＜財務省財務総合政策研究所「フィナンシャル・レビュー」July―2005＞

積極労働政策の評価―レビュー[*1]

黒澤　昌子[*2]

要　　約

長期雇用という体制が弱体化しつつある中，求職と求人の結びつきを支援する職業紹介や職業能力開発への支援を充実させる等の積極労働政策の重要性が高まっている。労働政策に大きな変革が求められている今，今後の具体的な施策を模索する上でも，これまで行われてきた数々の施策に対する評価を費用―便益分析の枠組みで検証することは有益であるが，わが国でのそうした分析の蓄積は少ない。その背景には，データの未整備だけでなく，セレクション・バイアスなど，分析に伴う労働政策特有の技術的問題の存在がある。そこで本稿では，主に米国での事例を踏まえながら，職業訓練や職業紹介等のプログラム評価における分析的枠組みとその方法論についてのレビューを行い，費用―便益分析の枠組みで労働政策を評価する上で留意すべき点について整理することを通して，今後の日本におけるデータ整備をはじめとする労働政策評価の円滑な実施に資することを目的とする。

Ⅰ．はじめに

わが国をとりまく経済社会環境の大きな変化のなかで，企業主導による豊富な人的資本投資に支えられた長期雇用という体制は弱体化しつつある。こうした状況においては，企業での再訓練を通して雇用維持を促進させる支援だけでなく，人々が企業に頼ることなく自らの能力に投資し，労働市場を介して円滑に転職できるような環境の整備として，求職と求人の結びつきを支援する職業紹介や，スキルを付加することを通して企業の人材ニーズと労働者の職業能力とのミスマッチを解消する職業能力開発への支援を充実させる等の積極労働政策が今まで以上に求められるようになっている。

労働政策に大きな変革が求められている今，今後の具体的な施策を模索する上でも，これまで行われてきた数々の施策に対する評価を費用―便益分析の枠組みで検証することは有益であるし，政策評価に基づき希少な財源を最も効率的に利用するような施策の取捨選択を行うことは，国民に対する説明責任という成熟した民主主義社会の要請を果たすためにも欠かせない。しかしながら，特にわが国ではそうした分析の蓄積は少ない。

検証を可能にするデータを入手することの難しさがその最たる要因であろう。しかしそれ以外にも，プログラムの効果を他の要因から識別することの難しさ，さらにはその効果がプログラムに参加した個人以外にも波及することなど，

＊1　井堀利宏氏，林正義氏には貴重なコメントをいただいたことをここに感謝する。
＊2　政策研究大学院大学教授

分析作業に立ちはだかる多くの技術的問題の存在も無視できない。とくに労働施策では提供されるプログラムへの参加が個人の選択に基づいて決められることから，参加者と非参加者の違いからプログラム効果を推定しようとする限り，推定にバイアスの生じる可能性が高く，政策論議に資する頑健な政策効果の推定を行うのは極めて難しい。

実際，職業訓練施策評価の歴史が長い米国では，施策効果の推定値が利用された推定手法や対照群の選択に敏感に反応し，整合的な結果を得られなかったことを受けて，訓練希望者を無作為に訓練に割り当てるという実験的方法によるデータ収集にまで踏み込んだ経緯がある。こうした「実験」も決して万能薬とはいえないが，実験データに基づく推定値をベンチマークとすることで，さまざまな非実験的推定方法や対照データの妥当性の検証が可能となった。本稿の第一の目的は，そうした施策評価方法を評価した一連の研究に基づき，職業訓練や職業紹介といった積極的労働政策の評価における分析的枠組みと方法論についてのレビューを行うことである。

第二の目的は，費用―便益分析の枠組みで労働政策を評価する上で留意すべき点を整理することである。こうした方法論に沿った分析から得られるのは，プログラム参加者における「平均的な効果」であり，その値に１人あたりの費用を対応させることによって，通常の費用―便益分析の枠組みでの労働政策の評価が可能となる。しかしながらプログラム効果は個人によって多様であるし，労働政策の目標は効率性だけではない。公平性という政策目標が達成されているかどうかを見極めるためには，参加者における平均値以外のプログラム効果の特性値にも注目する必要がある。また，労働政策は，それが多くの人々を対象にしたものほど，置き換えや代替効果，あるいは賃金や価格の変化を通して市場全体にも影響を与える可能性が高くなる。したがって，政策の包括的な効果を計測するためには施策の直接的効果だけでなく間接的効果についても計測する必要がある。

このように，そもそもどういった「成果」を対象に，誰に与える影響を測定するのか，そしてそのためにはどのようなデータを構築し，どのような分析手法をとるべきか。こうした観点からこれまで行われてきた施策実証研究を評価し，その問題点を整理することは，今後の日本におけるデータ整備をはじめとする労働政策評価の円滑な実施に資すると思われる。

まず第Ⅱ～Ⅲ節では，施策の提供するプログラムへの参加が参加者に与える平均的効果の測定に焦点をおき，その分析的枠組みと方法論を考察する。とくに第Ⅱ節では，計測上の主な問題であるセレクション・バイアスのメカニズムと実験的データを用いた対処法を明らかにし，次いで第Ⅲ節では非実験的データを用いた推定法として，操作変数法，２段階推定法，difference–in–differences 推定法，ならびに近年発展が目覚しいマッチング法の利点や弱点を，実証研究例を踏まえながら考察する。第Ⅳ節では，参加者における平均的効果に基づいた費用―便益分析の方法とその問題点を整理する。

Ⅱ．施策評価の枠組みとサンプル・セレクション

Ⅱ－１．施策効果指標

本節では，就業状態などにおいて一定の条件を満たしている者が，自発的意思に基づいて参加できる職業紹介や職業訓練などの積極労働政策の一環として提供されるプログラムを想定し，その効果分析の枠組みと方法について考察する。

賃金や就業状況に影響を与える政策は数多く存在するが，単に所得補助を行う政策とは異な

り，積極労働政策には求職者や未熟練層等の特定グループに求職支援や訓練サービス，就労の機会等を提供することを通して，就業率や賃金率を直接向上させる機能が求められる。例えば訓練プログラムが失業者に提供された場合，それは参加者の雇用可能性を高めて労働の質・生産性を向上させるから，失業期間の短縮と就業率の向上，そして就業後の賃金率の増加や福祉依存率の低下などをもたらすことが期待される。実際はこうした効果以外にも，訓練を受けた人々と労働市場で代替的あるいは補完的な立場にいる人々や彼らを雇用する企業，ならびにこうした施策のコストを負担する納税者にも訓練プログラムの影響は及んでいるのであるが，そうした間接的な効果をも含めた推計には一般均衡モデルを用いる必要がある。それについては第Ⅳ節で触れるとして，本節並びに次節では，まずプログラム参加者にもたらされる「直接的」な効果の計測のみに焦点を当てる。

プログラムの成果指標として何をとりあげるのかは，政策目標やデータの入手可能性等に依存する。たとえば訓練プログラムの成果指標としては，就業率や賃金率，ならびに福祉依存率などが考えられるが，これまでの研究では，就業と就業した場合の賃金率とを統合した指標としての収入がその成果指標とされることが多い[1)]。一方，求職支援プログラムについては，失業期間や就業率が成果指標とされることが多いようである。

施策評価の目的は，その成果指標が施策の一環として提供されるプログラムに参加することでどの程度高められるのかを測ることである。本稿では，プログラム参加者が「プログラム」に「参加」する場合の効果を参加しない場合との比較として評価するケース，すなわち施策を変数化すると二値変数となるような施策の効果計測に焦点を当てる[2)]。ここではそうした施策変数をプログラムに参加した場合は1，参加しなかった場合は0をとるダミー変数Dで表し，その値に対応する形でi番目の個人がプログラムに参加した場合の成果指標をY_{1i}，参加しない場合の成果指標をY_{0i}で表すとしよう。

すると，i番目の個人にとって，プログラムに参加することの効果は，そのひとがプログラムに参加することで生じる成果指標の変化$\alpha_i = Y_{1i} - Y_{0i}$によって計測される。ただしこの効果は個人によって異なるであろうから，プログラム評価を行うに当たり，$Y_1 - Y_0$の分布のどういった特性値を推計すべきかを決めなければならない。政策目標によっても有用な特性値は変わりうるが，最も頻繁に推定されるのが当該プログラムに実際に参加した人々におけるプログラムの平均的効果（average treatment effect on the treated）であり，それは次のように定義される[3)]。

$$ATT = E(Y_{1i} - Y_{0i} | D_i = 1) = E(\alpha_i | D_i = 1) \quad (1)$$

プログラムの間接的効果を無視した場合，この数値から参加者一人当たりのプログラム費用を差し引いたものが当該プログラムの純便益であり，いわゆる費用―便益分析で用いられる指標となる。

そのほか，対象となる母集団全体から任意抽出された人々における当該プログラムの平均的効果を示す特性値としてaverage treatment effectがある。

1）収入を成果指標とした場合は，訓練後1年程度の間に得られた総勤労所得を用いることが多い。訓練後全く就業していなかったサンプルの収入は0となる。

2）たとえば訓練プログラムといっても，訓練量や頻度，あるいは教室型訓練やOJTなどの訓練内容において多種多様のコースや形態が存在するかもしれない。その場合は，参加者の受ける訓練量や内容が互いに排反となるようにプログラムを分類し（すなわちDは二値変数ではなく，3以上の値をとる変数となる），それぞれに異なるYの値を対応させることで基本的には本節に示す分析と同じような枠組みによって対処することができる。マッチング推定法への適用については，例えばLechner（2002）を参照。

3）平均以外の特性値については本稿の第Ⅳ節を参照。

$$ATE = E(Y_{1i} - Y_{0i}) = E(\alpha_i) \quad (2)$$

プログラム拡充の是非を考えるのであれば，現行のプログラム対象外の人々も含めた母集団についてのATEを計測することも有用となろう。

ATTとATEの違いは，成果指標Yをその平均値μと誤差uとに分離するとわかりやすい[4]。

$$Y_{1i} = \mu_1 + u_{1i}$$
$$Y_{0i} = \mu_0 + u_{0i} \quad (3)$$

ここで$\mu_1 = E(Y_{1i})$，$\mu_0 = E(Y_{0i})$，すなわち$E(u_{1i}) = E(u_{0i}) = 0$である。すると，個人iについてのプログラム効果は，

$$\alpha_i = Y_{1i} - Y_{0i} = (\mu_1 - \mu_0) + (u_{1i} - u_{0i})$$
$$= ATE + u_{1i} - u_{0i} \quad (4)$$

