

全国消費実態調査の匿名データを用いた、 2人以上世帯の保険需要の分析

宇野 慧

アステラス製薬株式会社 開発本部 データサイエンス部

要旨：

貯蓄型保険と非貯蓄型保険両方の需要に対して、世帯属性の中でも特に就業状況が与える影響に着目した。データは平成16年度の総務省全国消費実態調査の匿名データを用いた。SUR(Seemingly Unrelated Regression) Tobitモデル推定の結果、就業状況が保険需要に与える影響は貯蓄型と非貯蓄型で大きく異なることが確認できた。

キーワード：全国消費実態調査匿名データ、保険需要、SUR Tobitモデル、NLMIXEDプロシジャ

発表の構成について

- 1: 分析の背景
- 2: 分析に用いたデータの説明、分布状況の確認
- 3: 推定に用いた理論モデルの説明、プログラムの概要
- 4: 推定結果および結果の解釈
- 5: まとめ、参考文献等

1: 分析の背景(1/3)

- 世帯の保険需要については、資産選択の問題として多くの実証分析の蓄積がある
 - しかし、**【貯蓄型】と【非貯蓄型】**という商品特性に着目したものは無い
- ⇒今回、貯蓄型と非貯蓄型の違いを明示的に扱った分析として、両保険の支出額を同時方程式モデルにて推定

1: 分析の背景(2/3)

【貯蓄型保険】

- 保険料を継続的に支払い、契約期間を満了すると満期保険金として一括金額が受領できる
- 満期時点での給付金額が決まっている確定給付型の保険が多く、原則的に満期までの長期的な資産運用の性質を持つ金融商品として、非貯蓄型と比べて月当たりの支払額が高額に設定されている
- 資産運用に関してリスク回避的な世帯の加入率が高く、支払額も高いと考えられる

1.分析の背景(3/3)

【非貯蓄型保険】

- 毎月の掛け捨て型であるため、貯蓄型のような金融資産的な性質は持たない。
- 月当たりの支払額は貯蓄型に比べて安価に設定されていることが多い
- リスクを選好する世帯では、非貯蓄型保険で支出を安価に抑え、その差額分を別の金融商品として資産運用に回す可能性が考えられる

2. 匿名データの特徴(1/2)

- 総務省全国消費実態調査(平成16年度)をベースに、一部加工を行ったものが提供される
- **2人以上世帯(43861世帯)**と、単身世帯(3936)で別の調査票
- 2人以上世帯の調査期間は9月～11月
⇒ボーナス月の影響は除外されている
- 2人以上世帯では、世帯属性データ、支出データ、耐久消費財データなど、1780系列のデータが格納

2. 匿名データの特徴(2/2)

- 匿名性を担保することを目的として、高額貯蓄のトップコーディングや、リサンプリングなど、一部データ加工を実施しているため、**元データの集計結果と完全一致しない**
- しかし、あくまで公的統計データであり、**このデータを用いた分析は正式な実証研究の成果として公表可能**

(疑似マイクロデータとの違い)

2. 匿名データ利用方法(1/2)

- 匿名データは、学術研究の発展又は高等教育の発展に資すると認める場合であって、以下の要件をすべて満たす場合に、提供可能
- 統計の作成又は統計的研究にのみ利用されること。
- 学術研究又は高等教育(学校教育法(昭和22年法律第26号)第1条に規定する大学又は高等専門学校における教育)の用に供することを直接の目的とするものであること。

2.匿名データの利用方法(2/2)

- 学術研究の成果又は教育内容が公表され、社会に還元されること。

【申請に要する期間、費用】

- 利用申請～受け取りまで、審査期間を含め1カ月強は必要
- 費用は1データセットで1万円程度
(※今回の申請ではSAS Institute Japan株式会社様にご負担頂きました)

2.分析に用いる変数(1/2)

- 就業属性(世帯主、世帯の2人目に着目)

変数名	詳細
自営業ダミー	世帯主/2人目が自営業者の場合に1をとるダミー
大企業ダミー	世帯主/2人目が大企業(500人以上)に勤務している場合に1をとるダミー
中小企業ダミー	世帯主/2人目が中小企業(500人未満)に勤務している場合に1をとるダミー
公務員ダミー	世帯主/2人目が公務員の場合に1をとるダミー

