

コンセンサス予測は“良い予測”なのか

—四半期実質 GDP 年率成長率予測による検証—⁽¹⁾

飯塚 信夫

**Are consensus forecasts “Good predictions” ?
— Verification by Quarterly real GDP Annual Growth Rate Forecast
in Japan —**
Nobuo Iizuka
Kanagawa University

【概要】 かねて“良い予測”と評価されているエコノミストの予測の平均または中央値である「コンセンサス予測」の精度を評価した。「ESP フォーキャスト集計」の四半期実質 GDP 年率成長率予測のパネルデータを用い、時系列でみたコンセンサス予測の予測誤差の特性、横断面、すなわち各時点の各フォーキャスターの予測の分布の特性を確認した。この結果、各時点の四半期実質 GDP 年率成長率予測値は、各フォーキャスターの予測のばらつき範囲に収まらないことが多く、予測平均は必ずしも最良の集約方法ではないことなどがわかった。一方、各フォーキャスターの予測のばらつきは景気後退期に大きくなる傾向があることも明らかになった。この特性を用いて、新しい予測モデルを構築したところ、コンセンサス予測の精度を高めることができた。

【キーワード】 経済予測、予測の組合せ、予測評価、リアルタイムデータ、Out-of-sample の予測精度

【Abstract】 We assessed the predictive accuracy of a ‘consensus forecast,’ which is the mean or median of those predictions of economists that have been rated better than others. First, using the quarterly annual GDP growth rate forecast panel data from the ESP Forecasts in Japan, we analyzed the characteristics of errors in consensus forecasts over time and those of the distribution of each forecaster's estimates at each point in time. As a result, we found that quarterly GDP annual growth

(1) 本研究は、神奈川大学経済貿易研究所共同研究調査助成「景気の現状把握精度を高めるための諸手法の検討」(2021~2022年度)の助成を受けている。本稿の作成にあたり、第3回 Economics & Econometrics 研究会(2022年9月28日開催、神奈川大学経済貿易研究所との共催)、第37回応用経済時系列研究会・研究報告会の参加者から貴重なコメントを頂いた。特に、第37回応用経済時系列研究会・研究報告会の討論相手の林田元就氏からは詳細なコメントを頂き、本稿の改善に結び付いたと考えている。深く感謝したい。なお、本稿に残された誤りはいうまでもなく筆者の責に帰するものである。

rate predictions at each point in time often do not fall within the range of variability in each economist's forecasts, and that taking their average is not necessarily the best aggregation method. Conversely, it also became clear that the variability in each forecaster's predictions tends to increase during recessions. Using this characteristic, we constructed a new prediction model, and were able to improve the accuracy of consensus prediction.

【Keywords】 Macroeconomic forecast, Forecast combinations, Forecast Evaluation, Real-Time Data, Out-of-Sample Forecast Accuracy

1. はじめに

エコノミストの予測の平均または中央値である「コンセンサス」は、かねて良い予測⁽²⁾と評価されている。日本のフォーキャスターの予測を毎月集計している「ESP フォーキャスト集計」を分析した河越・土屋（2022a）は、2021年度の予測においてコンセンサス予測（フォーキャスターの予測の総平均）がすべてのフォーキャスターの予測結果の中で「総合8位と18年連続で一桁のよい順位を維持している（平均では5.8位）」ことを明らかにしている。Chikamatsu et al. (2021) は、様々な計量モデルに基づく GDP ナウキャストと ESP フォーキャストのコンセンサス予測を組み合わせることで予測の精度が高まることを示している。一方、Manski (2011) は個々の予測の平均値が個々の予測に比べて予測精度が高いのは、数学理論から当たり前前に引き出されるものだとして指摘している。コンセンサス予測が大きく外れない予測であるとは言えるものの、過度にありがたがるべきではないということである。斎藤（2020）も「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」ことを、数式を用いて説明している。そのうえで、「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」と結論付けている。

ESP フォーキャスト調査は、「コンセンサス」をより精度の高いものにする工夫を始めている。河越・土屋（2022b）は ESP フォーキャスト調査における平均分布予測の過去14年間のデータを活用した新たな参考指標を開発した。ESP フォーキャスト調査では、平均値が公表されている各フォーキャスターの点予測値のほかに実質 GDP 成長率と消費者物価上昇率については各フォーキャスターの分布予測を聞き取りしている。この結果を用いて実績値が確率50%で収まると予想される幅などを示している。

河越・土屋（2022b）は調査設計の工夫により予測精度を向上させるものと言えるが、調査対象が年度予測値に限定されており、本稿が対象とする四半期実質 GDP 年率成長率予測の精度向上には利用できない。そこで、本稿では与えられたアンケート調査の結果の下で、コンセンサス

(2) 例えば、小峰（2022）は、ESP フォーキャスト調査の3つの定理の一つとして「エコノミストの予想の平均値（コンセンサス予想）は良い予想である」としている。

予測やその予測誤差の「クセ」を統計的に把握したうえで、四半期実質 GDP 年率成長率予測の精度向上に貢献したい。

本稿の構成は以下の通りである。第2節では本稿の分析のために（公社）日本経済研究センターから貸与を受けた「ESP フォーキャスト調査・パネルデータ」の概要を説明する。第3節では、時系列でみたコンセンサス予測の予測誤差の特性とその背景について述べる。第4節では横断面、すなわち各時点の各フォーキャスターの予測の分布などの特性について述べる。第5節では、第3節と第4節の考察を踏まえて2種類の新しい予測モデルを提案する。第6節ではコンセンサス予測と新しく提案したモデルの Out-of-sample の予測精度を比較する。第7節はまとめである。

2. データ

本稿では、各フォーキャスターの四半期実質 GDP 年率成長率の予測値として、（公社）日本経済研究センターから貸与された「ESP フォーキャスト調査・パネルデータ」を用いている。同機関が運営する「ESP フォーキャスト調査」について、調査開始の2004年4月から2021年3月までのフォーキャスター（予測者）ごと、予測時点ごとの予測が収録されている。本稿で扱う四半期実質 GDP 年率成長率のほか、四半期単位の予測としては、消費者物価（生鮮食品を除く総合）、鉱工業生産指数など13の経済データの予測が収録されている。また、年度単位の予測としては、実質 GDP 成長率、名目 GDP 成長率、GDP の需要項目、鉱工業生産、消費者物価など27の経済データが収録されている。

