

(調査研究報告書)

ビッグデータを用いた企業規模による倒産確率の算出と

適正な保険のありかた

石川 温 (金沢学院大学経営情報学部教授)

藤本 祥二 (金沢学院大学経営情報学部非常勤講師)

1.1 序論

企業の倒産確率の研究は、信用リスクの観点から、保険事業にとって非常に重要である。一般的に倒産確率は様々なモデルを利用して、過去に倒産した企業の特徴から統計的に算出され、より精緻なモデル構築と分析が研究されている。ここでは、倒産確率は非常に多くの説明変数の線形結合として記述できるとして、重回帰分析を用いて倒産確率を計算する[1]。重回帰分析を行うためには従属変数と独立変数の間に線形の関係がなければならない。しかしデータによっては非線形の関係が当てはまる場合がある。そのような場合、分析によって得られた指標が正しく計算されず、分析によって導き出される指標が母集団を的確に反映するものではなくなってくる。このような事態を避けるには従属変数と独立変数の間の線形関係を確かめる必要があるが非常に多くの説明変数全てにおいて線形性が成り立っている可能性はかなり少ないと考えられる。また、回帰係数の検定や信頼区間の推定には残差の正規性や分散の均一性が前提となっているが、説明変数の増やし方によってこの前提が成り立たなくなってしまうことがある。

本研究調査では、説明変数の線形結合を仮定せず、少数の説明変数で倒産確率を記述する。そして、倒産確率の変数依存性を分布という形で定性的・定量的に観測し、倒産の確率密度関数を明らかにしようとする。分布を直接見ることによって分析における仮定を少なくし、より信頼できる結果を得ることを目指す。説明変数には、企業の基本的な財務データであり、企業の規模の指標の一つである売上高、そして企業年齢を用いる。企業が倒産に至るには様々な過程がある。倒産原因を一概に特定することは難しい。しかし、経営が悪化する最初の原因に売上の減少が見られるため売上との関係を見ることは非常に重要である。さらに売上は従業員数や資産と並んで企業の規模を表す重要な変数である。売上との関係が分かればその他企業の規模を表す変数との関係も容易にわかる。また、倒産を免れてきた企業は必然的に企業年齢が高くなる。そのため年齢と倒産に関して高い相関があると考えられる。これらの説明変数の関数系が定性的および定量的に確定できれば、そこに

新たに説明変数を加えていくことにより、従来の倒産確率の算出手法を効率化することが可能になる。

近年、世界各国企業の網羅的な財務データが入手可能になってきたことにより詳細な分析が可能になってきた。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され、利用しやすい状況で管理されている。本研究では、この大規模データ(いわゆるビッグデータ)を分析することにより、この問題にアプローチする。大規模データを扱うことによって、自然科学の分野での統計物理学の手法を用いることができる。原子や分子の世界ではアボガドロ数と呼ばれる大量の粒子を扱う。個々の粒子は力学の原理に従いお互いに力を働かせ合う。この相互作用に関する力の方程式を立てることはできるが、大量の粒子に対する大量の方程式を解くことは難しい。しかし平均やゆらぎ等の集団的な性質に着目すれば大量の方程式をシンプルにまとめることができ、その性質を分析することが可能である。このような自然科学で用いられている手法を経済の分野に応用し、経済データを物理学的視点で分析する分野を経済物理学[2, 3]の分野が発展してきた。本研究では Bureau van Dijk 社が提供する世界中を包括した最大級のビッグデータを用いて分析を行う。大規模なデータによって統計的な特徴がこれまでよりクリアに得られることが期待できる。

我々の先行研究では、企業の成長法則に関する連続する2年の売上データに観測される拡張ジブラ(Gibrat)性[4-7]と拡張ジブラ則[8-12]の観測と導出を行っている。本研究では企業の倒産確率の分布の観測を行った後、その分布の特徴の導出を企業の成長法則から導くことを試みている。企業の成長法則を分析することは企業活動の時系列の分析を行っていることになる。各企業は年齢を重ねるごとに事業を多角化したり整理したりして活動を変化させている。企業毎に経営理念や目的があり複雑な活動を展開している。しかし、多数の企業活動の集団的な性質にはマクロな観点からのある種の制限が課せられる。そのような集団を統計的に見ることで全体を制限する何らかの法則があるのではないかと考えられる。例えば需要に関しては無限に存在するわけではなく、限りあるパイを多数の企業で奪い合うことになる。統計力学の分野では限りあるエネルギーを多数の粒子に分配するための法則を原理として様々な現象を解明している。我々はこれまでに企業を特徴づける様々な変数に対して統計的に観測される様々な特徴を分析してきた。倒産という現象は企業活動の時系列の末に起こる現象である。我々がこれまでに分析してきた手法やその結果を用いることで簡単な形式で倒産確率を計算することが可能になることが期待される。

本研究では倒産確率を企業年齢や企業規模を表す変数と関連付けて分析を行う。そして倒産確率を我々の先行研究である企業の成長法則と結びつ

けることによって倒産確率を導出するモデルを構築する。そして、得られたモデルを近似するによって解析な結果を導き出す。また近似に依らない解を導き出すために、今後に向けた数値シミュレーションの方法を考察する。

本稿の章立ては、以下のようにになっている。2章では、用いたデータの概略を紹介する。3章では、企業の倒産確率の年齢による分布を算出する。4章では、企業の倒産確率の売上による分布を算出する。5章では、連続する2時点の成長から1時点の分布を導く我々の先行研究について説明する。6章では企業の成長法則から企業の倒産確率の規模による分布を導出する。7章で企業年齢と企業売上の関係を分析する。8章を本稿のまとめと今後の課題に充てる。

2.1 データの詳細

分析を実行するデータには Bureau van Dijk 社の ORBIS データベース [13] を用いた。このデータベースは全世界 (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。日本では帝国データバンクや東京商工リサーチ等の信用調査会社が有名であるがそういった様々な調査会社との提携も行っている。我々は ORBIS の 2009 年版と 2014 年版を入手し分析を行った。用いたデータには企業の「財務状況データ」と「ステータス情報データ」の2種類がある。各企業には Bureau van Dijk 社が独自に割付けた ID (BvDID) が割り振られており、この ID を用いて2種類のデータベースの紐づけることが簡単に可能である。財務状況データは提供時の年版の直近 10 年の企業の財務状況 (売上、利益、従業員数、有形固定資) の数値データである。ここでの直近 10 年とは、確認された最後に記録が確認された年から遡って 10 年という意味である。ステータス情報データには企業の設立年とその年版時点での企業のステータス情報が記述されている。ステータス情報は Active (活動中), Bankruptcy (倒産), Dissolved (分解), In liquidation (清算中), Inactive (非活動中), Merger (併合), Unknown (不明) の項目で分類されたカテゴリーデータである。表 1 は入手したデータの中で日本企業に注目した財務データとステータスデータのレコード数である。

版	財務データ	ステータスデータ
2009 版	986,516	4,098,331
2014 版	980,213	5,223,310

表 2.1: 日本企業のデータ数

財務データのレコード数はステータスデータのレコード数に比べて約 5 分の

1 のボリュームである。日本の全ての株式会社には財務諸表の公開が義務付けられている。2015 年に導入されたマイナンバー制度によって日本中の会社に対して法人番号が付与された。国税庁のウェブサイトを利用することで各都道府県の株式会社数を知ることができる[14]。これによると 2016 年の株式会社の総数が約 190 万社であるので財務データは株式会社の半分強をカバーしていると考えられる。また、ステータスデータには株式会社ではないものも多数含まれている。

3.1 企業年齢の違いによる倒産確率の分析

まず初めに日本企業の倒産確率の年齢の違いによる分布を調査した。企業の年齢はデータベースにある設立年の情報から知ることができる。倒産した企業の判定にはステータス情報を用いる方法と財務データを用いる方法がある。本章ではステータス情報から得られた企業ステータスと年齢に注目して分析を行う。

3.2 データのステータス情報を用いた倒産確率の分析

まずはステータス情報を使って倒産企業を判定する。前章で説明したようにステータス情報には、Active, Bankruptcy, Dissolved, In liquidation, Inactive, Merger, Unknown の分類がある。ある時期に Active だった企業がその後 Active じゃない状況に変わった時その企業は倒産したと仮定する。我々は 2009 年版のステータス情報と 2014 年版のステータス情報を持っており。以下の定義

$$\frac{\text{2009 年が Active の企業でかつ 2014 年は Active ではない企業数}}{\text{2009 年が Active の企業数}} \quad (3.1)$$

を用いることによって 2009 年と 2014 年の間に倒産した企業の倒産確率を見積もることが可能である。ここでの定義では Unknown の企業も倒産した企業だとみなしている。

図 3.1 に日本とアメリカの企業の年齢分布と倒産確率を示した。左図は日本のデータ用いた結果であり、右図はアメリカのデータを用いた結果である。○印で表されている年齢分布は 2014 年版のデータを用いて、その当時の企業年齢 T としたときの Active な企業数 $N(T)$ を集計した結果である。×印で表されている倒産確率は、企業年齢 T 別に式 (3.1) の定義を用いて計算した倒産確率である。図 3.1 の左図から、日本のデータでは倒産確率に企業年齢依存性があまりないことが観測される。図 3.1 の右図のアメリカのデータでは企業年齢 20 歳以下の企業と、20 歳以上の企業では振る舞いが違うことが観測される。20 歳以下の企業に関して、年齢が若いほど倒産確率が高い傾向にある。20 歳以上の企業に関しては、倒産確率に企業年齢依存性はあまり

