科学や実験など大規模データが利用可能な分野でも応用されてきたが、近年マイクロデータの利用可能性が高まることによって、都市経済学以外の分野でも分析対象がパレートを等しく、傾きはほぼ-1にある種々の分布に従うかを否かを検証する研究が行われはじめた。齊藤・渡辺(2007)では、我が国の法人企業業の1/3をカバーする約82万社のデータを用いて企業間関係の分析を行っている。そこでは、企業間のネットワーク構造に着目して、当該企業と「仕入先」「販売先」「大株主」に関する企業間ネットワークに、パレートの性質をばらばら関係があることが発見されている。

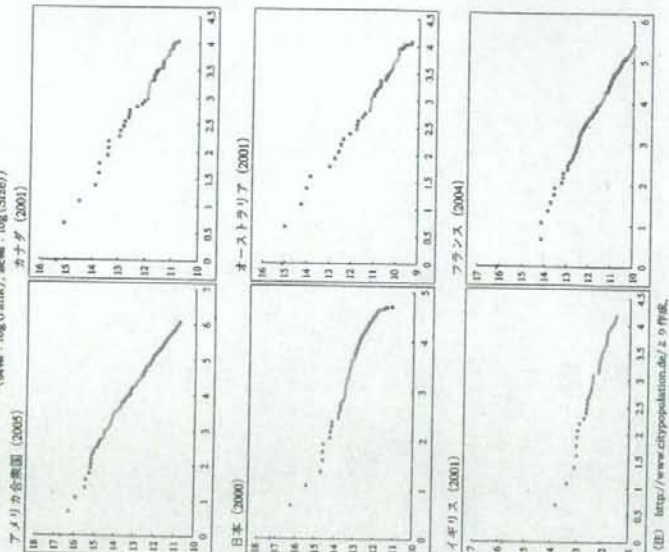
このようにランクサイズ回帰による分析が盛んな理由の一つは、最小二乗法中心の簡便な方法で分析ができることにある。

図1は各国の都市人口についての縦軸に人口の対数値、横軸にその大きさに対応するランクの対数値をとってプロットしたグラフである。グラフは概ね右下がりの直線になっている。実際、図1のデータで(1)式についての回帰を行った

ランクサイズ回帰の検定について

図1. 各国都市の人口規模と順位の数分布

(横軸: $\log(\text{rank})$, 縦軸: $\log(\text{Size})$)



場合、推定値が-1に非常に近い。都市経済学におけるランクサイズ回帰では、長い間通常の β 検定によって $\alpha = -1$ という帰無仮説が調べられてきたが、通常 β 値が非常に大きく、 α の推定値は-1にかなり近い実証研究が非常に多かった。この点に関して、数多くはないものの、先行研究が蓄積されてきている。詳しいことは後述するが、近年の研究から、都市のサイズがパレート分布に従っている場合にはランクサイズ回帰の β 統計量が漸近的に発散することがわかり、当然このような場合は、通常の β 検定の棄却域は使用できない。

他方、Rosen and Resnick(1980)等では、複合仮説がパレート分布に従っているかの簡便

的な検定として、(1)式の一次の項の二乗項を説明変数として加え、その推定値が0であれば、パレート分布に従っているとすると分析方法も提案されている。ここでは β 検定が用いられているが、この場合も上と同じ問題を含む。Nishiyama, Osada and Sato(2007)では、パレートの検定のために、一次の項と二次の項をそれぞれの帰無仮説の下で β 検定できるようにシミュレーションによって棄却域を求めた。しかし、この定式化で、パレートの性質を調べるのならば、一次項の係数が-1、二次項の係数が0という複合仮説が成立しているかを F 検定により調べるのが自然である。ただし、 β 検定の際に関連したように、真の分布がパレート分布であっても、サンプルサイズが大きくなると共に

F値が発散し、F検定を行うと帰無仮説を棄却しやすくなる懸念がある。

そこで、本稿では、同様に二乗の項を説明変数に考慮して推定を行い、F値の棄却域をシミュレーションによって構成し、ラングサイズ回帰のパラメータの検定手法として新たに提案する。

さらに企業の資産規模のデータを用いて、二次項も含めてラングサイズ回帰の実証分析を行う。また、本稿で得られたF値、F値の臨界値を用いて、パラメータの検定を行う。その際先行研究で指摘されていない、二次項を含んだ場合の説明変数間の相関の高さに着目した、一次項と二次項の相関係数について、漸近的にどのような挙動をとるのかを調べた新たな事実がわかった。

