

解説 ESP フォーキャスト調査の評価について

ESP フォーキャスト調査委員会委員 河越正明*

(社) 経済企画協会発行「ESP」2006年9月号 pp.60-64

1 はじめに

本稿の目的は、9月5日に(社)経済企画協会から発表された「ESP フォーキャスト調査の評価について」に関し、その評価方法を解説することである。この評価方法については、ESP フォーキャスト調査委員会(委員長:小峰隆夫法政大学教授)において議論され了承されたものである*¹。さらにESP フォーキャスト調査(以下、ESPFと呼ぶ。)の参加予測者の評価作業の延長として、ESPFが計算している予測総平均(いわゆるコンセンサス)のパフォーマンスを自己点検し、その総合評価がベスト10に入る精度であることを示す。

2 評価の概要

予測者 i が指標 k について時点 j に作成した t 期の予測値を \hat{y}_{ijt}^k とし、その実現値を y_t^k とすれば、その予測誤差は

$$e_{ijt}^k \equiv y_t^k - \hat{y}_{ijt}^k \quad (1)$$

とあらわすことができる。

評価対象として、調査項目の中から(a)年度予測として実質GDP成長率、(b)四半期予測として実質GDP成長率(季調済前期比年率)、CPI上昇率(前年同期比)及び失業率(季調値)の計4項目を取り上げる。予測時点は、(a)については、後述するデータの利用可能性から2004年度については2004年4月以降、2005年度については2005年1月以降とした。(b)については、公表直前から原則6ヶ月分とした。予測対象期間は、(a)については2004,05年度の2年度分、(b)は2004年4-6月期から2006年1-3月期までである。

予測の評価を具体的に数式で示そう。実質GDP成長率の年度値については、年度別に

$$RMSE_{it}^{GDPFY} = \left(\frac{1}{N_t} \sum_j (w_{jt}^{GDPFY} e_{ijt}^{GDPFY})^2 \right)^{1/2}, \quad t = FY2004, 05; \quad i = 1, \dots, 37 \quad (2)$$

のようにRMSE(Root Mean Square Error, 平均予測誤差自乗和の平方根)を計算する。四半期予測についても、

$$RMSE_{it}^k = \left[\frac{1}{4} \sum_{Q \in t} \left(\frac{1}{N_Q} \sum_j (w_{jQ}^k e_{ijQ}^k)^2 \right) \right]^{1/2}, \quad t = FY2004, 05; \quad k = GDPQ, CPIQ, UNQ \quad (3)$$

* 内閣府経済社会総合研究所特別研究員

*¹ 紙幅の制約から評価方法の詳細は割愛せざるをえないが、詳細についてはご興味がある場合は、経済企画協会までお問い合わせ下さい。

のように、年度別に計算する。式(2),(3)の w_{jt}^k は予測時点の違いを反映したウェイトである(後述)。また、調査対象月数 N_t 又は N_Q で除して、対象月数の違いを調整している。次にこれらの4指標を年度別に、

$$TTL_t = \frac{1}{4} \sum_k \left(50 + 10 \times \frac{RMSE_t^k - \text{Mean}(RMSE_t^k)}{\text{Std}(RMSE_t^k)} \right), \quad t = FY2004,05 \quad (4)$$

を計算し、それぞれの指標のいわば偏差値(平均50,標準偏差10)の平均値をもって総合評価とする。そしてこの年度毎に求めた平均偏差値を用いて、各予測者のランキングを行った。ただし、通常の成績の偏差値と異なり、偏差値が小さいほど成績が良い。以上の計算を行う上での詳細を以下で述べる。

実績値 式(1)の計算にあたり、実績値が後で改訂されるため、どの実績値を使うか実は自明ではない。そこで、まず一番最初に公表される値を「実績値」として用いることとした。さらに、GDP統計については、固定価格方式から連鎖価格方式への移行(2004年12月)、基準年の1995年から2000年への変更(2005年12月)といった大きなデータ系列の変更があって予測値との対応関係が複雑であるため、実質GDP成長率の年度値について表1のように整理した。

(表1 予測値と実績値の対応： 実質GDP伸び率の年度値)

