



京都大学大学院経済学研究科
ディスカッションペーパーシリーズ

因果性と異質性の経済学② : Causal Forest

石原 卓典 京都大学大学院経済学研究科・研究員
依田 高典 京都大学大学院経済学研究科・教授

No. J-20-004

2020 年 8 月

〒606-8501
京都市左京区吉田本町
京都大学大学院経済学研究科

因果性と異質性の経済学②: Causal Forest

京都大学大学院経済学研究科 研究員 石原 卓典¹
京都大学大学院経済学研究科 教授 依田 高典²

1. はじめに

近年、複数の経済学研究が、機械学習の手法を使って、複数の属性情報で条件づけた介入効果を予測し、その予測された介入効果を用いた分析を行っている。この属性情報で条件づけた介入効果のことを条件付平均介入効果 (Conditional Average Treatment Effect; CATE) や異質介入効果 (Heterogeneous Treatment Effect; HTE) と呼ぶ。

連載第 1 回目では限界介入効果 (Marginal Treatment Effect; MTE) を紹介した。MTE では観察不可能な変数を含めて介入効果の異質性を推定する。今回紹介する因果推論の機械学習では多数の観察可能な属性変数を使ってその異質性を推定するものである。

機械学習の手法は、元々は複数の属性情報を用いて結果を予測するために活用されてきた。しかし、近年は、結果の予測だけにとどまらず因果推論に機械学習を活用する研究が盛んに行われている (Athey and Imbens 2019, Mullainathan and Spiess 2017, 依田 2019)。中でも、決定木やランダムフォレスト等の木に基づく手法で著しい発展がある。Athey and Imbens (2016) は木に基づく手法を因果推論に導入し、さらに Wager and Athey (2018) がコウザルフォレストとしてその手法を発展させ、その手法によって予測される CATE が、幾つかの条件の下で一致性や漸近正規性を持つことを明らかにした。

また省エネ研究・貧困及び労働研究・開発研究・租税研究など、さまざまな領域の経済学研究がコウザルフォレストを使用して CATE を予測し、その分布の特徴やどのような属性がその異質性を説明するかといったことを検証している (Miller 2020, Carter, Tjernström, and Toledo 2019; De Neve et al. 2019; Farbmacher, Kögel and Spindler 2019; Hoffman and Mast 2019; O'Neill and Weeks 2018)。さらに、単一の介入効果を評価する省エネ研究や貧困及び労働研究が、CATE を使用した例証を通して、「誰に介入すべきか」という観点からターゲティングを実施することの有用性を提示している (Bertrand et al. 2017; Davis and Heller 2017; Knittel and Stolper 2019)。

本稿の構成は以下の通りである。まず第 2 節で統計的学習の基礎として、予測モデルの精度の評価について解説する。第 3 節では木に基づく予測モデルである決定木やランダムフォレストについて紹介する。また、決定木やランダムフォレストを因果推論に応用した因果木や

¹ ishihara.takunori.5n@kyoto-u.ac.jp

² ida@econ.kyoto-u.ac.jp

コウザルフォレストについて紹介する. このコウザルフォレストを応用した最近の実証研究について第4節で紹介を行う. 最後に第5節で本稿のまとめを行う.

2. 統計的学習の基礎

本題に移る前に, まず機械学習の考え方の基本であるモデルの精度の評価についてみていく. はじめに述べたように, 機械学習では属性 X を用いて結果 Y を予測する関数 $f(X)$ を見つけることが目標である. 機械学習では, ある予測モデル \hat{f} の性能の評価を行うために, そのモデルの予測値 $\hat{f}(x)$ が実際の観測データ Y_i とどれだけ近いかということを考える. その予測値と実際の観測データとの乖離度合いを表す指標として平均二乗誤差 (Mean Squared Error; MSE) が用いられる.

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{f}(X_i))^2$$

これはデータ $\{Y_i, X_i\}$ ($i = 1, \dots, N$)について, それぞれの予測値と実際の観測データの乖離度合いの平均を表したものである. この MSE が小さければモデルと現実のデータとのあてはまりがよいことを表す.

MSE を用いると, モデル f を選ぶ基準として MSE の期待値

$$\begin{aligned} EMSE(f) &= E \left[\frac{1}{N} \sum_{i=1}^N (Y_i - f(X_i))^2 \right] \\ &= E_X E_{Y|X} \left[\left(\frac{1}{N} \sum_{i=1}^N (Y_i - f(X_i))^2 \right) \middle| X_i = x \right] \end{aligned}$$

を用いることになる. 従って上記の EMSE を最小化するためには, それぞれの x について最小化を行えばよいため,

$$f(x) = \arg \min_c E_{Y|X} [(Y_i - c)^2 | X_i = x]$$

として求められる. 上記を解くと,

$$f(x) = E[Y_i | X_i = x]$$

として条件付期待値が得られる. 従って, MSE を基準とすると, $X_i = x$ における Y_i の最良の予測は条件付期待値ということになる.

また実際に予測したモデルの評価を行う場合, モデルの構築に使用したデータについてのあてはまりがよいことはそれほど重要ではない. むしろ重要となるのは, 未知のデータについてのモデルのあてはまりの良さである. この未知のデータに対するあてはまりの良さを測るために, 手元のデータを訓練データとテストデータに分割する. そして, 訓練データを使ってモデルの構築を行い, テストデータによりモデルの性能評価を行う. このテストデータでの MSE が最小となるようにモデルの調整を行い, 最終的な予測モデルを得る. 一般的に複雑なモデルを選択すると, 訓練データの MSE は減少する傾向がある. しかし, モデルをより複雑にしたとしても, テスト MSE が同様に減少するとは限らない. あるモデルの訓練 MSE が小さいにもかかわらず, テスト MSE が大きい場合, データに対し過学習であるという. これはモデルが訓練データに対して適合しすぎているために, 未知のモデル f の真の性質ではなく, たまたま得られた訓練データの性質をとらえているに過ぎない.

従って, 予測モデルを構築する際にはテストデータでの MSE が重要となってくる. それではこのテストデータの MSE はどういう構造をしているのだろうか. 言い換えると, どういった要因がこのテストデータの MSE の大きさを決めるのだろうか. 以下では, テストデータでの MSE の期待値を分解することでこのことを見ていく.

