

討議としての業績評価
外的要因を考慮した成果指標による公的機関の業績管理
独立行政法人 労働政策研究・研修機構
研究員 富岡 淳

《要旨》

公的組織の定量的な業績管理と比較分析の方法を提示する。とくに、外的要因の影響を考慮した「調整成果指標」に焦点をあてる。適用の実例として、公共職安の県別長期データを利用した実証分析を示す。応用が可能な分野は、労働、教育、医療、司法など幅広い。外的要因と不確実性について調整すると、成果指標の様相はまったく変わる。また、収益という一元的な基準がある企業部門に比べると、公的部門はしばしばステイクホルダーやタスクが多角的であり、情報の不完全性が著しいため、定量的なプライオリティがつけにくい。その限りでは、性急な成果主義の導入は予想外の弊害をもたらす、持続的でないこと、業績評価には多角的・定性的・暫定的な性格が必要であることを、内外の試行錯誤の事例をふまえて論じる。調整成果指標は、現場ヒアリングなど定性的な情報収集と、定量的なデータ分析との往復、というプロセスと適合的であるため、業務に関する発見的な討議のためのシグナルとして機能しつつ、個別文脈に沿って漸次的に質を改善することが可能である。

(備考) 本論文は、独立行政法人労働政策研究・研修機構としての見解を示すものではない。

富岡 淳

構成

1. はじめに：背景と全体の概要
2. 調整成果指標のアイデア
 - 2-1. 例と図解
 - 2-2. 拡張：複数の要因と政策効果
 - 2-3. 調整指標の特徴と含意：仮説生成とモデル評価の往復
3. データの種類に対応した分析手法
 - 3-1. 各組織レベルの一時点のデータにおける問題
 - 3-2. 各組織レベルのパネルデータの場合
4. 応用例：県別データによる実証分析
 - 4-1. データ
 - 4-2. 推定するモデル
 - 4-3. 推定結果
 - 4-4. 調整指標 \hat{a}_i の分布とは何か？
 - 4-5. 先行研究との比較
5. 指標の不確実性の問題
 - 5-1. 組織の規模と成果指標の変動
 - 5-2. 規模と変動：数値例
 - 5-3. 規模と変動：県別データ
 - 5-4. 点推定と区間推定
 - 5-5. 不確実性への対策
6. 指標の選定，情報，多元性，エージェンシー，費用
 - 6-1. 背景
 - 6-2. 業務の多元性と対立性
 - 6-3. 指標の多角性と有用性のトレードオフ
 - 6-4. 情報の非対称性と評価制度とのゲーム
 - 6-5. 切磋琢磨と非協調
 - 6-7. アカウンタビリティ
 - 6-8. 政策の費用と便益
 - 6-9. 評価の費用と公共化による費用削減
 - 6-10 成果指標のための倫理規定
7. 結語
8. 文献
- A 付録

* 構想と執筆の諸段階において、天利浩、稲上毅、大谷剛、奥田栄二、小倉一哉、小野旭、川口大司、坂田圭、周燕飛、立道信吾、中村良二、馬欣欣、勇上和史、吉田克己の各氏、立命館大学 BKC 社系研究機構セミナーおよび関西労働研究会の参加者各位、厚生労働省、都道府県労働局、職業安定所職員の方々から、貴重な示唆を頂きました。形式・内容上の欠点は著者の責に帰すものです。

- A-1 ミッション，タスク，指標，その多元性とコスト
- A-2 第3節への補論(1)：多次元(階層)データの活用
- A-3 第3節への補論(2)：調整成果指標における集計バイアス
- A-4 第4節への付図表
- A-5 第4-4節への補論：観測上の同値性

1. はじめに：背景と全体の概要

公的機関の定量的な業績管理および評価の方法を探究する。とくに、外的要因を考慮した比較分析・目標管理である「調整成果指標」の可能性に焦点をあてる。そのアイデアは、各オフィスが置かれた条件の違いの影響を調整した成果指標を作成し、目標設定と比較に用いるものである。外的条件を共有するオフィス群に対して、ある「期待される成果水準」を適宜計算し、それと実現した水準との差という尺度において、各オフィスの状況を吟味する。正あるいは負の差がとくに大きいオフィスについては、定性的な調査や議論を行なうことで、これまで注目されていなかった要因が浮かび上がるかもしれない。それを新たな調整要因として加えて成果指標を計算しなおすならば、より精緻で現場の納得度も高い業績評価へと近づけることが可能であろう。

この手法が応用可能な分野は、教育(学校評価)、医療(病院評価)、司法(裁判所評価)、労働など幅広いが、本稿はハローワークの業務統計の県別長期データを利用して、就職(マッチング)件数という成果指標に関して実証分析を行う。外的要因(労働市場の需給)や、標本サイズの影響(成果指標の不確実性の幅=信頼区間の存在)を考慮するか否かで、成果指標はまったく様相を変えることを示す。さらに論文の後半では、内外の先行事例をふまえて、公的機関における定量的な評価の陥穽と対策を検討する。

一般に、よい評価システムに必須の要件のひとつは、関係者への透明性とアカウントビリティである。本稿も、政策企画者、実施者、評価者、被評価者、利用者など、広く公共政策の評価に関心をもつ読者の参考に資することを第一の目的とし、直感的な説明に留意する。

各分野における政策・業績評価の動向と意義

長期雇用と組織内スキル形成が弱体化しつつある先進国では、近年、職業紹介や能力開発といった「積極的雇用政策」のプレゼンスが増している。¹一方、低成長期の経済において、

¹ OECDは雇用政策を3種に分類している。積極的政策：失業者や就職困難者の状況を、就業と賃金の点で改善することを目的とする。職業紹介、職業訓練、雇用助成金、障害者雇用など。消極的政策：失業保険など。その他：最低賃金制度、解雇規制など、対象者がより一般的なもの。OECD主要加盟国における平均的な積極的雇用政策支出は、80年代以降、GDPの1%弱の水準を維持している(OECD, 2006, Figure 3.4)。

限られた公的経済資源を効率的に配分するためには、複数の政策・制度を比較してプライオリティを付けていく必要がある。こうして、公的部門の活動の定量的な評価が盛んになってきた。

業務の定量的測定は、外部への説明責任の遂行に加えて、組織内部でも①透明性を高める②事例学習による業務改善③評定④賞罰（予算策定）、などに活用されている。そしておそらくこの順で、現場への正または負のインパクトは大きい（Bruijn, 2007）。近年はとくに、透明性と一貫性を高めた「目標設定」と「業績評価」の試みが内外で増えている。たとえば、学校は共通テストの平均点、警察は検挙件数、裁判所は判決数、大学は専門的論文の発表数、病院は手術後の生存率、といった成果指標を用いて、各組織の業務を事前および事後に吟味評価するなどである。

職業紹介行政においては、日本を含む主要先進国は、少なくとも何らかの目標管理制度を導入している。すなわち、全国、各行政区域、各職安のそれぞれのレベルにおいて、前期の実績や経済状況などを勘案して、就職率などの指標に今期の達成目標を設定、実績データを随時収集して目標達成度を確認し、必要に応じて関係組織との交渉・指導に利用している。

2

公的部門の定量的な業績管理は、需要側すなわち利用者にとっても重要になっている。利用者により豊かな情報と選択肢を提供すること、そこでの反応をふまえて供給者が一層の創意工夫・切磋琢磨を実現することが期待されている。

先行例の教訓

ただし、1990年代以降の内外の試行錯誤の実例を吟味すると、公的部門における定量的目標管理や組織間相対評価、民間委託などは、便利な魔法の杖ではないことがわかる。その成否の条件を、近年の行政学や経済学は理論的・定性的に、また（萌芽的だが）定量的に分析している。

定量的な業績評価が有効であるのは、業務の目的、内容、環境、影響が一元的・安定的であり、短期的に観察可能な場合である。その種の業務は、官民を問わずほとんどの組織に存在する。定量的管理の部分的な適用が、組織全体に規律の感覚を醸成する効果も期待できる。

そうでない分野・対象について、定量的評価や市場メカニズム（価格という公共的シグナ

² 詳細は OECD(2001)、日本については厚生労働省職安局 (2006) を参照。後者『平成18年度の雇用保険三事業による事業の目標設定について』には、「原則として全ての事業について目標を設定」、「目標設定事業数154事業（平成17年度）→168事業（平成18年度）」、「目標設定した全ての事業について、定量的なアウトカム目標を設定」、「個別事業に加え、一定の施策のまとまりごとに目標を設定」とあり、各目標の達成度を示している。

ルにもとづく分権的な需給調整)のアナロジーを用いる場合には注意が必要である。たとえば、制御や観測が難しい外的な要因の影響が大きい環境では、実績値と評定・賞罰を単純に結合するのは有害である。環境要因とデータの精度を考慮に入れた成果測定がのぞましい。また、容易に測定される業務指標で、制度設計者の期待に沿った変化があらわれたとき、観測しにくい他の面で、業務の質に劣化が生じている可能性は高い。さらに、目標水準やその達成・未達成を議論する前に、そもそもある成果指標を選択し、他の指標を選択しない理由、そして特定の水準を設定する理由について妥当性が確認されていなくては、論理的におかしい。

しかし、現実にはしばしばこれらの条件は満たされていない。そもそも外部の分析者は、現場の実情に疎い。現場では、「報告・評価疲れ」のために本来業務が支障をきたすケースが多い。極端な場合、業務全体が評価システムとの空虚なゲームと化す。それならば、「業務の定量的な測定は現状把握の一手段であって、付加価値の評定ではない」と割りきるほうが、組織にとっても社会にとっても、よほど合理的である。

探索的な討議のためのシグナル

一般に、収益という明快で無慈悲な尺度がある企業部門に比べると、公的部門の業務はその目的、種類、影響、ステイクホルダーの範囲が多角的・対立的・長期的であり、外部性をもつ。そのため、事前にも事後にもプライオリティがつけにくい。その種の組織を適確に管理・評価するには、多角的・定性的・暫定的・継続的なアプローチが必要になる。たとえばそれは、当事者、内外のステイクホルダーや観察者からの意見聴取などの定性的な情報収集と、定量的なデータ分析との往復という形をとろう。調整成果指標の手法は、このような探索的なフィードバックと適合的であり、個別文脈に沿って手法自体を漸次的に改善することが期待できる。

この業績評価は、不合意を含めて誰の声も排除しない、という意味で「公共的」な討議の性質をもつことができる。成員間の公共的討議は、部門間、階層間の正確な情報の伝達を促し、成員の帰属意識を醸成し、士気を高めて組織の改善をもたらすメリットがある。

本稿の構成は冒頭に掲げた。次節は、まず調整指標のアイデアを例示する。³

³ なお、本稿における業績評価の対象を定義するならば、「公金を用いて、自治体や政府の法政策的な裏付けと裁量と責任のもと、財・サービスの供給を遂行する組織」となるのか。ここでいう財・サービスの生産供給は、公務員が担う場合が多いが、民間部門、すなわち非営利組織（NPOその他）や営利組織に委託する場合もふくむ。政府部門と民間部門の境界は、概念上も実際上もさほど明快ではなく、固定的でもない。（ラフな目安は、公金の民間部門への分配や市場独占力である。その限りにおいて、納税者への説明責任は民間部門にも

2. 調整成果指標のアイデア

2-1. 例と図解

同じ業務を行なう一群のオフィスにおいて、なんらかの成果指標がすでに選定されている状況から議論をスタートする。(成果指標がいかに選択される(べき)かについては、第6節、および補論 A-1 で扱う。そこでは、組織にはまず大まかなミッションが与えられ、そこから具体的なタスクが定まり、次に成果指標、という順で選択がなされると想定する)。

たとえば、一般に各ハローワーク(以下、HW)の属性や置かれた状況はさまざまである。それらの違いを無視した目標水準設定や比較評価は、あらっぽい方法であり、またフェアでない。

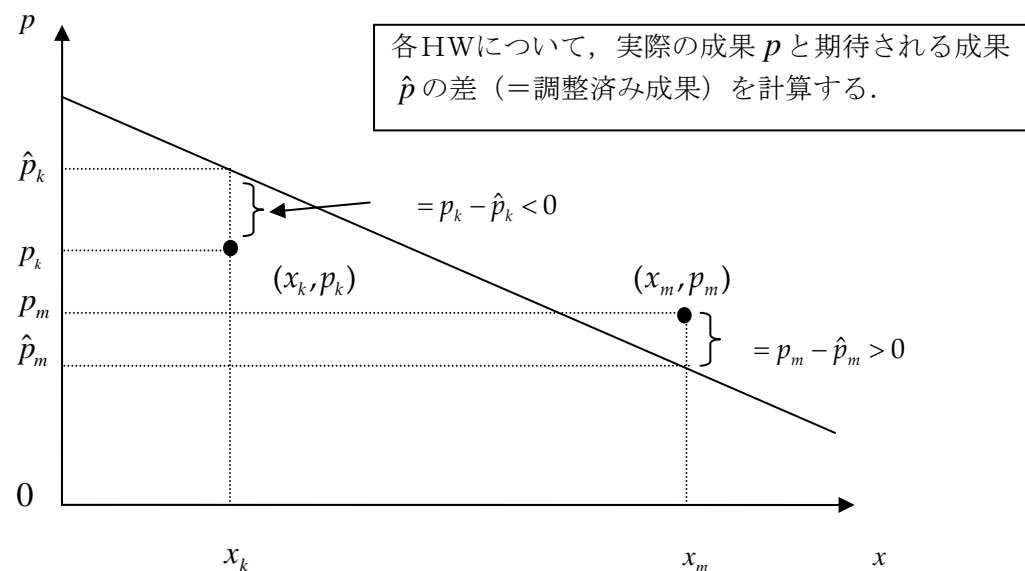
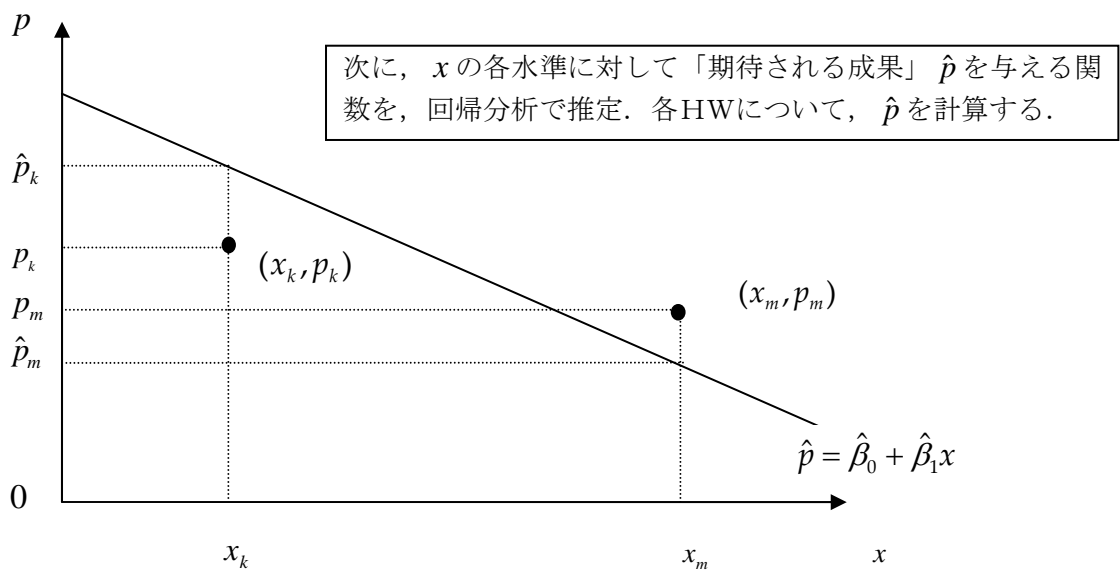
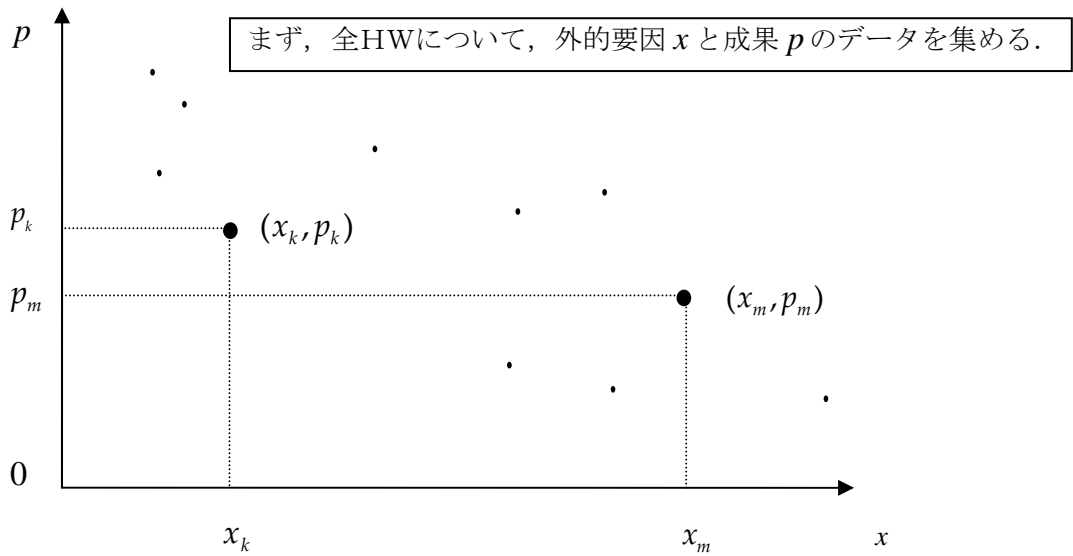
例 2-1-1. 仮に、求職者の平均年齢が高いほど、就職確率は低くなるとしよう。求職者の平均年齢が高いHWは相対的に不利な条件下にあるので、その分だけ成果のベンチマークを低めに定めるか、指標に割増の調整を施したうえで他のHWと比較するのが理にかなってよう。

そこで、各HWが置かれた条件の違いの影響を調整した「調整済み成果指標」を作成し、目標設定と比較に用いる。アイデアはごく単純である。所与の条件に対応して各HWに何らかの「期待される成果水準」を適宜計算し、それと実際の成果水準との差という尺度において、各HWの状況を吟味する。次頁の概念図 2-1 は、もっともシンプルな設定において、この指標の作成手順を示している。この例では、 k と m の2ヶ所のオフィスにおいて、外的要因の調整をおこなった結果、成果指標の水準が逆転している。

回帰分析の用語でいえば、この手法は残差分析の応用である。成果指標を被説明変数とし、各HWの外的環境や業務体制を説明変数として、回帰モデルを推定する。モデルから、各HWごとの成果の実績値と理論値の乖離を計算し、それを調整済み成果指標として利用する。

求められよう。)「評価手法の適用可能な範囲」についても同様である。じっさい、政府部門の理論的・定量的な研究が活発であるのは、業務水準が金銭単位で一元的に測定可能であり、何らかの最適化問題としてミクロ経済学の応用が比較的容易な分野である(たとえば企業への物品発注・調達や、規制の経済的影響の分析。Cf. Laffont and Tirole (1993))。ただし、その種の問題設定は、公共財的・非市場的な性質の強い公的部門においては普遍的ではない。本稿の分析枠組みは、さしあたり金銭単位での最適化問題として定式化されない業務評価を主な対象とする。

図 2-1. 調整済み成果指標の作成 (概念図)



肝心の、ある HW に期待される成果水準の定め方は、「HW i と同じ外的条件のもとにある HW 群について ($x_1 = x_m = x_n = \dots$)、成果の実績値 (p_1, p_m, p_n, \dots) を集め、成果の平均値を計算すると、この水準 \hat{p}_i である、それゆえこの成果水準 \hat{p}_i を HW i にも期待してよい」といった考えにもとづくものとする。これは、公的部門には企業組織における収益のような明快で根本的な評価尺度が存在しないため、環境が比較可能なオフィス群ごとに、成果の実態にもとづいたベンチマーク水準を定める、とも言える。

以下、調整指標の作成手順とその解釈を示す。(定性的には、概念図 2-1 と対応している。)

モデル設定 安定所番号などをもとに、全 HW に適当に識別番号をつける。任意の i 番の HW における成果指標 (就職率) を p_i であらわす。また、求職者の平均年齢を x_i であらわす。全国に 600 箇所の HW があるとして (母集団)、各 HW の成果指標 p_i について

$$p_i = \beta_0 + \beta_1 x_i + u_i, i = 1, 2, \dots, 600$$

という一次式を考える。 β_0 は切片。 β_1 は、求職者の平均年齢 x_i が 1 歳上がったときに、HW の指標 (就職率) p_i が平均的にどれだけ変化するかを示す。前出の概念図の例のように、中高年ほど就職率が低い場合は、 β_1 は負の値をとる。しかし、通例、母集団における二つの係数 (β_0, β_1) は未知であり、われわれは (p, x) の標本からその値を推測したい。 u_i は平均年齢以外の要因であり、データ上は観測できない。⁴