2004年4月から2021年3月までに一度でも参加したフォーキャスターの数は64人であり、上記期間の回答総数（同一時期でも予測対象が異なれば別の回答とカウントする）は、6万6161件である⁽³⁾。「ESP フォーキャスト調査」においては、基本的に毎年1月の調査で予測期間を1年（4四半期分）延ばしている。このため、各フォーキャスターの年度単位の回答数は最大98件になる⁽⁴⁾⁽⁵⁾。上記期間に一度も休まずに参加したフォーキャスターは21人である。

四半期実質 GDP 年率成長率の予測は、一次速報が公表される直前の調査（0ヵ月前）から、

(3) ただし、コロナ禍の下で、2020年4月に定例調査に追加して実施された臨時調査（パネルデータの date に202004_2 と書かれている）は除いて集計している。

(4) 2019年度を例にとれば、2019年第1四半期（2019年5月に1次速報が公表）の予測は2019年4月、5月の2回、第2四半期（2019年8月に1次速報が公表）の予測は2019年4～8月の5回、第3四半期（2019年11月に1次速報が公表）の予測は2019年4～11月の8回、第4四半期（2020年2月に1次速報が公表）の予測は2019年4～2020年2月の11回、2020年各四半期および2021年1～3月期の予測は2019年4～2020年3月の12回、2021年4～6月期から2022年1～3月期の予測は2020年1～3月の3回である。

(5) なお、2007年度（2007年4～08年3月）は99件、2018年度（18年4～19年3月）は101件、2020年度（20年4～21年3月）は103件である。この理由は下記の通り。2007年度では、実績値が判明しているタイミングの2007年6月調査においても2007年1～3月期の調査を行っている。2018年度では、年が改まる前の2018年11月調査から2020年4～6月期の調査を始め、2018年12月調査から2020年7～9月期の調査を始めている。2020年度では、年が改まる前の2020年9月調査から2022年4～6月期の調査を始め、2020年12月調査から2022年7～9月期の調査を始めている。

最大25ヵ月前までさかのぼることができる。例えば、2022年2月に公表された2021年10～12月期の実質 GDP 年率成長率の予測は、2020年1月から2022年2月まで調査されている。本稿では ESP フォーキャスト集計における優秀フォーキャスター選定の評価対象にならない、GDP の1次速報が公表される直前（0ヵ月前予測）から5ヵ月前までの予測を分析対象とする。2021年10～12月期の実質 GDP 年率成長率予測を例にとれば、2022年2月調査（0ヵ月前）から2021年9月調査（5ヵ月前）までとなる。

河越（2007）は、ESP フォーキャスト集計では「回答期限が毎月5日前後に設定され、10日前後に結果が公表される」と説明している。例えば、2022年2月調査では、2022年2月4日に回答期限が設定され、集計結果が2月9日に公表された。2021年10～12月期の GDP 速報が公表されたのは2022年2月15日だった。この2022年2月調査における2021年10～12月期の実質 GDP 年率成長率予測が0ヵ月前予測となるが、予測を行う時点で鉱工業生産、実質輸出入など予測に用いる主な月次経済統計の2021年10～12月期の実績値が揃っている。一方、3ヵ月前予測（2021年11月調査）では、予測対象となる2021年10～12月期の主な月次経済統計の実績値は得られず、前の四半期である2021年7～9月期の実績値も判明していない。2ヵ月前予測までと3ヵ月前予測以降では、予測対象の四半期の実質 GDP 年率成長率に関する情報量に大きな違いがあるといえる。

3. 時系列でみたコンセンサス予測の予測誤差の特性とその背景

（1）平均ゼロ、標準偏差は予測期間が長いほど大きく

コンセンサス予測の予測誤差を「実績値－コンセンサス予測」と定義して、その特徴をまず確認する。ここでの実績値は、河越・土屋（2022）にならって1次速報値とする。

図表1は予測誤差の平均、標準偏差などの基本統計量をまとめたものである。予測誤差の平均は0ヵ月前予測の0.2ポイントから5ヵ月前のマイナス0.9ポイントである。0ヵ月前予測から5ヵ月前予測のいずれの平均値も「0に等しい」という帰無仮説を棄却できない。2004年1～3月期から2020年10～12月までの17年間（68四半期）を均してみれば予測誤差はゼロと言って良いということになる。

予測誤差の標準偏差は予測期間が長くなるほど大きくなっている。0ヵ月前予測では1.5だが、5ヵ月前予測だと5.2となる。予測誤差の四分位範囲に注目すると、0ヵ月前予測ではマイナス0.7からプラス1.2の範囲に収まっているが、5ヵ月前予測ではマイナス2.5からプラス1.2と広がる。マイナス方向に広がっていることから、予測期間が長くなるほど過大な予測になる傾向があるといえる。

予測誤差がマイナス、過大な予測になるのは景気後退期に多いことが良く知られている。本稿の分析対象期間では、2008年9月のリーマンショック、2020年春以降のコロナ禍という予測しがたい大きな経済ショックもあった。そこで、景気後退期間とコロナ禍（2008年4～6月期から2009年1～3月期、2012年4～6月期から2012年10～12月期、2019年1～3月期以降）を除いたサンプルでも予測誤差の基本統計量を確認した。すると、予測誤差の標準偏差は2ヵ月前から5ヵ月前でほとんど変化がなく、第1四分位はマイナス1.5ポイント前後、第3四分位はプラス1.5ポイント前後で安定していることがわかる。第1四分位のマイナス幅は全サンプルよりも小

図表1 四半期実質 GDP 年率成長率のコンセンサス予測の予測誤差の基本統計量

(1) サンプル全体 (2004年1～3月期から2020年10～12月期)

	予測時点					
	0 ヶ月前	1 ヶ月前	2 ヶ月前	3 ヶ月前	4 ヶ月前	5 ヶ月前
平均	0.2	0.1	-0.2	-0.5	-0.7	-0.9
標準偏差	1.5	2.5	3.0	3.3	4.4	5.2
最小	-4.1	-7.6	-10.9	-12.1	-20.9	-30.1
第1四分位	-0.7	-1.4	-1.9	-2.3	-2.3	-2.5
第2四分位	0.0	0.1	-0.2	-0.3	-0.3	-0.3
第3四分位	1.2	1.6	1.5	1.3	1.3	1.2
最大	4.7	9.0	9.3	8.7	11.9	12.3
サンプル数	68	68	67	67	67	66