母集団における y と x の真の関係が $Y = f(X) + \epsilon$ であるとしよう. このとき, あるテスト点 x_0 が与えられた時, テスト MSE の期待値は以下のように分解することができる.

$$\mathbb{E}[(y_0 - \hat{f}(x_0))^2] = \text{Var}[\hat{f}(x_0)] + [\text{Bias}(\hat{f}(x_0))]^2 + \text{Var}[\epsilon]$$

ここで, $\text{Var}[\hat{f}(x_0)]$ はモデルに基づく予測値 $\hat{f}(x_0)$ の分散を表す. また $\text{Bias}(\hat{f}(x_0)) = f(x_0) - \mathbb{E}[\hat{f}(x_0)]$ は予測モデル \hat{f} が真のモデル f と異なることから生じるバイアスを表している. 最後に $\text{Var}[\epsilon]$ は真の関数形に含まれる誤差項 ϵ の分散である. 分散 $\text{Var}[\hat{f}(x_0)]$ は予測モデルを構築する際に異なる訓練データを使った時にどの程度 \hat{f} が変化するかを表したものと考えることができる. 一方, バイアス $\text{Bias}(\hat{f}(x_0))$ は複雑な実際の現象 f を単純なモデル \hat{f} で近似したことによる誤差として考えることができる.

誤差項の分散 $\text{Var}[\epsilon]$ は調整できないため, テスト MSE の期待値を小さくするためには, モデルの分散 $\text{Var}[\hat{f}(x_0)]$ とバイアス $\text{Bias}(\hat{f}(x_0))$ を同時に小さくするようなモデルを選択する必要がある. 一般的な傾向として, 複雑なモデルを選択すると分散は増加し, バイアスは減少する. この増減のどちらが大きいかによって, テスト MSE が増加するか減少するかが決まるのである.

3. コウザルフォレストの概要

3.1 木に基づく手法

Wager and Athey (2018) のコウザルフオレスト (causal forest; 以下, CF) は機械学習の手法の一つであるランダムフォレスト (random forest; 以下, RF) を因果推論の文脈に持ち込んだノンパラメトリックな推定手法である。そのため, CF を紹介するにあたり, まずはその基礎となる RF や決定木について補足しておく必要があるだろう。そのため, 以下ではまず決定木について, 次にランダムフォレストについて説明を行う。これらについてのより詳細な議論は, James et al. (2018) や Hastie et al. (2018) 等の機械学習についての入門書などを参照してほしい。

3.1.1 決定木

決定木では木の頂点から始めて, 複数回データを 2 分割していくことにより, 予測モデルを構築する。分類を目的とする場合には分類木を用い, 数値の予測など回帰を行うことが目的である場合には回帰木と呼ばれるものを用いる。こうした決定木を作成するアルゴリズムとしては, CART (Classification and Regression Tree: Breiman et al. 1984) などがよく用いられている。以下で決定木と呼ぶ場合は, CART アルゴリズムに基づく回帰木を指す。

今 $i = 1, \dots, N$ のデータの組 $\{Y_i, X_i\} \in \mathcal{R} \times \{0, 1\}^p$ が与えられた下で, 決定木を構築する場合を考えてみよう。実際には訓練データとテストデータに分けたうえでモデルの調整を行うが, ここでは簡単化のためにデータの分割は行わない。ここで Y_i は応答 (あるいは結果変数) を, X_i は特徴量 (あるいは共変量) を表すものとする。決定木の目標は, $X_i = x$ で条件つけた応答 Y_i の条件付期待値を予測するモデルを構築することである。応答 Y_i の条件付期待値 $\mu(x)$ を以下のようにあらわす。

$$\mu(x) = \mathbb{E}[Y_i | X_i = x]$$

< 図 1: 決定木の例 >

上記の $\mu(x)$ を予測するモデルを構築するために, 決定木では図 1 上図のような形でデータを分割することによりモデルの構築を行う。図 1 上図の分割が始まっていない頂点のことを親ノードと呼び, 一番下の R_1 や R_2 , R_3 のことをリーフまたは最終ノードと呼ぶ。図 1 のようにある値 c_1 , c_2 を決め, それに従いデータを分割していき, リーフ内のデータ数が特定の数より少なくなった時に分割を停止する。最終的に各分割領域での予測値 $\hat{\mu}_j$ は, その領域にあるデータの応答の平均として表される。 c_1 , c_2 のような閾値を定める基準として, 平均二乗誤差が用いられることが多い。MSE は以下のように定義される。

$$MSE = \frac{1}{N} \sum_{j=1}^J \sum_{i \in R_j} (Y_i - \hat{\mu}_j)^2$$

ここで $j \in \{1, \dots, J\}$ は分割領域の数を示している。また、 $\hat{\mu}_j$ はある分割領域 j での応答の平均を示している。

このように決定木では、**MSE** が最も小さくなる値を閾値として選択することにより、分割点を定める。実際には実行可能なすべての分割領域について、上記の **MSE** を計算するのは計算量が膨大になるため不可能である。そのため、再帰的な 2 分割法により **MSE** の計算を行う。この方法は上記のようにすべての分割を見据えた上で分割点を決めるのではなく、特定の分割点のみに着目して、その分割点ごとに最適な閾値を定め分割を行っていくものである。より具体的には、再帰的な 2 分割法では予測変数 X_j と分割点 s を選び、予測変数空間を $\{X|X_j < s\}$ と $\{X|X_j > s\}$ の領域に分ける際に、可能な限り 2 つの領域の **MSE** の合計が小さくなるようにする。例えばまず親ノードにおいてすべての j, s を考え、

$$R_1(j, s) = \{X|X < s\}, \quad R_2(j, s) = \{X|X > s\}$$

を定義し、以下を最小化するような j と s を探す。以下の $\#(\cdot)$ はデータ数を示している。

$$\frac{1}{\#(R_1)} \sum_{i: x_i \in R_1(j, s)} (Y_i - \hat{Y}_{R_1})^2 + \frac{1}{\#(R_2)} \sum_{i: x_i \in R_2(j, s)} (Y_i - \hat{Y}_{R_2})^2$$

その後、各ノードに対してそれ以上分割できなくなるまで同様の方法で分割点を探していく。この手順を図示したのが図 1 下図である。図 1 下図ではまず親ノードにおいて $X_1 = c_1$ を閾値として定め、分割を行っている。次に $X_1 > c_1$ となる領域について $X_2 = c_2$ となる点を閾値として定め再度分割を行っている。

