



RIETI Discussion Paper Series 24-J-005

滞在人口データを用いた取引先企業のリース需要予測

加藤 壘

三井住友ファイナンス&リース株式会社

宮川 大介

早稲田大学

柳岡 優希

株式会社東京商工リサーチ

雪本 真治

三井住友ファイナンス&リース株式会社



Research Institute of Economy, Trade & Industry, IAA

独立行政法人経済産業研究所

<https://www.rieti.go.jp/jp/>

滞在人口データを用いた取引先企業のリース需要予測*

加藤 壘 (三井住友ファイナンス&リース株式会社)

宮川 大介 (早稲田大学)

柳岡 優希 (株式会社東京商工リサーチ)

雪本 真治 (三井住友ファイナンス&リース株式会社)

要 旨

本稿では、近年利用が拡大している個人のGPS位置情報データ（滞在人口データ）を用いて工場レベルの滞在人口を計測した上で、計測対象工場を有する企業の「取引先」におけるリース需要の予測を行う。具体的には、第一に、製造業に属する企業の主要な工場における滞在人口の変動をGPS位置情報データによって計測する。第二に、この計測値を用いて、当該企業の仕入先に当たる取引先企業におけるリース需要の発生有無を予測するモデルを、国内最大のリース取引実績を有する三井住友ファイナンス&リース株式会社のリース契約データを用いて構築する。予測結果から、新型コロナウイルス感染症の拡大から企業活動が大きく攪乱された時期を除けば、滞在人口データが取引先企業のリース需要予測に貢献することが確認された。

キーワード: オルタナティブデータ、GPSデータ、リース、サプライチェーンネットワーク

JEL classification: G31, L14

RIETI ディスカッション・ペーパーは、専門論文の形式でまとめられた研究成果を公開し、活発な議論を喚起することを目的としています。論文に述べられている見解は執筆者個人の責任で発表するものであり、所属する組織及び（独）経済産業研究所としての見解を示すものではありません。

*本稿は、独立行政法人経済産業研究所（RIETI）におけるプロジェクト「企業ダイナミクスと産業・マクロ経済」の成果の一部である。本稿の原案は、経済産業研究所（RIETI）のディスカッション・ペーパー検討会で発表を行ったものである。検討会参加者からの有益なコメントに感謝したい。また、本稿の作成にあたっては、深尾京司教授（アジア経済研究所）、森川正之教授（RIETI）、和泉潔教授（東京大学）から多くの有益なコメントを頂いた。ここに記して、感謝の意を表したい。

1. イントロダクション

近年、経済学を中心とする社会科学分野で、公的統計や企業の決算情報などの伝統的なデータとは異なる情報源から取得した、所謂「オルタナティブデータ」に注目が集まっている。オルタナティブデータの例としては、クレジットカードの消費データや、消費電力のデータ、衛星画像のデータなど様々なデータが挙げられる。こうしたオルタナティブデータは、伝統的なデータと比較すると、速報性が高く細かな粒度でデータを利用できることから、とりわけマクロレベルの経済活動を予測する場面で活用が進んでいる。具体的な研究事例としては、バルチック海運指数やガソリン消費量など日次や週次で利用可能な速報性の高いデータを活かして世界の GDP をナウキャストした事例 (Jardet and Meunier(2020)) や、クレジットカードの利用状況データの粒度の高さを活かして地域ごとの小売業の売上高を予測した (Aladangady et al.(2019)) などが挙げられる。

こうした研究の近年の動向として、2020 年以降における新型コロナウイルス感染症の拡大及び対応する行動制限を受けて、オルタナティブデータの中でも地点ごとに細かい粒度で人流の動きを捉えることが出来る位置情報データに注目が集まっている。位置情報データを利用した研究は様々見られているが、特に製造業の企業が保有している工場やその周辺の滞在人口について位置情報データを用いて計測し、経済活動を予測する分析が進められている。例えば、鋳工業生産指数の予測に当たり工場の滞在人口の計測結果を利用することで伝統的な経済統計のみを利用した予測より精度を改善させた事例 (古川ほか(2022)) や、各自動車メーカーの工場の滞在人口を計測しそれぞれの自動車メーカーの自動車生産量を予測した事例 (水門・柳井(2020)) などが挙げられる。

このように位置情報データを有効に活用した事例が近年蓄積されている一方で、こうした情報を他の種別のデータ、例えば、企業間の取引関係を記述したデータと組み合わせる事例は乏しい。企業間の取引関係データを活用した事例には、とりわけサプライチェーンネットワークデータを用いた研究が数多く、最近の研究として、例えば、Carvalho et al.(2021)では、東日本大震災に起因するショックがサプライチェーンを通じて上流から下流へどのように波及したかを分析の対象としている。また、Arata and Miyakawa(2022)では、世界金融危機や新型コロナウイルス感染症のパンデミックによってマイナス成長となったセクターに属する下流の企業から上流のサプライヤーへ、需要ショックがどのように波及したかを実証的に分析している。以上の研究は、年次単位で計測された伝統的な企業レベルの財務情報とサプライチェーンネットワークデータを用いることで、ショックの伝播を実証的に検討したものであるが、人流データの様な高粒度・高頻度で継続されたオルタナティブデ

ータは用いていない。これらの研究を踏まえた自然な発想として、位置情報データを用いて計測した各企業（例：工場）単位の滞在人口の変動によって活動の変化を捉えた上で、当該企業とサプライチェーンで接続されている企業を特定した上でこれらの企業の動向を分析対象として定めた分析が展望されるだろう。

以上の議論を踏まえて、本稿では、位置情報データの一つである携帯電話の GPS データを用いて計測した各企業の工場滞在人口の変動を利用し、サプライチェーン上で当該計測対象企業の上流に位置する企業の活動を予測するモデルの構築を試みる。本稿の分析と密接に関係する既存研究として、位置情報データとサプライチェーンネットワークデータを結び付けることで、自動車産業のサプライチェーンの異常度合いを定量的に測る指数を構築した上田ら(2023)が存在する。彼らは、サプライチェーンネットワークデータを利用し、完成車メーカーと当該企業と直接的な取引関係を持つサプライヤを特定した上で、両社の工場間の二地点間の人流を GPS データに基づいて時系列データとして計測している。この計測結果を用いて、教師なし学習による異常検知モデルを構築し、異常スコアを算出することで、取引関係にある企業の工場間の人流の変動から自動車産業全体のサプライチェーンの異常度合いを計測している。本稿では、こうした新しい研究テーマに属するものであり、企業単位で計測した位置情報データを用いて、サプライチェーンで接続された取引先の企業の活動を予測するモデルを、教師あり学習によって構築する。

