

レビューの不正操作に関するサーベイ

坂口 洋英

March 4, 2022

Abstract

オンライン上の口コミや、消費者レビューは、消費者の意思決定に大きな役割を果たす。しかしながら、近年、企業が金銭と引き換えに偽のレビュー（フェイクレビュー）を執筆させることが社会問題化している。しかしながら、レビュー不正操作の実態については未だ不明瞭な点が多い。特に国内においては、取材レベルの調査が行われるのに留まっている。

本稿は、こうしたレビューの不正操作に関する先行研究を、米国における実証研究を中心に紹介したのち、国内におけるレビューの不正操作の実態について、Amazon.co.jp を対象とした実証分析により調査するものである。

1 はじめに

口コミ (Word of Mouth: WoM) は消費者が日常で財を選択するにあたり、大きな役割を果たす。特に近年では、高速インターネット回線とスマートデバイスの浸透を背景に様々なオンラインサービスが生まれ、インターネット上の口コミ (e-WoM)、その中でも消費者レビューが存在感を増している。例えば、Amazon や楽天のような EC サイトでは、レビューシステムやレーティングシステムが導入されており、販売される財への消費者からの評価が分かるようになっている。また、「食べログ」のような、特定の分野に特化した、レビューやレーティングを主目的とするようなサイトも、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、厚生労働省による「デジタル・プラットフォーム利用者の意識・行動調査」¹では、消費者の半数以上がレビューを参照していることが示されている。また、Chevalier and Mayzlin (2006) から連なる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、消費者レビューが需要に影響することから、レビューや口コミを悪用する企業も出現するようになった。国内では、2010年代前半から、「食べログやらせ事件」や「ペニーオークション詐欺事件」などをきっかけに、レビューや口コミを不正に操作することが、ステルスマーケティングの文脈で問題視されるようになった。

特に近年では、Amazon におけるレビューの不正操作が社会問題として注目を集めている。このレビューの不正操作においては、「フェイクレビュー」と呼ばれる偽のレビューが投稿される。これは、商品の販売者がレビューの執筆を消費者に委託して投稿させたレビューとなる。所謂「サクラややらせ」に該当するものであり、消費者による実際の使用感に基づいた”本物”のレビューではない”偽物”という点でフェイクといえるレビューとなる。このようなフェイクレビューにより、高評価のレビュー数を水増しして高評価を装うなど、レビューを不正に操作することが可能となる。

¹https://www.caa.go.jp/notice/assets/consumer_system_cms101_200520_03.pdf

米国においては、コロナ禍による Amazon の利用増加もあって増加傾向にあり、2020 年には全体のレビューのうち 40% ものレビューが偽装の疑いがもたれている。国内においても、多くのメディアで取り上げられているほか、Amazon のレビューについての検索エンジンの検索結果上位は、こうしたフェイクレビュー関連の記事が上位を占め、レビューの不正操作が横行していることが伺える。

このようなレビューの不正操作がもたらす被害については、まず、消費者が欺かれ、望まない消費を行うという直接的なものが想定される。加えて、消費者のレビューという仕組み自体への信頼を失わせる可能性がある。これは、財の選択に係るサーチコストを増大させ、間接的に、消費者だけでなく商品販売者やプラットフォーム全体に被害をもたらすことが懸念される。

米国では強く問題視されており、FTC(連邦取引委員会)は、金銭を受け取り公平な第三者と偽ってレビューを執筆することを明確に違法行為と定め、高い罰金を科している²。一方我が国では、2012 年の景品表示法のガイドラインの改訂に留まっている。これは、フェイクレビューが景品表示法の不当表示にあたる可能性について示したものにすぎず、法曹界からは規制強化を求める声もある³。

一方で、レビューシステムに関する規制は、レビューの投稿を減らし、レビューシステムを弱体化させるトレードオフを生じさせる可能性が高い。そのため、規制にあたっては、レビューの不正操作がもたらす被害を定量的に評価することが求められる。

このようなレビューの不正操作がもたらす被害は、不正を行う主体に影響される。フェイクレビューにより被害が生じるのは、低品質な財の供給者が、あたかも高品質なものに見せかけるようフェイクレビューを用いているケースである。この場合は、先述したように、消費者が欺かれて粗悪品を購入してしまうという形で被害が生じる恐れがある。

高品質な財の高級車がフェイクレビューによる不正操作を行っている場合、異なる様相を呈する。高品質な財に対して、財が高く評価されているようなフェイクレビューが投稿されていても、消費者を欺くことにはならない。倫理的には問題なものの、消費者への直接的な被害や、レビューシステムへの信頼の毀損といった事態にはつながりにくいと考えられる。

高品質な財の供給者がレビューの不正操作を行うというのは非直観的であり、想像し難い。しかしながら、高品質な財を供給しながらも、参入して間もないために知名度が低いような企業がフェイクレビューを使うというのは十分にあり得るだろう。また、理論研究の中には高品質な財を供給する企業ほど不正を行うという均衡が起こりうることを支持する結果を示すもの (Dellarocas, 2006; Yasui, 2020) も存在する。

また、レビューの不正操作がもたらす被害は、消費者の異質性と分布にも左右される。フェイクレビューの氾濫する現状に対し、その存在を割り引いて商品を選択するような消費者や騙されにくい消費者が多ければ、不正操作の被害は小さくなる。一方で、フェイクレビューの存在を割り引かずにレビューをそのまま評価するような消費者や、フェイクレビューに騙されやすい消費者が多ければ、不正操作の被害は大きくなる。

これらを考慮すると、レビューの不正操作がもたらす被害を把握するには、どのような企業がレビューの不正を行うのか、消費者がどのようにフェイクレビューに反応するのかについてのエビデンスを蓄積することが肝要となる。

しかしながら、レビュー不正操作の実態については未だ不明瞭な点が多い。特に国内においては、取材レベルの調査が行われるのに留まっており、実証分析によりこれらについて研究した例はない。

²<https://www.ftc.gov/news-events/press-releases/2021/10/ftc-puts-hundreds-businesses-notice-about-fake-reviews-other>

³日本弁護士連合会は 2017 年に「ステルスマーケティングの規制に関する意見書」を消費者庁に提出している。
https://www.nichibenren.or.jp/library/ja/opinion/report/data/2017/opinion_170216_02.pdf

一方で、米国では少ないながらも、先行研究が存在する。米国ホテルレビューサイトにおける不正操作対象とした Mayzlin et al. (2014) を嚆矢として、ボストンのレストランのレビューを分析した Luca and Zervas (2016) や、近年では Amazon.com(米 Amazon) のフェイクレビューを対象とした He et al. (2021) が存在する。

そこで、本研究は、こうしたレビューの不正操作に関わる実証研究について、先行研究のサーベイを行い、既存の実証研究により明らかになっていることと現状の課題について整理する。そのうえで、国内におけるフェイクレビューの実態について、実証研究を通じ調査する。

本稿の以降の構成は以下の通りである。二節ではレビューの不正操作に関する研究について、概観を解説したうえで、主要な三つの論文について詳細なサーベイを行う。三節では、日本のデータを用いた分析結果を報告し、判明したことについて述べる。四節では今後の展望について述べる。

2 先行研究のサーベイ

本節では、レビューの不正行為に関連する先行研究のサーベイを行う。まず、理論研究と実証研究の現状について簡易に述べたのちに、代表的な実証研究である Mayzlin et al. (2014) と Luca and Zervas (2016)、He et al. (2021) について紹介する。

2.1 理論研究の整理

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、こうしたシグナリングモデルが拡張される。消費者は、レビューをシグナルとして受け取る。この際、消費者は“本物”の消費者によるレビューと不正操作されたものの両方が存在することを理解し、後者の存在を考慮し割り引いて財の選択を行う。企業はこうした消費者を想定して、レビューの不正操作の程度を調整する。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フェイクを行い、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のレビューにより、消費者の需要が影響される理論モデルを構築し、分析を行った。このモデルにおいては、消費者は、ステルスマーケティングに関係なく財の品質を見分けられるタイプと、財の品質を見分けることができず、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消費者に分けられる。結果、均衡では、低品質な商品を供給する企業ほど多くのステルスマーケティングを行うことが示された。また、ステルスマーケティングによる厚生損失は、財の品質の差や、前者の品質を見分けられる消費者の数に依存して依存することを示している。

