

# みずほレポート

2020年10月20日

## 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

- ◆本稿では、慶應義塾大学「日本家計パネル調査」の個票データを利用し、日本の家計におけるリスク性資産の保有に対して、どのような要因が影響しているのか、機械学習の手法を活用することで、因果構造の推定を試みた。
- ◆分析結果からは、リスク性資産の保有に対して大きく2つの経路があることが確認された。1つ目の経路は、老後を見据えた安全資産（預貯金）に余剰分がある場合において、リスク性資産を保有するものである。預貯金には、所得、負債（住宅ローンなど）、年齢（ライフ・サイクル）が影響を与えており、これらの要因はリスク性資産の保有と間接的に関係している。2つ目は、家計のリスク選好が高い場合に、リスク性資産を保有する経路である。この選好の高さには、主に金融リテラシーの有無が影響している可能性が指摘された。
- ◆リスク性資産の保有率が低い主因は、老後の資金的な不安であると推察される。少子高齢化が進行する日本経済では、政策努力はすべきだが、その不安を完全に払拭することは困難であるため、2つ目の経路である金融リテラシーを高めることが重要になる。日本の金融リテラシーは国際的にみても低く、改善余地が大きいため、政府や金融機関などによる積極的な金融教育の推進が必要である。

---

経済調査部 主任エコノミスト 小寺信也  
03-3591-1435 shinya.kotera@mizuho-ri.co.jp

---

●当レポートは情報提供のみを目的として作成されたものであり、取引の勧誘を目的としたものではありません。本資料は、当社が信頼できると判断した各種データに基づき作成されておりますが、その正確性、確実性を保証するものではありません。本資料のご利用に際しては、ご自身の判断にてなされますようお願い申し上げます。また、本資料に記載された内容は予告なしに変更されることもあります。なお、当社は本情報を無償でのみ提供しております。当社からの無償の情報提供をお望みにならない場合には、配信停止を希望する旨をお知らせ願います。

---

## 目 次

---

1 : はじめに	1
2 : 先行研究	3
3 : データ	4
4 : 分析手法	7
(1) ランダム・フォレスト	7
(2) ベイジアン・ネットワーク	8
5 : 分析結果	10
(1) ランダム・フォレスト	10
(2) ベイジアン・ネットワーク	12
(3) 介入効果	16
(4) 頑健性	18
6 : まとめ	20
参考文献	22
補論 1 : PC アルゴリズムについて	24
補論 2 : ランダム・フォレストによる相対的重要度	25
補論 3 : ベイジアン・ネットワーク (PC アルゴリズム)	27
補論 4 : 条件付き確率 (事後分布)	27
補論 5 : 頑健性チェック	28

## 1 : はじめに

本稿は慶應義塾大学「日本家計パネル調査」の個票データを利用し、日本の家計におけるリスク性金融資産の保有には、どのような要因(変数)が影響しているのかについての実証分析を行う<sup>1</sup>。特に、本稿では機械学習の手法を活用することで、変数間の因果構造を推定し、どのような政策介入がリスク性資産保有率の向上に向けて効果的かについてシミュレーションを行うことを目的とする。

諸外国と比較した日本の家計の金融資産保有の特徴として、現預金の割合が高く、リスク性資産の保有割合が少ないことが指摘される。2020年3月末時点における資金循環統計による国際比較では、金融資産に占めるリスク性資産の割合は、日本が13.0%であるのに対し、米国は44.8%、欧州は25.9%となっており<sup>2</sup>、米国の割合は日本の3倍以上となっている。資金循環統計を用いた比較は定義などの違いもあって実勢より差が大きくなる傾向にあるが、それらを調整しても日米間の差は依然として2倍程度存在することが指摘されている(福原、2016)。また、総務省「全国消費実態調査」によると、2014年では有価証券を保有する世帯の割合(保有率)は、総世帯の23.5%にとどまっており、リスク性資産を保有している家計の方が少数派となっている。

こうした家計の低調なリスク性資産への投資は、政府も問題視している。金融庁(2017)は、米国や英国と比較して日本の家計金融資産の伸びが低い背景に、リスク性資産の割合が低く、運用リターンによる家計の資産増が限定的であることを示し、家計の金融資産増加や健全な成長資金の供給のためには、「現預金(貯蓄)から投資」への移行を促進することが重要であると指摘している。この目的のため、iDeCo(個人型確定拠出年金)の加入対象者の拡大、つみたて(積立)NISAの創設、「雇用本位の業務運営に関する原則」の公表など、これまでも政府は様々な取組を行っているが、上記のデータが示す通り、リスク性資産を保有する世帯が少ない状況に大きな変化はない。

家計のリスク性資産への投資を促進し、資金の好循環につなげるためには、そもそもの家計におけるリスク性資産保有の構造メカニズムを把握することが重要となる。Merton(1969)などの古典的な理論によると、リスク性資産の期待収益率が安全資産の利子率を上回る限り、いかに家計がリスク回避的であったとしても、リスク性資産を保有することが望ましい。しかし、実際には大多数の家計はリスク性資産を保有しておらず、このモデルでは現実を適切に説明しきれていない。理論モデルは簡素化された一定の仮定の下で成立するものであるため、現実との前提条件が異なれば、実際に起きている現象を説明できないのは当然である。

そこで、本稿では先験的な制約条件をできるだけ少なくし、データを機械に学習させることで、リスク性資産保有の有無に影響を及ぼしている要因を特定することを試みる。実際に観測されたデータからリスク性資産保有行動の「パターン」を機械に学習させることで、より現実に即したモデルの構築・要因の特定が可能になることが期待される。また、影響を及ぼす可能性のある変数には、これまでの実証研究でリスク性資産との関係が指摘された要因を可能な限り多く利用することを試みる。リ

<sup>1</sup> 本稿の分析に際しては、慶應義塾大学パネルデータ設計・解析センターによる「日本家計パネル調査(JHPS/KHPS)」の個票データの提供を受けた。記して感謝したい。もちろん、あり得べき誤りはすべて筆者の責任である。

<sup>2</sup> 日本銀行(2020)を参照。リスク性資産=株式+出資金+投資信託、と定義した。

リスク性資産に直接的（または間接的）に影響を与えている変数が特定できれば、政府や企業が介入可能な変数を変化させることで、どの介入がリスク性資産保有率の上昇に対して効果的かを試算することが可能となる。この因果構造の解明と介入効果の試算が本稿の最終目的となる。

具体的な機械学習の手法は、日本銀行（2017）を参考に2つのステップで行う。まず、最初のステップでは、ランダム・フォレストと呼ばれる手法を利用し、各変数の「重要度」を算出することを目的とする。この重要度とは、どの変数を利用するとリスク性資産の保有の有無をより正確に予測できるかを示したものであり、次のステップで行う因果構造の学習の際に、どの変数を利用すればよいか、当たりをつけることができる。なお、多くの説明変数を利用する場合、一般的な重回帰分析では多重共線性が問題となるが、ランダム・フォレストの手法を利用することで、こうした問題を回避しやすくなる。

ただし、ランダム・フォレストにより発見されるのは、あくまでも変数間のパターンであり、変数間の関係性（因果構造）は特定されない。そのため、次のステップでは、ベイジアン・ネットワークと呼ばれる機械学習の手法を利用し、変数間のネットワーク構造のグラフを推計する。これにより、どの変数がどのような経路を通してリスク性資産保有の有無に対して影響を与えているのかを推定することが可能になる。

なお、リスク性資産の分析は、家計がリスク性資産を保有するか否か、リスク性資産を保有している家計がどの程度のリスク性資産を保有するか、に分割して分析されるケースが多い（伊藤他（2017）、塩路他（2013）など）。後者の論点も非常に重要な点であるが、本稿が対象とする分析は前者のみである点には留意されたい。

分析の結果は、以下の通りである。リスク性資産の保有に対しては大きく2つの経路が影響を与えている。1つ目の経路は、老後を見据えた預貯金に余剰がある（と思う）場合に、リスク性資産を保有するものである。預貯金に対しては、所得水準、住宅ローンなどの負債、年齢（ライフ・サイクル）が影響を与えており、これらの要因はリスク性資産の保有と間接的に関係している。2つ目は、家計のリスクに対する選好が高い場合に、リスク性資産保有率が上昇する経路である。この選好の高さには、金融リテラシーの有無が影響している可能性が指摘された。

日本の家計によるリスク性資産への投資が少ない主因は、老後の資金的な不安である可能性が高い。ただし、少子高齢化が進行する日本経済では、政策努力はすべきだが、その不安を完全に払拭することは困難であるため、2つ目の経路である金融リテラシーを向上させることでリスク性資産に対する選好を高めることが重要となる。日本の金融リテラシーは国際的にみても低く改善余地が大きいいため、政府や金融機関などによる積極的な金融教育の推進が必要となる。

本稿の構成は以下の通りである。まず、2節では、先行研究の整理を行う。実証分析を中心に整理することで、本稿の分析に利用する変数の候補を探る。3節では、本稿で利用するデータの特性を概観した上で、2節を踏まえた利用する変数の一覧を提示する。4節では、本稿で利用する2つの機械学習の分析手法について概要を述べる。5節が分析結果であり、因果構造の推計、介入効果、頑健性チェックを行う。最後の6節で全体をまとめ、今後の課題を提示する。

## 2：先行研究

家計におけるリスク性資産の保有に関する理論的な背景は、祝迫（2012）、塩路他（2013）、伊藤他（2017）などで整理されているが、一般的な効用関数の仮定の下では、「リスク性資産の期待収益率 $>$ 安全資産の利子率」との関係性が成立する限り、保有量の違いはあるにせよ、必ずリスク性資産を保有することが望ましい。しかし、現実にはリスク性資産を保有しない世帯も多いため、こうした理論モデルと現実の乖離をもたらす要因として、労働所得、不動産、取引コストなどの存在が指摘されている。労働所得は、将来の所得に不確実性が存在する場合、予備的貯蓄動機の影響により、安全資産が選好されることの要因となる。所得に対する不確実性は、職業・就業形態などによって異なるため、それらが資産選択に影響を及ぼすことも想定される。また、不動産（住宅）は、流動性が低い資産であることに加え、高額な住宅ローンは家計にとって固定的な支払い義務となることから、リスク性資産の保有を抑制する方向に働く。最後の取引コストは、幅広い意味でのコストを意味しており、取引手数料などの金銭的なコストだけでなく、金融知識を得るためのコストや情報収集のコストなども含んでいる。

上記のような理論的背景を踏まえ、日本のデータを利用した家計の資産選択に関する実証研究を概観する。まず、本稿と同様に包括的な観点から家計のリスク性資産について分析した研究に伊藤他（2017）が挙げられる。彼らは大阪大学のパネルデータを利用した実証分析を行い、リスク性資産の保有の有無には、古典的理論ファクター<sup>3</sup>、流動性制約、予備的貯蓄動機（失業・将来不安）、参加コスト（金融知識など）の要因が有意な影響を及ぼしていることを報告している。彼らはリスク性資産の保有確率の国差の背景についても考察し、金融知識や構造的要因（制度面など）が両国の乖離を生んでいる主な要因である可能性を指摘している。

塩路他（2013）も多様な要因の検討を行っている。彼らは、逐次クロスセクション・データによる実証分析を行い、株式保有の有無に対して、特に頑健性の高い要因が金融資産残高であることを指摘している。その他には、株式投資を行う際の情報に関する要因（学歴、ネット取引の熟練度など）や、就業形態（自営業）などの要因も有意性が観察されたが、年齢効果の係数は時系列的に安定しなかったことを報告している。

北村・内野（2011）は、逐次クロスセクション・データを用いて、大卒家計と非大卒家計間のリスク性資産保有の差が、どの程度属性要因で説明可能かを検証した。両者の差の3～5割は属性の差（特に勤め先規模、所得水準、金融資産水準）で説明可能であることが判明したが、依然として説明されない部分が多いため、参加コスト（金融知識など）の存在が示唆されると結論付けている<sup>4</sup>。

木成・筒井（2009）は、郵政総合研究所の調査データを用いて、危険資産比率（0から100まで）を被説明変数としたトービット推計を行い、資本資産価格モデル（CAPM）が想定する要因（リスク許容