ない様子が観測される。年齢分布に関してはどちらの国も片対数表示で直線に近い単調減少のふるまいをしている。これは指数分布に近い分布である。ただし日本のデータでは 60 歳を境に単調減少の傾きが大きく変わっている。この 60 歳の境は第二次世界大戦に関係していると考えられる。アメリカのデータに関しては単調減少の傾きが緩やかに小さくなっていて、直線より少しだけ上に反った曲線になっている。

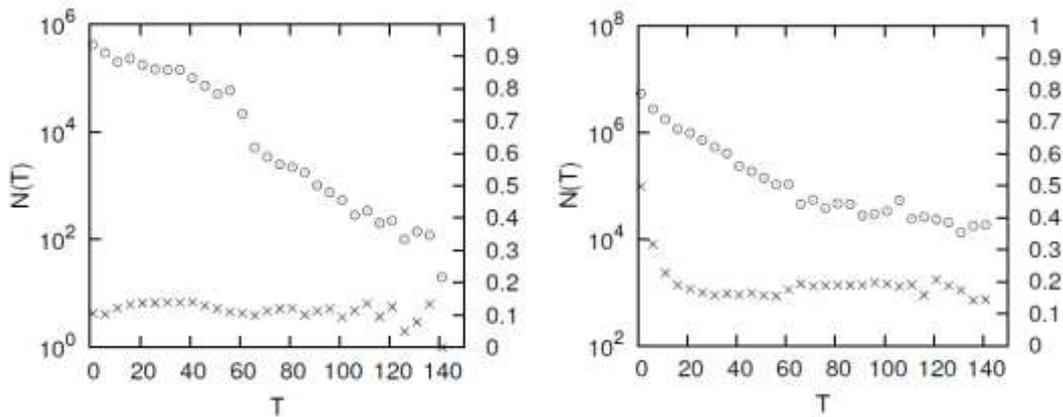


図 3.1: 日本とアメリカの企業の年齢分布と倒産率。○印は年齢 T と企業数 $N(T)$ の関係、×印は倒産率 (右軸目盛)。左図は日本のデータ、右図はアメリカのデータ。

ここでは日本とアメリカの結果を掲載したが、他にも様々な国の企業で年齢分布と倒産確率の調査を行った。倒産確率に関しては日本、スペインのようなタイプの年齢にほとんど依存しない国と、アメリカ、フランス、イギリス、イタリア、ドイツ、オランダ、韓国のようなタイプの、若い企業に年齢依存性がある国に分類できる。

グラフの形から倒産確率と企業の年齢分布に関して、次のようなモデルが考えられる。図 3.1 では 2009 年から 2014 年の 5 年間における倒産確率を観測したが、ここでのモデル構築には 1 年間での倒産確率 D_1 を考える。日本型タイプとアメリカ型タイプの両方を説明できるように倒産確率を

$$D_1 = \lambda_1 + \lambda_2 e^{-\kappa T} \quad (3.2)$$

で表現する。この式の第 2 項の $\lambda_2 = 0$ とすれば日本型の年齢に依らない一定の倒産確率 λ_1 を表現することができる。また、第 2 項の年齢依存性がアメリカ型の若い企業の倒産確率の年齢依存性を表している。この倒産確率が長期間持続すると仮定することで時点 T での Active な企業数 $N(T)$ が従う関数形を求めることができる。時点 T で Active な企業が時点 $T + 1$ でも Active である割合は

$$\frac{N(T+1)}{N(T)} = 1 - \lambda_1 - \lambda_2 e^{-\kappa T} \quad (3.3)$$

で表すことができる。この式は次の微分方程式

$$\frac{dN(T)}{dT} = -\lambda_1 - \lambda_2 e^{-\kappa T} \quad (3.4)$$

に書き換えられる。この微分方程式の解は

$$N(T) = N_0 \exp \left[-\lambda_1 T + \frac{\lambda_2}{\kappa} (e^{-\kappa T} - 1) \right] \quad (3.5)$$

の形式で求められる。ここで積分定数 N_0 は時刻 $T = 0$ で企業数である。毎年誕生する企業数が一定であると近似できる場合、 $N(T)$ は企業年齢 T の Active な企業数と解釈できる。企業年齢分布が(3.5)式のように表されるとすると

$$\frac{N(T+1)}{N(T)} = e^{-\lambda_1} \exp \left[\frac{\lambda_2}{\kappa} (e^{-\kappa(T+1)} - e^{-\kappa T}) \right] \quad (3.6)$$

が成り立つ。指数 $\lambda_1, \lambda_2, \kappa T$ が小さい時、

$$D_1 = 1 - \frac{N(T+1)}{N(T)} = \lambda_1 + \lambda_2 (1 - \kappa T) \quad (3.7)$$

となるので κT が小さい時の(3.5)式と矛盾しない。また、5年後の倒産確率を D_5 で表現すると、 D_5 は

$$D_5 = 1 - \frac{N(T+5)}{N(T)} \sim 5\lambda_1 + 5\lambda_2 (1 - \kappa T) = 5D_1 \quad (3.8)$$

の形式で表される。ここで展開したモデルを日本の場合とアメリカの場合に当てはめてみる。日本の場合は $\lambda_2 = 0$ であるとするればよい、アメリカの場合は(3.5)式で κT が小さい時の近似行い

$$N(T) = N_0 \exp \left[-(\lambda_1 + \lambda_2)T + \frac{\lambda_2 \kappa}{2} T^2 \right] \quad (3.9)$$

を用いて分析を行う。

まず、日本の場合の分析を行う。図 3.1 左図の×印より5年後の企業倒産確率は

$$D_5 = 0.111 \pm 0.029 \quad (3.10)$$

と見積もることができる。1年後の企業倒産確率は

$$D_1 = 0.022 \pm 0.006 \quad (3.11)$$

である。一方、図 3.1 左図の○印の60歳以下のデータを用いてその傾きから λ_1 を見積もると

$$\lambda_1 = 0.035 \pm 0.003 \quad (3.12)$$

となる。(3.11)式と(3.12)式の両者は近い値であることが確認できた。

次にアメリカの場合の分析を行う。図 3.1 右図の×印より(3.8)式を用いて各

パラメータを推定すると

$$\lambda_1 = 0.026 \pm 0.001 \quad (3.13)$$

$$\lambda_2 = 0.061 \pm 0.005 \quad (3.14)$$

$$\kappa = 0.14 \pm 0.01 \quad (3.15)$$

をえる。ここで λ_1 の推定には左 12 点の最小値を用いた。また、図 3.1 右図の○印のデータを用いて(3.9)式を使ってパラメータを推定すると

$$\lambda_1 = 0.068 \pm 0.006 \quad (3.16)$$

$$\lambda_2 = 0.074 \pm 0.024 \quad (3.17)$$

$$\kappa = 0.07 \pm 0.05 \quad (3.18)$$

となる。ここで λ_1 の見積もりには左 12 点の最小値を、 λ_2 の見積もりには κT が小さいとみなせる左 3 点のデータを用いた。両者の分析結果は近い値になっていることが確認できる。

日本型の企業では行の倒産確率は年齢に依存せずに一定であること、企業の年齢分布が指数関数で当てはまる。企業倒産確率が一定という性質が他の年でも成り立ち、かつ、年々誕生する企業数が大きく変化しない場合、この短期間の性質は企業の年齢分布が指数関数に従うことと整合性を持つ。

一方、アメリカ型の企業の倒産確率は特に若い年齢層で年齢に依存する。企業年齢が低くなるに従って倒産確率は高くなり、企業年齢が高くなるに従って倒産確率は高くなる。ある程度年齢が高くなるとそれ以降の年齢では倒産確率は一定値になる。アメリカ型の倒産確率は日本企業に倒産確率に於いて観測された性質を拡張したモデルを使って説明することができると考えられる。アメリカ型の場合でも短期間の倒産に関する性質が他の年でも成り立つとすると、企業の年齢分布関数を説明することができる。

上記の性質が定性的に成り立つだけでなく、定量的にも成り立つことを確認するために実際のデータを用いて検証を行った。見積もられてパラメータには誤差範囲で一致したものもあれば、一致しないものも含まれた。一致しないものには上述の仮定が必ずしも成立していない場合があることが原因であると考えられる。しかし、それほど大きな隔たりがあるわけではなく、仮定した近似はそれほど悪くないと考えられる。日本の場合、年々誕生する企業数は大きく変化しないことが報告されている[15]。

この分析で我々は倒産の定義として 2009 年データで Active で、2014 年データで Active ではない企業を倒産した企業だとした。Active ではない企業には Unknown や欠損データを表す NA の企業も含まれている。日本の場合は Unknown や NA データは全体数に比べて非常に小さいので両者の違いはそれほど問題にはならない。一方アメリカのデータでは全体の約 1/3 に相当するためこの違いには注意を払わなければならない。アメリカのデータでは Unknown や NA データを母数から除外すると倒産確率は小さくなる。そのた

め倒産確率からパラメータを見積もった値は変わってくる。しかしこの場合でも図 3.1 で観測された定性的な性質は変わらなかった。この理由で本稿には Unknown や NA データを含めた場合の倒産確率の分析結果を掲載した。

4.1 売上データを用いた倒産確率

企業の分布を測る指標として前章では企業年齢を用いていた。この章では各企業の売上を用いて倒産確率を分析することを考える。倒産企業を観測するためには、存在しなくなった企業を調査する必要がある。存在しなくなってしまったものを観測するというにはある種の矛盾が含まれている。そのような難しさがあるために一つの定義のみを用いるのには問題がある。複数の定義を用いて比較検討することが重要である。

