| | |
|---|---|
| Title | レビューの不正操作の現状と課題 : 実証研究のサーベイ |
| Sub Title | The landscape and challenges of online review manipulation : a survey of empirical studies |
| Author | 坂口, 洋英(Sakaguchi, Hirohide) |
| Publisher | 慶應義塾経済学会 |
| Publication year | 2025 |
| Jtitle | 三田学会雑誌 (Mita journal of economics). Vol.117, No.3 (2025. 2) ,p.451 (185)- 480 (214) |
| JaLC DOI | 10.14991/001.20250201-0185 |
| Abstract | 近年，「フェイクレビュー」と呼ばれる偽のレビューにより，レビューを中心とした評価システムが不正に操作されることが横行している。このレビューの不正操作は，消費者の意思決定を歪めるだけでなく，レビューシステム自体への信頼を毀損させることが懸念されている。そこで，本稿は，レビューの不正操作についての実証研究を中心にサーベイを行う。既存の研究を整理し，その限界を考察することで，今後の研究の方向性についての示唆を提供する。<br>Consumer reviews play significant roles in influencing consumer choices. But, in recent years, review-based evaluation systems are being increasingly manipulated through “fake reviews.” Such manipulation can mislead consumers’purchasing decisions and undermine trust in the review system. However, despite the growing concern, few studies have investigated their nature, impact, and the measures necessary to prevent consumer harm. Therefore, this paper surveys empirical research on review manipulation, incorporating experimental approaches and analysis of observational data. By examining the existing research on online review manipulation and highlighting its limitations, we aim to provide insights for future research directions in this field. |
| Notes | 展望論文 |
| Genre | Journal Article |
| URL | https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00234610-20250201-0185 |

「三田学会雑誌」117 巻 3 号（2025 年 2 月）

# レビューの不正操作の現状と課題：実証研究のサーベイ

坂口洋英 *

# The Landscape and Challenges of Online Review Manipulation: A Survey of Empirical Studies

Hirohide Sakaguchi *

**Abstract:** Consumer reviews play significant roles in influencing consumer choices. But, in recent years, review-based evaluation systems are being increasingly manipulated through "fake reviews." Such manipulation can mislead consumers' purchasing decisions and undermine trust in the review system. However, despite the growing concern, few studies have investigated their nature, impact, and the measures necessary to prevent consumer harm. Therefore, this paper surveys empirical research on review manipulation, incorporating experimental approaches and analysis of observational data. By examining the existing research on online review manipulation and highlighting its limitations, we aim to provide insights for future research directions in this field.

**Key words:** fake reviews, survey, consumer protection, online retail, digital platform

**JEL Classifications:** D18, L81

＊ 敬愛大学経済学部
Faculty of Economics, Keiai University
hhhsakaguchi@gmail.com

## 1　はじめに

消費者レビューおよびそれに基づくレーティングで構成される財の評価システムは，消費者の意思決定において不可欠となっている。Amazon や楽天のような EC サイトでは標準的に搭載されるほか，食べログのような，特定の分野に特化したレビューサイトも人々の日常生活を大きく支えている。

こうした消費者レビューが，財の需要にも大きく影響を与えていることは，様々な調査や実証研究にも支持されている。例えば，消費者庁（2020）は，半数以上の消費者がレビューを参照し，その 7 割以上がレビューを信頼している調査結果を示している。また，Chevalier and Mayzlin（2006）; Chintagunta et al.（2010）; Hollenbeck（2018）などの様々な実証研究において，レビューが消費者の需要に正の影響をもたらす結果が導かれている。

一方で，レビューを中心とした評価システムには，不正に操作される脆弱性が存在する。なんらかの形でレビューを水増しすることで，財の評価を意図的に歪めることが可能であり，いわば，“サクラ” や “ヤラセ” に弱いという欠点をもつ。

実際に近年では，主に Amazon において，このレビューの不正操作[(1)]ともいうべき行為が横行している。ここで投稿されるのが「フェイクレビュー」と呼ばれるもので，売り手が自作自演や金銭を介して消費者に投稿させた偽のレビューである。これは，本来利害関係のない消費者が投稿するべき “本物” のレビューとは異なり，利害関係が存在する点で “フェイク” といえるレビューである[(2)]。

このようなレビューの不正操作は，ステルスマーケティングの一種であり，競争上の公正性を損なう行為である。商品の実態と乖離したフェイクレビューが氾濫することで，消費者が誤った判断を下す可能性が高まるばかりでなく，プラットフォームそのものへの信頼も損なわれる恐れがある。

各国ではこれを脅威と見なし，規制が進行している。米国では FTC（連邦取引委員会）が規制を強化し，EU では DSA（Digital Service Act；デジタルサービス法）が発効している。我が国においても，2023 年 10 月に景品表示法の不当表示として明確に違法とされたが，法整備の遅れ[(3)]や，その有効性や強制力の弱さが指摘されている（カライスコス, 2023）。

しかしながら，規制の進展にもかかわらず，数多くの事例からフェイクレビューの氾濫が示唆さ

---

(1)　本稿では，レビューシステムの脆弱性を突き，評価を歪める行為を，Mayzlin et al.（2014）の review manipulation という表記に倣い，レビューの不正操作と呼称する。

(2)　フェイクレビューがフェイクとされるのは，消費者と売り手の利害関係によるものであり，必ずしも商品の実態と乖離しているとは限らない。仮に，高品質な財の売り手がレビューの執筆を依頼した場合でも，それはフェイクレビューとなる。

(3)　長らく 2012 年の景品表示法ガイドラインの改定にとどまり，明確な違法行為としては認定されていなかった。

図 1：SNS（LINE）におけるフェイクレビュー募集


れる[(4)]。米国においては，フェイクレビュー検出サービスである Fakespot が，2020 年の 3 月から 9 月に投稿されたレビューのうち 40%ものレビューがフェイクである疑いを指摘している[(5)]ほか，2022 年には Amazon がフェイクレビュー仲介業者を提訴している[(6)]。日本国内においても，フェイクレビュー検知サービスである「サクラチェッカー」が人気を博すほか，SNS 上でフェイクレビューの募集が依然として行われている。図 1 は，SNS 上での実際の募集の例である。

---

(4) レビューが実際にフェイクか否かの判別は困難であるため，フェイクレビューがどの程度割合を占めるかについては，明確なエビデンスが存在しない。

(5) https://www.bloomberg.com/news/articles/2020-10-19/amazon-fake-reviews-reach-holiday-season-levels-during-pandemic?embedded-checkout=true

(6) https://www.sankei.com/article/20220223-4LX3CNKVBVM3TAPV3XRHRL37G4/

以上に述べた背景から，レビューの不正操作は今なお脅威となっており，政府による実効性のある規制や，プラットフォーマーによる対策の強化が求められているといえる。

ただし，いくつかの点について考慮する必要がある。まず，レビューの不正操作は必ずしも害をもたらさないという点である。仮に，高品質な財の売り手が不正を行っているケースでは，フェイクレビューは実態と乖離しないため消費者被害につながりにくい。むしろ，品質のシグナルとして消費者にとって有益となる可能性すらある。こうしたケースは直観的ではないものの，いくつかの理論研究においても支持されている（Dellarocas, 2006; Yasui, 2020）。

さらに，レビューシステムへの介入は，レビューの投稿をも減少させるトレードオフを生じさせる可能性も高い。例えば，レビューの投稿に厳しい条件を課す場合，フェイクレビューでない真のレビューも減少してしまう。

したがって，レビューの不正操作は喫緊に対処すべき脅威でありながらも，規制には慎重な判断が求められる。有効な対策を講じるにあたっては，まずその実態を明らかにする必要がある。そのうえで，現状においてどれだけの被害が発生しているかを把握し，さらにどのような対策が被害抑制に効果的かを検証することが肝要である。このような背景から，明らかにすべき実証上のリサーチクエスチョンは，以下の 3 点として大別できる。

第一に，不正行為の詳細なパターンを明らかにする必要がある。具体的には，どのようなレビューがフェイクであり，どのような企業や売り手が不正に関与しているかなどが求められる。先述したように，不正に関与している売り手の属性や商品の品質の把握は，消費者被害に大きく影響を与えるため，特に重要となる。

第二に，消費者やプラットフォームが，レビューの不正操作によってどのような影響を受けるかを調査する必要がある。消費者がどの程度フェイクレビューを判別できるか，不正による被害がどれだけ生じているかについて明らかにするのは無論として，フェイクレビューの下で消費者の行動がどのように変化するかということも，消費者やプラットフォームの被害が生じる機序やメカニズムの解明にあたり必要となる。さらに，公平性を重視する立場からは，消費者の属性とフェイクレビューから受ける影響の関係が政策決定において大きなファクターとなる。例えば，学歴や所得の低い消費者ほどフェイクレビューに欺かれやすい場合，仮にレビューの不正操作が総余剰に与える影響が正であっても，規制は正当化されうる。

第三に，どのような介入やレビューシステム・レーティングシステムが，レビューの不正操作の被害を抑制できるかを明らかにする必要がある。新しい施策を勘案することは困難であるものの，既存の施策や現状のレビューシステムの設計について効果を検証することでも，今後の施策の方向性を決定づける点で有益である。

しかしながら，先行研究の蓄積は，これらのリサーチクエスチョンに十分に答えられる水準に達していない。特に，第三のリサーチクエスチョンにあたる，施策・制度の検証については蓄積が不

十分である。

そこで，本稿では，こうしたレビューの不正操作に関連する研究について，実証研究を中心にサーベイし，現在の課題と今後の研究の方向性を考察する。

まず第 2 節では，関連する理論研究について概説する。第 3 節では，観察データを用いた実証研究について整理し，代表的な実証研究である Mayzlin et al.（2014）と Luca and Zervas（2016），He et al.（2022b）について紹介する。第 4 節では，近年増加傾向にある，実験データを用いた実証分析について述べたうえで，Akesson et al.（2023）と Gandhi et al.（2024）を紹介する。第 5 節ではその他の関連研究について述べる。第 6 節では，既存研究の限界を整理し，今後の研究の展望について述べる。

