SBI Financial and Economic Research Institute

SBI-FERI Working Paper Series

SBI-FERI WP J-2

家計を投資から遠ざけるものは何か

― 時間選好・金融リテラシー・投資イメージの役割と政策的含意 ―

関口麻衣子
SBI 金融経済研究所

副島豊
SBI 金融経済研究所

2026 年 8 月 10 日

本稿は筆者らの個人的見解であり、筆者の所属組織や SBI 金融経済研究所の見解を示すものではない。

SBI-FERI Working Paper Series J-2
2026 年 8 月 10 日

家計を投資から遠ざけるものは何か[1]

― 時間選好・金融リテラシー・投資イメージの役割と政策的含意 ―

関口麻衣子[2]　副島豊[3]

要　旨

損失が生じない投資機会が目の前にあっても、なぜ人は投資しないのか。SBI 金融経済研究所が実施しているアンケート調査では、仮想的投資商品を用いてリスク回避度の把握（確実性等価の計測）を行っている。最低でも元本が保証される選択肢があるにもかかわらず回答者の 66%が「買わない」を選択した。この一見すると不合理な投資回避行動は株式市場参加パズルを解く手掛かりとなりうる。日本人 1 万人の個票データに対し、カイ二乗検定、機械学習手法（決定木、ランダム・フォレスト、勾配ブースティング）、ベイジアン・ネットワーク、ロジスティック回帰分析を適用し、投資回避の決定要因を重要度・依存構造・限界効果の各側面から検証した。分析の結果、投資回避を最も強く規定する要因は時間選好率の高さであり、次いで金融リテラシーの低さ、国内株式に対するポジティブイメージの欠如が重要な要因として特定された。一方、性別・年齢・学歴・居住地といった人口統計学的属性の直接的影響は相対的に小さく、クロス集計で観察される男女差は時間選好率や金融リテラシーの性差による交絡効果である可能性が示唆された。これらの知見は、金融リテラシー教育に加え、投資に対するポジティブな認識形成を促す施策の重要性、および時間選好率の決定要因の解明という新たな研究課題を提起するものである。

Keywords: 株式市場参加パズル、時間選好、金融リテラシー
JEL Classification: G11, G41, G53

---

[1] 本稿は筆者らの個人的見解であり、筆者らの所属組織や SBI 金融経済研究所の見解を示すものではない。本研究で利用した個票データは、SBI 金融経済研究所が 2022 年から実施している「次世代金融に関する一般消費者の関心や利用度に関するアンケート調査」の第 2 回目として 2023 年 8～9 月に実施されたものである。同個票データは研究目的のために利用申請方式で公開されている（https://sbiferi.co.jp/questionnaire/question20240425.html）。

[2] SBI 金融経済研究所

[3] SBI 金融経済研究所

１．はじめに

家計の金融資産投資に関する高いリスク回避特性は、経済学のパズルの一つであり、1980年代からエクイティプレミアム・パズルとして世界的に知られてきた。また、株式投資は長期的な資産形成に有利であるにもかかわらず、かなりの家計が株式保有をしていない（株式市場に参加していない）現象が世界的に観察されている。これは、株式市場参加パズル（Stock market participation puzzle）として知られており、その原因を検証する実証分析が 2010 年頃より多数行われている。実際、日本では金融資産保有の構成比が預金や現金など安全資産に偏っており、多くの家計が株式や投資信託などの金融資産に「直接」投資しないことが指摘されてきた。しかし、これが世界的な現象であることは一般には知られていない。

本研究では、SBI 金融経済研究所が実施しているアンケート調査を用いて、著しくリスク（不確実性）を回避する回答者がどのような属性を有しているかを検証することを試みた。同研究所は、2022 年から毎年、国内外の株式や暗号資産等に関する一般消費者の関心や投資動向に関するアンケート調査を日本ほか複数国において実施している。その集計結果は SBI 金融経済研究所（2022, 2024a, 2024b, 2025）に示されている。性別、年齢、学歴や年収、金融資産残高など個人・家計の属性に加えて、2023 年実施調査からはリスク回避度や時間選好率、金融リテラシーを計測する質問が導入されている。リスク回避度に関する質問では、以下のような確実性等価を測る質問を導入したところ、合理的とは言い難い結果が観察されており、特にその傾向が日本において著しかった。

SBI 金融経済研究所（2024a、2023 年実施調査の結果）の Q42 は、「50%の確率で 12 万円が得られ、50%の確率で 8 万円が得られる投資商品」を想定し、この投資商品がいくらなら買うか、払ってもよいと思う最大金額を選択する設問となっている。回答選択肢は、「10 万円より高くても買う」、10 万円、98,000 円から 2,000 円刻みで 80,000 円まで選択肢があるほか、「買わない」という選択肢も用意されている。その回答分布状況を表 1 に示している。日本では、「10 万円より高くても買う」（リスク選好的）を選択した回答者が 10%、期待値の 10 万円（リスク中立的）が 10%、10 万円未満（8 万円超）（リスク回避的）が 6%、8 万円（元本保証型）が 9%であったのに対し、「買わない」を選択した投資回避者が 66%と最も多い結果となっている。

投資額 8 万円を選択すれば、損失を出す可能性がないまま 50%の確率で 4 万円のリターンを得ることができる。にもかかわらず投資を回避する回答者がなぜこれほど多いのか、一種のパズルといえよう。こうした現象は日本以外でも観察されている。日本に比べると「買わない」選択者の水準こそ低いものの、米国 42%、英国 47%、ドイツ 44%、中国 11%、韓国 26%と、2023 年の調査対象 6 カ国の全てで、一見すると不合理な回答が多数観察されている。こうした現象は前出の株式市場参加パズルと整合的な結果となっており、マクロ集計

値で見た家計部門による株式保有額の少なさの背景には、株式資産運用を行っていない家計が多数あることが指摘できよう。これは、エクイティプレミアム・パズルが、株式投資を行う・行わないという二択問題と、行う場合の株式投資の金額（他の安全・危険資産との構成）という 2 つの問題に分解可能なことを示唆している。

本論文は、上述の仮想的な資産に対する投資意欲の分析から、株式市場参加パズルの理由を探る試みであり、適切な家計に適切な株式投資を促すための政策的手がかりを得ることを目指している。

表 1　仮想的資産で計測したリスク回避・選好性向

(%)

	10万円より高くても買う	10万円 [期待値]	8万円～10万円	80,000円 [最低保証値]	買わない
日本	10	10	6	9	66
米国	11	14	20	14	42
英国	9	13	17	14	47
ドイツ	14	13	19	11	44
中国	30	19	34	7	11
韓国	18	20	20	17	26

出所）SBI 金融経済研究所（2024a）。
注）期待値や保証値の文言はアンケートの質問文には含まれない。「8 万円～10 万円」は両端を含まない。また、四捨五入のため合計は 100%にならない。

本論文の構成は以下の通りである。第 2 節では、エクイティプレミアム・パズル型アプローチと株式市場参加パズル型アプローチの 2 つの視点から先行研究を整理し、日本を対象とした実証分析を概観したうえで、本稿のアプローチを示す。第 3 節では、アンケート調査の概要を紹介し、性別・年齢・学歴・年収・金融リテラシー・時間選好率など各属性別に投資回避者の割合をクロス集計により検証する。第 4 節では、カイ二乗検定、決定木やランダム・フォレスト、勾配ブースティングの機械学習手法、ベイジアン・ネットワーク、ロジスティック回帰分析の各分析手法とその結果を示し、投資回避行動の決定要因を重要度・依存構造・限界効果の各側面から定量的に評価する。第 5 節で結論と政策的含意を述べる。

2．先行研究

先行研究を以下の視点から 2 つのグループに大別してサーベイする。ひとつ目は、「株式投資の実態を前提とすると、家計は説明困難なほどに極端なリスク回避的選好を有している

ことになってしまうこと」、すなわち、「観測されるリスクプレミアムを説明するためには非現実的なリスク回避度が必要となる」というパラドックスを検証するアプローチである。効用関数ベースの資産価格モデルに現実的な仮定や制約を追加してパズルを説明しようという試みであり、これをエクイティプレミアム・パズル型アプローチと呼ぶ。もうひとつは、「株式投資を行っていない家計が多いこと」に注目したもので、株式市場参加パズル型アプローチと呼ぶ。これは、個人の意思決定の異質性を問う分析であり、なぜ長期的にみればリスク対比で高い収益が見込める株式投資に参加しないのか、その原因を探る研究である。

２．１　エクイティプレミアム・パズル型アプローチ

Mehra & Prescott (1985)が最初に指摘したエクイティプレミアム・パズルは、株式市場の実データから計測された要求リスクプレミアムを前提に効用関数の相対的リスク回避度を計測すると、説明できないほど極端なリスク回避性向が得られてしまうというパズルである。Julliard & Ghosh (2012)によれば、米国株式市場の要求リスクプレミアムは過去 100 年間で年率約 7%の高い水準にあり、標準的な消費ベースの資産価格モデル（C-CAPM）で観測されるリスクプレミアムを説明するためには、相対的リスク回避度が四半期データで 102〜146、年次データで 32〜49 という極めて高い値が必要となる。Mehra & Prescott (1985)やMankiw (1986)の議論では、合理的なリスク回避度の上限は年次データで 10 程度と考えられており、この大きなギャップがパズルとされてきた。日本においても同様な分析結果が得られている（例えば、Maki & Sonoda, 2002; Mehra, 2003）。

過度なリスク回避特性が計測されてしまう現象を説明する試みとして、様々なアプローチがなされてきた。以下のような拡張が代表的なものとして挙げられる。株式市場への限定的な参加(Mankiw & Zeldes, 1991; Vissing-Jørgensen, 2002)、経済ショックが投資家に及ぼす影響の異質性(Mankiw, 1986)、所得変動や恒常所得に関する不確実性(Heaton & Lucas, 1996)、効用が現在の消費だけでなく過去の消費水準（習慣）にも依存するという効用関数の修正(Campbell & Cochrane, 1999)、時間変動するリスク回避度(Guiso et al., 2018; Hanaoka et al., 2018)、取引コストや流動性制約の存在(Acharya & Pedersen, 2005)、政府による金融規制や税金(McGrattan & Prescott, 2003)、金融政策の影響(Bernanke & Kuttner, 2005)、新興市場における金融制度やグローバル市場分断(Bekaert & Harvey, 1995)などである。

このほか、行動経済学のプロスペクト理論のように得失に関する非対称な心理的評価や参照点依存性を想定する研究(Benartzi & Thaler, 1995; Barberis et al., 2001)や、家計の株式投資に関する認知バイアスに焦点を当てたものも存在する(Dominitz & Manski, 2007; Hurd et al., 2011)。市場崩壊や大恐慌といった稀だが極めて深刻な経済災害の可能性がリスク回避行動に大きな影響を与えるという研究もあり（レアイベント仮説：Rietz, 1988; Barro, 2006）、日本の 1990 年代初頭の株価バブル崩壊や東日本大震災を取り上げた研究がその一例として

ある(Hanaoka et al., 2018)。

２．２　株式市場参加パズル型アプローチ

もう一つのアプローチは、どのような家計が株式投資に参加しているか（していないか)、その背景にある理由は何かを問う研究群である。研究の入口として家計のミクロデータ、とくにサーベイが分析対象となっており、家計サーベイで「多くの家計が株式に直接投資していない」（年金や保険を通じた間接的な投資は含めない）という事実を整理したところから問題提起が始まっている。表２・表３は、株式市場やリスク資産市場への参加率を先進国について示したものである。サーベイ調査かマクロ集計値か、サーベイ調査の対象群の違いなどによって同じ国でも数値にばらつきがあり、厳密な国際比較は難しいが、先進国であってもリスク資産市場の参加率は 1〜4 割程度に留まっていることがわかる。日本と対比されることが多い米国においてすら、株式市場の参加率は低い水準にとどまっており、株式市場参加パズルはグローバルな現象として観察されている。

表２　ユーロ圏の株式市場参加率（下段は投資信託市場参加率）

（%）

ベルギー	ルクセンブルク	オランダ	ドイツ	オーストリア	フランス	イタリア	スペイン	ポルトガル	ギリシャ	フィンランド	スロベニア	スロバキア	キプロス	マルタ	ユーロ圏
10	8	5	11	5	12	3	12	4	1	21	6	2	16	10	9
21	12	13	16	9	9	6	7	2	0	31	6	4	1	9	10

出所）ECB の Household Finance and Consumption Survey (2023)中の Table C1。調査時点は 2017 年。保険や年金等を通じた株式投資は含まれない。

表３　その他のサーベイ調査：株式のみを対象

（%）

					Badarinza et al. (2016)							
米	日本	カナダ	英	インド	豪	カナダ	ドイツ	フランス	イタリア	オランダ	英	米
21	19, 15	36	21	5	35	10	11	15	5	10	18	15

出所）左から、米国は FRB の Survey of Consumer Finances (2022)、日本は前者の 19%が大阪大学のサーベイ（馮ほか, 2017)、後者の 15%が金融広報中央委員会のサーベイ（塩路ほか, 2013)、カナダは SIMA サーベイ(2025)、英は Financial Conduct Authority の Financial Lives (2022)、インドは Securities and Exchange Board of India のサーベイ(2025)である。その右列は国際比較のため調整を行った Badarinza et al. (2016)による（参加率比較 Table1 中の株式直接保有を参照)。同一国であっても出典により数値が異なりうるのは、i)対象が株式のみか他のリスク資産を含むか、ii)集計対象はマクロかサーベイのサンプル調査か、iii)調査年の相違に起因する。サーベイ調査では対象サンプルの違いでも水準が異なりうる。なお、日本では、リスク性金融資産の保有割合が、慶應義塾大学の JHPS/KHPS を用いた Alzuabi & Gray (2025)で 23%、有価証券保有世帯の割合が総務省統計局・全国消費実態調査 2014 年で 24%とされている。また、祝迫（2018）や小寺（2020）も日本家計の金融資産投資を分析している。

同パズルに注目した代表的な初期の研究が Haliassos & Bertaut (1995)である。アンケート調査を分析し、所得、資産、教育水準、年齢などの属性をコントロールしても、依然として株式を保有しない家計が大量に残ることを示した。例えば、所得や資産水準が高く、教育水準も高く、年齢も高いのに株式投資を行っていない家計が多く存在するという現象である。投資資金の不足だけでは説明できない本パズルは、世界的に共通して観察される現象として注目を集めるようになった。こうした初期の研究では、株式保有ダミー変数（質的選択変数）を被説明変数とし、所得・資産・教育・年齢など家計や個人の特性で Probit/Logit モデルを適用する分析が主流であった。

2000 年代になると、リスク回避の強度、信念や期待・情報・リテラシーの家計ごとの異質性(Guiso et al., 2003; Lusardi & Mitchell, 2011; van Rooij et al., 2011)、習慣の持続性や惰性・調整の鈍さ(Calvet et al., 2009; Bilias et al., 2010)、将来所得や出費にかかる不確実性(Heaton & Lucas, 2000; Viceira, 2001)、最低投資額や証券口座維持手数料などの市場参加制約(Luttmer, 1999)、参加にかかる情報収集コスト・メンタルコスト(Mulligan & Sala-i-Martin, 2000; Vissing-Jørgensen, 2002)などによりパズルが解明できないか、多くの仮説検証が行われた。例えば、労働所得の将来見通しや医療費ほか生活コストの不確実性、あるいは住宅ローンや教育費の出費などが、株式投資という追加リスクや追加支出（手元流動性の減少）を奪っているのではないかという仮説が指摘・検証されている(Polkovnichenko, 2004)。また、多くの家計は株式のリスクを過大評価し、期待リターンを過小評価しているのではないかという主観的期待（の実態データからの乖離）に注目したものもある(Dominitz & Manski, 2007; Kézdi & Willis, 2011)。類似の研究として、大恐慌などの大きなマクロ経済ショック時の経験が信念形成やリスクテイク・市場参加に長期にわたって影響を残すこと、直近の経験が強く影響すること（近時性効果：recency effects)、経験にもとづかない情報（統計データや第三者からの知識）には影響されにくいことなども検証されている(Malmendier & Nagel, 2011)。

このほか、データが信頼でき、株式市場全体の仕組みが公正である（マーケット・インテグリティが保たれている）という市場や金融機関、様々な他の市場参加者への信頼も重要であり、そうした信頼の欠如が不参加に影響しているのではないか（騙されて損をするのではないか）という仮説も検証されている(Guiso et al., 2008)。類似の視点から過去の歴史に根差した金融教育の背景や宗教的・文化的な環境を取り上げた研究もある(Guiso et al., 2003; Guiso et al., 2004)。

株式市場をどう理解（誤解）しているかという視点からの最新の研究もある(Duraj et al., 2024)。株式投資に参加しない人の心理的障壁の一つとして、「株式市場への参入や継続には膨大な知識と労力が必要である」という誤解が挙げられる。アクティブ運用を想定している潜在的な参加者は、株式発行企業や市況のモニタリングを重要視し、多数の株式をフォロー

する大変さから少数の銘柄に集中し、あるいは投資を手控えてしまうことすらありうる。これは、努力しないと投資リターンの成果が得られないという「金融市場の効率性（Market Efficiency）に対する理解不足」が関係している。インサイダー情報などを保有していない限り、一般的な個人投資家がいくら情報収集を行っても、市場が十分効率的であれば追加的に得られる超過リターンは平均的に見れば存在しない。価格発見機能を担う主役であるプロ投資家にとってもアルファ（超過リターン）を探すのは容易ではなく、その源泉となるアノマリーや秘匿情報が発見されてもすぐに市場で消化され、株価に反映されてしまう。こうした効率市場仮説に立つと、上述の誤解や懸念は杞憂に過ぎないのであるが、先行研究では一定の影響力を持つことが検証されている。

アンケート調査では、金融知識に関する自信過剰や知識不足への過度な不安が株式投資への参加率やパフォーマンスに影響することも検証されている。行動経済学の文脈で種々のランダム化比較試験（RCT）が実施されており、女性の株式市場参加率が低い理由の一つに、投資リテラシーの不足以上に投資リテラシーが低いと自認（誤認）していることも指摘されている(Bucher-Koenen et al., 2025)。逆に、男性は自信過剰（採点結果のリテラシースコアと自己認識の同スコアのギャップが大）が市場参加率の高さを後押ししているという研究もある。Bucher-Koenen et al. (2017)は、女性の金融リテラシーに関する包括的なレビューを提供している。複数国のデータを用いて、女性が男性に比べて金融リテラシーが体系的に低いこと、「分からない」と回答する傾向が強いことを示した。この差は年齢、教育水準、所得を考慮しても残り、若年層から高齢層まで一貫して観察されている。女性のほうが金融知識への自信も低く、知識不足と自信不足の両方が老後の資産形成や退職準備に悪影響を与えると指摘している。

隣人と交流したり教会に通ったりするという社会的なつながりが高い人は、株式市場に参加する可能性が高いこと(Hong et al., 2004; Brown et al., 2008)、家計内や親族友人に株式投資経験者がいる場合、参加可能性が高いことも検証されている(Li, 2014)。Li (2014)は、米国の家計動態パネル調査（PSID）を用い、観察可能な属性をコントロールしてもなお、親や子どもが過去 5 年以内に株式市場に参入した家計は、自身も 5 年以内に参入する確率が約 30%高いことを示し、拡張家族内での情報共有が株式投資の決定に重要な役割を果たすことを実証している。これらはピア効果と呼ばれており、株式市場参加パズルを解く一視点として注目されている。参加行為の情報伝播に注目するもの、参加後のリターンの情報伝播に注目するもの(Kaustia & Knüpfer, 2008)、伝播のネットワークを研究対象にするもの(Parakhoniak et al., 2026)などがある。ネットワーク分析からは、低所得層においては投資コストを軽減する実用的助言や模倣行為のつながりが参加の鍵となる一方、高所得層においてはリターンに関する情報を得るつながりが重要となることが指摘されている。なお、社会的な相互作用が強いと参加率が高まる以外に、ネガティブな情報の伝播により参加率が逆に低くなるという複数均衡の可能性も指摘されている。

対象データに関しては、家計サーベイデータの分析だけでなく、i)参加コスト等を組み込んだライフサイクル最適化モデルの構造推定(Gomes & Michaelides, 2005)や参加コストの推計(Alan, 2006)、ii)行政データや銀行等の口座データという個票ビッグデータ・レジストリデータを用いた精緻な計測(Calvet et al., 2007)、iii)限定参加を組み込んだうえでの資産価格モデル（C-CAPM）への回帰など幅広く展開している。