<sup>3</sup> リスク性資産の期待収益率、安全資産利子率、市場ボラティリティ、相対的リスク回避度から構成される理論上の最適リスク性資産比率。

<sup>4</sup> 日本の分析事例ではないが、金融知識の観点からは、金融リテラシーと株式保有の関連性を指摘した van Rooij et al. (2011) などの研究が挙げられる。

度、期待収益率)に加え、金融資産額、所得、年齢、銀行・証券会社への信頼、行動バイアス(自信過剰)などの変数も理論と整合的な形で有意に影響を与えていることを報告している<sup>5</sup>。

また、特定の要因に着目してリスク性資産との関係を分析したものに、Iwaisako et al. (2019)が挙げられる。彼らは、家計の資産の大部分を占める持ち家(不動産)が株式保有比率を低下させている可能性を日本のデータを利用して検証し、住宅価額の上昇は株式保有比率を増加させる一方、住宅ローンの増加が株式保有比率を低下させる関係にあることを示している<sup>6</sup>。

日本の各種データを整理することで、リスク性資産投資の背景を考察したものに中城他(2017)や福原(2016)がある。中城他(2017)は、リスク性資産の保有を妨げる要因として、日本の株式パフォーマンスが低調であること、老後の生活が具体的にイメージできないなど生活設計が不十分であることを指摘している。また、福原(2016)は、日米間でリスク性資産保有量が異なる背景として、金融資産保有額の世帯間の格差、不動産所有による流動性制約、確定拠出年金や投資信託を含む制度要因などの存在を指摘している。

以上の実証分析を整理すると、リスク性資産保有に関係する可能性のある要因としておおむね以下の12項目が指摘できる。1:基本属性(年齢・学歴など)、2:就業形態、3:金融資産残高、4:予備的貯蓄動機、5:流動性制約、6:(勤労)所得、7:不動産(住宅)、8:参入コスト(金融知識など)、9:金融機関への信認、10:リスク回避度、11:資産選択の制度要因、12:リスク性資産の期待収益率。次節では、本稿で利用するデータを概観するが、できるだけこれらの要因が捕捉できる(または、近似できる)変数を選択することを目標とする。

なお、本稿で利用するベイジアン・ネットワークの手法を観測データに活用した先行研究としては、日本銀行(2017)やDixit et al. (2020)などが指摘できる。日本銀行(2017)は短観のデータを利用した学習を行い、企業のインフレ予想形成メカニズムに関する因果構造の把握を試みている。Dixit et al. (2020)は、世界幸福度の構造をベイジアン・ネットワークにより学習させている。彼らは、GDPなどの各連続変数を3段階に離散変数化し、ブートストラップ(bootstrap)法により学習した1万個のネットワーク構造を平均することで、最終的なモデルの提示を行っている。

### 3: データ

本稿で分析に用いるデータは、慶應義塾大学パネルデータ設計・解析センターが提供する「日本家計パネル調査」である。同調査は、個別の調査として実施・管理されてきた(旧)「日本家計パネル調査(JHPS)」(2009年～)と(旧)「慶應義塾家計パネル調査(KHPS)」(2004年～)を2014年に統合したものであり、調査対象は、JHPSが20歳以上の男女、KHPSが20歳～69歳の男女である<sup>7</sup>。新規コホートの

<sup>5</sup> 証券会社への信頼という観点からは、森・坂口(2020)が、証券会社に対するイメージがネガティブな人ほど有価証券を保有していない、との関係性があることを統計的に確認している。

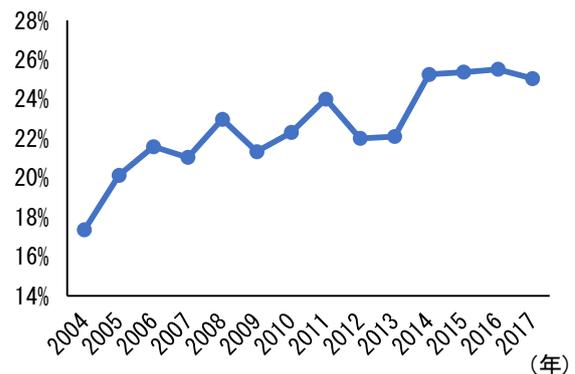
<sup>6</sup> 住宅保有に伴うリスク性資産の抑制については、徳田・齋藤(2014)でも分析されている。彼らは、OECD諸国のパネルデータによる推計から、日本のリスク性資産投資の低調さは、住宅保有要因により説明可能であるとしている。

<sup>7</sup> 調査対象者は層化2段階無作為抽出法により選定されている。調査方法の詳細などは、慶應義塾大学パネルデータ設計・解析センターのHPを参照されたい。

追加や脱落サンプルの補充を行いながら、各年1回実施されており、2017年時点のサンプル数は4,626である。調査項目も、個人属性、就業、所得、金融資産、負債、住宅など、包括的な内容となっている。調査対象者が有配偶の場合は、配偶者にも同一の調査項目が用意されている。サンプルの特徴（2017年調査時点）として、総務省「国勢調査」と比較して地域分布の乖離はないが、中高年・高学歴の世帯主が多く就業率が高いこと、総務省「家計調査」と比較して家計が保有する金融資産額は若干小さいことなどが報告されている（祝迫他、2018）。

本稿におけるリスク性資産は、日本家計パネル調査上における有価証券（株式、債券、株式投資信託、公社債投資信託、貸付信託・金銭信託など）と定義する。同調査により、リスク性資産を保有する世帯割合の推移をみたのが図表1である。中長期的には緩やかな上昇傾向で推移しているが、2014年以降は25%程度でほぼ横ばいである。総務省「全国消費実態調査」では、2014年の有価証券の保有率は23.5%であったため、日本家計パネル調査における保有率は若干高いものの、おおむね同程度である。

図表1 有価証券保有率の推移



(注) 無回答を除く。  
 (資料) 慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

本稿の分析で利用する変数は、前節で整理した要因を参考に、図表2で示した37変数を選択した。ただし、制度要因とリスク性資産の期待収益率については、適切な代理変数が見つからなかったことに加え、日本のみのデータの場合、すべての世帯に共通する要因とみなすことも可能であるため除外した。また、リスク許容度の変数は、報酬の受取り方 (Barsky et al, 1997) から計算されることが多いが、日本家計パネル調査では同様の質問がないため、リスク許容度に近い変数を選択した。図表2・33番の「選好」は、仮に300万円の資金的余裕があるときに、「株、株式投信、社債、国債、銀行預金、郵便貯金」の6種類の金融商品のうち運用したい商品の順番を記入する質問を利用し、株、株式投信、社債の順番の平均値を算出したものである。必ずしも一対一で対応しているわけではないが、本稿では同変数が高い（順番が低い）ことを、リスク選好が低いとみなす<sup>8</sup>。

日本家計パネル調査は、同一家計を追跡した調査であり動学的な分析が可能であるが、本稿では図表2で示したように各年の平均値を利用したクロスセクション・データとして利用する。これは、利用したい変数が毎年調査されていないことに加え、毎年調査されている項目でも欠損値（無回答）が少なからず確認されるため、平均値を利用した方がデータの安定性が高まることを考慮したためである。基本的には、各世帯における2015～2017年の3年間の平均値を利用するが<sup>9</sup>、一部の変数は2015・

<sup>8</sup> この質問の仕方だと、リスク許容度と同時に、金融商品「自体」に対する選好が入るため、リスク許容度のみを純粋に取り出した変数ではない。ただし、資金的余裕があるにもかかわらず、預貯金などの安全資産で運用したいとの回答はリスク許容度が低いとみなすことが可能であると考えられるため、本稿ではリスク選好と表現する。

<sup>9</sup> ただし、3年平均の変数であっても、無回答・脱落により2年平均や単年のみの値となることもある。

図表2 利用変数一覧

	変数名	説明	利用年	タイプ
目的変数	0 保有	リスク性資産(有価証券)の保有の有無(2区分)	2015～17年平均	F
基本属性	1 学歴	世帯主の最終学歴(6区分: 中学、高校、短大、大学、大学院、その他)	2016年	F
	2 年齢	世帯主の年齢(歳)	2016年	N
就業関連	3 就業	世帯主の就業(6区分: 非労、失業、自営業等、非正規、正規、役職・経営者)	2016年	F
	4 職業	世帯主の職業(7区分: 注1)	2016年	F
	5 産業	世帯主の産業(7区分: 注2)	2016年	F
	6 給与	世帯主の前年の年収(万円、税込み、主な仕事からの収入)	2015～17年平均	N
	7 労働時間	世帯主の労働時間(時間、1週間の平均)	2015～17年平均	N
予備的貯蓄動機	8 不安1	今の生活に不安がある(4階級)	2015～17年平均	O
	9 不安2	将来に不安を感じる(4階級)	2015～17年平均	O
	10 満足1	世帯所得の満足感(11階級)	2015～17年平均	O
	11 満足2	生活全般の満足感(11階級)	2015～17年平均	O
	12 余裕	定年後を問題なく生活するための所得・資産を保有(5階級)	2015～16年平均	O
	13 年金	老後の生活に十分な年金をもらえる(もらっている)(5階級)	2015～16年平均	O
金融資産	14 預貯金	預貯金額(万円)	2015～17年平均	N
所得	15 世帯所得	前年の世帯年収(万円、税込み、資産売却は除く)	2015～17年平均	N
流動性制約	16 負債	借入金残高合計(万円)	2015～17年平均	N
住宅関連	17 住宅環境	住宅の所有関係(3区分: 賃貸、持ち家、社宅など)	2016年	F
	18 住宅価値	持ち家の市場価格(万円)	2015～17年平均	N
	19 土地価値	自己保有の土地の市場価値(万円)	2015～17年平均	N
	20 資産価値	住宅価値+土地価値(万円)	2015～17年平均	N
	21 ローン	住宅ローン残高総額(万円)	2015～17年平均	N
	22 計画	今後の住宅購入の計画(4区分: なし、一戸建、マンション、その他)	2016年	F
参入コスト	23 知識1	利子率が10%、インフレ率が20%なら今お金を使うべき(5階級)	2015～16年平均	O
	24 知識2	1年後に1万円返ってくる国債の価格は1万円であるべき(5階級)	2015～16年平均	O
	25 ネット	インターネットを利用する機器数(4階級)	2015～16年平均	O
	26 証券1	(証券市場の印象) 確実なもうけが出せない(5階級)	2015～16年平均	O
	27 証券2	(証券市場の印象) 大損する可能性がある(5階級)	2015～16年平均	O
	28 証券3	(証券市場の印象) 一時の価格変化に反応して売買すべきでない(5階級)	2015～16年平均	O
信頼	29 証券4	証券会社はどの顧客に対しても役に立つアドバイスをする(5階級)	2012～14年平均	O
	30 証券5	証券会社はもうかりそうな株を紹介してくれる(5階級)	2012～14年平均	O
リスク	31 降水	初めて行く場所に出出する際、降水確率が何%以上なら傘を持っていか(4階級)	2015～17年平均	O
	32 保守的	何を買うにしても馴染みの店の方が安心である(4階級)	2015～16年平均	O
	33 選好	リスク性資産選好(ゆとりがある際の運用手段: 注3)	2015～16年平均	N
その他	34 把握1	所得の状況をどの程度把握しているか(5階級)	2015～16年平均	O
	35 把握2	金融資産の状況をどの程度把握しているか(5階級)	2015～16年平均	O
	36 合理的	自分は合理的な生き方をしている(3階級)	2015～16年平均	O
	37 割引	時間割引率(8階級: 注4)	2015～17年平均	O