次節では、先行研究のレビューを行う。3節ではラングサイズ回帰のシミュレーションを行う。4節は統計量とF統計量の棄却域を構成する。またそれに基づき実証分析を行う。4節では、結論と今後の課題、付録では実証結果で得られた知見より、ラングサイズ回帰の説明変数間の相関係数の挙動について調べている。

2. 先行研究

経済学でラングサイズ回帰のあてはまりが上がある。これは、都市の人口規模の対数値をその大ききのラングの対数値に回帰すると、定数はほぼ-1に等しくなるというものである。この関係は、都市規模がi.i.d.でパラメータの値が1のパラメータ分布に収まっているときに成立することが知られている。この文脈で行ったのはRosen and Resnick(1980)であり、現在もなおこの分野で必ず引用される文献である。Soo(2005)はそれを更新したデータについて調べている。これらの論文では、最小二乗推定(OLS)によりラングサイズ回帰を行った点推定値を得ている。それと同時に、1.ラングの対数の係数に関する検定に基づいて、傾きが-1であるという仮説の検定、および2.回帰モデルにラングの対数とその2乗項を説明変数に含めて後者

の係数がゼロかどうかをF検定で調べることにする。パラメータの検定が行われている。しかし、被説明変数は順序統計量であり、その結果、定義上被説明変数は、分散不均一と自己相関をもつ。そのため、古典的な回帰理論を適用するとはできない。それらを考慮して推定する性質を調べたのがGabaix and Ioannides(2004)、Gabaix and Ibragimov(2007)などであり、一貫性、漸近正規性が証明されている。また、Nishiyama and Osada(2005)、Nishiyama, Osada and Sato(2007)は、OLSよりも有効性のある推定方法としてtrimmed OLSとGLS法による推定方法を提案している。

それらの研究において、推定に関しては実証統計的にはあまり大きな問題は生じないことが示されているが、検定に関しては、Nishiyama and Osada(2004)、Nishiyama, Osada and Sato(2007)らが分散不均一と自己相関のために、傾き-1であるという帰無仮説を標準的なF検定で調べることができないことを示している。加えて、帰無仮説の下でも $\beta = \sum(\text{回帰残差})^2 / (n - \text{説明変数の数})$ が漸近的にゼロに収束するという問題が指摘されている。そのため、真の分布がパラメータ分布であっても、サンプルサイズが大きくなると共にF値が発散し、F検定を行うと帰無仮説を棄却しやすくなる。この問題を回避するため、それらの論文では修正したF検定が提案されている。そこでは、通常のF値が検定統計量として用いられているが、棄却域はシミュレーションによって構成されている。Nishiyama, Osada and Sato(2007)では、パラメータの検定のために、1.で述べたように説明変数に二乗の項も含み、F検定と二次の項をそれぞれ帰無仮説の下でF検定できるようにシミュレーションによって棄却域を求めた。しかし、この定式化で、パラメータを調べるならば、一次項の係数が-1、二次項の係数が0という複合仮説が成立しているかをF検定により調べるのが自然である。ただし、F検定の際に問題になったように、ここでは帰無仮説の下でも $\beta = \sum(\text{回帰残差})^2 / (n - \text{説明変数の数})$ が漸近的にゼロに収束し、真の分布がパラメータ

表1. シミュレーション結果: 係数推定 (a, a_1, a_2, b, b_1, b_2), 各F統計量 (a, a_1, a_2, b, b_1, b_2), 回帰残差の二乗和(SSR)の期待値と標準偏差

一次項に対するF統計量: (1)式の帰無仮説は $a_1 = -1$, (4)式の帰無仮説は $a_1 = -1$,
二次項に対するF統計量: (2)式の帰無仮説は $a_2 = 0$, (4)式の帰無仮説は $a_2 = 0$,
nはサンプル数、繰り返し回数は1000回。

	n=50					(3)式: $\log(\text{Size}) = c + a \cdot \log(\text{rank}) + a_1 \cdot \log^2(\text{rank})$					(4)式: $\log(\text{rank}) = c + b \cdot \log(\text{Size}) + b_1 \cdot \log^2(\text{Size})$				
	a	a ₁	a ₂	ssr	df	a	a ₁	a ₂	ssr	df	b	b ₁	b ₂	ssr	df
平均	-1.306	0.047	-1.688	1.746	-1.007	0.006	-0.313	3.306	0.014						
中央値	-1.191	0.029	-2.240	1.847	-1.007	0.015	-0.157	2.563	0.012						
標準偏差	0.748	0.130	8.127	6.769	0.027	0.100	0.651	7.711	0.008						
分散	0.560	0.017	66.048	45.814	0.001	0.073	0.421	59.457	0.000						
実数	1.492	1.414	0.843	0.513	66.892	0.474	7.235	0.053	5.430						
変数	-0.952	0.905	0.337	-0.276	0.055	-1.866	-0.153	0.411	1.743						
範囲	6.760	1.213	76.305	61.368	0.985	2.931	1.358	52.552	71.939	0.068					

