

大学中退の逐次意思決定モデルの構造推定

村澤 康友（甲南大学）

2022年3月7日

「実証的なモラルサイエンス」研究集会

報告の流れ

研究の動機と貢献

先行研究

大学中退モデル

構造推定

モンテカルロ実験

実証分析

反実仮想分析

研究の動機

(大学生の) 成績データの分析

- ・ 鹿野・高木・村澤 (2011)
- ・ 大学における $IR = EBPM$

動的離散選択モデルの構造推定

- ・ 2 値選択（在学／退学）の最適停止モデルが一番簡単
- ・ 退学=授業料収入の損失（大学にとって重要！）

論文の目的

- ・大学中退行動を最適停止問題として定式化し
- ・某大学某学部 2016 年 4 月入学者のうち男子 301 名の
入学後 4 年間の成績データを用いて
- ・意思決定モデルの構造母数の推定と反実仮想分析を試みる

「動的離散選択モデルの構造推定」の大学中退行動への応用

1. 「賃金構造基本統計調査」から計算した退学後の生涯賃金を観測可能な退学の価値とみなして短期パネルでの構造母数の識別を確保
2. CCP 法を修正し、2 値ロジット・モデルの完全分離の問題を回避
3. 個別の母数や母数間の関係について、モンテカルロ実験で推定精度を確認
4. (構造母数の推定精度が低いため) 効用関数の識別を必要としない反実仮想分析

報告の流れ

研究の動機と貢献

先行研究

大学中退モデル

構造推定

モンテカルロ実験

実証分析

反実仮想分析

海外の先行研究

Arcidiacono (2004, J. Econom)

- NLS72 の個票データ
- 高校卒業時・大学3年進級時・大学卒業後の3期間

Stange (2012, AEJ: Applied Economics)

- NELS の個票データ
- 高校卒業時から大学4年進級時までの4期間

Stinebrickner and Stinebrickner (2014, J. Labor Econ.)

- Berea College の 2001 年度入学生 341 名
- 1 年次前期から卒業までの 9 期間 (意思決定は 1 年次後期開始時から 4 年次後期開始時までの 7 回)
- 学期毎のアンケート調査と成績データを接合

