



日本銀行ワーキングペーパーシリーズ

海運データを用いた輸出のナウキャストイング

久野遼平*

hisanor@g.ecc.u-tokyo.ac.jp

古川角歩**

kakuho.furukawa@boj.or.jp

No.22-J-19
2022年11月

日本銀行
〒103-8660 日本郵便（株）日本橋郵便局私書箱 30号

* 東京大学

** 日本銀行調査統計局

日本銀行ワーキングペーパーシリーズは、日本銀行員および外部研究者の研究成果をとりまとめたもので、内外の研究機関、研究者等の有識者から幅広くコメントを頂戴することを意図しています。ただし、論文の中で示された内容や意見は、日本銀行の公式見解を示すものではありません。

なお、ワーキングペーパーシリーズに対するご意見・ご質問や、掲載ファイルに関するお問い合わせは、執筆者までお寄せ下さい。

商用目的で転載・複製を行う場合は、予め日本銀行情報サービス局 (post.prd8@boj.or.jp) までご相談下さい。転載・複製を行う場合は、出所を明記して下さい。

海運データを用いた輸出のナウキャストイング*

久野 遼平†・古川 角歩‡

2022年11月

【要 旨】

本稿では、これまで経済の動向を捕捉するのに活用されることが比較的少なかった、船舶の位置情報などを記録した海運ビッグデータ（AIS データ）を用いて、わが国の輸出動向をナウキャストイングする。その際、全国の港湾に関する地理データ等の公的統計や、機械学習の手法を活用しながら、予測精度の改善を図る。分析の結果、AIS データを用いることで、速報性が高い既存統計（貿易統計の上月速報値）と比較しても、高い精度でわが国の実質輸出の動向をナウキャストイングすることができた。特に、本稿で構築したナウキャストイングモデルは、輸出が大きく増減した場面（感染症が拡大した 2020 年春や、供給制約が顕在化した 2021 年央頃など）でも、その動きを相応にフォローできた。

JEL 分類番号： C49、C55、E27

キーワード：ナウキャストイング、オルタナティブデータ、AIS データ、輸出

* 本稿の作成にあたっては、青木浩介氏、城戸陽介氏、小出桂靖氏、中島上智氏、長野哲平氏、福永一郎氏、法眼吉彦氏、八木智之氏および日本銀行のスタッフから有益なコメントを頂戴した。記して感謝の意を表したい。ただし、残された誤りは筆者らに帰する。なお、本稿の内容や意見は、筆者ら個人に属するものであり、日本銀行の公式見解を示すものではない。

† 東京大学 (hisantor@g.ecc.u-tokyo.ac.jp)

‡ 日本銀行調査統計局 (kakuho.furukawa@boj.or.jp)

1. はじめに

近年、既存統計とは異なる情報源に基づく非伝統的なデータ、いわゆる「オルタナティブデータ」への関心が高まっている¹。オルタナティブデータを経済分析に活用するメリットは多岐に亘るが、その一つとして、多くのオルタナティブデータが既存統計と比較して速報性が高い（データが利用可能になるまでの時間が短い）点に着目し、足もとの経済情勢を迅速に捉えるためにこれらを活用するというものがある。この点、先行研究では、こうしたオルタナティブデータは、経済のナウキャストニング²において、パフォーマンス向上に資することも報告されている（Galbraith and Tkacz (2015)、Nyman-Andersen and Pantelidis (2018)、中澤 (2022)、大久保ほか (2022)、古川ほか(2022)）。

わが国経済のナウキャストニングについては、これまで、GDP (Hayashi and Tachi (2022)、Maehashi and Shintani (2020)、Hara and Yamane (2013)、Chikamatsu et al. (2021)、中澤 (2022))、個人消費（大久保ほか (2022)）、生産 (Shintani (2005)、古川ほか(2022)) などモデルの構築が行われてきた³。一方、GDP の約 2 割を占める輸出については、振れが大きいこともあり、それなりの精度でナウキャストニングすることが比較的困難であった。もともと、海外における最新の研究 (Arslanalp et al. (2019, 2021)、Cerdeiro et al. (2020)) では、これまで経済動向の捕捉に活用されることが比較的少なかった船舶の位置情報などを記録した海運ビッグデータ (AIS <Automatic Identification System>データ、詳細は 3 章) を活用することが、輸出の動向を早期に捕捉するうえで有益であると指摘されている。

そこで本稿では、既存研究を参考に、海運ビッグデータを活用してわが国輸出のナウキャストニングを行う。筆者らの知る限り、わが国で海運ビッグデータを輸出のナウキャストニングに活用した研究は本稿が初めてである。なお、わが国の輸出において海上貿易は名目金額の約 7 割を占めているため、海運ビッグデータを用いることで輸出の傾向を追うことができると考えられる (図表 1)。本稿の特徴は、AIS データと、全国の港湾に関する地理データ等を組み合わせ、わが国輸出に関連する船舶を識別しているところである。こうして絞り込まれた

¹ 本稿では、亀田 (2021) と同様に、伝統的な経済データ (月次や四半期のマクロ経済指標や上場企業の決算開示データ等) 以外のデータを総称して、オルタナティブデータと呼んでいる。

² ナウキャストニングとは、将来の予測をするのではなく、現在の経済動向を捕捉することを指す。経済活動に関する公的統計は公表されるまでにラグを伴うことが一般的であるため、よりリアルタイムに近い形で経済動向を捕捉するための研究が近年増加している。

³ 海外における経済活動のナウキャストニングの代表的な先行研究としては、Giannone et al. (2008)が挙げられる。

船舶を基に「輸出指数」を作成し、これをナウキャストリングモデル（線形回帰モデル）の変数として活用する。また、「輸出指数」の作成に機械学習のアプローチも活用することで、ナウキャストリングモデルの精度の向上を図る。

本稿の貢献は大別して三つある。第一には、本稿で作成した輸出指数は、実質輸出の動きを趨勢的に捉えていることである。輸出に関する公的統計としては財務省「貿易統計」と日本銀行「実質輸出入の動向」が存在し、それぞれ翌月の下旬ごろに公表されているが⁴、本分析の手法を用いることで、公的統計よりも3週間近く早く輸出の動向を捕捉することが可能となる。

第二には、本稿の輸出指数は、同等の速報性を持つ既存統計（貿易統計の上旬速報）と比較しても、良好なパフォーマンスを示す傾向があることである。特に、輸出が大きく増減した場面、例えば、感染症が拡大した2020年春や、供給制約が顕在化した2021年央頃などでも、その動きを相応にフォローすることができた。これらの結果は、これまでわが国の経済分析においてあまり活用されてこなかったAISデータが、わが国の輸出動向を捕捉するうえでも有用であることを示唆している。

第三には、輸出活動に従事している船舶を精緻に識別する手法を考案すると共に、「輸出指数」の作成に機械学習の手法を用いることで、「輸出指数」により輸出活動を予測する精度を改善する手法を提示したことである。この点は、三つの貢献のうち最も重要であると考えている。具体的には、国内港湾の境界線に関する地理データを用いることで、船舶がどの港に出入りしたかを正確に捕捉したほか、カーネル法や深層学習といった機械学習モデルを応用することで、輸出関数を推計し、「輸出指数」を作成している。実質輸出に対するナウキャストリングモデルの予測精度を平均平方二乗誤差（Root Mean Squared Error、RMSE）で検証した結果、輸出指数の作成に機械学習を活用しなかったモデルでは、既存統計のみを利用したモデルと比べ、若干の改善しかなかった一方、輸出指数の作成に機械学習を活用したモデルでは予測精度が大きく改善した（RMSE:既存統計 5.80、輸出指数<機械学習なし> 5.70、輸出指数<機械学習あり> 3.64）。これは、実質輸出の予測には、非線形的な要素が強く、そうした関係を捉えるためには機械学習の手法が有用である可能性を示唆している。

本稿の構成は以下のとおりである。まず、2節では先行研究を概観し、本稿の特徴を整理する。3節では、本分析で使用するデータを説明し、4節では、本稿

⁴ 例えば、2022年6月分の輸出に関する「貿易統計」と「実質輸出の動向」は、2022年7月21日に公表された。なお、貿易統計については、毎月上旬までの実績に関する速報値（月上旬速報）が当月末、毎月上中旬までの実績に関する速報値（上中旬速報）が翌月上旬に公表される。

で用いるナウキャストモデルを解説する。5節では、ナウキャストモデルの推計結果とその含意について整理し、6節はまとめである。

2. 先行研究

本稿の分析は、主として、AISデータの活用、オルタナティブデータを用いたナウキャストモデル、機械学習を用いた経済予測という3つの先行研究と関連している。以下、この順で詳しくみていく。

第一に、本稿はAISデータを活用した貿易ナウキャストに関する先行研究と関連している(Adland et al. (2017)、Arslanalp et al. (2019、2021)、Cerdeiro et al. (2020))。Adland et al. (2017)は、タンカーに関するAISデータを用いることで、産油国などにおける原油の貿易量をナウキャストしており、分析の結果、パイプラインを通じた取引が大きいような時期や地域間については予測精度が悪化する局面はありつつも、AISデータは原油の貿易量のトレンドを捕捉するうえで有用であると指摘している。Arslanalp et al. (2019)は、マルタの貿易量について、AISデータを用いて貨物船の数、積載量を集計することでナウキャストを行い、船舶位置情報が貿易量の予測指標として有益であることを示した。また、Arslanalp et al. (2021)は、Arslanalp et al. (2019)の分析手法を発展させ、太平洋諸島諸国を対象に、各国の港湾の地理情報や海上コンテナ船定期航路に関する情報を用いることで、貿易量のナウキャスト精度を向上させた。Cerdeiro et al. (2020)は、全世界をカバーするAISデータを用いて、世界各国における海上貿易の輸出入の動向を捕捉するための国別・船舶種別(自動車運搬船、コンテナ船等)の指標を作成し、AISデータが地域的な貿易量(例えば日本からの自動車の輸出量)を捕捉するのに有用であることを示した。これらの指標は定期的にアップデートされ、国連商品貿易統計データベースにおいて公開されている^{5,6}。このように、AISデータを用いた貿易量のナウキャストに関する研究は近年増加している一方、現状、わが国に特化した研究は、筆者らの知る限りみられていない。本稿では、わが国沿岸部におけるAISデータに加えて、全国の港湾に関する地理データや、各港湾にお

⁵ <https://comtrade.un.org/data/ais>

⁶ なお、同データベースにはわが国に関する指標も含まれているが、常に最新のデータに更新されているわけではないため、ナウキャストには使用しづらい。また、5-2節で確認するとおり、本分析の手法を用いることで、Cerdeiro et al. (2020)よりも高い精度でわが国の輸出動向を捕捉することが可能である。

