

# 主観的業績評価の理論と現実

— 大手自動車販売会社の人事・製品取引データと社員意識調査結果との  
接合データによる検証\* —

高橋新吾・都留康・上原克仁

この論文の目的は、主観的業績評価の功罪を実証的に明らかにすることにある。現代の職場の多くでは、客観的成果指標の利用可能性は限られており、主観的評価が用いられることが普通である。主観的評価には、①マルチタスクがある場合のインセンティブ強化に役立つ、②制御不能なリスクの軽減に利用することによりインセンティブの強化をもたらすというポジティブな側面がある一方で、③えこひいきなどのバイアスが伴い、④そのバイアスが生産性に悪影響を及ぼすというネガティブな側面もあることを理論研究は示唆してきた。本稿では、大手自動車販売会社の人事・製品取引データと社員意識調査データとの接合データを用い、そうした理論的予測の実証分析を行った。分析の結果、4つの命題はおおむね実証的に支持された。しかし、マルチタスクがある場合のインセンティブ強化とバイアスがある場合の生産性の低下への統計的支持は限界的なものにとどまった。

JEL Classification Codes: M50, M52, J31

## 1. はじめに——主観的業績評価に関する先行研究と本稿の目的

近年のインセンティブに関する経済分析は、賃金が客観的成果指標によって決められる状況設定におけるものが多い。たとえば、Lazear (2000)におけるフロントガラス取り付け工を対象とした出来高給導入のような状況がその典型例である。けれども、現代の多くの職場では客観的成果指標は利用可能ではなく、上司による主観的評価により査定がなされるのが普通である。

主観的業績評価はきわめて広範に用いられているゆえにポジティブな側面があることは当然である。だが、同時にネガティブな側面もある。ポジティブな側面とは主観的評価がインセンティブ強化に役立つということであり、ネガティブな側面とは主観的評価がバイアスを伴うという可能性である。本稿の目的は、主観的評価の功罪に関する実証研究を行うことである。

まず、主観的評価のポジティブな側面を取り上げよう。Holmstrom and Milgrom (1991)によれば、従業員が複数のタスクを行うとき、客観的指標のみに頼ってインセンティブを付与する

と、評価しにくいタスクの遂行がおろそかになる。たとえば、高校教員を生徒の試験成績のみで評価すると、生徒のコミュニケーション能力を培うためのタスクなどが等閑視されるというのがその例である。Baker, Gibbons and Murphy (1994)の場合は、従業員の行動が会社の価値を高めずに客観的指標のみを高めるような状態として、マルチタスク問題をモデル化している。Bakerらは、このような状況において、客観的指標のみを用いてインセンティブを付与すると、粉飾決算などの歪曲が起こるとしている。Bakerらの貢献は、マルチタスクの問題が、主観的評価と客観的評価とを組み合わせることであり、客観的指標のみを用いてインセンティブを付与すると、粉飾決算などの歪曲が起こるとしている。Bakerらの貢献は、マルチタスクの問題が、主観的評価と客観的評価とを組み合わせることであり、客観的指標のみを用いてインセンティブを付与すると、粉飾決算などの歪曲が起こるとしている。

客観的指標を用いたインセンティブ付与においては、成果が自分の制御できないリスクによって左右されることにより、従業員の努力水準がファーストベスト以下になる。もし、誤差が観察可能であれば、成果から誤差を足し引きしたものを基にインセンティブを付与し、インセ

ンティブ強度を高めることができる。もちろん、誤差は観察できないから客観的に誤差を足し引きすることはできない。

しかし、ある程度主観的にリスクを軽減することは可能である。たとえば、業績が制御不能なリスクによって悪くなったと評価者が判断した場合、評価基準を緩めるなどの措置を採ることは可能である。実際、Merchant(1987)やGibbs, Merchant, Van der Stede and Vargus(2004)などの会計学の文献では、事後的に評価基準を緩めるような行動を評価者が採る可能性を示唆している。

ここでひとつ付け加えたいのは、経済学においては、相対評価によりこの問題が解決できるという議論がなされてきたことである(たとえばGibbons and Murphy 1990など)。これによれば、制御不能のリスクがすべての従業員に均等に影響を与えるとすると、相対評価を使えばリスクに左右されないインセンティブ付与ができる。厳密にモデル化された相対評価の理論に比べると、主観的にリスクを除去するという論理は厳密ではない。けれども、先に述べたように、現実には主観的評価は広範に行われている。したがって、われわれはリスク軽減に関する両方の理論を検証する。

さて、主観的評価の実証研究は、客観的評価のそれに比べると数こそ少ないが近年徐々に増えてきている。だが、実証研究の多くはHolstrom(1979)のインフォーマティブ原理(Informativeness principle)を実証するにとどまる研究が大半を占めている。つまり、それらは主観的評価が客観的評価に含まれない重要な情報を含んでいるか否かの実証である。主観的評価がマルチタスクインセンティブの付与や、リスク軽減といった役割を担っているかどうかを実証している文献はGibbs, Merchant, Van der Stede, and Vargus(2004)やEderhof(2010)などごく一部である。

以上から、本稿の第1の目的は、日本の大手自動車販売会社の人事および製品取引データを使うことによって以下の3つの問題を検証することである。

(1)主観的評価はマルチタスクインセンティブを強化するために使われているか。

(2)主観的評価は評価の基準を事後に緩めるような形でリスクの除去に使われているか。

(3)相対評価の証拠はみられるか。

次に、主観的評価の2番目のテーマであるバイアスの問題に移ろう。主観的評価は先に述べたように、インセンティブ強化に役立つが、バイアスもしくは「えこひいき」という問題をはらむ。Prendergast and Topel(1996)は、「えこひいき」を特定の従業員に対して評価者がもつ嗜好としてモデル化し、「えこひいき」が常にインセンティブを弱めることはないとしている。これに対し、MacLeod(2003)においては、評価者の評価と被評価者の自己評価の違いをもってバイアスを定義している。MacLeodはこれを知覚されるバイアス(perceived bias)と呼んでおり、知覚されるバイアスがある場合は、従業員の努力量が減るであろうことを示している。