	n=500					(3)式: $\log(\text{Size}) = c + a \cdot \log(\text{rank}) + a_1 \cdot \log^2(\text{rank})$					(4)式: $\log(\text{rank}) = c + b \cdot \log(\text{Size}) + b_1 \cdot \log^2(\text{Size})$				
	a	a ₁	a ₂	ssr	df	a	a ₁	a ₂	ssr	df	b	b ₁	b ₂	ssr	df
平均	-1.137	0.012	-4.931	4.176	0.005	-1.027	0.010	-4.946	11.394	0.004					
中央値	-1.105	0.010	-6.539	5.834	0.004	-1.028	0.014	-4.700	9.696	0.003					
標準偏差	0.325	0.033	20.752	18.418	0.006	0.064	0.028	15.658	22.521	0.002					
分散	0.106	0.001	430.636	339.232	0.000	0.009	0.001	245.161	507.188	0.000					
実数	0.272	0.291	0.842	0.712	78.411	0.019	0.793	0.235	1.414	7.550					
変数	-0.440	0.441	0.453	-0.446	6.347	0.124	-0.300	-0.103	0.618	2.124					
範囲	2.445	0.254	197.942	169.172	0.144	0.712	0.308	146.596	254.013	0.022					

	n=3000					(3)式: $\log(\text{Size}) = c + a \cdot \log(\text{rank}) + a_1 \cdot \log^2(\text{rank})$					(4)式: $\log(\text{rank}) = c + b \cdot \log(\text{Size}) + b_1 \cdot \log^2(\text{Size})$				
	a	a ₁	a ₂	ssr	df	a	a ₁	a ₂	ssr	df	b	b ₁	b ₂	ssr	df
平均	-1.059	0.004	-8.881	9.065	0.001	-1.016	0.005	-12.316	20.961	0.001					
中央値	-1.054	0.004	-13.205	12.422	0.001	-1.016	0.005	-12.719	20.460	0.001					
標準偏差	0.171	0.013	43.271	39.578	0.001	0.043	0.012	33.020	44.795	0.001					
分散	0.029	0.000	1872.362	1566.387	0.000	0.002	0.000	1090.311	2006.587	0.000					
実数	0.112	0.131	1.010	0.840	41.000	-0.095	0.113	0.108	0.453	8.419					
変数	-0.216	0.231	0.453	-0.446	4.543	-0.012	-0.337	0.022	0.165	2.130					
範囲	1.350	0.103	461.735	418.064	0.028	0.348	0.068	271.673	436.014	0.008					

分布であっても、サンプルサイズが大きくなる

と共にF値が発散し、F検定を行うと帰無仮説を棄却しやすくなる懸念がある。

そこで、本稿では、同様に二乗の項を説明変数に考慮して推定を行い、F値の棄却域をシミュレーションによって構成し、パラメータの検定の手法として提案する。

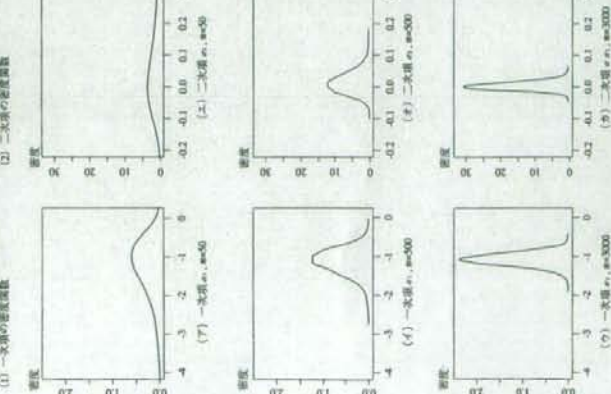
これらの提案が、非常に初歩的なことからわかるようにラングサイズ回帰は、汎用的で長い間多くの分野で利用されてきているが、その統計的性質が明らかになっていないのが現状である。

3. モンテカルロシミュレーションと実証分析

前節で述べたように、Gabaix and Ioannides(2004)、Gabaix and Ibragimov(2007)で、(2)