## 2　理論研究

レビューの不正操作に関連する理論研究は，Nelson（1970）や Milgrom and Roberts（1986）のシグナリングモデルの文脈に置かれる。彼らは，シグナリングモデルによる広告の分析において，高品質な商品を供給する企業だけが広告を行い，低品質な商品を供給する企業は広告を行わない分離均衡が存在しうることを示している。

レビューの不正行為の問題がモデル化される際には，ステルスマーケティングの一環として，こうしたシグナリングモデルが拡張される。ここで，フェイクレビューやそれに類するものは，企業による広告とは別のプロモーションの手段として行われる。さらに，これらとは別に，他の消費者による真のレビューが存在する。

消費者は，レビューをシグナルとして受けとって行動するものの，消費者はフェイクレビューなどの偽情報と，真のレビューなどの正しい情報を判別することができず，どちらのレビューがどれだけの量存在するかを観察できない。一方で，真のレビューとフェイクレビューの両方が存在することについては理解しており，後者の存在を考慮して財の選択を行う。企業はこうした消費者を想定して，フェイクレビューの量などを調整する。既存の理論研究はこのようなモデルを用い，均衡においてどのような主体が不正を行い，どのような結果をもたらすか，その結果はどのような要因に影響されるのかについて分析している。

Mayzlin（2006）は，企業のステルスマーケティングと純粋な消費者口コミによって消費者の需要がどのように影響されるかについての理論モデルを構築し，分析を行っている。このモデルにおいては，二企業 $A$ と $B$ が，互いに自社の財が他社の財より優秀だという口コミを流す。同時に，財の品質を見分けられる消費者 $D$ が，どちらの財が優れているかについて正しい口コミを流す。その後，財の品質の判別ができない消費者 $C$ が，ステルスマーケティングも含めた風評に依存して選択を行う。この際，$C$ は企業による口コミの操作を考慮し，これらを割り引いて口コミから品質を推

測する。企業は，$C$ の選択を想定して[(7)]，どれだけの量の口コミを流すかを決定する。

均衡では，低品質な商品を供給する企業の方が，多くの口コミを流す。また，ステルスマーケティングによる厚生損失は二つの財の品質の差や財の価格が大きいほど増加し，消費者 $D$ の数が大きくなるほど減少することが示されている。

一方で，Dellarocas (2006)は，高品質な財を供給する企業ほどより多く不正を行う均衡が存在しうることを，別の理論モデルにより示している。このモデルでは，質 $q$ の財を供給する独占企業の質が，フェイクレビューを投稿することで，質についてのシグナル $y$ を操作する。消費者は，$\theta = E[q|y]$ をもとに商品の質を判断し，需要は消費者の効用関数における質についてのコンポーネント $f(\theta)$ で決定される。

$q$ が高い高品質な財を供給する企業ほどより多く不正を行う均衡と，逆に $q$ が低い低品質な財を供給する企業ほど多く不正を行う均衡の両方が存在しうることが示されており，前者のパターンはレビューの不正操作が消費者に有益であることを示唆する。二つの均衡のうちどちらが成立するかは，$f(\theta)$ に依存する。

近年では，Yasui (2020)が，レビューシステムの特徴を反映した，より現実のレビューの不正操作の問題に対応した無限期間の連続時間モデルによる分析を行っている。均衡において質の高い商品の売り手ほど不正を行いやすい結果を示すだけでなく，フィルタリングの強化など，施策がもたらす影響についても明らかにされている。

当研究のモデルでは，1 つの売り手が，各時点 $t$ において，財を $q$ 個販売し[(8)]，フェイクレビューを $F_t$ 投入する。財の質 $\theta_t$ と，商品のレーティング $Y_t$ に対し，消費者は，レーティングに基づいた質の期待値として，レピュテーション $M_t \equiv E[\theta_t|Y_t]$ を形成する[(9)]。この際，消費者はフェイクレビューによりレーティングが歪められていることを認識しており，商品のレーティングを割り引いて商品の質を評価する。

$\theta_t$ は平均回帰過程に従う一方で，$Y_t$ は，過去のレーティングや，$\theta_t$ と販売量 $q$ により決定される本物のレビューの投稿のほかに，売り手の投稿するフェイクレビュー $F_t$ により変化する。売り手は，これらを考慮し，長期的な利潤最大化を目標に，$F_t$ の量を決定する[(10)]。

均衡では，$\theta_t$ が高いほど，$F_t$ が増加する。これは，$\theta_t$ が高いほど，将来的なレピュテーション $M_t$ が大きくなりやすく，$F_t$ を増加させることによる将来の利益が大きいことによる。一方で，$M_t$ が高い売り手[(11)]は，フェイクレビューのコストが大きくなり，$F_t$ が減少する。レーティングの情報

---

(7) $D$ の選択は企業の利潤に含まれず，口コミを介してのみ利潤に影響を与える。
(8) 販売量は固定。
(9) 均衡価格 $p_t$ は，このレピュテーションに等しくなる ($p_t = M_t$)。
(10) $t$ 時点でフェイクレビューの限界費用は，$M_t$ と $F_t$ の増加関数となる。
(11) レーティング $Y_t$ が高い売り手といえる。

量[12]という点で，規範的分析も行われており，フェイクレビューへのフィルタリングを強化するとレーティングの情報量が減少する可能性や，古いレビューの比重を上げたレーティングシステムが情報量を向上させうることも示されている。

この結果をベースラインに，ナイーブなタイプの消費者が混在するケースについても検証される。彼らは，売り手がフェイクレビューを利用してレーティングを歪めることを想定せず，レーティングをそのまま受け取り，商品の質を判断する。

均衡においては，ベースラインと同様に，$\theta_t$ や $M_t$ が高いほど，$F_t$ が大きくなる。ただし，フィルタリングの強化は，$\theta_t$ と $Y_t$ のバイアスを減少させ，ナイーブな消費者を利する結果となっており，消費者保護とレーティングシステムの情報量はトレードオフであることが示唆される。

これらの理論研究は，レビューの不正操作とその影響に関して，重要な知見を提供する。第一に，必ずしも低品質な財を供給する企業が不正を行うのではなく，Dellarocas（2006）や Yasui（2020）が支持するように，高品質な財を供給する企業が不正を行うような均衡が存在しうることである。こうしたケースでは，レビューの不正操作は有害とは限らず，消費者にとって有益なシグナルとなる可能性がある。消費者余剰に対する影響を，実証研究により定量的に評価することが求められる。

第二に，消費者のタイプの分布や効用関数が，レビューの不正操作の被害に影響を与える点である。Mayzlin（2006）では，財の品質を正しく見分けられる消費者が多いほど被害が減少し，Dellarocas（2006）では，消費者の商品の質への効用関数によって不正を行う企業のタイプが変化した。さらに，Yasui（2020）の結果は，ナイーブなタイプの消費者が多い場合，消費者保護の観点から規制が好ましい可能性を示している。消費者が実際にどれだけ騙されやすいのか，騙されやすい消費者はどの程度の割合なのかなど，フェイクレビューに対しての消費者異質性と分布について明らかにすることが必要となる。

## 3　観察データを用いた実証研究

### 3.1　概観

レビューの不正操作にについて，観察データを用いた実証研究は，米国ホテルレビューサイトにおける不正操作を対象とした Mayzlin et al.（2014）とボストンのレストランのレビューを分析した Luca and Zervas（2016）があげられる。近年では，Amazon（Amazon.com）を対象とした He et al.（2022b）が存在する。

レビューの不正操作に関して，実証面で常に付きまとう課題として，レビューがフェイクか否かが判別できないという問題がある。こうしたフェイクの識別について，Mayzlin et al.（2014）と Luca

（12）$Y_t$ と $\theta_t$ の相関係数を情報量の評価に用いる。

and Zervas (2016)は，フェイクレビューを直接観察することはできず，プロキシーを用いるアプローチをとっている。一方で，He et al. (2022b)は，Facebook のフェイクレビュー依頼コミュニティの調査を通じ，不正操作が行われている製品をある程度把握することに成功している。

Mayzlin et al. (2014)と Luca and Zervas (2016)は，どのような店舗や企業が不正に関与しているかに焦点を当てており，不正操作がもたらす影響を調査するには至っていない。

一方，He et al. (2022b)は，どのような不正に関与している商品かを明らかにしたうえで，間接的ながらも，不正が売上に与える影響や，消費者に与える被害についても分析を行っている。

### 3.2 Mayzlin et al. (2014)

Mayzlin et al. (2014)は，Tripadvisor と Expedia における 2011 年 10 月時点での米国ホテルのレビューを対象に，どのような経営形態やオーナーシップのホテルがレビューの不正操作を行っているかについて分析している。両サイト上でも，レビューがフェイクか否かは判別できない。そこで，彼ら[(13)]は2 つのレビューシステムの違いを利用している。

**プロキシー**

当時，Tripadvisor と Expedia は同じ経営母体であったものの，異なるレビューシステムを有していた。Expedia では同サイトを通じて予約した消費者のみがレビューを行うことができたのに対し，Tripadvisor ではどんな消費者でもレビューを行うことができた。そのため，Tripadvisor におけるフェイクレビューの投稿のコストは相対的に低く，ホテル経営者がレビューの不正操作を行う場合，Expedia よりも，Tripadvisor にフェイクレビューを投稿する可能性が高い。

このような背景から，Mayzlin et al. (2014)は，Tripadvisor と Expedia におけるレビューの分布の差を，どれだけ不正行為が行われているかについてのプロキシーとして扱っている。これは以下の(1)で表される。

$$Y_{ij}^{N} = \frac{NstarReviews_{ij}^{TA}}{TotalReviews_{ij}^{TA}} - \frac{NstarReviews_{ij}^{Exp}}{TotalReviews_{ij}^{Exp}} \tag{1}$$

ここで，$i$ はホテルを，$j$ は地域を意味する。$TA$ は Tripadvisor，$Exp$ は Expedia を示す。$Nstar\ Reviews$ は星 $N$ のレビューの数を，$TotalReviews$ はレビューの合計本数となる。そのため，右辺第一項は Tripadvisor における $N$ 点レビューの割合であり，第二項は Expedia における $N$ 点レビューの割合となる。