予測値 予測値については、調査対象の予測者のうち無回答数が多い者を2004,05年度それぞれ3つ除いた結果、両年度とも37機関(人)となった*2。難しいのがこれら37予測者の回答で散見される無回答の扱いである。無回答の回数が多い者が成績良好とならないためにはペナルティを課すことが望ましいが、それは一体どの程度とすべきか。ここでは、当該月の有効回答の $(w_{jt}^k e_{ijt}^k)^2$ の平均を計算し、その1.5倍の数値で代用した。

予測時点の違い ESPFでは毎月調査しているので、一つの実績値に対して予測時点の異なる複数の予測値が存在する。そして、予測時点と公表時点が近いほど予測者が直面する不確実性は小さく、精度の高い予測値となるはずであり、この点を調整する必要がある。例えば2005年度の実質GDP成長率を同じ0.1ポイントだけ外した2人の予測者がいた場合に、2005年1月時点(2004年7-9月期の実績まで利用可能)の予測で外した予測者と、2006年5月時点(2005年10-12月期の実績まで利用可能)で外した予測者を同じように扱うのは適当ではない。このウェイトは、次節で述べるように推計式を用いて求めた。

RMSE 予測の精度を評価するためにMSE(Mean Squared Error)を計算し、もとのデータと単位をあわせるために平方根をとったRMSEを指標として用いている。ただし、これ以外にも種々の指標がありえ*3、異なる指標を用いれば違った評価となる可能性がある点には注意が必要である。理論的には、指標は予測者の損失関数(またはコスト関数)次第であり、RMSE(またはMSE)を用いるのは、暗黙のうちに損失関数が、予測誤差の2次関数であることを仮定していることになる。

3 ウェイト計算の詳細

自己回帰モデルの活用 年度の実質経済成長率並びに四半期のCPI上昇率及び失業率については、実績値 y_t^k が次のような自己回帰過程に従って発生していると想定し、予測作成時点が異なる予測誤差に与えるウェイト

*2 ESPFでは、必ずしも「機関」として回答を得ているわけではなく、中には個人として回答している場合もある点に注意されたい。

*3 経済予測についてのテキスト、例えばDiebold(1998)にはMSE, RMSEのほか、予測誤差の絶対値に基づくもの(MAD)、誤差率に基づくもの(MAPD, MSPE, RMSPE)等が挙げられている。

トを計算した。

$$y_t^k = \alpha^k + \sum_m \beta_m^k y_{t-m}^k + u_t^k, \quad u_t^k \sim iidN(0, \sigma_k^2) \quad (5)$$

2005年度の実質経済成長率を例に考えると、年度の成長率は毎四半期の前年同期比の平均として計算でき、

$$y_{2005}^{GDPFY} \approx \frac{1}{4}(y_{2005Q2} + y_{2005Q3} + y_{2005Q4} + y_{2006Q1}) \quad (6)$$

と表せる(右辺の上添え字の GDPQ は簡単化のために省略)。予測時点 j で利用可能な実績値に基づく条件付き期待値が、予測値である。そこから求められる予測誤差は、 u_t^k の線形結合で表され、その係数は $\beta_m^k (m=1, 2, \dots)$ の関数である。その結果、予測時点 j の予測誤差の分散は一般的に、

$$Var(e_{j,2005}^{GDPFY}) = \Phi_j(\beta) \sigma_{GDPQ}^2 \quad (7)$$

と表せる。式(5)を推計して式(7)の β を $\hat{\beta}$ で置き換え、ウェイトを

$$w_{j,2005}^{GDPFY} \equiv (\Phi_j(\hat{\beta}))^{-1/2} \quad (8)$$

と定義する。これを予測誤差に乗じることで、予測時点にかかわらず予測誤差の分散を一定化することが可能となる。

式(5)を推計すると $m=2$ となり、その推計結果に基づき計算したウェイトは表2の通りである。また、以上と同様のやり方で CPI 上昇率及び失業率の四半期予測におけるウェイトを計算できる。ただし四半期値であるので、式(6)では月次データの3ヶ月分の平均をとることになる。それぞれの推計結果から、式(8)と同様の計算を行い、その結果を4-6月期を例に示せば、表3の通りである。

(表2 年度成長率に用いる予測時点別ウェイト)

(表3 CPI 上昇率及び失業率(四半期)に用いるウェイト: 4-6月期の例)