式のタイプのラングサイズ回帰の推定値の一致性と漸近正規性が証明された。Nishiyama, Osada and Sato(2007)は、パラメータの検証のために、Rosen and Resnick(1980)やSoo(2005)タイプの(4)式についてシミュレーションを行い、帰無仮説を一次の項が-1、二次の項が0としてそれぞれのF値の臨界値を求めている。さらにSoo(2005)がアダプテッドしたデータを用いてパラメータ推定検定を行っているが、その推定値の挙動や問題点については指摘していない。ここで(4)式は通常定義されるラングサイズ回帰の逆回帰となっていることに注意されたい。ただし、逆回帰であっても、一次の項に関して(2)式が有する統計的性質はおそらく変わらないと予想されるが、理論的な検証は今後の課題である。

図2. (2)式 $\log(\text{Size}) = c + \alpha \log(\text{rank}) + \alpha_1 \log^2(\text{rank})$ の推定値のシミュレーション結果、繰り返し回数: 10000回



てOLSを行った。繰り返し計算は10000回行っている。

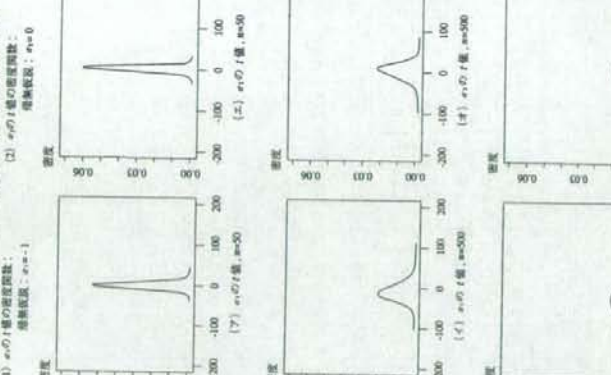
$$\log S_{0i} = c + \alpha \log i + \alpha_1 \log^2 i, \quad i = 1, \dots, n \quad (2)$$
$$\log S_{1i} = c + \alpha_1 \log i + \alpha_2 \log^2 i, \quad i = 1, \dots, n \quad (3)$$
$$\log S_{2i} = c + \beta_1 \log S_{0i} + \beta_2 \log^2 S_{0i}, \quad i = 1, \dots, n \quad (4)$$

表1はサンプルサイズごとのシミュレーション結果で、記述統計量である、 $\alpha, \alpha_1, \beta_1, \beta_2$ は推定値、 α_0 は帰無仮説が $\alpha = -1$ 、 β_1 は帰無仮説が $\alpha_0 = 0$ 、 β_2 は帰無仮説が $\beta_2 = 0$ の両側検定を行ったときのF統計量、SSRは回帰の残差二乗和である。

α については、サンプルサイズが50, 500, 3000と大きく異なる程、平均値が-1に近くなり、 α_1 は同様にサンプルサイズが大きくなると平均値が0に近くなり、両係数とも一貫性が成り立っているように見える。

図2は(3)式について $n = 50, 500, 3000$ のとき回帰を行い、その推定値の密度関数を描いたものである。左の列は一次項 α 、右の列は二次項 α_1 の密度関数である。サンプルサイズが大きくなるほど、一次項は-1、二次項は0の周りに推定値が集まっているのがわかる。(4)式の一次項 β_1 、二次項 β_2 (3)式の結果と同じ挙動になっている。一方、表1より(3)式、(4)式を通じて、サンプルサイズが大きくなる程F値の平均値が絶対値が大きくなり、範囲(レンジ)は広がっている。これは先行研究で指摘されているように、サンプルサイズが大きくなるとF値が均散傾向にあることと矛盾が大きい。図3は、左の列が一次項 α_0 の帰無仮説 $\alpha = -1$ の下での統計量の密度関数、右の列が二次項 α_1 の帰無仮説 $\alpha_0 = 0$ の下でのF統計量の密度関数である。サンプルサイズが大きくなるほど、分散が大きくなっている。Nishiyama, Osada and Sato (2007)で指摘されているように、F統計量の分母を構成する β が、 n が大きくなる程0に近づいていることが関係していると考えられる。

図3. (3)式 $\log(\text{Size}) = c + \alpha \log(\text{rank}) + \alpha_1 \log^2(\text{rank})$ のF値のシミュレーション結果、繰り返し回数: 10000回



係数に関して帰無仮説が棄却される。シミュレーションで得られた境界値は、2よりかなり大きく、サンプル数の増加に伴って大きくなっており、通常のF統計量の境界値を用いるより帰無仮説を棄却しにくくなっている。以上より、順帰帰(3)式であっても逆帰帰(4)式であっても、F検定を行う場合は通常の棄却域は用いることが適切でないことわかった。

