

経済危機が雇用に与える影響 — causal forest を用いた試論 —

The Impact of Economic Crisis on Regional Employment:
A Causal Forest Approach

安達 有 祐*

ADACHI Yusuke

小 川 光**

OGAWA Hikaru

津布久 将 史***

TSUBUKU Masafumi

This paper analyzes the impacts of economic shocks on Japanese regional employment using causal forest algorithms, and our data-driven findings are compared with those estimated under the standard OLS estimation. Our results are as follows. The largest declines in employment growth during the recessionary period associated with the crisis are in industries and regions where employment grew before the crisis, and in industries that have a large share of employment in the region. These results differ from the standard OLS results, which show that the negative impact of shocks is greater in regions with higher unemployment rates. In addition, testing the robustness with respect to outliers, which are samples that could be arbitrarily removed, we show that the results obtained by causal forests are robust, but those of OLS are not.

Keywords: Economic shocks, employment, regions, industries, Random Forest

* 國學院大學経済学部経済学科

Faculty of Economics, Kokugakuin University

** 東京大学大学院経済学研究科

Graduate School of Economics, The University of Tokyo

*** 大東文化大学経済学部

Department of Economics, Daito Bunka University

I. はじめに

近年、計量経済学では分析手法の発達が著しい。これまでの計量経済学では回帰分析から出発し、説明変数が被説明変数に及ぼす効果を明らかにしてきた。そこに因果推論を計量経済学に応用する動きが登場し、政策決定やビジネスへ活用されるようになった。さらにコンピュータの発達にともない、膨大なデータを読み込み数値計算や判別を行うアルゴリズムである機械学習も利用されるようになった。本研究では、Wagner and Athey (2018) により提案されたcausal forestと呼ばれる機械学習を応用した手法を使い、経済危機が日本国内の地域の雇用にどのような影響を与えたのか、その影響はどのような地域で顕著であったのかを明らかにする。これを通じて、経済危機の効果を検証する上でデータ駆動的な手法であるcausal forestが通常の最小二乗法に比べて有用となる可能性を示す。

雇用は一般市民にとって大きな関心事であると同時に、経済学者、政治家、政策実務家にとっても関心の高い問題である。人々に安定した生活を提供し、失業を減らすことは効率的な資源配分を促し社会の公平性に繋がるからだ。一方で、安定した雇いを脅かすものとして、多くの人が懸念しているのが経済危機とそれに伴う景気後退である。特に、企業活動の縮小をもたらす経済危機は、労働需要を減少させ、その結果、職を失う労働者の数を増加させる。近年、雇用を脅かした危機の代表例として、2008-2009年の世界金融危機と2020年以降のCOVID-19の世界規模での感染拡大が挙げられる。例えば総務省統計局の「労働力調査」によれば、世界金融危機の前である2007年の7月～9月期の完全失業率は3.8%であったが、世界金融危機の後の2009年7月から9月期の完全失業率は5.4%まで悪化した。

これらの危機が地域経済に及ぼした影響は多くの研究者によって解明を試みられてきた。例えば、Fingleton et al. (2012) はイギリスのデータをもとに、危機に伴う不況が地域によって異なることを見出し、Holl (2018) は世界金融危機時のスペインのデータを用いて、危機の雇用への悪影響が都市部より地方で大きかったことを明らかにしている。彼らの研究は伝統的に計量経済学で用いられてきた回帰分析を応用し、経済危機が地域に与える影響を明らかにしているものの、どのような要因が危機の影響の地域差をもたらすのかについて答えるには十分な分

析がなされていない。既存の分析方法に従ってこの間に答えるためには、データ分析を行う事前の段階で経済危機の影響を説明する理論的仮説が必要になる。しかし、経済危機は様々な変数に同時にまたは連続的にショックを与えるため、これらを含めた理論モデルを構築し検証可能な仮説を導くことは困難である¹⁾。

そこで、理論モデルから示される仮説がなくても分析を行う手法が提案されてきた。このような手法はデータ駆動的な推定手法と呼ばれ、例えばWager and Athey (2018) は、機械学習の手法である回帰木 (regression forest) と無作為比較実験 (RCT) を組み合わせることで因果推論を行うcausal forestと呼ばれる手法を提案した。それまでのRCTに代表される因果推論の手法では、経済危機の影響はコントロールグループとトリートメントグループの間での平均介入効果しか求められず、各地域に応じてどのように危機の影響が異なるのかを知ることはできなかった。しかし、この手法を用いれば条件付きの平均介入効果を求めることができ、どのような特徴を持つ地域が経済危機から大きな影響を受けたのかを推定することができる。

機械学習手法を用いた地域経済研究のうち、我々の関心に最も近いのは、イタリアのCOVID-19パンデミックが地域経済に与えた影響を分析したCerqua and Letta (2022) である。COVID-19の第一波におけるイタリアの地域経済への影響はその地域でどれだけ感染が広がっているのかという疫学的パターンとは相関が無く、むしろ社会的集約リスクが高い地域や事前の段階で労働市場に脆弱性が見られる地域で経済的ダメージが大きいことを明らかにした²⁾。

この手法だけでなく、地域経済学の多様なトピックに様々な機械学習の手法を適用する研究が増え始めている。Galdo et al. (2022) は、2011年のデータを用いて、インドにおける「都市部」の再定義に機械学習のアプローチを用いている。彼らは、標準的なアプローチと機械学習の一種であるランダムフォレスト・アプローチで定義された都市部を比較し、例えば、後者で推定される都市化率は、従来のアプローチで定義された都市部より6～9ポイント低いことを示した。機械学習による都市部の再定義は、他の国でも行われている。例えば、Arribas-Bel et al. (2022) は、密度ベースの機械学習アルゴリズムを適用してスペインの市街地を定義し、従来の行政

的に定義された市街地と比較している。2017年の1200万件の建物データを用いた分析を通じて、彼らは機械学習アルゴリズムを用いて画定された都市圏が、通勤型都市圏に近いことを示している。都市・地理学の分野では、Reades et al. (2019) が、ロンドンの4835の軌道における近隣の変化について、2001年から2021年までの社会・経済・環境要因のデータを用いて2021年の変化を予測する機械学習の分析を実施している。また、英国における近隣変化の分析は、Patias et al. (2022) により1971年から2011年という長い期間で試みられている。

以上を受けて本研究は経済的な負のショックに伴う景気後退が地域・産業ごとに与えた影響の異質性に着目し、どのような地域・産業において危機の影響が強かったのかを明らかにすることを目的とする。負のショックからの影響が異なる要因の分析には、最小二乗法と機械学習を用いた手法であるcausal forestによる推定を行う。具体的には、総務省の「労働力調査」の四半期ごとの地域・産業別雇用者数のデータを用いて、1985年から2020年までの間に起こった4つの経済的な負のショックにより雇用者数がどの程度減少するのか、またどのような特徴を持つ地域や産業において負の影響が強くなるのかを明らかにする。