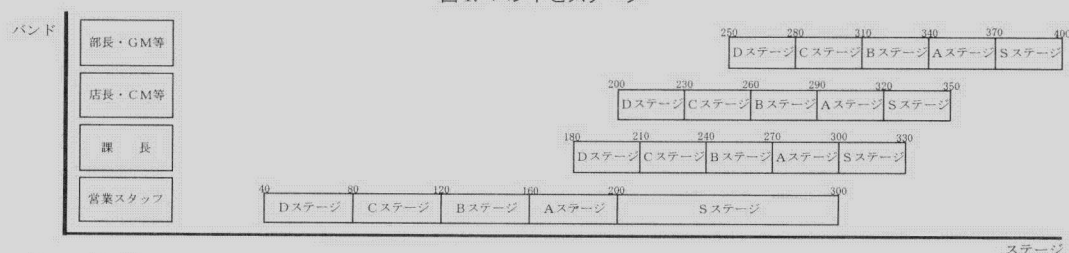
さて、主観的評価に関するバイアスの実証研究は、男女間賃金格差の研究など間接的に主観的評価のバイアスを示すものがほとんどである。直接に主観的評価のバイアスを示したものは、オーケストラ団員の採用など、普通の事業会社以外でのケースが多い。通常の企業におけるバイアスの存在を示したものは少数であって、評価者の人種が評価結果に与える影響を調べたElvira and Town(2001)にとどまる。また、評価のバイアスが生産性に影響を与える影響を調べたものはほとんどなく、管見の限りではEngelland and Riphahn(2011)を挙げるのみある。そこで、本研究の第2の目的として前記の問題に続く第4と第5の問題を検証する。

(4)主観的評価におけるバイアスの証拠はあるか。

(5)主観的評価におけるバイアスは生産性を下げるか。

以上の、5つの問題を分析する際に、われわれの用いる大手自動車販売会社のデータは最適

図1. バンドとステージ



注) 図中の数字は「職務給」(千円)を意味する。

表1. 業績考課シート  
(ある新車営業スタッフの例)

	考課項目	ウエイト	評価点	考課
量的業績	個人販売台数	70	30	B
	個人獲得利益	70	20	
	個人保険収入	20	20	
	目標面接	40	40	
質的業績	売掛金回収までの所要日数	50	20	
	過去販売車両車検掌握率	50	30	
	目標面接	100	30	

注) 業績考課は量的業績評価点の合計と質的業績評価点との合計を所定のマトリックス表を参照して決定される。

であろう。なぜなら、後に詳述するように、第1に、そのデータには主観的評価結果のみならず、獲得利益(自動車販売価格から車両原価と各種経費を差し引いて算出される値)などの客観的成果指標が含まれているためである。また第2に、そのデータは、人事評価に関する社員意識調査結果と接合されており、評価結果を社員がどう感じているかがわかるからである。

## 2. AutoJapanの賃金・人事制度

AutoJapan(仮称)は、従業員数約2000名の国内有数の大手自動車販売会社である。同社は、2000年に職能資格制度の全廃を含む人事制度改革を実施し、バンドと呼ばれる職位の序列が作られた(図1)。

AutoJapanの人事制度を理解するためには、ランク付けのシステムを理解する必要がある。バンドの序列は、部長・ジェネラルマネージャー(GM)、店長・チーフマネージャー(CM)、課長、営業スタッフの順である<sup>1)</sup>。なお、中古車営業スタッフも多数存在する。しかし、制度改革とほぼ時を同じくして、改革とは無関係な理由により、中古車営業スタッフが新車も臨時

に担当したり担当を外れるという不規則な変化があったため、ここでは考察の対象から外す。

図1にみるように、各バンド内の従業員は、5つのステージにランク付けされる。ステージの区分は、職務給<sup>2)</sup>を単純に5つのレンジに区分したものである。たとえば、職務給が月20万円を超えたらSステージになるなどである。

2000年の人事制度改革以降、毎年の職務給昇給額は年度末に行われる業績考課により決まることとなった。業績考課はS, A, B, C, Dの5段階評価になっている。業績考課は「量的業績」と「質的業績」を総合して評価するシステムになっている(表1)。量的業績は各個人の販売台数、個人獲得利益、個人保険収入手数料および目標面接の4項目からなっており、各項目にウエイトが決められている。質的業績は、新車営業においては売掛金回収までの所要日数、過去販売車両の車検掌握率、および目標面接によって決まり、それぞれにウエイトが与えられている。この各項目に第1考課者である各店舗の課長が評価点をつけ総合点が計算される。これを基に、上記の5段階の評価のいずれかが与えられる。このように、業績考課は客観的評価と主観的評価の両方を要素として含むことが容易に理解できよう。

業績考課に使われるアルファベットとステージに使われるアルファベットは同じである。これは偶然ではない。表2にみるように、職務ステージと業績考課とは、賃金昇給額決定マトリックスによって関係づけられている。たとえば、Aステージの人は自分のステージより高い評価(つまりA評価以上)を受けることが期待されており、実際A評価を得ないと昇給はない。

表2. 昇給決定マトリックス

(単位: 千円)

		職務給昇給額				
		業績考課結果				
		S	A	B	C	D
職務 ステ ージ	S	2.3	-2.2	-40.0	-50.0	-70.0
	A	5.6	2.1	-2.1	-40.0	-50.0
	B	40.0	5.2	2.0	-1.9	-40.0
	C	50.0	40.0	4.8	1.8	-1.8
	D	70.0	50.0	40.0	4.4	1.7