ける輸出活動などに関する公的統計も活用することで、わが国の輸出動向を高い精度でナウキャストする。

第二に、本稿は AIS データという（少なくとも経済動向の捕捉においては）非伝統的なデータを用いているという点で、オルタナティブデータを用いたナウキャストモデルに関連している。わが国に関するものとしては、例えば、中澤（2022）が GDP について、大久保ほか（2022）が個人消費について、古川ほか（2022）が製造業の生産活動について、それぞれオルタナティブデータを用いてナウキャストモデルを構築している。これらはいずれも、経済活動をリアルタイムに捕捉するうえでオルタナティブデータが有用であることを示している。これらの研究で用いられているデータ（インターネット上の検案件数（中澤（2022））、クレジットカードの使用履歴（大久保ほか（2022））、携帯電話の GPS データ（古川ほか（2022）））は、海外の先行研究でも多く使用されており、経済活動をナウキャストするうえでどのようなオルタナティブデータが有用かについて、急速に知見が蓄積されつつある。例えば、Galbraith and Tkacz（2015）は、カナダを対象にクレジットカード、小切手の取引データを用いた GDP のナウキャストモデルを構築し、取引データを含まないモデルと比べて予測精度が改善することを示した。また、Nyman-Andersen and Pantelidis（2018）は、欧州における新車販売（主要 10 개국）について、Google の関連カテゴリーの検索数（Google Trends）を含めたナウキャストモデルを構築し、Google Trends のデータを用いることで、予測精度が改善すると指摘している。本稿の分析からは、AIS データも経済活動のナウキャストに有用であることが分かった。

第三に、本稿は機械学習モデルを経済予測に応用している先行研究とも関連している。Bolhuis and Rayner（2020）は、伝統的な OLS による予測における、多重共線性や非線形性への対応といった欠点を指摘したうえで、こうした欠点への対処法の 1 つとして、機械学習モデルの活用が選択肢となると指摘している。オルタナティブデータを用いた分析においても、線形モデルに代えて機械学習を用いることで予測精度が改善したとの結果を得ている先行研究も多い（古川ほか（2022）、Anesti et al.（2021）、Ashwin et al.（2021）、Richardson et al.（2021））。例えば、Anesti et al.（2021）は、オルタナティブデータと一般的なマクロ統計を用いて、線形モデルと機械学習モデルを使用した場合の GDP 予測の結果を比較しており、特にオルタナティブデータを説明変数に含むときには機械学習モデルの導入による予測精度の改善が大きいとの結果を得ている⁷。本稿

⁷ この点について Anesti et al.（2021）は、オルタナティブデータを使用するとデータ量が増え、変数間の関係も複雑になることから、多重共線性や非線形性に対応できる機械学習を活用する

においても、AIS データを集計する過程でカーネル法や深層学習と呼ばれる機械学習を用いることで、ナウキャスト精度を向上できることが分かった。

3. 本分析で使用するデータ

3-1. AIS データ

AIS とは、船舶自動識別装置 (Automatic Identification System、AIS) と呼ばれ、洋上を航行する船舶同士が航行情報を相互に交換するための装置のことである。AIS からは、船舶の識別符号、船名、位置 (緯度・経度)、船底から船体と水面が交わる線までの距離を指す「喫水 (きっすい)」、目的地などを含んだデータが発信されており、付近を航行している他船のほか、陸上の受信局や衛星で受信することができる。AIS の船舶への搭載は、円滑な航行管制や海難防止などを目的に、国際条約によって一定の基準を満たす船舶について義務付けられている。わが国では、国際航海に従事している旅客船または 300 総トン以上の船舶、および 500 総トン以上の全船舶について搭載が義務付けられているため、輸出に従事するような大きさの船舶であれば原則全て AIS を搭載していると考えられる。

AIS が発した電波は、受信媒体 (陸上受信局や衛星) を持っていれば誰でも受信することができる。この点、近年は、小型衛星技術の発展等を受けて、AIS データを精度高く把握し提供する民間サービスが増えており、本分析では、VesselFinder 社が提供している日本周辺の AIS データを用いて分析を行う (2017 年 1 月 1 日~2022 年 3 月 31 日)⁸。その際、本稿では、輸出動向を捕捉するという本分析の目的を踏まえ、船舶の種類を自動車運搬船とコンテナ船に絞り込み、データ頻度を 6 時間単位に落とししたうえで取得した AIS データを使用する⁹。本邦の輸出においては、自動車運搬船とコンテナ船が輸出額に占める割合は 60%程度 (海上貿易額に占める割合は 85%程度) であることに加えて、輸出額の約 30%を占める航空輸送が海上コンテナ船による輸出と概ね同じ動き

ことのメリットが増加する可能性を指摘している。

⁸ サービス媒体によっては、最新の AIS データを数分~数十分程度のラグで提供しているケースもあり、オルタナティブデータの中でも特にリアルタイム性が強いデータである。

⁹ 日本周辺には旅客船など輸出に従事していない船舶も多数航行しているほか、これらも多くの場合 AIS は数秒~数十秒ごとに電波を発しているとみられ、この期間の全ての AIS データを取得すると膨大なデータ量となるため、本稿では、一定の絞り込みを行っている。また、わが国の財輸出の多くが本州、九州、四国で行われていることを踏まえて、分析対象エリアは北海道と沖縄を除く全国の沿岸部とした。本分析で使用したデータは、csv ファイルで 6,591,066 行、1.1GB のサイズとなっている。

をとる（後述）ことも踏まえると、今回分析に使用するデータは相応のカバレッジを有していると言える（図表 1）。本分析では、7,003 隻に関する AIS データを分析に使用する。今回取得したデータを基に、日本周辺における船舶の分布をみると、工業地帯近辺に多くの船舶が集まっている様子が窺える（図表 2）。

3-2. 港湾データ

AIS データをもとに、ある船舶が輸出活動に従事しているかどうかを識別するためには、その船舶が港湾内に位置しているかどうかを正しく判定する必要がある。本分析では、国土交通省が提供している「国土数値情報ダウンロードサービス」から、港湾法で指定された全国の港湾の境界線に関する地理データを取得し、各港湾のエリアを特定する。具体的には、同データとわが国の海岸線のデータと突合することで、港湾の境界線と海岸線に囲まれた港湾「エリア」を特定する（図表 3 (a)）。ただし、海峡や離島の存在によって港湾の境界線と海岸線のデータから港湾エリアを特定するのが困難なケースも一部には存在する。これらのケースのうち、貿易額が大きい港湾については手動でエリアを特定し（図表 3 (b)）、その他の小規模な港湾については分析対象から除外する。この結果、わが国の海上貿易による名目輸出額の 99%をカバー（2017～2020 年）する 302 港湾のエリアが特定された。

4. 「輸出指数」の算出

当節ではまず、AIS データと港湾データを用いて輸出動向を捕捉するための「輸出指数」を算出する方法について解説する。また、港湾別の統計データを活用し、輸出指数を改善する手法についても議論する。

4-1. 輸出指数の算出方法

（ステップ 1. 海外向け船舶の特定）

今回取得したデータの中には、国内から海外に向かって航行している船舶だけでなく、国内の港湾間を行き来している船舶等も含まれている。そのため、船舶の輸出活動を把握するために、まずはその船舶の次の目的地が海外かどうかを識別することが重要である。AIS データには次の目的地を入力した項目が含まれているものの、船員が手入力しているため、フォーマットが統一されておらず分析に活用するのが難しい¹⁰。そこで本分析では、各船舶の位置情報を継続的

¹⁰ 例えば、次の目的地が東京である場合、AIS データの目的地の項目には、「Tokyo」「JP TOKYO」「TYO」など様々なフォーマットで入力されていることがあり、実際の目的地と正しくマッチ

にトラッキングすることで、船舶が国内の港を出港した後に海外に向かっているか、あるいは国内の別の港に向かっているかを特定する。

具体的には、船舶が国内の港を出港したあとに、同船舶が発信する AIS データが、「①48 時間以上途絶える」、または「②24 時間以上途絶え、その期間の前後で喫水が変化している」のどちらか1つの条件でも満たす場合には、同船舶が「海外に向かった」と判断する。今回取得した AIS データは日本沿岸部を対象としているため、船舶の AIS データが長時間途絶えた場合には、同船舶が日本を離れ海外に向かったと考えることができる。①の条件では、その基準を 48 時間に設定した¹¹。ただし、例えば、北九州と韓国のように、国内外の港湾の距離が近い場合には、国内の港を出港してから 48 時間以内に再び国内に戻ってくるケースがある。このため、②の条件では、船舶の AIS データが途絶えた時間が短くてもその前後で喫水が変化していて、この間に何らかの積卸作業を行ったと考えられる場合には、その船舶が国内の港を出港した後に海外に向かったと識別している。

これらの条件に基づき、サンプル期間においてのべ 109,356 隻について次の目的地が海外であると識別された。

（ステップ 2. 貨物積載量の推計）

次に、AIS データに含まれている喫水に関する情報を用いることで、船舶の積載量（積荷の重量）を推計する。船舶の積載量が増すと、船がより水面深くに沈み喫水が深くなるため、喫水から船舶の積載量を逆算することが可能である。具体的には、「喫水は船舶の重量に比例する」と仮定することで、以下の式を用いて船舶 i の時点 d における積載量を計算することができる（図表 4）^{12,13}。

ングを行うことが難しい。

¹¹ なお、仮に国内の港から国内の港に向かっている場合でも、航路によっては一時的に分析対象エリアから離れることがあるほか、何らかの事情（一時的に AIS のスイッチを切るなど）で AIS データが短期的に途切れることがある。そのため、こういった船舶が海外に向かっていると誤って認識しないよう、十分に長いと思われる時間に設定した。

¹² 正確には、「船舶の重量は、船舶の水面下に沈んでいる部分の体積分の海水の重量に等しい」という法則が成り立つ（アルキメデスの原理）。ここで、船舶の形が直方体に近く、「船舶の水面下に沈んでいる部分の体積」が喫水に比例すると仮定すると、上記の通り「喫水は船舶の重量に比例する」という関係が得られる。

¹³ 実際には、積載量が小さいときには船体を安定させるために海水などの重しを積み込んで喫水をあえて深めることが多いため、積載量と喫水の関係が図表 4 のような比例関係にはなっていない可能性もある。

$$\text{積載量}_{i,d} = \text{最大積載量}_i \times \frac{\text{喫水}_{i,d} - \text{喫水}_{i,\min}}{\text{喫水}_{i,\max} - \text{喫水}_{i,\min}}, \quad (1)$$

ここで、喫水 $_{i,d}$ は船舶 i の時点 d における喫水であり、最大積載量 $_i$ 、喫水 $_{i,\max}$ 、喫水 $_{i,\min}$ はそれぞれ船舶 i が積載することができる貨物の最大重量、最大積載時の喫水、空荷時（荷物を積載していないとき）の喫水である。喫水 $_{i,d}$ はAISデータから取得し、最大積載量 $_i$ と喫水 $_{i,\max}$ は、VesselFinder社が管理している船舶データベースから取得した。喫水 $_{i,\min}$ については、サンプル期間における船舶 i の喫水の最小値（ \min_d 喫水 $_{i,d}$ ）を推計値として使用する。