大学中退行動の実証研究（分析対象による分類）

標本調査 大学中退行動の規定要因

大学・学部を観測単位 中退率の規定要因

集計時系列データ 中退率の変動要因

特定の大学・学部の学生 個別の学生の中退予測

報告の流れ

研究の動機と貢献

先行研究

大学中退モデル

構造推定

モンテカルロ実験

実証分析

反実仮想分析

大卒プレミアム

日本的一般労働者の生涯賃金（退職金を含めない）

	男	女
大卒	2億6910万円	2億1730万円
高卒	2億1280万円	1億5230万円
プレミアム	5630万円	6500万円

分析の枠組み

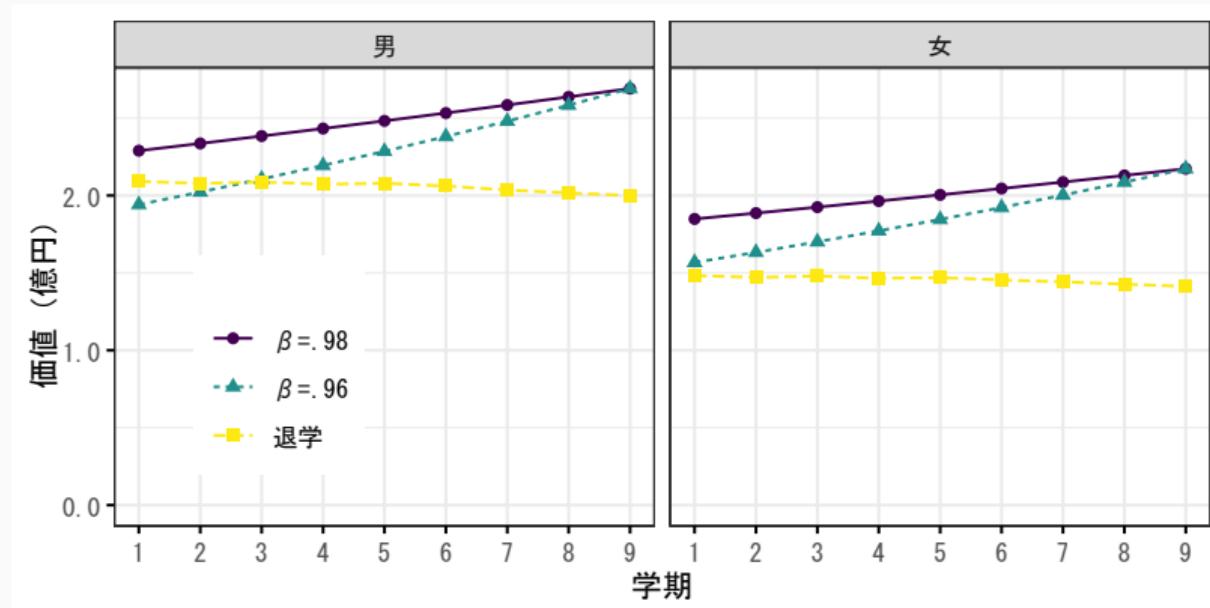
簡単な例

- ・ 入学者は退学しなければ 4 年後に自動的に卒業
- ・ 在学の効用／費用は 0
⇒在学の価値 = 大卒の生涯賃金の割引現在価値
- ・ 退学の価値 = その時点以降の高卒の生涯賃金
- ・ 割引因子 $\beta = .98$ (半期)

各学期の開始時に在学者は価値が高い方の行動を選択

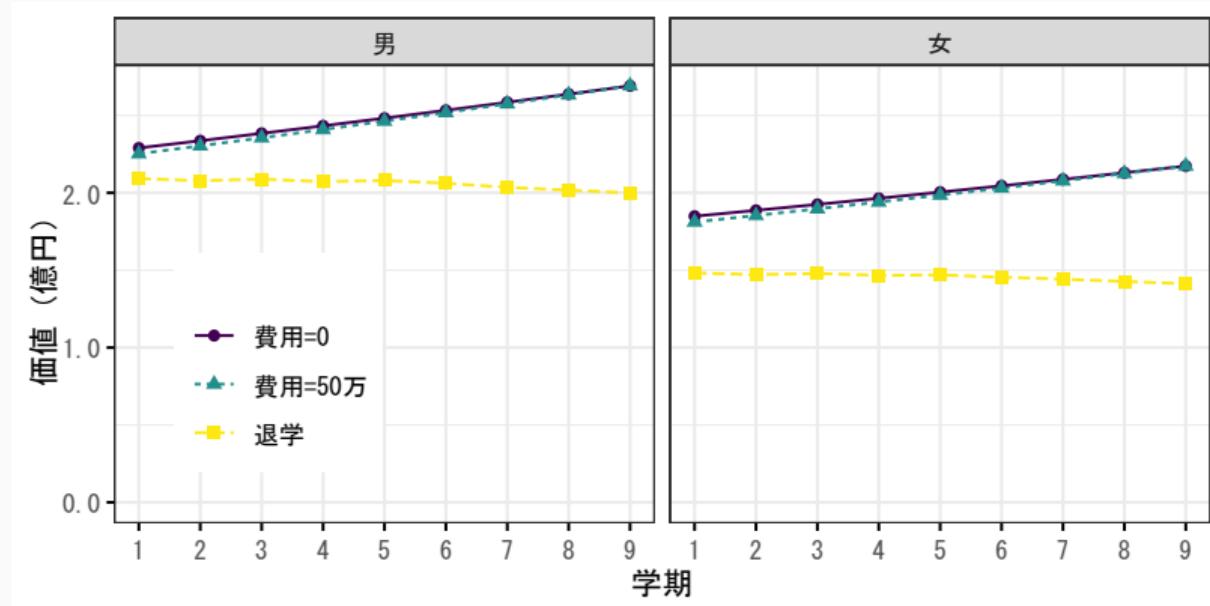
割引因子の低下の効果

男子は $\beta = .96$ なら 1 年次は退学を選択



在学費用の発生の効果

1学期当たり 50万円の費用なら意思決定は変化しない



最適停止モデル

意思決定者（学生）

- d_t : 在学ダミー
- s_t : 状態ベクトル
- $U_t(d_t; s_t)$: 効用
- β : 割引因子

分析者

- x_t : 分析者が観測する状態ベクトル
- $u_t(d_t; x_t) := E(U_t(d_t; s_t)|x_t)$
- $e_t(d_t) := U_t(d_t; s_t) - u_t(d_t; x_t)$ (観測されない状態)
- $e_t := (e_t(0), e_t(1))'$

