

財務データと ESG レーティングデータによる 株式時価総額の統計モデリング

地 道 正 行
阪 智 香

要 旨

本稿では、財務データと ESG レーティングデータを結合したものをを用いて、非対称分布族を考慮した株式時価総額の対数の分布を検討するとともに、非対称誤差をもつ両対数モデルを用いて株式時価総額の統計モデリングを行う。その際、探索的データ解析の視点から、データ可視化によって得られた知見を統計モデリングに利用し、さらに赤池情報量規準を利用することによってモデル選択を行う。なお、本研究は動的文書生成によって再現可能研究の立場から実施される。

キーワード：財務データ (Financial Data), ESG レーティングデータ (ESG Ratings Data), 非対称分布族 (Family of Skew Distributions), 探索的データ解析 (Exploratory Data Analysis), 再現可能研究 (Reproducible Research)

I はじめに

本稿では、Bureau van Dijk (BvD) 社¹⁾ から提供されるデータベース Osiris²⁾ から抽出された世界154カ国の上場企業の企業財務データと FTSE

1) ビューロー・ヴァン・ダイク社 <http://www.bvdinfo.com/ja-jp/home>

2) Osiris (オシリスまたはオサイリス) は世界の上場企業約 9 万社以上の情報が国際比較可能な統一のフォームで収録されたデータベースである。収録情報としては、世界の上場企業および上場廃止企業 (一般事業会社) の財務三表 (貸借対照表, 損益計算書, キャッシュフロー計算書) 等の財務情報が最長30年にわたって収録されている。

Russell 社³⁾ の ESG レーティングデータ⁴⁾ の両方を結合したものを利用し、株式時価総額の対数をとったもの（以後、「対数株式時価総額」と記す場合がある）の分布と両対数モデルによる統計モデリングを行う。その際、大鹿ら（2020）で検討されている財務指標と ESG 指標の両方を説明変数としてもつものを、誤差分布を非対称分布族（family of skew distributions）に拡張することによって、モデルの当てはめの精度向上をはかる。

本稿の構成は以下のようなものである。まず、本稿で扱う財務データと ESG レーティングデータ（非財務データ）が結合されたものについて解説するとともに（II 節）、そのデータ可視化を行う（III 節）。データ可視化は、株式時価総額とその対数をとったもののヒストグラムと正規 Q-Q プロットを描くことによって分布情報を得る。また、（単年度分の）全指標に対するデータの対散布図を描くことによって、（対数）株式時価総額とその他の指標の関係を探索的に調べる。次に、これらのデータ可視化によって得られた知見を利用し、統計モデリングを実施する。具体的には、対数株式時価総額に非対称正規分布（Skew-Normal: SN distribution）と非対称ティータ分布（Skew-t: ST distribution）を当てはめ（IV 節）、その結果も考慮して、株式時価総額に両対数モデルを当てはめる（V 節）。両対数モデルについては、誤差分布として、正規誤差（N error）、非対称正規誤差（SN error）、非対

る。

3) <https://www.ftserussell.com>

4) FTSE Russell ESG レーティングデータの評価モデルはボトムアップ・アプローチによる以下のような階層によって構成されている。まず、300以上の指標について、企業の開示情報（有価証券報告書、CSR 報告書、統合報告書、ウェブサイト等）に基づく調査が行われる（1社につき平均約125の指標が適用される）。この調査項目は、各テーマに関連する国際的な枠組みをもとに構成されている。この調査の精査後に、以下の手順によって主なスコアが算出される。

(1) 企業からのフィードバックを受け付け、追加調査を経た最終調査結果に基づき、「環境」が5テーマ、「社会」が5テーマ、「ガバナンス」が4テーマの合計14テーマの評価が行われる（テーマスコア）。

(2) 「環境」「社会」「ガバナンス」それぞれの総合評価が算定される（ピラースコア）。

(3) ESG 全体の総合評価が算出される（サマリースコア）。

上記の評価にあたっては、テーマ・ピラー段階でリスクエクスポージャーも考慮される。

称ティー誤差 (ST error) をもつものを仮定し, 当てはめを行う。次に, これらのモデルの当てはまり具合を赤池情報量規準⁵⁾ (Akaike Information Criterion: AIC) で評価する (VI 節)。さらに, モデルの当てはまりの経年変化を考察し (VII 節), 本稿で詳しく述べるできなかったデータ解析の結果を補足する (VIII 節)。最後に, 今後の展望を述べる (IX 節)。

追加情報として, 付録 A に, 本稿で扱うデータについての読み込みとデータラングリング (Wickham and Grolemund, 2016 参照), および欠測パターンについての情報が与えられている。また, 付録 B には正規分布 (Normal: N distribution) と対数正規分布 (Log-Normal: LN distribution) について, 付録 C には非対称正規分布 (Skew-Normal: SN distribution) と対数非対称正規分布 (Log-Skew-Normal: LSN distribution), および, 非対称ティー分布 (Skew-t: ST distribution) と対数非対称ティー分布 (LSkew-t: LST distribution) に関する尤度, 赤池情報量規準などの説明が与えられている。さらに, 本研究はデータ解析環境 R⁶⁾ を利用して行っており, 本稿で利用したスクリプトを付録 D に, R に関する情報を付録 E に与えている。

本研究は, Tukey (1977) による探索的データ解析 (Exploratory Data Analysis: EDA) の視点に立って行われる⁷⁾。また, 本稿の執筆には, Sweave⁸⁾ による動的文書生成によって再現可能研究を実施することを試みている⁹⁾。

II データとその要約

本稿では, BvD 社のデータベース Osiris から抽出したものにもとづくデー

5) 赤池情報量規準については, 例えば, Akaike (1973), Konishi and Kitagawa (2008) を参照されたい。

6) <https://www.r-project.org>

7) R による EDA の実行については, 例えば, 地道 (2018-c) を参照されたい。

8) <https://stat.ethz.ch/R-manual/R-devel/library/utils/doc/Sweave.pdf>

9) Sweave を利用した動的文書生成と再現可能研究については, 例えば, 地道 (2018-b) を参照されたい。

タセット（以後、OsirisC2020と表記）¹⁰⁾とFTSE Russell社から提供されるESGレーティングデータにもとづくデータセット（以後、FTSE2020と表記）¹¹⁾を結合したものの（以後、OsirisC-FTSE2020と表記）¹²⁾をベースとして利用している。また、モデリングを行う際に、対数を計算するために財務データについては正の値に限定しており、さらに決算月数については12ヵ月のものを選択している。データのRへの読み込みとデータラングリングについての具体的な操作については、付録Aを参照されたい。利用するデータの要約は以下のようなものである：

-
- 10) 本研究では、データベースOsirisの2020年3月版から連結（Consolidated）財務情報を上場企業（上場廃止企業含む）96,377社の主要財務情報（91項目）について抽出したデータを前処理したデータセットOsirisC2020を利用する。なお、前処理については、地道（2018-a）、地道、阪（2021）を参照されたい。特に、地道、阪（2021）ではGNU Parallel（cf. Tange, 2018）を用いて、前処理を並列化することについて議論されている。なお、規模が大きなデータの前処理とデータラングリングの並列化については、地道（2020-a, b）が参考になる。
 - 11) FTSE Russell社のESGレーティングは、毎年2回、6月末と12月末に算出される。本研究で用いるESGレーティングデータは、2015～2020年（6年間）の各年における6月末の評価を反映した後に7月に取得したデータである（最新のデータは2020年7月に入手）。2020年7月のESGレーティングは、FTSE Indexを構成する大型・中型株企業などの先進国25カ国2,167社・新興国24カ国1,778社の合計49カ国3,945社を対象としている。この対象企業は各国の株式時価総額の大きい順に選定されており、世界の全上場企業の株式時価総額（42,473社）に占める割合を算定したところ73%、世界の全上場企業の売上（52,495社）に占める割合を算定したところ69%である。したがって、FTSE Russell社のESGレーティングの対象企業数はOsirisと比較して少ないものの、株式時価総額と売上上の観点からはデータベースOsirisに収録されているの大部分の企業が含まれていることとなる（阪, 2021）。
 - 12) データセットOsirisC2020とFTSE2020の前処理と結合に関する詳細は、地道、阪（2021）を参照されたい。

データの要約

firmID	year	year_SPT	market_cap	shareholders
Length :14366	Min. :2013	Min. :2015	Min. : 32	Min. : 7874
Class :character	1st Qu.:2014	1st Qu.:2016	1st Qu.: 2749	1st Qu.: 1443453
Mode :character	Median :2016	Median :2018	Median : 5890	Median : 2997843
	Mean :2016	Mean :2018	Mean : 14724	Mean : 7285962
	3rd Qu.:2017	3rd Qu.:2019	3rd Qu.: 13060	3rd Qu.: 6746400
	Max. :2018	Max. :2020	Max. :1090308	Max. :294365704

net_income	log.mc	log.sh	log.ni	ESG
Min. : 74	Min. : 3.465	Min. : 8.971	Min. : 4.304	Min. :0.000
1st Qu.: 147567	1st Qu.: 7.919	1st Qu.:14.183	1st Qu.:11.902	1st Qu.:1.700
Median : 326486	Median : 8.681	Median :14.913	Median :12.696	Median :2.500
Mean : 921304	Mean : 8.758	Mean :14.975	Mean :12.760	Mean :2.473
3rd Qu.: 767808	3rd Qu.: 9.477	3rd Qu.:15.725	3rd Qu.:13.551	3rd Qu.:3.300
Max. :98616437	Max. :13.902	Max. :19.500	Max. :18.407	Max. :5.000
				NA's :885

E	S	G	type
Min. :0.000	Min. :0.000	Min. :0.000	Length:14366
1st Qu.:1.000	1st Qu.:1.200	1st Qu.:2.500	Class :character
Median :2.000	Median :2.300	Median :3.100	Mode :character
Mean :2.126	Mean :2.209	Mean :3.168	
3rd Qu.:3.100	3rd Qu.:3.200	3rd Qu.:4.000	
Max. :5.000	Max. :5.000	Max. :5.000	
NA's :885	NA's :885	NA's :885	

ここで、変数名（列名）は以下のようなものである。なお、株式時価総額、純資産、当期純利益は期末の数値である。

firmID: 企業名+BvD ID

year: 財務データセット OsirisC2020 の会計年度

year_SPT: ESG レーティングデータセット FTSE2020 におけるデータ公表年

market_cap: 株式時価総額（単位：100万 US ドル）

shareholders: 純資産（単位：1,000 US ドル）

net_income: 当期純利益（単位：1,000 US ドル）

log.mc: 株式時価総額の対数（単位：100万 US ドルの自然対数）

log.sh: 純資産の対数（単位：1,000 US ドルの自然対数値）

log.ni: 当期純利益の対数（単位：1,000 US ドルの自然対数）

ESG: ESG レーティングのサマリー（summary）指標

E: ESG レーティングの環境（Environment）指標

S: ESG レーティングの社会（Social）指標

G: ESG レーティングのガバナンス (Governance) 指標

type: 国のタイプ E: 新興国 (Emerging country), G: 先進国 (Global (developed) country)

このデータにおいて、株式時価総額 `market_cap` (とその対数 `log.mc`), 純資産 `shareholders` (とその対数 `log.sh`), 当期純利益 `net_income` (とその対数 `log.ni`) がデータセット `OsirisC2020` からのものであり、サマリー指標 ESG, 環境指標 E, 社会指標 S, ガバナンス指標 G, 国のタイプ `type` がデータセット `FTSE2020` からのものである。なお、2種類の年情報 `year, year_SPT`¹³⁾ は、財務データセット `OsirisC2020` の年情報 `year` については、会計年度が当てられているのに対して、ESG レーティングデータセット `FTSE2020` の年情報 `year_SPT` は、公表された年が当てられている。また、これらの年情報は表記上2年のラグがあることに注意する必要がある。この理由は、ESG レーティングデータが公表されるまでに約2年の時間を要するため¹⁴⁾、ESG レーティングの評価のもとになった企業のESG情報の対象年度は、2年前のものであると想定し、財務データの会計年度とESG情報の対象年度(公表年の2年前の年度)を対応させたからである(地道, 阪, 2021も参照)。

年度別の企業数は表1によって与えられる¹⁵⁾。データの要約結果からもわかるが、ESG レーティングデータに関しては、欠測値 (NA) が存在するため、実際にモデリングに利用する際はデータ数が減少する。欠測値除去後の

13) 年情報を与える変数名 `year_SPT` における `SPT` は、ESG 指標の3つの階層、すなわち、「サマリー (要約)」(Summary), 「ピラー (支柱)」(Pillar), 「テーマ」(Theme) の頭文字をとったものである。詳しくは、FTSE Russell (2019), 阪 (2021), 地道, 阪 (2021) を参照されたい。なお、本稿では、サマリーである ESG とピラーである E, S, G をデータとして利用している。

14) FTSE Russell 社が ESG レーティングデータを開示するまでには、企業から提供された ESG 情報を、収集・評価し、企業にフィードバック後、評価が確定し、公表という段階を経る必要がある。

15) 脚注11) で与えられている企業数は、(結合する前の) `FTSE2020` に関するものであり、表1で与えられるものよりも多くの企業が存在する。この差は、データセット `OsirisC2020` と `FTSE2020` を結合する際に生じたものである。

年度別企業数については、表 2 を参照されたい。なお、データの欠測パターンについては、付録 A を参照されたい。

表 1：年度別企業数

	会計年度	ESG 指標年度	先進国企業数	新興国企業数	総企業数
1	2013	2015	1466	629	2095
2	2014	2016	1489	647	2136
3	2015	2017	1463	647	2110
4	2016	2018	1538	698	2236
5	2017	2019	1625	1282	2907
6	2018	2020	1613	1269	2882

表 2：欠測値除去後の年度別企業数

	会計年度	ESG 指標年度	先進国企業数	新興国企業数	総企業数
1	2013	2015	1422	594	2016
2	2014	2016	1465	633	2098
3	2015	2017	1457	641	2098
4	2016	2018	1527	679	2206
5	2017	2019	1607	724	2331
6	2018	2020	1605	1127	2732

データの要約結果から、財務データである株式時価総額 `market_cap`、純資産 `shareholders`、当期純利益 `net_income` は全て右に歪んでいることがわかる。このため、これらの対数をとったもの (`log.mc`, `log.sh`, `log.ni`) をデータラングリングの段階で新たに列に追加している。具体的な処理については、付録 A を参照されたい。

本稿では、このデータを用いて株式時価総額 `market_cap` (とその対数 `log.mc`) を純資産 `shareholders` (とその対数 `log.sh`) と当期純利益 `net_income` (とその対数 `log.ni`)、さらに各種の ESG 指標 `ESG`, `E`, `S`, `G`、および国のタイプ `type` を用いて説明するための統計モデリングを行う。その際、次節ではこれらの多変量データを可視化することによってモデリングを行う上で重要な知見を探索する。なお、年情報は、財務データに関して議論するときは会計年度 `year` を利用し、ESG レーティングデータと財務

データとともに考えるときは、データセット FTSE2020 の年情報 `year_SPT` を利用している。

III データ可視化

この節ではデータを幾つかの観点から可視化する。

1. 株式時価総額の分布に関する可視化

本稿では株式時価総額 `market_cap` を幾つかの変量で説明するためのモデリングを行うことから、まずこの変量の分布に関する可視化を行う。そのために、会計年度が2018年の株式時価総額データのヒストグラムを与える。

ヒストグラム（図1）から、分布は極端に右に歪んでいることがわかる。Jimichi *et al.* (2018), 地道 (2019) 等でも指摘されているが、このような歪みを補正するための代表的な変換は対数をとることである (Tukey, 1977, Mosteller and Tukey, 1977, Fox and Weisberg, 2019 も参照のこと)。実際に、対数株式時価総額 $\log(\text{market_cap})$ のヒストグラムを描くと、図2のようになる。

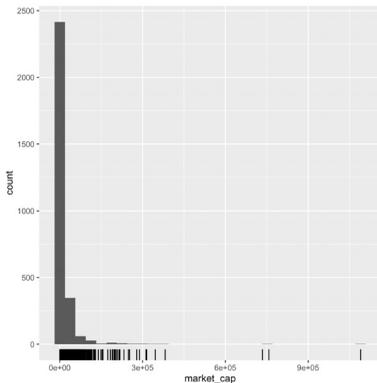


図1：株式時価総額のヒストグラム：ラグ付き

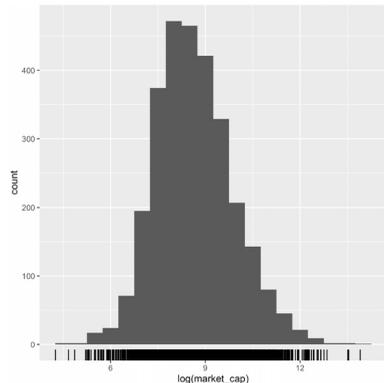


図2：対数株式時価総額のヒストグラム：ラグ付き

図2より、株式時価総額を対数変換したものの分布は、ほぼ対称の形状と

なるが、若干右に歪んだものとなっているように思われる。そこで、対数株式時価総額の正規 Q-Q プロットを描いたものが図 3 である。

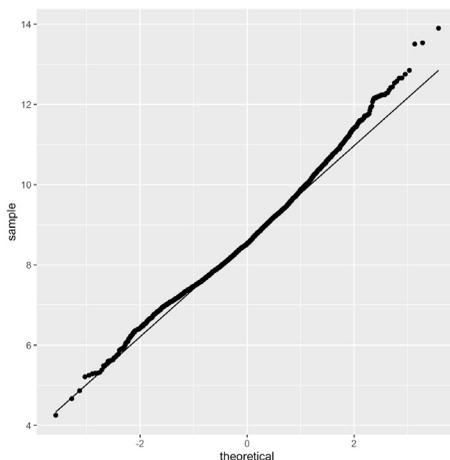


図 3：対数株式時価総額の正規 Q-Q プロット

正規 Q-Q プロットから、裾の部分での当てはまりが顕著に悪いことがわかる。なお、歪度は 0.39 であることから、右に歪んでいることがわかる。

以上の結果から、株式時価総額の対数をとったものの分布構造は、非対称正規分布、または非対称ティー分布などに代表される非対称分布族に属する分布に従っていることを検討する必要があることがわかる。非対称分布族についての詳細は、Azzalini and Capitanio (2014) を参照されたい。また、非対称正規分布、非対称ティー分布については、付録 C に簡単な解説を与えているので併せて参照されたい。

2. 対散布図

対散布図 (図 4) を描くことによって、2 変量間の関係とそれぞれの変量の分布に関する情報を探る。以下の説明では、対散布図のブロックを行列の成分に見立てて、ブロックの位置を成分 (i, j) , $i, j=1, \dots, 8$ と表記する。