(2) 景気後退局面などを除くサンプル

	予測時点					
	0 ヶ月前	1 ヶ月前	2 ヶ月前	3 ヶ月前	4 ヶ月前	5 ヶ月前
平均	0.1	0.1	0.0	-0.1	0.0	-0.1
標準偏差	1.4	1.9	2.2	2.3	2.3	2.3
最小	-4.1	-5.3	-5.7	-5.7	-4.6	-4.9
第1四分位	-0.6	-1.2	-1.5	-1.3	-1.5	-1.9
第2四分位	0.1	0.1	0.1	0.0	0.0	0.2
第3四分位	1.2	1.6	1.4	1.6	1.5	1.6
最大	3.1	3.8	4.0	4.2	5.1	5.5
サンプル数	53	53	52	52	52	51

出所：日本経済研究センター「ESPフォーキャスター調査・パネルデータ」をもとに作成

さい。

これは、景気後退期などを除いたことで、結果的に過大な予測をしてしまったサンプルが減ったためであろう。また、景気後退期ほど予測対象の四半期の主な月次経済統計が入手されることで予測誤差が縮まる度合いが大きいとも言える。

一方、第3四分位の値は全サンプルと景気後退期を除いたサンプルであまり変化がない。0 ヶ月前予測の標準偏差も全サンプルが1.5に対して、景気後退期間を除いたものでも1.4とほぼ変わらない。前述したように0 ヶ月前予測においては予測を行う時点で予測対象の四半期の主な月次経済統計が出揃っている。しかし、これらの月次経済統計と四半期実質 GDP 年率成長率の間に単純な線形関係があるわけではなく、そのあたりの不確実性がこの標準偏差に表れているとも考えられる。

また、本稿では四半期実質 GDP 年率成長率の1次速報値⁽⁶⁾を実績値として分析しているが、四半期実質 GDP 年率成長率は2次速報、第一次年次推計⁽⁷⁾など基礎統計が加わる中で改定が行

(6) 1次速報時点の実質 GDP の実績値は、(公財)東京財団政策研究所ウェブサイトの「リアルタイムデータベース」から取得した。

(7) 第一次年次推計は、1993SNA2005年度基準まで確報と呼ばれていた。

われる。2002年4～6月期から2015年1～3月期の実質 GDP の前期比成長率の改定幅を分析した飯塚（2017）によれば1次速報から第一次年次推計までの改定幅は0.461。さらに2015年4～6月期から2021年1～3月期について分析した山澤（2022）によれば1次速報から第一次年次推計までの改定幅は0.271である。

ESP フォーキャスト調査における実質 GDP 成長率は前期比年率であり、前期比成長率のほぼ4倍となる。0ヵ月前予測の標準偏差とこの改定幅の大きさはほぼ一致しており、埋めがたい誤差の可能性もあろう。

（2）振れが大きく、マイナス成長も珍しくない四半期実質 GDP 年率成長率の実績値

以上のような予測誤差の特性が生まれる理由は2つあると考えられる。

第1に、予測対象となる四半期実質 GDP 年率成長率の実績値の振れが大きく、マイナス成長になることが珍しくないことである。本稿の分析対象である2004年第1四半期から2020年第4四半期までの17年間（68四半期）の1次速報値の標準偏差は5.8である（図表2）。5ヵ月前予測の予測誤差の標準偏差に近い。四分位範囲をみるとマイナス0.7%からプラス3.4%と広い。景気後

図表2 四半期実質 GDP 年率成長率の実績値とコンセンサス予測の基本統計量

（1）サンプル全体（2004年1～3月期から2020年10～12月期）

	実績値	予測時点					
		0ヵ月前	1ヵ月前	2ヵ月前	3ヵ月前	4ヵ月前	5ヵ月前
平均	0.8	0.6	0.7	0.9	1.2	1.4	1.7
標準偏差	5.8	5.2	4.2	4.0	3.5	2.1	1.6
最小	-27.8	-26.6	-23.5	-23.0	-21.3	-6.9	-4.1
第1四分位	-0.7	0.0	0.5	0.6	1.1	1.1	1.2
第2四分位	1.5	1.5	1.4	1.5	1.4	1.5	1.6
第3四分位	3.4	2.5	2.5	2.3	2.2	2.2	2.2
最大	21.4	18.0	14.2	14.1	13.3	9.4	9.1
サンプル数	68	68	68	67	67	67	66

（2）景気後退局面などを除くサンプル

	実績値	予測時点					
		0ヵ月前	1ヵ月前	2ヵ月前	3ヵ月前	4ヵ月前	5ヵ月前
平均	1.7	1.5	1.5	1.6	1.6	1.6	1.7
標準偏差	2.6	2.1	1.6	1.6	1.4	1.4	1.3
最小	-6.8	-6.8	-4.9	-4.2	-3.8	-4.0	-4.1
第1四分位	0.3	0.8	0.8	1.0	1.3	1.3	1.3
第2四分位	1.9	1.7	1.8	1.7	1.6	1.6	1.7
第3四分位	3.7	2.8	2.5	2.3	2.2	2.2	2.2
最大	6.0	5.8	5.3	5.0	4.6	4.5	4.4
サンプル数	53	53	53	52	52	52	51

出所：内閣府「四半期別 GDP 速報」、日本経済研究センター「ESP フォーキャスト調査・パネルデータ」をもとに作成

退局面などを除くと振れの大きさは小さくなるが標準偏差は2.6ある。四分位範囲をみるとプラス0.3%から3.7%であり、第3四分位の値は全サンプルと変わりがない。

分析対象の68四半期のうちマイナス成長率だったのは21四半期であるが、このうち、景気後退期間で成長率がマイナスだったのは9四半期のみであり、景気拡張期間でもマイナス成長になるのは珍しくない。

第2に、振れの大きい実績値に対し、コンセンサス予測の標準偏差は、予測期間が長いほど小さい。3ヵ月前予測と4ヵ月予測の四分位範囲は1.1%~2.2%、5ヵ月前予測は1.2%~2.2%である。さらに、予測期間が長いほどコンセンサス予測がマイナスとなることはまれである。

3ヵ月前予測~5ヵ月前予測でコンセンサス予測の成長率がマイナスだったのは以下の通りである。

3ヵ月前予測：2008年10~12月期、**2009年1~3月期**、2011年4~6月期、**2014年4~6月期**、**2019年10~12月期**、2020年4~6月期（6四半期）

4ヵ月前予測：2009年1~3月期、2009年4~6月期、2011年4~6月期、**2014年4~6月期**、**2019年10~12月期**、2020年4~6月期（6四半期）

5ヵ月前予測：2009年1~3月期、2009年4~6月期、**2014年4~6月期**、**2019年10~12月期**（4四半期）

共通してマイナス成長率見通し（太字）なのは、消費税率が8%に引き上げられた直後の2014年4~6月期と10%に引き上げられた直後の2019年10~12月期である。駆け込み消費の反動減とほぼ間違いなく起きるであろうことを情報として織り込んだと考えられる。

