

基礎研 レター

経済予測はどのくらいはずれるのか(3) ～四半期別GDP速報の予測精度(個別機 関 vs コンセンサス予測)～

経済研究部 経済調査部長 斎藤 太郎
(03)3512-1836 tsaito@nli-research.co.jp

1—GDP速報の予測精度

本シリーズの[第1回](#)、[第2回](#)では主として年度ベースの実質GDP成長率の予測を扱ったが、今回は四半期ベースのGDP速報(QE:Quarterly Estimates)の予測精度を検証する。通常の経済見通しでは予測時点よりも先の経済変数を予測するが、四半期別GDP速報の場合はすでに公表されている様々な経済統計(多くは月次統計)を、内閣府が公表しているGDP速報の推計方法をもとに加工することにより、過去の実績値を予測する。たとえば、多くの民間調査機関が2019年10-12月期のGDP速報の直前予測を発表するのは2020年1月末になる(速報値は2/17の公表予定)。純粋な予測というよりは、内閣府が行っている推計方法をどれだけ正確に再現できるかが予測の優劣を決める。予測の結果がすぐに判明することも特徴のひとつである。年度見通しの結果が判明するのは、予測値を作成してから少なくとも1年以上先となるが、GDP速報は予測値を作成してから結果が公表されるまでの期間は2週間程度である(直前予測の場合)。

予測誤差の尺度としては、前回までと同様に平均誤差、平均絶対誤差の2つを用いる。

ある指標について、 t 期の予測誤差 e_t は、予測値を f_t 、実績値を y_t とすると、

$$e_t = y_t - f_t$$

と表され、

$$\text{平均誤差} = \frac{1}{k} \sum_{t=1}^k (y_t - f_t)$$

$$\text{平均絶対誤差} = \frac{1}{k} \sum_{t=1}^k |y_t - f_t|$$

となる。

年度ベースの予測は1980年度～2018年度の約30年間を扱ったが、民間調査機関によるGDP速報予測の歴史は比較的新しい。多くの民間調査機関がGDP速報の予測を行い、通信各社がそれを集計

するようになったのは1990年代半ば以降である。

ここでは、2000年1-3月期から2019年7-9月期までの約20年間（79四半期）の実質GDP（1次速報）の実績値（前期比年率）、民間調査機関の予測値を用いて予測精度を検証した。予測値はブルームバーグ、QUICKがGDP速報公表の直前に集計した各調査機関の予測値の平均（コンセンサス）である。

79四半期の平均絶対誤差は0.99%となった（図表1）。先行き1年間を予測する年度の実質GDP成長率の平均絶対誤差が1.25%であったこと、推計に用いられる基礎統計がほぼ揃った段階での予測であることを踏まえれば、約1%の誤差は大きいという印象があるかもしれない。

平均誤差（実績値－予測値）は0.19%となった。このうち実績値が予測値から上振れたことが45回、下振れたことが33回、予測値が実績値と完全に一致したことが1回である。上振れの割合は57.0%となっており、民間調査機関の予測値はやや過小となる傾向がある。

これを景気拡張期と景気後退期に分けてみると、景気拡張期の平均誤差は0.30%、上振れの割合は61.2%（下振れが38.8%）、景気後退期の平均誤差は▲0.38%、下振れの割合は66.7%（上振れが33.3%）となっている。今回の検証期間では景気後退期が12四半期と少ないことには留意が必要だが、景気拡張期には実績値が予測値から上振れ、景気後退期には下振れする傾向がある。

図表1 GDP速報の予測誤差(コンセンサス)

	(前期比年率, %)		
	全期間(00/1Q~19/3Q)	景気拡張期	景気後退期
平均成長率(1次速報・実績)	1.16	1.89	▲2.80
予測誤差(平均絶対誤差)	0.99	1.07	0.57
予測誤差(平均誤差)	0.19	0.30	▲0.38
上振れ割合	57.0%	61.2%	33.3%
下振れ割合	41.8%	38.8%	66.7%