対角ブロック：

(RD1) 財務データ (market_cap, shareholders, net_income) の推定された密度関数から、全て右に歪んだ分布に従っている：

$$(i, i), \quad i=1, 2, 3$$

(RD2) ESG レーティングデータ (ESG, E, S, G) の推定された密度関数から、国のタイプ type に関して異なった分布に従っている：

$$(i, i), \quad i=4, \dots, 7$$

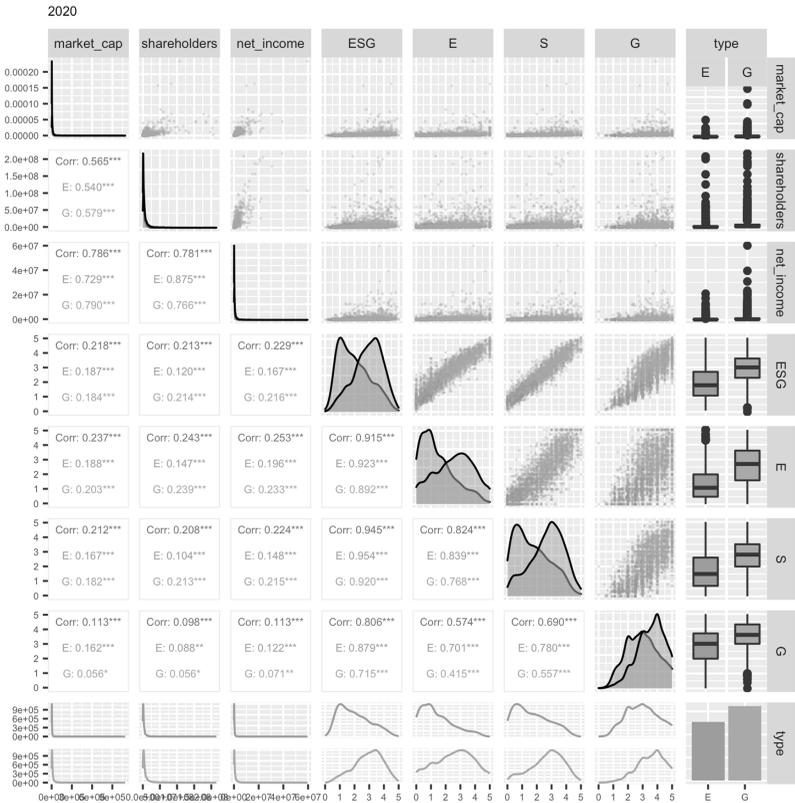


図 4：対散布図：粗データ

(RD3) 国のタイプ別 type の企業数の頻度がバーチャートから先進国に

属する企業数が多い：

$$(i, i), \quad i=8$$

上三角ブロック：

(RU1) 財務データ (market_cap, shareholders, net_income) の各ペアの散布図は 2 次元の意味で歪んでいる：

$$(i, j) = (1, 2), (1, 3), (2, 3)$$

つまり、原点付近にデータが集中しており、原点から離れるに従って分布状況は粗になっている。

(RU2) ESG レーティングデータ (ESG, E, S, G) のそれぞれのペアの散布図からそれぞれの指標は正の相関関係にある：

$$(i, j) = (4, 5), (4, 6), (4, 7), (5, 6), (5, 7), (6, 7)$$

特に、ESG 指標 ESG と (E, S, G) の各指標の間には、(E, S, G) 間の相関よりも強い正の相関関係があるように見受けられる：

$$(i, j) = (4, 5), (4, 6), (4, 7)$$

(RU3) 財務データ (market_cap, shareholders, net_income) と ESG レーティングデータ (ESG, E, S, G) のそれぞれのペアの散布図から線形関係は見受けられない：

$$(i, j), \quad i=1, 2, 3, \quad j=4, \dots, 7$$

(RU4) 財務データ (market_cap, shareholders, net_income) の国のタイプ別 type のボックスプロットから、分布は、それぞれのタイプ共、右に歪んでいることがわかる：

$$(i, j), \quad i=1, 2, 3, \quad j=8$$

(RU5) ESG レーティングデータ (ESG, E, S, G) の国のタイプ type 毎のボックスプロットから、分布の位置は新興国 (type=E) よりも先進国 (type=G) の方が高いことがわかる：

$$(i, j), \quad i=4, \dots, 7, \quad j=8$$

下三角ブロック：

- (RL1) 財務データ (market_cap, shareholders, net_income) の各ペアの相関から、それぞれ正の相関があり、国のタイプ type によって際だった差異は見られない：

$$(i, j) = (2, 1), (3, 1), (3, 2)$$

- (RL2) ESG レーティングデータ (ESG, E, S, G) のそれぞれのペアの散布図から、正の相関があることがわかる：

$$(i, j) = (5, 4), (6, 4), (6, 5), (7, 4), (7, 5), (7, 6)$$

特に、ESG 指標 ESG と (E, S, G) の各指標の間には0.8~0.9の相関があることがわかる：

$$(i, j) = (5, 4), (6, 4), (7, 4)$$

このことは、対称な位置で与えられる散布図の結果を肯定している。なお、この結果は、統計モデリングを行う際に、ESG 指標 ESG と (E, S, G) の各指標を説明変数として同時に採用することの是非を検討する必要があることを示唆する。つまり、多重共線性 (multicollinearity) の意味で、全ての指標を説明変数に入れることは得策ではないことを表している。

- (RL3) 財務データ (market_cap, shareholders, net_income) と ESG レーティングデータ (ESG, E, S, G) のそれぞれのペアの相関は無相関に近いことがわかる：

$$(i, j), \quad i=4, \dots, 7, \quad j=1$$

このことは、対称な位置で与えられる散布図の結果を肯定している。

- (RL4) 財務データ (market_cap, shareholders, net_income) の国のタイプ type の推定された密度からそれぞれのタイプ共、分布は歪んでいることがみてとれる：

$$(i, j), \quad i=8, \quad j=1, \dots, 3$$

- (RL5) ESG レーティングデータ (ESG, E, S, G) の国のタイプ type 毎の推定された密度から、新興国 (type=E) の分布は左に、先

進国 (type=G) の分布は右に、それぞれ歪んでいることがわかる：

$$(i, j), \quad i=8, \quad j=4, \dots, 7$$

これらの結果から、株式時価総額 market_cap とその他の財務データについては、ある程度の正の相関はあるものの、ESG 指標に関しては無相関に近い。ため、株式時価総額 market_cap をその他の変数で説明するための統計モデリングを単に線形回帰モデルで行っても、当てはまりが良いモデルを構築することが難しいことが予想される。この問題を考慮して、財務データに関する歪みを緩和するためにそれぞれの対数をとった変量

$$\begin{aligned} \log.mc &:= \log(\text{market_cap}), \\ \log.sh &:= \log(\text{shareholders}), \\ \log.ni &:= \log(\text{net_income}) \end{aligned}$$

と ESG レーティングデータとの対散布図を描き直したものが図 5 である。

粗データの場合 (図 4) から対数変換した場合 (図 5) との比較において主要な変化としては以下のようなものがある：

対角ブロック：

(LD1) 対数変換された財務データ ($\log.mc, \log.sh, \log.ni$) の分布は歪みがほぼなくなり、対称化されていることがわかる：

$$(i, i), \quad i=1, 2, 3$$

しかしながら、詳しく調べると III 節でも指摘したが、(国のタイプにかかわらず) 正確には若干歪んだものとなっていることがわかる。この結果は、これらのデータは非対称分布に従っている可能性があることを指摘するものである。

上三角ブロック：

(LU1) 対数変換された財務データ ($\log.mc, \log.sh, \log.ni$) の各ペアの散布図は 2 次元の意味で歪みが解消されているが、若干原点方向から右上がりの方向に「潰れている」ことがうかがえる：

$$(i, j) = (1, 2), (1, 3), (2, 3)$$

この結果からも、非対称分布に従っている可能性があることを指摘している。このことについては、次節で詳細に検討する。

(LU2) 対数株式時価総額データ $\log.mc$ と ESG レーティングデータ (ESG, E, S, G) のそれぞれのペアの散布図から緩やかな線形関係がみられる：

$$(i, j), i=1, j=4, \dots, 7$$

このことは、対数株式時価総額 $\log.mc$ を ESG レーティング指標 (ESG, E, S, G) で線形モデリングを行なう動機となる。

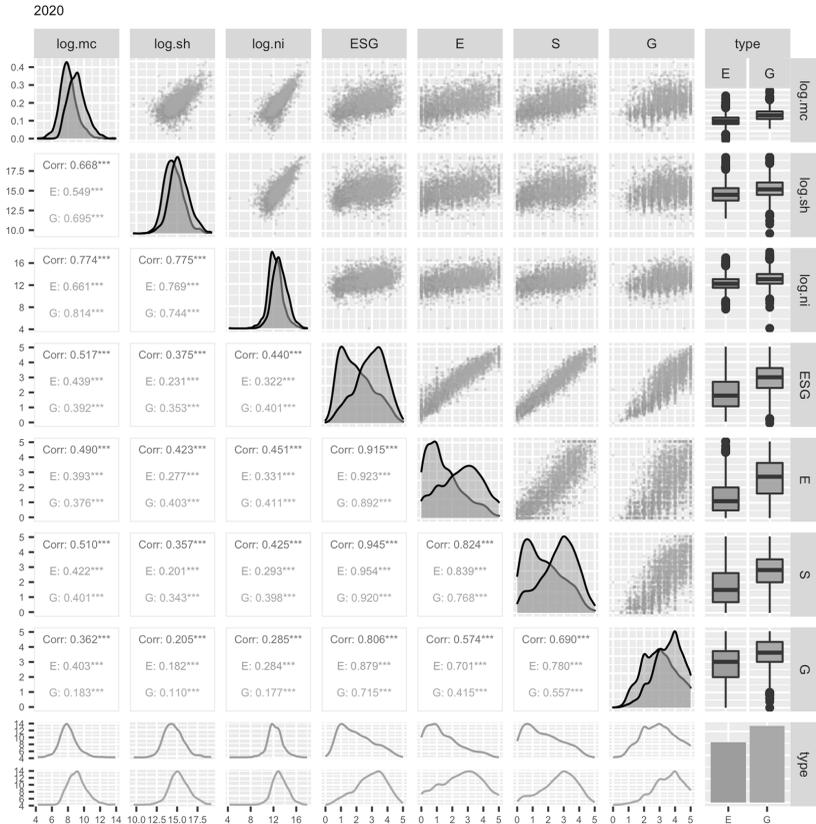


図 5 : 対散布図 : 対数スケール

下三角ブロック：

(LL1) 対数変換された財務データ ($\log.mc$, $\log.sh$, $\log.ni$) の各ペアの相関から、それぞれ正の相関があり、国のタイプ $type$ によって若干差異がみられる：

$$(i, j) = (2, 1), (3, 1), (3, 2)$$

この結果から、国のタイプの情報をダミー変数として統計モデリングを行うことを検討する。

(LL2) 対数株式時価総額 $\log.mc$ と ESG レーティング指標 (ESG , E , S , G) のそれぞれのペアの相関は緩やかな正の相関があることがわかる：

$$(i, j), \quad i=4, \dots, 7, \quad j=1$$

このことは、対称な位置で与えられる散布図の結果を肯定するとともに、対数株式時価総額 $\log.mc$ を ESG レーティング指標 (ESG , E , S , G) で線形モデリングを行なう動機となる。

(LL3) 対数株式時価総額 $\log.mc$ の国のタイプ $type$ 毎に推定された密度は、それぞれのタイプ共、右に歪んでいることがわかる：

$$(i, j) = (8, 1)$$

この結果は、前小節の結果と整合する。

以上の可視化の結果を踏まえて、次節で対数株式時価総額の分布に関するモデリングを考察するとともに、その結果を利用して、財務指標と ESG レーティング指標でモデリングを行う。

IV 対数株式時価総額の分布に関するモデリング

この節では、前節で与えた可視化（ヒストグラム（図1、2）、正規 Q-Q プロット（図3））と財務データの対数をとった場合の対散布図による可視化の結果（LD1）、（LL3）より、対数株式時価総額に非対称正規分布と非対称テーパー分布を当てはめる。なお、ここで利用するアイデアは、地道（2017-a, b）、Jimichi *et al.*（2018）で考察された結果を応用したものであり、非対

称分布の当てはめに関しては R の `sn` パッケージ¹⁶⁾ を利用している。また、(対数)非対称正規分布と(対数)非対称ティー分布については、付録 C を参照されたい。

1. 対数非対称正規分布の当てはめ

2018会計年度末の株式時価総額 `market_cap` が対数非対称正規分布 $LSN(\xi, \omega^2, \alpha)$ に従うと仮定し、その対数 $\log(\text{market_cap})$ に非対称正規分布 $SN(\xi, \omega^2, \alpha)$ を当てはめることを考える。母数 (ξ, ω, α) ¹⁷⁾ を最尤法¹⁸⁾ によって推定した結果は

$$(\hat{\xi}, \hat{\omega}, \hat{\alpha}) = (7.48, 1.70, 1.73)$$

であり、これより推定された確率密度関数 (統計モデル)

$$f_{SN}(\log(\text{market_cap}) | \hat{\xi}, \hat{\omega}, \hat{\alpha}) := \frac{2}{\hat{\omega}} \phi\left(\frac{\log(\text{market_cap}) - \hat{\xi}}{\hat{\omega}}\right) \\ \times \Phi\left(\hat{\alpha} \frac{\log(\text{market_cap}) - \hat{\xi}}{\hat{\omega}}\right)$$

で与えられる。ここで、

$$\phi(z) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), \quad \Phi(z) := \int_{-\infty}^z \phi(x) dx = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

は、それぞれ、標準正規分布 $N(0, 1)$ の確率密度関数と累積分布関数である。

この統計モデルを株式時価総額の対数 $\log(\text{market_cap})$ のヒストグラムに重ね書きしたものと標準化直接残差の 2 乗に関する P-P プロット¹⁹⁾ を、

16) <http://azzalini.stat.unipd.it/SN/>

17) 非対称正規分布に関する母数付け (parameterization) には、「直接母数」(Direct Parameter: DP) と「中心化母数」(Centered Parameter: CP) がある。ここで、 (ξ, ω, α) は非対称正規分布の直接母数である。なお、これらの母数の役割については Azzalini and Capitanio (2014) を参照されたい。

18) 非対称正規分布に関する母数の最尤推定についての詳細は、Azzalini and Capitanio (2014) の第 3 章を参照されたい。なお、本稿を通じて最尤推定値の推定・検定などの推測結果 (ゼット比検定表等) は、最尤推定量の漸近正規性を利用している。最尤法に関する文献としては、稲垣 (2003) 等を参照されたい。

19) 標準化 (正規化) 直接残差の 2 乗は近似的にカイ自乗分布に従うことから、その結果を使って経験的な確率点と理論的な確率点を対としてプロットしている。詳細は、Azzalini and Capitanio (2014) の p. 61、または地道 (2017-b) の付録 B を参照されたい。

それぞれ、図6と図7に与える。これらの図から、当てはまりはある程度よいと判断できる。

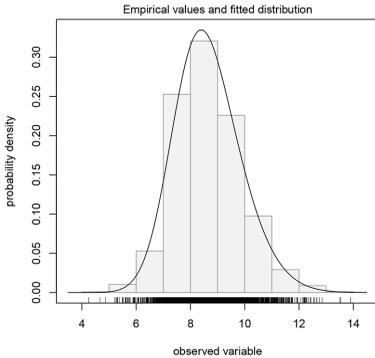


図6：株式時価総額の対数のヒストグラムと非対称正規分布にもとづく統計モデル

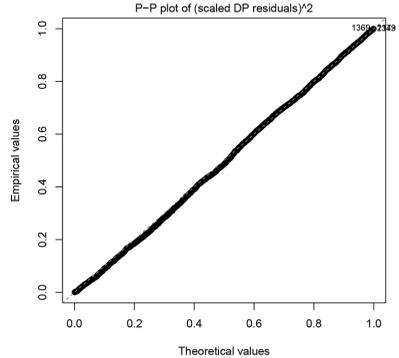


図7：株式時価総額の対数に非対称正規分布を当てはめたときの標準化直接母数残差の2乗のP-Pプロット

2. 対数非対称ティー分布の当てはめ

2018会計年度末の株式時価総額 market_cap が対数非対称ティー分布 $\text{LST}(\xi, \omega^2, \alpha, \nu)$ に従うと仮定し、その対数 $\log(\text{market_cap})$ に非対称ティー分布 $\text{ST}(\xi, \omega^2, \alpha, \nu)$ を当てはめることを考える。非対称ティー分布についての詳細については、Azzalini and Capitanio (2014) を参照されたい。

まず、母数 $(\xi, \omega, \alpha, \nu)$ を最尤法によって推定した結果は、

$$(\hat{\xi}, \hat{\omega}, \hat{\alpha}, \hat{\nu}) = (7.57, 1.55, 1.53, 19.54)$$

であり、これより推定された確率密度関数（統計モデル）は、

$$\begin{aligned} & f_{\text{ST}}(\log(\text{market_cap}) \mid \hat{\xi}, \hat{\omega}, \hat{\alpha}, \hat{\nu}) \\ &= \frac{2}{\hat{\omega}} f_t\left(\frac{\log(\text{market_cap}) - \hat{\xi}}{\hat{\omega}} \mid \hat{\nu}\right) \\ & \times F_t\left(\hat{\alpha} \frac{\log(\text{market_cap}) - \hat{\xi}}{\hat{\omega}} \sqrt{\frac{\hat{\nu} + 1}{\left(\frac{\log(\text{market_cap}) - \hat{\xi}}{\hat{\omega}}\right)^2 + \hat{\nu}}}\right) \hat{\nu} + 1 \end{aligned}$$

で与えられる。ここで、

$$f_1(z|\nu) := \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}} \left(1 + \frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad F_1(z|\nu) = \int_{-\infty}^z f_1(x|\nu) dx$$

は、それぞれ、自由度 ν のティー分布の確率密度関数と累積分布関数である。

統計モデルを株式時価総額の対数 $\log(\text{market_cap})$ のヒストグラムに重ね書きしたものと標準化された直接母数残差の2乗に関する P-P プロットを、それぞれ、図8と図9に与える²⁰⁾：

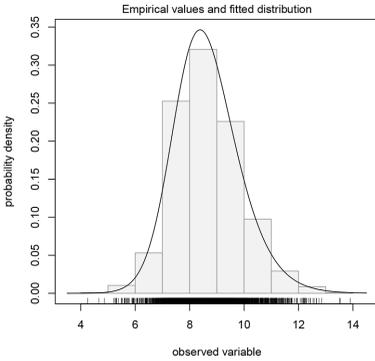


図8：株式時価総額の対数のヒストグラムと非対称ティー分布のもとづく統計モデル

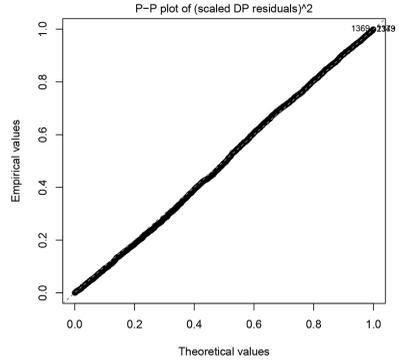


図9：株式時価総額の対数に非対称ティー分布を当てはめたときの標準化直接母数残差の2乗の P-P プロット

図8と図9よりデータはこのモデルにも当てはまっており、対数非対称正規分布の場合（図6，7）と比較しても、可視化した結果だけでは優劣の判別をつけることは難しい。よって、これらのモデル選択に関しては後の節で赤池情報量基準によって行う。

