

第12章 政策評価分析の手法：

職業訓練の効果

1 はじめに

経済学研究の中心的テーマに経済政策の効果を測定するということがある。しかし、政策を実施する前であれば、政策結果に関するデータはないので、一般には経済理論に基づく議論をするか、同じような状況にあった過去の事例を実証研究して、それに基づいて提言するかのいずれかになることが多い。よくある政策議論の問題点は、現状とは似ても似つかない過去の状況に関する実証結果をもって、政策議論をしてしまうことである。特に観察点が少ないマクロ時系列分析では、現状が過去のどの時点に類似しているかなどということを考えずに実証を行って、政策効果があたかも、どのような経済環境下でも不変であるかのような議論がしばしばなされてきたように思う。

Heckman and Smith (1998) が適切に指摘しているように、近年マイクロデータが広範に利用可能になるにつれて、政策評価を行う場合にも、できるだけ他の条件を一定にして、つまり、政策を行う場合と行わない場合の厳密な比較を行うことを心がけるようになってきた。また、ミクロ計量経済学の分野では、そのための手法が開発され、多くの実証研究が政策分析の方法を根本的に変えようとしている。

経済政策の分析で難しいのは、何を成果として計測するのか、あるいは、その効果は政策実施後どのぐらいで出てくるものなのか、といった成果に関する評価の問題である。この問題はマイクロデータを用いたミクロ計量経済学でも解決されてはいないが、同一主体を繰り返し調査するパネルデータが利用可能になれば、政策と成果の時間ラグの問題もある程度理解できるようになるだろう。

昨今、企業は業績評価、研究者も業績評価、政府も行政評価と総評価の時代に入りつつあるように思われる。経済学として、この評価の方法をどう考えているのかをここで紹介しておきたい。

政策評価あるいはプログラム評価に関する概説書としてはLee(2005)、Cameron and Tridevi (2005, Chapter 25)、Wooldridge (2003, Chapter 18) 等を参照されたい。

2 政策評価の基礎概念

政策評価の方法は科学の分野で用いられてきた実験計画法に大きな影響を受けている。もちろん、自然科学では実験環境を管理した管理実験を行うことが可能であり、その結果として、様々な因果関係や投入物の効果を厳密に測定できるのに対して、社会科学では実験環境を完全には管理できないばかりか、被験者が実験の内容を理解してそれに応じて行動を変えてしまうことも考えられる。とすれば、むしろ実験であると意識させないような自然実験 (natural experiments) が望ましいことになるが、この場合でも、すべての環境を管理することは出来ない。しかし、少なくとも実験の参加者を無作為に抽出しておくことによって、不参加者との差が、実験効果以外のもので説明される可能性を抑えることはできるだろう。

なんらかの政策あるいはプログラムの評価を行う場合には、その成果を表す変数 y を基に評価する必要がある。理論的に考えれば、ある政策あるいはプログラムに参加した場合の経済主体の成果は y_1 で表され、もし経済主体がそれに参加しなかった場合の成果を y_0 で表すと、その政策あるいはプログラムの純効果は $y_1 - y_0$ で表されるはずである。容易に想像がつくように、同一経済主体が政策プログラムに参加することと、参加しないことを同時に経験することはできないという問題が生じる。すなわち、あるトレーニングに参加した人の成果 y_1 は観測できるが、その人がトレーニングに参加しなかった場合の成果 y_0 は、実際には参加しているのだから、実測できないだろう。この実測できないケースをあたかも経験したかのように扱うことを仮想現実 (counterfactual) を設定すると言い、この仮想現実がいかに現実的であるかによって、政策評価の適切さが違ってくる。

基本的な考え方は、同一経済主体の個別評価 $y_1 - y_0$ は個々人の状況に応じて違ってくるだろうが、ここで我々が関心があるのは、社会全体の平均として、その政策やプログラムの効果がどれくらいあるかということを実測することにある。そこで、 d を政策処理 (treatment) を受けた指標 (indicator) であるとしよう、もし $d = 1$ であれば経済主体は処理群 (treatment group)、 $d = 0$ であれば、経済主体は処置を受けていない対照群 (control group) に属しているとしよう。