前述したとおり、(3)式と(4)式で、パラメータが1となるバレート性の成立を調べるのならば、一次項の係数が-1、二次項の係数が0という複合仮説が成立しているかを調べるのが自然である。ただし、F検定の際の問題になったように、 $S^2 = \sum(\text{回帰残差})^2 / (n - \text{説明変数の数})$ が漸近的にゼロに収束し、真の分布がバレート分布であっても、サンプルサイズが大きくなると共にF値が発散し、F検定を行うと帰無仮説を棄却しやすくなるということが懸念される。

図4は(3)式について、シミュレーションで発生させたデータを用いて帰無仮説 $\alpha = -1, \alpha_1 = 0$ (4)式について帰無仮説 $\beta_1 = -1, \beta_2 = 0$ の下で計算したF統計量の密度関数である。 n が大きくなるほど、F値が大きくなっていくことがわかる。表3はF統計量の棄却域である。

通常のF統計量は、制約の数が等しい場合、サンプルサイズが大きくなるほど、境界値が小さくなる。(3)式と(4)式で、パラメータ1のバレート性の成立を調べる場合、通常のF統計量の境界値は $n = 50$ 以上だとおおよそ3である。しかし、表3の結果では、F統計量と同じように、サンプル数が大きくなる程、境界値が大きくなっており、通常のF検定を行うと、帰無仮説 $\alpha = -1, \alpha_1 = 0$ 、帰無仮説 $\beta_1 = -1, \beta_2 = 0$ を棄却しやすくなってしまふ。よって、順帰帰(3)式であっても逆帰帰(4)式であっても、F検定を行う場合は通常の棄却域は用いることが適切でないことわかった。以上より、本稿では、シミュレーションで得たF値の棄却域

このようにバレート分布から発生させたデータを用いても、推定値に関してF検定を行うと、ランクサイズルールが成立している $\alpha = -1$ や $\alpha_1 = 0$ を棄却しやすくなってしまふ。そのため、シミュレーションによって得られた棄却域を用いることが解決策の一つになる。

表2は、(3)式、(4)式のランクサイズ回帰において、一次項、二次項の係数の有意性をF検定で検証するためのシミュレーションで得られた棄却域である。 α は-1、 α_1 は0が帰無仮説であり、各サンプルサイズに対する両側1%、5%、10%水準の境界値である。サンプル数が多くなるほど、境界値の絶対値は大きくなり、棄却域が狭くなっていく。F検定は、F値が絶対値で2より大きければ、当該

3.1 モンテカルロシミュレーション
各シミュレーションでは、パラメータが1のバレート分布から $n = 50, 100, 200, 500, 1000, 3000$ のデータを発生させ、(3)式、(4)式につい

表2. シミュレーションによるF値の漸近値(繰り返し回数は1000回)
 (1)式: $\log(\text{Size}) = c + a \cdot \log(\text{rank}) + \alpha \cdot \log^2(\text{rank})$, (4)式: $\log(\text{rank}) = c + \beta \cdot \log(\text{Size})$
 一次項に対する検定: (3)式の帰無仮説は $\alpha = -1$, (4)式の帰無仮説は $\beta = 0$ の両側検定。
 二次項に対する検定: (3)式の帰無仮説は $\alpha = 0$, (4)式の帰無仮説は $\beta = 0$ の両側検定。