このように、株式市場参加パズルを起点にした分析は、事実の計測、パズルの発生原因の探求、参入コストなど直接計測不能な数値の推計、資産価格モデルを基盤とするエクイティプレミアム・パズルとの橋渡し、行動経済学との接近という発展を辿った。こうした研究蓄積を通じて、より広範な家計の金融行動研究である Household finance の研究分野が注目を集めている(Gomes et al., 2021; Campbell & Ramadorai, 2026)。これは、家計が行う複雑かつ異質で、相互依存的な意思決定を研究する分野であり、デジタル化社会におけるビッグデータ、AI、パーソナライズド・マーケティング、各種 Fintech サービスなど、近年の情報技術革新や新サービス登場によって分析可能領域が急拡大したことを反映している。

同分野では、従来の分析では対象範囲に含まれなかったようなリアルな人間行動や思考を対象に実証分析が試みられている。例えば、限定的な資産市場参加、負債管理の非合理性（低利の預金残高を抱えながら高利のリボ払いを続ける）、金融制度設計におけるデフォルト設定のナッジ効果（年金や投信運用は商品ラインアップやデフォルト設定で選択が左右される）、借り換えの不作為（住宅ローン金利が低下しても、注意不足や手続きの複雑さから有利な借り換えを行わない不作為、あるいは類似ケースとして有利な税制度や社会保障制度を活用しない不作為）、ネットトレーディングなどでの過剰取引、自信過剰、過去の成功からの誤学習、不都合を直視したくないがゆえの現状認識回避・判断の先送り、SNS の影響などである。

Household finance は、単なる金融投資や借入行為を超え、人間行動、社会構造、そして技術革新が交差する現代金融における重要分野の一つになっている。また、家計の意思決定論にとどまらず、金融機関が消費者のバイアスにどう反応し、それが市場全体の均衡（価格設定や商品設計）や富の移転・格差にどう影響するかも研究対象となっている。マーケティングやプラットフォームビジネスなど産業組織論、金融制度論、行動経済学、認知科学、市場アノマリーとの親和性も高く、様々な経済ファイナンス理論のクロスフィールドにもなっている。複雑な最適化問題を解く計算能力の向上、大規模な行政データの利用可能性、因果推論のためのミクロ計量経済学の発展、そして行動経済学の新パラダイムなどがこうした研究を支えていると指摘されている(Campbell & Ramadorai, 2026)。

２．３　日本を対象とした先行研究

日本の先行研究はサーベイ調査を用いたものが多い。最初に、分析に活用されている代表的なサーベイ調査を紹介する。まず、J-FLEC（金融経済教育推進機構、2024 年までは金融広報中央委員会）による「家計の金融行動に関する世論調査」と「金融リテラシー調査」が挙げられる。このほか大阪大学による「くらしの好みと満足度についてのアンケート」や、慶應義塾大学（パネルデータ設計・解析センター）による JHPS/KHPS（日本家計パネル/慶應家計パネル）が実証分析に利用されている。これらは同一個人や家計を調査対象とするパネルデータとして作成されている[4]。また、2006 年から毎年実施されている日本証券業協会の「個人投資家の証券投資に関する意識調査」を用いた研究もある。本稿が利用した SBI 金融経済研究所の「次世代金融アンケート」は 2022 年から毎年実施され、これまで 4 回分のアンケートが蓄積されている。他の調査と異なり、暗号資産やステーブルコインなど新しい金融資産も調査対象に含まれており、株式や債券、FX への投資と組み合わせた分析が可能となっている点が特徴である。

以下では、日本を対象とした実証分析の結果を紹介する。金融広報中央委員会のサーベイを利用した詳細な分析として、塩路ほか（2013）が挙げられる（先行研究である Fujiki et al., 2012 も参照）。両研究は、「家計の金融行動に関する世論調査」の個票データを用い、家計が株式市場に参加するか否かと、参加者が金融資産のうちどの程度を株式に配分するかの決定要因をそれぞれ推定している。金融資産残高や学歴が高いほど株式市場への参加確率は高まる一方で、金融資産に占める株式の比率に対しては、金融資産残高は負の相関を持つことが示されている。また、年齢効果に関しては、高齢層ほど株式保有額が多いという観察事実があるものの、高齢層が多額の金融資産を保有していることによる部分が大きく、年齢そのものが参加確率を高める効果は限定的であった。

藤木（2019）では「金融リテラシー調査」が用いられ、株式、投資信託、外貨預金等の危険資産を購入した経験がある家計は調査サンプルの半分以下であり、購入経験者の中にも商品性を理解せずに購入している層が約 1 割存在していることが示されている。金融リテラシー指数が高いほど、金融商品の商品性を正しく理解し、複雑な仕組みの商品に対しても適切な判断を行う傾向があることが確認されている。しかし、3 種すべての商品を購入した経験がある層であっても、複雑な仕組みの商品に対して「理解していなくても金融機関が信用できれば買う」という判断も見られている。情報収集において、金融リテラシーが高い層はマスメディアやウェブサイトを多用する傾向がある。一方で、リテラシーが低い層や高齢層は「金融機関店舗のみ」あるいは「どこで情報を得ればよいか分からない」状態にある確率が高い。また、「店舗・セミナー・アドバイザー」を利用する層は、高齢で知識の自己評価

[4] 金融広報中央委員会と大阪大学のサーベイ調査の比較を伊藤ほか（2019）が行っている。

が高い一方、金融トラブルの経験を持つ傾向があることを明らかにしている。金融リテラシーの低さが情報取得や市場参加の障壁となっている可能性を指摘し、顧客の属性と情報源の組み合わせに応じた金融機関のマーケティング上の示唆を提示している。

大阪大学の「くらしの好みと満足度についてのアンケート」はパネルデータであり、同一個人の年齢や金融経済・社会環境の変化への対応などが観察できる。また、日本だけでなく米国も調査対象としている。同調査を用いた伊藤ほか（2019）は、リスク資産を保有するか否かには、予備的貯蓄動機（特に老後不安）や流動性制約、金融知識が強く影響する一方、リスク資産に投資するようになると、そのシェアは株式投資の期待収益率やボラティリティなど市場要因に応じて時間の経過とともに各々のリスク回避度に応じた適切なリスク資産割合に近づいていくことを発見している（Calvet et al., 2009 による部分調整モデルを適用）。大阪大学のアンケートにはクジや保険、報酬に関する仮想的な設問があり、これを用いて相対的リスク回避度を推定することが可能である。伊藤ほか（2019）の分析では、日本人が米国人よりリスク回避的であるという証拠は得られておらず、リスク資産への投資実体の相違（特に参加率の低さ）においては、市場参入時における金融知識（参入コスト）や資産形成を巡る制度的要因（モデル上は定数項の相違として計測）が重要な役割を果たしていると指摘されている。

同じアンケート調査を用いた Kadoya et al.（2017）は、年齢、資産、教育水準に加え、リスク回避度や他者への信頼といった心理的要因を制御した後でも、金融リテラシーの高さが株式市場への参加およびリスク性資産の保有割合を有意に高めることを示している。伊藤ほか（2019）と同様、リスク回避選好や保有総資産だけでなく、適切な投資判断を行うための知識不足に起因することを指摘している。関田（2020）も同じアンケート調査を用いて、金融リテラシーは株式市場参加率を高める効果を持っており、作成したリテラシー指数が 1 標準偏差上昇すると、参加率が 18%も高まるという推計結果を得ている。高い金融リテラシーは老後の貯蓄計画の策定も促進し、株式保有と計画策定の両面を通じて純資産の蓄積に寄与するというメカニズムも確認している。また、自身の金融知識に関して自信過少な人々はリスク性資産の保有割合が少ない（自信過剰な人々は保有割合が多い）という結果も得ている。

慶應義塾大学の JHPS/KHPS も大阪大学調査と同様にパネルデータとして作成されているため、市場参入だけでなく、投資の積み増しや取り崩し、退出という時系列変化をライフタイムイベントとあわせて分析可能となっている[5]。小寺（2020）は、同サーベイ調査に機械学習の手法を適用し、リスク性資産への市場参加に至る 2 つの因果経路を識別している。

[5] 例えば、Hanaoka et al.（2018）は JHPS を利用し、東日本大震災の体験が個人のリスク選好に変化をもたらしたことを発見している。具体的には、i)激しい揺れを経験した男性は震災後にリスク許容的になった一方、女性にはそうした変化が見られないこと、ii)仮想的なクジに関するアンケート調査結果から効用関数のパラメータ変化として観察された男性側の変化は、ギャンブルの頻度が増加するなど行動面でも観測されていることが紹介されている。

具体的には、ランダム・フォレストによる重要変数の選択とベイジアン・ネットワーク分析による変数間の因果経路の識別である。一つ目の経路は、年齢や世帯所得が負債や預貯金に影響し、預貯金の多寡が老後を見据えた余裕（流動性制約のなさ）を通じてリスク資産保有に向かうというものである。二つ目は、学歴や情報収集力、および金融知識がリスク選好を通じてリスク資産保有に向かうものである。経済成長や年金改革などを通じた将来不安の解消という政策対応が効果的であるが、それ以外にも金融リテラシーの向上を通じたリスク選好の高まりによってリスク資産投資を促せるため、家計が金融知識の向上を通じて適切なリスク許容度を把握し、そのうえで市場参加を進めることが重要と指摘している。なお、前述の伊藤ほか（2019）は、リスク選好が低いわけではなく金融リテラシーの不十分さに課題があるという解釈を行っており、政策課題の認識は同じでも市場参加誘因の構造理解には相違がある[6]。

Yamori & Ueyama (2022)は、投資経験が金融知識を高めるという逆方向の因果が存在するという内生性の問題に対処するため、操作変数法を用いた検証を行い、これを考慮したとしても金融知識の向上が株式市場参加率の向上に繋がる（その一方で、参入後の株式シェアの多寡には影響しない）という分析結果を得ている。また、金融リテラシーと内生性を適切に制御すると性別による参加率の差は消失する、すなわち女性の参加率の低さは金融リテラシーの不足に一因があることを示している。

Iwaisako et al. (2022)は、持ち家が金融資産選択に及ぼす因果効果を分析している。住宅価格の地域差を用いた操作変数法により、住宅資産と住宅ローン負債の影響を分析し、資産価値の上昇は資産効果を通じて株式保有割合を高める一方で、初期負債の大きさは抑制方向に働くことを示している。米国では住宅資産の価格変動リスクが株式保有を阻害する効果が検証されているが、日本ではそうした現象は確認されず、持家取得が必ずしも株式投資への参加を妨げる要因ではないとしている。

行動経済学の視点からは、末廣（2022）がメンタル・バジェッティング（心理的な予算割当）が家計の金融資産形成に与えるバイアスを分析している。メンタル・バジェッティングは、家計が予算を特定のカテゴリー（食費、被服費、娯楽費等）ごとに割り当てて管理する心理的な傾向のことを指す。Antonides et al. (2011)は、この傾向がある家計は、消費や預金だけでなく金融資産への投資にも予算管理を行うため、リスク資産形成にも有効である点を指摘している。一方、日本でアンケート調査を行った末廣（2022）は、メンタル・バジェッティングは予算の硬直化を招き、結果として資産保有に対して負の影響があることを指摘している。例えば、低金利の預金を多額に持ちながらリボ払いなど高金利の借金を抱えているといった思考の硬直化、非合理的選択が挙げられる。こうした問題を回避するためには、

[6] 標準的な理論モデルではリスク回避度など効用関数の特性は所与で不変と想定する場合が多く、伊藤ほか（2019)の発想はこれに従っている。一方で、上述の Hanaoka et al. (2018)のようにリスク回避度の変化を計測した研究もあり、小寺（2020）のように、因果的な関係を仮定した分析も行われている。

効果的な金融教育には単なる管理手法の推奨だけでなく、具体的な資産蓄積の目的設定を組み合わせることが不可欠であると結論付けている。

このほか、日本証券業協会の「個人投資家の証券投資に関する意識調査」を用いた顔ほか（2019）は、金融リテラシーの種類や情報源の違いを取り上げ、数学的な基礎知識よりも、投資信託の仕組みといった応用的リテラシーの高さがリスク資産の保有確率をより高めることを明らかにしている。また、専門家の助言を情報源とする投資家は、投資信託を保有する傾向が強く、運用リターンがプラスになる確率も高まる一方で、知人などの非専門家からの情報は投資行動を抑制する方向に働くことを指摘している。竹村ほか（2024）は、独自のアンケート調査データを用いて金融リテラシーとは別に 6 つのリスクリテラシー要素を計測し、ゼロリスク指向はリスク資産保有確率に影響せず、リスク・リターンのトレードオフや異なるリスク間のトレードオフに関するリテラシーが正の影響を有している点を指摘している[7]。

これら先行研究では、金融リテラシーほか広義の市場参加コスト（最低投資額、手数料、税コスト・税処理等だけでなく情報収集コストやメンタルコストなど）は、市場参加を阻む要因になっているとの結果が得られている。近年、NISA/新 NISA（少額投資非課税制度）導入に伴う株式市場参加促進効果が観察されており、税コスト削減や NISA が身の回りで話題となるピア効果を通じたナッジ（気づき・促し）効果があったと考えられる(Katauke et al., 2025)。また、ネット証券台頭による手数料引き下げや、ネットサービス普及による情報収集コストの低下、投資開始時のメンタルコストの低下（店頭に向かう際の心の負担とスマートフォンアプリの気楽さの相違)、ロボアドバイザーの普及なども寄与していることが考えられる。海外の証券市場分析では上記を支持する実証分析がなされている(Adams et al., 2024; Liu et al., 2024; Guo et al., 2025; Reher & Sokolinski, 2024)。Guo et al. (2025)は、デジタルプラットフォームを通じた金融知識の普及が中国の個人投資家の市場参入や運用成果に与える影響を分析し、デジタル技術による情報提供は株式市場への参加率を高め、ポートフォリオの分散投資を促進し、リスク調整後リターンを向上させることを明らかにしている。特に、従来は投資に消極的だった高齢層、低学歴層、および低資産層においてその効果が顕著に現れることを発見している。

[7] 金澤ほか（2020）は、リスクリテラシーを構成する 6 要因を計測している。具体的には、リスク対便益のトレードオフ、異なるリスク間のトレードオフ、リスクの基礎知識、ゼロリスク志向、特定リスクを過大・過少する認知バイアス、リスク認知のパラドックスであり、このうち前 3 者のリテラシーがリスク資産保有に正の効果を持つという分析結果を得ている。

２．４　本稿のアプローチ

本研究は株式市場参加パズルを直接分析するのではなく、仮想的な投資商品への投資選好や不参加の選択から原因を探るというアプローチを試みている[8]。具体的には、SBI 金融経済研究所が 2023 年に実施したアンケート調査（日本人 10,000 人）の個票データを用い、仮想的な投資商品に対する投資回避行動の決定要因を、複数の統計・機械学習の手法を組み合わせて分析するものである。①カイ二乗検定による各変数と投資回避・選択との関連性の探索的スクリーニング、②決定木、ランダム・フォレスト、勾配ブースティングの機械学習手法による重要変数の特定、③ベイジアン・ネットワークによる変数間の依存構造の可視化、④ロジスティック回帰分析による各変数の影響の方向性・強度の定量評価と介入効果の試算、という手法を用いている。

本稿のアプローチは、小寺（2020）と共通点を持ちつつも、いくつかの重要な相違がある。主な相違点は以下の通りである。第一に、被説明変数が異なる。小寺（2020）が実際のリスク性資産保有の有無を対象としたのに対し、本稿では仮想的な投資商品への投資意欲（確実性等価の質問における「買う・買わない」の選択）を分析対象としている。これにより、実際の市場参加の制約（口座開設コスト、最低投資額等）から切り離された、より純粋な投資回避性向を捉えることが可能となっている。第二に、説明変数の構成が異なる。小寺（2020）では預貯金・世帯所得・負債・住宅関連など経済的変数が中心であったが、本稿では時間選好率、金融リテラシー、国内株式のポジティブ・ネガティブ認識、暗号資産を含む投資経験など、行動経済学的・心理的要因をより豊富に取り込んでいる。第三に、分析手法の幅が広い。本稿ではカイ二乗検定、決定木、ランダム・フォレスト、勾配ブースティング、ベイジアン・ネットワーク、ロジスティック回帰分析の 6 手法を適用し、各手法の結果を横断的に比較検証している。特にロジスティック回帰分析による限界効果の推計を加えることで、各要因の影響の大きさを定量的に評価し、政策的介入の優先順位を示すことを試みている。

[8] 同じサーベイ調査を用いた山沖・副島（2025）は、暗号資産や株式債券などのリスク資産市場参加を質的選択モデルで直接推計しているほか、市場参加と参加後の保有額対数値を Generalized Tobit Model で FIML 同時推計することを試みており、暗号資産投資家の属性分析などを行っている。

３．投資回避者の属性

SBI 金融経済研究所によるアンケート調査（2023 年、第 2 回）は、日本・アメリカ・イギリス・ドイツ・中国・韓国を対象に、従来の金融商品である「株式、社債、FX」に加えて、新しいデジタル金融商品である「暗号資産、ステーブルコイン、セキュリティ・トークン、NFT」に関して調査を行っている[9]。調査項目については、回答者の属性として、性別・年齢・年収等のほか、金融商品についての認知度・投資経験・認識等、金融リテラシーやリスク回避度、時間選好率等となっている（金融リテラシー、リスク回避度、時間選好率の調査は第 2 回調査から追加されている)。対象者数は、日本人が 10,000 人、日本以外の国を含めて計 20,000 人であり、デジタル金融商品を対象としたものとしては日本では最大規模の調査となっている。また、デジタル金融商品と従前の金融商品を同時に対象としている点でも分析上の有用性が高いものとなっている。

同アンケート調査では、サンプル数の偏りを回避するため、日本人男性 5,000 人、同女性 5,000 人をそれぞれの年齢層（20～29 歳、30～39 歳、40～49 歳、50～59 歳、60 歳以上）で各 1,000 人となるように調査を行っている。

最初に、どのような属性・特徴を有する回答者ほど投資回避的であるかを、性別や年収水準別に投資回避割合を比較することで検証する。交絡効果が存在しうるため、この分析だけでは影響力の高低は判断できないが、シンプルな第一次接近法として有益である[10]。まず、投資回避者には明確なジェンダーギャップが存在しており、どの属性とのクロス分析を行っても女性の方が投資を回避しがちである。全体では、男性の 54%、女性の 72%が投資回避者であった（図１)。カテゴリー分類の際に男女差が大きくなると、性差の影響が反映されかねない。このため、以下では性別という交絡を避けるため男女別にカテゴリー比較を行っていく。

年齢別にみると、男女とも、世代が上がるにつれて投資回避者の割合が高まり、女性の 50 代以上では 8 割近くが投資回避者であった（図１)。男性は、50～60 代で 6 割程度、70 代で 7 割超が投資回避者である。最も低い 20 代の男性層でみても 4 割強が投資回避者である。居住地についてみると、大都市（政令指定都市等）ほど低く、町村部で高くなっている（図２)。学歴は高学歴ほど投資回避者が減少し、男性の大学院卒で 3 割台、女性の大学院卒で 5 割となっている（図３)。

---

[9] NFT は資金決済法・金融商品取引法等の規制が係る金融商品ではないが、便宜的に金融商品という総称に含めている。

[10] 結果 Y に対して要因 1 が強い影響力を有しているようにみえても、真の因果性をもつ要因 2 と要因 1 の間に正相関がある（例えば要因 2 が要因 1 を決定している）にすぎず、要因 1 は結果 Y に対して無関係であるケースもありうる。こうした見かけ上の関連を交絡効果という。要因 2 に気が付かない、もしくは計測が困難である場合、交絡効果の存在は因果性や関係性の誤認に繋がりやすい。