本稿で構築するモデルでは、工場の滞在人口を計測する企業の仕入先に当たる企業を、株式会社東京商工リサーチ（以下 TSR 社）が提供する企業相関データを利用して特定し、サプライチェーン上流に当たる企業と定義した。その上で、仕入先企業の活動の把握には、三井住友ファイナンス&リース（以下 SMFL 社）のデータを利用し、特定の期間内に SMFL 社と金融に当たる契約を合意したか否かについて予測する。予測に用いる説明変数には、TSR 社が提供する企業情報データから取得した基本的な企業属性情報を利用するほか、滞在人口の計測対象企業の前 6 か月の滞在人口データを加えて利用するパターンと利用しなかったパターンの 2 通りを設定した。この 2 通りのパターンの分析結果を比較することにより、滞在人口データの有用性について検討する。

複数の予測モデル構築と精度検証を行なった結果、トレーニングデータのデータ利用期間を十分に確保した場合、滞在人口データを利用することで精度の向上が見られた。また、トレーニングデータのデータ利用期間が短い場合は、データ利用期間に占める新型コロナウイルス感染症のパンデミック初期の社会情勢の変化が大きい期間の割合が低い場合に、精度の向上が見られた。以上の結果は、新型コロナウイルス感染症の拡大から企業活動が大

きく攪乱された時期を除けば、滞在人口データが取引先企業のリース需要予測に貢献することを示唆している。

本稿の構成は以下の通りである。まず第2節で、本稿で用いるデータについて概説する。第3節では、分析の手法について概説する。第4節では分析で得られた結果と考察について述べる。第5節は結びである。

2. データ

本節では、本分析で用いるデータについて概説する。

2-1. GPS データ

Orbital Insight, Inc.が提供するGPS計測数データ取得サービスを利用した。本データは、スマートフォン端末のアプリケーションが収集したGPS情報を利用して整備されたデータであり、2019年9月分から蓄積されている。これまで位置情報データを利用した多くの先行研究では携帯電話の基地局をベースとして数百メートル四方のメッシュ単位で取得されたデータを使用しているが、本データは、地図上から形状の制約なく指定した範囲でデータを取得することが出来る。また、データ取得の頻度は1時間毎もしくは1日毎の単位となり、この単位ごとに指定した範囲内にGPSデータ上滞在していた端末数を収集することが出来る。

本分析では、後述する滞在人口の計測対象となる製造業の企業22社に対して、後述するTSR社が提供する事業所データに基づきこれらの企業の工場の立地を特定し、推測される工場の敷地をGPSデータの計測範囲として指定した。さらに、工場の守衛室への立ち寄りなど極一時的な滞在による影響を排除するため、1日当たり最低5分以上滞在した端末のみを、その日に滞在したものと見做し、1日ごとに滞在した端末の数を収集した。本データの提供期間が2019年9月からとなることから、データの取得期間は2019年9月から2022年5月までとした。

2-2. TSR 事業所データ

TSR社が提供する事業所データを利用した。このデータは、TSR社の調査結果に基づき、工場を含む企業の事業所の住所や土地面積などの情報を取りまとめたデータである。

本分析では、工場の滞在人口の計測対象として 22 社の企業を選定した。これらの企業は業種が製造業に当たる企業を取り扱っている製品の種類を問わず幅広く選定したもので、また日本国内に主要な工場を持つ企業となっており、上場企業 17 社に加え未上場企業 5 社も含んでいる。これらの企業に対して、本データを用いて各企業の工場の立地する住所を調査した上で、滞在人口を計測する対象の工場の立地と敷地範囲の特定に利用した。

2-3. TSR 企業関連データ

TSR 社が提供する企業関連データを利用した。このデータは、TSR 社の調査結果に基づき、特定の 1 社に対し、その企業に対して仕入を行っている企業、その企業に対して販売を行っている企業、その企業の株主に当たる企業の情報を取りまとめたデータである。

本分析では、滞在人口の計測対象企業の仕入先として登録されている企業と、滞在人口の計測対象企業を販売先として登録している企業のリストを、本データを利用して抽出したうえで、抽出された 1322 社を滞在人口の計測対象企業に対する仕入先の企業として定義した。仕入先企業のうち 75 社が上場企業となっており、1247 社が非上場企業となっている。また、業種は製造業が 742 社で卸売業が 168 社と、仕入先企業の多くがこの 2 業種に属する企業となっている。

2-4. TSR 企業情報データ

TSR 社が提供する企業情報データを利用した。このデータは、TSR 社の調査結果に基づき、全国の企業を対象に企業の属性情報、簡易的な決算情報、仕入先や販売先の取引規模の順位などの情報などを取りまとめたデータである。

本分析では、滞在人口の計測対象企業と、その仕入先の企業について、説明変数で利用するための企業属性情報の収集と、これら両者間の取引規模が仕入先企業にとって何位の規模であるか特定するためのデータとして使用した。

2-5. SMFL 契約データ

リースの取引規模では国内最大の企業である SMFL 社の契約データを利用した。このデータは、SMFL 社でこれまで行われたリース等の契約情報を蓄積したデータである。

本分析では、目的変数として、仕入先の企業が SMFL 社と機械や設備などのリース契約

もしくは延払契約を合意したイベントを採用した。こうしたリース需要に係るイベントは、企業の活動水準と相関していると考えられる。本研究の目的であるリアルタイム性の高い予測モデルの構築にとって、細かな頻度で計測された目的変数の存在が重要となるため、本件ではこうした目的変数の設定を行っている。なお、延払は、リースと異なり物件の所有権は契約先側に属するものの、契約期間満了まで SMFL 社に所有権が留保され、リース取引と需要が発生する要因が共通することから分析対象に含める。これらの契約の合意が行われるタイミングは、企業が間接金融により機械や設備を導入することを決断するタイミングと同義であることから、契約の合意が取れた時点でもリース需要が発生した時点と捉え分析を行う。