本研究の分析により、以下の2点が明らかとなった。第一に、causal forestの分析により経済的なショックの影響が強くなる地域の特徴である。具体的には、危機前に雇用者数が増加しているような地域・産業や地域内の全雇用者に占める労働者のシェアが高い産業において経済危機を通じた雇用の成長率に負の影響が大きくなることが明らかとなった。特に後者の場合、地域内の全雇用者に占める労働者の割合が20%を超えるような産業は、そうでない産業と比較して負の影響が有意に大きくなるが示された。このように機械学習の手法を用いた場合は共変量で条件を付けたときの影響の大きさを統計的に比較できるため、より細かく分析することが可能となる。第二に、causal forestを用いることの有用性である。本論文において、最小二乗法で分析を行った場合は外れ値の影響により推定結果が安定しないが、causal forestを用いた場合は外れ値の影響をあまり受けないことが示された。これは、外れ値の処理方法によって結果が左右されないcausal forestの結果の頑健性を示すものである。

本稿の構成は以下のとおりである。2節は、負の

ショックの計測方法やcausal forestの推定手法について説明する。3節は、使用したデータについて述べる。4節は、分析の結果について説明をし、結語となる5節において本研究の分析結果をまとめ、今後の課題をまとめる。

II. 推定方法

本研究は、負のショックからの影響が地域や産業によって異なることに着目し、その要因を明らかにするために最小二乗法とcausal forestによる推定を行う。これらの要因分析を行うために、地域や産業ごとのショックからの影響を測る方法が必要となる。本節では、危機からの影響の測り方とショックの影響の異質性についての要因分析の手法について述べる。

1. 危機の影響の測定

負のショックからの影響を測る方法について、本研究ではFingleton et al. (2012) にならい見かけ上無相関な回帰方程式システム (Seemingly Unrelated Regression; SUR) を採用する。Fingleton et al. (2012) によれば、雇用者数の成長において地域間での相関関係が存在しており、通常の最小二乗法では推定値にバイアスを持つとされている。この相関関係を考慮した推定を行うことが必要であり、SURは事前に相関関係を仮定することなく推定することができるため、負のショックの影響を地域・産業別に推定する手法とした。SURの推定式は以下の式で表すことができる。

$$e_{rit} = \beta_{1ri}R_{1t} + \dots + \beta_{4ri}R_{4t} + const_{ri} + \varepsilon_{rit} \quad (1)$$

ただし e_{rit} は地域 r の産業 i における t 期の雇用者の増加率を表し、 R_{mt} は m 回目 ($m=1, \dots, 4$) の負のショックがあったことを示すダミー変数である。また、 β_{mri} は m 回目のショックが地域 r の産業 i の雇用の成長率に与える負の影響の大きさを表している。この係数は定数項である平常時を基準とし、0から離れた値をとるほど平常時と比較してどれほど負の影響があったのかを表している。(1)式を各地域の各産業ごとに推定することで得られた結果を比較することで、危機が各地域・各産業の雇用に対してどのような影響を与えたのかがわかる。例えば、 $\beta_{mli} = \dots = \beta_{mji}$ であれば m 回目のショックが雇用の

成長率に与える影響は地域間で差がないことを表しており、 $\beta_{mr1} = \dots = \beta_{mrj}$ であれば m 回目のショックが雇用の成長率に与える影響は産業間で差がないことを示している。

2. 異質性の要因分析

SURで推定した負のショックの影響は、地域・産業ごとに異なる。本研究の目的は、この異質性の要因を分析する手法を検討することにある。例えばHoll(2018)は、2008年の世界金融危機がスペインの雇用者数の成長率に与えた影響を市町村別に推定した。推定には、線形回帰モデル(OLS)を使用し、考えられる要因を説明変数に加え危機の前後で推定値に差が出るかを検証した。しかしながら、Martin and Sunley(2015)によれば異質性の要因は複雑に絡んでおり、さらにそれらの要因と雇用者数の成長率との関係はかなりのあいまいさが残っているとされている。考えられる要因と雇用の成長率の間の関数形を事前に仮定しておく必要があるOLSのような分析手法は、推定結果にバイアスが生じる可能性がある。加えて、サンプルをいくつかのサブサンプルに分割して条件付きの推定を行うさいにも、そのサンプルの分割に恣意性が入る可能性がある。そのため、負のショックの異質性の要因を明らかにするためには事前の仮定が少ない手法が望ましい。近年、発展してきた機械学習を用いた分析では事前の仮定が少なく、データ駆動型の分析が行われている。例えば、Cerqua and Letta(2022)はCovid-19が地域の雇用者数に与える影響を分析したが、反実仮想としてCovid-19が分析対象の地域に存在しなかったときの雇用者数をrandom forestなどの機械学習の手法から推定した。このように、要因が結果に与える影響が複雑な場合、機械学習の方法は有効な手段となる。そこで、本研究も機械学習の手法を用いてショックが地域や産業に与える影響の異質性についての要因を分析する。具体的には、Wager and Athey(2018)で提案されたcausal forestを採用する。

3. causal forestの分析方法

causal forestの目的は、処置効果(treatment effect)を推定することにある。ここで、サンプル $i = 1, \dots, n$ について、観察される共変量 X_i 、結果 Y_i と処置ダミー $W_i \in \{0, 1\}$ の変数が得られているとする。このとき、 x で条件付けた処置効果 $\tau(x)$

は以下のようにあらわされる。

$$\tau(x) = E[Y_i^{(1)} - Y_i^{(0)} | X_i = x]$$

$Y_i^{(1)}$ はサンプル i に処置がとられたときの結果、 $Y_i^{(0)}$ は処置されなかった場合の結果を表している。しかし、この2つの結果は同時に起こることはないので、一般的に処置効果 $\tau(x)$ は観察されない。しかし、unconfoundnessが成り立てば、この処置効果を推定することができる。ここで、unconfoundnessとは、結果と処置が変数で条件付けることで独立になることであり

$$\{Y_i^{(1)}, Y_i^{(0)}\} \perp\!\!\!\perp W_i | X_i$$

とあらわされる。この条件が成立するとき、 $X_i = x$ の平均処置効果は

$$\widehat{ATE}_x = \frac{1}{\sum_i 1[W_i = 1, X_i = x]} \sum_i 1[W_i = 1, X_i = x] Y_i - \frac{1}{\sum_i 1[W_i = 0, X_i = x]} \sum_i 1[W_i = 0, X_i = x] Y_i$$