ベルマン方程式

在学者／退学者の価値関数

- $V_t^j(x_t, e_t)$: 価値関数
- $\bar{V}_t^j(x_t)$: 積分した価値関数

積分したベルマン方程式

$$\bar{V}_t^1(x_t) = E \left(\max_{d_t \in \{0,1\}} \left\{ \underbrace{u_t(d_t; x_t) + e_t(d_t)}_{\text{今期の効用}} + \beta E \left(\underbrace{\bar{V}_{t+1}^{d_t}(x_{t+1})}_{\text{来期の価値}} \mid x_t; d_t \right) \right\} \mid x_t \right)$$

s.t. $x_{t+1} \sim F_{t+1}(\cdot | x_t; d_t)$

条件付き選択確率 (CCP)

選択肢別価値関数

$$v_t(x_t; d_t) := u_t(d_t; x_t) + \beta E \left(\bar{V}_{t+1}^{d_t}(x_{t+1}) | x_t; d_t \right)$$

最適な選択

$$d_t^* := \arg \max_{d_t \in \{0,1\}} \{v_t(x_t; d_t) + e_t(d_t)\}$$

$\{e_t(j)\}$ は iid で平均 0, 尺度母数 σ のガンベル分布にしたがう
⇒ 在学確率関数 (CCP 関数) は 2 値ロジット・モデル

$$p_t(1|x_t) = \frac{\exp(v_t(x_t; 1)/\sigma)}{\exp(v_t(x_t; 0)/\sigma) + \exp(v_t(x_t; 1)/\sigma)} = \Lambda \left(\frac{v_t(x_t; 1) - v_t(x_t; 0)}{\sigma} \right)$$

※ 後ろ向き帰納法で $v_t(\cdot, \cdot)$ を解く

報告の流れ

研究の動機と貢献

先行研究

大学中退モデル

構造推定

モンテカルロ実験

実証分析

反対仮想分析

ロジット・モデルの性質

退学確率

$$p_t(0|x_t) = \frac{\exp(v_t(x_t; 0)/\sigma)}{\exp(v_t(x_t; 0)/\sigma) + \exp(v_t(x_t; 1)/\sigma)}$$

期待最大効用

$$\bar{V}_t^1(x_t) = \sigma \ln \left(\exp \left(\frac{v_t(x_t; 0)}{\sigma} \right) + \exp \left(\frac{v_t(x_t; 1)}{\sigma} \right) \right)$$

第1式を対数変換し、第2式に代入すると

$$\underbrace{\bar{V}_t^1(x_t)}_{\text{在学状態の価値}} = \underbrace{v_t(x_t; 0)}_{\text{退学の価値}} - \underbrace{\sigma \ln p_t(0|x_t)}_{\text{退学を保留する価値}}$$

補正関数

識別制約

1. 在学時の効用 $u_t(1; \cdot)$ は時点 t に依存しない (除外制約)
2. 退学の価値 $v_t(\cdot; 0)$ は状態ベクトル x_t に依存せず既知

$y_t := v_t(\cdot; 0)$ とすると在学の価値は

$$\begin{aligned}v_t(x_t; 1) &:= u(1; x_t) + \beta E(\bar{V}_{t+1}^1(x_{t+1}) | x_t; 1) \\&= u(1; x_t) + \beta E(v_{t+1}(x_{t+1}; 0) - \sigma \ln p_{t+1}(0 | x_{t+1}) | x_t; 1) \\&= u(1; x_t) + \beta y_{t+1} - \underbrace{\beta \sigma E(\ln p_{t+1}(0 | x_{t+1}) | x_t; 1)}_{\text{補正関数}}\end{aligned}$$

補正関数

$$z_t(x_t) = \int \ln p_{t+1}(0 | x_{t+1}) dF_{t+1}(x_{t+1} | x_t; d_t = 1)$$

CCP 法

$u(1; \mathbf{x}_t) := \mathbf{x}'_t \boldsymbol{\theta}$ と定式化し, $\eta := \sigma^{-1}$ とする.

1. 状態遷移確率分布 $\{F_t(\cdot | \cdot; \cdot)\}$ を推定
2. 誘導形の CCP 関数 $\{p_t(\cdot | \cdot)\}$ をノンパラメトリック回帰で推定
3. $\{\hat{F}_t(\cdot | \cdot; \cdot)\}, \{\hat{p}_t(\cdot | \cdot)\}$ から補正関数 $\{\hat{z}_t(\cdot)\}$ を計算
4. 次式の 2 値ロジット・モデルを用いて $(\boldsymbol{\theta}, \beta, \eta)$ を最尤法で推定

$$p_t(1 | \mathbf{x}_{i,t}) = \Lambda \left(\underbrace{\mathbf{x}'_{i,t} \eta \boldsymbol{\theta} + \beta \eta y_{t+1} - \beta \hat{z}_t(\mathbf{x}_{i,t})}_{\text{在学の価値}} - \underbrace{\eta y_t}_{\text{退学の価値}} \right)$$