n	α		β				
	Lower	Upper	Lower	Upper			
n=50	10%	[-0.01, -13.74]	(2.42, 00)	[-0.01, -3.81]	(8.05, 00)	[-0.01, -7.9]	(17.05, 00)
	5%	[-0.01, -16.37]	(16.03, 00)	[-0.01, -11.82]	(10.56, 00)	[-0.01, -9.8]	(20.67, 00)
	1%	[-0.01, -23.07]	(32.23, 00)	[-0.01, -16.49]	(14.16, 00)	[-0.01, -13.98]	(28.3, 00)
n=100	10%	[-0.01, -18.17]	(16.16, 00)	[-0.01, -13.98]	(15.70, 00)	[-0.01, -11.08]	(24.79, 00)
	5%	[-0.01, -21.36]	(31.38, 00)	[-0.01, -17.81]	(17.86, 00)	[-0.01, -16.87]	(33.37, 00)
	1%	[-0.01, -28.55]	(51.54, 00)	[-0.01, -25.94]	(22.49, 00)	[-0.01, -21.95]	(43.14, 00)
n=200	10%	[-0.01, -24.4]	(21.25, 00)	[-0.01, -18.6]	(20.8, 00)	[-0.01, -19.93]	(33.48, 00)
	5%	[-0.01, -29.02]	(40.5, 00)	[-0.01, -24.30]	(32.67, 00)	[-0.01, -22.22]	(43.26, 00)
	1%	[-0.01, -38.65]	(77.76, 00)	[-0.01, -29.68]	(54.37, 00)	[-0.01, -25.6]	(65.48, 00)
n=500	10%	[-0.01, -36.7]	(30.98, 00)	[-0.01, -27.73]	(32.05, 00)	[-0.01, -26.64]	(49.88, 00)
	5%	[-0.01, -42.17]	(40.32, 00)	[-0.01, -36.05]	(37.18, 00)	[-0.01, -36.30]	(61.06, 00)
	1%	[-0.01, -55.09]	(59.80, 00)	[-0.01, -44.63]	(48.15, 00)	[-0.01, -46.92]	(85.13, 00)
n=1000	10%	[-0.01, -48.84]	(40.10, 00)	[-0.01, -36.28]	(43.24, 00)	[-0.01, -41.96]	(26.17, 00)
	5%	[-0.01, -57.83]	(50.79, 00)	[-0.01, -47.03]	(50.54, 00)	[-0.01, -49.54]	(52.94, 00)
	1%	[-0.01, -75.03]	(75.61, 00)	[-0.01, -60.23]	(60.43, 00)	[-0.01, -66.78]	(44.90, 00)
n=2000	10%	[-0.01, -75.09]	(66.25, 00)	[-0.01, -59.92]	(68.60, 00)	[-0.01, -65.91]	(42.09, 00)
	5%	[-0.01, -87.96]	(83.91, 00)	[-0.01, -76.78]	(79.53, 00)	[-0.01, -78.67]	(82.71, 00)
	1%	[-0.01, -115.14]	(126.25, 00)	[-0.01, -113.91]	(104.92, 00)	[-0.01, -97.50]	(72.97, 00)

表3. シミュレーションによるF値の漸近値(繰り返し回数は1000回)

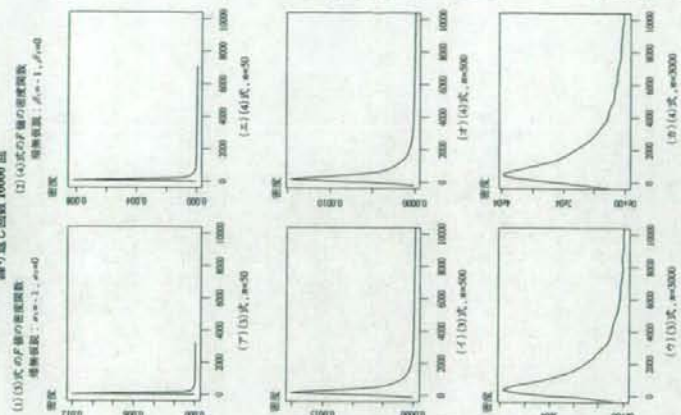
n	(1)式				(4)式			
	5%	1%	10%	5%	10%	5%	1%	
50	38.34	279.13	577.85	374.16	603.09	1407.78		
100	397.18	574.69	1209.84	776.63	1246.99	2638.46		
200	554.05	800.71	1575.5	1011.48	1584.85	3464.72		
500	1113.95	1621.68	3187.17	1853.66	2888.15	5888.67		
1000	2006.01	2922.18	5405.54	2945.61	4500.72	9021.17		
3000	4597.47	6715.73	12816.42	6197.66	8922.91	16318.68		

をパレート性の検定手法として提案する。

3.2 実証研究

ここでは、日経 NEEDS の2006年の上場企業の資産のデータを用いてパラメータ1のパレートの検定を行う。サンプル数は1736社で

図4. (3)式、(4)式のF値のシミュレーション結果: 繰り返し回数10000回



パラメータ1のパレート分布に従う場合、対数をとって回帰モデルの形にすると、一次項のパラメータが-1の(2)式の形になる。(2)式の推定値は-1.05で検定でも棄却されなかった。そこでさらに、二次項を加えその係数が0になるか否かを確認したのが(3)式である。検定結果では、F検定でパレート性は棄却されなかった。F検定においても、一次項、二次項それぞれの帰無仮説も棄却しなかったが、 α_1 は0.4で、符号が逆となり、値が大きく異なった。このように症状的な要因として多重共線性が考えられる。特にランク回帰では、線形(ランク)とlog(ランク)を説明変数に含むため多重共線性が起きやすい。実際サンプル1000では、説明変数間の相関は0.989と非常に1に近い。