注) 自分のステージ以上の考課結果を取らないと、職務給が上らない、それ未満だと職務給が下がる。

Aより下の評価を受けると賃金が下がるシステムになっている。これは、その他のステージでも同じである(より詳しい説明はTsuru 2008を参照)。

AutoJapanの業績考課は、制度設計上は絶対評価で行うことが原則である。しかし、現実問題としては、各従業員の所属ステージを判断基準として「よくやった者」はそのステージより高い考課を、「それなりの成績の者」はそのステージと同じ考課を上司が与えるという可能性がある。つまり、従業員の現在のステージを基準に考課を決める可能性に関して推定の際に考慮する必要がある。

### 3. AutoJapanにおける運用実態と仮説の設定

#### 3.1 新人の指導育成

ここからは、主観的評価がどのようにインセンティブ強化に役立っているかについての作業仮説を設定していく。まずAutoJapanでは、どのようなマルチタスクインセンティブが重要であろうか。AutoJapanは自動車販売会社ゆえ販売台数や獲得利益などが営業スタッフの会社への一番の貢献であって、マルチタスクの問題は他のホワイトカラーの仕事と比較して少ないと思われるかもしれない。しかしながら、以下の2つの理由でマルチタスクは重要である。

まず第1は新人の指導育成である。これは、職場で日常的かつ非公式的に行われる後輩への助言などを含む。われわれは本研究に先だって、いくつかの店舗で聞き取り調査を行ったが、ほとんどの店長や課長が指導育成の重要性を指摘

した。

「営業スタッフの人事考課は課長がつけてきますが、私も全部目を通しますので当然私も関与します。これは非常にむずかしい話になりますが、同じ台数を10台ずつ販売するA君とB君がいて、A君は後輩の面倒見がすごくよくて、同席もするし商談も手伝ういろいろなアドバイスもする、そして10台売ってくる。B君は自分のことしかしなくて同じ10台売っている。間違いなくA君の方が評価はぐっと高いです。ですから、実績面ではないところでの評価をしているつもりです。大事な要素だと思います。やがて自分が課長になっていくステップを踏んでいる営業スタッフの育成をしていくということで必ず必要です。」<sup>3)</sup>

指導育成の重要性の高い店舗では、主観的評価を用いて指導育成を評価することによりインセンティブを付与する可能性がある。これを検証するためには、指導育成の重要性を表す指標が必要である。この指標として、われわれは店舗ごとの営業スタッフ1人当たりの新人数を使う。主観的評価により新人育成に対するインセンティブを付与しているとすると、従業員はこのインセンティブに反応するはずであって、営業スタッフ1人当たりの新人数と業績考課結果との間には正の相関関係が生まれるはずである(ただし、新人訓練に手間取ると負の相関関係も考えられる)。その際、教育は先輩が後輩を指導するという形で行われるとすると、指導育成へのインセンティブは、各店舗において比較的地位の高い営業スタッフに対して特に強く付与されるものと考えられる。

**仮説1**「営業スタッフ1人当たりの新人数が業績考課結果に及ぼす影響は、各店舗でステージが相対的に高い社員で正となる。」

#### 3.2 長期的顧客関係の構築・維持

マルチタスクが重要と思われる第2の理由は長期的な顧客関係の構築・維持である。フォロ

ーアップなどを丁寧に行い長期的な関係を構築すると、顧客のリピーター需要の掌握につながる。しかしながら、販売台数のみで営業スタッフを評価すると、フォローアップなどのタスクに対するインセンティブの付与ができない。そこで上司は、長期的な顧客関係の構築を主観的に評価するであろう。

さて、株式会社などの法人客は個人客に比べてリピーター需要が高いはずだから、長期的顧客関係は、特に法人客において重要になると予測される。実際われわれが行った聞き取り調査によれば、法人客の場合は社長や発注担当者にいかにか会うか、または会えるかが、営業スタッフとして重要であり、個人客とは異なる仕方で顧客満足を向上するタスクが求められている。

それゆえ、長期的な顧客関係の向上に対するインセンティブは特に法人客の割合の高い店において強く付与されるであろう。このインセンティブに従業員が反応するとすれば、法人客割合と業績考課に正の相関関係が生まれるはずである。これから第2の仮説が設定される。

**仮説2**「店単位での年間総顧客数に占める法人客の割合と業績考課との間には正の相関関係がある。」

### 3.3 リスクの軽減

次に、主観的評価がリスクの軽減に使われているかどうかを考えてみる。この点に関しても、われわれは次の質問をした。「もし店舗単位で一時的なショックがあったら、評価の底上げなどで対応するか。」これに関しては意見が分かれた。ある管理職は一時的なショックも永続的な地理的状况も考課に際して考慮することはないとしている。これに対し、他の管理職は、一時的ショックがあれば店舗レベルでそれを考慮するであろうし、部下に対してそうせざるを得ないだろうと述べている。それゆえ、これもデータにより検証する必要がある。

この検証のために以下の変数を考えてみる。

生産性ギャップ=店舗の生産性-会社全体の生

産性

店舗の生産性は店舗における1人当たり年間獲得利益である。これに対し会社全体の生産性は、全社における1人当たりの年間獲得利益である。この差をここで生産性ギャップと呼んでいる。生産性ギャップは、店舗全体に影響を及ぼす何らかのショックの結果であると考えられる。

考課者が制御不能なリスクを主観的に除去しているとする、この生産性ギャップが負になった場合、事後的に評価基準が緩められるはずである。実際には、生産性が負になった場合は業績考課に上乘せされることが観察されるであろう。