※小標本だと完全分離が生じる

修正 CCP 法

対数オッズ比のノンパラメトリック推定値を従属変数として
 (θ, β, η) を非線形最小 2 乗法で推定

$$\ln \frac{\hat{p}_t(1|x_{i,t})}{\hat{p}_t(0|x_{i,t})} = \underbrace{x'_{i,t} \eta \theta + \beta \eta y_{t+1} - \beta \hat{z}_t(x_{i,t})}_{\text{在学の価値}} - \underbrace{\eta y_t}_{\text{退学の価値}}$$

※ Bajari, Chu, Nekipelov and Park (2016, Quant. Mark. Econ.) のアイデア

報告の流れ

研究の動機と貢献

先行研究

大学中退モデル

構造推定

モンテカルロ実験

実証分析

反実仮想分析

大学中退モデル

設定

- ・ 每学期 12 科目を履修
- ・ 8 学期の在籍と 64 科目 の修得が卒業要件
- ・ 観測期間は第 1~8 期の 8 期間

変数

- ・ $n_{i,t}$: 修得科目数
- ・ $N_{i,t}$: 累積修得科目数
- ・ $N_{i,t}^* := N_{i,t} - 8t$ (超過累積修得科目数)
- ・ $g_{i,t}$: GPA 水準 (4 つの水準に離散化)
- ・ $d_{i,t}$: 在学ダミー
- ・ $x_{i,t} := (N_{i,t-1}^*, g_{i,t-1})'$

実験デザイン

GPA 水準 $\{g_{i,t}\}$ の遷移確率行列

$$P := \begin{bmatrix} \pi & 1-\pi & 0 & 0 \\ 1-\pi/2 & \pi & 1-\pi/2 & 0 \\ 0 & 1-\pi/2 & \pi & 1-\pi/2 \\ 0 & 0 & 1-\pi & \pi \end{bmatrix}$$

修得科目数

$$n_{i,t}|g_{i,t} \sim \text{Bin}(12, p(g_{i,t}))$$

在学時の効用関数

$$u(1; x_{i,t}) := \alpha + \kappa N_{i,t-1}^* + \lambda g_{i,t-1}$$

母数の設定

退学の価値

- ・ 「賃金構造基本統計調査」から退学後の生涯賃金を計算
- ・ 卒業の価値も同様（推定には不要）
- ・ 留年の価値 = 入学 4 年後の退学の価値とする（推定には不要）

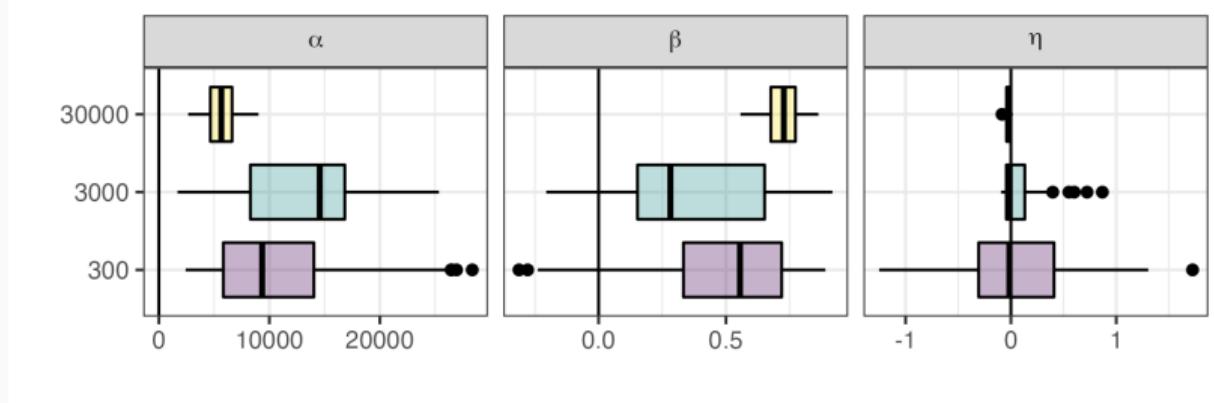
構造母数

- ・ 割引因子 : $\beta = .98$
- ・ 尺度母数 : $\sigma = 100$ (万円) すなわち $\eta := 1/\sigma = .01$
- ・ 効用関数の係数 : $\alpha = \kappa = \lambda = 0$
- ・ 同じ GPA 水準に留まる確率 : $\pi = .5$
- ・ 科目修得確率 : $p(1) = .6, p(2) = .7, p(3) = .8, p(4) = .9$

