

世帯主の就業状況が貯蓄性保険需要 に与える影響についての考察 ～NLMIXEDによるTobit/Hurdleモデル推定～

テルモ株式会社臨床開発部

宇野 慧

Satoshi_Uno@terumo.co.jp

コンテストについて

- SAS Institute Japan株式会社様主催
- 教育用擬似マイクロデータを利用した分析のコンテスト
 - 規定課題：事前に提示された度数表、集計表の結果を再現
 - 自由課題：自由な分析(用いるのは擬似マイクロデータのみ)
- 39団体45名がエントリー（年齢制限30歳まで）
 - カテゴリーA(使用経験5年以上)
 - カテゴリーB(使用経験2年以上5年未満)
 - カテゴリーC(使用経験2年未満)
- 審査プロセス
 - 論文審査：規定課題と自由課題の論文を非公開で審査
 - 口頭審査：優秀賞受賞者の中から最優秀賞を決定(公開審査)

自由課題

世帯の就業属性

■ 企業区分

⇒ 民間 or 公務員 (or 自営業、無職←今回は無し)

■ 企業規模

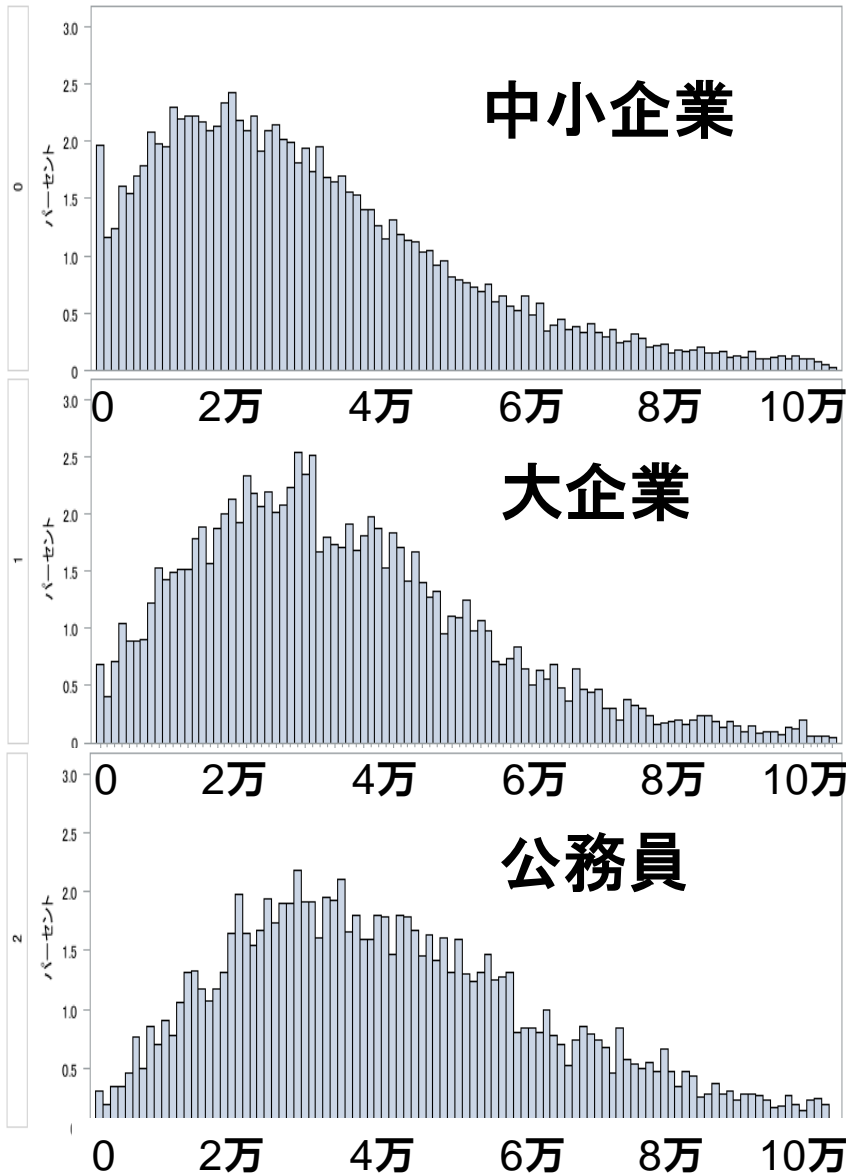
⇒ 中小企業ダミー(499人以下)
大企業ダミー(500人以上)
公務員ダミー

中小企業を基準に、
大企業・公務員の
特性を検証！

■ 産業区分

⇒ 不詳が多数のため使用せず

就業属性別保険料支払額の分布



単位:千円	全体	中小企業	大企業	公務員
世帯数	32027	15607	6999	6658
最小値	0.00	0.00	0.00	0.00
第1四分位	15.82	13.36	18.53	24.29
中央値	28.52	25.29	30.24	38.54
第3四分位	45.57	41.54	45.33	56.31
最大値	1020.35	840.79	707.02	1008.60
平均値	34.85	31.83	35.29	44.59
標準偏差	32.17	30.61	28.22	33.88

■支払額は
公務員 > 大企業 > 中小企業

■中小企業ではゼロに値が
集中している

Tobitモデル/Hurdleモデルの背景

- 保険の潜在需要は**マイナスの値**も考えられる
ex.怪我をした場合に通常額以上を支払い、
逆に平常時は保険料相当額を受け取るような保険
- しかし、実際の保険料支払額では**ゼロが下限値**
⇒以下のようにモデルを記述

$$\begin{cases} y^* = X\beta + u & u | X \sim N(0, \sigma^2) \\ y = \max(0, y^*) \end{cases}$$