図１　投資回避者属性：年齢別 Q2

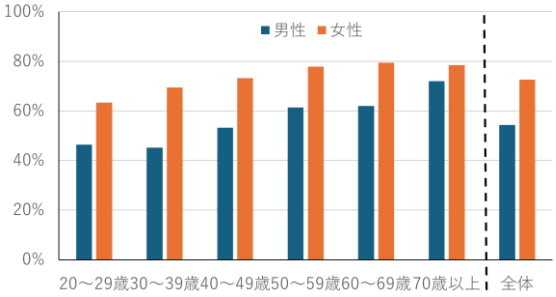
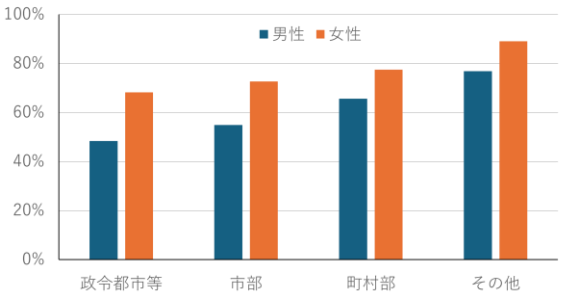
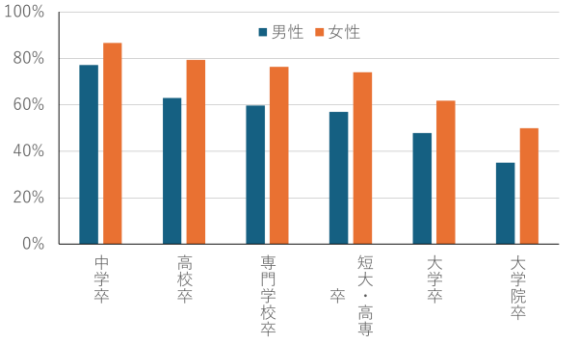


図２　居住地別 Q3



注）Q2 はアンケート調査の質問項目番号。以下同じ。各カテゴリーの構成比は、SBI 金融経済研究所（2024a）を参照。

図３　学歴別 Q4[11]



世帯年収（以下、年収）、世帯保有金融資産残高（以下、金融資産残高）についてみると、いずれも高くなるにつれて投資回避者が減少している（図４・図５）[12]。なお、これらのアンケート項目では「分からない」、「答えたくない」を選択した回答者が突出して多く、これは日本固有の現象である。詳細は SBI 金融経済研究所（2024b, 2025）を参照。アンケート選択肢の細かい刻み幅をそのまま定量分析に用いると、クロス集計のサンプル数が少なくなるため、年収・金融資産残高を、それぞれ４カテゴリーに集約した[13]（なお、図４・図５では、「分からない」と「答えたくない」の性質の違いを確認するため、分けて示している）。集約の基準は、投資回避者のサンプル数が概ね偏らないよう設定している。これは第 4 節

[11] 在学中・中退の場合は最終卒業学歴を示す。

[12] 以降の分析では用いていないが、年収 1,000 万円以上について投資回避割合をみると、男性 34%、女性 52%であり、さらに投資回避割合が低下している。

[13] 年収は無収入～300 万円未満、300 万円以上 600 万円未満、600 万円以上、「分からない・答えたくない」に、金融資産残高は 100 万円未満～300 万円未満、300 万円以上 1,000 万円未満、1,000 万円以上、「分からない・答えたくない」に区分している。

の定量分析（外部統計の平均値を参照した 2 カテゴリー集約）とは異なるが、本節は探索的観察を目的とするため、サンプル数を考慮した 4 分割で傾向を確認している。

図 4　世帯年収別 Q7

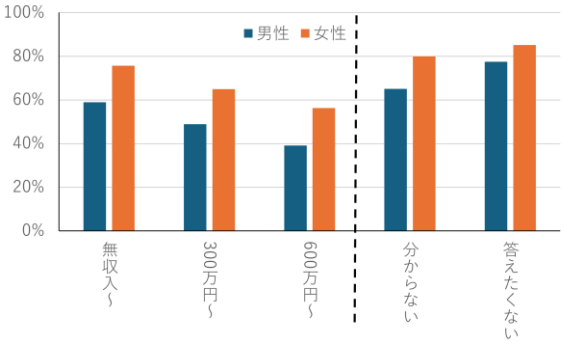
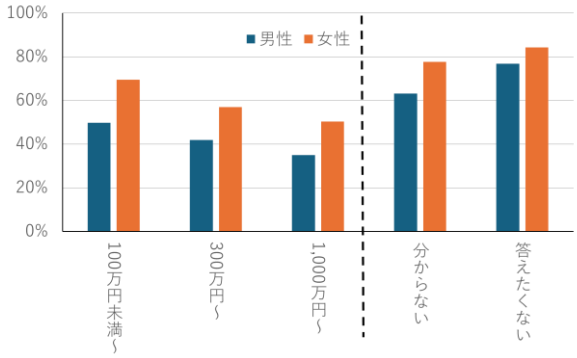


図 5　世帯保有金融資産残高別 Q8



次に、国内株式の認知度についてみると、国内株式について「（あまり）知らない/聞いたことがない」とした低認知度グループの方が投資回避的である。男性で 4 割対 6 割強と 2 割以上のギャップがあり、女性についても 2 割以上のギャップが存在している（図 6）。株式に関する知識の有無と投資行動の間には強い関連性が観察される。この関連性が因果性を示すものなのかについては第 4 節で検証する。

アンケート調査では、各金融資産別に「詳しく知っている、ある程度の認識はある」と回答した者のみを対象に、当該資産の投資経験を聞いている。こうした認知層に対する投資経験の質問についてみると、株式の投資経験がない方がより投資回避的であったが、投資経験がある者でも投資回避を選択しているものが男性で 4 割、女性で 5 割も存在している（図 7）[14]。投資経験の有無という体験型の金融リテラシーに注目しても、それだけでは投資回避行動は十分には説明できないことが示唆されている。また、認知度と投資経験の図 6・図 7 を比較すると、両者に大きな差は確認されないが、これは両図の対象集団が異なるため（脚注 14 参照）、単純比較は妥当でない。両者の独立的な影響については第 4 節で検証する。

国内株式に対する認識については複数回答可であるため、回答選択肢をその内容に応じて「ポジティブなイメージであるもの（例えば、利益が期待できる）」、「ネガティブなイメージであるもの（例えば、損失が生じることへの不安がある）」、「無関心」に分類し、投資回避者の割合をみた（図 8）[15]。ポジティブイメージ群が低いことは予想の通りであった。し

[14] 図 6 の認知有りのカテゴリーと図 7 は、計測対象者は同じであるが、分母が異なる（図 6 は調査対象者全員）ため、図 7 の加重平均が図 6 になるわけではない。

[15] 複数回答可のため、同じ回答者が複数のカテゴリーに含まれる場合がある。また、「（何らかの）関心あり」は「無関心」と対比するために集計しており、国内株式に対してポジティブイメージがある者とネガティブイメージがある者の双方が含まれる。

かし、ネガティブイメージ群と無関心群を比較すると、無関心のほうが回避割合が高く、特に男性においては、ポジティブ・ネガティブイメージ群間のギャップ以上に大きな格差となっている。仮説としては、リスク回避度が高い（仮想の投資質問における投資回避）と、金融投資について関心を持たなくなるという因果が考えられる。この点については、第 4.6 節のベイジアン・ネットワークにより、変数間の関係の方向性について示唆を得る。

図 6　国内株式の認知度 Q10.1

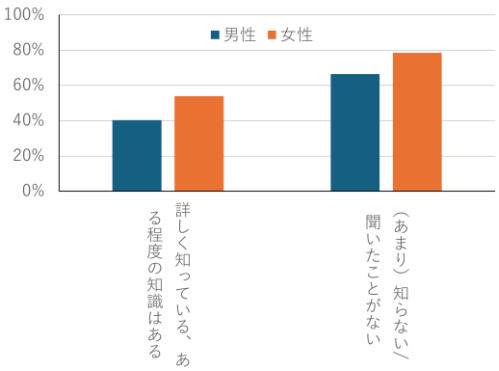


図 7　国内株式の投資経験 Q11.1

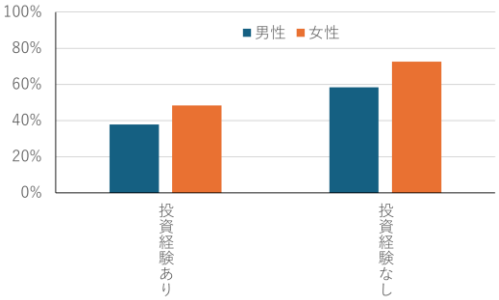
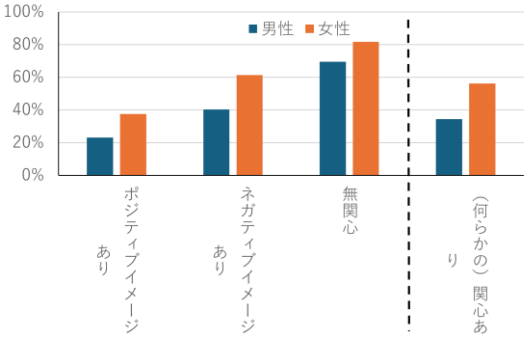


図 8　国内株式に対する認識 Q12



暗号資産の認知度では、株式と同様に認知度が低い方が投資回避的である（図 9）。また、男性の場合、株式認知度とほとんど相違がない結果となっており、認知に関するアセットクラスの相違は投資回避特性には影響を持っていないようである。女性の場合、両カテゴリー間でギャップが小さくなっているが、暗号資産の認知度について「詳しく知っている/ある程度知っている」と回答した者は全体の 12%に止まっており（株式は 37%）、特に女性においてサンプル数が少なく、このため解釈には留意が必要である。

認知層に対する暗号資産の投資経験では、株式と同様に投資経験がない方が投資回避的であるが、投資経験者の回避割合を株式と比較してみると男性で 2 割強、女性で 4 割まで低

下している（図 10）。暗号資産の投資経験者は全体の 4%であり、これらの層で投資回避割合が突出して低くなっていることは、暗号資産投資家のリスク選好の高さを反映したものと推測される。

図 9　暗号資産の認知度 Q18.1

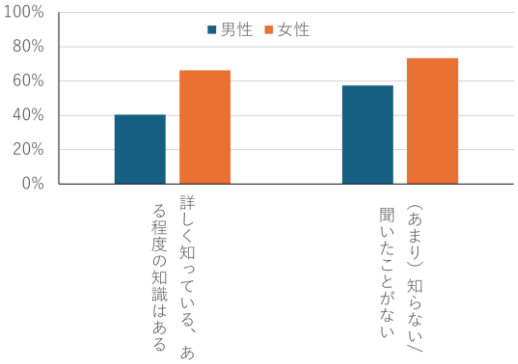
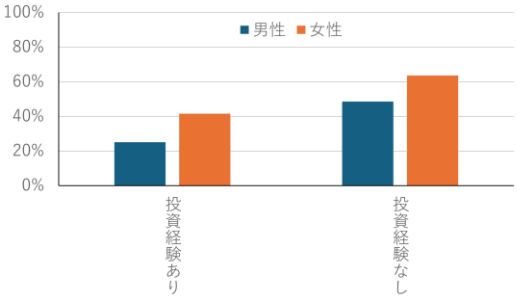


図 10　暗号資産の投資経験 Q19.1



金融リテラシー4 問（Q38：単利、Q39：複利、Q40：実質金利、Q41：ポートフォリオ分散効果）について、3 問以上正答と 3 問未満正答に分け、金融リテラシーの高低が及ぼす影響を観察した（図 11）[16]。高リテラシーのほうが投資回避の割合が低く、男性 3 割弱、女性 4 割と暗号資産投資経験に次いで低い水準となっている。また、男女ともに 40%弱の大きな選択格差が生じており、これは暗号資産投資経験の有無による格差より大きい。個別の質問ごとにみると、いずれも誤答のほうが強く投資回避的であるほか、より基本的な知識といえる「単利」の誤答群が最も高い投資回避割合（男性 8 割、女性 9 割弱）を示している（図 12）。正答群を比較しても、単利の正答群は男性で 4 割であるのに対し、実質金利や投資の分散効果で 3 割程度と低くなっている。女性についても同様である。これは「単利が分かる層」の中に「分かるが投資しない人」が混じっているのに対し、「実質金利・分散効果が分かる層」では投資意欲がより高いことを示唆している。これらより、金融リテラシーの高低は投資回避の特性と強い関係性があると推測される。前述の交絡効果の可能性もあるため、第 4.6 節のベイジアン・ネットワーク分析で検討する。

[16] 4 問正答（完答）とそれ以外に分類した場合、高リテラシーのサンプル数が非常に少なくなるため、3 問以上正答を高リテラシーとみなした。

図 11　金融リテラシー全体評価

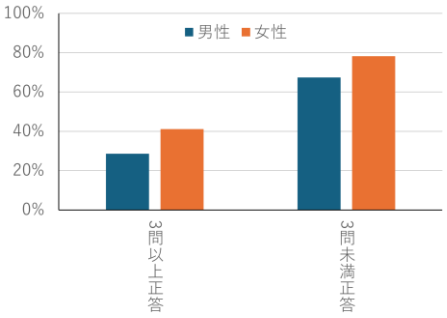
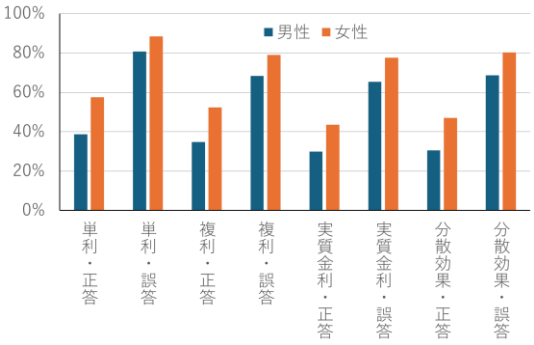
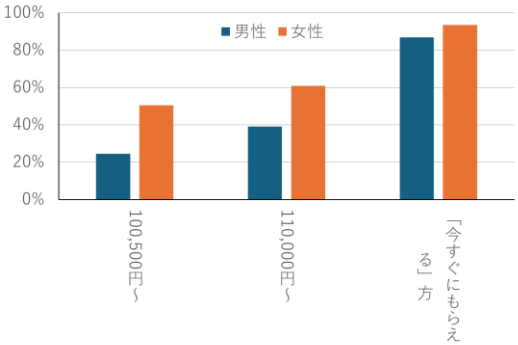


図 12　金融リテラシー・個別評価 Q38～41



最後に時間選好率との関係をみる。アンケートの Q45 では、「仮に、『今すぐ 10 万円をもらえる』か、または『1 年後に 10 万円より多い金額を確実にもらえる』という機会に恵まれたとお考え下さい。もしあなたが『1 年後にもらう』方を選ぶとしたら、いくら以上であれば『1 年後にもらう』方を選びたいと思いますか。」という調査が行われている。時間選好率を細かく計測するため、複数の選択肢を用意しており、ここでは「今すぐもらう」、「時間選好率 10%以上」、「同・10%未満」の 3 カテゴリーに集約した。投資回避割合は「今すぐもらう」で 87%（男性）・94%（女性）と最も多く、10%以上で 39%（男性）・61%（女性）、10%未満で 25%（男性）・51%（女性）となっている[17]。時間選好率が高い方が投資回避的であり、特に、「今すぐもらう」を選択したカテゴリーの投資回避割合は、他の属性分析に比べて突出して高い（図 13）。男女の格差が小さいのも特徴となっている。

図 13　時間選好率 Q45



注）選択肢の最低金額は 10 万 500 円（年率 0.5%）、最高金額は 15 万円（年率 50%）である。本文の「10%」は 11 万円（1 年後受取額）に対応する。

[17] 本図の割合は各時間選好カテゴリー内の割合であり、各カテゴリーの数値を合計しても 100%にはならない。

選択肢の最高金額は 15 万円（年率 50%）であり、「今すぐにもらう」に次いで多かったのが年率 50%であった。これは調査対象 6 カ国に共通した特徴である。金融資産への投資を考えた場合、「確実に」年率 50%で運用できる投資商品はハイパーインフレ国の銀行預金でもない限り存在しないであろう。時間選好率に関する調査結果は、本稿のリサーチ動機である不合理な投資選択（損失は生じず、悪くともゼロリターンである投資を多くの人が回避する）と並ぶパズルである。

行動経済学的な視点からの説明仮説としては、以下のような様々な仮説が考えられる。i)お金をもらえる行為と投資行為は同じではない（あるいはアンケート調査では両者を比較考慮することで合理的な選択を行うようなことはなさず、狭いフレーミングでの選択に陥りやすい[18]）、ii)後で確実にもらえると言われても信用できず、不確実性プレミアムを上乗せする（リスクや不信の混入）、iii)アンケートで仮想質問をすることに無理があり、真剣に考えない（仮想バイアス）、iv)年率換算を行って冷静に考えることができない・しない（一種の設問不理解・思考コスト節約）、v)お金は同一でなく文脈依存であり、臨時収入と勤労所得や資産投資は同一に認知されない（一種のメンタルアカウンティング）などである。これらに対し、経済合理的な仮説としては、vi)資金流動性に直面しており現在の資金入手に強いバイアスが働いているという理由も考えられる。

資金流動性制約仮説はアンケート調査の範囲内で検証可能である。年収、金融資産残高とクロス集計して投資回避者の割合を計測した（表 4）。仮説通り、年収や金融資産残高が低い方が投資回避的であった[19]。しかし、年収・金融資産残高の高低カテゴリー間のギャップは、ほとんどの場合、数%に止まっており、時間選好率のカテゴリー間のギャップに比べれば僅かである。女性・10%未満のように例外ケースも存在している。このことは、資金流動性制約の説明力は一部に止まるものであり、年収水準が高い、あるいは多くの金融資産残高を保有していても、高い時間選好率が示される原因は別にあることを示唆している。

---

[18] アンケート調査に限らず、現実の借入・投資行動においても、リボ払いなど高利の借入金（確定金利）を抱えたままで、比較的安全な投資機会（運用利回りは借入金利より低い）を探す行為は観察されている。資産利回りと負債利回りをトータルで考えず、別物として管理することはありうる行為である。

[19] 世帯年収、世帯保有金融資産残高の高低については、それぞれ、厚生労働省・国民生活基礎調査の世帯年収の平均値、金融広報中央委員会・家計の金融行動に関する世論調査の金融資産保有額の平均値を参考に設定した。本調査では、世帯年収および世帯保有金融資産残高を金額帯の選択肢として把握しているため、外部統計の平均値と厳密に一致する閾値を設けることは難しい。そこで、平均値を概ね上回る水準であり、かつ回答選択肢上の区切りが明確な金額帯として、世帯年収 600 万円以上、世帯保有金融資産残高 2,000 万円以上を高水準として区分した。

表４　年収・金融資産残高、時間選好率別の投資回避者割合

（%）

	時間選好率					
	男性			女性		
世帯年収	10%未満	10%以上	「今すぐ」	10%未満	10%以上	「今すぐ」
無収入～600 万円未満	26	39	85	47	59	92
600 万円以上	18	31	79	41	46	88
世帯保有金融資産残高						
100 万円未満～2,000 万円未満	21	33	81	39	51	90
2,000 万円以上	20	29	73	42	41	84

注）シャドーは、年収（金融資産残高）の高低カテゴリー間で投資回避割合が高い方に付している。

４．定量分析手法と結果

第 4 節では定量分析手法と結果を示す。第 3 節では記述的に男女別の比較を行ったが、機械学習手法およびロジスティック回帰分析において性別を説明変数として組み込むことで、他の要因を統制した上で男女差を評価している。まず、データの前処理を行い、各属性を本稿での統計処理に適した二値変数へと変換する。以下の 4 つの分析を行う。なお、従属変数は「買う（投資選択）＝1、買わない（投資回避）＝0」（買う・買わないはアンケートの表現）としている。①カイ二乗検定により各変数と投資回避・選択との関連性を計測（探索的スクリーニング）、②機械学習手法（決定木、ランダム・フォレスト、勾配ブースティング）により重要変数を特定、③ベイジアン・ネットワークにより変数間の条件付き依存構造を可視化、④ロジスティック回帰分析により各変数の影響の方向性・強度（限界効果）を定量評価し、介入効果を試算する[20]。

４．１　データ前処理

変数間の関係の解釈性を高め、セル内観測数を十分に確保することを目的に、すべての変数を二値変数化して説明変数として使用する（表５、統計処理の詳細は Appendix を参照）。カイ二乗検定では、多値カテゴリー変数においては Cramér's V 等の効果量を用いることが一般的であるが、本稿では全変数を二値に統一することで効果量を ϕ 係数に統一し、変数

[20] 本章における統計検定および各分析モデルの構築には、Python（ver. 3.13.13）を使用した。主要ライブラリとして、記述統計および統計検定には SciPy（ver. 1.16.3）、決定木、ランダム・フォレスト、勾配ブースティングには scikit-learn（ver. 1.7.2）、ベイジアン・ネットワーク（因果探索）には causal-learn（ver. 0.1.4.4）、ロジスティック回帰分析には statsmodels（ver. 0.14.5）をそれぞれ用いた。