平均処理効果 (average treatment effect : ATE) は次のように定義できる。

$$ATE = E(y_1 - y_0)$$

また、処理群の平均処理効果 (average treatment effect on the treated: ATT) は次のように表せる。

$$\begin{aligned} ATT &= E(y_1 - y_0 | d = 1) \\ &= E(y_1 | d = 1) - E(y_0 | d = 1) \end{aligned}$$

右辺の第一項は観察できるが、第二項は観察不可能である。観察可能なデータは $E(y_0|d=0)$ なので、 y_0 の期待値と d は独立であると仮定し、 ATT を次のように近似する。

$$ATT \approx E(y_1|d=1) - E(y_0|d=0)$$

ATE を推定するにはさらに強い仮定を置く必要がある。すなわち、 ATE は次のように書き換えることができる。

$$ATE = P(d=1)E(y_1 - y_0|d=1) + P(d=0)E(y_1 - y_0|d=0)$$

ここでは $E(y_0|d=1)$ に加えて、 $E(y_1|d=0)$ も観察不可能であり、これも仮定を置いて近似する必要がある。すなわち、 y_1 の期待値と d も独立であると仮定すると、 ATE は ATT と同じ近似を用いることができる。

$$ATE = ATT \approx E(y_1|d=1) - E(y_0|d=0)$$

この関係が成り立つためには、処理群および対照群の選択が全くランダムに行われていることが前提になる。この選択がランダムに行われていない場合には、セレクション・バイアスについて配慮する必要がある。

ここで、政策評価を行う上でのデータの性質について整理しておこう。実際のデータはいくつかのパターンに分類することができる。(1) 導入された政策が社会実験の中で行われたものであり、処理群と対照群があらかじめ設定されており、データも両方そろっている場合、さらにこれは処理を受ける前後の同一主体に対する観察をおこなったパネルデータになっている場合と1時点だけのクロスセクション・データである場合に分けられる。(2) 政策を実施した処理群のデータはあるが、対照群のデータは不在であり、外部情報から対処群をみつけてマッチングさせる場合、これも処理前後の観察が出来るパネルデータかクロスセクション・データかでできることが違ってくる。

実証分析する場合、データがどのような性質を満たしているかで分析方法が異なってくる。

3 政策評価の計量経済学

本節では前節で導入した概念をどのように具体的にどのように計測するかについて論じていこう¹。

まず、データが処理群と対照群ともにそろっておりしかも処理前後でパネルデータ化されている場合には、Differences-in-Differences (DID) 推定を

¹本節は主として Cameron and Trivedi (2005,25 章) を参照している。

用いればいい。ここでは時間表示として事前を b(=before)、事後を a(=after) とすると、処理群の時間を通じた処理効果 (TE) はつぎのように表すことができ、これを Before-After(BA) 推定とも言う。

$$TE_i = E_i(y_{1a} | d = 1) - E_i(y_{1b} | d = 1) = BA_i$$

同じく対照群の事前事後の違いは次のようになる。

$$TE_j = E_j(y_{0a} | d = 0) - E_j(y_{0b} | d = 0) = BA_j$$

ここで対照群の時間を通じた変化は実際は処理効果ではなく、無処理であっても変化した部分ということであり、その部分を含むと処理群の処理効果が過大評価されることになる。従って上の 2 式の差をとったものが DID 推定といわれるものである。

$$DID = TE_i - TE_j = BA_i - BA_j = E_i(y_{1a} - y_{1b} | d = 1) - E_i(y_{0a} - y_{0b} | d = 0)$$

ここでも処理群の選択がランダムに行われているのであれば、処理群の平均 $\overline{BA_i}$ から対照群の平均 $\overline{BA_j}$ を引くことによって平均 \overline{DID} を導くことができる。選択に何らかの変数が影響を与えているのであれば、その変数のもたらすバイアスは何らかの意味でコントロールする必要がある。よく知られている問題は、処理を受けることがわかった時点で y が事前に引き下げられる場合がある (Ashenfelter's dip)。この場合は処理群の一時的な y の低下を考慮して、過去の平均を用いるなどして一時的な低下分を識別する必要がある。

