

# 平均回帰とコントロール変数

福井義高

青山学院大学大学院国際マネジメント研究科

令和5年4月19日

---

# 1: 平均回帰

## 平均回帰は統計的必然

## 2: 凡庸への回帰

- Regression towards mediocrity (Francis Galton)
  - 背の高い親からは親より低い子が生まれ、背の低い親からは親より高い子が生まれる
- 経済データでも時間が経つと全体の平均に回帰していくのはよくあること、というより、ほぼそうなる
  - 企業の収益性はその典型例

### 3: 平均回帰する企業の収益性

財務業績の多くの測定値は平均回帰する傾向にあり、ROEもその例外ではない。．．． 企業を最低のROEのグループから最高のROEのグループに分類し、．．． それ以降の5年間でどのように推移したかをグラフに[すると]．．． 最も高いグループはその期間に低下し、最も低いグループではその期間に上昇しているので、平均回帰という概念に一致している。

Lundholm & Sloan (2019, 邦訳141頁)

## 4: 競争で平均回帰?

### ■ 米国のエライ先生によると

端的な答えは競争が原因である。もし企業の資本利益率が高ければ、このことにより、他企業の注目を集めることになる。既存のライバル企業は、その企業の価格よりも安く販売し、さらに新規企業が市場に参入する。初めに高い利益を得ていた企業も競争相手の脅威に応じて自社の価格を切り下げるために、収益性は悪化しROEは低下する。

Lundholm & Sloan (2019, 邦訳142頁)

## 5:別のエライ先生も

ROEが異常に高い(あるいは低い)企業は、利益が減少(あるいは増加)する傾向にある。... ROEやその他の投資利益率のこうした動きは、「平均回帰的(mean-reverting)」と特徴づけることができる。... これは競争の経済学が予言していることそのものである。高いROEが下降しやすいのは、儲かるところには競争が起きやすいことを反映している。低いROEが上昇しやすいのは、資本が非生産的な事業から収益性の高い事業に流れ出すことを反映している。

Palepu et al. (2013, 邦訳258-260頁)

## 6: 実証研究の常識

- 平均回帰するのはROEに限らず、売上高利益率も要するに、ROEと同様、売上高利益率も、競争の圧力により時が経つにつれて「正常な」水準に収束する傾向があるということである。

Palepu et al. (2013, 邦訳261頁)