自社を高評価に装うなら，Tripadvisor に対して高評価のフェイクレビューを相対的に多く投稿す

(13) 以後，論文の複数の筆者は全て彼らで統一する。

る。この場合，$N$ が 4 や 5 のときに右辺の第一項が，第二項と比べて大きくなり，$Y_{ij}^N$ は大きくなる。一方，競合他社から攻撃が行われる場合，Tripadvisor において低評価のレビューが相対的に多く投稿されることになる。この場合，$N$ が 1 や 2 のときに $Y_{ij}^N$ が大きくなる。

**モデル**

彼らは，Tripadvisor と Expedia のレビューとホテルの属性に関するクロスセクションデータを用い，(1)のプロキシー $Y_{ij}^N$ を，ホテルの属性に回帰すること[14]で，不正操作を行いやすい属性を調査している。モデルは以下の(2)で表現される。

$$Y_{ij}^N = X_{ij}B_1 + OwnAF_{ij}B_2 + Nei_{ij}B_3 + NeiOwnAF_{ij}B_4 + \gamma_j + \epsilon_{ij} \tag{2}$$

ここで，$X_{ij}$ はレーティング[15]や施設の有無などのホテルの属性である。$OwnAF_{ij}$ には，ホテルの保有と所属に関する変数が含まれ，ホテルが独立系（非チェーン）であることを示すダミー変数と，ホテルが複数のホテルを所有するオーナー（マルチユニットオーナー）に所有されているか否かを示すダミー変数である[16]。$Nei_{ij}$ は近隣にホテルがあるか否かを意味するダミー変数，$NeiOwnAf_{ij}$ は近隣のホテルについてのホテルの保有と所属に関する変数であり，非チェーンか否か，マルチユニットオーナーに所有されている否かを示すダミー変数である。$\gamma_j$ は地域（市）の固定効果であり，$\epsilon_{ij}$ は誤差項である。

主な関心の対象は，$OwnAF_{ij}$ や $NeiOwnAf_{ij}$ といった変数であり，これらはレビューの不正のインセンティブに関わってくる。例えば，マルチユニットオーナーが所有するホテルの場合，不正操作が露呈した場合の風評へのダメージは，他に所有するホテルに波及するため，フェイクレビューのコストが大きい。一方で，非チェーン系のホテルは，ブランドイメージの棄損を恐れる必要がないために，フェイクレビューのコストが小さい。また，競合他社が同じ地域に存在するホテルは，競合他社フェイクレビューでその競合相手を攻撃することで，自社の需要を増大させることができる。

**推定結果**

低評価のフェイクレビューによる攻撃のプロキシーとして，1 点と 2 点の低評価のレビューの割合の差を被説明変数とした場合 ($Y_{ij}^N|_{N=\in\{1,2\}}$) の推定の結果，近隣に他のホテルがあるホテルは有意に $Y_{ij}^N$ が約 3%ポイント分増加する。この近隣ホテルが非チェーンの独立系ホテルである場合，さ

---

(14) 論文中では差分の差分法（DID）を用いていると主張されているが，一般的な DID とは異なり，クロスセクションデータを用い得点分布の差をホテルの属性に回帰している。

(15) レビューによるレーティングとは異なる，3 つ星ホテルや 5 つ星ホテルといった，星の数によるレーティング。

(16) これらは排他的ではない。非チェーン系ホテルだがマルチユニットオーナーに所有されるケースも，チェーン系だがほかにホテルを所有しないオーナーに所有されるケースもある。

らに有意に 1.7%ポイント分大きくなる。ただし，近隣ホテルがマルチユニットオーナーにより所有されている場合，2.5%ポイント分少なくなる。これは，近隣地域に競合相手がいる場合，攻撃的なフェイクレビューを投稿しやすい可能性を意味する。加えて，競合相手が非チェーンの独立系ホテルだと攻撃が激化するが，競合ホテルがマルチユニットオーナー所有のものだと攻撃が緩和される。

次にフェイクレビューによる高評価の偽装について，5 点の高評価レビューの割合の差 ($Y_{ij}^N|_{N=5}$) を被説明変数とすると，ホテルが独立系の場合，$Y_{ij}^N$ が有意に 2.4%ポイント分増加する。マルチユニットオーナーにより所有されるホテルの場合は，有意に 3.1%ポイント分少なくなる。これは，独立系ホテルほどフェイクレビューにより自社の評価を高く装う傾向を示している。逆にマルチユニットオーナーは自社のレビューについて不正に操作しにくい傾向がみられる。他方で，低評価のケースとは対照的に近隣のホテルの存在は有意な影響を与えているとはいえない結果になった。自社への正のフェイクレビューは，競合他社の存在に関係なく行われている可能性がある。

以上の結果からは，ホテルを 1 つしかもたない非チェーンのホテルほど，フェイクレビューにより，他社を攻撃しやすく，自社の評価を高く装うことがうかがえる。ほかにホテルを保有しておらず，チェーンに所属していないため，ブランドイメージを棄損しにくく，フェイクレビューのコストが低い。これは，経済的インセンティブと整合的であるといえる。

Mayzlin et al.（2014）の結果は，レビューの不正操作がもたらす被害に関して肝要となる，企業が供給する財の品質[(17)]については直接的なエビデンスを示さない。ただし，非チェーンかつ個人経営のホテルは，必ずしも質が低いとはいえないものの，非ブランドのホテルと解釈可能である。このことは，近年の Amazon におけるレビューの不正操作について，非ブランド系やマーケットプレイスで販売される商品に対しフェイクレビューが投稿されている可能性が高いことを示唆する。

### 3.3 Luca and Zervas（2016）

Luca and Zervas（2016）は，米国のレストランレビューサイト Yelp に 2004 年から 2012 年にかけて投稿されたレビューを対象に，パネルデータを構築し，レビューの不正操作がどのような背景で行われているかについて分析している。パネルデータを用いることで，Mayzlin et al.（2014）と比べ，レビューの時系列に沿った変化を観察することができているといえる。しかしながら，同様に，レビューがフェイクか否かを判別できない問題を抱えている。彼らは異なるアプローチでこの問題に対処しており，レビューが Yelp にフィルタリングされたか否かをフェイクのプロキシーに利用する。

---

（17）ホテルのレーティングはコントロール変数として用いられているものの，レーティングが高いほど 5 点の高評価レビューの割合の差は有意に低くなる。また，レストランやコンベンションセンターのような施設があるほど，同じく高評価の割合の差を減少させる。

**プロキシー**

Yelp がフィルタリングしたレビューを，フェイクレビューのプロキシーとして扱う場合，フィルタリングアルゴリズムを分析しているという懸念が指摘できる。ただし，Yelp によるおとり捜査を用いた調査結果は，実際にフェイクレビューを利用している企業のレビューは高い確率でフィルタリングされていることを示す。具体的には，不正操作を行っていると判断された企業のレビューのフィルタリングされる割合は 79%であり，ボストンのレストランの平均である 19%と比べて極めて大きい。そのため，フィルタリングされたか否かをフェイクのプロキシーとして用いる妥当性は担保されているといえる。

まず，フィルタリングの傾向を把握するために，線形確率モデルにより推定を行っている。これは以下(3)のようなモデルとなる。

$$Fliterd_{ik} = b_i + z'_{ik}\beta + \epsilon_{ik} \tag{3}$$

ここで，$i$ はレストラン，$k$ はレビューを示す。$z_{ik}$ はレビューの属性に関する変数のベクトル，$b_i$ はレストランの固定効果となる。$Fliterd_{ik}$ はレビューが削除された場合 1 となるダミー変数である。

推定の結果，星 5 や星 1 のレビューは星 3 のレビューと比べ有意にフィルタリングされやすい傾向が示された。レーティングを操作するには中程度の評価を行う意味が薄いことと整合的である。また，長いレビューや，他に多くのレビューを投稿しているユーザーによるレビューは有意に削除されにくい傾向がみられた。

**モデル**

メインの分析にあたっては，フィルタリングされた高評価レビューや低評価レビューの投稿数をプロキシーとして用いる。これを被説明変数として，レビューの不正操作を行うインセンティブに関わる変数に対して，どのように反応するかを，パネルデータを用いた固定効果モデルにより推定している。この際，フィルタリングのプロセスをモデル化し，そこから誘導型モデルを導くことで，プロキシーを用いた分析を可能としている。これは以下のようになる。

$t$ 期のレストラン $i$ に対して投稿されるフェイクレビューの本数 $f^*_{it}$ を以下のように仮定する。

$$f^*_{it} = x'_{ij}\beta + b_i + \mu_t + \epsilon_{it} \tag{4}$$

$x_{ij}$ は時間で変動するレビューの不正操作のインセンティブに関わる変数のベクトルであり，$t-1$ 期の得点分布などや，競合レストランの出現といった変数が用いられる。$b_i$ はレストランの固定効果であり，$\mu_t$ は $t$ 期の固定効果，$\epsilon_{it}$ は誤差項となる。$f^*_{it}$ は観察できない潜在変数となり，観察できるのは各レビュー $k$ のフィルタリングの有無となる。これは(5)となる。

$$f_{ikt} = \alpha_0(1 - f^*_{ikt}) + (\alpha_0 + \alpha_1)f^*_{itk} + u_{itk} \tag{5}$$

$f_{ikt}$ はレビュー $k$ がフィルタリングされた場合 1 となる変数であり，$f^*_{ikt}$ はレビュー $k$ がフェイクの場合 1 となる観察できない潜在変数である。$u_{itk}$ は誤差項となる。ここで，$\alpha_0$ はレビューがフェイクでないものにもフィルタリングしてしまう偽陽性率，$\alpha_0 + \alpha_1$ はレビューがフェイクであり実際にフィルタリングする真陽性率となる。これらについて，$\alpha_0 \in [0, 1]$ と， $\alpha_1 \in (0, 1 - \alpha_0]$ を仮定する。これは，真陽性率が偽陽性率より大きいことを意味する。この仮定は，先述した Yelp の調査結果と整合的である。