20) ここで描かれている P-P プロットはエフ分布のもとづくものである。

V 株式時価総額の両対数モデリング

この節では、財務データとして純資産 `shareholders` と当期純利益 `net_income` を、ESG レーティングデータ（非財務データ）として、サマリー指標 ESG と環境指標 `E`、社会指標 `s`、ガバナンス指標 `G` を使って、株式時価総額 `market_cap` を説明するためのモデリングを検討する。なお、ここで利用するアイデアも、地道（2017-a, b）、Jimichi *et al.*（2018）の結果を応用しており、モデルの当てはめに関しては R の `sn` パッケージを利用している。

その際、これまで得られた可視化とモデリングの結果を以下のようにモデリングに利用する：

(M1) 対数株式時価総額 $\log(\text{market_cap})$ が非対称分布族に従うことを仮定

（可視化の結果（LD1）と IV 節の当てはめ結果の利用）

(M2) 純資産 `shareholders` と当期純利益 `net_income` の歪みを対数をとることによって解消し、対数株式時価総額 $\log(\text{market_cap})$ と線形関係があることを仮定

（可視化の結果（LD1）、（LU1）の利用）

(M3) 対数株式時価総額 $\log(\text{market_cap})$ と ESG レーティングデータ（`ESG`, `E`, `s`, `G`）のそれぞれのペアの間に線形関係があることを仮定（可視化の結果（LU2）、（LL2））

(M4) ESG 指標 `ESG` と（`E`, `s`, `G`）の各指標の間にある多重共線性を考慮してモデルには `ESG` または（`E`, `s`, `G`）のどちらか一方を説明変数として持つものを検討

（可視化の結果（RU2）、（RL2））

(M5) 国のタイプ `type` による若干の差異をダミー変数として統計モデリング

（可視化の結果（LL1））

以上の指針 (M1)～(M5) を適宜利用しながら、以下に実際にモデリングを行う。その際、指針 (M4) に従って、説明変数として ESG 指標 ESG を利用する場合と、ESG 指標 (E, S, G) を利用する場合に分けて議論する。

1. 両対数モデル：ESG 指標 ESG を利用する場合

説明変数として ESG 指標 ESG を利用し、モデリングの指針 (M1)～(M5) を反映した両対数モデルは以下のようなものである：

$$\begin{aligned} \log(\text{market_cap}_i) = & \alpha_0 + \alpha_1 \log(\text{shareholders}_i) + \alpha_2 \log(\text{net_income}_i) \\ & + \alpha_3 \text{ESG}_i + \beta D(\text{type}_i) + \log(\varepsilon_i) \end{aligned} \quad (1)$$

ここで、

$$D(\text{type}_i) := \begin{cases} 1, & \text{type}_i = \text{G}, \\ 0, & \text{type}_i = \text{E}, \end{cases}$$

はダミー変数であり、G は先進国、E は新興国を表すことを思い出そう。

誤差分布としては以下の仮定を検討する：

$$(N) \log(\varepsilon_i) \stackrel{\text{i.i.d.}}{\sim} \mathbf{N}(0, \sigma^2), \quad i = 1, \dots, n$$

$$(SN) \log(\varepsilon_i) \stackrel{\text{i.i.d.}}{\sim} \mathbf{SN}(0, \omega^2, \alpha), \quad i = 1, \dots, n$$

$$(ST) \log(\varepsilon_i) \stackrel{\text{i.i.d.}}{\sim} \mathbf{ST}(0, \omega^2, \alpha, \nu), \quad i = 1, \dots, n$$

ここで、“ \sim ” は独立に同一の分布に従う (independent and identically distributed: i.i.d.) ことを表す記号であり、仮定 (N), (SN), (ST) はそれぞれ誤差が独立に同一の正規分布, 非対称正規分布, 非対称テーパー分布に従うことを仮定している。

以下に、これらの誤差分布をもつモデルを順次当てはめる。

注意 V.1. 両対数モデル (1) を粗データのスケールでみた場合は以下のようなモデルを仮定していることになっている：

$$\begin{aligned} \text{market_cap}_i = & \gamma \times \text{shareholders}_i^{\alpha_1} \times \text{net_income}_i^{\alpha_2} \\ & \times \exp(\alpha_3 \text{ESG}_i + \beta D(\text{type}_i)) \times \varepsilon_i, \end{aligned}$$

$$(LN) \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \mathbf{LN}(0, \sigma^2), \quad i = 1, \dots, n$$

$$(LSN) \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \mathbf{LSN}(0, \omega^2, \alpha), \quad i = 1, \dots, n$$

$$(LST) \varepsilon_i \overset{i.i.d.}{\sim} LST(0, \omega^2, \alpha, \nu), \quad i=1, \dots, n$$

ここで、仮定 (LN), (LSN), (LST) は、それぞれ、誤差が独立に同一の対数正規分布, 対数非対称正規分布, 対数非対称ティー分布に従うことを仮定している。なお、これらの分布については、付録 B と付録 C を参照されたい。

正規誤差 (N) を仮定した場合

正規誤差を仮定した両対数モデル (1) を当てはめることによって得られる回帰係数のティー検定表 (表 3) の結果から、全ての回帰係数は有意となっている。

表 3 : ティー検定表 : 正規誤差

	Estimate	Std.Error	t value	Pr(> t)
(Intercept)	-0.4539	0.1679	-2.70	0.0069
log(shareholders)	0.1276	0.0172	7.43	0.0000
log(net income)	0.5111	0.0161	31.67	0.0000
ESG	0.1660	0.0143	11.63	0.0000
typeG	0.4612	0.0302	15.26	0.0000

この結果から、標本回帰超平面の方程式は以下のように与えられる：

$$\begin{aligned} \log(\text{market_cap}_i) &= \hat{\alpha}_0 + \hat{\alpha}_1 \log(\text{shareholders}_i) + \hat{\alpha}_2 \log(\text{net_income}_i) \\ &\quad + \hat{\alpha}_3 \text{ESG}_i + \hat{\beta} D(\text{type}_i) \\ &= -0.454 + 0.128 \log(\text{shareholders}_i) \\ &\quad + 0.511 \log(\text{net_income}_i) + 0.166 \text{ESG}_i + 0.461 D(\text{type}_i) \end{aligned} \tag{2}$$

誤差分散の推定値, 決定係数, 自由度調整済み決定係数はそれぞれ以下のように与えられる：

$$\text{誤差分散の推定値} : \hat{\sigma}^2 = 0.702^2$$

$$\text{決定係数} : R^2 = 0.679$$

$$\text{自由度調節済み決定係数} : \bar{R}^2 = 0.679$$

この結果から、決定係数と自由度調節済み決定係数は共に 67.9% である。また、回帰診断に関するプロット (図 10) における残差の正規 Q-Q プロット

を見ると、裾の部分が正規分布に当てはまっていないことがわかる。

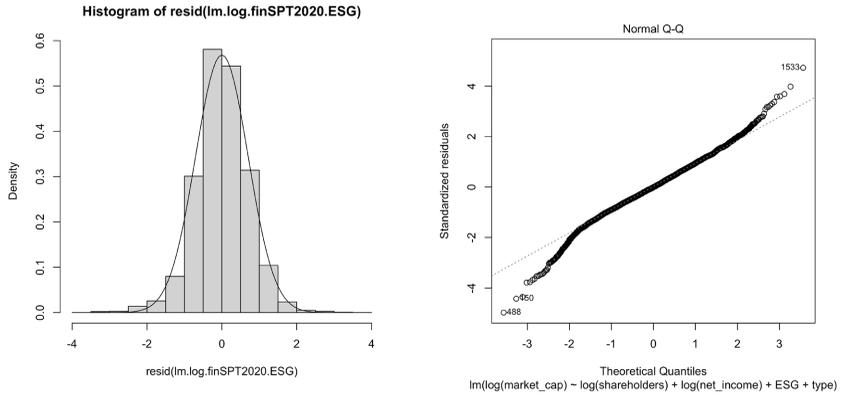


図10：正規誤差を仮定したモデル（1）の当てはめに伴う残差のヒストグラムと正規 Q-Q プロット

非対称正規誤差（SN）を仮定した場合

非対称正規誤差を仮定した両対数モデル（1）を当てはめることによって得られる回帰係数のゼット比検定表（表4）の結果から、切片項 α_0 以外の母数は有意となっている。

表4：ゼット比検定表：非対称正規誤差

	estimate	std.err	z-ratio	Pr {> z }
(Intercept.DP)	-0.0729	0.1646	-0.4427	0.6580
log(shareholders)	0.1394	0.0168	8.3038	0.0000
log(net_income)	0.5069	0.0155	32.7746	0.0000
ESG	0.1609	0.0142	11.3394	0.0000
typeG	0.4408	0.0302	14.5746	0.0000
omega	0.8502	0.0277	30.7284	0.0000
alpha	-1.0056	0.1205	-8.3449	0.0000

この結果から、修正標本回帰超平面の方程式（例えば、地道，2017-b, Jimichi *et al.*, 2018 参照）を求めると、

$$\log(\text{market_cap}_i) = \hat{\alpha}_0 + \hat{\omega} \hat{\delta} + \hat{\alpha}_1 \log(\text{shareholders}_i)$$

$$\begin{aligned}
 & +\hat{\alpha}_2 \log(\text{net_income}_i) + \hat{\alpha}_3 \text{ESG}_i + \hat{\beta} D(\text{type}_i) \\
 = & -0.073 + 0.85 \times 0.798 \times (-0.709) \\
 & + 0.139 \log(\text{shareholders}_i) + 0.507 \log(\text{net_income}_i) \\
 & + 0.161 \text{ESG}_i + 0.441 D(\text{type}_i) \\
 = & -0.554 + 0.139 \log(\text{shareholders}_i) \\
 & + 0.507 \log(\text{net_income}_i) + 0.161 \text{ESG}_i + 0.441 D(\text{type}_i)
 \end{aligned}
 \tag{3}$$

ここで、 $b := \sqrt{2/\pi}$, $\hat{\delta} := \hat{\alpha} / \sqrt{1 + \hat{\alpha}^2}$ である。

また、回帰診断に関するプロット (図11) における残差のヒストグラムと P-P プロットを見ると、当てはまりは良いと思われる。

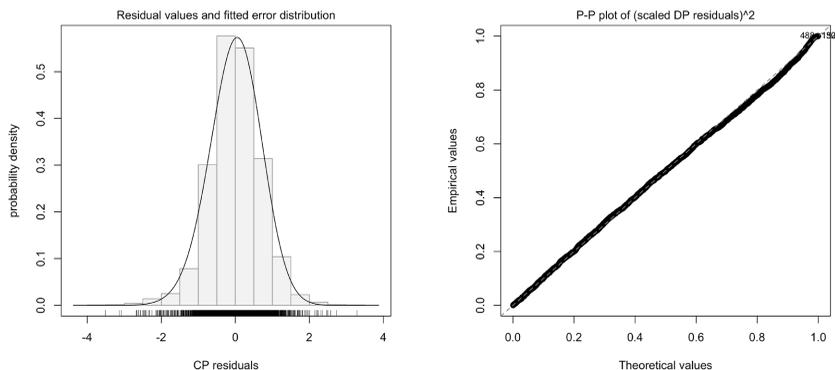


図11：非対称正規誤差を仮定した両対数モデル (1) の当てはめに伴う残差のヒストグラムと P-P プロット

注意 V.2. 表 4 において切片項 α_0 は有意になっていないが、この推定値は、直接母数の推定値になっており、それが有意でないことを表している。なお、回帰 (超) 平面を考えると、中心化母数の推定値 $\hat{\alpha}_0 + \hat{\omega} b \hat{\delta} = -0.554$ を考えることが適切であるので (Azzalini and Capitanio, 2014 の p. 67 参照)、この推定値を利用することとすると、この値は有意であることが別途示される。

非対称ティール誤差 (ST) を仮定した場合

非対称ティー誤差を仮定した両対数モデル (1) を当てはめることによって得られる回帰係数のゼット比検定表 (表 5) の結果から, 非対称母数 α 以外の母数は有意となっている。

表 5 : ゼット比検定表 : 非対称ティー誤差

	estimate	std.err	z-ratio	Pr {> z }
(Intercept.DP)	-0.4163	0.1741	-2.3908	0.0168
log(shareholders)	0.1069	0.0176	6.0863	0.0000
log(net_income)	0.5488	0.0173	31.7325	0.0000
ESG	0.1490	0.0136	10.9782	0.0000
typeG	0.3879	0.0292	13.2770	0.0000
omega	0.6018	0.0204	29.4416	0.0000
alpha	-0.2335	0.1676	-1.3937	0.1634
nu	6.8960	0.8565	8.0514	0.0000

この結果から, 修正標本回帰超平面の方程式 (例えば, 地道, 2017-b, Jimichi *et al.*, 2018 参照) を求めると,

$$\begin{aligned}
 \log(\text{market_cap}_i) &= \hat{\alpha}_0 + \hat{\omega} b_i \hat{\delta} + \hat{\alpha}_1 \log(\text{shareholders}_i) \\
 &\quad + \hat{\alpha}_2 \log(\text{net_income}_i) + \hat{\alpha}_3 \text{ESG}_i + \hat{\beta} D(\text{type}_i) \\
 &= -0.416 + 0.602 \times 0.9 \times (-0.227) \\
 &\quad + 0.107 \log(\text{shareholders}_i) + 0.549 \log(\text{net_income}_i) \\
 &\quad + 0.149 \text{ESG}_i + 0.388 D(\text{type}_i) \\
 &= -0.54 + 0.107 \log(\text{shareholders}_i) \\
 &\quad + 0.549 \log(\text{net_income}_i) + 0.149 \text{ESG}_i + 0.388 D(\text{type}_i)
 \end{aligned} \tag{4}$$

ここで, $b_i := \sqrt{\hat{\nu}} \Gamma((\hat{\nu}-1)/2) / (\sqrt{\pi} \Gamma(\hat{\nu}/2))$, $\hat{\delta} = \hat{\alpha} / \sqrt{1 + \hat{\alpha}^2}$ である。

また, 回帰診断に関するプロット (図12) における残差のヒストグラムと P-P プロットを見ると, 当てはまりは良いと思われる。

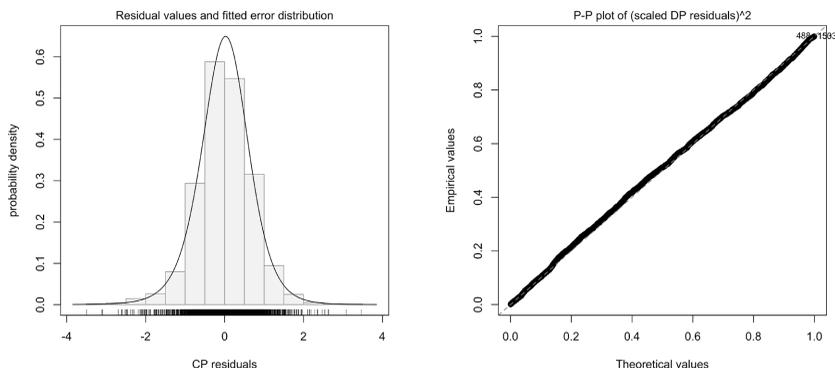


図12：非対称ティー誤差を仮定した両対数モデル（1）の当てはめに伴う残差のヒストグラムと P-P プロット

両対数モデル（1）の当てはめに関する考察

この小節では、両対数モデル（1）の誤差に関して3種類の分布を仮定し、当てはめた。それぞれの残差の可視化の結果から、正規誤差（SN）の場合は裾の部分で当てはまらな問題があることがわかり、非対称正規誤差（SN）と非対称ティー誤差（ST）の場合は、どちらが良いかは可視化結果だけでは、判断が付きかねるので、後節で赤池情報量規準によってモデル選択の観点から議論することにする。

2. 両対数モデル：ESG 指標（E, S, G）を利用した場合

説明変数として ESG 指標（E, S, G）を利用し、モデリングの指針（M1）～（M5）を反映した両対数モデルは以下のようなものである：

$$\log(\text{market_cap}_i) = \alpha_0 + \alpha_1 \log(\text{shareholders}_i) + \alpha_2 \log(\text{net_income}_i) + \alpha_3 E_i + \alpha_4 S_i + \alpha_5 G_i + \beta D(\text{type}_i) + \log(\varepsilon_i), \quad (5)$$

両対数モデル（1）と同様に誤差分布としては以下の仮定を検討する：

$$\begin{aligned} \text{(N)} \log(\varepsilon_i) &\sim \text{N}(0, \sigma^2), \quad i=1, \dots, n \\ \text{(SN)} \log(\varepsilon_i) &\sim \text{SN}(0, \omega^2, \alpha), \quad i=1, \dots, n \\ \text{(ST)} \log(\varepsilon_i) &\sim \text{ST}(0, \omega^2, \alpha, \nu), \quad i=1, \dots, n \end{aligned}$$

以下に、これらの誤差分布をもつモデルを順次当てはめる。

注意 V.3. 両対数モデル (5) を粗データのスケールでみた場合は以下のようなモデルを仮定していることになっている：

$$\text{market_cap}_i = \gamma \times \text{shareholders}_i^{\alpha_1} \times \text{net_income}_i^{\alpha_2} \\ \times \exp(\alpha_3 E_i + \alpha_4 S_i + \alpha_5 G_i + \beta D(\text{type}_i)) \times \varepsilon_i,$$

$$(\text{LN}) \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{LN}(0, \sigma^2), \quad i=1, \dots, n$$

$$(\text{LSN}) \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{LSN}(0, \omega^2, \alpha), \quad i=1, \dots, n$$

$$(\text{LST}) \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{LST}(0, \omega^2, \alpha, \nu), \quad i=1, \dots, n$$

正規誤差 (N) を仮定した場合

正規誤差を仮定した両対数モデル (5) を当てはめることによって得られる回帰係数のティー検定表 (表 6) の結果から、全ての回帰係数は有意となっている。

表 6 : ティー検定表 : 正規誤差

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.6171	0.1784	-3.4598	0.0005
log(shareholders)	0.1404	0.0173	8.1154	0.0000
log(net_income)	0.5078	0.0161	31.6262	0.0000
E	-0.0546	0.0178	-3.0625	0.0022
S	0.1715	0.0213	8.0463	0.0000
G	0.0453	0.0181	2.5081	0.0122
typeG	0.4727	0.0301	15.7215	0.0000