パネルデータとしての具体的分析方法としては、一般に対照群と処理群の事後の y の推定式を次のように定義する。

$$y_{i0a} = x_i' \beta + \gamma y_{ib} + \delta_a + \varepsilon_{ia}$$

$$y_{i1a} = x_i' \beta + \gamma y_{ib} + \delta_a + \alpha d_{ia} + \varepsilon_{ia}$$

ここで δ_a は時間効果 (time drift) を表し、 α は処理群の固定効果を表す。このようなパネルデータに対して固定効果推定を行うと固定効果 α が得られ、これは DID 推定と一致する²。

データがパネルデータではなく、1時点のみのクロスセクションデータであれば、Cross-Section (CS) 推定を用いる。

$$TE = E(y_{1a} | d = 1) - E(y_{0a} | d = 0) = CS$$

ここでの仮定は対照群に入った人の成果 $y_{0a} | d = 0$ が処理群に入った人が入らなかった場合に得られたであろう成果 $y_{0a} | d = 1$ と等しいということである。

² 2 期間モデルでは 1 階差分 (first difference: FD) 推定と固定効果推定は一致するので、DID 推定は FD 推定でも得られる。

ある。ここでは事前のデータは利用できないが、仮に $y_{1b}|d=1$ と $y_{0b}|d=0$ が等しければ、 $CS = DID$ となるが、一般には、等式は成り立たない。

次に、処理群の選択が内生的に決まっている場合のバイアスを考慮した推定方法を紹介したい。これはヘックマンの2段階推定法と同じ考え方で、まず、処理群に選ばれるためのプロビット推定を行い、そこで得られた逆ミルズ比を用いてセレクション・バイアスの修正を行う。

$$y_{i0} = x'_i \beta_0 + u_{i0}$$

$$y_{i1} = x'_i \beta_1 + u_{i1}$$

$$d_i^* = z'_i \gamma + \varepsilon_i$$

ここで、 d_i^* は次のように定義される潜在変数である。

$$d_i = \begin{cases} 1 & \text{iff } d_i^* > 0 \\ 0 & \text{iff } d_i^* \leq 0 \end{cases}$$

また $E(u_1|x, z) = E(u_0|x, z) = 0$ であると仮定する。この場合の処理効果は次のように表せる。

$$y_{i1} - E[y_{i0}|d_i = 1] = y_{i1} - x'_i \beta_0 + \sigma_{0\varepsilon} \frac{\phi(z'_i \gamma)}{1 - \Phi(z'_i \gamma)}$$

これは次のように書き換えることも出来る。

$$E[y_{i1}|d_i = 1] - E[y_{i0}|d_i = 1] = x'_i(\beta_1 - \beta_0) + (\sigma_{0\varepsilon} - \sigma_{01\varepsilon}) \frac{\phi(z'_i \gamma)}{\Phi(z'_i \gamma)}$$

ここで $(\sigma_{0\varepsilon} - \sigma_{01\varepsilon}) \frac{\phi(z'_i \gamma)}{\Phi(z'_i \gamma)}$ の項は選択バイアスを表している。

先に述べたように、政策を実施した処理群のデータはあるが、対照群のデータは不在であり、外部情報から対処群をみつけてマッチングさせる必要が出てくる場合には次のような手法を用いる。

実際に外部データが十分にあり、処理群に含まれる個別サンプルの全ての変数、属性にぴったり一致するような対照群サンプルを選ぶことが出来れば、これを完全一致マッチング (exact matching) と呼ぶが、分析に用いる変数が増えるに従って、個別変数をマッチングさせることは難しくなる。このような場合、変数一つ一つをマッチングさせるのではなく、ある程度変数を集約して表現した条件付き確率 (これを propensity score と呼ぶ) を処理群と対照群でマッチングさせるという方法が考えられる³。

具体的な考え方は、処理群に選ばれる確率 $[\text{Pr}[d_i = 1|x]]$ を全サンプルを用いたロジット推定によって求め、それを処理群と対照群に分け、さらに確