となる。ただし、 $1[\cdot]$ は指示関数、 X はすべて離散変数とする。平均処置効果(average treatment effect; ATE)はすべての x の平均処置効果に加重平均をとることで計算される³⁾。

$$\widehat{ATE} = \sum_x \widehat{ATE}_x Pr(X_i = x)$$

しかしながら、実験ではなく統計調査データを用いる場合はunconfoundnessが成り立つ可能性は低い。そこで、Athey and Imbens(2016)やWager and Athey(2018)は、random forestを応用しunconfoundnessを成立させる手法を提案した。Wager and Athey(2018)で提案された手法はcausal forestと呼ばれる。Athey and Imbens(2016)で提案されたcausal treeを複数回 ($B > 2$) 実行し毎回の平均処置効果 $\hat{\tau}_b(x)$ についての平均をとったもの ($\hat{\tau}(x) = B^{-1} \sum_{b=1}^B \hat{\tau}_b(x)$) がcausal forestによって推定される平均処置効果となる。このcausal treeは、classification and regression tree(CART)を基本としているため、ここではCARTの推定方法について説明する⁴⁾。

CARTの作成は、共変量で特徴づけられる空間を

複数の空間に区切り、その分割した空間内の予測誤差を最小にするように作られる。この区切られた空間 (leaf) の数を L 、空間を分割する仕切りを partition (Π) とする。このとき

$$\Pi = \{l_1, l_2, \dots, l_L\}, \cup_{j=1}^L l_j = X$$

であり、 $l(x; \Pi)$ を $x \in l$ となる区切られた空間 $l \in \Pi$ とする。いま、partition $\pi \in \Pi$ が与えられたときの条件付き期待値を

$$\mu = E[Y_i | X_i \in l(x; \Pi)]$$

とすると、実際のサンプルから得られる μ の推定量 (予測値) は

$$\hat{\mu}(x; \Pi) = \frac{1}{\sum_i \mathbb{1}[X_i \in l(x; \Pi)]} \sum_i \mathbb{1}[X_i \in l(x; \Pi)] Y_i$$

となる。partitionはこの予測値と実際の誤差の二乗の平均 (Mean squared error: MSE) を最小化するように作成される。サンプルをMSEを計算するテスト用 S^{te} と partitionの作成のためのトレーニング用 S^{tr} に分けた上で、leafごとのMSEを以下の式から求める。

$$\begin{aligned} & MSE\mu(S^{te}, S^{tr}, \pi(S^{tr})) \\ &= \frac{1}{|i: i \in S^{te}|} \sum_{j \in S^{te}} \{(Y_j - \mu(X_j, S^{tr}, \Pi))^2 - Y_j^2\} \end{aligned}$$

ただし、 $|i: S|$ は S を満たす i の数を表す。CARTは、以下の式を最大化するようにpartitionを置いていく

$$Q(\pi) = -E_{S^{tr}, S^{te}}[MSE\mu(S^{te}, S^{tr}, \pi(S^{tr}))]$$

causal forestはこのCARTの推定手法を利用し、unconfoundnessが成り立つように共変量に仕切りを置いていく手法である。Wager and Athey(2018)によればcausal treeの推定方法はdouble sample treeとpropensity treeの2つ存在するが、本研究におけるcausal tree推定はpropensity treeを使用する。propensity treeの推定手順は以下のとおりである。

0. サンプルサイズ n について (X_i, Y_i, W_i)

が観察されている。ここで、 X_i は変数、 Y_i は結果、 W_i は処置ダミーとする。また、各leafに k 個以上のサンプルが含まれているとする。

1. サンプルからランダムにサブサンプルを取り出す。このとき、取り出したサブサンプルを $\Gamma \in \{1, \dots, n\}$ とする。また、サブサンプルのサイズは $|\Gamma| = s$ で非復元抽出とする。
2. サブサンプル Γ を用いてtreeを作成する。ここで作成するtreeは、(X_i, W_i) のみを使用する。つまり、共変量を利用して処置を予測できるモデルであり、leaf内のサンプルがunconfoundnessを満たすようなpartitionを作成することができる。ちなみに、ここのtreeは上記のCART法で作成する。
3. 以下の式から x を含むleafの処置効果を推定する

$$\begin{aligned} \hat{\tau}(x) &= \frac{1}{|i: W_i=1, X_i \in L|} \sum_{i: W_i=1, X_i \in L} Y_i \\ &\quad - \frac{1}{|i: W_i=0, X_i \in L|} \sum_{i: W_i=0, X_i \in L} Y_i \end{aligned}$$

causal forestはこの手順で作成したcausal treeを複数回 (今回は1000回) 実行したものである。また、causal forestによって推定される平均処置効果 ($\hat{\tau}(x)$) は、causal treeによって推定された平均処置効果 ($\hat{\tau}_b(x)$) の平均をとったもの $\hat{\tau}(x) = B^{-1} \sum_{b=1}^B \hat{\tau}_b(x)$ となる。

III. データ

1. 地域と産業の分類

負のショックからの影響の要因を分析するためには、データの収集頻度が高く、複数の経済的な負のショックが含まれる長期の時系列データであること、細かく地域や産業が分かれていることが求められる。本論文では、これらの条件を満たす総務省の労働力調査のデータを使用した。このデータは、四半期別の地域・産業別の雇用者数を含む、1947年以降の長期時系列データとなっている。

本研究は、データの収集が可能な1985年の第1四半期以降の地域・産業別の雇用者数についてのデー

タを使用する⁵⁾。このデータセットには11の産業が含まれるが、漁業、鉱業、インフラ産業（電気・ガス・熱供給・水道業）の雇用者数が極端に少ないため今回の分析からそれらを外した8産業を対象とする。地域は、図1が示すように10の地域区分に分かれている⁶⁾。したがって、10の地域と8つの産業別に1985年の第1四半期から2020年の第1四半期までのデータが含まれている。しかしながら、2011年の第2四半期から2013年の第4四半期の11四半期分のデータは、東日本大震災のため欠損している。そのため、実際には36年間の四半期分のデータから11四半期を除いた133四半期分のデータを利用した。また、本研究の労働力調査のデータは季節調整が施されたものである。これら133四半期のデータからSURモデルを使用して経済的な負のショックからの影響を(1)式より推定した。

2. 危機の期間：処置群とコントロール群

表1は、内閣府がBry and Boschan (1971) にしがい設定した景気基準日付と呼ばれる景気循環の山と谷を示している。本研究の分析対象期間は第11循環から第14循環が該当し、この間に経済的な危機

は4回発生した。第11循環の後退期のはじまりは、バブル経済の崩壊である。次の第12循環は1997年に起こったアジア通貨危機をきっかけとして後退が発生した。第13循環の後退期は、アメリカで起こったITバブルの崩壊をきっかけとしている。第14循環の後退期のはじまりは、2008年にアメリカから世界に広がった世界金融危機である。