決定木は木を深くすることによって、データ内の複雑な構造に対処することができる。しかし、決定木による回帰や分類の予測結果は分散が大きくなることも知られている。

そのため、この決定木の予測の分散を抑えるために、複数の決定木の平均を取ることによってこの分散を低下させようというアイデアがランダムフォレストである。

3.1.2 ランダムフォレスト

RF は決定木にアンサンブル法を適用したものである。アンサンブル法とは簡単に言えば、複数のモデルを合わせて一つのモデルとすることである。つまり、**RF** では決定木を複数作成し、それらを統合することによって一つのモデルとして扱うのである。

この RF のアルゴリズムを表 1 にまとめている。RF のアルゴリズムでは、 B 本の木に対して、それぞれサイズ N の訓練データからブートストラップサンプリングによって、ブートストラップ標本 Z^* を取り出す³。取り出した標本 Z^* について、 p 個の特徴量の中からランダムに m 個の特徴量を取り出す。そして取り出した m 個の変数を用いて標本 Z^* に対し、MSEを小さくするように再帰的な2分割を行う。さらにこの分割を最小ノードサイズ n_{min} に到達するまで繰り返し行っていく。最終的に得られた B 本の決定木の予測結果を統合することにより、最終的な予測を行う。回帰を行う場合、この統合は各決定木の平均を取ることで行われる。

<表 1: ランダムフォレストのアルゴリズム>

また RF では、フォレストを育てる際に使用した変数の重要度表す変数重要度という指標がある。これはその特徴量での分割が応答への回帰にどれくらい寄与しているかを示すものである。

この RF アルゴリズムのアイデアを因果推論に持ち込んだのが、次節で紹介するコウザルフォレストである。CF では RF 同様、複数の因果木を作成し、それらの結果を統合することにより、条件付平均介入効果の予測値を行う。

3.2 木に基づく因果推論

CF では通常の RF と同様に複数の木からモデルを構築する。RF の目標は応答 Y_i と特徴量 X_i を用いて予測を行うことであった。一方、CF では応答 Y_i と特徴量 X_i に加え、介入の割当変数 W_i を用いて因果推論を行うことにその目的がある。

通常、観察データを用いて因果推論を行う場合、以下の潜在結果の条件付独立の仮定 (unconfoundedness) が必要である。

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

ここで $Y_i^{(0)}$ と $Y_i^{(1)}$ はそれぞれ個人 i が介入を受けなかったときの潜在結果と受けた時の潜在結果を表している。

さらに特徴量 X_i の任意の実現値 x について、介入群 $W_i = 1$ に割り当てられる人と統制群

³ RF では一本の決定木を育てるときに使用しなかった標本(テストデータ)のみを使った計算を行うことができる(抜取標本, Out-Of-Bag; OOB). この OOB 標本による推定値は N 分割交差確認で得られるものとほぼ同一である (Hastie et al. 2018). そのため、CF や RF ではフォレストを育てるのと並行して交差確認を行うことができる。後述する CF の R パッケージ `grf` ではこの OOB を利用して平均介入効果や CATE の予測値を計算することもできる。

$W_i = 0$ に割り当てられる人のどちらも含まれている必要がある。そのため、以下の重なり仮定 (overlap) を置く。

$$0 < Pr(W_i = 1 | X_i = x) < 1$$

これらの仮定の下で、以下の条件付平均介入効果 (conditional average treatment effect; CATE) を推定していく。

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

CF でも通常の因果推論の場合と同様に、上記の 2 つの仮定の下で、 $\tau(x)$ の推定を行うことが目的となる。

基本的なアルゴリズムは CART による決定木や RF と同様であるが、応答の予測値 $\hat{\mu}(x)$ ではなく、CATE の推定値 $\hat{\tau}(x)$ を求めることに目的が置かれているため、修正を行う必要がある。

また、通常の決定木や RF では手元にある標本を、モデルの構築を行う訓練データとモデルの評価を行うテストデータに分割する。一方、Athey and Imbens (2016)の因果木や Wager and Athey (2018)の CF の特徴の一つは、honesty という方法により訓練データをさらに 2 分割することがあげられる。この honesty を行うことにより、予測した $\hat{\mu}(x)$ や $\hat{\tau}(x)$ が一致性や漸近正規性といった性質を満たすことが知られている。Honesty では、一方の訓練データをモデルの分割点の選択に用い、もう一方のデータを予測値の推定のために用いる。

CF を構成する木は因果木 (causal tree) と呼ばれる。今、ある標本 \mathcal{S} について因果木を推定することを考えよう。ここで \mathcal{S}^{te} 、 \mathcal{S}^{tr} はそれぞれテストデータ、訓練データを表し、 \mathcal{S}^{est} は介入効果の推定のために用いられるデータを表している⁴。因果木 Π のリーフ $\ell(x; \Pi)$ での CATE は以下のようにあらわされる。

$$\tau(x; \Pi) \equiv \mathbb{E}[Y_i(1) - Y_i(0) | X_i \in \ell(x; \Pi)] = \mu(1, x; \Pi) - \mu(0, x; \Pi)$$

ここで、 $\mu(w, x; \Pi)$ は任意の $w \in \{0, 1\}$ について、 $\mu(w, x; \Pi) \equiv \mathbb{E}[Y_i(w) | X_i \in \ell(x; \Pi)]$ である。 $\tau(x; \Pi)$ の推定量は以下のようにあらわされる。

⁴ 後述する R パッケージ `grf` の `causal_forest` において、この分割はデフォルトの場合 $\mathcal{S}^{te} : \mathcal{S}^{tr} : \mathcal{S}^{est} = 2 : 1 : 1$ である。

$$\hat{\tau}(x; \Pi) = \frac{1}{|\{i: W_i = 1, X_i \in \ell(x; \Pi)\}|} \sum_{\{i: W_i=1, X_i \in \ell(x; \Pi)\}} Y_i - \frac{1}{|\{i: W_i = 0, X_i \in \ell(x; \Pi)\}|} \sum_{\{i: W_i=0, X_i \in \ell(x; \Pi)\}} Y_i$$

つまり因果木 Π のリーフ $\ell(x; \Pi)$ における CATE は、リーフ内の介入群の応答の平均と統制群の平均の差として求められる。こうしたリーフ内での CATE を推定するために、因果木を育てていく必要がある。

CART による決定木の場合と同様に、因果木についての MSE が最も小さくなるような分割点を探し、分割を行っていく。ここで各個人についての真の介入効果を τ_i とすると、介入効果についての MSE は以下のようにして表現される。

$$MSE_{\tau}(S^{te}, S^{est}, \Pi) \equiv \frac{1}{\#(S^{te})} \sum_{i \in S^{te}} (\tau_i - \hat{\tau}(X_i; S^{est}, \Pi))^2$$