（ステップ3. 輸出指数の計算）

最後に、ステップ1と2で得られたデータを使って輸出指数を計算する。一般的には、輸出を行う船舶は国内で複数の港に寄港して積込作業を行ってから海外に向かうため、輸出活動を正確に把握するためには各港における輸出品の積込状況を捕捉する必要がある。しかしながら、AISデータからは入港時と出港時の積載量しか推計できないため、各港における積込量を捕捉することはできない（図表5）。そこで、本分析では、海外に向かっている船舶が積載している貨物が全て輸出品であると仮定し、海外に向かっていると識別された船舶の出港時の積載量を合計することで、輸出指数を算出する。以下の式の通り、輸出指数は自動車運搬船とコンテナ船それぞれについて、月次で算出する。

$$\text{輸出指数}_{\text{type},t} = \sum_{\substack{i,d \\ i \in S_{\text{type}} \\ \text{目的地}_{i,d} = \text{海外}}} \text{積載量}_{i,d} \quad (2)$$

ここで t は月次の時点（例えば2022年1月）、 d は t 月内の時点、 type は船舶の種類（自動車運搬船またはコンテナ船）、 S_{type} は type に分類される船舶の集合、目的地 $_{i,d}$ は船舶 i の時点 d における次の目的地が海外か国内かを表す変数である¹⁴。

¹⁴ なお、(2)式では、船舶が国内の各港湾で積込んだ輸出品が全て、国内の最終寄港地を出港した時点で輸出指数に計上されている。貿易統計上は輸出品が各税関を通った時点で輸出額が計上されるため、本稿で算出している輸出指数と公的統計で輸出計上のタイミングが異なっている点には留意が必要である。

4-2. 輸出指数の算出結果

上節で作成した輸出指数を実質輸出（実績）と比較すると、自動車については前年比の相関係数が 0.89 となっており、高い精度で実質輸出の動きを捕捉できている（図表 6（a））。特に、本稿の輸出指数は、2020 年春の新型コロナウイルス感染症拡大に伴う輸出の大幅な落ち込みをはっきりと捉えることができおり、自動車運搬船の輸出指数が実際の自動車輸出の動向を捕捉するうえで有用であることを示唆している。

一方、コンテナ船の輸出指数については、実質輸出との前年比の相関係数が 0.42 となっており、大まかなトレンドは一致しているものの、自動車運搬船と比較するとあまり精度が高くない結果となった（図表 6（b））。特に、2021 年春以降は前年比が実質輸出を恒常的に下回った状態が続いている。

4-3. コンテナ船に関する輸出指数の改善

自動車運搬船と比較してコンテナ船の輸出指数の精度が高くない要因としては、コンテナ船で輸出される品目の種類が幅広く、積荷によって重量当たりの金額が大きく変動し得ることが考えられる。実際、(2) 式が示す通り、輸出指数は輸出品の重量に基づいて算出されているため、重量当たり金額の変動を捉えることはできない。今回の分析期間においても、特に 2021 年以降は、半導体関連の輸出が増加し、重量当たり金額が大きい品目の割合が高まっていた可能性がある（図表 7）。この間の輸出指数が実際の輸出金額と比較して低めの伸び率で推移していたことの背景には、こういった構成品目の変化による影響もあったと考えられる。

このほか、輸出品が収納されているコンテナ自身の重さがノイズとなって、コンテナ船の輸出指数の精度に影響している可能性もある¹⁵。輸出品が軽かったり、コンテナの稼働率が低かったりすると、船舶の積載量に占めるコンテナ自身の重さの割合が高くなり、輸出品の動向を正確に把握することが困難になる可能性もある¹⁶。

これらの点を踏まえると、コンテナ船による輸出動向を正確に捕捉するためには、「積荷の重量当たり金額」や「積荷に占める輸出品以外の重量」を考慮す

¹⁵ 最も標準的なサイズのコンテナ（20 フィートコンテナ）の場合、メーカーによって幅はあるものの、コンテナ自身の重さが 2t 程度、最大積荷重量が 20～30t 程度であり、コンテナ船の推計積載量の 1 割以上はコンテナ自身の重さだと思われる。

¹⁶ ロジスティクスの都合上、コンテナの中身が空のまま船舶に積込まれることが少なくない点にも留意が必要である。実際、港湾統計（国土交通省）によると、2020 年には、全国の港湾で積込まれたコンテナのうち約 40%が空コンテナだった。

ることが重要と考えられる。これらのデータは、そもそも船舶レベルでは存在しないため、正確な計数を得ることはできないが、船舶が積卸を行った港にある程度依存して決まっていると考えられる。実際、港湾別の「輸出量当たり輸出額」、「コンテナ当たり重量」、および「出荷量に占める輸出の割合」をみると、港別に特徴がある（図表 8）。このため、積荷に関する船舶レベルの詳細なデータがなくても、船舶がどの港に寄港したかをみることで、同船舶に積込まれた積荷の特徴（積荷の重量当たり金額等）を推測することが可能である¹⁷。

以上の考察を踏まえ、本分析では下記の「輸出関数」を港別に推計することで、コンテナ船による輸出額のナウキャスト精度を高めることを試みる。

$$\text{輸出額}_{i,p,d} = f(\text{入港時積載量}_{i,p,d}, \text{出港時積載量}_{i,p,d}, \text{寄港時間}_{i,p,d}, \theta_p) \quad (3)$$

ここで添え字 i, p, d は船舶 i が港 p を時点 d に出港したことを表し、輸出額 $_{i,p,d}$ 、入港時積載量 $_{i,p,d}$ 、出港時積載量 $_{i,p,d}$ 、寄港時間 $_{i,p,d}$ はそれぞれ出港時に積んでいる輸出品の金額（実質ベース）、入港時の積載量、出港時の積載量、および港 p に滞在していた時間を表す。説明変数に含まれている入港時と出港時の積載量の差分等は、船舶がその港で積込んだ積荷の量を表していると考えられる。ただし、船舶が積込と積降を同時に行った場合は、積載量の変化だけからは積込量を正確に予測することはできない可能性がある。そこで、積込と積降の量が多いほど寄港時間が長くなるとの想定のもと、寄港時間を追加的に考慮することで、こうした問題に対処している。 θ_p は港 p の特性を反映したパラメータであり、輸出額の精度が最も高くなるような値を推計する。港別にパラメータを推計することで、例えば「輸出量当たり輸出額が高い港を出港した船舶について積載量当たりの輸出額を多めに見積もる」といったことが可能となり、ナウキャスト精度の改善につながることを期待できる。

(3) 式の輸出関数が推計されたら、船舶レベルの輸出額を足し上げることで、コンテナ船に関する輸出指数（改良版）を算出する。

$$\text{輸出指数}_{\text{container},t}^r = \sum_{i,p,d} \text{輸出額}_{i,p,d} \quad (4)$$

ここで d は t 月内の時点である。

¹⁷ 例えば、東京港で積込まれた輸出品は川崎港で積込まれた輸出品と比べて重量当たり金額が高い傾向があるため、東京港を出港した船舶の積荷は、川崎港を出港した船舶の積荷と比較して、重量当たり金額が高いと推測することができる。

なお、(3) 式の輸出関数がどのような形の関数であるかは先験的に判断できないため、非線形な関数を扱うことができる「カーネル法」と「深層学習」を使用して (3) 式を推計する。推計には 2017 年 1 月～2020 年 12 月のデータを使用する。以下では、両手法の概要を解説する。

4-3-1. カーネル法を用いた輸出関数の推計

カーネル法による (3) 式の推計では、説明変数 \mathbf{x} をより高度な次元におけるベクトル $\phi(\mathbf{x})$ に変換し、輸出関数が $\phi(\mathbf{x})$ の線形関数で表せると仮定する。本分析のケースでは、輸出関数は以下のように書き表すことができると仮定する。

$$\text{輸出額}_{i,p,d} = \boldsymbol{\theta}_p^T \phi(\text{入港時積載量}_{i,p,d}, \text{出港時積載量}_{i,p,d}, \text{寄港時間}_{i,p,d}) \quad (5)$$

ここで、上添え字の T は転置を意味する。通常では、 $\text{輸出額}_{i,p,d}$ の予測誤差が最小になるように輸出関数を推計するが、本分析ではそもそも $\text{輸出額}_{i,p,d}$ のデータが存在しない。そこで、以下のような損失関数を設定し、港レベルの輸出額の予測誤差を最小化することで、 ϕ が与えられた時の輸出関数を推計する¹⁸。

$$L_p = \frac{1}{2} \sum_t \left\{ \sum_{i,d} \boldsymbol{\theta}_p^T \phi(\mathbf{x}_{i,p,d}) - \text{輸出額}_{p,t} \right\}^2 + \frac{\lambda}{2} \boldsymbol{\theta}_p^T \boldsymbol{\theta}_p \quad (6)$$

ここで $\mathbf{x}_{i,p,d} = (\text{入港時積載量}_{i,p,d}, \text{出港時積載量}_{i,p,d}, \text{寄港時間}_{i,p,d})$ であり、 $\text{輸出額}_{p,t}$ は港 p における t 月の海上コンテナ輸出額（実質ベース）である¹⁹。(6) 式は、 t 月に港 p を出港した全ての船舶についてその輸出額を合計した値と $\text{輸出額}_{p,t}$ の誤差が最も小さくなるように、輸出関数を推計することを意味している（ d は t 月内における時点）。右辺第 2 項は、パラメータ $\boldsymbol{\theta}_p$ の値が大きくなり過ぎないように調整する正則化項である。当小節では、カーネル法の概要と、推計結果について解説する。詳細な推計方法については補論を参照されたい。

カーネル法では、(6) 式を変形することで、輸出関数を元のベクトルである $\phi(\mathbf{x})$ ではなく、その内積 $\phi(\mathbf{x})^T \phi(\mathbf{x}')$ で書き表す。 $\phi(\mathbf{x})^T \phi(\mathbf{x}')$ はカーネル関数と呼ばれ、

¹⁸ なお、Python や R など多くのプログラミング言語において (5) 式を推計するためのパッケージや関数が用意されているが、上述の通り、 $\text{輸出額}_{i,p,d}$ のデータが存在しないため、本分析ではこれらの機能を使うことができない。補論では、本分析の文脈において、これらの機能を使わずにカーネル法を用いて輸出関数を推計する方法について解説する。

¹⁹ 港 p における t 月の海上コンテナ輸出額（実質ベース）は、 t 月の海上コンテナ輸出額（実質ベース、全国分）の値を、港 p の同年の輸出額（名目ベース、海上コンテナ以外の輸送手段も含む）で案分することで算出した。

慣例的に $k(\mathbf{x}, \mathbf{x}')$ と表記される。 $\phi(\mathbf{x})$ が複雑でも $k(\mathbf{x}, \mathbf{x}')$ は比較的簡単に計算できることがあるため、カーネル法を用いることで、様々な ϕ のパターンについて計算量を抑えながら推計し、精度が高くなるよう輸出関数の形を効率的に探索することが可能となる。図表 9 では、様々なカーネル関数を試行したときの輸出関数の精度を示している。本稿では、アウトオブサンプル期間における精度を踏まえて、「ガウシアンカーネル関数」を用いた輸出関数を採用した。