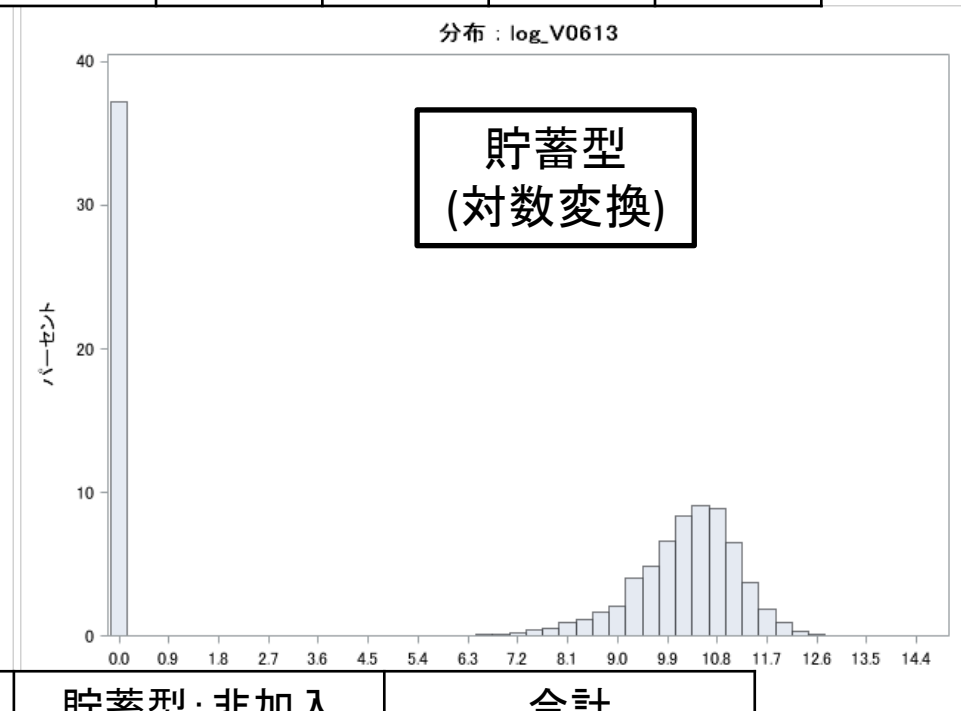
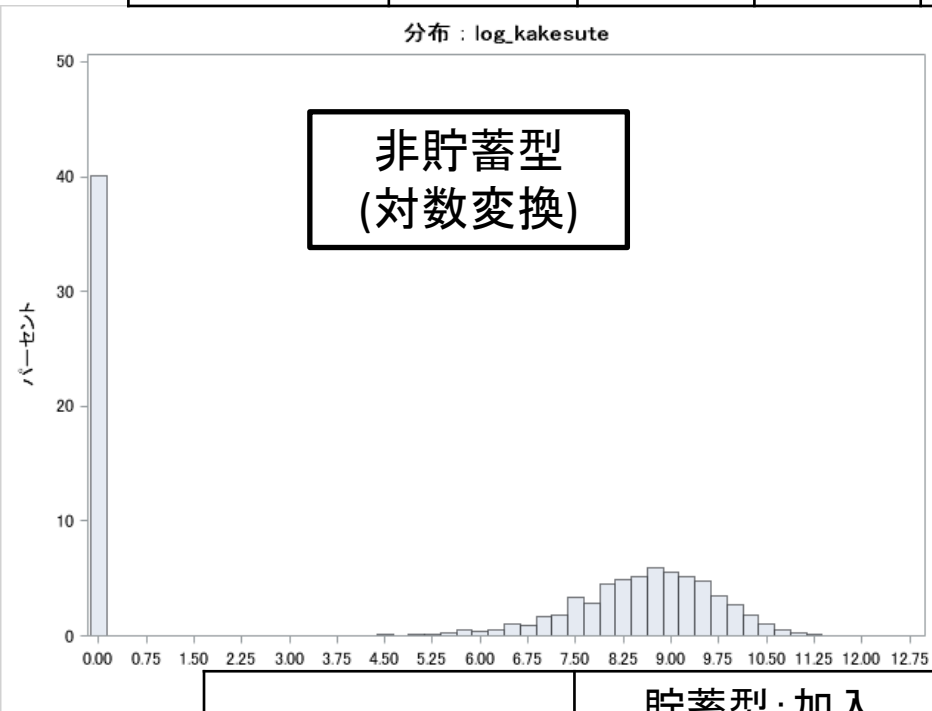
2.分析に用いる変数(2/2)