間での関連性の強さを直接比較可能にしている。ベイジアン・ネットワークは連続変数を扱うことも可能ではあるが、本稿では条件付き確率表（CPT）による直接的な確率推定と、条件付き独立性の検定の安定化を目的として変数を離散化する。例えば、「認知→投資経験→ポジティブ認識」といった依存構造を推定する際、二値変数であれば各ノードの状態数が少なく、CPT のパラメータ数が抑えられるため、限られたサンプル数でも推定が頑健となる。機械学習では連続変数を扱え、分岐の閾値もモデル推計で決定できるが、本論文では事前に閾値で分割することで結果の解釈が容易となることを優先した（詳細後述）。二値変数化によりシンプルなツリー分岐として提示でき、投資回避者を分かつ属性の境界が明確になる。ロジスティック回帰分析の説明変数は連続変数も扱えるが、二値化することで「当該属性の有無による確率変化」として限界効果を直感的に解釈できる利点がある。ただし、二値化は情報量の損失を伴うため、本稿では理論的に線形関係が想定される年齢については連続変数として投入している。また、変数を統一することで異なる分析手法を横断的に観察しやすくなる。ただし、連続変数を用いた分析による頑健性検証は今後の課題である。

表 5　質問項目と変数名

質問項目	変数名（カイ二乗検定、機械学習手法）	変数名（ロジスティック回帰分析）
Q1 性別	性別	同左
Q2 年齢	年齢	年齢_L(中間値)
Q3 居住地	居住地	同左
Q4 学歴	学歴	同左
Q7 世帯年収	年収・低、年収・高、年収・NR*	年収・高_L(ダミー)、年収・低_L(ダミー)
Q8 世帯保有金融資産残高	資産残高・低、資産残高・高、資産残高・NR	資産残高・高_L(ダミー)、資産残高・低_L(ダミー)
Q10.1 国内株式の認知度	内株認知度	同左
Q11.1 国内株式の投資経験	内株投資経験	同左
Q12 国内株式に対する認識	内株認識・P、内株認識・N、内株認識・I**	内株認識・P_L(ダミー)、内株認識・N_L(ダミー)
Q18.1 暗号資産の認知度	暗号認知度	同左
Q19.1 暗号資産の投資経験	暗号投資経験	同左
Q38～Q41 金融リテラシー	FL・単利、FL・複利、FL・実質金利、FL・分散効果、FL・全体	同左（FL・全体は除く）
Q42 リスク回避・選好度	従属変数	同左
Q45 時間選好率	時間選好率	時間選好率_L(反転)

注）ベイジアン・ネットワークでは、カイ二乗検定および機械学習手法で用いた変数のうち主要なものを採用している（後出）。また、変数名末尾の「_L」は、ロジット分析用に整理した変数であることを示し、括弧内は各加工内容（中間値利用、ダミー変数化、解釈のための符号反転）を指す。詳細は後述の分析セクションにて記す。

*) NR (Non-Response)は「分からない」および「答えたくない」。

**) P はポジティブイメージ、N はネガティブイメージ、I は無関心。

また、日本のアンケート調査固有の事情もある。年収や資産額に対して無回答が突出して多く、これらを連続変数として扱うと欠損値となり分析から除外されてしまう。無回答自体を独立したカテゴリーとしてダミー変数化することで、「自己の経済状況の把握や開示への消極性」という情報を分析に組み込める。この変数が投資回避行動と関連する可能性は、本稿の分析で実際に確認されている。このほか、年収や金融資産残高については、投資行動への影響が線形または対数線形であるとは限らず、一定水準を境とした閾値効果（例：資金流動性制約が緩む水準で投資が可能になる）の存在が想定される。連続変数として線形関係を仮定するとこうした非線形性を捉えにくいが、二値化は閾値を挟んだ平均的な差分として効果を把握でき、非線形性に対してより頑健な計測となりうる。これらの実データ特性も前処理における検討要素である。

変数の二値化にあたっては、年齢は 50 歳、年収は 600 万円、金融資産残高は 2,000 万円をそれぞれ閾値として設定した。これらの閾値は、総務省統計局・国勢調査、厚生労働省・国民生活基礎調査、金融広報中央委員会・家計の金融行動に関する世論調査における年齢・世帯年収・金融資産保有額の平均的な水準を参考に、各アンケート選択肢の区切りとの整合性を考慮して設定している。詳細は Appendix を参照。

４．２　カイ二乗検定

まず、投資の選択と回避を分かつ要因を識別するために、投資選択を従属変数（買う＝1、買わない＝0）とし、各説明変数との関連を Pearson の独立性検定により評価する。本検定は予測精度の評価ではなく探索的スクリーニングを目的とするため、全サンプル（10,000）を用いる。全変数が二値変数であるため、統計量は以下で定義される。

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

ここで i, j は分割表の各セルを表す添字、O_{ij} は観測度数、E_{ij} は期待度数であり、自由度は 1(=(2-1)×(2-1))である。

本分析には 2 つの統計的課題がある。第一に、サンプルサイズが大きいため、実質的にはごく小さな関連しかない変数であっても、p 値が低く算出され有意と判定されやすい。第二に、多数の単変量検定を並列して行うため、多重比較による偽陽性（Type I error）の増大が懸念される[21]。前者に対しては効果量である ϕ 係数を関連性の強さの指標として用い、後者

[21] 複数の検定を独立に並列実施すると、検定数が増えるほど少なくとも 1 回は偽陽性が発生する確率（Family-Wise Error Rate：FWER）が指数関数的に増大する。最も保守的な Bonferroni 補正では各検定の有意水準を厳しく設定するが、ここでは「有意と判定されたものの中に偽陽性が混じる割合（FDR）を制

に対しては False Discovery Rate (FDR)を制御する Benjamini–Hochberg 法に基づく q 値を統計的有意性の指標として用いる。ϕ 係数の絶対値は 0〜1 の値をとり、投資回避・選択と各説明変数の関連性が強いほど 1 に近い値をとる。可視化の際に、有意性の差を識別しやすくするため、$-\log_{10}(q)$と対数変換を用いる。値が大きいほど、多重比較を考慮しても統計的有意性が高い（偽陽性である可能性が低い）ことを示す。

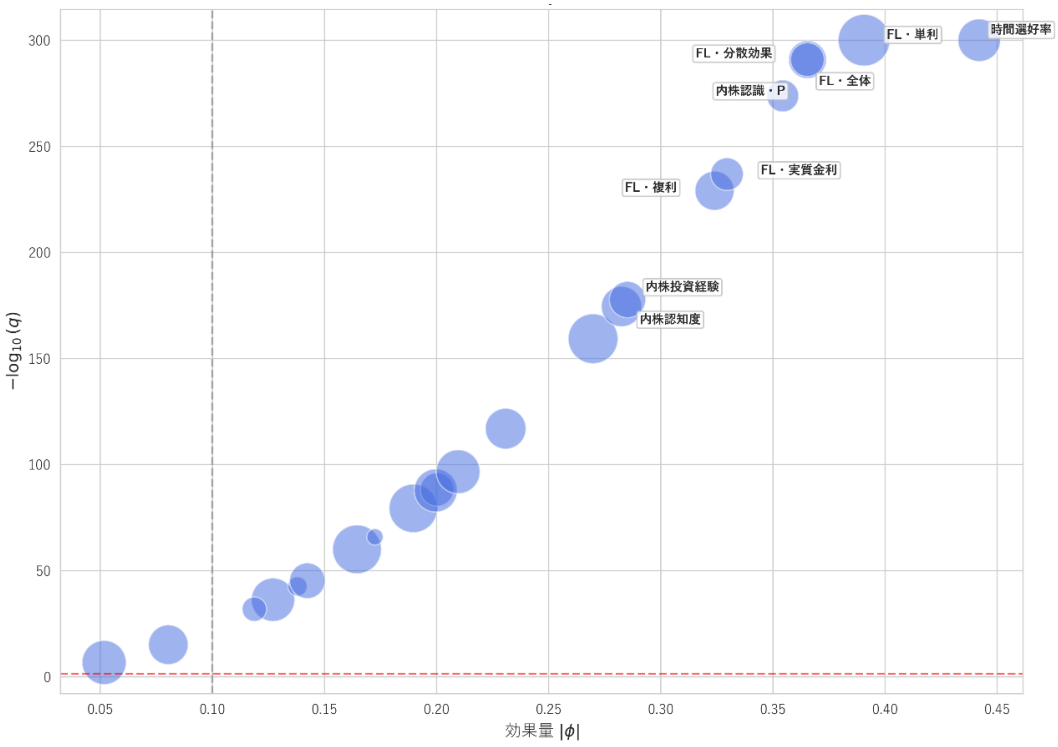
図 14 では、各変数の効果量（ϕ 係数：x 軸）と$-\log_{10}(q)$：y 軸を同一グラフ上に示し、関連性の強さと統計的信頼性を確認した。破線は $q = 0.05$ および$|\phi| = 0.10$（Cohen (1988)の small effect の目安）の参照線であり、効果量と有意性の大小を視覚的に把握するために付している[22]。また、各サークルの円の面積は、全回答者における各属性の出現率（サンプル内における当該属性の構成比）を示している。これにより、関連性の強さ（ϕ 係数）だけではなく、その属性が調査対象全体においてどの程度の広がり（マジョリティ性）を持っているかを視覚的に把握できる。

図 14 では、多くの変数が縦軸の信頼度で有意となっているため、ϕ 係数に着目して評価すると、時間選好率の同係数が最も大きく、金融リテラシー（FL・単利、FL・全体、FL・分散効果、FL・実質金利、FL・複利）は、いずれも同係数全般的に大きい値を示している。次いで、国内株式に対するポジティブ認識と投資経験、認知度（それぞれ、変数名は内株認識・P、内株投資経験、内株認知度）が高い値となっている。なかでも、ポジティブ認識は金融リテラシーと同様な水準となっており、認知や投資経験と差が存在している。単なる認知よりもポジティブ認識の方が関連性が強いという結果は、いくつかの解釈可能性を持つ。第一に、過去に投資を経験していても、その結果が低パフォーマンスであった場合、その後の投資回避行動に繋がる可能性がある（投資損益に関する設問はアンケート調査に含まれるが、本稿では分析に用いていない)。第二に、経験の有無という事実よりも、「投資にはメリットがある」という能動的な信念のほうが、投資行動を強く規定している可能性がある。ただし、本検定は単変量の関連性を示すものであり、変数間の関係の方向性の解釈については第 4.6 節のベイジアン・ネットワークで示唆を得る。ポジティブ認識が強い影響力を持っている点は後述する他の分析手法でも観察されている。

御する」という Benjamini-Hochberg 法を採用している。p 値を Benjamini-Hochberg 法で補正したものが q 値である。

[22] 図中に示した破線は、分析対象を選別・棄却するための閾値ではなく、効果量としては小さいものの無視できない最低限の水準（$|\phi| = 0.10$）と、多重比較を考慮した場合に関連が偶然である可能性が相対的に低い水準（q＝0.05）を視覚的に把握するための参照線である。

図 14　効果量 ϕ 係数と q 値による有意度



注）統計的に極めて高い有意性を有する一部変数（FL・単利、時間選好率）は、有意確率（q 値）がコンピュータの浮動小数点演算の限界に達し実質的に 0 として算出される。そのため、可視化における順序の整合性を保つ目的で、$-\log_{10}(q)=300$ を上限（天井値）としてプロットしている。

実質的な関連性が大きい変数のうち、特に高い効果量を示した客観的知識指標（FL・単利）および意思決定特性指標（時間選好率）を代表的要因として取り上げ、投資回避・選択の二値（回避がゼロ、投資が 1）に対して、単利の正答・誤答（1・0）や時間選好率の高低（1・0）の組み合わせ 2×2 のセルのうち、どの組み合わせがカイ二乗統計量を押し上げているのかを確認する。各セルについて以下のピアソン残差（Pearson Residuals）に基づく 2×2 のヒートマップを作成した（図 15）。

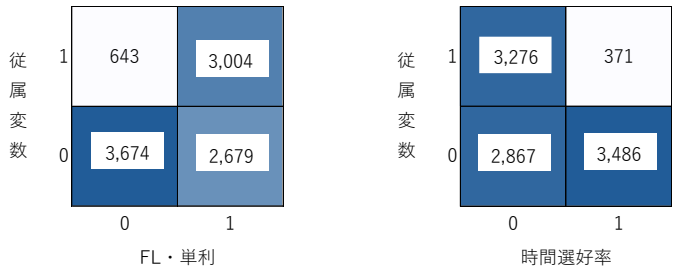
$$r_{ij} = \frac{O_{ij} - E_{ij}}{\sqrt{E}_{ij}}$$

ここで i,j は分割表の各セルを表す添字、O_{ij} は観測度数、E_{ij} は変数同士が独立であった場合に予想される期待度数を示す。

各セルのピアソン残差は、観測度数と期待度数の差を標準化したものであり、カイ二乗統計量は全セルのピアソン残差の二乗和（$\chi^2=\Sigma r^2$）として表現できる。したがって、ピアソン残差の絶対値が大きいセルは、カイ二乗統計量への寄与が大きい。慣習的には $|r|>2$ が顕著

な偏りの目安とされる（標準正規分布の両側 5%水準に対応）。単利を理解していない群（0 値）では、投資回避（0 値）が期待度数を大きく上回り（濃い青）、投資（1 値）が期待度数を下回っている。一方、単利を理解している群（1 値）では逆のパターンが観察されており、両群の分布構造が大きく異なっていることがわかる。時間選好率については、高い（今すぐ 10 万円を受け取る群）ほど、（両者が独立であった場合に比べて）過剰に投資回避的であることを示している。図 15 を比較すると、両変数ともに濃い青のセル（期待度数より過剰に投資回避的）が観察されているが、時間選好率の「高時間選好×投資回避」セルはサンプル数 3,486 と最多であり、時間選好率が投資回避者のマジョリティを特徴づける変数として機能していることが視覚的に示されている。

図 15　ピアソン残差のヒートマップ図



注）青色が濃いほどピアソン残差が大きい。

ただし、カイ二乗検定は単変量手法であり、他の属性による交絡を統制できていない。そのため、ここで観察された関連は因果構造や調整済み効果を直接反映するものとは限らない。本段階はあくまで「どの変数に関連がありそうか」、「どの変数で偏りが生じているか」を把握する探索的スクリーニングである。次節以降では、まず、機械学習手法により変数の重要度を多変量で評価し（第 4.3〜第 4.5 節）、次にベイジアン・ネットワークで変数間の依存構造を検討し（第 4.6 節）、最後にロジスティック回帰分析で各変数の影響の方向性と大きさを推定する（第 4.7 節）。

4．3　決定木

機械学習とは、データから規則性（パターン）を学習し、予測・分類・推薦・意思決定などの性能を改善していく計算手法・研究分野である。学習は、古典的には「ある課題に対して、経験（データ）を通じて性能指標が向上するなら、そのシステムは学習している」とされる

ことが多い[23]。データ量と計算能力の向上を背景に、機械学習は経済学や金融研究を含む様々な分野で活用されている。決定木やその応用（ランダム・フォレストや勾配ブースティングなど多数）は機械学習の代表的な手法であり、現在も広く利用されている。これらの技術的詳細は例えば毛利（2023）や坂本（2019）を参照。

前小節では単変数の影響をみたが、本節では変数全体の情報を同時に利用し、どの変数が投資回避に影響を与えるかを評価する。まずは、この問題に対して決定木（Decision Tree）を適用する。決定木は、強い影響力を持つ変数を順に取り上げ、これに沿って分類していくことでターゲットをできるだけ高い精度で識別する手法である。分岐条件を層状に積み重ねて枝を分岐させていくため決定木と呼称されている。具体的には、ノードとして選ばれた説明変数の値（分岐条件、閾値）によって、全サンプルを小さなノードに分割していくことを繰り返す。どの説明変数でサンプルを分割するかが重要となり、その決定には、ジニ不純度が用いられる[24]。ジニ不純度とは、分割された群内の混ざり具合（不純度が高いほど混ざり合いが高い）を示す指標で、分割後の群がなるべく同じ特徴を持つ人の純粋な状態（不純度が低い状態）になるように決定木が構成される。これは変数の選択順と分割の区切り目を最適に選択するアルゴリズムによって実行される[25]。

以下、本稿で適用した分析手順を示す。

（1）従属変数を投資回避・選択とし、表 5 に示した変数を説明変数（特徴量）とする。

（2）データをトレーニングセット 8 割とテストセット 2 割に分割する[26]。

---

[23] AI はより広い概念であり、知的に見えるタスク（推論、計画、認識、言語理解、意思決定など）を機械に行わせる技術・研究分野を指すため、機械学習は広義 AI に含まれる。第一世代の AI は専門家による If-then ルールで動くものが中心であり、データから規則性を学ぶ機械学習ではなかった。将棋や経路探索のように、状態空間を探索して最適解を見つけるものも古典的 AI であるが機械学習ではない。

[24] ノード内のクラスの総数を K（本論文では K＝2）、ノード内のデータがクラスiに属する確率を p_i とすると、ジニ不純度は Gini=$1-\sum_{i=1}^{K} p_i^2$ で定義される。その値が 0 に近いほど当該領域は純粋であり、本稿のように 2 クラスの場合は 0.5 に近いほど不純であることを示す。なお、分割指標にはターゲットの再現率（捕捉能力）がより高かったジニ不純度を採用した。後続のアンサンブル学習等で用いるエントロピー指標による分析も並行して実施したが、抽出される主要な変数、主要な分岐構造、および各ノードにおけるクラス構成比に大きな相違がないことを確認しており、本結果の頑健性は確保されている。また、ジニ不純度はエントロピーよりも、ノード内の多数派クラスを純化して分離する傾向がやや強く、投資回避層というマジョリティの特徴把握という本研究の目的に適している。

[25] これは、全変数の組み合わせを網羅的にクロス分析することで、説明対象（ここでは投資回避者）を効果的に分離し、どの属性の組み合わせに投資回避者が集中的に所属しているかを検証する方法である。2 値をとる変数が 20 タイプあった場合、組み合わせの総数は 2 の 20 乗で約 105 万にも上る。細かすぎると疎（所属する観察データがゼロのセルが多数）となるため、対象を多くカバーするような「変数と分割基準の組み合わせ」を広範に探索する必要がある。そのアルゴリズムを計算量が増えないように考える必要があり、決定木はシンプルだが計算節約的な手法となっている。決定木だけの単体では精度が上がりにくいため、予測が目的の場合には、ランダム・フォレストや勾配ブースティングのように決定木を部品として組み合わせるアンサンブル学習が利用される。

[26] 決定木による分割の結果、ツリーの最上位にある根ノードにおけるサンプル数は 8,000 となる（図 16）。

（3）決定木モデルの構築に際しては、モデルの視認性と解釈性を優先し、決定木の最大深度を 3 に制限する。モデル性能の評価ではトレーニングセットとテストセットのいずれも活用し、インサンプルとアウトサンプルでの予測力評価を行う[27]。テストセットを用いた検証結果は、正解率が 0.75 であった。特に、分析の主眼である投資回避層については適合率が 0.87、F1 スコアが 0.78 と高い水準にあり、簡素なモデルながら十分な判別能力を有していることが確認された。なお、従属変数が「買う 3〜4 割、買わない約 6 割」とサンプルがやや不均衡であることから、比率の大きいデータの影響を小さく（比率の小さいデータの影響を大きく）するためにデータに重み付け（クラスウェイト）を行った。以降の機械学習手法でもサンプル不均衡に配慮し、適切な重み付けを適用している。

（4）説明変数の与え方で結果が異なるため、最終的なモデルでは重要な説明変数のみが利用されるよう段階的な絞り込みを行う。これは、モデルの複雑さを減らし解釈を容易にする効果もある。具体的には、分割前後のジニ不純度の減少量から算出される各説明変数の重要度合い（特徴量重要度）[28]を用いた。特徴量重要度の 0.01 は、慣習的に無視できないほどの寄与とされる目安であり、これを超える変数のみを最終モデルの説明変数として残した[29]。図 16 は絞り込み後の最終モデルである。

上記の分析手続きにより、最上部の根ノードから始まり、全体のサンプルを各ノードの分岐条件に従って枝の方向に分割し、最終的に 1 番下部の葉ノードに分類した決定木が得られる。各ノードには以下 4 つの情報を示している。ノード内の上から順に、ジニ不純度、当該ノードに分類されたサンプル総数、各クラスの重み付きサンプル数内訳（左から「買わない」、「買う」）、ノード内の多数派である。葉ノードでは、投資回避を選択したサンプル数が多かったものに赤丸を付している（図 16）。

