

# 給付型奨学金の効果検証

Kei Ikegami

2017 年 1 月 20 日

## 1 事実

- Jasso の予約採用募集時期は 5,6 月に 1 種 2 種、10,11 月に 2 種のみ。在学採用は毎年春。ただし詳しい時期や回数は高校に一任されている。

## 2 モデルの設定

### 2.1 モデル 1：池上/12/22

ナイーブなモデルを設定する。

全国の高校生は 3 年生の 5、6 月に、以下のプロセスをバックワードに意思決定する。

1. 進学するか否か
2. 受験する学校（職場）のセットは何か

受験する学校のセットは以下のように決定される。 $A$  を国公立 4 年生大学、 $B$  を私立 4 年生大学、 $C$  を専門学校の集合とする。 $A, B, C$  の各学校は奨学金ありと奨学金なしで 2 パターンずつ存在する。受験生は  $A$  の中から効用を最大化する選択肢を 1 つ選び、 $B, C$  については効用の高い順にランク付けする。 $A$  から選んだものも含めてすべての選択肢についてのランキングを作成した後、受験生が住んでいる都道府県での受験校平均だけ上から取り出したものを受験校のセットとする（ただし正の値のもののみを含むようにする）。

生徒  $i$  が学校  $n$  の奨学金なしパターンを選択する効用を  $u_n^i$  とすると、

$$u_n^i = (I_n - H^i - T_n)r_n^i$$

ただし、 $I_n$  は学校  $n$  を卒業した人の平均生涯賃金、 $H^i$  は  $i$  さんが住んでいる都道府県の高卒者平均生涯賃金、 $T_n$  は学校  $n$  の学費の割引現在価値、 $r_n^i$  は  $i$  さんが学校  $n$  に受かる確率である。

奨学金ありパターンを選択する効用を  $u_n^{i*}$  とすると、

$$u_n^{i*} = (u_n^i + S_n^i - P^i - C - \infty * \mathbf{1}(X_i \in Q))r_n^i d^i$$

ただし、 $S_n^i$  は  $i$  さんが学校  $n$  に進学した際にもらえる奨学金の合計金額の割引現在価値、 $P^i$  は  $i$  さんが返済する金額の割引現在価値、 $C$  は奨学金に申し込むコスト、 $X_i$  は  $i$  さんの demographic で  $Q$  が奨学金をもらえない demographic の集合である。また、 $d_n^i$  は  $i$  さんの奨学金採用確率である。

上で述べた通り、学生は各学校について上の二つの効用を比べ、効用最大化問題を解いている。

進学せずに就職するやつの数を、「受験校のセットが空集合の受験生の数」+「(各学校に対する合格率で当落を判定し、すべて不合格になった受験生の数)  $\times w$ 」(ただし  $w$  は全落ちしたやつが浪人せずに就職する確率)で推定する。上の効用をパラメトライズして、この高卒就職組の数を高校別 or 県別に出して MLE をすればいい？

### 2.2 モデル 1 改良:池上 12/24,5

モデル 1 をもう少し specific にする。

- 意思決定に自宅通学か下宿かの分岐を入れ込んだら
- 大学の characteristics として「関関同立」のような世間一般に受け入れられている格付けをを追加
- specification として各大学の受験生獲得シェアの学費弾力性のようなものを推定するようにパラメトライズする。  
→この文脈だと国公立に絞った方が嬉しいかも？（一つしか選べないので）

疑問としては次のものがある。

- 流動性制約をこのモデルで扱っているか？
- そもそも奨学金もらっている人が少ないし、torbit とかのほうがいいのか？

以下で具体的に書く。 $s = 0$  で自宅通学、 $s = 1$  で下宿を表す。 $a = 0$  で奨学金なし、 $a = 1$  で奨学金受給を表す。 $i$  さんが学校  $n$  に  $s, a$  の状態で通うことの効用を以下のように書く。

$$u_{i,n,s,a} = \{y_n \beta_1 + \alpha_i (T_n - \mathbf{1}(a = 1)A_i) + \gamma_i H_n + \delta_i \mathbf{1}(s = 1)S_n + \eta_i R_i - \mathbf{1}(a = 1)(\mathbf{1}(X_i \in Q_n) * \infty)\} \\ * r_n(X_i) * \{p(X_i)\mathbf{1}(a = 1)\}$$