- その他の世帯属性

変数名	詳細
世帯主年齢	世帯主の年齢(5歳刻みの区分データ)
世帯人員数	世帯人員数
18歳未満人数	世帯員のうち18歳未満の人数
65歳以上人数	世帯員のうち65歳以上の人数
大都市圏ダミー	3大都市圏に居住している場合に1をとるダミー
世帯主女性ダミー	世帯主が女性の場合に1をとるダミー
経常所得対数値	経常所得を自然対数変換した値
貯蓄対数値	貯蓄額を自然対数変換した値

2.基本統計量(支出額>0の部分)

variable	N(>0 sub)	mean	std	min	q1	median	q3	max
非貯蓄型	26289	8.59	1.13	1.20	7.94	8.69	9.37	12.77
貯蓄型	27527	10.23	0.96	5.04	9.71	10.35	10.87	14.83



	貯蓄型:加入	貯蓄型:非加入	合計
非貯蓄型:加入	18906世帯	7383世帯	26289世帯
非貯蓄型:非加入	8621世帯	8951世帯	17572世帯
合計	16334世帯	27527世帯	43861世帯

2.世帯主の就業区分別の基本統計量

- 貯蓄型保険に関する、自営業の顕著な傾向

非貯蓄型保険	世帯数	加入世帯数	加入率(%)	平均値(>0)	標準偏差(>0)
非就業・その他	9855	4757	48.3	8.47	1.12
自営業	7060	3192	45.2	8.67	1.17
大企業	6479	4923	76.0	8.60	1.12
中小企業	15025	9280	61.8	8.60	1.11
公務員	5442	4137	76.0	8.59	1.15

貯蓄型保険	世帯数	加入世帯数	加入率(%)	平均値(>0)	標準偏差(>0)
非就業・その他	9855	6076	61.7	10.05	1.02
自営業	7060	6	0.1	9.88	0.35
大企業	6479	5567	85.9	10.26	0.92
中小企業	15025	11046	73.5	10.22	0.94
公務員	5442	4832	88.8	10.47	0.92

2. 基本統計量まとめ(1/3)

- 非貯蓄型・貯蓄型の両方に加入している世帯が40%程度、非貯蓄型・貯蓄型のいずれか一方に加入している世帯がそれぞれ20%程度、いずれの保険にも加入していない世帯が20%
- 非貯蓄型保険では平均値が8.59(指数変換すると約5400円)の比較的正規分布に近い分布となる。また、貯蓄型保険についても、平均値が10.23(指数変換すると約27700円)の比較的正規分布に近い分布となる

2.基本統計量まとめ(2/3)

【非貯蓄型保険】

- 就業状況に応じて加入率が異なり、非就業や自営業に比べ、大企業、中小企業、公務員の加入率が高い
- 加入世帯の平均値に着目すると、非就業に比べ他の世帯が若干高い傾向

2.基本統計量まとめ(3/3)

【貯蓄型保険】

- 自営業世帯が突出して加入率が低い
- 加入世帯の平均値についても、非就業と比較して大企業、中小企業、公務員の世帯では支払額が若干高い傾向

⇒上記の傾向を、モデル推定で検証する

3. 推定モデル(1/4)

- 保険料は負の値を取らないため、0での打ち切り分布となる
⇒貯蓄型、非貯蓄型それぞれの決定を扱う、連立Tobitモデルで分析
- ただし、同時方程式モデルではなく、誤差の分散構造のみに関係性を盛り込んだSUR(Seemingly Unrelated Regression) Tobitモデルで分析
- 負の値を取りうる潜在的な需要(y^*)を用いて、次ページのようにモデル設定を記述。

3. 推定モデル(2/4)

$$\begin{cases} y_1^* = X\beta + \varepsilon_1 \\ y_1 = \max(0, y_1^*) \end{cases}$$

$$\begin{cases} y_2^* = X\gamma + \varepsilon_2 \\ y_2 = \max(0, y_2^*) \end{cases}$$

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{pmatrix} \right)$$

3. 推定モデル(3/4)