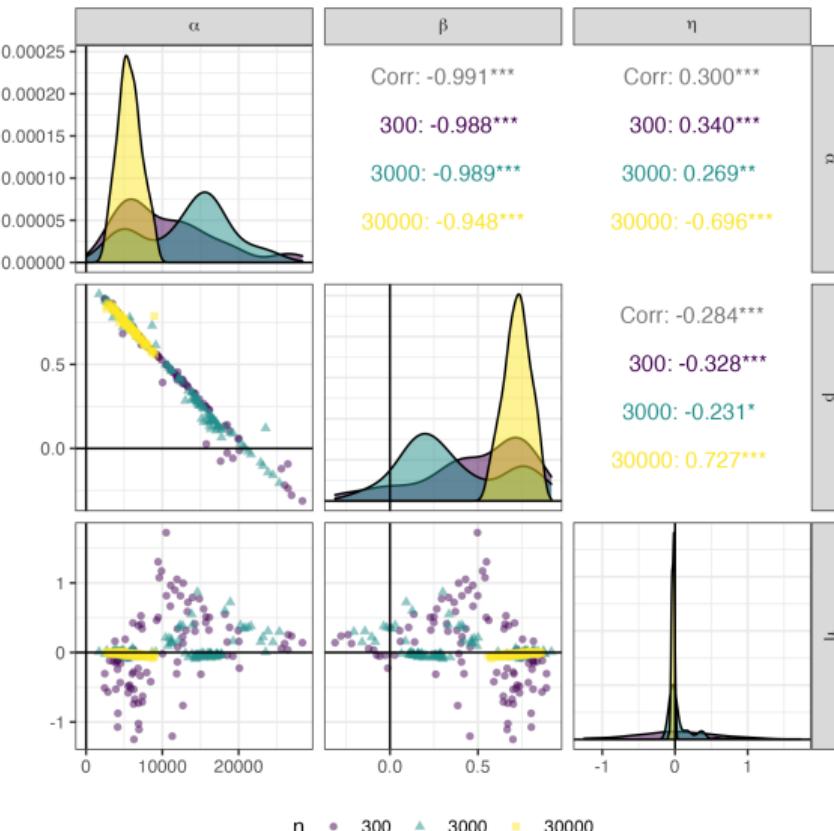
推定手順

- $\kappa = \lambda = 0$ は既知とする
- 修正 CCP 法
 1. 状態遷移確率分布を **学期別の経験分布** で推定
 2. 誘導形の CCP 関数をノンパラメトリックなロジスティック回帰で推定
(超過累積修得科目数と GPA 水準ダミーのみを説明変数として **学期別** に推定)
 3. 補正関数を計算
 4. 在学／退学の対数オッズ比のノンパラメトリック推定値を従属変数とした線形モデルに補正項を加えて構造母数を推定 (母数の範囲は制約しない)
- 入学者数 $n = 300, 3000, 30000$ を比較
(ノンパラメトリックな CCP 関数の定式化は同じ)

構造母数の推定値の分布



構造母数の推定値の散布図行列



考察

実験結果

- ・ n が大きいと推定値のばらつきは小さくなる
- ・ $n = 30000$ でも真の値に十分に近づいていない
- ・ (α, β) の推定値は強い負の相関

考察

1. 退学の観測数が少ないため、構造母数の推定精度は低い.
2. 構造母数の推定値に偏りは、おそらく誘導形の CCP 関数のノンパラメトリック推定値の偏りが原因.
3. β が小さいと卒業の価値が大きく割り引かれるため、 α が大きくないと誰も在学を選択しない. このトレードオフ関係は正確に推定される.

報告の流れ

研究の動機と貢献

先行研究

大学中退モデル

構造推定

モンテカルロ実験

実証分析

反実仮想分析

使用データ

某大学某学部 2016 年 4 月入学者の 2020 年 3 月末時点の中途退学・卒業・留年状況

	入学者数	退学者数	卒業者数	留年者数
男	301	12	242	47
女	101	2	94	5
計	402	14	336	52

※分析対象は男子のみ

設定

- ・ 每学期 24 単位を履修
- ・ 8 学期の在籍と 128 単位の修得が卒業要件

変数

- ・ $N_{i,t}^* := N_{i,t} - 16t$ (超過累積修得単位数)
- ・ $g_{i,t}$: GPA 水準 (4 つの水準に離散化)