3.分析手法

本節では、本研究で行った分析の手法について概説する。

3-1. GPS データ計測数の補正

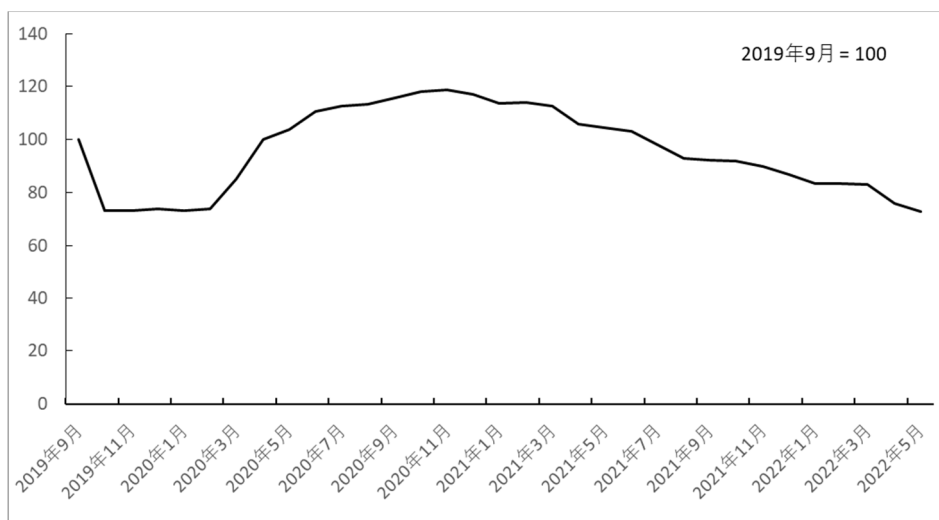
本分析で使用した GPS データについては、スマートフォンのアプリケーションを利用したユーザーから送信された位置情報を収集したものとなっている。そのため、長期的なデータの推移を確認したい場合、ある地点の実際の滞在人口が長期的に変化しなかった場合であっても、対象のアプリケーションのアクティブユーザー数が変動すれば、GPS データの計測数も合わせて変動してしまう。実際にこのような変動が発生しているか確認するため、日々の滞在人口の変動が小さいと考えられる住宅街の地域の GPS データの計測数を、データの利用期間である 2019 年 9 月から 2022 年 5 月までの期間、月単位で集計した結果を図 1 に示す。なお、計測の対象とした住宅街の所在地域は表 1 に示す 6 地点である。

表 1. GPS データ計測対象の住宅街

地点名	所在地
村山団地	東京都武蔵村山市
木曾団地	東京都町田市
岩倉団地	愛知県岩倉市
高蔵寺ニュータウン	愛知県春日井市
香里団地	大阪府枚方市
狭山住宅	大阪府大阪狭山市

(注) 表は GPS データを計測した対象の住宅街の一覧を示している。住宅街地域は大規模団地を選定しており、地点名には団地の名称、所在地には団地が所在する都道府県名と市町村名を示している。

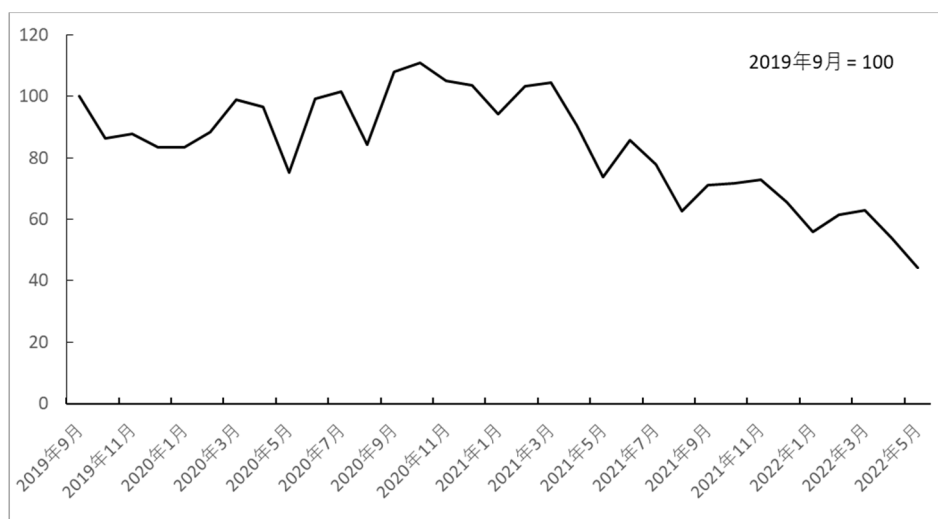
図 1. 住宅街の GPS データ計測数の長期変動



(注) 図は表 1 に示した住宅街を対象とした GPS データの計測数を各日ごとに単純に合計した値の各月ごとの平均値を集計し、2019 年 9 月時点の値を 100 とした上でそれ以降の月の値の比率を、2019 年 9 月から 2022 年 5 月までの各月ごとに示したものである。

住宅街の滞在人口は、実際には長期的に変動しないものと思われるが、図 1 の結果を見ると GPS データ計測数が大きく増減するような動きを見せており、前述したアプリケーションのユーザー数の変動の影響を受けているものとみられる。そこで、本分析の対象とする工場の GPS データ計測数も同様の影響を受けているか確認するため、滞在人口の計測対象企業 22 社の GPS データ計測数の変動を平均したものを図 2 に示す。