係数推定 できるだけ多くの HW について就職率 p_i と求職者の平均年齢 x_i のデータ (標本) を入手し、上のモデルに最小二乗法などを適用して、係数 (β_0, β_1) の推定値

⁴ 我々が把握したいのは、外的要因の水準 x に対して成果の期待値を与える関数 $E(p|x)$ であった。ところで、任意の確率変数は、条件付期待値関数と、それに直交する残差項 (平均 0) とに分解できる (Goldberger (1991, ch.5))。すなわち、 $p = E(p|x) + u$ 。一般に $E(p|x)$ は非線形だが、以下では説明の簡単化のため、一次関数によるその最良近似を推定する (両者は $E(p|x)$ が線形の場合、一致する。なお、最良は最小二乗誤差の意味)。線形最良近似は、大標本のもとでは必ず一致推定できる。それゆえここでの問題は、推定の技術的な可否ではなく、 $E(p|x)$ の具体的な定式が十分興味深いか、とくに政策的・実務的な観点から意味があるかどうかである。このような、成果指標と外的要因のボトムアップな選び方については、2-3 節で詳述する。一方、第 4 節では標準的な労働経済理論の示唆にもとづき選定したモデルを分析する。

$$(\hat{\beta}_0, \hat{\beta}_1)$$

を得る．ここでは仮に， $(0.4, -0.002)$ という推定値を得たとしよう．⁵

期待水準の設定 全HWに共通なこの係数の推定値と，各HWに固有な x_i の値から，各HWの成果の予測値，あるいは「期待される」成果水準 \hat{p}_i を計算する：

$$\hat{p}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, i = 1, 2, \dots, N$$

ここで N は，データを手に入れた HW の数（サンプルサイズ）を表す．さきほどの推定値の例を使うなら，

$$\begin{aligned} \hat{p}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i \\ &= 0.4 - 0.002x_i, i = 1, 2, \dots, N \end{aligned}$$

\hat{p}_i の解釈は次の通り．外的条件（求職者の平均年齢）が同一である複数のHWの間では，期待される成果は同一の水準となる．実際，係数 (β_0, β_1) の推定値は全HWで共通であるから， x_i の値が等しい複数のHWの間では \hat{p}_i も共通の値となる．

したがって逆に， \hat{p}_i の値が異なるHWの間では， x_i の値も異なっている．すなわち，各HWの置かれた条件（求職者の年齢）に有利不利の違いがある場合は，その違いを考慮したうえで，期待される成果の水準を個別に設定している．たとえば， x_i の係数の推定値が負だとすれば，求職者の平均年齢 x_i が高い HW については，その分だけ期待される成果 \hat{p}_i が低く計算される．⁶

実績値との比較 各HWの実際の成果指標の値 p_i と，期待成果指標の値 \hat{p}_i との差 $p_i - \hat{p}_i$ を計算する：

$$p_i - \hat{p}_i, i = 1, 2, \dots, N$$

あるHWではこの値がプラスであれば，そのHWは自身が置かれた状況のもとで期待される水準をこえた成果をあげていると解釈できる．

⁵ ここでは，これらのパラメータが定量的に重要であること，とくに推定値（の絶対値）が十分にゼロよりも大きいことを仮定して話を進める．実際の分析では，その点は統計的検定によって判断する．

⁶ 前述のとおり， \hat{p} は，ある x の水準を共有するすべての HW における p の平均値に準ずるものとみなして大過ない．

仮に β_1 の推定値が負の値だとすれば、求職者の平均年齢 x_i が高い HW については、期待される成果 \hat{p}_i が低いため、 $p_i - \hat{p}_i$ は高めに計算される。この意味で $p_i - \hat{p}_i$ は、いわば条件調整済み成果指標である。

組織間比較 この $p_i - \hat{p}_i$ をもとに、全HWの調整指標の一覧を作成する。

ここで手順④と⑤の意味を数値例で考えてみたい。各 HW の期待される成果は、推定式

$$\begin{aligned}\hat{p}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i \\ &= 0.4 - 0.002x_i, \quad i = 1, 2, \dots, N\end{aligned}$$

において x_i を代入することで求まる。仮想的な二つの HW, k と m の比較を考える。実際の成果が、 $p_k = 0.34$ と $p_m = 0.32$ であるとしよう。すなわち、実際の就職率では、HW k の方が HW m よりもパフォーマンスがよい。

しかし、いま求職者の平均年齢のデータをみると、 $x_k = 25$ と $x_m = 50$ であったとしよう。すなわち、HW k では若年の求職者が多く、HW m では中高年の求職者が多い。われわれは、この条件の違いを考慮した上で、両 HW のパフォーマンスを吟味したい。そのため、これらのデータと推定式から、両HWの期待される成果と調整済み成果指標を計算してみる：

HW k

実際の成果（就職率） $p_k = 0.34$

求職者平均年齢 $x_k = 25$

期待される成果 $\hat{p}_k = 0.4 - 0.002(25) = 0.35$

調整済み成果 $p_k - \hat{p}_k = 0.34 - 0.35 = -0.01$

HW m

実際の成果（就職率） $p_m = 0.32$

求職者平均年齢 $x_m = 50$

期待される成果 $\hat{p}_m = 0.4 - 0.002(50) = 0.30$

調整済み成果 $p_m - \hat{p}_m = 0.32 - 0.30 = 0.02$

求職者の平均年齢の違いを考慮したうえで成果を比較すると、調整なしの場合とは成果指標の水準が逆転している。

HW k は、一見したところのパフォーマンス p_k ではすぐれているが、実は求職者の若さという恵まれた条件のわりには、期待に応えるだけの十分な成果をあげていない（調整済み成果指標がマイナスである）。

逆に HW m は、一見したところのパフォーマンス p_m はふるわないが、多くの求職者が中高年であるという厳しい条件のもとで期待される以上の成果をあげている（調整済み成果指標がプラスである）。

同様の計算を全 HW について行えば、外的条件の違いを考慮しつつ、全 HW のパフォーマンスを吟味できる。

2-2. 拡張：複数の要因と政策効果

上でみたモデルは、いくつかの方向に拡張できる。現実には複雑であるから、そうする必要もある。

まず、成果指標を左右する外的要因は一つではありえない。また、現場からのフィードバックによって、それまでの分析では考慮されていなかった新たな要因が発見されてゆく可能性は高い（次の 2-3 節参照）。

さらに、上述した分析の枠組みからは、シンプルな政策的含意を導き得ない。説明変数が、政策的に動かし得ない要因であったためである。政策当局が水準をコントロールできる要因を分析しない限り、政策効果は議論できない。

複数の要因を導入し、政策的含意の明快な分析を行うためには、分析の枠組みをつぎのように拡張すればよい。

(I) 政策当局にとって動かさない与件である諸要因の組（行ベクトル） \mathbf{x} と、政策当局の裁量の範囲内である諸要因（政策変数）の組 \mathbf{z} をリストアップする。

\mathbf{x} の構成としては、求人・求職の属性（年齢や学歴）、地域の景気指標の組 (x_1, x_2, x_3) などが考えられる。また \mathbf{z} の構成としては、求人検索用コンピュータの台数、紹介担当職員の人件数、業務時間の組 (z_1, z_2, z_3) などが考えられる。

(II) \mathbf{x} および \mathbf{z} を説明変数として、次の回帰モデルを考える。全部で600箇所のH

Wがあるとする、

$$p_i = \beta_0 + \mathbf{x}_i\beta_1 + \mathbf{z}_i\beta_2 + u_i, i = 1, 2, \dots, 600$$

すなわち、任意の i 番のHWの成果指標 p_i は、政策当局に動かせない与件である諸要因 \mathbf{x}_i と、政策当局の裁量の範囲内である諸要因 \mathbf{z}_i 、そしてデータとして入手できない他の諸要因 u_i によって決定されている。係数の組（列ベクトル） β_2 は、政策効果を表している。

(Ⅲ) 以後は、説明変数が単一の場合と同様である。データを集め、係数 $(\beta_0, \beta_1, \beta_2)$ を最小二乗法などで推定する。すべてのHWについて、期待される成果と、その実績値との差 $p_i - \hat{p}_i$ を計算する。

以上の方法によって、各HWの置かれた状況や業務体制に関する様々な違いを調整した上で、パフォーマンスを吟味できる。また、どのような業務体制や施策がどの程度成果に結びつくのか、つまり β_2 の大きさが推定できる。⁷

ただし、係数の推定値について因果関係の解釈が可能であるためには、 $(\mathbf{x}_i, \mathbf{z}_i)$ の外生性 (u_i との無相関) という条件が必要である。たとえば、一般にHWはクライアントを選別せず、とくに年齢差別を禁じているため、求職者の平均年齢 x は各HWにとって外生的に与えられる要因と想定してよいであろう。また、各HWには高度な設備を独自に選定・購入する予算上の裁量がないとすれば、求人検索機の台数 z も各HWにとって外生的と想定できよう。⁸

Wooldridge(2006)によると、同種の手法は米国で「御値打ちな」不動産物件の選定や、学校教育の質をめぐる司法判断で利用されている。行政学の初等的解説に Rubenstein, Schwartz and Stiefel (2003), Courty and Marschke (2004), Grubb (2004) がある。

⁷ 各係数の解釈は次の通り。簡単化のため、各変数は1次元だとすれば、係数 β_1 は、政策変数 z_i が一定のときに、外的要因 x_i が1単位増えることで（たとえば当該HWの求職者の平均年齢が1歳上昇することで）、成果指標がどれだけ変化するかを表す。また係数 β_2 は、外的要因 x_i が一定のときに、政策変数 z_i を1単位増やすことで（たとえば求人検索機を1台増やすことで）成果指標をどれだけ変化させるかを表す。

⁸ とくに、政策のインパクト β_2 を正しく推定する条件（政策変数 \mathbf{z}_i の外生性）は、関連する他の外的要因 \mathbf{x}_i を同時に考慮することで達成される可能性が高まる。これが、非実験的なデータにもとづく政策分析において多変数モデルを採用することの主な理由である（いわゆる観測可能な変数 \mathbf{x}_i にもとづく \mathbf{z}_i のセレクション、あるいは p_i と \mathbf{z}_i の条件付独立の前提。cf. Wooldridge (2001))。

2-3. 調整指標の特徴と含意：仮説生成とモデル評価の往復

この段階にいたって浮上するひとつの興味は、調整後になお残っている成果水準のばらつきは何に由来するものか、であろう。様々な解釈がありうる。しかし、 u_i がデータ上観測できない以上、それらは定量的データの裏づけがないという意味ではすべて同等である。

そこで、次の段階の作業は、新たな定性的な仮説の生成、そしてモデルの修正・再推定・検証である。当該組織が存続し、環境が変化する限り、この往復プロセスに論理的な終点はない。本節はその点を中心として調整成果指標の特徴をあげる。

(1) まず、上記の手法は、比較的シンプルであるという利点がある。最小二乗法の知識さえあれば手法を理解できるし、実際の計算もエクセルなどで容易に実行できる⁹。また、データが利用者レベルのマイクロデータでなく、各組織や行政区域レベルの集計データであっても、一定程度に有用な分析が可能である（第4節で分析例を示す）。

(2) 調整成果指標のばらつきを、個々のオフィスの「努力」や「技能」のばらつきと同一視することはできない。ばらつきの原因は、「調整に利用した要因以外の何か」であり、それ以上のことを所与のデータから確言することは不可能である。¹⁰

(3) この手法は、データ分析と現場の調査とを往復することで、漸進的に分析の質を向上させてゆくことが可能である。実績値と期待される値の差 $p_i - \hat{p}_i$ (=調整済み成果指標の値) が大きくプラスであったHWについては、個別にヒアリング調査などを行うことによって、他所のパフォーマンス向上につながる要因（いわゆるベスト・プラクティス）が見出される可能性がある。また逆に、実績値が期待される値に遠く及ばなかったHWを調査すれば、見過ごされていた外的ないし構造的な問題が明らかになるかもしれない。

こうして、いわば正負の「外れ値」を調査することで発見された新たな要因は、次回の分析に説明変数として導入できる。この「モデル設定・データ分析→外れ値のオフィスを調査

⁹ ただし、以上はもっとも単純なケースであった。データの量が大きくなると、表計算ソフトでは対応できない。データの質が低い場合、素朴な最小二乗法では精確な推定は難しい。より洗練されたモデルと、強い仮定が必要となる（第3節で後述）。とはいえ、調整指標のアイデアと手順に本質的な変更は生じない。

¹⁰ マクロ経済学でいえば、実物的景気循環モデルは残差項の動きを技術ショックと同一視し、サプライサイドの効率性を論じる。つまり、本来「モデル外の諸要素」として消去法的に規定されるにすぎない残差項を、積極的に特定の要因と同一視してモデル全体の解釈の礎としている。一方、調整成果指標にとって、残差項のばらつきは解釈のゴールではなく、そこに隠されている要因を発掘するためのヒントである。次の(3)を参照。

(4) この往復プロセスが実現するためには、新たに見いだされた要因（の候補）について、データを収集する体制が整備されていくことが当然必要である。

(5) 一度分析に使用した説明変数が、成果指標にとって定量的に重要な規定要因であるか否かは、係数の推定後に統計的検定を行うことによって判断する。そして、ある政策の効果が実質的にゼロに等しい場合は、当該政策の停止を検討すべきであるし、効果がゼロではないにせよかなり小さい場合は、当該政策の費用との兼ね合いで純便益の大きさを確認することが望ましい（第 6-6 節および 6-7 節で詳述）。

(6) この手法は、時系列でデータが入手しうる場合は、前期比でみた成果について検討することも可能である。ただし、一般に、成果の変化分は成果の絶対水準よりも組織間でのばらつきが少ないため、限られた標本から成果水準の有意な差を検出することは難しくなる（第 5 節で詳述）。

(7) 上記の手法は事後評価であった。事前に目標を設定したい場合、たとえば t 期において $t+1$ 期の目標水準を設定するには、まず t 期のデータから係数 β を推定し、次に何らかの方法で次期の要因 x_{t+1} を予測すれば、 \hat{p}_{t+1} を t 期において計算できる。この事前の設定水準と、 $t+1$ 期の実現値との間にどれだけ乖離が生じるかは、 x_{t+1} の予測の精度、およびモデルのあてはまりのよさと安定性次第である。

(8) 成果を規定する要因はあまりに多く、どんなに費用をかけても、すべての要因のデータを集めるのは不可能である。ではこの不可能性をどう考えるか。一つには、これは例の「コップに水が半分入っているのをどう見るか」の問題と同じ構造だと考えられる。つまり、「半分しか入っていない」と思うか、「半分も入っている」と思うかの違いである。

執筆者自身は、「半分は入っている」派、あるいは、「少しずつ、もっと注いで行くべき」派になる。前者は、「条件の違いをまったく考慮せずに成果のみを比較する方法に比べるなら、調整成果指標の方法は改善であるはず」との立場である。後者は、現場調査とデータ分析を往復し、埋もれていた要因を掘り起こしていく志向を指している。

ル間の識別、新たな理論仮説の生成、そしてアカデミックな研究と社会的ニーズの接合の契機となる（中馬（2002）、Bewley（1995、1998、1999）、Blinder et al.（1998））。

(9) 最後に、回帰分析で使っている諸要因が、成果のデータの動きを何%くらい説明できるか、つまり「説明力」の大きさは、一応数字としてわかる（決定係数 R^2 ）。いわば、コップには半分水が入っているのか、三分の一入っているのか、にあたる情報である。この情報を使うと、要因の探索に目途がつく。

ただし、何%の説明力に到達したら満足すべきかについて、理論的、先験的に正しい基準はない。ある程度の説明力を達成することは望ましいが、直感的に妥当でない説明変数（因果関係の意味で成果を左右する要因とは考えがたいもの）を採用して説明力を上げて意味はない。とくに、政策変数の係数（政策効果）の推定値の信頼性は、モデル全体のあてはまりの良さとは必ずしも関係ない。業務改善のヒントを得ることを目的とする以上は、第一義的なのはアウトプットを規定する原因の探索である。

例 2-3-1. HW の成果水準とのら猫の数に正の相関が見られるとする。後者を変数として右辺に加えるならば、モデルの説明力をあげることはできる。しかし、分析者はむしろ成果水準と猫の数の両方を規定する原因（繁華街の景気など）を探索すべきであろう。政策当局が地域猫を保護するなどして猫を増やしても、おそらく HW の成果水準は改善しないからである。猫の数は労働市場や景気循環のシグナルではあるが、政策変数として有用な因果性をもたない。

例 2-3-2. 予測（Forecasting）と因果モデル推定とは目的が異なる。予測は、因果のプロセス理解についてはまったく教えるところがないブラックボックスであるとしても、予測としては正確・有用でありうる。たとえば、天気予報、あるいは傘を携帯した通行人の数は、降雨の原因ではない。しかし、それらは降雨と相関があり、したがって予測（降雨に備える）には有用である。

3. データの種類に対応した分析手法

上述したアプローチは、問題設定と利用可能なデータの種類に対応して拡張できる。

3-1. 各組織レベルの一時点のデータにおける問題

前出の単回帰モデルの設定に戻る。簡単化のため切片を省略すると、各組織の調整成果指標は残差 \hat{u}_i （観測不可能な誤差項の推定値）として次のように求められるのだった：

$$p_i = \beta x_i + u_i, \quad i = 1, 2, \dots, N$$

$$\hat{p}_i = \hat{\beta} x_i$$

$$p_i - \hat{p}_i = p_i - \hat{\beta} x_i = \hat{u}_i$$

いま、データ上観測不可能な要因 u_i は、独立に分布する二つの要素の和で構成されるとしよう。一つは各組織に固有の（内部的あるいは外部的な）要因 a_i であり、もう一つはそれ以外の純粋なノイズ v_{it} である。つまり、 a_i は組織改革・業務改善のヒントになりうる情報だと考える。そこで上の手順を書き直すと、

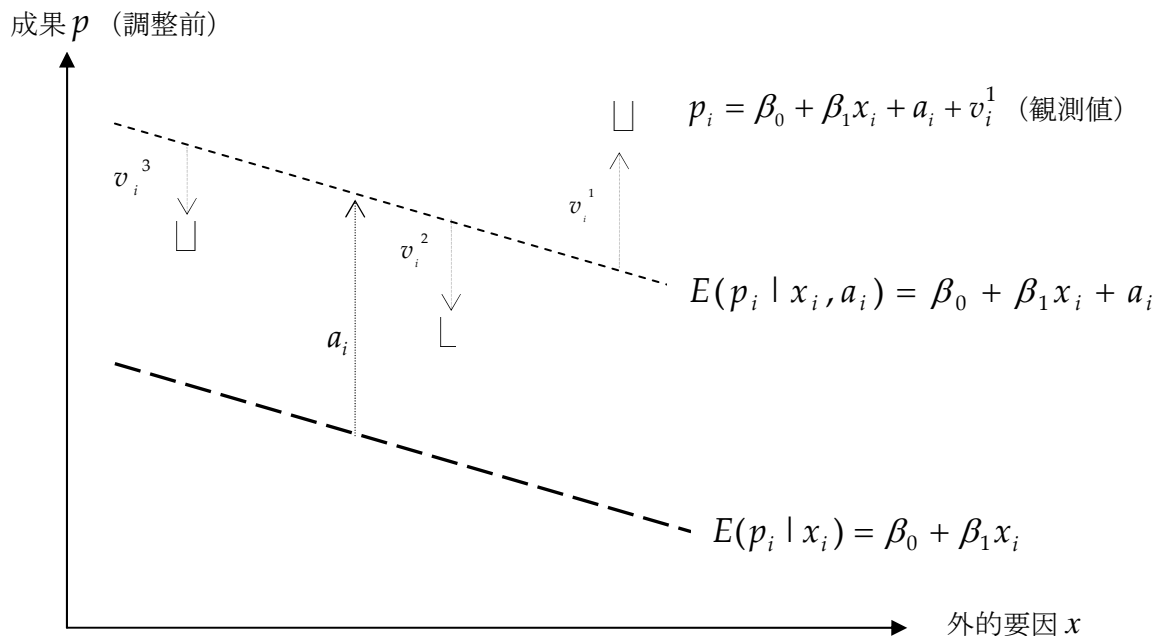
$$p_i = \beta x_i + u_i = \beta x_i + a_i + v_i$$

$$\hat{p}_i = \hat{\beta} x_i$$

$$p_i - \hat{p}_i = p_i - \hat{\beta} x_i = \hat{a}_i + \hat{v}_i$$

1 時点のデータからは、調整成果指標の水準に然るべき理由がある (a_i) 場合を単なるノイズ (v_{it}) から識別できない (図 3-1 参照)。また a_i と x_i に相関がある場合は、最小二乗推定量は係数 β を正しく推定できない。すなわち、データ上で観察が不可能な各組織の性質の影響と、観察可能な要因の影響とを分離して推定できない (脱落変数によるバイアス)。¹³