図 4.1 の×印は 2009 年の年齢に関して 5 歳幅の階級で調べた倒産確率である。年齢に依らず 10%前後であることが分かる。この倒産確率に関して前章で詳しく調べた。

本章では、財務データを用いることで活動を停止した企業を推定することを考える。企業活動は商品やサービスを売り上げることによって成り立っている。ここでは売上データが存在しなくなった企業を活動停止した企業とみなして、倒産確率を計算した。以下の定義が売上データを用いた倒産確率である。本稿では売上データを変数 Y で記述する。

$$\frac{\text{2008 年の } Y \text{ が存在する企業でかつ 2013 年の } Y \text{ が存在しない企業数}}{\text{2008 年の } Y \text{ が存在する企業数}} \quad (4.1)$$

売上が存在しないという状況には、記録が存在しない場合と、計上された売り上げが 0 という 2 種類の状況が存在する。今回の定義にはどちらの状況でも売り上げが存在しないという定義にしている。この定義により、2008 年から 2013 年の間に企業活動がなくなった確率を計算することができる。2008 年の売上と 2013 年の売上について調べているのは前章のステータスデータとの整合性を考える為である。2009 年版のステータス情報は主に 2008 年に調査された情報であり、2014 年版のステータス情報は主に 2013 年に調査されたものである。財務データの方は決算日の情報であるのでステータスデータに合わせる形で 2008 年の売上と 2013 年の売上を調べている。

図 4.1 の▲印は 2008 年の年齢に関して 5 歳幅の階級で調べた倒産確率である。0-5 歳階級は 50%弱の倒産確率である。企業年齢階級が大きくなるに従って倒産確率は徐々に落ちる傾向にある。

これらの 2 種類の定義による違いは、企業の実態活動を表す売上活動が存在するのかが存在しないのかに関する問題である。売上活動率を調べるために次の量を計算した。

2009年がActiveの企業でかつ2008年のYが存在しない企業数 (4.2)
2009年がActiveの企業数

図 4.1 の*印は上記の量を年齢に関して 5 歳幅の階級で調べた結果である。20 歳未満の企業では 50%未満で年齢階級が上がるにつれて上がっていく傾向にある。表 2.1 より財務諸表の調査はステータスの調査より難しいと考えられる。そのためデータの網羅性の観点から実際に売上は存在していても調査から漏れている企業もある可能性がある。しかし、(4.1)式の定義による倒産確率の計算には上位を形成するであろうと考えられる 20%以上の日本企業が含まれている。売上にはパレートの法則[16]が成り立つことが知られているのでこのデータで日本の全企業の売上のほぼ 100%に近い値を構成すると考えられる。

また、売上の値は企業の規模を表す重要な指標である。(4.1)式の倒産確率を使うことで企業規模による倒産確率の算出を実行することが可能になる。倒産確率に関して年齢の違いによる分析だけではなく、企業規模を表す変数である売上の違いによる分析も重要である。

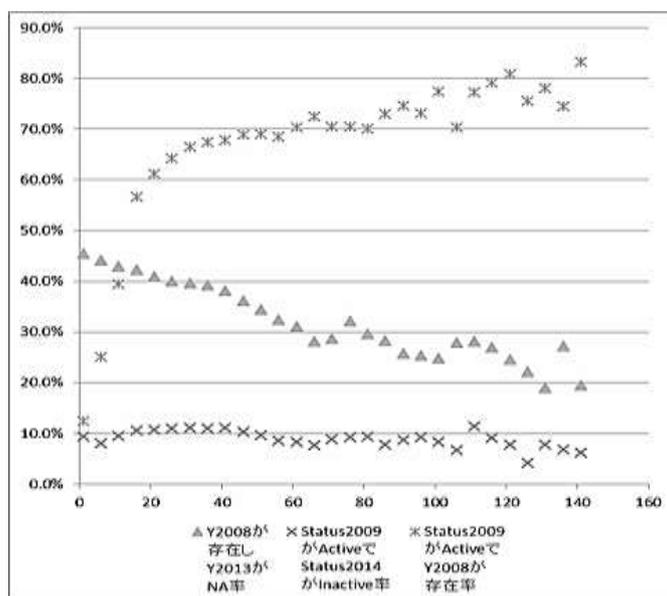


図 4.1: 倒産確率の企業年齢による分布。横軸は企業の年齢、縦軸は倒産確率を表す。×印は(3.1)式による定義を使って求め倒産確率。▲印は(4.1)式による定義を使って求めた倒産確率。*印は(3.3)式による定義を使って求めた売上活動率。

4.2 倒産確率の売上依存性の分析

売上データに関しては直近 10 年のデータを所持しているので 1 年間隔の細かな範囲で倒産企業の分析が可能である。また、我々の先攻研究[8-12]とリンクした分析が可能となる。

ある年の企業の売上を変数 Y_T で表す。次の年の企業の売上を変数 Y_{T+1} で表す。ある年 T から次の年 $T + 1$ の間の企業の倒産確率は次の式である。

$$D = \frac{Y_T \text{ 存在する企業でかつ } Y_{T+1} \text{ 存在しない企業数}}{Y_T \text{ が存在する企業数}} \quad (4.3)$$

図 4.2 は日本企業に関する 2011 年から 2012 年にかけての倒産確率を売上を横軸として観測した分布である。横軸の 2011 年の売上 Y_{2011} を $Y_T \in [10^{2+0.1(n-1)}, 10^{2+0.1n})$ のビンに分けて各領域で(4.3)式の倒産確率 D を計算している。図 4.2 より両対数表示で直線的な特徴を捉えることができる。この特徴から倒産確率の売上の違いによる分布は

$$D \propto Y_T^{-\lambda} \quad (4.4)$$

のべき関数で表すことができる。ここで λ は図の傾きから求められるべき指数である。他の年や他の国でもべき指数の大きさが違うだけで同様の特徴がみられることから(4.4)式の形は年や国に依らない普遍的なものであると考えられる。またべき指数 λ がその国の倒産確率の分布の特徴を表すものである。国によるべき指数の値を、後の表 6.1 に示す。

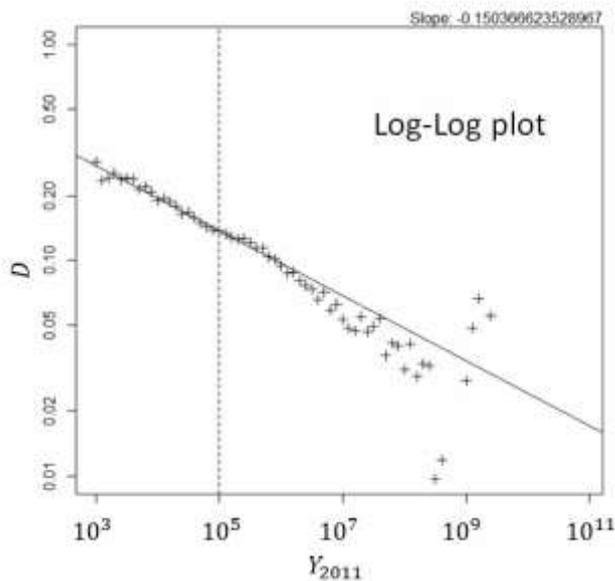


図 4.2: 日本企業に関する 2011 年から 2012 年の間の倒産確率の売上規模による分布。横軸は企業の売上、縦軸は倒産確率を表す。両対数プロットで $10^3 \sim 10^5$ 円の領域の傾きを回帰分析で求めた。

分析結果から得られる特徴は企業の成長と絡めて分析することができると考えられる。倒産しなかった企業では売上は 1 年間で Y_T から Y_{T+1} に成長する。 Y_{T+1} が存在しない企業には様々な原因が考えられるが、その中には十分な成長を果たせなかったために次の年の売上が存在しない企業が多くあると考えられる。我々のこれまで先行研究に於ける企業の成長に関する部分を拡張することで図 4.2 の倒産確率の特徴を導くことが可能である。

5.1 企業の規模に関する分布

これまでの先行研究で企業の規模に関する様々な性質が分かっている。本節ではその様々な性質のうち本研究に大きくかかわる法則について説明する。ある1時点に観測される企業サイズ(規模)分布の法則と、短期間の2時点で観測される反転対称則とジブラ則である。

まずは企業のサイズ(資産・売上・利益・所得・従業員数など)に関する分布の法則について説明する。ある時点 T に於ける企業のサイズを表す変数 x_T に関しての確率分布 $P(x_T)$ に関して、ある閾値 x_{th} より大きいサイズの企業では

$$P(x_T) \propto x_T^{-\mu-1} \quad \text{for } x_T > x_{th} \quad (5.1)$$

のベキ分布に従い、閾値 x_{th} より小さいサイズの企業では

$$P(x_T) \propto x_T^{-\mu-1} \exp\left[-\alpha \ln^2 \frac{x_T}{x_{th}}\right] \quad \text{for } x_T < x_{th} \quad (5.2)$$

の対数席部分布に従うという法則である。図 5.1 左は日本企業の 2005 年の利益の分布である。

反転対称性は短い間隔で続いた2時点の企業サイズに関する変数 x_1, x_2 に対するもので、同時分布 $P_{12}(x_1, x_2)$ に関して時間反転を行っても同時分布の形は変わらない

$$P_{12}(x_1, x_2) = P_{12}(x_2, x_1) \quad (5.3)$$

という法則である。これは時間反転 $x_1 \Leftrightarrow x_2$ の入れ替えで系が対称であることを表している。確率分布でこの条件が成り立っているということは、熱力学に於ける熱平常状態を表していることになる。図 5.1 の右図は日本企業の 2004 年と 2005 年の利益の散布図である。反転対称性に従っている。