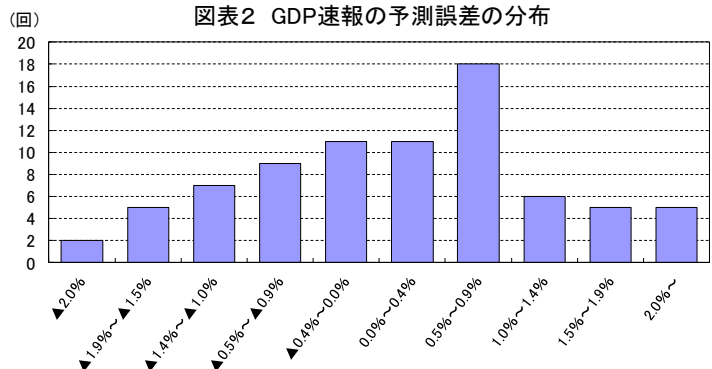
(注)コンセンサスはブルームバーグ、QUICK調査による

前回、年度ベースの成長率の予測値は景気拡張期には過小、景気後退期には過大となりやすいことを確認した。四半期ベースのGDP速報の予測は、年度の予測と異なり経済の先行きを予想するものではない。あくまでも公表済みの経済統計を加工することによる推計であるため、景気に対する定性判断などは予測値に影響しないはずであるが、実際には景気の局面と予測誤差の方向には一定の関係が見られる。

79四半期の予測誤差の分布を0.5%刻みで見ると、0.5%～0.9%が79回のうち18回、全体の2割強（22.8%）と最も多くなっている（図表2）。予測誤差の絶対値が0.5%未満であれば、実績値がほぼ事前の市場予想（コンセンサス）通りだったと判断してもよいと考えられるが、そういうケースは22回と全体の3割弱（27.8%）にとどまっており、逆に1割弱（8.9%）は予測誤差の絶対値が2%を超えている。

2000年1-3月期以降で実績値が予測値から最も大きく上振れたのは2002年10-12月期の3.5%で、事前予測のコンセンサスは前期比年率▲1.5%のマイナス成長だったが、実績値（1次速報）は前期比年率2.0%と明確なプラス成長となった。逆に、最も大きく下振れたのは2014年4月の消費税率引き上げ（5%→8%）から2四半期目に当た

図表2 GDP速報の予測誤差の分布



(注)予測誤差(前期比年率)＝実績値(1次速報)－予測値(コンセンサス)
対象期間：2000年1-3月期～2019年7-9月期(79四半期)

る 2014 年 7-9 月期の▲3.6%である。事前予測のコンセンサスは前期比年率 2.0%のプラス成長だったが、実績値（1次速報）は同▲1.6%のマイナス成長となった¹。2014 年 11 月に公表されたこの結果を受けて、安倍首相は 2015 年 9 月に予定されていた消費税率の引き上げ（8→10%）を 2017 年 4 月に延期することを表明した（その後、2016 年 11 月に 2019 年 10 月への再延期を決定）。

2—需要項目別の予測誤差

次に、需要項目別の予測誤差を見てみよう。GDP速報の予測は、実質GDPについては前期比、前期比年率の両方が集計されているが、需要項目別では前期比のみの集計となっている。そこで前期比の平均絶対誤差を見ると、公的固定資本形成が 1.22%と最も大きく、設備投資が 0.92%、住宅投資が 0.89%とそれに続いている（図表 3）。予測誤差が大きいのはいずれも投資関連で、実績値の変動（標準偏差）が大きい点が共通している。したがって、単純な予測誤差で民間消費など比較的安定的な動きをするものとその精度を比較することはできない。

公的固定資本形成の予測誤差の大きさは、1次速報では推計の主要な基礎統計である「建設総合統計」の 3 ヶ月目が公表されていないことに原因があると考えられる。「建設総合統計」の 3 ヶ月目は別の統計を用いた回帰式で推計することによって補外しているが、回帰式自体は公表されていないため、予測することが難しい。また、設備投資は推計方法が複雑であることに加え、1次速報の時点では需要側推計に用いられる法人企業統計が未公表（需要側推計値は内閣府による仮置き値が用いられる）であることも予測が難しい理由となっている。

公的固定資本形成は予測誤差が大きいですが、実質GDPに占めるウェイトが最近では 5%程度と小さいため、予測誤差がGDP全体に与える影響はそれほど大きくない（前期比年率・寄与度で 0.29%）。一方、民間消費は前期比の平均絶対誤差は 0.28%と比較的小さいがウェイトが高いため、寄与度では 0.64%（年率換算）と実質GDP成長率の予測誤差に与える影響は大きい。需要項目別の予測誤差を寄与度ベース（年率換算）で大きい順に並べると、民間消費（0.64%）、設備投資（0.56%）、民間在庫（0.53%）となる。

