

# 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

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。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。

## 1. はじめに

長期雇用と組織内スキル形成が弱体化しつつある先進諸国では、公的な職業紹介や能力開発支援といった「積極的雇用政策」のプレゼンスが増している。しかし、低成長期の経済において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・制度の効果を比較してプライオリティを付けていく必要がある。そこで、公的機関の業務を定量的に管理・評価することの必要性が広く認識されるようになった。<sup>1</sup> 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（Brujin, 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、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で、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、関係組織との交渉・指導に利用している。<sup>2</sup>

また、公的部門の定量的な業績管理には、サービス利用者（需要者）により豊かな情報と選択肢を提供すること、そして利用者の反応をふまえて供給者が一層の創意工夫を実現することが期待されている。

しかし、成果水準の安易な定量的評価は公的機関の本来の目的にとって有害ともなりうる。たとえば、成果に影響を及ぼす外的な環境が組織間で異なる場合、環境の違いを考慮せずに各組織を評価付けることは、事実認識として誤っているのみならず、現場の士気を不必要に下げる恐れがあろう。

本稿は、公的機関の定量的な業績管理および評価の方法として、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」を提案する。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

\* 青森公立大学講師

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。

## 2. 調整成果指標のアイデア

### 2-1. 基本的なアイデア

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。職業紹介を例にとると、成果指標には就職件数（ないし率）が考えられる。ただし、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、現実には一般的であるが、きわめてあらっぽい方法であり、またフェアでない。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はいよう。

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。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。この手法

は回帰分析と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。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。

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 $(x_1 = x_{i1} = x_{i2} = \dots)$ 、成果の実績値 $(p_1, p_{i2}, p_{i3}, \dots)$ 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」といった考えの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。形式的にいえば、我々が求めたいのは、外的要因の水準 $x$ に対して成果 $p$ の期待値を与える関数 $E(p|x)$ である。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。これを $p = E(p|x) + u$ と表すことにする。以下、調整指標の作成手順とその解釈を示す。

①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 $p_i$ であらわし、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 $p_i = E(p_i | x_i) + u_i, i = 1, 2, \dots, 600$ を考える。 $u_i$ は平均年齢以外の要因であり、データ上は観測できない。定義より $E(u_i | x) = 0$ である。

条件付き期待値 $E(p|x)$ は、母集団においてある $x$ の水準を共有するすべてのHWにおける $p$ の平均値である。したがって、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。また、各HWの置かれた条件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、 $x$ と $p$

の相関が負だとすれば、求職者の平均年齢が高いHWについては、その分だけ期待される成果が低く計算される。

②実績値との比較 各HWの実際の成果指標の値  $p_i$  と、期待成果指標の値との差を計算する： $p_i - E(p_i | x_i), i = 1, 2, \dots, 600$ 。

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。また、 $x$  と  $p$  の相関が負だとすれば、求職者の平均年齢  $x_1$  が高いHWについては、期待される成果が低いため、 $p_i - E(p_i | x_i)$  は高めに計算される。このような意味で、この値は条件調整済み成果指標とよぶことができる。

③組織間比較 全HWの調整成果指標の  $p_i - E(p_i | x_i)$  を求め、ランキングなどを作成し、調整していない成果指標  $p_i$  のランキングと比較する。

④推定 通例、条件付き期待値関数  $E(p|x)$  そのものは未知であるので、われわれは  $(p_i, x_i)$  の標本を用いてこの関数を推測したい。すなわち、 $p_i = E(p_i | x_i) + u_i$  に最小二乗法などを適用して  $E(p|x)$  を推定する。とくに、 $E(p|x)$  が線形であると仮定するか、あるいは  $E(p|x)$  の線形最良近似を推定するならば、この作業は  $p_i = \beta_0 + \beta_1 x_i + u_i$  のパラメータの推定値  $(\hat{\beta}_0, \hat{\beta}_1)$  を求めることに等しい。一般に  $E(p|x)$  は非線形だが、通例、変数変換などを行って線形関数によるその最良近似を推定することが多い。<sup>3</sup>