図 2. 滞在人口の計測対象企業の工場の GPS データ計測数の長期変動

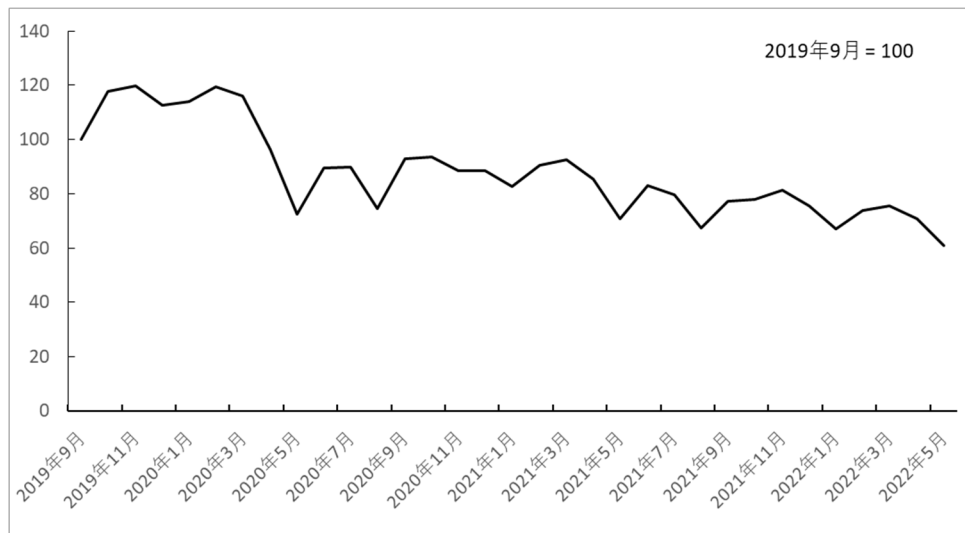


(注) 図は、滞在人口計測対象の各社ごとに、各月の GPS データ計測数を日平均した値を求め、2019 年 9 月を 100 とした場合の各月の比率の値を算出し、最後に滞在人口計測対象の 22 社で平均値を求めた結果を、2019 年 9 月から 2022 年 5 月までの各月ごとに示したものである。

図 2 の結果を見ると、住宅街地域の GPS データ計測数と同様に 2020 年後半以降の GPS データ計測数が下降傾向となっているほか、2019 年 11 月から 2020 年 3 月にかけての期間が、その後の期間と比較して GPS データ計測数が少なくなっている。こうした点が共通していることを考えれば、工場の GPS データ計測数についても、アプリケーションのユーザー数の変動の影響を受けているとみられる。したがって、工場の GPS データ計測数に補正を加えないまま分析に利用すると、実際の工場の滞在人口の変動と乖離したまま利用してしまうものと考えられる。

そこで、アプリケーションのユーザー数の変動の影響を緩和するため、1 日単位で取得した滞在人口の計測対象企業の工場の GPS データ計測数を、表 1 で示した住宅街の GPS データ計測数の合計値で割った値を、補正後の GPS データ計測数として利用することを考える。この補正を行った結果となる、滞在人口の計測対象企業 22 社の補正後の GPS データ計測数の変動を図 3 に示す。

図 3. 滞在人口の計測対象企業の工場の補正後の GPS データ計測数の長期変動



(注) 図は、滞在人口計測対象の各社ごとに、工場のGPSデータ計測数を住宅街のGPSデータ計測数で割って算出した補正後のGPSデータ計測数を各月ごとに日平均した値を基に、2019年9月を100とした場合の各月の比率の値を算出し、最後に滞在人口計測対象の22社で平均値を求めた結果を、2019年9月から2022年5月までの各月ごとに示したものである。

図3の結果を見ると、2020年後半以降の補正後のGPSデータ計測数が横ばい傾向になっているほか、2019年11月から2020年1月にかけての期間とその後の期間を比較すると、前者の期間のほうが補正後のGPSデータ計測数が多くなっていることが分かる。また、補正後のGPSデータ計測数は、補正前のGPSデータ計測数と異なり、2020年3月から減少傾向を示している。2020年3月ごろから新型コロナウイルスのパンデミックの影響を受け、人の移動制限や在宅勤務の普及が起こったことを考慮すれば、これは実際の工場の滞在人口の変動とより近い変動となっていることが考えられる。したがって、本分析ではこの補正後のGPSデータ計測数を工場の滞在人口の計測結果として利用し、分析を行う。

3-2. 予測モデル構築

本研究では、2-2項で取り上げた滞在人口計測対象の企業22社と、その仕入先企業の企業情報データおよび3-1項の方法で補正を行ったGPSデータ計測数に基づく工場の滞在人口データを利用する。その上で、滞在人口の計測対象企業の仕入先に該当する企業が、予測対象の期間の間にSMFL社と金融に関わる契約を新たに合意したか否かを予測するモデル

を構築する。モデル構築に当たっては、標準的な線形モデルや機械学習モデルとして、表 2 に示す手法を用い、それぞれ前処理やパラメータの異なる複数のパターンでモデルを構築する。

表 2. モデル構築で用いた手法

モデル手法名	概要
ElasticNet	回帰手法の一つであり、L1 正則化と L2 正則化をパラメータにより線形結合で組み合わせる手法。
Logistic Regression	二値分類に用いる回帰手法で、シグモイド関数をモデルの出力に利用する。
RandomForest	機械学習アルゴリズムの一つで、決定木を弱学習器として用いる。生成した多数の決定木を組み合わせる一つのモデルとする。
XGBoost	機械学習アルゴリズムの一つで、決定木に基づく勾配ブースティング手法。決定木の分岐に Level-wise と呼ばれる手法を採用し、決定木の深さを優先している。
LightGBM	機械学習アルゴリズムの一つで、決定木に基づく勾配ブースティング手法。決定木の分岐に Leaf-wise と呼ばれる手法を採用し、決定木の幅を優先している。
Support Vector Machine	2 つのクラスのデータ群を分割するような境界線を決定することで分類を行うアルゴリズム。

(注) 表は、モデル構築に利用したモデル構築手法の一覧を示す。モデル手法には用いた手法の一般的な呼称を、概要にはそれぞれの手法の簡単な解説を示す。

さらに、これらの手法を用いて作成したモデルについて、最大 3 モデルまでの複数のモデルを、予測結果の平均に基づくアンサンブルで組み合わせたモデルを追加し、これを含めテストデータの AUC に基づく精度が最も高かったモデルを最終的に選択する。後述する分析期間の条件ごとに、工場の滞在人口の計測結果を使用したモデルと使用しなかったモデルの間で、選択したモデルのテストデータに対する AUC を比較し、滞在人口データが予測に寄与するか検証する。

3-2-1. 目的変数の設定

予測対象は仕入先企業のリース需要であるが、リース需要の測定には 2-5 項で説明した SMFL 社の契約データを使用した。データを半月単位で区分し、仕入先企業が半月の間に SMFL 社と金融に当たる契約の合意を行った場合は”1”、行わなかった場合は”0”とした変数を、契約有無フラグと設定した上で、二値分類の形で分析を実施した。期間を区分する単位は月の上旬と下旬とし、上旬の場合 1 日から 15 日、下旬の場合は 16 日から月末日と定義した。