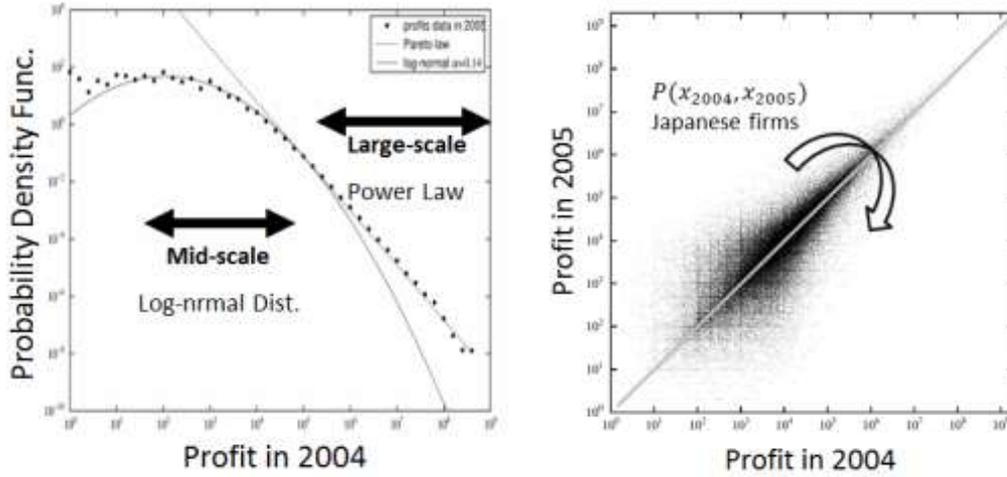


図 5.1: 左図は日本企業の 2004 年の利益分布。大きなサイズの企業ではベキ分布に、中間サイズの企業では対数正規分布に従う。右図は日本企業の 2004 年と 2005 年の利益の散布図。反転対称性に従っている。

5.2 ジブラ則からベキ法則の導出

過去の研究では 2 時点の成長に関するジブラ則と反転対称性の組み合わせから、大規模サイズのベキ分布を解析的に導き出すことがなされている [6, 7]。ジブラ則は短い間隔で続いた 2 時点の企業サイズに関する変数 x_1, x_2 に対する法則である。成長率 $R = x_2/x_1$ に関して、初期値 x_1 を狭い範囲に限った条件付き確率分布 $Q(R|x_1)$ を考える。この条件付き確率分布の形が初期値の値である x_1 の大きさに依存しない

$$Q(R|x_1) = Q(R) \quad (5.4)$$

というのがジブラ則である。変数 x_1, x_2 に関する同時分布 $P_{12}(x_1, x_2)$ に関して、変数変換を行い変数 x_1, R に変数変換した同時分布関数を $P_{1R}(x_1, R)$ 、変数 R^{-1}, x_2 に変数変換した同時分布関数を $P_{R^{-1}2}(R^{-1}, x_2)$ とする。積分測度の関係から次の関係

$$x_1 P_{12}(x_1, x_2) = P_{1R}(x_1, R) \quad (5.5)$$

$$x_2 P_{12}(x_1, x_2) = P_{R^{-1}2}(R^{-1}, x_2) \quad (5.6)$$

が成り立つ。成長率 R とその逆数 R^{-1} に関する条件付き確率の定義式

$$Q_R(R|x_1) = \frac{P_{1R}(x_1, R)}{P_1(x_1)} \quad (5.7)$$

$$Q_{R^{-1}}(R^{-1}|x_2) = \frac{P_{R^{-1}2}(R^{-1}, x_2)}{P_2(x_2)} \quad (5.8)$$

に於いてジブラ則 $Q_R(R|x_1) = Q_R(R)$, $Q_{R^{-1}}(R^{-1}|x_2) = Q_{R^{-1}}(R^{-1})$ が成り立つ場合以下の関係

$$P_{1R}(x_1, R) = P_1(x_1)Q_R(R) \quad (5.9)$$

$$P_{R^{-1}2}(R^{-1}, x_2) = P_2(x_2)Q_{R^{-1}}(R^{-1}) \quad (5.10)$$

が導かれる。(5.7)式から(5.10)の4式から

$$\frac{P_1(x_1)}{P_2(x_2)} = \frac{1}{R} \frac{Q_{R^{-1}}(R^{-1})}{Q_R(R)} \quad (5.11)$$

が導かれる。この式の右辺は R のみで書かれた関数なので右辺は $x_1 \rightarrow Ax_1$, $x_2 \rightarrow Ax_2$ の変数変換で変わらない、すなわち

$$\frac{P_1(x_1)}{P_2(x_2)} = \frac{P_1(Ax_1)}{P_2(Ax_2)} \quad (5.12)$$

であることを示すことができる。この関係式は関数 $P_1(x_1)$ と関数 $P_2(x_2)$ が同じスケーリング則を満たすことを示している、ベキ分布に従う確率密度関数 $P(x) = Cx^{-\mu-1}$ に於いて $x \rightarrow Ax$ の変数変換を行うと $P(x) \rightarrow P(Ax) = A^{-\mu-1}P(x)$ に変換される。つまり同じベキ指数 μ のベキ関数は(5.12)式を満足する関数になっている。

ジブラ則に加えて詳細釣り合い則(5.3)式を導入することで、関数 $P_1(x_1)$ と関数 $P_2(x_2)$ の具体的関数形がベキ関数になることを示すことができる。詳細釣り合い則が満たされる時、関数 $P_1(x_1)$ と関数 $P_2(x_2)$ は同じ関数形に、また関数 $Q_{R^{-1}}(R^{-1})$ と関数 $Q_R(R)$ も同じ関数形になる。このことを考量すると、(5.11)式は

$$\frac{P(x_1)}{P(x_2)} = \frac{1}{R} \frac{Q(R^{-1})}{Q(R)} \quad (5.13)$$

と表すことができる。(5.13)式の右辺を $G(R)$ で表すことにすると

$$P(x) = P(Rx)G(R) \quad (5.14)$$

を得る。この式を $R = 1 + \epsilon$, ($\epsilon \ll 1$)で展開すると、 ϵ の1次の項の比較から次の微分方程式

$$G'(1)P(x) + xP'(x) = 0 \quad (5.15)$$

を得る。この微分方程式を解くことで確率密度関数がベキ関数 $P(x) = Cx^{-\mu-1}$ であることが導かれる。また、ベキ指数 μ は成長率分布の $R = 1$ 付近の値や傾きを用いて

$$\mu = -2 - \frac{\lim_{R \rightarrow 1+0} Q'(R) + \lim_{R \rightarrow 1-0} Q'(R)}{Q(1)} \quad (5.16)$$

という式で得られることが分かる。

5.3 拡張ジブラ則から対数正規分布の導出

次に拡張ジブラ則について説明する。拡張ジブラ則とはジブラ則が規則的に緩やかに破れている状態のことを表す。図 5.2 は日本企業に関する 2007 年と 2008 年に関する売上の成長率 R に関する条件付き分布 $Q(R|x_T)$ である。左図は x_T の値が大きい場合の分布で x_T の値に依らずに同じ曲線に重なっている。この性質がジブラ則を表している。右図は x_T の値が小さい場合の分布で x_T の値が変わると少しずつ分布の形が変わっていく。この性質が拡張ジブラ則を

表している。

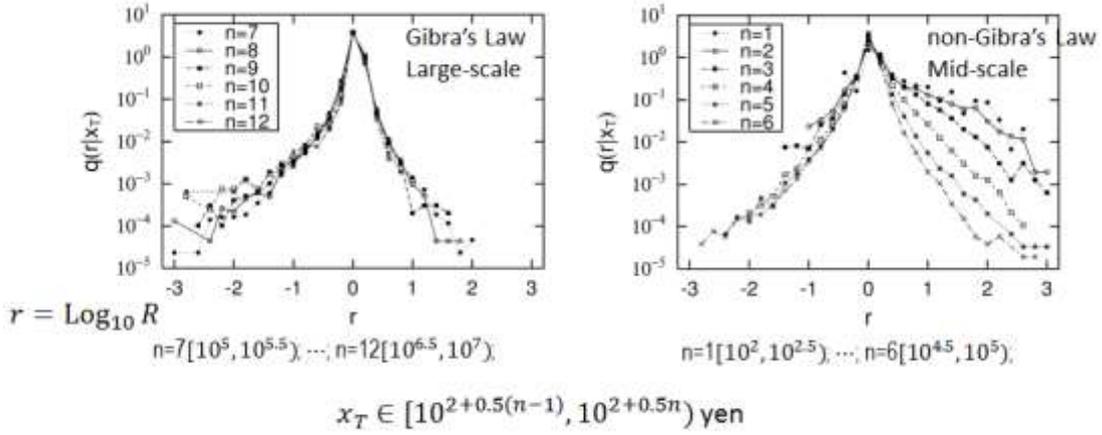


図 5.2: 日本企業に関する 2007 年と 2008 年に関する売上の成長率 R に関する条件付き分布 $Q(R|x_T)$ 。左図はジブラ則を表している。右図は拡張ジブラ則を表している。