- (注) 1. 職業は、販売従業者・サービス職従業者、管理的職種、事務従業者、運輸通信従業者・保安職従業者、採掘従業者・製造などの作業従業者、専門的・技術的職業従業者、その他の7区分。  
 2. 産業は、建設業、鉱・製造業、卸小売業・飲食業・宿泊業・運輸業、金融・保険業、不動産業、通信情報業、医療・福祉・教育学習支援業・公務、その他の7区分。  
 3. 選好は、仮に300万円の資金的余裕があるときに、運用したい金融商品の順番を記入する質問を利用し、株、株式投信、社債の3つの回答(順番)の平均値を算出した。  
 4. 割引は、1カ月後に1万円もらうかわりに、13カ月後に最低いくら貰えれば満足するかの回答。  
 5. タイプは、N(numeric): 実数、O(ordered): 順序あり因子、F(factor): 順序なし因子  
 6. 世帯主とその配偶者の両方のデータが利用可能な変数は、世帯主の回答を優先した。  
 7. 一部にサンプル数が非常に少ない選択肢が確認されたため、内容の近い選択肢と統合している。

(資料) 慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

16年の2年平均や2015年以前の調査項目を利用しているほか、職業など順番で表現できない（カテゴリ）変数については2016年の値を利用する<sup>10</sup>。なお、実数でない（順序あり因子）の変数の平均は、四捨五入により値を調整している。

その他のデータ処理については、以下の通りである。1点目に、世帯属性は、世帯主の属性で定義される場合が多いため、本稿でもそれを採用しているが、日本家計パネル調査の対象者は世帯主とは限らない。同調査で就業状態などの詳細な項目が取得できるのは、調査対象者とその配偶者のみであるため、分析対象とするサンプルは、調査対象者、または、調査対象者の配偶者が世帯主である家計に限定した。2点目に、本稿の関心事項である「保有」が不明であるサンプルは分析対象から除外した。3点目に、後述する機械学習の手法を適用する際には、データに欠損値があると分析ができない。そこで、欠損値の補完推計では最も精度が高いとの実例が報告されている（Waljee et al. 2013）、MissForest<sup>11</sup>により欠損値の補完を行った<sup>12</sup>。同手法のイメージとしては、次節で取り上げるランダム・フォレストにより、欠損値を除くデータから予測モデルを構築し、そのモデルを利用して欠損値の予測を行っていく手法である。ただし、欠損値が非常に多いサンプルの場合、それらを補完するための十分な情報が得られず、欠損値の予測精度が低下することが懸念されるため、図表2で示した37変数のうち、4変数以上が欠損値となるサンプルは分析対象から除外した。以上の処理を行った最終的なサンプル数は4,267である。

## 4：分析手法

機械学習の手法は、インフレ予測形成メカニズムの機械学習分析を行った日本銀行（2017）と同様のアプローチで行う。本稿の目的は、ベイジアン・ネットワーク（BN: Bayesian Network）と呼ばれる機械学習の手法を利用し、リスク性資産保有に対する因果構造を学習することであるが、図表2で示した全ての変数を利用すると、複雑かつ解釈性の乏しい構造が推計され、因果構造を適切に把握できないことが懸念される。そこで、BNの分析前に、リスク性資産保有に対して重要度の高い変数を選定するステップを挟む。具体的には、機械学習の分野で一般的によく利用されているランダム・フォレスト（RF: Random Forest）と呼ばれる手法を活用する。以下では、日本銀行（2017）をベースに、それぞれの分析手法を概観する<sup>13</sup>。

### （1）ランダム・フォレスト

RFは、機械学習の分野では広く知られている手法であり、被説明変数に対する予測力が高いことで

<sup>10</sup> 2年平均の値でも、どちらかの年が無回答であれば回答があった年だけの値となる。また、2016年の値を利用する変数（平均でない）変数についても、2016年が無回答のサンプルについては、2015年や2017年に回答があれば、その値を利用する。ただし、2016年が無回答で、2015・17年の両年で値はあるが一致しない場合は、欠損値とした。

<sup>11</sup> パッケージは、RのrandomForestSRCを利用した。多変量バージョンのMissForest（Stekhoven and Bühlmann, 2012）を適用した。

<sup>12</sup> MissForestを行う際には、上記37変数以外にも補完推計に役に立つ可能性のある変数を加えて行った。具体的には、住宅の購入時期、住宅の購入価格、敷地面積、駅からの距離、消費額などの変数を加えている。ただし、これらは、あくまでも補完推計の精度を上げることを目的に加えた変数であるため、最終的な分析には利用していない。

<sup>13</sup> 日本銀行（2017）は、各手法について分かりやすく解説しているので、詳細は同論文も参照されたい。

有名である<sup>14</sup>。また、特定の関数式を仮定しないこと、非線形の関係も把握できること、多重共線性の問題が起りにくいことなどの利点もある。推計方法としては、まず、説明変数の中からランダムにいくつかの変数を選択する。一般的には、 $k$ 個の説明変数の中から、 $\sqrt{k}$ 個の変数を選択する機会が多いため、本稿でもこれを踏襲する<sup>15</sup>。このランダムに選択した変数から「決定木 (decision tree)」と呼ばれるツリー構造をしたモデルを作成する。決定木とは、サンプルを2つに分割することを繰り返し、被説明変数の予測誤差を小さくしていく分析手法である<sup>16</sup>。この「ランダムに変数を選定⇒決定木の推計」というプロセスを繰り返すことで、大量の決定木を作成する。一部の説明変数だけを利用して、個々の決定木の予測精度は高いものではないが、これらの決定木を組み合わせる（決定木による多数決により値を予想する）ことで、予測精度が高くなることが知られている<sup>17</sup>。

本稿でRFを利用する背景には、同分析手法が重要度 (variable importance) の算出が可能であることも理由となっている。この重要度を利用することで、リスク性資産保有の有無には、どの変数との関係性が強いのか把握することが可能になる。本稿で計算する重要度は、permutation importanceと言われ、ある変数をサンプル間でランダムに入れ替えた際に、予測誤差がどの程度上昇するかにより計算される。リスク性資産保有の有無に対して、ある説明変数との関係性が非常に高かった場合、その変数をサンプル間でランダムに並び替えるとモデルの説明力は大幅に低下する。一方、ほぼ無関係の説明変数であれば、同変数のシャッフル前後で予測精度にはほとんど変化が観察されないため、重要度が低いことを確認できる。

## (2) ベイジアン・ネットワーク

RFは変数の重要性を指摘できるが、変数間の因果構造まではわからない。例えば、年齢の重要度が高いとの分析結果が得られても、年齢がどのような経路でリスク性資産の保有に影響を与えているかは不明である。そこで、RFで得られた重要度の高い変数を利用し、BNを学習させることが次のステップとなる。これにより、リスク性資産の保有に対してどの変数が直接的（間接的）な影響を与えているのか（因果構造）が推計される。構造が把握できれば、政府や企業が介入可能な変数の値を変化させることで、介入によるリスク性資産保有率の変化を試算することが可能となる。

BNについても概観しておきたい<sup>18</sup>。BNとは、有向非巡回グラフ (DAG : Directed Acyclic Graph) を用いて、変数間の条件付き確率の関係性を示したものである<sup>19</sup>。例えば、A~Fの6つの確率変数の同

<sup>14</sup> RFの分析手法の詳細は、James et al. (2013) なども参照されたい。

<sup>15</sup> 本稿であれば、説明変数は37個（図表2）であるため、1つの決定木に利用する変数は6個である。

<sup>16</sup> 例えば、「預貯金」・「負債」の2変数を利用してリスク性資産保有率の決定木を作ることを考える。まず、任意の閾値で貯蓄額のサンプルを2つに分割し（貯蓄額が高いサンプルと低いサンプルに分割し）、分割したサンプル間で「保有」の予測誤差の分散を計算する。この時、この予測誤差が最小になる閾値がサンプルを2つに分ける候補となる。同様のプロセスを「負債」でも行うことで、どちらの変数を利用して2分割する方が、予測誤差が小さいか判明する。例えば、「貯蓄」の方が小さければ、決定木の最初の分岐点は貯蓄となる。このようなプロセスを繰り返すことで、決定木が学習される。

<sup>17</sup> このような学習手法は、アンサンブル学習と言われる。

<sup>18</sup> 本稿でのBNに関する記載は、Scutari and Denis (2014)、Hunermund and Bareinboim (2019)、Kalisch et al. (2012) を参考にした。

<sup>19</sup> DAGは名前の通り、変数間の関係性に向きがあり (Directed)、巡回しない (Acyclic) グラフである。例えば、「 $X \rightarrow Y \rightarrow Z \rightarrow X$ 」は巡回してしまうので、Acyclicの条件を満たさない。

時分布が以下のように示される場合、DAGは図表3のように示すことが可能である。

$$P(A) P(C) P(B|A, C) P(D|B) P(E|B) P(F|E)$$

仮に政府がDに影響を与えたいが、介入可能な変数がCとEであった場合、ランダム・フォレストではCもEも、Dへの重要度が高い（Dと相関がある）と学習される可能性が高いが、BNにより因果構造が特定できれば、実際にはCに介入しないと期待する効果は得られないことが判明する。なお、図表3における各変数の丸（○）をノード（node）、ノード間を結ぶ線をエッジ（edge）と呼ぶ。

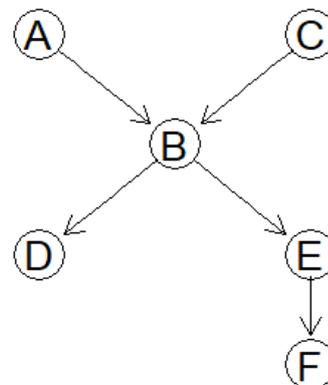
本稿でのBNの学習は、観測データを利用して図表3のようなネットワーク構造を学習し（structure learning）、学習したネットワーク構造を所与として各変数の条件付き確率を学習する（parameter learning）、という2つのステップを踏む。

まず、最初のネットワーク構造の学習手法（structure learning）については、①Constraint-based algorithms（変数間の条件付き独立を統計的に検定することで構造を学習）、②Score-based algorithms（BIC(Bayesian Information Criterion)などデータの適合度(goodness of fit)を表すスコアを最大化するように構造を学習）、③Hybrid algorithms（①と②の手法を組み合わせることで構造を学習<sup>20</sup>）の3種類のアプローチが存在する。本稿で主に利用するのはPC（Peter & Clark）アルゴリズムと呼ばれる手法であり、上記の①に分類される。同アルゴリズムは、特に変数間の因果構造を把握する際に利用される手法であり、本稿の目的に合致すると想定される。

PCアルゴリズムの概要については補論1を参照されたいが、同アルゴリズムは全部で2つのステップから構成されている。第1のステップでスケルトン(skeleton)と呼ばれる各ノードを矢印のないエッジでつないだグラフを推定し、第2のステップでそのスケルトンを所与としてエッジの矢印の方向性を推計する。ただし、第2のステップでは、すべての矢印の方向性が決定できるとは限らないという推計上の制約が存在する点には留意が必要である。

また、PCアルゴリズムにおけるネットワーク構造は、条件付き独立の検定を繰り返し行うことで推定されるが、連続変数と離散変数が混合する場合には学習が非常に困難となるため<sup>21</sup>、通常、各変数を離散変数又は連続変数のどちらかに統一することが必要となる。連続変数の場合、各変数の分布がGaussianであることの仮定が一般的であるが、本項の関心事項である「保有」の変数が既に2値変数であることから、この仮定を満たすとは想定しにくい。そのため、本稿のBNの分析では、すべて離

図表3 DAGの例



(資料) みずほ総合研究所作成

<sup>20</sup> まず、条件付き独立の検定により、条件付き独立の関係性を部分的に把握する。これによりサーチ・スペース (search space) を狭めた上で、score-based algorithms による学習を行う。

<sup>21</sup> 連続変数と離散変数が混合する場合も、学習が可能な R のパッケージ (deal) も存在するが、「連続変数ノード → 離散変数ノード」との関係性は不可との制約条件が存在する。