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示している。これは、同じくステルスマーケティングを対象に異なるモデルを設定しており、ステルスマーケティングの存在を把握し、ステルスマーケティングを割り引いて財の質を予想する消費者が仮定される。均衡では、高品質な財を供給する企業ほどより多く不正を行うケースと、逆に低品質な財を供給する企業ほど多く不正を行うケースの両方が存在しうる可能性を示され、どちらの均衡が成立するかが、予想された質を評価する消費者の効用関数により決定される需要関数に依存することが示されている。

近年では、Yasui (2020) がよりレビューシステムの特徴を反映した理論モデルによる分析を行っている。ここでは、企業が財の質や商品の現在の評価に応じてその都度ダイナミックにフェイクレビューの量を調整する。消費者は、こうした企業のフェイクレビューを考慮して財の質を予想する。均衡では、高品質な財を供給する企業ほどより多く不正を行うケースを行う上に、それが高品質な企業のシグナルとなるために、フェイクレビューはむしろ消費者に便益をもたらすという結果が導かれている。一方で、騙されやすいナイーブな消費者の存在も考慮され、フェイクレビューの規制が彼らを利することにも言及している。

高品質な商品を供給する企業ほどレビューの不正操作に手を染めるとするのは直観的でない。しかしながら、Dellarocas (2006) や Yasui (2020) はそうした可能性を支持している。加えて、Nelson (1970) や Milgrom and Roberts (1986) は、広告の分析ではあるものの、レビューの不正操作は一種のプロモーションとして考えることもできる。これらの結果からは、レビューの不正操作は一概に有害とはいえないため、実証研究による定量的な評価の必要性がいえる

また、Mayzlin (2006) や Yasui (2020) は、レビューの不正操作がもたらす被害が、消費者の異質性に依存することを指摘している。また、Dellarocas (2006) は消費者の予想された質への評価が、需要関数に影響し、どちらの均衡が成立するかを決めることを示している。これらの結果からは、消費者がどのようにフェイクレビューに反応するかについてのエビデンスの必要性が伺える。

2.2 既存の実証研究とその概観

レビューの不正に関わる実証研究は、米国において、米国ホテルレビューサイトにおける不正操作対象とした Mayzlin et al. (2014) を嚆矢として、ボストンのレストランのレビューを分析した Luca and Zervas (2016)、近年では米 Amazon のフェイクレビューを対象とした He et al. (2021) が存在する。これらは、レビューの不正操作の主体がどのようなもので、どのようなインセンティブに基づいて不正操作を行っているかについて、実証分析を通じ調査している。これらの研究について簡易にまとめたのが以下の表 1 となる。

表 1: レビューの不正操作に関する実証研究

	Mayzlin et al. (2014)	Luca and Zervas (2016)	He et al. (2021)
対象サイト	Tripadvisor&Expedia	Yelp	Amazon
産業	ホテル	レストラン	特定産業でない
データの種類	クロスセクション	パネル	パネル
フェイクの把握	代理変数	代理変数	製品レベルで把握
不正操作の種類	自社高評価偽装・他社攻撃	自社高評価偽装・他社攻撃	自社高評価偽装
研究の範囲	不正の主体	不正の主体	不正の主体・不正の影響

レビューの不正操作に関して、実証面で常に付きまとう問題として、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 and Zervas (2016) は、フェイクを直接観察することはできず、代理変数を用いるアプローチをとっている。一方で、He et al. (2021) 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。

一方で、He et al. (2021) は、自社製品を高く装うために行われる正のフェイクレビューのみ扱っている。対し、Mayzlin et al. (2014) と Luca and Zervas (2016) は、競合相手から投稿されていると思われる、負のフェイクレビューについても扱っている。これは、後者の 2 つが対象とするのが、ホ

テルやレストランといった、競合相手が同じ地域に限定されやすいことに起因する。一方で、前者は Amazon を対象としているため、潜在的に多くの競合相手が存在するうえ、どの商品が競合関係にあるか定かではない⁴。

これらの研究は、レビューの不正操作の主体とそのインセンティブに焦点を当てており、不正操作がもたらす被害を定量的に評価するまでには至っていない。ただし、He et al. (2021) は、フェイクレビューの投稿が開始されてからセールスランクが上昇することや、欺かれて望まぬ購入をしたと思わしき低評価レビューが増加することを示しており、限定的ではあるものの、レビューの不正操作がもたらす影響について推定している。

全体としては、フェイクレビューの全貌を把握するには至っておらず、エビデンスの蓄積が十分な水準にあるとはいえない。特に、また、これらの研究は米国を対象としており、特に日本におけるレビューの不正操作の実態については未だ不明瞭なことが多いのが現状である。

2.3 Mayzlin et al. (2014)

Mayzlin et al. (2014) は Tripadvisor と Expedia における米国ホテルのレビューを対象に、どのような企業が、特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、彼らは二つのレビューシステムの違いを利用している。

代理変数

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 Expedia を通じて予約した消費者のみがレビューが可能であったのに対し、Tripadvisor においては、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 にフェイクレビューを投稿するよりも、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不正行為が行われている度合いの代理変数として扱っている。これは以下の (1) で表される。

$$Y_{ij} = \frac{NstarReviews_{ij}^{TA}}{TotalReviews_{ij}^{TA}} - \frac{NstarReviews_{ij}^{Exp}}{TotalReviews_{ij}^{Exp}} \quad (1)$$

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。 $NstarReviews$ 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。

Tripadvisor に自社を高く評価するフェイクレビューを投稿することで高評価を装うなら、 N が 4 や 5 の時に第一項が高くなることが考えられる。一方、競合他社から攻撃目的での低評価のフェイクレビューが投稿された場合、 N が 1 や 2 の時に第一項が増えることが考えられる。

⁴He et al. (2021) は Amazon においては、競合相手を攻撃するレビューはコストが大きく、効果が限定的であることも示している。ただし、フェイクレビューによる攻撃を受けた事例は存在し、例えば、Anker はフェイクレビューによる攻撃についてプレスリリースを出している (現在は削除。以下で確認は可能。https://news.infoseek.co.jp/article/prtimes_000000227_000016775/)

モデル

彼女らは、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) の代理変数 Y_{ij} を、ホテルの属性に回帰することで⁵、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。

$$Y_{ij} = X_{ij}B_1 + OwnAF_{ij}B_2 + Nei_{ij}B_3 + NeiOwnAF_{ij}B_4 + \sum \gamma_j + \epsilon_{ij} \quad (2)$$

ここで、 X_{ij} はレーティング⁶や施設の有無などのホテルの属性、 $OwnAF_{ij}$ はホテルが独立系（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⁷。 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であり、 $NeiOwnAF_{ij}$ は近隣のホテルが非チェーンか否か、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。 γ_j は地域の固定効果であり、 ϵ_{ij} は誤差項である。

主な関心の対象⁸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代理変数として、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の推定の結果、近隣に他のホテルがあるホテルは有意に約3%ポイント分 Tripadvisor での低評価レビューが多くなっている。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、さらに有意に1.7%ポイント分差が大きくなる。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る場合、2.5%ポイント分差が少なくなる。

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された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を被説明変数とした場合、ホテルが独立系の場合有意に2.4%ポイント分 Tripadvisor での高評価レビューが多くなっている。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、3.1%ポイント分差が少なくなる。

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みせかける傾向を示している。逆にマルチユニットオーナーは自社のレビューを不正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言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いえ。