なお、需要項目別の平均絶対誤差を合計すると 2.98%となり、実質GDPの平均絶対誤差 0.99%を大きく上回る。ある期のGDP速報の予測値は需要項目毎に過小予測、過大予測が混在しており、プラスとマイナスが打ち消されることでGDP全体の予測誤差が小さくなるが多いためである。需要項目別の予測値が大きく外れても、逆方向に外れたためにGDP全体の予測誤差が小さくなる、逆

図表 3 需要項目別の予測誤差（平均絶対誤差）

	前期比	〈寄与度〉	〈年率寄与度〉
実質GDP	0.25	〈0.25〉	〈0.99〉
民間消費	0.28	〈0.16〉	〈0.64〉
住宅投資	0.89	〈0.03〉	〈0.11〉
設備投資	0.92	〈0.14〉	〈0.56〉
民間在庫変動	—	〈0.13〉	〈0.53〉
政府消費	0.24	〈0.04〉	〈0.14〉
公的固定資本形成	1.22	〈0.07〉	〈0.29〉
外需	—	〈0.11〉	〈0.45〉
輸出	0.64	〈0.09〉	〈0.38〉
輸入	0.68	〈0.08〉	〈0.33〉
需要項目の合計	—	〈0.74〉	〈2.98〉

（注）寄与度は公表時の需要項目別のウェイトをもとに計算

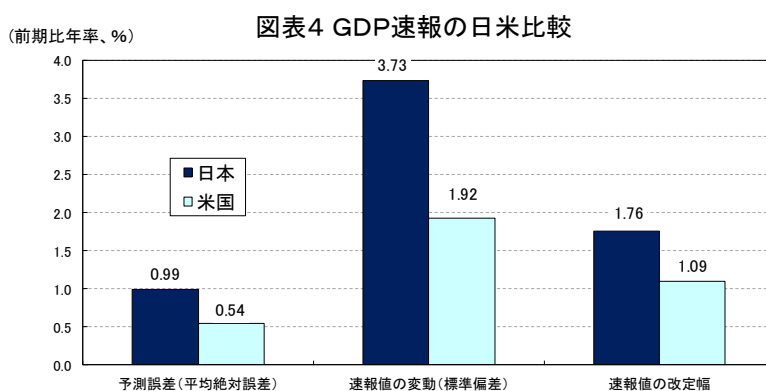
¹ 2014 年 7-9 月期の実質GDP成長率はその後上方修正され、現在は前期比年率 0.4%のプラス成長となっている

に需要項目毎の予測誤差は比較的小さかったが、同じ方向に外れたためGDP全体の予測値が大きく外れることも少なくない。

3—GDP速報予測の日米比較

日本のGDP速報の予測誤差が比較的大きい一方、米国ではGDP速報が事前の市場予想（コンセンサス）からあまり大きく外れない印象が強い。このことを実際のデータで確かめてみると（コンセンサスはブルームバーグ集計による）、2000年1-3月期から2019年7-9月期までの平均絶対誤差は0.54%となり、日本の予測誤差の約半分であった（図表4）。

ただし、予測誤差の大きさだけで日本のエコノミストの予測能力が米国に劣っていると結論づけることはできない。日本の予測誤差が相対的に大きくなる理由のひとつとして、米国に比べて実績値の変動が大きいことが挙げられる。速報段階における実質GDP成長率（前期比年率）の標準偏差を比較すると、日本の3.73%に対して米国は1.92%となる。日本の成長率は四半期毎の振れがかなり大きく、このことが予測を難しくしている一因とも言えるだろう。いったん発表された速報値がその後の改定で大きく修正されることが多いのも日本の特徴である。最初に公表された1次速報から最新のデータまでの改定幅を平均すると、米国の1.09%に対して日本は1.76%と大きい。



(注)速報値の改定幅は速報値と最新値との差(絶対値)を平均したものの2000年1-3月期~2019年7-9月期の平均。ただし、速報値の改定幅は2000年1-3月期~2019年1-3月期の平均
(資料)内閣府、米国商務省