図 3-1 組織固有の要因 a_i とノイズ v_{it} (概念図)



¹³ ひとつの対処方法は、操作変数を探すことである (e.g. Wooldridge (2006)). 他には、本文に述べるようにパネルデータを利用する方法がある。

3-2. 各組織レベルのパネルデータの場合

データに新たな次元（たとえば時系列）が加わると、情報量が増す¹⁴。各ユニットを追跡調査したパネルデータがある場合、

$$p_{it} = \beta x_{it} + a_i + v_{it}, i = 1, 2, \dots, N, t = 1, 2, \dots, T$$

もしも x_{it} が a_i および v_{it} と無相関ならば、サイズ $N \times T$ の標本に最小二乗法を適用して、係数 β を推定しつつ、 a_i を v_{it} から識別できる。こうして把握された各組織固有の、期間を通して不変な要因 a_i の情報は、中長期的観点からの業務改革に有用かもしれない。¹⁵

時系列データがあっても、 a_i と x_{it} に相関がある場合は、最小二乗法では係数 β は（それゆえ a_i も）正しく推定できない。しかし、一定の条件のもとで、固定効果モデルを利用するならば、各変数の期間平均からの乖離を利用してモデルを変換して観測不可能な a_i を消去し、

$$\begin{aligned} p_{it} &= \beta x_{it} + a_i + v_{it}, i = 1, 2, \dots, N, t = 1, 2, \dots, T \\ \bar{p}_i &= \beta \bar{x}_i + a_i + \bar{v}_i & (\bar{x}_i = T^{-1} \sum_{t=1}^T x_{it}) \\ p_{it} - \bar{p}_i &= \beta (x_{it} - \bar{x}_i) + v_{it} - \bar{v}_i \end{aligned}$$

これを $\dot{p}_{it} = \beta \dot{x}_{it} + \dot{v}_{it}$ と書きなおすならば、第 2-1 節でみた設定と形式上等しく、したがって同様の仮定のもとで \dot{p}_{it} を \dot{x}_{it} に回帰する最小二乗法により係数 β を正しく推定できる¹⁶。次に係数の推定値を利用して全体の理論値を求め、 a_i の推定値を計算する。

$$\hat{a}_i = \bar{p}_i - \hat{\beta} \bar{x}_i$$

第 2 節以降みてきた一時点データの場合には、調整指標は

¹⁴ マクロとミクロの意味（たとえば地域レベルと各オフィスレベル）の多次元データを活用するアプローチとして、調整成果指標における集計バイアスの検知がある。付録 9-4 節を参照。

¹⁵ $a_i = p_{it} - \beta x_{it} - v_{it}$ より、時系列で期待値をとると

$$E(a_i) = E(p_{it} - \beta x_{it} - v_{it}) = E(p_{it}) - \beta E(x_{it}) - E(v_{it}) = E(p_{it}) - \beta E(x_{it})$$

$E(p_{it})$ と $E(x_{it})$ は、 T が大きければ標本の期間平均 \bar{p}_i と \bar{x}_i で推定できる。それゆえ $E(a_i)$ の推定の成否は、 β の推定値の正確さ次第となる。

¹⁶ 第 2 節とアナログ的に、 $E(v_{it} | x_{it}, a_i) = 0$ は仮定する。標本において $N \rightarrow \infty$ あるいは $T \rightarrow \infty$ のとき、 \dot{p}_{it} を \dot{x}_{it} に線形回帰する最小二乗推定量が β に近づくために必要であるのは $E(\dot{v}_{it} | \dot{x}_{it}) = E(v_{it} - \bar{v}_i | x_{it} - \bar{x}_i) = 0$ 、すなわち母集団において誤差項 v_{it} の平均が全期間の説明変数 $(x_{i1}, x_{i2}, \dots, x_{iT})$ から独立という条件である。それゆえ過去の p が今期の p を直接に左右している場合、この条件は成立しない (Arellano (2003), Wooldridge (2001, ch.10))。

$$\hat{u}_i = p_i - \hat{\beta}x_i = p_i - \hat{p}_i$$

であった。一方、調整指標 \hat{a}_i は、各変数の期間平均から計算されている。これはパネルデータを利用した、長期的な調整成果指標のひとつと考えることができる。じっさい、外的要因 x_{it} の期間平均が等しい所同士は、期間平均で同水準の実績値が期待されている。¹⁷

4. 応用例：県別データによる実証分析

本節は、HWの就職件数、有効求職者数、有効求人数を県単位で集計した時系列データに対して、これまで検討してきたアプローチを適用する。

4-1. データ

データは、厚生労働省の職業安定業務統計のうち月次・県単位で公表されている「一般職業紹介状況」からとる。成果の調整要因としては月間有効求職者数と月間有効求人数、そして成果指標としては就職件数をもちいる。定義は以下の通り。「新規求職申込件数」は、期間中にあらたに受けた求職申込の件数をいう。「月間有効求職者数」は、「前月から繰り越された有効求職者数」と当月の新規求職申込件数の合計数をいう。「求人数」についても同様である。「就職件数」は、自安定所の有効求職者が、自安定所の紹介により就職したことを確認した件数をいう。¹⁸

¹⁷ なお、直接的に各組織をあらわすダミー変数 $d_i, i=1,2,\dots,N$ を導入して

$$p_{it} = \beta x_{it} + \delta_1 d_1 + \delta_2 d_2 + \dots + \delta_N d_N + v_{it}, i=1,2,\dots,N, t=1,2,\dots,T$$

の各係数を推定する方法もある。両者の推定値は等しい。明らかに、組織の数 N が大きい一方で時系列の次元 T が小さい場合は、標本サイズに比して説明変数の数が大きすぎるために、 a_i または $\delta_i, i=1,2,\dots,N$ の正確な推定は不可能である。(標本中の HW が全部で 600 箇所あるとして、3 時点のパネルデータの場合、サンプルサイズは $N \times T = 1800$ だが、推定すべきパラメータは少なくとも 601 個にのぼる。) 見方を変えると、この推定量は標準的な線形回帰モデルの仮定のもとで不偏性を満たすが、一致性のためには $N \rightarrow \infty$ でなく $T \rightarrow \infty$ が必要である ($i=1,2,\dots,N$ であるからクロスセクション単位の増加には情報価値がない。Wooldridge(2006, ch.14), Cameron and Trivedi(2006, ch.11)). むろん、 $T \rightarrow \infty$ の場合は時系列分析に特有の問題(非定常性など)が現れる(Arellano(2003), Kano and Ohta (2004) 参照)。

¹⁸ 対象となる雇用カテゴリーの定義は以下の通り。「一般」は常用及び臨時季節を合わせたものをいう。「常用(労働)」は雇用契約において雇用期間の定めのない仕事(労働)、又は 4 か月以上の雇用期間が定められている仕事(季節労働を除く.)をいう。「臨時」とは、雇用契約において 1 か月以上 4 か月未満の雇用期間が定められている仕事(労働)をいい、「季節」とは季節的な労働需要に対し、又、季節的な余暇を利用して一定の期間(4 か月未満、4 か月以上の別を問わない。)を定めて就労する仕事(労働)をいう。

クロスセクションは都道府県 ($N=47$), 時系列は月次データ 13 年間分 ($T=13 \times 12=156$) からなるパネルデータを作成した. 標本サイズは $NT=7332$ である. 付録 A-4 節に, その記述統計 (表 A-1) を, また, 全国レベルに集計した時系列データ 2 種類をグラフに示している. 図 A-2 は総計, 図 A-3 は中位数 (50%目の値=メディアン) である. 次頁の図 4-1 と 4-2 には, 各県ごとの求人求職件数の全期間の分布を箱ひげ図 (ボックス・プロット) で示した. 各箱内の白線は中位数をあらわす. 地域によって変動幅はおおきく異なる. 大都市圏では変動が大きい. 北海道の中位数は主要大都市のそれと近いが, 同時に, 求職者数の幅が非対称的に広い.

図 4-1 求人数, 求職者数 (千人) の分布と中位数: 北海道から長野 (1992-2004)

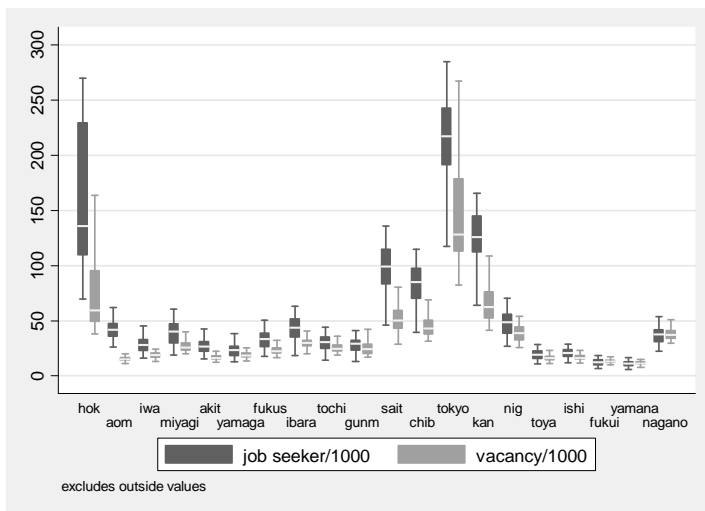
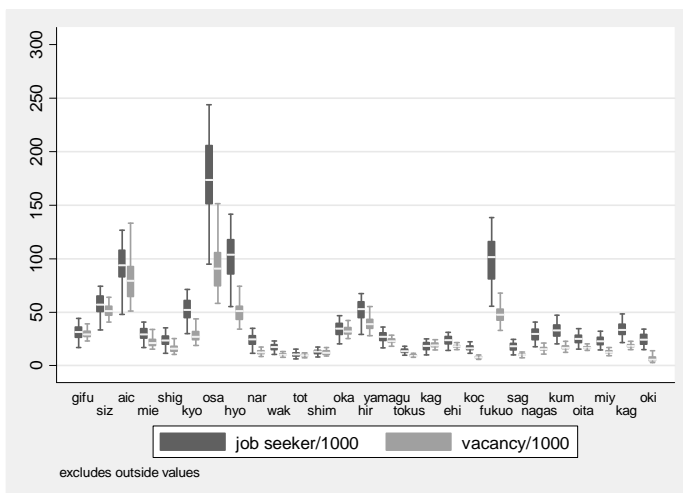


図 4-2 求人数, 求職者数 (千人) の分布と中位数: 岐阜から沖縄 (1992-2004)



4-2. 推定するモデル

以上のパネルデータにおいて、就職件数を従属変数として、有効求人数と有効求職者数に回帰する。成果指標の調整要因と関数形を選定するにあたって、モチベーションを労働経済学に求める。形式的にはこのモデルは標準的なマッチング関数と同値である（関連する研究との比較は第4-10節で後述）。すなわち、就職件数 M は求職者数 U と求人件数 V の関数、とくにコブ＝ダグラス型関数

$$M = F(U, V) = U^{\beta_u} V^{\beta_v}$$

であるとひとまず想定する。 $0 < \beta_u < 1$ の場合、求職者数が追加されるたびに（他の事情が一定ならば）就職件数は増加するが、同じ勢いでは伸びずに逡減する。係数に関するこの当否は、推定結果から検証できる。 β_v と求人件数についても同様である。なお、両辺の対数をとって線形に変換した式 $\ln M = \beta_u \ln U + \beta_v \ln V$ の係数をデータから推定するのが便利である。¹⁹

この節では、3-2節でみた固定効果推定量を利用する。すなわち、各県には期間を通じて観測不可能で安定的な個別要因の水準 a_i が存在するものと想定し、かつそれが求職や求人の数と相関している可能性を考慮したうえで、その個別効果 a_i を含めた各パラメータを推定する。たとえば、非都市圏など、社会経済の流動性が少なく、求人求職の性質が均質的であり、地元企業や教育機関との信頼・連携関係や求人開拓などのマッチングのノウハウが継

¹⁹ したがって、両説明変数の係数は弾力性の値として解釈できる。以下では、都道府県パネルデータの構造を利用するために、さらに次のようにモデルの特定を試みる。

$$M_{i,t} = F_{i,t}(U_{i,t}, V_{i,t}) = B_{i,t} U_{i,t}^{\beta_u} V_{i,t}^{\beta_v} = B \exp(a_i + w_t + z_{i,t}) U_{i,t}^{\beta_u} V_{i,t}^{\beta_v}$$

ここで、 a_i はクロスセクションの個別効果、 w_t はクロスセクションに共通の期間効果、 $z_{i,t}$ は純粋なノイズである。 w_t は一般に、第 t 期に 1、それ以外は 0 の値をとるダミー変数、あるいは四半期ダミーやトレンド項を含むが、本稿では各期のダミー変数を用いる。（もっとも、北海道と東京と沖縄など地域によってトレンドや季節性が異なる可能性は高い。季節調整の是非や、期間の設定は、分析目的に依存する（田中（2002, 第5章））。季節性込みで「次期の成果」を予測したい場合は、過去の季節調整前のデータ（原数値）を用いることには合理性があろう。経済学でいう「消費の平滑性仮説」を、年末年始の消費データから検定する場合も同様）。この非線形関数のパラメータは、対数変換した線形関数

$$m_{i,t} = b + \beta_u u_{i,t} + \beta_v v_{i,t} + a_i + w_t + z_{i,t}$$

から推定するのが便利である。標準化の仮定 $\sum_{i=1}^N a_i = 0 = \sum_{t=1}^T w_t$ を加えて、固定効果モデルのパラメータを推定するのが、この節でのアプローチである。あるいは、確率変数の分布 $a_i \square iid(0, \sigma_a^2)$, $w_t \square iid(0, \sigma_w^2)$, $E(a_i + w_t + z_{i,t} | u_{it}, v_{it}) = 0$ を想定し、多次元モデルをあてはめることもできる。本稿は、第5節で指標の不確実性を議論する際に多次元モデルを利用する。

承されている地域では、そうでない地域に比べると、求人件数や求職者数も多い、といった可能性がある。²⁰

スケール・メリットが一定のときは、就職率や充足率は労働市場の有効求人倍率（需給逼迫率）の関数となる（規模に関する収穫一定は、マッチング機構に混雑効果などの外部性が存在しないことと解釈できる）。後に見るようにこの点はデータで検証可能である。²¹ a_i の違い以外にも、県（クロスセクション）で求人と求職にばらつきがあるのは、地理的、あるいは行政区分上の外生的事情によるものとみなす。また、マクロの期間ダミー変数によって、全国レベルの変動（景気循環など）の影響を吸収する。

なお、第2節で扱った1時点データの分析方法は、ここでの長期データにも応用できる。たとえば、過去にさかのぼって各期間ごとに調整成果指標を計算し、その時間的な推移を観察する。あるいは、実際に有効求人倍率などについて行われているように、每期（月次や四半期、年次）のデータを用いてリアルタイムで調整指標を計算し、「足元」の状況を速報的に把握しつつ、次期のベンチマーク水準を設定する、といった利用方法も検討に値しよう。

ただし、紙幅の都合もあるので、以下ではあえて対極的なアプローチのみを実行する。すなわち、過去のデータを一度に利用して、中長期的なスパンで各地域の調整指標を求め、その地域間での分布を吟味する。

4-3. 推定結果

各期・各県の就職件数は、求職者数、求人数、各県の個別効果、各期のダミー変数、誤差項と定数項からなる関数と考える。対数線形式

$$m_{i,t} = b + \beta_u u_{i,t} + \beta_v v_{i,t} + a_i + w_t + z_{i,t}$$

に第3-2節で見たデータ変換をほどこし推定量を計算した結果を表4-1に示す。期間ダミーの係数は省略した。有効求職者数と有効求人数は、ともに就職件数に対して有意な正の弾性値が推定されている。両推定値の合計は1.009であり、ほとんどの先行研究と同様に、マッ

²⁰ 第3-2節の表記を使うならば、 $p_{it} = \beta x_{it} + a_i + v_{it}$ において x_{it} と a_i に正の相関がある。

²¹ マッチング関数 F が規模について収穫一定（一次同次）ならば、求人充足率について $M/V = F(U, V)/V = F(U/V, 1)$ 、また就職率について $M/U = F(U, V)/U = F(1, V/U)$ がなりたち、いずれも需給逼迫率の関数である。一次同次性は、次の4-3節でみるようにデータから検証可能である。もっとも、本来はここで求職者数と求人数だけではなく、HWの紹介担当スタッフ数をも考慮すべきであろう。すなわち、求職、求人、スタッフというインプットを定数倍したときに、マッチングというアウトプットは等しく定数倍されるか否かを検証したい。しかし、これはスタッフ数のデータがないために不可能である。

チング関数は規模に関して収穫一定であると解釈できる²²。モデル全体の説明力 (R-sq) も比較的に高い。²³

表 4-1 固定効果モデルの推定 被説明変数：対数就職件数

説明変数	係数推定値	標準誤差	T 値	P>t	[95%信頼区間]	
対数求職者数	0.452	0.081	5.610	0.000	0.290	0.614
対数求人数	0.557	0.233	2.390	0.021	0.087	1.027
固定効果 (within)回帰分析			サンプルサイズ	7332		
			クラスター (グループ) サイズ	47		
R-sq: within (グループ内)	0.798	グループ規模: 最小値		156		
Between (グループ間)	0.899	平均		156		
overall (全体)	0.877	最大値		156		
F 値(157,46)	279.790					
Prob > F	0.000					

(標準誤差は 47 のクラスター内相関について頑健な推定量)

次に、第 3-2 節で触れた方法によって、各県の固定効果 a_i の推定値を求めた。図 4-3 は、その点推定値を各県が所属する地域ブロック名とともにプロットしている。中位層の諸地域は水準に大きな差がない。ひとまず上位層 (0.5 の近傍) と下位層 (-0.5 の近傍) を比較してみると、興味深いことに、一般に高い失業率や低い求人倍率で知られる地域 (北海道・東北、沖縄九州) は、調整成果指標 \hat{a}_i で見ると上位にいる。一方、南関東、東海など、経済の中心地は下位である。

表 4-2 は、就職率および求人充足率の原数値の期間平均と、調整成果指標とで、順位にどの程度変化があるかを示している。③の調整後順位の列が、調整指標の図 4-3 に対応している。調整による変動はしばしば大きく、とくに北海道・東北、および九州・沖縄の諸県に顕著である。最大のケースでは、調整によって 28 位も上昇しており、順位表が示唆する印象はまったく変わってしまう。

²² 両推定値の和と 1 の差がゼロであるとの帰無仮説は統計的検定で棄却される。なお、各推定値の標本分散は、地域間の不均一分散と各地域内の自己相関に対して頑健な (cluster-robust) 推定量による (Cameron and Trivedi, (2005, ch.21))。

²³ もっとも、長期時系列のデータにマクロ期間ダミーをあてはめるとき「説明力」が高くなるのは珍しくも興味深くもない。第 2 節で触れたように R-sq は優先的な基準ではない。