これまでの研究で拡張ジブラ則と反転対称性の組み合わせから中規模サイズの対数正規分布を解析的に導き出すことがなされている[8,9]。ジブラ則の(5.4)式を緩やかに破るような次の形の成長率分布

$$Q(R|x_1) = d(x_1)R^{-t_+(x_1)-1} \quad \text{for } R > 1 \quad (5.17)$$

$$Q(R|x_1) = d(x_1)R^{t_-(x_1)-1} \quad \text{for } R < 1 \quad (5.18)$$

を仮定する。ここで $d(x_1)$ は図 5.2 の成長率分布の $\log_{10} R = r = 0$ に於ける値に相当する。 $t_+(x_1)$ は成長率分布の $r > 0$ の領域を直線とみなした時の勾配に相当する。 $t_-(x_1)$ は成長率分布の $r < 0$ の領域を直線とみなした時の勾配に相当する。ジブラ則が成り立たないが詳細釣合則が成り立つ場合の(5.11)式に対応する式は

$$\frac{P(x_1)}{P(x_2)} = \frac{1}{R} \frac{Q(R^{-1}|x_2)}{Q(R|x_1)} \quad (5.19)$$

である。(5.17)式と(5.18)式を(5.19)式に適用すると

$$\tilde{P}(x) = \tilde{P}(Rx)R^{t_+(x)-t_-(Rx)+1} \quad (5.20)$$

を得る。ここで $\tilde{P}(x) = P(x)d(x)$ である。(5.20)式を $R = 1$ の周りで展開することによって次の微分方程式

$$[1 + t_+(x) - t_-(x)]\tilde{P}(x) + x\tilde{P}'(x) = 0 \quad (5.21)$$

を得る。この微分方程式の解は

$$\tilde{P}(x) = Cx^{-1}e^{-H(x)} \quad (5.22)$$

$$h(x) = t_+(x) - t_-(x) \quad (5.23)$$

$$H(x) = \int \frac{h(x)}{x} dx \quad (5.24)$$

である。これを(5.20)式に代入することによって次の式

$$-H(x) + H(Rx) = [t_+(x) - t_-(Rx)] \ln R \quad (5.25)$$

を得る。(5.25)式を $R = 1$ の周りで展開することによって次の微分方程式

$$x[t_+''(x) + t_-''(x)] + t_+'(x) + t_-'(x) = 0 \quad (5.26)$$

$$2t_+'(x) + t_-'(x) - 3t_-''(x) - x^2[t_+^{(3)}(x) + 2t_-^{(3)}(x)] = 0 \quad (5.27)$$

を得る。これらの連立微分方程式の解は

$$t_+(x) = -\frac{C_{-2}}{2} \ln^2 x + (C_{+1} - C_{-1}) \ln x + (C_{+0} - C_{-0}) \quad (5.28)$$

$$t_-(x) = \frac{C_{-2}}{2} \ln^2 x + C_{-1} \ln x + C_{-0} \quad (5.29)$$

である。ここで C は積分定数を表す。これらの解は(5.25)式を満たすべきであるという条件から $C_{-2} = 0$, $C_{+1} = 0$ である。得られた解を(5.22)式から(5.24)式に用いると、最終的に次の解

$$\tilde{P}(x) = Cx^{-(\mu+1)}e^{-\alpha \ln^2(x/x_0)} \quad (5.30)$$

を得る。ここで $\mu = 2C_{-0} - C_{+0}$, $\alpha = C_{-1}/2$ である。 x_0 は(5.24)式を積分する際の境界の値である。この解において x_0 はベキ分布に従う領域と対数正規分布に従う領域の閾値と考えることができる。閾値よりも大きな $x_0 < x$ の領域ではジブラ則が成り立ち $\alpha = 0$ であると考えられ、その領域では確率分布はベキ分布になる。図 5.1 のデータではある閾値以降の領域の分布はベキ分布になっている、その領域に対応する図 5.2 の左図の成長率分布にはジブラ則が成り立っていることが分かる。閾値よりも小さな $x < x_0$ の領域ではジブラ則が破れているため $\alpha \neq 0$ であると考えられ、そのような領域では確率分布は対数正規分布になる。図 5.1 のデータではある閾値より小さな領域の分布は対数正規分布になっていて、その領域に対応する図 5.2 の右図の成長率分布にはジブラ則が破れていることが分かる。そして、成長率分布の傾きの変化から(5.30)式の α の値を見積もることができて、結果は対数正規分布の α の値を矛盾なく説明するものになっている。また、 $d(x_1)$ の値は $R = 1$ に於ける成長率分布の値であるので、図 5.2 の $r = 0$ に於ける成長率分布の値に対応する。この値は n を変えてもほとんど変わらないため $d(x_1)$ は x_1 依存性がほとんどない定数であると考えられる。

5.4 詳細準釣合則と分布の指標の関係

次に詳細準釣合則について説明する。確率密度関数を分析する際に(5.3)式の詳細釣合則を導入することで対称性が高くなることによって方程式に制限が加わり解析会を得ることが容易になった。しかし一般的に時間反転対称な平衡状態に系があるとは限らず、そのような場合には

詳細釣合則を修正する必要がある。ここでは次式で定義される詳細準釣合則

$$P_{12}(x_1, x_2) = P_{12} \left(\left(\frac{x_2}{a} \right)^{1/\theta}, ax_1^\theta \right) \quad (5.31)$$

を導入する。詳細釣合則では図 5.1 の右図のように $x_1 = x_2$ の直線に対称軸を入れていた。詳細準釣合則では対数プロットで切片 a 、傾き θ の直線

$$\log x_2 = \theta \log x_1 + \log a \quad (5.32)$$

で対称軸を入れる。この操作によって $ax_1^\theta \Leftrightarrow x_2$ の入れ替えに関する対称性、系の静的な対称性に準ずる対称性があると仮定している。この仮定の下で理論計算をやり直すには成長率 R の定義を形式的に次の式

$$R = \frac{x_2}{ax_1^\theta} \quad (5.33)$$

で置き換えればよい。(5.19)式に対応する式は

$$\frac{P(x_1)}{P((x_2/a)^{1/\theta})} = \frac{1}{R} \frac{Q(R^{-1}|(x_2/a)^{1/\theta})}{Q(R|x_1)} \quad (5.34)$$

に置き換わる。この系に関しても(5.17)式や(5.18)式のような成長率分布の直線近似を行うと(5.20)式と全く同じ形式を得る。確率密度関数の解は

$$\tilde{P}(x_1) = C x_1^{-(\mu+1)} e^{-\alpha \ln^2(x_1/x_0)} \quad (5.35)$$

$$\tilde{P}((x_2/a)^{1/\theta}) = C ((x_2/a)^{1/\theta})^{-(\mu+1)} e^{-\alpha \ln^2((x_2/a)^{1/\theta}/x_0)} \quad (5.36)$$

である。 $d(x)$ の x 依存性が無視できるとして、 x_1 の確率密度関数 $P_1(x_1)$ と x_2 に関する確率密度関数 $P_2(x_2)$ を次のように

$$P_1(x_1) = C_1 x_1^{-(\mu_1+1)} e^{-\alpha_1 \ln^2(x_1/x_{01})} \quad (5.37)$$

$$P_2(x_2) = C_2 x_2^{-(\mu_2+1)} e^{-\alpha_2 \ln^2(x_2/x_{02})} \quad (5.38)$$

と表すことにする。連続時期のベキ指数 μ と対数分散の逆数 α との間には

$$\frac{\mu_1}{\mu_2} = \frac{\alpha_1}{\alpha_2} = \theta \quad (5.39)$$

の関係があることが分かる。このことから同時分布に於ける時間反転の対称軸の傾きが分布の特徴であるベキ指数や分散の変化を説明する量となっていることが示された。

5.5 空間的準反転対称性と分布の指標の関係

これまでは $x_1 \rightarrow x_2$ のように時間発展する量について考えてきた。ここでは違う次元の2変量 x, y が x_1, x_2 に於ける(5.31)式と似たような関係

$$P_J(x, y) = P_J \left(\left(\frac{y}{ax} \right)^{1/\theta}, ax^\theta \right) \quad (5.40)$$

にある場合を考える。この関係式を空間的準反転対称性と呼ぶ。ここで

$P_j(x, y)$ は確率変数 x, y に関する同時確率密度関数を表す。比率 $R = y/ax^\theta$ にする、条件付き確率分布 $Q(R|x)$ に対して(5.4)式のジブラ則に対応する式

$$Q(R|x) = Q(R) \quad (5.41)$$

が成り立っているとす。これら(5.40)式と(5.41)式から、 x, y に関する確率密度関数

$$P_x(x) = C_x x^{-(\mu_x+1)} \quad (5.41)$$

$$P_y(y) = C_y y^{-(\mu_y+1)} \quad (5.42)$$

が導かれべき指数の間には

$$\frac{\mu_x}{\mu_y} = \theta \quad (5.43)$$

の関係が成り立つ。

6.1 倒産確率の分布の導出

前節の2時点の成長を用いて分布を導出するという考えを倒産確率の分布に応用する。倒産した企業は2時点の当期、次期データの次期に当たるものが存在しない。この存在しないという確率を企業の成長率分布を用いて推定する。