## 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。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。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せない与件である諸要因の組(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、紹介担当職員の人数、業務時間などが考えられる。

(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HWがあるとすると、 $p_i = E(p_i | x_i, z_i) + u_i, i = 1, 2, \dots, 600$ 、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  $\mathbf{u}_i$  によって決定されている。

(III) データを集め、条件付き期待値関数  $E(p_i | x_i, z_i)$  を最小二乗法などで推定し、すべてのHWについて、期待される成果と、その実績値との差を計算する。

以上の方法で、各HWの置かれた状況や業務体制に関する違いを調整した上でパフォーマンスを吟味できる。残差分析の応用は、米国で「御値打ちな」不動産物件の選定や、学校教育をめぐる司法判断で利用されているという(Wooldridge(2006))。また、Rubenstein, Schwartz and Stiefel (2003), Courty and Marschke (2004), Grubb (2004) は行政学分野での活用を示唆している。関連文献との詳細な比較は第4-5節で行う。

## 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、 $\mathbf{u}_i$  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) 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。<sup>4</sup>
- (2) 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、実際の計算も表計算ソフトで容易に実行できる。<sup>5</sup> また、データが利用者レベルのミクロデータでなく、各組織や行政区域レベルの集計データであっても、有用な分析が可能である。
- (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（＝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問題が明らかになるかもしれない。

こうして、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→新要因の発掘→モデル設定・データ分析→…」というフィードバックのプロセスを繰り返すことで、より精緻で、現場の人々の納得度も高い評価へと、質を改善してゆくことが可能となる。いかなる分野

であっても、現場で評価される側は、成果指標の結果にかかわらず、分析者・評価者に対して主張したいことを多く抱えているものであろう。この点で、先に述べたように調整成果指標の手法は理解と実践が容易であるので、各オフィス自身が重要と考える外的要因を追加して自ら分析をやり直して新たな結果を提示したり、自己管理に利用したりすることも可能である。いずれ組織の決定に服するとしても、一方的に業績を評定されて終わるか、あるいは考慮されるべき外的要因について発言する機会や応答を受ける機会があるかによって、成員の士気は変わってくるはずである。<sup>6</sup>

- (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ることが当然必要である。
- (5) 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で判断する。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。
- (6) 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、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) 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、「半分も入っていない」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。

(8) 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」は決定係数  $R^2$  として把握される。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、因果関係の意味で成果を左右するとは考えがたい要因を採用して説明力を上げてても意味はない。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り、「説明力」の高さはその結果として達成すべきものである。

### 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。簡単化のため線形モデルを仮定し、切片を省略するなら、各組織の調整成果指標は残差（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：

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

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

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ているでしょう。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  $v_i$  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、

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

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

1時点のデータからは、調整成果指標の水準に於けるべき理由がある ( $a_i$ ) 場合を単なるノイズ ( $v_i$ ) から識別できない。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  $\beta$  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）。ひとつの対処方法は、操作変数を探すことである。他には、次に述べる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

各ユニットを追跡調査したパネルデータがある場合、

$$p_{it} = E(p_{it} | x_{it}, a_i) + v_{it} = \beta x_{it} + a_i + v_{it},$$

$$i = 1, 2, \dots, N, t = 1, 2, \dots, T$$

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  $\beta$  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。推定は、 $a_i = p_{it} - \beta x_{it} - v_{it}$  より、時系列で期待値をとる：

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

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

$E(p_{it})$  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、 $\beta$  の推定値の正確さ次第となる。

パネルデータがあっても  $a_i$  と  $x_{it}$  に相関がある場合は、最小二乗法では係数  $\beta$  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、