ここで、 $y_n$  は学校  $n$  の、就職という選択肢と比べられる特徴量（例えば生涯賃金）。 $T_n$  は学校  $n$  の学費の現在価値、 $A_i$  は  $i$  さんが貰える奨学金の現在価値である。 $H_n$  は学校  $n$  の「格」を表す指標で、「東大・京大」「早稲田・慶應」などの大学間のランク付けが社旗通念として一般に受け入れられていることから取り入れる価値があると判断した。これは順序付きのカテゴリ変数でもいいし、格ごとのダミー変数でもいい。 $S_n$  は学校  $n$  に下宿で通う場合の家賃である。学校周辺のワンルームの平均家賃をデータとして用いる（通っている間の割引現在価値を入れる）。 $R_i$  は  $i$  さんが返済する奨学金の金額の割引現在価値である。現在でも返済する金額には差異が存在しているようなのでその変動を利用する。 $X_i$  は  $i$  さんの demographic で、 $Q$  が奨学金の対象である個人の demographic 空間ないの部分集合である。 $r_n(X_i)$  は  $i$  さんの学校  $n$  に対する合格率で、 $X_i$  の関数で書く。 $p(X_i)$  は  $i$  さんの奨学金採用率で、これも  $X_i$  の関数で書く。式中の  $\mathbf{1}$  は indicator function である。ただし、上式での各パラメータは個人に依存した形にしているので、実際に計算するときは BLP みたいに mean utility と hetero な部分に分解して計算することになるので注意（めんどくさいしとりあえず homogenetic なモデルでやるのがいいかも）。

先のモデルと同様に、 $i$  さんの県内での平均受験校数を  $N_i$  として、国公立から最大の効用をもたらす学校の一つ、私立と専門から効用の大きい順に  $N_i - 1$  個選び、 $N_i$  この受験校 bundle を作る。それと、就職する効用である  $u_{i,0}$  とを比べて、就職するシェアを離散選択モデルの文脈で計算する。そのシェアに、bundle に全落ちする学生のうちから何割かを浪人せずに就職する人達として、そのシェアに追加し、全体としての高卒就職率を算出する。

浪人せずに就職する割合を beta 分布などからシミュレーションして、Farmanian and Salanie (2004) に基づいて likelihood function を作成し、MLE を求める。（それかそれもパラメータとしてただの MLE でやる）

## 2.3 モデル 2:池上 12/28

意思決定の過程は以下のとおり。主体は高校 3 年生とその世帯とする。

まず、高校 3 年生は以下の 2 パターンに分類される。

1. 奨学金なしで通うことのできる学校が存在する
2. 奨学金なしではどの学校にも通うことができない

ここで、「通うことのできる学校が存在する」とは実現可能性集合に高校以降の学校が少なくとも一つ含まれている状態を指す。学校に関する情報は完全であることを仮定し、存在するすべての学校に対して世帯ごとに通悪可能かいないかが判断されている状態を想定する。下宿も自宅通学もどちらの可能性も考慮して実現可能か否かは判断されることとする。

パターン 1 の学生はそれぞれ「通うことのできる学校リスト ( $A$  と呼称)」を入手する。「奨学金がもらえる時に通うことのできる学校リスト ( $B$  と呼称)」も同時に作成され、その際に以下の 3 パターンに分割できる。

- 1-a  $A$  の中で効用最大化を達成する選択肢が  $B$  の中で効用最大化を達成する選択肢と一致する。
- 1-b  $A$  と  $B$  とで効用最大化を達成する選択肢が異なる。
- 1-c 奨学金制度をよく知らず、 $A$  しか見ることができない。

簡単のために  $1 - a$  は奨学金に申し込まないとする。 $1 - b$  は  $B$  における最大の効用が受験しないという選択肢の効用を上回った時のみ奨学金への申し込みを行う。奨学金の採択確率は  $X_i$  を  $i$  さんの demographic として  $q(X_i)$  でかけるとする。 $1 - c$  は奨学金を申し込まない。

また 2 のパターンの学生は以下の 2 パターンに分類できる。

2-a 奨学金を知っている。それによって「奨学金があったら通えるリスト ( $C$  と呼称)」が与えられる。

2-b 奨学金を知らない。

この時 2-a は  $C$  における最大効用が受験しないという選択肢の効用を上回った時のみ勉強を始め、奨学金に申し込む。2-b は受験することはできない。

上で、各学生は自分のリストから最大の効用を与える学校と受験しないという選択肢を比べて「勉強を始める」という意思決定を行い、パターンによっては「奨学金に申し込む」という意思決定も行う。次に、彼らは受験をするか否かという意思決定を行うために、自分の受験校の bundle を決定する。この bundle の決定は受験というくじの価値を決めるものである。bundle の決定は単純に、「仮にその学校に受かったとして、受験しないで就職するよりはいいことがありそうだ」と思えないものは bundle に入れない問いうルールに従うとする。 $i$  さんの bundle にはいる学校の個数は「 $i$  さんの在住県の学生が受験する学校の平均数」と「受験しないという選択肢よりも大きな効用をもたらす、 $i$  さんのリスト内にある学校の数」の小さい方と等しいとする。