以上の結果からは、低評価のフェイクレビューにより近隣の競合他社を攻撃するホテルの存在が示唆される。さらに、非チェーンやかつ他にホテルを所有しないオーナーのホテルほどフェイクレ

⁵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。

⁶レビューとは異なる、所謂“星”にあたる公式のレーティング。

⁷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、チェーン系だが他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。

⁸ホテルの属性は Tripadvisor と Expedia の得点分布の差をコントロール変数として用いられる。

ビューによる攻撃を行いやすい。また、こうしたホテルは、フェイクレビューにより自社を高評価の偽装することも行きやすいことが示唆される。これは、非チェーンのホテルやオーナーが他にホテルを所有しない場合、チェーンやオーナーの他に所有するホテルの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とと整合的である。

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⁹については直接的なエビデンスを示さない。しかしながら、近年の Amazon におけるレビューの不正操作について、高品質な財の供給者がフェイクレビューを利用していたとしても、フェイクレビューのコストが高いと考えられる有名企業は不正操作に関わりにくいことが示唆される。

2.4 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 のパネルデータを用い、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する。彼らは、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代理変数に利用する。

代理変数

これは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 がフィルタリングとは別に行ってい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代理変数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。

$$Fliterd_{ik} = b_j + z'_{ik}\beta + \epsilon_{ik} \quad (3)$$

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、 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、5 点や 1 点のレビューほど有意にフィルタリングされやすい傾向が示された。こうした傾向は、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代理変数として用いる。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

⁹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。レーティングが、得点分布の差に影響を与える観察されない要素と送選しない場合、質の高いホテルほど、自社を高く評価されているように装うレビューの不正操作を行いにくい傾向がみられる。

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、代理変数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。

$$f_{it}^* = x'_{ij}\beta + b_i + \mu_t + \epsilon_{it} \quad (4)$$

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であり、 $t-1$ 期の得点分布等や、競合レストランの出現といった変数が用いられる。 i はレストランの固定効果であり、 μ_t は t 期の固定効果、 ϵ_{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。

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

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、 α_0 はレビューがフェイクでない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。

$$\begin{aligned} f_{it} &\equiv \sum_k f_{itk} = \sum_k [\alpha_0(1 - f_{itk}^*) + (\alpha_0 + \alpha_1)f_{itk}^* + u_{itk}] \\ &= \alpha_0 n_{it} + \alpha_1 f_{it}^* + u_{it} \end{aligned} \quad (6)$$

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

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

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 β 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。 α_1 は正のため、 β 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係る誤差項である ϵ_{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 μ_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) ように定式化する。

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

ここで、 z'_{ikt} はレビューのフィルタリングされやすさに係る属性であり、 \tilde{u}_{ikt} は独立な誤差項である。(8) を代入して、以下の (9) が得られる。

$$f_{it} = \alpha_0 n_{it} + \alpha_1 (x'_{it}\beta + b_i + \mu_t + \epsilon_{it}) + \sum_k z'_{ikt}\gamma + u_{it} \quad (9)$$

(9) は、(7) にフィルタリングに係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 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 5 の低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しき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¹⁰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、有意な影響をあた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。自社への正の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伺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、質の低い財を供給する企業ほどレビューの不

¹⁰ 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た下では、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。

正行為に関わりやすいことが示唆される。

2.5 He et al. (2021)

He et al. (2021) 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(Ammazon.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. (2021) 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¹¹。

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。

フェイクレビュー商品の特徴

まず、こうしたフェイクレビューを募集している商品について、そうでない通常の競合商品¹²と比べ、どのような傾向があるか、記述統計を示している。競合商品の商品年齢¹³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、安価な商品な傾向がうかがえる。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少雨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 73292 位なのに対し、後者は約 90000 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¹⁴も、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。これは、後述する、本稿における調査結果と整合

¹¹ 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。

¹² ここで競合商品は、フェイクレビュー商品一つにつき、同じ検索ページにある上位 2 つの商品が選ばれた、合計約 2700 個の商品となっている。ただし、フェイクレビュー商品と比較されるのは、20 万個以上の商品プールから得られた記述統計量となっている。He et al. (2021) はワーキングペーパーの段階であり、前のバージョンのワーキングペーパーから変更箇所の更新を忘れるなど、何らかのミスと思われる。

¹³ 発売からの経過週数となる。

¹⁴ 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。

的である。また、他に販売する商品の価格やレビューの数などは、フェイクレビュー商品と近い値となっている。

Amazon 側の対応

He et al. (2021) 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 は不正と思わしき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後に削除される確率は 43% と高い削除率を示した。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 5 点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、文章が長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については、後手に回っていることが分かる。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変化

レビューの不正操作の影響を推定するにあたり、まず、前後比較による大まかな把握を行っている。具体的には、フェイクレビューの募集の前と後の商品のセールスランクやレーティングの変化をみる。これにあたっては、フェイクレビューが募集された週を 0 として、その前後週における商品ごとのそれらの変数の平均をとる。この際、短期的な影響と長期的な影響の両方を調査するため、前者についてはフェイクレビューの募集が始まった週を、後者についてはフェイクレビューの募集が終わった週を、それぞれ基準として、前後の週における変化をみる。

まず、レビューとレーティングについて、これを図示したのが、図 (1) と図 (2) である。横軸がフェイクレビューの募集開始/終了した週を 0 とした週左縦軸が投稿されたレビューの点、中央縦軸が投稿されたレビュー数、右縦軸がレビューの累積のレーティングとなっている。フェイクレビューの募集開始からしばらくはすべて大幅に増加するものの、フェイクレビューの募集が終了するといずれも減少する。これらのフェイクレビュー募集のタイミングと、自然なレビューのトレンドが相関する可能性があり、フェイクレビューによる影響だとは一概にいえぬものの、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。

特に、レビュー募集終了後に、平均的な点数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伺える。フェイクレビューによる被害の推定は困難だが、低評価のレビューは、低評価の消費者が望まぬ購入をした消費者と想定されるため、ある種の被害代理変数として用いることができる。

こうした代理変数として、星 1 のレビューの割合を縦軸にして図示したものが、以下の図 3 となる。フェイクレビューの募集が終わってから、星 1 レビューの割合は大きく増加する。このことから、フェイクレビューが実際に消費者を欺く被害が生じていることが伺える。

セールスランクなどについて、これを図示したのが、図 (4) と図 (5) である。左縦軸がセールスランクの対数、中央縦軸が売上量¹⁵、右縦軸が検索順位となっている。フェイクレビューの募集が始

¹⁵厳密な売上ではなく、Amazon.com における在庫の変動から推定されたものとなる。

Figure 4: 7-day average ratings, 7-day average number of reviews, and cumulative average ratings before and after fake reviews recruiting begins. The red dashed line indicates the last week of data before we observe Facebook fake review recruiting.

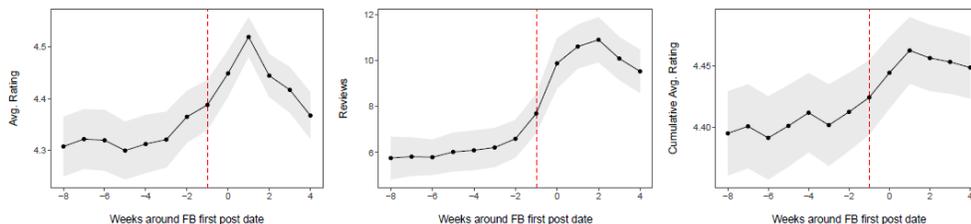


図 1: フェイクレビュー募集開始前後のレビューの変化 I(He et al., 2021, p19, Figure 4.)

Figure 8: 7-day average number of average ratings, reviews, and average share of one-star reviews before and after fake reviews recruiting stops. The red dashed line indicates the last week of data in which we observe Facebook fake review recruiting.