在学時の効用関数

$$u(1; x_{i,t}) := \alpha + \kappa N_{i,t-1}^* + \lambda g_{i,t-1}$$

修正 CCP 法

1. 状態遷移確率分布を**学期別の経験分布**で推定
2. 誘導形の CCP 関数をノンパラメトリックなロジスティック回帰で推定
(超過累積修得科目数・GPA 水準ダミー・**学期ダミー**を説明変数)
3. 補正関数を計算
4. 在学／退学の対数オッズ比のノンパラメトリック推定値を従属変数とした線形モデルに補正項を加えて構造母数を推定

推定結果

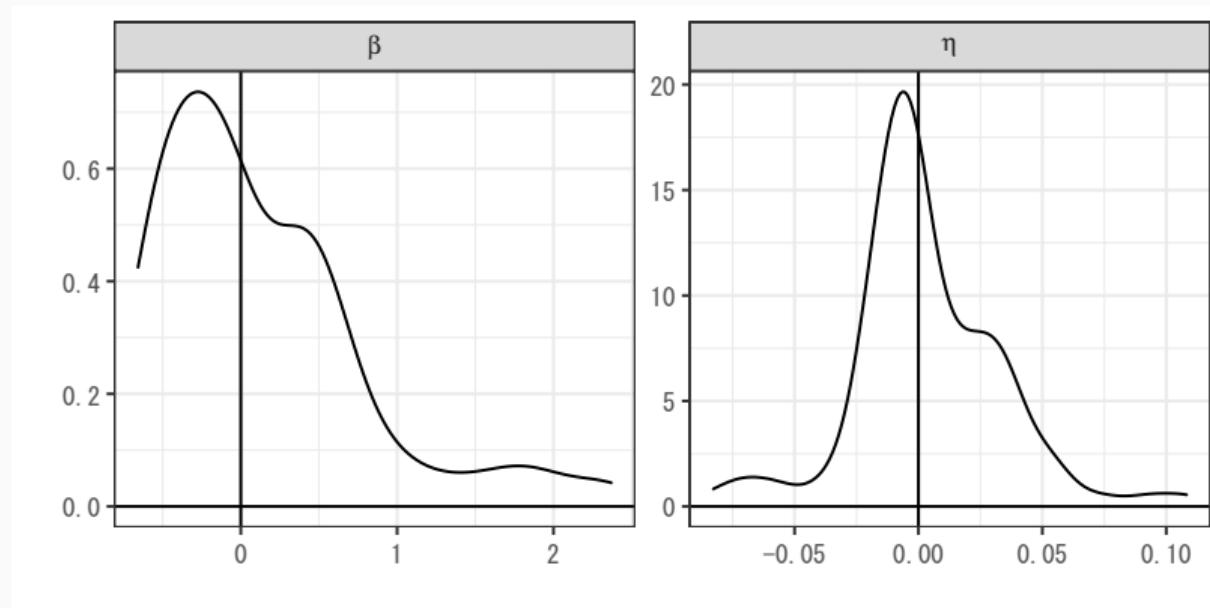
割引因子・尺度母数の制約の有無による比較

	制約なし	$\eta \geq 0$	$\beta \in [0, 1], \eta \geq 0$
割引因子 (β)	1.082	1.030	1.000
1/尺度母数 (η)	-0.003	0.000	0.000
定数項	170.580	-97223.960	-95227.040
超過累積修得単位数	8.093	-248.174	-226.913
GPA 水準	-598.198	35284.470	36578.480

※端点解となる。

推定量の分布

100回のブートストラップ



推定結果

割引因子・尺度母数を先驗的に与えた場合

	(1)	(2)	(3)	(4)
割引因子 (β)	0.98	0.98	1.00	1.00
1/尺度母数 (η)	0.01	0.05	0.01	0.05
定数項	32.756	454.067	-381.340	44.130
超過累積修得単位数	1.279	2.239	1.332	2.356
GPA 水準	183.477	16.334	177.106	13.833

※ (4) が最も自然な結果（有意ではない）

推定結果のまとめ

1. 割引因子 β は 1 に極めて近い
2. 尺度母数 σ は 20 万円程度
3. 超過累積修得単位数・GPA 水準が在学の効用に直接的に影響する証拠はない

$\Rightarrow \beta \in [.95, 1], \sigma \in [20, 100]$ の範囲で効用関数を特定化しない反実仮想分析を行う

報告の流れ

研究の動機と貢献

先行研究

大学中退モデル

構造推定

モンテカルロ実験

実証分析

反実仮想分析

効用関数の一時的な変化

記号の定義

- $\Delta_t(\cdot, \cdot)$: 効用関数の変化, ただし $t > T$ では $\Delta_t(\cdot, \cdot) = 0$
- $u_t^*(\cdot, \cdot) := u_t(\cdot, \cdot) + \Delta_t(\cdot, \cdot)$
- $v_t^*(\cdot, \cdot)$: 反実仮想の選択肢別価値関数
- $p_t^*(\cdot | \cdot)$: 反実仮想 CCP
- $\bar{V}_t^{j*}(\cdot)$: 反実仮想の積分した価値関数