$f_{ikt}$ の和をとることで，フィルタリングされたレビューの数 $f_{it}$ となる。これは以下の(6)となる。

$$\begin{aligned} f_{it} \equiv \sum\nolimits_k f_{itk} &= \sum\nolimits_k [\alpha_0(1 - f^*_{itk}) + (\alpha_0 + \alpha_1)f^*_{itk} + u_{itk}] \\ &= \alpha_0 n_{it} + \alpha_1 f^*_{it} + u_{it} \end{aligned} \tag{6}$$

ここで，$n_{it}$ は $i$ の $t$ 期のレビュー投稿数，$u_{it} = \sum_k u_{ikt}$，$f^*_{it} = \sum_k f^*_{ikt}$ となる。(4)を代入することで，以下の(7)が得られる。

$$f_{it} = \alpha_0 n_{it} + \alpha_1(x'_{it}\beta + b_i + \mu_t + \epsilon_{it}) + u_{it} \tag{7}$$

焦点となるのは，レビューの不正操作のインセンティブに関わる変数のベクトルである $x_{it}$ である。(7)から，インセンティブに関わる(4)のパラメータ $\beta$ は直接推定できないものの，誘導系パラメータ $\alpha_1\beta$ を推定することができる。$\alpha_1$ は正のため，$\beta$ の符号については把握することが可能となる。

推定にあたっては，想定される内生性が 2 点残る。第一に，フェイクレビューの投稿に関わる誤差項である $\epsilon_{it}$ と $x_{it}$ の相関である。これについては，レストラン $i$ と時期 $i$ の固定効果 $b_i$ と $\mu_t$ を説明変数に含めることで，レストランや時期に特有なショックをコントロールすることができる。

第二の内生性として，フィルタリングプロセスの誤差項である $u_{it}$ と $x_{it}$ の相関が考えられる。これは，レビューのフィルタリングのされやすさに関する観察されない属性と，$x_{it}$ の相関となる。これに対処すべく，$u_{itk}$ を以下の(8)のように定式化する。

$$u_{itk} = z'_{ikt}\gamma + \tilde{u}_{ikt} \tag{8}$$

ここで，$z_{ikt}$ はレビューのフィルタリングされやすさに関わる属性であり，$\tilde{u}_{ikt}$ は独立な誤差項である。(8)を用いて，以下の(9)が得られる。

$$f_{it} = \alpha_0 n_{it} + \alpha_1(x'_{it}\beta + b_i + \mu_t + \epsilon_{it}) + \sum\nolimits_k z'_{ikt}\gamma + \tilde{u}_{it} \tag{9}$$

(9)は，(7)にフィルタリングに関わるレビューの属性の和を加えた形となる。$z'_{ikt}$ に用いる具体的な変数は，(3)の推定結果から，レビューの長さや，レビュアーの他のレビューの数などを用いる。

対象としては Mayzlin et al.（2014）と同様に，自社の評価を高く装うための高評価のフェイクレビューと，他社を攻撃するための低評価のフェイクレビューの両方を扱う。前者については，フィルタリングされた星 5 レビューの投稿数を被説明変数として設定し，後者については，フィルタリングされた星 1 レビューの投稿数を被説明変数とする。

**推定結果**

フィルタリングされた星 5 レビューの投稿数を被説明変数とした場合，前期の星 1 や星 2 の低評価のレビューの投稿数は，有意に正の影響を与える。逆に，前期の星 4 や星 5 の高評価のレビューの投稿数は，有意に負の影響を与える。この結果は，フェイクレビューと思われる高評価のレビューが，レストランの評価が低くなると増加し，高くなると減少することを意味する。自社の評価を上げるというインセンティブに沿ってレビューの不正操作が行われていることがうかがえる。

また，前期までに投稿された合計レビュー数は，有意に負の影響を与えている。これは，レビュー数が増えるほど，フィルタリングされた高評価レビューが減少することを意味する。これも，レビューの全体数が増加すると，レーティングを操作するのに必要なフェイクレビューの数が増加し，追加的なフェイクレビューの効果が薄れることから，経済的インセンティブと整合的な結果といえる。

Mayzlin et al.（2014）と同様，レストランがチェーン店[(18)]の場合，有意にフィルタリングされた高評価レビューが減少する。チェーン店はフェイクのコストが大きいことと整合的である。加えて，同じ地域の競合レストランの存在は，有意な影響を与えているとはいえない結果になっており，この点についても同様である。自社の高評価を装うようなフェイクレビューは，競合他社と関係なく行われている可能性が示されている。

フィルタリングされた星 1 レビューの投稿数を被説明変数とする場合は，前期のレビューの投稿数は，いずれの点についても有意な影響を与えているとはいえない結果になっている。時系列に沿った点数の変化は，攻撃的なフェイクレビューにインセンティブを与えない傾向がうかがえる。一方で，前期のレビューの総数は有意に負の影響をもたらす。高評価のフィルタリングされたレビューを被説明変数としたケースと同様に，レビューの全体数が増加すると追加的なフェイクレビューの効果が薄れることから，経済的インセンティブと整合的な結果といえる。

他方，近隣の競合店は有意な影響を与える。近隣に同じ種類の料理を出す独立系のレストランが存在する場合，有意にフィルタリングされた低評価レビューが増加する。逆に，同じ料理を出すレストランでも，チェーン店の存在は，フィルタリングされた低評価レビューを減少させる。これらは，高評価レビューを被説明変数とする場合とは対照的ではあるものの，Mayzlin et al.（2014）と

（18）これはチェーン店を示すダミーを用いて推定を行っている。店舗の固定効果をモデルに組み込んでいる下では，多重共線性が発生し推定が行えないため，チェーン店の効果を計る際には，店舗の固定効果を外したランダムエフェクトモデルにより推定を行っている。

同様に，独立系の競合他社からの攻撃がなされていることがうかがえる結果になっている。また，フェイクレビューのコストが高いチェーンの場合は攻撃的なレビューが減少する点も共通している。

以上の Luca and Zervas（2016）の導いた結果からは，Mayzlin et al.（2014）と共通した結果がみてとれる。独立系の競合店舗ほどレビューの不正行為を行いやすく，これらが近隣に存在する場合は低評価のフェイクレビューにより攻撃を受けやすい。これらの点に関しては，Mayzlin et al.（2014）と同様に，企業が供給する財の品質に関連したエビデンスが直接的には示されていない。一方で，特に正のフェイクレビューについて，時系列に沿った得点分布の変化に，自社の評価を高くするインセンティブに沿う形で反応する。このことからは，質の低い財を供給する企業ほどレビューの不正行為に関わりやすいことが示唆される。

### 3.4 He et al.（2022b）

He et al.（2022b）は，近年問題となっている Amazon におけるレビューの不正操作を対象とする。彼らは，米 Amazon（Amazon.com）から取得したパネルデータの分析を通じ，レビューの不正操作について，その主体だけでなく被害についてもある程度のレベルで明らかにしている。最大の特徴は，Mayzlin et al.（2014）や Luca and Zervas（2016）とは異なり，フェイクレビューについてある程度判別を可能としている点である。

**フェイクの判別**

Amazon.com におけるフェイクレビューは，Facebook におけるフェイクレビューを募集するコミュニティで行われる。この Facebook のコミュニティにおいて，レビューの不正行為を行う商品の販売者は，商品代金のキャッシュバックや金銭と引き換えに，消費者に自社製品を購入し高評価のレビューを書くように依頼する。He et al.（2022b）は，リサーチアシスタントを用い，この Facebook におけるフェイクレビューを依頼するコミュニティへの投稿を調査することで，フェイクレビューの募集が行われている商品と，フェイクレビューの募集が開始された時期を把握している。(19)

この商品レベルでのフェイクレビューの把握により，レビューの不正操作を行う企業の特徴のより正確な推定が可能となっているだけでなく，フェイクレビューの募集前後の時系列に沿った商品の売上などの変化をみることができる。

**データ**

まず，2019 年 10 月から 2020 年 6 月にかけて Facebook からデータを収集し，おおよそ 1,500

---

（19）フェイクレビューの募集を行っている商品と時期は把握しているが，その商品に投稿される個別のレビューがフェイクか否かは判別できていない。

のフェイクレビューを募集した商品（フェイクレビュー商品[20]）を特定している。さらに，同期間に，Amazon の同じカテゴリにおける商品検索ページにおいて，フェイクレビュー商品と同じページにある商品について，レーティング[21]や価格などの属性を日次で取得する。加えて，2020 年 8 月から 2021 年 1 月にかけて，競合商品の中から選ばれた約 2,700 商品とフェイクレビュー商品について，レーティングなどを日次で，詳細なレビューデータを月に二回のペースで取得する。

**フェイクレビュー商品の特徴**

こうしたフェイクレビュー商品について，そうでない通常の競合商品と比べ，どのような傾向があるか，記述統計[22]が示されている。競合商品の商品年齢[23]は約 758 日なのに対し，フェイクレビュー商品の商品年齢は，平均約 230 日であり，かなり若い商品となっている。投稿されているレビューの数についても，こうした商品年齢の差に応じて，前者が平均約 450 件であるのに対し，後者は平均約 183 件となっている。また，価格は競合商品が平均約 45 ドルであるのに対し，フェイクレビュー商品は 33.4 ドルと安い。フェイクレビュー商品は相対的に新しく参入した，安価な商品である傾向がうかがえる。ただし，新商品とは言い難い水準であり，フェイクレビューを販売から 1 か月以内に行っている商品は全体の 1 割に満たない。

レビューのレーティングや，セールスランク，検索順位などについては，フェイクレビュー商品が競合商品を上回る傾向にある。平均的なレーティングは，競合商品が 4.2 点であるのに対し，フェイクレビュー商品は 4.4 点となっている。セールスランクは，前者が約 73,292 位なのに対し，後者は約 90,000 位となっている。検索順位についても，前者が約 21 位，後者が約 28 位となっており，全体的にフェイクレビュー商品が売上や評価の面では相対的に高い成績を残している。