同関数に基づいて、積載量と輸出額の関係をみると、出港時の積載量が入港時の積載量を下回るときは輸出をほとんど行っておらず、出港時の積載量が入港時の積載量を上回るときはその差分が増えるにつれて輸出額も増加する、という関係が成り立っている（図表 10 (a)）。また、寄港時間と輸出額の関係をみると、寄港時間が長いほど輸出額も増加している（図表 10 (b)）。

4-3-2. 深層学習法を用いた輸出関数の推計

当小節では (3) 式における輸出関数の推計において深層学習を用いる手法について解説する。まず、深層学習とは、機械学習の一手法であり、神経細胞を模したニューロンの組み合わせによって関数を表現するモデルである（図表 11）。一般的に、中間層の数が多く複雑なものを深層学習モデルと呼ぶ。深層学習は説明変数と被説明変数の関係について先験的な仮定を置くことなく推計できるため、カーネル法と比較しても汎用性が高い手法であると言える。もっとも、その代わりにパラメータの数も非常に大きなものとなりやすく、データ数が少ないと推計結果が安定しないため、アウトオブサンプルの予測精度が必ずしも改善するとは限らない。本分析では、カーネル法と同様に、輸出額 $i_{p,d}$ が入港時積載量 $i_{p,d}$ 、出港時積載量 $i_{p,d}$ 、寄港時間 $i_{p,d}$ の関数で表されるものとして、 t 月に港 p を出港した全ての船舶の輸出額を合計した値と輸出額 p,t の誤差が最も小さくなるように深層学習モデルを推計した^{20,21}。

深層学習を用いて推計した輸出関数に基づいて、積載量や寄港時間と輸出額の関係をみると、カーネル法と同様に、「出港時の積載量が入港時の積載量を下回るときは輸出をほとんど行っておらず、出港時の積載量が入港時の積載量を上回るときはその差分が増えるにつれて輸出額も増加する」、「寄港時間が長い

²⁰ 本稿では、多層パーセプトロンの中間層数や中間層の次元数の組み合わせについて様々なパターンを推計し、アウトオブサンプルの予測精度が最も高かった、中間層数が 6 層、中間層の次元数が 300 の単純な深層学習モデルを採用した。同モデルでは約 60 万個のパラメータを使用している。

²¹ 深層学習の具体的な推計方法を解説することは本稿の範疇を超えるため、ここでは省略する。詳細については、例えば Goodfellow et al. (2016) を参照。

ほど輸出額も増加する」といった関係が成立している（図表 12）。

4-3-3. 輸出指数（改良版）の算出結果

輸出指数（改良版）について、アウトオブサンプル期間（2021年1月～2022年3月）における実質輸出とのフィットをみると、カーネル法、深層学習どちらの手法を用いた場合でも、改良前よりも改善している（図表 13）。港別に輸出関数を推計することで、輸出品の構成比の変化等の影響も捕捉しながら、実質輸出の動きをより上手く捉えられるようになった可能性がある。なお、カーネル法と深層学習を比較すると、2021年前半の回復局面ではカーネル法の方が、2021年後半の減少局面では深層学習の方が、それぞれ実質輸出を上手く捉えられており、一概に優劣が言えない結果となった。

5. ナウキャストモデルの計算結果

5-1. ナウキャストモデルの精度

当節では、本稿で作成した船舶種類別の輸出指数を用いて、わが国全体の実質輸出のナウキャストを試みる。ナウキャストは以下の 2 ステップに分けて行う。第 1 ステップでは、実質輸出（原計数）の前年比を、各輸出指数の前年比で OLS 推計する。第 2 ステップでは、第 1 ステップで得られた実質輸出（原計数）の前年比の推計値を、実質輸出の季節指数を用いることで季節調整済前期比の値に変換する²²。また、比較対象として、輸出指数と同様に月末時点で利用可能な既存統計である貿易統計・上旬速報を、為替変動等を反映して実質化したうえで使用し、上記と同様のステップで実質輸出をナウキャストする検証用のモデルも推計する。

まず、全てのサンプル（2018年1月～2022年3月）を使ったステップ 1 の OLS の推計結果をみると、輸出指数を使ったモデルの方が既存統計（貿易統計の上旬速報、実質ベース）を使ったモデルよりも自由度調整済決定係数が高く、実質輸出（前年比）に対する説明力が高いことが確認できる（図表 14）²³。

²² 経済情勢を判断するうえでは季節調整済の系列を確認することが一般的であるため、本分析でも季節調整済ベースで実質輸出の動向を捉えることを一つの焦点とした。今回使用するオルタナティブデータの時系列は短く、輸出指数の季節調整を直接かけることができないため、まずは原計数ベースでモデルを推計するという 2 ステップを踏んでいる。

²³ なお、頑健性チェックの一環で、図表 14 の各定式化において実質輸出（原計数）の前年比の 1 期ラグ項を説明変数に追加して推計したところ、同ラグ項の係数は有意にならず、他の変数の有意性にも影響がなかったため、本分析では同ラグ項を含めない定式化を採用することと

次に、2021年1月～2022年3月をアウトオブサンプル期間として、この間のナウキャストモデルの予測精度について、機械学習の手法なども利用した以下の4つのモデル（以下のモデル①～④）で比較する。具体的に、モデル①は、既存統計（貿易統計の上旬速報、実質ベース）のみを使用したもの、モデル②は機械学習を活用せずに計算した輸出指数（4-2節で算出）を使用したもの、モデル③はカーネル法で計算したコンテナ船の輸出指数を使用したもの、モデル④は深層学習で計算したコンテナ船の輸出指数を使用したものを表す。モデル①と②の比較することで、より期近の情報を含むモデル②で予測精度がどの程度改善するかを検証できる。この点、モデル①のRMSEは5.80、モデル②は同5.70となっており、これを踏まえると、より直近の情報を利用することによる予測精度の改善は限定的であることが示唆された。一方、機械学習を使用したモデル③と④をみると、モデル③のRMSEは5.04、モデル④は同3.64と、深層学習を用いることで実質輸出の予測精度が大幅に向上した（図表15）^{24,25}。また、感染症が拡大した2020年春や、供給制約が顕在化した2021年央頃の動きをみると、モデル④では輸出の大幅な減少を相応にフォローできていた一方、モデル①では予測を大きく外していた。モデル①については、既存統計に当月の中旬以降の情報が含まれていないことが予測精度の低さの主な要因と考えられる。そのうえで、仮に中旬以降の情報があっても、機械学習の手法等で非線形性を相応に考慮しないとパフォーマンスが改善しないということが今回の含意でもある。

これらの結果は、輸出指数がわが国の輸出動向をリアルタイムに捕捉するうえで有用であることを示している。また、実質輸出の予測は、非線形的な要素が強く、そうした関係を捉えるためには機械学習の手法が有用である可能性を示唆している。この点、Anesti et al. (2021)は、オルタナティブデータを使用するとデータ量が増え、変数間の関係も複雑になることから、機械学習を活用することで、非線形性などに対応できるメリットがあるとの考え方と整合的な結果と解釈することができる。

した。

²⁴ アウトオブサンプル期間における予測値は、ローリングで推計した。すなわち、例えば2021年1月の予測値は2020年12月までの実績データを用いて推計したモデルを使って計算し、2021年2月の予測値は2021年1月までの実績データを用いて推計し直したモデルを使って計算した。この作業を2022年3月分まで繰り返すことで、各ナウキャストモデルの予測精度を比較した。

²⁵ 本分析では、機械学習の手法を用いて輸出指数を算出する際に十分なサンプル数を確保する観点から、2021年1月以降をアウトオブサンプル期間に設定した。この点、頑健性チェックのため2020年1月以降をアウトオブサンプル期間に設定してモデル①～④を推計し直しても、モデル①から④にかけて予測精度が改善していく姿に変わりがないことを確認している。

5-2. 先行研究との比較

上述の通り、AIS データを使用した先行研究の 1 つである Cerdeiro et al. (2020)は、わが国を含む世界各国について、同データを用いて船舶の種類別（自動車運搬船、コンテナ船、等）に輸出入の動向を捕捉するための指標を作成している。そこで当節では、Cerdeiro et al. (2020)の指標と本分析で算出した「輸出指数」のパフォーマンスを比較し、本分析で提示した手法を用いることでナウキャスト精度が向上しているかを確認する。

まず、Cerdeiro et al. (2020)が算出した、わが国における自動車運搬船及びコンテナ船に関する輸出指標を実際の輸出額と比較すると、相関係数はそれぞれ 0.70、0.39 となっており、本分析の「輸出指数」よりも低くなっている（図表 16）。次に、わが国全体の実質輸出額（前年比）に対する OLS 推計結果をみても、自由度調整済み決定係数は本分析の方が高い（図表 17）。また、2021 年 1 月～2022 年 3 月をアウトオブサンプル期間としてナウキャストモデルの予測精度を評価しても、本稿の輸出指数を用いたモデルの方がパフォーマンスが良好である（図表 18）。これらの結果は、本分析で算出した「輸出指数」が、Cerdeiro et al. (2020)と比較しても、わが国の輸出動向をナウキャストするうえでより有用であることを示唆している。

この背景には、本稿と Cerdeiro et al. (2020)の分析アプローチの次のような違いがあると考えられる。まず、港湾のエリアを特定するにあたって、本稿では公式な地理情報データを活用し精緻に行っている一方で、Cerdeiro et al. (2020)は船舶の位置情報を基に機械的な推計を行っている点である²⁶。この結果、本稿の方がより正確に、かつ高いカバレッジでわが国の港湾のエリアを特定することができていると考えられる。次に、コンテナ船に関する輸出指数について、本稿では港別の輸出関数を推計することでその精度を高めているが、Cerdeiro et al. (2020)ではそういった工夫は取られていない点である。このように、公的統計を活用したり、機械学習の手法も用いて船舶種類に応じて指標を精緻化する工夫をしたりしたことが、ナウキャスト精度の向上につながったと考えられる。

²⁶ 具体的には、停泊している船舶の位置情報を基に、多くの船舶が集まっているエリアを港湾と識別するような教師無し学習を行っている。

6. まとめ

本分析では、船舶の位置情報などを記録したオルタナティブデータである AIS データを用いて「輸出指数」を算出し、わが国の輸出動向をナウキャストした。分析結果をみると、本稿の輸出指数は実質輸出の動きを趨勢的に捉えており、同等の速報性を持つ既存統計（貿易統計の上旬速報）と比較しても、良好なパフォーマンスを示した。特に、輸出が大きく増減した場面、例えば、感染症が拡大した 2020 年春や、供給制約が顕在化した 2021 年央頃などでも、その動きを相応にフォローすることができた。また、本分析では、輸出指数を算出するにあたって、全国の港湾に関する地理データや各港湾における輸出活動などに関する公的統計を用いながら、機械学習モデルを応用する手法を提示し、先行研究対比で、ナウキャストの精度をはっきりと改善した。これらの結果は、これまでわが国経済分析においてあまり活用されてこなかった AIS データが、わが国の輸出動向を捕捉するうえで有用であることを示唆している。