図 4-3 調整指標の分布（県別，地域ブロック表示）

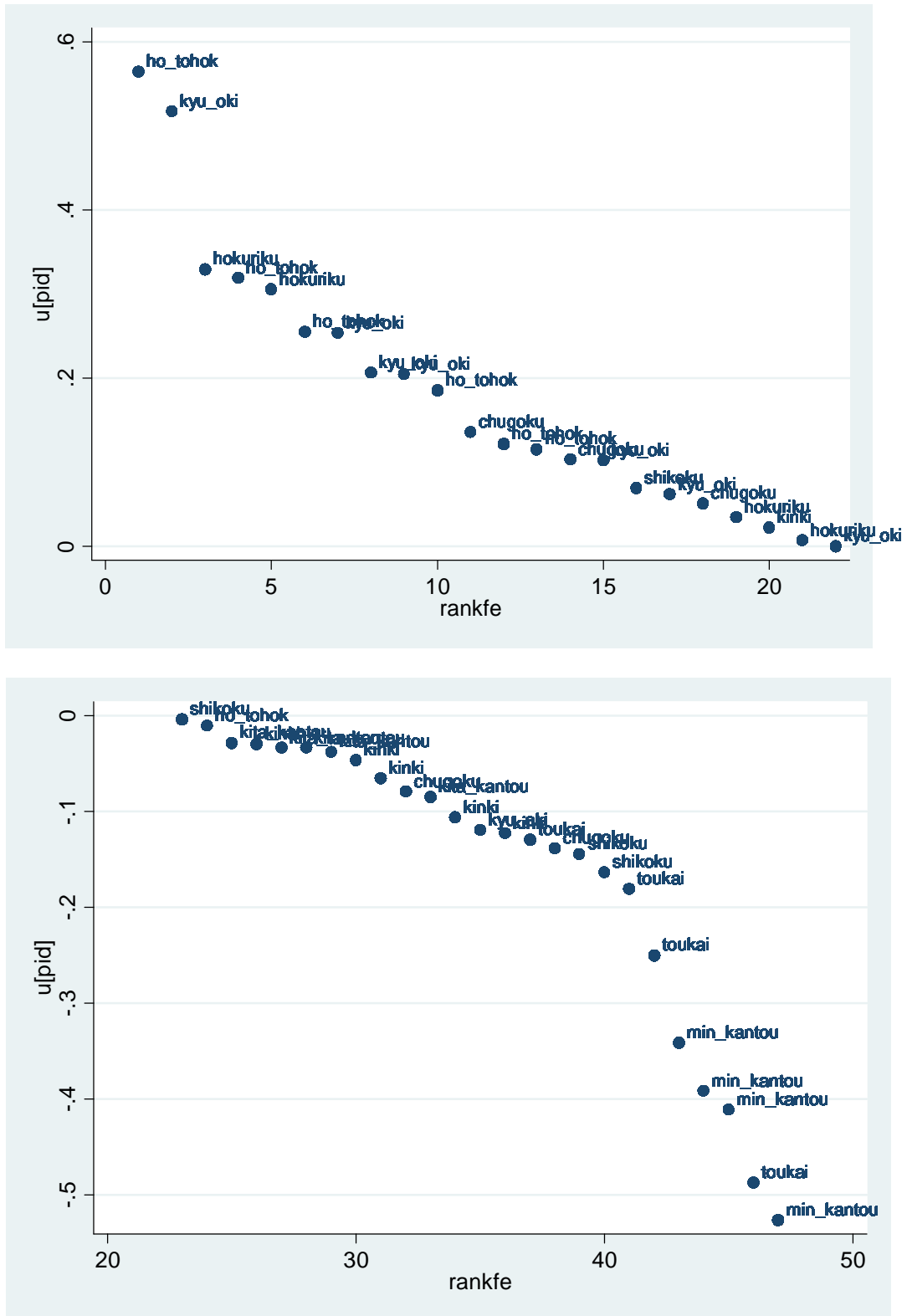


表 4-2 求人充足率・就職率の期間平均，調整成果指標 \hat{a}_i による順位表（県別，地域ブロック表示）

県（所属地域ブロック）	①充足率順位	②就職率順位	③調整後順位	②-③	① ③
北海道・東北	3	1	1	0	2
九州・沖縄	1	30	2	28	-1
北陸	8	2	3	-1	5
北海道・東北	2	26	4	22	-2
北陸	10	3	5	-2	5
北海道・東北	4	4	6	-2	-2
九州・沖縄	5	16	7	9	-2
九州・沖縄	7	20	8	12	-1
九州・沖縄	6	19	9	10	-3
北海道・東北	9	15	10	5	-1
中国	23	6	11	-5	12
北海道・東北	14	14	12	2	2
北海道・東北	18	10	13	-3	5
中国	25	7	14	-7	11
九州・沖縄	15	18	15	3	0
四国	32	5	16	-11	16
九州・沖縄	11	29	17	12	-6
中国	28	12	18	-6	10
北陸	27	13	19	-6	8
近畿	13	31	20	11	-7
北陸	36	8	21	-13	15
九州・沖縄	12	35	22	13	-10
四国	30	23	23	0	7
北海道・東北	24	25	24	1	0
北関東	40	9	25	-16	15
近畿	16	38	26	12	-10
北関東	26	28	27	1	-1
北関東	34	17	28	-11	6
北関東	37	11	29	-18	8
近畿	17	37	30	7	-13
近畿	29	32	31	1	-2
中国	31	27	32	-5	-1
北関東	39	21	33	-12	6
近畿	21	39	34	5	-13

九州・沖縄	20	41	35	6	-15
近畿	19	40	36	4	-17
東海	42	22	37	-15	5
中国	43	24	38	-14	5
四国	22	43	39	4	-17
四国	33	36	40	-4	-7
東海	38	34	41	-7	-3
東海	45	33	42	-9	3
南関東	35	44	43	1	-8
南関東	41	45	44	1	-3
南関東	44	47	45	2	-1
東海	47	42	46	-4	1
南関東	46	46	47	-1	-1

4-4. 調整指標 \hat{a}_i の分布とは何か？

そもそも、これらの図表に現れた \hat{a}_i のばらつきは、何を意味しているか。ラフにいえば、それはマッチングと相関のある要因のうちで、マクロ的な変動（景気や法改正）および地域内の需給条件（求人・求職状況）以外の、「残りの要素」の分布である。すなわち、 \hat{a}_i とは、各地域において期間を通して安定的であり、直接的には測定されていないすべての要因の影響力をまとめたものである。

見方を変えると、それはアウトプットとインプットの因果関係についての既存の仮説をしぼりこみ、新たな仮説を生成するための参考情報となる。われわれの文脈では、たとえば、

- ①競合性： 地方の労働市場においては、建設業などの比率が高い。職業紹介事業に収益性がないため、民間企業による紹介機関があまり存在しない。逆に、ホワイトカラーの求人・求職が多い大都市圏では、HW に求職者登録をして求職者手当や情報検索などのサービスを受けつつ、最終的には代替的なルートで再就職する例が多い
- ②情報の質： 図 4-1 と 4-2 が示唆するように、地方では人口の流出入や技術革新をふくむ社会経済上の変化が激しくない。そのため、求人側・求職側・紹介機関・教育機関のネットワークが安定的である。長期的関係に由来する信頼感もあって、情報の信頼性が高い
- ③業務の多様性： 地方では社会経済が均質的であり、HW のクライアントの性質やタスクに対立的な多元性が少ないため、HW の業務コストが低い

といった仮説が考えられる。

また、一般に若年層にくらべて、中高年層は IT に通曉していない。そのため、人口が高

齢化した地方では、求人側・求職側の双方にとってHW（そして窓口での相談・紹介）の重要性が高いのかもしれない。²⁴

このほかにも、さまざまな因果的解釈が付与できるであろう。しかし、今回のデータのみを所与とすれば、それらの解釈はすべて仮説として同等である。²⁵因果的解釈を選別するための、より正統的かつ強力な方法は、第2-3節で触れたように、関係者からの定性的なフィードバックを受けて、新たな変数のデータを追加し、再び調整指標を求めることである。われわれの文脈では、各県の労働需給状況に加えて、業種・学歴・年齢のばらつきや、開廃業率、人口移動、インターネット普及率などの時系列データをもとに、上記の仮説の検定を行なうことが考えられる。ただし、それは「第2ラウンド」の作業であり、本稿の範囲をこえる。

4-5. 先行研究との比較

ここまでの議論と、既存の文献の関係に触れる。労働経済学のマッチング関数の理論と実証の展望論文に Petrongolo and Pissarides (2001)がある。日本のデータを使った分析としては、マクロデータを用いた中村(2002)、都道府県パネルデータを用いた佐々木(2004, 2007)、神林・上野(2006)、Kano and Ohta(2004, 2005)などがある。通常のマッチング関数の研究では、因果関係を表現する経済モデル（インプット＝求人・求職と、アウトプット＝マッチの安定的な関係）を想定し、それを特徴付けるパラメータをデータから推定して、理論仮説の妥当性や労働市場の特性を検討することに主目的がある。たとえば市場にスケール・メリットや混雑効果があるか、求人・求職というインプットのうちストック（繰越分）とフロー

²⁴ 業務統計の地域別時系列データを見るにあたっては、HWのIT化と広域紹介拡大にともなって、業務統計と実態との地理的ギャップが広がった可能性に注意したい。90年代から求人求職情報についてIT化が進み、全国のHWの情報ネットワークが拡充した。来所者は所内で検索機が利用できるほか、インターネットでも同種の情報にアクセスできる。紙ベースの求人票で求人情報を公開していた過去に比べると、情報の量および流通範囲が格段に増したことは間違いない。その結果、HW間ではむしろのこと、都道府県間でも、①求人や求職を受理した所、②マッチングに貢献した所、③実際の職場、④データ上でカウントされている所、の間でギャップが拡大している可能性がある。さらに、若年層を中心として派遣や個人請負の比率が増加するにしたがって、登録求人企業の本社所在地と、労働者が実際に働く事業所所在地とが離れている（東アジアなどの海外を含む）事例が話題になっている。

²⁵ 統計的な相関は、特定の因果関係を証明するものではない。ある観測データと整合的な因果関係のモデルは複数存在するため、どのモデルが「良い」かを、所与のデータ自体を根拠に証明することはできず、別の基準（簡便性や直感的妥当性、一般性、審美性）が必要になることは、社会科学では日常的な現象である。（計量経済学における「観測上の同値性」、またはシステム論のいう「モデルの同型性」、付論A-5、またはワインバーク(1979 [1975] 第4章参照)。）

(新規流入分)では性質に差異があるか、経年変化はあるか、などである。

調整成果指標の研究にとっては、それらは第一段階の作業である。第二段階として、個々の組織や地域単位での調整成果指標の分布を把握する。第三段階として、調整指標のばらつき由来を探索するため、定性的な情報を関係者のヒアリングなどで収集して仮説を生成するとともに、定量的データの種類を充実させ、仮説を検証する。(本稿は、第二段階と第三段階の間までを扱う。)

就職率と充足率の二乗平均平方根をマッチング指標とした例に、労働政策研究・研修機構(2006)がある。²⁶周(2007a, 2007b)は「マッチング効率性」の指標として就職率、充足率などを採用し、それぞれの決定要因を都道府県別データから多角的に推定している。一方、本稿の調整指標は、求人の充足と求職者の就職という二つの次元を同時に考慮した指標だといえる。

Kano and Ohta(2005)は、本稿と同じ公共職業紹介の県単位データを、各地域の労働市場の情報と解釈して、「地域労働市場の効率性」を議論している。著者たちが注意している点とおり、解釈にあたっては、地域のすべての求人求職・マッチングがHWの業務データに入ってくるわけではない(とくに都市部ではその乖離が甚だしいであろう)点には注意が必要である。また、行政経営的に県(労働局・HW群)の状況を主に吟味する本稿とは視角が異なる。ただし、固定効果でクロスセクションの差異を吟味する点は同様である。なお、Kano and Ohta(2004)は、長期パネルデータの時系列面での性質に慎重な技術的注意を払っている。

経済理論の枠組みにこだわらず、行政評価的関心から、外的要因を考慮した公共職安の業績評価が、豪州や欧州で始まっている。ドイツはMosley and Mueller(2007)、オランダはvan Donk and de Koning(2007)、スイスおよびドイツはHilbert(2007)、豪州はAccess Economics(2002)を参照。これらは、本稿第2節と似たアプローチで、組織や地域レベルの1時点のデータを用いている。調整成果指標として実績値と予測値の差ではなく比を利用する点は本稿第2節と異なるが、テクニカルには見るべきものはない(ただし、各国の試行錯誤がもたらした政治経済学的な知見は興味深い。第6節で紹介する)。

5. 指標の不確実性の問題

²⁶それぞれ2乗した上で相加平均し平方根をとる($\sqrt{[(M/U)^2 + (M/V)^2]}/2$)。

5-1. 組織の規模と成果指標の変動

現実の成果指標は、標本からの推定値である。標本分散、すなわち推定における不確実性は排除できない。したがって評価に際しては、点での推定のみならず幅を考慮したい。

例 5-1-1. 共通テストの成績による学校の比較を考える。ある学校における生徒の成績の平均値は、潜在的な生徒という母集団から抽出されたその学校における生徒たちという標本から計算される。この標本のサイズが小さい場合、推定値の分散は大きくなる。たとえば小規模な学校における成績の平均値は、少数の突出した生徒の成績によって大きく影響を受けるため、時系列での変動が大きい。逆に、大規模な学校における成績の平均値は、標本が大きい分だけ変動が少ない。

成果の絶対水準でなく、前期からの変化分を指標とする場合は、標本分散の影響は一層小さくなる。変化分は絶対水準よりも小さいため、学校間の差異はもともと少ない。また、生徒（の質）がある iid の分布からの無作為抽出の場合は、成果が改善した次ぎの期には成果は悪化する可能性が高い（平均への回帰、regression to the mean）。実際、大規模校に比べると、小規模校は上位と下位とを頻繁に行き来する。それは標本の大きさが原因である部分が多い（Kane and Staiger (2002)）。

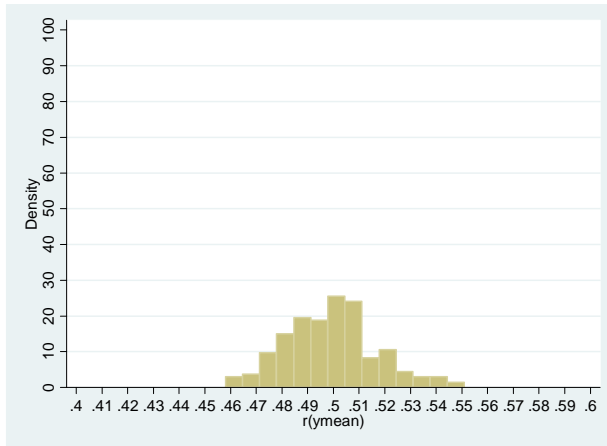
5-2. 規模と変動：数値例

利用者という標本のサイズが成果指標の平均値の分散に与える影響は、単純な人工的データで視覚的に確認できる。いま、利用者のサイズのみが異なるオフィス群があるとしよう。オフィスの効率性に差はない。利用者が 1,000 人のオフィスと 5,000 人のオフィスと 10,000 人のオフィスの三種類が、全国に 200 箇所ずつ存在する。成果指標は、各利用者の「成功／失敗」をあらわす 1 / 0 の二値変数（受験合格、就職など）とする。利用者たちにとって、成功／失敗の確率はどの組織に参加しようと同じく五分五分である。

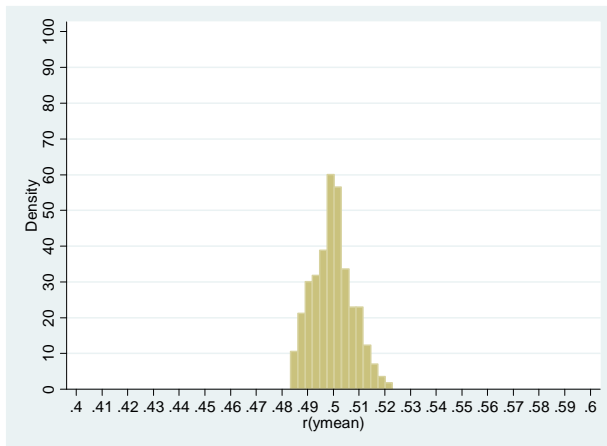
利用者の成果の平均値を、各オフィスごとに計算し、三つの規模ごとに平均値の分布を図 5-1 に描いてみる。オフィスの規模に関わらず分布の中心は 0.5 であるが、規模が大きい組織ほど標本分散は小さくなっている。つまり小規模組織の方が極端に良い（あるいは悪い）成果を得やすい。教訓は、規模という要因を考慮せずに一喜一憂しても仕方がないということであろう。²⁷

²⁷ 図①の作成手順は、各人の成果を潜在的に規定する要因（運や能力）の連続的な分布である $[0, 1]$ 区間の一様分布から、まず 1,000 個の乱数を生成する。次に、それぞれ 0.5 未満ならば 0 に、0.5 以上ならば 1 に変換し、0 と 1 からなる無作為標本を作成する（成功例 = 1，失敗例 = 0 の総計からなり、標本サイズ = 1000）。最後に、その標本の平均を計算する。こ

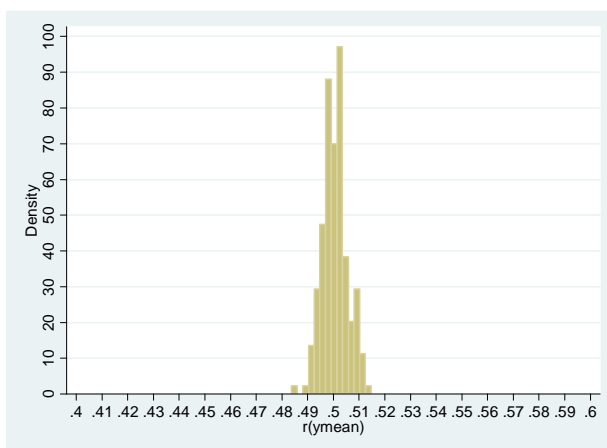
図 5-1 利用者数の差に由来する平均的成果のばらつき



組織数 = 200 , 各組織の利用者数 = 1,000



組織数 = 200 , 各組織の利用者数 = 5,000



組織数 = 200 , 各組織の利用者数 = 10,000

の手順を 200 回繰り返して、得られた 200 個の標本平均の分布を上図①に示している。図②は 5,000 個の乱数について、図③は 10,000 個の乱数について、同様の手順を適用。

5-3. 規模と変動：県別データ

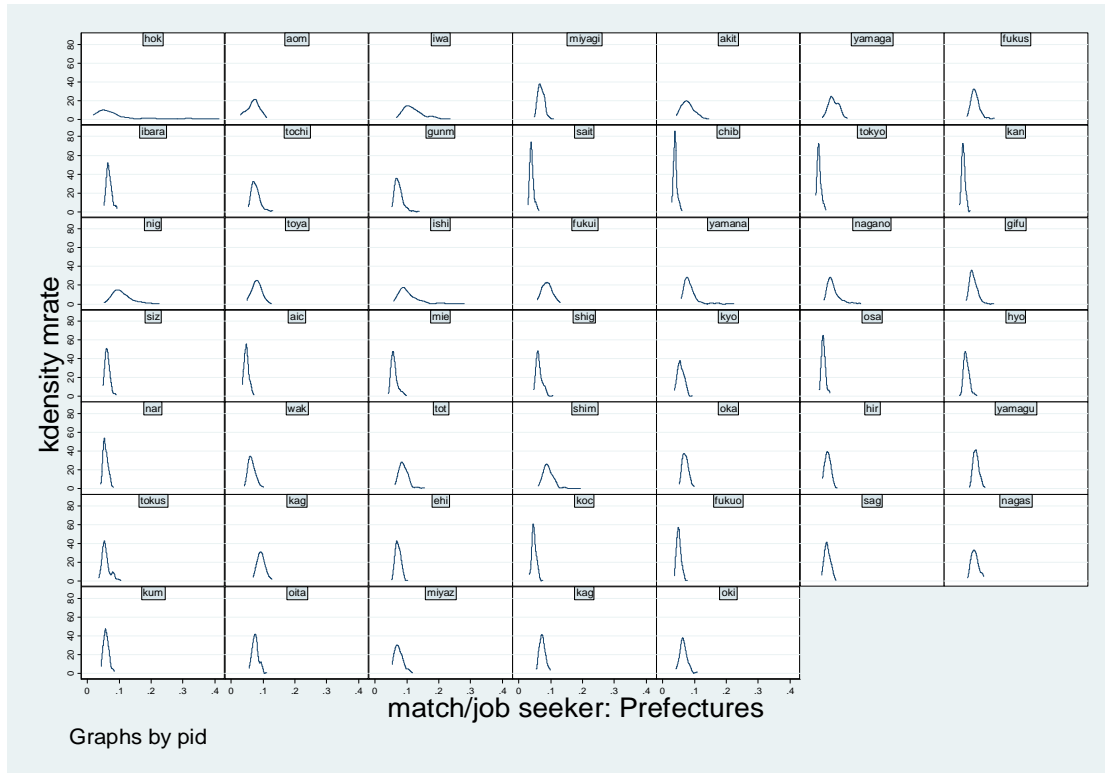
上の議論のアナロジーを、第4節で使用した県別職業紹介データを用いて考えてみる。

図4-1で確認したように、地方のHWと、都心の大規模なHWとでは、利用者という標本のサイズは著しく異なるからである。

全期間の月次就職率の分布を頻度数の平滑近似（カーネル分布）で図5-2に示した。大都市圏は、経済社会的な変動が大きいというイメージがあるが、HWの就職率は、大都市ほど中央値まわりに集中的に分布している。地方は、件数の場合とことなり、分散がきわめておおきい。ひとつの理由として、利用者規模の著しい差があるものと思われる。