と表される。つまり，ATEはプログラム効果の母集団全体における平均値$\mu_1 - \mu_0$であり，$u_{1i} - u_{0i}$はi番目の個人に特有なプログラム効果をその平均値からの乖離として示した部分である。(4)のD＝1についての条件付き期待値をとるとATTを得ることができる。

$$ATT = (\mu_1 - \mu_0) + E(u_{1i} - u_{0i} | D_i = 1)$$
$$= ATE + E(u_{1i} - u_{0i} | D_i = 1) \quad (5)$$

したがってATTとATEとの差は，プログラム参加者に特有なプログラム効果の平均になっていることがわかる。するとATT＝ATEとなるのは$u_{1i} = u_{0i}$，すなわち全ての人にとってプログラム効果が同じであるか，あるいは$E(u_{1i} - u_{0i} | D_i = 1) = 0$，すなわちプログラム効果に個人的な差があったとしても参加者においてその差は平均的に0になる，のいずれかを満たす場合である。後者はプログラムに参加するかどうかの決定が個人特有なプログラム効果$u_{1i} - u_{0i}$の大きさに影響されないことを仮定するものである。

あらゆる個人がプログラムから同一のインパクトを受けるという仮定は制約的であり，一般的にはプログラムに参加することの効果が大きいと期待される者ほど参加する確率は高いと考えられるが，後にみるように，このことは推定方法に大きな影響を与える。

Ⅱ－2．サンプル・セレクションの問題

プログラム効果の推計における問題は，同一人物についてY_0とY_1を同時に観察することができない点にある。我々が観測できる成果指標Yの構造は次のようになっている。

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i} = Y_{0i} + D_i (Y_{1i} - Y_{0i}) \quad (6)$$

プログラムの効果は，それに参加することで，参加しなかったら得られたであろう場合と比べてYがどれだけ変化するかによって測られる。ところが参加者について，参加しなかったら得られたであろうYの値は観察できないし，参加しなかった人について参加したら得られたであろうYの値も観察できない。そこでプログラム評価では，個別のα_iからαの分布の特性値，とくにその平均値の推定に焦点が移される。

通常は，観察されないY（仮想現実＝counter-factual）の平均値を観察される比較対照グループのYの平均値で代替するという方法がとられる。たとえばATTの場合，参加者が参加しなかったら得たであろう成果Y_0は観察されないから，その平均$E(Y_{0i} | D_i = 1)$を参加しなかったプログラム対象者のYの平均値$E(Y_{0i} | D_i = 0)$で代替し，$E(Y_{1i} - Y_{0i} | D_i = 1)$を

$$E(Y_{1i} | D_i = 1) - E(Y_{0i} | D_i = 0) \quad (7)$$

として推定する。しかしながらこれが(1)で定義されたATTと一致するのは，次の(8)式の後半部分（バイアス）が0になるときだけであり，実際にはそれが満たされないケース，すなわちプログラムに参加しなくても参加者と非参加者のYに違いのあることが多い。

4）以下はHeckman, Lalonde, and Smith（1999）に拠る。

$$ATT = E(Y_{1i} - Y_{0i} | D_i = 1) = E(Y_{1i} | D_i = 1) - E(Y_{0i} | D_i = 0) + \{E(Y_{0i} | D_i = 0) - E(Y_{0i} | D_i = 1)\} \quad (8)$$

たとえば訓練プログラムの場合，就業機会の確保が困難な者ほど訓練を受講しようとするかもしれないし，訓練機関もそうした人々にはより積極的に受講を斡旋するかもしれない。すると訓練受講者は，熟練度の低い者，つまり訓練をしなくても，訓練を受講しない者と比べて生産性の低い者であることが多く（$E(Y_{0i}|D_i = 1) < E(Y_{0i}|D_i = 0)$），その場合，(7)に基づく訓練効果は(8)よりも小さくなるであろう。反対に，訓練修了者の収入などで測られた訓練の事後評価が訓練機関ごとに義務付けられているような場合には，訓練の評価を高めるために，高い収入が得られそうな，つまりそもそも能力の高い者を選んで訓練を受講させるかもしれない[5]。あるいはやる気のある者ほど訓練受講を希望するかもしれない。こうした状況では，(7)に基づく訓練効果の方が(8)よりも大きくなってしまう。

このように，プログラムへの参加についての選択（セレクション）が，個人あるいはプログラムの割り振りを行う運営機関によって合理的に行われた結果，参加者と非参加者とが同じ母集団から抽出されなくなることによってもたらされるバイアス（セレクション・バイアス）の生じる可能性がある。こうしたバイアスの問題を回避しながらプログラムの平均的効果を推定する方法として，次に述べる実験的データの構築に基づく推定がある。

Ⅱ－３．実験的方法とその問題点

実験的方法は，プログラムの運営機関が，参加資格を持つ参加希望者を，参加者（トリートメント）と非参加者（コントロール）とに無作為に割り当て，それら２つのグループについて収集したデータに基づいて訓練の効果を計測するという方法である。このようにして収集されたデータは，「実験的データ」と呼ばれる[6]。こうしたデータに基づく分析手法は，従来から医学などの分野では頻繁に用いられてきたが，80年代後半より，とくに米国において経済学の分野でも応用されはじめている。

実験的データを用いると，なぜセレクション・バイアスを回避できるのであろうか。プログラムへの参加を希望し，通常であれば参加する個人 i に対して無作為割り当てが行われるとする。コントロールに割り当てられた場合は $T_i = 0$，トリートメントに割り当てられた場合は $T_i = 1$ であるとしよう。

無作為割り当てが行われる母集団と，通常の運営における参加者の母集団（$D = 1$）とが等しい限り，コントロール・グループのメンバーは，コントロールに割り当てられなければ参加していたであろう者によって構成されることになるから，そのグループの平均収入 $E(Y_i|D_i = 1, T_i = 0)$ は，参加者が参加していなかったら得られたであろう収入の平均値 $E(Y_{0i}|D_i = 1)$ に等しくなる。一方，トリートメント・グ

5）これはいわゆるクリーミングといわれる現象である。

6）訓練プログラムにも，内容や形態によって幾つかのコースが存在する場合，プログラム全体に対する受講希望者を各コースに無作為に割り当てるのが最も理想的な実験的データであろう。こうした実験的データがあれば，コースによる効果の違いについても推定することができる。しかしたとえば米国の訓練施策，JTPA の場合，受講希望者は通常，運営機関の職員によって審査され，適性に沿ったコースの勧奨を受ける。したがって，職員によるスクリーニング以前の段階で受講希望者を各コースへ無作為に割り当てても，そのような実験的データからは，典型的な JTPA プログラムの効果は計測されないであろう。そこで JTPA では，各コースへの勧奨が行われた後の段階でトリートメントとコントロールへの無作為な割り当てが行われた。すなわち，各コースについて，それぞれトリートメントとコントロールグループのデータが構築された。ただし，こうした実験的データより得られる各コースの効果は，そのコースに適していると判断された人々に対する効果であるから，コース別の効果を，同一人物に対する効果として比較することはできない（Bloom et al.（1997））。

ループの平均収入 $E(Y_i|D_i=1, T_i=1)$ は，通常の運営における参加者の平均収入 $E(Y_{1i}|D_i=1)$ に等しい。したがって，実験データに基づくコントロールとトリートメントとの標本平均の差は，プログラム参加者に対する平均的効果（ATT）の不偏推定量になる[7]（Heckman, Ichimura, and Todd，以下 HIT（1997））。

$$\hat{\alpha}_{EX}=\overline{Y}|_{T=1}-\overline{Y}|_{T=0} \qquad (9)$$

ところが，実験的方法にもいくつかの問題点がある。第一に，通常のプログラム運営であれば訓練を受講したであろう者を無作為にコントロール・グループに割り当てることの倫理上の問題や，割り当てのプロセスが無作為に行われることを監視するためにかかる多額の費用などの問題がある（Burtless and Orr（1986），Burtless（1995））。

第二の問題は，コントロール・グループに属する者が，公的なプログラムと代替的な他のサービスを受けることによって生じるバイアス（代替バイアス）やトリートメント・グループの成員全てが割り当てられたプログラムに参加しないことによって生じるバイアス（dropout バイアス）である（Heckman and Smith（1996））。こうした状況があると，コントロールとトリートメント・グループとの標本平均の差が示すのは，プログラムを「割り当て」られることの効果であって，それは必ずしもそのプログラムに「参加」することの効果にはならない。プログラムの ATT を推定するためには，非実験的手法を用いて代替プログラムや dropout をコントロールする必要がある[8]（Heckman, Hohmann, Khoo and Smith（2000））。

プログラム参加者を参加希望者の中から無作為に決定するという方法を適用すること自体が，人々の参加意欲，あるいは運営機関の割り当てプロセスに影響を与える場合，プログラム参加者の特徴が以前と違ったものになるという問題もある。たとえば通常の運営では，特定の基準を満たすと思われる者のみに対して訓練の受講勧奨が行われるとしよう。通常の運営においては受講勧奨を受ける確率が高く，受講を希望する者も，無作為に受講が割り当てられるとなると，受講できなくなる確率が高まるために，受講を希望しなくなるかもしれない。すると無作為割り当てが行われる場合の参加希望者は，通常のプログラム運営における参加希望者とは異なるために，推定される訓練効果が通常のプログラムの効果を示すとは必ずしも言えなくなる。これはランダム化バイアス（randomization bias）と呼ばれる問題である[9]（Heckman and Smith（1996））。

Ⅲ．非実験的方法

Ⅲ－１．非実験的方法[10]

非実験的方法における参加者の比較対照サンプルは，参加者のプログラム前の状況，あるいは調査地域や実施時期の等しい外部調査から再抽出された「参加資格を有するがプログラムに参加しなかった」サンプルの用いられることが

7）割り当ての対象者を「母集団」とする限り，(9)式はそこから任意抽出された人々におけるプログラムの平均的効果（ATE）の不偏推定量でもある。

8）あるいは，追加的仮定の下で，コントロール・グループにおける代替的プログラムを受けた比率，ならびにトリートメント・グループにおける dropout 比率を用いてプログラムの ATT を推定する方法もある（Bloom（1984））。

9）ただし，こうした状況の下であっても，ATE=ATT になる条件が満たされる限りバイアスは生じない。

10）本節の議論は Heckman, Lalonde, Smith（1999）に依拠している。

多い。ところが実験的に構築されたデータと違って参加者と非参加者の母集団は異なり，参加するかどうかの決定が必ずしも無作為ではないために，プログラム効果の識別は困難を極める。

ただし，参加者と非参加者について，その性別や学歴，ならびにプログラムに参加する前の職歴や収入など，Y に影響を与えると考えられる属性情報（X）が入手できるのであれば，D と Y との間に相関が存在しても，X をコントロールすることでプログラム効果が識別される可能性もある。