3-2-2. 説明変数の設定

表 3 に示す通り、説明変数として日付情報と、滞在人口の計測対象企業と仕入先企業それぞれに関する 2-4 項の企業情報データを由来とする属性情報を利用した。また、滞在人口の計測結果を使用することによる効果を検証するため、滞在人口の計測結果を加工して作成した変数を追加で使用した滞在人口データ有モデルと、これらの変数を追加しなかった滞在人口データ無モデルとしてそれぞれ構築した。

滞在人口の計測結果を加工することで、表 3 の下 5 行に示す通り 5 種類の変数を作成した。まず(1)の変数は、各データ時点の直近 6 か月間の滞在人口を利用して、直前の 1 ヶ月間の滞在人口を、2 ヶ月前の滞在人口で割って算出した滞在人口の前月比を、前の 5 か月分にわたり作成したものである。次に(2)の変数は、新型コロナウイルスのパンデミックの影響が日本国内で大きく広がる前である、2019 年 9 月 1 日から 2020 年 3 月 15 日までの平均の滞在人口を 1 とした場合の、前 1 か月間から前 6 か月間までの滞在人口の比率を 1 か月単位で作成したものである。続いて(3)の変数には、滞在人口の計測対象企業と仕入先企業の関係性が強いほど、目的変数に与える工場の滞在人口の変動の影響が大きいと考えられることから、両社のペアごとに相互の関係性の強さが強いほうが、(2)で算出した変数の変動の振れ幅が大きくなるよう補正したものを作成した。ここでは、両社関係性の強さを特定する方法として、仕入先企業の TSR 企業情報データに掲載される、仕入先企業にとっての販売先の販売規模順位を利用した。最後に(4)と(5)の変数では、滞在人口の計測対象企業 22 社中の滞在人口のシェアを利用した。これは、製造業全体として工場の滞在人口の変動トレンドが変化している中でも、相対的に特定の企業の工場の滞在人口が大きく変動しているかどうか、予測に反映させることを目的としている。この滞在人口シェアは、データの各行ごとに、各社の工場の滞在人口を、同じ期間の滞在人口の計測対象企業 22 社の滞在人口の合計値で割ることで算出した。(4)の変数では、(1)の変数作成で使用した滞在人口を、滞在人口のシェアに置き換えて同様の方法で前月比を算出したものを利用しており、(5)の

変数でも、(2)の変数作成で使用了滞人口を滞人口のシェアに置き換えて同様に算出したものを利用した。

表 3. 説明変数の設定

変数	データ元	滞人口有 モデル	滞人口無 モデル
データ対象月	(日付情報)	○	○
月上旬・下旬フラグ	(日付情報)	○	○
対象データ期間の総日数	(日付情報)	○	○
対象データ期間の平日日数	(日付情報)	○	○
滞人口計測企業の業種	TSR 企業情報	○	○
仕入先企業の業種	TSR 企業情報	○	○
滞人口計測企業の売上高	TSR 企業情報	○	○
仕入先企業の売上高	TSR 企業情報	○	○
滞人口の直近 5 か月間における前月比 (1)	GPS データ	○	
特定期間からの滞人口変化率 (2)	GPS データ	○	
両社間の関係性の強さで補正した特定期間からの滞人口の変化率 (3)	GPS データ TSR 企業情報	○	
滞人口の計測対象企業間での滞人口シェアの直近 5 か月間における前月比 (4)	GPS データ	○	
特定期間からの滞人口の計測対象企業間での滞人口シェアの変化率 (5)	GPS データ	○	

(注) 表は、モデル構築に利用した変数の一覧を示す。滞人口の計測結果を利用するモデルで用いた変数は滞人口有モデルで○を付けた変数であり、滞人口の計測結果を利用しないで作成するモデルで用いた変数は滞人口無モデルで○を付けた変数となる。

3-2-3. トレーニングデータ・テストデータの設定

滞在人口の計測元のGPSデータが利用可能となった期間が2019年9月からであること、また説明変数では目的変数の期間から直近6か月前までの滞在人口の計測結果を利用して、目的変数の対象期間が2020年3月から2022年5月までとなるデータを用いてトレーニングデータとテストデータを設定した。本分析においては、トレーニングデータを利用してモデル構築を行ったうえで、テストデータを利用して精度を確認した。トレーニングデータとテストデータの設定内容は表4と図4に示す。

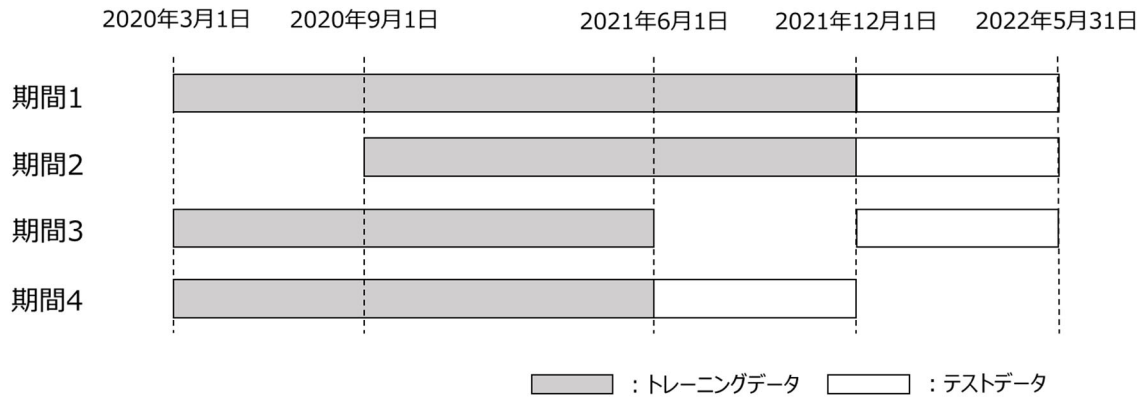
トレーニングデータとテストデータの設定に当たっては、テストデータの期間幅を全て6か月間と統一したうえで、利用可能なデータをすべて利用するパターンの期間1を設定した。続いて、期間1と比較しトレーニングデータの期間幅の長さによる影響について確認するため、期間1のトレーニングデータの初期の6か月間を削減した期間2を設定した。続いて、期間2と比較しトレーニングデータの設定時期の違いによる影響について確認するため、期間2とは異なり期間1のトレーニングデータの最後の6か月間を削減した期間3を設定した。最後に、期間3と比較することでテストデータの設定時期の違いによる影響について確認するため、期間3のテストデータの設定時期を6か月前倒しした期間4を設定した。