4. おわりに

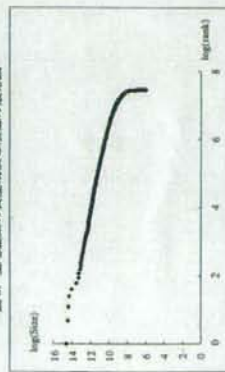
本稿では、古くから様々な分野で応用されてきたランクサイズ回帰のパレート性の検定手法を提案した。多くの実証研究では都市サイズ、企業の資産や売上高の規模などの研究対象がパレート性を持つことを、ランクサイズ回帰を行うことで観察してきた。具体的なには、規模対数値にその観念の対数値を回帰することによって、その係数が-1になることを調べる。また、パレート性の有無には、二次項の係数が0であることも条件になるため、本稿では、二次項を加えて推定を行った。パレート性の検定には、一次項、二次項それぞれ検定し、一次項の係数が-1、二次項の係数が0という複合仮説が成立しているかをF検定で調べる方法がある。しかし、分析対象がパレート分布に従う時、サンプルサイズが大きくなると、F値は発散してしまうため通常の検定を行うことはできないことがわかっており、F検定でも同様の問題が観察された。

そのため本稿では、F値の棄却域をシミュレ

表4. 上場企業の資産データ(2006年)を用いたランクサイズ回帰結果とパレート性の検定
サンプル数: 175, 5%有意水準で行っている。

式番号	検定結果		通常の重関数		シミュレーションによる重関数	
	一次項 (F値)	二次項 (F値)	F検定 (F検定)	一次項 (F検定)	二次項 (F検定)	F検定
(2)	0.401 (31.40)	-0.132 (-6.18)	棄却されない	棄却されない	棄却されない	棄却されない
(3)	0.401 (31.40)	-0.132 (-6.18)	棄却されない	棄却されない	棄却されない	棄却されない
(4)	2.143 (139.30)	-0.152 (-132.78)	棄却	棄却	棄却	棄却

図5. 上場企業の資産規模と順位の数直図



シミュンによって構成し、一次項の係数が-1, 二次項の係数が0という複合仮説を検証可能とし、パレート性の検定の新たな手法として提案した。これが本稿の主な貢献である。

実証研究では、本稿で得られたF値、F値の臨界値を用いて二次項も含めたランクサイズ回帰を行ったところ、検定によりパレート性があることが検証されたが、二次項を含んだ場合に多重共線性の症状がみられた。このことより、一次項と二次項の相関係数について調べ、漸近的に相関係数が1となることが明らかになり、このことが、実証結果で推定値の大きさに影響を与えている可能性があることがわかった。

今後の課題は、二次項とF統計量に関する統計的性質を明らかにすること、一次項と二次項の相関係数が漸近的に1になり完全な多重共線性を起こすことが推定値や検定統計量にどのような影響を与えるのかを調べることである。

付録 説明変数間の相関について

以下ではnが大きくなったときの相関係数を調べる。証明には以下のレンマを用いる。各レンマの証明は、Nishiyama and Osada (2004), Nishiyama, Osada and Sato (2007) で示されている。

$$+3 \sum \left(\frac{\log i}{i} - \frac{\log i}{n} + \sum \left(\frac{1}{n} - \frac{3}{2i^2} \right) \right)$$

$$= n \log^2 n - \left[3(n \log^2 n - 2n \log n + 2n) - \frac{3}{2} \left(\frac{\log^2 n}{n} + 3 \left(\frac{\log^2 n}{n} \right) \right) \right]$$

$$= n \log^2 n - 3n \log^2 n + 6n \log n - 6n$$

$$\text{よって、分子第1項は以下のようになる。}$$

$$\frac{1}{n} \sum \log^2 i = \log^2 n - 3 \log n + 6 \log n - 6$$

分子の第2項は、レンマ(a), (b)より

$$\left(\frac{1}{n} \sum \log i \right) \left(\frac{1}{n} \sum \log^2 i \right) \approx \frac{1}{n} (n \log n - n$$