Y が観察可能なベクトル X の関数 μ と観察不可能な誤差変数 u とに加算分離されるとしよう。このときの説明変数 X は D に影響を受けない変数であり，$E(u_{1i}|X_i)=E(u_{0i}|X_i)=0$ であるとする[11]。すなわち，

$$E(Y_{ji}|X_i)=\mu_j(X_i) \quad \text{for } j=0,1$$

これを⑹式に代入すると，

$$\begin{aligned} Y_i &= \mu_0(X_i)+D_i\{(\mu_1(X_i)-\mu_0(X_i))+(u_{1i}-u_{0i})\}+u_{0i} \\ &= \mu_0(X_i)+D_i\ \alpha(X_i)+\{D_i(u_{1i}-u_{0i})+u_{0i}\} \end{aligned} \quad ⑽$$

を得る。ここで D の係数 $\alpha(X_i)$ は属性 X_i をもつ人をランダム抽出した場合のプログラムの平均的な効果，すなわち $ATE(X_i)$ である。

一方，実際に参加した属性 X_i をもつ人々におけるプログラムの平均的効果（ATT）は次のように表される。これには ATE に加えて個人に特有なプログラム効果（$u_{1i}-u_{0i}$）の平均値が含まれる。

$$ATT(X_i)=E(Y_{1i}-Y_{0i}|X_i,D_i=1)=\mu_1(X_i)-\mu_0(X_i)+E(u_{1i}-u_{0i}|X_i,D_i=1)$$

最終的に必要となるのは，対象となる母集団における X についての平均をとった値，

$ATT=E_X\{ATT(X_i)|D_i=1\}$ である。

さて，⑽式を ATT の関数として書き直すと，

$$Y_i=\mu_0(X_i)+D_iATT(X_i)+\{u_{0i}+D_i[u_{1i}-u_{0i}-E(u_{1i}-u_{0i}|X_i,D_i=1)]\} \quad ⑾$$

となるから，プログラム効果が個人 i によって異なる可能性（heterogeneous effect）を考慮した場合，誤差項が D に依存するという特殊な形をもつことがわかる。⑾式より，X をコントロールした上で，非参加者の Y_0 を参加者の Y_0 の仮想現実とみなした推定量から識別されるのは，

$$E(Y_i|X_i,D_i=1)-E(Y_i|X_i,D_i=0)=ATT(X_i)+\{E(u_{0i}|X_i,D_i=1)-E(u_{0i}|X_i,D_i=0)\} \quad ⑿$$

である。したがって，誤差項と参加状況について $E(u_{0i}|X_i,D_i)=0$ が満たされるか，あるいは満たされないとしても⑿式の後半の括弧内が 0 になり参加者と非参加者におけるバイアスがバランスしない限り，$ATT(X_i)$ はバイアスをもって推定されることになる。

同様に⑽式より，$ATE(X_i)$ のバイアスについても明らかにすることができる。

$$E(Y_i|X_i,D_i=1)-E(Y_i|X_i,D_i=0)=ATE(X_i)+\{E(u_{0i}|X_i,D_i=1)-E(u_{0i}|X_i,D_i=0)+E(u_{1i}-u_{0i}|X_i,D_i=1)\} \quad ⒀$$

この式より，たとえ $E(u_{0i}|X_i,D_i)=0$ が満たされていても，$ATE(X_i)$ は必ずしも識別されないことがわかる。u_0 とプログラム参加の決定（D）が無相関であっても，u_1-u_0 と D とが相関している可能性は高い。したがって，プログラム効果の多様性を考慮した場合，ATE の識別はより難しくなる（Blundell and Dias（2002））。

この点を，μ を X の線形関数（$\mu_j(X_i)=X_i\beta_j$）として表し，重回帰分析の枠組みで考えてみよう。その場合，⑽式は次のように表される。

11）このような加算分離性はマッチング法では必要ない。なお，こうした X の外生性は厳密には必要ではないが，単純化のため仮定する（Heckman（2001））。

$$Y_i = X_i\beta_0 + D_iX_i(\beta_1 - \beta_0) + D_i(u_{1i} - u_{0i}) + u_{0i} \quad (14)$$

このモデルの推定方法は，①$u_{1i}=u_{0i}$，あるいは②$u_{1i}\neq u_{0i}$であっても$E(u_{1i}-u_{0i}|X_i, D_i=1)=0$が満たされているかどうかに大きく依存する。最も制約的ではあるが，まず①，すなわち訓練の効果が同じXをもつ個人について同一（homogeneous effect）であることを仮定すると，⑭式は以下のように表される。

$$Y_i = X_i\beta_0 + D_iX_i(\beta_1 - \beta_0) + u_{0i} \quad (15)$$

プログラム効果はXの値に依存せず，Yの水準をシフトするだけであるという仮定を追加すると，⑮式はさらに単純なダミー変数モデルの形になる[12]。

$$Y_i = X_i\beta_0 + D_i\alpha + u_{0i} \quad (16)$$

ここで$\alpha=\alpha(X_i)=\mu_1(X_i)-\mu_0(X_i)$である。①が満たされるモデルの場合，ATTとATEは等しく，$E(u_{0i}|X_i, D_i)=0$である限り，そのOLS推定量は不偏となる。ただし属性Xが同じであれば個人によるプログラム効果に違いがないというのは，検定するとほとんどの場合棄却されるほどの強い制約であることが知られている（Heckman, Smith and Clements（1997））。

一方②が満たされる場合，⑭式はいわゆるランダム係数モデルと呼ばれる形になる[13]。これはプログラム効果に個人的な差があったとしても，それがプログラムに参加するかどうかの決定には影響を与えないことを仮定したもので，この場合もATTとATEは等しくなる。誤差項の不均一分散を考慮する必要があるが，①のケースと同様，u_0とDとの関連がOLS推定量の不偏性の鍵となる（Heckman, Lalonde and Smith（1999））。

推定上，最もやっかいなのは上記の①も②もあてはまらない場合である。すなわち，プログラム効果に個人による差があり，プログラムへの参加は個々に特有なプログラム効果に基づいて決められているケースである。このような状況においては，操作変数法をはじめとする多くの非実験的推定法が妥当性を失う。

以下では，$E(u_{0i}|X_i, D_i)=0$が満たされない状況においてもATTおよびATEの一致推定量を得るための代表的な非実験的推定法として，操作変数法，2段階推定法，difference-in-differences，ならびにマッチング法を考察する。どの手法を用いるかは，プログラム効果の多様性を仮定するかどうか以外に，最終的に必要な施策効果指標や利用可能なデータの種類にも依存するが，その基準については第Ⅲ－6節で触れる。

Ⅲ－2．操作変数法

手元にあるのが一時点のクロス・セクションデータである場合に利用される代表的な手法に，操作変数法がある。この方法では，成果指標Yには影響を与えないがプログラムへの参加に影響を与える操作変数Zが少なくとも1つ必要となる。まずは$u_{1i}=u_{0i}$，すなわちプログラム効果が一定であるケースを考えよう。

この場合，①Xをコントロールした上でも，Zはプログラム参加の決定に影響を与える（すなわちXをコントロールした参加関数におけるZの係数はゼロではない），ならびに②$E(u_0|X, Z)=E(u_0|X)$，すなわちXをコントロールした上で，Zはu_0と平均独立（mean independent），を満たすことが操作変数Zの条件となる（Heckman, Navarro-Lozano（2003））。そのような操作変数が存在する場合，ZはDを通してのみYに影響を与えるから，Zに起因するYの変化からプログラム参加の効果を識別することができる。このように考えると，プログラムへの参加の割り当てが任意に行われる実験的データを用いた推定も，割り当てを操作変数とし

12）切片以外について$\beta_1=\beta_0$を仮定することを意味する。
13）こうしたモデルを用いた実証研究には，たとえばBjorklund and Moffit（1987）がある。

た操作変数推定法の一種と考えることができる。

たとえば⑯式のダミー変数モデルの場合，仮定①と②の下で⑩式の $\{X, Z\}$ の条件付き期待値をとると，

$$E(Y|X, Z) = \mu_0(X) + \alpha P(D = 1|X, Z)$$

となることから，操作変数が母集団をいくつかのグループに分けるようなカテゴリカル変数であれば，そのうちZのとる2つの値（z^*とz^{**}とする）をとりあげて，αの一致推定量を次のように求めることができる（Heckman, Lalonde, Smith（1999））。

$$\hat{\alpha}_{IV} = \frac{E|_X\{E(Y|X, z^*) - E(Y|X, z^{**})\}}{E|_X\{P(D = 1|X, z^*) - P(D = 1|X, z^{**})\}} \quad (17)$$

このような操作変数としては，外生的な政策や制度の変化を表す変数などが考えられる。たとえばA地区とB地区の居住者が母集団で，プログラムの参加資格がA地区に限定されているとしよう。性別や学歴構成などの属性がコントロールされる限り，Y_0の平均値がB地区でもA地区でも変わらないと想定されるならば，居住地区は有効な操作変数として機能する。もちろん，上記の仮定①と②を満たす操作変数が複数存在すれば，通常の二段階推定法（2SLS）によってαの一致推定量を求めることもできる。

米国では福祉受給家庭出身者，高校中退者などの「不利な立場」にある若年層（16－24才）を対象に，最長2年間，合宿しながら規律の習得から職業訓練までを行うJob Corpsという職業訓練プログラムがあるが，その効果を分析したMallar et al.（1982）では，自宅から訓練所までの距離が操作変数として用いられた。訓練所からの距離が長いほどD＝1になる確率は低いが，訓練所からの距離が直接的にプログラム効果に影響を与えるとは考えにくいため，上記の操作変数の条件を満たしているとされたのである。

しかしながら通常，①と②の条件を満たす適切な操作変数を見つけるのは極めて困難である。①の条件は検定できるが，②については，複数の操作変数が存在する場合の過剰識別検定から間接的に妥当性を推測できるにとどまる。①が満たされているとしても，その相関が弱いとパラメータの漸近バイアスが大きくなるという問題もある。操作変数推定量が一致性をもつためには，大きなサンプルサイズが不可欠であることはいうまでもない。

操作変数法のさらなる問題は，その妥当性が，プログラム効果が同一である，あるいはプログラム効果に多様性があったとしても，人々のプログラム参加の決定が自らに特有なプログラム効果（$u_{1i} - u_{0i}$）に基づいて行われていない，という場合に限定されることである。しかしながら，人々が自らのプログラム効果についてある程度正確に予測できるのであれば，プログラムへの参加を決める際にそうした私的情報を考慮に入れるのは自然であろう。