この MSE の S^{te} と S^{est} に対する期待値は以下のようにあらわされる。

$$EMSE_{\tau}(\Pi) \equiv \mathbb{E}[MSE_{\tau}(S^{te}, S^{est}, \Pi)]$$

通常の CART の場合とは異なり、実際に各対象についての τ_i を観察することはできない。そのため、上記の MSE をそのまま使用することはできない。従って、MSE の推定量を求めていく必要がある。介入効果の MSE の推定量は以下ようになる⁵。

$$-\widehat{MSE}_{\tau}(S^{tr}, S^{tr}, \Pi) \equiv \frac{1}{N^{tr}} \sum_{i \in S^{tr}} \hat{\tau}^2(X_i; S^{tr}, \Pi)$$

従って、因果木を育てるにあたっては、訓練データでの CATE の分散が最大になるような分割点で分割を行う。また、Athey and Imbens (2016)の因果木では honesty アプローチを採用しており、これによって MSE の期待値の推定量を以下の形で得ることができる。

$$-\widehat{EMSE}_{\tau}(S^{tr}, S^{est}, \Pi) \equiv \frac{1}{N^{tr}} \sum_{i \in S^{tr}} \hat{\tau}^2(X_i; S^{tr}, \Pi)$$

⁵ この部分に関する詳細な議論は Athey and Imbens (2016)を参照してほしい。

$$-\left(\frac{1}{N^{tr}} + \frac{1}{N^{est}}\right) \sum_{i \in S^{tr}} \left(\frac{S_{S^{tr}}^2(\ell)}{p} + \frac{S_{S^{control}}^2(\ell)}{1-p} \right)$$

ここで $S_{S^{tr}}^2(\ell)$ と $S_{S^{control}}^2(\ell)$ はそれぞれリーフ ℓ 内の介入群と統制群の応答 Y_i の分散を表している。また p は介入群に割り当てられている個人の割合である。第一項は MSE の推定量と同様、訓練データでの CATE のばらつきを表しており、そこから第二項でリーフ内の介入群と統制群の分散を罰則項として付け加えている。この EMSE 推定量では、介入効果に強い異質性をもたらす分割には報酬を与え、逆にリーフ内の推定量の分散を大きくするような分割に対して罰則を与えている。

こうして求めた因果木を複数作成し、その結果の統合を行ったものが CF である。今回の説明では CF で求められる CATE を、同一リーフ内の介入群と統制群の応答の差として説明したが、実際の CF アルゴリズムでは傾向スコアを予測し、それを使って介入効果を推定することも可能である (propensity tree)。さらに、Athey, Tibshirani and Wager (2019) では一般化ランダムフォレスト (generalized random forest) として、一般化モーメント法を使う形でランダムフォレストの一般化を行っている。また、同手法についての推定パッケージは R の grf パッケージとして公開されている (Tibshirani et al. 2020)。

4. 実証研究紹介

4.1 CF を用いた実証研究

この CF アルゴリズムを用いた実証研究について紹介していく。著者らは実証研究での CF の使い方について、大きく分けて①条件付平均介入効果を推定し、介入効果の分布の特徴や異質性について検証する目的で使用する形の研究と、②条件付平均介入効果を推定し、「誰に介入すべきか」という観点でターゲティングを行うために使用するタイプの研究があると考えている。

前者のタイプの実証研究としては、Miller (2020) や Carter, Tjernström, & Toledo (2019), De Neve et al. (2019), Farbmacher, Kögel & Spindler (2019), Hoffman & Mast (2019), O'Neill & Weeks (2018) などがあり、労働政策や納税行動などの文脈での介入効果の異質性を推定している。また、後者の実証研究としては Bertrand et al. (2017) や、Davis & Heller (2017), Knittel & Stolper (2019) があり、労働政策や節電行動の文脈で介入効果の異質性を求め、ターゲティングを行っている。表 2 ではこれらの研究でどのような情報が報告されているかということをもとめている。

<表 2:CF を用いた実証研究>

特に Knittel & Stolper (2019)では介入効果の異質性やターゲティングについて豊富なデータを用いて多面的に検討しており、CF を今後の実証研究に活用するために有用であると著者は考えている。そのため、以下ではまずこの Knittel & Stolper (2019)の紹介を行う。その他の研究については、次節でそれぞれ簡単に紹介する。

Knittel & Stolper (2019)では 90 万件以上の世帯を対象に、節電行動にホームエナジーレポート(HER)が与える効果の検証を行っている。この HER では、①需要家の電力消費量と②近隣世帯の消費量を比較した情報、③エネルギーの節約方法に関する情報を記載している。

彼らはまず一月当たりの電力消費量に対する HER の平均介入効果を求めている。それによると HER の平均介入効果は-0.085 kWh(-1%)であった。その後、介入効果の異質性を調べるために、過去の電力消費量や家の資産価値、敷地面積、所得など 13 個の特徴量を用いて CF により、異質介入効果の推定を行っている。その推定結果を示したものが図 2 である。

<図 2:介入効果の予測値の分布>

図 2 では、1 年目～3 年目の異質介入効果の分布を示している。ここから、1)同一年度内でも介入効果の異質性がみられること、2)その異質性は 2 年目以降も観察され、効果の散らばりの程度が拡大していることが見て取れる。具体的には、一年目の介入の結果、介入効果-10kWh 周辺で大きなピークがあり、効果ゼロ付近についてもピークがみられる。2 年目・3 年目については、その 1 年目にみられたピークが徐々に外側に向かって拡大していることが分かる。ここから、介入効果のあった世帯については時間が経過するにつれて節電方法について学習し、より節電するようになるが、他方そもそも節電効果のなかった約 18%の世帯については、時間の経過により節電をするようになるのではなく、むしろ電力消費量を増加させていることが分かる。

次に彼らの分析では、節電効果がみられたグループ(reducers)と節電効果がみられず、逆に電力消費量を増加させたグループ(increasers)に分け、その人たちの属性の違いについて調べている(図 3)。

<図 3:”reducers”と”increasers”の属性比較>

図 3 より、多くの属性について平均的に節電行動をとる世帯とそうでない世帯で、属性の違いがみられることが分かる。特に、過去の電力消費量(baseline usage)については著しい差がみられ、節電行動をとる世帯の過去の電力消費量はそうでない世帯よりも平均的にみて、はるかに高くなっていることが分かる。