この結果から、標本回帰超平面の方程式は以下のように与えられる：

$$\begin{aligned} \log(\text{market_cap}_i) &= \hat{\alpha}_0 + \hat{\alpha}_1 \log(\text{shareholders}_i) + \hat{\alpha}_2 \log(\text{net_income}_i) \\ &\quad + \hat{\alpha}_3 E_i + \hat{\alpha}_4 S_i + \hat{\alpha}_5 G_i + \hat{\beta} D(\text{type}_i) \\ &= -0.617 + 0.14 \log(\text{shareholders}_i) \\ &\quad + 0.508 \log(\text{net_income}_i) - 0.055 E_i + 0.171 S_i \\ &\quad + 0.045 G_i + 0.473 D(\text{type}_i) \end{aligned} \quad (6)$$

誤差分散の推定値、決定係数、自由度調整済み決定係数はそれぞれ以下の

ように与えられる：

誤差分散の推定値： $\hat{\sigma}^2=0.698^2$

決定係数： $R^2=0.683$

自由度調節済み決定係数： $\bar{R}^2=0.683$

この結果から、決定係数と自由度調節済み決定係数は共に68.3%である。また、回帰診断に関するプロット（図13）における残差の正規 Q-Q プロットを見ると、裾の部分が正規分布に当てはまっていないことがわかる。

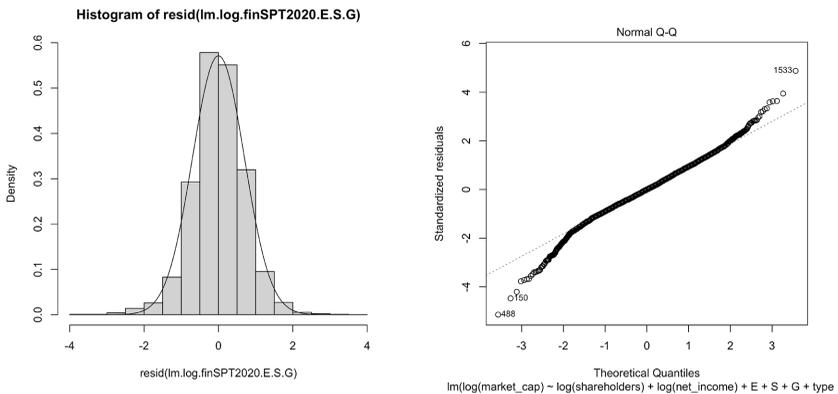


図13：正規誤差を仮定したモデル（5）の当てはめに伴う残差のヒストグラムと正規 Q-Q プロット

非対称正規誤差（SN）を仮定した場合

非対称正規誤差を仮定した両対数モデル（5）を当てはめることによって得られる回帰係数のゼット比検定表（表7）の結果から、切片項 α_0 以外の母数は有意となっている。

表7：ゼット比検定表：非対称正規誤差

	estimate	std.err	z-ratio	Pr {> z }
(Intercept.DP)	-0.2525	0.1748	-1.4444	0.1486
log(shareholders)	0.1523	0.0169	9.0191	0.0000
log(net_income)	0.5042	0.0154	32.8237	0.0000
E	-0.0553	0.0175	-3.1505	0.0016
S	0.1658	0.0211	7.8730	0.0000
G	0.0476	0.0178	2.6717	0.0075
typeG	0.4520	0.0301	15.0230	0.0000
omega	0.8456	0.0274	30.8904	0.0000
alpha	-1.0086	0.1199	-8.4144	0.0000

この結果から、修正標本回帰超平面の方程式を求めると、

$$\begin{aligned}
 \log(\text{market_cap}_i) &= \hat{\alpha}_0 + \hat{\omega} b \hat{\delta} + \hat{\alpha}_1 \log(\text{shareholders}_i) \\
 &\quad + \hat{\alpha}_2 \log(\text{net_income}_i) + \hat{\alpha}_3 E_i + \hat{\alpha}_4 S_i + \hat{\alpha}_5 G_i + \hat{\beta} D(\text{type}_i) \\
 &= -0.253 + 0.846 \times 0.798 \times (-0.71) \\
 &\quad + 0.152 \log(\text{shareholders}_i) + 0.504 \log(\text{net_income}_i) \\
 &\quad - 0.055 E_i + 0.166 S_i + 0.048 G_i + 0.452 D(\text{type}_i) \\
 &= -0.732 + 0.152 \log(\text{shareholders}_i) \\
 &\quad + 0.504 \log(\text{net_income}_i) - 0.055 E_i + 0.166 S_i + 0.048 G_i \\
 &\quad + 0.452 D(\text{type}_i) \tag{7}
 \end{aligned}$$

ここで、 $b = \sqrt{2/\pi}$, $\hat{\delta} = \hat{\alpha} / \sqrt{1 + \hat{\alpha}^2}$ であることを思い出そう。

また、回帰診断に関するプロット (図14) における残差のヒストグラムと P-P プロットを見ると、当てはまりはある程度良いと思われるが、ヒストグラムにおける原点付近での統計モデルとの間に隔たりがある。

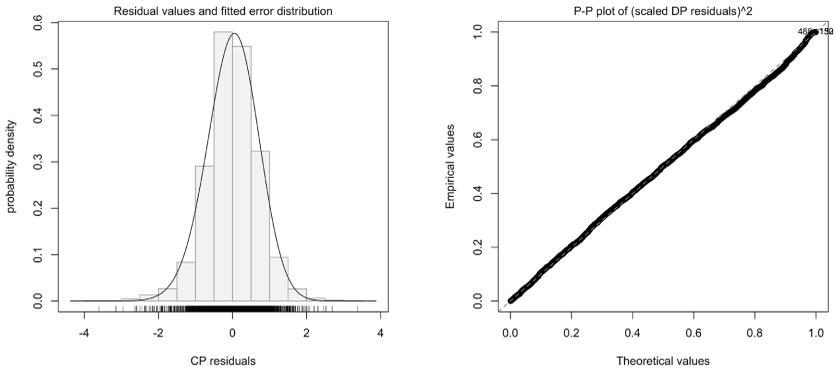


図14：非対称正規誤差を仮定した両対数モデル (5) の当てはめに伴う残差のヒストグラムと P-P プロット

注意 V.4. 注意 V.2 でも述べたが、表 7 において切片項 α_0 は有意になっていないが、この推定値は、直接母数の推定値になっており、それが有意でないということを表している。なお、回帰（超）平面を考えると、中心化母数の推定値 $\hat{\alpha}_0 + \hat{\omega}b\hat{\delta} = -0.732$ を考えることが適切であるので、この推定値を利用することとすると、この値は有意であることが別途示される（注意 V.2 と同様に、Azzalini and Capitanio, 2014 の p. 67 参照）。

非対称ティール誤差（ST）を仮定した場合

非対称ティール誤差を仮定した両対数モデル (5) を当てはめることによって得られる回帰係数のゼット比検定表（表 8）の結果から、非対称母数 α 以外の母数は有意となっている。

表 8 : ゼット比検定表 : 非対称テューン誤差

	estimate	std.err	z-ratio	Pr {> z }
(Intercept.DP)	-0.6002	0.1832	-3.2760	0.0011
log(shareholders)	0.1239	0.0178	6.9716	0.0000
log(net_income)	0.5423	0.0172	31.5455	0.0000
E	-0.0611	0.0167	-3.6552	0.0003
S	0.1640	0.0201	8.1764	0.0000
G	0.0413	0.0170	2.4230	0.0154
typeG	0.3979	0.0290	13.7371	0.0000
omega	0.5957	0.0201	29.5845	0.0000
alpha	-0.2339	0.1649	-1.4181	0.1562
nu	6.7519	0.8256	8.1786	0.0000

この結果から、修正標本回帰超平面の方程式を求めると、

$$\begin{aligned}
 \log(\text{market_cap}_i) &= \hat{a}_0 + \hat{\omega} b_{\hat{\nu}} \hat{\delta} + \hat{a}_1 \log(\text{shareholders}_i) + \hat{a}_2 \log(\text{net_income}_i) \\
 &\quad + \hat{a}_3 E_i + \hat{a}_4 S_i + \hat{a}_5 G_i + \hat{\beta} D(\text{type}_i) \\
 &= -0.6 + 0.596 \times 0.903 \times (-0.228) \\
 &\quad + 0.124 \log(\text{shareholders}_i) + 0.542 \log(\text{net_income}_i) \\
 &\quad - 0.061 E_i + 0.164 S_i + 0.041 G_i + 0.398 D(\text{type}_i) \\
 &= -0.613 + 0.124 \log(\text{shareholders}_i) \\
 &\quad + 0.542 \log(\text{net_income}_i) - 0.061 E_i + 0.164 S_i + 0.041 G_i \\
 &\quad + 0.398 D(\text{type}_i) \tag{8}
 \end{aligned}$$

ここで、 $b_{\hat{\nu}} = \sqrt{\hat{\nu}} \Gamma((\hat{\nu}-1)/2) / (\sqrt{\pi} \Gamma(\hat{\nu}/2))$, $\hat{\delta} = \hat{\alpha} / \sqrt{1 + \hat{\alpha}^2}$ であることを思い出そう。

また、回帰診断に関するプロット (図15) における残差のヒストグラムと P-P プロットを見ると、当てはまりは良いと思われる。

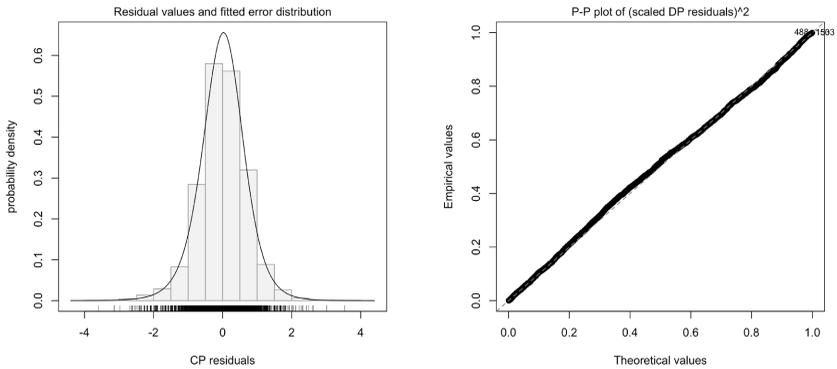


図15：非対称ティー誤差を仮定した両対数モデル (5) の当てはめに伴う残差のヒストグラムと P-P プロット

両対数モデル (5) の当てはめに関する考察

この小節では、両対数モデル (5) の誤差に関して 3 種類の分布を仮定し、当てはめた。それぞれの残差の可視化の結果から、正規誤差 (SN) の場合は裾の部分で当てはまらな問題があることがわかり、非対称正規誤差 (SN) と非対称ティー誤差 (ST) の場合は、どちらが良いかは可視化結果だけからは、判断が付きかねるので、後節で赤池情報量規準によってモデル選択の観点から議論することにする。

VI 赤池情報量規準によるモデル選択

ここでは、まず対数株式時価総額に非対称正規分布と非対称ティー分布を当てはめた結果を赤池情報量規準を使って比較・検討する。次に、2 種類の両対数モデルに対する 3 種類の誤差分布 (正規分布, 対称正規分布, 非対称ティー分布) を仮定したものから選択するために赤池情報量規準を利用する。なお、一般に赤池情報量規準の値が小さい方が当てはまりが良いと考えられる。

1. 対数株式時価総額の分布に関するモデル選択

対数株式時価総額に、非対称正規分布 (selm.log.mc2018.SN), 非対称ティー分布 (selm.log.mc2018.ST) を当てはめたときの, それぞれのモデルに対する母数の次元 (dim) と AIC の値を表 9 に与える.

表 9 : AIC 表 : 対数株式時価総額の分布に関する比較

	dim	AIC
selm.log.mc2018.SN	3	9306.21
selm.log.mc2018.ST	4	9299.10

対数株式時価総額に非対称正規分布と非対称ティー分布を当てはめた可視化の結果 (非対称正規分布の場合 : 図 6, 7, 非対称ティー分布の場合 : 図 8, 9) と同様に, 表 9 で与えられる AIC の値からは, ほぼ同様の結果となったが若干非対称ティー分布が良いことがわかった.

2. 株式時価総額の両対数モデルに関する選択

対数株式時価総額に正規誤差 (lm.log.finSPT2020.ESG), 非対称正規誤差 (selm.log.finSPT2020.ESG.SN), 非対称ティー誤差 (selm.log.finSPT2020.ESG.ST) をそれぞれもつ両対数モデル (ESG 指標として ESG を利用) を当てはめたときの, それぞれのモデルに対する母数の次元と AIC の値を表 10 に与える.

表 10 : AIC 表 : 両対数モデル (ESG 指標 : ESG) に関する比較

	dim	AIC
lm.log.finSPT2020.ESG	6	5830.30
selm.log.finSPT2020.ESG.SN	7	5817.00
selm.log.finSPT2020.ESG.ST	8	5726.53

この結果から, 非対称ティー誤差をもつ両対数モデル (selm.log.finSPT2020.ESG.ST) が最も良いことがわかった.

対数株式時価総額に正規誤差 (lm.log.finSPT2020.E.S.G), 非対称正規誤差

(selm.log.finSPT2020.E.S.G.SN), 非対称ティー誤差 (selm.log.finSPT2020.E.S.G.ST) をそれぞれもつ両対数モデル (ESG 指標としてピラーの E, S, G を利用) を当てはめたときの, それぞれのモデルに対する母数の次元と AIC の値を表11に与える.

表11: AIC 表: 両対数モデル (ESG 指標: E, S, G) に関する比較

	dim	AIC
lm.log.finSPT2020.E.S.G	8	5800.70
selm.log.finSPT2020.E.S.G.SN	9	5787.14
selm.log.finSPT2020.E.S.G.ST	10	5692.65

この結果からも, 非対称ティー誤差をもつ両対数モデル (selm.log.finSPT2020.E.S.G.ST) が最も良いことがわかった.

さらに, 説明変数として ESG 指標 ESG を持つ場合と, (E, S, G) を持つ場合の両対数モデルの比較としては, 後者の非対称ティー誤差をもつ両対数モデル (selm.log.finSPT2020.E.S.G.ST) がこれらのモデルの中で最も良いことがわかる. ただし, これらのモデルは, 対数株式時価総額への ESG 指標の寄与がどのようなものかを知るという観点から考えると, 利用上の用途が異なっているため, 単純に比較することは注意が必要であろう.

VII 経年変化を考慮した対数株式時価総額の分布と両対数モデルの選択

ここまでは, ESG 指標の公表年が2020年 (対応する会計年度は2018年) のデータについての結果を詳細に考察してきたが, ESG 指標の公表年が2015年から2020年までの, それぞれの時点でモデルを当てはめた結果を, 赤池情報量規準の経年変化をプロットすることによって考察する. その際, 各年毎に解析対象となる企業が異なっていることからデータ数も異なっており (表1参照), 単純に赤池情報量規準を比較することが難しいため, ベンチマークとして赤池情報量規準をデータ数で割った値 (AIC/n) (1社あたりの AIC 値) も同時にプロットする.

1. 経年変化を考慮した株式時価総額の分布に関するモデル選択

株式時価総額の分布に正規分布と非対称正規分布、非対称ティー分布を各年毎に当てはめ、経年変化をみる。なお、本稿では、これまで株式時価総額の分布に正規分布を当てはめてこなかったが（IV節）、ここではベンチマークのため含めることにする。図16は、2015～2020年（ESG指標の公表年）のそれぞれの時点における企業数 n （上段）と3種類の分布を当てはめたときのAIC値 AIC （中段）、1社あたりのAIC値 AIC/n （下段）の経年変化のプロットである。なお、プロットされている系列の記号は、それぞれ、1m（正規分布）、1mSN（非対称正規分布）、1mST（非対称ティー分布）を表す。図16から読み取れる結果を以下に与える：

- (DT1) 3種類の分布を当てはめるために利用したデータ数（企業数 n ）は各年毎に同一であるため、企業数のプロット（上段）は一本の系列がプロットされている。なお、企業数は増加傾向にある。これは、新興国の企業数が増加しているためである（表1も参照）。
- (DT2) AIC値（中段）は、その構造上、データ数の増加に伴って大きくなることが予想される²¹⁾。また、各年毎に観測された企業が異なっているために、経年変化を単純に比較することはできない。よって、各年毎にモデルの当てはまりをみる必要がある。その結果として、各年度とも、若干であるが、非対称ティー分布（1mST）の当てはまりが良いことがわかる。
- (DT3) 1社あたりのAIC値（下段）をみると、各年毎の当てはまりは、正規分布（1m）よりも、非対称正規分布（1mSN）と非対称ティー分布（1mST）が良いことがわかる。この結果からも、本稿で考察対象とした、非対称分布族を対数株式時価総額に当てはめることの重要性がわかる。さらに、若干であるが、各年度とも非対称ティー分布

21) 例えば、データ $\{x_1, \dots, x_n\}$ が正規分布 $N(\mu, \sigma^2)$ に従って生成されているとき、 $AIC = n \log(2\pi\sigma^2) + n + 4$ で与えられ、 σ^2 （：データの分散）の挙動にも依存するが、データ数 n の増加に伴ってAICの値も増加する傾向にある（付録Bも参照）。

($lmST$) が良いことがわかる。なお、1社あたりの AIC 値は、ある意味での「基準化」がされているので、モデルの当てはまりの経年変化をみることができる。すなわち、経時観測された個々の企業の追跡結果ではないが、モデルの当てはまりの傾向としては、2018年度を境に当てはまりが悪くなってきていることがうかがえる。このことは、新興国の企業数の増加と関係しており、新興国における企業の対数株式時価総額の分布構造が先進国のものと若干異なっているためと考えられる。

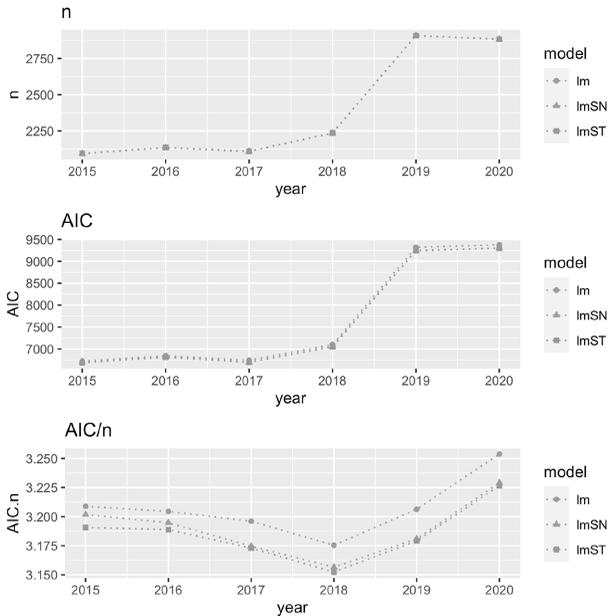


図16：2015～2020年（ESG指標の公表年）のそれぞれの時点における企業数 n （上段）と3種類の分布を当てはめたときの AIC 値 AIC （中段）、1社あたりの AIC 値 AIC/n （下段）の経年変化のプロット： lm （正規分布）、 $lmSN$ （非対称正規分布）、 $lmST$ （非対称ティータ分布）