若干の拡張 四半期の実質 GDP 成長率については、式(5)の右辺に他の説明変数 $x_{t-j} (j=0, 1, \dots)$ も加えたモデルを考えた。こうした若干の拡張形を用いるのは、自己回帰モデルを用いてウェイトを計算すると、ウェイトの減少スピードが極めて緩やかで、予測実務の実感からすると緩やか過ぎるという感が否めないからである*4。GDP 統計は二次統計という性格上、同時期の結果を示すさまざまな統計が出た後に公表される。予測の実務においては、こうした種々の統計を使って、四半期の実質 GDP 成長率を推計しているのが実態だと推測される。つまり時間の経過とともに、他の指標が利用可能となることから、当該四半期の結果についての不確実性は減少していくが、自己ラグのみを説明変数としている自己回帰モデルでは、使用する情報に対する制約が厳しすぎるために、こうした点を考慮できない。

そこで具体的には以下のように、 t 期に観察可能である x_t を y_t の予測に使うことができると想定する(式(9)の上添え字 GDPQ は簡単化のために省略)。

$$y_t = \alpha_y + \sum_{m=1}^M \beta_m y_{t-m} + \sum_{n=0}^N \zeta_n x_{t-n} + u_t, \quad u_t \sim iidN(0, \sigma_y^2) \quad (9)$$

$$x_t = \alpha_x + \sum_p \eta_p x_{t-p} + v_t, \quad v_t \sim iidN(0, \sigma_x^2) \quad (10)$$

*4 例えば4-6月期の予測については、8月時点のウェイト1.0から半年前の2月時点でのウェイトは0.962である。

より具体的に4-6月期の実質GDP成長率の予測する場合について検討する。 x_t として鉱工業生産指数(季調済前期比伸び率)を用いることを考えると、予測誤差の分散は、

$$\begin{aligned} \text{Var}(e_{j,2005Q2}^{GDPQ}) &= \Psi_j(\beta)\sigma_{GDPQ}^2 + \Theta_j(\beta, \zeta, \eta)\sigma_{IIPQ}^2 \\ &= \Omega_j(\beta, \zeta, \eta, r)\sigma_{GDPQ}^2, \quad r \equiv \sigma_{IIPQ}^2/\sigma_{GDPQ}^2 \end{aligned} \quad (11)$$

と計算される。式(9)及び(10)の推計結果から $\Omega_j(\hat{\beta}, \hat{\zeta}, \hat{\eta}, \hat{r})$ を計算し、 $w_{jt}^{GDPQ} = [\Omega_j(\hat{\beta}, \hat{\zeta}, \hat{\eta}, \hat{r})]^{-1/2}$ を求める。この結果が表4である。2,5,8月のウェイトは推計結果から直接から与えられ、その間の月のウェイトは一定の比率で補間している。1995年基準の連鎖価格、固定価格の系列の場合も同様に計算できる。

(表4 四半期実質GDP成長率に用いるウェイト: 4-6月期の例)

留意点 以上のようなウェイトの計算方法は、次のような点から必ずしも正確ではない点に留意が必要である。まず、推計に用いたデータは、データの事後的な改定によって、毎月の調査時点で予測者が予測を作成するのに用いたデータとは異なっていることである。二番目に、上の方法はパラメータの推計誤差、さらには定式化のミスに起因する予測誤差を無視していることである。三番目に、予測時点の違いから生じるウェイトの違いから明らかなように、四半期予測におけるCPI上昇率や失業率では、公表直前月の誤差が極めて重要である。しかし、ESPFでは予測値を小数点第1位までしか求めていないために、大まかな計算となってしまっている。

4 コンセンサスの評価

今回行った各予測者の評価作業の延長として、同様のやり方で、ESPFで毎回算出している「予測総平均」、いわゆるコンセンサス予測のパフォーマンスを評価してみよう。その結果は表5が示す通り、個別の指標では概ね10位前後であり、総合評価では2004年度5位、2005年度8位となる。なお、評価対象期間中ずっと強気派(または弱気派)であったならどのようなパフォーマンスとなるかは、同表の「上位平均」(または「下位平均」)の数字を見ることでわかる。

(表5 コンセンサスの評価)

