

集計データを用いた 鉄道運輸収入変動要因の影響の定量化

中川 伸吾¹・松本 涼佑²・鈴木 崇正³・渡邊 拓也³・深澤 紀子³

¹正会員 公益財団法人鉄道総合技術研究所 情報通信技術研究部 情報解析
(〒185-8540 東京都国分寺市光町 2-8-38)
E-mail: nakagawa.shingo.39@rtri.or.jp

²非会員 公益財団法人鉄道総合技術研究所 情報通信技術研究部 情報解析 (同上)

³正会員 公益財団法人鉄道総合技術研究所 情報通信技術研究部 情報解析 (同上)

鉄道事業者が旅客輸送によって得る収入には変動が見られる。この変動をもたらす要因は様々に考えられるが、それらの要因が収入に与える影響は定量化できていない。この影響を評価できれば、将来の営業施策の検討や、外部要因による将来の収入変動の見通しを立てることの一助となる。そこで、運輸収入データが集計データであることをふまえ、Causal Impact を適用して、収入を増減させると考えられる要因の影響の定量化を図った。実データを用いたいくつかのケーススタディにおいて、影響の程度や有意性について自然な考察が可能な結果を得て、運輸収入データに対する妥当な分析ができることを確認した。

Key Words: railway revenue, aggregate data, Causal Impact, increase in revenue

1. はじめに

旅客鉄道事業者にとって、旅客輸送によって得る運輸収入は、事業運営の根幹であり、その変動は経営に直結する重要な指標である。しかし、この変動をもたらす要因は定性的には様々に考えられるものの、定量的な評価、すなわち要因と考えられる事象が運輸収入に与えている影響の評価はできていない。このような評価、たとえば割引きっぷ販売などの営業施策の効果や、鉄道沿線での集客施設の開業などの効果の評価ができれば、将来の営業施策の検討や、外部要因による将来の運輸収入変動の見通しを立てることの一助となり、運輸収入データを営業施策・経営戦略により活用できることにつながる。

このような定量的な評価を難しくしている原因に、鉄道事業者が継続的に蓄積している運輸収入データが集計データであることが挙げられる。これは、集計データには要因の影響を受けるデータと受けないデータがひとまとまりになって集計されている場合があること、また、このような因果関係を評価する手法も大部分は集計前の個票データへの適用を前提としているためである。そこで、本研究では、このような実際の運輸収入データの特徴をふまえた定量的評価の方法について検討し、因果推論手法を適用することにより、データ整備や実験計画等

で時間的・金銭的コストのかかる個票データ分析を行うことなく、様々な収入変動要因の影響を定量化できることを示した。

2. 既往研究

運輸収入の変動を扱った既往の文献類は少なく、路線別の営業係数（100 円の収入を得るためのコスト、すなわちコスト／収入×100）を試算してその推移および路線間比較について考察した例¹などにとどまる。この背景としては、運輸収入は鉄道事業者の経営の根幹に関わる重要な情報であるため、公表されている運輸収入データが、1 年間や四半期ごと、全社合計や路線・地域での合計など、大きな単位で集計されたデータにとどまっており、詳細な分析に用いるのが難しいという点が考えられる。この点は、本研究のねらいである収入変動要因の影響の評価においても課題となる。

「事象 X が運輸収入に影響を与えている」、すなわち「事象 X が収入変動の要因である」という因果関係の有無を示すためには、因果推論の手法を適用すればよい。このような因果関係の分析手法として最も有効なものとしては、ランダム化比較試験が挙げられる²。これは観測対象を無作為に複数のグループに分けて、それぞれに

対して異なる処置を行い、結果を比較してその処置の効果を推定するものである。しかし、この方法はあらかじめ計画された社会実験で用いられるものであり、実験計画や実験結果の分析に、時間・金銭の面で莫大なコストがかかることが多い。また、運輸収入変動の要因分析に用いる場合、一部の旅客には事象 X がある状況、別の旅客には事象 X が無い状況で、それぞれきっぷを販売する必要があるため、事象 X の内容によっては旅客に不公平をもたらしてしまう。

このように社会実験が困難である場合には、既に社会に存在している実験のような状況を見つけ出す自然実験の手法が適用できる。考えられる自然実験手法として、操作変数法、回帰不連続デザイン、傾向スコアマッチングなどが挙げられる。しかし、これらを含め、因果推論の手法の多くは、個票データの活用を前提としている²⁾。中長期的な効果の考察に活用可能な、鉄道事業者が継続的に蓄積している運輸収入データは、基本的に集計データであるため、これらの手法を用いた分析は困難である。