ただし、逆に生産ギャップが正であった場合に評価基準を辛くすると、従業員はこれを見越して努力を最低限度に抑えるような行動に出る可能性がある。つまりラチェット効果が働く恐れがある。このため、主観的評価によるショックの吸収は、主に負のショックがあった場合に行われると考えられる。以上から仮説3を設定する。

**仮説3**「生産性ギャップが負の場合には、事後的に考課結果に上乘せがなされる。しかし生産性ギャップが正の場合には、考課結果の差引きは行われない。」

### 3.4 相対評価の混入

このデータで相対評価の問題を考える場合、2つの準拠集団における位置を考える必要がある。まず1つめは店舗内での各従業員の相対的位置で、2つめは各ステージ内での相対的位置である。これを表現するために、各準拠集団における各人の年間獲得利益ランキングを考えてみよう。すなわち、店舗内獲得利益ランキングとステージ内獲得利益ランキングがそれである。

もし相対評価が混入しているとすれば、獲得利益をコントロールした後でも、それらの変数は業績考課結果に影響を及ぼすであろう。つまり、ランキングがよくなればなるほど(1位に近づくほど)考課結果は高くなることが予想さ

れる。

ただし、ランキングが獲得利益に与える効果は、準拠集団の規模に依存すると考えられる。たとえば、大きい店舗内でランキングを上げるのは大変であろう。とはいえ、昇給額は全社共通で決められている。したがって、ランキングを上げることのうまみは大きい店舗ほど小さいといえる。つまり、相対評価によるインセンティブ効果は大きい店舗のほうが弱いであろう。これから、小さい店舗の方がより相対評価にウェイトを置いている可能性がある。以上から仮説4を設定する。

**仮説4**「店舗内獲得利益ランキングまたはステージ内獲得利益ランキングが高くなればなるほど業績考課は上がる。また、ランキングが業績考課に与える影響は店舗の規模またはステージの規模により異なる。」

#### 4. データ

以下の分析では AutoJapan から入手した以下の3つのデータを用いる。まず、従業員人事データである。これで年齢、入退社状況、異動と毎月の賃金台帳の詳細がわかる。年度ごとに行われる業績考課結果は2000年度から2003年度まで利用可能である。したがって、分析はこの期間に絞られる。次に、製品取引データで、いつ誰がどの車種を誰にいくらで販売し、どの程度の利益を上げたのかがわかる記録である。

さらに、われわれは、AutoJapan 社員に対して2000年の人事制度改革に関する従業員意識調査を2005年に行っており、このデータも合わせて使用する。意識調査では回答者に社員ID番号を記入して頂いたため、これにより人事・製品取引データと従業員意識調査との接合が可能となる。要約統計量は表3の通りである。B欄でサンプル数が激減するのは、①A欄に含まれている従業員のうち意識調査に回答していない新車営業スタッフがかなりいるため、そのサンプルがデータ接合の際落とされたこと、また②B欄はクロスセクションデータであるためである。C欄でサンプルが激減するのは、

これが店単位での分析だからである。なお、分析対象は経験2年以上の新車営業スタッフに絞る。経験1年以下の新人は売り上げがゼロの場合がほとんどだからである。また、課長に昇進する1年前の者も除外した。昇進直前には一般スタッフとは異なる評価が与えられる(「手心が加えられる」)可能性があるためである。

### 5. 推定モデルと分析結果

#### 5.1 主観的評価とインセンティブ

以上の仮説を検証するために以下の線形モデルを推定する。

$$\begin{aligned} (\text{業績考課結果または昇給額})_{it} = & \beta_0 + \beta_1(\text{個人客年間獲得利益}) + \beta_2(\text{法人客年間獲得利益})_{it} + \beta_3 \\ & (\text{個人客販売台数})_{it} + \beta_4(\text{法人客販売台数}) + \beta_5 \\ & (\text{ステージダミー})_{it} + (\text{マルチタスク関連変数})_{it} + \gamma(\text{生産性ギャップ変数})_{it} + \delta(\text{相対評価変数})_{it} \\ & + (\text{店舗固定効果}) + u_{it} \end{aligned}$$

被説明変数として業績考課結果はアルファベット考課を数字に直したものを使った。S評価が5でD評価が1である。また、代替的に、個人別昇給額を用いた。

業績考課の重要な決定要因は年間獲得利益と年間販売台数である。法人客は個人客に比べ顧客当たりの売り上げが大きい場合があり、それゆえ売り上げの分散も大きい。このため、個人客と法人客とでは考課結果に異なる影響をもちるので別個に入れた。各従業員の現在のステージが考課の基準になる可能性があることはすでに述べた。それゆえ、ステージダミーを入れた。

マルチタスク関連変数として、各店舗における「営業スタッフ1人当たりの新人数」が入る。新人指導に対するインセンティブは店舗内でより地位の高いものにより強く与えられるはずであるから「店舗内第1~3位ステージダミー」との交差項も合わせて入れる。「店舗内第1~3位ステージダミー」とは各店舗にいる営業スタッフの上位3つのステージのダミーである。これらの変数により指導育成にインセンティブが