むしろ、規模以外の要因も作用しているのは間違いない。北海道は、人口規模は南関東の諸県に並ぶにもかかわらず、分散が大きいからである。この背景には、建設業などの季節的なマッチングの占める割合が高く、かつその変動がとくに大きい点が考えられる。同地では、建設業での年間を通した雇用の確保が、労使と行政の間で重要なトピックとなつて久しい。

図5-2 就職率の分布（県別，1992.1 2004.12，カーネル分布推定）



5-4. 点推定と区間推定

政策論に登場する数字は一人歩きしやすいが、点としての予測値が的中するか否かにはそもそも意味はない。過去の出来事についても、幅をもった推定しかできない。将来の出来事の予測値がきわめて大きな幅をもつのは不可避である。

例 5-4-1. 最近、厚生労働省の研究班が国公立のがん専門病院の入院治療から5年後の生存率を比較した結果が公表された。入院時のがんの進行度や年齢など外的要因の影響を回帰分析で調整したうえで各病院の実績を比較している（「がん生存率，最大 28 ポイント差／専門病院，治療5年後／病状・年齢のぞき調査」『朝日新聞』2006年7月11日朝刊）。調整をしないまま原数値で比較評価をくだすよりは、情報の内実，構成がいくらか明らかになった点で前進である。しかし，病院規模などに由来する不確実性の幅は示されていない。それゆえ，たとえ点推定値で最下位と最高位の組織であっても，評価に順位で差をつけることの意義は疑わしい。

例 5-4-2. 日本政府は1960年の長期経済計画において，策定段階では10年間の平均成長率として6.5%から8.0%の間での成長を見込んでいた。しかし，いつしかその幅の中間の7.2%という数字が選ばれ，「国民所得倍増論」のスローガンとなり，予測の当否はもっぱらこれに基づき行なわれたという（西尾（2001），第15章）。7.2%というGNP平均成長率は，ちょうど1970年に「倍増」が達成される数値だった。いずれにせよ現実にはその予測値は大幅に外れ，GNP，国民所得ともに1960年代半ばには「倍増」していた（中村（2007），第5章）。

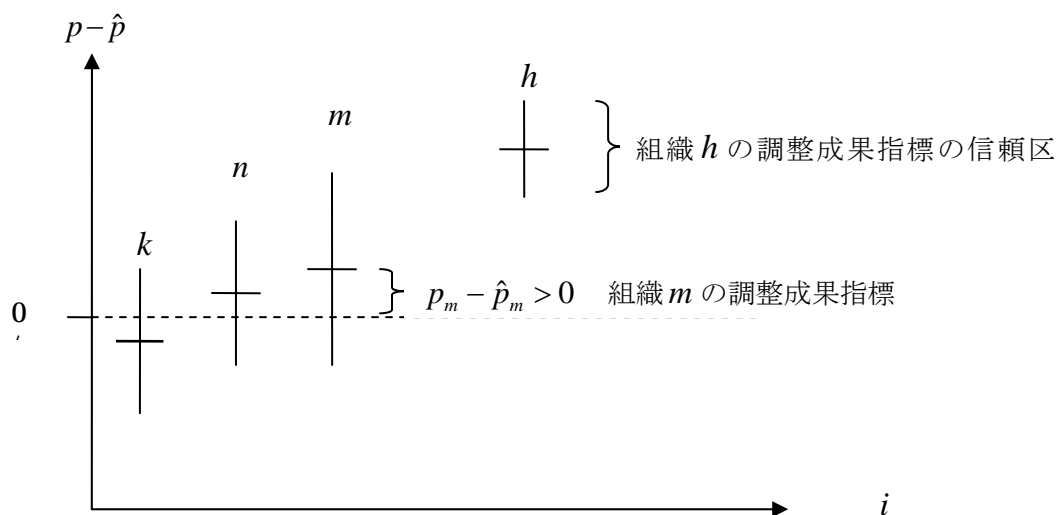
例 5-4-3. Manski (1995)によると，米国ワシントン DC の政治家も，世間の目を引く知見に注目し，留保なしに宣伝に利用する。その数字が一人歩きをする場合も多い。1960年代，ホワイトハウスで経済のアドバイザーとして経済指標の予測を担当した学者が，予測値の点推定にくわえて，それがとりうる不確実性の幅（range）を報告した。このときジョンソン大統領は「ranges（放牧場）は牛のものだ。数字はひとつ，よこしたまえ」（"Ranges are for cattle. Give me a number"）と言ったという（Manski (1995), p.8）。

統計データには常に不確実性の幅が不可避であることを自覚し，耐性を培うことが必要であろう。したがって，成果指標を比較するにあたっては，Goldstein and Spiegelhalter (1996) が学校評価の文脈で主張するように，成果の区間推定を明示することが望ましい²⁸。二つ

²⁸ ラフにいえば「95%信頼区間」とは，ある母集団の変数の平均値を推定するため，あ

の組織の成果の点推定値の差は、実はサンプルの小ささに由来する誤差の範囲内にあるのかもしれない。たとえば図 5-3 の概念図では、点推定は 1 位から 4 位まで序列を付ける（すべてのペアにおいて高低が決まる）。しかし、区間推定で重ならずに順序が付けられるペアは h と n 、および h と k のみである。

図 5-3 調整成果指標によるランキングと信頼区間(概念図)



そこで、前節で利用したHWデータにおいて、調整成果指標とその区間推定を試算してみる。「期間-県-地域ブロック」という3次元の階層（多次元）モデルを適用し、県と地域ブロックというふたつのレベルの指標を同時に推定する。なお、前節のモデルとは式は同一であるが、変数の分布に関する仮定と係数推定のアルゴリズムが異なるため、県の点推定値は異なる。結果を図 5-4、図 5-5 に示す。明らかに、区間推定を考慮すると、多くの地域の調整指標は互いに大差がなく、大小関係が決まるペアの数は著しく減っている。今回のような限られた標本にもとづく点推定（図中の黒点）から決めた順位は、不確実性を無視したきわめて危うい推論であることが了解できよう。²⁹

るサイズの標本を無作為に 100 回抽出し、毎回標本平均を計算した場合に、うち 95 回は母集団の平均値がその範囲に含まれている様な区間を意味する。標本サイズが増すと信頼区間は縮小する。区間の幅が大きい場合は、標本ごとに点推定値が大きく異なるため、ひとつの標本から強い結論を導くのは危険である。区間がゼロをまたいでいる場合は、平均値の正負すら判明しない。

²⁹ 個別効果の信頼区間を求めるには、個別効果 a_i と誤差項に平均ゼロの正規分布を仮定し、最尤法で推定した a_i の分散と平均ゼロの仮定を用いて事前確率分布を計算、それを変数 (p, x) の実現値にもとづく尤度関数とあわせて、 a_i の「更新」された事後確率分布を求める。この事後確率分布の平均と標準偏差（エンピリカル・ベイズ予測）から、各 a_i の信頼区間

図 5-4 県別調整成果指標と 95%信頼区間

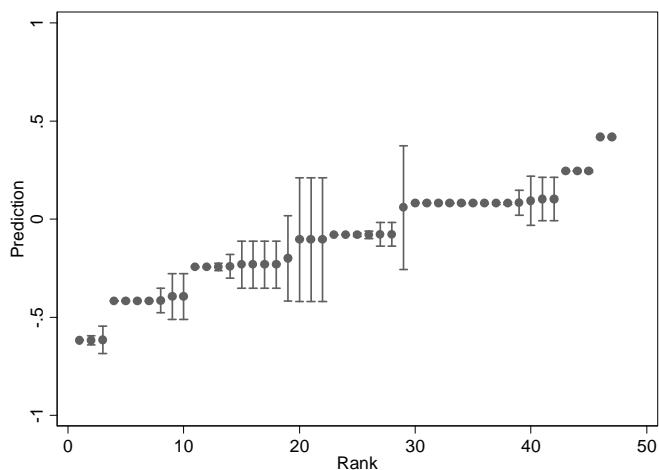
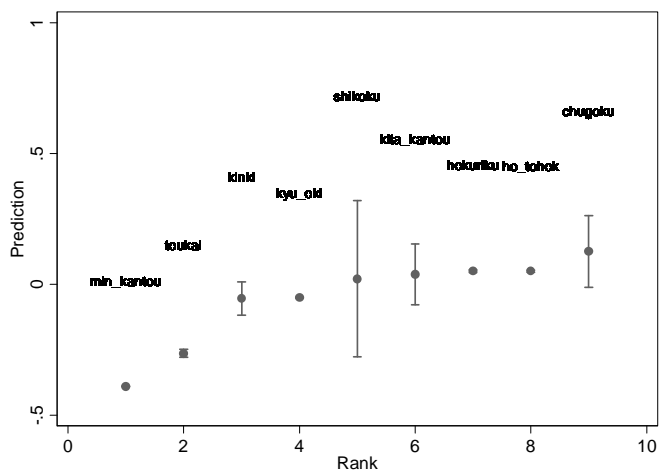


図 5-5 地域ブロック別調整成果指標と 95%信頼区間



注：データは第 4 節と同じ。

5-5. 不確実性への対策

各組織の成果指標の信頼性を増す一つの方法は、標本単位の集計である。たとえば、単年度で評価するのではなく、複数年度の成果指標の平均値をとる（時系列での集計）。これまで計算してきた県別パネルデータの調整指標は、期間平均である点で、この考えに沿って

が求まる。本文の二つの図は、この手順で作成されている（Rabe-Hesketh and Skrondal (2005) に平易な解説とプログラムがある。または本稿 A-2 節を参照）。

る。³⁰

また、HWをふくむ公的組織は、本体と附属施設とに分かれている場合がある。業務統計上の単位として本体と附属施設が分かれているが、現実の指揮系統は一体化しているならば、分析上も一体化することに問題は少なく、利用者のサイズを増すことができよう（クロスセクションでの集計）。逆に、分析対象となるカテゴリーを細分化すると、標本は縮小する。

例 5-5-1. 米国の教育論ではエスニック・マイノリティの成績改善が重視されているが、各学校におけるマイノリティの数が（文字通り）少ないために、信頼しうる評価が難しいと指摘されている（Kane and Staiger (2002)）。

例 5-5-2. 障害者というカテゴリーを取りあげて、職業紹介機関における就労マッチングを分析する際、障害者の標本サイズが小さいであろう。実現値の変動は大きく、成果指標としての信頼性はそれだけ低い。

なお、このほかに標準的な回帰分析の作業、たとえば新たな外的要因の導入、関数形の再検討、データの測定誤差の縮小は、モデル全体の適格性を改善し、調整成果指標の精度をシャープにする。すなわち、第 2-3 節で論じたフィードバックのプロセスは、成果の点推定のみならず、区間推定をも改善する。

6. 指標の選定，情報，多元性，エージェント，費用

ここまでは、成果指標の種類は所与だと仮定してきた。現実には、ある組織にはミッション、タスク、成果指標、成果の解釈、という各次元が存在し（詳細な定義は付論 A-1 を参照）、そこでいかなる指標がいかに選定されるか、すなわち選定のプロセス自体が、その後の成果水準を大きく左右する。たとえば、データ上カヴァーしえない要因が多い場合、少数の成果指標の導入と性急な外部からの評価、意味づけは、期待に反した反応を現場スタッフレベルにおいてひきおこす。その種の経験をふまえて、現在先進諸国では「評価制度の評価」

³⁰ 形式的には、これはある種の人事管理論や情報の経済理論と類比できる。エージェントの真の生産性をランダムなノイズから識別するために、短期的成果と報酬を連動させず、長期間パフォーマンスを観察し、真の生産性が正確に推定できるようになってから報酬（再就職先や退職金をふくむ）に格差をつけ始める、といった説明である。青木・奥野（1996）、Dixit(2002)を参照。

というメタ分析が始まった段階である³¹。日本は周回遅れという面があるが、見方を変えれば、海外の先行事例から教訓を得たうえで制度をデザインすることが可能である。³²

6-1. 背景

1990年代から、アングロ・サクソン諸国において広がった新公共経営 (NPM, New Public Management)の考え方は、政府を政治と行政に区分する従来の政官関係論のなかに、企画と実施という概念を導入し、「企画と実施を分離して実施を独立させることで、効率化をねらう」ものだとされる (飯尾 (2007, 158 頁))。さらに、競争市場による効率性確保といった伝統的な経済学のアナロジーを行政学に導入し、過去 20 年間の行政改革をバックアップした。³³

成果主義の例としては、米国の連邦政府が 1993 年以降、職業訓練施策などに目標管理制度を義務付けている (Government Performance and Results Act, GPRA)。各組織に広く裁量を認めるかわりに、結果責任と説明責任とを強化し、成果水準を次期予算額にリンクする。米国の公的職業訓練 (Job Training Partnership Act, JTPA) や特許庁は、この種の原理を組織単位で、また英国の特許庁は職員単位にまで適用した例であり、成果にもとづく組織 (Performance Based Organizations, PBOs) と呼ばれる (Courty and Marchke (2004))。この流れの背景にあるのは、成果の「効率的」生産のために制度を変えること、なかでも業

³¹ 行政学の概観に Bruijn (2007), Ferlie, Lynn Jr. and Pollitt (2005), 欧州の積極的雇用政策については De Koning (2007)。関連する経済分析に Blank(2000), Hart(2003)がある。

³² この点で、1990 年代の日本企業における成果主義や定量的目標管理の導入ブームとその後の反省 (ゆり戻し) と似た経緯を辿っている。ある大手商社は 2004 年末に製品の性能データ捏造が発覚したことを契機に、「業績目標の達成度合いという定量評価のみだった人事評価制度が業績偏重主義の温床になったとみて、働く姿勢など定性評価が八割を占める制度に大転換した。(中略) 利益なしには企業は存続しえないが、利益のために社会を裏切れば存在理由が根底から揺らぐ。それを防ぐにはトップ自らが内向きになりがちな組織を開き、風を通す努力を続けるしかない」(日本経済新聞 2007 年 12 月 26 日夕刊)。定量的成果主義の圧力が、情報の不完全性という条件を媒介としてひきおこす歪みは、官民「連携」の画期的な試みである介護保険制度や、一種の職業紹介機能をもつにいたった労働者派遣事業においても表面化した。公的部門における例は後述する。

³³ とくに英語圏の先進諸国では近年、職業紹介制度の規制改革や官民の「効率性」比較も注目を集めている (樋口・児玉・阿部 (2005), de Koning(2007), OECD, 2001, 2005)。ただし、しばしば官と民では利用者 (求人・求職) の母集団が異なる。公共職安はクライアントを選別しない・してはいけない (社会政策・労働法政策上の起源は濱口 (2004) が記述している)。民間業者は基本的に、収益という行動基準にしたがってクライアントへの対応を決定する。このような行動原理の違いに加えて、データの制約もあり、利用者の属性の差を十分に考慮に入れた官民比較は意外に困難である。他方、本稿の主題はいわば官官比較、そして官の目標管理制度である。

務の定量的情報を利用してスタッフのインプットを操作するという考えである。

一方、情報・制度・契約の経済理論は、完全競争市場という理論的ベンチマークから離脱し、経済的効率性を改善するための仕組みを探求してきた。とくに、情報の不完全性下にある経済主体間の誘因と制度設計の議論、「プリンシパル（依頼人）－エージェント（代理人）問題」という観点は、公的部門の検討にあたっても応用できる。企業部門では「株主－経営者－労働者」、公的部門では「国民・各種団体－政治家－行政官－ストリート・レベル職員（ケースワーカー、求人開拓員他）」、行政による民間委託や経営の監督は「行政－企業・NPO」という設定になるろうか。いずれにせよ、公的部門がかかわる場合は、純粋な企業部門の場合よりも理論的に複雑であり、未整理、未開拓の部分が大きい。以下では、内外の事例をふまえて主要な論点をスケッチしてみる。

6-2. 業務の多元性と対立性

「議論の本位を定めざれば、その利害得失を談ずべからず。」³⁴

一般に公的組織は、プリンシパル（依頼人）、ステイクホルダー（利害関係者）の数が多い。また、事前規制のため、それらを主体的に選別しにくい。結果として、異質な多数のミッションやタスクを遂行することになり、それが業務の効率性に影響する。³⁵

例 6-2-1. 多くの公立小中学校では、生徒の親、地域社会、教育委員会、首長、教員組合、文部科学省、政治家、経済界、メディアがそれぞれ異なる（かつしばしば対立する）要求をする。学力、進学実績、生活習慣、社会性、道德教育、公民教育、英語教育、学童保育機能、犯罪被害の回避、費用・人員の削減など。周知のとおり、数年単位で世論や政府方針が大きく変わることも珍しくない。また、これらの成果水準は、費用や利潤などと同等に定量的か

³⁴ 「城郭は、守る者のために利なれども、攻る者のためには害なり。敵の得は味方の失なり。往者の便利は来者の不便なり。故にこれらの利害得失を談ずるには、先ずそのためにする所を定め、守る者のためか、攻る者のためか、敵のためか、味方のためか、何れにてもその主とする所の本を定めざるべからず」（福澤（1995[1875]、16頁））。

³⁵ 本来は、これらの特徴こそが公的部門の存在事由であり、その境界を定義づける（べきである）、といえよう。行政学の教科書は、「公的」官僚制組織の特殊性として次の6点をあげている。①規模が大きい。②業務が多種多様である。③企業における収益のような統一的な評価尺度がない。④業務が独占的で競争に直面していない。⑤活動に多くの制約が定められ、柔軟性に欠ける。⑥活動を究極的に規律するのは政治のメカニズムである（西尾（2001）、第4章）。また、これらの特徴は相互に関連している。たとえば①、③、④が顕著であるほど、⑤（事前規制）と⑥（政治的調整による決定）が要請されるであろう。Wilson（2000）をも参照。

つ短期的に把握できるものばかりではない。測定できるものについても、効率性、平等性、公平性のどの次元を重視するかについて、ステイクホルダー間で合意が成立しにくい。さらに生徒は地域からの全入であるから、地域社会内部や各生徒・家庭の属性にばらつきが大きい場合は、学校職員のタスクの多元性と対立は深刻になる。³⁶

例 6-2-2. 一方、私立の小中学校は、入学時に一定の基準から選抜を行なうことで、生徒の均質性を確保し、対立的な多元性を回避している。タスクに多元性があるとしても、裁量によってある程度、相互補完的なものに制限しうる。その場合、金銭的ないし非金銭的なインセンティブを明確に設定することは合理的になる。とくに進学塾や大学受験予備校はミッションがシンプルであり、習熟度や目標校別に生徒をグループ化し、模擬試験や合格実績といった明快な短期指標をターゲットにしているため、この点は一層あてはまる。

このように、公的組織と民間組織とで、業務の多元性・対立性の程度に差がある場合、両者を単純な成果指標で比較することには合理性が乏しい。また、ミッションやタスクが対立的に多元的である組織に関しては、多角的な成果指標が考えられるべきである。³⁷

6-3. 指標の多角性と有用性のトレードオフ

一般に、最適化問題は1次元の指標を必要とする（いわゆる「最大多数の最大幸福」は論理的に不可能である）。したがって、複数の指標を同時に利用したい場合は、加重平均その他の方法で、ひとつの総合指標へ変換するのが一つの方法である。

しかし、複数の指標をひとつの総合指標に変換すると、重要な情報が失われる場合もある。

例 6-3-1. Hilbert(2007)は、スイスの公共職安の相対評価システムにおける総合指標を紹介している。合成された指標は、解釈が曖昧であること、各構成要素のウェイトのとり方に大きく左右されること、そしてウェイトの付け方自体に恣意性が多く、関係者間で合意が成立しにくいこと、などの問題点が指摘された。（このほかの点でも、評価手法とシステムの導入に

³⁶ 文部科学省（2006）の『義務教育諸学校における学校評価ガイドライン』が掲げるタスクのリストは長大である。付論 A-1 節を参照。

³⁷ なお、以上の議論のコロラリーとして、組織の設計および評価に関して示唆を得る。タスクやミッションは多元的であっても、相互に対立しない（補完的である）限りは、大きなコストは生じない。その場合は、成果指標も基本的に同じ方向に動くため、少数の指標を観測すれば済む。それゆえ、補完的なタスクは同一組織に集約し、代替的対立的なタスクは分離する、という組織の設計理念が導かれる。マイクロ経済学的な概説として Dixit (2002), Courty, Heinrich and Marchak(2004), 青木 (1995) を参照。

あたって問題点が多く提出され、中央政府が各地域と事前に十分な議論を行わなかったため、1990年代の改革は関係者間で衝突をうみ、けっきょく十分には実現しなかった。OECD(2001)をも参照。)