本研究では、これらの経済危機に関連する景気循環に基づき、危機の影響を受けた期（処置群）と経済危機の影響がない期（コントロール群）を設定する。処置群は、経済危機の生じたとされる山の期からその後経済が拡張期に転じる谷Ⅱまでの期間とし、これに対するコントロール群は経済危機が生じる直前の経済拡張期（谷Ⅰから山までの期間）からランダムに5期間を選択した。具体的に第11循環を例にとれば、処置群は1991年の第1四半期から1993年の第3四半期の32期間とし、コントロール群は1986年の第4四半期から1990年の第4四半期の17期間から5期をランダムに選択した。

IV. 分析の結果

1. SURモデルの推定結果

(1)式より、負のショックが雇用者数の増加率に与える影響をSURモデルから推定した $\hat{\beta}_{mir}$ が得られた。この推定値は、4つのショック、8の産業、10の地域ごとに得られるため、320個の推定値が存在する。また、コントロール群の推定も同様に行った。図2は β の推定結果をまとめたものである。この図は箱ひげ図となっており、線分の両端は外れ値を除くデータの最小値と最大値を表している。また、箱の両端はそれぞれ25%分位点、75%分位点の値を示し、箱内の線は中央値を表している。図の上側はコントロール群の推定結果を、下側は処置群の結果をまとめたものである。処置群の $\hat{\beta}$ の中央値はおよそ -0.003 、コントロール群の $\hat{\beta}$ の中央値はおよそ

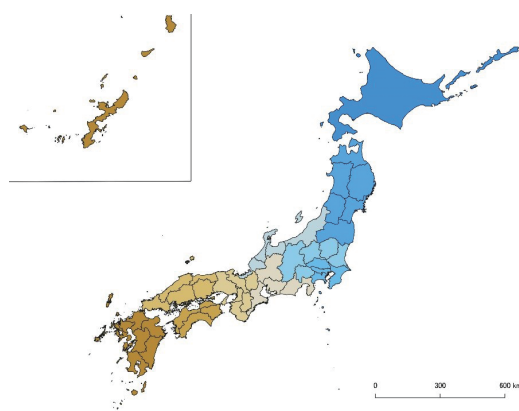


図1 地域分類

表1 景気基準日

	谷Ⅰ	山	谷Ⅱ	期間(月)		
				拡張	後退	全期間
第11循環	1986年11月	1991年2月	1993年10月	51	32	83
第12循環	1993年10月	1997年5月	1999年1月	43	20	63
第13循環	1999年1月	2000年11月	2002年1月	22	14	36
第14循環	2002年1月	2008年2月	2009年3月	73	13	86

出所：内閣府 (<https://www.esri.cao.go.jp/jp/stat/di/hiduke.html>) より作成

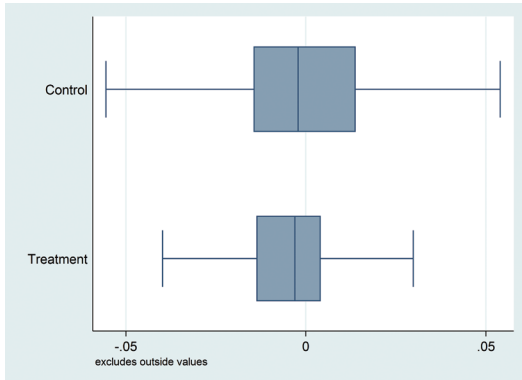


図2 経済危機からの影響

注：外れ値を除いた箱ひげ図になっており、線分の両端は外れ値を除くデータの最小値と最大値を表している。また、箱の両端はそれぞれ25%分位点、75%分位点の値を示し、箱内の線は中央値を表している図の上側はコントロール群の推定結果を、下側は処置群の結果をまとめたものである。

-0.021である。したがって、中央値で見た場合は処置群のほうがコントロール群よりも負の影響が大きいがわかる。一方で、分散で見た場合コントロール群のほうが処置群よりも大きくなっている。つまりSURモデルを用いた分析を通じて、経済危機の影響は地域や産業ごとにばらつきは見られるものの単純な中央値の比較でも雇用の成長率に対して負の影響があることが分かった。

2. 異質性の要因分析

(1)式をSURモデルにより推定した結果によれば、負のショックからの影響や回復は地域や産業によって異なる可能性が示された。この異質性の要因を明らかにするために最小二乗法とcausal forestにより推定を行った。ここで、異質性の要因として挙げられるのは先行研究を踏まえ以下の4つとした。

1. 危機の直前1年間の雇用成長率 (*emp.growth*)
2. 危機時の失業率 (*umemp.rate*)
3. 地域の全雇用に対する産業別雇用比率 (*emp.share*)
4. 地域の人口 (*lnpop*)

1.と2.は、危機発生時の地域や産業の雇用動向や水準によって、危機の影響が異なるかどうかを捉えるために用いる。3.は、特定の産業への依存度が、

危機の地域に対する影響とどのように関係しているかを捉えるために使用される。これは、危機の際に地域特化の経済が有効に機能するかどうかを捉える上でも有効である。最後に、4.は地域の規模に関する経済が働くかどうかを捉えるものである。大きな地域は、ある産業の雇用ショックを地域内の他の産業の雇用で吸収することができる可能性がある。また、産業ごとの特性をとらえるために産業ダミー変数に加えた。

(1) 最小二乗法の推定結果

まずは最小二乗法を用いて負のショックからの影響に異質性が生じる要因について行った分析について説明をする。危機の影響を捉えた $\hat{\beta}$ ((1)式で得られたSURモデルの推定値)が、どのような要因に左右されるのかを明らかにするために以下の式を推定する。

$$\hat{\beta}_{mir} = b_0 + b_1 X_{ir} + b_2 X_{ir} \times dum_m + \epsilon_{mir} \quad (2)$$

ただし、 X_{ir} は異質性について考えられる要因として先に挙げられた変数 (*emp.growth*, *umemp.rate*, *emp.share*, *lnpop*)と産業ダミー変数を、*dum*は危機のときに1をとるダミー変数とする。(2)式を推定することで、平常時においてこれらの要因が雇用の成長率に与える影響である(b_1)に加えて、これらの要因が与える影響は危機時にはどう変化するのか(b_2)を明らかにすることができる。ここで重要なのは b_1 や b_2 の間に差があるかどうかであり、 $b_1 = b_2$ という帰無仮説が棄却されるかどうかの