時点 $T+1$ の問題は変化しないので, $j = 0, 1$ について

$$\bar{V}_{T+1}^{j*}(\cdot) = \bar{V}_{T+1}^j(\cdot)$$

反実仮想の選択肢別価値関数

時点 T では $j = 0, 1$ について

$$\begin{aligned} v_T^*(x_T; j) &= u_T^*(j; x_T) + \beta E \left(\bar{V}_{T+1}^{j*}(x_{T+1}) | x_T; d_T = j \right) \\ &= u_T(j; x_T) + \Delta_T(j; x_T) + \beta E \left(\bar{V}_{T+1}^j(x_{T+1}) | x_T; d_T = j \right) \\ &= v_T(x_T; j) + \Delta_T(j; x_T) \end{aligned}$$

在学／退学の価値の差は

$$\begin{aligned} v_T^*(x_T; 1) - v_T^*(x_T; 0) &= v_T(x_T; 1) - v_T(x_T; 0) \\ &\quad + \Delta_T(1; x_T) - \Delta_T(0; x_T) \end{aligned}$$

在学／退学の対数オッズ比は

$$\begin{aligned} \ln \frac{p_T^*(1|x_T)}{p_T^*(0|x_T)} &= \frac{\nu_T^*(x_T; 1) - \nu_T^*(x_T; 0)}{\sigma} \\ &= \frac{\nu_T(x_T; 1) - \nu_T(x_T; 0)}{\sigma} + \frac{\Delta_T(1; x_T) - \Delta_T(0; x_T)}{\sigma} \\ &= \ln \frac{p_T(1|x_T)}{p_T(0|x_T)} + \frac{\Delta_T(1; x_T) - \Delta_T(0; x_T)}{\sigma} \end{aligned}$$

σ が既知なら $p_T(\cdot|\cdot)$ と $\Delta_T(\cdot, \cdot)$ のみから $p_T^*(\cdot|\cdot)$ は求まる。

(β, σ) が既知なら $p_T^*(\cdot|\cdot)$ を所与として $\{p_t^*(\cdot|\cdot)\}$ は逆順で逐次的に求まる（詳細は略）。