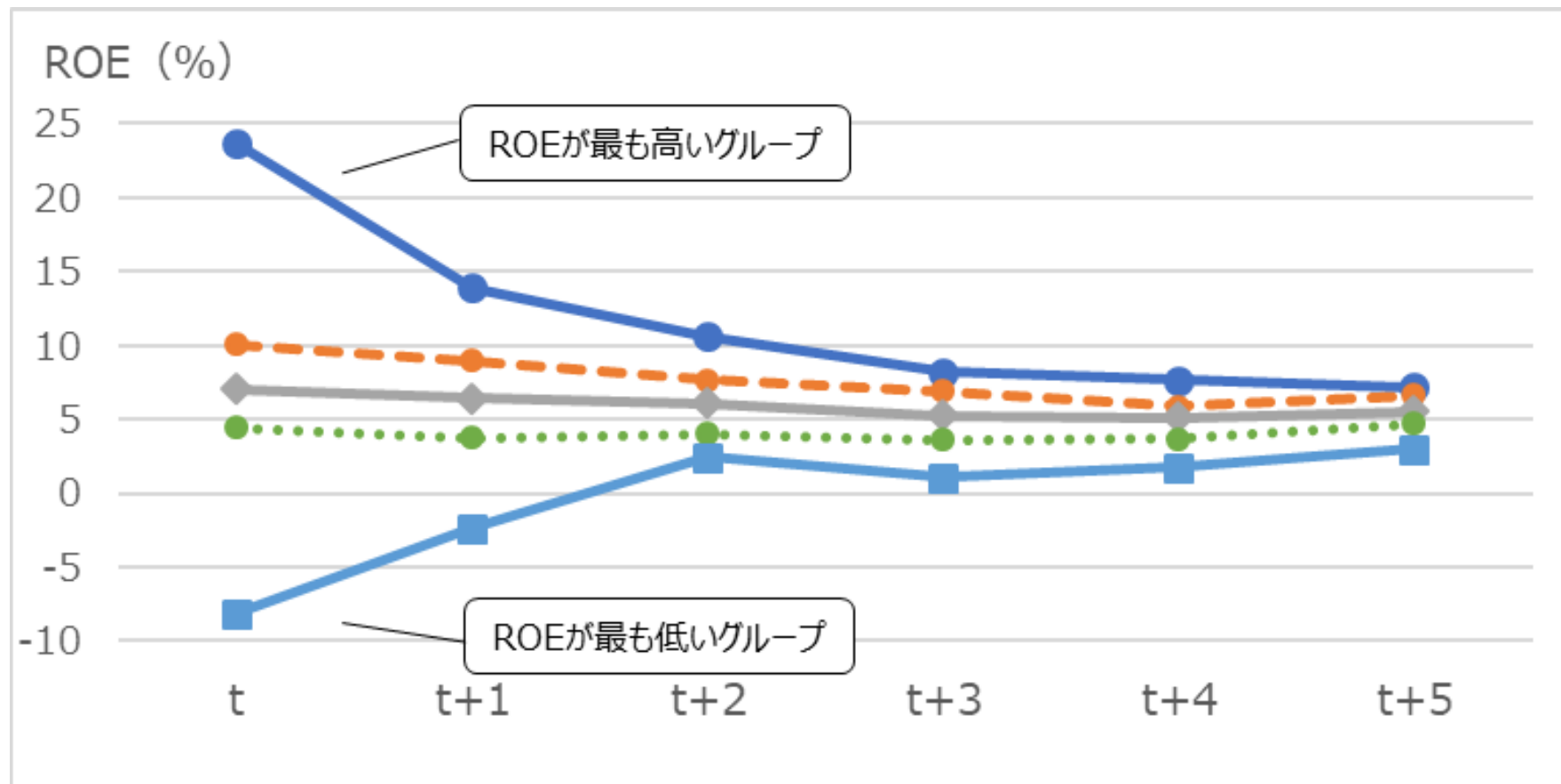
---

## 7: 日本企業はどうなっているのか

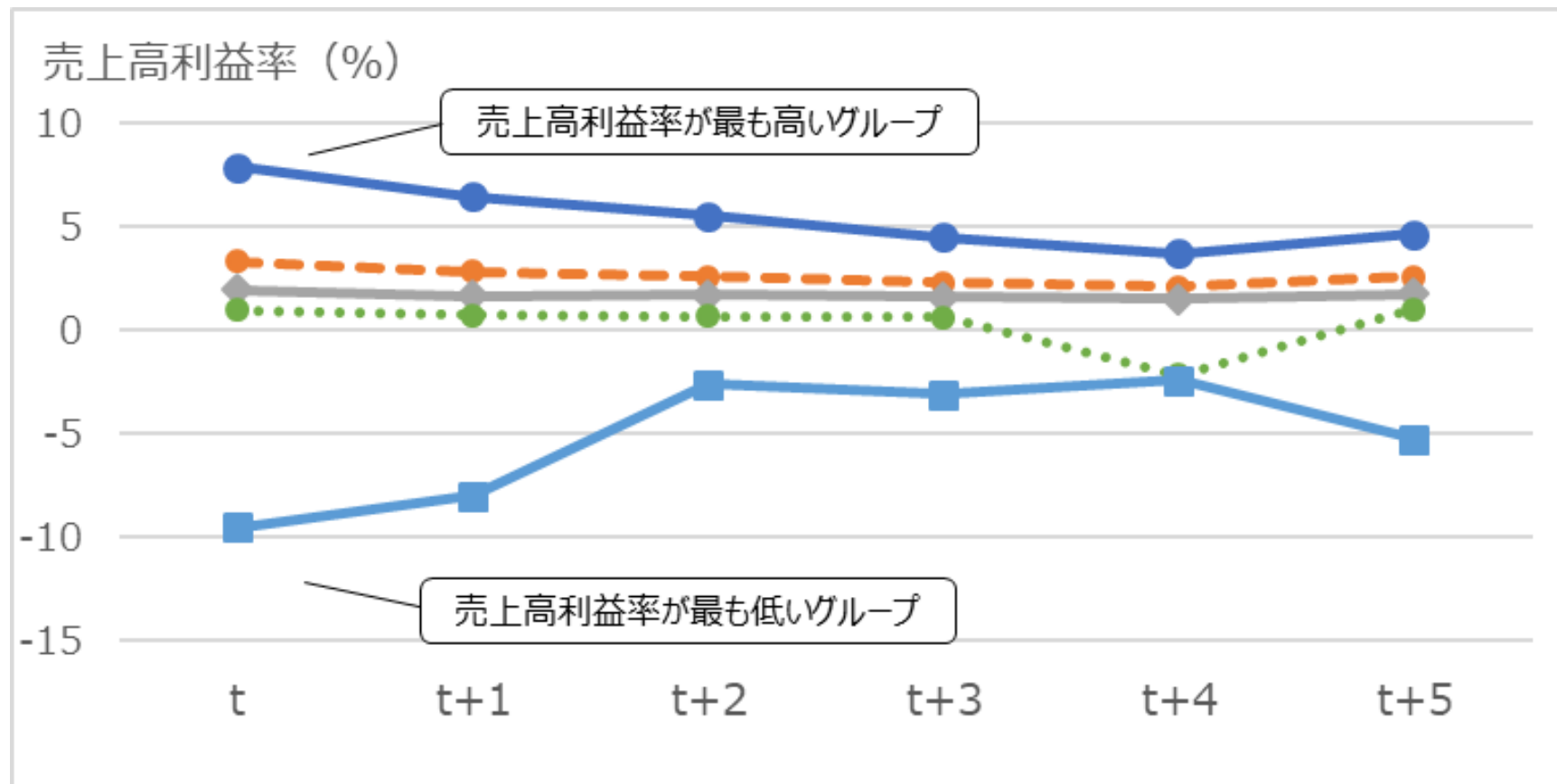
- 1975～2020年のデータで同様のグラフを作ってみると、米国と同じような結果が得られた