このような集計データに適用できる自然実験の手法として、差の差分析（以下、DID という）³⁾がある。DID は、19 世紀に既に用いられた考え方であり、ある事象の影響を受けたと考えられる時系列データと、受けなかったと考えられる時系列データの推移を比較して、その事象が時系列データに与えた影響を推定する手法である。しかし、影響を受けなかったと考えられる時系列データとしてどのような時系列データを用いるかについて、分析者の恣意性が入ってしまうことが短所である。この短所を改良した因果推論手法として、2015 年に発表された Causal Impact（以下、CI という）⁴⁾がある。そこで本研究では、この CI を運輸収入の分析に活用することとした。なお、CI の鉄道分野への応用例は、管見の限りでは、我々のこれまでの取り組み⁵⁾以外に公表されていない。

3. 鉄道運輸収入データの特性

本研究では、ある鉄道事業者（以下、 α 社という）から運輸収入データを提供いただいた上で、このデータから収入変動要因の影響の定量化を図った。本章では、このデータの特性についてまとめる。

提供いただいたデータは以下の 2 種類である。いずれも、新型コロナウイルス感染症拡大前の、約 10 年分のデータである。

(a) エリア別月次収入データ。 α 社での様々なきっぷの売上額を、エリア・月・収入区分ごとに合計したものである。エリアとは α 社の路線や駅を概ね都道府県単位に区分したものである。路線や駅に属さない箇所（ α 社のインターネット予約サイトなど）で販売されたきっぷは特定のエリアに計上する。月とはきっぷが販売された時

表-1 エリア別月次収入データの収入区分

収入区分	きっぷの種類	きっぷの区間
定期券	定期券	全て
近距離	定期券以外(〇駅から〇円区間という形のきっぷ)	一定距離以内
中長距離	定期券以外	一定距離以上 (Z社内の区間)
他社関係	定期券以外	一定距離以上 (Z社内で完結しない区間)

点の月であり、実際にきっぷが利用されたタイミングとは必ずしも一致しない。また、 α 社で販売したきっぷであれば、他事業者のきっぷ、他事業者にまたがるきっぷも計上される。逆に、他事業者が販売した α 社のきっぷは計上されない。収入区分とは、きっぷの種類・区間に基づく区分であり、表-1 に示す 4 種類である。

(b) 券売機月次収入データ。駅の自動券売機での近距離きっぷの売上額を、駅・月ごとに合計したものである。ただし、無人駅などの小規模な駅については、その駅を管理する近隣の駅に統合されて計上される場合がある。また、月はきっぷが販売された時点の月だが、売上集計の締切のタイミングにより、月末の売上が翌月に計上される場合がある。

本研究は、これらのデータを分析対象とすることを前提とする。 α 社以外の事業者がもつ運輸収入データも、具体的な区分や粒度は様々にありうるが、同様に集計されたデータであると考えられる。

4. CI について

2 章で述べた CI は、複数の時系列データがあつて、そのうち 1 つに事象の影響があると考えられる状況において適用可能な因果推論手法である。事象の影響があると考えられる 1 つの時系列データを処置群と呼び、それ以外の時系列データを対照群と呼ぶ。処置群や対照群は全て同一時期・同一粒度の時系列データである必要がある。

CI は、事象が無かった場合の処置群の推移（実際には事象は起こっているため、これは反実仮想である）を、対照群の推移から推定して、実際の処置群の推移と比較することで、事象の発生が処置群に及ぼした影響の程度や、その影響が続いた期間を推定する手法である。推定のイメージを図-1 に示す。この反実仮想の推定は、事象発生前の推定対象時系列を対照群で説明するベイズ構造時系列モデルをつくり、これを事象発生後のデータにあてはめることで行う。対照群の時系列データ全てを推定に用いるのではなく、変数選択のシミュレーションを多数回重ねて、その結果よく選択された時系列を説明変数として採択する仕組みを採ること、2 章で述べた DID の短所を改良しているほか、対照群が処置群と同じ種類の時系列データでなくてもよい、という特長を得て