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

$$\bar{p}_i = \beta \bar{x}_i + a_i + \bar{v}_i \quad (\bar{x}_i = T^{-1} \sum_{t=1}^T x_{it}, \text{etc.})$$

$$p_{it} - \bar{p}_i = \beta(x_{it} - \bar{x}_i) + v_{it} - \bar{v}_i$$

これを  $\hat{p}_i = \beta x_i + v_i$  と書きなおすならば、形式上は標準的な単回帰モデルに等しく、したがって同様の仮定のもとで  $\hat{p}_i$  を  $x_i$  に回帰する最小二乗法により係数  $\hat{\beta}$  を正しく推定できる。<sup>7</sup> この推定値  $\hat{\beta}$  を利用して全体の理論値を求め、 $\hat{u}_i$  の推定値を計算する： $\hat{u}_i = \hat{p}_i - \hat{\beta} x_i$ 。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  $\hat{u}_i = p_i - \hat{\beta} x_i = p_i - \hat{p}_i$  であった。一方、調整指標  $\hat{u}_i$  は、各変数の期間平均から計算されており、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。実際、外的要因  $x_i$  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。<sup>8</sup>

#### 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

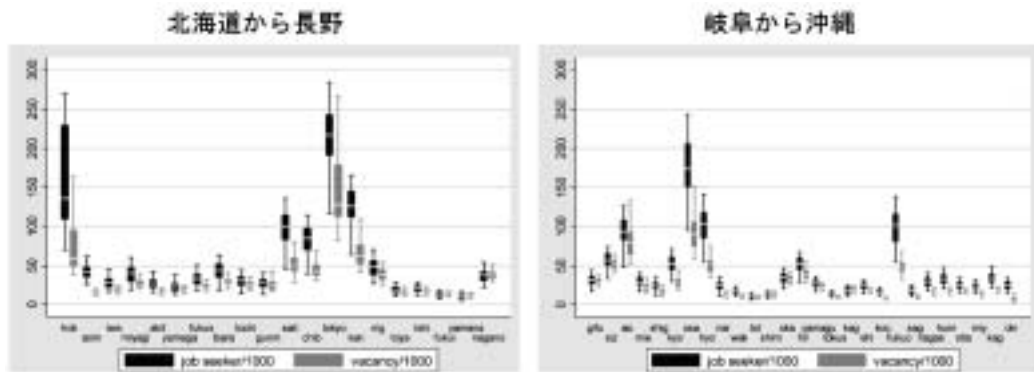
データは、厚生労働省の職業安定業務統計の

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、成果指標としては就職件数をもちいる。「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。<sup>9</sup>

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=13×12=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。

図1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1 求人数(job seeker), 求職者数(vacancy)の分布と中位数 (千人、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4-5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  $M = F(U, V) = U^{\beta_U} V^{\beta_V}$  であると想定する。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。  $\beta_U + \beta_V = 1$ 、すなわちマッチング関数  $F$  が一次同次のときは、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。(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  $M/U = F(U, V)/U =$

$F(L, V/U)$ 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、次の4-3節でデータから支持されることを確認する。

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、さらに次のように特定する。

$$M_{i,t} = F_{i,t}(U_{i,t}, V_{i,t}) = B_{i,t} U_{i,t}^{\alpha} V_{i,t}^{\beta}$$

$$= B \exp(\alpha_i + w_t + z_{i,t}) U_{i,t}^{\alpha} V_{i,t}^{\beta}$$

ここで $\alpha_i$ 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。<sup>10</sup>この非線形関数のパラメータは、対数変換した線形関数  $m_{i,t} = b + \beta_u u_{i,t} + \beta_v v_{i,t} + \alpha_i + w_t + z_{i,t}$  から推定するのが便利である。標準化の仮定  $\sum_{i=1}^n \alpha_i = 0 = \sum_{t=1}^T w_t$  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  $\alpha_i \sim iid(0, \sigma_{\alpha}^2)$ ,  $w_t \sim iid(0, \sigma_w^2)$ ,  $E(\alpha_i + w_t + z_{i,t} | u_{i,t}, v_{i,t}) = 0$  を想定し、多次元モデルをあてはめることもできる。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。)

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 $\alpha_i$ 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 $\alpha_i$ 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承されている地域では、求人数や求職者数も多い、といった可能性がある。第3-2節の表記を使うならば、 $p_{i,t} = \beta x_{i,t} + \alpha_i + v_{i,t}$ において $x_{i,t}$ と $\alpha_i$ に正の相関がある。