³この方法は Rosenbaum and Rubin (1983) によって開発され、現在では以下で紹介するように沢山の拡張が行われている。

率を均等な階層に分け、同じ階層に入るもの同士をマッチングさせ、処理効果の平均を求めるといふものである。ここでマッチングをどうするかということが問題になる。すなわち、一度マッチングに使った対照サンプルを再び使うことを認めるかどうか、比較対照するために処理サンプルに対していくつの対照サンプルを割り当てるのか。最も propensity score が近いもの一つを選べばいいのか (caliper matching とする)、それともその周辺の対照サンプルを複数割り当てるのがいいのか、また、具体的なマッチングの方法としてどのようなものを用いるのか、といった問題がある⁴。

マッチングの詳細な手法はかなり技術的に高度になるので、ここでは扱わないが、主要な手法の基本的な考え方を紹介しておきたい⁵。一般に処理効果は次のように表すことができる。

$$\Delta^M = \frac{1}{N_T} \sum_{i \in \{d=1\}} [y_{i1} - \sum_j w(i, j) y_{j0}]$$

ここで $\sum w(i, j) = 1$ 、 $0 < w(i, j) \leq 1$ となるマッチング・ウェイトである。 N_T は処理群のサンプル数を表す。

対照群としてどのようなサンプルを処理群にマッチさせるかという事であるが次のようなマッチング手法が提案されている。(1) 最近隣マッチング (nearest-neighbor matching) の考え方では、全ての処理サンプル i に対して、次のような条件を満たす集合 $A_i(x) = \{j | \min_j \|x_i - x_j\|\}$ を対照群として選択する⁶。(2) カーネル・マッチング (kernel matching) では、ウェイトを次のように定義する $w(i, j) = K(x_j - x_i) / \sum_{j=1}^{N_{ic}} K(x_j - x_i)$ 、ここで K はカーネル関数を表す。(3) 層化マッチング (stratification matching) とは propensity score を均等に層化し、層内で処理群と対照群が同じスコアになるようにした後で、処理効果を推定する。同じスコアのペアが組めない場合には、その層内での処理効果は計算されない。(4) 半径マッチング (radius matching) では対照群集合を次のように定義する。 $A_i(p(x)) = \{p_j | \|p_i - p_j\| < r\}$ すなわち、propensity score の差が半径 r 以内であればペアとしてマッチングするという方法である。

4 職業訓練の賃金効果の推定

これまで述べてきた政策評価の手法を具体的なデータに当てはめてみよう。ここでは、Lalonde (1986) によって使われ、その後、Dehejia and Wahba (1999, 2002) によって再検討され、Cameron and Trivedi (2005, Chapter 25)

⁴実際にこれらの問題にどう対処するかということは研究者の判断にゆだねられている。逆に、決定的に正しい方法が知られているわけではなく、試行錯誤するしかない。

⁵最新のマッチング手法のアルゴリズムに関しては、Abadie et al (2004)、Becker and Ichino (2002)、Becker and Caliendo (2007) 等を参照されたい。

⁶ $\|\cdot\|$ はベクトル間のユークリッド距離を表す。

によってデータが公開された、the National Supported Work (NSW) による 1970 年代の職業訓練調査を用いる。この NSW は処理群と対照群を含んだ社会実験であるが、Lalonde(1986) は対照群を外部データと差し替えて、非実験の場合の処理効果を測ることによって、実験そのもののもつバイアスについて指摘した政策評価に関する重要な研究である。

ここで用いるデータセットには、1976-77 年に職業訓練を受けた 185 人の男性と、その処理群に対して Panel Study of Income Dynamics (PSID) から 2490 人の 55 歳以下の家計主である男性が対照群として選ばれている。このデータに含まれる変数の定義と平均値は表 1 に載せてある。処理群と対照群を比べると、対照群の方が年齢も高いし、学歴も高い。高卒以下の学歴の人は処理群では 71% であるのに対して、対照群では 30% に過ぎない。黒人も処理群が断然高い。失業経験も処理群の方がはるかに高い。逆に結婚率は対照群が高く、処理群では低い。この結果は、処理群の選択は明らかに無作為抽出ではなく、かなりの偏りをもって選ばれていることがわかる。職業訓練の賃金効果を測定する場合にはこの点に注意する必要がある。