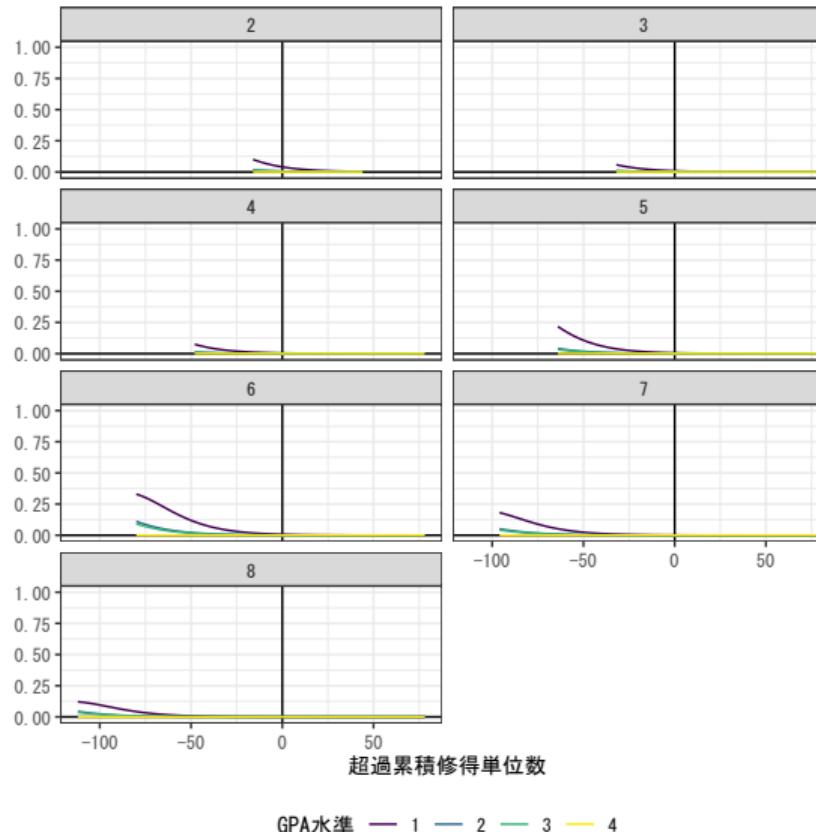
例：学資補助

構造母数の設定

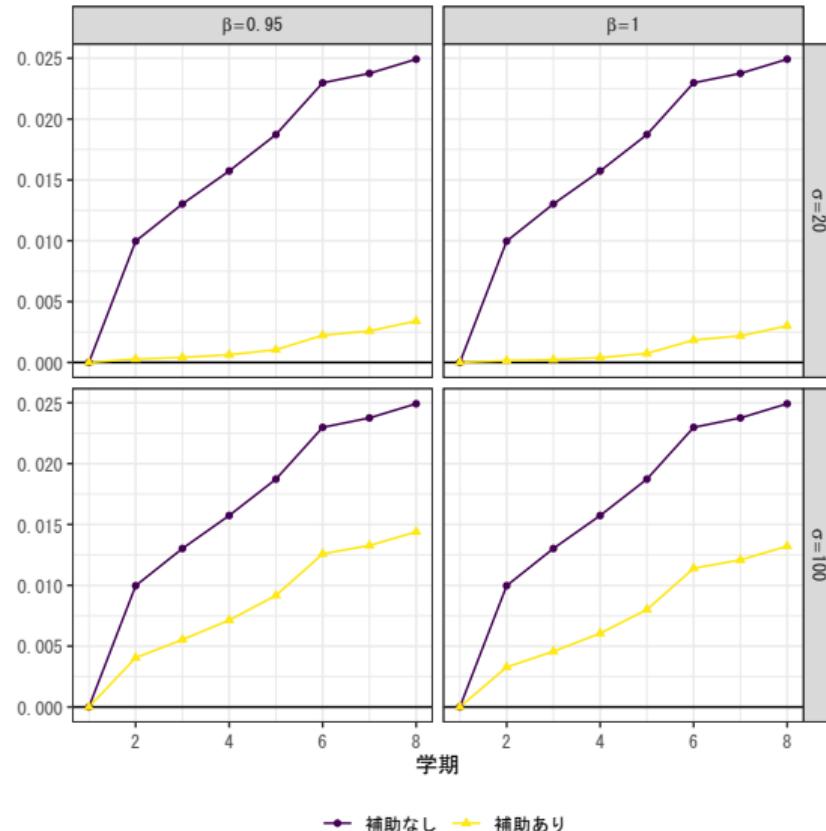
- ・ 割引因子 $\beta = 1$
- ・ 尺度母数 $\sigma = 20$ (万円)
- ・ 効用関数の特定化は不要

反実仮想：在学中 4 年間の毎学期 10 万円の学資補助

在学確率関数を引き上げる処置効果



累積中退確率を引き下げる処置効果



学資補助の処置効果

1. 每学期 10 万円の学費補助は、某学部 2016 年 4 月入学者の男子の 4 年間の累積中退確率を約 2.2% 引き下げる。
特に 2 年次までの中退はほぼ 0 になる。
2. $\beta \in [.95, 1]$, $\sigma \in [20, 100]$ の範囲では、累積中退確率は 1.0~2.2% 低下。
3. 学費補助は成績に直接的に影響しない。
⇒ 累積中退確率の低下は卒業確率の上昇を意味せず、
退学の意思決定を遅らせる効果と解釈すべき。
4. もともと卒業できる学生が最大の利益を得る。大学の利益にはならない。

参考文献

- Arcidiacono, P. 2004. Ability sorting and the returns to college major. *Journal of Econometrics*, 121, 343–375.
- Bajari, P., C. S. Chu, D. Nekipelov, M. Park 2016. Identification and semiparametric estimation of a finite horizon dynamic discrete choice model with a terminating action. *Quantitative Marketing and Economics*, 14, 271–323.
- Stange, K. M. 2012. An empirical investigation of the option value of college enrollment. *American Economic Journal: Applied Economics*, 4, 49–84.
- Stinebrickner, R., T. Stinebrickner 2014. Academic performance and college dropout: using longitudinal expectations data to estimate a learning model. *Journal of Labor Economics*, 32, 601–644.
- 鹿野繁樹・高木真吾・村澤康友 2011.「経済学の成績に対する数学学習の効果：コントロール関数アプローチによる推定と予備検定」,『統計数理』,, 59, 301–319.