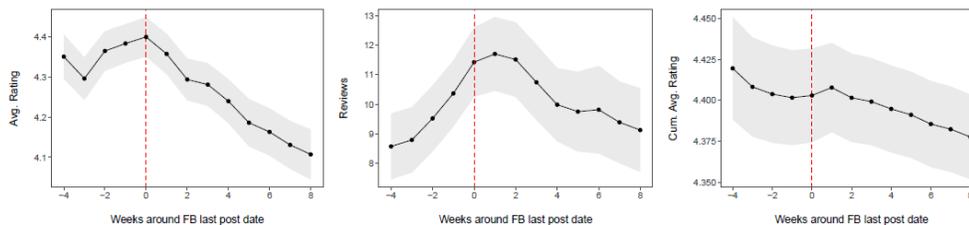


図 2: フェイクレビュー募集終了前後のレビューの変化 (He et al., 2021, p23, Figure 8.)

Figure 14: 7-day average share of one-star reviews before and after fake reviews recruiting stops. The red dashed line indicates the last time we observe Facebook fake review recruiting.

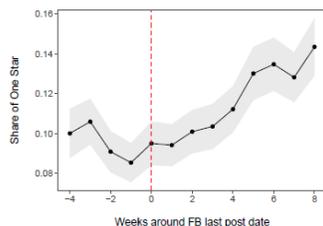


図 3: フェイクレビュー募集終了前後の低評価レビュー割合の変化 He et al., 2021, p38, Figure 14.)

Figure 5: 7-day average sales rank before and after fake reviews recruiting begins (left), sales in units (center), and keyword search position (right) before and after fake reviews recruiting begins. The red dashed line indicates the last week of data before we observe Facebook fake review recruiting.

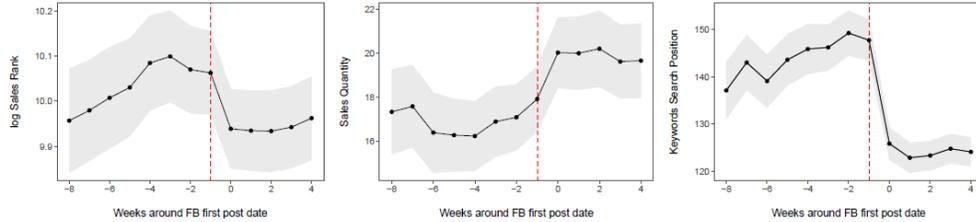


図 4: フェイクレビュー募集開始前後の売上の変化 II(He et al., 2021, p21, Figure 5.)

Figure 9: 7-day average sales rank, sales in units, and keyword rank before and after fake review recruiting stops. The red dashed line indicates the last week of data in which we observe Facebook fake review recruiting.

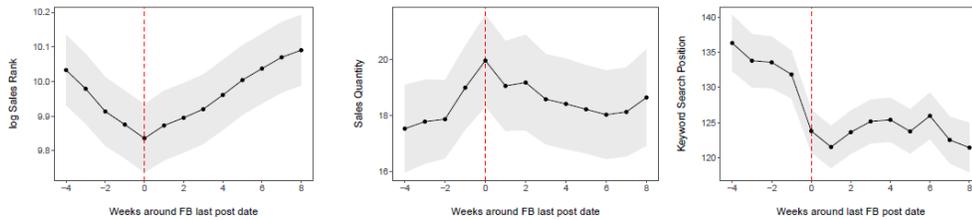


図 5: フェイクレビュー募集終了前後の売上の変化 II(He et al., 2021, p24, Figure 9.)

まるのは、セールスランクや検索順位が落ちているタイミングとなり、募集により大幅に順位を改善させ、売上を伸ばしている。しかしながらフェイクレビューの募集が終わると、セールスランクと売上は大幅に悪化する。検索順位は大きく反転せず、セールスランクほど大きく変化しない。フェイクレビュー自体が売上を伸ばす働きを持つ¹⁶ため、レビューが操作された影響と識別できないものの、フェイクレビューにより、一時的に売上が上昇していることが伺える。

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、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. (2021) 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しき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

¹⁶消費者に商品を購入させたうえでレビューを投稿させるプロセスを踏む。

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。

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

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。 α_i と τ_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 $After_{it}$ の係数である β_1 と $Treated_i \times After_{it} + \alpha_i$ の係数である β_2 である。 β_1 はフェイクレビューを募集することによる効果で、 β_2 は DID 推定量に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の結果、 β_2 は有意に正、 β_1 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がレビュー投稿されていることが伺える。

セールスランクの対数を被説明変数とした場合、 β_1 は有意に正、 β_2 は有意に負となっており、 $\beta_1 + \beta_2$ 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 β_2 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伺える。

He et al. (2021) が示した結果は、レビューの不正行為を行う企業がどんなものかについて、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。特に、フェイクレビューの募集が終わったのちに、フェイクレビュー商品が評価を落とし低評価のレビューが増加していることから、低品質な商品が高く評価されているように装うことが目的であると考えられる。また、フェイクレビューがもたらす影響についてもエビデンスを示している点でも差別化される。特に、分析は限定的ながらも、フェイクレビューによる被害を定量的に示している。

3 日本における分析結果

前述したように、特に国内におけるレビューの不正操作については不明瞭なことが多い。そこで、本節では、国内における実態を明らかにすべく、Amazon.co.jp を対象に実証分析を行う。分析にあたっては、Luca and Zervas (2016) の代理変数を用いる手法を踏襲する。

3.1 データ

Amazon レビュー

Amazon におけるレビューは、レビュー文と評価点、他のユーザーからの投票に加え、投稿日付や Amazon での購入の有無など、様々なメタ情報が不随する。これらのレビューは集計され、総合点とレビュースコアの分布が表示される。総合点は、単純なレビュースコアの平均ではなく、Amazon 側のアルゴリズムで計算されている¹⁷。

Amazon においてレビューを投稿する¹⁸ためには、Amazon のアカウントと、一定以上の累計購入額 (一年間につき五千円) が必要となる。そのため、多数のアカウントを作成して大量のフェイクレビューを投稿することは、コスト面で非現実的である。

¹⁷ 購入者のレビューや文章付きのレビューのウェイトが大きい。

¹⁸ レビュー文章なしの点数のみのレビューも投稿可能。

において多く、代理変数としての妥当性が担保されている。

Yelp と異なり、Amazon においては削除されたレビューが観察できなくなる。そのため、データの取得間隔が長期間に渡る場合、レビューの削除を観察できない可能性がある。具体的には、データを観察していない期間において、レビューが投稿され、その期間中に削除が行われた場合、レビューの削除が観察できない。故に、高頻度で商品データを取得する必要がある。

こうした高頻度でレビューを取得することを前提とした場合、あまりにも多くの商品を取得対象とすることは、時間的制約から困難である。そこで、本研究では、Amazon の売上上位と考えられる商品を対象とする。具体的には 11 カテゴリにおける、Amazon の売れ筋ランキングに掲載されている商品に加え、検索順位が上位なものを対象とした。

取得プロセスは、以下のように行われる。まず日次で Amazon の売れ筋ランキングならびに検索順位 20 ページ分の商品チェックし、これらに一度でも入った商品をデータ取得対象²⁰に加える。それと並行して、データ取得対象に含まれている商品のレビューページを随時取得する。²¹

本稿では、2021 年 1 月から 8 月までこれらのプロセスを繰り返し、レビューデータを取得した。これらに加え、2021 年 9 月に商品データならびにレビュアーのデータ、2021 年 11 月に個別のレビューの削除状況²²を一括で取得した。最終的に、約 12000 の商品に対し、約 118 万のユニークなレビュー²³を取得した。

レビューについて取得したのは、レビューの点数や投稿日時といったメタデータに加え、レビューテキストやタイトル、写真の有無といったレビューの属性。さらにはレビュアーのレビュー投稿数などのレビュアーの統計である。商品データについては、メーカーが Amazon に専用ストアをもつか否かや、セラーの情報等について取得した²⁴。