ここで職業訓練の成果として用いるのは 1982 年のドル価値で測った 1978 年の実質賃金 (RE78) である。表 2 には 4 種類の評価方法で計算した処理効果が記載されている (処理効果は網掛けで表示してある)。最も簡単な処理効果の測定は 1978 年における処理群の平均から対照群の平均を引いたものである (treatment-control comparison) が、これは \$-15205\$ となっている。容易にわかるように、これは職業訓練の結果伸びた実質賃金を比較しているのではなく、1 時点における 2 つのサンプルの平均を比較しているにすぎない。次に推定すべきは主要な変数をコントロールした上で、処理効果をみる方法である (control function estimator と呼ばれている)。この OLS 推定式は次のように書くことができる。

$$RE78_i = x_i' \beta + \alpha d_i + u_i$$

表 2 から明らかなように $\alpha = \$217.944$ であり、処理効果はプラスになっている。

3 番目の方法は before-after comparison (BA) である。これは処理群に関して RE78 と RE75 の差をとって平均することによって求めることができる。すなわち、表 1 より $\$6349 - \$1532 = \$4817$ となる。4 番目は differences-in-differences (DID) である。対照群の RE75 から RE78 への伸びを計算し ($\$21553.920 - \$19063.340 = \$2490.58$)、それを先ほどの処理群の伸びから引くと DID が求まる ($\$4817 - \$2491 = \$2326$)。これは次の OLS 推定式によっても求まる。

$$RE_{it} = \phi + \delta D78_{it} + \gamma \alpha d_i + \alpha D78_{it} * d_i + u_i$$

ここで、D78 は 1978 年ダミー、 d = 処理効果を表すダミー。 $t = 1975, 1978$ 。

係数 α が求める DID 推定と一致する。

表 3 は propensity score を求めるための準備としてのロジット推定の結果が載せてある。この推定結果を用いて層化マッチングを行い、平均処理効果を推定すると\$995 であることがわかった。前節で論じたようにマッチングには様々な方法があり、結果も一様ではないが (Cameron and Trivedi (2005, p.895) 参照)、総合的に判断すると平均処理効果は\$1000-\$2000 程度であったようだ。

5 おわりに

本章では政策評価の方法について解説したが、政策評価を厳密に行うと、効果は低く出てしまうと思われた方もいるかもしれない。確かに、条件をコントロールしない比較では政策効果を過大評価しがちであり、完全ではないにしても条件を出来る限りコントロールすれば、平均処理効果を見る限り効果は低く出がちである。本章では明示的には議論しなかったが、政策効果は強く出る主体と、ほとんど効かない主体に分かれることも事実である。平均処理効果ではなく、個別の階層、地域、主体属性に効く政策というものがあるのか、あるとすればそれはなぜかといった問題もおもしろい課題である。医薬品や医療処置であれば、万能薬や万能療法があるわけではなく、患者別に効果が違うことは容易に理解できだろう。経済政策もさらに細かいミクロレベルに降りていって政策の効き方がなぜ違うのか、そのような政策は意図して特定の人に効果があったのだろうか、それとも偶然だろうか、といった問題が解明できれば、経済政策に対する理解も質も大幅に改善されるのではないだろうか。

6 STATA コード

本章で用いたデータは Lalonde(1986) で用いられたものであるが、Cameron のホームページ [www.econ.ucdavis.edu /faculty/cameron](http://www.econ.ucdavis.edu/faculty/cameron) に入っている `nswpsid.dal` としてダウンロードできる。以下のプログラムは Cameron and Trivedi (2005, Chapter 25) で用いられたもの (`MMA25P1TREATMENT.DO`) を踏襲しているが、編集している。

```
set more off
using nswpsid.dal
```

```
/*Data Generation*/
```

*もとのデータでは 1974 年の失業率 (U74) と 1975 年の失業率 (U75) が逆になっているので、それを再定義する。

```
drop U74 U75
gen U74 = cond(RE74 == 0, 1, 0)
gen U75 = cond(RE75 == 0, 1, 0)
```