図 6.1 の左図は日本企業の売上の 2011 年と 2012 年の散布図である。この図のデータを元に右図の成長率分布を描くことができる。横軸の $T = 2011$ 年の売上 Y_T を細かなビン $Y_T \in [10^{3+0.2(n-1)}, 10^{3+0.2n})$ に分割する。 $n = 1$ に相当するビンの領域は $Y_T \in [10^3, 10^{3.2})$ であり、この領域は図 6.1 の左図の灰色縦線の最初の 2 本の間の領域に相当する。この領域内のデータを用いて成長率 $R = Y_{T+1}/Y_T$ の条件付き確率密度関数 $Q(R|Y_T)$ を求めた。図 6.1 右図の $n = 1$ の折れ線グラフは求めた条件付き確率密度関数に対して、 $r = \text{Log}_{10} R$ 、 $q(r|Y_t) = \text{Log}_{10} Q(R|Y_T)$ という変換を行ったものをプロットした図である。同様に $n = 2$ に相当する売上ビンの領域は $Y_T \in [10^{3.2}, 10^{3.4})$ であり、図 6.1 の左図の灰色縦線の 2 本目と 3 本目の間の領域に相当する。この領域のデータを使って求めた成長率分布 $q(r|Y_t)$ を表示したものが図 6.1 右図の $n = 2$ の折れ線グラフである。同様にして $n = 3, 4, 5$ の領域に関しての成長率を求めて図 6.1 右図に折れ線グラフを重ねて表示した。

倒産企業に関するデータは図 6.1 の左図の散布図に現れない。なぜなら Y_{2011} のデータは存在するが Y_{2012} のデータが存在しないような企業を倒産企業だと考えたからである。このように倒産企業は散布図には直接表れない企業であるが、次のような1つの仮定を置くことによってその確率を見積もることが可能である。倒産した企業のほとんどは図 6.1 左図の下の灰色角丸四角形の

領域に入ってしまった企業で十分な成長ができなかったために Y_{2012} の売上データは計上できない状況に陥ってしまったのであると仮定する。この仮定の下で図 6.1 右図の $n = 1$ の確率密度関数を見直す。 $n = 1$ の確率密度関数は $r = 2$ 付近で値の最小値を取っており、それ以下にはデータが存在していない。つまり売上が計上できたと仮定すれば $r = -2$ 以下の部分に入っている企業がある程度の割合で存在していたと考えられる。この領域の存在している企業の割合は $r = -2$ 以下の領域の確率密度関数の積分を使って見積もることが可能である。つまり、図 6.1 右図のグラフの左下に上書きした三角形の領域の積分を行えば $n = 1$ の領域の倒産企業の確率を見積もることができると考えられる。積分の下限値は $-\infty$ で、積分の上限値では $Y_{T+1} = 10^{-2}$ とすると $r = -\text{Llog}_{10} Y_T / 10^2$ が積分の上限値になる。図 6.1 の右図を見ると $r < 0$ の領域の $q(r|Y_t)$ を見ると n に依らずほぼ重なっているが、線の左裾が n が大きくなるにしたがって左下に伸びて行くことが分かる。そのため n の値にかかわらず同じ関数形を用いることにする。また、 $q(r|Y_t)$ の形は緩やかにカーブしているが、簡単のため直線で近似して $q(r|Y_t) = 10^{(c_1 + c_2 r)}$ とする。

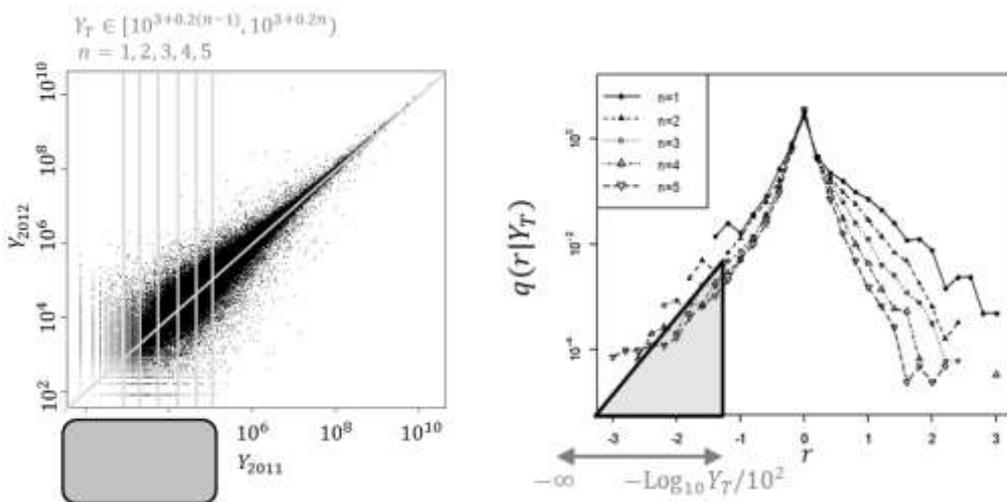


図 6.1: 左図は日本企業の売上の 2011 年と 2012 年の散布図。右図は $n = 1, 2, 3, 4, 5$ の領域で測定した成長率分布。左図の灰色角丸四角形の領域が右図の $n = 1$ の灰色三角形の底辺の領域になる。

これらの考察をまとめると以下の式

$$d \propto \int_{-\infty}^{-\frac{\text{Llog}_{10} Y_T}{10^2}} q(r|Y_t) dr = \int_{-\infty}^{-\frac{\text{Llog}_{10} Y_T}{10^2}} 10^{(c_1 + c_2 r)} dr \propto Y^{-c_2} \quad (6.1)$$

倒産確率 d を見積もることができると考えられる。

ここで(4.2)式のデータの観測から求めた倒産確率 D とこの節の仮定の下で成

長率分布から見積もった倒産確率 d の形を見比べると、どちらもベキ関数である点は同じであることが分かる。倒産確率の売上サイズ依存性がベキ関数になるという理由は成長率分布を用いた議論から定性的に説明できることになる。

さらに定量的な説明が可能であるか確かめるために様々な国で倒産確率のデータから(4.2)式のベキ指数 λ と、成長率分布の傾きから(6.1)式のベキ指数 c_2 を調べ比較を行った。2004, 2005, ..., 2012 年のデータを使って1年間隔の倒産確率、成長率分布を求めベキ指数の平均値を求めた結果を表 2 に示す。定量的にも近い値を取っているが、一致はしていない。

国	λ	c_2
日本	0.17 ± 0.05	0.95 ± 0.21
フランス	0.37 ± 0.08	0.78 ± 0.05
イタリア	0.31 ± 0.07	0.81 ± 0.03
スペイン	0.17 ± 0.06	0.82 ± 0.08

表 6.1: 倒産確率のベキ指数 λ と成長率分布から見積もったベキ指数 c_2 の比較

7.1 企業年齢と売上規模の関係

本稿のここまでの分析で倒産確率の企業年齢依存性と売上規模による依存性を調べた。この章では企業の年齢と売上にどのような関係があるのか、その関係性の分析を行う。年齢と売上の関係を調べることで企業の成長に関する特徴を得ることができる。企業は成長や衰退を繰り返して最終的に活動停止の状況に至る。企業規模の成長と年齢を様々な企業に於いて調査し、分布レベルで関係を見ることは重要である。

図 7.1 は 2008 年の日本企業の企業年齢と売上の関係である。点印は散布図を表している。年齢を固定して分布を観測すると、ベキ分布と対数正規分布とどちらにでも取れるような分布になっている。ベキ分布ではモーメント発散という理論的な特徴がある。ベキ指数が 2 以下の分布では分散が発散し、ベキ指数が 1 以下の分布では平均が発散する。企業の売上の分布ではベキ指数は 1 に近い値になるため算術平均をそのまま計算するのは危険である。試しに乱数シミュレーションを用いてベキ指数 1 以下の分布に対応する乱数を有限個発生させて平均値を計算すると次のような現象が起こる。有限個のデータを用いるので形式的に算術平均を計算することは可能で、算術平均の値は有限な値になる。しかし、同じ設定で乱数シミュレーションを繰り返し行って算術平均を測定すると、計算するたびに桁違いの平均値が出てきてしまう。そのため誤差がとてつもなく大きくなり何回シミュレーションを繰り返しても確定的な値にならない。そのため、図 7.1 では幾何平均を取ったものを黒丸印で表示している。幾何平均を取ることはデータの対数を取って算術平均を計

算し指数を使って元に戻すことに相当する。ベキ分布に従う変数に対して対数を取る変数変換を行うと、変換後の変数の分布は指数分布になる。指数分布の算術平均の値は有限で発散しないため、乱数シミュレーションを繰り返し行って算術平均を計算すると、計算するたびに設定に近い平均値が得られ、繰り返しの測定数を増やすことによって誤差が小さくなっていき、確定値を得ることになる。また売上の分布が対数正規分布であったとしても対数正規分布を特徴付けるパラメータは対数平均であるため、幾何平均がその対数平均に対応する。このような理由で幾何平均が年齢を固定した時の代表値として扱いやすいものになっている。この図 7.1 に示した幾何平均を調べることによって企業年齢と売上の関係にある程度見積もることが可能である。