3.2 レビューレベルの分析

まずレビューレベルデータの分析により、Amazon における削除の傾向をつかむ。このデータは、各レビューについて観測された最後の時点のもので構成されたレビューレベルのクロスセクションデータとなる。これは、Luca and Zervas (2016) と同様に、のちに削除プロセスのコントロールに用いられる。

レビューの記述統計

取得したレビューについて、レビューレベルの記述統計は以下の表 2 のようになる。得点については各点数のダミーとなっている。取得したレビューデータは、観測期間より前に投稿されたものも含まれる。そのため、これらのデータについては、サバイバルバイアスが生じている可能性がある

²⁰ Amazon における商品 ID で管理する

²¹ ランキングや検索順位については日次で取得するのに対し、レビューは随時取得している。これは具体的にはすべての取得対象商品のレビューを順に取得次第、また取得対象の商品のデータを取得することを繰り返す。この取得感覚はおおよそ日次となっている

²² 本稿ではレビューの取得を製品ページレベルで取得している。時にレビューは削除されていないにも関わらずレビューページから消える場合がある。そのため、別個に削除状況を取得している。この際、レビューの個別ページが 404 ステータスコードを返した場合に削除とみなす。

²³ 取得対象となった ID 数自体は約 24000ID である。ただし、Amazon においては色違いやセット商品などのバリエーション商品も異なる ID により扱う。レビューは一つの商品 ID に紐づけられるのではなく、このような複数のバリエーション商品に紐づけられる。本稿では、複数にわたるバリエーション商品商品については、代表的なもの、具体的には最も多くのレビューが紐づけられているもののみ扱う。

²⁴ 写真の有無については取得を途中から始めたためいくつか欠損している。また、セラーについても、商品によってはデータが得られず欠損している。

る。そこで、レビューがのちに削除されるか否かだけでなく、観測期間内に投稿されたか否かで分けて統計量を推定している。

観測期間に投稿されたものは全体の約 32% を占める約 37 万 5 千レビューである。全レビューのうち、約 13% が削除されている一方で、観測期間内に投稿されたレビューは約 20% が削除されており、サバイバルバイアスが生じているのが分かる。そのため、観測期間内に投稿されたレビューについてのみ扱う。

削除されたレビューと削除されていないレビューで大きく差があるのは、まず、得点があげられる。削除レビューは極端に 5 点が多く、74% を占める。非削除レビューでも 5 点が最も多いが 55% にとどまっている。一方で 1 点のレビューについては削除されているものの 6% を占めており、これは削除されていないものの 11% より少ない。ただし、5 点のレビューほど極端な傾向はしていない。He et al. (2021) と同じ傾向であり、自社商品の評価を高く装う目的でフェイクレビューが投稿されていることが示唆される。

写真の有無については、削除されたものの 18% が写真つきであるのに対し、削除されていないものは約 10% となっている。He et al. (2021) と類似した結果となっている。一方で、テキストやタイトルの長さに大きな違いはない。レビューアの統計情報については、削除レビューのほうが大幅に投稿数も得票数も少ない結果になっている。Luca and Zervas (2016) の回帰分析によりしめされた結果と同様である。

購入者の比率は両者で高いものの、削除レビューのほうが 3% 多くなっている。ただし、非購入者のレビューは反映される事前のフィルタリングが厳しい可能性があるため、一種のセクションが働いている可能性がある。デフォルトアイコンについても両者で高い。ただし、削除レビューのほうが若干割合が少なくなっている。

投稿時発売からの経過週数²⁵に関しては削除レビューが非削除レビューの半分ほどとなっている。削除されているのは販売初期に投稿されたレビューが多い。

表 2: レビューレベル記述統計

	削除		非削除		削除 (観測期間内)		非削除 (観測期間内)	
	平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
*5	0.69	0.46	0.52	0.50	0.74	0.44	0.55	0.50
*4	0.15	0.36	0.23	0.42	0.12	0.33	0.21	0.41
*3	0.06	0.23	0.09	0.29	0.04	0.20	0.08	0.28
*2	0.03	0.18	0.05	0.22	0.03	0.16	0.05	0.22
*1	0.07	0.26	0.10	0.30	0.06	0.24	0.11	0.31
得票数	2.46	22.28	3.19	25.67	1.66	16.07	1.89	14.95
購入者	0.94	0.24	0.91	0.29	0.95	0.22	0.92	0.28
写真の有無	0.15	0.35	0.06	0.25	0.18	0.38	0.10	0.30
レビューア投稿数	34.19	250.15	87.46	423.63	29.00	258.28	80.37	445.00
レビューア得票数	93.95	768.16	239.33	1250.78	72.57	683.34	186.09	1154.90
アイコンがデフォルトのもの	0.86	0.34	0.88	0.32	0.87	0.34	0.90	0.31
レビューテキストの長さ	118.71	137.52	132.43	184.60	115.06	125.73	118.13	161.99
レビュータイトルの長さ	9.84	6.42	10.29	7.10	9.78	6.28	10.05	6.97
レビューテキストひらがな率	0.51	0.11	0.50	0.12	0.51	0.12	0.49	0.13
投稿時の発売からの経過週数 (レビュー基準)	40.54	61.15	89.75	111.48	33.59	49.46	74.06	103.90
投稿時の発売からの経過週数 (取り扱い開始日基準)	37.91	61.73	71.00	120.64	31.63	44.72	64.82	95.03
N	158562		999467		75541		299095	

²⁵ Amazon における取り扱い開始日は発売日とずれている場合や取得できないケースがある。そのため、最も古いレビューを発売日とした場合と、取り扱い開始日を発売日とした場合両方について検討する。

レビューレベルの特徴

こうしたレビューの属性について、Luca and Zervas (2016) と同様に、商品の固定効果をコントロールしたうえで、削除されやすい属性について回帰分析を通じ明らかにする。モデルは (3) と同様に、削除されたか否かを被説明変数とした線形確率モデルとなる。推定結果は以下の表 (3) のようになる。標準誤差は商品をクラスターとしたクラスターロバストな標準誤差から計算されている。

固定効果をコントロールしたうえで、高得点レビューほど削除されやすい。こうしたことから、フェイクレビューにより評価を高く装う目的がうかがえる。加えて、その他多くの変数についても、記述統計と同様の傾向を示している。他に活発に活動しているレビュアーは有意に削除されにくく、発売から間もないレビューほど有意に削除されやすい。

特徴的なのは、レビューテキストの長さをコントロールしたうえで、ひらがな率が高いものが有意に削除されやすい点である。独特のテキストの癖（稚拙な日本語等）に関連している可能性がある。

また、デフォルトアイコンのレビューが有意に削除されにくいのも注目になる。直観的には、Amazon におけるデフォルトアイコンの方が、一種のいわゆる”捨てアカウント”のように感じられる。このことから、あえてフェイクレビューでないように見せかけるように、フェイクの手段が巧妙化しており、そのうえで Amazon 側もそれに対応している可能性が考えられる。

同様なことがレビューテキストの長さや、写真の有無にもいえる。写真がなく短いレビューのほうがフェイクと受け取られやすいというのが一般的であると考えられるため、その裏を突く形でフェイクレビューが行われている可能性が示唆される。また、これらの結果は、He et al. (2021) とも共通している。

3.3 プロダクトレベルの分析

次に、プロダクトレベルでの傾向をみる。このデータは、観測期間内に投稿されたレビューに限定したレビューレベルデータを商品ごとに集計したうえで、セラーの属性を中心とした商品属性を接続したものである。ここでは、Mayzlin et al. (2014) を参考にした代理変数を用い分析を行う。

記述統計

レビューレベルの記述統計が以下の表 (4) である。セラー²⁶データに関しては得られない商品が 3 割ほどあるため、それぞれ記述統計を記す。ただし、全般的に大きな違いはない。レビューの点については、削除/非削除と 4-5 点/1-3 点のレビューの数を合計している。特に点 4-5 の削除レビューがフェイクレビューの疑いが高く肝要となる。