学生たちは次に「受験するかしないか」の意思決定を行う。bundle の決定によりくじの価値が決定したのでこの意思決定を考えることができる。 $i$  さんの bundle を  $B_i$  とする。 $B_i$  には  $1, 2, \dots, m_i$  だけ学校が入っており、それぞれに  $r_{i,1}, r_{i,2}, \dots, r_{i,m_i}$  が  $i$  さんの合格率として紐付いており、それぞれの合格した際の効用を  $u_{i,1}, u_{i,2}, \dots, u_{i,m_i}$  と表記することとする。

bundle ないの学校は数字が小さいほど  $i$  さんにとっての効用が大きいとして、より効用が大きい学校に受ければそれ以下の学校には進学しないとして、すべての学校に落ちた時の効用を  $b_{i,0}$  とすると、 $B_i$  の持つ期待効用は以下のように書ける。

$$u_i^* = r_{i,1}b_{i,1} + (1 - r_{i,1})r_{i,2}b_{i,2} + \dots + \prod_{j=1}^{m_i-1} (1 - r_{i,j})r_{i,m_i}b_{i,m_i} + \prod_{j=1}^{m_i} (1 - r_{i,j})b_{i,0}$$

今、 $u_i$  を  $i$  さんが就職する際の効用として、誤差項に Type 1 extreme value distribution を仮定すれば離散選択モデルとして高卒就職を意図的に選択する人のシェアを出すことができる。

最後に今まで扱ってきた効用を以下のようにパラメトライズする。 $i$  さんが学校  $j$  に奨学金なしで進学し、そこで学位をとることの効用を  $u_{i,j}$  とし、 $i$  さんが学校  $j$  に奨学金付きで進学しそこで学位をとることの効用を  $u_{i,j}^*$  とする。この時以下のようなものである。

$$u_{i,j} = \alpha_i \sum_{n=1}^{L_j} \{h^n(y_i^n - t_j - w_i^n)\} + \alpha_i \sum_{n=L_j+1}^{\infty} \{h^n(y_i^n + z_j^n - w_i^n)\} + \beta_i d_j + \delta_i C_{i,j}$$

$$u_{i,j}^* = \alpha_i \sum_{n=1}^{L_j} \{h^n(y_i^n - t_j - w_i^n + s_i)\} + \alpha_i \sum_{n=L_j+1}^{\infty} \{h^n(y_i^n + z_j^n - w_i^n - r_i^n)\} + \beta_i d_j + \delta_i C_{i,j}$$

ここで、

$y_i^n$  世帯  $i$  の  $n$  年後の世帯収入

$w_i^n$   $i$  さんが高卒で就職した際の  $n$  年後の収入

$r_i^n$   $i$  さんが大学進学して  $n$  年後に返済する奨学金の額

$z_j^n$   $j$  校を卒業した学生が  $n$  年後に稼いでいる年収

$d_j$   $j$  校の「格」を示すカテゴリー変数（ダミーにしても良い）

$C_{i,j}$   $i$  さんが  $j$  校に「合格」し、「学位を所得」するまでに必要な勉強コスト（必要な時間でバイトした時の収入増加分の割引現在価値）

$L_j$   $j$  校で学位をとるまでに必要な年限

$h$  割引率

である。

## 2.4 モデル 2 改良：池上 12/29

浪人するという意思決定をモデルに組み込む。まず  $i$  さんの浪人可能年数を  $Y_i$  とする。 $Y_i$  は

$$Y_i = \#e \text{ s.t. } \exists j \text{ s.t. } b_{i,j}^e \geq u_{i,0}^e$$

で定義される。すなわち、「 $i$  さんが  $e$  浪して合格した時の効用が  $e$  回受験に失敗した時に就職した時の効用よりも大きくなる学校  $j$  が少なくとも一つ存在している年数」である。