*** 追加的なデータを作る。**

```
gen AGESQ = AGE*AGE
gen EDUCSQ = EDUC*EDUC
gen NODEGREE = 0
replace NODEGREE = 1 if EDUC < 12
gen RE74SQ = RE74*RE74
gen RE75SQ = RE75*RE75
gen U74BLACK = U74*BLACK
gen U74HISP = U74*HISP
sum AGE EDUC NODEGREE BLACK HISP MARR U74 U75 RE74
RE75 RE78 TREAT AGESQ EDUCSQ RE74SQ RE75SQ U74BLACK
U74HISP
/*表 1 */
bysort TREAT: sum AGE EDUC NODEGREE BLACK HISP MARR
U74 U75 RE74 RE75 RE78 TREAT AGESQ EDUCSQ RE74SQ RE75SQ
U74BLACK
```

/*表 2 Treatment-control comparison */

```
regress RE78 T
regress RE78 TREAT, robust
```

/*表 2 Control function estimator */

```
regress RE78 TREAT AGE AGESQ EDUC NODEGREE BLACK HISP
RE74 RE75
regress RE78 TREAT AGE AGESQ EDUC NODEGREE BLACK HISP
RE74 RE75, robust
```

***処理効果との交叉項を作る**

```
gen TAGE = TREAT*AGE
gen TAGESQ = TREAT*AGESQ
gen TEDUC = TREAT*EDUC
gen TNODEGREE = TREAT*NODEGREE
gen TBLACK = TREAT*BLACK
gen THISP = TREAT*HISP
gen TRE74 = TREAT*RE74
gen TRE75 = TREAT*RE75
```

```
regress RE78 TREAT AGE AGESQ EDUC NODEGREE BLACK HISP
RE74 RE75 TAGE TAGESQ TEDUC TNODEGREE TBLACK THISP
TRE74 TRE75
```

```
/*表 2 Differences-in-differences*/
```

```
gen id = _n
label variable id "id"
gen EARN1 = RE75
gen EARN2 = RE78
reshape long EARN, i(id) j(year)
gen dyear2 = 0
replace dyear2 = 1 if year==2
gen Tdyear2 = TREAT*dyear2
regress EARN Tdyear2 TREAT dyear2
regress EARN Tdyear2 TREAT dyear2, robust
```

```
/*表 2 Before-after comparison*/
```

```
regress EARN Tdyear2 if TREAT==1
regress EARN Tdyear2 if TREAT==1, robust
```

```
/*表 3 Propensity score の計算*/
```

```
logit TREAT AGE AGESQ EDUC EDUCSQ MARR NODEGREE BLACK
HISP RE74 RE75 RE74SQ RE75SQ U74BLACK
predict PSCORE
```

* 以下では propensity score の計算プログラムを示す。

```
sum PSCORE if TREAT==1
scalar PTMIN = r(min)
scalar PTMAX = r(max)
sum PSCORE if TREAT==0
scalar PCMIN = r(min)
scalar PCMAX = r(max)
drop if PSCORE < PTMIN
drop if PSCORE < PCMIN
drop if PSCORE > PTMAX
drop if PSCORE > PCMAX
sum PSCORE
gen PSCORESQ = PSCORE*PSCORE
regress RE78 TREAT PSCORE PSCORESQ
```

