The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Propensity Score 及び Causal Tree を用いた LGBT 施策と収益の因果推定

Causal Inference between LGBT Measures and Profit using Propensity Score and Causal Tree

田原 万悠子[*1]　長谷川 泰大[*1]　古庄 涼花[*1]　村井 千恵[*1]　森 峻人[*2]　星野 崇宏[*1]
Mayuko Tahara　Yasuhiro Hasegawa　Ryoka Furusyo　Chie Murai　Takato Mori　Takahiro Hoshino

[*1] 慶應義塾大学 経済学部　[*2] 慶應義塾大学 文学部
Faculty of Economics, Keio University　Faculty of Letters, Keio University

This paper aims to examine the effect of LGBT measures on profit. Previous researches are limited to the effects of general diversity and company-internal LGBT measures, and no mention have been made toward the effect of general LGBT services provided by companies. On the other hand, various studies have revealed that demand indeed exists for LGBT-friendly goods and services from the community. Therefore, this paper carries out an empirical analysis under the hypothesis that LGBT measures do have effect in raising firm revenue.
Rental property data of real estate information site has been used for demonstration. Given LGBT-friendly tags to some of its rental properties, the paper compares the effect on the number of browsing and inquiries between the dwellings with tags and those without. However, because assigning the tag is not conducted randomly, endogeneity concerning the other factors of the dwelling, such as its age and rent fee, exists. Thus, the effect of the tag cannot be obtained by simple comparison between the two groups. Therefore, the paper analyzes the effect in consideration of endogeneity, using IPW estimator and doubly robust estimator derived from propensity score. Specifically, by consolidating multiple covariates into one variable known as the propensity score, the paper successfully estimates the pure effect of the tag. In addition, a machine learning technique called Causal Tree was used to infer the heterogeneity. Our hypothesis is supported and revealed that LGBT-friendly tag gives a positive effect on both the number of views and number of inquiries.

1. はじめに

近年、LGBT に対する社会的な関心が高まっていて、電通の調査によれば、当事者の割合は 9%とも推計されている。その一方、LGBT に対する差別・スティグマは依然として存在している。その影響として、LGBT 当事者のいじめ体験率や自殺未遂のリスクが格段に高いことが示されていて、ホモフォビアの声も強い[1]。

LGBT に対するスティグマ・差別の是正が進まない理由の 1 つとして、民間企業が積極的に LGBT フレンドリーな施策を進められていないことがあげられる。LGBT フレンドリーな施策が進められていない理由の一つとして、施策が専ら人権的なインセンティブのみで捉えられ、経済的なインセンティブが弱いと捉えられていることが考えられる。そこで、本研究では LGBT 施策と経済的な利益の関係性について調査することにした。

従来の、ダイバーシティが企業に与える経済的な影響に関する研究は、労働経済学的な観点から、ダイバーシティ全般と経済的利益に関する研究が多く[2]、LGBT に特化した経営学的な施策の純粋な効果に関する研究は希少である。この理由として、①アウティングの懸念から、対面でのヒアリングに正しく回答しない、②企業の取り組みが始まったのがここ数年であるため詳細な統計が取られていないなどが挙げられる。

また、同様の理由から、企業内部のダイバーシティの度合いと経済的利益の関係性を調べた研究が多く[3]、企業によるダイバーシティを促すような財と経済的利益の関係性を調べた研究は希少であった。

以上の問題意識より、本研究では、「LGBT 施策の収益率は正である」との仮説を立て、研究によるアウティングの可能性を排除しつつ、LGBT フレンドリーを促す財と経済的利益の関係性を調べることにした。

具体的には、LGBT フレンドリーを促す財・サービスに関する既存のデータを使用し、LGBT フレンドリーな財・サービスかどうかには内生性があることを踏まえて、その財・サービスが「LGBT フレンドリーを促している」という理由によって、どの程度収益に影響が及んでいるのかについて統計的な因果推定をした。

2. 解析手法及び解析結果

本研究では、賃貸物件情報サイトに関するオンラインデータを使用した。当賃貸物件情報サイトでは、住戸が LGBT フレンドリーであることを示す、LGBT フレンドリータグが設けられている。そして、データ上に、そのタグの有無に関するカラムが含まれていたので、今回の解析に最適だと考えられた。