Tobitモデル

- y の条件付き密度関数を、標準正規分布の累積分布関数、密度関数、インディケータ関数を用いて以下のように構成

$$f(y_i | X_i) = \left\{1 - \Phi\left(\frac{X_i\beta}{\sigma}\right)\right\}^{1[y_i=0]} \times \left\{\left(\frac{1}{\sigma}\right)\phi\left(\frac{y_i - X_i\beta}{\sigma}\right)\right\}^{1[y_i>0]}$$

※ Φ, ϕ は標準正規分布の累積分布関数、確率密度関数

※※0時点でのみ左側打切りが生じるパラメトリック生存時間モデルで、正規分布を仮定したものと等価

- モデルの欠点：

保険加入の2値選択と支払額の量的選択を同一のパラメータによって推定

⇒制約を緩めたHurdleモデルで検討

Hurdleモデル

■ Hurdle(正規)モデル

$$f(y_i | X_i) = \{1 - \Phi(X_i\gamma)\}^{1_{[y_i=0]}} \times \left\{ \left(\frac{1}{\sigma} \right) \left(\frac{\Phi(X_i\gamma)}{\Phi\left(\frac{X_i\beta}{\sigma}\right)} \right) \phi\left(\frac{y_i - X_i\beta}{\sigma}\right) \right\}^{1_{[y_i>0]}}$$

■ Hurdle(対数正規)モデル (今回の主たるモデル)

$$f(y_i | X_i) = \{1 - \Phi(X_i\gamma)\}^{1_{[y_i=0]}} \times \left\{ \left(\frac{1}{\sigma} \right) \left(\frac{\Phi(X_i\gamma)}{y_i} \right) \phi\left(\frac{\log(y_i) - X_i\beta}{\sigma}\right) \right\}^{1_{[y_i>0]}}$$

Hurdleモデル推定上の工夫

■ 尤度式は1本だが、

- ・ 加入/非加入の2値選択
- ・ 加入後の支払額(連続量)の選択

2つの選択間の独立性を仮定

■ ⇒ ① γ をプロビットモデルで推定

② $y > 0$ のサブサンプルについて、 β, σ^2 を推定

①・②から、最尤推定量が求まる

(詳細はWooldridge(2010)などを参照)

その他の世帯属性(①所得・住宅関連属性)