そのほかは、リーマンショック（2008年9月）、東日本大震災（2011年3月）、コロナ禍による緊急事態宣言の発令（2020年4月）というショックを織り込めた時点でマイナス成長予測になっている。逆に、2009年4~6月期はV字回復という情報を織り込めなかった4ヵ月前予測と5ヵ月前予測がマイナス成長と読み間違えている。

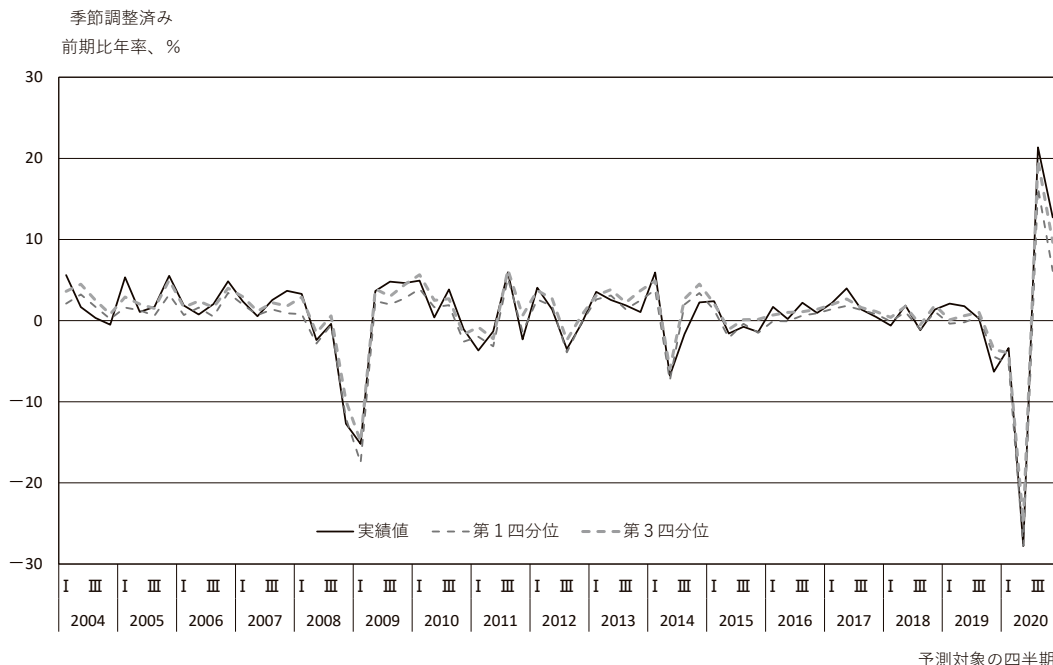
一方、景気後退期間を除いたサンプルを見ると、コンセンサス予測の平均値が実績値の平均値にほぼ等しいことが注目される。3ヵ月前予測から5ヵ月前予測の四分位範囲はすべて1.3%~2.2%の範囲となっている。3ヵ月前予測から5ヵ月前予測は、予測対象の四半期に関する情報がないため、日本経済のトレンド成長率にプラスマイナスの判断を加えて、各フォーキャスターが予測を行っている可能性もあろう。

4. 横断面でみた各フォーキャスターの予測と実績値

(1) フォーキャスター予測の四分位範囲に収まらない実績値

次に各四半期について各フォーキャスターが行った予測と実績値の関係を確認する。図表3は、四半期実質GDP年率成長率の0ヵ月前予測について、各予測対象の四分位範囲と実績値の推移を示したものである。四分位範囲は、各フォーキャスターの予測の低い方から4分の1のところの予測（第1四分位）と4分の3のところの予測（第3四分位）の範囲であり、フォーキャスターの予測の中央50%部分の範囲を指す。2004年第1四半期から2020年第4四半期の68四半期

図表3 四半期実質 GDP 年率成長率の実績値とフォーキャスターの0 ヶ月前予測の四分位範囲



出所：内閣府「四半期別 GDP 速報」、日本経済研究センター「ESP フォーキャスト集計」をもとに作成

のうち、この予測中央50%の部分の範囲に実績値が収まっていたのは16四半期（23.5%）である。29四半期（42.6%）は実績値が第3四分位より大きく、23四半期（33.8%）は実績値が第1四分位より小さい。予測中央50%に収まる確率は、予測期間が長くなるほど低くなり、5 ヶ月前予測においては13.6%まで低下する。各四半期についてフォーキャスターが行った予測の大勢という情報は、実績値の事前予測にあまり役に立たないと考えられる。

（2）予測の最大値と最小値の範囲に実績値が収まる確率は5 ヶ月前予測でも5 割以上

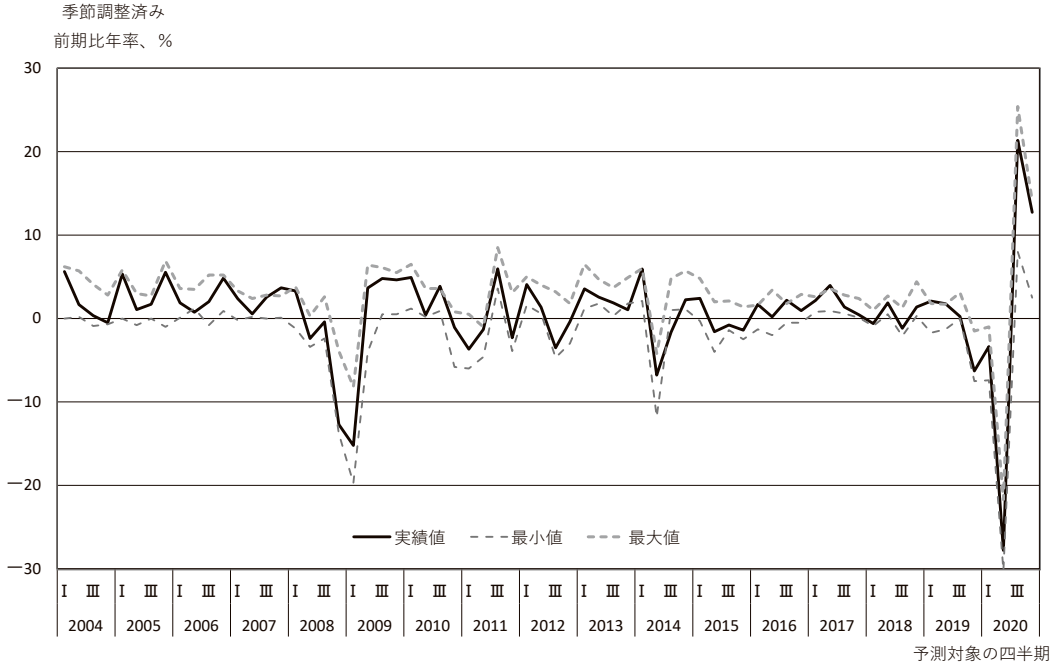
図表4は、四半期実質年率 GDP 成長率の0 ヶ月前予測について、各四半期の予測の最大値、予測の最小値と実績値の推移を示したものである。2004年第1四半期から2020年第4四半期の68四半期のうち、予測の最大値と最小値の範囲に実績値が収まっていたのは58四半期（85.3%）と四分位範囲に比べて大きく増える。予測の最大値と最小値の範囲に収まる確率は、予測期間が長くなるほど低くなるが、5 ヶ月前予測でも53%である。言い換えれば、ESP フォーキャスト予測の最大値と最小値の範囲に四半期実質 GDP 年率成長率が収まると予測すれば、5 割の確率で当たるということである。