同じ属性を持つ者であっても，大きなプログラム効果が期待される者ほどプログラムに参加する確率が高いと，Zとプログラム効果α_iとに相関が生じることになる。訓練所までの距離が操作変数に用いられたケースのように，確かにプログラムへの参加行動モデルを考えると，参加費用には影響を与えるが参加の収益には影響を与えない変数は①と②の条件を満たす変数としてふさわしい。しかし，人々が自らのプログラム効果をある程度予測でき，それに基づいてプログラムに参加するかどうかを決められる状況では，訓練所から遠い場所に住んでいるにもかかわらず実際に訓練を受けた人々は，高い参加コストを相殺しても余りある大きなプログラム効果を期待できたから参加したのではないかと考えられる。他方，訓練所に近い受講者にはそれほど高いプログラム効果を期待していなかった者も含まれるのではないだろうか。

すると訓練所からの距離が短いほど訓練を受ける確率（D）は高まるが，受けた人々の中では，距離が短いほどプログラム効果の平均値は低くなるという関係の生じる可能性がある。す

なわちZは参加確率Dだけでなく，参加者あるいは非参加者の構成（観察できない特徴について）にも影響を与えている。訓練所からの距離（Z）は，このようにして観察できない変数を通してもプログラム効果に影響を与えることになるから，訓練所への距離はATE，ATTのいずれの操作変数としても機能しないことになる（Heckman（1997））。

この問題点の解決策のひとつは，操作変数を局所的（locally）に適用することである。Imbens and Angrist（1994）によると，①の条件に参加関数がZについて単調（monotonic）であるという仮定を追加した場合，それと②を満たすZを操作変数として利用した⒄式で示される推定量は，Zのz^*からz^{**}への変化に伴って参加行動を変えるという仮想的母集団におけるプログラムの平均的な効果（local average treatment effect＝LATE）の推定量として解釈されるという。ただし，⒄式と違って，LATEは操作変数Zの特定な値，z^*やz^{**}に依存する。Zが訓練所への距離を表す例を考えると，この効果は距離がz^*からz^{**}に減ることによって（観察不可能な属性を含む他の条件が一定のもとで）訓練に参加しはじめる人々における訓練の平均的効果を示すことになる[14]。

LATEは用いられた操作変数Zの特定範囲における効果として限定されているため，常にその値から政策上有意義な解釈ができるとは限らない。しかし，制度変更の変数がZとしての条件を満たしている場合には政策的にも興味深い含意が得られる。たとえばDが訓練プログラムの受講を示す場合，LATEは「政策がAからBに変更された場合にその変更によって訓練を受講するようになった人々におけるプログラム効果の平均値」といった意味をもつからである。

z^*とz^{**}との差を0に収束させると，LATEはある政策のレベルz^*において参加・不参加が無差別であるような人々における参加の平均的効果を示すことになり，これはMTE（marginal treatment effect）と呼ばれる。定義されるZの範囲を調整することにより，ATEやATT，LATEはいずれも，このMTEの加重平均として表すことができる（Heckman and Vitacyl（2000））。

Ⅲ－3．2段階推定法

一時点のクロス・セクションデータが利用可能な場合，操作変数法から一歩踏み込んで，プログラムへの参加行動，ならびにY関数の誤差項uと参加行動との関係を明示的にモデル化することを通して，プログラム効果の不偏推定量を得るというのがこの方法の基本的な考え方である[15]。ただしこの方法においても，前節と同じような変数，すなわち参加関数には含まれるがY関数には含まれない説明変数の存在が重要となる。

プログラムへの参加確率を示す関数として以下のように単純なモデルを考えよう。

$$D^*_i = f(Z'_i) + \varepsilon_i \qquad ⒅$$

ここでZ'_iは個人の参加確率に影響を与える属性を表す外生変数のベクトル，D^*_iはプログラムに参加する可能性を示す潜在変数であり，$D^*_i > 0$であれば$D_i = 1$（参加する），$D^*_i \leq 0$であれば$D_i = 0$（参加しない）であるとする。また，誤差項εはZ'と独立であり，Z'とXには共通の変数も含まれるとする。

まず，$u_{1i} = u_{0i}$すなわちプログラム効果が一定で，$\alpha(X_i) = \alpha$であるダミー変数モデル⒃式を考えよう。

X，Z'そしてDについての条件付期待値をとると，

14）遠くても参加している人は，近くても必ず参加するのであるから，$Z = z^*$では参加しないが$Z = z^{**}$で参加しはじめた者のY_1の平均値は，$Z = z^{**}$を満たす参加者の平均値から$Z = z^*$でも訓練に参加している人々を除外して平均を求め直せばよい。Y_0についての平均値も同様にして求めることができる。

15）最近は参加関数や結合分布をセミ・パラメトリックに特定化する方法も開発されている（Heckman and Vytlacil（2005）.

$$E(Y_i|X_i, Z'_i, D_i) = \mu_0(X_i) + D_i\alpha + E(u_{0i}|X_i, Z'_i, D_i) \quad (19)$$

誤差項の条件付期待値は次のように分解されるから，

$$E(u_{0i}|X_i,Z'_i,D_i) = D_iE(u_{0i}|X_i,Z'_i,D_i=1) + (1-D_i)E(u_{0i}|X_i,Z'_i,D_i=0)$$
$$= D_iE(u_{0i}|X_i,\varepsilon_i > -f(Z'_i)) + (1-D_i)E(u_{0i}|X_i,\varepsilon_i \leqq -f(Z'_i)) \quad (20)$$

Y 関数と参加関数の誤差項，u_0と ε との結合分布を仮定すれば，⑳式に示される u_0と D との相関を示す部分を推定することができる。その結合分布が Barnow et al.（1980）のように２変数正規分布に従うと仮定すると，E（$u_{0i}|X_i,Z'_i,D_i$）は $E(\varepsilon_i|X_i,Z'_i,D_i)$ に比例するから，まずは⑱式をプロビット・モデルとして推定し，その推定値より計算される E（$\varepsilon_i|X_i,Z'_i,D_i$）の推定値を独立変数に追加して⑯式を推定するという２段階推定法を適用することによって，α の一致推定量を得ることができる（Heckman（1978））。ただし，Z'と X に共通する変数が多いと，E（$\varepsilon_i|X_i,Z'_i,D_i$）の推定値と X との間に高い多重共線性が生じやすくなるために，Z'には含まれるが X には含まれない説明変数の存在が実際上は不可欠となる[16]。

プログラム効果の多様性を考慮する場合には，u_0と ε だけでなく，u_1-u_0も含めた結合分布を仮定する必要がある（Blundell and Dias（2002））。ただしその場合，この推定法で ATE を識別することはできない。

Ⅲ－４．Difference-In-Differences 推定法

これまでの推定法は，プログラム実施後の１時点におけるクロスセクション・データを用いたものであったが，もしここに各個人 *i* について，プログラム参加前後についての，少なくとも二時点についてのパネル・データ，あるいは同一母集団から抽出されたプログラム実施前後の数年分のクロス・セクションデータが存在する場合には，u と ε の結合分布や適切な操作変数の存在を仮定しなくても，α の一致推定量を求めることができる[17]。

まず，プログラムは k 時点で１度だけ起こり，参加後の観察時点を t とし，それを明示したダミー変数モデル⑯式を次のように書き直そう。

$$Y_{it} = \mu_0(X_{it}) + D_{it}\alpha + u_{0it} \qquad t > k \quad (21)$$

ここで，D_{it}が１の値をとるのは，⑱式の $D^*_{it} > 0$ が t>k において満たされるときである。この誤差項 u_{0it}が個人 *i* に特有で t については不変である部分 Φ_i と個々に共通のマクロ効果 θ_t，そして誤差項 v_{it} とに加法分離され（$u_{0it} = \Phi_i + \theta_t + v_{it}$），その誤差項 v_{it} がプログラムへの参加とは独立，すなわち $E(v_{it}|D_i) = 0$ であるならば，difference–in–differences 推定法（DID）によってプログラム効果の一致推定量を得ることができる（Blundell and Dias（2002））[18]。Y 関数における加法分離性と誤差項 v と D との独立性さえあれば，属性 X でコントロールするまでもなく，

16）Z'に X の変数以外の変数が含まれていない場合，このモデルにおける訓練効果の識別は $E(\varepsilon_i|X_i,Z'_i,D_i)$ の非線形性のみに依存する。

17）ここでいう数年分のクロスセクション・データとは，プログラム実施前の s 時点においても，その後の k 時点でプログラムに参加する人と参加しない人の母集団が限定でき，それぞれからランダム抽出されたデータである。

18）個別効果が０の平均を持つ確率変数である場合を含む。ここでの個別効果は，時間と共に変化しない個人の先天的能力や訓練等のプログラムに対する選好，あるいは参加者と非参加者についてのデータが異なる調査票を用いて収集されていることから生じる違いなどを表していると考えられる。

$$\{E(Y_{it}|D_i=1)-E(Y_{is}|D_i=1)\}-\{E(Y_{it}|D_i=0)-E(Y_{is}|D_i=0)\}$$
$$=\alpha+E(v_{it}-v_{is}|D_i=1)-E(v_{it}-v_{is}|D_i=0)=\alpha \quad s<k<t \quad (22)$$

となり，これは参加者，非参加者サンプルについての標本平均として次のように推定することができる。

$$\hat{\alpha}_{DID}=[\overline{Y}_t|_{D=1}-\overline{Y}_s|_{D=1}]-[\overline{Y}_t|_{D=0}-\overline{Y}_s|_{D=0}]$$

プログラム効果に多様性がある場合でも，同様の仮定のもとで，この方法から ATT の一致推定量を得ることができる[19]。

特殊な例として，u_{0it} にマクロ効果 θt が含まれないと想定される場合には，プログラム参加者のみのデータを用いて，プログラム効果 ATT を識別することもできる。これはいわゆる before-after 推定量と呼ばれるものである。その場合の推定量は次のように求めることができる。

$$\hat{\alpha}_{BA}=\overline{Y}_t|_{D=1}-\overline{Y}_s|_{D=1}$$

たとえばわが国のように，公共職業訓練の効果を計測しようとしても，プログラム参加者に対する事後調査は存在するが適当な比較対照データが存在しない場合には，こうした手法を利用せざるをえない。たとえば黒澤（2003）は，東京都の公共職業訓練校の修了生に対して実施されたデータを利用して，訓練前後の差分をとることを通して訓練が修了生の収入に与える効果を推定している。その結果，訓練の効果は女性にはおおむねプラスであるが中高年男性には有意にマイナスであるとの結果が示された。調査設計上，訓練前の収入の観測時点 s が各人によって異なることを利用して，年齢の効果から訓練効果を識別することはできたが，それでも説明変数に含まれる労働市場の需給状況などで把握されない年次効果が存在すると，その効果を訓練効果から分離することはできず，訓練受講者のみのデータを用いた分析の限界が示される形となった。