変数名	変数の定義	効果の予想
経常収入	Youto004を1000円単位に変換	+
住宅ローン 支払額	Youto178を1000円単位に変換	+
住宅ローン 完済ダミー	持家世帯(Shoyuuが1もしくは2)で、かつ住宅ローン 支払額がゼロの場合1を取るダミー変数	-
持家 戸建住宅 ダミー	持家世帯(Shoyuuが1もしくは2)で、かつ戸建住宅に 居住している世帯(Tatekata=1)	+
持家 共同住宅 ダミー	持家世帯で(Shoyuuが1もしくは2)で、かつ共同住宅 に居住している世帯(Tatekata=2, 3, 4, 5, 6)	±

その他の世帯属性(②世帯人員属性)

変数名	変数の定義	効果の予想
女性ダミー	S1_Sexが2の場合に1を取るダミー変数	±
年代ダミー	世帯主年代(S1_Age)を各10歳刻み(60歳以上は1区分)にし、各年代でダミー変数を作成	+
就業人員数 ダミー	就業人員(ShuugyouJinin)が1名、2名以上の世帯を1区分とするダミー変数をそれぞれ作成	+
非就業人員数 ダミー	世帯人員-就業人員を非就業人員として定義し、0名、1名、2名以上の各区分のダミー変数を作成	+

推定プログラムの概要

【Tobitモデル】

```
PROC NLMIXED TECH=NEWWRAP ;
```

```
/*パラメータ設定*/
```

```
PARMS b0 – b14 sigma ;
```

```
BOUNDS sigma > 0;
```

```
/*デザイン行列設定*/
```

```
xbeta      = b0+b1*Daikigyout+ ... + b14*NoLoan;
```

```
xgamma     = g0+g1*Daikigyout+ ... + g14*NoLoan;
```

```
/*尤度関数*/
```

```
【Hurdle(対数正規モデル)】
```

```
IF Youto174m=0 THEN ll = log(1-CDF('NORMAL',xgamma,0,1));
```

```
ELSE ll = log( CDF('NORMAL',xgamma,0,1))
```

```
+ log( PDF('NORMAL', log(Youto174m), xbeta,sigma))
```

```
- log(sigma) - log(Youto174m) ;
```

```
MODEL Youto174m ~ general(ll); RUN;
```

$$f(y_i | X_i) = \{1 - \Phi(X_i \gamma)\}^{1[y_i=0]} \times \left\{ \left(\frac{1}{\sigma} \right) \left(\frac{\Phi(X_i \gamma)}{y_i} \right) \phi \left(\frac{\log(y_i) - X_i \beta}{\sigma} \right) \right\}^{1[y_i>0]}$$

推定結果(就業属性)

	Tobit	Hurdle (Normal)	Hurdle (Log-Normal)
大企業ダミー	-2.889	-3.083	-0.006
	(0.307)	(0.360)	(0.009)
	<.0001	<.0001	0.495
公務員ダミー	4.929	5.711	0.209
	(0.311)	(0.350)	(0.009)
	<.0001	<.0001	<.0001
観察数	20393	20269	20393
対数尤度	-153883	-152952 <small>(部分尤度)</small>	-78340

- 3モデル中では、Hurdle(Log-Normal)が最もフィット
- 公務員世帯の高い保険需要
- 大企業世帯の特性は不明瞭

(中小企業と差が無い可能性)

推定結果(収入・支出)

	Tobit	Hurdle (Normal)	Hurdle (Log-Normal)
経常収入	0.051	0.058	0.001
	(0.001)	(0.001)	(0.000)
	<.0001	<.0001	<.0001
住宅ローン 支払額	0.141	0.154	0.002
	(0.003)	(0.003)	(0.000)
	<.0001	<.0001	<.0001
住宅ローン 完済ダミー	8.097	9.239	0.083
	(0.464)	(0.531)	(0.013)
	<.0001	<.0001	<.0001

- ・ ローン完済世帯が予想に反して高い保険需要
- ・ 経常収入・ローン支払額については予想と整合

推定結果(居住住宅属性)