散変数に揃えて（連続変数は離散変数に変換した上で）行うこととする<sup>22</sup>。

以上のプロセスにより、図表3のようなネットワーク構造が学習されるが、ここで十分に注意が必要なことは、推計された構造が真の因果関係を示していると安易に解釈すべきではないことである。これは、PCアルゴリズムにより学習されたネットワークが因果構造であると明言するには、潜在変数（latent variables）がないとの条件を満たす必要があるためである<sup>23</sup>。ただし、通常、潜在変数の有無は不明であることに加え、特に観測データを利用している場合は、潜在変数をコントロールする手段もない。真の因果構造は未知ではあるものの、因果構造を適切に表現している確率の高いBNの特徴として、相当にまばら（fairly space）であり、変数間の解釈性が明瞭かつ有意義なものであることが指摘されている（Scutari and Denis, 2014）。学習結果を鵜呑みにするのではなく、専門家の知見なども借りながら、学習された因果構造の妥当性を検討することが重要である。

妥当性の高いDAGが推計された後には、各ノードの条件付き確率の推計（parameter learning）を行う。本稿では一般的な最尤推定（maximum likelihood estimation）を利用する<sup>24</sup>。離散変数の場合は、観測データの頻度を元に各ノードにおける確率が計算される。この計算された確率を用いて介入によるリスク性資産保有率の変化をシミュレーションする。

## 5：分析結果

### （1）ランダム・フォレスト

RFの分析による重要度の結果を確認する<sup>25</sup>。被説明変数は、リスク性資産の保有の有無であり、説明変数は図表2で示した37変数である。また、頑健性の観点から、全サンプル（サンプル数：4,267）と世帯主を就業者に限定したサンプル（サンプル数：3,298）の2種類で分析を行った。決定木の個数は1,500個である。また、2種類のサンプル間で比較が行いやすいように、重要度は最も高い値を1とした場合の相対的重要度（relative importance）で評価する。

図表4が分析結果であり、2種類の相対的重要度の平均値が高い順番に上位15位をプロットしている<sup>26</sup>。図表からは「選好」と「預貯金」の相対的重要度が突出して高いことが確認できる。次に「余裕」がやや高めであり、4位以降の重要度はほぼ同程度となっている。この分析からは、安全資産である預貯金とリスク性資産に対する選好が、有価証券保有の有無を予想する上では非常に有用な変数であることが確認できる。塩路他（2013）の分析結果では、家計の株式保有を予測する上では、金融資産残高が重要な変数であることが指摘されていたが、本稿の分析でも同様のインプリケーションが得られたと考えられる。リスク選好の重要性が高いとの結果は、当然と言えるかもしれないが、先行研究

<sup>22</sup> 離散変数の条件付き独立の検定にはピアソンのカイ二乗検定（Pearson's  $\chi^2$  test）を利用した。また、検定時における有意水準は5%とした。

<sup>23</sup> この条件以外にも、「faithfulness 仮定」及び「causal Markov 仮定」と呼ばれる2つの仮定が必要となるが、latent variables がある場合、この2つの仮定を自動的に満たさない。

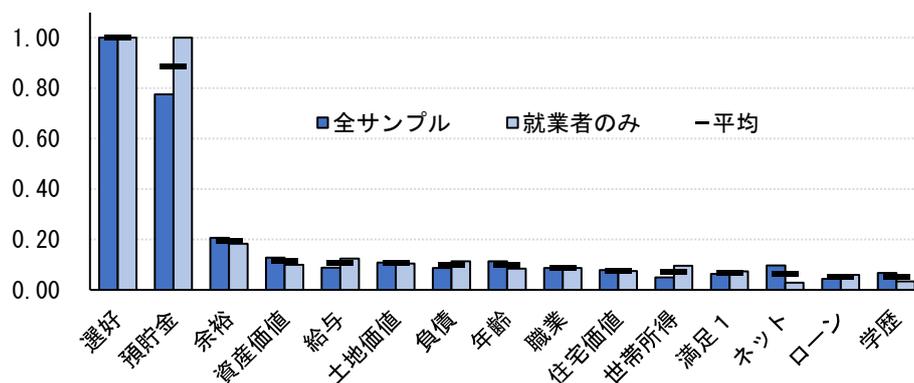
<sup>24</sup> その他にも、Bayesian Estimation の手法も一般的に利用される。

<sup>25</sup> 分析には、R の randomForestSRC のパッケージを利用した。

<sup>26</sup> 37変数すべての重要度は補論2を参照のこと。

(伊藤他 (2017) など) でも重要な要素であると指摘されてきた相対的リスク許容度の重要性を示唆する結果である。また、先行研究では、不動産、参入コスト、学歴、年齢などの要因もリスク性資産に対する有意性が指摘されていたが、RFの分析結果からは、リスク性資産保有の有無に対する直接的な予測力はあまり高くないことが示された<sup>27</sup>。

図表 4 相対的重要度 (被説明変数:「保有」)



(注) 平均ベースで上位 15 位をプロットしている。

(資料) 慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

次に、突出して重要度が高かった「選好」と「預貯金」の関係性を詳しく分析する。図表 4 で重要度が高くなかった変数でも、この 2 変数に影響を与えることで、間接的にリスク性資産保有に対して影響を与えている可能性が考えられる。そこで、この 2 変数をそれぞれ被説明変数とし、「選好」と「預貯金」を除く残りの 35 変数を説明変数とした分析 (RF) を行った。上記同様、全サンプルと世帯主が就業者の 2 種類で分析を行い、決定木の数は 1,500 個とした。「選好」と「預貯金」に対する (相対的) 重要度の結果をプロットしたのが、それぞれ図表 5 と図表 6 である<sup>28</sup>。

「選好」に対する重要度では、「証券 1~3、ネット、知識 1・2、学歴、年齢、世帯所得」などの変数の重要度が高いことが確認できる。参入コストに関連する変数の重要度が全般的に高いことが特徴であり、リスク選好の高さには、証券市場に対する認識、金融リテラシー、ネットなどからの情報収集手段などが影響している可能性が考えられる。先行研究 (北村・内野 (2011)、van Rooij et al. (2011) など) では、金融リテラシーなどの参入コストの要因も指摘されていたが、この要因はリスク性資産保有の有無に対して直接的に影響するわけではなく、リスク選好などの要因を通して間接的に影響している可能性が想定される。

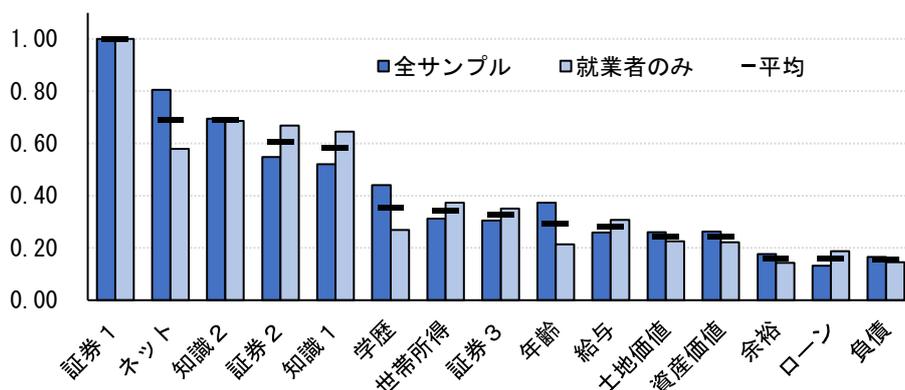
「預貯金」を被説明変数とした場合、重要度の高い変数に「余裕、世帯所得、負債、年齢」などが指摘できる。平均ベースで重要度が最も高い変数は「余裕」であり、定年退職後の生活に資金的ゆとりがあるとの認識は、預貯金の残高と密接に関係していることが確認できる。なお、「余裕」は、図表

<sup>27</sup> ただし、ランダム・フォレストの重要度は統計的有意性をみるものではないため、これらの変数の有意性を否定するものではない。

<sup>28</sup> 35 変数すべての重要度は補論 2 を参照。

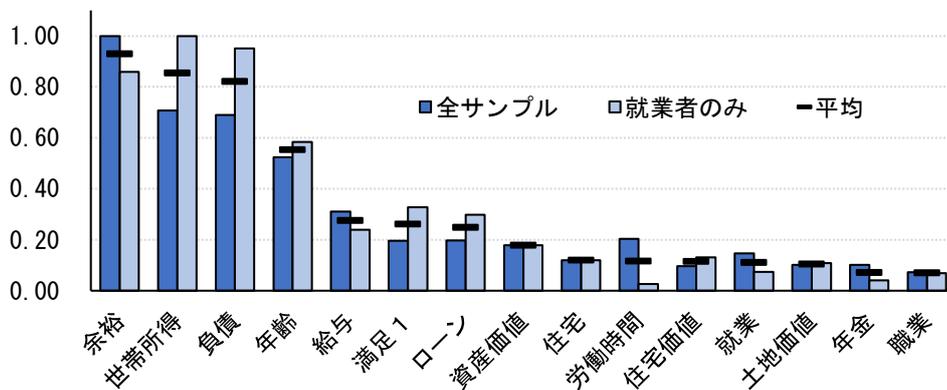
4でも3位であり、4位以下の変数と比較してやや高い重要度が観察されていた。また、世帯所得と負債は就業者のみのデータにおいて特に高い値となっているが、世帯所得の高い家計は貯金が貯まりやすい、負債を持つ家計は貯金が貯まりにくいとの関係性が示唆される。住宅ローンの存在がリスク性資産の保有を抑制するとの指摘があるが (Iwaisako et al. (2019) など)、借入金の存在により、貯金が貯まらないためリスク性資産も保有しないとの関係性が示唆される。なお、住宅ローンより概念の広い負債の方が預貯金に対する重要度は高いため、より重要なのは借入金の金額であり、借入金の目的・種類ではないと想定される。

図表5 相対的重要度 (被説明変数:「選好」)



(注) 平均ベースで上位15位をプロットしている。  
 (資料) 慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

図表6 相対的重要度 (被説明変数:「預貯金」)



(注) 平均ベースで上位15位をプロットしている。  
 (資料) 慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

## (2) ベイジアン・ネットワーク

次に、重要度の高い変数を利用して、変数間の因果構造の推計を試みる<sup>29</sup>。重要度の高い変数をどのように定義するかは難しい問題であるが、本稿では便宜上、図表4～6における相対的重要度の平均

<sup>29</sup> PCアルゴリズムによる学習はRのpcalgのパッケージを利用し、それ以外のBNの分析はRのbnlearnのパッケージを利用した。

が0.3以上の変数と定義する。該当する変数は、図表4では、「選好、預貯金」の2変数、図表5では、「証券1、ネット、知識2、証券2、知識1、学歴、世帯所得、証券3、年齢」の9変数、図表6では、「余裕、世帯所得、負債、年齢」の4変数である。これに図表4の被説明変数である「保有」を加え、重複を除くと14変数となる。ただし、因果構造の解釈性の向上と議論の簡素化を図るため、似た要因をキャプチャーしている「証券1と証券2」、「知識1と知識2」は、より重要度が高い変数（「証券1」、「知識2」）を利用する。以上のプロセスを踏まえ、最終的なBNの分析に利用する変数は12変数とする。

前述の通りBNの分析では、連続変数が離散変数に統一する必要があるが、データの特性上、離散変数に統一する。また、離散変数の階層が多いとサンプル数の問題から検定が適切に行えないことに加え、学習後の議論を簡素化するため、離散化に際しては、「保有」の変数を除くすべての変数を3段階（高・中・低など）で表現する<sup>30</sup>。具体的には、預貯金、世帯年収、負債の連続変数は、Hartemink(2001)のアルゴリズム<sup>31</sup>により3分割した。それ以外の変数はサンプル数の分布や変数の意味を考慮して3段階に整理した。離散化・整理後の最終的なBNの分析に利用した変数の概要は図表7で示している。

図表7の変数を利用し、PCアルゴリズムによる学習を行ったが、前述の通り同アルゴリズムでは、すべてのエッジの方向性を決定できるとは限らず、事実、8個のエッジの方向性が不明との結果が得られた（推計されたBNは補論3を参照）。うち5個のエッジは、学歴と年齢が関係しているが、図表7の変数のうち、年齢と学歴に関しては、各世帯の属性を表す変数であり、外生的な性質が強い。そのため、他の要因からこの2つの変数にエッジの矢印が向くことは通常あり得ず<sup>32</sup>、これらのエッジの方向性が決定される。年齢・学歴以外の変数で矢印の向きが不明なエッジは、「預貯金 — 負債」、「世帯所得 — 負債」、「預貯金 — 世帯所得」の3種類である。うち世帯所得関連のエッジは、所得水準が高いので預貯金も多い、所得水準が高いので多くの借入が可能との関係性を想定する方が自然であると思われるため、「世帯所得→預貯金」及び「世帯所得→負債」であると考えられる<sup>33</sup>。一方、預貯金と負債の関係性は必ずしも明確ではない部分もあるが、一般的な家計は、貯蓄より借金の返済を優先する傾向があると考えられることから、本稿では負債（返済義務）があるので預貯金が増えない（「負債 → 預貯金」）との関係性を想定する<sup>34</sup>。