米国のGDP統計のほうが総じてユーザーフレンドリーとなっていることも予測精度の高さにつながっていると考えられる。たとえば、米国のGDP統計は需要項目別の推計方法が詳細に解説されていることに加え、実績値についても個人消費、設備投資といった需要項目別だけでなくその内訳についてもかなり細かく公表されている(たとえば、個人消費の内訳：家具、自動車、ガソリン、食料、衣料、ヘルスケア、娯楽サービスなど、設備投資の内訳：建築物、機械、知的財産生産物など)。このため、予測者は需要項目の内訳をかなり細かい部分までブレイクダウンした上で予測値を作成することが可能である。これに対して、日本でもGDP速報の推計方法のマニュアルは内閣府から公表されているが、推計の考え方は示されているにもかかわらず具体的な推計方法まで示されていない、あるいは民間では入手できないデータによって推計されている部分も多い。

また、実績値については、以前は全く公表されていなかった需要項目の内訳もある程度公表されるようになったが、米国に比べるとまだ見劣りがする。たとえば、家計消費の内訳は耐久財、半耐久財、非耐久財、サービスの区分に限られ、総固定資本形成については、住宅、その他の建物・構築物、輸送用機械、その他の機械設備等、知的財産生産物に分かれているが、民間、公的の区別がないため、利用しにくい。

4—コンセンサス予測が優秀となる理由

ここまで、GDP速報の予測精度をコンセンサス（各機関の予測値平均）でみてきたが、コンセンサス予測は個別機関の予測と比べて相対的に優秀な成績となることが知られている。コンセンサス予測のパフォーマンスは、直感的には予測機関全体の中位程度になるように思われるかもしれないが、実はコンセンサス予測の誤差が各機関の平均的な予測誤差よりも小さくなることは以下の数式から明らかである。

予測機関を*i*とすると、

$$t \text{ 期におけるコンセンサス予測の絶対誤差} = \left| y_t - \frac{1}{n} \sum_{i=1}^n f_{it} \right| \cdots \textcircled{1}$$

$$\text{一方、各機関予測の絶対誤差の平均} = \frac{1}{n} \sum_{i=1}^n |y_t - f_{it}| \cdots \textcircled{2}$$

で表される。

①式を変形すると、

$$\frac{1}{n} \left| n \cdot y_t - \sum_{i=1}^n f_{it} \right| = \frac{1}{n} |(y_t - f_{i1}) + (y_t - f_{i2}) + (y_t - f_{i3}) + \cdots + (y_t - f_{in})| \cdots \textcircled{1}'$$

②式を変形すると、

$$\frac{1}{n} (|y_t - f_{i1}| + |y_t - f_{i2}| + |y_t - f_{i3}| + \cdots + |y_t - f_{in}|) \cdots \textcircled{2}'$$

となる。

①'式右辺と②'式右辺の違いは各項を合計したものに絶対値がついているか、各項それぞれに絶対値がついているかである。各項が全て同じ符号の場合、①' = ②' となり、各項に異なる符号が存在する場合、①' < ②' となる。つまり、①' ≤ ②' が必ず成立する。

具体的には、実績値が市場予測のレンジから外れた場合（市場予測の上限値、下限値の場合も含む）、①'式右辺の各項は全て同じ符号となり、①'式は②'式と一致する。すなわち、コンセンサス予測の誤差と個別機関の予測誤差の平均は等しくなる。実績値が市場予測のレンジ内となった場合には、①'式右辺の各項はプラスとマイナスが混在し、これらが互いに打ち消し合うことにより、①'式は②'式よりも小さくなる。すなわち、コンセンサス予測の誤差は個別機関の予測誤差の平均よりも小さくなる。