表2 OLS 推定による経済危機の影響

	b_1 (s.d)	b_2 (s.d)	F-value (p-value)
(i) <i>emp.growth</i>	-.0339 (.0333)	-.0931 (.0657)	2.01 (.15)
(ii) <i>unemp.rate</i>	.0021 (.0016)	-.0036 (.0021)	2.75* (.09)
(iii) <i>emp.share</i>	-.0864 (.0457)	.0662 (.0698)	.90 (.34)
(iv) <i>lnpop</i>	.0006 (.0026)	.0039 (.0037)	1.12 (.28)

注：*は10%水準で有意であることを示す。どの推定結果もサンプルサイズは619で、産業の固定効果がコントロールされている。

検定を行う必要がある。もし棄却されれば異質性の要因がショックの時の雇用者数の増加率に影響することになる。

OLS推定の結果は表2にまとめられている。各列には、行に示された説明変数についての b_1 と b_2 の推定値および帰無仮説 $b_1 = b_2$ について F 検定を行った結果が示されている。F 検定の結果、失業率についてのみ、F 値が2.75となり10%の有意水準でこの帰無仮説が棄却された。これは平常時における失業率の推定係数は.0021であり経済危機のダミーと失業率の交差項の推定係数は-.0036であるため、危機の前に失業率が高い地域や産業ほどショックにより雇用者数の成長率が低下すること意味する。

(2) causal forestによる推定結果

次にWager and Athey (2018) により提案されたcausal forestを用いて、地域や産業の属性といった共変量がSURモデルから推定した $\hat{\beta}$ に与える影響を明らかにする。まずは、共変量で条件付けた平均の経済危機が雇用の成長率に及ぼす影響を推定し、これらが地域や産業について異質性を持つことを確認する。図3はcausal forestを用いた反実仮想により推定された負のショックが雇用者数の増加率に与える影響をヒストグラムで表したものである。負のショックの後の雇用成長率は、一部の産業や地域でプラスになっているもののほとんどの場合でマイナスになっている。またこれらの負の影響が-0.015から0の間にあることから、産業や地域の属性によって負のショックからの影響について異質性が存在していることがわかる。

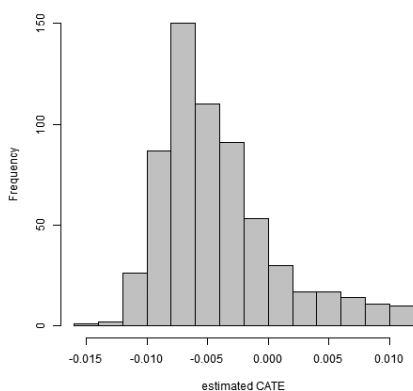


図3 causal forestによる分析結果

注：横軸は処置効果の水準を縦軸はその度数を示す。

本研究の目的は、causal forestを用いてこれらの異質性を生み出す要因を解明することである。先に挙げられた4つの要因を用いて条件付けた経済危機の影響が、どのような地域・産業で大きかったのかを明らかにする。図4は経済危機によって雇用の成長率にどのような影響があったのかを、地域や産業の特徴を表す共変量に応じてどのように異なるのかを表している。横軸は各共変量の値、縦軸はそれらの共変量で条件付けられたときの平均処置効果を表している。これらの図において、点線の端は最大値と最小値、箱の端は25%分位点と75%分位点を表している。箱内の実線は中央値を表している。

図4 (a)と(b)から、経済危機を通じた雇用の成長率に与える負の影響が大きくなるのは、危機前の雇用成長率が高いもしくは地域内の全雇用者に占める労働者のシェアが高いことが要因であることがわかる。これは以下のような検定からも明らかにすることができる。危機前の雇用成長率であれば、この変数の値が1以上（成長率がゼロ以上であった地域・産業のグループとそれ以外のグループに分け、それぞれについて平均と標準誤差を求める。1以上のグループでは平均値は-.008であり、標準誤差は.0022、1よりも小さいグループの平均値は-.0005、標準誤差は.0030となり、この2つのグループ間の差を検定するとz値が10.133であった。そのため経済危機の影響が1%の有意水準で2つのグループの間に差が無いという仮説は棄却できる。したがって、経済危機の前に雇用成長率が高い地域や産業は負のショックが大きかったことがわかる。

また地域内における全雇用者のうち産業の労働者が占める割合についても、20%以上とそれ未満のグループに分けると割合の大きいグループの平均値は-.0086、標準誤差は.0015、小さいグループの平均値は-.0035、標準誤差は.0023である。危機前の雇用成長率のケースと同様にこの2つのグループ間の差を検定すると、z値は7.228となりグループ間に差がないという仮説は有意水準1%で棄却される。したがって、地域内の全雇用者のうち産業の労働者が占める割合が大きい地域や産業は、負の経済的ショックの影響を強く受けることが示唆される。なお、図4 (c)と(d)に見られるように、地域人口と失業率の地域による違いが危機の影響の地域差を生む要因になっている可能性は低いことがわかる。