なお，これら一連の推定量には，その推定値が訓練前のデータの観測時点に左右されやすいという共通の問題がある（Bassi（1983））。特に訓練プログラムの参加者については，訓練受講の前年の収入がそれ以前のトレンドよりも落ち込む傾向が見られることから（Ashenfelter（1978）），適切な s の選択は，その落ち込みが一時的なのか恒久的なのかに応じて定められなければならない。その落ち込みが一時的なものであれば，訓練受講者の Y は訓練を受けなくても非受講者より大きく上昇するであろうから，s を訓練受講の直前にすると，(22)式の 2 行目後半部分が 0 にならず，α が過大推定されることになる。

また，(22)式の推定量をもってしても，プログラム参加者と非参加者のマクロショックに対する反応が異なると，プログラム効果が識別されなくなるという問題もある。これに対処する方法としては，s から t までに生じたマクロ変動と同じようなマクロ変動が起こった 2 時点間のデータを追加し，その間の変化と s から t までの変化とを比較するというやり方がある（Bell, Blundell and Van Reenen（1999））。

Ⅲ－5．マッチング法

ATT のバイアスが 0 になる条件に基づいて α の一致推定量を計測しようとする方法に，マッチング法がある。これはプログラム参加者の各サンプルに観察可能な属性について似通った非参加サンプルをマッチさせ，双方の Y の差を測ることによってプログラム効果を推定するという方法である。

同一データに同じ説明変数が利用された場合，マッチング手法と重回帰分析とは極めて似た結

19）(11)式の誤差項から，$E[u_{1i}-u_{0i}-E(u_{1i}-u_{0i}|X_i,D_i=1)|X_i,D_i=1]=0$ となるから，$E(Y_{it}-Y_{is}|X_{it},X_{is},D_i=1)-E(Y_{it}-Y_{is}|X_{it},X_{is},D_i=0)=\alpha(X_{it})+E(u_{1i}-u_{0i}|X_{it},D_i=1)=ATT(X_{it})$ となることがわかる（Heckman, Lalonde, Smith（1999））。この推定法では ATE を識別することはできない。

果をもたらすことも示されているが（Friedlander and Robins（1995）），この方法と重回帰分析との最も大きな違いは，成果関数や $E(u_0|X)$ の形（とくに誤差項との加算分離性）を仮定する必要がないために，誤った定式化によって生じる推定値の偏りを回避できることにある（Friedlander, et al.（1997））。しかしながら一方で，マッチング法の仮定する制約は強く，バイアスを最小限にとどめるには参加者・非参加者双方について多くの情報が必要となる。最近では，純粋にノンパラメトリックなマッチング法に重回帰によるコントロールやDID，ブロッキング等の要素を取り入れることによって，より弱い仮定のもとで一致推定量を得る方法も開発されている（Imbens（2003））。

成果指標Yおよび参加変数Dに影響を与える変数ベクトルを統合してXと表すと，マッチング法は，属性Xをコントロールすると，成果変数Yの分布はプログラム参加変数Dと独立（⊥）であることを仮定することからはじまる。つまり，

$$\{Y_0, Y_1\} \perp D \mid X \qquad (23)$$

さらに，参加者に対する仮想現実が対照グループに常に存在するためには，全ての属性Xについて，参加者と非参加者の双方が存在している必要がある。

$$0 < P(D=1 \mid X) < 1 \qquad (24)$$

これら2つが満たされると，

$$E(Y_0|D=1, X) = E(Y_1|D=0, X) = E(Y_0|X) \qquad (25\text{-}1)$$

$$E(Y_1|D=1, X) = E(Y_1|D=0, X) = E(Y_1|X) \qquad (25\text{-}2)$$

となり，(8)式の後半部分に示されるプログラム効果のバイアスはゼロになる。(23)，(24)式を満たすXについて同一の値をもつ参加者と非参加者は，完全にランダムにD＝1あるいはD＝0に割り振られていることになり，同じXをもつ参加者と非参加者の差は純粋にプログラム効果を示すものになる。このときATTはATEと等しい。

ATTを推定するだけであれば，上記(23)の独立性は Y_1 に対しては必要ない。また参加者に対する仮想現実の存在のみを保証すればよいので，(24)式の代わりに必要な仮定は

$$P(D=1 \mid X) < 1 \qquad (26)$$

となる（Smith and Todd（2003））。したがって，(25-1）式が満たされるようなコントロール変数Xが存在する限り，似たような属性Xをもつプログラム参加者と非参加者との比較からATTの一致推定量を得ることが可能となる。

しかし属性Xの数が多いと，すべての属性Xについて「似たような」比較対照サンプルを見つけることは困難となる。この問題を克服したのがpropensityスコアに基づくマッチング法である。この方法は，(23)式の関係が満たされると，Xの代わりにプログラムに参加する条件付確率，P(D＝1|X）を条件付けた場合にも同様の独立性が満たされることを利用し，属性が似たもの同士ではなく，P(D＝1|X）が似たもの同士についてマッチングを行うというものである（Rosenbaum and Rubin（1983））[20]。条件付参加確率P(D＝1|X）が同じであれば，実際にプログラムに参加するかどうかは完全にランダムに決定されるのであるから，参加者（D＝1）と非参加者（D＝0）の母集団においてXは同じように分布されていることになる。このP(D＝1|X)＝Pはpropensityスコアと呼ばれ，通常はプロビットあるいはロジットモデルとして推定されることが多い。

$$E(Y_0|D=1, P(X)) = E(Y_0|D=0, P(X)) \qquad (27)$$

すると(26)－(27)式さえ満たされれば，ATTを次

20）ATTの推定においては，(25-1)式を仮定すれば Y_0 とDの独立性を仮定する必要はないが，その場合，Pの条件付独立性は別途仮定する必要がある。

のように表すことができる。

$$ATT = E(Y_1 - Y_0 | D = 1) = E|_{P(X)|D=1}\{E(Y_1 | D = 1, P(X)) - E(Y_0 | D = 1, P(X))\}$$
$$= E|_{P(X)|D=1}\{E(Y_1 | D = 1, P(X)) - E(Y_0 | D = 0, P(X))\}$$

したがって，プログラム参加確率の一致推定量$\hat{P}$について似たような値をもつ非参加者サンプルを各参加者サンプルにマッチさせ，それらの差に基づいてATTの一致推定量を得ることができることになる。

ただし，この推定量を実際のデータを用いて計算する上ではいくつかの問題点がある。第一はいわゆるコモン・サポートにかんする問題である。㉖式は全ての参加者に対応する非参加者の存在を母集団において保証する仮定であるが，実際のサンプルでは満たされないことが多い。㉖式が標本でも満たされるためには，Pについて似通った非参加者サンプルが存在しない（すなわち比較すべき対照サンプルが存在しない）参加者サンプルを分析から除外することが必要である。ただしそうした場合に推定されるのは「propensityスコアの似通った非参加サンプルの存在するXの範囲（コモン・サポート）内」の参加者におけるプログラムの平均的効果となる。プログラム効果が誰にでも一定ならばこれは「参加者全体」におけるATTの一致推定量になるが，効果が個人によって異なる場合には必ずしもそうはならない。

第二は参加関数の特定化，とくに説明変数の選択基準についてである。この点はマッチング法に限られた問題ではないが，前述したように，マッチング推定量の場合，propensityスコア関数に含まれる説明変数Xが㉗式を満たすように選ばれることが一致性に不可欠であることから，説明変数の選択はとりわけ重要である。理論モデル上，参加関数ならびにY関数に必要とされる変数を含めることはもちろんであるが，HIST（1997）によると，とりわけプログラム参加前の就業状況や収入情報が推定量のバイアスを軽減させる上で有効であることが示されている。

ただし，情報が多ければ多いほどよいということでもない。説明変数を増やすことによって㉗式が満たされなくなる，あるいはコモン・サポート問題を悪化させる可能性もある。Dehejia and Wahba（1999，2002）では，balancing propertyに基づく検定をその選択基準として適用している[21]。その方法とは，①理論的にプログラムへの参加確率Pならびに成果指標Y_0に影響を与えると考えられる説明変数を含めてpropensityスコア関数を推定する，②$\hat{P}$の値が近いサンプル同士をいくつかのグループ（strata）に分ける（このプロセスは脚注23と同様），③参加サンプルと非参加サンプルについて，strata毎にXの平均値の同一性を検定する，というものである[22]。③の検定が棄却されれば，説明変数間の交差項や二乗項などを含めて①～③のプロセスを繰り返すことになる。この基準が満たされないとpropensityスコア関数の特定化ミスが示唆されることから，こうした方法を適用することによってマッチング推定量のバイアスの低減につながると考えられる。

第三は推定における非参加者サンプルへのウェイトの付け方である。各参加者サンプルのPと完全に同一のPをもつ非参加者サンプルを見つけることは不可能であろうから，それに近いPの値をもつ非参加者サンプルに基づいてマッチするY_0を推計しなければならない。Pについての参加者と非参加者間との「距離」をどのように定義するか，ならびにどの程度の「距離」にある非参加者をマッチの対象とするかに

21）これは，Pが同じであれば，プログラム参加者と対照群のXは同じように分布されている（$X \perp D|P$）という命題に基づく（Rosenbaum and Rubin（1983））。本来ならば平均値だけでなく，第2，第3モーメント等についても検定する必要がある。

22）Becker and Ichino（2002）を参照。

よって推定量にもいくつかの種類があるが，一般的にマッチング推定量は次のような形で表すことができる（Smith and Todd（2003））。これは後述の㉚式に示される推定量に対してクロスセクション・マッチング推定量とも呼ばれる。

$$\hat{\alpha}_{CM}=\frac{1}{n}\sum_{i\in I}\left[Y_{1i}-\sum_{j\in I_{0i}}W(i,j)Y_{0j}\right] \qquad (28)$$

I はコモン・サポート内の参加者サンプルの集合で n はその数，I_{0i} は各参加者サンプル i の対照群となる非参加者サンプルの集合を表す。各参加者サンプル i の Y_1 にマッチさせる非参加者の Y_0 は，非参加者サンプルの加重平均として推定される。