	Tobit	Hurdle (Normal)	Hurdle (Log-Normal)
持家戸建 住宅ダミー	-1.531	0.354	0.229
	(0.446)	(0.567)	(0.013)
	0.001	<.0001	<.0001
持家共同 住宅ダミー	-10.012	-10.475	-0.046
	(0.903)	(1.132)	(0.026)
	<.0001	<.0001	0.071

- ・戸建てがTobitとHurdleで符号逆転

⇒保険加入選択と支払額選択の間で異なる方向性

(加入しにくい、した場合は手厚く保険をかける)

推定結果(世帯主属性)

	Tobit	Hurdle (Normal)	Hurdle (Log-Normal)
女性ダミー	2.646	3.009	-0.183
	(0.653)	(0.780)	(0.018)
	<.0001	<.0001	<.0001
40代ダミー	6.749	8.877	0.215
	(0.356)	(0.427)	(0.010)
	<.0001	<.0001	<.0001
50代ダミー	11.105	13.553	0.274
	(0.388)	(0.461)	(0.011)
	<.0001	<.0001	<.0001
60代以上 ダミー	12.752	15.476	0.226
	(0.536)	(0.639)	(0.015)
	<.0001	<.0001	<.0001

推定結果(世帯人員属性)

	Tobit	Hurdle (Normal)	Hurdle (Log-Normal)
就業2人以上 ダミー	3.175 (0.282) <.0001	3.932 (0.326) <.0001	0.122 (0.008) <.0001
非就業1人 ダミー	-4.296 (0.314) <.0001	-5.262 (0.366) <.0001	-0.169 (0.009) <.0001
非就業2人 以上ダミー	6.277 (0.413) <.0001	7.387 (0.472) <.0001	0.201 (0.012) <.0001

- ・ 非就業人員に関する効果の非単調性は解釈困難

(1人の場合は高齢者で、2人以上は子供など、何らかの特殊事情の可能性)

まとめと今後の課題

■ まとめ

- 公務員世帯のリスク回避的な特徴が、モデル解析によって示唆された
- 世帯特性の情報が不十分で、結論の妥当性は検証困難
例：
 - ・ 預貯金、住宅評価額等のストックに関する情報
 - ・ 世帯人員構成(年齢、就労状況)、最終学歴など

■ 今後の課題

- 実データでの検証
世帯属性など、今回のデータに含まれない情報の影響を考慮

Hurdle(対数線形モデル) SASプログラム

```
/*Hurdle model(Two-part model) Log-Normal dist. at y>0*/
/*TechでのNewton-Raphson法の指定は必須。デフォルトでは収束しない*/
proc nlmixed data=data_mix MAXIT=1000 TECH=NEWWRAP QTOL=1e-14 ABSGCONV=1e-14 GCONV=1e-14;

/*初期値の設定*/
parms
/*対数線形回帰の結果から指定*/
  b0 1.021125344 b1 -0.05916142 b2 0.209038522 b3 0.001270675 b4 -0.182680838 b5 0.228598369
  b6 -0.046155271 b7 0.214494366 b8 0.274243173 b9 0.225452828 b10 0.121997928 b11 -0.168474518
  b12 0.201046897 b13 0.002220714 b14 0.082524129
/*プロビット回帰の結果から指定*/
  g0 -0.8694 g1 -0.1636 g2 -0.313 g3 -0.0025 g4 0.3169 g5 -0.405 g6 0.4648 g7 -0.2194
  g8 -0.5269 g9 -0.2278 g10 -0.1532 g11 -0.1998 g12 -0.5694 g13 0.0016 g14 0.8906
/*分散はある程度適当に指定*/
  sigma 0.4;
/*分散に非負制約を付ける*/
  bounds sigma > 0;

/*デザイン行列を作成*/
  xbeta = b0 + b1*Daikigyou + b2*Kankouchou + b3*Youto004m + b4*Women + b5*MochiieKodate +
    b6*MochiieKyoudou + b7*Age40s + b8*Age50s + b9*Age60o + b10*ShuugyouJinin2o + b11*huyou1 +
    b12*huyou2o + b13*Youto178m + b14*NoLoan;
  xgamma = g0 + g1*Daikigyou + g2*Kankouchou + g3*Youto004m + g4*Women + g5*MochiieKodate +
    g6*MochiieKyoudou + g7*Age40s + g8*Age50s + g9*Age60o + g10*ShuugyouJinin2o + g11*huyou1 +
    g12*huyou2o + g13*Youto178m + g14*NoLoan;
/*尤度式を指定(今回のモデルでは、最初から対数尤度で記述した方が良い)*/
  IF Youto174m=0 THEN ll = log(1-CDF('NORMAL',xgamma,0,1));
    ELSE ll = log( CDF('NORMAL',xgamma,0,1))+ log( PDF('NORMAL', LOG(Youto174m), xbeta,sigma))
    -log(sigma) -log(Youto174m) ;
  MODEL Youto174m ~ general(ll); run;
```