- 各保険への支出額(0、正)によって、次ページの4通りの尤度関数となる
- Φ 、 ϕ はそれぞれ標準正規分布の累積分布関数、密度関数

3. 推定モデル(4/4)

if $y_1 = 0$ and $y_2 = 0$ then

$$L = \Phi_2 \left(\frac{X\beta}{\sigma_{11}}, \frac{X\gamma}{\sigma_{22}}, \frac{\sigma_{12}}{\sigma_{11}\sigma_{22}} \right)$$

if $y_1 = 0$ and $y_2 > 0$ then

$$L = \frac{1}{\sigma_{22}} \phi \left(\frac{y_2 - X\gamma}{\sigma_{22}} \right) \left[1 - \Phi \left(\frac{1}{\sqrt{1 - (\sigma_{12}/\sigma_{11}\sigma_{22})^2}} \left(\frac{X\beta}{\sigma_{11}} + \frac{\sigma_{12}}{\sigma_{11}\sigma_{22}} \frac{y_2 - X\gamma}{\sigma_{22}} \right) \right) \right]$$

if $y_1 > 0$ and $y_2 = 0$ then

$$L = \frac{1}{\sigma_{11}} \phi \left(\frac{y_1 - X\beta}{\sigma_{11}} \right) \left[1 - \Phi \left(\frac{1}{\sqrt{1 - (\sigma_{12}/\sigma_{11}\sigma_{22})^2}} \left(\frac{X\gamma}{\sigma_{22}} + \frac{\sigma_{12}}{\sigma_{11}\sigma_{22}} \frac{y_1 - X\beta}{\sigma_{11}} \right) \right) \right]$$

if $y_1 > 0$ and $y_2 > 0$ then

$$L = \frac{1}{\sqrt{\sigma_{11}^2 \sigma_{22}^2 - \sigma_{12}^2}} \phi_2 \left(\frac{y_1 - X\beta}{\sigma_{11}}, \frac{y_2 - X\gamma}{\sigma_{22}}, \frac{\sigma_{12}}{\sigma_{11}\sigma_{22}} \right)$$

3. 推定プログラム概要(1/2)

- 詳細は別添を参照されたい

```
proc nlmixed data=data tech = NEWRAP;
/*パラメータの初期値設定*/
parms b100 ~ b230 s11 ~ s22; bounds s11 s22 > 0; pi=atan(1)*4;
y1 = log_kakesute; y2 = log_V0613;

/*右辺を定義する*/
xbeta = b100 + b101*Jiei1 +... + b130*log_saving ;
xgamma = b200 + b201*Jiei1 +... + b230*log_saving ;

/*誘導系のモデル式を定義する*/
e1 = y1 - xbeta; e2 = y2 - xgamma;
/*分散共分散行列の行列式を定義する*/
det_sig = s11**2*s22**2 - s12**2;
```


3. 推定プログラム概要(2/2)

```
/*尤度関数を定義*/  
if y1 = 0 and y2 = 0 then  
    ll = log(probbnrm(xbeta/s11,xgamma/s22,s12/(s11*s22)));  
else if y1 = 0 and y2 > 0 then  
    ll = -log(s22) + log(PDF('NORMAL',e2/s22, 0,1))  
        + log(1-CDF('NORMAL',1/(1-(s12/(s11*s22))**2)**0.5  
        * (xbeta/s11+(s12/(s11*s22))*e2/s22),0,1));  
else if y1 > 0 and y2 = 0 then  
    ll = -log(s11) + log(PDF('NORMAL',e1/s11, 0,1))  
        + log(1-CDF('NORMAL',1/(1-(s12/(s11*s22))**2)**0.5  
        * (xgamma/s22+(s12/(s11*s22))*e1/s11),0,1));  
else if y1 > 0 and y2 > 0 then  
    ll = -1/2*log(2*pi) -1/2*log(det_sig) -1/det_sig  
        *(s22*e1**2 - 2*s12*e1*e2 + s11*e2**2);  
model log_V0613 ~ general(ll);  
run; /*尤度関数の最大化を定義*/
```

推定上のTips

- 初期値は、連立方程式を構成する各式に対してTobit推定した推定値を用いる
- デフォルトの収束基準ではLocal maxに落ちてしまう可能性が高いので、QTOLなど厳し目の収束基準を設定する
- 推定が上手くいかない場合、スケールパラメータの役割をしている分散パラメータの値を変更することで、経験上うまくいくことが多い
- 2変量正規分布の累積分布はprobbnrmで計算可能だが、3変量以上の場合には自前で数値積分の計算が必要なため、難易度が上がる