---

[27] 決定木の評価指標として、正解率（Accuracy）、適合率（Precision）、再現率（Recall）、F1 スコア（F1-score）等が使用される。

正解率：全体のサンプルのうち、正しく予測されたサンプルの割合を示す。モデルが全体としてどれだけ正確に予測できているかを評価する指標。

適合率：モデルが正例と予測したサンプルのうち、実際に正例であるサンプルの割合を示す。誤検出を減らすことが重要な場合に重視される指標。

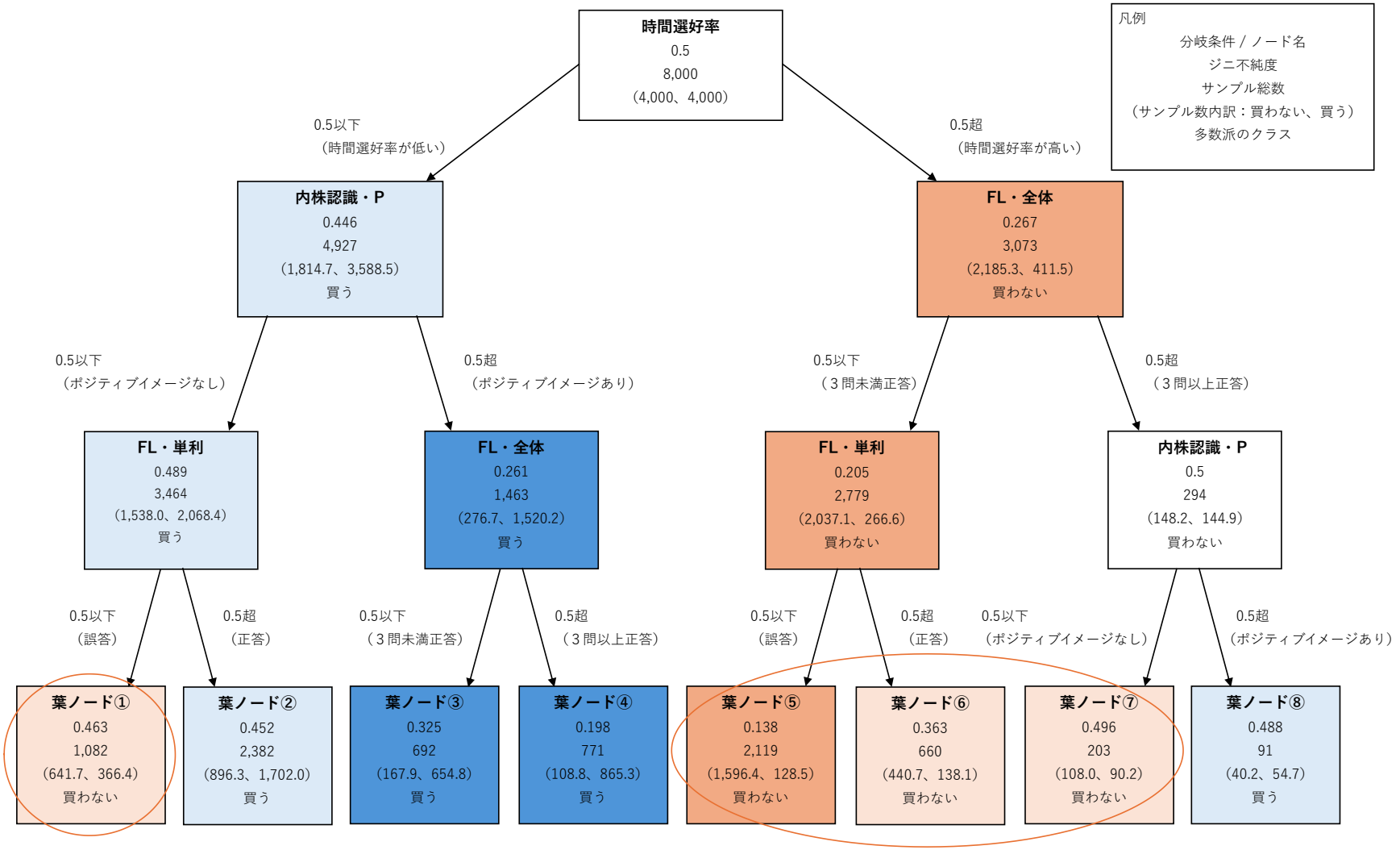
再現率：実際の正例のうち、正しく正例と予測されたサンプルの割合を示す。見逃しを減らすことが重要な場合に重視される指標。

F1 スコア：適合率と再現率の調和平均を示す。適合率と再現率のバランスを評価する指標。なお、全クラスの指標を単純平均したものは、マクロ平均（Macro Average）と称される。

[28] 特徴量重要度には、不純度の減少量、分割に使用された回数、網羅したデータの範囲などに基づく算出方法がある。本稿では、分割前後のジニ不純度の減少量に基づく方法（平均不純度減少）を採用した。重要度算出の考え方は、ランダム・フォレスト、勾配ブースティングでも共通である。

[29] 二段階の推定手順は、ノイズとなる変数を排除し、真に支配的な分岐ルールを抽出するための変数選択プロセスとして採用したものである。一段階目の全変数モデルと二段階目の絞り込みモデルにおいて、上位変数の選択に変化がないことが確認されている。

図 16　投資回避に影響を与える説明変数の決定木



決定木分析の結果を解釈すると以下の通り。

- 投資回避の選択を説明するのに最も重要な説明変数は時間選好率であり、決定木の推計では根ノード（第 1 階層）に同変数が選択された。図中では、時間選好率が低い場合は左の枝に進み、高い場合は右の枝に進む。全体 8,000 人のうち、4,927 人が時間選好率が低いノードに、3,073 人が高いノードに振り分けられている。右側の高時間選好率ノードでは、重み付きサンプルによる内訳をみると「買わない」が 2,185 人、「買う」の 412 人より圧倒的に多くなっている[30]。また、第 3 階層以下の各ノードの（買わない・買う）の多数派をみても、葉ノード⑧以外ではすべて「買わない」が多くなっている。このため、時間選好率が高いほど投資を回避しやすいといえる。

- 時間選好率以外の変数と組み合わせることで、買わないグループの特徴を更に明確化していく。まず、右側第 2 階層に登場する変数は FL・全体である。金融リテラシーが低いと左側の枝に、高いと右側の枝に進む。左に進んだ人数は 2,779 人、右は 294 人となっており、時間選好率が高いグループには低金融リテラシーが多いことがわかる。その内訳をみると、「買わない」が 2,037 人と圧倒的に多数派である。一方、時間選好率は高いが金融リテラシーは高いグループは、買うと買わないが 140 人台で拮抗している（ノードには単純に数値比較した結果として「買わない」を付している）。拮抗しているとは、ジニ不純度による分離が十分に進んでいないことを意味しており、「このノードにおける高金融リテラシー」は分離効果（説明力）が低いといえよう。

- 第 3 階層では、高時間選好率・低金融リテラシーのグループを分割する変数として FL・単利が採択されている。正誤どちらの枝に分岐しても葉ノード⑤と⑥はともに「買わない」が多数派であるが、内訳をみると誤答の葉ノード⑤のほうが「買わない」選択肢の構成比率が顕著に高くなっている（1,596 人対 129 人）。第 2 階層と第 3 階層が類似した変数であるため、高時間選好率・低金融リテラシーほど投資回避的という結論に変わりはない。

- 一方、高時間選好率であるが金融リテラシーは高いというグループに対して、第 3 階層で選択された変数は内株認識・P である。同グループは少数ながら買う・買わないが拮抗していたが、国内株式にポジティブイメージがある右側の枝では、葉ノード⑧で「買う」がやや多くなっている。逆にポジティブイメージがない枝側の葉ノード⑦では、「買わない」がやや多くなっている。ジニ不純度の視点からは、ここでの内株認識・P は分離効果が低いといえよう。

---

[30] 左が「買わない」を示す。重み付きサンプルを使用していることから、小数点を含む値になることがあるほか、クラス毎のサンプル数を加算してもサンプル総数（重みなしの実サンプル数）と合わない場合がある。

- 次に、低時間選好率の枝の分岐をみていく。第 2 階層の変数として内株認識・P が採択されている。「買わない」方は低金融リテラシーによる識別が効果的であったが、決定木全体の左側の「買う」が主力の枝では、ポジティブイメージがあることの「買う」への効果が重要と判断されている。図中第 3 階層の左から 2 番目のノードではポジティブイメージの効果をみることができる。買わないは 277 人に対して買う 1,520 人と非常に多く、このノードと分岐先の葉ノード③と④の 3 つでは、買うの構成比率が順に、85%、80%、89%となっている。表 1 でみたように全体の買わない選択者が 66%もいる（買う選択は僅かに 34%）なかで、特徴的なグループになっていることがわかる。これに対し、ポジティブイメージがない第 3 階層の最左ノードでは、買うが多数派ではあるが、買う 2,068 人に対して買わない 1,538 人と、その差が小さくなっている。
- 低時間選好率・ポジティブイメージありのグループを更に純化分割する変数として、FL・全体が採用されている。高低リテラシーのいずれの分岐先の葉ノードも買うが多数派となっているが、高リテラシーの葉ノード④のほうが買うの構成比がより高くなっている。
- これに対して低時間選好率・ポジティブイメージなしのグループでは、FL・単利が選択されており、誤答したグループの葉ノード①では、決定木全体の左側半分で唯一、買わないが多数派になっている。時間選好率が低くとも、ポジティブイメージがなく、単利に誤答するグループでは、低時間選好率の「買う」効果が打ち消されているといえよう。一方、ポジティブイメージがなくとも単利に正答している葉ノード②では、買うが多数派になっており、大まかな構成比をみると葉ノード①の 2：1 から 1：2 に反転している。これは、ポジティブイメージがなくとも、最も基礎的な金融リテラシーである単利を理解しているだけで、投資選択行動が大きく転換し得る可能性を示唆している。

上記の分析より、本データにおいて投資回避を最も強く分離する変数は時間選好率であることが確認された[31]。第 2 階層の右ノードでは、買う・買わない構成が 412 人対 2,185 人となっており、最初の枝分かれだけで買う層は 3,000 人中の 400 名強まで絞り込まれてしまっている。「買わない」選択を最も強く特徴づける変数となっている。

このグループは 1 年後の高リターンより現在の現金を好む層であり、表 4 で確認したように流動性制約が主因でもないことを考えると、個人の時間選好に働きかけるのは容易ではない。しかし、第 2 節でみたように Q45 の回答で「今すぐもらう」が最大グループであることに鑑みれば、株式等への投資が適切な家計も一定数含まれているはずである。少なくと

[31] ただし、決定木は予測・分類モデルであり、時間選好率への介入により投資回避が変化するかどうかは、後出のロジスティック回帰分析（介入効果の試算）で検証する。なお、変数間の依存構造については、ベイジアン・ネットワーク分析で検討する。

も、時間選好率の高さが何に由来するものかが投資回避の原因を探る重要な研究テーマであることが確認された。

次に、ノードのサンプル人数の大きさからして効果的な方策は、低時間選好率のグループに対して株式投資にポジティブイメージを持ってもらうことである。これにより「買う」の選択が上昇する（図 16 の青色の葉ノード③・④）。金融リテラシーの向上も、これに次いで効果的である。葉ノード①から②への移動（多数派反転）や、③から④への移動（買う比率上昇）により投資回避を抑制する効果を持つことが分かった。

まとめると、金融投資の回避性向の改善には金融リテラシーの向上が重要という指摘がなされているが、それ以上に、現在を重視する時間選好率の高さの克服が重要であり、投資にポジティブイメージを持ってもらうことも金融リテラシー向上と同じかそれ以上に重要であることが判明した。

ここまでは、解釈の容易さを優先し、クラスの分離を視覚的に把握しやすいジニ不純度（ノード内の同質性を高める指標）を用いた単一の決定木による分析をみてきた。しかし、単一の木は、視認性が良いという長所がある一方で、明示的に現れない説明変数の影響、特に交絡効果を検証しにくいほか、過学習し易いといった側面もある[32]。次に、これらの点を改善する観点から、情報理論に基づき不確実性を最小化するエントロピーを指標に用いたアンサンブル学習による頑健性検証を試みた。

４．４　ランダム・フォレスト

ランダム・フォレスト（Random Forest）とは、複数の決定木を組み合わせたアンサンブル学習手法である。単一の決定木は、データセットの違いで構造が大きく変わるというサンプル依存や過学習しやすい傾向を抱えている。これに対し、ランダム・フォレストでは、元データから復元抽出により作成されたサブセット（ブートストラップ標本）で多数の決定木を学習し、得られた推計結果について多数決や平均を用いて結果を算出することで、決定木の課題を克服する。具体的には、各決定木はサブセットを使用して学習され、最終的な予測は各決定木の予測結果の多数決（分類問題では多数決、数値予測の回帰では平均）によって決定される。また、フルサンプルの単一の決定木では明示的に現れなかった説明変数も分析対象とすることができ、より詳細な分析結果を得ることができる。

以下、ランダム・フォレストをここでの分類問題（投資回避・選択）に適用した分析手続きを示す。決定木との違いは 3 番目以降のステップとなる。

---

[32] トレーニングセットを機械が学習しすぎた結果、同セットに過剰に適合し、未知のテストセットには適合できていない（汎用性がなくなった）状態を示す。

（3）ランダム・フォレストモデルの構築では、トレーニングセットのサンプル数は 8,000 に固定し、そこからの抽出をランダムに行うことで複数のサブセットを作成した（サブセット数については、グリッドサーチを用いた交差検証により予測精度を最大化する最適値として複数候補から選定している）。これは、ブートストラップ・サンプリングであり、復元抽出ゆえに同一サンプルがサブセットに複数回含まれることが生じる。

また、決定木の推定においてもランダムさが活用されており、各ノードで分岐を決定する変数の候補をランダムに選択することが行われている。これは、各決定木の構造が似通うことを防ぎ、特定のデータに過度に適合することを抑えるための仕組みであり、多数の決定木を組み合わせることで結果的に変数情報は網羅的に活用される。

ただし、決定木の階層の深さや、決定木の総数（サンプルのサブセット数）、各分岐で見る説明変数の数（もしくは最大数）などのハイパーパラメータは、一つのランダム・フォレストモデルを構成するすべての決定木の間で統一されている。ハイパーパラメータの適切な設定の模索においてはグリッドサーチを実施している。すなわち、木の深さなどについて様々な組み合わせを設定し、個々の分岐ではなくモデル全体の設計において総当たりで全組み合わせを検証する手法を採っている[33]。

また、特定のトレーニングセットにたまたまフィットした設定が選ばれないよう、トレーニングデータ全体を 5 つに分け、4 つを学習用・1 つを評価用にしてモデルを推計することをサブセットデータの役割を入れ替えながら（どれを評価用に使うか）5 回繰り返す作業を行っている。これは 5 分割交差検証法と呼ばれている。その際の予測精度評価にはマクロ平均 F1 スコアを用いている。これは、各クラス（買う・買わない）の F1 スコアを単純平均したものであり、クラスサイズに依存せず両クラスを均等に評価できるため、本分析のようにやや不均衡なデータに適している。

（4）設定されたハイパーパラメータに基づき、トレーニングデータ全体から最終的なランダム・フォレストモデルを構築する。具体的には、サブセットに対して決定木を学習する。予測においては、テストデータに対して各決定木を適用し、その結果を多数決することで予測結果を得る。

次に、各決定木において各説明変数が分岐にどの程度寄与したか（学習時の貢献）をもとに、ランダム・フォレスト全体の特徴量重要度を評価する。本分析では、不純度指標にエントロ

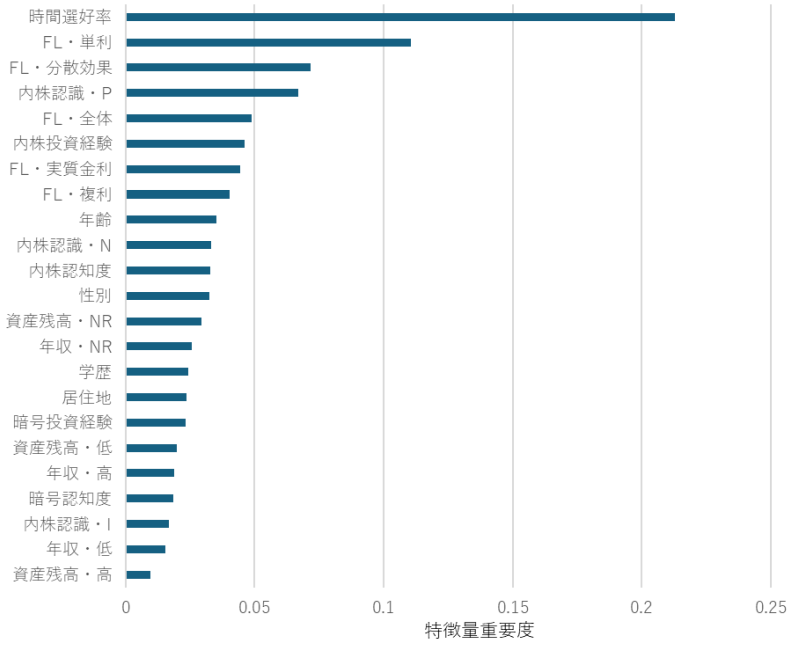
---

[33] ハイパーパラメータの最適化は、5 分割交差検証を用いたグリッドサーチにより実施した。評価指標にはマクロ平均 F1 スコアを用い、探索範囲は、計算効率とモデルの収束傾向を確認するため、決定木の数（n_estimators）を代表的な 3 点{100, 200, 500}、木の最大深さ（max_depth）をモデルの複雑度に応じた 5 段階{None, 5, 8, 10, 12}とした。分割に用いる変数の最大数（max_features）を、標準的な基準とされる sqrt（全変数の平方根、本分析では全変数の約 0.21 倍に相当）および log2（同対数、同約 0.20 倍に相当）周辺を網羅しつつ、より多くの変数情報を保持する場合の性能を確認するため、全変数の{0.2, 0.3, 0.4, 0.5, 0.6}倍という広範な試行を行った。最終的に選定された最適パラメータは、n_estimators: 500, max_depth: 10, max_features: sqrt であり、このモデルによる最良の F1 スコアは 0.7657 であった。

ピーを用い、各分岐における不純度減少量を全決定木で集計した平均不純度減少（Mean Decrease Impurity：MDI）に基づき算出する。一般にこの手法はカーディナリティの高い変数（取り得る値が多い変数）を過大評価する傾向があるが、本分析では全変数を二値化しているため、この影響は限定的である。

以上の分析手続きによって投資回避・選択に影響を与える説明変数とその重要度合い（特徴量重要度）を推定した（図 17）。各説明変数がモデルの予測に対して与える影響を示した重要度（全体合計が 1 になるよう各説明変数に割り当てている）をみると、時間選好率が 0.213 と突出して大きくなっている。これは、前小節の決定木 1 つの分析結果の発見と同じである。次いで FL・単利、FL・分散効果、内株認識・P、FL・全体、内株投資経験、FL・実質金利、FL・複利となっている。これらの説明変数を合わせると全体の 64%を占める。決定木 1 つの分析では深度 3 に制限したため、時間選好率、FL・単利、内株認識・P、FL・全体の 4 つしか重要変数として登場しなかった。ランダム・フォレストでは各分岐の説明変数候補をランダムに絞り込むことで、単一木の分岐に登場しなかった変数も多数の木の中で評価される。その結果、FL・分散効果や FL・複利、FL・実質金利など、より幅広い変数の相対的重要度を計測することができている[34]。

図 17　特徴量重要度：ランダム・フォレスト



[34] 特徴量重要度は時間選好率を最大に、上位 8 変数で全体の 64%を占めるに留まっており、特定の少数変数に過度に依存せず、多様な変数が寄与していることが確認できる。これは、ランダム・フォレストが特定の変数への過度な依存を避け、特徴量重要度の評価を多様化させる特性によるものである。

全体の傾向としては、年齢や年収、性別より、時間選好率や金融リテラシー、株式投資の認識や投資経験がより重要となっていることがわかる。これは、財務余力がないから・若いから・女性だから、投資を回避するという説明よりも、時間選好率が高いから・金融リテラシーが低いから、株式投資にポジティブイメージや投資経験がないから投資を回避するという説明力の方が強いことを示唆している。なお、第 3 節のクロス集計分析では、男女間の投資回避格差が明確に観察されていたが、性別が投資回避・選択への影響度合いを他の変数と比較すると、相対的に小さいことがわかる。これは、性別が真の要因ではなく、時間選好率や金融リテラシーに性差があり、それが交絡効果として性別を重要変数のように見せてしまっている可能性を示唆している。このほか、年収や金融資産残高の高低より、回答を回避した年収・NR、資産残高・NR が上位に来ており、この点は第 3 節の分析と整合的な結果となっている。