このように介入効果の異質性について言及したのちに、彼らは分析の焦点をターゲティング

に当てている。ここで彼らは HER による節電効果の高い人に限って介入を行うことで、介入の有効性を高めることを目的としている。具体的には、HER を行うことにより生じる費用を考慮した上で、節電効果を最大にするケースを考えている⁶。また彼らは発電に伴う限界費用を\$7 と仮定したうえでシミュレーションを行っている。それぞれの年度の HER による純便益の分布を描いたものが図 4 である。

<図 4:HER による純便益の分布>

図 4 では純便益が高い世帯から順に横軸に並べている。縦軸は節電による余剰を金銭換算した純便益を表している。図 4 より、仮にすべての人に対して HER を送る場合を考えると、HER による総余剰は 1 年目で\$5.5M、2 年目で\$9.1M、3 年目で\$11.5M というように、時間が経過するにつれて増加していることが分かる。これは介入効果の異質性が時間を経過するにつれて拡大していることに対応する。次にターゲティングを行うケースを考える。限界費用を超える世帯に対してのみ介入を行うことを考えると、一年目に HER が送付されるのは約 72% の世帯である。この時、総余剰は\$6.4M となり、全体に介入を行う場合と比べ 14% 総余剰が増加する。二年目についても、約 73% の世帯が介入を割り当てられ、総余剰は\$10.4M となり、14% の総余剰の増加が見込まれる。最後に三年目については、同様のターゲティングを行うことによって、総余剰は\$13.0M となり、12% の増加が見込まれる。このように事前に介入による効果が予見できている場合、ターゲティングを行うことで総余剰の改善が見込まれることが分かる。

このように、Knittel & Stolper (2019)では実際の社会実験から得られたデータをもとに、CF の手法を用いて介入効果の異質性を調べ、異質性を考慮した上でターゲティングを行うことの有用性について例証している。

4.2 その他の実証研究

ここでは、CF を用いて CATE の推定を行っている実証研究について簡単に紹介していく⁷。まず Carter, Tjernström, and Toledo (2019)では、ニカラグアの小農家を対象にした所得増進のための開発プログラムの効果の異質性を検証している。その結果、特に初期の土地資産や技術効率・教育水準が低く、農業経験の浅い農家に対してプログラムの効果が大きくみられることを報告している。

De Neve et al. (2019)では、納税率にナッジが与える効果の異質性について検証を行ってい

⁶ 彼らの論文の中では他に、発電費用や環境への外部性を電力の限界費用として考慮し、それを踏まえたうえで節電効果を最大化することや、需要家の HER への支払意思額 (willingness to pay: WTP) を考慮した上で総余剰をシミュレートしている。

⁷ ここで紹介するもの以外に Grimm et al. (2020)でも分析に CF を用いているが、そこでは異質性を推定することは行わず、平均介入効果の頑健性確認のために CF を使用している。

る。彼らの研究では 3 種類のナッジについての効果検証を行っている。一つ目は簡素化であり、納税の手続き方法を記載した手紙の内容を簡素にしたものである。二つ目では簡素化に加え、納税が遅れることによる罰則や増税を明記したメッセージを記載している(抑止)。三つでは簡素化に加え、納税者の社会規範に訴えかける内容を付け加えている(社会規範)。こうした介入を行った結果、介入の内容ごとに介入効果の異質性が存在し、簡素化では正の領域で介入効果のばらつきが大きく、抑止や社会規範で介入効果のばらつきが小さく効果が正と負の両方を取りうることを示している。さらに納税者の属性と介入効果の関係性についても調べており、高齢の納税者には簡素化だけでは効きにくいことや、若者に対しては抑止的なメッセージが最も効果的であること、子供のいる納税者に対しては簡素化が効果的であることを報告している。

Farbmacher, Kögel and Spindler (2019)では、貧困状況と認知力についての異質性に焦点を当てている。彼らは Carvalho et al. (2016)のデータを用いてこれを検証している。Carvalho et al. (2016)の研究ではアメリカの低所得者の標本を無作為に給料日前と給料日後に認知テストを行うように割り振り、給料日前後での認知機能の差が、見られないことを報告している。Farbmacher, Kögel and Spindler (2019)では、CATE を推定しより個人に近い形で効果の異質性を見ることで、貧困状況と認知力についての効果検証を行っている。彼らの分析の結果、実験時の収入が低く、若年層と高齢層の人について認知パフォーマンスが悪化したことを報告している。

Hoffman and Mast (2019)では、犯罪の発生率に政府支出が与える効果を検証している。彼らの分析ではセンサスショックを介入としてそれが犯罪発生率に与える効果の異質性を地域ごとに推定している。この結果、低所得や雇用増加率の低い地域で効果が大きくなり、犯罪発生率が抑制されていることを報告している。

O'Neill and Weeks (2018) では、時間帯別電気料金が節電行動に与える効果の検証を行っている。彼らの研究ではアイルランドエネルギー規制委員会(CER 2011)が行った Electricity Smart Metering Customer Behavioral Trial のものを使用している。彼らはこのデータを用いて、時間帯別電気料金の導入による節電効果に異質性がみられることを、モデルに含める変数の組み合わせを変えて検討している。また、変数重要度を精査することを通して、過去の電力使用量が異質性を記述するために重要であることを報告している。また属性情報と介入効果の関係についてもみており、例えば電力消費量の大きい若く高学歴な世帯ほど介入効果が大きくなるといったことを報告している。

Miller (2020)では、漁場崩壊に個別の漁獲量割当制度が与える効果について異質性を検討している。彼らは政策効果が動学的に変化する点に着目し、CF を修正して推定を行っている。これにより、彼らは割当制度による介入効果が漁業間での不均一性を持つだけでなく、時間経過に対しても不均一性を持つことを報告している。

Bertrand et al. (2017)では、コートジボワールで行われた公共事業に対する臨時雇用が、雇用状態に与える効果の異質性を検証している。彼らの分析の結果、プログラム期間中の収入

への異質性は大きく、プログラム終了後にはより小さな異質性がみられることを報告している。また、このプログラムはコストが高く費用対効果が悪いが、自己選抜ではなく異質介入効果に基づいたターゲティングを行うことで、費用対効果を高めることができることを示唆している。

Allcott and Kessler (2019)では節電効果に HER が与える効果を検証している。そこからさらに、心理的なコストを加味した形での厚生評価を行っている。その厚生評価を行う際に異質介入効果を推定し、節電量や HER への支払意思額、社会厚生のをそれぞれを最大化するターゲティングを行っている。