経済危機が雇用に与える影響

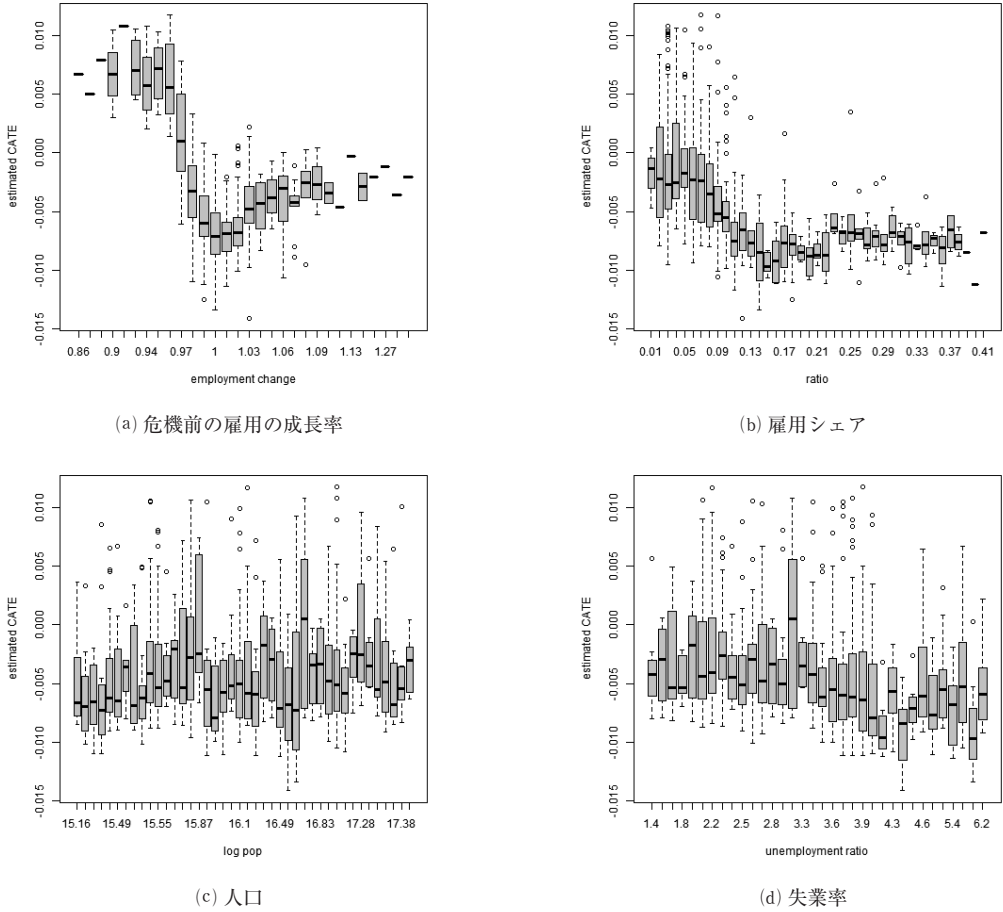


図4 危機時の条件付き平均処置効果

注：横軸は各変数の値、縦軸は条件付きの平均処置効果を表す。白丸は外れ値を表す。(a)は危機前の四半期の雇用成長率、(b)は全労働者に占めるある産業の雇用者の割合、(c)は当該地域の人口の対数、(d)は当該地域の失業率である。

3. 外れ値による推定結果の違い

推定結果が外れ値に大きく左右されるような推定手法は、結果が安定しないことから良い手法とは言えない。先に示した結果のサンプルサイズは619であるが、ショックの直前の雇用成長率が1.5を超えるサンプルがあり、そのサンプルは雇用者数が少ない地域・産業であったため分析から外していた。そこで、本節ではこのサンプルを含めたものを使用し同じ推定を行ったとき、これまでの推定結果と今回の結果がどのように違うのか確認する。

(1) 最小二乗法と外れ値

まずは、外れ値を含めたデータを使用し最小二乗法で(2)式を推定した。表3は、推定の結果をまとめ

表3 OLS推定による経済危機の影響:外れ値を含めたケース

	b_1 (s.d)	b_2 (s.d)	F-value (p-value)
(i) <i>emp.growth</i>	.0131 (.0237)	-.1402 (.0615)	5.19 (.02)
(ii) <i>unemp.rate</i>	.0019 (.0016)	-.0034 (.0021)	2.51 (.11)
(iii) <i>emp.share</i>	-.0859 (.0459)	.0657 (.0699)	.88 (.34)
(iv) <i>lnpop</i>	.0004 (.0026)	.0041 (.0037)	1.24 (.26)

注：**は5%水準で有意であることを示す。どの推定結果もサンプルサイズは620で、産業の固定効果がコントロールされている。

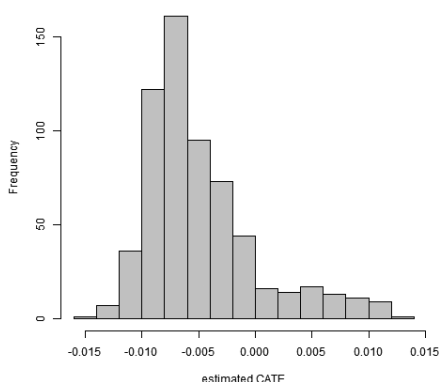


図5 causal forestによる雇用の成長に対する影響：外れ値を含めたケース

注：横軸は処置効果の水準を縦軸はその度数を示す。

たものである。外れ値を含めた推定の場合、ショック期において直前の雇用成長率が有意に影響するものの、外れ値を含めず推定した際に有意となっていた失業率は有意にならなかった。

推定の結果は、表3にまとめた。今回の帰無仮説は $b_1 = b_2$ であり、この仮説を検定するためにF検定を行った。ショック期において直前の雇用成長率 (*emp.growth*) のみ帰無仮説が棄却された。具体的には、平常時における直前の雇用成長率の推定係数は.0131であり、負のショック期のダミーと直前の雇用成長率の交差項の推定係数は-.1402となった。 $b_1 = b_2$ を帰無仮説としてF検定を行った結果、F値が5.19となり5%の有意水準でこの帰無仮説を棄却できる。したがって、危機の前に雇用の成長率が高い地域や産業ほどショックにより雇用の成長率が低下することが示唆される結果となった。

また、外れ値を含めず推定した場合の結果で有意に差が存在するとされた失業率は、外れ値を含めると有意にならなかった。平常時における失業率の推定係数は.0019であり、負のショック期のダミーと失業率の交差項の推定係数は-.0034となった。 $b_1 = b_2$ を帰無仮説としてF検定を行った結果、F値が2.51となり10%の有意水準でもこの帰無仮説を棄却できない。

(2) causal forest と外れ値

最後に、causal forestを用いて推定を行った結果について説明する。図6は、ショックに伴う不況時

に発生した雇用の成長率の低下と地域や産業の特徴を表す共変量の関係を示したものである。外れ値を含めた推定と同じように分析を行った結果、causal forestの推定結果に大きな変化はなかった。つまり、外れ値に対して頑強な推定手法であることが示された。

具体的には、成長率を示す変数の値が1以上（成長率がゼロ以上）の大きい地域・産業における負の影響の平均は-.0082であり、標準誤差は.0023、その値が1よりも小さい（成長率が負の）地域・産業における負の影響の平均は0、標準誤差は.0030となる。したがって、この2つのグループ間の差を検定するとz値が11.869となるため、影響が1%の有意水準で2つのグループの間に差がないという仮説は棄却できる。したがって、前節と同様にショックの前に成長率が高い地域や産業はショックによる負の影響を受けやすいといえる。

また、地域内における全雇用者のうち産業の労働者が占める割合が20%を超える場合の負の影響の平均は-.0081、標準誤差は.0015となる。一方で、この割合が20%を下回る場合の負の影響の平均は-.0036、標準誤差は.0023である。この2つのグループ間の差を検定するとz値は6.637となり、負の影響が2つのグループ間で差がないという仮説は有意水準1%で棄却される。したがって、こちらの結果も同様に地域内の全雇用者のうち産業の労働者が占める割合が大きい地域や産業は、負の経済的ショックの影響を強く受けることが示された。

V. おわりに

本研究は、世界金融危機などの負の経済的なショックに伴う景気後退が地域や産業の雇用者数の変化に与える影響を分析した。このような負の経済的なショックは地域や産業によって異なる影響を与えることがFingleton et al. (2012)などの研究により明らかになっているが、なぜ地域や産業によって影響が異なるのかを明らかにすることは難しい。その理由としては、分析を行うために必要な理論的な仮説を立てることが難しいからである。例えば、負のショックが生産サイドに与える影響だけでなく、労働者の期待やリスクに対する態度が組み合わさっていることや様々な要因が相互作用しており、回帰分析のような手法を用いた場合に推定値にバイアスが生じる可能性がある。さらに、異質な影響を明ら