以上のプロセスを踏まえ、最終的なDAGをプロットしたのが図表8となる。分析手法でも言及した通り、同グラフの因果構造の解釈には留意が必要であるが、以下で論じるように、各変数間の関係性

<sup>30</sup> 先行研究で紹介したDixit et al. (2020)や、Scutari and Denis(2014)の分析例でも3段階でBNの分析が行われている。

<sup>31</sup> 連続変数の離散化に利用されるアルゴリズムであり、変数間の相互情報量 (pairwise mutual information) をできる限り維持したまま、各変数を離散化する手法である。

<sup>32</sup> 例えば、「年齢が高くなったので給与水準が上昇する」との関係性は理解できるが、「給与水準が上昇したので年齢が高くなる」との関係性はあり得ない。

<sup>33</sup> 保有する預貯金残高や借入金残高により、個々の所得水準が決定されることは通常あり得ない。なお、世帯所得と預貯金（世帯所得と負債）の間には正の相関関係があることがデータからは確認される。

<sup>34</sup> 預貯金と負債残高にはデータ上、負の相関関係が観察される。預貯金が少ないので借入をするという関係性も考えられるものの、この関係性が借入を行う時にのみ成立するのに対し、借入の存在により預貯金が貯まらないとの関係性はある程度の期間に渡って預貯金に影響を及ぼすと考えられるため、本稿では「負債→預貯金」との関係性を想定する。

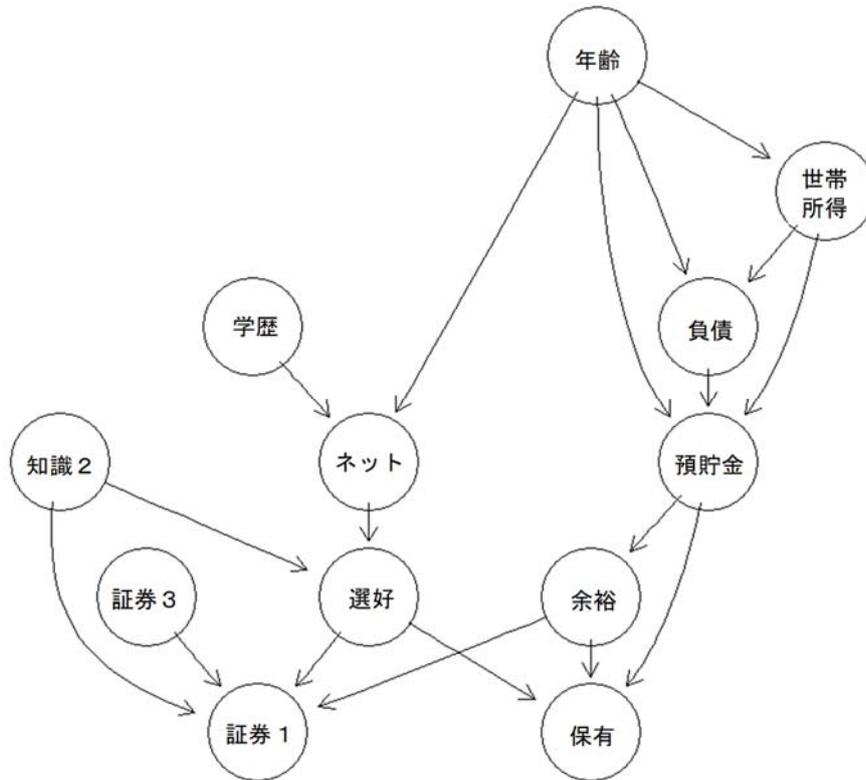
図表7 整理後の変数一覧

変数	離散変数後	備考
保有	①保有あり<1122>、②保有なし<3145>	有価証券の保有の有無
学歴	①中卒・高卒<2254>、②短大・その他<540> ③大卒・院卒<1473>	世帯主の最終学歴
年齢	①～39歳<444>、②40～64歳<2308>、③65歳～<1515>	世帯主の年齢
余裕	①高<878>、②中<1295>、③低<2094>。	定年後の資金的余裕：「1(高い)⇔5(低い)」の5段階の質問に対して、1～3:高、4:中、5:低
預貯金	①高(940万～)<1288>、②中(157～940万)<1582>、 ③低(～157万)<1397>	Hartemink(2001)による分割
世帯所得	①高(653万～)<1702>、②中(373～653万)<1499>、 ③低(～373万)<1066>	Hartemink(2001)による分割
負債	①高(3220万～)<107>、②中(140～3220万)<1492>、 ③低(～140万)<2668>	Hartemink(2001)による分割
知識2	①高<522>、②中<1566>、③低<2179>	国債価格の問題：「1(不正解)⇔5(正解)」の5段階の評価に対して、1～3:低、4:中、5:高
ネット	①利用なし<962>、②1つ<1276>、③2つ以上<2029>	インターネットを利用する機器数
証券1	①そう思う<991>、②まあまあそう思う<1592>、 ③それ以外<1684>	「(証券市場の印象) 確実なもうけが出せない」への回答
証券3	①思う・まあまあそう思う<1828>、②どちらともいえない<2058>、 ③それ以外<381>	「(証券市場の印象) 一時の価格変化に反応して売買すべきでない」への回答
選好	①リスク選好・高<1048>、②リスク選好・中<1668>、 ③リスク選好・低<1551>	株・株式投信・社債の運用順位の平均が、3.83以下:高、3.83超4.7以下:中、4.7超5以下:低

(注) <>内はサンプル数。

(資料) 慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

図表8 ベイジアン・ネットワーク (PC アルゴリズム)



(資料) 慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

はおおむね妥当な範囲に収まっていると考えられる。また、金融機関で対顧客業務を担当した経験のある実務家にも同グラフの因果構造の検証を依頼したが、違和感はないとの回答が得られた。評価には多分に主観が含まれるため、これをもって学習されたグラフに問題がないとは断言できないものの、少なくとも学習された DAG はあり得る因果構造の 1 つを捕捉していると考えられる。

図表 8 からは、本稿の関心事項である「保有」に直接的な影響を与えているのは「選好」、「預貯金」、「余裕」の 3 変数であることが確認できる。リスク性資産への選好と預貯金の 2 変数は RF の分析でも重要度が高かった変数であるが、同分析で 3 位であった「余裕」も直接的にリスク性資産の保有に影響を与えている可能性が指摘できる。また、同 DAG を所与として、各ノードの条件付き確率を最尤推定し、主要な確率を補論 4 でプロットしている。条件付き確率を確認すると、これら 3 変数が「高」の場合、リスク性資産保有確率は約 47~48% であるのに対し、「低」の場合では同確率は約 7~16% にとどまり、非常に大きな差が観察された。

「余裕」には預貯金額が影響しており、預貯金額には世帯所得と負債が影響を与えていることが確認される。条件付き確率からは、世帯所得が多いほど預貯金も多い、負債が多いほど預貯金が低いとの傾向にあることが観察される<sup>35</sup>。これらの構造を踏まえると、祝迫 (2012) でも指摘されている通り、リスク性資産の購入は家計に余裕資金がある場合に行われており、住宅ローンなどの借入金の返済が完了し、老後の生活に向けた十分な安全資産を確保する（又は確保したと認識する）までは、リスク性資産の投資を行わない傾向にあることが想定される。所得が高い家計や、負債がない（返済が完了した）家計の方が十分な安全資産を確保できる確率が高く、先行研究でリスク性資産保有に対して所得・負債関連の変数の有意性が確認された背景である可能性が想定される。また、年齢は「保有」に対して直接的な影響力はないが、世帯所得・負債・預貯金のすべてに影響を与えているため、年齢階級（ライフ・サイクル）とリスク性資産保有率との間に正の相関（より正確にはコブ型の関係性）が確認できる（塩路他、2013）背景になっていると想定される。

「保有」に対して影響力のあるもう一つの要因がリスク選好であるが、この選好には、金融リテラシーと、インターネットを利用する機器数が影響を与えているとの結果が得られた。ランダム・フォレストの分析では「証券 1」の重要度が高かったが、同変数は「選好」に影響を与えているのではなく、逆に「選好」から影響を受けている変数であると学習された。「選好」について金融リテラシーの条件付き確率をみると、知識が「高」の場合、「選好」が高い確率は約 40% であるが、知識が「低」の場合、同確率は約 19% と半減する。金融市場に対する正しい知識を持っていることが、リスク性資産に対する選好の高さにつながっていることが想定される。「ネット」の条件付き確率も確認すると、リスク選好が高い確率は、利用機器数が 2 つ以上で約 28%、1 つで約 24%、0（なし）で約 17% である。リテラシーと比較すると「ネット」による「選好」の差異は小さいが、幅広く情報収集が可能

<sup>35</sup> 補論 4 では、預貯金の条件付き確率を示しているが、預貯金が「高」の確率は、世帯所得「低：25%、中：27%、高：37%」、負債「低：39%、中：14%、高：27%」である。預貯金「高」の確率は、負債「高」が負債「中」よりも高くなっているが、この背景には借入金を多く借り入れるためには、ある程度の貯蓄額が必要であるためと想定される。ただし、預貯金「高」の確率が最も高いのは、負債が「低」の世帯であり、負債の存在により預貯金が貯まりづらいとの関係性が示唆される。

な者は、リスク性資産に対する選好が高くなることが示唆される。なお、「ネット」には、年齢と学歴が影響しているとの結果が確認された。高齢者の場合「ネット」利用がない確率が高く、高学歴であればネットを利用する機器数が2つ以上の確率が高くなっている。また、図表8からは、「学歴」が影響を与えている変数は「ネット」のみであり、一般的に想定されることの多い、学歴から所得や知識などへの影響は観察されなかった。

以上を総合すると、リスク性資産の保有の有無には大きく2つのチャンネルがあることが推察される。1つは老後に備えて十分な安全資産（預貯金）が確保できた際の余裕資金をリスク性資産の投資へ向けるチャンネルである。安全資産の確保には、世帯の所得や負債が影響を与えており、所得が低い場合や住宅ローンを含む負債の存在は、預貯金が貯まらない要因となるため、リスク性資産の保有確率も低くなる。2つ目は、リスク選好を通じたチャンネルである。金融リテラシーや情報収集の能力が高い場合、より自身にとって合理的なリスク選好を形成することが可能になり、リスク性資産の保有率も高まることが想定される。

### **(3) 介入効果**

上記の分析を踏まえて、リスク性資産保有率を高めるため、どのような政策対応が効果的かについてシミュレーションを行う。具体的な介入の前提条件は図表9に示しているが、「知識2」、「ネット」、「余裕」の変数の確率分布を変更した場合に、それぞれリスク性資産保有率がどのように変化するかを確認する。介入1と2はリスク選好を高めることで、保有率の向上を狙うものであり、介入1では金融リテラシー、介入2では情報リテラシー（インターネットの利用）を向上させる。介入3と4は、預貯金をコントロールすることはできないため、定年後の資金的ゆとり（「余裕」の変数）に対する認識を変化させることを目的とする。介入3は、例えば、抜本的な構造改革により年金不安の解消や期待成長率の上昇が実現し、全員が「余裕＝高」と認識すると仮定する。介入4は少なくとも「余裕＝低」の家計がなくなる（「低」の家計は全員「中」になる）との前提である。例えば、中城他（2017）が指摘するような長期的な生活設計の促進（老後の生活を具体的にイメージさせる）などの対応により、老後の資金不足を過大評価する家計を減らす対応を想定する。