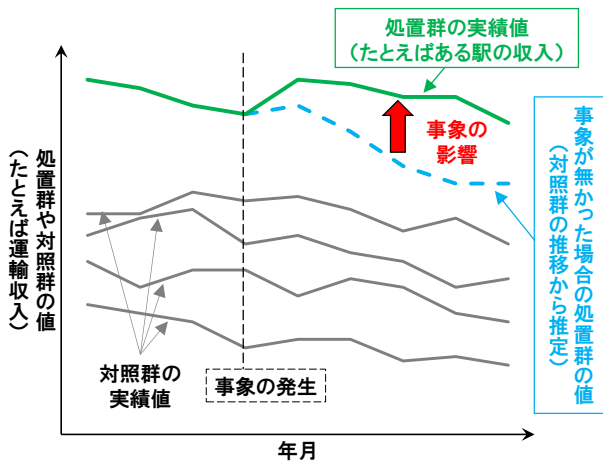


図-1 CIでの影響推定の考え方

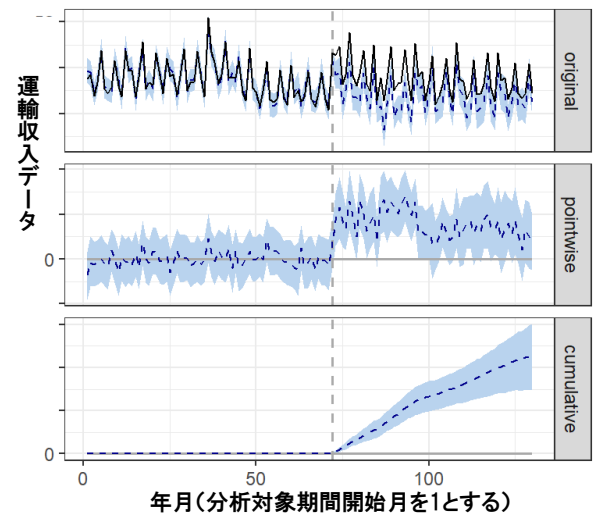


図-2 Rで出力されるCIの計算結果の例

いる。

5. 事象の影響の定量化（ケーススタディ）

3章で述べた、 α 社から提供いただいた運輸収入データに対して、データの対象時期の間に発生した事象についての影響の定量化を、CIを用いて実施した。

CIの適用手順は以下のとおりとした。

- ①分析対象の事象Xについて、それが影響を与えると考えられる収入を処置群と設定する。
- ②処置群がエリア別月次収入データである場合は、他エリアのエリア別月次収入データを対照群と設定する。処置群が券売機月次収入データである場合は、同エリアの他駅の券売機月次収入データを対照群と設定する。ただし、事象の特性上この設定が不相当である場合は、妥当な対照群を設定する。
- ③①②の処置群・対照群を用いてCIの計算を行う。計算に際しては、統計ソフトR、およびそのパッケージであるCausalImpactを用いる。ベイズ構造時系列モデルの構築には事象発生の前月までのデータを用い、そのモデルを用いて事象発生月以降の効果を定量化する。ただし、事象発生が月の終わりに近く、事象発生月の処置群への影響が小さいと想定される場合は、事象発生月のデータもモデル構築に用いて、効果の定量化の対象期間を事象発生月の翌月以降とする。
- ④③の計算で推定されたベイズ構造時系列モデルにおいて、対照群のうち、モデル中で用いられた割合（Inclusion確率という）が5%未満であるデータを対照群から除く。この操作は、モデルへの寄与がほぼ無いデータを対照群から除いて統計的誤差を小さくするために行う。
- ⑤④で抽出されたデータのみを対照群として、再度③と同様にCIの計算を行い、得られた結果を、事象

Xが処置群に与える影響とする。

出力される結果の例を図-2に示す。横軸は月であり、分析対象の開始月を1とした通算の月数である。縦破線は事象Xがあった月である。「original」欄（図-2上段）は、実線が処置群の実績値、破線が事象Xが無かった場合の処置群の推定値である。「pointwise」欄（図-2中段）の破線は、各月における処置群の実績値と推定値の差である。これが各月の、事象Xが運輸収入に与えた影響額である。「cumulative」欄（図-2下段）の破線は、pointwise欄の破線の値を事象Xがあった月から累積した値、つまり、事象Xが運輸収入に与えた通算の影響額である。CIが推定するのはoriginal欄の破線の値であり、pointwise欄とcumulative欄の破線の値は、original欄の実線と破線の値から算出される値である（したがって推定値である）。各欄の薄青塗りの部分は、破線で示す推定値の90%信用区間であり、推定値がこの範囲にある確率は90%であるということを意味する。

CIにおいては、影響程度の信用区間の下限が正、つまりpointwise欄の薄青塗り部分全体がプラスの領域にあれば、その月において、事象Xの影響は有意にプラスであると言える。マイナスについても同様である。信用区間としては一般には95%信用区間が用いられることも多いが、本研究では、収入変動要因の把握を将来の営業施策につなげたいという目的をふまえ、できるだけ多くの事象を有意な要因として抽出できるように、やや緩い90%信用区間を用いることとした。