例 6-3-2. HWの業務評価者が、成果主義的な指標として就職率を採用する一方で、再就職後の定着率や雇用形態は採用しなかったとしよう。両指標が補完的であるよりは対立的である限りにおいて、これは各HWが定着率や雇用の安定性を重視する暇をなくし、短期的で質の低いマッチングを増やす誘因あるいはプレッシャーとなりうる（指標と行動にそのような関係が存在しないならば、そもそも成果指標の影響を考察する必要はない）。他方、両者を加重平均などによってひとつの総合指標にまとめると、就職率の高さが定着率の低さを隠してしまい、総体としてミスリーディングな指標となる可能性がある。この種のありうべき歪みを排除するには、時系列あるいはレーダーチャートの、就職率と定着率、求人充足率などの動きを、個別かつ同時に観察する必要がある。

職業紹介の例では、就職率、求人充足率、賃金変化率、求職期間、就職後定着率、生涯所得などを把握したい。

評価指標のこのような多角性は、情報を処理する能力に余裕がある場合は有用である。しかし現実には情報処理のコストは大きいため、多角性と有用性にはトレードオフがある（費用については6-8節で再論する）。ミッション、タスク、指標が多角的になると、巨大な組み合わせの集合が生まれ、因果関係の推定はたちまち人間の認知能力をこえる。また、それがコンピュータの処理能力の範囲内であったとしても、情報の処理の内実は、当事者たちにとってブラックボックスとなる。しかし、ブラックボックスが出力する規範的な評価に、どれほどの説得力があるだろうか。

比較的にバランスのとれた一元的指標の例として、OECD(2005)が職業紹介の分野で提案する「節約された求職者手当額+ 再就職後賃金からの雇用保険料収入や税収」がある。実際に有用な精度でこの指標を利用するには、おそらく税・保険行政のマイクロデータによる中長期間の追跡調査が必要であり、その実施コストは大きい。とはいえ概念上のベンチマークとしては、明快さと包括性をそなえた指標である。³⁸

³⁸ OECD(2005)は、 $B + tW$ を提唱している。ここで B は施策対象者の再就職により節約された各種手当の額、 t は税率、 W は再就職後の雇用収入。失業だけでなく、雇用、収入の評価も可能であり、政府の財政バランス改善の観点も含まれている。（予算規模の維持拡大を目的とする行政モデルとは対極的である。その種の対比の実例として、米国の職業訓練機関を分析した Heckman et al. (2002)がある）。

6-4. 情報の非対称性と評価制度とのゲーム

業績評価のシステムが、各組織の成員に対してどのような行動の誘因を提供するか、逆効果が生じないか、のミクロ的な検証が必要である。実際に事例を吟味すると、公的機関の定量的業績評価は、OECD や英・米・豪政府が（おそらく暗黙のうちに完全情報の完全競争市場をイメージしつつ）主張してきたほどには単純でも容易でもない。

成果指標の観察期間が短期的であれば、短期間で常に成果を出すような圧力がかかる。それゆえ、効果が現れるまでにかかなりの費用と時間がかかる種類の業務や、不確実性をともなう根本的な組織改革は実行しにくくなる。また、前期の実績に自動的に上乘せして次期の目標値が設定される場合は、スタッフが事前に「がんばりすぎない」ようになるとしても無理はない。³⁹

さらに、会計上の費用は把握できるが、サービスの品質は測定困難である場合は、費用削減（効率化）の背後で品質の低下が生じるであろう。これらの問題は、利用者の離脱・退出というオプションが存在しない独占的な供給機関の場合は、とくに感知されがたく、深刻な帰結をもたらす。それゆえに、利用者および成員の「発言」オプションによるチェックと修正というフィードバックの仕組みが一層重要になる。

例 6-4-1. 米国では、職業訓練機関の成果指標（就職と賃金）に次年度の予算をリンクさせた成果主義的な試みがある（JTPA）。観測可能な少数の次元について、強いインセンティブを設計したものといえる。その結果、評価制度に対応して、明らかに戦略的な行動が観察された。本来のターゲットであるべき生涯所得などの長期的利益は、観測が困難であるため、軽視される。たとえば、訓練機関は当初、各受講者の訓練終了後 3 ヶ月以内の任意の時点における雇用状況について政府に報告する規定であった。すると、各訓練機関は就職が決まった受講者についてはその時点で報告し（その後の就業継続には関知せず）、また、就職先が決まって

³⁹ これらの点を調整成果指標の枠組みに統合することは今後の課題である。たとえば、近年は求人・求職の件数にも目標管理が導入されており、求人と求職が外生的なインプットだという第 4 節の想定には疑問の余地がある。管理当局が設定したノルマのデータがあるならば、それを新たな説明変数として吟味するべきである。もしも今期の成果のノルマが、過去の実績値の関数であり、かつ（今期の求人・求職件数に関わらず）今期の実績値を左右している場合、第 4 節の表記に沿うならば、 $m_{i,t} = b + \delta m_{i,t-1} + \beta_u u_{i,t} + \beta_v v_{i,t} + a_i + w_t + z_{i,t}$ において、説明変数 $m_{i,t-1}$ と観測不可能な a_i に相関が生じる（前期に $m_{i,t-1}$ が被説明変数として a_i の影響を受けている）。それゆえ、仮に今期の求人・求職と a_i とが無相関であっても、これまで見た手法では係数は正確に推定できず、ダイナミック・パネルの設定になる。（Arellano (2003), Cameron and Trivedi (2005, ch.22), Wooldridge (2001, ch.10)）。

いない受講者については期限まで報告しない、という傾向が発生した。これを受けて、訓練終了後3ヶ月の時点の状況についても報告するよう、政府が規定を改めたところ、訓練機関は、3ヶ月がすぎる直前まで（のみ）元受講者の状況を把握し、非就労者に様々な就労支援を行なうようになったという（Dixit (2002), Heckman et al. (2002), Courty and Marschke (2004) 参照）。

もちろん、評価システムの改革が、誘因の変化を通して行動や成果水準を変化させること自体は問題ではなく、その変化が本来のミッションにてらして望ましいかどうかの問題である。まず状況の変化を実証的に把握し、次にそれをミッションにてらして規範的に評価する、という二つの段階を区別しつつ両立させることが必要となる。

例 6-4-2. 期間内の実績件数の最大化を目指す組織があるとすれば、測定されない成果の質、安定性、将来性は軽視されるであろう。この点は、いわゆる市場化テストのように、官民の「効率」を一定期間内の成果指標で比較するといった統制実験的な試みがなされる場合に、重大な問題を提起する。通常の公的機関（HW、教育機関その他）は、正しくも、短期的な試験点数や就職件数の最大化のみを行動原則としてはいない。一方、通常の企業組織は、中長期的な収益最大化を行動原則としている。すると、通常的环境と行動原則から乖離した設定に官民双方を置き、一定期間、両者を観測して効率性を比較することからえられる知見は、その設定内では明快かつ正確であっても、その他の設定には一般化できない（予想が外れる）可能性が高い。⁴⁰

成果主義の導入は、行動を変えることで成果水準を変えることもあれば、単純に、現場でのデータ作成時に不正をひきおこす可能性もある。

⁴⁰ このように、ある知見が他の時と場所では成立しない（一般化できない）状況を、政策評価論の用語では、分析結果に内的妥当性はあるが外的妥当性がない、という。この問題は、統制実験によって発展した科学技術においてより鮮明に現れている。「もともと近代自然科学とりわけ物理学や化学は法則の確立を目的としていたが、しかしその法則というのは、直接的な応用とは無関係な天文学をのぞいては、まわりの世界から切り離され純化された小世界、すなわち環境との相互作用を極小にするように制御された自然の小部分のみに着目し、そのなかで人為的・強制的に創出された現象によってはじめて認められるものである。自然科学はそのような法則の体系として存在し、実際にはかなり限られた問題にたいしてのみ答えてきたのであるが、そのような科学にもとづく技術が、生産の大規模化にむけて野放図に拡大されれば、実験室規模では無視することの許された効果や予測されなかった事態が顕在化するの避けられない。そしてそのような効果や事態は往々にしてネガティブな結果をもたらすことになる。」（山本（2007））。

例： 民間保険会社の出身者が公的な年金保険庁の長官に就任し、保険料納付率向上のため、きびしく数値目標を掲げたところ、成果指標の操作（納付率を上げるため分母の対象者を減らす、すなわち不正に納付を免除した）が発生した事例は、われわれの記憶に新しい。これは近年の民間保険会社の「未払い問題」とは方向が逆である。保険会社にとっては、保険料納付と保険金未払いが収益増に貢献する。この差は官民の行動原則の差を明快に反映している。しかし、性急な目標管理主義が背景にある点は共通している。

例： 英国でサッチャー政権時から継続している教育改革については、多くの知見が得られている。政府が共通テストの成績の平均値によって学校のランキングを作成し、公表した結果、多くの学校は平均点上昇に貢献すると思われる一部の生徒にのみ指導のエネルギーを注ぐようになった。成績の悪い生徒をリストから外す例さえあった。こうして教育改革の結果、全体的な平均値は上昇したかに見えるが、同時に教育の格差が深まった、という見方も根強い⁴¹。その結果、2007年以降、政府は単純な結果のランクの発表を中止し、平均値の信頼区間の明示、外的環境要因による調整などの修正を行なっている⁴²。近年、日本の政府は英国の教育改革に注目してきたが、かの国の経験したこうした負の側面にも注意を払う必要がある（この点、政権与党内部でも、慎重な異論が出ているのは興味深い。河野・後藤田・上野・篠田・橋本・山内（2006）を参照）。⁴³

6-5. 切磋琢磨と非協調

業績の組織間比較は、組織間の切磋琢磨を促進する効果がある一方で、協力意識を阻害する可能性もある（Myers and Goldstein, 1997）。他の組織に対してもっとも友好的で協力的であった組織が、業績評価に際してもっとも損をする事態が生じかねない。その結果、優れた事例についての相互情報提供や、緊急時の協力関係が失われたり、足を引っ張り合う状況

⁴¹ 英国の高級紙には、教育専門家の批判や特集報道が掲載されない日はないかのようである。阿部（2006）、山口（2005）、Myers and Goldstein（1997）をも参照。

⁴² 2007年、英国ガーディアン紙は、教員や児童への悪影響を憂慮して、全国の小学校の成績をランキング形式で報道することを止めると宣言した。ただし、アルファベット順、および地域別の結果は報道している（Cf., “Why ranking has reached the end of the road” *EducationGuardian.co.uk*, December 6, 2007）。一方、タイムズ紙は現在も順位形式での公表を続けている（“The complete ranked schools's league tables based on GCSE and A-level results 2007,” *TimesOnline*, February 7, 2008）。

⁴³ なお、本節の議論をうらがえすならば、定量的な業績管理と評価は、業務の目的、内容、環境、結果、影響が一元的・安定的であり短期的に観察可能な分野・対象においては容易であり有効である。その種の業務は、分野をとわずほとんどの組織に存在するであろう。定量的管理を一部に適用することで、組織全体に規律の感覚を醸成する効果も期待できる（Brujin（2007））。

が生じるとすれば、社会総体として損失は大きい。「自由な競争の限界」を示す例として知られるゲーム理論の「囚人のジレンマ」を想起させる状況である。

例 6-5-1. 豪州では、2000 年前後に公共職業紹介システムが、競争的な民間委託（実績の格付け評価と入札にもとづく）に移行した。すると、規定に反して、各紹介組織が質の高い求人情報を秘匿し、他の組織に提供しない傾向が現れた。このため、雇用サービス関連の NPO が構成する団体 Jobs Australia(2006)は、個別組織のベスト・プラクティスを調査し紹介する仕組みの導入を提唱している。しかし、相対評価が入札資格を左右する限り、個別組織は自らの成功の秘訣をあえて公開する誘因をもたない。日本でも、HW と異なり、民間紹介会社は「非公開求人」をむしろセールス・ポイントにしている。⁴⁴

6-6. 相対評価と絶対評価

また、調整成果指標による評価は一種の相対評価である。相対評価で最下位であっても、何らかの絶対評価の観点からすれば十分な業績を残している可能性はある。下位であることは、社会的に非効率で不要な組織であることを必ずしも意味しない（たとえば、どのようなクライアント（求職者、求人企業）であっても、コストの観点にもとづいて選別せずに、必ず対応する、といった社会的に有用なセーフティネットの機能を果たしているかもしれない。逆にいえば、相対評価で上位であるオフィスが、絶対評価の観点で改善の余地がない、とも確言できない）。一見した「非効率」からすぐに存在を否定するのではなく、では具体的によりよい供給者・供給方法はなぜ見当たらないのか、またいかにして成立するか、と問うてみる価値はあろう（Dixit(2002), Wilson(2000)）。その際、他地域や海外での事例研究を参照するのも有用である。

6-7. アカウンタビリティ

官民を問わず、具体的な文脈に即して優れた評価システムを構築するのは容易ではない。また、環境が変化する限り、業績評価システム自体に試行錯誤があるのは当然である（Courty and Marschke (2003)は評価システムの変化のプロセスを理論化している。行政学では Bruijnen (2007, ch.3)を参照）。とはいえ、大幅な振幅にともなう甚大な調整コストは抑

⁴⁴ なお、この背後にはまた、求人情報を公開すると、求人企業に直接アクセスする求職者が現われることを紹介組織が懸念している面があるかもしれない。実際、HWでは、所内に求人情報の検索機を導入して以降、相談窓口を経由せずに企業に直接連絡をとる求職者（「直行組」）が増大したといわれている。その場合、現在の HW の業務統計上は、マッチングの件数としてカウントされない。

えたい。先行事例を見ると、その際に鍵となるのは、小規模でのシステム試行や、ステイクホルダーへのアカウントビリティのようである。

例 6-6-1. 豪州政府は 90 年代後半から、職業紹介分野に改革を行い、営利・非営利組織への一定の業務委託を導入した。委託期間に連動して、各組織の定量的な業績評価を行い、結果は一般にも公開している。複数の指標を標準化した総合的な成果指標について、各種外的要因の影響を回帰分析で調整している模様である。この定期的な格付けシステム (star rating とよばれる) は、各組織のイメージに影響するうえに、次期の入札の参加条件を自動的に左右する相対評価である (評価が下位 10 数%であった組織は、次期に参加不可にするなど)。紹介組織にとっては死活問題である。しかし、その評価手法の内実がきわめて不透明であり、予測はもちろん事後的な理解も難しい、と当初から現在に至るまで各紹介組織や研究者から批判があがっている。ある就職支援企業の幹部は、経営戦略をたてるうえで、現行の格付けシステム自体が迷惑な不確実性の源であり、プロセスの透明化を望む、と語っている (豪州の企業に対する筆者のヒアリング (2006 年 10 月) にもとづく)。じっさい、紹介組織を顧客として、「格付けシステムの理解を助け、次回の格付けを予測する」ことをサービスとする企業が現れている。⁴⁵あるシンクタンクは、豪州政府の委託で格付けシステムの評価を行っているが (Access Economics (2002))、その報告書にはデータと分析手法の詳細が開示されていない。なお、「独立評価委員会」は、現行システム全体を批判的に検討し改善を提言している (Productivity Commission (2002))。豪州では 2007 年に保守連合の長期政権が終わり、労働党が与党となったこともあり、官民連携のあり方に変化が訪れるかもしれない。なお、オランダは豪州型に似た職業紹介の民間委託システムを導入したが、振幅の大きい試行錯誤を続けている (De Konig (2007), Struyven (2007), OECD(2001)参照)。またスイス政府は、公共職安の比較評価手法と管理システムを開発したが、導入が性急であったこともあり、分権的伝統をもつ各自治体の反対が強く、けっきょく実現には至らなかった。Hilbert (2007)は、スイス政府の構想した分析手法を、代わりにドイツの職安データにあてはめて、その性能を吟味している。

例 6-6-2. 企業部門は元来、行政のようなアカウントビリティを課されていないため、官民連携がアカウントビリティを低下させるケースが生じている。近年、中東・アフリカ・中南米に大規模に介入している軍事請負会社は、欧米の正規の軍隊と補完的な任務を遂行している。企業であるため、正規軍とは異なり委託元である欧米諸国の行政法や戦時国際法の統制は受けず、議会への説明責任もなく、活動実態は不明朗である。まさにその点に、委託元政

⁴⁵ <http://www.jnsolutions.com.au/>

府にとっての利用価値がある。もっとも、しばしば捕虜や現地民間人に対する重大な人権侵害が判明した結果、最近では委託元政府の責任が国際的に問題視されつつある (Singer (2005))。

例 6-6-3. 英国では、初等中等教育の評価手法は、政府のウェブサイトである程度公開されるようになった。いまだにその複雑さや妥当性に対する批判は多いとはいえ、ブラックボックスよりはるかに望ましい。このような改革の背景には、政府から最優秀と評価された学校の校長が「我々は年々厳格になる政府の管理介入に批判的であって、経験とコモン・センスにもとづく独自のやり方で学校を運営している」と声明を出して注目を集めるなど、自由で多角的な討議が続いてきたことが作用している (“Top heads say scrap school tables” (BBC News, 2005/12/02)).⁴⁶

以上、公的部門の業績評価にあたって留意すべきポイントをいくつかあげてきた。性急で一方的な評価とラベリングは、組織の成員や利用者に社会的スティグマを与えるリスクがある。それは中長期的にはスタッフの質の低下にもつながる。たとえば、共通テストの平均点にもとづく評価で「最低ランク」の学校の生徒や教職員は、社会的な負のイメージを甘受すべきなのか。本稿の観点からするならば、各学校をテストの成績ですぐに序列づけるのはナンセンスに近い。成績を左右する諸要因、たとえば入学時の成績、家庭の経済状況、地域の治安や文化的環境、生徒と教職員の人数比、学校の規模などについて学校間でばらつきが存在する可能性は高い以上は、これら環境要因の影響を調整したうえで成果を比べるべきである。さらに、共通テストの点数という指標が、学校本来のミッションを十分に代表しているという暗黙の前提についても検討の必要がある (第 6-2 節および付論 A-1 を参照)。

このように、評価の公開方法と当事者の反論可能性には注意する必要がある。とはいえ、民主的な社会において、強制的に徴収された公金を利用する独占的な供給機関である以上は、何らかのかたちで自己の存在事由を外部に立証する責務があろう。それゆえ、業務の基礎情報の公開と、それにもとづいた関係者・第三者による業績評価 (政策情報) の創出、という方向性は、基本的には促進されるべきであろうし、現にそうなりつつある。

すると重要になるのは、第 2-3 節で触れたように、評価する側、される側を含めて、誰もが評価に対する異論を提出しうる言論の公共的空間をつくり、討議の継続 (再審の可能性) を保証することである (Ingram and Schneider (2006), Myers and Goldstein (1997), Bruijn (2007, ch.7), Stiglitz (1999), 大森 (2002, 259 頁) を参照)。⁴⁷

⁴⁶ <http://news.bbc.co.uk/1/hi/education/4491680.stm>

⁴⁷ 「合理的選択を行なう利己的経済人」を想定する伝統的な経済理論の観点からしても、

6-8. 政策の費用と便益

政策当局が利用しうる資源は有限であるから、各施策がもたらす費用（人件費、賃料、その他）も考慮に入れて、純便益としての成果指標を考慮することが望ましい。

費用も成果も、金額単位のデータが入手できる場合は、純便益の計算は明快である。一方、費用と成果指標とで単位が異なる場合は、「成果／費用」すなわち「単位金額あたりの成果指標」を新たな成果指標と考えることができる。

ただし、費用の定義自体に選択肢がある。まず、会計的な支出概念による費用は、もっとも明快で入手しやすいデータであろう。これに対して便益は、定義自体に議論の余地が多い。（国民のうちの誰に対する、どの期間の、どのような尺度での影響を考慮すべきか？公的な供給財にはしばしば外部性があるため、便益は一層測定しがたい。）

便益に比べると、支出額は一見明快であるため、公的機関の「効率性」をめぐる議論では、予算や支出額自体が支配的な成果指標となりがちである。すなわち、外部は組織の支出額の削減を要請し、組織の内部はその維持拡大を迫及する傾向がある。⁴⁸

公的部門におけるより基本的で適切な問題設定は、「所与の予算額で国民の便益を最大にする政策は何か」である（「何もしない」を含む）。そこでは機会費用の概念が必要となる。「機会費用」は、「他の用途で得られたであろう最大の便益」を意味する。それは会計上の費用と異なって、反事実的な想定であり、実質的には便益の推定と同様、正確な把握は難し