このような商品の属性に加え，フェイクレビュー商品の販売者についても[24]，記述統計が示されており，これらの 84%が中国企業となっている。

**Amazon 側の対応**

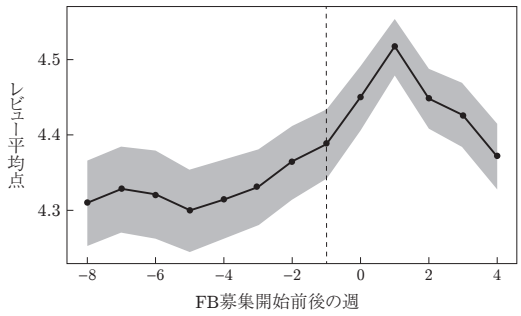
He et al.（2022b）は，フェイクレビューの削除状況についても，貴重な知見を提供している。Amazon は不正と思われるレビューの削除を行うが，通常の商品のレビューは 23%程度しか削除されないのに対し，フェイクレビュー商品のレビューがのちに削除される確率は 43%と高い削除率を

---

（20）論文中では，focal product とされる。
（21）ここでは，Amazon によるレビューの点数を集計した総合評価を指す。
（22）この記述統計に記載はなく，具体的な数値は提示されていないものの，ほとんどのフェイクレビュー商品が知名度の低い非ブランド品であることも言及されている。
（23）発売からの経過週数となる。
（24）これは競合商品の販売者と比較されず，フェイクレビュー商品の販売者だけに絞られている。

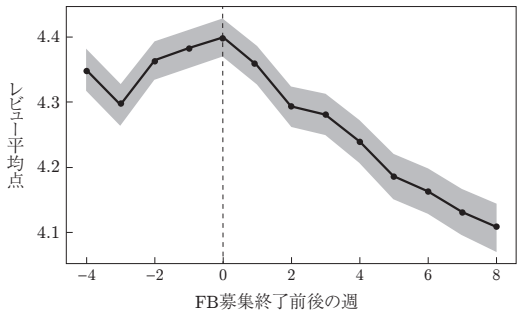
図 2：レビュー平均点のトレンド

(a) フェイクレビュー募集開始前後のトレンド
(He et al. (2022b), Figure 4-1.)




レプリケーションコード（He et al., 2022b）より筆者作成

(b) フェイクレビュー募集終了前後のトレンド
(He et al. (2022b), Figure 8-1.)




レプリケーションコード（He et al., 2022b）より筆者作成

示した。これは，Amazon のフェイクレビュー検知がある程度効果的に働いていることを意味する。

こうして削除されたレビューは，削除されていないレビューと比べ，平均的に高い評価をもつ。これは，削除されたレビューのほとんどが星 5 のレビューであることによる。また，文章が長く，タイトルが短く，写真が添付されている傾向にある。Luca and Zervas (2016)における Yelp のフィルタリングでは，文章が短いレビューほど削除されやすいことが示されており，対照的な結果といえる。直観的には文章が長いレビューや写真つきのレビューはフェイクである可能性が低いと考えられ，フェイクレビューが巧妙化しており，それに Amazon 側も対応している可能性が示唆される。

一方で，削除されるタイミングからは，Amazon の対処が後手に回っていることがわかる。削除されたレビューが削除されるまでの日数は，投稿されてから平均 100 日，中央値 53 日となっている。

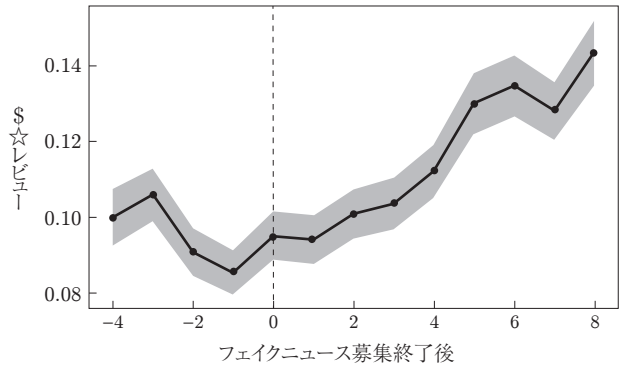
**フェイクレビュー募集後のトレンド**

レビューの不正操作の影響を推定するにあたり，フェイクレビュー商品について，フェイクレビューの募集開始前後のトレンドを確認する。図 2 (a)は，フェイクレビューの募集開始した週を 0 として，レビュー平均点の推移をプロットしたものであり，図 2 (b)は募集終了週を基準に同様のグラフを描いたものである。募集開始から平均点が上昇したのち，募集終了後は急激に下降している。フェイクレビューにより一時的に高評価に装われていたものが，実態が明らかになるにつれ評価を落としていることが示唆される。[(25)]

特に，レビュー募集終了後にレビューの平均点が減少することから，低評価のレビューが増加していることがうかがえる。低評価のレビューは，望まぬ購入をした消費者と想定されるため，被害の

(25) セールスランクなどの他の指標についても，フェイクレビュー募集開始と募集終了を機に，急激に変化し，フェイクレビューによる高評価の偽装がうかがえる結果になっている。

図 3：フェイクレビュー募集終了前後の星 1 レビュー割合のトレンド（He et al.（2022b), Figure 14.）




レプリケーションコード（He et al., 2022b）より筆者作成

プロキシーとして用いることができる。フェイクレビューの募集終了週を基準に，星 1 のレビューの割合のトレンドをプロットしたものが，図 3 となる。フェイクレビューの募集が終わってから，星 1 レビューの割合が大きく増加しており，少なくない数の消費者が実際にフェイクレビューに欺かれ，被害が生じていることがうかがえる。

**フェイクレビューの因果効果**

フェイクレビューの投稿される時期は，商品の固定効果をコントロールしたうえでも，他の観察されないショックと相関する可能性が高い。また，その期間に価格が低下したり，プロモーションが行われる可能性もある。

こうした内生性の問題に対処すべく，He et al.（2022b）は自然実験的な状況を利用している。Amazon.com はフェイクレビューと思われるレビューの削除を行うが，分析期間中のある一時期において削除されたレビューが極端に多い。この大量削除が予想されていなかった場合，この時期にフェイクレビューを募集した商品は，フェイクレビューにより高評価を水増ししにくい。

そこで，大量削除の時期にフェイクレビューを募集した商品をコントロールグループとして扱い，他の時期にフェイクレビューを募集した商品をトリートメントとして扱う。これらのアウトカムの変化を比較することで，差分の差分法（DID）を用いることができる。モデルは以下の(10)となる。

$$y_{it} = \beta_1 Treated_i + \beta_2 After_{it} + \beta_2 Treadted_i \times After_i + \alpha_i + \tau_t + X'_{it}\gamma + \epsilon_{it} \qquad (10)$$

ここで，$t$ 期の商品 $i$ に対し，$y_{it}$ がアウトカム，$After_{it}$ はフェイクレビューの募集開始後を示す。$\alpha_i$ と $\tau_t$ は $i$ と $t$ の固定効果を示す。$X_{it}$ は時間で変動する商品の属性となる。

関心の対象は $\beta_2$ と $\beta_3$ である。$\beta_2$ はフェイクレビューを募集すること自体による効果で，$\beta_3$ がフェイクレビューによりレビューが操作されたことによる効果といえる。

まず，累積レビュー数の対数を被説明変数とした推定では，$\beta_3$ は有意に正，$\beta_2$ は 0 に近く有意とはいえない結果となり，トリートメントでのみ有意にレビューが増加しているといえる。フェイクレビューの大量削除により，コントロールではレビューが増加していない一方で，トリートメントではフェイクレビューが投稿されていることがうかがえる。

セールスランクの対数を被説明変数とした場合，$\beta_2$ は有意に正，$\beta_3$ は有意に負となっており，$\beta_2 + \beta_3$ は負となる。フェイクレビューの募集によりコントロールはセールスランクを落とすものの，トリートメントはセールスランクを上昇させていることがわかる。また，DID 推定量である $\beta_3$ が有意に負であることから，フェイクレビューの募集が全体としてセールスランクを改善する因果効果がうかがえる。

He et al.（2022b）が示した結果は，どのような商品の売り手が不正を行うかについて，Mayzlin et al.（2014）や Luca and Zervas（2016）と比べ，より詳細で直接的な結果を示している。これらは，フェイクレビューの募集が終わったのちに急激に評価を落としていることから，本来は低品質である商品と判断できる。さらに，限定的ながらもフェイクレビューがもたらす影響を明らかにしている点で，これら二つの先行研究のギャップを埋めるものといえる。

## 4　実験データを用いた実証研究

### 4.1　概観

近年では，フェイクレビューについての実験を行い，その影響や頑健なレビューシステムを模索する例（Ananthakrishnan et al., 2020; Akesson et al., 2023; Huang et al., 2023）や，実験の結果と観察データを組み合わせてフェイクレビューの影響を調査する Gandhi et al.（2024）がある。

Ananthakrishnan et al.（2020）は，レストランのレビューサイトを模した実験サイトを用い，5 つのレビューシステムのパフォーマンスについて検証している。フェイクと判断されたレビューを削除するのではなく，それが疑わしいレビューであることを表示するシステムの方が，消費者のプラットフォームへの信頼を向上させるとの結果が示されている。

Akesson et al.（2023）は，EC サイトを模した web ページで実際に商品を選択させる実験を行っている。この実験を通じ，消費者はフェイクレビューをうまく判別できず，その結果質の低い商品に誘導されることや，その厚生損失が消費額の約 10%に及ぶことを示している。また，フェイクレビューの存在を警告するプラットフォーム側の教育的介入が，この厚生損失を約半数に減少させると推定している。

Huang et al.（2023）は，属性が同一で評価分布が異なる商品を被験者に提示する実験により，フェイクレビューへの疑念が低分散の商品に消費者を誘導し，商品の WTP（支払い意思額）を減少させることを明らかにしている。この際，被験者はフェイクレビューの存在についての警告が行われる