なお、分析対象データや出力結果に示される金額は、本稿ではデータの都合上秘匿する。事象Xが運輸収入に与えた影響の程度は、影響額の代わりに、影響による増減率、具体的には（処置群の実績値－処置群の推定値）／（処置群の推定値）で表記する。

(1) 新線開業

α社での新線開業の影響を評価する。この新線は、3つのエリア A,B,Cにまたがる路線であり、他の鉄道事業者とも直通運行している。

まず、エリア A の収入について効果を計算した結果、開業当月を事象発生月とした場合、収入への影響は有意でない、もしくは有意にマイナスであるという結果になった。一方で、開業前月を事象発生月とした場合、図-3 に示すように他社関係収入で、単月では有意とは言えないもののおよそ 7 年間、平均で+6.6%の影響が見られ、累積では有意なプラスであった。この開業によって、他の鉄道事業者への直通列車が大きく増えて他社にまたがる区間の利便性が高まったことから他社関係収入にプラスの影響があり、また、当該新線のきつぶの販売開始が開業 1ヶ月前だったため、影響が開業前月から現れたと考えられる。

一方、エリア B とエリア C では、図-4～図-7 に示すように、エリア A と同様に開業前月から他社関係収入が有意にプラスとなっているだけでなく、中長距離収入についても、開業当月から有意にプラスとなった。特に影響が顕著であったエリア C の中長距離収入 (図-7) では、開業直後の 1 年間平均で+30%となった。エリア B,C では、エリア A との間の需要が多く、この開業によって利便性が大きく上がったことから、中長距離収入にも有意な影響が現れたと考えられる。

これらの結果において、事後裾野確率 (Posterior tail-area probability, 結果が偶然に得られる確率と解釈できる) はいずれも 5%未満であり、ここからも、事象の影響が有意であったと言える。

(2) 高速道路開通

分析対象期間中に、α社営業範囲で高速道路の開通がいくつかあった。その影響を評価する。高速道路の開通は特定の駅だけでなく広い範囲に影響を及ぼすこと、および、高速道路の対抗交通機関となるのは主に都市間の優等列車であると考えられることから、本分析ではエリア別月次収入データを用いた。

まず、あるエリア D での高速道路開通について述べる。この開通では、エリア D の中心都市と、隣接エリアの主要都市との間が全通した (ただし、この中心都市と逆方向の主要都市との間には未開通部分が残っている。以下、この開通を 1 回目開通という)。

4 種類の収入区分を合計した額を対象にした分析結果 (図-8) では、1 回目開通前後での差はほぼ見られず、少なくとも有意な増減は全く無い (事後裾野確率も 46% ある)。ただし、これを収入区分ごとに分析すると、近距離収入 (図-9) では有意に増加、中長距離収入 (図-10) では減少となっており、その増減がほぼ相殺されて