表4. トレーニングデータとテストデータの期間分割表

期間名	トレーニングデータ	テストデータ
期間1	2020年3月～2021年11月(21ヶ月)	2021年12月～2022年5月(6ヶ月)
期間2	2020年9月～2021年11月(15ヶ月)	2021年12月～2022年5月(6ヶ月)
期間3	2020年3月～2021年5月(15ヶ月)	2021年12月～2022年5月(6ヶ月)
期間4	2020年3月～2021年5月(15ヶ月)	2021年6月～2021年11月(6ヶ月)

(注) 表は、期間ごとに設定したトレーニングデータ及びテストデータで利用したデータの利用期間を示しており、括弧内はデータの利用期間の長さを示している。

図4. トレーニングデータとテストデータの期間分割図



(注) 図は、表 4 で示したトレーニングデータとテストデータの利用期間について、目視で確認できるような図で表したものである。

3-2-4. データの絞込条件

SMFL 社の取引先の中には、同じ種類の物件を四半期毎や毎月といった頻度で定期的に新規リース契約を行う企業も多い。これらのリース契約の大半は既に導入している設備の定期的な取り換えに起因するもので、企業の活動状況の変化とは異なる理由でリース契約が行われていると考えられる。このようなリース契約を行う企業のデータは、工場の滞在人口の変動とリース需要との関連性を検討する上では、利用することが適切ではないため、一定の基準をもって分析データから除外する処理を行った。本分析では、概ね四半期に一度以上の頻度で新規リース契約を行う顧客について定期的に新規リース契約を行う顧客と見做し、SMFL 契約データの取得期間である 2020 年 3 月から 2022 年 5 月にかけての期間中、通算 11 回以上新規に契約を行った仕入先企業をデータから除外した。

また、仕入先企業に対して滞在人口の計測対象企業との取引規模が相対的に小さい場合は、工場の滞在人口の変動と仕入先企業の活動の関係性も小さくなることが考えられる。そこで、仕入先企業の TSR 企業情報データに掲載される、仕入先企業にとっての販売先の販売規模の大きさ順の情報を利用し、仕入先企業から見て滞在人口の計測対象企業への販売規模の大きさが 1 位もしくは 2 位のペアのみ残るようにデータの絞込を行った。

3-2-5. 精度の評価手法

本分析においては、特定の閾値の設定に依存しない精度評価の指標として、ROC 曲線に基づく AUC を参照する。各分析期間において、構築した滞在人口有モデルと滞在人口無モ

デルをテストデータに適用し、両者間の AUC を比較した上で、滞在人口の計測結果の有用性を評価する。

4. 分析結果

本節では、本研究で行った分析の結果と、結果の解釈について述べる。

4-1. 各分析期間の結果

分析期間と滞在人口の計測結果の利用有無ごとに、各手法にてモデル構築を行い、テストデータに適用した場合の AUC は表 5 の通りとなった。表には、各手法により生成されたモデルのうち最も AUC が高いパターンの結果を示す。

表 5. 分析期間と滞在人口データ利用有無ごとの手法別 AUC 比較

手法	期間 1	期間 1	期間 2	期間 2
	滞在人口有	滞在人口無	滞在人口有	滞在人口無
ElasticNet	0.6959	0.6843	0.6909	0.6756
Logistic Regression	0.6910	0.6837	0.6869	0.6769
RandomForest	0.6677	0.6571	0.6701	0.6565
XGBoost	0.6408	0.6099	0.6355	0.6118
LightGBM	0.6690	0.6593	0.6716	0.6726
Support Vector Machine	0.6944	0.6534	0.6877	0.6683
手法	期間 3	期間 3	期間 4	期間 4
	滞在人口有	滞在人口無	滞在人口有	滞在人口無
ElasticNet	0.6683	0.6758	0.5898	0.6135
Logistic Regression	0.6580	0.6858	0.6073	0.6155
RandomForest	0.6380	0.5894	0.5651	0.6131
XGBoost	0.5975	0.6509	0.5798	0.6099
LightGBM	0.6602	0.6663	0.5898	0.6091
Support Vector Machine	0.6007	0.5925	0.5259	0.6007

(注) 表は、各期間と滞在人口データの利用有無別に構築したモデルを、テストデータに対して適用した場合の AUC と、各モデルで最終的に選択されたモデルで利用している手法を示す。利用手法の概説は表 2 に示す通り。

続いて、各手法により生成したモデルを基に、予測結果の平均に基づくアンサンブルモデ

ルを構築した結果を表 6 に示す。アンサンブルモデルは多数のパターンに基づき作成しているため、その中で最も高い AUC を得られた結果を示している。

表 6. 分析期間と滞在人口データ利用有無ごとのアンサンブルモデル

条件	AUC	モデル手法の組み合わせ
期間 1 滞在人口有	0.6965	ElasticNet (2 パターン) + Support Vector Machine
期間 1 滞在人口無	0.6849	ElasticNet (2 パターン) + Logistic Regression
期間 2 滞在人口有	0.6927	ElasticNet (2 パターン) + Logistic Regression
期間 2 滞在人口無	0.6830	ElasticNet (2 パターン) + Logistic Regression
期間 3 滞在人口有	0.6666	ElasticNet (3 パターン)
期間 3 滞在人口無	0.6829	Logistic Regression (3 パターン)
期間 4 滞在人口有	0.6073	Logistic Regression (2 パターン) + LightGBM
期間 4 滞在人口無	0.6241	ElasticNet + Logistic Regression + XGBoost

(注) 表は、分析期間と滞在人口の計測結果の利用有無ごとに、各手法にて構築した最大 3 モデルまでの複数のモデルを利用して、予測結果の平均に基づき構築したアンサンブルモデルをテストデータに適用した場合の、AUC とアンサンブルモデルを構築した元となるモデル手法の組み合わせを、最も AUC が高くなった場合について示す。

続いて、表 5 に示した単体のモデルと、表 6 に示したアンサンブルモデルの両者を含めた各モデルを比較し、分析期間と滞在人口の計測結果の利用有無ごとに、AUC に基づき最終的なモデル選択を行った結果を表 7 に示す。選択したモデルの利用手法の項目で、複数の手法名または手法名の後にパターン数を記載している条件ではアンサンブルモデルを選択しており、単体の手法名を記載している条件では、アンサンブルを行わない単体のモデルを選択している⁴。