トリートメントグループ[26]と，警告が行われないコントロールグループに分けられ，フェイクレビューの存在についての疑いがコントロールされる[27]。

これらの研究は，個人レベルでの実際の選択行動を把握しており，フェイクレビューの下で消費者が受ける影響の分析が充実している。観察データを用いる場合，調査に限界のあるリサーチクエスチョンであり，そのギャップを補うものといえる。

また，制度や介入を評価するにあたっては，データバリエーションが観察されにくい関係上，観察データの分析では困難であり，実験を用いたアプローチが適しているといえる。他方，Gandhi et al.（2024）は実験によりフェイクレビューに関する消費者の商品に対する信念を抽出し，それを用いた需要推定を行っている。これにより，レビューの不正操作がもたらす影響を，フェイクレビューにより消費者が欺かれる被害と，レビューシステムへの不信による需要減少に分けたうえで，定量的に評価している。実験データを組み合わせることで，観察データのみでは分析が困難な点を明らかにしているといえる。

### 4.2 Akesson et al.（2023）

Akesson et al.（2023）は，Amazon を模した実験サイトで実際に商品を選択させる実験を行い，フェイクレビューの下での被験者の選択行動を観察することで，レビューの不正操作が消費者に与える影響について分析している。さらに，教育的介入を行った場合に，被害をどれだけ抑制できるのかについて検証している。

#### 実験設計

彼らは，英国で 10,000 人の成人を被験者として募集し，Amazon を模した実験サイトにおいて実際に商品を選択させる実験を行った。ここでは，ヘッドホン，車載カメラ，コードレス掃除機の 3 つの商品カテゴリのうちからいずれかを選び，さらにその商品カテゴリの中で 5 つの商品から 1 つを選択する。被験者は抽選に当選した場合，その選択した商品を受け取ることができる。これらの商品は，1 つが外部機関 Which?[28]によって認定された優秀な商品（Best Buy），1 つが買うべきではないと認定された商品（Don't Buy）となる。実験サイトでは，商品の価格[29]と説明のほかに，商品についてのテキスト付きレビューが表示され，レーティングとレビュー得点の分布が表示される。

---

（26）フェイクレビューの氾濫についての Economist の記事が表示される。

（27）単にトリートメント効果が計られるだけではなく，被験者のフェイクレビューへの疑念の強さに対して，操作変数としても用いられる。

（28）イギリスの消費者団体である，Consumers' Association により運営されるブランドであり，消費者保護のために，商品をテストする。

（29）価格はカテゴリごとに固定となっている。

表 1：Akesson et al.（2023）での介入 [a]

| | G1 | G2 | G3 | G4 | G5 | G6 |
|---|---|---|---|---|---|---|
| インフレしたレーティング分布 | | ✓ | ✓ | ✓ | ✓ | ✓ |
| フェイクレビューのテキスト | | | ✓ | | ✓ | ✓ |
| より疑わしいフェイクレビューテキスト | | | | ✓ | | |
| プラットフォームの推奨マーク | | | | | ✓ | ✓ |
| 警告バナー | | | | | | ✓ |

[a] Akesson et al.（2023）より筆者作成

被験者は，商品の選択において，真のレビューのみが表示されるコントロールグループ（グループ 1）と，フェイクレビューについての介入を受ける 5 つのトリートメントグループ（グループ 2〜5），合計 6 のグループにランダムに振り分けられる。これらのグループと介入の関係を示したものが表 1 である。

グループ 1 では，Amazon に実際に投稿された本物と思われるレビューと，実際のレーティングが表示される。グループ 2 からグループ 5 までは，フェイクレビューの存在を再現した介入が行われ，どのグループにおいても，Don't Buy 商品に，フェイクレビューで操作されたレーティングを模した，高得点のレーティングと星 5 が多い極端な分布が表示される。このとき Don't Buy 商品のレーティングは平均して 1.46 ポイント分上昇する。

さらに，グループ 3 と 4 では，グループ 2 の処置に加え，レーティングだけでなくレビューについてもフェイクと疑わしいものが表示される。グループ 3 では Don't Buy 商品における否定的なレビューを，反復的なフレーズや誇張的な表現が用いられたテキストなど，フェイクレビューの特徴をもつ高評価のレビューに置き換える。グループ 4 では，金銭を提供されたことを認める内容など，さらにフェイクとして疑わしい特徴があるテキストのレビューが表示される。

グループ 5 と 6 では，グループ 3 の処置に加え，フェイクレビューの下での教育的介入が行われる。グループ 5 では Amazon における推奨商品のマークである Amazon's choice を再現したマークが特定の商品に表示され，グループ 6 ではこの推奨マークに加え，実験サイトのページ上部に，フェイクレビューの存在を警告するメッセージが表示される。

**WTP 推定のための付随実験**

この主実験に追加で，1,000 人の成人を被験者として募集し，商品への WTP を申請させる付随実験も行っている。この実験の結果を利用することで，商品への WTP を推定し，分析のアウトカムとして用いる。これは，主実験で提示された商品について，商品の属性やレーティング，Which での評価を見せたうえで，各商品に対して支払い意思額（WTP）を尋ねるものである。

この実験は，被験者が WTP を正直に申告するように，BDM 法（Becker et al., 1964）を用いる。こ

の際，BDM 法での商品の贈呈は被験者全員に行われるわけではなく，抽選の当選者に対してのみ，300 ポンドの贈呈のうえで行われる。これは以下のようなプロセスである。まず，当選した被験者が WTP を申請した商品から 1 つの商品がランダムに選ばれ，ランダムな価格が設定される。当選者の申請した WTP がこの価格以上である場合，その商品が贈呈され，さらに 300 ポンドと価格の差額が賞金として贈呈される。申請した WTP がこの価格未満であるときは，300 ポンドの賞金のみが贈呈される。実験の事前にこの BDM 法のプロセスと，WTP を正直に申告することが合理的であることを伝えている。

この方法で申告された WTP は，フェイクレビューがなく，正しい情報が提示された下での WTP といえる。この WTP と，付随実験で申請した被験者の属性，さらに主実験の被験者の属性を用いることで，主実験での被験者についても商品の WTP[(30)] を予測することが可能となる。この予測された WTP がアウトカムとして用いられる。

**フェイクレビューの影響**

各種アウトカムをトリートメントグループのダミーに回帰することで，フェイクレビューの下において被験者がどのような影響を受けるか明らかにする。ここで，Don't Buy や Best Buy 商品を選ぶか否かが商品の質についてのアウトカム[(31)]となり，選択した商品への WTP が厚生についてのアウトカムとなる。

コントロールにおける Don't Buy 商品を選ぶ確率は約 10%ポイント，WTP は約 57.1 ポンドとなっている。一方，グループ 2 は Don't Buy 商品を選ぶ確率が有意に約 5.8%ポイント増加し，選択した商品の WTP が有意に 2.9 ポンド減少する。Best Buy 商品を選ぶ確率は有意な影響を受けているとはいえない。レーティングの操作の下では，質が低く，本来低い WTP しかもたない商品に誘導され，消費者余剰が減少していることを示す結果といえ，トリートメントと比べて，約 5.1%の厚生損失が生じている。

さらに，グループ 3 については，Don't Buy 商品を選ぶ確率が有意に約 12.6%ポイント増加し，Best Buy 商品を選ぶ確率が有意に約 5.2%ポイント減少する。さらに選択した商品の WTP が有意に 6.6 ポンド減少する。レーティングだけでなく，レビューにフェイクが混在する状況下では，より質の低い商品に誘導され，より大きい被害が生じている結果となっている。ここでは，トリートメントと比べ，約 11.7%の厚生損失が生じている。

グループ 4 についても，同様の結果となっており，Don't Buy 商品を選ぶ確率が有意に約 11.0%ポイント増加し，Best Buy 商品を選ぶ確率が有意に約 4.3%ポイント減少する。さらに選択した商品

---

（30）WTP は予測に用いる変数により 2 種類の予測値があり，本稿では被験者の年齢から予測された WTP についての結果を紹介する。どちらの WTP についても結果は大きく変化しない。

（31）線形確率モデルとなる。

の WTP が有意に 6.6 ポンド減少する。Don't Buy 商品の選択確率や厚生損失について，グループ 3 と比較して被害は少ないものの，その差は非常に小さい。顕著な特徴があったとしても，消費者はフェイクレビューをあまり見抜くことができないことがうかがえる。

グループ 5 については，Don't Buy 商品を選ぶ確率が有意に約 14.3%ポイント増加し，Best Buy 商品を選ぶ確率が有意に約 3.6%ポイント減少する。さらに選択した商品の WTP が有意に 7.2 ポンド減少する。プラットフォームの推奨マークの導入は，むしろ Don't Buy 商品の選択確率を上げ，WTP を減少させる効果となっている。ただし，Best Buy 商品を選ぶ確率については，被害が緩和されている。ここでは，トリートメントと比べ 12.7%の厚生損失が生じている。

**教育的介入の効果**

グループ 5 と，同様の処置に加えてフェイクレビューについての警告表示を受けたグループ 6 のアウトカムを比較することで，教育的介入の効果がどれだけあるのかを明らかにしている。グループ 6 は，グループ 5 と比べ，Don't Buy 商品を選ぶ確率が有意に約 5.8%ポイント減少し，選択した商品の WTP が有意に 3.0 ポンド増加する。Best Buy 商品を選ぶ確率は有意な影響を受けているとはいえない。質が低い商品に誘導されにくくなり，消費者余剰が増加するといえる結果であり，教育的介入はフェイクレビューの被害を緩和できていると判断できる。これは，フェイクレビューの下で生じる厚生損失について，42.2%分の被害が減少している。

Akesson et al.（2023）は，レビューの不正操作による被害がどの程度のものなのかを明らかにしている。観察データを用いた He et al.（2022b）は限定的な分析にとどまっており，先行研究のギャップを埋めるものといえる。さらに，観察データを用いた研究では困難な対象である，教育介入の効果を検証しており，第 1 節で述べた第三のリサーチクエスチョンである制度や施策の評価という点において，貴重な知見を蓄積するものといえる。