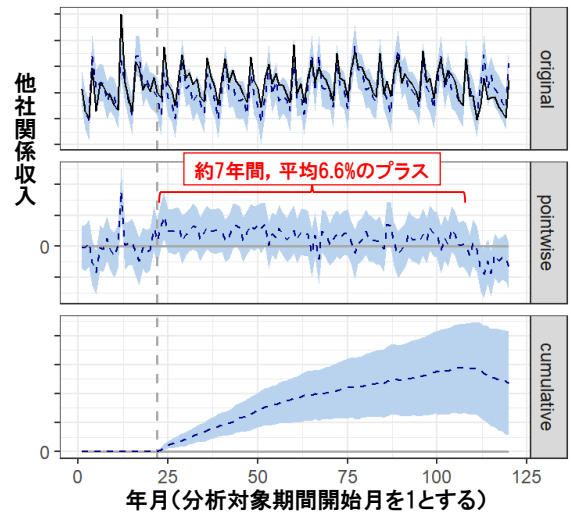


図-3 新線開業がエリア A の他社関係収入に与えた影響

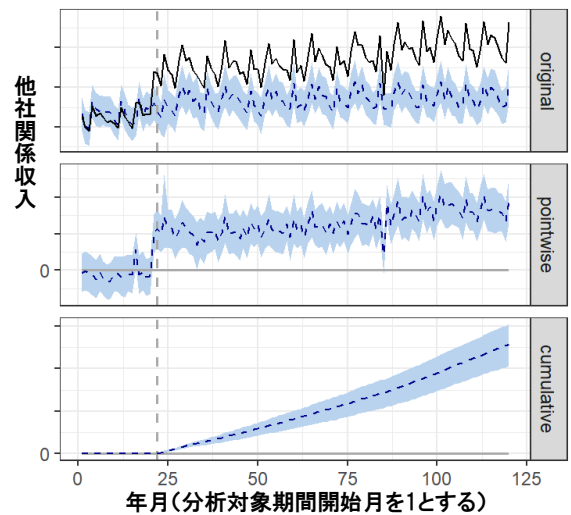


図-4 新線開業がエリア B の他社関係収入に与えた影響

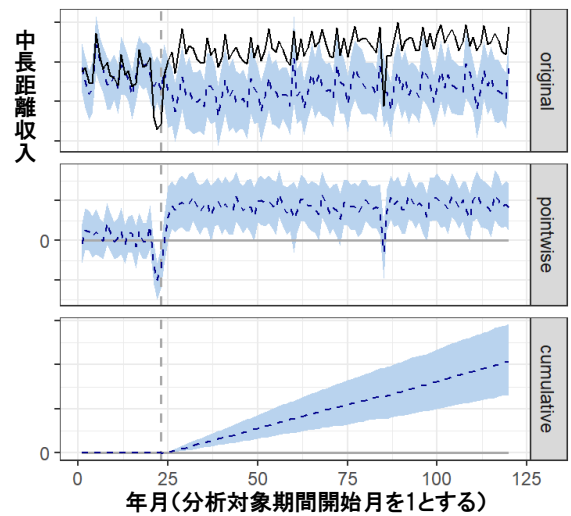


図-5 新線開業がエリア B の中長距離収入に与えた影響

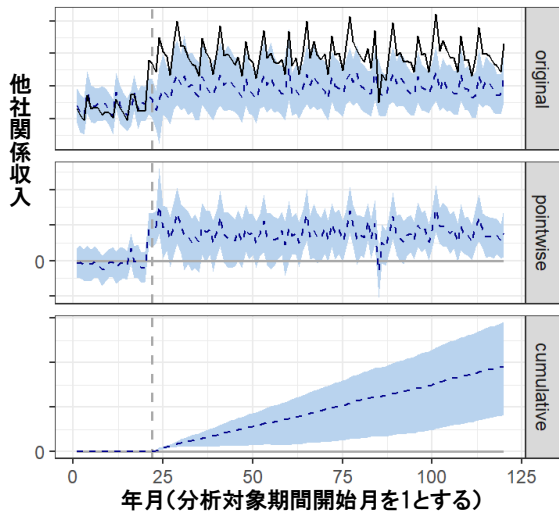


図-6 新線開業がエリア C の他社関係収入に与えた影響

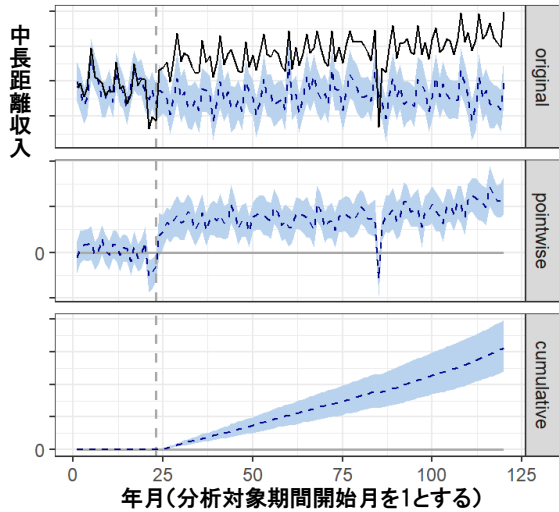


図-7 新線開業がエリア C の中長距離収入に与えた影響

いることが確認された。近距離収入については、ほぼ同時に集客施設が開業したことの効果と考えられる。一方、中長距離収入は、単月ではほとんどの月で有意とは言えないものの、累積効果では 90%信用区間の上限がほぼ 0 となっており、事後裾野確率も 3%と十分に小さい結果となった。影響の程度は平均-2.8%であった。

次に、その後逆方向の主要都市との間も全通したため、その時点での効果を同様に計算した結果を図-11、図-12 に示す（以下、この開通を 2 回目開通という）。1 回目開通と同様に、ほとんどの月で有意とは言えないが、近距離収入、中長距離収入ともマイナスとなっている。ただし、中長距離収入の減少の程度は 1 回目開通と比べて小さく、開通直後の 1 年間の減少額を比較すると 2 回目開通は 1 回目開通の 53%である。2 回目開通の区間は、1 回目開通の区間と比べて、距離が短く（半分以下である）、都市部で並行する一般道路がある程度あることから、鉄道への影響が小さかったものと考えられる。事後