以上、みてきたようにランダム・フォレストには特徴量重要度の多角的評価という利点があるものの、i)単一の決定木のような直感的な分岐構造の可視化が困難、ii)特徴量重要度は変数間の相互作用や因果関係を直接表現するものではない、という限界がある。また、具体的な分岐条件やグループ分けを視覚的に観察できるという決定木の利点を活用しにくいという課題もある[35]。これらの問題への対応を考える前に、次節では別のアンサンブル手法（勾配ブースティング）による補完的な特徴量重要度評価を行い、ランダム・フォレストの頑健性を確認しておく。

４．５　勾配ブースティング

機械学習では、勾配ブースティング（Gradient Boosting）も頻繁に用いられる。ランダム・フォレストがサブセットを用いて複数の決定木を独立に学習し、その多数決や平均によって予測を行うのに対し、勾配ブースティングは、複数の弱学習器（ランダムな予測よりわずかに良い精度を持つ単純なモデル、典型的には深さ 3〜5 程度の浅い決定木）を逐次的に追加しながら、前段階の予測誤差を補正するように学習させるアンサンブル学習手法である。予測誤差を段階的に減少させることで予測精度の向上を図る。高い予測性能を示すことが多い一方、学習率、階層の深さ、逐次追加で補正する回数などのハイパーパラメータに敏感であり、設定によっては過学習が生じやすいという特徴がある[36]。前述の決定木やランダム・

[35] 具体的な分岐条件を確認するためには、複数の決定木を 1 つずつ確認する必要がある。そのうえで共通のパターン性を見出すというのは現実的なアプローチではない。

[36] 決定木、ランダム・フォレストと比較してハイパーパラメータの数が多く、調整の難易度が高い。本分析では以下のとおり、ハイパーパラメータの最適化を行うことで、過学習を抑制し、モデルの信頼性を高めている。

手法はランダム・フォレスト同様、5 分割交差検証を用いたグリッドサーチにより実施した。評価指標には

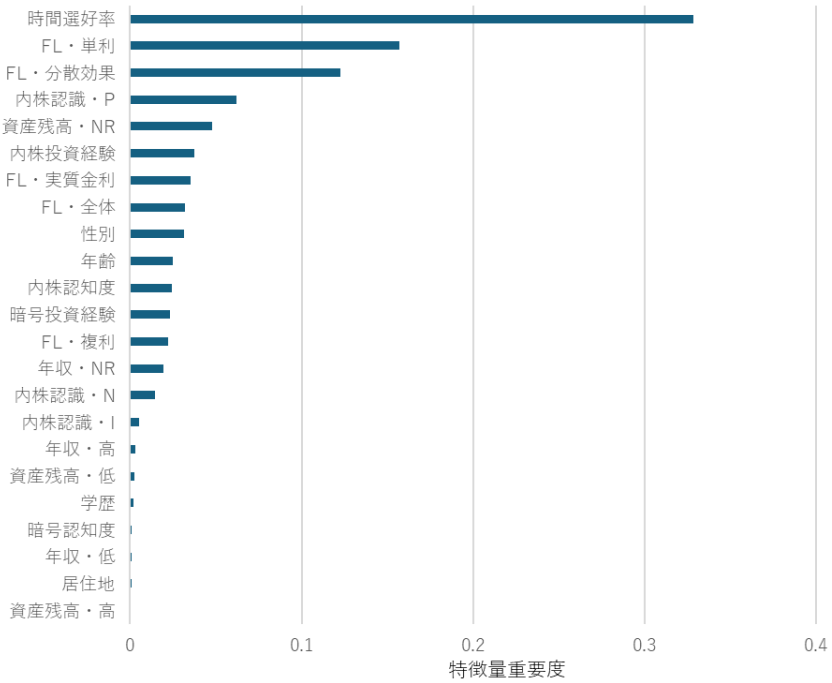
フォレストとの手法間の一貫性を考慮し、Gradient Boosting Classifier を採用した。これにより、探索範囲や評価指標を揃え、手法間での比較の厳密性を図っている。

勾配ブースティングの推計においては、上述の推計方法の違い以外はランダム・フォレストと同様の設定としている。学習が終了した最終モデルにおける各説明変数の相対的な重要度（特徴量重要度）を算出した（図 18）。

時間選好率の寄与は 0.329 とランダム・フォレストによる推計結果以上に高い重要度を示している。次いで金融リテラシーや国内株式のポジティブ認識・投資経験が上位に位置しており、ランダム・フォレストの結果と主要な変数の順位構造はほぼ同内容となっている。相違点は、資産残高・NR がより上位に推計されている点、下位の説明変数の寄与度がより小さく推計されている点である。前者は、勾配ブースティングが逐次的に予測誤差を補正する過程で、無回答層の予測における資産残高・NR の重要性が浮き彫りになった可能性を示唆している。後者は、勾配ブースティングがより集中的な変数選択を行う特性によるものと解釈される（特徴量重要度の分布を確認すると、勾配ブースティングでは上位 8 変数で 82% と、ランダム・フォレストとの対比でより少数の特定変数に重要度が集中する傾向がみられた）。しかし、いずれの手法においても時間選好率が第 1 位であり、次いで、FL・単利、FL・分散効果が続くという上位構造は完全に一致している。分析手法に依存せず主要な変数が共通していることは、本分析から得られた知見の高い頑健性を示すものである。

マクロ平均 F1 スコアを用い、探索範囲は、決定木の本数（n_estimators）を{100, 200, 500}、木の最大深さ（max_depth）をモデルの複雑度を抑制する観点から{3, 5, 8}とした。また、各分割で用いる説明変数の最大数（max_features）についても、前節と同様の探索範囲（sqrt、log2、および全変数の 0.2〜0.6 倍）を設定した。加えて、勾配ブースティング特有の学習率（learning_rate）については{0.01, 0.05, 0.1}を探索対象とした。最適化の結果、learning_rate: 0.1、n_estimators: 100、max_depth: 3、max_features: sqrt が選定され、最良のマクロ平均 F1 スコアは 0.7855 であった。勾配ブースティングは浅い決定木を用いて誤差を逐次的に補正する特性を持つが、本分析においても相対的に少数の決定木（100 本）と浅い階層（3 層）で性能が収束しており、モデルの理論的特性を反映した妥当な結果が得られたといえる。

図 18　特徴量重要度：勾配ブースティング



以上、決定木、ランダム・フォレスト、勾配ブースティングの 3 手法の推計結果をみてきた。投資回避・選択に影響を及ぼす変数の特定に有効であった一方で、説明に用いた変数間の依存関係を示すものではなく、単一決定木でみたように変数の組み合わせ方や、組み合わせの並べ方（分岐の順番）で、投資回避・選択を分類するものであった。

4．6　ベイジアン・ネットワーク

性別が交絡効果である可能性を指摘したように、各変数が投資回避・選択にどのような依存構造をもって影響しているかも重要な情報である。本研究が対象とする株式市場参加パズルの文脈においては、単一の要因のみならず、複数の行動特性や認知要因が連鎖的に関与している可能性が高い。そこで本節では、機械学習手法で特定された主要な変数（影響が強いと推計された説明変数）を対象に、ベイジアン・ネットワーク（以下、BN）を用いて、変数間の依存構造を可視化することを試みた。BN は、変数間の確率的依存関係を有向非巡回グラフ（Directed Acyclic Graph：DAG）と条件付き確率表（CPT）で表現するモデルである（Pearl, 1988）。なお、機械学習手法およびロジスティック回帰分析では、予測性能の評価および比較可能性を確保するため、サンプルの 8 割を学習データ、2 割を検証データとして用いている。一方、BN 分析は予測精度の評価を目的とするものではなく、変数間の依存構造の把握を目的とするため、全サンプルを用いて推定を行っている。

変数としては、機械学習手法で重要度の高かった説明変数群から、時間選好率（最重要変数）に加え、i)金融リテラシーは個別項目よりも体系知を表す「FL・全体」に集約[37]、ii)国内株式関連は認知・投資経験・ポジティブ認識の連鎖構造を検証するため 3 変数すべてを保持、iii)暗号関連は対象者数が限定的（投資経験者は全体の 4%）のため除外、という基準で計 5 つの説明変数を選定した。変数の数を限定したのは BN の特性による。BN の構造学習は、変数を増やすと指数関数的に複雑になる。可能な DAG の数は、例えば変数が 5 個で約 3 万通り、10 個で約 4×10^{18} 通りに達する。サンプル数にかかわらず、変数間の依存関係の候補が膨大になると推定が不安定になるため、主要な 5 つの説明変数に限定して推計を行っている。

本研究で用いるデータは単一時点で収集された横断データであり、変数間の時間的な先後関係を直接識別することはできない。そのため、短期的な経験や学習によって変化しやすい要因と、性向的で比較的安定していると考えられる要因とを区別することが考えられる。本分析では、経済学の標準的な考え方に従い、時間選好率を性向的特性として上流に位置付け、他変数からの影響を受けない要因として扱った。ただし、時間選好率は真に独立で先天的なものかどうかについては様々な研究がなされており、所得・教育などの社会経済属性や、所属社会の平均寿命や年齢・健康、認知能力、幼少期環境、文化・歴史的環境などが影響力を持つという指摘も多い[38]。また、過度な投資回避傾向に対して政策目的で介入を行いたい場合、時間選好率の決定要因の分析は効果的な介入法の発見において重要となる。

BN の推計には、条件付き独立性にもとづく PC アルゴリズムを用いた。変数の特性に基づく背景知識を探索空間に組み込むことで、理論的に不適切な関係を除外している。依存の方向については変数間の条件付き独立性の有無等に基づいて推定されている。分析結果から方向を一意に定めることができなかったケースについては、矢印のない実線（無向エッジ）として示した[39]。

---

[37] 本節では構造の可視化を優先して統合指標を用いるが、次節のロジスティック回帰分析では、各知識要素の固有の影響を精緻に捉えるため、指標を「FL・全体」から各個別項目へと差し替えている。その際、変数間の多重共線性による推定の不安定化を避けるため、全体指標はモデルから除外して分析を行った。

[38] 時間選好率は、本調査に含まれる質問項目から「現在志向性」という性向的特性として測定しており、行動経済学の先行研究でも、時間選好が短期的な行動によって大きく変化しにくいことが指摘されている（Frederick et al., 2002）。この点を踏まえ、時間選好率を他変数からの影響を受けない上流ノードとして固定し、流入を禁止する制約を設けた。なお、上流を固定しない推定も試みたが、国内株式の投資経験から認知へ向かうなど時間的順序と整合しにくい向きが生じたため、本稿では上流を固定した方式を採用した。

[39] PC アルゴリズムを直感的に説明すると次のようになる。真に関係がある変数同士なら様々な変数で条件付けても独立にはなりにくい。逆に、間接的な関係しかない変数同士であれば適切な条件づけを行うことで独立になりうる。まず、全変数が全て接続している完全グラフから開始し、条件付き独立性が見つかったリンク（ノードを繋ぐ線）を順に排除していく。こうして得られたスケルトン（因果の方向性なしの骨格）に対し、分離集合の情報を用いてリンクの方向性を決定できる箇所に因果の方向性を付していく。こうして得られたグラフを部分的有向非巡回グラフ（Partial DAG）と呼ぶ。詳細な計算法については、例えば小川（2020）を参照。完全グラフからスタートするため、変数（グラフのノード）が多いと計算負荷が大きなものとなる。なお、本稿のように機械学習手法とベイジアン・ネットワークを組み合わせて経済データの依存構造を探る先行研究として、日本銀行（2017）や小寺（2020）がある。

図 19 は、以上の方法により推定された BN を示している。同図では、エッジの太さ（線の幅）は ϕ 係数に基づく関連の強さを示し、矢印の向きは条件付き依存関係の方向を表している。すなわち、条件要素の状態が与えられることで、結果要素の生起確率が影響を受けるという順序性が観察できる。これは厳密には因果関係とは区別されるものである。BN は変数間の確率的依存関係を有向非巡回グラフ（DAG）で表したモデルであり、厳密には因果関係を直接示すものではない。ただし、因果的 Markov 条件・忠実性・十分性などの仮定を置けば、BN は因果モデルとして解釈可能となる(Pearl, 2009)。本研究ではこれらの仮定の妥当性を厳密に検証していないため、推定された構造はあくまで変数間の依存関係を示すものとして解釈する。

推定されたネットワーク図より、投資回避・選択の従属変数に対して、大きく二つの影響経路が確認された（図 19 の従属変数に向かう太矢印)。

第一に、「資質と知識による直接的な関連」である。時間選好率から従属変数への太い矢印は、他変数を介さずに成立する強い関連性を示しており、現在志向性という性向的特性が、知識や認知などの状態とは独立に投資回避と強く結び付いている重要な要因であることを示唆している（ただし、時間選好率を上流ノードに固定したことが影響している可能性もある)。同様に、FL・全体から従属変数への太い矢印も、一般的な金融知識水準そのものが投資行動と直接的に関連している可能性を示している。ただし、FL・全体には内株認知度からの入力経路もあり、認知が金融リテラシーを介して従属変数に間接的に影響する経路も存在する点に留意が必要である。

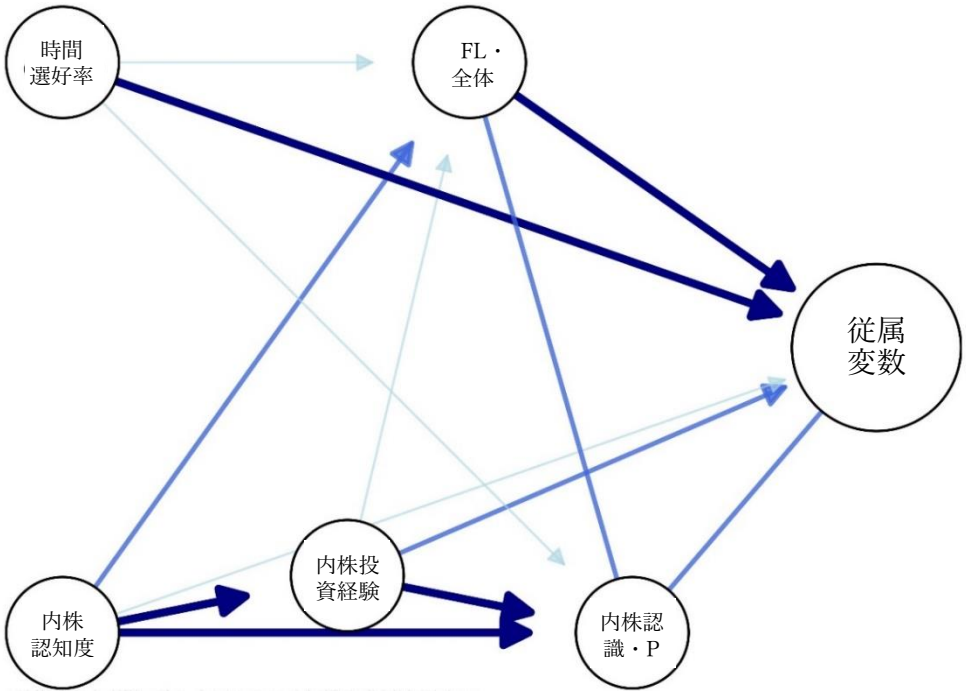
第二に、「プロセス的な経路を通じた段階的な関連」である。本分析では、国内株式の「認知」が、「投資経験」を介して「ポジティブな認識」へと至る段階的な経路に加え、「認知」から「ポジティブな認識」へ直接向かう経路も確認された。内株認識・P から従属変数へのリンクについては、無向エッジとなっており方向性は確認されておらず、関連性の強さを示す ϕ 係数もやや低く推計されている。一方で、内株投資経験から従属変数へ向かう矢印が別の経路として観測されている（やや弱い有向エッジ)。これらより、ポジティブ認識の最終的な投資行動への影響は、BN 分析では確定的に位置付けることはできない。

また、内株認知度から FL・全体へ向かう矢印は、代表的な金融資産である国内株式に関する認知が金融リテラシーに影響する要素となっており、間接的に従属変数に影響する経路を作り出している。現実的には、金融リテラシーが高い人ほど認知が高いという因果方向も考えられる。ここでの推計結果は、BN モデルの変数選択や時間選好率に関する想定のほか、推計の頑健性に依存している。

このほか、FL・全体と内株認識・P の間には無向エッジが確認された。これは、両者の依存関係が確認されたものの、データから条件付き独立性の検定のみでは方向を一意に定められなかったことを示す。

時間選好率から、国内株式の認知・投資経験・ポジティブ認識への矢印は、内株認識・P に一本の弱い関係が観察されただけであり、時間選好率がこれらと直接強く結び付いているわけではないことが推測される。

図 19　ベイジアン・ネットワーク



注）矢印は ϕ 係数（絶対値）の 0.285（33 パーセンタイル点）、0.365（66 パーセンタイル点）を境に 3 区分している。

以上の分析から、投資回避・選択には個々の要因が独立に作用するのではなく、認知・投資経験・認識・知識・性向特性が複合的に関連し合う依存構造の中で形成されている可能性が示された。このように、機械学習や BN は、重要変数の特定や複雑な連鎖構造の可視化において有用な探索的手法である。しかし、それらは個々の要因が投資行動に与える「純粋な影響（他の要因を一定に制御した上での寄与）」を定量的に評価するものではない。そこで次節では、探索段階で得られた知見を踏まえつつ、説明要因が投資行動に与える影響を、方向性と大きさ（係数符号条件や限界効果）、推計の確実性（統計的有意性）の両面から定量的に評価するため、ロジスティック回帰分析を適用する。

４．７　ロジスティック回帰分析

最後にロジスティック回帰分析により、各説明変数が従属変数に与える影響の大きさと方向を定量評価する[40]。具体的には、係数の推定値に加え、限界効果（説明変数が二値の場合は１単位変化に伴う確率変化）を算出し、各要因の影響度を比較可能な形で提示する。表５で示される変数を説明変数とし、機械学習手法との比較可能性を確保するため、ロジスティック回帰分析についても、サンプル全体を８：２に分割した際の学習データ（８割）を推計対象とした（残る２割は予測精度検証に用いている）。なお、従属変数の構成比にやや偏りがあることから、前述の機械学習手法と同様に重み付けを行っている。

本推計では一部の変数について異なる前処理を行っている。

- 年齢（Q2）：アンケート調査では 10 代ごとの離散選択肢となっている。予測精度の向上が見込まれるよう、各カテゴリーの中間値を用いることで年齢を連続的な数値変数として近似的に扱った（例えば、20～29 歳を 24.5 歳とした）。
- 世帯年収（Q7）、世帯保有金融資産残高（Q8）：本推計では、基準カテゴリーを設定したダミー変数を用いた。具体的には、年収・高_L(ダミー)は高年収層である場合を１(それ以外を 0）とし、年収・低_L(ダミー)は低年収層である場合を１（それ以外を 0）とした（資産残高についても同様）（表６）。これら 2 つの変数を同時に説明変数に加えることで、両ダミーが 0 となる年収無回答層を比較基準として、年収情報を回答した各層に属することが投資行動とどのように関連しているかを識別している。この比較は、年収水準の違いという側面に加え、情報開示の側面を反映した効果（開示姿勢の違い）も含まれていると考えられる。本モデルにおいて、各係数は無回答層との差分を表すが、さらに、年収・高_L(ダミー)と年収・低_L(ダミー)の係数を比較することで、年収額を回答している層における年収水準の差（高年収効果）を評価することが可能となっている（資産残高についても同様)。モデル２では、年収・低_L(ダミー)と資産残高・低_L(ダミー)の 2 変数を落とし、低年収層と無回答層を同一カテゴリーとして扱う。これは、モデル 1 において無回答層を比較基準とした際に観測される効果が、純粋な情報開示姿勢の違いか、低年収層と無回答層の混合効果かを識別するための頑健性チ

---

[40] 本分析では、説明変数が投資選択確率に与える影響を直感的に把握するために、確率そのものの変化量（平均限界効果：dy/dx）により影響力を定量評価する。平均限界効果（Average Marginal Effects：AME）は、各観測点における限界効果を算出し、それらを単純平均したものである。これは、平均的な観測点における限界効果（MEM）とは異なり、サンプル全体の平均的な反応を表す。説明変数が二値変数の場合、dy/dx は、説明変数が 0 から 1 に変化した時に投資選択確率が平均して何%ポイント変化するかを示し、確率の変化として直観的に把握しやすい。