$\alpha_i$ 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  $m_{i,t} = b + \beta_u u_{i,t} + \beta_v v_{i,t} + \alpha_i + w_t + z_{i,t}$  に第3-2節で見たデータ変換をほどこし固定効果推定量を計算した結果を表1に示す。期間ダミーの係数は省

表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)

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<sup>11</sup>。

次に、第3-2節で触れた方法によって、各県の固定効果 $\alpha_i$ の推定値を求めた。図2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。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、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 $\hat{\alpha}_i$ 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す。③の調整後順位の列が、

図2 調整指標の分布（県別、地域ブロック表示）

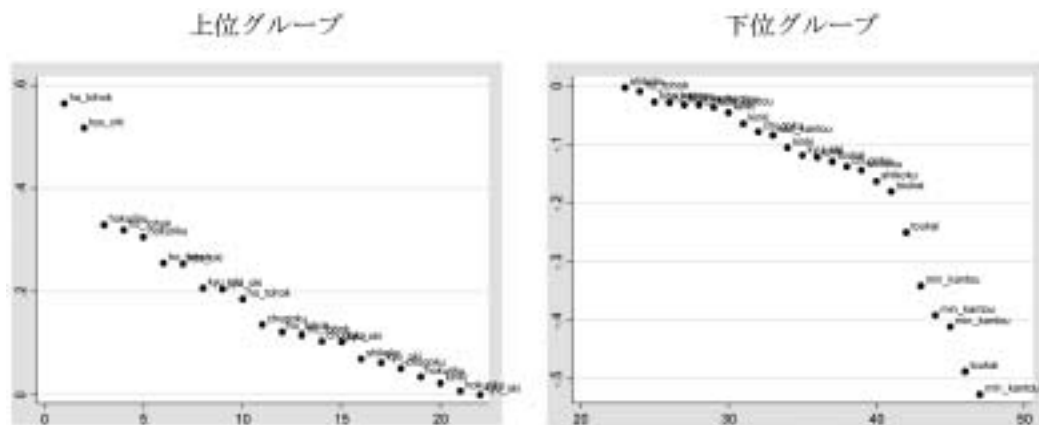


表2 求人充足率・就職率の期間平均、調整成果指標 $\hat{\alpha}_i$ 順位表（県別、地域ブロック表示）

| 県（所属地域ブロック） | ①充足率順位 | ②就職率順位 | ③調整指標順位 | 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  |

調整指標の図2に対応する。調整による変動は大きく、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。

#### 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。ラフに言えば、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、新たな仮説を生成するための参考情報となる。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がないため、民間企業による紹介機構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1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

層に比べて中高年層は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。<sup>12</sup>

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。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)などがある。これらの標準的な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（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、ばらつきの由来を探索

するため、定性的な情報をヒアリングなどで収集して仮説を生成し、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中間までを扱う。）

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。<sup>13</sup>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、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とおり、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）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、業績評価をめぐる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、その紹介は別の機会に譲る）。

## 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

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。

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。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影響を受けるため、時系列での変動が大きい。大規模な学校における成績の平均値は、標本が大きい分変動が少ない。