データの期間は 2018 年 5 月 1 日〜7 月 31 日で、地域は東京都内 52 市区町村に絞った。

解析を行う前に、住戸ごとに日別で集計されていた元データを加工する必要があった。加工手順は以下の図 2 の通りである。

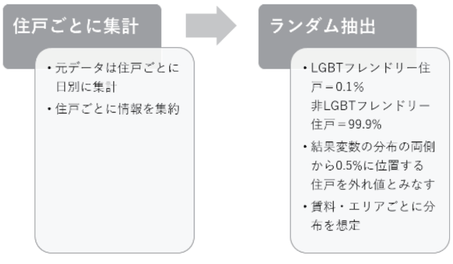
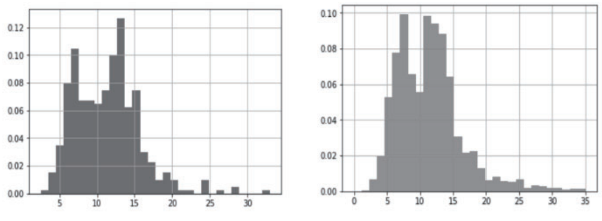


図 2. データの加工手順

住戸の抽出に当たっては、LGBT 住戸を賃料 5 分位で分け、エリアごとに分布を想定し、その分布に合わせ 非 LGBT 住戸を抽出した。その賃料分布は図 3 の通りである。

連絡先：森 峻人, taka.m.617-11-038@keio.jp

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019



[1] LGBT 住戸　　[2] 非 LGBT 住戸

図 3. 住戸の賃料分布

本研究では、データ上の制約及び広告業界では閲覧数やお問い合わせ数が収益と密接に関係しているという理由から、収益の代理変数を閲覧数及びお問い合わせ数に設定して解析を進めた。

そして、ルービンの因果モデル[4][5][6]を利用し、平均処置効果 ATE (average treatment effect)を推定した。更に、住戸全体ではなく、LGBT 住戸群 ($Z=1$) のみにおける平均処置効果 ATT (average treatment effect on the treated) も推定した(図 4 参照)。

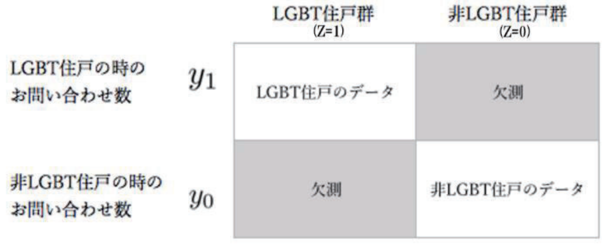


図 4. 因果推論における欠測データ

2.1 傾向スコアの算出

住戸 i に対して LGBT フレンドリータグ($Z=1$)に割り当てられる予測確率を、共変量 X に回帰するロジスティックモデルで、傾向スコアの推定値を算出した。選択した共変量は、表 1 に表記のある 9 つである[1]。算出に当たっての式は以下の通りである(x_i= 共変量 、α_i= 共変量の偏回帰係数、$\hat{e}_i$ = 傾向スコア)。

$$\hat{e_i} = \frac{1}{1+\exp(-\hat{\boldsymbol{a}}_i \boldsymbol{x}_i)}$$

偏回帰係数とその有意性を確認するための Wald 検定の検定統計量と p 値については、それぞれ表 1 の通りになった。

表 1. 共変量と検定統計量

共変量	偏回帰係数	Wald 検定統計量	p 値
面積	1.957e-02	3.257	0.001126
建物種別	2.148e-01	1.709	0.087444
築年数	1.355e-02	3.131	0.001745
賃料	-5.874e-03	-0.240	0.810625
部屋数	-4.558e-01	-3.617	0.000298
構造種別	-3.350e-02	-1.247	0.212345
礼金	-4.140e-08	-0.051	0.959401
敷金	-1.219e-06	-1.532	0.125582
相場	-2.980e-02	-1.713	0.086789

2.2 IPW 推定量

本研究では、最初に IPW 推定量を用いて因果推定を行った。y_1及び y_0の周辺期待値の差が平均処置効果の推定値なので、IPW 推定値は、以下の式で表される(Z_i = 住戸の割付、W_i =傾向スコアの逆数の重みの和)。

$$\hat{\tau} * = \frac{1}{W_1}\sum_{i-1}^{N}\frac{z_i y_i}{e_i} - \frac{1}{W_0}\sum_{i-1}^{N}\frac{(1-z_i)y_i}{1-e_i}$$

計算の結果、IPW 推定量を用いた因果推定の結果は ATT で平均及び標準偏差に関して、表 2 の通りになった。

表 2. ATT の IPW 推定量と標準誤差
(標準化した値であり実数値でない。以下の図表式も全て同様)