2. 経年変化を考慮した株式時価総額の両対数モデルに関するモデル選択

まず、株式時価総額に正規誤差と非対称正規誤差、非対称ティー誤差を持つ両対数モデル (1) を各年毎に当てはめ、経年変化をみる。図17は、2015年から2020年 (ESG 指標の公表年) までのそれぞれの時点における企業数 n (上段) と 3 種類の誤差分布をもつ両対数モデル (1) を当てはめたときの AIC 値 AIC (中段), 1 社あたりの AIC 値 AIC/n (下段) の経年変化のプロットである。なお、プロットされている系列の記号は、それぞれ、 lm (正規誤差), $lmSN$ (非対称正規誤差), $lmST$ (非対称ティー誤差) をもつ両対数モデルを表す。

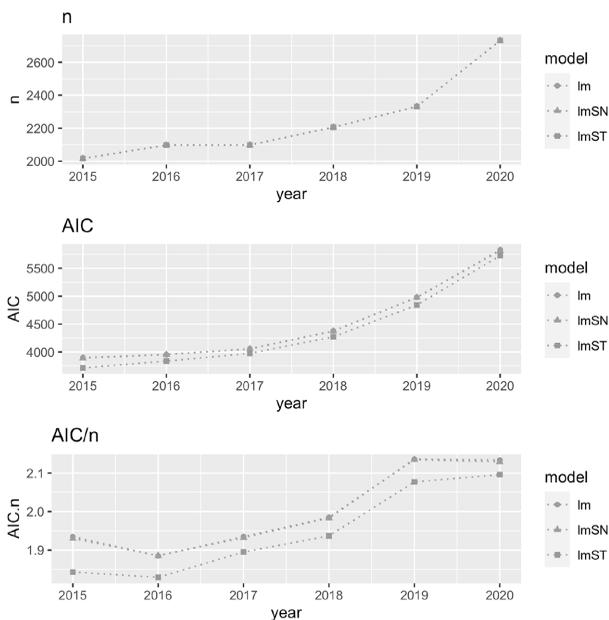


図17：2015～2020年 (ESG 指標の公表年) のそれぞれの時点における企業数 n (上段) と説明変数に ESG 指標 ESG を持つ両対数モデル (1) を当てはめたときの AIC 値 AIC (中段), 1 社あたりの AIC 値 AIC/n (下段) の経年変化のプロット： lm (正規誤差), $lmSN$ (非対称正規誤差), $lmST$ (非対称ティー誤差)

図17から読み取れる結果を以下に与える：

- (LT1) 3種類の誤差分布を持つ両対数モデル (1) を当てはめるために利用したデータ数 (企業数 n) は各年毎に同一であるため、企業数のプロット (上段) は一本の系列がプロットされている。ここで、注意が必要なことは、表1で与えた企業数と当てはめに利用されたデータ数が欠測値の関係で一致しないことである。つまり、新興国の企業数が増加しているように見えるが、ESG 指標が納められたデータはそれよりも少ないことに注意が必要である。なお、実際にモデルの当てはめに利用できる年度別の企業数については表2を参照されたい。
- (LT2) AIC 値 (中段) は、対数株式時価総額に分布を当てはめた場合の考察 (DT2) と同じ理由で、経年変化を単純に比較することはできないため、各年毎にモデルの当てはまりをみる必要がある。その結果として、各年度とも、正規誤差と非対称正規誤差を持つ両対数モデル ($1m$, $1mSN$) はほとんど AIC 値が近く、若干であるが、非対称ティー誤差を持つ両対数モデル ($1mST$) の AIC 値が小さいという意味で、当てはまりが良いことがわかる。
- (LT3) 1社あたりの AIC 値 (下段) をみると、各年毎の非対称ティー誤差をもつ両対数モデル ($1mST$) の当てはまりが良いことが顕著になっている。この結果からも、本稿で考察対象とした、非対称分布族を対数株式時価総額に当てはめることの重要性がわかる。

次に、株式時価総額に正規誤差と非対称正規誤差、非対称ティー誤差を持つ両対数モデル (5) を各年毎に当てはめ、経年変化をみる。両対数モデル (1) との相違点は、モデルに含める説明変数として、ESG 指標がサマリーの ESG からピラーの (E, S, G) に変更されたという点だけである。よって、図18は、図17と同様に解釈できる。

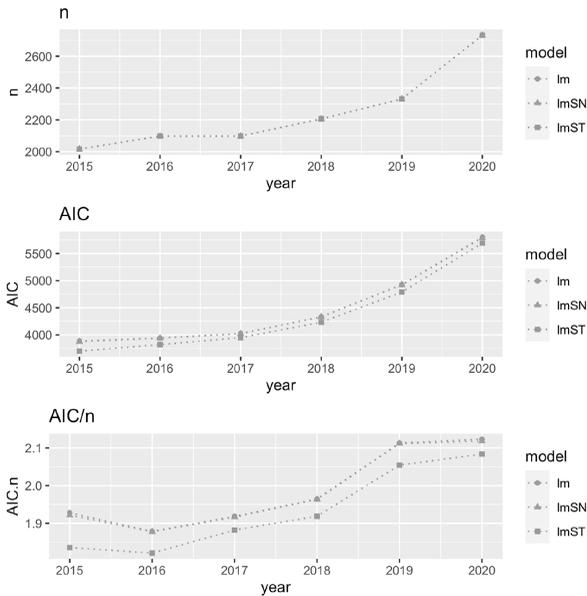


図18：2015～2020年（ESG指標の公表年）のそれぞれの時点における企業数 n （上段）と説明変数に ESG 指標（ e, s, g ）を持つ両対数モデル（1）を当てはめたときの AIC 値 AIC （中段），1社あたりの AIC 値 AIC/n （下段）の経年変化のプロット：lm（正規誤差），lmSN（非対称正規誤差），lmST（非対称ティール誤差）

図18から読み取れる結果を以下に与える：

- (L2T1) 両対数モデル（1）を当てはめるために利用したデータ数（企業数 n ）と同一であるため，考察（LT1）と同様の考察が可能である。
- (L2T2) 両対数モデル（1）を当てはめた場合の考察（LT2）と同様の考察が可能である。つまり，各年度とも，正規誤差と非対称正規誤差を持つ両対数モデル（lm, lmSN）はほとんど AIC 値が近く，若干であるが，非対称ティール誤差を持つ両対数モデル（lmST）の AIC 値が小さいという意味で，当てはまりが良いことがわかる。なお，プロットの結果（図17と図18）からはわかりにくいですが，全ての場合において両対数モデル（1）を当てはめた場合よりも，両対数モデル（5）

の方が AIC 値が小さい。

(L2T3) 両対数モデル (1) を当てはめた場合の考察 (LT3) と同様の考察が可能である。つまり、1社あたりの AIC 値は、各年毎の非対称ティー誤差をもつ両対数モデル (1mST) の当てはまりが良いことが顕著になっている。この結果からも、本稿で考察対象とした、非対称分布族を対数株式時価総額に当てはめることの重要性がわかる。

VIII その他のデータ解析結果

ここでは、実際に行った他のデータ解析の結果の一部を以下に補足として与える：

- 先進国と新興国の企業を分けて、本稿で検討したモデルを当てはめ、種々検討してみたところ、総じて先進国へのモデルの当てはまりは良いが、新興国へのモデルの当てはまりが悪いという結果が得られた。
- ESG 指標の公表年が2020年 (対応する会計年度が2018年) のデータに対して、ESG 指標として、ESG と (E, S, G) を説明変数として全て含むモデルを当てはめてみた結果として、ESG に対する回帰係数の推定値が正であるのに対して、それ以外の指標の回帰係数が全て負となる結果が得られた。これは、ESG と (E, S, G) の間に存在する多重共線性が影響しているものと思われる (可視化の結果 (RL2) も参照)。簡単な考察として、ESG 指標の公表年が2020年 (会計年度が2018年) のデータに対して、多重共線性を検出するための指標の一つである分散膨張因子 (Variance Inflation Factor: VIF) (例えば、Jimichi, 2016 参照) を計算してみたところ、ESG の VIF 値は213.5となった。通常、VIF 値は10より高いと多重共線性の影響を受けるといわれているので、特に ESG に対する値は高いものと思われる。
- 非対称ティー誤差を仮定した両対数モデル (1) の当てはめ結果として得られる母数に対するゼット比検定表 5 では、非対称母数 α が有意となっていないため、この結果を受け入れるならば誤差分布は「ティー分

布」であると考えることができる。ただし、本稿で考えている2015年から2020年の経年で全てこの結果が成り立つかどうかは、この期間の経年変化をみる必要がある。図19は、非対称母数 α に関する（最尤）推定値（上段）と標準誤差（Standard Error: SE）（中段），検定統計量（Test Statistics: TS）（下段）の推移をプロットしたものである。ここで、検定統計量のプロットに関しては棄却限界を表す破線を追記している²²⁾。この結果から、年度によっては、非対称ティー分布を考える必要があることを示している²³⁾。

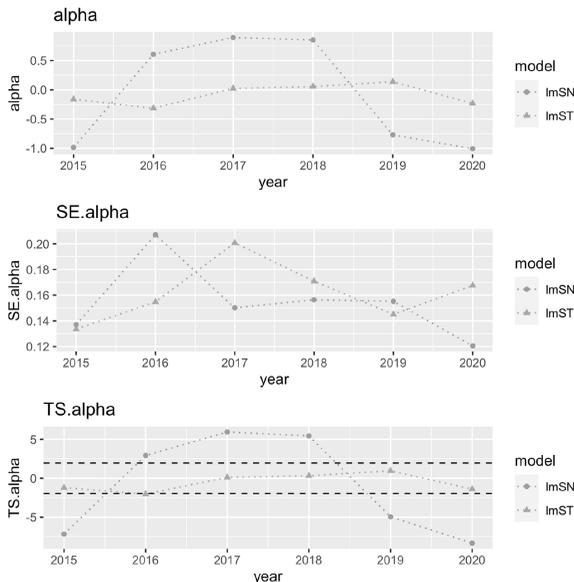


図19：2015～2020年（ESG 指標の公表年）のそれぞれの時点で両対数モデル（1）を当てはめたときの非対称母数 α に関する（最尤）推定値（上段）と標準誤差（中段），検定統計量（下段）の経年変化のプロット：lmSN（非対称正規誤差），lmST（非対称ティー誤差）

22) 最尤推定量の漸近正規性と一致性を利用して、ゼット比統計量が帰無仮説 $H_0: \alpha=0$ のもとで標準正規分布に従うことから、標準正規分布の両側5%点（ ± 1.96 ）が有意水準5%の両側検定で有意となる棄却限界となる。

23) 図から判断することは難しいが、2016年のゼット比の値は、 -2.04 となり、5%有意となる。

- 両対数モデル (1) を当てはめた場合と同様に、非対称ティー誤差を仮定した両対数モデル (5) のゼット比検定表 8 は、非対称母数 α が有意となっていないため、この結果を受け入れるならば誤差分布は「ティー分布」であると考えられることができる。ただし、図19と同様に、2015年から2020年の経年変化をみた図20から、2016年は $\alpha \neq 0$ が検定の結果として得られて、非対称ティー分布を考える必要があることがわかる。

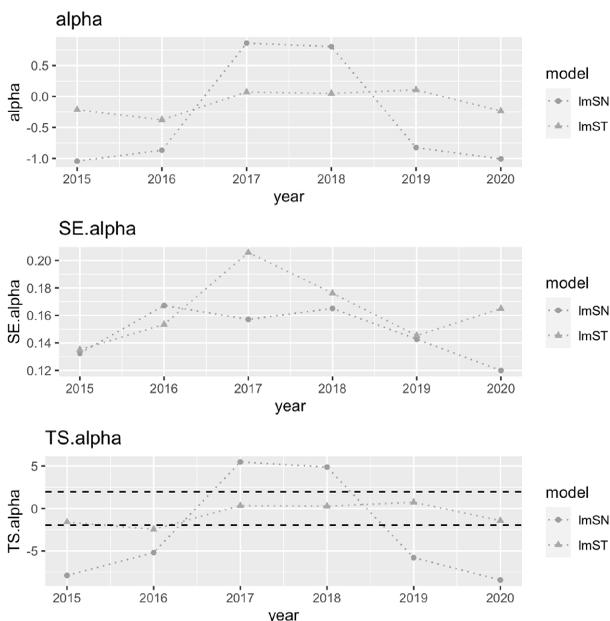


図20：2015～2020年（ESG 指標の公表年）のそれぞれの時点で両対数モデル (5) を当てはめたときの非対称母数 α に関する（最尤）推定値（上段）と標準誤差（中段）、検定統計量（下段）の経年変化のプロット：lmSN（非対称正規誤差）、lmST（非対称ティー誤差）

IX おわりに

本稿では、非対称分布族を考慮した分布や両対数モデルを対数株式時価総額に当てはめることに関して議論した。結果としては、2015年から2020年の

期間（ESG 指標の公表年）にわたって、対数株式時価総額の分布としては、若干非対称ティール分布が当てはまることがわかった。また、同じ期間に関して両対数モデルに関しては非対称ティール誤差を仮定した場合が当てまることがわかった。

最後に、今後のデータ解析・統計モデリングの視点として重要と思われる事項について述べる。

- (P1) 通常の重回帰分析では、モデル評価基準として決定係数や自由度調整済み決定係数が利用されるが²⁴⁾、本稿で利用した正規誤差の場合の両対数モデルを当てはめた場合の決定係数は67.9~68.3%である。決定係数はあくまでも相対的な指標であるため、単純に高低を述べるのが難しいが、このモデルの改良を（例えば説明変数を増加させる方法で）行う余地はあると思われる。ただし、本稿で行った誤差分布を非対称分布族に拡張することによって赤池情報量規準では正規誤差の場合よりも改良されていることは注目に値する²⁵⁾。
- (P2) 本稿では、財務データの対象年度と、それに対応させる ESG レーティングデータの公表年に2年のラグを設定した。これは、ESG レーティングの評価の基礎として用いられる企業の ESG 情報の開示・調査・評価から、レーティングの公表までに、企業 ESG 情報の対象年度から約2年の時間を要するためであるが、相互相関および効率的市場仮説の観点から他のラグを検討することも考えられる。
- (P3) III 節で与えた対数株式時価総額のヒストグラムと、IV 節で考察した結果から、データは非対称分布、特に、非対称ティール分布に当てはまっていることがわかったが、対散布図5において、対数株式時価総額は先進国と新興国（国のタイプ）に関して、それぞれ、非対称分布に従っていることが見て取れる。このことは、対数株式時価総額を国のタイ

24) 通常の重回帰分析で決定係数が利用される理由は、最小自乗法の本質である正射影によって導かれる「平方和の分解」にもとづいて決定係数が構成されることによる。

25) 誤差分布が非対称分布族にしたがう場合は、推測に最尤法を用いることから、モデル評価基準として決定係数を用いることは理にかなわない。

ぶに分類（分割）したものが、それぞれ非対称分布に従うことを示唆するように思われる。また、逆に、それぞれ非対称分布に従う独立な変量（今の場合は対数株式時価総額）を合併したものが、再び非対称分布に従うことが予想されるかもしれない。これらのことは、理論的には確率変数（確率分布）に関する無限分解可能性（infinitely divisible）や再生性（reproductive property）などと関係が深い話題であるが、非対称正規分布が無限分解可能でないことが Kozubowska and Nolanb (2008) の Proposition 2.1 に与えられており、再生性を保たないことが Azzalini and Capitanio (2014) の p. 27 で示されている。よって、この現象をこのような観点から理論的に説明することは難しい。

これらの問題について引き続き検討する予定である。なお、問題（P3）についての一つのアプローチとしては、二つの非対称分布の混合分布の観点から検討することも可能であると思われる。

（筆者（地道）は関西学院大学商学部教授）

（筆者（阪）は関西学院大学商学部教授）

参考文献

- [1] Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle, *Proceedings of the 2nd International Symposium on Information Theory*, Petrov, B. N., and Csaki, F. (eds.), Akadimiai Kiado, Budapest: pp. 267-281.
- [2] Azzalini, A. (1985) A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, Vol. 12, No. 2, pp. 171-178.
- [3] Azzalini, A. with the collaboration of A. Capitanio (2014) *The Skew-Normal and Related Families*, Cambridge University Press, Institute of Mathematical Statistics Monographs.
- [4] Fox, J. and S. Weisberg (2019) *An R Companion to Applied Regression, Third Edition*, Sage.
- [5] FTSE Russell (2019) *ESG Data Model 6th Research Cycle, Methodology (2019/20)*, <https://qsd.ftserussell.com/Docs/ESG/FTSE%20Russell%20ESG%20Data%20Model%20Methodology%20-%20April-2019-March-2020-Rev2.pdf>
- [6] 稲垣宣生 (2003) 『数理統計学－改訂版－』, 裳華房.
- [7] Janssens, J. (2014) *Data Science at the Command Line*, O'Reilly Media.
(太田満久, 下田倫大, 増田泰彦監訳, 長尾高弘訳 (2015) 『コマンドラインでは

- じめるデータサイエンス：分析プロセスを自在に進めるテクニック』, オライリー・ジャパン.)
- [8] Jimichi, M. (2016) *Shrinkage Regression Estimators and Their Feasibilities*, Kwansai Gakuin University Press.
- [9] 地道正行 (2017-a) 『Rによる対数非対称正規線形モデルによる財務データの統計モデリング』, 商学論究, 第64巻, 第5号, pp. 159-185, 関西学院大学商学研究会.
- [10] 地道正行 (2017-b) 『Rを利用した非対称分布族にもとづく財務データの統計モデリング』, 経済学論究, 第71巻, 第2号, pp. 141-174, 関西学院大学経済学部研究会.
- [11] 地道正行 (2018-a) 『探索的財務ビッグデータ解析—前処理, データラングリング, 再現可能性—』, 商学論究, 第66巻, 第1号, pp. 1-32, 関西学院大学商学研究会.
- [12] 地道正行 (2018-b) 『探索的財務ビッグデータ解析—データ可視化, 統計モデリング, モデル選択, モデル評価, 動的文書生成, 再現可能研究—』, 商学論究, 第66巻, 第2号, pp. 1-41, 関西学院大学商学研究会.
- [13] 地道正行 (2018-c) 『データサイエンスの基礎：Rによる統計学独習』, 裳華房.
- [14] 地道正行 (2019) 『変換による財務データの統計解析—売上高の場合—』, 商学論究, 第67巻, 第1号, pp. 17-46, 関西学院大学商学研究会.
- [15] 地道正行 (2020-a) 『探索的財務ビッグデータ解析—前処理の並列化—』, 商学論究, 第67巻, 第3号, pp. 1-19, 関西学院大学商学研究会.
- [16] 地道正行 (2020-b) 『探索的財務ビッグデータ解析—PG-Stromによるデータラングリングの並列化—』, 商学論究, 第68巻, 第1号, pp. 1-34, 関西学院大学商学研究会.
- [17] Jimichi, M., D. Miyamoto, C. Saka, and S. Nagata (2018) Visualization and statistical modeling of financial big data: Double-log modeling with skew-symmetric error distributions, *Japanese Journal of Statistics and Data Science*, Vol. 1, No. 2, pp. 347-371, <https://doi.org/10.1007/s42081-018-0019-1>
- [18] 地道正行, 阪智香 (2021) 『財務データとESGレーティングデータの事前処理と結合』, 商学論究, 第69巻, 第3号, pp. 79-116, 関西学院大学商学研究会.
- [19] Konishi, S. and G. Kitagawa (2008) *Information Criteria and Statistical Modeling*, Springer.
- [20] Kozubowska, T. J. and J. P. Nolanb (2008) Infinite divisibility of skew Gaussian and Laplace laws, *Statistics & Probability Letters*, Vol. 78, pp. 654-660.
- [21] Mosteller, F. and J. W. Tukey (1977) *Data Analysis and Regression: A Second Course in Statistics*, Addison-Wesley, Reading Mass.
- [22] 大鹿智基, 阪智香, 地道正行 (2020) 『「社会にとってよい企業」への市場の評価とサステナビリティ』, 企業会計, 第72巻, 第1号, pp. 74-80.
- [23] 阪智香 (2021) 『ESG情報と企業価値』, 商学論究, 第68巻, 第4号, pp. 149-170, 関西学院大学商学研究会.
- [24] Tange, Ole, (2018) *GNU Parallel 2018*, ISBN: 9781387509881, DOI: 10.5281/zenodo.1146014, URL: <https://doi.org/10.5281/zenodo.1146014>, Mar,

2018.