2.3 節とほぼ同じ記号を使って、以下のように「 $i$  さんが  $j$  校に  $e$  浪して合格した時の効用」を新しく定義する。

$$u_{i,j}^e = \alpha_i \sum_{n=1}^e \{h^n(-f_j - w_i^n)\} \alpha_i \sum_{n=1+e}^{L_j+e} \{h^n(-t_j - w_i^n)\} + \alpha_i \sum_{n=L_j+e+1}^{\infty} \{h^n(z_j^{n-e} - w_i^n)\} + \beta_i d_j + \delta_i C_j + \eta_i g_j + \gamma e \quad (1)$$

$$u_{i,j}^e = \alpha_i \sum_{n=1}^e \{h^n(-f_j - w_i^n)\} \alpha_i \sum_{n=1+e}^{L_j+e} \{h^n(-t_j - w_i^n + s_i)\} + \alpha_i \sum_{n=L_j+e+1}^{\infty} \{h^n(z_j^{n-e} - w_i^n - r_i)\} + \beta_i d_j + \delta_i C_j + \eta_i g_j + \gamma e \quad (2)$$

ただし

$f_j$   $j$  校に受かるための浪人費用 1 年分

$C_j$   $j$  校の偏差値

$g_j$   $j$  校を卒業した場合の配偶者の期待所得

である。

上の効用を用いて、 $i$  さんの bundle の持つ価値が以下のように書ける。

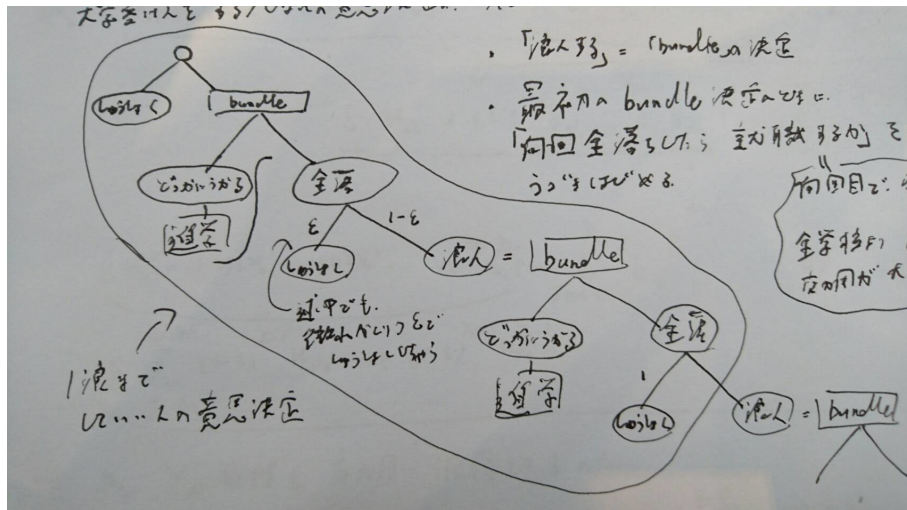
$$u_i^* = \sum_{e=0}^{Y_i} (1-\epsilon)^e u_i^e + \epsilon \sum_{e=0}^{Y_i-1} (1-\epsilon)^e u_{i,0}^e + (1-\epsilon)^{Y_i} u_{i,0}^{Y_i} \quad (3)$$

ただし、

$$u_i^e = r_{i,1}^e b_{i,1}^e + \dots + \Pi_{j=1}^{m_i^e} (1 - r_{i,j}^e) r_{i,m_i^e}^e b_{i,m_i^e}^e \quad (4)$$

$$u_{i,0}^e = \alpha_i \sum_{n=0}^{\infty} (h^n w_i^{n-e}) \quad (5)$$

また、 $b_{i,j}^e = \max(u_{i,j}^e, u_{i,j}^{e-1})$ 、 $r_{i,j}^e$  は  $i$  さんが  $e$  浪で  $j$  校に受かる確率、 $m_i^e$  が  $i$  さんの  $e$  浪目の bundle には入っている学校である。 $\epsilon$  はすべての志望校に不合格となった場合に次年度浪人することが可能なにもかかわらず就職することを選んでしまう確率である。(3) の構成は以下の図を参考にする。(  $e=3$  とかで試すとわかる )



### 3 分析

#### 3.1 分析 1 : RD (徳田)

#### 3.2 分析 1 改良 : RD (池上)

#### 3.3 分析 2 : ヘテロな離散選択 (池上)

$n_1$  を 1 種奨学金申請者数、 $n_2$  を 2 種奨学金申請者数、 $n_3$  を 1 種 2 種併用奨学金申請者数、 $n_4$  を奨学金に申請せずに進学を希望する人数、 $n_5$  を進学を希望しない人数、とする。 $n_1, n_2, n_3$  は Jasso のデータからわかる。 $n_4$  は各大学の倍率と合格者数のデータから合計志望者数を計算し  $n_1, n_2, n_3$  の合計をそこから引くことで計算する。 $n_5$  はデータがあるので簡単にわかる。