	LGBT 住戸群	非 LGBT 住戸群	差
PV 平均	0.04891	-0.02813	0.07704
PV 標準誤差	0.01701	0.01701	0.02406
CV 平均	0.06302	-0.01809	0.08110
CV 標準誤差	0.01887	0.01887	0.02669

2.3 分位点での因果効果の比較

次に、賃料を三分位に分け、各分位での因果効果について比較した。この時、各分位での住戸の割合は、全体の割合に対して 3 分の1となるようにした。分位での比較を行うことで、特定の集団に対する LGBT フレンドリータグの効果を検証した。つまり、LGBT 住戸がどの賃料価格帯に所属していることで閲覧数・お問い合わせ数により影響があるかを分析した。

結果は表 3 の通りで、閲覧数の比較を行うと、賃料三分位目で最も値が大きく、賃料が高い住戸にタグが付与されている方で閲覧数が上昇している。一方、お問い合わせ数では、賃料二分位目で最も値が大きく、賃料が低い住戸の方がお問い合わせされやすい傾向にあった。

表 3. 賃料分位ごとの ATT の IPW 推定量

	閲覧数	お問い合わせ数
賃料一分位目	-0.04176	0.04961
賃料二分位目	0.11394	-0.03144
賃料三分位目	0.01753	0.02112

2.4 賃料による異質性

次に、賃料を独立変数と置いて、閲覧数・お問い合わせ数を説明するモデルを構築し、そこから LGBT フレンドリータグの有無による因果効果を推定した。

まず、LGBT 住戸・非 LGBT 住戸それぞれを賃料で結果変数(閲覧数・お問い合わせ数)について単回帰分析を行った。LGBT 住戸(Z=1)と非 LGBT 住戸(Z=0)は、それぞれ

$$\begin{cases} Y_1 = \beta_{10} + \beta_{11}W + \epsilon_1 \\ Y_0 = \beta_{00} + \beta_{01}W + \epsilon_0 \end{cases}$$

の式で表現される。IPW 推定量で重み付けすることで β_0と β_1 それぞれの推定量を求めた。そして、β_0 と β_1 の差を求めることにより、閲覧数・お問い合わせ数の処置群での平均介入効果(ATT) を検証した。結果は表 4・5、図 5 の通りである。

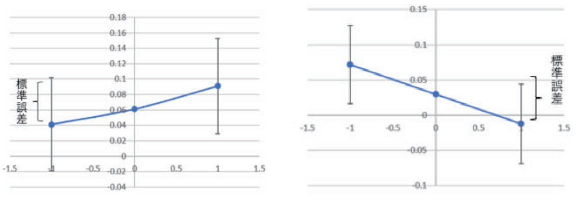
[1] C 統計量は 0.59 であった。

表 4. 閲覧数の結果

	因果効果	標準誤差
w=-1	0.04108	0.06044
w=0	0.06099	0.00129
w=1	0.09110	0.06178

表 5. お問い合わせ数の結果

	因果効果	標準誤差
w=-1	0.07186	0.05538
w=0	0.02982	0.00129
w=1	-0.01222	0.05675



[1] 閲覧数の群間差　[2] お問い合わせ数の群間差

図 5. 閲覧数・お問い合わせ数の群間差結果

2.5 Causal Tree による推定

なお、因果効果の異質性の推定には、機械学習的手法の利用も可能である。Wager と Athey が提案した、Breiman のランダムフォレストを拡張した推定法"propensity tree"[7]における決定木のコスト関数は、施策対象の割り当ての予測誤差を最小化するよう設計されている。従って、本研究のように施策の割り当てを無作為に行えない場面での使用が適している。この決定木を用いて、因果効果の異質性の推定を行った。ここには結果の概要のみを示し、詳細な決定木及び結果の標準化値等については当日発表する。

閲覧数の決定木の構造によれば、閲覧数における因果効果の大小に影響を与える変数は賃料と築年数ということがわかった。最初の分岐点は、賃料の価格帯である。低価格帯ならば、築年数が短く賃料が安い物件ほどプラスの効果となる。しかし、築年数が長い、あるいは低価格帯の中でも賃料が高い物件の場合、マイナス効果となる。一方、中・高価格帯ならば、極端に築年数が長い物件のみがマイナス効果である。賃料と築年数が共に中間地帯に位置する物件ほど、効果は高くなる。