$$+ \frac{1}{2} \log n) \frac{1}{n} (n \log^2 n - 2n \log n + 2n)$$

$$\approx (\log n - 1) (\log^2 n - 2 \log n + 2)$$

$$= \log^2 n - 3 \log^2 n + 4 \log n - 2$$

となる。よって分子は、 $2 \log n - 4$ である。レ

ンマにより、分子は以下の近似が成立する。

$$\frac{1}{n} \sum \log^2 i - \left(\frac{1}{n} \sum \log i \right) \left(\frac{1}{n} \sum \log^2 i \right)$$

$$= \log^2 n - 3 \log^2 n + 6 \log n - 6 + o(1)$$

$$= -2 \log^2 n - 6 \log n + 6 + o(1)$$

$$= 4 \log^2 n - 16 \log n + 20 + o(1)$$

次に、分母 $\frac{1}{n} \sum \log^2 i - \left(\frac{1}{n} \sum \log i \right)^2$ と $\frac{1}{n} \sum \log^2 i$

$$- \left(\frac{1}{n} \sum \log i \right)^2$$

レンマ(b)より、

$$\frac{1}{n} \sum \log^2 i - \left(\frac{1}{n} \sum \log i \right)^2$$

$$= \log^2 n - 2 \log n + 2 + \frac{\log n}{2n} - \log n - \log n + O(n^{-1})$$

$$= -\left[\log n - 1 + \frac{\log n}{2n} + O(n^{-1}) \right]$$

$$= 1 - \frac{\log n}{2n} + O(n^{-1})$$

$\frac{1}{n} \sum \log^2 i$ の近似で用いたのと同様の要請に

$$\text{より、} \sum \log^2 i = n \log^2 n - 4n \log n - 3n \log n$$

$$+ 6n \log n - 6n + O(\log n)$$

従って、レンマ(b)を用いると

$$\frac{1}{n} \sum \log^2 i - \left(\frac{1}{n} \sum \log i \right)^2$$

$$= \log^2 n - 4 \log^2 n - 3 \log^2 n + 6 \log n - 6 + o(1)$$

$$= -\log^2 n - 2 \log n + 2 + o(1)$$

$$= 4 \log^2 n - 16 \log n + 20 + o(1)$$

となる。よって、相関係数の分母は

$$\sqrt{1 - \frac{\log^2 n + O(n^{-1})}{4 \log^2 n - 16 \log n + 20 + o(1)}}$$

$$= \sqrt{4 \log^2 n - 16 \log n + 20 + o(1)}$$

となり、相関係数の2乗は以下のようになる。

$$\rho^2 = \frac{4 \log^2 n - 16 \log n + 16 + o(1)}{4 \log^2 n - 16 \log n + 20 + o(1)}$$

$$\approx 1 - \frac{\log^2 n - 4 \log n + 5}{2 \log^2 n - 16 \log n + 20 + o(1)}$$

テイラー展開を用いて、相関係数は

$$\rho \approx 1 - \frac{1}{2} \frac{\log^2 n + O(\log^2 n)}{\log^2 n}$$

と近似される。以上より、ランクサイズ回帰に

おいては、説明変数間の相関係数が漸近的に1に

なることがわかった。

(独立行政法人経済産業研究所・
京都大学経済研究所)

注

* 一橋大学経済研究所、定例会報告(2008年2月27日)に対する討論者であった野口輔子氏のコメントに感謝します。また、同定例会の参加者から多くの有意義なコメントを頂きました。記して感謝致します。

参考文献

- 阿部(雄野)希子・渡辺(秀) (2007) 「企業関係と企業規模」『経済研究』第53巻第4号, pp.302-313.
Gabaix X. and R. Ibragimov (2007) "Log (Rank-1/2): A Simple Way to Improve the OLS Estimation of Tail Exponents," Harvard Institute of Economic Research Discussion Paper No. 2106.
Gabaix X. and Y. M. Ioannides (2004) *The Evolution of City Size Distributions*, Handbook of Urban and Regional Economics, Vol.4, Chap.53.
Nishiyama Y. and S. Osada (2004) "Statistical Theory of Rank Size Rule Regression under Pareto Distribution," CAEA Discussion Paper 009, Kyoto University.
Nishiyama, Y., S. Osada and Y. Sato (2007) "OLS Estimation and the t Test Revisited in Rank-size Rule Regression," forthcoming in *Journal of Regional Science*.
Rosen K. T. and M. Rensick (1980) "The Size Distribution of Cities: An Explanation of the Pareto Law and Primacy," *Journal of Urban Economics*, Vol.8, Issue 2, pp.145-186.
Soo K. T. (2005) "Zipf's Law for Cities: A Cross Country Investigation," *Regional Science and Urban Economics*, Vol.35, Issue 3, pp.239-263.