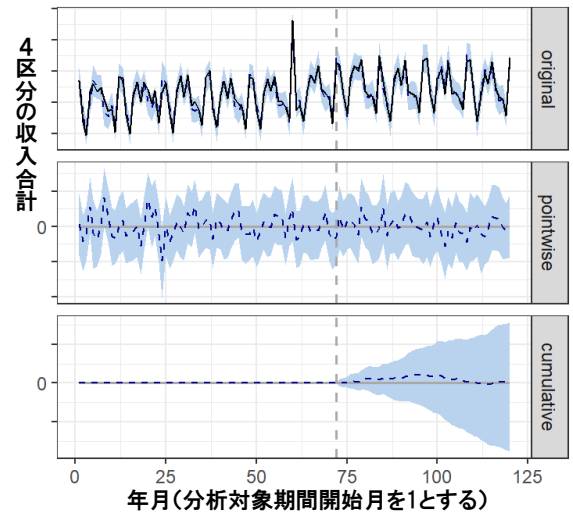


図-8 1 回目開通がエリア D の収入合計に与えた影響

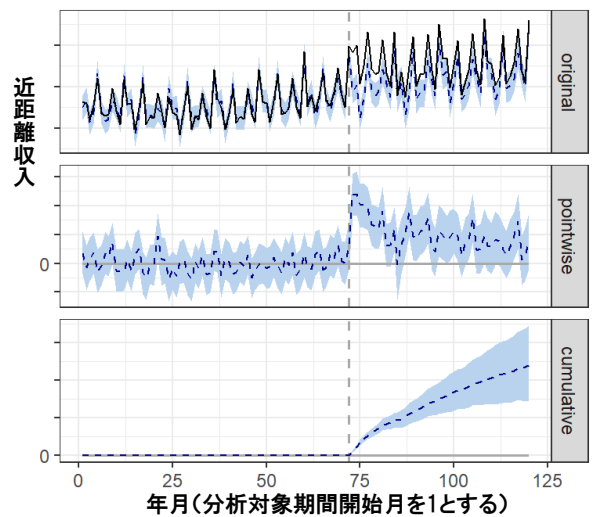


図-9 1 回目開通がエリア D の近距離収入に与えた影響

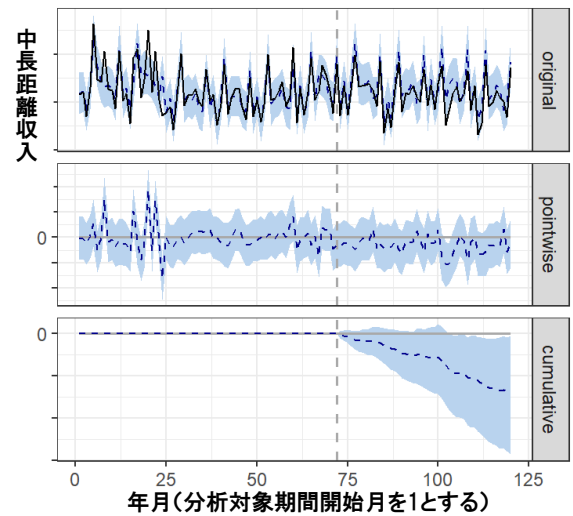


図-10 1 回目開通がエリア D の中長距離収入に与えた影響

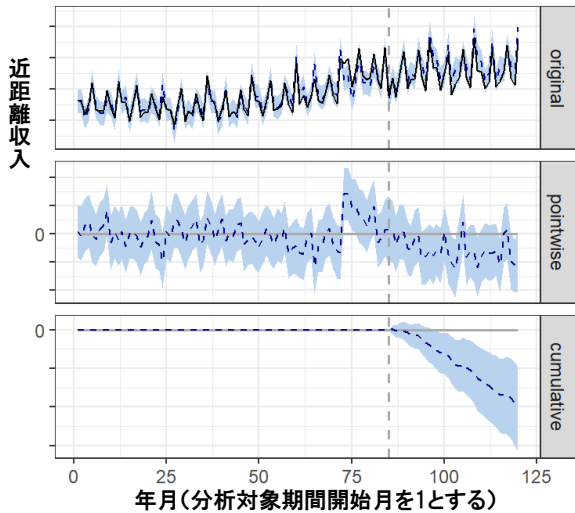


図-11 2回目開通がエリア D の近距離収入に与えた影響

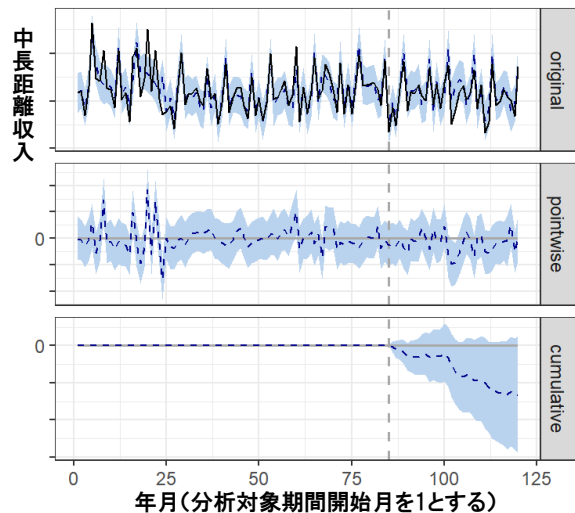


図-12 2回目開通がエリア D の中長距離収入に与えた影響

裾野確率は、近距離収入（図-11）で 1%未満，中長距離収入（図-12）で 9%であった。

これらの分析事例は、高速道路開通の影響が有意だったと示すものではないが、影響は総じてマイナスであり、その影響の程度が開通する道路の長さや周辺の道路事情によることを示唆するものである。また、図-8～図-10 に示す 1 回目開通の分析結果は、収入全体を分析してもわからない影響を、収入区分ごとの分析で定量化できることを示すものであり、事象に応じて分析対象データを細分化して処置群や対象群を設定することの必要性を示すものである。

(3) 世界遺産登録

分析対象期間中に、α社営業範囲で世界遺産の新規登録があったため、これが最寄駅である S 駅の券売機収入に与えた影響を評価する。この世界遺産は S 駅から数 km 離れているが、アクセス手段として自家用車ではな