なお、本稿では、各港湾における船舶の出入りと輸出額の関係がある程度安定していることを前提としている。例えば、産業構造の変化等を背景として、ある港から輸出される品目の内訳が変わり、その港から出港する船舶の積載量と輸出額の関係が変われば、本稿で算出した輸出指数の前提が変化し、ナウキャストモデルの精度が低下する可能性がある点には留意していく必要がある。今後、AIS データを貿易ナウキャストに活用していく際には、こうした可能性にも配慮しつつ、必要に応じて輸出指数の算出に用いる公的統計をアップデートするなど、分析精度を高めるための工夫を続けていくことが重要であると考えられる。

本稿で用いた手法は自動車運搬船やコンテナ船以外にも適用することができるため、石油タンカーや液化天然ガス（LNG）運搬船等のデータも使ってわが国の輸入動向に関するナウキャストモデルを作ることは、今後の有望な分析課題であると考えられる。輸出動向と輸入動向のナウキャストモデルが揃えば、貿易収支に関するナウキャストを行うことも可能となる。

【参考文献】

- 大久保友博・高橋耕史・稲次春彦・高橋優豊 (2022)、『『オルタナティブデータ消費指数』の開発：オルタナティブデータを用いた個人消費のナウキャストリング』、日本銀行ワーキングペーパーシリーズ、No. 22-J-9
- 亀田制作 (2021)、「オルタナティブデータを用いた日銀リサーチの紹介」、日銀レビュー・シリーズ、No. 2021-J-16
- 中澤崇 (2022)、「オルタナティブデータを用いたGDPナウキャストリングモデルの構築」、日本銀行ワーキングペーパーシリーズ、No. 22-J-1
- 古川角歩・箕浦征郎・久野遼平・八木智之 (2022)「オルタナティブデータと機械学習アプローチを用いた鉱工業生産ナウキャストリングモデルの構築」、日本銀行ワーキングペーパーシリーズ、No. 22-J-15
- Adland, R., H. Jia and S.P. Strandenes. (2017), "Are AIS-based Trade Volume Estimates Reliable? The Case of Crude Oil Exports," *Maritime Policy & Management*, vol. 44(5), pp. 657-665.
- Anesti, N., E. Kalamara, and G. Kapetanios (2021), "Forecasting UK GDP Growth with Large Survey Panels," Bank of England Staff Working Paper, No. 923
- Arslanalp, S., M. Marini, and P. Tumbarello (2019), "Big Data on Vessel Traffic: Nowcasting Trade Flows in Real Time," IMF Working Papers, 19/275.
- Arslanalp, S., R. Koepke, and J. Verschuur (2021), "Tracking Trade from Space: An Application to Pacific Island Countries," IMF Working Papers, 21/225.
- Ashwin, J., E. Kalamara, and L. Saiz (2021), "Nowcasting Euro Area GDP with News Sentiment: A Tale of Two Crises," ECB Working Paper Series, No. 2616.
- Bolhuis, M. A. and B. Rayner (2020), "Deus ex Machina? A Framework for Macro Forecasting with Machine Learning," IMF Working Papers, Vol. 2020, Issue 45.
- Cerdeiro, D. A., A. Komaromi, Y. Liu, and M. Saeed (2021), "World Seaborne Trade in Real Time: A Proof of Concept for Building AIS-based

- Nowcasts from Scratch," IMF Working Papers, 20/57.
- Chikamatsu, K., N. Hirakata, Y. Kido, and K. Otaka (2021), "Mixed-frequency Approaches to Nowcasting GDP: An Application to Japan," *Japan and the World Economy*, vol.57: 101056.
- Galbraith, J. W. and G. Tkacz (2015), "Nowcasting GDP with Electronic Payments Data," ECB Statistics Paper, No. 10.
- Giannone, D., L. Reichlin, and D. Small (2008), "Nowcasting: The Real-time Informational Content of Macroeconomic Data," *Journal of Monetary Economics*, vol. 55(4), pp. 665-676.
- Goodfellow, I., Y. Bengio, and A. Courville (2016), *Deep Learning*, MIT Press.
- Nymand-Andersen, P. and E. Pantelidis (2018), "Google Econometrics: Nowcasting Euro Area Car Sales and Big Data Quality Requirements," ECB Statistics Paper No. 30.
- Hara, N., and S. Yamane (2013). "New Monthly Estimation Approach for Nowcasting GDP Growth: The Case of Japan," Bank of Japan Working Paper Series, No. 13-E-14.
- Hayashi, F. and Y. Tachi (2022), "Nowcasting Japan's GDP," *Empirical Economics*, 2022.
- Maehashi, K. and M. Shintani, (2020), "Macroeconomic Forecasting Using Factor Models and Machine Learning: An Application to Japan," *Journal of the Japanese and International Economies*, vol. 58(C).
- Nymand-Andersen, P. and E. Pantelidis (2018), "Google Econometrics: Nowcasting Euro Area Car Sales and Big Data Quality Requirements," ECB Statistics Paper No. 30.
- Richardson, A., M. T. van Florenstein, and T. Vehbi (2021), "Nowcasting GDP Using Machine-learning Algorithms: A Real-time Assessment," *International Journal of Forecasting*, vol. 37(2), pp. 941-948.
- Shintani, M. (2005), "Nonlinear Forecasting Analysis Using Diffusion Indexes: An Application to Japan," *Journal of Money, Credit and Banking* vol. 37(3) pp. 517-538.

(補論) カーネル法の推計方法

当補論では、(6) 式からカーネル法を用いて輸出関数を推計する方法について解説する。まず、損失関数 L_p をパラメータ $\boldsymbol{\theta}_p$ で微分することで、以下の式を得る。

$$\frac{\partial L_p}{\partial \boldsymbol{\theta}_p} = \sum_t \left\{ \sum_{i,d} \boldsymbol{\theta}_p^T \phi(\mathbf{x}_{i,p,d}) - \text{輸出額}_{p,t} \right\} \cdot \sum_{i,d} \phi(\mathbf{x}_{i,p,d}) + \lambda \boldsymbol{\theta}_p \quad (\text{A1})$$

パラメータ $\boldsymbol{\theta}_p$ が損失関数を最小化するとき、上式は 0 に等しくなることから、以下の式を得る。

$$\boldsymbol{\theta}_p = -\frac{1}{\lambda} \sum_t \left\{ \boldsymbol{\theta}_p^T \psi(\mathbf{x}_{p,t}) - \text{輸出額}_{p,t} \right\} \cdot \psi(\mathbf{x}_{p,t}) \quad (\text{A2})$$

ただし、 $\psi(\mathbf{x}_{p,t}) = \sum_{i,d} \phi(\mathbf{x}_{i,p,d})$ (d は t 月内の時点) である。ここで、 $-\frac{1}{\lambda} \sum_t \left\{ \boldsymbol{\theta}_p^T \psi(\mathbf{x}_{p,t}) - \text{輸出額}_{p,t} \right\}$ を t 番目の要素に持つベクトルを \mathbf{a} 、 $\psi(\mathbf{x}_{p,t})^T$ を t 番目の行に持つ行列を $\boldsymbol{\Phi}$ とすると、上式は $\boldsymbol{\theta}_p = \boldsymbol{\Phi}^T \mathbf{a}$ と書き直すことができる。これを (6) 式の損失関数 L_p に代入することで、以下の式を得る。

$$L_p = \frac{1}{2} \mathbf{a}^T \boldsymbol{\Phi} \boldsymbol{\Phi}^T \boldsymbol{\Phi} \boldsymbol{\Phi}^T \mathbf{a} - \mathbf{a}^T \boldsymbol{\Phi} \boldsymbol{\Phi}^T \mathbf{y} + \frac{1}{2} \mathbf{y}^T \mathbf{y} + \frac{\lambda}{2} \mathbf{a}^T \boldsymbol{\Phi} \boldsymbol{\Phi}^T \mathbf{a} \quad (\text{A3})$$

ただし、 $\mathbf{y}^T = (\text{輸出額}_{p,1}, \text{輸出額}_{p,2}, \dots)$ である。ここで、 $\boldsymbol{\Phi} \boldsymbol{\Phi}^T$ を \mathbf{K} と表すと、(A3) 式はさらに以下のように書くことができる。

$$L_p = \frac{1}{2} \mathbf{a}^T \mathbf{K} \mathbf{K} \mathbf{a} - \mathbf{a}^T \mathbf{K} \mathbf{y} + \frac{1}{2} \mathbf{y}^T \mathbf{y} + \frac{\lambda}{2} \mathbf{a}^T \mathbf{K} \mathbf{a} \quad (\text{A4})$$

(A4) 式は、損失関数 L_p を新しいパラメータ \mathbf{a} で書き表した形になっている。そこで、同式をパラメータ \mathbf{a} で微分することで、 L_p を最小化する \mathbf{a} を次のように求めることができる。

$$\mathbf{a} = (\mathbf{K} + \lambda \mathbf{I})^T \mathbf{y} \quad (\text{A5})$$

ここで \mathbf{I} は単位行列である。行列 \mathbf{K} の n 行 m 列目の要素は以下のように書くことができる。

$$\begin{aligned} K_{nm} &= \psi(\mathbf{x}_{p,n})^T \psi(\mathbf{x}_{p,m}) \\ &= \sum_{i,n'} \phi(\mathbf{x}_{i,p,n'})^T \sum_{j,m'} \phi(\mathbf{x}_{j,p,m'}) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i,j,n',m'} \phi(\mathbf{x}_{i,p,n'})^T \phi(\mathbf{x}_{j,p,m'}) \\
&= \sum_{i,j,n',m'} k(\mathbf{x}_{i,p,n'}, \mathbf{x}_{j,p,m'}) \tag{A6}
\end{aligned}$$

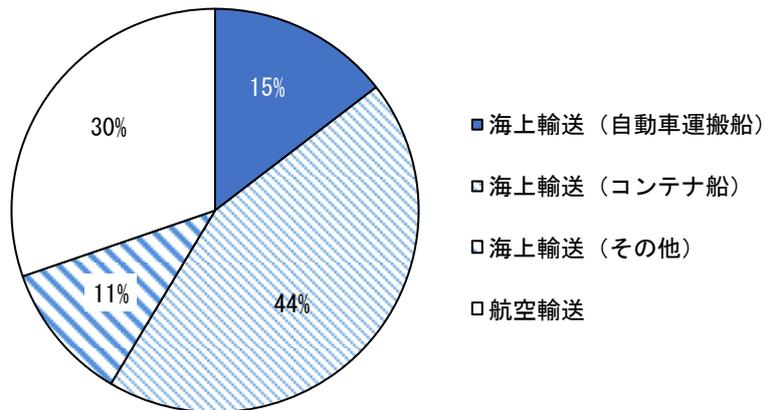
ここで、 n' 、 m' はそれぞれ n 月、 m 月内における時点である。(A6) 式における $k(\mathbf{x}_{i,p,n'}, \mathbf{x}_{j,p,m'}) = \phi(\mathbf{x}_{i,p,n'})^T \phi(\mathbf{x}_{j,p,m'})$ はカーネル関数と呼ばれるものである。以上の計算結果を組み合わせることで、以下の予測関数が得られる。

$$\begin{aligned}
\text{輸出額}_{i,p,d} &= \boldsymbol{\theta}_p^T \phi(\mathbf{x}_{i,p,d}) \\
&= \mathbf{a}^T \boldsymbol{\Phi} \phi(\mathbf{x}_{i,p,d}) \\
&= \mathbf{k}(\mathbf{x}_{i,p,d})^T (\mathbf{K} + \lambda \mathbf{I})^T \mathbf{y} \tag{A7}
\end{aligned}$$