（3）各フォーキャスターの予測値の集約方法による予測誤差の違いは大きくない

以上のような横断面の特性を踏まえると、各フォーキャスターの予測値の総平均であるコンセンサス予測が、必ずしも最良の予測値の集約方法ではないと予想される。実績値がフォーキャスターの予測中央50%に収まる確率は0 ヶ月前予測であっても決して高いものではないためだ。

そこで、各時点の予測の平均値（コンセンサス予測）と、最小値、第1四分位、中央値（第2

図表4 四半期実質 GDP 成長率の実績値とフォーキャスターの0カ月前予測の最大値、最小値



出所：内閣府「四半期別 GDP 速報」、日本経済研究センター「ESP フォーキャスト集計」をもとに作成

図表5 フォーキャスター予測の集計方法の違いによる平均絶対誤差（MAE）の違い

	コンセンサス 予測	最小値	第1四分位	中央値	第3四分位	最大値	最大値と 最小値の平均
0カ月前予測	1.19	2.66	1.34	1.14	1.20	2.38	1.36
1カ月前予測	1.88	2.92	1.92	1.92	1.94	2.72	1.85
2カ月前予測	2.15	3.02	2.19	2.18	2.25	3.08	2.14
3カ月前予測	2.37	3.22	2.38	2.40	2.46	3.73	2.49
4カ月前予測	2.73	3.31	2.81	2.75	2.76	3.49	2.70
5カ月前予測	2.99	3.45	3.09	3.00	2.98	3.80	2.88

四分位)、第3四分位、最大値を予測値とした場合の予測誤差を、平均絶対誤差（MAE）という尺度で比較した（図表5）。MAEは実績値と予測の差の絶対値を平均したものであり、最も数値が小さい予測値の精度が高いと考えられる。0カ月前予測は中央値のMAEが最も小さいが、1カ月前予測から4カ月前予測はコンセンサス予測のMAEが最小である。5カ月前予測では第3四分位にわずかに負けているものの、最小値圏にあると言って良いだろう。確かに、河越・土屋（2022）が言うようにコンセンサス予測の「優れた成績の原因は大きく外れない」ことにあるといえる。

ただし、各集約値間のMAEの差はさほど大きいものではない。例えば、0カ月前予測においては各時点で最小値の予測を選び続けるとMAEは2.66となり、コンセンサス予測との差は1.48ポイントとなる。予測対象の実質GDP成長率は前期比年率であり、この差は前期比に換算する

と0.4ポイント弱である。これは、前述した実質 GDP 成長率の1次速報から年次推計（確報）までの改定幅とほぼ同等である。しかも、最も誤差が大きい予測とコンセンサス予測の差は5ヵ月前予測では0.81ポイントまで縮小する。

興味深いのは、参考のために算出した予測の最大値と最小値の平均値を予測とした場合、コンセンサス予測と遜色ないパフォーマンスとなっていることである。MAEで比較すると、1ヵ月前予測、2ヵ月前予測、4ヵ月前予測、5ヵ月前予測においてコンセンサス予測より小さい値になっている。コンセンサス予測より精度の高い予測値を作る余地があることが確認できる。

（4）各フォーキャスターの予測値のばらつきは景気後退局面で高まる傾向

各フォーキャスターの予測値のばらつきは、予測対象の四半期ごとに異なる。図表6は0ヵ月前予測について、予測対象の四半期ごとの各フォーキャスターの予測値のばらつきを標準偏差で示したものである。2008年第1四半期を景気の山として始まった景気後退局面では標準偏差が次第に大きくなり、景気の谷を付けた2009年第1四半期をピークに低下に転じている。2012年からの長期の景気拡張局面にあった2015年近辺から2017年までは標準偏差が徐々に低下している。そして、景気の山をつけた2018年第4四半期あたりから標準偏差が徐々に上昇し、コロナ禍の2020年で急上昇した。標準偏差の前期差を見ても、景気後退期間においてその前より大きくなっていることが確認できる。

（5）コンセンサス予測の誤差と各フォーキャスターの予測値のばらつきの関係

前述したように、コンセンサス予測の誤差は景気後退期にマイナス幅が大きくなる（予測値が過大になる）傾向がある。コンセンサス予測の予測誤差を景気後退期に1を取り、その他の期間には0をとる景気ダミーで回帰すると、0ヵ月前予測を除き、景気ダミーの係数は有意にマイナスとなる。これは欧米におけるコンセンサス予測の先行研究と同様の結果である。

一方、景気後退期に入っているかどうかを事前に知ることは難しい。そこで、各フォーキャスターの予測値のばらつき（標準偏差）を景気ダミーの代わりに用いた推計を行った（図表7）。コンセンサス予測の予測誤差を定数項と標準偏差、もしくは標準偏差の前期差で回帰をした。標準偏差もしくは標準偏差の前期差の符号はマイナスとなることが期待される。符号条件が合っていて係数が有意なのは、1ヵ月前予測、2ヵ月前予測、3ヵ月前予測、4ヵ月前予測で標準偏差の前期差を用いたケースになる。