注目すべきは、コンセンサス予測が割合よい成績であることである。なぜ、37者中18位程度より良い成績を収めることができるのであろうか。その理由は、直観的には、どこにも完全な予測はないということに求められる。各予測は実態のある一面を捉えているが、見逃している点もあるに違いない。種々の予測を集めてきて平均をとるということは、個々の予測が見逃しているリスクの分散化を図ることに他ならず、その有効性はポートフォリオ理論の分散投資のアナロジーから考えることが可能である。コンセンサスを計算する際のウェイトの付け方については、理論的に最適ウェイトを求めることができるが、既存の研究によれば*5、最適ウェイトによるコンセンサスが、単純な算術平均より優れたパフォーマンスとなるとは限らず、むしろ後の方が優れている場合も少なくない。

以上から私見であるが、予測総平均とは決して単なる平均ではなく、その総合評価がベスト10に入るそれ自体が一つの優れた予測である。毎月40弱の予測者の予測を集約して公表するということはそれ自体大きな意義のある事業であるけれども、さらに調査結果として出される「予測総平均」にも積極的な意義を認めることができると考えている。

*5 Clemen (1989)などを参照せよ。

参考文献

Clemen, Robert T. (1989) "Combining Forecasts: A Review and Annotated Bibliography." *International Journal of Forecasting*. Vol. 5. pp. 559–583.

Diebold, Francis X. (1998) *Elements of Forecasting*. Cincinnati, Ohio: South-Western College Publishing.

表1 予測値と実績値の対応：実質GDP伸び率の年度値

予測時点	予測している系列	2004年度		2005年度	
		実績	実績	実績	対応
2004年4-12月	1995年基準1995年固定価格	○	n.a. (注1)		評価せず
2005年1-12月	1995年基準連鎖価格(2000年参照価格)	○ (1-5月)	n.a. (注2)		(A)で評価
2006年1月以降	2000年基準連鎖価格(2000年参照価格)(A)	-	○		

注1: 2005年4-6月期で実績値の公表が終了

注2: 2005年7-9月期で実績値の公表が終了

表2 年度成長率に用いる予測時点別ウェイト

予測時点	2000年基準 連鎖価格	1995年基準 連鎖価格	1995年基準 固定価格
t年3-5月	1.000	1.000	1.000
t-1年12月～t年2月	0.433	0.443	0.482
t-1年9-11月	0.270	0.280	0.309
t-1年6-8月	0.200	0.209	0.225
t-1年3-5月	0.177	0.186	0.200
t-2年12月～t-1年2月	0.171	0.180	0.191

表3 CPI上昇率及び失業率(四半期)に用いるウェイト: 4-6月期の例

予測時点	CPI上昇率	失業率
7月	1.000	1.000
6月	0.495	0.480
5月	0.303	0.297
4月	0.243	0.241
3月	0.205	0.209
2月	0.180	0.189

表4 四半期実質GDP成長率に用いるウェイト

予測時点	2000年基準 連鎖価格	1995年基準 連鎖価格	1995年基準 固定価格
8月	1.000	1.000	1.000
7月	0.957	0.955	0.952
6月	0.915	0.913	0.907
5月	0.876	0.872	0.863
4月	0.837	0.831	0.809
3月	0.800	0.792	0.759
2月	0.765	0.755	0.712

表5 総平均及び上位・下位平均の評価

	2004年度			2005年度		
	上位平均	総平均	下位平均	上位平均	総平均	下位平均
年度予測 実質GDP成長率	35	10	14	6	16	34
偏差値平均	60.63	44.57	45.23	41.60	46.92	56.32
四半期予測 実質GDP成長率	26	8	27	2	14	37
偏差値平均	52.47	41.70	55.63	38.76	45.15	66.81
CPI上昇率	34	16	17	6	6	35
偏差値平均	62.15	44.53	46.19	40.47	42.10	66.82
失業率	35	11	15	27	13	31
偏差値平均	63.75	43.71	45.26	51.69	44.80	55.37
総合評価	36	5	15	5	8	35
偏差値平均(*)	59.75	43.63	48.05	43.13	44.74	61.33

*: ベスト5の優秀フォーカスターの平均は2004年度41.54、2005年度41.81である。