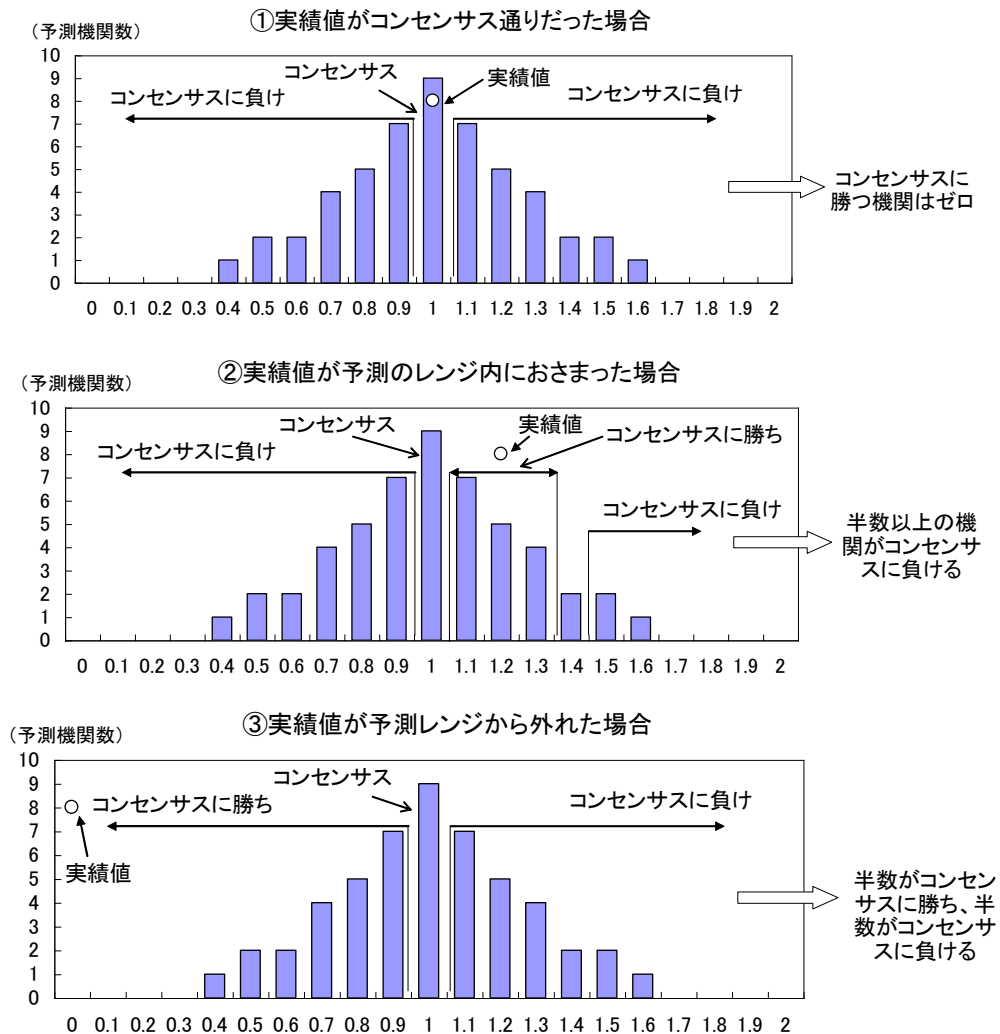
このことを概念図で示すと図表5のようになる²。まず、実績値がコンセンサス予測通りだった場合（図表5の①）、コンセンサス予測と同じ予測値を出した機関は引き分け、それ以外の機関は負けとなり、コンセンサス予測に勝つ機関はゼロとなる。次に、実績値が予測値のレンジ内におさまった場合

² ここでは「コンセンサス=平均値=中央値」のケースで説明する

(図表5の②)、半数以上の機関がコンセンサス予測に負ける。最後に、実績値が予測値のレンジから外れた場合(図表5の③)、半数の機関がコンセンサス予測に勝ち、半数の機関がコンセンサス予測に負ける。どのケースにおいてもコンセンサス予測が半数以上の機関に負けることはない。

結局、コンセンサス予測の誤差が個別機関の予測誤差の平均よりも大きくなることは原理的にありえない。このような試行を繰り返すことにより、期間が長くなるほどコンセンサス予測は相対的に優秀な成績をおさめることになるのである。

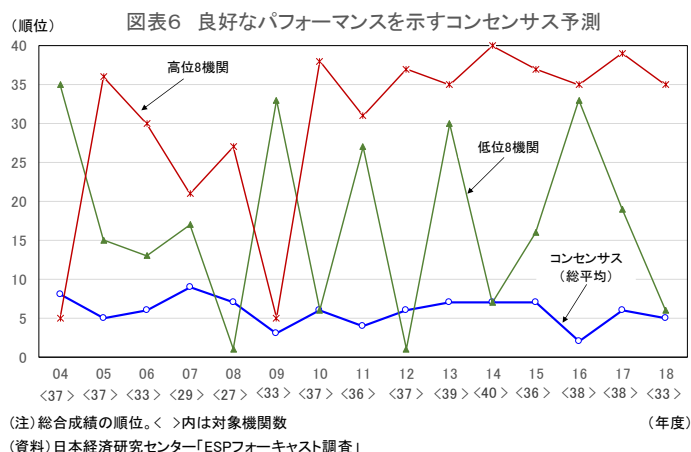
図表5 コンセンサス予測が優秀となる理由(概念図)