たとえば Nearest–Neighbor マッチングでは，各参加者サンプル i の参加確率 $\hat{P}$ との差が最も小さい参加確率をもつ非参加サンプルを１つだけマッチさせるものである（つまり最も近いサンプルには W＝１，あとは W＝０を当てはめる）。Stratified マッチングでは，全サンプルを $\hat{P}$ の値に基づきいくつかのグループ（strata）に分割し，参加者と非参加者サンプルの比較はグループ毎に行われる[23]。一方，Kernel マッチングでは，各参加者サンプル i に対してある一定距離内の非参加サンプルの加重平均をマッチとし，そのウェイトは $\hat{P}$ についての参加者 i と非参加者 j との距離に反比例させる。このときのウェイト W は次のようになる。

$$W(i,j)=G\left(\frac{\hat{P}_j-\hat{P}_i}{a_n}\right)\Big/\sum_{k\in I_0}G\left(\frac{\hat{P}_k-\hat{P}_i}{a_n}\right)$$

ここで G（・）は非参加サンプルの範囲を規定する Kernel 関数，a_n はそのパラメータを示す。

P の値が近い相手とのみマッチされた方が質の高い比較ができるが，そのように限定すると比較できる非参加者サンプルの数が犠牲になる。こうしたトレードオフがあるために，どの推定法が最も望ましいか一概には言えないが，いくつかの推定法を用いることにより，推定値の頑健性を総合的に判断することができると考えられる。

さて，以上のマッチング推定量が一致性を満たすための条件は㉗式，すなわち条件付参加確率 P（すなわち属性 X）が同じであれば，プログラム参加者と参加しなかった者の「プログラムに参加しなかったら得られたであろう Y（Y_0）」は平均的に同じになる，というものであった。しかしこれまでの推定方法においても考慮されてきたように，X では説明できない要因がプログラムへの参加に影響を与え，それが参加者と非参加者の成果指標にも影響を及ぼしている可能性は高い。その場合㉗式は満たされず，㉗式にもとづくクロスセクション・マッチング推定量も一致性をもたなくなる。

ただし，そうした違いがたとえば個人のもつ潜在能力や人的ネットワークの違いなどの観察できない時間を通して一定な要因，あるいは共通のマクロ効果にもとづくのであれば，参加者と非参加者とのマッチをプログラム参加前後の Y の差について行うことによって，一致性を満たすマッチング推定量を得ることができる。参加者と非参加者が異なる地域や労働市場に属している，あるいは参加者と非参加者のデータが異なる調査から得られているために生じうる参加者と非参加者の Y_0 の違いもこの方法で取り除くことができる。

参加前の s 時点と参加後の t 時点についてこの推定法が一致性を満たすための条件は，

$$E(Y_{0t}-Y_{0s}|D=1,P(X))$$
$$=E(Y_{0t}-Y_{0s}|D=0,P(X)) \qquad (29)$$

となる。前節のような誤差項の分解を仮定する

23）この strata は，同一 strata 内の参加者と非参加者サンプルの P の平均が同一であるという仮説が棄却されないように構築される。ただし，マッチする相手のいない strata のサンプルは分析から除外される。この場合の推定量は，各 strata 内の参加者と非参加者の $\overline{Y}$ 差の加重平均として求められる。このときのウェイトは，各 strata に含まれる参加者サンプルが全参加者サンプルに占める比率である。

ならば，これは v_{it} についての条件となる。この仮定に基づくマッチング推定量，

$$\hat{\alpha}_{DIDM}=\frac{1}{n}\sum_{i\in I}[(Y_{1ti}-Y_{0si})-\sum_{j\in I_{0i}}W(i,j)Y_{0tj}-Y_{0sj} \quad (30)$$

はDIDマッチング推定法といわれる（HIT（1997））。

前節のDID推定量との違いは，各参加サンプルを参加確率Pの値に応じた加重平均と比較させることを通して，参加関数の説明変数Xの分布が参加者と対照群とで等しくなるようにしている点，ならびにコモン・サポート内での効果測定に限定している点である。次節に示すように，こうした措置はDIDマッチング推定量のバイアスを抑えるのに役立っていると考えられる。

わが国において，この方法を公共職業安定所（職安）における職業紹介機能の評価に応用した研究では，年齢や学歴，前職の企業規模や収入等の属性をpropensityスコア（職安利用関数）の説明変数としたところ，職安を利用して転職すると収入が有意に低下し，その傾向はとくに首都圏で大きいという結果が男性について示されている（黒澤（2005））。データの制約上，転職に成功した者だけを分析対象としているため，職安の利用が就職確率に与える影響についても分析する必要があるが，この結果は職業紹介の代替手段の多い首都圏において，職安の有効性が問われることを示唆するものといえよう。

Ⅲ－6．非実験的方法の取捨選択

非実験的推定法の多くはuとDとの関係や分布，Yの特定化等において何らかの仮定を前提としており，利用される推定方法によって，同じデータでも施策効果の推定値に大きなばらつきの生じることが多い。しかしながら，各推定法が前提としている仮定は仮想現実を含んでおり，データを用いてその妥当性を検証するのは困難である。

バイアスの最も小さな推定方法はどれか。この問に答えるひとつの方法が実験的データと非実験的に収集されたデータとの比較である。実験的手法にも，第Ⅱ節で述べたような問題は存在するが，実験的手法に基づくATT推定値と，実験的データのトリートメント・サンプルと非実験的に収集された非参加者サンプルに基づくATT推定値との比較を通して，様々な非実験的推定法の評価を行うことができる。実験的データのコントロール・サンプルと，非実験的に収集された非参加者サンプルとのATTを推定することを通して各推定法のバイアスを推定するという，より直接的なやり方もある[24]。

Lalonde（1986）は前者のやり方で推定方法を比較した先駆的研究である。この研究では，NSWデモ調査の実験的データから得られたATTの推定値と，CPSやPSIDといった外部調査から抽出された非実験的データを用いたDIDや2段階推定法などによる推定値との比較が行われたが，いずれの非実験的推定法も実験的手法に基づく推定値との乖離が大きく，非実験的推定方法のパフォーマンスの悪さが強調される結果となった。

その後やはり前者の方法を用いたDehejia and Wahba（1999，2002）では，propensityスコアに基づくクロスセクション・マッチング法によって実験的手法による推定値とほぼ変わらない推定値を得ることができるという結果が示され，マッチング法の有効性が強調された。しかし後者の方法を用いたSmith and Todd（2003）は，その結果が比較対照グループの限定方法やpropensityスコア関数の説明変数に大きく依存することを指摘し，推定バイアスはDIDマッチング法において最小になることを示している。同じく後者の方法を用いたHIT（1997）やHeck-

24）すなわち，実験的データのコントロールにD＝1，非実験的データの非参加者にD＝0を当てはめ，それに伴うATTを推定するという方法である。

man, Ichimura, Smith and Todd（以下 HIST, 1998）でも同様の結果が示されている。

HIT（1997）や HIST（1998）は，非実験的データを用いた推定法のもたらすバイアス，$E(Y_0|D=1)-E(Y_0|D=0)$を，①コモン・サポートにあてはまらない X に起因するバイアス，②コモン・サポート内での X の分布の違いに起因するバイアス，③訓練への受講プロセスにおけるセレクション（いわゆるセレクション・バイアス）という 3 つの要素に分解している。それぞれのバイアス要素を JTPA のデータを用いて推定したところ，そのかなりの部分が③ではなく①や②から生じているということが明らかにされた。また，プログラム参加者と対照群の地域やデータ収集に用いられた調査票が異なることも，かなりのバイアスをもたらすことが示された。

これら一連の研究結果が一般化できるとすれば，参加者・非参加者双方について職歴等の豊富な属性情報がある場合には，①と②に起因するバイアスを除去することができるマッチング法を用いる，あるいはそのほかの方法を用いる場合でも，分析をコモン・サポート内に限定し，非参加者サンプルにウェイト付けすることが望ましいということになる[25]。また，プログラム参加前のデータが入手可能な場合には，時間的に不変なバイアスについても除去することができる DID マッチング法を適用すれば，より一層良好なパフォーマンスが期待できるといえる。ただし，利用できる説明変数が少ない場合や，コモン・サポートの範囲が極度に限定されるようであれば，マッチング法は不適切であろう。参加者と非参加者サンプルが地理的に離れていたり，変数の測定方法が異なるのであれば，時間的に不変なバイアスに対応した推定方法が望ましい。このように，どういった推定方法を用いるべきかは，利用可能なデータの種類や施策の運営実態などに応じて決められるべきである。

前述のように実験的データとの比較ができない場合でも，プログラム実施前後のパネル・データが利用可能であれば，推定方法の前提となる仮定の妥当性を間接的に推測することができる。Heckman and Hotz（1989）は，それぞれの推定方法が前提とする仮定について，①プログラム実施前のデータに当てはめて非参加者と参加者との差を検定する，あるいは②過剰識別制約が満たされているかどうかを検定するという，2 通りの検定方法を提案している。たとえばクロスセクション・マッチングの場合，propensity スコア関数における説明変数 X がコントロールされれば，参加者と非参加者のプログラム前の成果指標に有意な違いは見られないはずであるから，もし違いがあるとすれば，(27)式の妥当性は疑わしいということがいえる。ただし，これらの特定化検定を行うには，そもそも適当なデータが整備されている必要がある。また，検定の検出力が低く，棄却されないモデルの推定値にばらつきが残った場合，最終的にプログラム効果の推定値についての合意が得られず，施策評価としては役に立たない可能性もある（Heckman and Smith（1996），Burtless（1995））。

データの特徴によっては，分析において別途調整が必要な推定法もある。たとえば訓練プログラムについての非実験的データは，訓練機関などに登録された受講者のリストから抽出された受講者サンプルと，外部調査から再抽出された非受講者サンプルとをプールして構築されることが多いが，このようなデータでは，受講者サンプルが過剰抽出されていることが多い。すなわち母集団における参加者の割合よりも，データにおける参加者の割合の方が大きくなるために（choice-based sampling），参加確率の推定においては適切なウェイトを付けて調整する必要が生じる。2 段階推定法や固定効果モデルなどは，D の値に条件付けされたモーメントに基づく推定法であるから，このようなデータに適用してもプログラム効果の推定量は一致性を保つ。しかしたとえば異なる時点のクロスセクシ

25）対照サンプルの P の分布が参加サンプルの分布に等しくなるようにウェイトづけする。

ョン・データをプールしたものを利用する場合や，操作変数法を用いる場合に一致推定量を得るためには，データを調整する必要がある（Heckman and Robb（1985））。propensity スコアを用いたマッチング法の場合には，P の代わりにそのオッズ比 P/(1－P)を基準にすればよい（Smith and Todd（2003））。

Ⅳ．社会への効果：費用－便益分析

Ⅳ－１．費用―便益分析

労働政策プログラムの実施に伴って社会全体にもたらされる便益あるいは損失を貨幣価値に換算し，社会全体の受ける純便益を評価するのが費用―便益分析である。これは施策に対する評価を，経済効率性の観点から行うものである。