*Propensity score を 10 層に分ける

```
global cut1 = 0.1
global cut2 = 0.2
global cut3 = 0.3
global cut4 = 0.4
global cut5 = 0.5
global cut6 = 0.6
global cut7 = 0.7
global cut8 = 0.8
global cut9 = 0.9
gen STRATA = 1
replace STRATA = 2 if PSCORE > $cut1 & PSCORE <= $cut2
replace STRATA = 3 if PSCORE > $cut2 & PSCORE <= $cut3
replace STRATA = 4 if PSCORE > $cut3 & PSCORE <= $cut4
replace STRATA = 5 if PSCORE > $cut4 & PSCORE <= $cut5
replace STRATA = 6 if PSCORE > $cut5 & PSCORE <= $cut6
replace STRATA = 7 if PSCORE > $cut6 & PSCORE <= $cut7
replace STRATA = 8 if PSCORE > $cut7 & PSCORE <= $cut8
replace STRATA = 9 if PSCORE > $cut8 & PSCORE <= $cut9
replace STRATA = 10 if PSCORE > $cut9
tab STRATA T
* 同一層内での平均の比較を行う
tab STRATA TREAT, sum(AGE) nostand nofreq
tab STRATA TREAT, sum(EDUC) nostand nofreq
tab STRATA TREAT, sum(MARR) nostand nofreq
tab STRATA TREAT, sum(NODEGREE) nostand nofreq
tab STRATA TREAT, sum(BLACK) nostand nofreq
tab STRATA TREAT, sum(HISP) nostand nofreq
tab STRATA TREAT, sum(RE74) nostand nofreq
tab STRATA TREAT, sum(RE75) nostand nofreq
tab STRATA TREAT, sum(U74BLACK) nostand nofreq

* 同一層内での平均差の検定を行う
bysort STRATA: oneway EDUC T
#delimit ;
global sum = 0 ; /* Sums the estimate of interest over strata ;
global sumwgt = 0 ; /* Sums the number of treated obs over strata */
global count = 0 ; /* This gives the number of Strata used */
global numcut = 10;
global XLIST AGE AGESQ EDUC NODEGREE BLACK HISP RE74
RE75;
```

```
forvalues i = 1/$numcut { ;
  global addon = 0 ; /* Within strata estimate of interest */
  global tobs = 0 ; /* Within strata number of treated obs */
  capture { ;
    quiet regress RE78 TREAT $XLIST if STRATA == 'i' ;
    global addon = _b[TREAT] ;
    quiet summarize TREAT if TREAT==1 & STRATA=='i' ;
    global tobs = _result(1) ; * # of treatment observations ;
  } ;
  di "'i' estimate = $addon Top cut = ${cut'i}' #treat obs = $tobs" ;
  if $addon ~ = 0 { ;
    global sum = $sum + $addon * $tobs ;
    global sumwgt = $sumwgt + $tobs ;
    global count = $count + 1 ;
  } ;
} ;
#delimit cr ;

* Prppensity score の加重平均をとる
di $sum / $sumwgt " Count = " $count
```

参考文献

- [1] 北村行伸 (2005) 『パネルデータ分析』、岩波書店
- [2] Abadie, Alberto, Drukker, David, Herr, Jane Leber, and Imbens, Guido W.(2004) "Implementing Matching Estimators for Average Treatment Effects in Stata", *The Stata Journal*, 4(3), pp.290-311.
- [3] Becker, Sascha O. and Ichino, Andrea.(2002) "Estimation of Average Treatment Effects Based on Propensity Scores", *The Stata Journal*, 2(4), pp.358-377.
- [4] Becker, Sascha O. and Caliendo, Marco. (2007) "mhbounds-Sensitivity Analysis for Average Treatment Effects", *The Stata Journal*, 7(1), pp.71-83.
- [5] Cameron, A.C. and Trivedi, P.K.(2005) *Microeconometrics: Methods and Applications*, Cambridge University Press.

- [6] Dehejia, R.H. and Wahba, S. (1999) "Reevaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, 94, pp.1053-1062.
- [7] Dehejia, R.H. and Wahba, S. (2002) "Propensity Score-Matching Methods for Nonexperimental Causal Studies", *The Review of Economics and Statistics*, 84(1), pp.151-161.
- [8] Heckman, J.J. and Robb, R. (1985) "Alternative Methods for Estimating The Impact of Interventions", in J. Heckman and B. Singer (eds), *Longitudinal Analysis of Labor Market Data*, Cambridge University Press.
- [9] Heckman, J.J. and Smith, J.A. (1998) "Evaluating the Welfare State", in Strøm, S. (ed) *Econometrics and Economic Theory in the 20th Century: The Ragner Frisch Centennial Symposium*, Cambridge University Press., pp.241-318.
- [10] Holland, P.W. (1986) "Statistics and Causal Inference", *Journal of the American Statistical Association*, 81, pp.945-960.
- [11] Lalonde, R. (1986) "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", *American Economic Review*, 76, pp.604-620.
- [12] Lee, Myoung, Jae. (2005) *Micro-Econometrics for Policy, Programm, and Treatment Effects*, Oxford University Press.
- [13] Moffitt, R. (1991) "Program Evaluation with Nonexperimental Data", *Evaluation Review*, 15., pp.103-120.
- [14] Rosenbaum, P. and Rubin, D.B. (1983) "The Central Role of Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70, pp.41-55.
- [15] Rubin, D.B. (1974) "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Education Psychology*, 66, pp.688-701.
- [16] Winkelmann, Rainer and Boes, Stefan. (2006) *Analysis of Microdata*, Springer.
- [17] Wooldridge, Jeffrey. M. (2003) *Econometric Analysis of Cross Section and Panel Data*, The MIT Press