質的選択モデルとしてロジットモデルとプロビットモデルのいずれも適用可能である。本分析では、説明変数が投資選択確率に与える影響を平均限界効果として評価し、その結果を確率の変化として直感的に把握できる形で提示することを重視している。この点において、ロジットモデルは係数をオッズ比として解釈できる特徴を有するため、本稿ではロジットモデルを採用し、ロジスティック回帰分析を行った。

ェックである。モデル 1 をメインモデルとし、モデル 2 は補完的な分析として位置付ける。

- 国内株式に対する認識（Q12）：複数回答可の質問であり、ポジティブ・ネガティブ・無関心の各カテゴリーを独立に判定可能である。本推計では、年収や資産残高と同様に基準カテゴリーを設定したダミー変数を用いた。具体的には、内株認識・P_L(ダミー)はポジティブイメージがある場合を 1（それ以外を 0）とし、内株認識・N_L(ダミー)はネガティブイメージがある場合を 1（それ以外を 0）とした（表 6）。これら 2 つの変数を同時に説明変数に加えることで、両ダミーが 0 となる無関心層を比較基準として、特定の認識を持つことが投資行動とどのように関連しているかを識別している。本モデルにおいて、各係数は無関心層との差分を表すが、さらに、内株認識・P_L(ダミー)と内株認識・N_L(ダミー)の係数を比較することで、何らかの認識を持っている層における、認識の内容の差が投資行動に与える影響を評価することが可能となっている。これにより、単なる対象の認知にとどまらず、さらにその認識内容をポジティブな内容へと転換させることが、投資確率をどの程度追加的に押し上げるかという効果を識別している。
- 金融リテラシー（Q38～41）：本推計では、多重共線性による推定の不安定化を避けるため、複合指標である FL・全体を除外し、各知識要素（単利、複利等）の個別の影響を識別するモデルとしている[41]。
- 時間選好率（Q45）：従属変数において「買う＝1、買わない＝0」としていることから、時間選好率_L(反転)において 1 と示された特徴を持つ（時間選好率が低い）ことが従属変数の 1 と整合的になるよう調整した。これにより、係数の符号が正であれば投資促進要因、負であれば投資阻害要因として一貫した解釈を可能にしている。

[41] カイ二乗検定、機械学習手法では、マクロな知識水準とミクロな個別知識のいずれが識別力を持つかをアルゴリズムに選択させるため、意図的に重複を含む全変数を投入した。カイ二乗検定は単変量の独立性検定であり多重共線性の問題は生じないこと、また、決定木、ランダム・フォレスト、勾配ブースティングといった木系手法は変数間の相関に比較的ロバストであることによる。一方、ロジスティック回帰分析は変数間の線形従属性に弱く、重複指標を同時に投入すると係数符号の逆転（抑制効果）など推定が不安定になるため、変数を取捨選択した。

表６　ロジスティック回帰分析におけるダミー変数の構成と比較基準

年収・資産残高			内株認識		
	年収/資産残高・高_L(ダミー)	年収/資産残高・低_L(ダミー)		内株認識・P_L(ダミー)	内株認識・N_L(ダミー)
高額層	1（＝該当する）	0（＝該当せず）	ポジティブイメージがある層	1	0
低額層	0	1	ネガティブイメージがある層	0	1
金額無回答層	0	0	無関心層	0	0

注）シャドーは、各変数の比較基準となる層を示す。

これを踏まえ、ロジスティック回帰式における各説明変数の係数が示す内容をまとめると以下のとおり（表７）。

表７　ロジスティック回帰式における係数の解釈

説明変数	係数の解釈
年齢_L(中間値)	正の係数は、年齢が高くなるほど投資が選択される確率が増加。
上記以外	正の係数は、説明変数が 1 の場合に投資が選択される確率が増加。 （例 1）性別の正の係数は、説明変数が 1（男性）の場合に投資が選択される確率が増加。 （例 2）時間選好率の正の係数は、説明変数が 1（時間選好率が低い）の場合に投資が選択される確率が増加。

ロジスティック回帰分析において説明変数がすべて二値変数である場合、完全分離または準完全分離が生じやすく、推定の不安定化や収束不良に注意が必要となる。また、サンプル数が少ない場合、説明変数の数が増えると組み合わせパターンが急増するため観測数の少ないセルが生じやすく、多重共線性の確認が重要となる。VIF（Variance Inflation Factor）を確認したところ、全ての変数において目安となる 10 を大幅に下回る数値が確認された。さらに、最尤推定が正常に収束しており、推計係数や標準誤差に不自然な値がみられないことを確認している。なお、交差項を用いなかったのも説明変数の組み合わせ数の増加による推定の不安定化を避けるための設計上の配慮によるものである。適合度を Pseudo R^2 で確認したところ、モデル 1 で 0.342、モデル 2 で 0.336 と、いずれも McFadden (1979)の経験則では良好な適合度の目安を示す 0.2 以上の値となっている。

ロジスティック回帰分析の推計結果をまとめると以下のとおりである（表８）。本研究の目的は回避する属性の検証であるため、以下では、推計結果を「投資を行う確率」の増減に注

目し、推計結果を、投資回避行動を説明する視点に引き直して解釈する。なお、特に断りがない限りモデル 1 の数値を紹介している（有意性については概ね同様であるが、限界効果に若干の相違が出ている程度である）。

表８　ロジスティック回帰分析の推計結果

	モデル 1			モデル 2		
説明変数	β	β (s.e.)	dy/dx	β	β (s.e.)	dy/dx
定数項	-2.374***	0.130	---	-2.170***	0.124	---
性別	0.490***	0.061	0.072	0.538***	0.060	0.079
年齢_L(中間値)	-0.022***	0.002	-0.003	-0.022***	0.002	-0.003
居住地	-0.030	0.062	-0.004	-0.012	0.062	-0.002
学歴	0.076	0.061	0.011	0.064	0.060	0.009
年収・高_L(ダミー)	0.370***	0.094	0.054	0.312***	0.070	0.046
年収・低_L(ダミー)	0.199**	0.087	0.029	---	---	---
資産残高・高_L(ダミー)	0.219*	0.126	0.032	0.031	0.114	0.005
資産残高・低_L(ダミー)	0.381***	0.079	0.056	---	---	---
内株認知度	0.089	0.078	0.013	0.076	0.078	0.011
内株投資経験	0.553***	0.081	0.081	0.550***	0.081	0.081
内株認識・P_L(ダミー)	0.706***	0.077	0.103	0.731***	0.076	0.108
内株認識・N_L(ダミー)	0.374***	0.060	0.055	0.391***	0.060	0.058
暗号認知度	-0.107	0.107	-0.015	-0.093	0.106	-0.014
暗号投資経験	1.053***	0.156	0.154	1.122***	0.155	0.166
FL・単利	0.864***	0.072	0.126	0.883***	0.072	0.130
FL・複利	0.244***	0.068	0.036	0.258***	0.068	0.038
FL・実質金利	0.504***	0.074	0.074	0.515***	0.074	0.076
FL・分散効果	0.473***	0.067	0.069	0.497***	0.066	0.073
時間選好率_L(反転)	1.643***	0.071	0.240	1.665***	0.071	0.246
正解率		0.79			0.79	
Pseudo R^2		0.342			0.336	
対数尤度		-3,648			-3,682	
サンプルサイズ		8,000			8,000	

注 1）*** $p < 0.01$、** $p < 0.05$、* $p < 0.1$ を示す。
注 2）機械学習の分析と同様、サンプルに重み付けを行っているが、dy/dx は各回答者（各観測点）について算出した限界効果を単純平均したものであり、平均化の段階では重み付けは反映されていない。

- クロス集計および機械学習手法による分析では、性別の影響は時間選好率や金融リテラシーの性差を介した交絡効果である可能性が示唆されていた。一方、ロジスティック

回帰分析では、性別は他の説明変数の影響を考慮しても有意となっており、限界効果をみると男性である場合、投資を行う確率が 7.2%ポイント高い。すなわち、男性は女性に比べて投資を回避する確率が相対的に低い。本分析結果は、時間選好率や金融リテラシーを統制した後でも性別固有の効果が一定程度残ることを示しており、性別の影響は完全には交絡効果のみで説明できないことが判明した。

- 年齢係数は負値で有意であり、年齢が高いほど投資を回避する傾向があることが計測されている。1 歳増加することの限界効果は－0.3%ポイントであり、10 歳増加ごとに投資確率が 3%ポイント低下する（なお、年齢はアンケートでは 10 代ごとの離散選択肢であるが、本推計では各区分の中間値を用いて連続的な線形変数として近似的に扱っている）。
- 居住地と学歴は統計的に有意ではなく、限界効果も小さい。これらが投資選択に関係している可能性も事前には想定していたが、機械学習手法とロジスティック回帰分析のいずれでも重要性は小さいと推計された。
- 年収は有意に推計されており、限界効果をみると、無回答層と比較して高年収層は 5.4%ポイント、低年収層は 2.9%ポイント、投資を行う確率が高い。このことから、年収を回答しない層（秘匿する層）ほど相対的に投資を回避する傾向が強いといえる。なお、低年収と無回答を同一カテゴリーとしたモデル 2 においても、高年収層の限界効果は 4.6%ポイントと有意に推計されており、年収階層による投資意欲の差が頑健に確認された。
- 資産残高についても、モデル 1 においては有意に推計されており、限界効果をみると、無回答層と比較して低資産残高層は 5.6%ポイント、高資産残高層は 3.2%ポイント、投資を行う確率が高い。本モデルでは年収等の他要因を制御しているため、資産残高のみが投資に与える純粋な押し上げ効果としては、低資産残高層の方が無回答層との差が相対的に大きく現れる結果となった。高年収層が同時に高資産残高層である傾向が強いため、高資産残高が持つ影響力の多くが年収変数によって説明されているためと考えられる。また、年収と同様に、資産残高を回答しない層（秘匿する層）ほど相対的に投資を回避する傾向が強いといえる。日本のアンケート調査では年収や資産残高を回答しないサンプルが高いことが知られているが、その含意を検証した研究は少ない。ここでは、「分からない」、「答えたくない」の合算ベースであるが、無回答ほど投資を回避する傾向があることが確認された。
- モデル 2 において、年収が有意である一方、資産残高は有意でない。これを、投資意思決定においては資産残高よりも年収が強く影響すると解釈した場合、時間選好率の大きな影響力とあわせて考えると、手元資金の余裕度が重要な要因となっていることが考えられる。また、資産残高が高いからといって投資が選好されるわけではないこと

は、俗に言うタンス預金の大量の存在と整合的に解釈可能である[42]。表 4 のクロス分析では、時間選好率の高さは流動性制約だけでは説明できない点を指摘したが、年収と金融資産残高と時間選好率が投資回避・選択にどのような影響をおよぼしているかは更なる検証の余地がある。

- 国内株式の認知関連では、認知度自体は有意性がない一方で、国内株式に対して何らかのイメージを持つことは投資選択を有意に促進している。無関心層を基準とした場合、ポジティブイメージがある層は 10.3%ポイント、ネガティブイメージがある層であっても 5.5%ポイント、投資を選択する確率が高い。これは、本研究の重要な知見の一つであり、リスク面を意識する慎重な認識を持っていたとしても、無関心でいるよりは投資に対して能動的であることを示している。何らかの認識（関心）を持つこと自体が投資行動への第一歩となる可能性が示唆される。さらに、無関心層との比較では、投資回避の解消には「ネガティブな印象の解消」よりも、「ポジティブな評価の形成」の方が限界効果の大きさという点で約 2 倍（10.3%対 5.5%）大きいことが示唆される。
- 国内株式の投資経験も有意に推計されており、投資経験の有無の限界効果は 8.1%ポイントと推計されている。第 4.6 節のベイジアン・ネットワークでは認知から投資経験への経路が確認されているが、ロジスティック回帰分析では認知度の係数は有意でない。これは、認知度の効果が主に投資経験を介して発現していることを示唆している。
- 暗号資産の認知度と投資経験についても、認知度は有意でなく投資経験は有意となっている。ただし、サンプル全体に占める投資経験者の割合は 4%程度であり、サンプル全体への影響度は小さい。
- 金融リテラシー変数については全てが有意に推計されており、誤答の場合が投資回避的となっている。限界効果をみると、単利が最も大きく 12.6%ポイントとなっている。
- 時間選好率は有意であり、低いほど投資積極的、すなわち、高いほど投資回避的となっている。限界効果をみると、時間選好率が低い場合には、投資を選択する確率は 24.0%ポイントも高く、本研究で用いた説明変数の中でも影響の大きさは最も大きい。これは機械学習の分析と整合的な結果となっている。他の属性要因を制御した上でも、時間選好率と投資回避・選択との間には強い関連が認められる。

5．結論

本研究では、SBI 金融経済研究所が 2023 年に実施したアンケート調査（日本人 10,000 人）の個票データを用い、仮想的な投資商品に対して投資を回避する個人の特性を、カイ二乗検

[42] 1990 年代後半以降、銀行券の発行残高が増大しており、名目値での経済活動とは全く不整合な状態となっている。検証は困難であるが、両者のギャップは巨額のタンス預金が存在していることを示唆している。

定、決定木、ランダム・フォレスト、勾配ブースティング、ベイジアン・ネットワーク、ロジスティック回帰分析の 6 手法を組み合わせて多角的に分析した。本稿が対象とした仮想的投資商品は、最低でも元本が保証される（損失が生じない）設計であるにもかかわらず、日本の回答者の 66%が「買わない」を選択しており、この一見不合理な投資回避行動の背景要因を探ることが本研究の目的であった。本研究は、2024 年から導入された新 NISA（少額投資非課税制度）の直前にあたる 2023 年時点の行動を捉えたものであり、今後の制度定着過程を検証する際に参照されうる一時点のデータとして位置づけられる。分析の結果、以下の知見が得られた。

第一に、投資回避行動を最も強く規定する要因は時間選好率の高さであった。すなわち、将来よりも現在を重視する傾向が強いほど、投資を回避する確率が高い。この結果は、カイ二乗検定における最大の ϕ 係数、決定木における根ノードへの選択、ランダム・フォレストおよび勾配ブースティングにおける突出した特徴量重要度、ロジスティック回帰分析における最大の限界効果（24.0%ポイント）のいずれにおいても一貫して確認された。「1 年後に確実にもらえる金額よりも今すぐ 10 万円を受け取る」と回答した高時間選好率グループでは、投資回避割合が男性で 8 割超、女性で 9 割超に達しており、他の属性による差異を大きく上回っていた。世帯年収や世帯保有金融資産残高とのクロス集計からは、この高時間選好率が流動性制約のみでは説明できないことも確認されており、現在志向性という性向的特性が投資回避に対して独立かつ強力な影響を持つことが示された。ベイジアン・ネットワーク分析では、時間選好率を性向的特性として上流に位置付けた構造の下で、時間選好率から従属変数への直接的な影響経路が確認された。

第二に、金融リテラシーの水準が投資回避行動と強い関連性を有していた。金融リテラシー 4 問（単利、複利、実質金利、ポートフォリオ分散効果）の正誤はいずれも有意な影響を示し、特に最も基本的な知識である単利の理解が最大の効果量を持っていた（ロジスティック回帰分析の限界効果：12.6%ポイント）。単利を理解していない層では投資回避割合が男性 8 割・女性 9 割弱に達しており、金融リテラシーの欠如が投資回避の重要な背景要因となっていることが確認された。ベイジアン・ネットワーク分析では、金融リテラシーから従属変数への直接的な影響経路がみられ、時間選好率とは独立した第二の直接的な関連が確認された。小寺（2020）が指摘した「金融リテラシーがリスク選好を通じてリスク資産保有に至る」という間接的経路と比較すると、本稿では金融リテラシーが投資回避・選択に対して直接的な経路を有しつつ、同時に国内株式の認知から金融リテラシーへの影響経路も確認され、両者が複合的に作用する構造が観察された点が特徴的である。

第三に、国内株式に対するポジティブイメージの有無が、金融リテラシーと同程度かそれ以上の影響力を持つことが明らかになった。決定木分析では、低時間選好率グループにおいてポジティブイメージの有無が第 2 階層の分岐変数として選択され、ポジティブイメージがあるグループでは「買う」の構成比が 85〜89%に達した。ロジスティック回帰分析におけ

るポジティブ認識の限界効果は 10.3%ポイントであり、単利の理解（12.6%ポイント）に次ぐ大きさであった。注目すべきは、ポジティブイメージがある層だけではなく、ネガティブイメージがある層であっても、無関心層に比べれば 5.5%ポイント投資確率が高いことが示された点である。これは、リスク面を意識する慎重な認識を持っていたとしても、無関心でいるよりは投資に対して能動的であることを示しており、何らかの認識（関心）を持つこと自体が投資行動への第一歩となる可能性を示唆している。さらに、無関心層と比較した場合に、ポジティブ認識の形成はネガティブな印象の解消と比較して投資確率を約 2 倍大きく押し上げる効果があり、投資回避の解消には投資に対する否定的認識の軽減よりも肯定的認識の形成がより重要であることが示唆された。ベイジアン・ネットワーク分析では、内株認知度から投資経験を経てポジティブ認識に至るプロセス的な段階構造も確認された一方、ポジティブ認識から投資行動への方向性は一意に定まらなかった。もっとも、ロジスティック回帰分析の結果と合わせると、認知・経験・認識の連鎖的経路の存在が強く示唆される。

第四に、性別、年齢、学歴、居住地といった基本的な人口統計学的属性の直接的影響力は、時間選好率や金融リテラシーと比較して相対的に小さいことが確認された。特に、クロス集計分析では男女間に明確な投資回避格差（男性 54%対女性 72%）が観察されていたにもかかわらず、機械学習手法による多変量分析では性別の特徴量重要度は低い水準にとどまった。これは、性別が真の要因ではなく、時間選好率や金融リテラシーにおける性差が交絡効果として性別を重要変数のように見せている可能性を示唆しているが、ロジスティック回帰分析では他変数をコントロールしても性別は有意であり（限界効果 7.2%ポイント）、性別固有の要因の存在も否定はできない。もっとも、時間選好率（24.0%ポイント）や単利の理解（12.6%ポイント）と比較すれば性差の影響は限定的であった。同様に、年齢の限界効果も 10 歳あたり −3%ポイントとさほど大きくはなく、学歴・居住地は統計的に有意ではなかった。モデル 2 では、年収は有意に推計された一方、資産残高は有意ではなかった。これらの結果は、投資回避行動が「財務余力がないから」、「高齢だから」、「女性だから」といった属性的説明よりも、「時間選好率が高いから」、「金融リテラシーが低いから」、「株式投資にポジティブイメージがないから」といった行動経済学的・認知的要因によってより強く説明されることを意味している。

以上の知見は、株式市場参加パズルの政策的対応に関して重要な含意を持つ。第一に、金融リテラシー教育の推進は引き続き有効な政策手段であるが、本稿の分析からは、単なる知識の伝達にとどまらず、「無関心層」を関心層へと引き上げ、さらに投資に対するポジティブイメージの形成を促す施策が重要であることが示された。具体的には、投資の成功体験の共有やリスクとリターンの適切な理解を促進するプログラムの設計が求められる。第二に、時間選好率の高さが最大の阻害要因であるという発見は、従来の金融教育中心の政策アプローチの限界を示唆している。時間選好率は性向的特性として短期的な介入による変化が困難である可能性が高いが、その決定要因（将来不安、社会経済環境、認知能力等）の解明は、

効果的な介入策の設計において重要な研究課題となる。新 NISA のような制度的ナッジや少額からの積立投資の普及といった施策は、高い時間選好率を直接変化させるものではないが、投資開始時の心理的障壁を低下させることで、間接的にこの問題を緩和する可能性がある。第三に、年収や資産残高そのもの以上に、これらの情報を回答しないこと自体が投資回避と関連している点は、自身の経済状況の把握や開示に消極的な層が投資からも遠ざかっている可能性を示しており、家計の経済状況の「見える化」を促す施策の意義を示唆している。