職業紹介や訓練プログラムの場合，その主な便益は，参加者がプログラムを受けることで得る生涯所得の増分の現在価値である[26]。これは追加的な生産額であり，非参加者に何らの追加的費用も科さないから，そのまま社会全体への便益になる。社会における便益の配分や以下に述べる間接的（スピルオーバー）効果を考慮しないのであれば，参加者一人あたりの便益（ATT）から施策実施に伴う参加者一人あたりのコストを差し引いたものにプログラムへの参加者数を掛けたものが，このプログラムの社会全体に対する総純便益，つまり社会厚生への効果を表す値になる。したがって，一人あたりの純便益が正の値であれば，プログラムの実施は経済効率的に望ましいということになる。ただし，これまで労働の分野でなされてきた施策評価分析のほとんどは，参加者一人あたりの便益として就業率や収入への ATT を前節に示されたような分析方法を用いて計測することにとどまり，施策実施に伴って発生する費用までを考慮したものはそれほど多くない。

たとえば米国の歴代職業訓練施策の評価研究から得られた共通の見解は，若年層に対する収入への効果はゼロあるいはマイナスであるが，成人，とりわけ女性には有意な収入増加効果があるというものである（LaLonde（1995），Friedlander, et al.（1997））[27]。そのうち JTPA については Bloom et al.（1997）が実験的データに基づく費用―便益分析を行っている。そこでも若年層については粗便益でさえもマイナスであるが，成人へのプログラムは参加者１人当たり＄500程度の純便益をもたらしており，とくに非自発的失業者や低所得層の女性に対して費用効果的なのは求職支援であるという結果が示されている[28]。

しかしながら，こうした方法にもいくつかの問題がある。

Ⅳ－２．一般均衡的な効果

費用―便益分析の便益部分を，プログラム参加者における便益の平均的効果（ATT）を合算したものとして計算している背景には，プログラムに参加しない場合の成果指標 Y_0はプログラムの存在に影響を受けないという仮定がある

26）訓練受講中に賃金の助成などが行われたのであれば，その額は受講者の便益，非受講者の費用となる。また，米国の Job Corps のような若年層向けの訓練施策には，犯罪率を引き下げる効果も期待される。この便益の金銭的価値は，主に訴訟費用や犯罪被害額などの減少として推定されている。

27）若年層の収入を高めるという効果が示された数少ないプログラムに米国の Job Corps がある。その費用―便益分析については Maller et al.（1982）を参照。

28）訓練プログラムの効果は長期に渡ってもたらされるであろうが，この研究では不確実な数値を外挿せずに，データの存在する訓練後30ヶ月以内の効果だけを対象に，費用―便益分析を行っている。

（Heckman and Smith（1998））。しかし現実にはそうした仮定が満たされず，プログラムの存在が，プログラムに参加していない人々の厚生にも間接的に影響を及ぼす可能性は十分ある。

たとえばLaLonde（1995）によると，米国の訓練施策が参加者の収入を引き上げる効果の多くは，就業率の高まりによってもたらされているという。賃金の下方硬直性がある限り，訓練参加者の就業率の高まりは，訓練参加者と代替的な非参加者の就業率の低下あるいは解雇という犠牲のもとに起こっていることもありうる。訓練プログラムがなければ雇われていたであろう非参加者を「代替」あるいは「置き換え」て訓練参加者が雇用されるとすれば，このプログラムの社会全体への効果は，訓練参加者の得る便益から，代替・置き換えられた非参加者の被る損失を差し引いたものとして評価されなくてはならない[29]。

これらの効果を把握するためには，そのプログラムが就業率に与える影響と，就業を条件とした賃金率に与える影響とを分けて推定する必要がある。このように就業確率と賃金率への効果を分解して推定することは，施策の機能するメカニズムをより深く理解するためにも必要である。最近では，プログラムがある時点tでの参加者の就業率を押し上げるとしても，それが失業から就業への移行を促進するからか，あるいは一度得た就業機会を維持させるからかを区別するために，プログラムへの参加が将来の就業期間や失業期間に与える影響を，生存分析の手法（survival analysis）を用いて分析した研究もなされている[30]。労働者個人についての就業歴やプログラムへの参加歴を含んだデータを用いて，就業から失業，あるいは失業から就業への移行プロセスに，施策が及ぼす影響をモデル化するには，competing risksという枠組みでの生存分析が最も適切であり，こうした視点からの施策効果の分析は，今後ますます必要になるであろう。

また，規模の大きな訓練や職業紹介プログラムであれば，長期的には労働供給量を増やすことを通して賃金に下方圧力を与えるであろうし，プログラムが租税で賄われている限り，政府による支援策の拡大は人々の租税負担を高め，それが人々の消費や投資行動に影響を与える可能性もある[31]。このように，ある施策が社会全体に与える効果を包括的に測定するには，これまでの部分均衡的視点だけでなく，より長期にわたって市場全体にもたらされる間接的な効果をも考慮した一般均衡的な視野に基づく費用―便益分析を実施する必要がある。

プログラムの間接的効果の推計は困難を極めるが，最近はそうした間接的効果の可能性をも組み込んだ一般均衡モデルに基づくシミュレーションを用いた推計の試みもなされている。たとえばHeckman, Lochner and Taber（1998）は合理的期待形成を仮定した重複世代モデルに基づき，米国における大学への学費補助政策の効

29）ただし，たとえば雇用助成制度が長期失業者の短期失業者への代替をもたらしていたとしても，長期的には長期失業者を一度でも労働市場に戻すことによって有効労働供給を増やし，それが賃金を低め，結果的には自然失業率を引き下げる効果もあると考えられる（Bell, Blundell, and Van Reenen（1999））。

30）米国の非実験的データ（NLSY）を用いた分析にはGritz（1993），JTPAの実験的データを用いた分析にはHam and LaLonde（1996）及びEberwein et al.（1997），さらに，フランスにおける公的雇用創出施策についての非実験的データを用いた分析にBonnal et al.（1997）がある。

31）これらの効果以外にも，entry effect（Moffit（1992））や，dead weight loss（DWL）などもある。entry effectとは，たとえば訓練プログラムがなければ失業しなかったであろう者が，その受講資格を得るために仕事を辞めて失業者になるといった効果である。訓練プログラムを受けるためにわざわざ失業した者の方が，それ以外の訓練受講者よりも就業能力が高いとしよう。その場合，そうした人を受講者に含めて訓練が訓練後の就業率に与える効果を推定しても，その推定値は過大になる。こうした効果の有無を検証するためには，プログラムの存在が，失業率を高める効果があるかどうかをテストする必要がある。また，DWLとは，訓練施策がなくても同様の訓練が行われていたような場合に生じる。

果をシミュレートしている。そうした政策は人々の大学進学率を高めることが期待されるが，それが労働市場における高スキル労働の供給量を増やし，長期的には高スキル労働者の賃金を相対的に引き下げる可能性もある。その可能性が人々の大学進学行動に与える効果をも考慮した場合，学費補助政策の社会全体への効果は直接的効果の１／10に減少するという。ただし，この方法ではモデルにおける多くのパラメータ値を既存研究から当てはめなければならず，その特定化や制約に伴うミスが施策効果の推定値にバイアスをもたらすという問題がある。

Ⅳ－３．費用―便益分析に欠ける視点

プログラム参加者一人当たりの純便益に基づく費用―便益分析は，経済効率の観点から施策を評価するものであるが，このような分析では把握することの困難な指標もある。たとえば訓練プログラムを受講することによる余暇の減少など，貨幣価値に換算することの困難な便益や費用を費用―便益分析に取り入れることは難しい。

費用―便益分析では，費用が同じであれば，プログラム参加者への平均的効果（ATT）の総額が大きいほど効率的であるとされるが，積極労働政策には効率性のみならず公平性の維持あるいは向上という目的もある。所得への限界効用は所得水準によって異なるのであるから，同じ金額でも低所得者に対しては補正係数を乗じ，高所得者よりも大きな価値をもたせることを通して公平性の観点を費用―便益分析に導入することもできる（Boardman et al.（1996））。

公平性の観点からすれば，プログラムに参加しなければ最も困窮する人々，すなわちY_0の低い人々に対してプログラムが提供されることが望ましいかもしれない。その場合，プログラム評価を行う際には，参加者がプログラムから得た便益の大きさだけでなく，参加者がプログラムに参加しなかったら得たであろう成果Y_0の値にも注目する必要がある（Cobb–Clark & Crossley（2003））。とくに，プログラムに参加しなくても困らない人ほどプログラムから得られる便益が大きい，すなわち$\mathrm{cov}(Y_0,\alpha)>0$であると，効率性と公平性のトレード・オフが生じるため，Y_0を把握することはプログラムを評価する上でも重要となる。

プログラム効果が個人によって多様である限り，便益の分布状況の把握も政策上有用な課題になる（Heckman, Smith and Clements（1997））。たとえば，プログラムは参加者全般に平均的な便益をもたらしたのか，それともほんの一部が大きな便益を獲得したのかといったことは効果の平均値からは明らかにされない。プログラムの平均的効果が等しくても，便益を得る参加者の数が多い方が公平性の観点からすればより望ましいかもしれない。そのためには，たとえば便益を得た参加者の割合の計測といったことが必要になろう（$P(Y_1>Y_0 \mid D=1)$）。また，参加者の受けたプログラム効果の分布をより詳細に知るためには，効果の上位25％分位や下位25％分位点を計測する必要がある。

ただし，プログラム参加者と非参加者のデータに基づいてY_0とY_1の周辺分布を構築し，そこから結合分布についての上限と下限を統計的に規定することはできても，その範囲は広すぎて実用的ではないという問題がある。JTPAの実験的データを用いたHeckman, Smith and Clements（1997）では，Y_0とY_1の分布についての仮定を設けること等によって，より狭い結合分布の区間推定に基づいた分位点を推定している。

V．結びにかえて

本稿では，米国やわが国での事例を踏まえながら，職業訓練や職業紹介等のプログラムが参加者に与える平均的効果の推定における分析的枠組みとその方法論についてのレビューを行った上で，その費用―便益分析への適用と限界について考察を加えた。

希少な予算を有効に利用し，施策をより望ましいものへと改善してゆくためには，既存の施策についても，できるだけ客観的な費用と便益の比較に基づいた評価を行うことが不可欠である。ただし，評価を包括的なものにするためには，プログラムへ投入した資源の収益率だけに注目するのではなく，そのプログラムがどういった経路で収入に影響を及ぼすのか，ならびにそのプログラムが長期的に社会全体に与える影響のメカニズムをも考慮し，可能な限りその度合いを明らかにする必要がある。