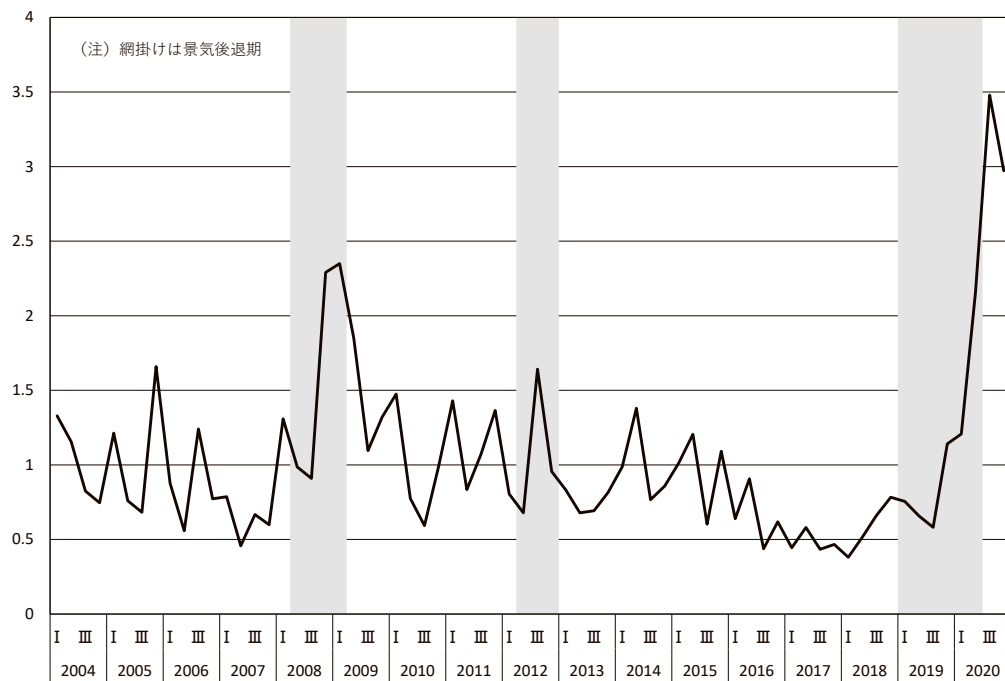
5. 2種類の予測モデルを提案

以上の考察をもとに、2種類の予測モデルを提案したい。

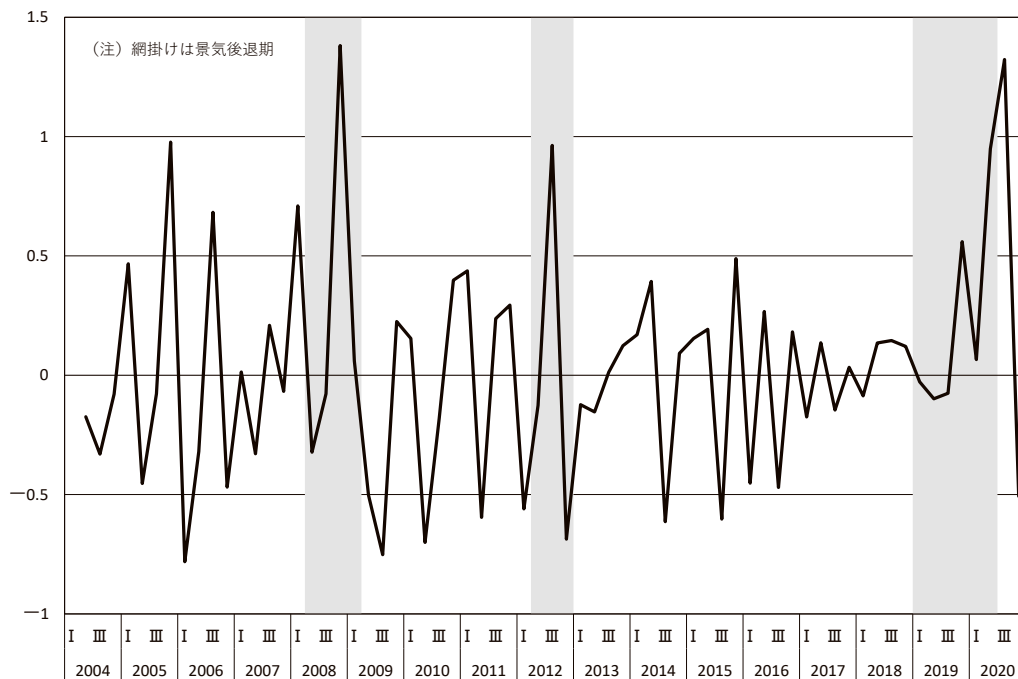
第1のモデルは、コンセンサス予測と予測対象の四半期ごとのフォーキャスターの予測の標準偏差の前期差を用いて、四半期実質 GDP 年率成長率を予測するモデルである。図表8に示したように、四半期実質 GDP 年率成長率の実績値を定数項とコンセンサス予測のみで回帰した model 1 と、標準偏差の前期差を加えた model 2 の推定結果を比べると、自由度修正済み決定係数（Adj-R）は0ヵ月前予測では model 1 と model 2 で大きな違いはないが、1ヵ月前予測から5ヵ月前予測では model 2 の自由度修正済み決定係数が大きくなっている（ただし、5ヵ月前予

図表6 0カ月前予測におけるフォーキャスター予測のばらつき

(1) 標準偏差



(2) 標準偏差の前期差



図表7 コンセンサス予測の予測誤差の回帰結果

被説明変数：コンセンサス予測の予測誤差
推定期間：全期間（2004年1～3月期～2020年10～12月期）

	0 ヶ月前予測		1 ヶ月前予測		2 ヶ月前予測	
定数項	-0.526 (0.438)	0.140 (0.182)	0.374 (0.788)	0.037 (0.306)	-0.134 (0.546)	-0.168 (0.368)
標準偏差	0.687 (0.444)		-0.281 (0.900)		-0.055 (0.584)	
Δ 標準偏差		-0.007 (0.443)		-2.745*** (0.619)		-2.189*** (0.550)
Adj-R	0.054	-0.015	-0.010	0.260	-0.015	0.191
D.W.	1.891	1.668	1.392	1.543	1.377	1.649

	3 ヶ月前予測		4 ヶ月前予測		5 ヶ月前予測	
定数項	-0.147 (0.469)	-0.438 (0.411)	-0.787 (0.964)	-0.588 (0.434)	-2.083** (0.797)	-0.908 (0.567)
標準偏差	-0.291 (0.518)		0.103 (1.119)		1.198 (0.856)	
Δ 標準偏差		-1.641*** (0.401)		-3.187*** (0.871)		-0.745 (0.881)
Adj-R	-0.008	0.234	-0.015	0.221	0.019	-0.002
D.W.	1.344	1.526	1.829	2.025	1.831	2.122

(注) 係数が0という帰無仮説を、***は1%有意で、**は5%有意で、*は10%有意で棄却することを示す

測では標準偏差の前期差は有意ではない)。インサンプルでの比較ではあるものの、コンセンサス予測だけを用いるより、標準偏差の前期差を加えることで予測精度が高まる可能性がある。