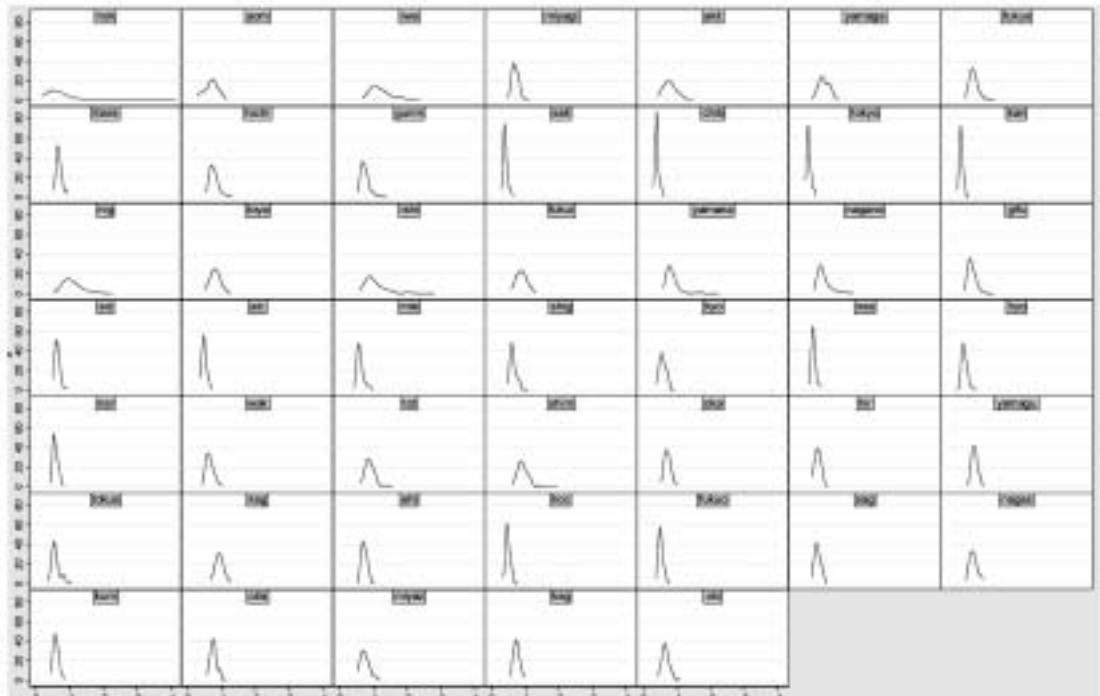
成果の絶対水準でなく、前期からの変化分を指標とする場合は、標本分散の影響は一層大きくなる。変化分は絶対水準よりも小さいため、学校間の差異はもともと少ない。また、生徒(の質)があるiidの分布からの無作為抽出の場合

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))。

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3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、大都市ほどむしろ中央値まわりに集中している。地方は、件数の場合とことなり、分散が大きい。この背景には、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

図3 就職率の分布(県別, 1992.1 - 2004.12, カーネル分布推定)



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って久しい。

## 5-2. 点推定と区間推定

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。二つ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。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。

信頼区間の推定にあたっては、時間、地域ブロック、都道府県という本稿のデータの三つの次元の構造を利用する。いま、地域ブロックと都道府県の二つのレベルに着目するなら、同じブロック内の都道府県のスコアは独立に分布していない（都道府県の母集団から無作為に抽出する場合に比べ、同一ブロック内でのスコアが均一的となる）可能性がある。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。両者の違いは、前者は変数  $a_i$  や  $\mu_i$  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、小規模な地域ブロックなど、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、という仮定のもとでは、他の大きなクラスターの情報量を

活用しうる。<sup>14</sup>

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  $i$  は各ブロック（クラスター）の、 $s$  は各ブロック内の県のインデックス、 $x_{is}$  はブロック  $i$  に参加している県  $s$  の属性をあらわす変数。固定効果モデルは、 $a_i$  の推定値を成果の比較に用いる。一方、本節で利用する多次元モデルは個別効果と誤差項に正規分布を想定し、比較の指標に信頼区間を明示するエンピリカル・ベイズ予測を利用する。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  $\text{Prior}(a_i)$  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：すなわち、 $\text{Posterior}(a_i | p_{is}, x_{is}) \propto \text{Prior}(a_i) \times \text{Likelihood}(p_{is}, x_{is} | a_i)$ 。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。