ただし、算出される介入効果は、あくまでも観測データによる条件付き確率から算出されるものであることに加え、各ノードで利用している変数は代理変数も含んでいる点には留意が必要である。例えば、金融リテラシーは金利・リスクなどの様々な知識が含まれる概念であるが、データの制約上、本稿では国債価格の知識のみを代理変数として利用している。そのため、同変数は金融リテラシーの一部しか把握できておらず、シミュレーションは実際の効果より過小評価となる可能性がある。同様の指摘は「余裕」や「ネット」などの他の変数にも当てはまるため、各介入による保有率上昇幅の絶対値よりも、上昇幅を相対的に比較することで、どの介入がより効果的な可能性が高いかに注目する方がより適切と思われる。

分析結果をプロットしたものが図表10（1）である。介入1と2はリスク選好の変化を通じた効果であるため、介入後のリスク選好の分布を図表10（2）でプロットしている。分析結果をみると、一番効果が高いのは介入3の場合であり、リスク性資産保有率の増加幅は10%Ptを超える。ただし、老

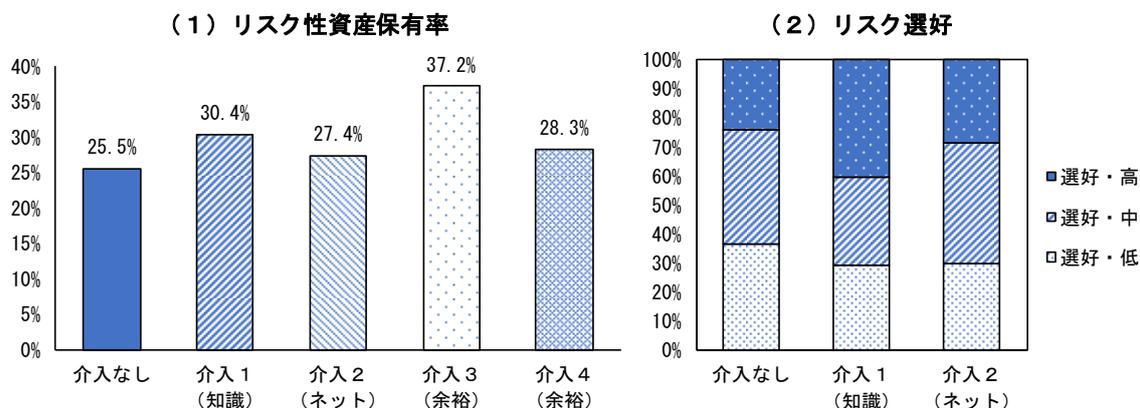
後不安の解消などが必要であり、政策の難易度は非常に高い。次に効果的な介入は、金融リテラシーの向上であり、5%Pt程度の保有率の上昇が確認される。現状の分布では同問題を正確に回答した人は12%と非常に限定的であるため、政府や金融機関などによる金融教育の促進余地は大きいことが指摘できる。3番目には、少なくとも定年後の資金的ゆとりが全くないとする層をなくすことであり、3%Pt程度の増加となる。必要以上に老後の資金不足を想定している家計が存在する場合、生活設計のシミュレーションなどにより、全く足りない状況ではないことを明確にすることなどが考えられる。また、効果の程度は最も低い、インターネットの利用機器数を増やすことでも2%Pt程度の保有率の上昇が期待できる。

図表9 前提条件

	介入1: 金融知識			介入2: インターネット			介入3・4: 定年後の余裕				
	高	中	低		2以上	1	0		高	中	低
現状	12%	37%	51%	現状	48%	31%	22%	現状	21%	30%	49%
介入1	100%	0%	0%	介入2	100%	0%	0%	介入3	100%	0%	0%
								介入4	21%	79%	0%

(資料) 慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

図表10 介入効果



(資料) 慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

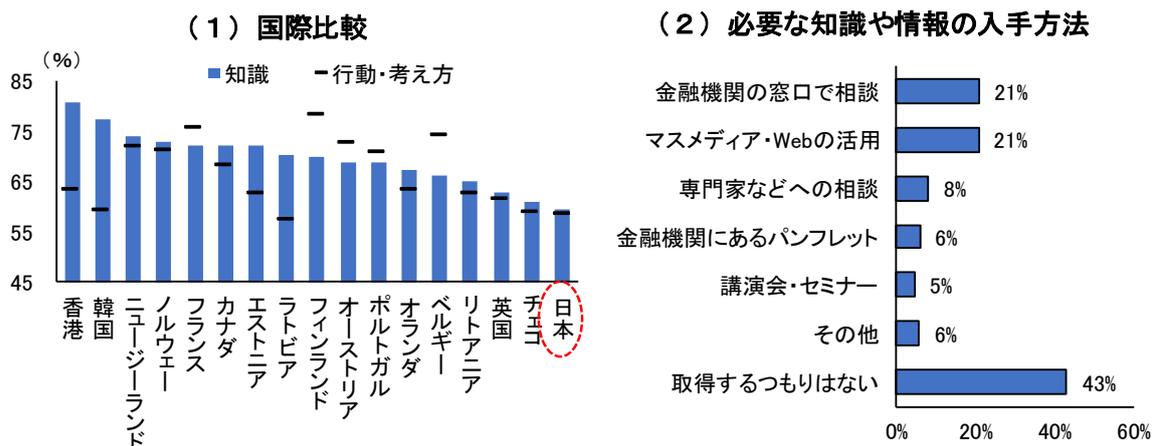
本稿の「はじめに」では、日本の現預金が投資に向かわないことが問題視されていると指摘したが、これは老後の資金余剰がない（十分な預貯金が貯まるまで投資しない）と考える人が多いことが主因となっている可能性が考えられる。事実、上記シミュレーションでも、老後に向けて十分な資金があると認識させることができれば、リスク性資産の保有率は他の介入より大きく上昇している。ただし、少子高齢化により、中長期的に潜在成長率が低下する可能性が高い日本経済では、老後の資金面での不安を和らげるには、年金改革や潜在成長率の底上げなどの抜本的な対応が必要であり、実現に向けた難易度は非常に高い。

そのため、効果の点では次善となるが、実現可能性の観点からは、金融リテラシーを向上させることで、人々のリスク性資産に対する考え方を変革していくことが、リスク性資産保有率の向上には有

効であると考えられる。金融広報中央委員会「金融リテラシー調査（2019年）」によると、OECD調査と比較した日本の金融知識に関する正答率（金利・リスクなど5項目の平均）は、データが利用可能な30か国中22位であり、データを先進国・地域に限定した場合、17か国中最下位となる（図表11（1））<sup>36</sup>。同調査では、望ましい金融行動・考え方を選択した割合（長期計画・貯蓄重視など6項目の平均）についても国際比較を行っているが、日本の順位は17か国中16位であり、国際的にみても金融リテラシーの低さが問題点として指摘できる。

今回の分析に利用した「日本家計パネル調査」の2017年調査では、投資の際に必要な知識や情報の入手先について調査されているが（図表11（2））、インターネットの活用と並んで金融機関の窓口との回答が多く、金融機関も積極的にリテラシー向上に向けた取り組みを行うことが重要であると考えられる。ただし、同調査では「取得するつもりがない」との回答が約4割を占め、金融リテラシーの向上に対して消極的な家計が多いことも確認される。より合理的なリスク選好を形成することは、家計にとっても社会にとっても望ましい変化であるため、こうした消極的な層に対するアプローチの手段が課題となろう。

図表11 金融リテラシー



(注) (1) 先進国・地域のみを表示。(2) 複数回答。

(資料) 金融広報中央委員会「金融リテラシー調査（2019年）」、慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

#### (4) 頑健性

最後にPCアルゴリズムで学習したネットワーク構造の頑健性について確認したい。具体的には、他のネットワーク構造を学習する（structure learning）アルゴリズムを利用して、PCアルゴリズムと学習結果がどこまで異なるかを確認する。ただし、前述の通り学習された構造が因果構造を示していると安易に解釈するのは適切でないことに加え、学習アルゴリズムはエッジの向きを断言できないものも含まれるため、ここでの頑健性の検証はスケルトン（エッジの矢印がないグラフ）を確認するにとどめる。これにより、例えば「保有」と「預貯金」の間に（矢印の向きは不明だが）エッジは存在

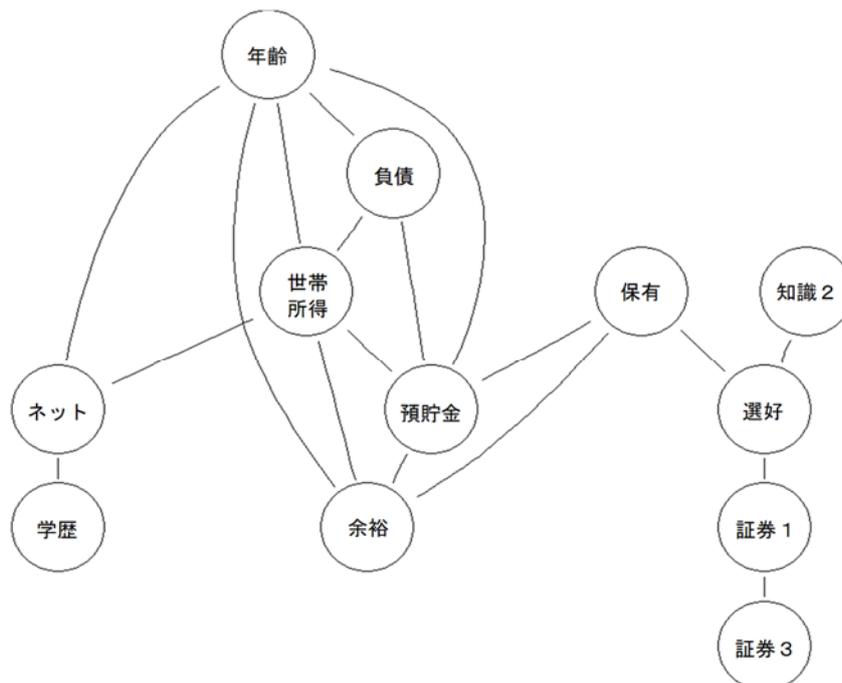
<sup>36</sup> OECD調査は、OECDの金融教育に関する国際ネットワーク会議参加国に対する調査。先進国・地域の定義はIMFによる区分を利用した。

することの頑健性が確認できる。

利用したアルゴリズムは、PCアルゴリズムと同じカテゴリーに属するConstraint-basedアルゴリズムより3種類 (Hilton-PC、Grow-Shrink (GS)、Incremental Association (IAMB))、Score-basedアルゴリズムより2種類 (Hill-Climbing (HC)、Tabu Search (Tabu))、Hybridアルゴリズム1種類 (Restricted Maximization (RMAX2)) の合計6種類である。本稿では、それぞれのアルゴリズムの学習結果より得られた6個のネットワーク構造のスケルトンのうち、4個以上の構造でエッジが確認されたものを頑健性が高いものとして扱うこととする。なお、各アルゴリズムの設定などは補論5を参照されたい。

頑健性の高さが確認されたエッジをプロットしたものが図表12である。PCアルゴリズム (図表8) のスケルトンと比較すると、PCアルゴリズムで確認されたエッジは18本であったが、図表12もエッジの本数は18本であり総数は同じである。うち15本のエッジは両者で共通しており、全般的にPCアルゴリズムの結果は頑健性が高いことが示唆される。ただし、PCアルゴリズムでは確認されていた3本のエッジが図表12では観測できず、逆に新しい関係性が3本確認されている。

図表12 他のアルゴリズム学習結果の平均



(資料) 慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

PCアルゴリズムと異なり観測されなかった関係性は、「選好—ネット」、「知識2—証券1」、「余裕—証券1」の3種類であるが、前述のインプリケーションに対する変更点となりうるのは、「選好」と「ネット」の関係性であろう。PCアルゴリズムでは「ネット→選好」との関係性が指摘されたが、ここでの分析結果では、6個の学習結果のうち2個のみで確認されるにとどまった。頑健性が全くないわけではないが、前掲の図表10でみた介入効果でも一番効果が小さかった関係性でもあるため、やや