経済危機が雇用に与える影響

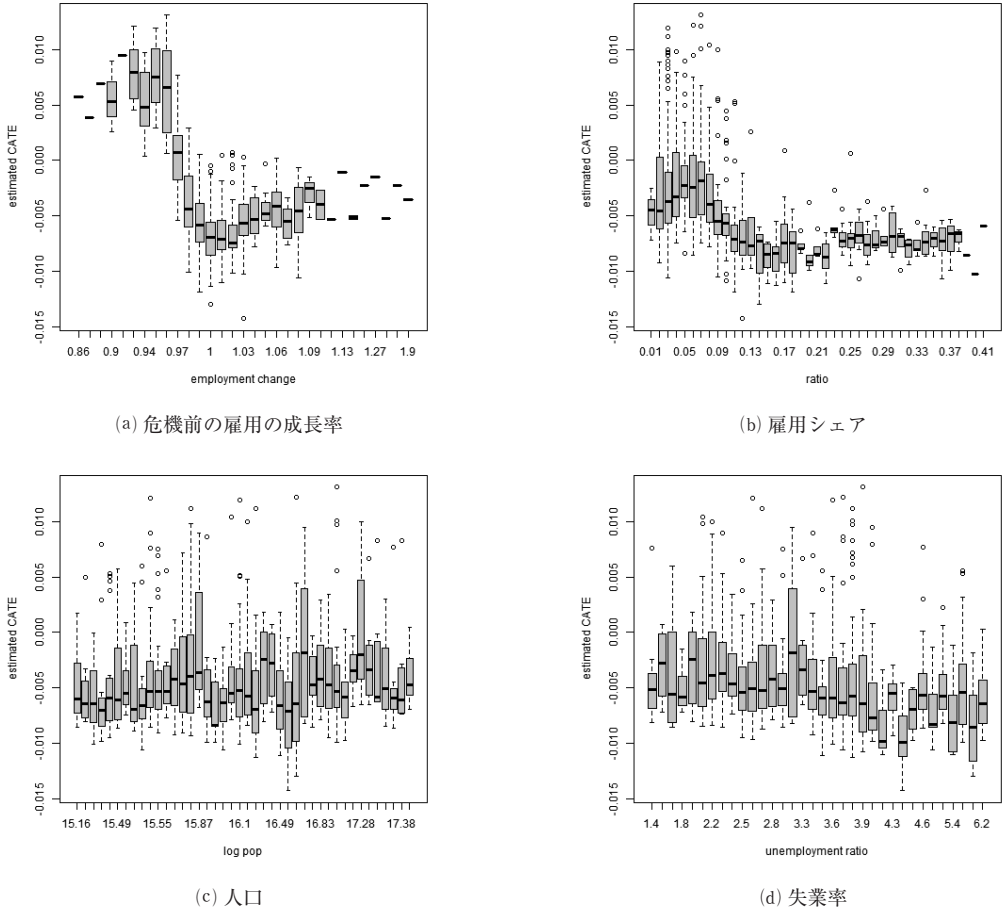


図6 危機時の条件付き平均処置効果：外れ値を含むケース

注：横軸は各変数の値、縦軸は条件付きの平均処置効果を表す。白丸は外れ値を表す。(a)は危機前の四半期の雇用成長率、(b)は全労働者に占めるある産業の雇用者の割合、(c)は当該地域の人口の対数、(d)は当該地域の失業率である。

かにするためには、データをいくつかのサブグループに分けて分析することが考えられるが、理論的な背景がないため分析者の恣意性が入り込む可能性がある。

本研究は、これらの問題を緩和するための試論として、負の経済的なショックが地域・産業別の雇用者数の変化に与える影響をWager and Athey (2018)で提案されたcausal forestを用いて分析した。この手法は、理論的な背景がない場合でもどのような特徴を持つ地域・産業が負の経済的なショックから受ける影響が強く出るのかを明らかにすることができる。総務省の「労働力調査」の四半期ごとの地域・産業別の雇用者数のデータを用いて1985年から2020年までの間に起こった4つの負の経済的なショック

の影響を本研究の分析の対象とした。

まずは、負の影響の大きさを計測するためにFingleton et al. (2012)で提案された見かけ上無相関な回帰方程式システム (SUR) を採用した。このSURから推定された、負のショックの影響は地域・産業ごとに異なっており先行研究と同様に異質性を持つことが示された。本研究では、地域・産業別の異質性に関して、その要因を分析する手法を検討した。具体的には、伝統的な計量経済学的手法である最小二乗法と機械学習を応用したcausal forestを用いた。causal forestの分析の結果として次の2つことが明らかとなった。1つは、危機前に雇用者が増加しているような地域・産業において負の経済的なショックにより雇用者数が減少すること、もう1

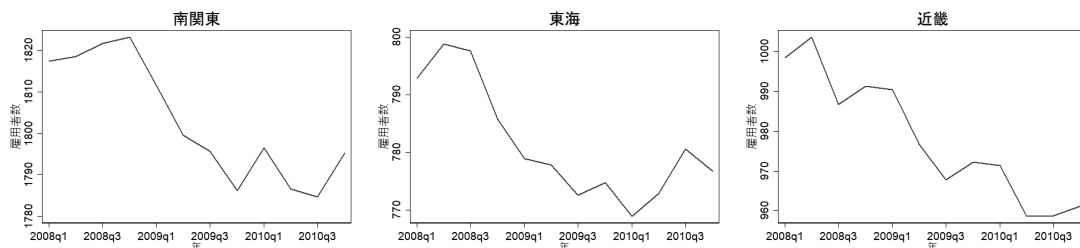


図7 世界金融危機後の雇用者数の変化

注：縦軸は地域内の総雇用者数を表す。

つは地域内の残雇用者に占める労働者のシェアが高い産業、特にシェアが20%を超えるような産業は雇用者数の減少が大きかった。

また、最小二乗法の結果と比較した際にcausal forestは外れ値による影響を受けにくいことも明らかとなった。統計調査データは外れ値を含む可能性が高く、これらの影響を受けやすい推定手法はその推定結果の信頼性も低くなる。負の経済的なショックが地域・産業の雇用に与える影響を分析するためには統計調査データを用いる必要があり、このような統計調査データを用いる上で外れ値の影響を受けにくいことから、本論文の結果は、causal forestの手法は有用な推定手法の一つとなる可能性を示すものといえよう。