使えるデータは「いずれかの奨学金に申請した家計に関する諸変数」、「いずれかの奨学金に申請した家計がどの奨学金を選択したかの離散データ」、「各奨学金がいくらを支給しいくら返済を求めるか」である。「奨学金に申し込まなかった家計に関する変数」は知りようがないことに注意する。これらのデータから上の 5 つの選択肢に関する離散選択モデルの文脈で、ヘテロな効用関数を推定することを考える。

$j = 1, 2, 3, 4, 5$  であり、 $T$  年分のデータが得られているとする。 $p_{jt}$  を選択肢  $j$  の  $t$  年における割引現在価値とする。ここで、割引現在価値は「払う総額の割引現在価値からもらう総額の割引現在価値を引いたもの」と「大卒資格の有無による生涯所得の変化」、「大学に通う間労働しない分の機会費用」から計算する。また各選択肢がこの  $p_{jt}$  で完全に特徴付けられているとする。 $y_i$  を家計  $i$  の世帯収入、 $Z_i$  を家計  $i$  についての諸変数をまとめたベクトル、 $\xi_{jt}$  を家計に共通の観測できないショック、 $\epsilon_{ijt}$  を家計レベルの観測できないショックとする。

この時家計  $i$  の  $j$  に対する  $t$  年の効用 ( $u_{ijt}$ ) を以下のように書く。ただし、 $a_i$  は家計  $i$  の変数に依存したパラメータであり、 $a_i = a + \Pi_a Z_i + \Sigma_a v_i$  でかけるものとする。この時  $v_i$  は標準正規分布からの乱数として、 $a$  と  $\Pi_a$  と  $\Sigma_a$  が推定したいパラメータとなる。

$$\begin{aligned} u_{ijt} &= a_i(y_i - p_{jt}) + Z_i' \beta + \xi_{jt} + \epsilon_{ijt} \\ &= a_i y_i + Z_i' \beta + (-a p_{jt} + \xi_{jt}) + (-\Pi_a Z_i - \Sigma_a v_i) p_{jt} + \epsilon_{ijt} \\ &= a_i y_i + Z_i' \beta + \delta_{jt} + \mu_{ijt} + \epsilon_{ijt} \end{aligned} \quad (6)$$

ここで  $\epsilon_{ijt}$  が第 1 種極値分布に従っているとすると、

$$q_{ijt} = \frac{\exp \delta_{jt} + \mu_{ijt}}{\sum_k \exp \delta_{kt} + \mu_{ikt}}$$

として、

$$q_{ijt}^* = \int_{v_i} q_{ijt} dP_{v_i}(v_i) \quad (7)$$

が家計  $i$  の  $j$  に対する  $t$  年の選択確率である。これは  $a, \xi_{jt}, \Pi_a, \Sigma_a$  を未知の値とする関数としてかけている ( $\xi_{jt}$  は  $j$  ごとに別々のパラメータとすればいい?)。この選択確率を利用して奨学金申請者のデータから尤度関数を作成し、パラメータの最尤推定量を求める。これを初期値として以下の BLP っぽい推定を行う (BLP は推定の精度がクソらしいのでそれっぽい初期値を持ってきてあげると良くなりそうだなと思ってる)。

各選択肢のシェアは以下のように計算できる。

$$s_{jt} = \int_Z \int_{v_i} q_{ijt} dZ dv_i \quad (8)$$

ここで以下のループを回す。

1. パラメータを given として (最初は上で求めた MLE) シェアを (8) 式に従って計算する。
2. 計算したシェアに基づいて  $\delta$  に関する以下の縮小写像を使い  $\delta$  を更新する。更新式は  $\delta_t^{h+1} = \delta_t^h + (\log(S_t) - \log(s_t))$  であり、 $S_t$  は実際の各選択肢のシェアで  $s_t$  がシミュレーションしたシェアである。それぞれの変数が選択肢ごとの値を含むベクトルとなっていることに注意。

3. 更新した  $\delta$  を  $p_{jt}$  と定数項に回帰し、 $a$  と  $\xi_{jt}$  を更新する。
4. 更新した  $a$  と  $\xi$  に基づいて、 $\omega = \delta - (-ap_{jt} + \xi_{jt})$  を計算し、 $\omega'(E(\omega'\omega))^{-1}\omega$  を最小化するような  $\Pi_a, \Sigma_a$  を numerical に求める
5. 更新されたすべてのパラメータを使ってループの 1 に戻る。

これが収束するまで回してパラメータを推定する。