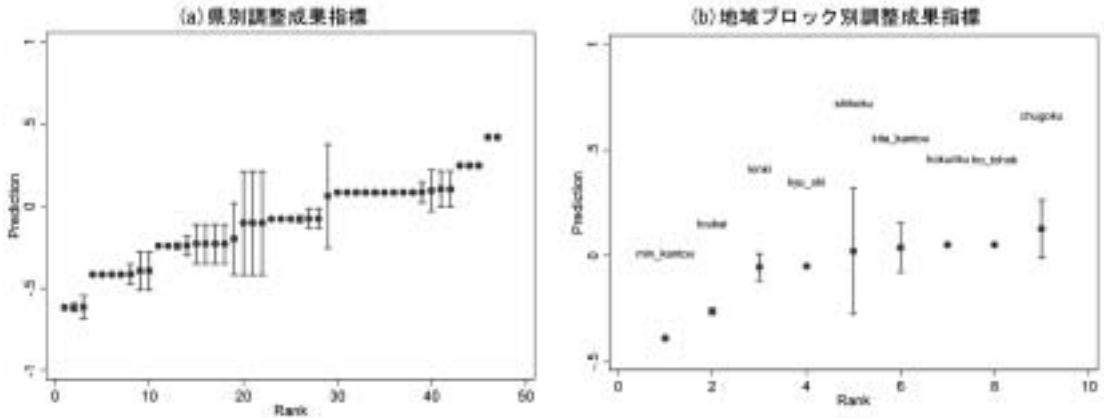
一時点での都道府県、組織、利用者といった3次元のモデルも可能である。またはパネルデータで組織  $i$ 、利用者  $s$ 、時点  $t$  として

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

といった3次元分析も可能である。本節は、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。なお、前節の固定効果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4に示す。パネル(a)は県単位の、パネル(b)は地域ブロック単位の区間推定である。明らかに、区間推定では多くの地域の調整指標は互いに重なっており、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。

図4 95%信頼区間と調整成果指標



注：データは第4節と同じ。

### 5-3. 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）。本稿が前節で計算し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いる。<sup>15</sup> また、HWをふくむ公的組織は、本体と付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付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、分析上の一体化にも問題は少なく、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。<sup>16</sup>

## 6. 結 語

本稿は公的機関の定量的な業績管理および評価の方法論を提案し、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。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て、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、それは回帰分析という残差分析

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地域における各期の労働市場の需給条件（求人、求職）、そして各期のマクロ要因（景気など）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争性、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。

最後に、実際の業績管理に重要と思われる他の論点に触れる。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（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す機能となりうる。むしろ、公的部門の多面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義なものにするには、一定の修練と時間が必要であろう。これらは実際に業績管理を行う際に決定的に重要な論点だと思われるが、本稿の範囲を超える。しかし、調整成果指標は、このような評価と討議のプロセスのなかでも有意義なシグナルの機能を果たしうるように思われる。

(2012年11月30日受付、2013年1月30日受理)

#### (謝 辞)

本稿の構想と執筆にあたって、労働政策研究・研修機構、立命館大学、横浜市立大学、関西労働研究会、厚生労働省職業安定局の関係者・セミナー参加者各位より貴重な示唆を頂きました。残る欠点は著者の責に帰すものです。

#### 7. 文 献

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  
 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  
 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.

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いて」平成18年10月改訂版  
 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  
 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」『地域雇用創出の新潮流 — 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  
 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  
 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  
 Access Economics (2002), *Final Report: Independent Review of the Job Network Provider Star Ratings Method*.  
 Bewley, T. (1998), "Why Not Cut Pay?" *European Economic Review*. vol. 42, issue 3-5, pages 459-490.  
 Blundell, R. and Windmeijer, F. (1997), "Cluster effects and simultaneity in multilevel models," *Health Economics*, vol. 6(4), pages 439-443.  
 Bruijn, H. (2007), *Managing Performance in the Public Sector*, Second edition, Routledge.  
 Cameron, A. and Trivedi, P.(2006), *Microeconometrics*, Cambridge University Press.  
 Courty, P. and Marschke, G.(2004), "Benchmarking Performance," *Public Finance and Management*, 4(3), p. 288-316.  
 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
- Goldstein, H. (1995), *Multilevel Statistical Models*, 2nd ed., Oxford University Press.
- 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.
- Hilbert, C. (2007), "Implementation of Performance Measurement in Public Employment Services in Switzerland," in De Koning ed. (2007), ch.13.
- 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.
- 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*.
- 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.
- Snijders, T. and Bosker R. (1999), *An Introduction to Basic and Advanced Multilevel Modeling*, Sage Publications Ltd.
- Struyven, L. (2007), "Between Efficiency and Equality: New Public-Private Arrangements in Employment Assistance for the Unemployed," in De Konig ed. (2007), Ch.8.
- 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.
- Wooldridge, J. (2006), *Introductory Econometrics: A Modern Approach*, 3rd ed., Thomson Higher Education.