表 3. 要約統計量

	サンプル数	平均値	標準偏差	最小値	最大値
<b>A 考課昇給決定式で使われた変数</b>					
業績考課結果	2061	2.653	0.834	1.000	5.000
昇給額(円)	2061	3629.209	11534.250	-50000.000	50000.000
年間獲得利益(万円)	2061	2066.534	788.871	137.284	6157.831
年間販売台数	2061	62.501	23.063	7.000	177.000
年間獲得利益(個人客からのみ)(万円)	2061	1649.153	-637.496	408.815	4345.844
年間獲得利益(法人客からのみ)(万円)	2061	417.381	336.072	-22.689	2623.720
年間販売台数(個人客からのみ)	2061	44.906	17.072	4.000	120.000
年間販売台数(法人客からのみ)	2061	17.595	14.886	0.000	158.000
ステージ S ダミー	2061	0.025	0.157	0.000	1.000
ステージ A ダミー	2061	0.098	0.297	0.000	1.000
ステージ B ダミー	2061	0.257	0.437	0.000	1.000
ステージ C ダミー	2061	0.309	0.462	0.000	1.000
ステージ D ダミー	2061	0.310	0.463	0.000	1.000
営業スタッフ 1 人当たり新人数	2061	0.151	0.351	0.000	2.800
店舗内第 1 位ステージダミー	2061	0.203	0.403	0.000	1.000
店舗内第 2 位ステージダミー	2061	0.247	0.432	0.000	1.000
店舗内第 3 位ステージダミー	2061	0.249	0.432	0.000	1.000
年間総顧客数における法人客の割合	2061	0.077	0.048	0.010	0.247
生産性ギャップ(万円)	2061	-0.684	296.530	-757.361	1237.802
店舗内獲得利益ランキング	2061	4.533	2.545	1.000	13.000
ステージ内獲得利益ランキング	2061	81.925	55.377	1.000	250.000
店舗内営業スタッフ数	2061	8.003	1.955	3.000	13.000
ステージ内営業スタッフ数	2061	162.678	52.030	2.000	250.000
営業スタッフの勤続年数	2061	12.766	9.412	1.083	42.167
課長の勤続年数	2061	26.376	5.124	10.833	37.167
<b>B 残差推定式で使われた変数</b>					
個人平均残差	266	0.011	0.238	-1.139	0.752
1日に上司と話す平均的な時間数(分)	266	28.316	22.676	0.000	180.000
1日に部下と話す平均的な時間数(分)	266	4.267	12.883	0.000	120.000
「業績考課は公平ではない」という回答ダミー	266	0.015	0.122	0.000	1.000
「アドバイスをする人は以前も今もいる」という回答ダミー	266	0.211	0.408	0.000	1.000
「以前も今も助け合いしながら仕事をしている」という回答ダミー	266	0.782	0.414	0.000	1.000
<b>C 店舗生産性推定式で使われた変数</b>					
各店舗生産性(1人当たり獲得利益)(万円)	212	2076.224	311.661	1416.450	3411.614
各店舗残差標準偏差	212	0.407	0.136	0.124	1.054
店舗営業スタッフ数	212	7.514	1.919	4.000	12.000
店舗平均営業スタッフ勤続年数	212	12.998	3.255	5.479	21.667
営業スタッフ 1 人当たり新人数	212	0.218	0.474	0.000	2.800
課長の勤続年数	212	26.453	5.244	12.167	36.167

与えられていたか否かを検証する。次のマルチタスク変数として店単位での総顧客に占める法人客割合を入れる。もし長期的顧客関係にインセンティブが付与されていれば、法人客割合の係数は正になるはずである。

次に、リスクが主観的に除去されているか否かを検証するために、生産性ギャップをモデルに入れる。生産ギャップは店舗全体に影響を及ぼす何らかのショックの結果である。評価者がリスクを主観的に除去する場合、生産性ギャッ

プが負であれば、事後的に評価基準が緩められる。しかし、生産性ギャップが正の場合に評価を辛くするとラチェット効果の恐れがある。したがって、リスク除去は生産性ギャップが負の場合のみにみられるであろう。

相対評価関連変数は、各店舗内での獲得利益ランキング、およびこれと店舗の営業スタッフ数との交差項、各ステージ内での獲得利益ランキング、および各ステージ内での営業スタッフ数との交差項である。これらの変数によって、

表 4. 業績考課結果・昇給額の規定要因、店舗固定効果

	業績考課結果	昇給額(円)
年間獲得利益(1万円, 個人)	0.0002397*** (0.0000525)	3.96133*** (1.370951)
年間獲得利益(1万円, 法人)	0.0000647 (0.0000851)	1.77975 (1.690976)
年間車両販売台数(個人)	0.110027*** (0.0014876)	184.0225*** (31.59292)
年間車両販売台数(法人)	0.008907*** (0.0024177)	117.2553*** (29.47316)
営業スタッフ1人当り新人数	-0.1096461** (0.0483089)	-1175.0410 (946.4669)
営業スタッフ1人当り新人数×店舗内第1位ステージダミー	0.2378252** (0.0756873)	1559.695 (1245.858)
営業スタッフ1人当り新人数×店舗内第2位ステージダミー	0.1612159** (0.0636189)	1862.729** (817.848)
営業スタッフ1人当り新人数×店舗内第3位ステージダミー	0.2104604*** (0.0441807)	4342.466*** (1150.628)
全顧客に占める法人客割合	0.2051373 (0.7178178)	15667.84 (11913.78)
生産性ギャップ(正)	0.0001338 (0.0001141)	2.223939 (2.323556)
生産性ギャップ(負)	-0.0003343** (0.0001278)	-6.461996** (1.995984)
店舗内ランキング	-0.0944348* (0.521877)	-1632.739 (1052.903)
店舗内ランキング×店舗内営業スタッフ数	0.0090756* (0.0051362)	122.8782 (94.1289)
ステージ内ランキング	-0.2793351*** (0.0414544)	540.5018 (1015.242)
ステージ内ランキング×ステージ内営業スタッフ数	0.0009159*** (0.0001858)	-1.751775 (4.290776)
勤続年数・ステージダミー等のコントロール変数	Yes	Yes
サンプル数	2061	2061
F値	247.95	34.51
自由度修正済決定係数	0.7562	0.4869

相対評価が混入していたかどうかを検証できる。実際の推定では、ランキング変数を正規化することにより、ランキングが平均から1標準偏差上がった時効果がどのように変化するかを調べる。