メーカーストアの割合は 43% 程になっている。大手ブランドほど、専用のストアがあることが多いためこれは一種のブランド品の代理変数となる。また、セラーの情報が得られた商品のうち、中国系セラー²⁷が 42% となっている。

プロダクトレベルの特徴

レビューの不正操作の被害は、不正を行う企業が供給する商品の質に依存する。これを明らかにするにあたって、直観的に考えられる方法は、商品の削除されたレビューの数を、削除されていないレビューに回帰する方法である。ただし、これには、レビューの数が多き商品は両者が高くなる可能性が高いという内生性の問題がある。

²⁶商品ページにおいて示される販売元をセラーとしている。

²⁷セラーの電話番号と住所から判断する。

表 3: レビューレベルの回帰分析結果

	(1)	(2)
	deleted	deleted
*5	0.04*** (14.93)	0.04*** (15.73)
*4	0.01*** (3.31)	0.01*** (3.48)
*2	-0.00 (-0.72)	-0.00 (-1.62)
*1	0.00 (0.29)	0.00 (0.08)
得票数	-0.00*** (-5.29)	-0.00*** (-4.56)
購入者	0.03*** (6.60)	0.03*** (7.40)
写真あり	0.04*** (9.26)	0.04*** (10.14)
レビュアー投稿数	-0.00* (-2.08)	-0.00* (-1.96)
レビュアー得票数	-0.00*** (-6.55)	-0.00*** (-6.83)
レビューテキストひらがな率	0.03*** (8.29)	0.03*** (8.42)
アイコンがデフォルトのもの	-0.03*** (-13.64)	-0.03*** (-14.50)
レビューテキストの長さ	0.00*** (8.93)	0.00*** (9.06)
レビュータイトルの長さ	0.00*** (5.38)	0.00*** (4.78)
投稿時の発売からの経過週数 (レビュー基準)	-0.00*** (-3.79)	
投稿時の発売からの経過週数 (取り扱い開始日基準)		-0.00*** (-3.82)
切片	0.15*** (13.59)	0.15*** (13.80)
N	308146	343244

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

表 4: プロダクトレベルの記述統計

	全商品		セラーデータあり	
	平均	標準偏差	平均	標準偏差
* 4-5(削除)	6.11	43.20	6.71	48.19
* 4-5(非削除)	21.06	68.88	22.87	73.15
* 1-3(削除)	0.92	6.38	0.97	6.79
* 1-3(非削除)	6.77	17.55	7.19	17.43
メーカーストアあり	0.43	0.49	0.43	0.49
中国系セラー			0.42	0.49
セラー評価 (点)			92.79	10.10
N	10748		7619	

そこで、Mayzlin et al. (2014) を参考にした代理変数を作成し、分析を行う。商品 i に対する代理変数は以下の (11) のようになる。

$$m_i = \frac{NstarDeletedReviews_i}{DeltedReviews_i} - \frac{NstarUndeletedReviews_i}{UndeletedReviews_i} \quad (11)$$

$NstarDeletedReviews_i$ は特定の点の削除レビューを、 $DeltedReviews_i$ は削除レビュー全体を表す。 $NstarUndeletedReviews_i$ は特定の点の非削除レビューを、 $UndeletedReviews_i$ は非削除レビュー全体を示す。Mayzlin et al. (2014) と同様に正のフェイクレビューでレビューを操作している場合、削除されたものにフェイクレビューが多いならば、 N を 4 や 5 としたときに右辺第一項が大きくなりやすい。これを商品の属性に回帰することで、不正を行い易い商品の属性を明らかにすることができる。

推定結果は以下の (5) となる。ゼロ除算が発生しうる関係上、サンプルサイズは小さくなっている。1-3 点の非削除レビューは有意に正、4-5 点の非削除レビューは有意に負となっている。低評価の非削除レビューが多い低品質な商品の供給者ほど不正を行い易く、高評価の非削除レビューが多いような高品質な商品の供給者は不正を行いにくことが示唆される。

メーカーストアは有意に負となっており、専用のストアがあるようなブランド品ほど不正行為に関わりにくいことが示唆される。Mayzlin et al. (2014) と Luca and Zervas (2016) が示すような、不正操作が露呈した場合にブランドへのダメージがあるような企業はフェイクレビューのコストが大きく不正を行いにくことと整合的である。また、中国系セラーは有意に正となっている。これは各種報道や、He et al. (2021) の結果とも整合的である。

表 5: パネルデータ記述統計

	(1)	(2)
	mayzline	mayzline
* 1-3(非削除)	0.00*	
	(2.29)	
* 4-5(非削除)		-0.00***
		(-3.48)
メーカーストアあり	-0.08***	-0.07***
	(-5.35)	(-4.84)
中国系セラー	0.07***	0.07***
	(4.63)	(4.64)
セラー評価 (点)	-0.00	-0.00
	(-0.92)	(-0.65)
切片	0.18*	0.18*
	(2.35)	(2.29)
N	2539	2539

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.4 パネルデータ分析

次に、パネルデータを用いて、レビューの不正操作が、時系列で変動するどんなインセンティブに基づくか調査する。このパネルデータは、観測期間内に投稿されたレビューに限定したレビューレベルデータを商品と週ごとに集計した週次のパネルデータとなっている。これらはサバイバルを排

すため、観測期間に発売された商品のみを対象としている。先述したように、発売日が正確に把握できないため、レビューベースの発売日と取り扱い開始日ベースの発売日を代理変数として用いる。

記述統計

どちらの発売日を採用しても大きく変化はない。全体で各商品は週あたり約 1.5 本のレビューが投稿される。そのうち、約 0.4 本が削除された高評価レビューとなる。おおよそ 20% のレビューが削除対象となっている。商品年齢については、取り扱い開始日ベースの場合、平均約 18.3 週となり、レビューベースの場合は約 15.3 週となる。

表 6: パネルデータ記述統計

	取り扱い開始日ベース		レビューベース	
	平均	標準偏差	平均	標準偏差
* 4-5(削除)	0.40	3.22	0.32	2.81
* 4-5(非削除)	0.93	4.20	0.76	3.80
* 1-3(削除)	0.04	0.39	0.03	0.34
* 1-3(非削除)	0.23	0.87	0.20	0.81
取り扱い開始日ベース商品年齢	18.28	8.40		
レビューベース商品年齢			15.33	9.35

時系列での変化

Luca and Zervas (2016) と同様に、商品の固定効果をコントロールしたうえで、時系列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レビューがどのように反応するかみる。モデルは (8) と同様であり、削除レビューの投稿数を被説明変数として、レビューの属性の和をコントロールとして用い推定を行う。インセンティブに関わる変数としては、非削除レビューの投稿数の 1 週ラグと消費年齢を用いる。

推定結果は以下の表 (7) のようになる。標準誤差は商品をクラスターとしたクラスターロバストな標準誤差から計算されている。固定効果ならびにコントロール変数となるレビュー属性は省略している。どちらの商品年齢を用いたパターンでも結果に大差はない。被説明変数には、削除された 4-5 点レビューの投稿数を用いている。加えて、比較のために削除されていない 4-5 点レビューの投稿数についても被説明変数としている。

推定の結果、前期に高評価の 4-5 点の非削除レビューが増加すると、有意に 4-5 点の高評価の削除レビューが減少する。言い換えれば、前期にフェイクレビューである可能性が低い高評価のレビューが減少するようなときに、フェイクレビューと疑わしい高評価のレビューが増加する。つまり、評価の低下に対して、経済的インセンティブに沿う形で、フェイクレビューと思わしきレビューが投稿されている可能性が示唆される。これは、Luca and Zervas (2016) と整合的である。