このことは2004年度から実施されている日本経済研究センターの「ESPフォーキャスト調査」³でも実証されている。ESPフォーキャスト調査は年度毎に、実質GDP、消費者物価、失業率などの予測誤差をもとにパフォーマンス評価を行い、優秀フォーキャスターを選出しているが、それと同時に、総平均、低位8機関、高位8機関の順位を公表している。

³ 2012年度までは経済企画協会が実施していた

2004～2018年度の15年間、総平均（コンセンサス）の成績は27～40機関（15年間の平均は35機関）の中で上位10位以内（平均6位）に必ず入っている（図表6）。一方、低位8機関の成績は2008年度、2012年度は1位となったものの、それ以外の年度は総平均を下回り、平均では17位となっている。また、高位8機関が総平均の成績を上回ったのは2004年度の1回だけで、平均では30位と予測期間の中位を大きく下回る結果となっている。



5—個別機関の予測誤差

最後に、GDP速報に関する個別機関の予測誤差をコンセンサス予測と比較しながら確認する。QUICKの集計では、毎四半期30前後の機関が回答しているが、今回は継続的に予測値を回答している機関に限定して予測誤差を計算した。具体的には対象期間中に調査回数の8割以上回答している16機関（ニッセイ基礎研究所を含む）、当該四半期終了から概ね1ヵ月後に各機関から公表される直前予測を対象とした。

前述したように、コンセンサス予測は相対的に優秀な予測となるため、個別機関がコンセンサス予測に勝つことは難しい。今回の検証期間（2000年1-3月期～2019年7-9月期）においても、実質GDP（前期比年率）の平均絶対誤差は、最もパフォーマンスの良い機関でも1.00%とコンセンサスの0.99%をわずかながら上回り、16機関全てがコンセンサスよりも予測精度が劣るという結果となった。最も予測誤差が大きい機関の平均絶対誤差は1.33%であった（図表7）。予測誤差を5年毎に区切ってみると、短期間ではコンセンサスよりも優秀な機関も少数存在するが、勝ち続けている機関はない。16機関中8機関はすべての期間でコンセンサスに負けている。

各機関の予測誤差がコンセンサスの予測誤差よりも小さかった場合に勝ち、大きかった場合に負け、同じだった場合に引き分けとし、引き分けを除いた上で勝率（勝ち数/（勝ち数+負け数））を計算すると、全期間で50%を上回っているのは3機関、最低は29.5%であった。16機関の平均勝率は39.2%と50%を大きく下回っている。上位3機関は平均絶対誤差ではコンセンサスに劣っているものの、勝率で見ればコンセンサスに勝っている。

図表7 個別機関の予測誤差と順位（実質GDP・年率換算）

	予測誤差 全期間	全期間	順位				勝率 全期間
			00～04年	05～09年	10～14年	15～19年	
機関A	1.00%	1	1	1	6	5	52.1%
機関B	1.01%	2	2	7	2	3	53.8%
機関C	1.03%	3	6	6	1	9	54.7%
機関D	1.06%	4	3	3	7	4	36.5%
機関E	1.14%	5	7	11	4	13	36.2%
機関F	1.15%	6	12	10	5	6	36.2%
機関G	1.20%	7	8	5	8	16	36.2%
機関H	1.20%	8	4	15	15	1	39.5%
機関I	1.21%	9	14	9	9	2	44.4%
機関J	1.22%	10	15	12	3	8	34.8%
機関K	1.23%	11	13	2	14	10	29.5%
機関L	1.23%	12	5	14	13	11	35.0%
機関M	1.26%	13	11	4	16	12	34.3%
機関N	1.30%	14	16	8	10	7	37.3%
機関O	1.31%	15	10	13	11	15	35.4%
機関P	1.33%	16	9	16	12	14	30.9%
コンセンサス	0.99%	1	2	1	7	3	—