懐疑的に評価する方が適切かもしれない。一方、図表12で新しく観測された関係性は、「余裕—世帯所得」、「年齢—余裕」、「ネット—所得」の3種類であるが、これらのエッジが存在することは明らかに不自然とは言い切れない。

矢印の方向性は対象外ではあるものの、PCアルゴリズムで学習されたエッジのうち、約7割のエッジは頑健性が高いものであることが確認された。例えば、本稿の主な分析結果である「保有」と「余裕、選好、預貯金」の関係性、「預貯金」と「世帯所得、負債、年齢」の関係性、「選好」と「知識2」間の関係性などである。ただし、残り3割の関係性、例えば、「ネット」と「選好」間の関係性などは、必ずしも頑健性が高いとは言い切れない分析結果となった。

## 6：まとめ

本稿では、日本の家計におけるリスク性資産保有の有無に関係する因果構造の推定を試みた。どのような要因がリスク性資産の保有に対して重要な要因となるのかを調べるため、できるだけ多様な変数を利用するとともに、機械学習の手法を用いることで、先験的な制約条件を少なくし、データから各変数間の因果構造を推定した。

分析結果からは、リスク性資産の保有に対しては、安全資産（預貯金）の余裕度と、リスクに対する選好の2つの経路が影響を与えていることが確認された。前者の経路は、老後を見据えた預貯金に余剰分がある（と思う）場合に、リスク性資産を保有するものである。低所得や住宅ローンを含む負債の存在などは、預貯金が貯まらない要因となるため、間接的にリスク性資産保有を抑制する要因となる。また、所得・負債・預貯金は、年齢からも影響を受けているため、年齢階級（ライフ・サイクル）とリスク性資産保有率が正の相関を示すことにつながっていると考えられる。後者の経路は、よりリスクに対する選好が高い場合に、保有率が上昇するとの構造である。リスク選好の形成に対しては、金融リテラシーの高さやインターネットの利用状況が影響している可能性が示唆された。ただし、インターネット利用とリスク選好の関係性は、必ずしも頑健性が高いとは言い切れない結果も示された。

資金の好循環に向けたリスク性資産保有率の増加には、期待成長率の上昇や抜本的な年金改革など、老後に備えた資金的余裕があるとの認識につながる政策対応が最も効果的である。現状、日本の家計資産が投資に向かわない主因は、こうした老後の資金不安である可能性が高い。もちろん政策努力はすべきだが、少子高齢化が進む日本経済では、老後の資金的な懸念を完全に払拭することは困難であろう。そのため、次善ではあるが、金融リテラシーを高めることでリスク性資産に対する選好を高めることが重要になる。日本の金融リテラシーは国際的にみても低く改善余地が大きいいため、政府や金融機関などによる積極的な金融教育の推進が必要であろう。正確な金融知識を身に着けることで、各家計に自身にとって最適なリスク許容度を把握させることが重要である。金融知識の必要性を感じていない消極的な家計も多いため、こうした家計に対するアプローチの仕方も大きな課題となる。

なお、効果の面では金融リテラシーより低いかもしれないが、生活設計の促進などもリスク性資産の保有を高める可能性がある。老後を見据えた生活設計の促進などにより、老後に必要な預貯金額を

明確にすることで、安全資産の必要額を過大評価している家計を減らすことができれば、リスク性資産保有率の上昇につながる可能性がある。

最後に今後の課題について簡単に述べる。第1に、本稿は因果構造の推計を機械学習の手法を用いて行ったが、同手法には潜在変数がないとの条件が必要である。観測データでは分析できることに限度があるかもしれないが、例えばランダム化比較試験など、実験データを利用した分析を行うことで、リスク性資産の保有に対する因果構造をより明確に把握することが可能になると想定される。第2に、本稿の分析対象は、リスク性資産保有の有無のみであり、保有した家計がどの程度の額を保有するのかという点については分析していない。この点についても、機械学習の手法などを活用することで、より現実を反映した要因を特定できる可能性がある。ただし、本稿のデータでもリスク性資産を保有するサンプル数は1,100程度であり、サンプル数が限定されてしまうことは、特に機械学習の手法を利用する場合に大きな課題となる。第3に、今回の分析で重要性が指摘された金融リテラシーの向上方法についても、より精緻な分析が求められる。特に金融教育に対して消極的な家計に対して、どのようなアプローチが効果的かを検証することは、重要なテーマであろう。

## 参考文献

- 伊藤雄一郎・瀧塚寧孝・藤原茂章 (2017) 「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—」 日本銀行ワーキングペーパーシリーズ、No. 17-J-2
- 祝迫得夫 (2012) 『家計・企業の金融行動と日本経済』 日本経済新聞出版社
- 祝迫得夫・植杉威一郎・小野有人・清水千弘・直井道生・堀雅博 (2018) 「家計の住宅投資、貯蓄、金融資産選択:2017年「日本家計パネル調査」の結果概要」 Grant-in-Aid for Scientific Research, HIT-REFINED Working Paper Series No. 78
- 北村行伸・内野泰助 (2011) 「家計の資産選択行動における学歴効果」 『金融経済研究』 第33号
- 木成勇介・筒井義郎 (2009) 「日本における危険資産保有比率の決定要因」 『金融経済研究』 第29号
- 金融庁 (2017) 「つみたてNISAについて」 平成29年7月  
<https://www.fsa.go.jp/policy/nisa/20170614-2/12.pdf> (accessed on 25 September 2020)
- 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』 第32巻第2号
- 徳田秀信・齋藤周 (2014) 「住宅保有に伴うリスク資産投資の抑制効果と制度的背景—OECD諸国のパネルデータに基づく実証分析—」 『みずほ総研論集』 2014年II号、1-30
- 中城瑞希・篠潤之介・今久保圭 (2017) 「わが国家計の金融資産選択における行動特性」 日銀レビュー、2017-J-7
- 日本銀行 (2020) 「資金循環の日米欧比較」 2020年8月21日 日本銀行調査統計局
- 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、No. 17-J-4
- 福原敏恭 (2016) 「日米家計のリスク資産保有に関する論点整理」 B0J Report & Research Papers
- 森俊介・坂口純也 (2020) 「顧客本位・金融教育・ナッジで促す家計の資産形成」 『大和総研調査季報』 2020年春季号 vol. 38
- Barsky, R. B., Juster, F. T., Kimball, M. S., and Shapiro, M. D. (1997), "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study", *Quarterly Journal of Economics*, 112(2), pp. 537-579.
- Dixit, S., Chaudhary, M., and Sahni, N. (2020), "Network Learning Approaches to Study World Happiness", *ArXiv, abs/2007.09181*.
- Hartemink, A. J. (2001), "Principled Computational Methods for the Validation Discovery of Genetic Regulatory Networks", Doctoral Dissertation, Massachusetts Institute of Technology
- Hünemann, P., and Bareinboim, E. (2019), "Causal Inference and Data-Fusion in Econometrics", *arXiv: Econometrics*.
- Iwaisako, T., Ono, A., Saito, A., and Tokuda, H. (2019), "Disentangling the Effect of Home Ownership on

- Household Stock-Holdings: Evidence from Japanese Microdata", *RIETI Discussion Paper Series*, 19-E-007
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013), *An Introduction to Statistical Learning*, New York: Springer
- Kalisch, M., Mächler, M., Colombo, D., Maathuis, M., and Bühlmann, P. (2012), "Causal Inference Using Graphical Models with the R Package pcalg", *Journal of Statistical Software*, 47(11), 1 – 26
- Merton, R. C. (1969), "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case", *Review of Economics and Statistics*, 51(3), pp. 247-257.
- Scutari, M. (2020), "Package 'bnlearn'", Reference Manual  
<https://www.bnlearn.com/documentation/bnlearn-manual.pdf> (accessed on 18 September 2020)
- Scutari, M., and Denis, J. B. (2014), *Bayesian Networks with Examples in R*, Chapman and Hall/CRC
- Stekhoven, D. J., and Bühlmann, P. (2012), "MissForest—Non-Parametric Missing Value Imputation for Mixed-Type Data", *Bioinformatics*, 28(1), 112-118.
- van Rooij, M., Lusardi, A., and Alessie, R. (2011), "Financial Literacy and Stock Market Participation", *Journal of Financial Economics*, Elsevier, vol. 101(2), pages 449-472
- Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., Marrero, J., Ji Zhu, J., and Higgins, P. D. (2013), "Comparison of Imputation Methods for Missing Laboratory Data in Medicine", *BMJ open*, 3(8).

## 補論 1 : PC アルゴリズムについて

PCアルゴリズムは、2つのステップから構成されており、スケルトン(skeleton)と呼ばれる各ノードを矢印のないエッジで表現したグラフを推定することが第1のステップとなる。この目的のため、初めにすべてのノードを他の全てのノードとエッジで結び、条件付き独立が検定されたエッジを外していくという手法を採る。本文の図表3のAを例にとると、まず、AとB~Fをすべてエッジで結ぶ。このときAとBは直接的な結び付きがあり、C~Fをどのように条件付けしたとしても、両者は相関がある(独立ではない)と検定されるため、エッジは残したままとなる。一方、AとDの間であれば、Bを条件付けした場合に、両者に相関はなくなる(独立となる)との検定結果が得られるため、AとDの間には直接的な結び付きはないことが判明し、AとDを結ぶエッジは削除できる。以上のプロセスを繰り返すことで、スケルトンが推定される。

次に、エッジの矢印の推計を行う。この目的のため、「 $X - Z - Y$ 」となっている3組の変数を利用し、この3者の関係がcolliderか否かを検定することで行う。通常、3組の変数の関係性として、以下の3つが想定される。

- i) chain :  $X \rightarrow Z \rightarrow Y$  (又は、 $X \leftarrow Z \leftarrow Y$ )
- ii) fork :  $X \leftarrow Z \rightarrow Y$
- iii) collider :  $X \rightarrow Z \leftarrow Y$

このとき、i)とii)には「 $X$ と $Y$ は独立ではないが、 $Z$ を条件付けにした場合は独立 ( $X \not\perp Y$  and  $X \perp Y | Z$ )」、iii)には「 $X$ と $Y$ は独立だが、 $Z$ を条件付けにした場合は独立でない ( $X \perp Y$  and  $X \not\perp Y | Z$ )」との確率的性質が存在することが知られている。iii)のみ性質が異なるため、観測データからiii)の性質を検定することができれば、3組の変数の関係性はcolliderであることが判明し、矢印の向きが決定できる。一方、i)とii)の性質が確認された場合は、chainとforkの双方の可能性があり、矢印の向きは不明である。そのため、同アルゴリズムではすべての因果構造が把握できるとは限らない。

ただし、本文の図表3の例では、観測データから矢印の向きを一様に決定することが可能である。例えば、「 $A - B - D$ 」の関係性を観測データより検定するとcolliderではないと判断されるため、「 $A \rightarrow B \rightarrow D$ 」、「 $A \leftarrow B \leftarrow D$ 」、「 $A \leftarrow B \rightarrow D$ 」の3つの可能性が残ってしまう。一方で、「 $A - B - C$ 」の関係性はcolliderであると検定されるため、AとBの関係性は「 $A \rightarrow B$ 」であることが判明する。この条件を満たす関係性は「 $A \rightarrow B \rightarrow D$ 」のみであるため、BとDの間の矢印の方向性も決定される<sup>37</sup>。

このように、PCアルゴリズムでは、条件付き独立の検定が非常に重要であるが、通常、離散変数又は連続変数への統一が必要である。本稿ではRの「pcalg」のパッケージを利用したが、一般的には連続変数の検定にはスチューデントのt検定 (Student's t-test)、離散変数の検定にはピアソンのカイ二乗検定 (Pearson's  $\chi^2$  test) を利用することが多い。

<sup>37</sup> 同様の考え方により、BとEの間の矢印の方向性は「 $B \rightarrow E$ 」であると決定される。「 $B - E - F$ 」の関係性についても、colliderではない、かつ、 $B \rightarrow E$ である、との条件を満たす関係性は「 $B \rightarrow E \rightarrow F$ 」である。