誰もが財・サービス・制度の正確な情報にアクセスできることは社会的に望ましい。それは完全競争市場均衡の効率性（厚生経済学の基本定理）の必要条件である。情報の過少供給は「市場の失敗」の典型例であり、市場メカニズムの性能はそれだけ劣化する。また、情報の不完全性をそのままにして、他の面で市場の競争性を高める試みが、社会的な厚生を改善する保証はない（旧共産圏の「移行経済」の惨状を想起されたい。セカンド・ベストの環境における漸進的政策の経済理論は、福島（1993）を参照）。一度「生産」され供給された情報は、公共財の条件（消費における非排除性・非競合性、追加的供給の費用極小）を満たし、外部性をもつ。しかし、個別組織や経済主体は、自己の便益の減少につながるかぎり、真の私的情報を開示する誘因をもたない。情報という公共財の過少供給は、政策形成にも反映する。情報と公共的討議へアクセスする資源自体が均等に分配されていない社会では、特殊利害を共有する少数者集団が政策形成に過剰な影響をおよぼすにもかかわらず、政策の費用負担は諸個人に拡散する。相互に分断され集積の利益を享受しない個別主体にとっては、情報の収集分析や意見伝達のコストは大きい（これは投票行動にも現れる。Dixit（2002））。その結果、多様な観点という財が過少供給され、政策決定が十分な多角的検討をへず、脆弱な根拠にもとづく危ういものとなる。多様な観点が現れる空間、開かれた討議のアリーナという「メタ公共財」の安定的な供給を制度的に担保することが社会的に望ましい（cf.スティグリッツ（1999）、Stiglitz（1999））。

⁴⁸ 一方、企業部門には、市場が競争的である限りで効率化（支出最小化）へのインセンティブがある。もっとも、それは自己目的ではなく、収益という明快な指標の最大化と表裏一体である。

い。しかし、概念上のベンチマークとしては有用である。

6-9. 評価の費用と公共化による費用削減

プラクティカルな観点からすると、業績評価という作業自体の（機会）費用について、これまで以上に留意するべきだと思われる。すなわち、はたして業績評価に投入する膨大な資金と人的資源に見合う結果が得られているのか。

現実には、行政のスタッフは通常業務に加えて業績管理と評価という負担に見舞われ、時間もスキルも不十分なままに内実に乏しい評価書や資料を作成提出し、本来の業務にも支障が出る、という悪循環の例が目につく（山谷（2006）、Bruijn(2007)）。そして、組織のミッション、タスク、指標が多角的で対立的であるほど、この種のコストは加速的に大きくなる（補論A-1参照）。たとえば英国でも日本でも、90年代から初等教育の現場は過労による心身の疾患が急増している。英国では教員の給与を上げたが、過労による退職者が増え、とくに校長など管理職が危機的に供給不足となった。医療機関でも、同様の事例が指摘されている。ともに成果のランキング表（league table とよばれる）を導入し、政府による目標管理と介入を強化した分野である（6-4節の文献を参照）。このような実情を見る限り、現状では、費用対効果には懐疑的にならざるをえない。本来のミッション遂行を妨害するような、本末転倒、内容空疎な業績評価は、停止するのがもちろん正しい。

この文脈で注目したいのは、評価の社会的な費用は、データの整備・開示体制と密接な関係があるらしいという経験則である。たとえば、周知の通り、日本のマクロ経済政策については、外国人研究者を含めて、きわめて活発で多角的な議論が行われている。その大きな要因は、マクロ統計は内外を問わずほぼ誰にでもアクセスが可能であり、また発言機会も開かれていることにある。

これに対して、労働、医療、教育などの個別分野では、基礎的な行政の統計へのアクセスが閉ざされている。諸外国の先進事例にならい、マイクロデータの継続的収集と、必要な匿名化を施した上での開示を進めるならば、第三者をふくめた研究と論争が活発となり、質の低い意見は淘汰され、中長期的には政策研究の質は格段に向上するものと思われる。いわば「知の公共化と漸次的改善」（山本（2007））である。そこでは、当事者組織は素材となる業務上の基礎的な情報（HWにおける一般職業紹介状況など）を広く収集し、内外に提供することが重要な責務となる。そして、政策的な争点と解決策の提案につながりうる情報（業務の社会的な付加価値の推定を含む政策評価など）は、当事者、関係者、第三者が公共的な討議の

空間で供給することが、評価という作業の機会費用削減の観点からも合理的である。⁴⁹

6-10 成果指標のための倫理規定

本稿全体の議論を要約する意味をかねて、英国の学校間比較の研究者が提案している「成果指標のための倫理規定」7箇条を紹介する（Myers and Goldstein (1997)）。教育以外の分野における業績評価のあるべき姿を考える上でも、簡明な叩き台となろう。

1. 不当な損害（Unwarranted harm） 成果指標による評価の公表が当事者に不当な損害を与えないこと。
2. 情報への権利（The right to information） 十分に正確な評価が実施された場合は、公表されるべきであること。
3. 文脈の考慮（Contextualization） 各組織にとって外生的な要因が成果を左右していると思われる場合は、その要因の影響を調整すること。また、評価の公表にあたっては、外的要因の調整方法を明示すること。
4. 不確実性の推定（Uncertainty estimation） いかなる成果指標も、標本分散や分析手法に由来する統計的な不確実性を免れず、成果の水準は厳密には確定できない。信頼区間などで、不確実性の程度を示すべきである。
5. 複数の指標（Multiple indicators） 複数の指標を採用すること。単一の（あるいは少数の）指標では、特定の局面だけが重視され、組織運営にゆがみが生じる可能性が高い。
6. 制度化された対応（Institutional response） 評価を実施した研究機関は、被評価者や第三者によるチェックや再評価を可能とするために、匿名化のうえでデータを開示すること。
7. 説明責任（The responsibility of agencies publishing information） 評価を実施した研究機関は、内容を普及する責任を負う。データ収集と分析の方法についても説明する。

そして本稿の観点からは、第8項として、「評価自体の機会費用」と「基礎的情報と評価の公共化による費用の削減」への留意をここに追加すべきかと思われる。

⁴⁹ 70年代以降の自治体経営改革の経験を要約した松下（1999）の定義では、自治体の発信しうる情報は、伝統的な事後的・一方的「広報情報」と、専門性と経験をもった市民が政策形成的議論に貢献するために有用な「政策情報」に大別される。後者はさらに「争点情報」「基礎情報」「専門情報」からなる。

7. 結語

本稿は、関連諸分野を横断的に展望しつつ、公的機関の定量的な業績管理および評価の方法論を検討し、HWの集計データで応用例を示した。

パフォーマンスを示す何らかの成果指標を、組織（オフィス）間で比較するという問題設定では、外的要因の影響を考慮して調整することが望ましい。そのうえで、なお残る調整成果指標のばらつきについては、仮説生成の参考情報としつつ、関係者、観察者で探索的な対話、自由な討議を行うならば、各オフィスのベスト・プラクティスや構造的障害が発見されるであろう。

技術的には、それは回帰分析でいう残差分析の応用である。成果指標を被説明変数とし、各組織の外的環境や業務体制を説明変数として、モデルを推定する。モデルから、各組織ごとの成果の実績値と理論値の乖離を計算し、それを調整済み成果指標として利用する。本稿はこの考えを拡張し、HWの県別パネルデータを用いて、中長期的に安定的な調整指標の分布を推定した。外的規定要因として調整に利用したのは各期の労働市場の需給条件（求人、求職）、そしてマクロの景気である。

調整の影響は大きいことが判明した。また、大都市圏と非大都市圏とでは、システムティックに後者の調整指標が大きい。解釈としては、地方における代替的なマッチングルートとの非競合性、社会経済の安定性に由来する情報の質の高さ、タスクやクライアントの均質性に由来する業務コストの低さ、などの仮説を生成した。関係者からの定性的なフィードバックを受けて次回の分析を行い、新たな調整変数を導入することで、漸進的に正確さと関係者の納得度が高まることが期待できる。調整成果指標の形式は、そのような発見的で探索的な作業に適合的である。第5節では、成果指標の不確実性を考慮して、点推定とともに区間推定を明示することの意義を計算例とともに示した。端的に言って、外的要因と不確実性を調整するか否かで、成果水準の解釈はまったく変わる。

6節で見たように、収益という基本的な尺度がある企業部門に比べると、公的部門はしばしばその業務の目的、種類、影響と説明責任が多面的・多義的・長期的であり、事前にも事後にも明確なプライオリティがつけにくい（あるいは、それらの特徴がそもそも公的部門の境界を設定するべきであるともいえよう）。その種の組織の業績管理・評価は、多面的・定性的・暫定的・継続的な性質をもつ必要がある。それは内外のステイクホルダーや観察者からの意見聴取など定性的な情報収集と、定量的なデータ分析との往復、という形をとろう。組織内に意見のアリーナ、討議の空間という公共財が安定供給されることは、意思決定者にとっても、情報の不完全性への対策、経営の透明化、現場スタッフの積極的参加と忠誠をも

たらず機能となりうる。

また、安易な定量的業務評価の弊害を防ぎつつ、アカウントビリティを推進すること自体は、官民に共通する課題である。元来、多義的・多元的な環境下にある限りにおいて、民間部門と行政部門の業務には共通点があるし、諸外国で行政が企業・NPOとの新しいパートナーシップを模索していることは、第6節で触れたとおりである

むろん、公的部門の多元的なステイクホルダーを含んだ形での情報整備、討議、交渉、合意のプロセスを有意味なものにするには、一定の修練と時間が必要であろう⁵⁰。調整成果指標の考え方は、そのプロセスのなかでもシグナルの機能を果たしうるように思われる。

いずれにせよ、今後、職業紹介のみならず、教育、医療その他の分野において、現場スタッフ、利用者、政策当局、研究者間の定性的・定量的なフィードバックが活発となり、公共的な討議としての業績評価が発展することを期待したい。

8. 文献

青木昌彦（1995）『経済システムの進化と多元性：比較制度分析序説』東洋経済新報社

青木昌彦・奥野正寛編著（1996）『経済システムの比較制度分析』東京大学出版会

阿部菜穂子（2006）「安倍政権は、問題の多いイギリス「教育改革」に追随するのか」『世界』11月号、岩波書店

飯尾潤（2007）『日本の統治構造』中公新書

上野 有子、神林 龍（2005）「求人規模と提示賃金の関係 ～日本におけるサーチフリクションの源泉～」ESRI Discussion Paper, No.156

大森彌（2006）『官のシステム』東京大学出版会

神林 龍・上野 有子（2006）「求人集中度とマッチングの効率性」ESRI Discussion Paper, No.160.

カント（2006 [1784]）「啓蒙とは何か」『永遠平和のために／啓蒙とは何か 他3編』中山元訳 光文社古典新訳文庫

河野太郎、後藤田正純、上野賢一郎、篠田陽介、橋本岳、山内康一（マネジメントの観点からの教育再生研究会）「教育改革の改革を一教育再生会議への7つの疑問」『世界』2007年6月号 http://www.kou1.info/pdf/education_paper.pdf

厚生労働省職業安定局（2006）「平成18年度の雇用保険三事業による事業の目標設定につ

⁵⁰ このためのファシリテーター育成や特別予算の確保などの取り組み事例については、Suskind (2006)の文献を参照。

いて」平成18年10月改訂版

<http://www.mhlw.go.jp/general/seido/anteikyoku/koyou06/dl/01.pdf>

齋藤純一（2000）『公共性：思考のフロンティア』岩波書店

櫻井稔（2006）『内部告発と公益通報』中公新書

佐々木勝（2004）「年齢階級間ミスマッチによるUV曲線のシフト変化と失業率」『日本労働研究雑誌』4月号

佐々木勝（2007）「ハローワークの窓口紹介業務とマッチングの効率性」『日本労働研究雑誌』10月号

周燕飛（2007a）「公的職業紹介におけるマッチング効率の地方・都市間格差とそのゆくえ」『ビジネス・レーバー・トレンド』12月号，（独）労働政策研究・研修機構

<http://www.jil.go.jp/kokunai/blt/bn/2007-12/index.html>

周燕飛（2007b）「公的職業紹介におけるマッチング効率の地域間格差」『地域雇用創出の潮流 — 統計分析と実態調査から見えてくる地域の実態』第6章，（独）労働政策研究・研修機構 <http://www.jil.go.jp/institute/project/h15-18/01/index.htm>

労働政策研究・研修機構（2006）「マッチング指標」『業務統計を活用した新規指標 2006』JILPT 調査シリーズ No.18, <http://www.jil.go.jp/kokunai/statistics/gyomu/>

スティグリッツ，J.（1999）「国家の役割の再定義」青木昌彦・奥野正寛・岡崎哲二編著『市場の役割 国家の役割』東洋経済新報社，第2章

高間邦男（2005）『学習する組織：現場に変化のタネをまく』光文社新書

田中勝人（2002）『経済統計：第2版』岩波書店

中馬宏之（2002）「労働経済学的なアプローチの限界と課題」『日本労働研究雑誌』No.500

中村二郎（2002）「転職システムとしての公的職業紹介機能」『日本労働研究雑誌』No.506，8月号

中村隆英（2007）『昭和経済史』岩波現代文庫

西尾勝（2001）『行政学』新版，有斐閣

ハーバード・ビジネス・レビュー編（2001）『業績評価マネジメント』ダイヤモンド社

ハーシュマン，A.（2005[1970]）『離脱・発言・忠誠：企業，組織，国家における衰退への反応』矢野修一訳，ミネルヴァ書房

濱口桂一郎（2004）『労働法政策』ミネルヴァ書房

樋口美雄・児玉俊洋・阿部正浩編著（2005）『労働市場設計の経済分析』東洋経済新報社

福澤諭吉（1995 [1875]）『文明論之概略』松沢弘陽校注，岩波文庫

- 福島隆司 (1993) 『漸進的政策勧告の経済学：租税と関税改革の厚生経済的評価』 創文社
- 松下圭一 (1999) 『自治体は変わるか』 岩波新書
- 文部科学省 (2006) 「義務教育諸学校における学校評価ガイドライン」
http://www.mext.go.jp/b_menu/houdou/18/03/06032817/003.htm
- 山口二郎 (2005) 『ブレイク時代のイギリス』 岩波新書
- 山本義隆 (2007) 『16世紀文化革命 2』 みすず書房
- 山谷清志 (2006) 『政策評価の実践とその課題：アカウンタビリティのジレンマ』 萌書房
- ワインバーグ, G. M. (1979 [1975]) 『一般システム思考入門』 松田武彦監訳・増田伸爾
 訳, 紀伊国屋書店
- Access Economics (2002), *Final Report: Independent Review of the Job Network Provider Star Ratings Method*.
- Aitkin, M. and Longford, N. (1986), "Statistical Modelling Issues in School Effectiveness Studies," *Journal of the Royal Statistical Society. Series A (General)*, Vol. 149, No. 1, pp. 1-43
- Akerlof, G., "Identity and Economics," *Journal of Economic Literature*.
- Arellano, M. (2003), *Panel Data Econometrics*, Oxford University Press.
- Productivity Commission (2002), *Independent Review of Job Network*, Report No. 21, AusInfo, Canberra. <http://www.pc.gov.au/inquiry/jobnetwork/docs/finalreport>
- Bewley, T. (1995), "A Depressed Labor Market as Explained by Participants." *American Economic Review*, 85(2): 250-254.
- Bewley, T. (1998), "Why Not Cut Pay?" *European Economic Review*. vol. 42, issue 3-5, pages 459-490.
- Bewley, T. (1999), *Why Don't Wages Fall During a Recession?* Harvard University Press.
- Blank, R.M. (2000), "When Can Public Policy Makers Rely on Private Markets? The Effective Provision of Social Services," *The Economic Journal*, Vol. 110, No. 462, pp. C34-C49.
- Blinder, A., Canetti, E., Lebow, D., and Rudd, J. (1998), *Asking about Prices: A New Approach to Understanding Price Stickiness*. Russell Sage Foundation.
- Blundell, R. and Windmeijer, F. (1997), "Cluster effects and simultaneity in multilevel models," *Health Economics*, vol. 6(4), pages 439-443.
- Brujin, H. (2007), *Managing Performance in the Public Sector*, Second edition, Routledge.
- Cameron, A. and Trivedi, P. (2006), *Microeconometrics*, Cambridge University Press.

- Courty, P., Heinrich, C. and Marschke, G. (2005), "Setting the Standard in Performance Measurement Systems," *International Public Management Journal*, 8(3), 1-27.
- Courty, P. and Marschke, G. (2003), "Dynamics of Performance-Measurement Systems," *Oxford Review of Economic Policy*, Oxford University Press, vol. 19(2), pages 268-284.
- Courty, P. and Marschke, G. (2004a), "Benchmarking Performance," *Public Finance and Management*, 4(3), p. 288-316.
- Courty, P. and Marschke, G. (2004b), "An Empirical Investigation of Gaming Responses to Explicit Performance Incentives," *Journal of Labor Economics*, Vol. 22, No. 1, pp. 23-56.
- De Koning, J. ed. (2007), *The Evaluation of Active Labour Market Policies: Measures, Public Private Partnerships and Benchmarking*, Edward Elgar Publishing.
- Dixit, A. (2002), "Incentives and Organizations in the Public Sector: An Interpretative Review," *The Journal of Human Resources*, Vol. 37, No. 4., pp. 696-727.
- Dryzek, J. (2006), "Policy Analysis as Critique," Moran, M., Rein, M. and Goodin, R. eds., *The Oxford Handbook of Public Policy*, Oxford University Press.
- Ferlie, E., Lynn Jr. L. and Pollitt, C. eds. (2005), *The Oxford Handbook of Public Management*, Oxford University Press.
- Forester, J. (2006), "Policy Analysis as Critical Listening," in Moran, M., Rein, M. and Goodin, R. eds., *The Oxford Handbook of Public Policy*, Oxford University Press.
- Goldberger, A.S. (1991), *A Course in Econometrics*, Harvard University Press.
- Goldstein, H. (1995), *Multilevel Statistical Models*, 2nd ed., Oxford University Press.
- Goldstein, H. (1998), "Interpreting Aggregate Level Models," mimeo.
- Goldstein, H., and Spiegelhalter, D. (1996) "League tables and their limitations: Statistical issues in comparisons of institutional performance." *Journal of the Royal Statistical Society*, A, 159, 385-443.
- Grubb, D. (2004) "Principles for the Performance Management of Public Employment Services," *Public Finance and Management*, 4(3), 352-398.
- Hart, O. (2003), "Incomplete Contracts and Public Ownership: Remarks, and an Application to Public-Private Partnerships," *Economic Journal*, 113, C69-C76.
- Heckman, J., Heinrich, C. and Smith, J. (2002), "The Performance of Performance Standards," *Journal of Human Resources*, 37(4).
- Hilbert, C. (2007), "Implementation of Performance Measurement in Public Employment

- Services in Switzerland," in De Koning ed. (2007), ch.13.
- Ingram, H. and Schneider, A. (2006), "Policy Analysis for Democracy," in Moran, M., Rein, M. and Goodin, R. eds., *The Oxford Handbook of Public Policy*, Oxford University Press.
- Jobs Australia (2006), "Submission on the Star Ratings for Disability Employment Network," <http://www.ja.com.au/attachments/11136/DENStarRatingsSubmission.pdf>
- Jones, A. (2000), "Health Econometrics," in A.J. Culyer and J.P. Newhouse eds., *Handbook of Health Economics*, vol.1A, ch.6, Elsevier Science.
- Kane, T. and Staiger, D. (2002) "Volatility in School Test Scores: Implications for Test-Based Accountability Systems" *Brookings Papers on Education Policy*, 2002.
- Kano, S. and Ohta, M. (2004), "Long-Run Matching Relationship in the Japanese Labor Market: A Panel Cointegration Approach", *Empirical Economics*, 29(4), pp921-937.
- Kano, S. and Ohta, M. (2005), "Estimating a Matching Function and Regional Matching Efficiencies: Japanese Panel Data for 1973-1999", *Japan and the World Economy* 17(1), pp25-41.
- Laffont, J. and Tirole, J. (1993), *A Theory of Incentives in Procurement and Regulation*, MIT Press.
- Manski, C. (1991), "Regression," *Journal of Economic Literature*, 29(1).
- Manski, C. (1999), *Identification Problems in the Social Sciences*, Harvard University Press.
- Moran, M., Rein, M. and Goodin, R. eds. (2006), *The Oxford Handbook of Public Policy*, Oxford University Press.
- Myers, K. and Goldstein, H. (1997) "Failing Schools or Failing Systems?" Chapter 6. In A. Hargreaves (Ed.), *Rethinking Educational Change with Heart and Mind: 1997 ASCD yearbook*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Mosley, H. and Mueller, K. (2007), "Benchmarking Employment Services in Germany," in de Koning ed. (2007), ch.11.
- OECD (2001), *Labour Market Policies and the Public Employment Service*.
- OECD (2005a), *Employment Outlook 2005*.
- OECD (2005b), *Modernising Government: The Way Forward*.
- OECD (2006), *Employment Outlook 2006*.
- Petrongolo, B. and Pissarides, C. (2001), "Looking into the Black Box: A Survey of the Matching Function," *Journal of Economic Literature*, American Economic Association, vol.