しかしながら、負の経済的なショックが雇用に与える影響を計測する上でいくつかの課題が残されている。1つは、負の経済的なショックが雇用に与える影響の期間も異質性を持つ可能性がある。例えば、図7は世界金融危機後の南関東、東海、近畿地方の総雇用者数の変化を表している。これらの図によると、東海地方では2010年の第1四半期に雇用者数が最小値を取っておりここまでの負の影響と考えられる。一方で、雇用者数が最小となる期までを負の影響とするのであれば、南関東と近畿地方は2010年の第3四半期までを負の影響ととらえることになる。本研究では負の影響を受ける期間を内閣府が設定した景気基準日付と呼ばれるものを利用し、すべての地域において負の影響を受ける期間を同一として分析を行ったが、この期間の長さも地域・産業によって異なる可能性がある。したがって、今後の課題として負の影響の長さも同時に分析できる手法を提案することがあげられる。

謝辞

3名の著者は名古屋大学在学中に根本二郎教授から指導を賜りました。特に、本稿を含めて安達の研究は、指導教員として常に前向きな言葉を掛けてくださった根本教授のご指導が拠り所になっております。根本教授から賜った言葉は安達の宝物であり、ご指導を頂いたことに心から感謝を申し上げます。

注

- 1) 経済に対する負のショックが生産サイドの効率性や生産性に影響を与えることはよく知られている (Canello and Vidoli, 2020; Adachi et al., 2022)。また、企業間の生産サプライチェーンにも影響を与える (Brandenburg, 2016; Kawakubo and Suzuki, 2022)。深刻な危機は、少なくとも短期的には労働者の期待やリスクに対する態度および生活様式等に影響を与える (Hoffmann et al., 2013; Cassar et al., 2017; Hanaoka et al., 2018)。これらが組み合わせられて労働需給に影響を与えるため、他の変数が変化しないことを前提に、ある変数の変化を通じてショックの雇用への影響を抽出することは困難である。さらに、危機が単純な経路で影響するのでなければ、様々な要因が相互作用する可能性も否定できず、それらの複雑な関係を分解した上で原因を特定することは難しい (Martin and Sunley, 2015)。
- 2) Cerqua and Letta (2022) によれば社会的集約リスクが高くなるのは当該地域において、接客業務のような不特定多数の人々との接触が不可避な産業 (小売業、宿泊業、飲食サービス業、娯楽業等) が大きな割合を占める場合とされている。
- 3) ただし、連続変数を離散化することは恣意性が入り込む可能性がある。さらに、すべてのグループに処置 $W_i=1$ とコントロール $W_i=0$ が存在しないとこの計算はできない。そこで、傾向スコア法を用いられることがしばしばある。
- 4) CARTは、Breiman et al. (1984) で提案された手法

である。

- 5) 今回の分析対象となる産業は以下のとおりである。
 1 : 農林業, 2 : 建設業, 3 : 製造業, 4 : 情報通信業・運輸業, 5 : 卸売り・小売業, 6 : 金融・保険・不動産業, 7 : サービス業, 8 : その他 (公務を含む)。
 また, 分析対象の期間に4回産業の改定があったが, すべての期間で産業分類が統一された分類にした。
 6) 労働力調査は日本を11の地域に分類しているが, 本研究では九州と沖縄をまとめ1つの地域にした。

参考文献

- Adachi, Y., Ogawa, H., & Tsubuku, M. (2022), Measuring productivity dynamics in Japan: a quantile approach. *Empirical Economics*, 63(1), pp.201-242
- Athey, S., and Imbens G. (2016), Recursive partitioning for heterogeneous causal effects, *Proceedings of the National Academy of Sciences*, 113(27), pp.7753-7360.
- Arribas-Bel, D., Garcia-Lopez, M.A., & Viladecans-Marsal, E. (2021), Building (s and) cities: Delineating urban areas with a machine learning algorithm, *Journal of Urban Economics*, 125, 103217.
- Brandenburg, M. (2016), Supply chain efficiency, value creation and the economic crisis: An empirical assessment of the European automotive industry 2002-2010. *International Journal of Production Economics*, 171, pp.321-335.
- Breiman, L., Friedman J. H., Olshen R. A., Stone C. J. (1984), *Classification and Regression Trees*, Routledge
- Bry, G. & Boschan, C. (1971), Cyclical analysis of time series: Selected procedures and computer programs, National Bureau of Economic Research. <https://www.nber.org/books-and-chapters/cyclical-analysis-time-series-selected-procedures-and-computer-programs>
- Canello, J. & Vidoli, F. (2020), Investigating space-time patterns of regional industrial resilience through a micro-level approach: An application to the Italian wine industry. *Journal of Regional Science*, 60(4), pp.653-676.
- Cassar, A., Healy, A., & Von Kessler, C. (2017), Trust, risk, and time preferences after a natural disaster: experimental evidence from Thailand. *World Development*, 94, pp.90-105.
- Cerqua, A., & Letta, M. (2022), Local inequalities of the COVID-19 crisis. *Regional Science and Urban Economics*, 92, 103752.
- Fingleton, B., Garretsen, H., & Martin, R. (2012), Recessionary shocks and regional employment: evidence on the resilience of UK regions. *Journal of Regional Science*, 52, pp.109-133.
- Galdo, V., Li, Y., & Rama, M. (2021), Identifying urban areas by combining human judgment and machine learning: An application to India. *Journal of Urban Economics*, 125, 103229.
- Hanaoka, C., Shigeoka, H., & Watanabe, Y. (2018), Do risk preferences change? evidence from the Great East Japan earthquake. *American Economic Journal: Applied Economics*, 10, pp.298-330.
- Holl, A. (2018), Local employment growth patterns and the Great Recession: The case of Spain. *Journal of Regional Science*, 58, pp.837-863.
- Hoffmann, A. O., Post, T., & Pennings, J. M. (2013), Individual investor perceptions and behavior during the nancial crisis. *Journal of Banking & Finance*, 37, pp.60-74.
- Kawakubo, T. & Suzuki, T. (2022), Supply chain dynamics and resilience of the economy during a crisis, RIETI Discussion Paper 22-E-070. <https://www.rieti.go.jp/jp/publications/summary/22080003.html>.
- Martin, R., & Sunley, P. (2015), On the notion of regional economic resilience: conceptualization and explanation. *Journal of Economic Geography*, 15, pp.1-42.
- Patias, N., Rowe, F., & Arribas-Bel, D. (2022), Trajectories of neighbourhood inequality in Britain: Unpacking inter-regional socioeconomic imbalances, 1971-2011. *The Geographical Journal*, 188, pp.150-165.
- Reades, J., De Souza, J., & Hubbard, P. (2019), Understanding urban gentrification through machine learning. *Urban Studies*, 56, pp.922-942.
- Wager, S., & Athey, S. (2018), Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), pp.1228-1242.