これは、高評価の非削除レビューを被説明変数としたケースと比較した場合、より顕著である。この場合、前期に高評価の非削除レビューが増加すると、高評価の非削除レビューが有意に増加する。さらに、前期の低評価の評価の非削除レビューが増加すると、高評価の非削除レビューは有意に減少する。高評価が高評価を、低評価が低評価呼ぶ形で前期のレビューに反応する。これは経済的インセンティブと関係なくレビューに反応しているといえる。

また、年齢についても、レビューベース開始年齢の場合 10% 有意水準で、取り扱い開始日ベースの場合は有意水準 5% で、それぞれ有意に負となっている。商品が若い、つまり発売から日が経つ

ていないほど高評価の削除レビューが投稿されやすい。これは商品の販売初期に評価を高く装うことを目的としていることが伺える。

表 7: パネルデータ推定結果

	(1)	(2)	(3)	(4)
	* 4-5(削除)	* 4-5(非削除)	* 4-5(削除)	4-5(非削除)
* 4-5(非削除・1週ラグ)	-0.26*** (-4.09)	0.28*** (4.66)	-0.29*** (-4.31)	0.30*** (4.99)
* 1-3(非削除・1週ラグ)	0.00 (0.08)	-0.16*** (-4.73)	0.03 (0.79)	-0.18*** (-4.25)
対数レビュー本数 (非削除・ラグ)	0.05 (0.43)	-0.16 (-1.50)	0.04 (0.46)	-0.16 (-1.70)
レビューベース商品年齢	-0.01 (-1.95)	0.01*** (3.83)		
取り扱い開始日ベース商品年齢			-0.01** (-2.71)	0.01*** (5.13)
切片	-0.11 (-0.42)	0.16 (0.64)	-0.14 (-0.45)	0.17 (0.57)
<i>N</i>	68298	68298	50098	50098

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

投稿タイミングの考察

パネルデータ分析の結果からは投稿初期にフェイクレビューを用いて商品を高評価に見せかけている可能性が示唆されている。また、レビューレベルの分析においても、発売から間もないレビューほど削除されやすいという結果が見出されている。

そこで、不正に“手を染めた”タイミングを考える。しかしながら、どの時点でフェイクレビューの募集などを始めたのかは定かではない。そこで、本稿では、代理変数として、4-5の削除レビューの投稿数が一定の閾値を超えたタイミングを、不正なレビューの操作を始めたタイミングとして扱う。

閾値は以下の表 (8) による。これらはそれぞれ、パネルデータにおける、4-5 点の削除レビューの投稿数の平均、平均 +1 標準偏差分、平均 +2 標準偏差分に対応する。

表 8: 不正操作開始の閾値

	実数値	投稿開始年齢 (平均)	投稿開始年齢 (SD)
閾値 1(μ)	1	6.86	5.78
閾値 2($\mu + \sigma$)	4	5.95	4.71
閾値 3($\mu + 2\sigma$)	7	5.88	4.62

図 (8)、図 (9)、図 (10) は、これらの閾値に対応する、不正を始めたタイミングにおける商品年齢²⁸のヒストグラムである。縦軸が商品数、横軸が商品年齢となっている。いずれも比較的販売初期に集中していることがわかる。このことから、販売初期に評価を高く装うことが目的であることが示唆される。

これらの不正の開始前後におけるレビューの投稿量を示したものが、図 (11)、図 (12)、図 (13) である。青の線が削除された 4-5 点の高評価レビューの投稿数の平均を示し、オレンジの線が削除さ

²⁸ 取り扱い開始日ベースのもの。

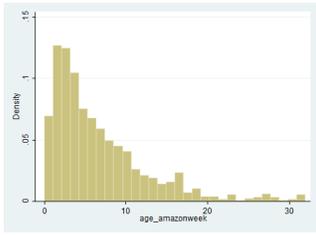


図 8: 不正操作開始週 (閾値 1)

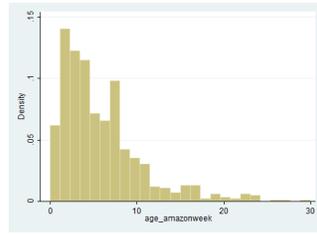


図 9: 不正操作開始週 (閾値 2)

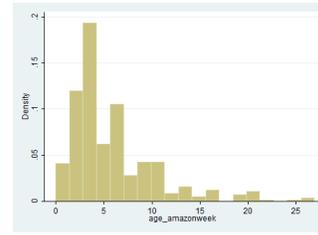


図 10: 不正操作開始週 (閾値 3)

れていない 4-5 点レビューの投稿数の平均を意味する。いずれも削除された高評価レビューは、削除されていないものと比べ開始週に集中している。

図 (14)、図 (15)、図 (16) は、縦軸を投稿量の割合の商品ごとの平均、つまり密度の商品ごとの平均としたものである。具体的には、青の線は、削除された 4-5 点の高評価レビューの分析期間中の投稿数の合計に対し、各週の投稿数の割合を商品ごとに平均をとったものを示す。オレンジの線は分析期間中の 4-5 点の削除されていないレビューの投稿数の合計に対する投稿数の割合の平均を意味する。より顕著に、削除された高評価レビューが不正開始週に集中して投稿されていることがわかる。

これらの結果は、削除された高評価レビューが人為的なものである可能性を支持する。加えて、多くの商品において不正開始時期が販売初期なことから、早期に評価を高く装うことが目的であることが伺える。

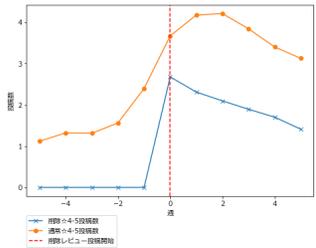


図 11: 不正開始前後投稿量 (閾値 1)

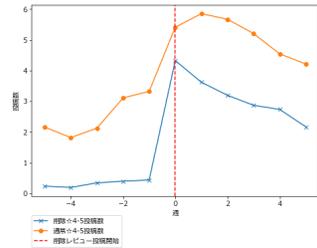


図 12: 不正開始前後投稿量 (閾値 2)

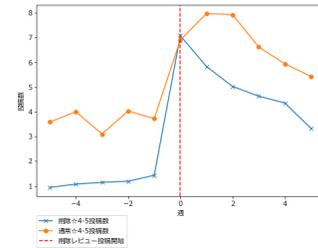


図 13: 不正開始前後投稿量 (閾値 3)

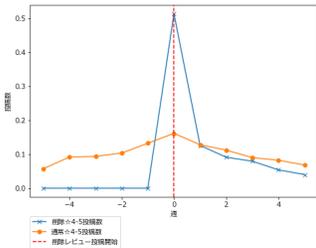


図 14: 不正開始前後投稿割合 (閾値 1)

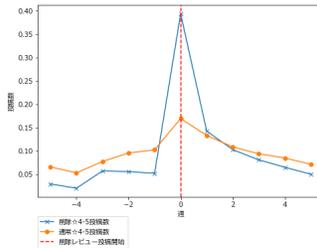


図 15: 不正開始前後投稿割合 (閾値 2)

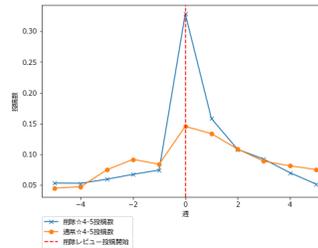


図 16: 不正開始前後投稿割合 (閾値 3)

被害の推定

削除された高評価レビューが早期に極端に集中していることから、特にフェイクレビューが大量に投入されるタイミング以降に高評価により欺かれた消費者が被害にあう可能性がある。実際に He

et al. (2021) でも、フェイクレビューの募集後に欺かれて望まぬ購入をしたと思わしき消費者が増加していた。

そこで、本稿では、He et al. (2021) と同様に、低評価のレビューを被害の代理変数として分析を行う。モデルは (9) に、不正開始タイミング以降の変数を追加した以下の (12) のようになる。

$$y_{it} = \alpha_0 n_{it} + \alpha_1 (x'_{it} \beta_1 + b_i + \mu_t + \epsilon_{it}) + \beta_2 \text{delafter}_{it}^w + \sum_k z'_{ikt} \gamma + u_{it} \quad (12)$$