## 8. ROEの推移(過去から将来)



## 9. 売上高利益率の推移（過去から将来）



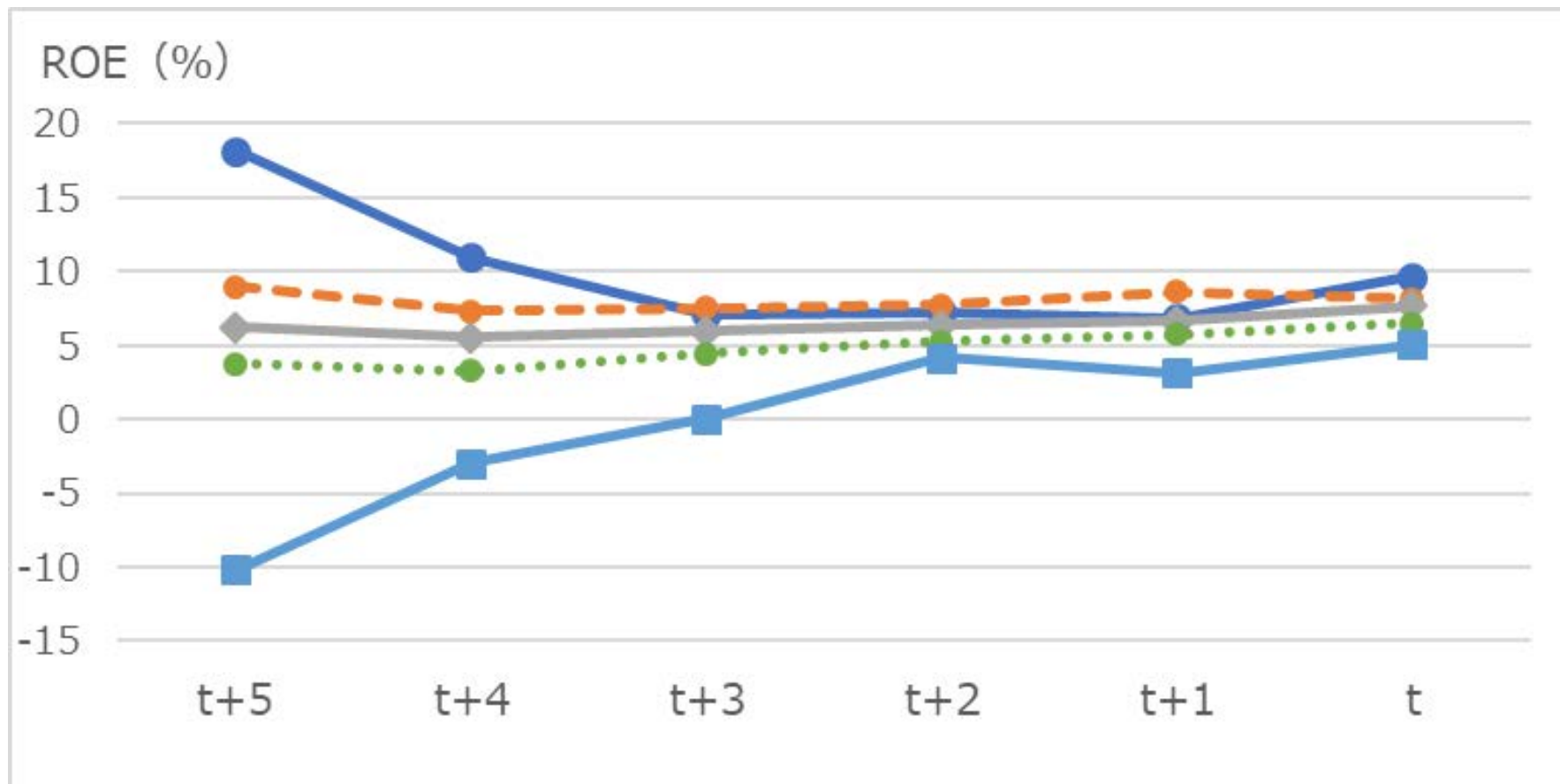
# 10: 市場競争⇒平均回帰は自明？

- 日本でも米国でも、時間が経つにつれ、市場競争によって企業の利益率が市場全体の平均値に収斂するというのは、自明のようにみえる
- しかし、本当に自明なのか、平均回帰という現象に関する必読文献であるにもかかわらず、経済経営実証研究ではほとんど参照されていないCampbell & Kenny (1999)に拠りながら、以下、平均回帰という現象に対する別の見方を提供する