---

<sup>1</sup> 2010年のOECD主要加盟国における平均的な積極的雇用政策支出はGDPの0.66%、最高はデンマークの1.91%である。一方日本は0.28%で最も低い部類である (OECDStat)。先進諸国における積極的雇用政策の現状を把握するにはOECD発行のEmployment Outlook各年版が便利である。

<sup>2</sup> 詳細はOECD(2001)、日本については厚生労働省職安局 (2006) を参照。

<sup>3</sup> 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である。誤差を  $w = E(p|x) - (b_0 + b_1x)$  とするとき、 $E(w^2)$  を最小にする線形近似は  $\beta_0 + \beta_1x$ 、 $\beta_1 = C(x,p)/V(x)$ 、 $\beta_0 = E(p) - \beta_1 E(x)$  で与えられる。

<sup>4</sup> マクロ経済学でいえば、教科書的な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。

<sup>5</sup> 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く、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。

<sup>6</sup> 調整成果指標でえられた知見(正負の外れ値のオフィスなど)をもとに、フィールドに出てコンタクトをとることは、社会科学者にとっても貴重な機会である。研究者共同体をこえて、研究対象となっている人々の交わす意見の空間に自らも参加することは、複数のモデル間の識別、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(1998), Blinder et al.(1998))。

<sup>7</sup>  $E(v_{it} | x_{it}, a_{it}) = 0$  は仮定する。標本において  $N \rightarrow \infty$  あるいは  $T \rightarrow \infty$  のとき、 $\bar{p}_N$  を  $\bar{x}_N$  に線形回帰する最小二乗推定量が  $\beta$  に近づくために必要であるのは  $E(v_{it} | x_{it}) = E(v_{it} - \bar{v}_{it} | x_{it} - \bar{x}_{it}) = 0$ 、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  $\bar{p}$  が今期の  $\bar{p}$  を直接に左右している場合、この条件は成立しない。

<sup>8</sup> 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、 $\hat{\beta}_i$  または  $\hat{\delta}_i, i=1, 2, \dots, N$  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、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。むしろ、 $T \rightarrow \infty$  の場合は時系列分析に特有の問題(非正常性など)が現れる(Kano and Ohta (2004)参照)。

<sup>9</sup> 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。

<sup>10</sup> 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。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。

<sup>11</sup> 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))。

<sup>12</sup> 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労働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。

<sup>13</sup> それぞれ2乗した上で相加平均し平方根をとる  $(\sqrt{((M/U)^2 + (M/V)^2)/2})$ 。

<sup>14</sup> 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。

<sup>15</sup> 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）に格差をつける、といった説明である。青木・奥野 (1996)、Dixit(2002)を参照。

<sup>16</sup> 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。

# Evaluation of Public Sector Organizations Using Adjusted Performance Measures

Jun TOMIOKA

## **Abstract**

Ever since demand has grown for accountability and user choice in relation to public sector institutions, improving performance of such institutions has become the central concern for policy makers. But evaluation of organizational performance in public sector poses difficult issues, due to the lack of unified criteria such as the profit-maximization in private sector. One promising approach is to use adjusted performance measures (APM), which capture factors that influence the organizational performance but are outside of the organization's control. This article illustrates basic ideas of APM and discusses appropriate statistical modeling techniques depending on the types of dataset and substantive context. It then reports findings from an empirical application to the public employment service. Since effectiveness of each public employment office is likely affected by a number of external factors, such as the characteristics of job seekers and vacancies, local labor market tightness and macroeconomic conditions, it is important for accurate assessment to properly consider the influence of these factors. This article also illustrates the importance of interval estimation of APM to deal with the uncertainty due to small sample sizes in the typical application. APM can be used in the wide range of fields, including health, education and policing.