ただし、 $\mathbf{k}(\mathbf{x}_{i,p,d})$ は n 番目の要素が $\sum_{j,n'} k(\mathbf{x}_{j,p,n'}, \mathbf{x}_{i,p,d})$ となるようなベクトルである。

(A7) 式が示す通り、カーネル法では輸出関数を元のベクトルである $\phi(\mathbf{x}_{j,p,m'})$ ではなくカーネル関数 $k(\mathbf{x}_{i,p,n'}, \mathbf{x}_{j,p,m'})$ で書き表す。 $\phi(\mathbf{x}_{j,p,m'})$ が複雑でも $k(\mathbf{x}_{i,p,n'}, \mathbf{x}_{j,p,m'})$ は比較的簡単に計算できることがあるため、カーネル法を用いることで、様々な ϕ のパターンについて計算量を抑えながら推計し、精度が高くなるよう輸出関数の形を効率的に探索することが可能となる。また、カーネル関数 $k(\mathbf{x}, \mathbf{x}')$ は、2つのベクトル \mathbf{x} と \mathbf{x}' が類似しているときに大きな値を取るという性質がある。このため、カーネル法を用いて推計した輸出関数は、「積載量や寄港時間の分布が類似している月の輸出額に基づいて足もとの輸出額を予測する」関数であると解釈することができる。

図表 1. 輸送手段別の輸出ウェイト



(注1) 2017年～2021年の名目輸出額における割合。
(注2) 自動車の輸出は全て自動車運搬船で行われると仮定。
(出所) 財務省

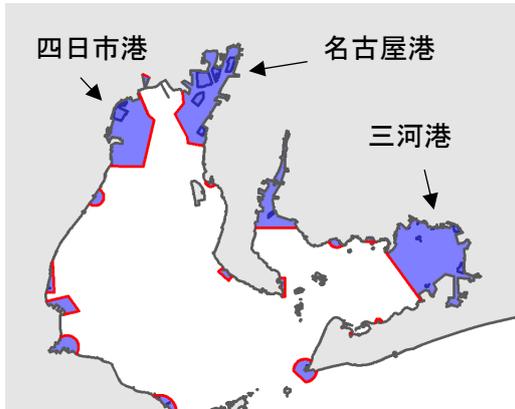
図表 2. 日本周辺における船舶の分布



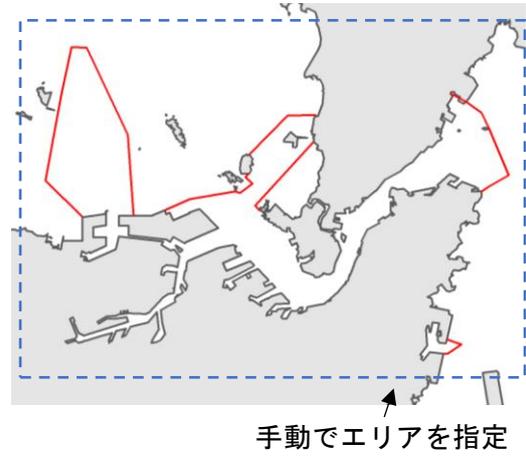
(注1) 2017年1月1日の船舶の分布。
(出所) VesselFinder 社

図表 3. 港湾データの概観

(a) 名古屋周辺

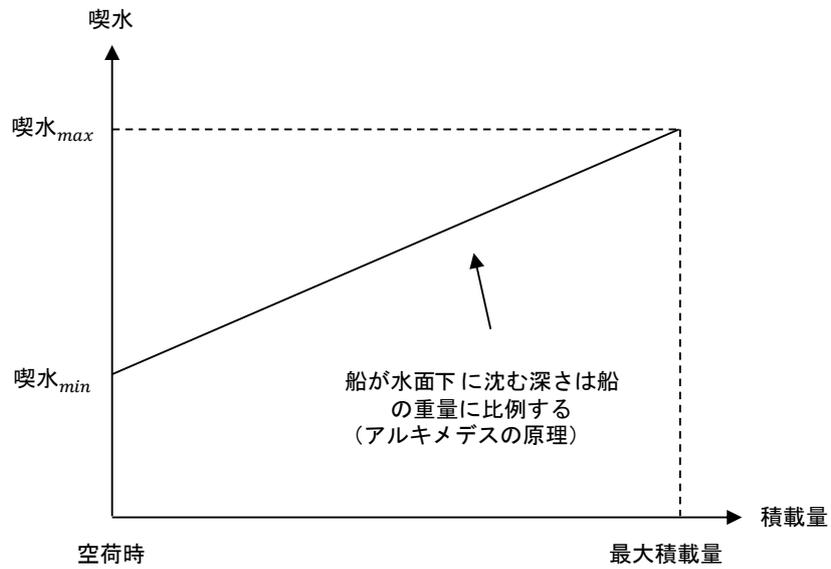


(b) 北九州周辺



(注 1) (a) は港湾エリアを自動で特定できる例、(b) は境界線が複雑なため港湾エリアを手動で特定する必要がある例。
 (注 2) 赤線は各港湾の境界線データ。特定された各港湾のエリアを青塗りで示している。
 (出所) 国土交通省

図表 4. 積載量と喫水の関係



図表 5. 船舶の積卸動向（イメージ）

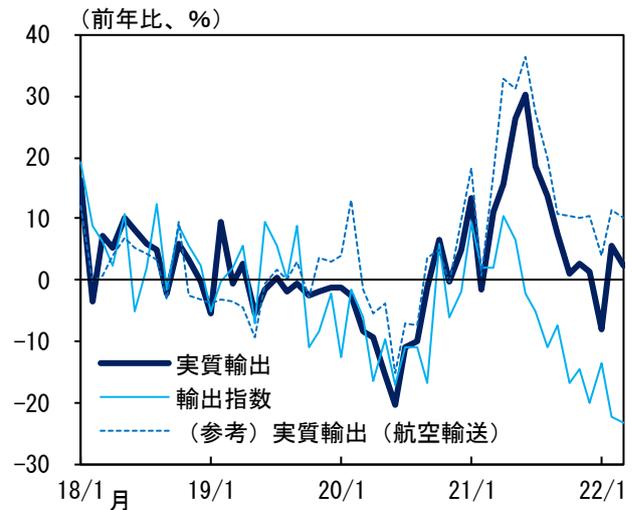
出港日時	港	入港時の積載量	積降量	積込量	(うち輸出向け)	出港時の積載量	次の目的地
4/1	鹿島	100	20 +	30	(10)	110	国内
4/4	東京	110	30 +	20	(20)	100	国内
4/7	名古屋	100	10 +	50	(40)	140	国内
4/15	博多 ：最終寄港地	200	20 +	30	(30)	210	海外

(注) 青塗りされた列が AIS データから取得できる情報。海外に向かっている船舶が積載している貨物が全て輸出品であると仮定すると、最終寄港地を出港した際の積載量（この例では 210）は国内各港で積込んだ輸出品の合計（この例では 10+20+40+...+30）に等しくなる。

図表 6. 輸出指数と実質輸出

(a) 自動車

(b) 海上コンテナ

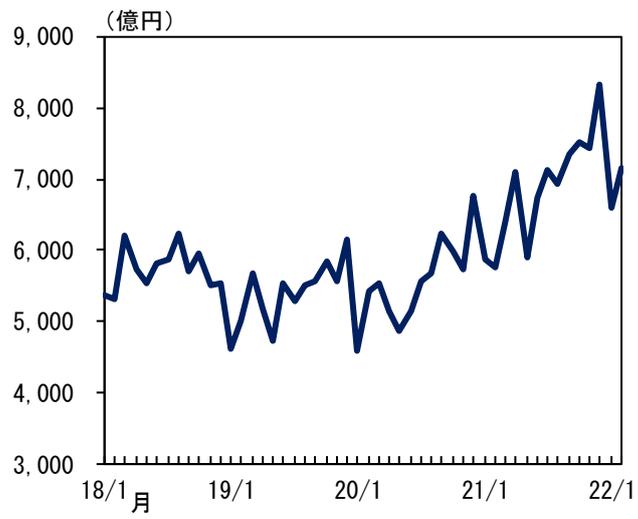


(注 1) 左図では自動車運搬船の輸出指数を、右図ではコンテナ船の輸出指数を示している。

(注 2) 海上コンテナと航空輸送による実質輸出額（右図）は、それぞれの運送形態で輸出された財の名目金額を財別の輸出物価指数で実質化したもの。

(出所) 財務省

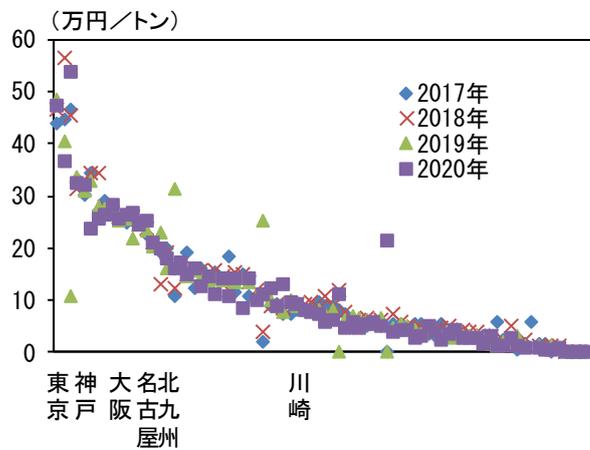
図表 7. 半導体関連輸出の動向



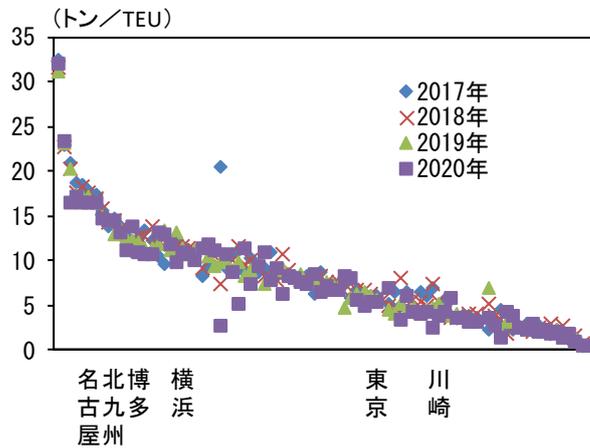
(注) 半導体と半導体製造装置の輸出額。
(出所) 財務省

図表 8. 港湾別の特徴

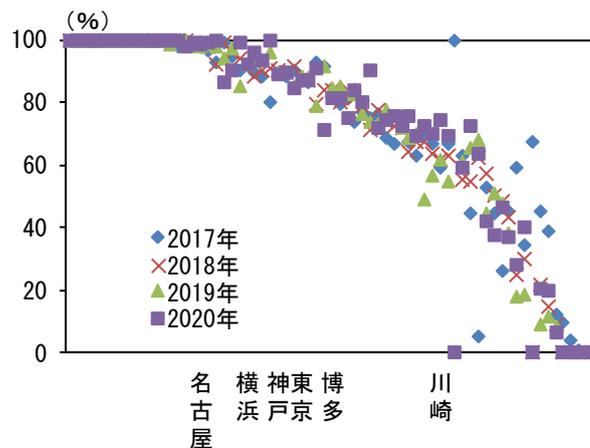
(a) 港別・輸出量当たり輸出金額



(b) 港別・コンテナ当たり重量



(c) 出荷量に占める輸出の割合



(注1) TEU (Twenty-foot Equivalent Units) は、20 フィートで換算したコンテナ個数を表す単位。
 (注2) 横軸には国内の港湾を並べており、主な港湾についてラベルを表示している。
 (出所) 財務省、国土交通省