⁴ 最も高い精度を実現した、「期間 1」かつ「滞在人口利用有」のケースに関して、各変数の **variable importance** (予測への貢献度合い) を確認すると、全 36 変数のうち上位 10 変数の中で 4 変数が滞在人口を用いて構築した変数となっており、当該変数群が予測に有用なシグナルを含むものであったことが示唆される。

表 7. 分析期間と滞在人口データ利用有無ごとの選択モデル

期間名	滞在人口 利用有無	AUC	選択したモデルの利用手法 (複数パターンの場合はアンサンブルモデル)
期間 1	有	0.6965	ElasticNet (2 パターン) + Support Vector Machine
	無	0.6849	ElasticNet (2 パターン) + Logistic Regression
期間 2	有	0.6927	ElasticNet (2 パターン) + Logistic Regression
	無	0.6830	ElasticNet (2 パターン) + Logistic Regression
期間 3	有	0.6683	Elastic Net
	無	0.6829	Logistic Regression (3 パターン)
期間 4	有	0.6073	Logistic Regression
	無	0.6241	Elastic Net + Logistic Regression + XGBoost

(注) 表は、各期間と滞在人口データの利用有無別に構築したモデルの AUC と、各モデルで最終的に選択されたモデルで利用している手法を示す。選択したモデル手法に複数のモデルを記載している場合は、複数のモデルをアンサンブルしていることを示す。

さらに、分析期間ごとに滞在人口の計測結果を利用した場合と利用しなかった場合の精度の比較は表 8 の通りとなった。表の①-②が取引先工場の滞在人口の計測結果の利用有無による AUC の差分を表しており、差分がプラスとなっている条件では滞在人口有モデルが滞在人口無モデルの精度を上回っていることを示す⁵。

この結果について確認すると、全体のデータを利用して分析を行った期間 1 の条件では、滞在人口の計測結果を利用することによる AUC の改善幅が大きく、予測精度の改善に貢献することが分かった。続いて、トレーニングデータの期間の長さに着目して比較すると、期間 1 と期間 2 の比較では、どちらも滞在人口の計測結果の利用により精度が改善しているものの、トレーニングデータの利用期間の長い期間 1 のほうが精度の改善幅が若干大きくなっている。このことから、トレーニングデータの利用期間が長い場合のほうが、より滞在人口の計測結果が予測に貢献していると考えられる。一方で、トレーニングデータの期間の

⁵ AUC の水準は、任意の 2 サンプルを取り出した場合にそのスコアの序列が真の序列を示している確率として解釈することが出来る。この意味で、今回の実証結果における AUC の改善度合いは定量的に見て限定的なものであると考えられる。一方で、予測モデルの精度に関する定量的な評価は AUC の様な指標のみではなく、その予測結果を実務的な目的に利用した場合にどの程度の経済的利益を実現できるかという観点から評価すべきという見方もある。

幅が同様で期間 2 より設定時期を 6 カ月前倒しした期間 3 においては、滞在人口の計測結果を利用しない場合で精度が高くなっている。さらに、期間 3 と比較してテストデータの設定期間をトレーニングデータの設定期間と連続するよう 6 カ月前倒しした期間 4 においても、滞在人口の計測結果を利用しない場合で精度が高くなっており、AUC の差分のマイナス幅は期間 3 や期間 4 とともに大きな差は生じていない。これらの結果から、滞在人口の計測結果の貢献に、テストデータの設定時期の違いは大きく影響しないものの、トレーニングデータの設定時期の違いが影響を与えた可能性が考えられる。

表 8. 分析期間ごとの滞在人口の計測結果の利用有無による精度比較

期間名	滞在人口有モデル AUC ①	滞在人口無モデル AUC ②	①－②
期間 1	0.6965	0.6849	0.0116
期間 2	0.6927	0.6830	0.0097
期間 3	0.6683	0.6829	-0.0146
期間 4	0.6073	0.6241	-0.0168

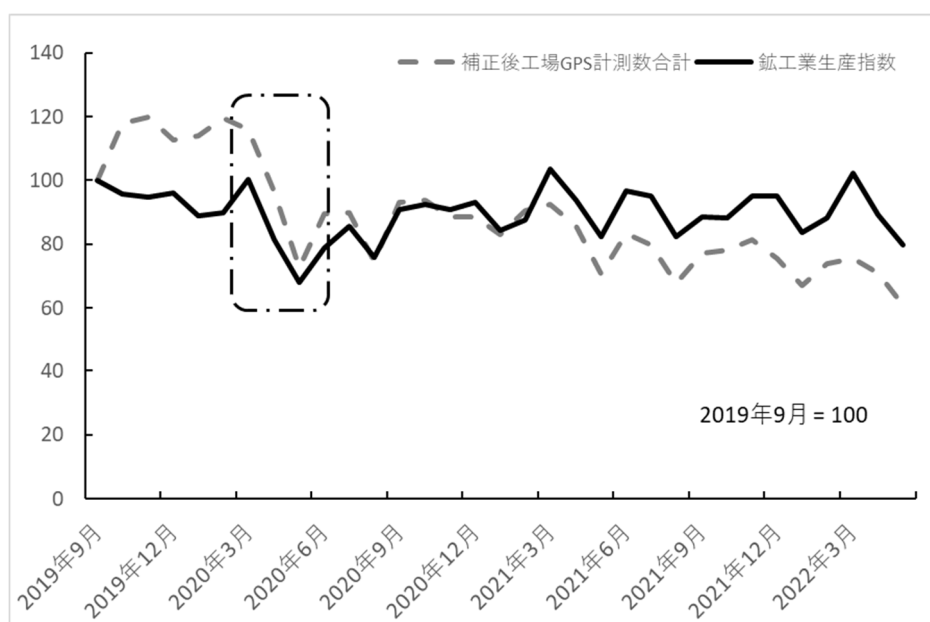
(注) 表は、期間と滞在人口の計測結果の利用有無ごとにトレーニングデータを利用して構築したモデルをテストデータに対して適用した場合の AUC を示している。滞在人口有モデル AUC①は滞在人口の計測結果を利用して構築したモデルの AUC を、滞在人口無モデル AUC②は滞在人口の計測結果を利用せずに構築したモデルの AUC を、①－②は両者の差分をそれぞれ表したものである。①－②がプラスの場合は滞在人口の計測結果を利用して構築したモデルの AUC が滞在人口の計測結果を利用しないで構築したモデルの AUC を上回ることを示しており、①－②がマイナスの場合はその逆を示している。