Davis and Heller (2017)では、就業プログラムがフルタイム雇用と犯罪率に対して与える効果の検証を行っている。それにより、雇用についての介入効果は平均的にはみられないが、異質性は確認され、特に若くて熱心に学校に通っていたタイプの人で効果がみられている。また、犯罪率について平均介入効果はみられるが、異質性はみられていない。また彼らの研究ではターゲティングにも言及している。ターゲティングを実施することで、就業プログラムのようなコストがかさむプログラムであっても、費用対効果を高めることができることを示唆している。

4.3 ロバストネス・テスト

推定した CATE の異質性が正確に推定できているかどうか、そして推定した CATE が十分なばらつきを持っているかということは重要である。これを確かめる方法として、Chernozhukov et al. (2018)では Best Linear Prediction(BLP)という方法を提唱している。

彼らの方法では、介入効果の異質性が線形であるとして、実際のアウトカムと予測値がどれくらい相関しているかということ計算する。具体的な計算手順は、以下の通りである。

- ① データを訓練データとテストデータに分割する。
- ② 訓練データを用いて予測モデルを育てる。
- ③ 予測モデルにテストデータを当てはめ、介入効果の予測値 $\hat{t}(x)$ を計算する。
- ④ 以下の回帰式を計算する。

$$Y = \alpha \hat{m}(X) + \beta_1(W - \hat{p}(X)) + \beta_2(W - \hat{p}(X))(\hat{t}(x) - \bar{t}) + \varepsilon,$$

$$E[w(X)\varepsilon X] = 0$$

ここで、 \bar{t} は $\hat{t}(x)$ の期待値、 $\hat{p}(X)$ は介入を受けることに対する傾向スコア、 $\hat{m}(X)$ は介入が与えられていないときの X で条件付けた Y の予測値である。また、 $w(X) = \{p(X)(1 - p(X))\}^{-1}$ である。

この①～④までの手順を複数回繰り返し、求めた $\hat{\beta}_1$ と $\hat{\beta}_2$ の係数値と標準誤差を用いて信頼区間を作り、それぞれの係数の中央値と信頼区間の上限・下限の中央値を報告する。

上記の式は真の介入効果と予測された介入効果が以下のような関係にあると仮定している。

$$E[Y_1 - Y_0 | \hat{t}] = \beta_1 + \beta_2(\hat{t}(x) - E[\hat{t}(x)])$$

β_1 は平均介入効果を表し、 β_2 は $\hat{\tau}$ が 1 変化することに対して真の介入効果がどれだけ変化するかということを表している。この β_2 が 1 であればモデルの異質性と真の異質性が完全に線形であることを示す。

R パッケージの `grf` 内の `test_calibration` でも同様のものを計算することができる。この関数では、抜取標本を用いて評価を行う。具体的には、抜取標本を用いて作成した予測値を使用して、 $C = \bar{\tau}(W - \hat{p}(X))$ と $D = (\hat{\tau}(X) - \bar{\tau})Y - \hat{m}(X)$ に回帰する。 D の係数は BLP の β_2 に相当し、異質性が正しく推定できているかを評価するものである。また、 C の係数は実際の平均介入効果と予測モデルから得られた平均介入効果の相関を表すものである。

5. 結論

本稿では近年研究が進んでいる因果推論の機械学習、特に CF についての紹介および CF を用いた実証研究についての紹介を行った。

著者らも寄付についてのオンライン実験を行い、そこで介入効果の異質性の検証を行っている(佐々木 他 2019)。著者らは国外の植林活動への支援と国内における被災者活動への支援という異質な寄付先を 2 つ設定し、2 水準のマッチング寄付、社会比較、それらの組み合わせ介入の計 5 つの介入を行い平均的にどれくらいの介入効果がみられるのかを明らかにした。そこから CF の手法を用いて、それぞれの寄付先における介入効果の異質性を検証している。さらに現在は、その介入効果の異質性に基づき、「誰にどの介入を与えるべきか」というテーラーメイドなターゲティングの有用性について検証を行っている。

また、著者らは節電行動についての異質性・ターゲティングの研究も行っている(村上 他 2020)。この研究では、節電努力に応じたリポートと社会比較ナッジを与え、介入効果が異質性をもつことを示した。さらに、ターゲティングを行うことでより大きな節電行動を引き出せることを検証している。

このように、CF をはじめとした機械学習の手法を因果推論に用いることで、介入効果の異質性だけではなく、個人・世帯ごとの個別介入規則 (Individual Treatment Rule; ITR) を求めることができる。そしてそれに基づくターゲティングを行うことで、より大きな社会厚生が達成される可能性がある。これは連載第 3 回目に予定している経験厚生最大化 (Empirical Welfare Maximization; EWM) の話題とも似ているが、CF や関連する因果推論による機械学習の方法では MSE を最小化することが目的であったが、EWM では厚生損失を最小化することに主眼が置かれる点に違いがある。

こうした手法の発展は著しいが、きれいなランダム化されたデータが必要であるという点では従来と変わりはない。むしろこうして手法が発展してきたからこそ、きれいにランダム化されたデータの価値はますます高まってきていると著者は考えている。