図表 9. カーネル法による輸出関数の推計結果

名称	$k(x, x')$	特徴	相関係数	
			訓練データ	テストデータ
多項式 (1次)	$x^T x'$	$\phi(x) = (x_1, x_2, \dots, x_n)$ に対応	0.83	0.82
多項式 (2次)	$(x^T x')^2$	$\phi(x) = (x_1, x_2, \dots, x_1^2, \dots, x_1 x_2, \dots)$ (x の各要素の2次項までの組み合わせ)に対応	0.84	0.86
多項式 (3次)	$(x^T x')^3$	x の各要素の3次項までの組み合わせに対応	0.89	0.24
ガウシアン	$\exp\left(\frac{-\ x - x'\ ^2}{2}\right)$	対応する ϕ は無次元であり、明示的に書き表すことができない	0.99	<u>0.91</u>

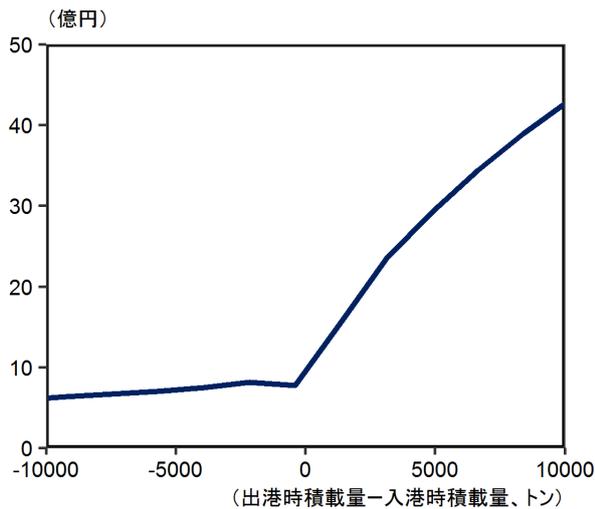
(注1) x は x_1, x_2, \dots を要素に持つベクトル

(注2) 相関係数は、各カーネル関数を用いて算出した輸出指数と、海上コンテナ船による実際の輸出額(実質ベース)の、前年比の相関係数。

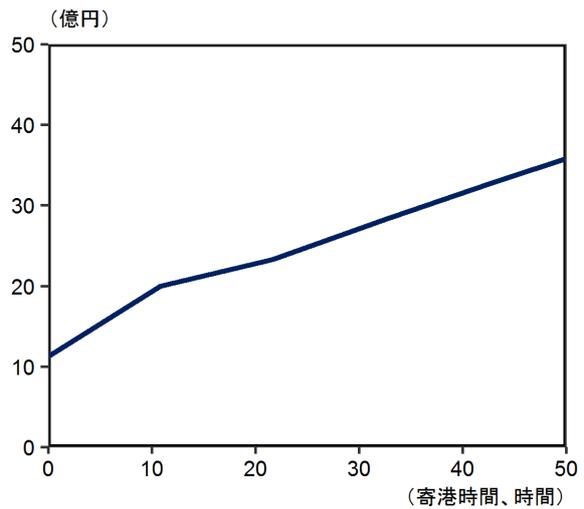
(注3) 訓練データは2017年1月~2020年12月、テストデータは2021年1月~2022年3月の期間のデータ。訓練データに基づいて輸出関数を推計し、両データ期間における輸出指数と実際の輸出額(実質ベース)の前年比の相関係数を計算した。

図表 10. ガウシアンカーネルを用いた輸出関数の推計結果

(a) 積載量と輸出額の関係

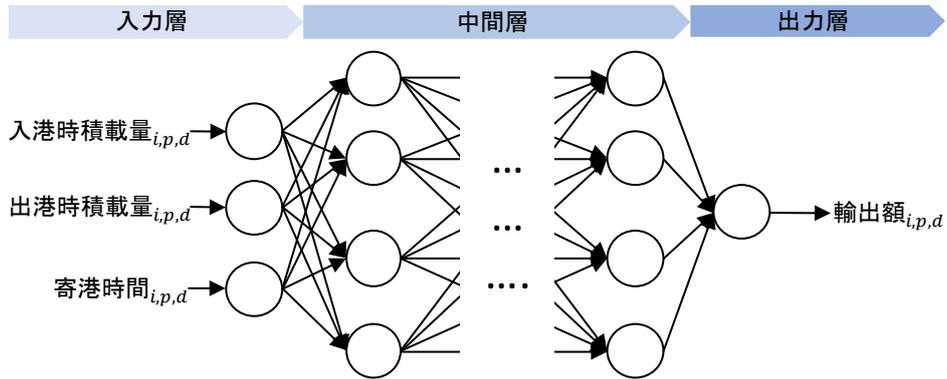


(b) 寄港時間と輸出額の関係



(注) 一般化加法モデルによる近似曲線。

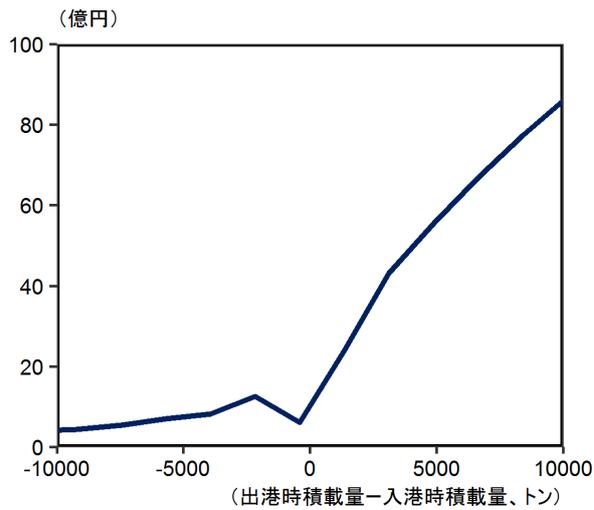
図表 11. 深層学習のイメージ



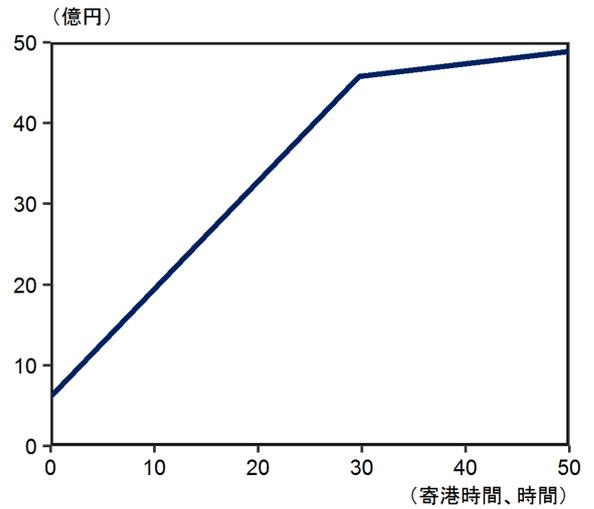
(注) 図内の丸はニューロンを表し、矢印はあるニューロンの出力結果が次の層のニューロンへの入力値となっている様子を表している。詳細は Goodfellow et al. (2016) を参照。