ここで、 y_{it} は、削除されていない 1-3 点の低評価のレビューである。 delafter_{it}^w は、削除されたレビューの投稿数が閾値を超えた商品かつ、その商品が閾値を超えて w 週以降を示す変数である。関心の対象は delafter_{it}^w の係数 β_2 であり、これが正になる場合、欺かれて望まぬ購入をした被害者の存在が示唆される。 $w = 2$ での推定結果²⁹は表 (9) となる。

全ての閾値で有意に正となっている。これはどの閾値を基準とした不正行為の開始から 2 週間以降は低評価の非削除レビューが増加することを意味する。これは、直接的にレビューの不正行為がもたらす被害を推定したわけではないものの、フェイクレビューの可能性の高いレビューが実際に望まぬ購入をする消費者を増やしていることを示す。

表 9: 被害推定結果

	(1)	(2)	(3)
	* 1-3(非削除)	* 1-3(非削除)	* 1-3(非削除)
フェイク投稿開始 2 週間以降 (閾値 1)	0.09** (3.00)		
フェイク投稿開始 2 週間以降 (閾値 2)		0.15*** (3.68)	
フェイク投稿開始 2 週間以降 (閾値 3)			0.36*** (5.98)
* 4-5(非削除・1 週ラグ)	-0.00 (-0.42)	-0.00 (-0.42)	-0.00 (-0.46)
* 1-3(非削除・1 週ラグ)	0.14*** (6.51)	0.14*** (6.49)	0.14*** (6.40)
対数レビュー本数 (非削除・ラグ)	0.06*** (4.17)	0.06*** (3.90)	0.05*** (3.48)
取り扱い開始日ベース商品年齢	-0.00*** (-5.48)	-0.00*** (-5.30)	-0.00*** (-5.27)
切片	-0.03 (-0.58)	-0.09 (-1.41)	-0.28*** (-3.39)
<i>N</i>	50098	50098	50098

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

本稿の分析は、先行研究と同じく、レビューの不正操作がどのような主体で、どのようなインセンティブに基づくかについて調査するものである。得られた結果は、レビューの不正操作を行う企業が、低品質な財の供給者であり、自社製品を高く評価されているように装うことが目的としており、特に販売初期にフェイクレビューを集中して投稿している可能性が高いことが示されている。

こうした結果は、Mayzlin et al. (2014) や Luca and Zervas (2016) の間接的に示された推定結果をより具体的に指示するものであり、He et al. (2021) と共通の傾向を示す。

²⁹ $w = 3$ や $w = 4$ でも有意に正となる。ただし $w = 1$ の時確は有意とならず、タイムラグがあることが考えられる。

4 今後の展望

本稿は、レビューの不正操作に関連する先行研究を、実証研究を中心にサーベイしたうえで、国内データを用いた Amazon.co.jp におけるレビューの不正操作に関する実証分析の結果を報告した。

こうしたレビューの不正操作がもたらす被害については、不正行為を行っている企業などの主体が重要になる。仮に、高品質な財を供給する企業がレビューの不正を行っている場合、消費者が欺かれる被害は起きにくい。

また、消費者の異質性も被害の大きさに関わってくる。レビューの不正操作を割り引いて考える消費者や、そもそもフェイクレビューに欺かれない消費者が多ければ、被害は小さくなる。

既存のレビューの不正操作に関わる実証研究は、主に前者についてのエビデンスをある程度蓄積してきており、現状、主に質の低い商品を供給する企業がフェイクレビューを行っていることが支持されている。本稿における結果も、同様の傾向を示す。

しかしながら、レビューの不正操作の被害の大きさを直接定量的に評価しているわけではなく、規制にあたってのエビデンスは不足しているというのが現状である。今後の方向性については、いくつかのポイントがあげられる。

消費者の調査

レビューの不正操作がもたらす被害は、先述したように、消費者にも依存する。しかしながら消費者が実際にフェイクレビューにどのように反応するかについての調査は、既存の実証研究が未だに及ばない範囲となっている。

また、消費者の実態の調査は、公平性の意味でも重要になってくる。仮にフェイクレビューに騙されやすい消費者が、低所得者や社会的弱者³⁰だった場合は、フェイクレビューが総合的には有害ではなかったとしても、規制が必要となるだろう。

このように、レビューの規制を議論するにあたって重要なトピックなものの、実証上の困難も大きい。例えば、消費者がフェイクレビューをどれだけ考慮しているか調査する場合、個々の消費者が参照したレビューに加え、それらを用いて実際に選択した商品のデータといった、消費者の行動についての詳細なデータが必要となることが考えられる。

一つの考えられる解決策はラボ実験であり、実際に、フェイクレビューに関するラボ実験を行うことで、最適なレビューシステムについての知見を得ようという先行研究 (Ananthakrishnan et al., 2020) も存在する。

テキスト分析

レビューと商品には豊富なテキストデータが不随する。しかしながら、現状は He et al. (2021) が限定的な分析³¹を行うのみである³²。

こうしたテキストデータの利用は、どのようなレビューが、人を欺きやすく、どのようなレビューが人を欺きにくいといった点を検証することができる。これは間接的にフェイクレビューの被害を評価することに繋がりをうる。

また、Amazon におけるフェイクレビューについては、He et al. (2021) や本稿における分析は、自社製品を高く装うようなフェイクレビューを扱うのに注力している。これは、先述したように、

³⁰一部報道では、フェイクレビューを行う消費者も、情報商材などの被害者である可能性がある。

<https://diamond.jp/articles/-/280327>

³¹低評価レビューに多い単語を機械学習モデルにより抽出している。

³²コンピューターサイエンスの分野では、フェイクレビューか否かの判別を機械学習モデルにより判別させようという試みもあり、例えば Feng et al. (2012) があげられる

Aamazon のような EC サイトにおいては商品同士の競合関係が不明瞭であることに起因する。しかしながら、Amazon においてフェイクレビューにより商品を攻撃されたという例はいくつか報告されており、こうした攻撃的なフェイクレビューの調査についても急務となっているといえる。

テキスト分析の手法³³は、商品の類似性などを商品ページのテキストから抽出することで、競合関係を把握できる可能性があり、こうした問題に対する解決策となりうる。そのため、今後のフェイクレビューの実証分析にあたって、テキスト分析が強力なツールになる可能性がある。

References

- Judith A Chevalier and Dina Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3):345–354, 2006.
- Chrysanthos Dellarocas. Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management science*, 52(10):1577–1593, 2006.
- Yuta Yasui. Controlling fake reviews. *Available at SSRN 3693468*, 2020.
- Dina Mayzlin, Yaniv Dover, and Judith Chevalier. Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8):2421–55, 2014.
- Michael Luca and Georgios Zervas. Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science*, 62(12):3412–3427, 2016.
- Sherry He, Brett Hollenbeck, and Davide Proserpio. The market for fake reviews. *Available at SSRN 3664992*, 2021.
- Phillip Nelson. Information and consumer behavior. *Journal of political economy*, 78(2):311–329, 1970.
- Paul Milgrom and John Roberts. Price and advertising signals of product quality. *Journal of political economy*, 94(4):796–821, 1986.
- Dina Mayzlin. Promotional chat on the internet. *Marketing science*, 25(2):155–163, 2006.
- Uttara M Ananthakrishnan, Beibei Li, and Michael D Smith. A tangled web: Should online review portals display fraudulent reviews? *Information Systems Research*, 31(3):950–971, 2020.
- Song Feng, Ritwik Banerjee, and Yejin Choi. Syntactic stylometry for deception detection. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 171–175, 2012.
- Matthew Gentzkow, Bryan Kelly, and Matt Taddy. Text as data. *Journal of Economic Literature*, 57(3):535–74, 2019.

³³Gentzkow et al. (2019) がこれらの手法について詳細なサーベイを行っている。