謝辞

本稿の執筆にあたって、佐々木周作氏、伊東慧介氏、平井直知氏から有益なコメントを頂いた。

参考文献

- [1] Allcott, H., & J.B., Kessler. (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics*, 11(1), 236-76.
- [2] Athey, S., & G.W., Imbens. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27), 7353-7360.
- [3] Athey, S., & G.W., Imbens. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11, 685-725.
- [4] Athey, S., J., Tibshirani, & S., Wager. (2019). Generalized random forests. *The Annals of Statistics*, 47(2), 1148-1178.
- [5] Bertrand, M., B., Crépon, A., Marguerie, & P., Premand. (2017). Contemporaneous and Post-Program Impacts of a Public Works Program: Evidence from Côte d'Ivoire. *World Bank*.
- [6] Breiman, L., J., Friedman, C.J., Stone, & R.A., Olshen. (1984). *Classification and regression trees*. CRC press.
- [7] Carter, M.R., E., Tjernström, & P., Toledo. (2019). Heterogeneous impact dynamics of a rural business development program in Nicaragua. *Journal of Development Economics*, 138, 77-98.
- [8] Carvalho, L.S., S., Meier, and S.W., Wang. (2016). Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday. *American Economic Review*, 106(2), 260–284.
- [9] CER (2011), Electricity smart metering customer behaviour trials (cbt) findings report, Technical report, Commission for Energy Regulation.
- [10] Chernozhukov, V., M., Demirer, E., Duflo, & I., Fernandez-Val. (2018). *Generic machine learning inference on heterogeneous treatment effects in randomized experiments*. National Bureau of Economic Research.
- [11] Davis, J.M., & S.B., Heller. (2017). Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *Review of Economics and Statistics*, 1-47.
- [12] De Neve, J.E., C., Imbert, J., Spinnewijn, T., Tsankova, & M., Luts. (2019). How to Improve Tax Compliance? Evidence from Population-wide Experiments in Belgium. Evidence from Population-Wide Experiments in Belgium. *Saïd Business School WP*, 7.
- [13] Farbmacher, H., H., Kögel, & M., Spindler. (2019). Heterogeneous Effects of Poverty on Cognition. *Munich Center for the Economics of Aging*.
- [14] Grimm, V., S., Kretschmer, & S., Mehl. (2020). Green innovations: The organizational setup of pilot projects and its influence on consumer perceptions. *Energy Policy*, 142, 111474.
- [15] Hastie, T., R., Tibshirani, J., Friedman. (2018). 統計的学習の基礎。(杉山将, 井手剛, 神尾敏弘, 栗田多喜夫, 前田英作監訳). 共立出版.

- [16] Hoffman, I., & E., Mast. (2019). Heterogeneity in the effect of federal spending on local crime: Evidence from causal forests. *Regional Science and Urban Economics*, 78, 103463.
- [17] James, G., D., Witten, T., Hastie, and R., Tibshirani. (2018). R による統計的学習入門. (落海浩, 首藤信道訳). 朝倉書店.
- [18] 依田高典. (2019). 経済分析のツールとしての機械学習. *経済セミナー*, 711, 23-27.
- [19] Knittel, C.R., & S., Stolper. (2019). Using Machine Learning to Target Treatment: The Case of Household Energy Use (No. w26531). National Bureau of Economic Research.
- [20] Miller, S. (2020). Causal forest estimation of heterogeneous and time-varying environmental policy effects. *Journal of Environmental Economics and Management*, 102337.
- [21] 村上佳世, 寫田栄樹, 牛房義明, 依田高典. (2020). ナッジとリベートの異質介入効果: 因果的機械学習の節電フィールド実験への応用. *京都大学経済学研究科* No. J-20-003.
- [22] Mullainathan, S., & J., Spiess. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87-106.
- [23] O'Neill, E., & M., Weeks. (2018). Causal Tree Estimation of Heterogeneous Household Response to Time-Of-Use Electricity Pricing Schemes. *arXiv preprint arXiv:1810.09179*.
- [24] 佐々木周作, 石原卓典, 木戸大道, 北川透, 依田高典. (2019). 寄付行動を促すための介入の効果は寄付先活動によって異なるか?: 全国規模オンライン実験と機械学習に基づく検証. *行動経済学*, 12(Special_issue), S14-S17.
- [25] Tibshirani, J., S., Athey, R., Friedberg, V., Hadad, D., Hirshberg, L., Miner, E., Sverdrup, S., Wager, & M., Wright. (2018). Package 'grf'.
- [26] Wager, S., & S., Athey. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228-1242.

図表

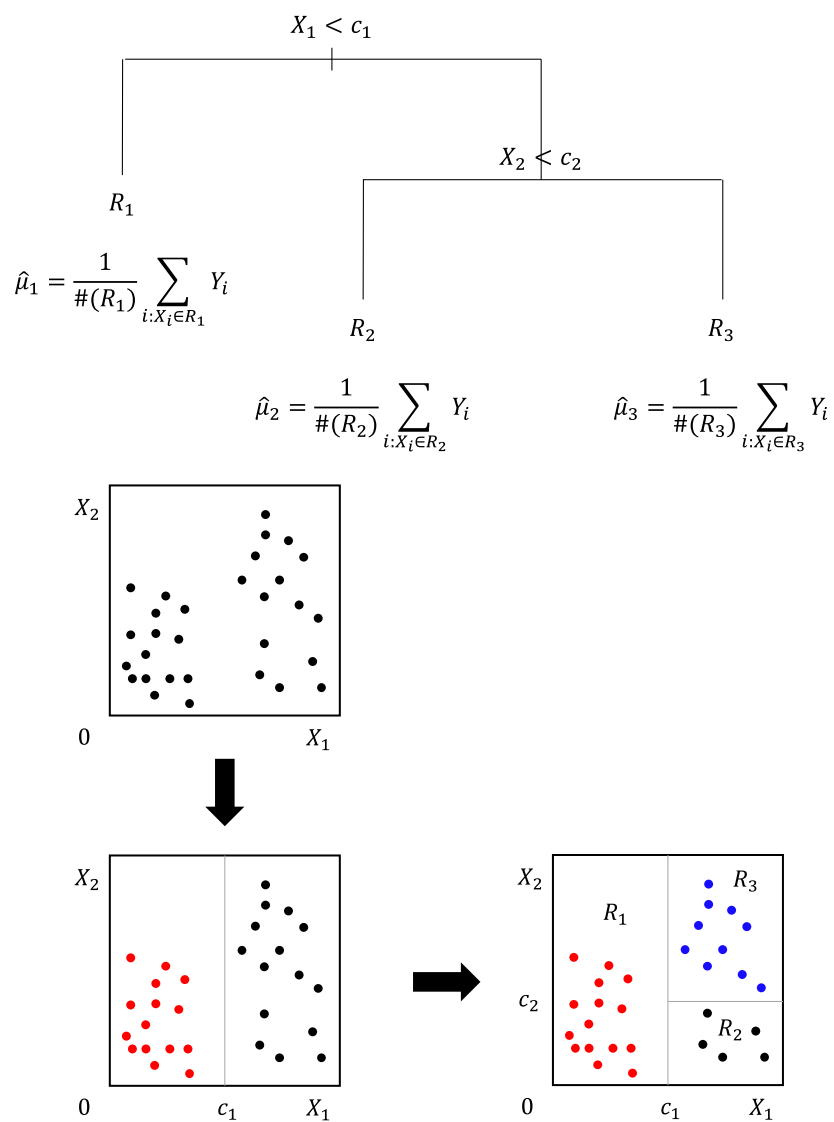


図 1: 決定木の例

表 1: ランダムフォレストのアルゴリズム

アルゴリズム: ランダムフォレスト

- Step 1 各 $b = 1, \dots, B$ について:
- a. サイズ N の訓練データからブートストラップ標本 Z^* を取り出す.
 - b. 取り出したブートストラップ標本 Z^* に基づき, 決定木 T_b を育てる. ここで, 最小ノードサイズ n_{min} に到達するまで, 再帰的に以下の手順を繰り返す.
 - i. p 個の特徴量の中からランダムに m 個の変数を選び出す.
 - ii. m 個の変数の中から最も応答と相関の高い変数を選び出し, その分割のための閾値を決める.
 - iii. ノードを 2 つの子ノードへと分割する.
- Step 2 上記の Step 1 で求めた各決定木 $\{T_b\}_1^B$ についての予測結果を統合する. 回帰木の場合テスト点 x についての予測は,

$$\hat{\mu}(x) = \mathbb{E}[Y_i | X_i = x] = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

として与えられる.