本稿では、店舗の属性がどのように業績考課に影響を与えるかを調べることにより、主観的評価がインセンティブに及ぼす効果を検証している。つまり、評価者は店舗の属性を基にどのようなタスクにどの程度のインセンティブを付与するかを決定する。このウェイトを知って、従業員は努力量を決定する。結果として、店舗の観察される属性と業績考課の間に相関関係が生じるはずだから、これを計測するというのが

本稿の基本方針である。もちろん、われわれが観察できない属性に依存して、新人の数の配置などが行われている可能性がある。また、生産性ギャップは、一時的なショック(たとえば仮店舗での営業)などのほか、永続的な立地条件なども影響する。したがって、そのような影響を係数から除去するために、店舗固定効果を入れている。

それでは結果をみよう(表4)。まずマルチタスクに関連する結果である。営業スタッフ1人当たり新入社員数は、ステージの高い営業スタッフに限って業績考

課に正の影響を及ぼしている。つまり、被説明変数が業績考課結果の場合と個人昇給額の場合とで若干結果が異なるが、新人数とその店舗内第1~3位ステージダミーとの交差項の合計値は、おおむね正の効果をもつ。しかしながら、F検定の結果によれば、新人教育の影響はすべてのモデルにおいて第3位ステージのみで有意である。つまり、新人教育は、店舗内での営業トップ層が行うというよりは、中間層の営業スタッフによって主に行われているということを示唆している。本結果は、すべての従業員が平等に新人教育をしているわけではないが、指導育成に対するインセンティブを与えるために、主に高ステージの従業員に主観的評価による考



課結果の上乗せがなされている仮説と合致する。

もうひとつのマルチタスク要因である長期的顧客関係にインセンティブが与えられているか否かに関しては、全顧客に占める法人客割合の係数は正ではあるが有意ではなかった。したがって、主観的評価が長期的顧客関係に対するインセンティブ付与に使われているという仮説は支持されない。もっとも、全顧客に占める法人客割合や法人の年間獲得利益は、われわれが意図する長期的顧客関係の重要性を正しく数量化できているとはいえず、今後改善の余地がある。

次にリスク除去に関する結果をみる。これに関しては生産性ギャップが負の場合、これによって業績考課にペナルティがつかないように考課の底上げが行われている。つまり、Merchant(1987)などが主張しているように、主観的評価を使ってリスクを軽減する行動を評価者が行っていることが確認された。

さらに、相対評価が混入しているか否かをみる。まず、店舗内でのランキングの係数が負になっており、6%で有意である。また、ランキングと店舗の人数の交差項が有意ではないが正になっている。つまり、獲得利益をコントロールした後でも、店舗内ランキングが上がると(つまり1位に近づく)業績効果が上がるという結果であるが、この影響は小さい店舗の方が大きい。平均的な規模の店舗でいうと(営業スタッフ数8人)、順位が1標準偏差上がると、業績考課が0.02上昇する計算になる。ステージ内でのランキングも重要なことがわかった。100人営業スタッフがいるステージでは、ランキングが1標準偏差上がると業績考課が0.18上がる。このことから、相対評価が混入している様子がうかがわれる。

以上の結果をまとめると、主観的評価がマルチタスクのインセンティブに使われているという仮説は、新人の育成指導の面に絞っていえば支持が得られた。しかしながら、長期的顧客関係の構築に関するインセンティブが与えられているという仮説は支持されなかった。主観的評価がリスクを軽減するために使われている仮説に対しては支持が得られた。さらに、絶対評価

のシステムにもかかわらず相対評価が混入しているという仮説も支持されることがわかった。

## 5.2 主観的評価とバイアス

ここまでは、主観的評価のポジティブな側面をみてきた。次に、バイアスというネガティブな側面に視点を移す。すでに述べたように、通常の事業会社を対象として、主観的評価にバイアスがあるか否かという実証分析は少ない。また、評価バイアスが生産性にどのような影響を及ぼすかに関する実証分析も乏しい。

そうした中で、Engellandt and Riphahn(2011)の研究は貴重である。彼らは大手企業の人事パネルデータを使って、「えこひいき」が社員の努力度に及ぼす影響をチェックしている。彼らの仮定は、「えこひいき」があれば同じ人が同じような業績考課を毎年取るであろうというものである。彼らは、各部署での業績考課の経年変化の標準偏差を取ることによって、「えこひいき」の指標とした。この指標を使って、「えこひいき」の高い部署、つまり標準偏差の小さい部署では、努力量が減ることを示した(努力量はサービス残業の量で捉えられている)。

しかし、主観的評価のバイアスは、真のパフォーマンスと実際の評価との差、もしくは評価者による評価と社員の自己評価との差として定義されるべきである。Engellandt and Riphahnの指標はそのいずれでもない。たとえば、同じ人が同じ業績考課を毎年得るのは、その人のパフォーマンスが変わっていないからかもしれない。また、業績考課が毎年変わることをもってバイアスがないというのも奇妙である。

したがって、われわれは考課の経年変化を使う代わりに、表4の考課推定式から得られる残差を使ってバイアスを計測する。

$$(\widehat{\text{残差}})_{it} = (\text{考課結果})_{it} - (\widehat{\text{考課結果}})_{it}$$

まず、従業員が考課結果の予測値をもって自己評価を行うとすると、残差は自己評価と実際に受けた評価との差を表す、つまり知覚されるバイアスを表すであろう。

次に、予測値が真のパフォーマンスを表すと

表5. 個人ごとの残差の時系列平均の規定要因, 店舗固定効果

評価は公正でない	-0.151594*** (0.0549268)
1日に部下と話す時間(分)	0.0024352* (0.0014293)
1日に上司と話す時間(分)	-0.0013039 (0.0007945)
アドバイスをする人・される人が以前も今もいる	-0.003374 (0.0424878)
以前も今も助け合って仕事を進めている	0.1393093*** (0.0522288)
定数項	-0.0679627* (0.0405248)
サンプル数	266
F値	5.58
自由度修正済決定係数	0.0885