4-2. 結果の解釈

結果から、期間 1 と期間 2 では滞在人口の計測結果が予測に貢献したものの、トレーニング期間が共に 2020 年 3 月から 2021 年 5 月までとなる期間 3 と期間 4 の条件においては、滞在人口の計測結果が予測に貢献しなかった。期間 3 と期間 4 のトレーニング期間は、新型コロナウイルスのパンデミック初期で第一回目の緊急事態宣言が発令された時期を含む期間であり、社会的な情勢の変化がとりわけ大きい期間であったため、外部からの来訪者の制限や在宅勤務の増加などにより、直前の期間と比較して工場の生産量の減少規模以上に大きく工場の滞在人口が減少したことが考えられる。そこで図 5 の通り、経済産業省の

鉱工業生産指数の製造工業の値と、滞在人口の計測対象企業 22 社の工場の滞在人口の計測結果である、補正後 GPS データ計測数の推移を比較する。図を見ると、2020 年 6 月以降では工場の GPS データ計測数と鉱工業生産指数の変動幅が同様に近い推移をたどっている。一方で、点線枠で示す 2020 年 3 月から 2020 年 6 月までの期間については、鉱工業指数の減少率に比べて GPS データ計測数の減少率が顕著に大きいことが分かる。このことから、実際に前述したような事態が発生したと考えられる。このことが期間 3、期間 4 で滞在人口の計測結果が予測精度の改善に貢献しなかった要因として考えられる。

図 5. 鉱工業生産指数と補正後の工場の GPS データ計測数の関係



(注) 図は、実線で示す図 3 で示したものと同様の補正後の工場の GPS データ計測数の変動と、破線で示す鉱工業生産指数の変動の関係性について、2019 年 9 月から 2022 年 5 月までの期間で示したものである。鉱工業生産指数については、2019 年 9 月を 100 とした場合の比率となるよう各月の値を割合で示している。一点鎖線で囲んだ時期においては、補正後の工場の GPS データ計測数の減少幅が鉱工業生産指数の減少幅を大きく上回る。

(出典) 経済産業省「鉱工業指数」

一方で、期間 1 は期間 2 と比較すると、期間 2 では含んでいなかったパンデミック初期のデータも利用する形でトレーニングデータの期間幅を延長することにより、期間 2 よりも滞在人口の計測結果が予測により大きく貢献している。これはトレーニングデータの利用期間の長さと同様、パンデミック初期のデータの利用とのトレードオフの関係で、トレーニン

グデータの利用期間の長さが優位に働いたためと考えられる。

5. 結語

本稿では、GPS データを用いて計測した製造業の企業の工場の滞在人口データを利用することで、その仕入先企業の企業活動の予測に寄与するか否かを検証した。実証結果から、長期間のデータを用いて予測を行う場合か、データの期間が短い条件下で新型コロナウイルスのパンデミックの影響による社会情勢の変動が大きい期間を除いた期間で予測を行う場合において、滞在人口データが予測に寄与する可能性があるとの示唆が得られた。こうした実証結果については、現時点で利用可能な滞在人口データの期間が 2 年半分のみであったことから、分析の妥当性を検証する上ではさらなるデータ蓄積を図る必要がある。また、新型コロナウイルスのパンデミック初期の社会情勢の変化による影響がより小さいデータを用いて予測を行うことで、滞在人口データを利用することによる精度向上幅が大きくなるという推測を確認することも有用だろう。滞在人口データは蓄積が始まってから日が浅く、データを蓄積した上で本稿で取り組んだ実証分析を継続していくことが今後の最も重要な課題と言える。

なお、今回の実証分析では、予測の対象となる企業のリース需要を SMFL 社の契約データから、リース資産の種別について特段の区別をせずに計測した。この点に関して、第一に、リース資産の種別を明示的に取り扱った分析が有益と考えられる。例えば、生産に直接利用される設備がリースされている場合は、今回予測対象としたリース契約は将来の企業活動を前もって示す先行指標として位置づけられる。一方で、リース契約の対象が事業規模の拡大に伴って導入される資産である場合も考えられる。こうしたリース資産の種別に関する丁寧な取り扱い、本稿で構築した予測モデルの実務的な価値を高める上で重要である。第二に、複数のリース会社の契約データや銀行借入れを伴う自社での資産取得など、企業活動に対応した代理変数をより充実させたいうでアウトカム変数として用いることも重要であろう。こうしたデータは、本稿の目的である、オルタナティブデータを十全に活用した予測モデルの構築という目的に留まらず、昨今その重要性が認識されているエビデンスに基づいた政策立案（EBPM）の観点からも重要な分析の材料を与えるものである。

【参考文献】

Aditya Aladangady, Shifrah Aron-Dine, Wendy Dunn, Laura Feiveson, Paul Lengermann, Claudia Sahm (2019). “From Transactions Data to Economic Statistics: Constructing Real-time, High-frequency, Geographic Measures of Consumer Spending”, National Bureau of Economic Research Working Paper Series No.26253.

Jardet, C., and B. Meunier (2020). “Nowcasting world GDP growth with high-frequency data,” Banque de France Working Paper, No. 788.

王悠介, 須合智広, 高橋耕史, 松村浩平 (2021). “位置情報データによる経済活動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.21-J-2.

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-14.

水門善之, 柳井都古杜 (2020). “携帯電話GPSデータに基づく自動車生産量のナウキャストリングと株式投資戦略”, 人工知能学会第二種研究会資料, FIN-025, pp. 9-12.

Vasco M Carvalho, Makoto Nirei, Yukiko U Saito, Alireza Tahbaz-Salehi, “Supply Chain Disruptions: Evidence from the Great East Japan Earthquake”, The Quarterly Journal of Economics, Volume 136, Issue 2, May 2021, Pages 1255–1321

Arata Yoshiyuki, Miyakawa Daisuke “Demand Shock Propagation Through an Input-output Network in Japan” RIETI Discussion Paper, 22-E-027

上田翼, 和泉潔, 坂地泰紀 (2023). “人流データを用いた自動車サプライチェーン異常指数の構築”, 人工知能学会第二種研究会資料, FIN-030, pp. 40-44.