(注) 予測誤差は2000年1-3月期から2019年10-12月期(79四半期)の平均絶対誤差
シャドーはコンセンサスを上回っているもの
勝率は対コンセンサス

予測誤差を需要項目別にみると、設備投資、民間在庫変動、政府消費についてはコンセンサスよりも優秀な機関は存在しないが、民間消費は5機関がコンセンサスの予測精度を上回っている（図表8）。

実質GDP全体の予測誤差と需要項目別の予測誤差は必ずしも連動しない。各機関の需要項目別の平均絶対誤差（寄与度ベースに換算）を積み上げて求めた需要項目計の予測誤差と実質GDP全体の予測誤差を比較すると、実質GDPの順位と需要項目計の順位は必ずしも一致していないことが確認された。

たとえば、機関Eは実質GDPの順位は5位だが、需要項目別の順位は低く需要項目計の順位は13位と下位に位置する。需要項目別の予測誤差は大きいですが、それぞれが逆方向に外れることにより実質GDPの予測誤差が小さくなるが多かったということになる。

一方、機関Cは実質GDPの順位は3位だが、需要項目別にみると8項目のうち6項目の順位が1位で、このうち、5項目はコンセンサスよりも誤差が小さかった。さらに、需要項目別の予測誤差を積み上げて求めた需要項目計の予測誤差は2.68%とコンセンサスの2.98%よりも小さく、順位は1位となった。

図表8 個別機関の需要項目別予測順位

	需要項目計	民間消費	住宅投資	設備投資	民間在庫変動	政府消費	公的固定資本形成	外需	輸出	輸入
機関A	2	2	2	2	3	8	2	6	2	4
機関B	3	3	7	5	1	1	3	1	3	3
機関C	1	1	1	1	4	2	1	4	1	1
機関D	6	9	5	9	7	3	4	3	5	9
機関E	13	8	9	16	13	11	8	11	11	13
機関F	5	6	3	6	8	5	5	7	11	5
機関G	4	4	6	4	2	13	10	12	6	7
機関H	11	7	10	10	9	16	14	9	8	10
機関I	7	10	4	8	14	9	9	5	4	2
機関J	9	13	13	7	10	10	11	10	9	11
機関K	14	12	8	11	11	7	15	14	14	12
機関L	12	5	11	14	12	14	13	13	12	14
機関M	10	15	12	3	15	15	6	8	10	6
機関N	8	16	14	15	5	4	7	2	7	8
機関O	15	14	15	12	16	12	12	16	13	15
機関P	16	11	16	13	6	6	16	15	15	16
コンセンサス	2	6	2	1	1	1	2	2	3	2

(注) 需要項目別の順位は各需要項目の平均絶対誤差による
 需要項目計の順位は各需要項目の平均絶対誤差(寄与度ベース)の合計による

6—まとめ

GDP速報（1次速報）における実質GDP成長率（前期比年率）の予測誤差（民間調査機関の平均、コンセンサス）は、2000年1-3月期から2019年7-9月期（79四半期）の平均で0.99%（絶対誤差）である。予測誤差の分布をみると、実績値が予測値から±0.5%以内におさまる確率は3割弱で、1割弱は予測誤差が±2%を上回っている。需要項目別には、公的固定資本形成、設備投資、住宅投資の予測誤差が大きいですが、実質GDP成長率への寄与度でみると、大きい順に民間消費、設備投資、民間在庫変動となる。日本のGDP速報の予測誤差は米国（平均絶対誤差は0.54%）の2倍近いが、その原因として、日本のGDP統計の振れが大きいことや推計方法の開示が不十分であることなどが挙げられる。

個別機関の予測精度を確認したところ、需要項目別にはコンセンサスよりも予測精度が高い機関が存在したが、実質GDP成長率については約20年間の平均でコンセンサス予測よりも良いパフォーマンスをあげている機関がひとつもなかった。

第1～3回では主としてGDP統計（第1回、第2回は年度ベースの実質GDP、第3回は四半期別GDP速報（QE））に関する予測精度の検証を行ったが、第4回は月次指標を取り上げる予定である。

(お願い) 本誌記載のデータは各種の情報源から入手・加工したものであり、その正確性と安全性を保証するものではありません。また、本誌は情報提供が目的であり、記載の意見や予測は、いかなる契約の締結や解約を勧誘するものではありません。