表6. 店舗生産性の規定要因, 店舗固定効果

残差の店舗単位標準偏差(当期)	-165.8863 (116.1772)
残差の店舗単位標準偏差(1期前)	-76.540280 (106.1895)
営業スタッフの平均勤続年数	-4.692841 (8.939548)
営業スタッフの数(店舗のサイズ)	-111.4562*** (23.82305)
営業スタッフ1人当たり新人数割合	111.9989*** (25.13879)
課長の勤続年数	-2.167124 (2.986282)
定数項	3080.306*** (270.4649)
年ダミー	Yes
サンプル数	212
F値	19.05
自由度修正済決定係数	0.5156

考えれば、残差は評価者の嗜好に基づくバイアスを表すと考えられる。したがって、残差は Engellandt and Riphahn の指標よりも、よりいっそう主観的評価におけるバイアスの定義に近いと考えられる。しかし、残差は単なる統計的誤差である可能性もある。したがって、残差がバイアスを含んでいるかどうかを調べることが必要になる。先に述べたように、われわれは2000年に実施された人事制度改革に関する社員意識調査を2005年に行った。これを基に、以下のダミーを作成した。最初のダミーは、業績考課が公正でないと考えていた従業員のダミーであり、これは①自分の能力が評価に反映さ

れていない、②1つのミスで全体の評価が下がった、③自分の責任でない原因で評価が下がった、という設問のうち、少なくとも2つ選んでいる人のダミーである。

考課推定式で使われた673人のうち267人からの回答が得られた。以下の推定式を考える。(残差の個人時系列平均)<sub>i</sub>

$$= \beta_0 + \beta_1(\text{評価は公正でない})_i$$

$$+ \beta(\text{その他変数})_i + (\text{店舗固定効果}) + u_i$$

この式により、被評価者が残差をバイアスと認識しているかどうか分かる。つまり、これは知覚されるバイアスがあるかどうかの検証である。

表5がその推定結果である。「評価は公正でない」と考えている人は、残差が小さくなる(負の方向に向かう)傾向がある。つまり、評価が公正でないと考えている人は、実際に与えられた評価が自己評価を下回ることが多くなっている。この結果は、誤差項が知覚されるバイアスを表しているというわれわれの仮定と整合的である。そのほかに、部下と話す時間が長い人は、残差が大きい。これは部下の指導育成が業績考課で評価されているという仮説と整合的である。おそらく、われわれのメインの推定で補足されていなかった教育の影響が残差に残っていたのであろうと推測される。

以上の結果により、残差が従業員によってバイアスと認識されていることがわかった。次にバイアスが生産性に悪影響を与えるかどうかをチェックする。残差にバイアスが含まれていることがわかったので、店舗ごとにバイアスの度合いを測る最も自然な方法は、各店舗ごとに(また各年ごとに)残差の標準偏差を取ることである。標準偏差が高ければ高いほど、バイアスが大きいと解釈できる。そして、このバイアスの指標が高いとき、はたして店舗の生産性が落ちるであろうか。

推定結果は表6の通りである。被説明変数は営業スタッフ1人当たりの平均獲得利益で、分析単位は店舗である。残差の標準偏差のほかに、店舗生産性に影響を与えそうな変数をいくつか入れている。測定誤差の標準偏差が大きいと、

当年の生産に負の影響があるという結果が出ている。この結果は両側検定では有意ではないが、片側検定で7%有意であり、バイアスが生産性を落とすという仮説と整合的である。しかし、前年度の測定誤差の標準偏差が生産性に有意な影響を及ぼすという結果はみられない。

だが、他の解釈も可能である。単なる統計的誤差も、報酬と努力との関係を弱め、インセンティブを削ぐから生産性を落とすという解釈も可能である。バイアスが生産性を落としたのか、それとも単なる統計誤差が生産性を落としたのかという疑問が出てくる。たしかに、純粋な統計的誤差が生産性を落としたという可能性は否定できないが、われわれはバイアスが生産性を落としたと解釈する方が自然であると考え、理由は以下の通りである。

第1に、業績考課は、年度末に行うので、実際の結果がわかるのは、次の年度に入ってからである。したがって、純粋な統計的誤差がわかるのは次の年度に入ってからである。それゆえ、インセンティブ削減効果が単なる統計的誤差から生まれるとすると、ラグ変数の方が有意に効いてくるのではないかと考えられる。

第2に、ラグ変数は有意な影響がない。これに対し、誤差項が評価者のバイアスのみを含んでいると仮定すると、従業員は自分の仕事ぶりが過小評価されているかどうかは、日々上司と向き合っているうちにある程度分かるはずである。したがって、もしバイアスがあるとする、実際の考課結果をみる前にすでにインセンティブ削減効果が現れることが考えられる。以上から、当年の残差の分散のみに(限界的とはいえ)有意な結果がみられるということは、評価者のバイアスによって生産性が落ちていると考えるほうが自然であろう。

## 6. おわりに

現代の多くの職場では、客観的成果指標による従業員の評価が行われる例は少数であり、通常は主観的評価による査定がなされている。先行の理論研究は、主観的評価には、①マルチタスクがある場合のインセンティブ強化に役立つ、