### 4.3 Gandhi et al.（2024）

Gandhi et al.（2024）は，実験データと観察データを併用することで，レビューの不正操作がもたらす影響を明らかにしている。まず，実験を通じ，フェイクレビューに関する消費者の信念を抽出する。実験においては，被験者は Amazon の商品ページを提示され，その商品がフェイクレビューを利用しているかについての信念や，フェイクレビューがどれだけ存在するかについての信念などを回答する。

抽出された信念から，フェイクレビューが混在しうる状況下での，商品の品質[(32)]についての信念が推定される。これは，Amazon で販売される商品の需要推定に用いられる。需要推定の結果から，

（32）品質は，真のレビューがポジティブな評価となる確率として定義される。

フェイクレビューの有無や，消費者の信念を変化させる形で，反実仮想シミュレーションが行われ，レビューの不正操作がもたらす影響が様々な角度から明らかにされる。

当研究は，これらの分析により，消費者がフェイクレビューという誤情報に欺かれることによる被害と，消費者のレビューシステムに対する不信がもたらす被害を，分離して推定することに成功している。

**データ**

彼らは，He et al.（2022b）と He et al.（2022a）のデータ，ならびに He et al.（2022a）が開発したアルゴリズムからフェイクレビューの利用に関するデータセットを構築する。約 7,500 個の商品のうち，約 1,500 個のフェイクレビューを募集した商品が特定されており，各レビューについてフェイクレビューの判別[(33)]が行われている。各商品の週次レベルの売上も取得されており，需要推定に用いられる。

フェイクレビューの利用，ならびにレビューレベルでのフェイクレビューの判別が行われているため，フェイクレビューを除外した場合に，レーティング[(34)]や，レビューの数がどのように変化するかをシミュレートすることが可能となっている。

**信念についての実験**

被験者は，まず第一のタスクとして，Amazon においてフェイクレビューを利用している商品の割合について，どの程度だと考えているか，0%から 100%の値を選択する。第二に，メインタスクとして，被験者が最もよく利用している 5 つのカテゴリの中から選ばれた 10 個の商品について，Amazon の商品ページを見せられ，その商品がフェイクレビューを利用しているか否かの確率がどの程度だと考えているか，実験サイトに設置されたスライダーを操作して，0%から 100%の値を選択する。その後に第三のタスクとして，メインタスクで最後に表示される商品について，この商品がフェイクレビューを利用していると仮定して，レビューのうちのフェイクレビューの割合を，0%から 100%の値を選択するよう要請される。最後に，理解度チェックのための質問がなされ，実験は終了する。これらの回答は，被験者の不正に対する信念の表明といえる。

被験者が閲覧する Amazon の商品ページは，ほぼ実物そのもの[(35)]であるが，商品のレビューやレーティング（⋆）の分布がランダムに変更されている。これにより，レビューの数やレーティングの分布に条件づけられた，被験者の不正に対する信念を推定することが可能である。

実験にあたっては，被験者が自分の信念を正直に答えることが最適な戦略であるように報酬が与

---

（33）アルゴリズムの予測結果にはとどまる。
（34）ここでは，He et al.（2022b）と同様，Amazon によるレビューの点数を集計した総合評価を指す。
（35）個別のレビューについては表示されない。

えられる。メインタスク[36]の各回答において，被験者の信念の正しさに応じて，最大 1 ドルの報酬を受け取ることができる。具体的には，被験者の信念と正解の差の分だけ，報酬が 1 ドルから 1 セントずつ減少する。例えば，フェイクレビューを使用していない商品（正解 0%）に対して 30%を回答した場合には，0.7 ドルを獲得でき，フェイクレビューを使用している商品（正解 100%）に対して 30%を回答した場合は，0.3 ドルを獲得できる。実験を通じ，被験者は参加に対しての報酬 1 ドルに加え，メインタスクのパフォーマンスに応じて最大 10 ドルを獲得することができる。

この信念の回答と報酬額の関係は，被験者が操作するスライダーに併記され，選択に応じて自動で更新される。選択した確率に応じて，商品がフェイクレビューを利用している場合に得られる報酬額と，利用していない場合に得られる報酬額が表示される。被験者が自分の選択と報酬額の関係を理解可能なように設計されており，自分の信念を正直に答えることが最適な戦略であるということを把握できるようになっている。

実験の結果，第一タスクから推定された，Amazon 全体での不正の割合についての信念は，平均して約 31%となった。これは，彼らのデータセットにおける割合（約 20%）と比較して大きい数値である。第二タスクにおける，各商品が不正を行っていることへの信念は，実際に不正を行っている商品については平均は 42%，行っていない商品については平均 39%となった。実際の不正の有無で大きな差はなく，消費者はフェイクレビューの利用をよく見分けられないことが示唆される[37]。

メインタスクから推定される，レビューの数やレーティングに条件づけられた信念については，レビューが少なくレーティングが高い商品が特に不正が行われていると疑われている。一方で，レビューが少なくレーティングが低い商品については疑われにくい。レビュー数が多い商品では，レーティングと信念の関係には顕著な関係がみられない。

**観察データを用いた需要推定**

次に，BLP モデル（Berry et al., 1995）により，需要推定が行われる。商品の間接効用関数は以下の(11)のようにモデル化される。

$$u_{ijt} = \beta_{0i} E(q^*_{jt}|r_{jt}, N_{jt}) - \alpha_i p_{jt} + \beta X_{jt} + \xi_j + \lambda_t + \epsilon_{ijt} \tag{11}$$

ここで，$i$ は個人，$j$ は商品，$t$ は週を表す。$p_{jt}$ は商品の価格，$X_{jt}$ は商品の属性である。$\xi_j$ は商品についての固定効果，$\lambda_t$ は週の固定効果である。キーとなるのは，$E(q^*_{jt}|r_{jt}, N_{jt})$ であり，商品

(36) 第一のタスクと第三のタスクについては，具体的な仕様について記述がないものの，第三タスクについては誘因両立であると言及されており，おそらくメインタスクと同様の設計がなされていると考えられる。

(37) 彼らは，別の 100 人の被験者を対象に，レビューテキストが閲覧可能な形に拡張した同様の実験を行っているが，むしろレビューを読んだ被験者の方が不正を判別する確率が減少する結果になっている。

表 2：Gandhi et al.（2024）での反実仮想シミュレーション結果 [a]

| | シナリオ 1 | シナリオ 2 | シナリオ 3 | ベースライン |
|---|---|---|---|---|
| 消費者余剰（$） | 61,791,060 | 61,688,828 | 60,581,713 | 60,665,290 |
| プラットフォーム利潤（$） | 19,143,666 | 19,247,171 | 18,285,577 | 18,446,978 |
| FR 商品 [b] 平均価格（$） | 31.30 | 31.68 | 31.18 | 31.62 |
| NFR 商品 平均価格（$） | 37.86 | 37.79 | 37.82 | 37.75 |
| FR 商品 売上個数 | 645,421 | 901,136 | 568,366 | 866,698 |
| NFR 商品 売上個数 | 4,569,797 | 4,395,619 | 4,408,504 | 4,213,520 |
| FR 商品 利潤（$） | 4,412,089 | 6,269,425 | 3,843,358 | 5,979,472 |
| NFR 商品 利潤（$） | 28,575,985 | 27,136,755 | 27,425,096 | 25,850,417 |

[a] Gandhi et al.（2024）より筆者が抜粋・作成
[b] FR 商品はフェイクレビューを利用した商品，NFR 商品はフェイクレビューを利用していない商品を指す

の $\star$ のレーティングである $r_{jt}$ と，レビューの数である $N_{jt}$ に条件づけられた，真の品質である $q^*_{jt}$ への消費者への信念となる。

$E(q^*_{jt}|r_{jt}, N_{jt})$ を，需要パラメータと同時に識別することは不可能である。そこで，彼らは実験で推定された信念から，$E(q^*_{jt}|r_{jt}, N_{jt})$ を計算[(38)]し，外挿している。

**反実仮想シミュレーション**

推定された需要パラメータを用い，反実仮想シミュレーションを行うことで，レビューの不正操作の影響を明らかにする。ここでは，3 つのシナリオでのシミュレーションが行われ，新たな均衡での商品の価格や消費者余剰[(39)]の，プラットフォーマーならびに売り手の利潤などが比較される。シナリオごとのシミュレーション結果を示したものが，上の表 2 である。

まず，フェイクレビューが存在せず，消費者がレビューを完全に信頼している状況（シナリオ 1）と現実（ベースライン）の比較を行う。ここでは，フェイクレビューによる不正が行えなくなった影響で，不正を行っていた商品の価格・売上がともに下がる一方で，不正を行っていなかった商品の価格・売上がともに上昇する。結果的に，消費者余剰とプラットフォームの利潤の両者が改善する。

次に，誤情報と不信による消費者被害を分離すべく，シナリオ 1 と，シナリオ 2・3 の比較を行う。シナリオ 2 は，誤情報による被害に対応したもので，フェイクレビューが存在しないという消費者の信念の下で，現実と同様にフェイクレビューが存在するケースである。シナリオ 3 は，フェイクレビューが存在するという消費者の信念の下で，フェイクレビューが存在しない場合のシミュ

（38）計算にあたっては，フェイクレビューの利用に条件づけられた商品の品質についての信念（事前分布）も必要となるが，消費者が正しくこの品質の分布を把握していることなどが仮定され，データから推定される。

（39）商品の選択における期待効用と，実際の経験的効用が異なるため，消費者余剰は log-sum value から直接計算されず，真の品質を用いて調整されたものから計算される。

レーション結果を示す。

シナリオ 2 においては，シナリオ 1 と比較し，消費者余剰が減少するものの，その減少幅は小さい。不正を行っている商品の価格・売上がともに伸びる一方で，不正を行っていない商品が価格を下落させるため，総合的な消費者余剰は小さくなる。一方で，シナリオ 3 における消費者余剰の減少幅は相対的に大きい。価格の変動は小さいものの，売上が全体的に減少する。これらの結果は，誤情報による消費者被害よりも，不信による被害の方が大きい可能性を示唆する。