第2のモデルは、コンセンサス予測の代わりに、各四半期の予測の最大値と最小値の加重平均を用いる方法である。前述したように、最大値と最小値の単純平均値を予測した場合のMAEが、コンセンサス予測と遜色ないパフォーマンスとなっているためである。四半期実質GDP年率成長率の実績値を定数項と予測の最大値、最小値で回帰したmodel3とmodel3に標準偏差の前期差を説明変数に加えたmodel4を推定した。ただし、多重共線性の影響を考慮して、最大値と最小値の係数は合計で1になるように制約をかけて推定した。model3は0ヶ月予測から5ヶ月予測のすべてについて、model1より自由度修正済み決定係数が小さいが、model4は1ヶ月前予測、2ヶ月前予測において、model1よりも自由度修正済み決定係数が大きくなった（図表8）。

6. Out-of-sample の予測力比較

Out-of-sample の予測力を比較するために、以下のシミュレーションを行った。

図表8 四半期実質 GDP 年率成長率の予測モデルの比較

被説明変数：実質 GDP 年率成長率の実績値

推定期間：全期間（2004年1～3月期～2020年10～12月期）

	0 ヶ月前予測				1 ヶ月前予測			
	model 1	model 2	model 3	model 4	model 1	model 2	model 3	model 4
定数項	0.122 (0.161)	0.081 (0.173)	-0.566 (0.553)	-0.840 (0.632)	-0.127 (0.285)	-0.071 (0.267)	0.125 (0.603)	-0.805 (0.758)
コンセンサ 予測	1.087*** (0.027)	1.088*** (0.028)			1.266*** (0.066)	1.141*** (0.072)		
予測の最大値			0.641*** (0.126)	0.695*** (0.139)			0.511*** (0.152)	0.690*** (0.182)
Δ標準偏差		0.195 (0.413)		-0.691 (0.632)		-2.085*** (0.643)		-2.784*** (0.817)
Adj-R	0.938	0.939	0.918	0.921	0.845	0.868	0.814	0.863
D.W.	1.783	1.719	1.727	1.678	1.375	1.399	1.482	1.611

	2 ヶ月前予測				3 ヶ月前予測			
	model 1	model 2	model 3	model 4	model 1	model 2	model 3	model 4
定数項	-0.462 (0.348)	-0.235 (0.359)	-0.518 (0.490)	-1.498** (0.592)	-0.993** (0.392)	-0.528 (0.450)	-0.061 (0.401)	-1.332** (0.555)
コンセンサ 予測	1.286*** (0.084)	1.069*** (0.133)			1.420*** (0.105)	1.070*** (0.202)		
予測の最大値			0.571*** (0.122)	0.775*** (0.128)			0.411*** (0.098)	0.645*** (0.097)
Δ標準偏差		-1.835** (0.874)		-2.315*** (0.605)		-1.429** (0.711)		-2.181*** (0.323)
Adj-R	0.779	0.790	0.741	0.783	0.735	0.747	0.630	0.721
D.W.	1.395	1.608	1.320	1.553	1.455	1.519	1.338	1.610

	4 ヶ月前予測				5 ヶ月前予測			
	model 1	model 2	model 3	model 4	model 1	model 2	model 3	model 4
定数項	-2.479*** (0.540)	-2.035*** (0.569)	-0.740 (0.988)	-2.000** (0.754)	-2.407*** (0.902)	-2.514*** (0.909)	-0.609 (0.832)	-1.662* (0.864)
コンセンサ 予測	2.255*** (0.216)	1.985*** (0.247)			1.874*** (0.386)	1.968*** (0.392)		
予測の最大値			0.509** (0.237)	0.786*** (0.121)			0.395** (0.220)	0.635*** (0.173)
Δ標準偏差		-1.616** (0.762)		-3.654*** (1.095)		-1.066 (0.786)		-2.088** (0.626)
Adj-R	0.622	0.641	0.466	0.610	0.258	0.269	0.193	0.238
D.W.	1.814	1.928	1.827	2.004	1.946	2.161	1.873	2.136

(注) 係数が0 という帰無仮説を、*** は1%有意で、** は5%有意で、* は10%有意で棄却することを示す

- ①2004年第1四半期から2009年第4四半期のデータを用いて model 1～model 4 を推定
- ②推定された式を用いて2010年第1四半期を予測
- ③推定期間を1四半期延ばし、①と②の動作を繰り返す

シミュレーションの結果、model 1～model 4 のそれぞれについて、2010年第1四半期～2020年第4四半期まで44四半期の四半期実質 GDP 年率成長率の予測値が作成される。これらの予測値と、コンセンサス予測の MAE を比較したのが図表 9 である。コンセンサス予測の MAE は、0 ヶ月前予測と3 ヶ月前予測において最も小さく、最良の予測になっている。しかし、0 ヶ月前予測、1 ヶ月前予測、4 ヶ月前予測、5 ヶ月前予測は、model 1 が最良の予測である。Model 1 は四半期実質 GDP 年率成長率を定数項とコンセンサス予測で回帰することで、コンセンサス予測の予測誤差のクセを調整しているモデルとみなすことができるが、それにより予測精度を高める可能性があることが確認できた。さらに、model 1 にフォーキャスター間の予測のばらつきの前期差を加えた model 2 が2 ヶ月前予測においては最良の予測となっていた。

図表10は、44四半期において、コンセンサス予測と各モデルの予測誤差の絶対値の小さい順に順位付けした結果を確認したものである。コンセンサス予測が1位だったのは9回（3 ヶ月前予測）から17回（0 ヶ月前予測、2 ヶ月前予測）であり、半分にも満たないことが確認できる。コンセンサス予測の予測誤差のクセを調整したモデルとみなせる model 1 と合わせると18回（3 ヶ月前予測、5 ヶ月前予測）から26回（0 ヶ月前予測）と“勝率”があがるが、ようやく5割程度の勝率である。

一方、0 ヶ月前予測においても、model 2、model 3、model 4 の予測誤差が最も小さい四半期が存在している。具体的には、以下の四半期である。

- ・ model 2 が1位（4四半期）：2010年10～12月期、2015年7～9月期、2018年1～3月期、2018年7～9月期
- ・ model 3 が1位（9四半期）：2011年4～6月期、2012年1～3月期、2012年10～12月期、2015年1～3月期、2017年1～3月期、2017年4～6月期、2018年4～6月期、2019年1～3月期、2020年1～3月期
- ・ model 4 が1位（5四半期）：2011年1～3月期、2014年10～12月期、2016年10～12月期、2019年10～12月期、2020年10～12月期