- [25] Tukey, J. W. (1977) *Exploratory Data Analysis*, Addison-Wesley Publishing Co.
- [26] van der Loo, M. and E. de Jonge (2018) *Statistical Data Cleaning with Applications in R*, Wiley.
- [27] Wickham, H. and G. Grolemund (2016) *R for Data Science*, O'Reilly.

謝辞

本研究の一部は以下の助成を得ている。

科研費 科学研究費 基盤研究 C:「グラフィカル・データ・アナリシスによる格差研究と社会環境会計による解決方法の提案」(2016年～2019年), 課題番号: 16K04022

科研費 科学研究費 基盤研究 C:「共有価値創造 (CSV) のための社会環境会計の構築」(2019年～2021年), 課題番号: 19K02006

 2019年度 学際大規模情報基盤共同利用・共同研究拠点 (JHPCN) 課題: 財務ビッグデータの可視化と統計モデリング, 課題番号: jh191002-NWJ

 2020年度 学際大規模情報基盤共同利用・共同研究拠点 (JHPCN) 課題: 財務ビッグデータの可視化と統計モデリング, 課題番号: jh201003-NWJ

 2021年度 学際大規模情報基盤共同利用・共同研究拠点 (JHPCN) 課題: 財務ビッグデータの可視化と統計モデリング, 課題番号: jh211001-NWJ



関西学院大学 図書館 図書費 B, 研究設備費 (III), 個人研究費

また, 九州大学の増田弘毅教授には, Kozubowska and Nolanb (2008) の論文の存在をご教示いただいた。さらに, BvD 社の増田歩氏にはデータ抽出に関してご足労を賜った。ここに感謝の意を表する。

付録 A データラングリングとデータの欠測パターン

本稿では、データセット OsirisC2020 とデータセット FTSE2020 を結合することによって得られるデータセット OsirisC-FTSE2020 (ファイル名: `finSPT2020.csv`) を利用している (地道, 阪, 2012参照). このデータセットを R の `readr` パッケージに付属する関数 `read_csv` を利用して読み込んだ後, `dplyr` パッケージを利用して, 財務データに関しては全て正の値を選択し (`market_cap > 0`, `shareholders > 0`, `net_income > 0`), 決算月数が12ヵ月のものを選択している (`month == 12`). なお, この処理に利用されたスクリプトは以下のようなものである:

パッケージとデータの読み込み, データラングリング

```
> # パッケージの読み込み
> library(readr)
> library(dplyr)
> # データの読み込み
> finSPT <- read_csv("../finSPT2020.csv")
> # データラングリング
> finSPT.sub <- finSPT %>%
+   filter(market_cap > 0, shareholders > 0, net_income > 0, month == 12)%>%
+   mutate(log.mc = log(market_cap), log.sh = log(shareholders), log.ni = log(net_income))%>%
+   select(firmID, year, year_SPT, market_cap, shareholders, net_income,
+         log.mc, log.sh, log.ni, ESG, E, S, G, type)
```

この処理によって得られたオブジェクト `finSPT.sub` には, 財務データに関しては欠測値は存在しないが, ESG レーティングデータに関しては欠測値が存在する²⁶⁾. `VIM` パッケージに付属する関数 `aggr` を利用することによって, 欠測率や欠測パターンを以下のように調べるができる:

欠測パターンの可視化

```
> aggr(finSPT.sub, col = "gray", cex.axis = 0.7)
```

26) 欠測値を含むデータの扱いについての包括的な文献としては, van der Loo and de Jonge (2018) が詳しい.

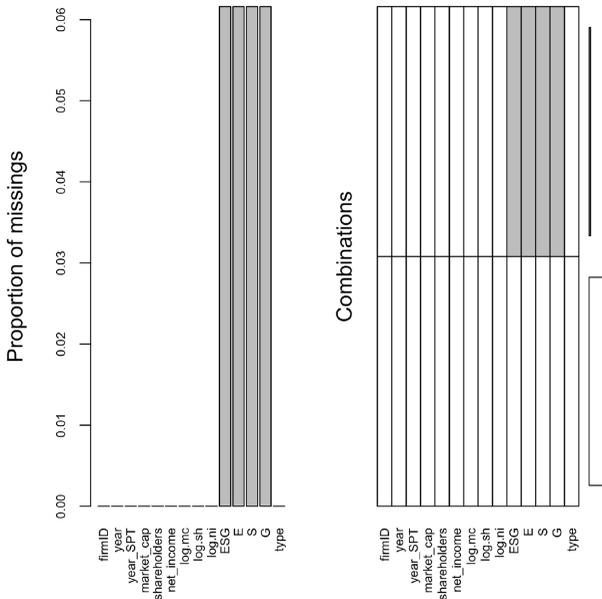


図21：オブジェクト `finSPT.sub` の欠測率と欠測パターン：ESG 指標 ESG, E, S, G のそれぞれの欠測率は約 6% (左図)、ESG 指標 ESG, E, S, G の全てが欠測 (：灰色) するパターンのみが存在 (右図)

結果として出力されるプロット (図21) において、「灰色」の部分欠測を表しており、ESG 指標 ESG, E, S, G の欠測率が、それぞれ、約 6% であることがわかり (左図)、欠測パターンとしては ESG 指標 ESG, E, S, G が全てが欠測する場合のみであることがわかる (右図)。なお、欠測件数は 885 である。

付録 B 正規分布と対数正規分布

ここでは、正規分布と対数正規分布についての基本的な事項を与える。

B.1 正規分布

確率変数 X が確率密度関数 (probability density function: p.d.f.) :

$$f_N(x | \boldsymbol{\theta}_N) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, x \in \mathbb{R} := (-\infty, \infty) \quad (9)$$

をもつとき、確率変数 X は正規分布 $N(\mu, \sigma^2)$ に従うと呼ばれ、 $X \sim N(\mu, \sigma^2)$ と書かれる。ここで、 $\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+ := (0, \infty)$ は未知母数であり、 $\boldsymbol{\theta}_N := [\mu, \sigma^2]'$ は正規分布の母数ベクトル²⁷⁾である。

B.2 対数正規分布

確率変数 X に対して、その対数 $\log X$ が正規分布 $N(\mu, \sigma^2)$ に従うとき、 X は対数正規分布 $LN(\mu, \sigma^2)$ に従うといわれる：

$$X \sim LN(\mu, \sigma^2) \stackrel{\text{def.}}{\iff} \log X \sim N(\mu, \sigma^2)$$

対数正規分布の p.d.f. は以下のように与えられる：

$$f_{LN}(x | \boldsymbol{\theta}_{LN}) = \frac{1}{\sqrt{2\pi\sigma^2} x} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, x \in \mathbb{R}^+ \quad (10)$$

対数正規分布の母数ベクトル $\boldsymbol{\theta}_{LN} = [\mu, \sigma^2]'$ は、正規分布の母数ベクトル $\boldsymbol{\theta}_N = [\mu, \sigma^2]'$ と同一であるが、分布が異なるため推定量は異なることに注意しよう。

B.3 対数正規分布と正規分布の対数尤度、赤池情報量規準の関係

対数正規分布と正規分布の p.d.f. には以下の関係が成り立つことに注意しよう：

$$f_{LN}(x | \boldsymbol{\theta}_{LN}) = f_N(\log x | \boldsymbol{\theta}_{LN}) \frac{1}{x} \quad (11)$$

この関係を利用すると、無作為標本 $\{X_1, \dots, X_n\}$ が対数正規分布 $LN(\mu, \sigma^2)$ に従うとき、その同時確率密度関数が、

$$\prod_{i=1}^n f_{LN}(x_i | \boldsymbol{\theta}_{LN}) = \prod_{i=1}^n f_N(\log x_i | \boldsymbol{\theta}_{LN}) \frac{1}{x_i} \quad (12)$$

となることがわかる。

さらに、この結果から、対数正規分布に従う無作為標本 $\{X_1, \dots, X_n\}$ にもとづく対数尤度は、

$$\begin{aligned}
 \ell_{\text{LN}}(\boldsymbol{\theta}_{\text{LN}} | \mathbf{x}) &= \log \prod_{i=1}^n f_{\text{LN}}(x_i | \boldsymbol{\theta}_{\text{LN}}) = \log \prod_{i=1}^n f_{\text{N}}(\log x_i | \boldsymbol{\theta}_{\text{LN}}) \frac{1}{x_i} \\
 &= \ell_{\text{N}}(\boldsymbol{\theta}_{\text{LN}} | \log \mathbf{x}) - \sum_{i=1}^n \log x_i
 \end{aligned} \tag{13}$$

となる。ここで、

$$\begin{aligned}
 \ell_{\text{N}}(\boldsymbol{\theta}_{\text{LN}} | \log \mathbf{x}) &= \log \prod_{i=1}^n f_{\text{N}}(\log x_i | \boldsymbol{\theta}_{\text{LN}}) = \sum_{i=1}^n \log f_{\text{N}}(\log x_i | \boldsymbol{\theta}_{\text{LN}}) \\
 &= \sum_{i=1}^n \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(\log x_i - \mu)^2}{2\sigma^2} \right\} \right) \\
 &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (\log x_i - \mu)^2
 \end{aligned}$$

であり、(13) 式は、対数正規分布に従う無作為標本 $\{X_1, \dots, X_n\}$ にもとづく対数尤度 $\ell_{\text{LN}}(\boldsymbol{\theta}_{\text{LN}} | \mathbf{x})$ と対数変換された無作為標本 $\{\log X_1, \dots, \log X_n\}$ が正規分布に従う場合の対数尤度 $\ell_{\text{N}}(\boldsymbol{\theta}_{\text{LN}} | \log \mathbf{x})$ の間の関係を表すことに注意しよう

このことから、

$$\max_{\boldsymbol{\theta}_{\text{LN}}} \ell_{\text{LN}}(\boldsymbol{\theta}_{\text{LN}} | \mathbf{x}) = \max_{\boldsymbol{\theta}_{\text{LN}}} \ell_{\text{N}}(\boldsymbol{\theta}_{\text{LN}} | \log \mathbf{x}) - \sum_{i=1}^n \log x_i \tag{14}$$

が成り立つので、最尤推定値ベクトルに関して以下の関係が成り立つ：

$$\hat{\boldsymbol{\theta}}_{\text{LN}} := \arg \max_{\boldsymbol{\theta}_{\text{LN}}} \ell_{\text{LN}}(\boldsymbol{\theta}_{\text{LN}} | \mathbf{x}) = \arg \max_{\boldsymbol{\theta}_{\text{LN}}} \ell_{\text{N}}(\boldsymbol{\theta}_{\text{LN}} | \log \mathbf{x}) \tag{15}$$

尤度方程式

$$\frac{\partial \ell_{\text{LN}}(\boldsymbol{\theta}_{\text{LN}} | \mathbf{x})}{\partial \boldsymbol{\theta}_{\text{LN}}} = \begin{bmatrix} \frac{\partial \ell_{\text{LN}}(\boldsymbol{\theta}_{\text{LN}} | \mathbf{x})}{\partial \mu} \\ \frac{\partial \ell_{\text{LN}}(\boldsymbol{\theta}_{\text{LN}} | \mathbf{x})}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} \left(\sum_{i=1}^n \log x_i - n\mu \right) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (\log x_i - \mu)^2 \end{bmatrix} = \mathbf{0}$$

を μ, σ^2 に関して解くことによって、最尤推定値は

$$\hat{\boldsymbol{\theta}}_{\text{LN}} := \begin{bmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} \overline{\log x} \\ \frac{1}{n} \sum_{i=1}^n (\log x_i - \overline{\log x})^2 \end{bmatrix} \tag{16}$$

となる。ここで、 $\overline{\log x} := \sum_{i=1}^n \log x_i / n$ とおいた。

これらの結果から、赤池情報量規準に関して以下のことが成り立つ：

27) プライム (') は行列・ベクトルの転置を表す。

$$\text{AIC}_{\text{LN}} := -2\ell_{\text{LN}}(\hat{\boldsymbol{\theta}}_{\text{LN}} | \mathbf{x}) + 2 \dim \boldsymbol{\theta}_{\text{LN}} = \text{AIC}_{\text{N}}(\log \mathbf{x}) + 2 \sum_{i=1}^n \log x_i \quad (17)$$

ここで,

$$\begin{aligned} \text{AIC}_{\text{N}}(\log \mathbf{x}) &:= -2\ell_{\text{N}}(\hat{\boldsymbol{\theta}}_{\text{LN}} | \log \mathbf{x}) + 2 \dim \boldsymbol{\theta}_{\text{LN}} \\ &= n \log(2\pi\sigma^2) + \frac{\sum_{i=1}^n (\log x_i - \hat{\mu})^2}{\sigma^2} + 4 \\ &= n \log(2\pi\sigma^2) + n + 4 \end{aligned}$$

は正規分布に従う無作為標本 $\{\log X_1, \dots, \log X_n\}$ にもとづく赤池情報量規準であり, $\dim \boldsymbol{\theta}_{\text{LN}} (=2)$ は, 母数ベクトル $\boldsymbol{\theta}_{\text{LN}}$ の次元 (dimension) である.

付録 C 非対称分布と対数非対称分布

ここでは, 非対称分布と対数非対称分布についての基本的な事項を与える. なお, 非対称分布族に関する詳細は Azzalini and Capitanio (2014) を参照されたい.

C.1 非対称正規分布

確率変数 X が p.d.f.

$$f_{\text{SN}}(x | \boldsymbol{\theta}_{\text{SN}}) = \frac{2}{\omega} \phi\left(\frac{x-\xi}{\omega}\right) \Phi\left(\alpha \frac{x-\xi}{\omega}\right), x \in \mathbb{R} \quad (18)$$

をもつとき, 確率変数 X は非対称正規分布 $\text{SN}(\xi, \omega^2, \alpha)$ に従うと呼ばれ, $X \sim \text{SN}(\xi, \omega^2, \alpha)$ と書かれる. ここで, $\xi (\in \mathbb{R}), \omega (\in \mathbb{R}^+), \alpha (\in \mathbb{R})$ は未知母数であり, $\boldsymbol{\theta}_{\text{SN}} := [\xi, \omega, \alpha]'$ は非対称正規分布の母数ベクトルである. また,

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), \quad \Phi(z) = \int_{-\infty}^z \phi(x) dx \quad (z \in \mathbb{R})$$

は, それぞれ, 標準正規分布 $\text{N}(0, 1)$ の p.d.f. と c.d.f. (累積分布関数) である. なお, (ξ, ω^2, α) は直接母数 (direct parameters) と呼ばれる.