## 補論2：ランダム・フォレストによる相対的重要度

被説明変数：保有

順位	変数	全サンプル	就業者のみ	平均
1	選好	1.00	1.00	1.00
2	預貯金	0.78	1.00	0.89
3	余裕	0.21	0.18	0.19
4	資産価値	0.13	0.10	0.11
5	給与	0.09	0.12	0.11
6	土地価値	0.11	0.10	0.11
7	負債	0.09	0.11	0.10
8	年齢	0.11	0.08	0.10
9	職業	0.09	0.09	0.09
10	住宅価値	0.08	0.07	0.08
11	世帯所得	0.05	0.10	0.07
12	満足1	0.06	0.07	0.07
13	ネット	0.10	0.03	0.06
14	ローン	0.04	0.06	0.05
15	学歴	0.07	0.03	0.05
16	労働時間	0.07	0.01	0.04
17	知識1	0.03	0.05	0.04
18	証券1	0.03	0.04	0.04
19	年金	0.05	0.02	0.03
20	証券5	0.03	0.04	0.03
21	産業	0.04	0.02	0.03
22	証券2	0.02	0.04	0.03
23	満足2	0.02	0.03	0.03
24	証券3	0.02	0.03	0.03
25	就業	0.04	0.01	0.02
26	知識2	0.02	0.03	0.02
27	不安1	0.02	0.02	0.02
28	不安2	0.02	0.02	0.02
29	住宅	0.02	0.01	0.02
30	証券4	0.01	0.02	0.02
31	把握1	0.01	0.02	0.01
32	把握2	0.01	0.01	0.01
33	保守的	0.01	0.01	0.01
34	合理的	0.00	0.01	0.01
35	計画	0.01	▲ 0.00	0.00
36	割引	0.00	0.00	0.00
37	降水	▲ 0.00	▲ 0.01	▲ 0.00

被説明変数：選好

順位	変数	全サンプル	就業者のみ	平均
1	証券1	1.00	1.00	1.00
2	ネット	0.81	0.58	0.69
3	知識2	0.69	0.69	0.69
4	証券2	0.55	0.67	0.61
5	知識1	0.52	0.65	0.58
6	学歴	0.44	0.27	0.35
7	世帯所得	0.31	0.37	0.34
8	証券3	0.30	0.35	0.33
9	年齢	0.37	0.21	0.29
10	給与	0.26	0.31	0.28
11	土地価値	0.26	0.23	0.24
12	資産価値	0.26	0.22	0.24
13	余裕	0.18	0.14	0.16
14	ローン	0.13	0.19	0.16
15	負債	0.17	0.15	0.16
16	計画	0.10	0.17	0.14
17	職業	0.15	0.09	0.12
18	割引	0.04	0.19	0.12
19	就業	0.13	0.09	0.11
20	満足1	0.10	0.11	0.10
21	住宅価値	0.10	0.11	0.10
22	産業	0.13	0.05	0.09
23	満足2	0.08	0.10	0.09
24	労働時間	0.13	0.00	0.07
25	不安1	0.02	0.09	0.06
26	合理的	0.06	0.05	0.06
27	証券4	0.07	0.03	0.05
28	不安2	0.04	0.06	0.05
29	把握2	0.02	0.07	0.05
30	住宅	0.03	0.04	0.04
31	年金	0.02	0.04	0.03
32	保守的	0.04	0.02	0.03
33	証券5	0.01	0.05	0.03
34	把握1	0.03	0.01	0.02
35	降水	▲ 0.02	▲ 0.01	▲ 0.02

被説明変数：預貯金

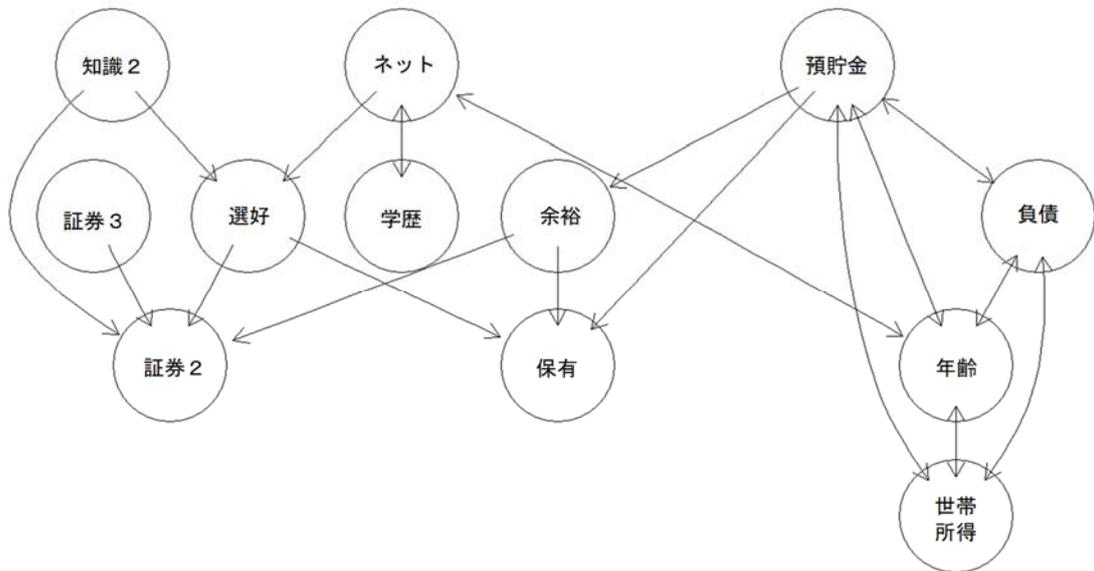
順位	変数	全サン プル	就業者 のみ	平均
1	余裕	1.00	0.86	0.93
2	世帯所得	0.71	1.00	0.85
3	負債	0.69	0.95	0.82
4	年齢	0.52	0.58	0.55
5	給与	0.31	0.24	0.28
6	満足 1	0.20	0.33	0.26
7	ローン	0.20	0.30	0.25
8	資産価値	0.18	0.18	0.18
9	住宅	0.12	0.12	0.12
10	労働時間	0.20	0.03	0.12
11	住宅価値	0.10	0.13	0.11
12	就業	0.15	0.07	0.11
13	土地価値	0.10	0.11	0.10
14	年金	0.10	0.04	0.07
15	職業	0.07	0.07	0.07
16	満足 2	0.05	0.07	0.06
17	ネット	0.07	0.05	0.06
18	学歴	0.05	0.04	0.05
19	証券 5	0.02	0.04	0.03
20	割引	0.02	0.04	0.03
21	産業	0.04	0.01	0.03
22	不安 1	0.02	0.02	0.02
23	不安 2	0.02	0.02	0.02
24	計画	0.02	0.01	0.02
25	知識 2	0.01	0.02	0.01
26	降水	0.01	0.01	0.01
27	保守的	0.02	0.00	0.01
28	把握 2	0.01	0.01	0.01
29	合理的	0.02	▲ 0.00	0.01
30	証券 4	0.00	0.01	0.00
31	証券 1	0.00	0.00	0.00
32	知識 1	▲ 0.00	0.00	▲ 0.00
33	証券 3	0.00	▲ 0.01	▲ 0.00
34	把握 1	▲ 0.00	▲ 0.01	▲ 0.01
35	証券 2	▲ 0.01	▲ 0.00	▲ 0.01

(注) 1. 順位は平均ベース。

2. 重要度がマイナスとなる場合は、同変数の順番を並び替えて値を予測することで、予測誤差がマイナスとなったことを示す。

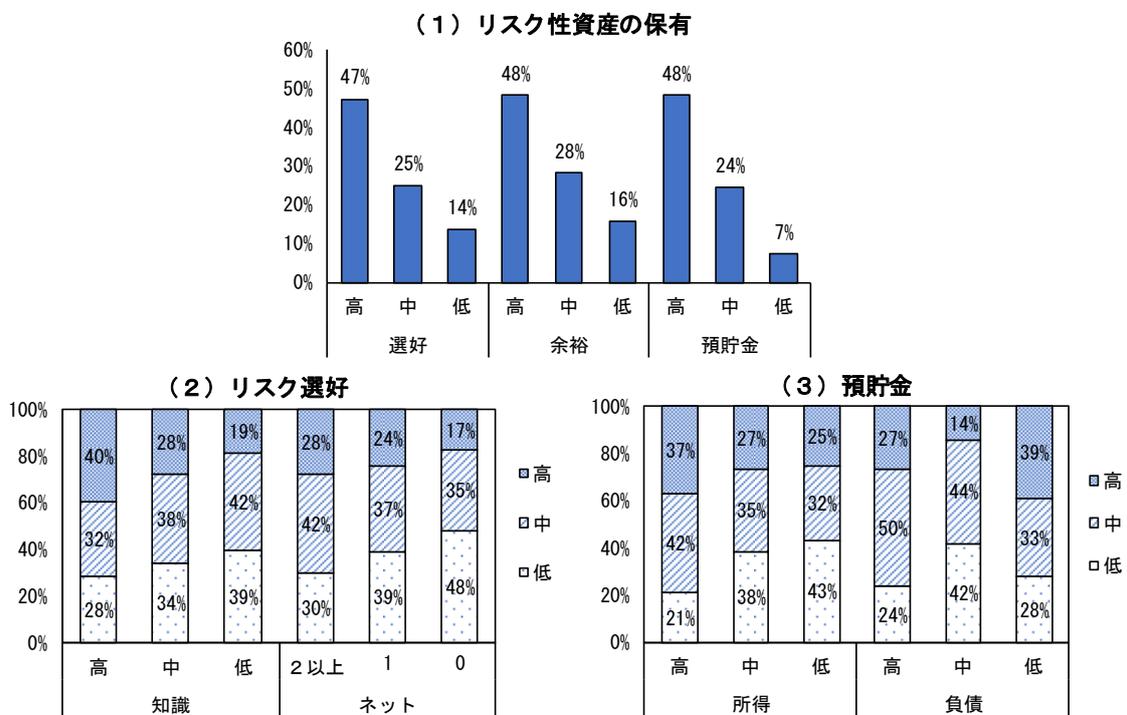
(資料) 慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

### 補論3：ベイジアン・ネットワーク（PC アルゴリズム）



(注) 双方向のエッジは矢印の向きが不明であることを示す。  
 (資料) 慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

### 補論4：条件付き確率（事後分布）



(資料) 慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成

## 補論5：頑健性チェック

本稿では、頑健性チェックのため、6種類のStructure learningのアルゴリズムにより学習したスケルトンを平均し、6モデル中4つ以上のモデルでエッジが確認されたものを頑健性が高い関係とした。分析にはRの「bnlearn」のパッケージを利用した。本稿で利用した6種類のアルゴリズムを整理したのが以下の表になる。

アルゴリズム	タイプ	テスト又はスコア	エッジ	PCアルゴリズムとの比較			備考
				TP	FP	FN	
Hilton Parents and Children	Constraint-based	Test: Mutual Information (Monte Carlo Permutation)	17	14	4	3	alfa=0.05, B=5000
Grow=Shrink	Constraint-based	Test: Mutual Information (Monte Carlo Permutation)	22	16	2	6	alfa=0.05, B=5000
Incremental Association	Constraint-based	Test: Mutual Information (Monte Carlo Permutation)	21	15	3	6	alfa=0.05, B=5000
Hill-Climbing	Score-based	Score: Bayesian Dirichlet (BDe)	27	17	1	10	iss=200, piror="marginal", restart=1500
Tabu Search	Score-based	Score: Bayesian Dirichlet (BDe)	27	18	0	9	iss=200, piror="marginal", tabu=500
Restricted Maximization	Hybrid	Test: Mutual Information Score: BIC	14	14	4	0	Hilton-PC+Tabu Search
平均	—	—	18	15	3	3	threshold=0.65

(注) alfa: type 1 error threshold、B: number of permutations for Monte Carlo tests、iss: imaginary sample size、“marginal”: marginal uniform priors、restart: the number of random restarts、tabu: the length of the tabu list  
(資料) Scutari (2020)、慶應義塾大学「日本家計パネル調査」より、みずほ総合研究所作成