図表 9 モデルによる MAE の比較

	コンセンサス予測	model 1	model 2	model 3	model 4
0 ヶ月前予測	1.15	1.15	1.16	1.36	1.34
1 ヶ月前予測	1.81	1.74	1.76	1.93	1.84
2 ヶ月前予測	2.01	1.99	1.97	2.09	2.10
3 ヶ月前予測	2.26	2.48	2.56	2.69	2.59
4 ヶ月前予測	2.65	2.38	2.63	2.86	2.96
5 ヶ月前予測	2.96	2.88	3.25	3.53	3.48

図表10 予測対象の四半期における各モデルの予測精度の順位

0 ヶ月前予測					
	コンセンサス予測	model 1	model 2	model 3	model 4
予測 1 位の回数	17	9	4	9	5
予測 2 位の回数	4	7	17	5	11
予測 3 位の回数	11	19	7	4	3
予測 4 位の回数	2	4	11	9	18
予測 5 位の回数	10	5	5	17	7
1 ヶ月前予測					
	コンセンサス予測	model 1	model 2	model 3	model 4
予測 1 位の回数	12	9	8	9	6
予測 2 位の回数	7	14	7	11	5
予測 3 位の回数	11	10	9	5	9
予測 4 位の回数	4	5	12	7	16
予測 5 位の回数	10	6	8	12	8
2 ヶ月前予測					
	コンセンサス予測	model 1	model 2	model 3	model 4
予測 1 位の回数	17	8	5	3	11
予測 2 位の回数	3	7	12	15	7
予測 3 位の回数	8	10	7	11	8
予測 4 位の回数	7	8	18	9	2
予測 5 位の回数	9	11	2	6	16
	44	44	44	44	44
3 ヶ月前予測					
	コンセンサス予測	model 1	model 2	model 3	model 4
予測 1 位の回数	9	9	9	7	10
予測 2 位の回数	11	10	8	6	9
予測 3 位の回数	14	8	5	12	5
予測 4 位の回数	4	4	17	4	15
予測 5 位の回数	6	13	5	15	5
4 ヶ月前予測					
	コンセンサス予測	model 1	model 2	model 3	model 4
予測 1 位の回数	15	8	5	6	10
予測 2 位の回数	6	11	15	7	5
予測 3 位の回数	8	15	7	6	8
予測 4 位の回数	2	5	10	20	7
予測 5 位の回数	13	5	7	5	14
5 ヶ月前予測					
	コンセンサス予測	model 1	model 2	model 3	model 4
予測 1 位の回数	13	5	6	10	10
予測 2 位の回数	5	11	12	8	8
予測 3 位の回数	5	15	14	8	2
予測 4 位の回数	5	8	8	10	13
予測 5 位の回数	16	5	4	8	11

以上の18四半期のうち9四半期の実績値がマイナスである。また、2014年4月の消費税率引き上げ後の2015年にかけての景気低迷期に model 2 から model 4 は強みを見せていることが確認できる。2019年10～12月期も消費税率引き上げ直後にあたるし、2020年10～12月期は緊急事態宣言明けで成長率が急回復していた時期にあたる。

model 3 と model 4 は過去のデータから予測の最大値と最小値の合成ウエイトを決め、1期先予測に用いている。このため、予測が大きく外れることもあり、予測精度は総じてみれば悪く、モデル間の比較でも予測誤差が大きい四半期が多い（予測4位や5位の回数が多い）。しかし、景気判断が難しい時期などには威力を発揮することも確認できた。

7. おわりに

本稿では、(公社)日本経済研究センターから貸与された「ESP フォーキャスト調査・パネルデータ」を用いて、四半期実質 GDP 年率成長率予測におけるコンセンサス予測（フォーキャスターの予測の総平均）が最良の予測かどうかを検証した。その結果、以下のことがわかった。

- ①予測対象の各四半期において、フォーキャスターの予測のばらつきは、四半期実質 GDP 年率成長率の実績値の変動の大きさをカバーしきれていない。
- ②実績値がフォーキャスターの予測の範囲を外れる（上振れ、下振れ）ことも珍しくない。この結果、総平均以外の予測値の集約を行って比較すると、コンセンサス予測の精度とさほど違くない。
- ③一方、景気後退期にコンセンサス予測は過大（予測誤差がマイナス）になる傾向があるが、予測対象の各四半期におけるフォーキャスターの予測のばらつき（標準偏差）は、景気後退期に大きくなることが確認できた。
- ④コンセンサス予測にフォーキャスター予測のばらつきの情報を加える予測モデルや、各四半期におけるフォーキャスター予測の最大値、最小値といった極端な予測値を活用した予測モデルを構築したところ、景気判断が難しい時期などにコンセンサス予測を上回る予測精度を発揮することが確認できた。

本稿ではフォーキャスターの予測のばらつきという情報を生かす可能性について、パネルデータを用いることで検証した。一方、パネルデータから得られる情報には、分布のばらつきだけでなく、歪度などもある。また、最大値や最小値という外れ値を生かすモデルの構築も様々な方法が考えられる。今後、これらについてさらに研究を進めたい。

●参考文献

-
- 飯塚信夫（2017）「GDP 速報改定の特徴と、現行推計の課題について」、『日本経済研究』、No.74、2017年3月
- 河越正明・土屋陽一（2022a）「2021年度 ESP フォーキャスト調査の評価に関するテクニカルノート」、日本経済研究センター、2022年10月21日
- 河越正明・土屋陽一（2022b）「ESP フォーキャスト調査の新たな参考指標について」、日本経済研究センター、2022年6月15日

- 小峰隆夫 (2022) 「小峰隆夫の私が見てきた日本経済史 (第100回) —私が見てきた ESP フォーキャスト調査」、日本経済研究センター、2022年1月18日
- 斎藤太郎 (2020) 「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) ~」、ニッセイ基礎研究所、2020年1月28日
- 山澤成康 (2022) 「GDP 速報、年次推計との改定幅は改善—法人企業統計の速報化に課題」、東京財団政策研究所 Review、R-2021-062、2022年3月2日
- Kyosuke Chikamatsu, Naohisa Hirakata, Yosuke Kido, Kazuki Otaka (2021) "Mixed-frequency approaches to nowcasting GDP: An application to Japan", *Japan & The World Economy* 57 101056
- Charles F. Manski (2011) "Interpreting and combining heterogeneous survey forecasts" In *The Oxford Handbook of Economic Forecasting*, ed. Michael P. Clements and David F. Hendry, Chapter 16, OXFORD university press