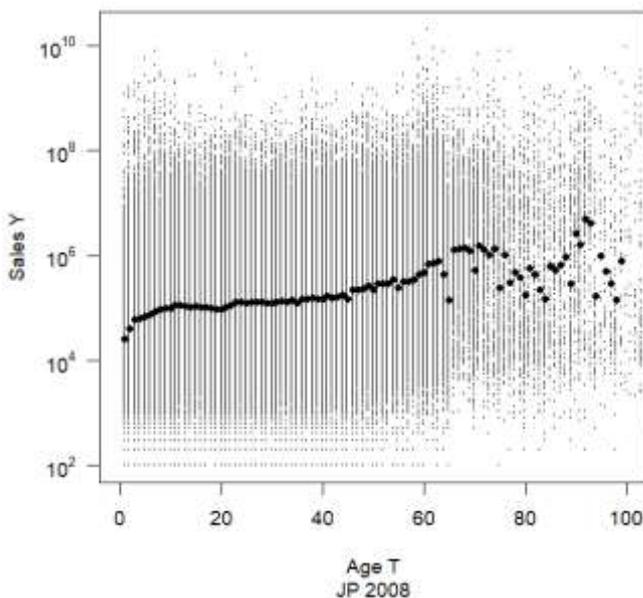


図 7.1: 2008 年日本企業の企業年齢 T と売上 Y の関係。点印は散布図に相当する。黒丸印は企業年齢を固定した条件下での売上の幾何平均を表す。

図 7.2 は図 7.1 の●印に当たる企業年齢を固定した条件下での売上の幾何平均を再掲して分析を行った図である。図 7.2 の左図は片対数表示であり、図 7.1 と同じ方法で表示したものである。年齢 20 歳以上の企業の売上の幾何平均は片対数表示で直線とみなすことができる。この結果は 20 歳以上の企業の売上は平均的に指数関数の関係にあり、年齢が高くなるに従って平均的な売上は大きくなるということを表している。図 7.2 の右図は左図と同じデータを両対数表示でプロットした図である。年齢 20 歳以下の企業の売上の幾何平均は両対数表示で直線とみなすことができる。の結果は 20 歳以下の企業の売上は平均的にベキ関数の関係にあり、年齢が高くなるに従って平均的

な売上は大きくなるということを表している。

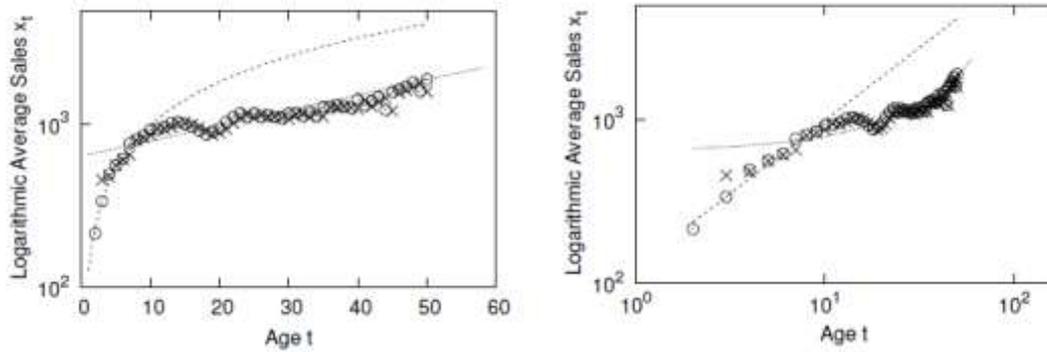


図 7.2: 企業年齢を固定した条件下での売上の幾何平均。左図は片対数表示、右図は両対数表示。

このような調査を様々な国のデータで行った。結果は定性的にはどの国でも似たような傾向があることが観測された。この定性的な性質は企業のサイズ分布のベキ分布がどのように形成されていくのかを調べる際に大変重要な性質であると考えられる。横軸は時間経過ではなく企業の年齢であるので注意する必要があるが、この横軸が時間経過系列であったとしても同様な傾向にあることを示すことが可能である。

7.2 拡張ジブラ則と企業成長の関係

時点 t に於ける成長率を $R_t = x_{t+1}/x_t$ で表現する。拡張ジブラ則を

$$q(r_t|x_t) = \frac{f(r_t)}{x_t^{1/\gamma}} \quad \text{for } r_t > 0 \quad (7.1)$$

$$q(r_t|x_t) = g(r_t) \quad \text{for } r_t < 0 \quad (7.2)$$

の形式で導入する。これは 5 章の(5.17)式、(5.18)式に対応する式であり、 $r_t = \log_{10} R_t$, $q(r|x_t) = \ln 10 \log_{10} R_t Q(R_t|x_t)$ を表している。(5.17)式、(5.18)式では具体的に直線の関数形を仮定していたが、(7.1)式、(7.2)式では図 5.2 の一般的な性質のみを仮定している。 $f(r)$ は単調減少の関数を仮定していて、 $g(r)$ には単調増加の関数を仮定している。 γ はある定数を表している。この仮定は図 5.2 の形から推測したものである。この設定の下で対数成長率の期待値は

$$\langle r_t \rangle = \int_{-\infty}^{\infty} r_t q(r_t|x_t) dr_t = \frac{A}{x_t^{1/\gamma}} - B \quad (7.3)$$

で表すことができる。ここで $A = \int_0^{\infty} r_t f(r_t) dr_t$, $-B = \int_{-\infty}^0 r_t f(r_t) dr_t$ は有限の値を持つ定数であると仮定している。(7.3)式の積分は成長率分布のデータを用いて数値的に実行することが可能である。図 7.3 の●印は図 5.2 に表示されて

いる成長率分布をそれぞれの n に関して実行した結果である。縦棒は r_t に関する 2 次のモーメントを計算して標準偏差を求めて図示したものである。●印の部分は(7.3)式を矛盾なく表現することが可能である。

(7.3)式に於いて第 2 項の B の寄与が第 1 項の寄与に比べてほとんどないと考えられる領域、つまり x_t が小さくて B が無視できる領域について考える。この領域では $A = \gamma D^{1/\gamma}$ と置くことで

$$\langle r_t \rangle = \gamma \left(\frac{D}{x_t} \right)^{1/\gamma} \quad (7.4)$$

に書き直すことができる。 $\langle r_t \rangle = \langle \log_{10} x_{t+1}/x_t \rangle = \log_{10} x_{t+1}/x_t$ の最後の等式を書き換えが許されると仮定すると(7.4)式は

$$x_{t+1} = x_t 10^{\gamma(D/x_t)^{1/\gamma}} \quad (7.5)$$

に書き換えることができる。 t が 1 よりも十分大きい時(7.5)式の微分方程式の解は

$$x_t = x_0 (\ln 10)^{\gamma} D t^{\gamma} \quad (7.6)$$

と書くことができる。ここで $\ln(1 + 1/t) \sim 1/t$ のマクローリン展開を使った。この(7.6)式はベキ指数 γ のベキ成長を表している。 x_t がある閾値 x_0 を超える領域では(7.3)式の右辺は定数 $E \equiv A/x_0^{1/\gamma} - B$ であるとみなすことができる。この時の(7.5)式に対応する微分方程式は

$$x_{t+1} = x_t E \quad (7.7)$$

である。この解は

$$x_t = C e^{Et} \quad (7.8)$$

で表すことができる。この(7.8)式は指数 E の指数成長を表している。

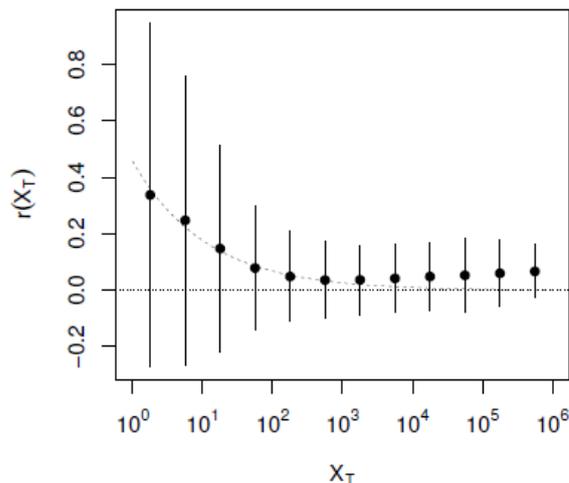


図 7.3: 数式(7.3)の積分を数値データを用いて行ったもの。黒丸印は各 x_t の領域で数値計算した $\langle r_t \rangle$ の値。縦棒の長さは 2 次モーメント $\langle r_t^2 \rangle$ の数値計算を元にした標準偏差の大きさを表す。

前段落の(7.4)式以降では、仮定をいくつか重ねる強引な論理展開を行っている。しかし、ベキ成長と指数成長の両方をうまく説明できる可能性が表れている。仮定の中で一番大きな問題は $(\log_{10} x_{t+1}/x_t) = \log_{10} x_{t+1}/x_t$ と書き換える部分であり x_t の確率的な揺らぎの効果がどのような形で最終形に残ってくるのか明らかではない。このような効果を取り込むための一番簡単な方法は乱数シミュレーションを行うことである。今現在、図 6.1 の左図のような散布図を再現するような条件付き確率の乱数の発生法を開発している最中である。開発中のものを使って適当な初期条件から企業を成長させ、十分に成長しきれなかった企業を倒産させるシミュレーションを行うと、図 7.2 の定性的な性質を再現することが可能である。

7.3 乗算確率過程を用いた企業成長のシミュレーション

ベキ分布を導き出すモデルに乗算的確率過程がある。最もシンプルな乗算確率過程はジブラ過程と呼ばれるもので

$$x_{t+t} = a_t x_t \quad (7.9)$$

の式で与えられる。ここで a_t は乗算ノイズ(乗算的確率係数)と呼ばれる確率変数である。 $t = 0$ で初期値 x_0 を適当に与えて(7.9)式を用いて確率変数 x_t の値を逐次的に更新していく。この過程での乗算ノイズ a_t は x_t とは全く独立の確率変数である。(7.9)式は $x_{t+1}/x_t = a_t$ 書き換えることができ、 a_t は確率変数 x_t の成長率を表していることが分かる。しかも a_t と x_t は互いに独立した確率変数であるため成長率分布が x_t に依存しないジブラ則を満足する確率過程になっている。しかし、この過程で x_t の乱数を多数発生させると x_t が従う確率分布は平均値がどんどん大きくなっていく安定しない対数正規分布になる。ところがこのジブラ過程を少し修正するだけで x_t が従う確率分布を安定したベキ分布に従うように変更することが可能である。その1つの方法は確率変数 x_t がある値より小さくなりそうとき、逆数規則で値を変えてしまう反射壁を設ける方法[17]である。別の方法は加算ノイズ(加算的確率係数)と呼ばれる確率変数 b_t を加えた確率過程

$$x_{t+t} = a_t x_t + b_t \quad (7.9)$$

のモデル[18]である。他にも参入と退出のメカニズムを乗算過程に組み合わせることでベキ則現象を説明するモデルが多く提案されている[19-24]。参入と退出のメカニズムの導入とは、まさに倒産した企業や新たに創業した企業の効果を導入するという本稿の目的と合致したモデルである。このモデルの乗算確率 a_t に x_t 依存性を図 5.2 の形で導入した乗算ノイズを用いればよい。このモデルに於いて図 7.2 のベキ成長、指数成長を確認し、乗算ノイズに関するどのようなパラメータがベキ成長や指数成長のふるまいを決定づけるのか分析を実行する予定である。