C.2 対数非対称正規分布

確率変数 X に対して, その対数 $\log X$ が非対称正規分布 $\text{SN}(\xi, \omega^2, \alpha)$ に

従うとき、 X は対数非対称正規分布 $\text{LSN}(\xi, \omega^2, \alpha)$ に従うといわれる：

$$X \sim \text{LSN}(\xi, \omega^2, \alpha) \stackrel{\text{def.}}{\iff} \log X \sim \text{SN}(\xi, \omega^2, \alpha)$$

(Azzalini and Capitanio (2014) の p. 53 を参照のこと。)

対数非対称正規分布の p.d.f. は以下のように与えられる：

$$f_{\text{LSN}}(x | \boldsymbol{\theta}_{\text{LSN}}) = \frac{2}{\omega x} \phi\left(\frac{\log x - \xi}{\omega}\right) \Phi\left(\alpha \frac{\log x - \xi}{\omega}\right), x \in \mathbb{R}^+ \quad (19)$$

ここで、 $\boldsymbol{\theta}_{\text{LSN}} := [\xi, \omega, \alpha]'$ は対数非対称正規分布の母数ベクトルである。

C.3 対数非対称正規分布の対数尤度と非対称正規分布の対数尤度の関係

ここで扱う対数尤度の関係は、B.3 で考察した、対数正規分布の対数尤度と正規分布の対数尤度の関係と同様の結果が、対数非対称正規分布と非対称正規分布の間にも成り立つことであり、ほぼ、同様の議論展開であるが、結果をたどることにする。

まず、対数非対称正規分布と非対称正規分布の p.d.f. には以下の関係が成り立つことに注意しよう：

$$f_{\text{LSN}}(x | \boldsymbol{\theta}_{\text{LSN}}) = f_{\text{SN}}(\log x | \boldsymbol{\theta}_{\text{LSN}}) \frac{1}{x} \quad (20)$$

この関係を利用すると、無作為標本 $\{X_1, \dots, X_n\}$ が対数非対称正規分布 $\text{LSN}(\xi, \omega^2, \alpha)$ に従うとき、その同時確率密度関数が、

$$\prod_{i=1}^n f_{\text{LSN}}(x_i | \boldsymbol{\theta}_{\text{LSN}}) = \prod_{i=1}^n f_{\text{SN}}(\log x_i | \boldsymbol{\theta}_{\text{LSN}}) \frac{1}{x_i} \quad (21)$$

となる。

さらに、この結果から、対数非対称正規分布に従う無作為標本 $\{X_1, \dots, X_n\}$ にもとづく対数尤度は、

$$\begin{aligned} \ell_{\text{LSN}}(\boldsymbol{\theta}_{\text{LSN}} | \mathbf{x}) &= \log \prod_{i=1}^n f_{\text{LSN}}(x_i | \boldsymbol{\theta}_{\text{LSN}}) = \log \prod_{i=1}^n f_{\text{SN}}(\log x_i | \boldsymbol{\theta}_{\text{LSN}}) \frac{1}{x_i} \\ &= \ell_{\text{SN}}(\boldsymbol{\theta}_{\text{LSN}} | \log \mathbf{x}) - \sum_{i=1}^n \log x_i \end{aligned} \quad (22)$$

となる。ここで、

$$\begin{aligned}
\ell_{\text{SN}}(\boldsymbol{\theta}_{\text{LSN}} | \log \mathbf{x}) &= \log \prod_{i=1}^n f_{\text{SN}}(\log x_i | \boldsymbol{\theta}_{\text{LSN}}) = \sum_{i=1}^n \log f_{\text{SN}}(\log x_i | \boldsymbol{\theta}_{\text{LSN}}) \\
&= \sum_{i=1}^n \log \left(\frac{2}{\omega} \phi \left(\frac{\log x_i - \xi}{\omega} \right) \Phi \left(\alpha \frac{\log x_i - \xi}{\omega} \right) \right) \\
&= -\frac{n}{2} \log(2\pi\omega^2) - \frac{1}{2\omega^2} \sum_{i=1}^n (\log x_i - \xi)^2 + \sum_{i=1}^n \log 2\Phi \left(\alpha \frac{\log x_i - \xi}{\omega} \right) \\
&= -\frac{n}{2} \log(2\pi) - n \log \omega - \frac{1}{2\omega^2} \sum_{i=1}^n (\log x_i - \xi)^2 \\
&\quad + \sum_{i=1}^n \log 2\Phi \left(\alpha \frac{\log x_i - \xi}{\omega} \right)
\end{aligned}$$

であり、(22) 式は、対数非対称正規分布に従う無作為標本 $\{X_1, \dots, X_n\}$ にもとづく対数尤度 $\ell_{\text{LSN}}(\boldsymbol{\theta}_{\text{LSN}} | \mathbf{x})$ と対数変換された無作為標本 $\{\log X_1, \dots, \log X_n\}$ が非対称正規分布に従う場合の対数尤度 $\ell_{\text{SN}}(\boldsymbol{\theta}_{\text{LSN}} | \log \mathbf{x})$ の間の関係を表すことに注意しよう

このことから、

$$\max_{\boldsymbol{\theta}_{\text{LSN}}} \ell_{\text{LSN}}(\boldsymbol{\theta}_{\text{LSN}} | \mathbf{x}) = \max_{\boldsymbol{\theta}_{\text{LSN}}} \ell_{\text{SN}}(\boldsymbol{\theta}_{\text{LSN}} | \log \mathbf{x}) - \sum_{i=1}^n \log x_i \quad (23)$$

が成り立つので、最尤推定値ベクトルに関して以下の関係が成り立つ：

$$\hat{\boldsymbol{\theta}}_{\text{LSN}} := \arg \max_{\boldsymbol{\theta}_{\text{LSN}}} \ell_{\text{LSN}}(\boldsymbol{\theta}_{\text{LSN}} | \mathbf{x}) = \arg \max_{\boldsymbol{\theta}_{\text{LSN}}} \ell_{\text{SN}}(\boldsymbol{\theta}_{\text{LSN}} | \log \mathbf{x}) \quad (24)$$

なお、対数非対称正規分布の母数ベクトルに対する最尤推定値ベクトル $\hat{\boldsymbol{\theta}}_{\text{LSN}} := [\hat{\xi}, \hat{\omega}, \hat{\alpha}]'$ は陽に表現することができないので、尤度方程式 $\partial \ell_{\text{LSN}}(\boldsymbol{\theta}_{\text{LSN}} | \mathbf{x}) / \partial \boldsymbol{\theta}_{\text{LSN}} = \mathbf{0}$ を満たす $\boldsymbol{\theta}_{\text{LSN}}$ を数値的に求める必要がある。

以上の結果から、赤池情報量規準に関して以下のことが成り立つ：

$$\begin{aligned}
\text{AIC}_{\text{LSN}} &:= -2\ell_{\text{LSN}}(\hat{\boldsymbol{\theta}}_{\text{LSN}} | \mathbf{x}) + 2 \dim \boldsymbol{\theta}_{\text{LSN}} \\
&= -2\ell_{\text{SN}}(\hat{\boldsymbol{\theta}}_{\text{LSN}} | \log \mathbf{x}) + 2 \sum_{i=1}^n \log x_i + 2 \dim \boldsymbol{\theta}_{\text{LSN}} \\
&= \text{AIC}_{\text{SN}}(\log \mathbf{x}) + 2 \sum_{i=1}^n \log x_i \quad (25)
\end{aligned}$$

ここで、

$$\begin{aligned}
\text{AIC}_{\text{SN}}(\log \mathbf{x}) &:= -2\ell_{\text{SN}}(\hat{\boldsymbol{\theta}}_{\text{LSN}} | \log \mathbf{x}) + 2 \dim \boldsymbol{\theta}_{\text{LSN}} \\
&= n \log(2\pi) + n \log \hat{\omega}^2 + \frac{1}{\hat{\omega}^2} \sum_{i=1}^n (\log x_i - \hat{\xi})^2 \\
&\quad - 2 \sum_{i=1}^n \log 2\Phi \left(\hat{\alpha} \frac{\log x_i - \hat{\xi}}{\hat{\omega}} \right) + 6
\end{aligned}$$

は非対称正規分布に従う無作為標本 $\{\log X_1, \dots, \log X_n\}$ にもとづく赤池情報量規準であり, $\dim \theta_{\text{LSN}} (=3)$ は, 母数ベクトル θ_{LSN} の次元である.

C.4 非対称ティー分布

確率変数 X が p.d.f.:

$$f_{\text{ST}}(x | \theta_{\text{ST}}) = \frac{2}{\omega} f_t\left(\frac{x-\xi}{\omega} \mid \nu\right) F_t\left(\alpha \frac{x-\xi}{\omega} \sqrt{\frac{\nu+1}{\left(\frac{x-\xi}{\omega}\right)^2 + \nu}} \mid \nu+1\right), x \in \mathbb{R} \quad (26)$$

をもつとき, 確率変数 X は非対称ティー分布 $\text{ST}(\xi, \omega^2, \alpha, \nu)$ に従うと呼ばれ,

$$X \sim \text{ST}(\xi, \omega^2, \alpha, \nu)$$

と書かれる. ここで, $\xi (\in \mathbb{R}), \omega (\in \mathbb{R}^+), \alpha (\in \mathbb{R}), \nu (\in \mathbb{R}^+)$ は未知母数であり, $\theta_{\text{ST}} := [\xi, \omega, \alpha, \nu]'$ は非対称ティー分布の母数ベクトルである. なお, $(\xi, \omega^2, \alpha, \nu)$ は直接母数と呼ばれる. また,

$$f_t(z | \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}} \left(1 + \frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad F_t(z | \nu) = \int_{-\infty}^z f_t(x | \nu) dx$$

は, それぞれ, 自由度 ν のティー分布の p.d.f. と c.d.f. であり, $\Gamma(\cdot)$ はガンマ関数である.

C.5 対数非対称ティー分布

確率変数 X に対して, その対数 $\log Y$ が非対称ティー分布 $\text{ST}(\xi, \omega^2, \alpha, \nu)$ に従うとき, X は対数非対称ティー分布 $\text{LST}(\xi, \omega^2, \alpha, \nu)$ に従うといわれる:

$$X \sim \text{LST}(\xi, \omega^2, \alpha, \nu) \stackrel{\text{def.}}{\iff} \log X \sim \text{ST}(\xi, \omega^2, \alpha, \nu)$$

対数非対称ティー分布の p.d.f. は以下のように与えられる:

$$f_{\text{LST}}(x | \boldsymbol{\theta}_{\text{LST}}) = \frac{2}{\omega x} f_1\left(\frac{\log x - \xi}{\omega} \mid \nu\right) \\ \times F_1\left(\alpha \frac{\log x - \xi}{\omega} \sqrt{\frac{\nu+1}{\left(\frac{\log x - \xi}{\omega}\right)^2 + \nu}} \mid \nu+1\right), x \in \mathbb{R}^+$$

ここで、 $\boldsymbol{\theta}_{\text{LST}} := [\xi, \omega, \alpha, \nu]'$ は対数非対称ティー分布の母数ベクトルである。

C.6 対数非対称ティー分布の対数尤度と非対称ティー分布の対数尤度の関係

ここで扱う対数尤度の関係も、B.3で考察した、対数正規分布の対数尤度と正規分布の対数尤度の関係と同様の結果が、対数非対称ティー分布と非対称ティー分布の間にも成り立つことであり、ほぼ、同様の議論展開であるが、結果をたどることにする。

対数非対称ティー分布と非対称ティー分布の p.d.f. には以下の関係が成り立つことに注意しよう：

$$f_{\text{LST}}(x | \boldsymbol{\theta}_{\text{LST}}) = f_{\text{ST}}(\log x | \boldsymbol{\theta}_{\text{LST}}) \frac{1}{x} \quad (27)$$

この関係を利用すると、無作為標本 $\{X_1, \dots, X_n\}$ が対数非対称ティー分布 $\text{LST}(\xi, \omega^2, \alpha, \nu)$ に従うとき、その同時確率密度関数が、

$$\prod_{i=1}^n f_{\text{LST}}(x_i | \boldsymbol{\theta}_{\text{LST}}) = \prod_{i=1}^n f_{\text{ST}}(\log x_i | \boldsymbol{\theta}_{\text{LST}}) \frac{1}{x_i} \quad (28)$$

となることがわかる。

さらに、この結果から、対数非対称ティー分布に従う無作為標本 $\{X_1, \dots, X_n\}$ にもとづく対数尤度は、

$$\ell_{\text{LST}}(\boldsymbol{\theta}_{\text{LST}} | \mathbf{x}) = \log \prod_{i=1}^n f_{\text{LST}}(x_i | \boldsymbol{\theta}_{\text{LST}}) = \log \prod_{i=1}^n f_{\text{ST}}(\log x_i | \boldsymbol{\theta}_{\text{LST}}) \frac{1}{x_i} \\ = \ell_{\text{ST}}(\boldsymbol{\theta}_{\text{LST}} | \log \mathbf{x}) - \sum_{i=1}^n \log x_i \quad (29)$$

となる。ここで、

$$\begin{aligned}
 \ell_{\text{ST}}(\boldsymbol{\theta}_{\text{LST}} | \log \mathbf{x}) &= \log \prod_{i=1}^n f_{\text{ST}}(\log x_i | \boldsymbol{\theta}_{\text{LST}}) = \sum_{i=1}^n \log f_{\text{ST}}(\log x_i | \boldsymbol{\theta}_{\text{LST}}) \\
 &= \sum_{i=1}^n \log \left(\frac{2}{\omega} f_i \left(\frac{\log x_i - \xi}{\omega} \mid \nu \right) \times F_i \left(\alpha \frac{\log x_i - \xi}{\omega} \sqrt{\frac{\nu+1}{\left(\frac{\log x_i - \xi}{\omega} \right)^2 + \nu}} \mid \nu+1 \right) \right) \\
 &= \sum_{i=1}^n \log \left(\frac{1}{\omega} f_i \left(\frac{\log x_i - \xi}{\omega} \mid \nu \right) \right) + \sum_{i=1}^n \log 2F_i \left(\alpha \frac{\log x_i - \xi}{\omega} \sqrt{\frac{\nu+1}{\left(\frac{\log x_i - \xi}{\omega} \right)^2 + \nu}} \mid \nu+1 \right)
 \end{aligned}$$

であり、(29) 式は、対数非対称テーパー分布に従う無作為標本 $\{X_1, \dots, X_n\}$ にもとづく対数尤度 $\ell_{\text{LST}}(\boldsymbol{\theta}_{\text{LST}} | \mathbf{x})$ と対数変換された無作為標本 $\{\log X_1, \dots, \log X_n\}$ が非対称テーパー分布に従う場合の対数尤度 $\ell_{\text{ST}}(\boldsymbol{\theta}_{\text{LST}} | \log \mathbf{x})$ の間の関係を表すことに注意しよう

このことから、

$$\max_{\boldsymbol{\theta}_{\text{LST}}} \ell_{\text{LST}}(\boldsymbol{\theta}_{\text{LST}} | \mathbf{x}) = \max_{\boldsymbol{\theta}_{\text{LST}}} \ell_{\text{ST}}(\boldsymbol{\theta}_{\text{LST}} | \log \mathbf{x}) - \sum_{i=1}^n \log x_i \quad (30)$$

が成り立つので、最尤推定値（ベクトル）に関して以下の関係が成り立つ：

$$\hat{\boldsymbol{\theta}}_{\text{LST}} := \arg \max_{\boldsymbol{\theta}_{\text{LST}}} \ell_{\text{LST}}(\boldsymbol{\theta}_{\text{LST}} | \mathbf{x}) = \arg \max_{\boldsymbol{\theta}_{\text{LST}}} \ell_{\text{ST}}(\boldsymbol{\theta}_{\text{LST}} | \log \mathbf{x}) \quad (31)$$

なお、対数非対称テーパー分布の母数ベクトルに対する最尤推定値ベクトル $\hat{\boldsymbol{\theta}}_{\text{LST}} := [\hat{\xi}, \hat{\omega}, \hat{\alpha}, \hat{\nu}]'$ は陽に表現することができないので、尤度方程式 $\partial \ell_{\text{LST}}(\boldsymbol{\theta}_{\text{LST}} | \mathbf{x}) / \partial \boldsymbol{\theta}_{\text{LST}} = \mathbf{0}$ を満たす $\boldsymbol{\theta}_{\text{LST}}$ を数値的に求める必要がある。

以上の結果から、赤池情報量規準に関して以下のことが成り立つ：

$$\begin{aligned}
 \text{AIC}_{\text{LST}} &:= -2\ell_{\text{LST}}(\hat{\boldsymbol{\theta}}_{\text{LST}} | \mathbf{x}) + 2\dim \boldsymbol{\theta}_{\text{LST}} \\
 &= -2\ell_{\text{ST}}(\hat{\boldsymbol{\theta}}_{\text{LST}} | \log \mathbf{x}) + 2\sum_{i=1}^n \log x_i + 2\dim \boldsymbol{\theta}_{\text{LST}} \\
 &= \text{AIC}_{\text{ST}}(\log \mathbf{x}) + 2\sum_{i=1}^n \log x_i \quad (32)
 \end{aligned}$$

ここで,

$$\begin{aligned} \text{AIC}_{\text{ST}}(\log \mathbf{x}) &:= -2\ell_{\text{ST}}(\hat{\boldsymbol{\theta}}_{\text{LST}} | \log \mathbf{x}) + 2 \dim \boldsymbol{\theta}_{\text{LST}} \\ &= -2 \sum_{i=1}^n \log \left(\frac{1}{\hat{\omega}} f_i \left(\frac{\log x_i - \hat{\xi}}{\hat{\omega}} \middle| \hat{\nu} \right) \right) \\ &\quad - 2 \sum_{i=1}^n \log 2F_1 \left(\hat{\alpha} \frac{\log x_i - \hat{\xi}}{\hat{\omega}} \sqrt{\frac{\hat{\nu} + 1}{\left(\frac{\log x_i - \hat{\xi}}{\hat{\omega}} \right)^2 + \hat{\nu}}} \middle| \hat{\nu} + 1 \right) + 8 \end{aligned}$$

は非対称テイー分布に従う無作為標本 $\{\log X_1, \dots, \log X_n\}$ にもとづく赤池情報量規準であり, $\dim \boldsymbol{\theta}_{\text{LST}} (=4)$ は, 母数ベクトル $\boldsymbol{\theta}_{\text{LST}}$ の次元である.

付録 D R スクリプト

本稿で利用した R スクリプトを以下に与える. なお, VII 節と VIII 節でみた赤池情報量規準と検定統計量の経年変化を描くためのスクリプトは冗長であるため割愛している.

```
#####
## 環境設定パート
#####
## R パッケージの読み込み
#####
library(readr)
library(dplyr)
library(ggplot2)
library(GGally)
library(xtable)
library(e1071)
library(car)
library(sn)
library(VIM)
#####
## データの読み込み
#####
finSPT <- read_csv("./finSPT2020.csv")
#####
## データラングリング
#####
finSPT.sub <- finSPT %>%
filter(market_cap > 0,shareholders > 0,net_income > 0,month == 12)%>%
  mutate(log.mc = log(market_cap), log.sh = log(shareholders), log.ni = log(net_income)) %>%
  select(firmID, year, year_SPT,
         market_cap, shareholders, net_income, log.mc, log.sh, log.ni, ESG, E, S, G, type)
#####
## 企業数をカウントする関数を定義
#####
count.obs <- function() {
  require(dplyr)
  a.T <- finSPT.sub %>% group_by(year) %>% summarise(n = n())
  a.G <- finSPT.sub %>% filter(type == "G") %>% group_by(year) %>% summarise(G = n())
}
```

```

a.E <- finSPT.sub %>% filter(type == "E") %>% group_by(year) %>% summarise(E = n())
a <- a.G %>% left_join(a.E)
a <- a %>% left_join(a.T)
a <- a %>% mutate(year_SPT = seq(2015, 2020)) %>% select(year, year_SPT, G, E, n)
print(a)
}
#####
### 欠測値を除去した企業数をカウントする関数を定義
#####
count.obs.omit.NA <- function(){
  require(dplyr)
  a.T <- finSPT.sub %>% group_by(year) %>% na.omit() %>% summarise(n = n())
  a.G <- finSPT.sub %>% filter(type == "G") %>% na.omit() %>% group_by(year) %>% summarise(G = n())
  a.E <- finSPT.sub %>% filter(type == "E") %>% na.omit() %>% group_by(year) %>% summarise(E = n())
  a <- a.G %>% left_join(a.E)
  a <- a %>% left_join(a.T)
  a <- a %>% mutate(year_SPT = seq(2015, 2020)) %>% select(year, year_SPT, G, E, n)
  print(a)
}
#####
### 2018 年度の株式時価総額に分布の当てはめ
#####
# 正規分布
lm.log.mc2018 <- filter(finSPT.sub, year == 2018) %>% lm(log.mc ~ 1, data = .)
coef.lm.log.mc2018 <- coef(lm.log.mc2018, param.type = "DP")
# 非対称正規分布
selm.log.mc2018.SN <- filter(finSPT.sub, year == 2018) %>% selm(log.mc ~ 1, family = "SN", data = .)
coef.selm.log.mc2018.SN <- coef(selm.log.mc2018.SN, param.type = "DP")
# 非対称テーパー分布
selm.log.mc2018.ST <- filter(finSPT.sub, year == 2018) %>% selm(log.mc ~ 1, family = "ST", data = .)
coef.selm.log.mc2018.ST <- coef(selm.log.mc2018.ST, param.type = "DP")
#####
### 対散布図を描くための関数を定義
#####
finSPT.ggpairs <- function(df = finSPT.sub, yr = 2020, type = "raw")
{
  require(GGally)
  if(type == "raw")
  {
    p <- df %>% filter(year_SPT == yr) %>%
      select(market_cap, shareholders, net_income, ESG, E, S, G, type) %>%
      ggpairs(upper = list(continuous = wrap("points", size = 0.1, alpha = 0.2), combo = "box"),
              lower = list(continuous = wrap("cor", size=2), combo = wrap("facetdensity")),
              mapping = aes(color = type),
              diag = list(continuous = wrap("densityDiag", alpha = 0.5)))+
      ggtitle(as.character(yr)) +
      theme(axis.text= element_text(size = 5),
            legend.title = element_text(size = 7),
            legend.text = element_text(size = 7),
            axis.title = element_text(size = 7),
            plot.title= element_text(size = 7),
            strip.text = element_text(size = 7))
  }
}
if(type == "log")
{
  p <- df %>% filter(year_SPT == yr) %>%
    select(log.mc, log.sh, log.ni, ESG, E, S, G, type) %>%
    ggpairs(upper = list(continuous = wrap("points", size = 0.1, alpha = 0.2), combo = "box"),
            lower = list(continuous = wrap("cor", size=2), combo = wrap("facetdensity")),
            mapping = aes(color = type),
            diag = list(continuous = wrap("densityDiag", alpha = 0.5)))+
    ggtitle(as.character(yr)) +
    theme(axis.text= element_text(size = 5),
          legend.title = element_text(size = 7),
          legend.text = element_text(size = 7),