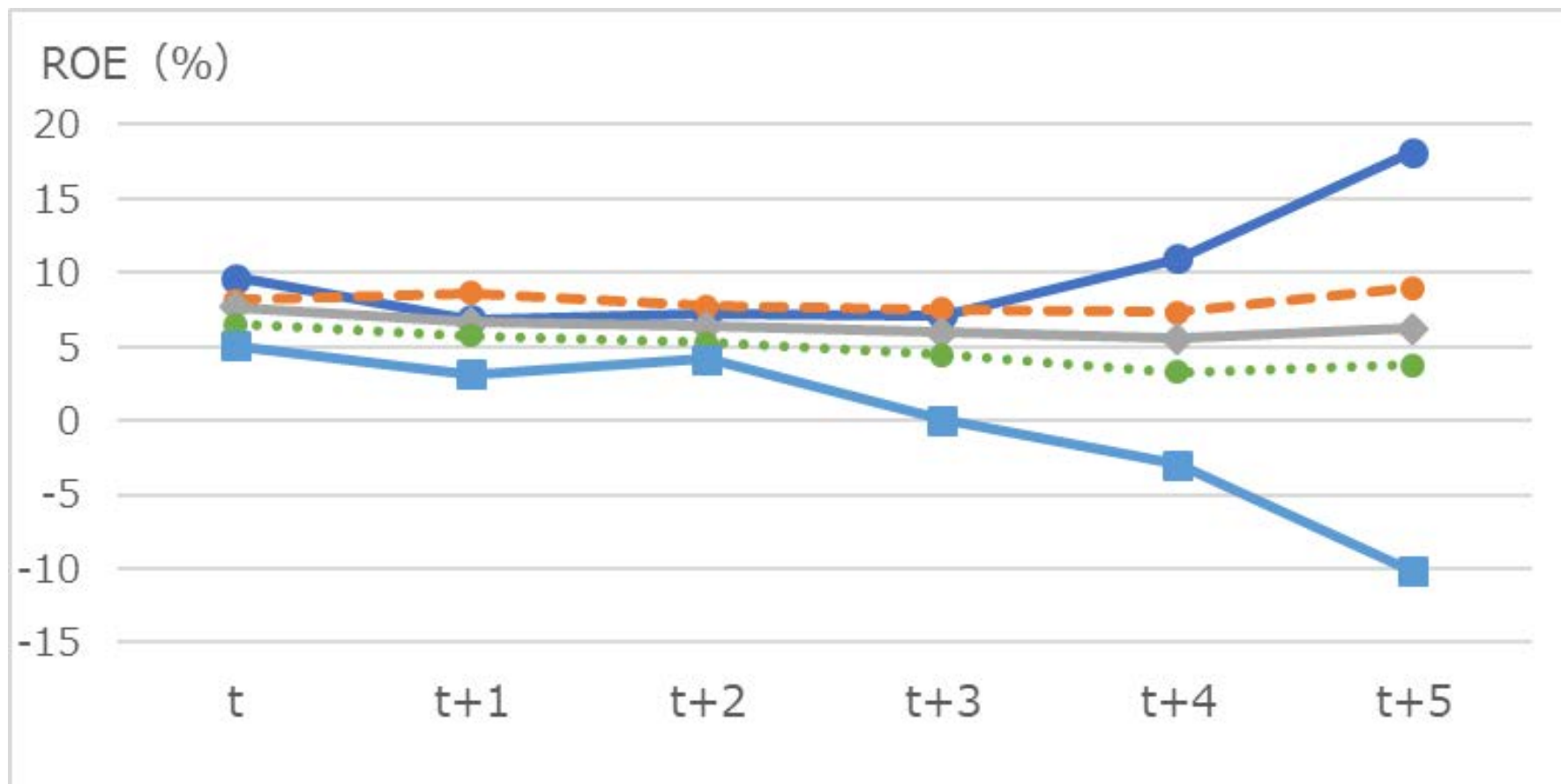
# 11: 後ろ向きにやってみる

- 背の高い親からは親より低い子が生まれ、背の低い親からは親より高い子が生まれるということは、背の高い子の親は子より低く、背の低い子の親は子より高いということ ⇒
- 親から子へと時間を前向きに見ても、子から親へと時間を後ろ向きに見ても平均回帰
- 利益率でも同じことをやってみる