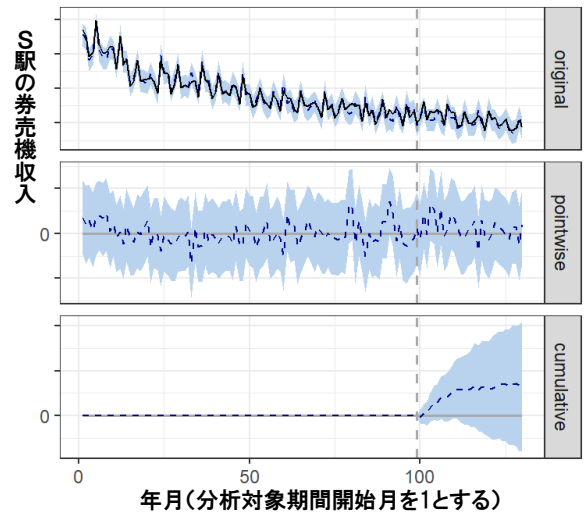


図-13 世界遺産登録が S 駅の券売機収入に与えた影響

く S 駅からのバスまたはタクシー利用が推奨されているため、当該世界遺産への訪問者増加が S 駅の券売機収入にプラスの効果をもたらす可能性が考えられる。

計算結果を図-13 に示す。登録後の月のうち 61%で、pointwise の破線の値はプラスとなっており、累積での影響は+2.2%となった。ただし、90%信用区間の下限が全ての月でマイナスであり、事後裾野確率も 22%と高く、影響が有意であるとは言えない。

なお、この結果はあくまでも、S 駅の券売機収入に関するものである。当該世界遺産および S 駅が所在する都市の統計データによれば、この都市の「歴史・文化」カテゴリの観光地点における年間入込観光客数は、登録年にその前年と比べて 5%増加している。これをふまえると、遠方からの観光客の動向がわかる運輸収入データを分析した場合には、より大きな影響が見られる可能性がある。

(4) 自然災害

分析対象期間中に、α社営業範囲のある都市で自然災害が発生し、通行止や大渋滞が起こるなど道路事情が大幅に悪化した日があった。これが当該都市にある T 駅の券売機収入に与えた影響について考察する。なお、道路事情は数日で回復したため、この災害の影響は発生当月のみに表れると想定される。

計算結果を図-14 に示す。災害発生当月に、+14%の影響が有意に観測された一方、前述の想定のとおり、翌月以降には有意な影響は見られない。道路事情悪化によって道路から鉄道への一時的な需要転移が起こり、これが T 駅の券売機収入にプラスの影響をもたらしたと考えられる。