8.1 まとめ

本研究では、企業の倒産確率の分布についての分析を行った。倒産の定義にステータスデータを使う方法と売上データの欠損を使う方法の2種類の方法で分析を行った。ステータスデータを使う方法ではステータス情報がアクティブであった企業が一定の時期が経過した後アクティブではなくなるという条件で倒産を定義した。売上データを使う方法では売上を計上していた企業が何らかの理由で一定の時期が経過した後売上を計上しなくなるという条件で倒産を定義した。ステータス情報を使った定義ではペーパーカンパニーのような活動実態のない企業も母数に入れてしまっている可能性が高い。また、株式会社には業績データの開示義務が存在するが、そうではない企業には義務はない。そのため業績データのボリュームはステータス情報のボリュームに比べて少ない状況になっている。ステータス情報を使った場合、実態のない企業を母数に入れているため、倒産確率を過小評価している可能性が高い。また業績データを使った分析では企業活動は行っているが業績データを開示しなくなった企業があるために倒産確率を過大評価している可能性がある。そのため両方の結果を観測しておく必要がある。

いずれにしても、倒産確率の年齢分布は年齢の小さい企業から年齢の高い企業に向けて指数関数的に落ちていく傾向がある。国による違いも指数関数を補正することによって吸収できると考えられる。倒産確率の年齢分布を用いることによって生存企業の年齢分布を再現することができることが分かった。この分析には倒産確率の年齢分布が十数年のスパンでほとんど変化しないという仮定を用いて行っている。日本の場合は企業年齢 60 歳の所で年齢分布の指数の大きさが大きく変化している。第2次世界大戦前後で倒産確率の年齢分布の形が大きく変わったと考えられる。アメリカの分析では生存企業の年齢分布に日本のような不連続な変化はない。しかし倒産確率の年齢分布は日本に比べて少し分析が難しい形になっている。そのため倒産確率の年齢分布を用いることによって生存企業の年齢分布を再現する試みは年齢の若い企業について合わせられる程度になっている。

倒産確率の売上規模による分布は売上が大きい企業の倒産確率がベキ関数で落ちていく傾向にあることが分かった。国によって多少のベキ指数は違う者の定性的な性質は変わらない。この定性的な性質は売上データの欠損を使う方法で再現することが可能である。企業の成長に関して、我々はこれまでに様々な分析を行ってきた。その手法を倒産確率の売上分布の再現に用いて倒産確率の分布を説明することを試みた。何らかの理由で十分な企業活動ができなくなり、企業活動が継続できなくなった結果、業績報告ができなくなった企業が倒産企業の大半を占めると仮定する。成長率分布の下限以下

に現れるはずであった十分に成長しきれなかった企業が倒産した企業だという仮定を置いて倒産確率の計算をおこなったところ、倒産確率分布のベキ分布を定性的に説明することができた。定量的には十分に説明できる結果には至らなかった。簡単のため成長率分布の形を直線で近似していたが、実際は裾が広がっているなのでこの部分を改良すればもう少し精度はよくなるだろうと考えられる。また売上が計上できなくなる企業には今回の仮定以外にも何らかの理由があると考えられる。今回の分析で企業の成長と倒産の関係を定性的にでも評価できたことには非常に大きな意味があると考えられる。企業の倒産を分析する一番の目的はリスク評価にある。通常の企業は売上が伸びなくなり、その結果利益が上がらなくなり、最終的には資金繰りがつかなくなったところで倒産する。今回の分析で財務データを得ることができない数多くの企業が存在することが分かった。本稿の分析では売上が計上している企業の成長率を用いて倒産企業の確率分布を分析していることになる。データを積み重ねることによって売上の成長率の段階で会社のサイズによる倒産確率をある程度見積もることができるようになる。個別の企業にとっては、成長率分布のどの位置に存在するのかによって今後の倒産確率を見積もることができるようになる。保険事業にとっては全体の分布を知ることによって適正な保険のあり方に有効な指標を与えることになる。

8.2 今後の展望

今後は分析の精度を上げるだけでなく、他の財務データと倒産確率のリンクを進めていきたい。財務データの中には売上と同じように企業の規模に関する変数がある。例えば従業員数や資産等がそれにあたる。このような変数に関して値が大きな大規模領域に関しては 5.3 節で説明したようなベキ指数が違うだけのベキ分布の関係にあることが分かっている。そのため大規模領域では他の財務データを使っても定性的にはそれほど違いはない結果が得られると考えられる。問題はベキ分布では表すことができない領域でそのような領域では綺麗な対称性などが見つかっていない。倒産確率に寄与しそうな企業は中規模以下の領域に多いと考えられるので慎重に分析を進める必要がある。

また 7.2 節で行った分析は解析的に進めていきたいという理由のために多くの近似や仮定を使ってしまっている。この節では 1 年間隔の成長率分布から時系列の推移を再構築して企業年齢に関する分布を導き出す試みを行っている重要な部分である。成長率分布を詳しく見ることは時間に関する微分解析的な発想に基づいており、時系列の推移を再構築することは積分解析的な発想に基づいている。物事の全体を知るためには微分解析的な分析だけでは十分ではなく、積分解析的な分析も加えた両方向からの分析が重要である。売上の成長率に関する分布では入手したデータの直近 10 年程を使って調査を行ったが、年次の違いによる大きな違いはほとんどなかった。ただし図

7.2 の売上の企業年齢別による平均に関して 20 歳付近の所で大きく上下に動いている部分が気になるところである。こういった現象は他の国のデータでは見つかっていない。2008 年の 15 から 20 年前に生まれた企業はちょうどバブル景気とその崩壊に何らかの形でかかわっている企業であると考えられる。その時代のビッグデータを入手して成長率分布を観測することで図 7.2 の大きく上下してるように見える部分の理由が分かるかも知れない。現在では 20 年前に比べてビッグデータの入手が簡単になっている。今後も継続的に世界中のビッグデータの分析を続けていくことが重要である。

<参考文献>

1. J. C. Neves and A. Vieira, “Improving bankruptcy prediction with Hidden Layer Learning Vector Quantization”, *European Accounting Review*, vol. 15 , (2006) issue 2.
2. H. E. Stanley, “Introduction to Phase Transitions and Critical Phenomena”, (Clarendon Press, Oxford, 1971).
3. A. Saichev, Y. Malevergne, and D. Sornette, “Theory of Zipf’s Law and Beyond”, (Springer, Berlin, 2009).
4. R. Gibrat, “Les Inegalites Economique”, (Sirey, Paris, 1932).
5. J. Sutton, “Gibrat’s Legacy”, *J. Econ, Lit.* 35 (1997) 40.
6. Y. Fujiwara, et al., *Physica A* 321 (2003) 598.
7. Y. Fujiwara, et al., *Physica A* 335 (2004) 197.
8. A. Ishikawa, *Physica A* 367 (2006) 425.
9. A. Ishikawa, *Physica A* 383 (2007) 79.
10. M. Tomoyose, S. Fujimoto, and A. Ishikawa, *Prog. Theor. Phys. Supple.* 179 (2009) 114.
11. A. Ishikawa, S. Fujimoto, and Mizuno, *Physica A* 390 (2011) 4273.
12. 石川 温, 藤本 祥二, 水野 貴之, 渡辺 努, 金沢学院大学紀要 第 14 号(2016) 経営経済情報自然科学編 13-21.
13. Bureau van Dijk, <http://www.bvdinfo.com/Home.aspx>.
14. 上場企業サーチ <http://上場企業サーチ.com/articles/0143582616>
15. Small and Medium Enterprise Agency, http://www.chusho.meti.go.jp/sme_english/index.html.
16. V. Pareto, *Cours d'Economie Politique* (Macmillan, London, 1897).
17. X. Gabix, *The Quarterly Journal of Economics*, vol. 144, (1999) 739-767.

18. H. Takayasu, A. Sato and M. Takayasu, *Phys. Rev. Lett.* 79 (1997), 966.
19. B. A. Hubermann and L. Adamic, *Nature*, vol. 401, (1999) 131.
20. W. J. Reed, *Economics Letters*, vol. 74, (2001) 15-19.
21. W. J. Reed and B. D. Hughes, *Physical Review*, E66, (2002) 067103.
22. Y. Malevergne, A. Saichev and D. Sornette, (2008) available at <http://ssrn.com/abstract=1083962>.
23. A. Saichev, Y. Malevergne and D. Sornette “Theory of Zipf's Law and Beyond “ (2010) Springer.
24. A. Coad, “The Growth of Firms”, (2009) Edward Elgar.