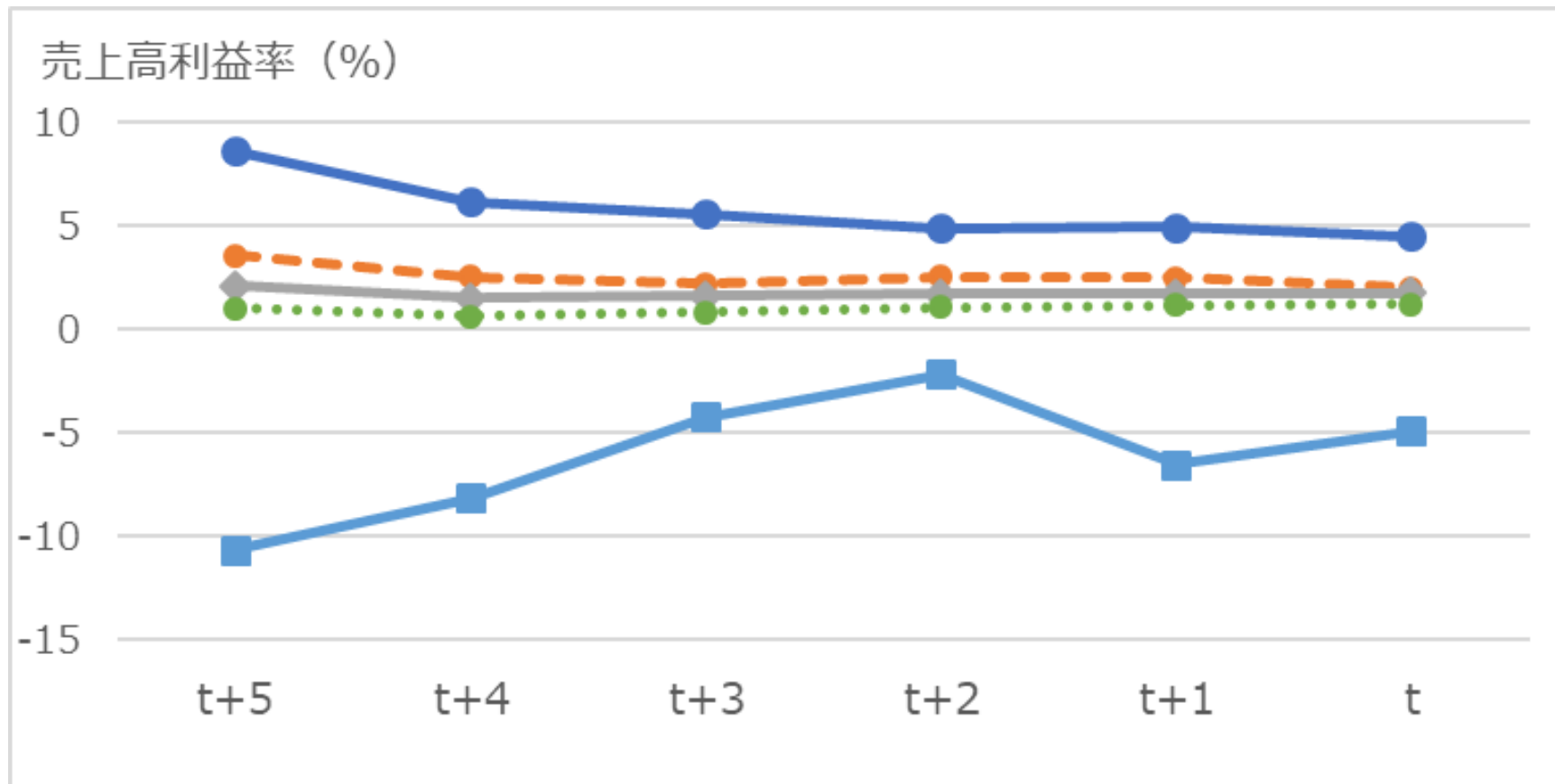
# 12: ROEの推移(将来から過去)