- 39(2), pages 390-431, June.
- Rabe-Hesketh, S. and Skrondal, A. (2005), *Multilevel and Longitudinal Modeling Using Stata*, Stata Press.
- Rice, N. and Jones, A. (1997), "Multilevel Models and Health Economics," *Health Economics*, 6.
- Rubenstein, Schwartz and Stiefel (2003) "Better Than Raw: A Guide to Measuring Organizational Performance with Adjusted Performance Measures" *Public Administration Review*, September/October, 63, 5: 607-615.
- Singer, P. (2005), "Outsourcing War" *Foreign Affairs*, Vol 84, Number 2, March/April. <http://www.foreignaffairs.org/20050301faessay84211/p-w-singer/outsourcing-war.html>
- Snijders, T. and Bosker R. (1999), *An Introduction to Basic and Advanced Multilevel Modeling*, Sage Publications Ltd.
- Stiglitz, J. (1999), "On Liberty, the Right to Know and Public Discourse: The Role of Transparency in Public Life," Oxford Amnesty Lecture.
- Stiglitz, J. (2000), *Economics of the Public Sector*, Third edition, Norton.
- Struyven, L. (2007), "Between Efficiency and Equality: New Public-Private Arrangements in Employment Assistance for the Unemployed," in De Konig ed. (2007), Ch.8.
- Suskind, L. (2006), "Arguing, Bargaining and Getting Agreement," in Moran et al. (2006).
- Van Donk, L. and de Koning, J. (2007), "Mediation Services and the Outflow from Short-term Unemployment: Average and Relative Effectiveness of Public Employment Offices," in de Koning ed. (2007), ch.12.
- Wilson, J. Q. (2000), *Bureaucracy: What Government Agencies Do and Why They Do It*. Basic Books.
- Woodhouse, G. and Goldstein, H. (1988), "Educational Performance Indicators and LEA League Tables," *Oxford Review of Education*, Vol. 14, No. 3, pp. 301-320
- Wooldridge, J. (2001), *Econometric Analysis of Cross Section and Panel Data*, MIT Press.
- Wooldridge, J. (2006), *Introductory Econometrics: A Modern Approach*, 3rd ed., Thomson Higher Education.

A 付録

A-1 ミッション，タスク，指標，その多元性とコスト

基本的な概念をやや詳しく解説する。

公立の学校や病院，大学，HWといった行政組織は，それぞれ何らかの**ミッション**を与えられており，それにもとづいて**タスク**が設定されていると考える．このような様々な候補から一組を選定する手続きを

$\{\text{ミッション}\} \mapsto \{\text{タスク}\}$

あるいは $M \mapsto T$ と書く．

例： 「職業安定法やILO条約の定めるところにより，HWのミッションは職業安定であり，そのタスクは求職・求人のマッチング支援である」など．もちろん，複数のミッションとタスクが存在すると考えなくては現実的ではない．失業給付，自治体とくに福祉行政や教育機関との連携，職業指導，適性検査，差別の禁止など．

例： 公立小学校ならば，「学校運営の自律的・継続的な改善・充実」「保護者・地域住民の学校運営への参画を促進」「全国的に一定水準の教育の質を保証しその向上を図る」といった三つのミッションが考えられる．そのためのタスクは，「①教育課程・学習指導，②生徒指導，③進路指導，④安全管理，⑤保健管理，⑥特別支援教育，⑦組織運営，⑧研修，⑨保護者・地域住民との連携，⑩施設・設備」などが考えられる（文部科学省（2006））．

ミッションとタスクのこのような多元性を， $M_1 \times M_2 \mapsto T_1 \times T_2 \times T_3$ などとあらわす．ただし，説明の簡単化のため，以下では主にタスクが単一であると想定して話を進める．

つぎに，タスクの実施状況，すなわち**アウトプット**の指標である「**成果指標**（Performance measure）」を選択したい．この選定手続きを $\{\text{タスク}\} \mapsto \{\text{アウトプット}\}$ ，あるいは $T \mapsto P$

と書く。ただ、じっさいは、ミッションとタスクは複数あるのがふつうである。したがって、成果の指標も複数必要である。

例： HWならば、指標として就職率（＝就職者数／求職者数）、そして充足率（＝就職件数／求人件数）その他が考えられよう。

例： 前出の公立小学校のタスクのひとつである「①教育課程・学習指導」については、評価指標として、次のような例が考えられている。「指導目標、指導計画、授業時数などの教育課程の編成・実施の状況。児童生徒の観点別学習状況の評価及び評定の結果。学力調査等の結果。運動や体力に関する調査の結果。児童生徒による授業評価の結果。説明、板書、発問など、各教員の授業の実施方法。（…）」⁵¹

指標のこのような多元性を、 $M_1 \times M_2 \mapsto T_1 \times T_2 \times T_3 \mapsto P_1 \times P_2 \times P_3 \times P_4$ などと書く。活動が多元的な対象を評価する際には、評価視点の多角性は重要である。ただし、その場合「明快さ」という定量的測定の魅力との間にはトレードオフが生じる。また忘れてならないのは、現場で情報を収集し、総括し報告するスタッフのコスト（本来業務から離れるという機会費用を含めて）も増加することである（第6節参照）。

⁵¹文部科学省（2006）では、①に関する指標の提案は次のように続く。「（…）視聴覚教材や教育機器などの教材・教具の整備と活用状況。体験的な学習や問題解決的な学習、児童生徒の興味・関心を生かした自主的、自発的な学習の促進状況。個に応じた指導の充実状況（個別指導やグループ別指導、学習内容の習熟の程度に応じた指導、児童生徒の興味・関心等に応じた課題学習、補充的な学習や発展的な学習などの学習活動を取り入れた指導、教師の協力的な指導等）。授業や教材の開発における外部人材の活用状況。地域の自然や文化財、伝統行事などの教育資源の活用状況。学校図書館の計画的利用及び読書活動の状況。観点別学習状況の評価及び評定の客観性、信頼性の確保状況。授業研究の実施状況。」さらに、「食育、人権教育、環境教育については、それぞれ食育基本法、人権教育及び人権啓発の推進に関する法律、環境の保全のための意欲の増進及び環境教育の推進に関する法律、及びそれらに基づいて策定された計画等の趣旨を踏まえて、各学校での指導計画の策定、指導、評価等がなされることが重要である。」最後に、「各学校の事情等に応じて、部活動の状況についても、評価を行うことが考えられる。」次の項目「②生徒指導」についても諸指標の提案があり、以下同様に項目⑩まで続く。

業績評価のシステムが時とともに自己増殖的に複雑化する現象は「マッシュルームの法則」とも呼ばれる (Bruijn(2007)). 辞書によると, mushrooming には「(キノコのように) 急成長する, 成り上がりの, 長続きしない」といった意味がある. 前出の学校評価の例からも想像できるように, 多角的であるために指標の次元を増やせば, それは容易に人間の認知能力をこえる. ミッション, タスク, 指標の集合のそれぞれの元来の要素が, 仮に離散的であったとしても, 直積集合は巨大になる. 何らかの規範的判断がコンピュータの処理能力の範囲内であったとしても, そのプロセスは関係者にとってブラックボックスとなる.

A-2 第3節への補論(1): 多次元(階層)データの活用

各組織レベルの集計変数は, データ収集のコストが低い. しかし, 情報量は個人レベルのマイクロデータに遥かに劣る. 他方, 各組織の個人レベルでのデータは, 整備されていたとしても, 情報保護の問題もあって, 外部の分析者はアクセスできない場合が多い. 集計データとマイクロデータには, この種のトレードオフがある.

仮に, 組織と個人の二つのレベルでデータが入手可能な場合, その豊かな情報量は以下のように活用できる. 各レベル内ではユニットは独立に分布していない可能性がある. 共通テストの平均値による学校比較を考えると, 生徒と学校という二つのレベルがあり, 同じ学校内の生徒たちの成績は独立に分布していない(生徒の母集団から無作為に抽出する場合に比べ, 同一学校内での生徒の成績が均一的となる). いわゆる「朱に交われば赤くなる」である. この場合, 統計学でいう多次元レベルモデルあるいは階層モデルを利用すると, 推定の正確さが増す(Goldstein (1995), Jones (2000), Rice and Jones (1997)は疫学, 犯罪学, 社会学, 地理学, 医療経済学における応用例を紹介している). 計量経済学でこれに近いものは, 観測されないクラスター効果のモデルと呼ばれる(Wooldridge (2001, ch.11)). 両者の違いは, 前者は変数 a_i や v_i の確率分布をそれぞれ iid の正規分布などと想定して, 分散行列の性質と各クラスター内の情報を利用するのに対し, 後者は確率分布の種類に関わらず頑健な分散行

列の推定量を利用する点である。前者には確率分布の特定化の失敗というリスクがあるが、特定が正しい場合は、より正確な推定が可能となる。たとえば、小規模校など、標本サイズの小さなクラスターは情報量が小さい。が、どのクラスターも同じ確率分布からの標本である、という想定のもとでは、他の大きなクラスターの情報量を活用しうる。⁵²

式で表わすならば、いずれも

$$p_{is} = \beta x_{is} + a_i + v_{is}, i = 1, 2, \dots, N, s = 1, 2, \dots, S$$

ここで i は各組織（クラスター）の、 s は各組織内の参加者のインデックス。 x_{is} は組織 i に参加している個人 s の属性をあらわす変数。固定効果モデルは、 a_i の推定値を成果の組織間比較に用いる。一方、多次元モデルは個別効果と誤差項に正規分布を想定し、比較の指標 a_i に信頼区間を明示するエンピリカル・ベイズ予測が利用できる。すなわち、最尤法で推定した分散と平均ゼロという個別効果の正規分布の仮定を用いて事前確率分布 $Prior(a_i)$ を計算、それを (p_{is}, x_{is}) の実現値にもとづく尤度関数とあわせて、 a_i の「更新」された事後確率分布を求める： $Posterior(a_i | p_{is}, x_{is}) \propto Prior(a_i) \times Likelihood(p_{is}, x_{is} | a_i)$ 。この事後確率分布の平均と標準偏差から、各 a_i の信頼区間が求まる（計算例は本稿第 5-3 節参照）。

同様に、一時点での都道府県、組織、利用者といった 3 次元のモデルも可能である。またはパネルデータで組織 i 、利用者 s 、時点 t として

$$p_{ist} = \beta x_{ist} + a_i + d_s + v_{ist}, i = 1, 2, \dots, N; s = 1, 2, \dots, S; t = 1, 2, \dots, T$$

など。本稿 5-3 節は、県、地域ブロック、時点という 3 次元での計算例を示している。

A-3 第 3 節への補論 (2): 調整成果指標における集計バイアス

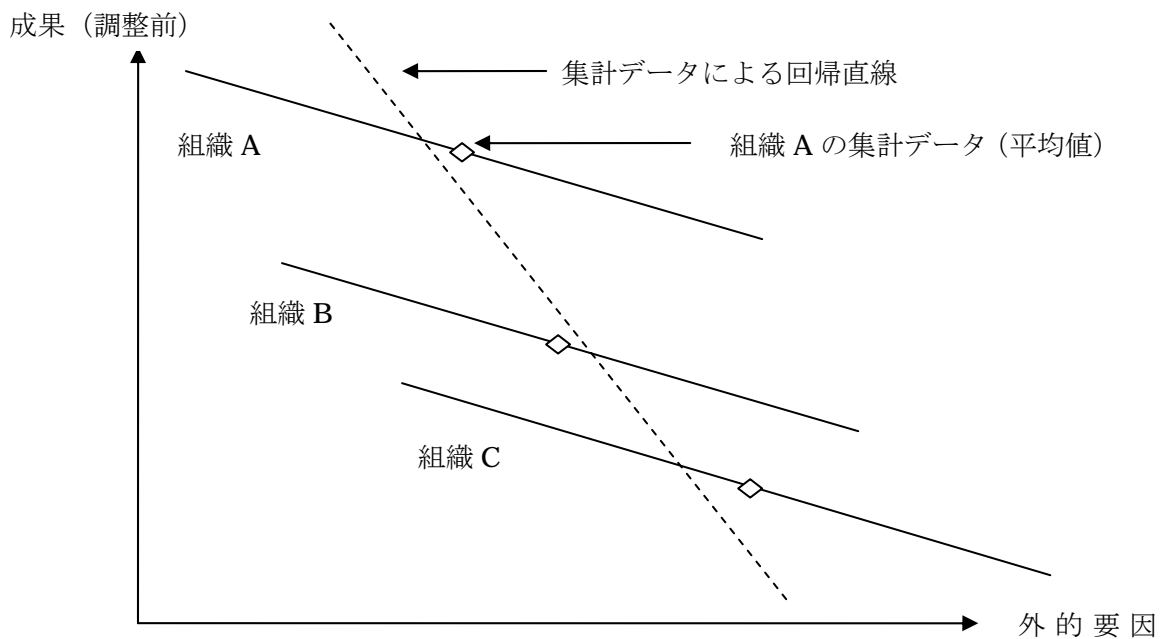
各組織レベルでの集計データの分析は、参加者レベルのマイクロデータの分析と非整合的な結果をもたらす場合もある（Goldstein (1998)）。1 時点のデータの場合を考える。次の図

⁵² ただし、Blundell and Windmeijer (1997)によれば、 $s \rightarrow \infty$ のとき固定効果モデルと多次元モデルは同値である。すなわち、各クラスター内の標本サイズが十分大きければ、「クラスター間の情報量の差の活用」には意味がなくなる。

A-1は、外生的要因 x と成果指標（調整前） p の平面に、三つの組織の「真のモデル」（ x と p の因果関係）を描いたものである。外的要因 x のどの水準においても、組織 A がもっともすぐれた成果を得ている。組織 B は第二位であり、組織 C は最下位である。

同じ図 A-1 において、三つの菱形は、各組織における外生的要因と成果の平均値をあらわす。この平均値のデータを用いて線形回帰分析を行うと、点線のごとき回帰直線（各 x に対して期待される成果水準を与える）が推定されるであろう。このとき、期待される成果水準と実際の成果水準との乖離としての調整成果指標（菱形と点線との垂直距離）を計算すると、組織 C はもっとも大きな正の値を得る。組織 B は負の値であり、最下位となる。つまり、マイクロデータを利用した場合とは順位が変わっている。これは調整成果指標における集計バイアスである。

図 A-1 マイクロデータでの因果関係（実線）と集計データによる推定（点線）



マイクロデータがある場合は、集計データを作成できるので、それぞれを用いた分析の結論を比較できる。逆は不可能である。ただし、これは二者択一ではなく、マイクロデータをもとに複数のレベルを同時に考察する方法もある（第5節およびA-2節を参照）。

A-4 第4節への付図表

変数の定義は4-1節を参照.

表 A-1 記述統計

変数	観測数(=N×T)	平均	標準偏差	最小値	最大値
対数就職件数	7332	7.728	0.676	5.905	11.405
対数求職数	7332	10.423	0.768	8.674	12.559
対数求人件数	7332	10.049	0.708	8.002	12.608

図 A-2 求職件数，求人件数，就職件数（月次全国総計：1992.1 2005.1）

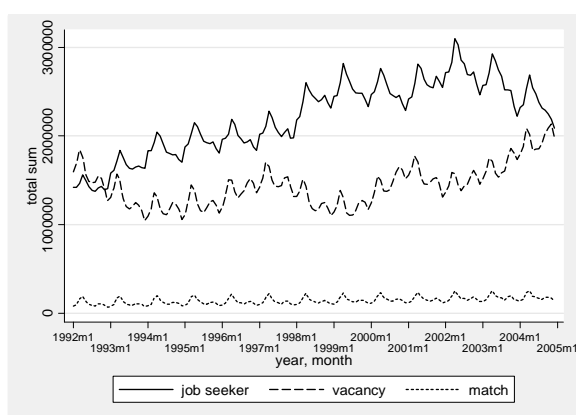
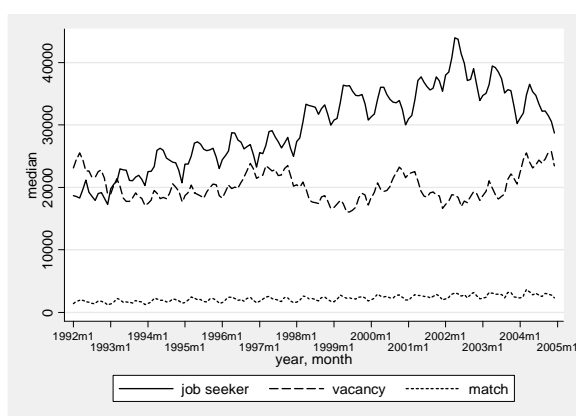


図 A-3 求職件数，求人件数，就職件数（月次中位数（メディアン）：1992.1 2005.1）



A-5 第4-4節への補論：観測上の同値性

ある観測データと整合的な因果関係の数学的モデルが多数存在することは、社会科学では日常的な現象である。計量経済学や一般システム論では「観測上の同値性」「モデルの同型性」などとよばれる。ある因果関係のモデルが、変数の実現値を生み出す、と考えることは、モデルの集合の各要素から、観測値の集合の各要素へ対応をつけることである。これを

$$\{\text{因果モデル}\} \leftrightarrow \{\text{観測値}\} \text{ あるいは } \{f\} \leftrightarrow \{(p_i, x_i; i=1, 2, \dots, N)\}$$

とあらわすことにする。問題は、この対応付けが一对一でなく、多対一の場合に生じる。たとえば、ある標本サイズ N のデータ $(p_i, x_i; i=1, 2, \dots, N)$ について、3種類の因果解釈がいずれも整合的ならば、

f_1

□

$$f_2 \leftrightarrow (p_i, x_i) \quad i=1, 2, \dots, N \quad \text{あるいは} \quad f_1(x_i) = f_2(x_i) = f_3(x_i) = p_i, \quad i=1, 2, \dots, N$$

□

f_3

この場合、どのモデルが「良い」かを所与のデータ自体を根拠に証明なり検定なりすることはできず、別の基準（簡便性や直感的妥当性、一般性、審美性その他）が必要になる（e.g., Cameron and Trivedi (2005), Manski (1999)）。以下は、ワインバーグ(1979[1975], 第4章)の数値例を若干修正したものである。

例：いま、成果指標 p を左右する要因の候補に、3種類の変数 (x, y, z) を思いついたとする。そしてそれらの関係（モデル）として、 $f_3(x, y, z) = p$ 、そして $f_2(y, z) = p$ という二つを考える。前者は3種の規定要因を取りいれて視野が広い、しかし後者はデータ収集コストが低く簡潔である、といえるかもしれない。

どちらのモデルがより「良い」かを実証的に判断するために、データを収集したところ、次の3通りの実現値を観測したとする（標本サイズ=3）。

$$(x_1, y_1, z_1, p_1) = (0, 3, 8, 10)$$

$$(x_2, y_2, z_2, p_2) = (4, 3, 12, 10)$$

$$(x_3, y_3, z_3, p_3) = (-4, 1, 12, 10)$$

たとえば、3変数の線形モデル

$$f_3(x, y, z) = (z - x)/2 + 2y = p$$

がこの標本データと整合的であることは、簡単な計算で確認できる。しかし、2変数の非線形モデル

$$f_2(y, z) = 6(y - 2)^2 + (z - 10)^2 = p$$

もこの標本と整合的である。したがって、整合性という点では両者に優劣はつけられない。 $f_2(y, z) = p$ は、データ収集のコスト節約の点でより優れている。しかし、将来、インプット x の値が成果指標 p を左右している例が観測され、 $f_3(x, y, z) = p$ のみがデータと整合的となるかもしれない。

あるいは、両モデルともにあてはまらず、別の要因 u だけが p を左右している（ように見える）例が観測される可能性もある。たとえば、これまでの実現値が $u_1 = u_2 = u_3 = 3$ で、4個目の観測値が $(x_4, y_4, z_4, u_4, p_4) = (-1, 2, 9, 4/3, 5)$ であったとすると、1変数のモデル

$$f_1(u) = 1 + 3u = p$$

は(3種類のモデルのうちで)唯一、データと整合的である。もしもデータの質が高く、サイズ4の標本で5個の変数 (x, y, z, u, p) を観測できていれば、この判断がくだせた。

しかし、 u が観測できない要因である場合もあろう。さらに、以上のいずれよりも社会科学の意味で魅力的な他のモデルが見つかる可能性は高い。

定性的および定量的な情報の量と質が改善すれば、これまで識別が難しかった複数のモデル(因果解釈の候補)の数を絞ることができる。調整成果指標における討議とフィードバックを論じた第2節を参照。