図表 12. 深層学習を用いた輸出関数の推計結果

(a) 積載量と輸出額の関係

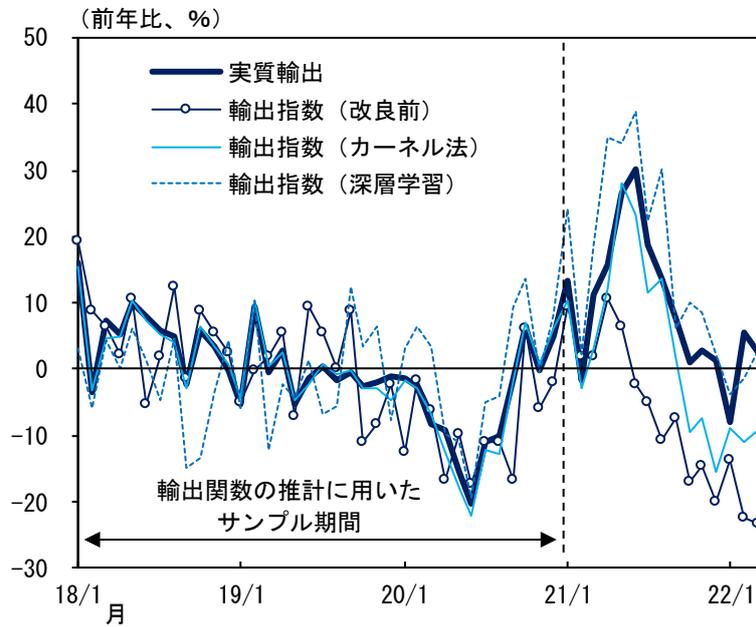


(b) 寄港時間と輸出額の関係



(注) 一般化加法モデルによる近似曲線。

図表 13. コンテナ船の輸出指数
(a) 時系列推移

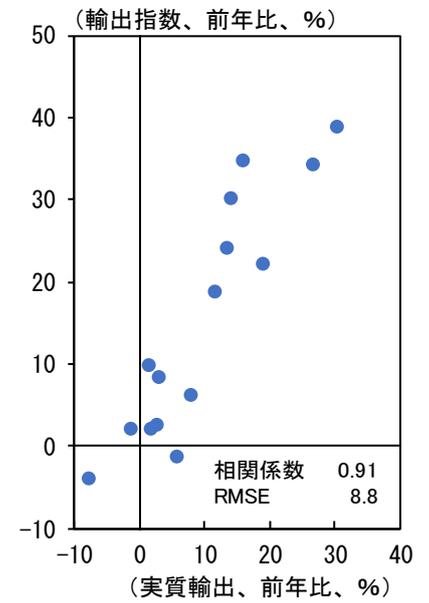
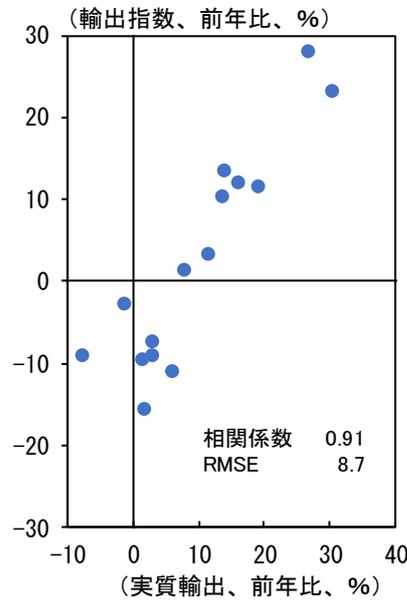
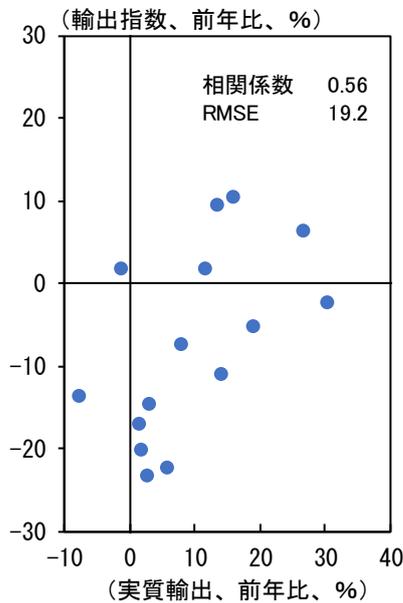


(b) アウトオブサンプル期間

改良前

カーネル法

深層学習



(注) 「実質輸出」は、海上コンテナによる輸出額(実質ベース)。(b)は、2021年1月~2022年3月(アウトオブサンプル期間)のデータ。

(出所) 財務省

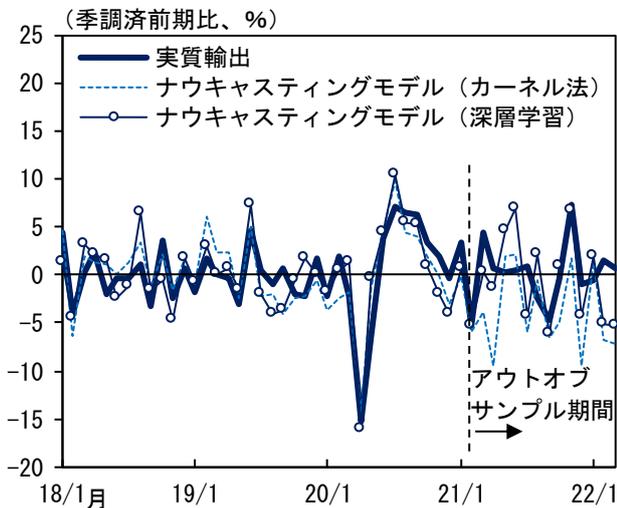
図表 14. OLS 推計結果

	被説明変数：実質輸出		
	(1) カーネル法	(2) 深層学習	(3)
定数項	1.38 ** (0.54)	0.51 (0.50)	1.70 *** (0.50)
輸出指数 (自動車運搬船)	0.31 *** (0.03)	0.28 *** (0.02)	
輸出指数 (コンテナ船)	0.18 * (0.10)	0.24 *** (0.05)	
貿易統計 (上旬速報)			0.59 *** (0.03)
修正R2	0.88	0.91	0.66

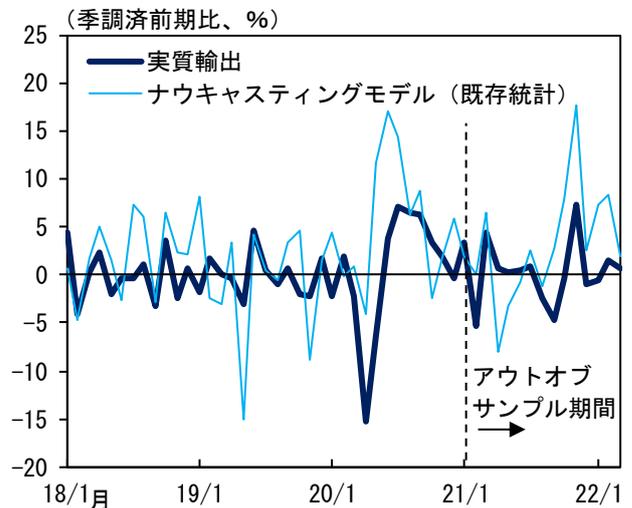
(注1) 各変数は前年比の値を使用。推計期間は2018年1月～2022年3月。
 (注2) 括弧内は標準誤差を表す。*は10%水準、**は5%水準、***は1%水準で統計的に有意であることを示す。
 (注3) (1)と(2)列は、それぞれカーネル法と深層学習を用いて計算した輸出指数(コンテナ船)を説明変数として使用した。
 (注4) なお、同推計期間における実質輸出額(前年比)のAR(1)モデルの自由度調整済み決定係数は0.67。

図表 15. ナウキャストモデルの計算結果

(a) 輸出指数を用いたモデル



(b) 既存統計を用いたモデル

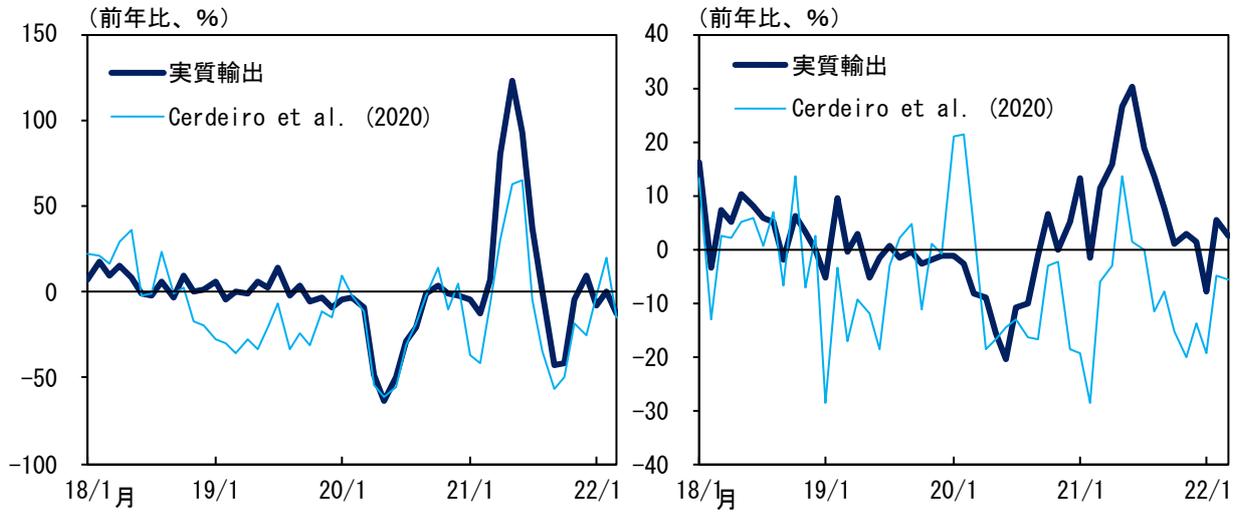


(注1) ナウキャストモデル(カーネル法)とナウキャストモデル(深層学習)はそれぞれ、輸出指数(自動車運搬船)と、カーネル法または深層学習を用いて計算した輸出指数(コンテナ船)を使用して推計したモデル。ナウキャストモデル(既存統計)は貿易統計の上旬速報を使用して推計したモデル。
 (注2) アウトオブサンプル期間(2021年1月～2022年3月)におけるRMSEは、ナウキャストモデル(カーネル法)が5.04、ナウキャストモデル(深層学習)が3.64、ナウキャストモデル(既存統計)が5.80。なお、機械学習を活用せずに計算した輸出指数を用いて同様の計算を行うと、RMSEは5.70となった。

図表 16. Cerdeiro et al. (2020)の輸出指標

(a) 自動車

(b) 海上コンテナ



(出所) Cerdeiro et al. (2020)、財務省

図表 17. Cerdeiro et al. (2020)の輸出指標に関する OLS 推計結果

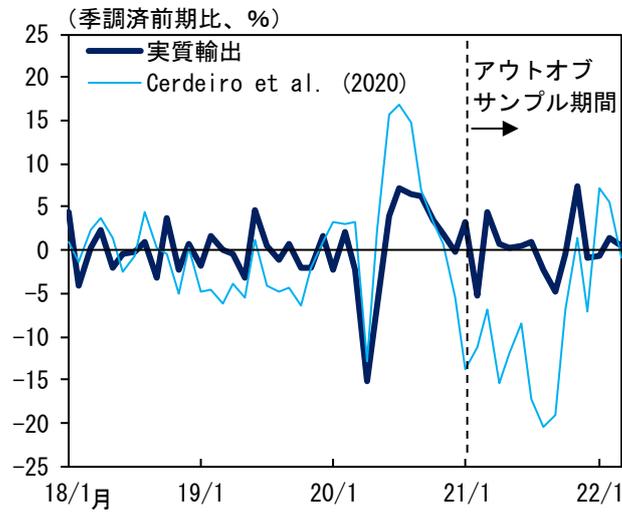
	被説明変数：実質輸出					
	(1) Cerdeiro et al. (2020)		(2) 本分析 (カーネル法)		(3) 本分析 (深層学習)	
定数	4.57	***	1.38	**	0.51	
	(1.25)		(0.54)		(0.50)	
輸出指数 (自動車運搬船)	0.28	***	0.31	***	0.28	***
	(0.05)		(0.03)		(0.02)	
輸出指数 (コンテナ船)	-0.01		0.18	*	0.24	***
	(0.12)		(0.10)		(0.05)	
修正R2	0.47		0.88		0.91	

(注1) 各変数は前年比の値を使用。推計期間は2018年1月～2022年3月。

(注2) 括弧内は標準誤差を表す。*は10%水準、**は5%水準、***は1%水準で統計的に有意であることを示す。

(注3) (1)列はCerdeiro et al. (2020)が算出したわが国における自動車運搬船とコンテナ船の輸出指標を説明変数として用いており、(2)と(3)列は、それぞれカーネル法と深層学習を用いて計算した輸出指数(コンテナ船)を説明変数として使用した。

図表 18. Cerdeiro et al. (2020)の輸出指標を用いた
 ナウキャストモデルの計算結果



(注) アウトオブサンプル期間 (2021年1月～2022年3月) における RMSE は 7.44。