6. おわりに

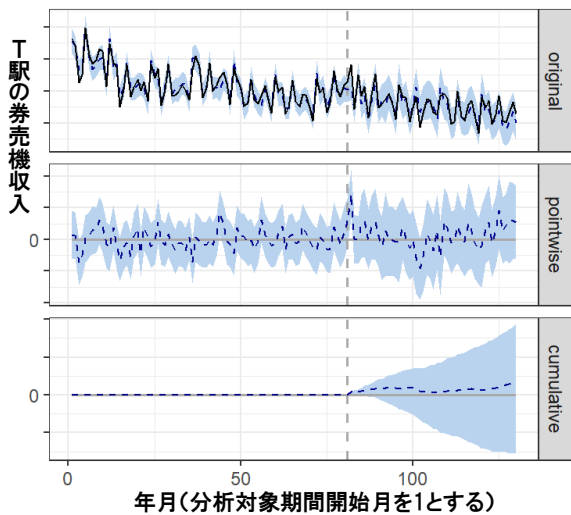


図-14 自然災害がT駅の券売機収入に与えた影響

本研究では、実際の運輸収入データに対して CI を適用した分析を行い、様々な事象が運輸収入に与える影響の定量化を図った。その結果、影響を定量化するとともに、影響の程度や有意性について自然な考察が可能であり、妥当な分析ができることを確認した。この方法は、ケーススタディ対象とした Z 社のデータだけでなく、実務で広く活用されている、期間や地域などで集計された様々な運輸収入データに対しても適用可能である。

今後の課題として、以下 3 点を挙げる。第一は、類似の事象でも、あるいは同一事象でも収入区分によって、影響が様々である理由を精査することである。本稿でも自然な考察はできているが、さらに分析事例を積み重ねることで、より妥当性の高い知見につなげることができると考えている。第二は、分析対象期間中に起こった複数の事象の影響を切り分ける方法の構築である。それぞれの事象に対して異なる対照群を設定できる場合はよいが、そうでない場合、分析事例を積み重ねた上で、それ

ぞれの事象と類似する事象から推定できるようにするのがよいと考えている。第三は、営業施策への活用可能性を高めるため、鉄道事業者が施策としてコントロール可能な内部要因の分析事例を重ねることである。本稿ではデータの制約上、ケーススタディで分析対象とした内部要因は、(1)の新線開業のみであった。たとえば、きっぷの種類ごとの運輸収入データから割引きっぷ発売の影響を定量化することなどが考えられる。

謝辞：運輸収入データを提供いただいた α 社に、深く御礼申し上げます。

REFERENCES

- 1) 梅原淳：全国の鉄道「営業係数」大公開，週刊東洋経済，第 6775 号，pp. 70-77，東洋経済新報社，2018. [Umehara, J.: *Zenkoku no tetsudo "Eigyō Keisu" daikokai*, Weekly Toyo Keizai, No. 6775, pp. 70-77, Toyo Keizai inc., 2018.]
- 2) 大久保将貴：因果推論の工具箱，理論と方法，Vol. 34, No. 1, pp. 20-34, 2019. [Okubo, S.: *Tools of Causal Inference*, Sociological Theory and Methods, Vol. 34, No. 1, pp. 20-34, 2019.]
- 3) 中室牧子，津川友介：「原因と結果」の経済学，pp. 98-104，ダイヤモンド社，2017. [Nakamuro, M. and Tsugawa, Y.: *Causal Inference in Economics*, pp.98-104, Diamond Inc., 2017.]
- 4) Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. : Inferring causal impact using Bayesian structural time-series models, *The Annals of Applied Statistics*, Vol. 9, No. 1, pp.247-274, 2015.
- 5) 松本涼佑，中川伸吾，鈴木崇正，渡邊拓也，深澤紀子：大型商業施設の開業等が鉄道輸送の収入に与える効果に関する一考察，第 28 回鉄道技術連合シンポジウム講演論文集，SS5-1-1，2021. [Matsumoto, R. et al.: *Effect of shopping complex openings on nearby railway station revenue*, Proceedings of the 28th jointed railway technology symposium, SS5-1-1, 2021.]

QUANTIFICATION OF THE IMPACT OF EVENTS ON RAILWAY REVENUE USING AGGREGATE DATA

Shingo NAKAGAWA, Ryosuke MATSUMOTO, Takamasa SUZUKI,
Takuya WATANABE and Noriko FUKASAWA

Railway operators' revenue from passenger transport fluctuates in accordance with various factors. However, the impact of those factors on the fluctuation is unclear. Therefore, we try to quantify such impact of some events e.g. opening of new line. For this purpose, "Causal Impact," a time series causal inference method, is applied to aggregated revenue data. As a result of some case studies for some events, we have quantified the level and significance of the impact reasonably and have confirmed that the simple method enables appropriate analysis.