さらに，シナリオ 3 においてはプラットフォームの利潤が大きく減少する。プラットフォーマーがフェイクレビューの削除を進める場合に，消費者の信念を変えない形で行ってしまうと，不信が維持され，逆効果になってしまう可能性がある。

Gandhi et al.（2024）は，He et al.（2022b）においては限定的であったレビューの不正操作がもたらす消費者被害だけでなく，プラットフォームの利潤への影響などについて，明確に定量化した形で明らかにしている。不正がもたらす影響という先行研究のギャップを，実験データと観察データを組み合わせた分析により，より進んだ形で埋めるものといえる。

## 5　その他の関連研究

機械学習の文脈においては，Ott et al.（2011, 2013）など，機械学習モデルによりレビューがフェイクである否かを判別することを試みる研究が数多く行われている。Ott et al.（2011, 2013）はクラウドソーシングを用い執筆させたフェイクレビュー[(40)]と，レビューサイトに投稿された実際のレビューを用い，学習用のデータセット[(41)]を作成し，レビューがフェイクか否かを判別する機械学習モデルを構築している。

近年においては，He et al.（2022a）は，He et al.（2022b）と同様のデータを用い，不正を行っている商品を判別したうえで，レビュアーのネットワークを把握し，そのネットワークの特徴を用いた機械学習モデルを提案している。これは，テキスト情報などからフェイクの判別を行う既存のモデルと比較し，高い精度を示している。

直接フェイクレビューを扱うわけではないものの，関連性が高い実証研究については，Anderson and Simester（2014）は，アパレル産業の小売りサイトにおけるレビューを対象に，実際に商品を購入していない消費者によるレビューについて分析している。これらが平均的に評価が低く，商品のフィット感や素材感について言及することが少ないうえに，嘘だと思われる文章上の特徴を含みやすいことを示している。

---

（40）フェイクレビューを書く業務を命じられたことを想定して執筆するよう依頼している。
（41）このデータセットは the gold standard dataset として，現在でも様々な研究で利用されている。

Park et al.（2023）はインセンティブレビュー[(42)]がもたらす影響について調査している。彼らは，インセンティブレビューが一般のレビューと比較して商品を高く評価し，売上を増加させる傾向を明らかにし，商品の売り手に依頼されるインセンティブレビューよりも，Vine プログラムのようなプラットフォームがレビューのインセンティブを与えるような仕組みが好ましいことを示している。

## 6　おわりに

第一のリサーチクエスチョンにあたる，不正の詳細に関しては，He et al.（2022b）が，フェイクレビューを募集している商品を判別することに成功しており，品質の低い商品の売り手が不正に関与する傾向にあることを明らかにしている。

しかしながら，個別のレビューがフェイクか否かは判別できていない。また，把握されたフェイクレビュー商品は約 1,500 個であり，同時に収集された 20 万個の商品と比べると小さい割合となっており，不正に関与している相当数の商品が見落とされている可能性がある。

第二のリサーチクエスチョンにあたる，レビューの不正操作の影響については，Gandhi et al.（2024）が観察データを用いたうえで，消費者やプラットフォーマー，売り手の受ける影響について定量的な評価を行っている。しかしながら，消費者の信念が観察データに一貫した形で推定されておらず，外挿される形となっている。加えて，レビュー単位でのフェイクの判別はアルゴリズムに依存しているため，反実仮想シミュレーションの信頼性についても疑問が残る。

第三のリサーチクエスチョンにあたる，施策の検証については，Akesson et al.（2023）が教育的介入の大きな効果を検証している。しかしながら，当研究がフェイクレビューの特徴として用いるもののいくつかは，He et al.（2022b）の調査結果と異なっており，現実に即していない可能性がある。さらに，厚生損失の計算に用いた，被験者の商品への WTP が，被験者本人の誘引両立的な選択から推定されたものではなく，付随実験の結果から予測されたものを用いており，信頼性に疑問が残る。

これらの限界を踏まえ，今後の研究の展望としては，以下があげられる。第一に，レビューがフェイクか否かの判別や，不正に関与している商品や売り手の特定について精緻化を進めることである。これは，He et al.（2022b）だけでなく，Gandhi et al.（2024）の分析の信頼性の担保にもつながる。

この際，Luca and Zervas（2016）の手法はプロキシーにとどまるものの，不正を行っている商品のカバレッジという範囲では，He et al.（2022b）を上回る可能性がある。また，特に Amazon にお

---

（42）かつて Amazon は，ユーザーが利害関係を明示することを条件に，商品の割引や無償提供と引き換えにレビューを投稿することを許可していた。これをインセンティブレビューという。現在では禁止されているものの，Vine プログラムとして，Amazon がレビュアーにインセンティブを与えレビューを執筆させるシステムがある。

いては，レビューの削除だけでなく，商品自体の削除や，常識では考えにくい長期間の在庫切れなど，別途もしくは組み合わせて利用できるプロキシーの候補が存在する。これらを用いた実証分析についても検討の価値がある。

第二に，レビューの不正操作に直面する消費者の異質性を調査することも重要性が高い。仮に，所得や学歴の低い，社会的弱者にあたる消費者について，不正操作による被害が大きい場合，不正が総余剰を増加させるとしても，公平性の面で規制の強化は正当化されうる。

Gandhi et al.（2024）の手法をベースに，この点を明らかにするには，個票レベルのデータが必要となると考えられるが，これは，消費者の信念を観察データと一貫した形で推定することにつながる可能性がある。

第三に，さらなる実験データを用いた研究の蓄積である。Akesson et al.（2023）の結果の頑健性を確かめるためにも，他の考えられうる施策の検証のためにも，実験を用いたエビデンスの蓄積が求められる。

考えられうる一つの案としては，商品表示のシステムの検証である。He et al.（2022b）の結果が示すように，フェイクレビューの募集により，売上順位とレーティングがともに上昇する。この際に，不正が行われた商品は，Amazon のシステム上優先して表示される可能性が高く，既存の商品表示システムが，フェイクレビューの被害を増幅している恐れがある。

これらは，プラットフォーマーの保有する詳細なデータなくしては解決が困難なものも多い。Amazon は官民連携の姿勢を示しているが[(43)]，今後は官学民での連携が求められる。

## 参 考 文 献


Akesson, J., Hahn, R. W., Metcalfe, R. D., and Monti-Nussbaum, M. (2023). The impact of fake reviews on demand and welfare. *Available at NBER 31836*.

Ananthakrishnan, U. M., Li, B., and Smith, M. D. (2020). A tangled web: Should online review portals display fraudulent reviews? *Information Systems Research*, 31(3):950–971.

Anderson, E. T. and Simester, D. I. (2014). Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research*, 51(3):249–269.

Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3):226–232.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Chevalier, J. A. and Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3):345–354.


（43） https://www.aboutamazon.jp/news/innovation/a-blueprint-for-private-and-public-sector-partnership-to-stop-fake-reviews

Chintagunta, P. K., Gopinath, S., and Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science*, 29(5):944–957.

Dellarocas, C. (2006). Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science*, 52(10):1577–1593.

Gandhi, A., Hollenbeck, B., and Li, Z. (2024). Misinformation and mistrust: The equilibrium effects of fake reviews on amazon. com. Technical report Technical report, UCLA Anderson School.

He, S., Hollenbeck, B., Overgoor, G., Proserpio, D., and Tosyali, A. (2022a). Detecting fake-review buyers using network structure: Direct evidence from amazon. *Proceedings of the National Academy of Sciences*, 119(47):e2211932119.

He, S., Hollenbeck, B., and Proserpio, D. (2022b). The market for fake reviews. *Marketing Science*, 41(5):896–921.

Hollenbeck, B. (2018). Online reputation mechanisms and the decreasing value of chain affiliation. *Journal of Marketing Research*, 55(5):636–654.

Huang, Y., Villas-Boas, J. M., and Zhao, M. (2023). Unmasking the deception: The interplay between fake reviews, rating dispersion, and consumer demand. *Available at SSRN 4621736*.

Luca, M. and Zervas, G. (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science*, 62(12):3412–3427.

Mayzlin, D. (2006). Promotional chat on the internet. *Marketing Science*, 25(2):155–163.

Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8):2421–2455.

Milgrom, P. and Roberts, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy*, 94(4):796–821.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78(2):311–329.

Ott, M., Cardie, C., and Hancock, J. T. (2013). Negative deceptive opinion spam. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 497–501.

Ott, M., Choi, Y., Cardie, C., and Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. *arXiv preprint arXiv:1107.4557*.

Park, S., Shin, W., and Xie, J. (2023). Disclosure in incentivized reviews: Does it protect consumers? *Management Science*, 69(11):7009–7021.

Yasui, Y. (2020). Controlling fake reviews. *Available at SSRN 3693468*.

カライスコス，アントニオス (2023). 諸外国におけるステルス・マーケティング規制の実効性確保. 国民生活研究 *Journal of Research on Social and Economic Life*, 63(1):1–22. [Antonios Karaiskos (2023). Shogaikoku ni okeru Stealth Marketing Kisei no Jikkosei Kakuho, *Kokumin Seikatsu Kenkyu (Journal of Research on Social and Economic Life)*, 63(1):1–22.]

消費者庁 (2020). デジタル・プラットフォーム利用者の意識・行動調査, [Shohishacho (2020). Digital Platform Riyosha no Ishiki-Kodo Chosa.] https://www.caa.go.jp/notice/assets/consumer_system_cms101_200520_03.pdf.

**要旨**: 近年,「フェイクレビュー」と呼ばれる偽のレビューにより，レビューを中心とした評価システムが不正に操作されることが横行している。このレビューの不正操作は，消費者の意思決定を歪

めるだけでなく，レビューシステム自体への信頼を毀損させることが懸念されている。そこで，本稿は，レビューの不正操作についての実証研究を中心にサーベイを行う。既存の研究を整理し，その限界を考察することで，今後の研究の方向性についての示唆を提供する。

**キーワード**: フェイクレビュー，サーベイ，消費者保護，デジタル・プラットフォーム，オンライン・リテール