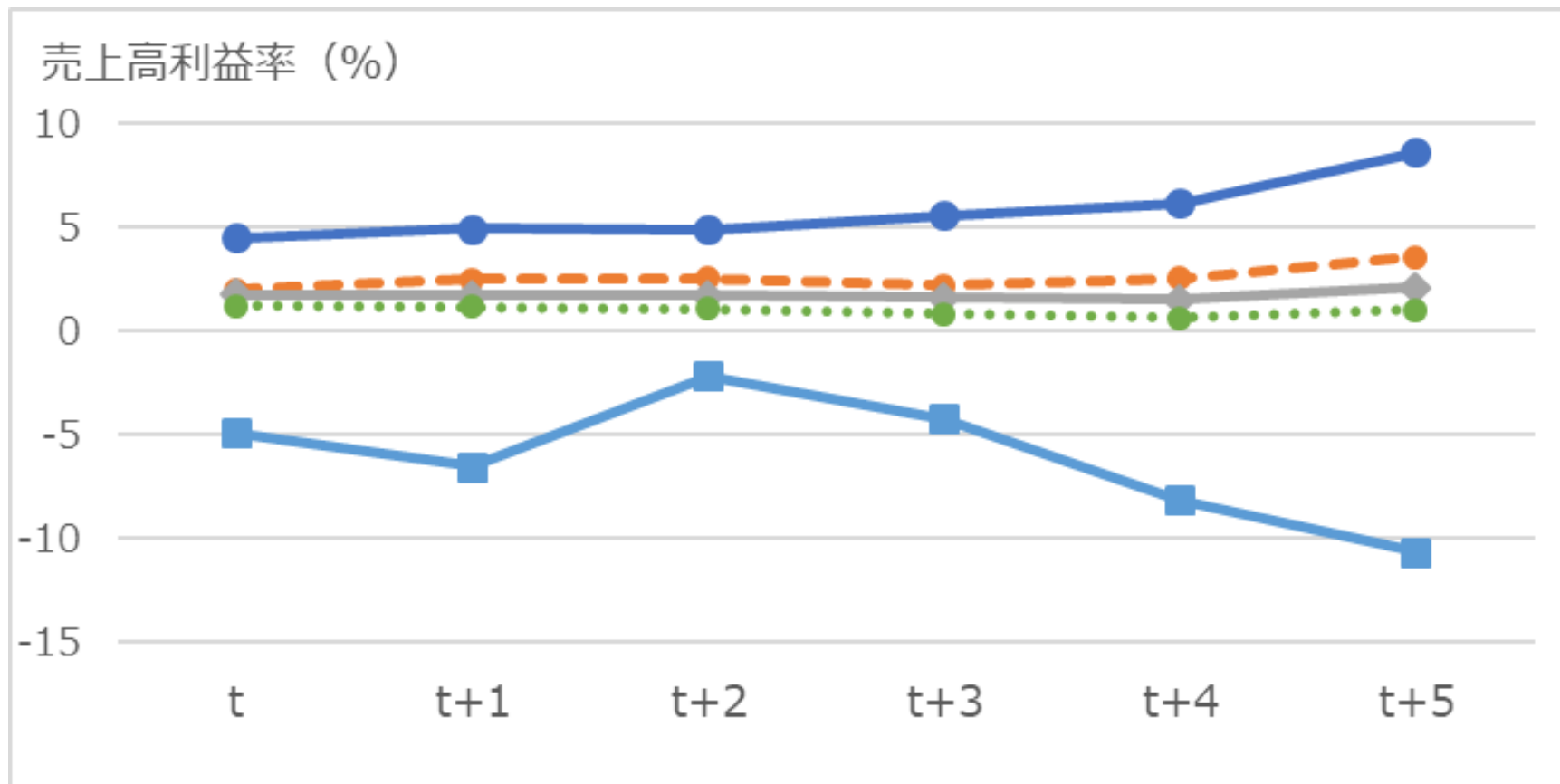
# 13: ROEの推移(時間軸を反転させ、過去から将来)



# 14: 売上高利益率の推移(将来から過去)



# 15: 売上高利益率の推移 (時間軸を反転させ、過去から将来)





## 16: 統計的必然としての平均回帰

- 身長であれ利益率であれ、時系列データの時点毎の数値の散らばりが時期によって変化せず一定であるかぎり、ある時点のデータと別の時点のデータを単回帰する場合、説明変数と被説明変数を入れ替えても、係数は同じ

## 17: 簡単な数学で確認(1)

二つの確率変数 $X$ と $Y$ を標準化(平均ゼロ、分散1)して $x$ と $y$ とし、 $y$ を $x$ で回帰すれば、推計式(条件付期待値)

$$y = bx$$

の係数は

$$b = \hat{\sigma}_{xy} / \hat{\sigma}_x^2 = \hat{\sigma}_{xy}$$

標準化しているので、相関係数(絶対値は1以下)

$$r = \hat{\sigma}_{xy} / (\hat{\sigma}_x \cdot \hat{\sigma}_y) = \hat{\sigma}_{xy}$$

すなわち

$$y = rx$$

## 18: 簡単な数学で確認(2)

逆に $x$ を $y$ で回帰しても、推計式(条件付期待値)

$$x = b'y$$

の係数は

$$b' = \hat{\sigma}_{xy} / \hat{\sigma}_y^2 = \hat{\sigma}_{xy}$$

やはり相関係数

$$r = \hat{\sigma}_{xy} / (\hat{\sigma}_x \cdot \hat{\sigma}_y) = \hat{\sigma}_{xy}$$