Hurdle(対数線形モデル) Rプログラム

β (対数線形回帰)パラメータ、 γ (プロビット)パラメータ、 σ (分散)パラメータの初期値を予め指定した上で、最適化関数optimを用いる。

```
opt1=optim(p,function(p){ #pベクトルで $\sigma$ 、 $\beta$ 、 $\gamma$ パラメータを格納
  sigma <- p[1]
  beta <- matrix(p[c(2,3,4,5,6,7,8,9,10,11,12,13,14,15)], ncol=1)
  gamma = matrix(p[c(16,17,18,19,20,21,22,23,24,25,26,27,28,29)], ncol=1)
  xb = X1%*%beta
  xg = X1%*%gamma
  l1=(y<=0)*log(1-pnorm(xg))
  l1=l1[!is.nan(l1)]
  l2= (y>0)*log(dnorm((log(y) - xb)/sigma)*pnorm(xg)/(y*sigma))
  l2=l2[!is.nan(l2)]

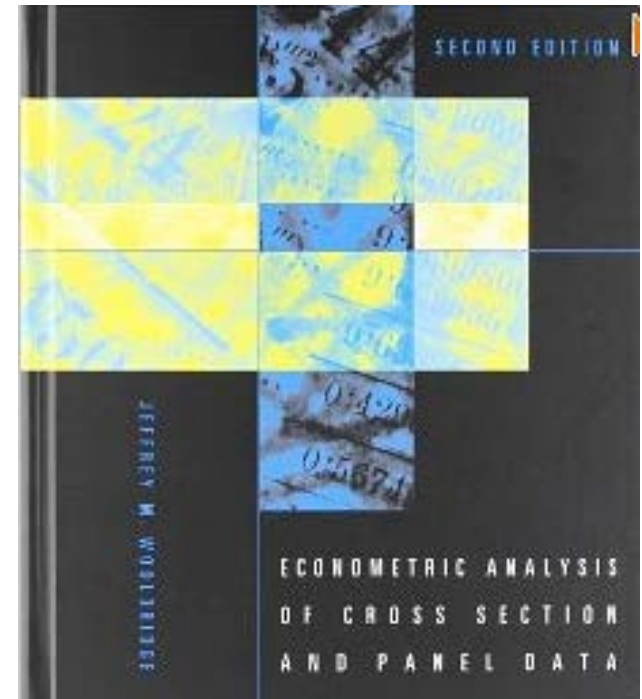
  ll=-1*sum(l1) -1*sum(l2) #対数尤度関数
  return(ll)
},method="Nelder-Mead", hessian=TRUE)

se1=sqrt( diag(solve(opt1$hessian)) ) #推定したヘッセ行列から、標準誤差を計算
par1=opt1$par
z1=par1/se1
```

※optim関数はパラメータ推定値しか返さないため、検定統計量の計算は自分で行う必要がある。また、収束の精度は不明(SASと同様に、初期値を変更する等の試行をしたが、収束せず)。

参考文献

- ▪ Liu, W. S. and Cella, J., (2008), Count Data Model in SAS® SAS Global Forum, Paper 371-2008
- ▪ Wooldridge, J., (2010), *Econometric Analysis of Cross Section and Panel Data* (2nd ed.), MIT Press





Thank you !