```

```

axis.title = element_text(size = 7),
plot.title= element_text(size = 7),
strip.text = element_text(size = 7))
}
if(type == "both")
{
  p <- df %>% filter(year_SPT == yr) %>%
  select(log.mc, log.sh, log.ni, market_cap, shareholders, net_income, ESG, E, S, G, type) %>%
  ggpairs(upper = list(continuous = wrap("points", size = 0.1, alpha = 0.2), combo = "box"),
  lower = list(continuous = wrap("cor", size=2), combo = wrap("facetdensity")),
  mapping = aes(color = type),
  diag = list(continuous = wrap("densityDiag", alpha = 0.5)))+
  ggtitle(as.character(yr)) +
  theme(axis.text= element_text(size = 5),
  legend.title = element_text(size = 7),
  legend.text = element_text(size = 7),
  axis.title = element_text(size = 7),
  plot.title= element_text(size = 7),
  strip.text = element_text(size = 7))
}
print(p)
}
}
#####
## モデル当てはめパート
#####
## 両対数モデルの当てはめ : ESG
# -----
# log(market_cap) ~ log(shareholders) + log(net_income) + ESG + type
# -----
#####
# 正規誤差 (lm)
lm.log.finSPT2020.ESG <- lm(log(market_cap) ~ log(shareholders) + log(net_income) + ESG + type,
data = filter(finSPT.sub, year_SPT == 2020))
coef.lm.log.finSPT2020.ESG <- coef(lm.log.finSPT2020.ESG)
# 非対称正規誤差 (lmSN)
selm.log.finSPT2020.ESG.SN <- selm(log(market_cap) ~ log(shareholders) + log(net_income) + ESG + type,
family = "SN", data = filter(finSPT.sub, year_SPT == 2020))
coef.selm.log.finSPT2020.ESG.SN <- coef(selm.log.finSPT2020.ESG.SN, "DP")
# 非対称テイエ誤差 (lmST)
selm.log.finSPT2020.ESG.ST <- selm(log(market_cap) ~ log(shareholders) + log(net_income) + ESG + type,
family = "ST", data = filter(finSPT.sub, year_SPT == 2020))
coef.selm.log.finSPT2020.ESG.ST <- coef(selm.log.finSPT2020.ESG.ST, "DP")
# 中心化母数を求めるための関数の定義
bnu <- function(nu) sqrt(nu)*gamma((nu-1)/2) / (sqrt(pi)*gamma(nu/2))
delta <- function(alpha) alpha/sqrt(1+alpha^2)
omega.bnu.delta <- function(omega, alpha, nu) omega*bnu(nu)*delta(alpha)
#####
## 両対数モデルの当てはめ : (E, S, G)
# -----
# Model : log(market_cap) ~ log(shareholders) + log(net_income) + E + S + G + type
# -----
#####
# 正規誤差 (lm)
lm.log.finSPT2020.E.S.G <- lm(log(market_cap) ~ log(shareholders) + log(net_income) + E + S + G + type,
data = filter(finSPT.sub, year_SPT == 2020))
coef.lm.log.finSPT2020.E.S.G <- coef(lm.log.finSPT2020.E.S.G)
# 非対称正規誤差 (lmSN)
selm.log.finSPT2020.E.S.G.SN <- selm(log(market_cap) ~ log(shareholders) + log(net_income) + E + S + G + type,
family = "SN", data = filter(finSPT.sub, year_SPT == 2020))
coef.selm.log.finSPT2020.E.S.G.SN <- coef(selm.log.finSPT2020.E.S.G.SN, "DP")
# 非対称テイエ誤差 (lmST)
selm.log.finSPT2020.E.S.G.ST <- selm(log(market_cap) ~ log(shareholders) + log(net_income) + E + S + G + type,
family = "ST", data = filter(finSPT.sub, year_SPT == 2020))
coef.selm.log.finSPT2020.E.S.G.ST <- coef(selm.log.finSPT2020.E.S.G.ST, "DP")
load("Double-log-Modelling.RData")

```

```

#####
## データ解析パート
#####
## データ 요약
summary(finSPT.sub)
#####
## 企業数カウント
#####
tmp <- count.obs()
colnames(tmp) <- c("会計年度", "ESG 指標年度", "先進国企業数", "新興国企業数", "総企業数")
#####
## 企業数表出力
#####
print(
  xtable(
    tmp,
    digits = c(0,0,0,0,0,0),
    display = c("s",rep("f",5)),
    floating = FALSE,
    caption = c("年度別企業数"),
    label = "table :number.firms",
    caption.placement = "top",table.placement="H",
    size = "\\setlength{\\tabcolsep}{2pt}"
  )
)
#####
## 企業数表出力
#####
print(
  xtable(
    tmp.na.omit,
    digits = c(0,0,0,0,0,0),
    display=c("s",rep("f",5)),
    floating = FALSE,
    caption = c("欠測値除去後の年度別企業数"),
    label = "table :number.firms.na.omit",
    caption.placement = "top",table.placement="H",
    size="\\setlength{\\tabcolsep}{2pt}"
  )
)
#####
# 2018 年度の株式時価総額のヒストグラム描画
#####
finSPT.sub %>% filter(year == 2018) %>% select(market_cap) %>%
  ggplot(aes(market_cap)) + geom_histogram() + geom_rug()
#####
## 2018 年度の対数株式時価総額のヒストグラム描画
#####
finSPT.sub %>% filter(year == 2018) %>% select(market_cap) %>%
  ggplot(aes(log(market_cap))) + geom_histogram(binwidth = 0.5) + geom_rug()
#####
## 2018 年度の対数株式時価総額の正規 Q-Q プロット描画
#####
finSPT.sub %>% filter(year_SPT == 2020) %>% select(market_cap) %>%
  ggplot(aes(sample = log(market_cap))) + geom_qq() + geom_qq_line()
#####
## 対数株式時価総額の歪度計算
#####
skewness.log.mc2018 <- finSPT.sub %>%
  filter(year == 2018) %>%
  pull(market_cap) %>% skewness()
#####
## 対散布図の描画: 粗データ
#####
finSPT.ggpairs()
#####
## 対散布図の描画: (財務データのみ) 対数スケール

```

```
#####
finSPT.ggpairs(type = "log")
#####
### 非対称正規分布の当てはめ結果 (ヒストグラム, P-Pプロット)
#####
plot(selm.log.mc2018.SN, which = 2)
plot(selm.log.mc2018.SN, which = 4)
#####
### 非対称正規ティーの当てはめ結果 (ヒストグラム, P-Pプロット)
#####
plot(selm.log.mc2018.ST, which=2)
plot(selm.log.mc2018.ST, which=4)
#####
### ティー検定表: 両対数モデル (ESG, 正規誤差)
#####
print(
  xtable(lm.log.finSPT2020.ESG,
    floating = FALSE,
    caption = c("ティー検定表: 正規誤差"),
    label = "table.t.lm.log.finSPT2020.ESG"),
  caption.placement = "top",
  table.placement = "H")
#####
### 残差のヒストグラム, 正規 Q-Qプロット: 両対数モデル (ESG, 正規誤差)
#####
hist(resid(lm.log.finSPT2020.ESG), xlim = c(-4, 4), freq = F)
curve(dnorm(x, mean = 0, sd = sd(resid(lm.log.finSPT2020.ESG))), -4, 4, add = T)
lm.log.finSPT2020.ESG %>% plot(which = 2)
#####
### ゼット検定表: 両対数モデル (ESG, 非対称正規誤差)
#####
print(
  xtable(summary(selm.log.finSPT2020.ESG.SN, "DP")@param.table,
    digits = c(0,4,4,4,4),
    floating = FALSE,
    caption = c("ゼット比検定表: 非対称正規誤差"),
    label = "table.z.selm.log.finSPT2020.ESG.SN"),
  caption.placement = "top",
  table.placement = "H")
#####
### 残差のヒストグラム, P-Pプロット: 両対数モデル (ESG, 非対称正規誤差)
#####
selm.log.finSPT2020.ESG.SN %>% plot(which = 2)
selm.log.finSPT2020.ESG.SN %>% plot(which = 4)
#####
### ゼット検定表: 両対数モデル (ESG, 非対称ティー誤差)
#####
print(
  xtable(summary(selm.log.finSPT2020.ESG.ST, "DP")@param.table,
    digits = c(0,4,4,4,4),
    floating = FALSE,
    caption = c("ゼット比検定表: 非対称ティー誤差"),
    label = "table.z.selm.log.finSPT2020.ESG.ST"),
  caption.placement = "top", table.placement="H")
#####
### 残差のヒストグラム, P-Pプロット: 両対数モデル (ESG, 非対称ティー誤差)
#####
selm.log.finSPT2020.ESG.ST %>% plot(which = 2)
selm.log.finSPT2020.ESG.ST %>% plot(which = 4)
#####
### ティー検定表: 両対数モデル (E, S, G, 正規誤差)
#####
print(
  xtable(lm.log.finSPT2020.E.S.G,
    digits = c(0,4,4,4,4),
```

```

floating = FALSE,
caption = c("ティ-検定表: 正規誤差"),
label = "table.t.lm.log.finSPT2020.E.S.G",
caption.placement = "top",table.placement="H")
#####
### 残差のヒストグラム, 正規-Q-Qプロット: 両対数モデル (E, S, G, 正規誤差)
#####
hist(resid(lm.log.finSPT2020.E.S.G), xlim = c(-4, 4), freq = F)
curve(dnorm(x, mean = 0, sd = sd(resid(lm.log.finSPT2020.E.S.G))), -4, 4, add = T)
lm.log.finSPT2020.E.S.G %>% plot(which = 2)
#####
### ゼット検定表: 両対数モデル (E, S, G, 非対称正規誤差)
#####
print(
xtable(summary(selm.log.finSPT2020.E.S.G.SN, "DP")@param.table,
digits = c(0,4,4,4,4),
floating = FALSE,
caption = c("ゼット比検定表: 非対称正規誤差"),
label = "table.z.selm.log.finSPT2020.E.S.G.SN"),
caption.placement = "top",table.placement="H")
#####
### 残差のヒストグラム, P-Pプロット: 両対数モデル (E, S, G, 非対称正規誤差)
#####
selm.log.finSPT2020.E.S.G.SN %>% plot(which = 2)
selm.log.finSPT2020.E.S.G.SN %>% plot(which = 4)
#####
### ゼット検定表: 両対数モデル (E, S, G, 非対称ティ-誤差)
#####
print(
xtable(summary(selm.log.finSPT2020.E.S.G.ST, "DP")@param.table,
digits = c(0,4,4,4,4),
floating = FALSE,
caption = c("ゼット比検定表: 非対称ティ-誤差"),
label = "table.z.selm.log.finSPT2020.E.S.G.ST"),
caption.placement = "top",table.placement="H")
#####
### 残差のヒストグラム, P-Pプロット: 両対数モデル (E, S, G, 非対称ティ-誤差)
#####
selm.log.finSPT2020.E.S.G.ST %>% plot(which = 2)
selm.log.finSPT2020.E.S.G.ST %>% plot(which = 4)
#####
### AIC 表: 株式時価総額へ非対称分布の当てはめ
#####
AIC.log.mc <- AIC(selm.log.mc2018.SN, selm.log.mc2018.ST)
colnames(AIC.log.mc) <- c("dim", "AIC")
#####
### AIC 表出力
#####
print(
xtable(AIC.log.mc,
digits = c(0,0,2),
floating = FALSE,
caption = c("AIC 表: 対数株式時価総額の分布に関する比較"),
label = "table.distribution.AIC"),
caption.placement = "top",
table.placement = "H")
#####
### AIC 表: 両対数モデルの当てはめ: ESG
#####
AIC.log.mc.ESG <- AIC(lm.log.finSPT2020.ESG,
selm.log.finSPT2020.ESG.SN,
selm.log.finSPT2020.ESG.ST)
colnames(AIC.log.mc.ESG) <- c("dim", "AIC")
#####
### AIC 表出力

```

```
#####
print(
xtable(AIC.log.mc.ESG,
digits = c(0,0,2),
floating = FALSE,
caption = c("AIC 表: 両対数モデル (ESG 指標: ESG) に関する比較"),
label = "table:AIC.log.mc.ESG"),
caption.placement = "top",
table.placement = "H")
#####
### AIC 表: 両対数モデルの当てはめ: E, S, G
#####
AIC.log.mc.E.S.G <- AIC(lm.log.finsPT2020.E.S.G, selm.log.finsPT2020.E.S.G.SN, selm.log.finsPT2020.E.S.G.ST)
colnames(AIC.log.mc.E.S.G) <- c("dim", "AIC")
#####
### AIC 表出力
#####
print(
xtable(AIC.log.mc.E.S.G,
digits = c(0,0,2),
floating = FALSE,
caption = c("AIC 表: 両対数モデル (ESG 指標: E, S, G) に関する比較"),
label = "table:AIC.log.mc.E.S.G"),
caption.placement = "top",
table.placement = "H")
#####
### 分布当てはめ結果の AIC の経年変化のプロット
#####
plot.transition.log.mi(dist.log.mi, type = "AIC")
#####
### 両対数モデル (ESG) 当てはめ結果の AIC の経年変化のプロット
#####
plot.transition1(dlm.ESG, type = "AIC")
#####
### 欠測値除去後の企業数カウント
#####
tmp.na.omit <- count.obs.omit.NA()
colnames(tmp.na.omit) <- c("会計年度", "ESG 指標年度", "先進国企業数", "新興国企業数", "総企業数")
#####
### 両対数モデル (E, S, G) の当てはめ結果の AIC の経年変化のプロット
#####
plot.transition2(dlm.dlm.E.S.G, type = "AIC")
#####
### 2016 年度のデータに両対数モデルを当てはめ
#####
selm.log.finsPT2016.ESG.ST <- selm(log(market_cap) ~ log(shareholders) + log(net_income) + ESG + type,
family = "ST", data = filter(finsPT.sub, year_SPT == 2016))
#####
### 両対数モデル (ESG) における非対称母数に関する推定値等の経年変化
#####
plot.transition.alpha(dlm1.ESG.alpha)
#####
### 両対数モデル (E, S, G) における非対称母数に関する推定値等の経年変化
#####
plot.transition.alpha(dlm2.ESG.alpha)
#####
### オブジェクト finsPT.sub の欠測情報の可視化
#####
aggr(finsPT.sub, col = "gray", cex.axis = 0.7)
#####
```

付録 E R の環境情報

R 関数 `sessionInfo` を実行することによって、本稿を執筆すること
に利用した R に関する環境情報を以下に与える：

sessionInfo の実行結果

```
> sessionInfo()
R version 4.1.0 (2021-05-18)
Platform: x86_64-apple-darwin17.0 (64-bit)
Running under: macOS Big Sur 10.16

Matrix products: default
BLAS: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib

locale:
[1] ja_JP.UTF-8/ja_JP.UTF-8/ja_JP.UTF-8/C/ja_JP.UTF-8/ja_JP.UTF-8

attached base packages:
[1] stats4 grid stats graphics grDevices utils datasets
[8] methods base

other attached packages:
[1] gridExtra_2.3 sn_2.0.0 VIM_6.1.0 colorspace_2.0-1
[5] car_3.0-10 carData_3.0-4 e1071_1.7-6 xtable_1.8-4
[9] GGally_2.1.1 ggplot2_3.3.3 dplyr_1.0.6 readr_1.4.0

loaded via a namespace (and not attached):
[1] Rcpp_1.0.6 vcd_1.4-8 lattice_0.20-44
[4] class_7.3-19 ps_1.6.0 zoo_1.8-9
[7] assertthat_0.2.1 digest_0.6.27 lmtest_0.9-38
[10] utf8_1.2.1 R6_2.5.0 ranger_0.12.1
[13] cellranger_1.1.0 plyr_1.8.6 MatrixModels_0.5-0
[16] pillar_1.6.1 rlang_0.4.11 curl_4.3.1
[19] readxl_1.3.1 SparseM_1.81 rstudioapi_0.13
[22] data.table_1.14.0 Matrix_1.3-3 labeling_0.4.2
[25] foreign_0.8-81 munsell_0.5.0 proxy_0.4-25
[28] numDeriv_2016.8-1.1 compiler_4.1.0 pkgconfig_2.0.3
[31] mnormt_2.0.2 tmvnsim_1.0-2 nnet_7.3-16
[34] tidyselect_1.1.1 tibble_3.1.2 matrixStats_0.58.0
[37] rio_0.5.26 reshape_0.8.8 fansi_0.4.2
[40] conquer_1.0.2 crayon_1.4.1 laeken_0.5.1
[43] withr_2.4.2 MASS_7.3-54 gtable_0.3.0
[46] lifecycle_1.0.0 DBI_1.1.1 magrittr_2.0.1
[49] scales_1.1.1 zip_2.1.1 cli_2.5.0
[52] stringi_1.6.2 farver_2.1.0 sp_1.4-5
[55] robustbase_0.93-7 ellipsis_0.3.2 generics_0.1.0
[58] vctrs_0.3.8 boot_1.3-28 openxlsx_4.2.3
```

```
[61] RColorBrewer_1.1-2  tools_4.1.0          forcats_0.5.1
[64] glue_1.4.2           DEoptimR_1.0-8       purrr_0.3.4
[67] hms_1.1.0            abind_1.4-5          haven_2.4.1
[70] quantreg_5.85
```