本研究にはいくつかの限界と今後の検討課題があり、これを最後に指摘しておく。第一に、本稿で用いた仮想的な投資商品への回答と実際の株式市場参加行動の間には乖離が存在しうる。仮想バイアスの影響を完全に排除することはできず、実際の投資行動を対象とした検証が今後求められる。第二に、横断データによる分析であるため、変数間の時間的先後関係を直接識別することはできない。そのため、本研究におけるベイジアン・ネットワーク分析は、他の分析手法による結果を補完する目的で、因果関係の確定ではなく、変数間の条件付き依存構造の把握に焦点をあてている。この枠組みの下で、時間選好率については仮定的に性向的特性として上流に位置付けて分析を行っている。パネルデータを用いた動学的分析や、RCT による因果効果の検証が今後の課題となる。SBI 金融経済研究所のアンケート調査では 2025 年に、同じ質問を用いて投資を回避した回答者に対する RCT を行っている（SBI 金融経済研究所, 2025）。この個票データが手掛かりとなろう。第三に、時間選好率の決定要因の分析が十分ではない。本稿では時間選好率を所与の性向的特性として扱ったが、その背景にある要因（将来不安、認知能力、文化的環境等）の解明は、政策的介入の設計において不可欠であり、今後の重要な研究課題である。第四に、本稿の分析は日本の回答者に限定されているが、表 1 で示したように投資回避現象は国際的に観察されている。各国データを用いた比較分析により、文化的・制度的要因の影響を識別することが可能となろう。

Appendix　質問項目に対する統計処理

質問項目	統計処理	変数名
Q1 性別	男性を 1、女性を 0 とした。	性別
Q2 年齢	20～49 歳、50 歳以上に分類の上、50 歳以上を 1、20～49 歳を 0 とした。区分については、総務省統計局・国勢調査の平均年齢（令和 2 年：47.7 歳）を参考にした。	年齢
	20～29 歳を 24.5（中間値）、30～39 歳を 34.5、40～49 歳を 44.5、50～59 歳を 54.5、60～69 歳を 64.5 とした上で、男性は 70 歳以上を 77.825 歳、女性は 70 歳以上を 79.980 歳とした。70 歳以上の年齢については、厚生労働省・簡易生命表の 70 歳の平均余命（令和 5 年：男性が 15.65、女性が 19.96）の 2 分の 1 を 70 歳に加算した（計算誤差を避けるため端数処理を行わずそのまま用いた）。	年齢_L(中間値)
Q3 居住地	政令指定都市（東京都 23 区を含む）、府県庁所在地を 1、市部（政令指定都市、府県庁所在地を除く）、町村部、その他を 0 とした。	居住地
Q4 学歴	短大・高専卒、大学卒、大学院卒を 1、中学卒、高校卒、専門学校卒を 0 とした。	学歴
Q7 世帯年収	無収入を含む 600 万円未満（低）、600 万円以上（高）、分からない・答えたくない（NR）に分類の上、それぞれ、該当する場合を 1、該当しない場合を 0 とする 3 つの説明変数を作成した。区分については、厚生労働省・国民生活基礎調査の世帯年収の平均値（令和 5 年：524.2 万円）を参考にした。	年収・低、 年収・高、 年収・NR
	分からない・答えたくない（無回答）を基準カテゴリーとし、600 万円以上（高）、無収入を含む 600 万円未満（低）の 2 つの説明変数を作成した。年収・高_L(ダミー)では、600 万円以上に該当する場合を 1、該当しない場合を 0 とし、年収・低_L(ダミー)では、無収入を含む 600 万円未満に該当する場合を 1、該当しない場合を 0 とした。	年収・高_L(ダミー)、 年収・低_L(ダミー)
Q8 世帯保有金融資産残高	100 万円未満～2,000 万円未満（低）、2,000 万円以上（高）、分からない・答えたくない（NR）に分類の上、それぞれ、該当する場合を 1、該当しない場合を 0 とする 3 つの説明変数を作成した。区分については、金融広報中央委員会・家計の金融行動に関する世論調査の金融資産保有額の平均値（令和 5 年：1,758 万円）を参考にした。	資産残高・低、 資産残高・高、 資産残高・NR
	分からない・答えたくない（無回答）を基準カテゴリーとし、2,000 万円以上（高）、100 万円未満～2,000 万円未満（低）の 2 つの説明変数を作成した。資産残高・高_L(ダミー)では、2,000 万円以上に該当する場合を 1、該当しない場合を 0 とし、資産残高・低_L(ダミー)では、100 万円未満～2,000 万円未満に該当する場合を 1、該当しない場合を 0 とした。	資産残高・高_L(ダミー)、 資産残高・低_L(ダミー)
Q10.1 国内株式の認知度	他人に教えられるほどではないが、ある程度の知識はある、他人に教えられるほど、詳しく知っているを 1、知らない/聞いたことがない、聞いたことはあるが、あまり知らないを 0 とした。	内株認知度

Q11.1 国内株式の投資経験	投資したことはあるが、現在は保有していない、現在保有しているを（投資経験ありとして）1、投資したことはないを 0 とした。	内株投資経験
Q12 国内株式に対する認識	回答選択肢を以下のとおり、ポジティブイメージ (P)、ネガティブイメージ（N)、無関心（I）に分類の上、それぞれ、そうした認識がある場合を 1、ない場合を 0 とする 3 つの説明変数を作成した。 ポジティブイメージ群：利益が期待できる、投資対象を値動きの異なる金融商品に分散することで、投資のリスクを低減する効果がある、利便性が高い、商品に対する好奇心がある。 ネガティブイメージ群：損失が生じることへの不安がある、商品の仕組みやリスクなどが理解できない、投資方法／利用方法が難解、使いこなせないことへの不安がある、投資家保護・消費者保護が十分でないおそれがある、発行体や仲介業者が信用できない、個人情報の漏洩が不安、本人認証の悪用・なりすまし詐欺等への不安がある、税制上不利な扱いをされる、規制が厳しくて使いにくい、その他。 無関心：特に印象はない。	内株認識・P、 内株認識・N、 内株認識・I
	無関心（I）を基準カテゴリーとし、ポジティブイメージ（P)、ネガティブイメージ（N）の 2 つの説明変数を作成した。内株認識・P_L(ダミー)では、ポジティブイメージがある場合を 1、それ以外を 0 とし、内株認識・N_L(ダミー)では、ネガティブイメージがある場合を 1、それ以外を 0 とした。	内株認識・P_L(ダミー)、 内株認識・N_L(ダミー)
Q18.1 暗号資産の認知度	他人に教えられるほどではないが、ある程度の知識はある、他人に教えられるほど、詳しく知っているを 1、知らない/聞いたことがない、聞いたことはあるが、あまり知らないを 0 とした。	暗号認知度
Q19.1 暗号資産の投資経験	投資したことはあるが、現在は保有していない、現在保有しているを（投資経験ありとして）1、投資したことはないを 0 とした。	暗号投資経験
Q38～Q41 金融リテラシー	各質問について、正答を 1、誤答を 0 とした。	FL・単利、 FL・複利、 FL・実質金利、 FL・分散効果
	Q38～41 のうち 3 問以上正答を 1、3 問未満正答を 0 とする説明変数を作成した。	FL・全体
Q42 リスク回避・選好度	いずれかの価格選択肢で買うを 1、買わないを 0 とした。	従属変数
Q45 時間選好率	「1 年後にいくらもらえるとしても、今すぐにもらえる」方を選ぶ（時間選好率が高い）を 1、それ以外（100,500 円～150,000 円以上のいずれかで 1 年後にもらうを選択）を 0 とした。	時間選好率
	1 年後にもらうを 1、「1 年後にいくらもらえるとしても、今すぐにもらえる」方を選ぶを 0 とした。	時間選好率_L(反転)

注）各質問項目において「不明」との回答はなかったため、欠損データ処理は行っていない。

参考文献

Acharya, V. V., & Pedersen, L. H. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics, 77*(2), 375–410.

Adams, S. W., Kasten, C., & Kelley, E. K. (2024). How free is free? Retail trading costs with zero commissions. *Journal of Banking & Finance, 165*, 107226.

Alan, S. (2006). Entry costs and stock market participation over the life cycle. *Review of Economic Dynamics, 9*(4), 588–611.

Alzuabi, R., & Gray, D. (2025). Household portfolio allocation and stock market beliefs: Evidence from Japanese households. *Journal of Financial Research.* Advance online publication. https://doi.org/10.1111/jfir.70028

Antonides, G., de Groot, M., & van Raaij, W. F. (2011). Mental budgeting and the management of household finance. *Journal of Economic Psychology, 32*(4), 546–555.

Badarinza, C., Campbell, J. Y., & Ramadorai, T. (2016). International comparative household finance. *Annual Review of Economics, 8*, 111–144.

Barberis, N., Huang, M., & Santos, T. (2001). Prospect theory and asset prices. *The Quarterly Journal of Economics, 116*(1), 1–53.

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics, 121*(3), 823–866.

Bekaert, G., & Harvey, C. R. (1995). Time-varying world market integration. *The Journal of Finance, 50*(2), 403–444.

Benartzi, S., & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics, 110*(1), 73–92.

Bernanke, B. S., & Kuttner, K. N. (2005). What explains the stock market's reaction to Federal Reserve policy? *The Journal of Finance, 60*(3), 1221–1257.

Bilias, Y., Georgarakos, D., & Haliassos, M. (2010). Portfolio inertia and stock market fluctuations. *Journal of Money, Credit and Banking, 42*(4), 715–742.

Brown, J. R., Ivković, Z., Smith, P. A., & Weisbenner, S. (2008). Neighbors matter: Causal community effects and stock market participation. *The Journal of Finance, 63*(3), 1509–1531.

Bucher-Koenen, T., Lusardi, A., Alessie, R., & van Rooij, M. (2017). How financially literate are women? An overview and new insights. *Journal of Consumer Affairs, 51*(2), 255–283.

Bucher-Koenen, T., Alessie, R., Lusardi, A., & van Rooij, M. (2025). Fearless woman: Financial literacy, confidence, and stock market participation. *Management Science, 71*(9), 7414–7430.

Calvet, L. E., Campbell, J. Y., & Sodini, P. (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy, 115*(5), 707–747.

Calvet, L. E., Campbell, J. Y., & Sodini, P. (2009). Fight or flight? Portfolio rebalancing by individual investors. *The Quarterly Journal of Economics, 124*(1), 301–348.

Campbell, J. Y., & Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy, 107*(2), 205–251.

Campbell, J. Y., & Ramadorai, T. (2026). Household finance in retrospect and prospect. NBER Working Paper No. 34621.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Dominitz, J., & Manski, C. F. (2007). Expected equity returns and portfolio choice: Evidence from the Health and Retirement Study. *Journal of the European Economic Association, 5*(2–3), 369–379.

Duraj, K., Grunow, D., Haliassos, M., Laudenbach, C., & Siegel, S. (2024). Rethinking the stock market participation puzzle: A qualitative approach. IMFS Working Paper Series, No. 210.

European Central Bank. (2023). Household Finance and Consumption Survey.

Federal Reserve Board. (2022). Survey of Consumer Finances.

Financial Conduct Authority. (2022). Financial Lives Survey.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature, 40*(2), 351–401.

Fujiki, H., Hirakata, N., & Shioji, E. (2012). Aging and household stockholdings: Evidence from Japanese household survey data. IMES Discussion Paper Series, 2012-E-17.

Gomes, F., & Michaelides, A. (2005). Optimal life-cycle asset allocation: Understanding the empirical evidence. *The Journal of Finance, 60*(2), 869–904.

Gomes, F., Haliassos, M., & Ramadorai, T. (2021). Household finance. *Journal of Economic Literature, 59*(3), 919–1000.

Guiso, L., Sapienza, P., & Zingales, L. (2003). People's opium? Religion and economic attitudes. *Journal of Monetary Economics, 50*(1), 225–282.

Guiso, L., Sapienza, P., & Zingales, L. (2004). The role of social capital in financial development. *American Economic Review (Papers and Proceedings), 94*(2), 526–556.

Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *The Journal of Finance, 63*(6), 2557–2600.

Guiso, L., Sapienza, P., & Zingales, L. (2018). Time varying risk aversion. *Journal of Financial Economics, 128*(3), 403–421.

Guo, X., Huang, Y., Sun, Q., & Yeung, B. (2025). Digital transmission of financial knowledge: Evidence from stock market investment. CEPR Discussion Paper No. 20651.

Haliassos, M., & Bertaut, C. C. (1995). Why do so few hold stocks? *The Economic Journal, 105*(432), 1110–1129.

Hanaoka, C., Shigeoka, H., & Watanabe, Y. (2018). Do risk preferences change? Evidence from the Great East Japan Earthquake. *American Economic Journal: Applied Economics, 10*(2), 298–330.

Heaton, J., & Lucas, D. (1996). Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy, 104*(3), 443–487.

Heaton, J., & Lucas, D. J. (2000). Portfolio choice and asset prices: The importance of entrepreneurial risk. *The Journal of Finance, 55*(3), 1163–1198.

Hong, H., Kubik, J. D., & Stein, J. C. (2004). Social interaction and stock-market participation. *The Journal of Finance, 59*(1), 137–163.

Hurd, M. D., van Rooij, M., & Winter, J. (2011). Stock market expectations of Dutch households. *Journal of Applied Econometrics, 26*(3), 416–436.

Iwaisako, T., Ono, A., Saito, A., & Tokuda, H. (2022). Disentangling the effect of home ownership on household stockholdings: Evidence from Japanese micro data. *Real Estate Economics, 50*(1), 268–295.

Julliard, C., & Ghosh, A. (2012). Can rare events explain the equity premium puzzle? *The Review of Financial Studies, 25*(10), 3037–3076.

Kadoya, Y., Khan, M. S. R., & Hamada, T. (2017). Does financial literacy affect stock market participation? *Journal of Behavioral and Experimental Finance, 14*, 1–10.

Katauke, T., Khan, M. S. R., & Kadoya, Y. (2025). What determines investment in the Nippon Individual Savings Account? An investigation of Japan's tax-exempt investment account. *PLOS ONE, 20*(2), e0313433.

Kaustia, M., & Knüpfer, S. (2008). Do investors overweight personal experience? Evidence from IPO subscriptions. *The Journal of Finance, 63*(6), 2679–2702.

Kézdi, G., & Willis, R. J. (2011). Household stock market beliefs and learning. NBER Working Paper No. 17614.

Li, G. (2014). Information sharing and stock market participation: Evidence from extended families. *The Review of Economics and Statistics, 96*(1), 151–160.

Liu, C.-W., Mithas, S., Pan, Y., & Hsieh, J. J. P.-A. (2024). Mobile apps, trading behaviors, and portfolio performance: Evidence from a quasi-experiment in China. *Information Systems Research, 36*(2), 828–846.

Lusardi, A., & Mitchell, O. S. (2011). Financial literacy and retirement planning in the United States. *Journal of Pension Economics & Finance, 10*(4), 497–514.

Luttmer, E. G. (1999). What level of fixed costs can reconcile consumption and stock returns? *Journal of Political Economy, 107*(5), 969–997.

Maki, A., & Sonoda, T. (2002). A solution to the equity premium and risk-free rate puzzles: An empirical investigation using Japanese data. *Applied Financial Economics, 12*(8), 601–612.

Malmendier, U., & Nagel, S. (2011). Depression babies: Do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics, 126*(1), 373–416.

Mankiw, N. G. (1986). The equity premium and the concentration of aggregate shocks. *Journal of Financial Economics, 17*(1), 211–219.

Mankiw, N. G., & Zeldes, S. P. (1991). The consumption of stockholders and nonstockholders. *Journal of Financial Economics, 29*(1), 97–112.

McFadden, D. (1979). Quantitative methods for analysing travel behaviour of individuals: Some recent developments. In D. A. Hensher & P. R. Stopher (Eds.), *Behavioural travel modelling* (pp. 279–318). Croom Helm.

McGrattan, E. R., & Prescott, E. C. (2003). Average debt and equity returns: Puzzling? *American Economic Review, 93*(2), 392–397.

Mehra, R. (2003). The equity premium: Why is it a puzzle? *Financial Analysts Journal, 59*(1), 54–69.

Mehra, R., & Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of Monetary Economics, 15*(2), 145–161.

Mulligan, C. B., & Sala-i-Martin, X. (2000). Extensive margins and the demand for money at low interest rates. *Journal of Political Economy, 108*(5), 961–991.

Parakhoniak, A., Balakina, O., & Bäckman, C. (2026). Beyond connectivity: Stock market participation in a network. *Journal of Economic Dynamics and Control, 183*, 105242.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.

Polkovnichenko, V. (2004). Limited stock market participation and the equity premium. *Finance Research Letters, 1*(1), 24–34.

Reher, M., & Sokolinski, S. (2024). Robo advisors and access to wealth management. *Journal of Financial Economics, 155*, 103829.

Rietz, T. A. (1988). The equity risk premium: A solution. *Journal of Monetary Economics, 22*(1), 117–131.

Securities and Exchange Board of India. (2025). Investor Survey 2025.

Securities and Investment Management Association (SIMA). (2025). Survey of Canadian investors.

van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics, 101*(2), 449–472.

Viceira, L. M. (2001). Optimal portfolio choice for long-horizon investors with nontradable labor income. *The Journal of Finance, 56*(2), 433–470.

Vissing-Jørgensen, A. (2002). Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy, 110*(4), 825–853.

Yamori, N., & Ueyama, T. (2022). Financial literacy and low stock market participation of Japanese households. *Finance Research Letters, 47*, 102797.

伊藤雄一郎・瀧塚寧孝・藤原茂章（2019）. 家計の資産選択行動―動学的パネル分析を用いた資産選択メカニズムの検証―. 金融研究, 38(3), 43–81.

祝迫得夫（2018）. 家計による情報取得と金融資産投資. トラスト未来フォーラム(編), 『人口減少・高齢化・少子化の同時進行の下での日本経済と金融』(トラスト未来フォーラム研究叢書).

顔菊馨・白須洋子・近藤隆則・三隅隆司（2019）. 日本の個人投資家のリスク資産投資：金融リテラシーの種類や情報源の違いはどのような影響を与えるのか？. 経営財務研究, 39(1,2), 86–103.

SBI 金融経済研究所. (2022). 「次世代金融に関する一般消費者の関心や利用度に関するアンケート調査」結果.

SBI 金融経済研究所. (2024a). 「次世代金融に関する一般消費者の関心や利用度に関するアンケート調査、第 2 回」の結果.

SBI 金融経済研究所. (2024b). 「次世代金融アンケート 2024」の結果.

SBI 金融経済研究所. (2025). 「次世代金融アンケート 2025」の結果.

小川雄太郎（2020）.『つくりながら学ぶ！Python による因果分析―因果推論・因果探索の実践入門』マイナビ出版.

小寺信也（2020）. 家計のリスク性資産保有の決定要因―機械学習による因果構造の推定. みずほリポート. みずほ総合研究所.

金澤伸浩・田中豊・小山浩一・内藤博敬・伊川美保・中山由美子（2020）. リスク教育のためのリスクリテラシー測定尺度. 日本リスク学会誌, 29(4), 243–249.

坂本俊之（2019）.『作ってわかる！アンサンブル学習アルゴリズム入門』C&R 研究所.

塩路悦朗・平形尚久・藤木裕（2013）. 家計の危険資産保有の決定要因について：逐次クロスセクション・データを用いた分析. 金融研究, 32(2), 63–103.

末廣徹（2022）. メンタル・バジェッティングの傾向は個人の金融資産残高の蓄積を促すのか. 生活経済学研究, 56, 33–48.

関田静香（2020）. 国民の資産形成と金融リテラシー. フィナンシャル・レビュー, 2020(1), 23–41.

竹村敏彦・神津多可思・武田浩一・末廣徹（2024）. 行動ファイナンスの視点を踏まえた個人のリスク資産保有比率の決定要因に関する実証分析～2023 年度金融行動調査の結果から～. 経済経営紀要, (47), 31–48.

日本銀行（2017）. 企業のインフレ予想形成に関する新事実：Part II ―機械学習アプローチ―. 日本銀行ワーキングペーパーシリーズ, 17-J-4.

藤木裕（2019）. 家計の金融知識と金融資産選択:「金融リテラシー調査」による実証研究. TCER Working Paper, J-17.

毛利拓也（2023）.『LightGBM 予測モデル実装ハンドブック』秀和システム.

山沖義和・副島豊（2025）. 暗号資産・株式等の投資決定要因：6 ヶ国検証－SBI 金融経済研究所アンケート調査（第 2 回）の個票分析. SBI-FERI ワーキングペーパーシリーズ, WP J-1.

馮譞・チュアイシリ・木成勇介（2017）. 金融資産選択における行動経済学的要因の影響. 個人金融, 12(2), 10–16.