次に、お問い合わせ数の決定木の構造によれば、お問い合わせ数における因果効果の大小に影響を与える変数は礼金と部屋数ということがわかった。また、礼金が中価格帯に属している、あるいは礼金が低く部屋数が少ない物件でプラスの効果があると分かった。礼金は低いが部屋数が多い物件は、これらの物件内の礼金価格帯で 3 つに分類され、中価格帯は効果なし、低・中価格帯はマイナス効果であることが分かった。対して、礼金が高い物件はすべてマイナス効果であった。

以上のように、決定木の利用によって、施策効果が大きい・小さい物件の特徴が推定された。

2.6 二重にロバストな推定法

結果変数の周辺期待値や因果効果を正しく推定できているかを検証するために、二重にロバストな推定法を用いて、分析した。y_1 及び y_0 における二重にロバストな推定量は、以下の式で導出される。

$$\hat{E}^{DR}(y_1) = \frac{1}{N}\sum_{i}^{N}\left[\frac{z_i y_{i1}}{e(\boldsymbol{x_i},\hat{\alpha})} + \left(1 - \frac{z_i}{e(\boldsymbol{x_i},\hat{\alpha})}\right) g(x_i,\hat{\boldsymbol{\beta}}_1)\right]$$

$$= \frac{1}{N}\sum_{i}^{N}\left[y_{i1} + \frac{z_i - e(\boldsymbol{x_i},\hat{\alpha})}{e(\boldsymbol{x_i},\hat{\alpha})}\left(y_{i1} - g(x_i,\hat{\boldsymbol{\beta}}_1)\right)\right]$$

$$\hat{E}^{DR}(y_0) = \frac{1}{N}\sum_{i}^{N}\left[y_{i1} + \frac{z_i - e(\boldsymbol{x_i},\hat{\alpha})}{e(\boldsymbol{x_i},\hat{\alpha})}\left(y_{i0} - g(x_i,\hat{\boldsymbol{\beta}}_1)\right)\right]$$

結果として、閲覧数については、

$$\hat{\Delta}_{DR} = 0.08683 - (-0.00202) = 0.08885$$

お問い合わせ数については、

$$\hat{\Delta}_{DR} = 0.02698 - (-0.00106) = 0.02804$$

となり、二重にロバストな推定量でも、閲覧数・お問い合わせ数で、収益に対して正の影響が見られた。

3. 結論

LGBT フレンドリータグを付けることによって、住戸に対する閲覧数、お問い合わせ数を増大させる効果をもたらすことから、LGBT タグと企業の収益性の間に正の関係があることがわかった。

賃料を3分位に分け各分位で因果効果を改めて測定すると、閲覧数で LGBT フレンドリータグの効果が最も出たのは3分位目の住戸群であり、お問い合わせ数で LGBT フレンドリー タグの効果が最も出たのは1分位目の住戸群であった。このことから上記の価格帯で LGBT フレンドリータグをつけることで閲覧数、お問い合わせ数それぞれに対して相対的に大きな正の影響をもたらすと言及できる。

賃料の異質性を考慮し、賃料を独立変数とおいたモデルで因果効果を測定すると、閲覧数は賃料が高い住戸に対しての LGBT フレンドリータグの付与がより効果的であることが分かる。一方、お問い合わせ数の効果は賃料が上がると逓減するため、賃料が低い住戸に比べ高い住戸はお問い合わせまで相対的に繋がりにくいことがわかった。

また、機械学習的手法を用いて因果効果の異質性を推定したところ、決定木の構造から、賃料だけでなく、築年数・礼金・部屋数といった変数も因果効果の大小に影響を与えることが分かった。

二重にロバストな推定量のモデルで因果効果を再度測定したところ、結果変数の周辺期待値や因果効果を正しく推定でき、より頑健な LGBT フレンドリータグによる因果効果の結果を得ることができた。

4. 参考文献

[1] 日高庸晴. (2017). 学校に求められる LGBT の児童生徒への理解と支援 (特集 多様性). 日本教育, (468), 6-9.

[2] 山本勲. (2014). 上場企業における女性活用状況と企業業績との関係-企業パネルデータを用いた検証. RIETI Discussion Paper Series 14-J-016.

[3] Ellison, S. F., & Mullin, W. P. (2014). Diversity, social goods provision, and performance in the firm. Journal of Economics & Management Strategy, 23(2), 465-481.

[4] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of educational Psychology, 66(5), 688.

[5] Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

[6] 星野崇宏. (2009). 調査観察データの統計科学-因果推論選択バイアス データ融合. 確率と情報の科学.

[7] Wager, S., & Athey, S. (2017). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, (just-accepted).