②制御不能なリスクの軽減に利用することによりインセンティブの強化をもたらすというポジティブな側面があることを明らかにしてきた。しかしながら過去の実証研究は、主観的評価がこのような役割を担っているかどうかを検証してこなかった。また、先行の理論研究は、①主観的評価にはバイアスが伴い、②そのバイアスが生産性に悪影響を及ぼすというネガティブな側面があることも指摘している。この点に関する過去の実証研究も非常に不十分であった。こうした理論的予測の実証を行うことは、効果的なインセンティブの理解を深めるためにきわめて重要であると考えられる。本稿は大手自動車販売会社の人事データと社員意識調査データとの接合データを用い、上に述べた理論的予測の実証を行った。分析の結果、以下のことが明らかになった。

第1に、主観的評価はマルチタスクがある場合のインセンティブを強化するために使われているという仮説は部分的に支持された。マルチタスクとして、新人の指導育成と長期的顧客関係の構築を取り上げた。前者は業績考課結果に影響していることが明らかにされた一方で、後者の影響は確認できなかった。

第2に、制御不能なリスクにより業績が悪化した場合、主観的評価を使って評価基準を甘くすることにより、リスクが軽減されているという仮説が支持された。店舗に負のショックが発生した場合、考課結果に上乘せがなされるが、正のショックがあった場合には考課結果の削減はなされないことが明らかになった。

第3に、主観的評価においてバイアスが生じていることも見受けられた。考課結果の推定式からの残差がバイアスを含むか否かの検証として、社員意識調査データを用いて分析すると、「評価が公正ではない」という回答や部下と話す時間数などが残差と相関しており、バイアスがあることの傍証を得た。

第4に、バイアスの程度の強い店舗は生産性が低くなるという仮説にも一定の支持があった。店舗の生産性を店舗単位の残差の標準偏差(当期および前期)などの変数と回帰させると、店

舗単位の残差の標準偏差(前期)はまったく有意ではないのに対して、店舗単位の残差の標準偏差(当期)は限界的な水準ではあるが負の効果が確認された。

(国際大学大学院国際関係学研究所・一橋大学経済研究所・青山学院大学国際政治経済学部)

## 注

\* 本稿の研究に際し、都留は日本経済研究奨励財団奨励金(2008年度)の研究支援を受けた。上原は科学研究費補助金・基盤研究(C)(22530216)および経済研究所内戦略経費(2009年度)の研究支援を得た。本稿の草稿は、経済研究所定例研究会(2011年6月15日)で読まれ、討論者の久保克行教授(早稲田大学)はじめ、参加者各位のコメントにより改善できた。また、大湾秀雄教授(東京大学)、町北朋洋研究員(アジア経済研究所)ならびに黒崎卓・岩崎一郎両教授(一橋大学)の助言も有益であった。以上の機関、個人に深甚の謝意を表す。

1) 本社にはサービス部門もありエンジニア層も存在するが、本稿の分析対象外なので捨象する。

2) 同社の人事制度改革後の基本給は「職務給」と「功績給」とからなる。功績給とは、制度改革前の職能資格制度における各人の「職能給」の到達水準を反映させたものであり、激変緩和措置といえる。

3) この聞き取り調査は、2006年5月に、ある店長に対して実施した。

## 参 考 文 献

- Baker, George, Gibbons, Robert and Murphy, Kevin J. (1994) "Subjective Performance Measures in Optimal Incentive Contracts," *Quarterly Journal of Economics*, Vol. 109, No. 4, pp. 1125-1156.
- Ederhof, Merle (2010) "Discretion in Bonus Plans," *Accounting Review*, Vol. 85, No. 6, pp. 1921-1949.
- Elvira, Marta and Town, Robert (2001) "The Effects of Race and Worker Productivity on Performance Evaluations," *Industrial Relations*, Vol. 40, No. 4, pp. 571-590.
- Engellandt, Axel and Riphahn, Regina T. (2011) "Evidence on Incentive Effects of Subjective Performance Evaluations," *Industrial and Labor Relations Review*, Vol. 64, No. 2, pp. 241-257.
- Gibbons, Robert and Murphy, Kevin J. (1990) "Relative Performance Evaluation for Chief Executive Officers," *Industrial and Labor Relations Review*, Vol. 43, No. 3, pp. 30S-51S.
- Gibbs, Michael, Merchant, Kenneth A., Van der Stede, Wim A. and Vargus, Mark E. (2004) "Determinants and Effects of Subjectivity in Incentives," *Accounting Review*, Vol. 79, No. 2, pp. 409-436.
- Holmstrom, Bengt (1979) "Moral Hazard and Observability," *Bell Journal of Economics*, Vol. 10, No. 1, pp. 74-91.
- Holmstrom, Bengt and Milgrom, Paul (1991) "Multi-task Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, Vol. 7, Special Issue, pp. 24-52.
- Lazear, Edward P. (2000) "Performance Pay and Productivity," *American Economic Review*, Vol. 90, No. 5, pp. 1346-1361.
- MacLeod, W. Bentley (2003) "Optimal Contracting with Subjective Evaluation," *American Economic Review*, Vol. 93, No. 1, pp. 216-240.
- Merchant, Kenneth A. (1987) "How and Why Firms Disregard the Controllability Principle," in William J. Bruns, Jr. and Robert S. Kaplan (eds.) *Accounting & Management: Field Study Perspectives*, Chap. 12, Boston, Massachusetts: Harvard Business School Press, pp. 316-338.
- Prendergast, Canice (1999) "The Provision of Incentives in Firms," *Journal of Economic Literature*, Vol. 37, No. 1, pp. 7-63.
- Prendergast, Canice and Topel, Robert H. (1996) "Favoritism in Organizations," *Journal of Political Economy*, Vol. 104, No. 5, pp. 958-978.
- Tsuru, Tsuyoshi (2008) "Transforming Incentives: Analysis of Personnel and Employee Output Data in a Large Japanese Auto Sales Firm," *Hitotsubashi Journal of Economics*, Vol. 49, No. 2, pp. 109-132.