欧米諸国における最近の研究で明らかにされてきたことは，適切なデータ構築の重要性である。プログラムの平均的効果を推定するという目的のみを考えると，統計的厳密性という観点からは，やはり実験的データに基づく推定に勝るものはない。しかし理想的な実験的データの構築には，さまざまな困難が伴う。そこで非実験的方法が現実性を持つのであるが，その場合でも，プログラム参加者および非参加者についての良質なデータを構築することが，正確なプログラム効果の推定を行うためには極めて重要になる。

まず，それが長期にわたって収集されるパネル・データであるほど，プログラムへの参加プロセスについての特定化検定を行う余地が生まれ，施策の長期的効果の測定が可能になるという意味で望ましい。また，そうしたデータには，公的に提供されるプログラムへの参加状況だけでなく，個人や企業によって行われた他の教育・訓練や求職支援の参加状況や参加費用についての情報や参加前の職歴・就業状態や収入についての情報，ならびにプログラムの成果指標として，賃金率だけでなく，就業状態や労働時間の情報も含めるべきである。さらに，HIT（1997）に示されるように，対照群を地域や参加資格等の面でプログラム参加者と同じ労働市場に直面する人々にすること（つまり母集団を等しくする），ならびに参加者と全く同じ調査票に基づいてデータを収集することが重要となる。地域や調査票のミスマッチを排除し，少なくとも観察できる変数については参加者に似通った非参加者を比較対照とするだけでも，プログラム効果を測る上でのバイアスの多くを回避することができるからである。

プログラム効果をコモン・サポート内に限定することや，比較対照をウェイト付けすることなどがプログラム効果のバイアスを大きく低減させるという観察事実は，そうしたメカニズムを推定方法に組み込んだマッチング法，とりわけ時間的に不変なバイアス要因についても除去することができるDIDマッチング法の有効性を示唆するものである。しかしながら，マッチング法の適用には詳細な情報がデータに含まれていることが大前提となる。

もちろん，理想的なデータをもってしても，置き換え効果の存在や，貨幣価値に換算することの困難な便益や費用の存在，ならびに間接効果の存在など，施策効果の測定には，さまざまな問題がつきまとう。しかしこうした問題の存在は，常に念頭に置かれるべきものであっても，評価の有用性を低下させるものではないはずである。

90年代後半からの雇用情勢の急速な悪化を背景に，わが国の労働政策のあり方には大きな変革が求められてきた。98年には労働者個人への

訓練費用の補助制度が開始され，同じ頃から公共訓練の受講者枠も民間教育機関への委託の形で急速な拡大をとげてきた。最近は長期失業者に対する職業紹介の民間委託や，構造改革特区における官民共同による職業紹介窓口の開設などの新たな取り組みも始められ，今年度からは企業への人材投資減税も施行される。政策の方向性に大きな変化が現れはじめている時こそ，従来の取り組み，そして新しい試みについても速やかに評価を行い，その情報をベースにして，望ましい職業紹介システムや職業訓練施策のあり方を着実かつ早急に定めてゆく必要がある。

参 考 文 献

Ashenfelter, Orley. 1978. “Estimating the Effect of Training Programs on Earnings,” *Review of Economics and Statistics*, Feb. 60(1)：47－57.

Barnow, B. G. Cain, and A. Goldberger. 1980. “Issues in the Analysis of Selectivity Bias,” in O. Ashenfelter and R. LaLonde eds.. *The Economics of Training Volume I : Theory and Measurement*. Vermont : Edward Elgar.

Bassi, Laurie. 1983. “The Effect of CETA on the Postprogram Earnings of Participants,” *Journal of Human Resources*, Fall, 18(4)：539－556.

Becker S. and A. Ichino. 2002. “Estimation of Average Treatment Effects Based on Propensity Scores,” *The Stata Journal*, 2, No. 4：358－377.

Bell, B. R. Blundell, and J. Van Reenen, 1999, “Getting the Unemployed Back to Work : An Evaluation of the New Deal Proposals,” *International Tax and Public Finance*, 6：339－360.

Bjorklund, Anders and Robert Moffitt. 1987. “The Estimation of Wage Gains and Welfare Gains in Self–selection Models,” *Review of Economics and Statistics*, Feb. 69(1)：42－49.

Bloom, Howard. 1984. “Accounting for No–Shows in Experimental Evaluation Designs,” *Evaluation Review*, 8(2)：225－246.

Bloom, H. et al. 1997. “The Benefits and Costs of JTPA Title II–A Programs : Key Findings from the National Job Training Partnership Act Study,” *Journal of Human Resources*, 32(3)：549－576.

Blundell, R and M. Dias. 2002. “Alternative Approaches to Evaluation in Empirical Microeconomics,” *Portuguese Economic Journal*, Vol. 1, Issue 2.

Boardman, Anthony et al. 1996. *Cost–benefit Analysis : Concepts and Practice.* Upper Saddle River, NJ : Princice Hall.

Bonnal, L., D. Fougere, and A. Serandon. 1997. “Evaluating the Impact of French Employment Policies on Individual Labour Market Histories,” *Review of Economic Studies*, 64：683－713.

Burtless Gary. 1995. “The Case of Randomized Field Trials in Economic and Policy Research,” *Journal of Economic Perspectives*, 9(2)：63－84.

Burtless Gary and Larry Orr. 1986. “Are Classical Experiments Needed for Manpower Policy?,” *Journal of Human Resources*, 21(4)：606－639.

Cobb–Clark, D. and T. Crossley. 2003. “Econometrics for Evaluations : An Introduction to Recent Developments,” *The Economic Record*, vol. 79, No. 247：491－511.

Dehejia, R., Wahba, S.（1999）“Causal Effects in Nonexperimental Studies : Reevaluating the Evaluation of Training Programs,” *Journal of the American Statistical Association*, 94（448）, 1053－1062.

Dehejia, R., Wahba, S.（2002）“Propensity Score Matching Methods for Nonexperimental Causal Studies,” *Review of Economics and Statistics* 84(1), 151－161.

Eberwein, Curtis, Joh Ham, and Robert Lalonde. 1997. “The Impact of Being Offered and Re-

ceiving Classroom Training on the Employment Histories of Disadvantaged Women : Evidence from Experimental Data," *Review of Economic Studies*, 64 : 655－682.

Friedlander, D. and P. Robins. 1995. "Evaluating Program Evaluations : New Evidence on Commonly Used Nonexperimental Methods," *American Economic Review*, Sept. 85(4) : 923－937.

Friedlander, D. and D. Greenberg and P. Robins. 1997. "Evaluating Government Training Programs for the Economically Disadvantaged," *Journal of Economic Literature*, Dec. 35 : 1809－1855.

Gritz, Mark. 1993. "The Impact of Training on the Frequency and Duration of Employment," *Journal of Econometrics*, 57 : 21－51.

Ham, John and Robert LaLonde. 1996. "The Effect of Sample Selection and Initial Conditions in Duration Models : Evidence from Experimental Data on Training," *Econometrica*, 64 (1) : 175－205.

Heckman, James. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46(3) : 931－59.

_____. 1997. "Instrumental Variables : A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32 : 441－62.

_____. 2001. "Micro Data, Heterogeneity, and the Evaluation of Public Policy : Nobel lecture," *Journal of Political Economy*, vol. 109, no. 4 : 673－748.

Heckman, James, N. Hohmann, M. Khoo, and J. Smith. 2000. "Substitution and Dropout Bias in Social Experiments : A Study of an Influential Social Experiment", *The Quarterly Journal of Economics*, vol. 115 (2), 651－694.

Heckman, James and Joseph Hotz. 1989. "Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programme," *Journal of American Statistical Association*, 84 (408) : 862－874.

Heckman, James, H. Ichimura, J. Smith, and P. Todd. 1998. "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66(5) : 1017－1098.

Heckman, James, H. Ichimura, and P. Todd. 1997. "Matching as an Econometric Evaluation Estimator : Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, 64 : 605－654.

Heckman, James, R. Lalonde, and J. Smith 1999, "The Economics and Econometrics of Active Labor Market Programs," in Orley Ashenfelter and David Card eds., *Handbook of Labor Economics*. Amsterdam : North–Holland.

Heckman, James. L. Lochner and C. Taber. 1998 "General Equilibrium Treatment Effects : A Study of Tuition Policy," *American Economic Review*, 88(2) : 381－86.

Heckman, James, S.Navarro–Lozano 2003. "Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models," *National Bureau of Economic Research Working Paper* 9497.

Heckman, James and Richard Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions," in James Heckmand and Burton Singer eds. *Longitudinal Analysis of Labor Market Data*. Cambridge : Cambridge University Press.

Heckman, James and J. Smith. 1996. "Experimental and Nonexperimental Evaluation," in Schmid G. O'Reilly J. and Schomann K. eds.. *International Handbook of Labour Market Policy and Evaluation*. Brookfield : Edward Elgar.

_____. 1998. "Evaluating the Welfare State", Frisch Centenary, Econometric Monograph Series, Cambridge : Cambridge University Press.

Heckman, James, J. Smith, and N. Clements. 1997. "Making the Most out of Programme Evaluations and Social Experiments : Accounting For

Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64：487－535.

Heckman, James. E. Vytlacil. 2000 "Local Instrumental Variables," *National Bureau of Economic Research Technical Working Paper* 252.

_____. 2005 "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, vol. 73：669－738.

Imbens, Guido, 2003. "Nonparametric Estimation of Average Treatment Effects Under Exogeneity : A Review," *National Bureau of Economic Research Technical Working Paper* 294.

Imbens, Guido and J. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62：467－75.

黒澤昌子2003「公共職業訓練の収入への効果」『日本労働研究雑誌』514号

_____2005「公共職業事業の評価」『「官製市場」改革』八代尚宏編著，日本経済新聞社，所収予定.

LaLonde, Robert. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review*, 76：604－620.

_____. 1995. "The Promise of Public Sector–Sponsored Training Programs," *Journal of Economic Perspectives*, 9(2)：149－168.

Lechner, Michael. 2002. "Program Heterogeneity and Propensity Score Matching : An Application to the Evaluation of Active Labor Market Policies," *The Review of Economics and Statistics*, 84 (2)：205－220.

Mallar, Charles, et al. 1982. *Evaluation of the Economic Impact of the Job Corps Program : Third Follow–up Report*. Princeton : Mathematica Policy, Inc.

Moffitt Robert. 1992. "Evaluation Methods for Program Entry Effects," in Charles Manski and Irwin Garfinkel eds.. *Evaluating Welfare and Training Progra*., Cambridge : Harvard University Press：231－252.

Rosenbaum, P., Rubin, D. 1983 "The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70, 41－55.

Smith, Jeffrey and Petra Todd, 2005 "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, v 125, n1－2：305－53.