表1 変数の定義と平均値の比較

変数	定義	全体 Mean	処理群 Mean	対照群 Mean
AGE	年齢	34.226	25.816	34.851
EDUC	教育年数	11.994	10.346	12.117
NODEGREE	教育年数が12より小さいダミー	0.333	0.708	0.305
BLACK	黒人ダミー	0.292	0.843	0.251
HISP	ヒスパニックダミー	0.034	0.059	0.033
MARR	結婚ダミー	0.819	0.189	0.866
U74	1974年の失業ダミー	0.129	0.708	0.086
U75	1975年の失業ダミー	0.135	0.600	0.100
RE74	1982年のドル価値で1974年の実質賃金	18230.000	2095.574	19428.750
RE75	1982年のドル価値で1975年の実質賃金	17850.890	1532.056	19063.340
RE78	1982年のドル価値で1978年の実質賃金	20502.380	6349.145	21553.920
TREAT	処理群=1、対照群=0	0.069	1.000	0
AGESQ	年齢の二乗	1281.610	717.395	1323.530
EDUCSQ	教育年数の二乗	153.186	111.060	156.316
RE74SQ	1982年のドル価値で1974年の実質賃金の二乗	5.21E+08	2.81E+07	5.57E+08
RE75SQ	1982年のドル価値で1975年の実質賃金の二乗	5.11E+08	1.27E+07	5.48E+08
U74BLACK	黒人で1974年に失業しているダミー	0.055	0.600	0.014
U74HISP	ヒスパニックで1974年に失業しているダミー	0.006	-	-
Sample Size		2675	185	2490

表2 職業訓練の賃金効果の推定

Dependent Variable: RE78	Treatment-control comparison		Control function estimator		Before-after comparison		Differences-in-differences	
	Coefficient	Robust z-ratio	Coefficient	Robust z-ratio	Coefficient	Robust z-ratio	Coefficient	Robust z-ratio
TREAT	-15205	-23.18	217.944	0.28			-17531.280	-48.62
AGE			158.506	1.05				
AGESQ			-3.233	-1.54				
EDUC			564.624	4.64				
NODEGREE			502.091	0.79				
BLACK			-699.335	-1.62				
HISP			2226.535	1.83				
RE74			0.279	4.51				
RE75			0.568	8.56				
Tdyear2					4817.09	7.71	2326.505	3.11
dyear2							2490.585	6.01
_cons	21553.92	69.13	-2836.703	-0.97	1532.056	6.47	19063.340	69.95
Number of observation	2675		2675		370		5350	
R-squared	0.061		0.586		0.139		0.087	
Root MSE	15152		10075		6010.8		14185	

表3 Propensity Score のためのロジット推定

Dependent Variable: TREAT	Coefficient	Robust z-ratio
AGE	0.331	2.75
AGESQ	-0.006	-3.42
EDUC	0.825	2.33
EDUCSQ	-0.048	-2.60
MARR	-1.884	-6.29
NODEGREE	0.130	0.30
BLACK	1.133	3.22
HISP	1.963	3.46
RE74	0.000	-2.95
RE75	0.000	-5.23
RE74SQ	0.000	3.59
RE75SQ	0.000	0.24
U74BLACK	2.137	5.00
_cons	-7.552	-3.08
Number of observation	2675	
Log Likelihood	-204.9295	
LR chi2(13)	935.44	
Prob>chi2	0.000	
Pseudo R2	0.695	