と同じ、すなわち

$$x = ry$$

## 19: 簡単な数学で確認(3)

$y$ を $x$ で回帰しても、 $x$ を $y$ で回帰しても、絶対値1以下の係数は同じ

$y$ を $x$ で回帰すると

$$y = bx$$

だったからといって、代数的一次関数ではないので

$$x = (1/b)y$$

とはならない

## 20: 統計的必然の視点から再検討

- 時系列で正の相関があるデータを標準化した場合、回帰を時間軸上、前向きに行っても後ろ向きに行っても、完全に相関(相関係数が1)していない限り、係数は1未満となるので、必ず平均回帰
- 利益率の場合、年度によって分散は大きくかわらないので、標準化しなくても、5分位に振り分けられる企業の構成が前向き回帰と後ろ向き回帰では異なるものの、 $t$ 年のデータを $t-1$ 年のデータで回帰した場合と、逆に $t-1$ 年のデータを $t$ 年のデータで回帰した場合で、1より小さい係数が大きく変わることはない

## 21: 統計的必然性はすでに指摘済み

- 利益率の平均回帰を市場競争の結果と解釈することの問題点は前世紀の終わりに指摘済み
- 米統計学会が発行する一般向けの雑誌 (Smith 1997) で、ノーベル経済学賞受賞者が執筆した教科書 (Sharpe 1985) に、同様のグラフとともに、「最終的には経済的圧力によって異なる企業の利益率と成長率は収斂する」と書かれていることに対して、この平均回帰は経済的現象ではなく統計的現象であり、平均回帰に関する解釈の誤りが繰り返されていると記されている
  - 後年に出た版では、当該箇所削除 (Sharpe et al. 1995)

## 22:「統計的必然としての平均回帰」をデフォルト仮説に

- 平均回帰という現象の背後に何らかの実質的内容(市場競争や遺伝など)があるかもしれないけれど、一般論として、単なる統計的必然としての平均回帰という仮説は、「最も破壊的かつ理にかなったライバル仮説」(the most pernicious plausible rival hypothesis)

Campbell & Kenny (1999, 51頁)

## 23: まずは後ろ向きにやってみる

- 経済経営研究においても否定し難い、「あまりにも頻繁に、平均回帰という統計的事実に根拠なき実質的意味が与えられている」現状

Campbell & Kenny (1999, 19頁)

- 時系列で平均回帰現象を見つけたら、結論を急ぐ前にひとまず後ろ向きに同じことをやってみる
- 前向きでも後ろ向きでも同じような結果が得られたら、統計的必然としての平均回帰を発見、というより確認しただけのことではないかと、疑ってみる必要



---

## 24:コントロール変数

観察データでモデル検証はほぼ不可能

## 25: 観察研究における回帰モデル

- 以下、変数はすべて標準化されていると仮定
- $x$ 、 $y$ 、 $z$ の三つの確率変数が観察されたとする
- この場合、 $y$ を被説明変数とする三つの回帰式あるいは条件付期待値関数はどれも正しい

$$y = \beta_1 x + \varepsilon_x \qquad E(y|x) = \beta_1 x$$

$$y = \beta_2 z + \varepsilon_z \qquad E(y|z) = \beta_2 z$$

$$y = \beta_1' x + \beta_2' z + \varepsilon_{xz} \qquad E(y|x, z) = \beta_1' x + \beta_2' z$$

観察研究で用いられる回帰モデルには、真のモデルとか誤ったモデルという概念が存在しないのである

宮川(2004, 51頁)

## 26:重回帰分析誤用の絶望的状况

- しかし、理論から変数を特定し、モデルが真のデータ生成過程を記述している、あるいはしていないという議論は可能
- ただし、真のモデルを識別することは実験研究より格段に困難

重回帰分析における最も多い誤用は、基準[被説明]変数に対する予測[説明]変数の単純な影響力の指標として偏回帰係数を解釈すること... 研究のプロが書く査読論文にも誤った解釈は多く、その誤りの頻度の惨状は絶望的なほど...

## 27: 予測と理論検証の越え難い溝

「偏回帰係数の正負や有無を基準変数への寄与として単純に解釈できないなら、そもそも重回帰分析という手法は研究の役に立たないのでは？」という声が聞こえてきそう... はい、その指摘は半分当たっています... 重回帰分析は基準変数の予測には役立ちます。でも標準偏回帰係数の単純な解釈ができないのですから、基準変数の学問的な説明にはほとんど役に立ちません

豊田秀樹(2017, iii-iv頁)

## 28: コントロール変数がひとつの場合

被説明変数を $y$ 、ターゲット変数を $x$ 、コントロール変数を $z$ とすると、単回帰

$$E(y|x) = \beta x$$

のターゲット係数 $\beta$ と重回帰

$$E(y|x, z) = \beta_1 x + \beta_2 z$$

のターゲット係数 $\beta_1$ の関係は

$$E(z|x) = \gamma x$$

とすれば、

$$\beta = \beta_1 + \gamma \cdot \beta_2$$

- ターゲット係数はコントロール変数との相関次第

## 29: コントロール変数が複数の場合

- Gordon (1968) を参考にした数値例

ターゲット変数1とコントロール変数2~5の相関係数行列

$$\begin{pmatrix} 1 & 0.8 & 0.2 & 0.2 & 0.2 \\ 0.8 & 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.8 & 0.8 \\ 0.2 & 0.2 & 0.8 & 1 & 0.8 \\ 0.2 & 0.2 & 0.8 & 0.8 & 1 \end{pmatrix}$$