推定結果を考察する際の注意点

- 線形モデルであれば、被説明変数が対数値の場合の推定値は
 - 離散変数: カテゴリーの基準に対する比
 - 連続変数: 弾力性(説明変数が対数値の場合)
変化率(説明変数が非対数値の場合)
- Tobitモデルでは加入/非加入の2値選択と、加入額の量的選択を同一パラメータで推定しているため、線形モデルのような数値の解釈が出来ない
- NLMIXEDはt分布に基づくp値を返してくるため、小標本の場合は調整が必要

4. 推定結果(1/2)

- 世帯主、2人目の就業属性の影響

(※比較対象は無職世帯)

	非貯蓄型保険			貯蓄型保険		
	Estimate	Standard Error	Pr > t	Estimate	Standard Error	Pr > t
世帯主自営業ダミー	2.238	0.146	<.0001	-3.350	0.113	<.0001
世帯主大企業ダミー	1.225	0.089	<.0001	0.305	0.070	<.0001
世帯主中小企業ダミー	0.502	0.079	<.0001	-0.132	0.062	0.033
世帯主公務員ダミー	1.224	0.091	<.0001	0.485	0.071	<.0001
2人目自営業ダミー	-0.165	0.118	0.163	0.369	0.093	<.0001
2人目大企業ダミー	0.326	0.095	0.001	0.046	0.075	0.538
2人目中小企業ダミー	0.169	0.054	0.002	0.114	0.042	0.007
2人目公務員ダミー	0.447	0.085	<.0001	0.380	0.066	<.0001

4. 推定結果(2/2)

- その他世帯属性の影響

	非貯蓄型保険			貯蓄型保険		
	Estimate	Standard Error	Pr > t	Estimate	Standard Error	Pr > t
世帯主年齢	0.050	0.013	<.0001	0.046	0.010	<.0001
世帯人員数	0.127	0.027	<.0001	0.173	0.021	<.0001
18歳未満人数	-0.138	0.035	<.0001	0.034	0.027	0.207
65歳以上人数	-0.380	0.038	<.0001	-0.021	0.029	0.483
大都市圏ダミー	0.100	0.044	0.023	-0.387	0.035	<.0001
世帯主女性ダミー	-0.582	0.085	<.0001	-0.273	0.067	<.0001
経常所得対数値	-0.027	0.010	0.005	0.433	0.008	<.0001
貯蓄対数値	0.018	0.005	0.000	0.022	0.004	<.0001

4. 結果のまとめと解釈(1/4)

【世帯主就業属性】

- 非貯蓄型保険では、無職世帯以外は高い保険需要で、特に自営業が突出して高い。
⇒ 自営業で貯蓄型を敬遠する傾向
- 貯蓄型保険では逆に、自営業が著しく低い。
⇒ 掛け捨てタイプの方が何らかの制度上で優遇される可能性が考えられる
- 全般的に、公務員世帯は高い保険需要。

4. 結果のまとめと解釈(2/4)

【2人目就業属性】

- 世帯主の就業状況と比べると影響は軽微
- 非貯蓄、貯蓄とも、公務員共働き世帯では高い保険需要。
- 世帯主の結果とは異なり、自営業世帯では貯蓄型の需要がやや高まる。

4. 結果のまとめと解釈(3/4)

【その他の世帯属性】

- 年齢につれ保険料支払い額が高まり、その傾向は非貯蓄型、貯蓄型で同程度
 - 世帯人数が多いほど、保険支出は多い
 - 子供が多い世帯では、非貯蓄型の支払額が少ない。また、老人が多い世帯では掛捨ては少なくなるが、貯蓄性については顕著な傾向なし
- ⇒ 保険商品以外で、長期的な資産形成を図る可能性
- 大都市圏では貯蓄型の需要が低い

4. 結果のまとめと解釈(4/4)

【その他の世帯属性】

- 世帯主が女性の世帯では、非貯蓄・貯蓄とも低水準
- 高所得世帯では、掛捨てよりも貯蓄型を志向
- 貯蓄額は非貯蓄、貯蓄のいずれも有意

5. 今後の展望

- 今回の分析で明らかになった、自営業世帯の保険加入行動の独自性について、さらに検証を行いたい。
- 今回の分析ではSUR Tobitモデルで推定を行ったが、以下の問題点がある
 - 同時決定モデルでは無いことから、資産選択の同時性を考慮出来ていない点
 - 加入・非加入の2値選択と、加入額の量的選択の傾向を同一パラメータで推定している点
 - 人数や年齢など、カテゴリー変数にすべきところを連続量として分析
- これらを考慮する方法はあるが、さらに複雑な尤度関数を記述することになり、推定が困難なため、今後の課題として