# 30: 被説明変数と説明変数の相関

被説明変数と説明(ターゲット1・コントロール2~5)変数の相関と回帰係数の関係

被説明変数との相関

回帰係数

1	2	3	4	5	1	2	3	4	5
0.60	0.60	0.60	0.60	0.60	0.27	0.27	0.19	0.19	0.19
0.60	0.60	0.55	0.55	0.55	0.28	0.28	0.17	0.17	0.17
0.55	0.55	0.60	0.60	0.60	0.24	0.24	0.19	0.19	0.19
0.55	0.60	0.60	0.60	0.60	0.13	0.38	0.19	0.19	0.19
0.50	0.60	0.60	0.60	0.60	-0.01	0.49	0.19	0.19	0.19

## 31 : 回帰係数はコントロール変数次第

- コントロール変数との相関、変数の数に左右されるターゲット変数の回帰係数

The whole point of *multiple* regression as contrasted with simple regression is to *try* to isolate the effects of the individual regressors, by “controlling” on the others. Still, when orthogonality is absent the concept of the contribution of an individual regressor remains inherently ambiguous.

Goldberger (1964, 201頁)

- 実証分析とは条件付期待値計算と割り切るのも一案



# おまけ:2011年医師国家試験問題43

43 新しく発売された抗菌薬 A の肺炎に対する治療効果を調べるために、新たに入  
院する肺炎患者を対象として、抗菌薬 A を投与した群(A 群)と既存の抗菌薬 B を  
投与した群(B 群)とに割り付けて、治療効果を入院期間で比較検討した。得られた  
結果を表に示す。

	A 群	B 群	P 値
対象者数	198 人	201 人	
入院期間(平均)	8.1 日	9.6 日	0.036

この結果の解釈について正しいのはどれか。

- a A 群は B 群に比べて入院期間が平均で 3.6 % 短い。
- b A 群の入院期間の平均値の誤差は 3.6 % 以内である。
- c A 群の方が B 群よりも入院期間が短くなる確率は 3.6 % である。
- d A 群の 96.4 % の患者は入院期間が B 群の平均入院期間よりも短い。
- e A 群と B 群とで入院期間に差がないのに、誤って差があるとする確率は 3.6 % である。

# 解答なし誤出題

医学界、および久留米大学医学部の数人の重鎮と目される医師に誤出題であることを説明したが、理解してもらえずに愕然とした。解答枝eが正解でないことを理解するには統計的基礎知識を必要とするが、彼らはその知識を持ち合わせていなかったからである。重鎮と目される医師たちとの対話の中で、医・歯・薬学系における統計学講義の貧弱さを思い知り肌寒い思いがした。

柳川(2017)

- ABSは大丈夫ですね

# 主な参考文献

## 平均回帰

福井義高・高橋美穂子(2022)「平均回帰—経済的実質なき統計的必然」『企業会計』75巻1号97-103頁。

Campbell, D. T., and D. A. Kenny. 1999. *A Primer on Regression Artifacts*. Guilford Press.

Galton, F. 1886. Regression towards Mediocrity in Hereditary Stature. *Journal of the Anthropological Institute of Great Britain and Ireland* 15: 246-263.

Lundholm, R., and R. Sloan. 2019. *Equity Valuation and Analysis with e-Val*, Fifth Edition. Kindle Direct Publishing. [深井忠他訳『企業価値評価: eValによる財務分析と評価』(第3版)マグローヒル・エデュケーション]

Palepu, K. G., et al. 2013. *Business Analysis and Valuation*, Fifth Edition, International Edition. South-Western. [斎藤静樹監訳『企業分析入門』東大出版会]

Sharpe, W. F. 1985. *Investments*, Third Edition. Prentice-Hall.

Sharpe, W. F., et al. 1995. *Investments*, Fifth Edition. Prentice-Hall.

Smith, G. 1997. Do Statistics Test Scores Regress toward the Mean? *Chance* 10 (4): 42-45.

## コントロール変数

豊田秀樹編著(2017)『もうひとつの重回帰分析』東京図書。

宮川雅巳(2004)『統計的因果推論』朝倉書店。

Goldberger A. S. 1964. *Econometric Theory*. Wiley.

Gordon, R. A. 1968. Issues in Multiple Regression. *American Journal of Sociology* 73 (5): 592-616.

## おまけ

柳川堯(2017)「 $p$ 値は臨床研究データ解析結果報告に有用な優れたモノサシである」『計量生物学』38巻2号153-161頁。