Hastie et al. (2018) P676 アルゴリズム 15.1 をもとに作成.

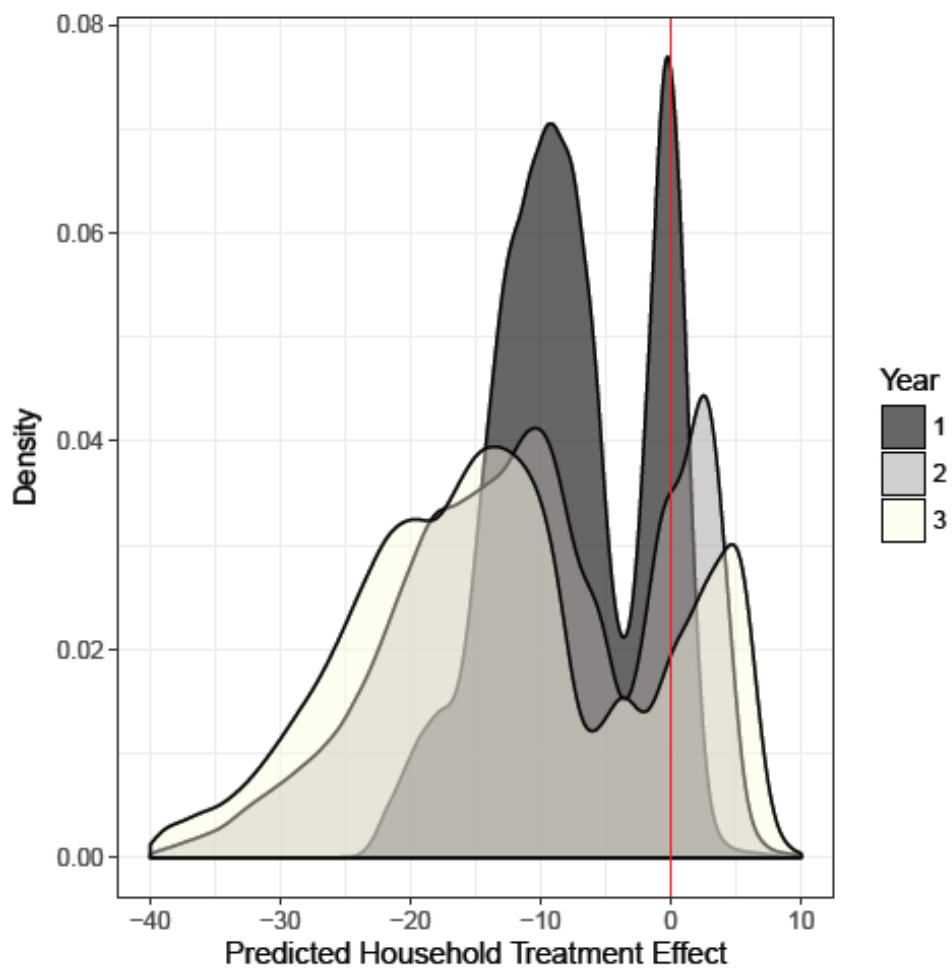


図 2: 介入効果の予測値の分布

Knittel & Stolper. (2019) Figure 7 より引用.

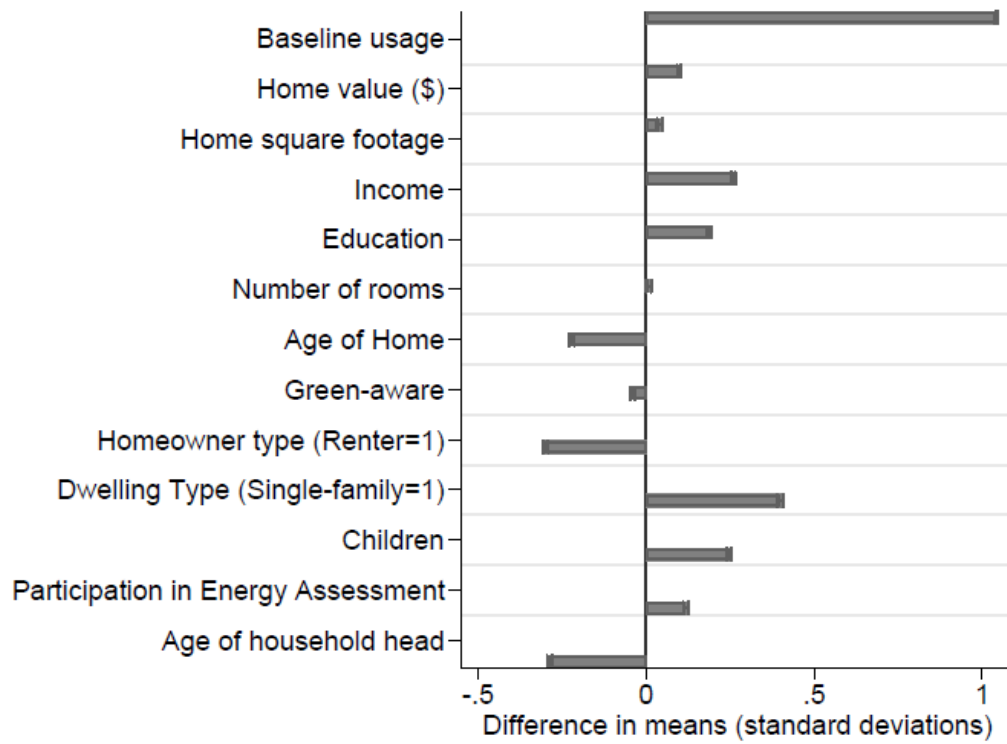


図 3: "reducers"と"increasers"の属性比較

Knittel & Stolper. (2019) Figure 8 より引用.

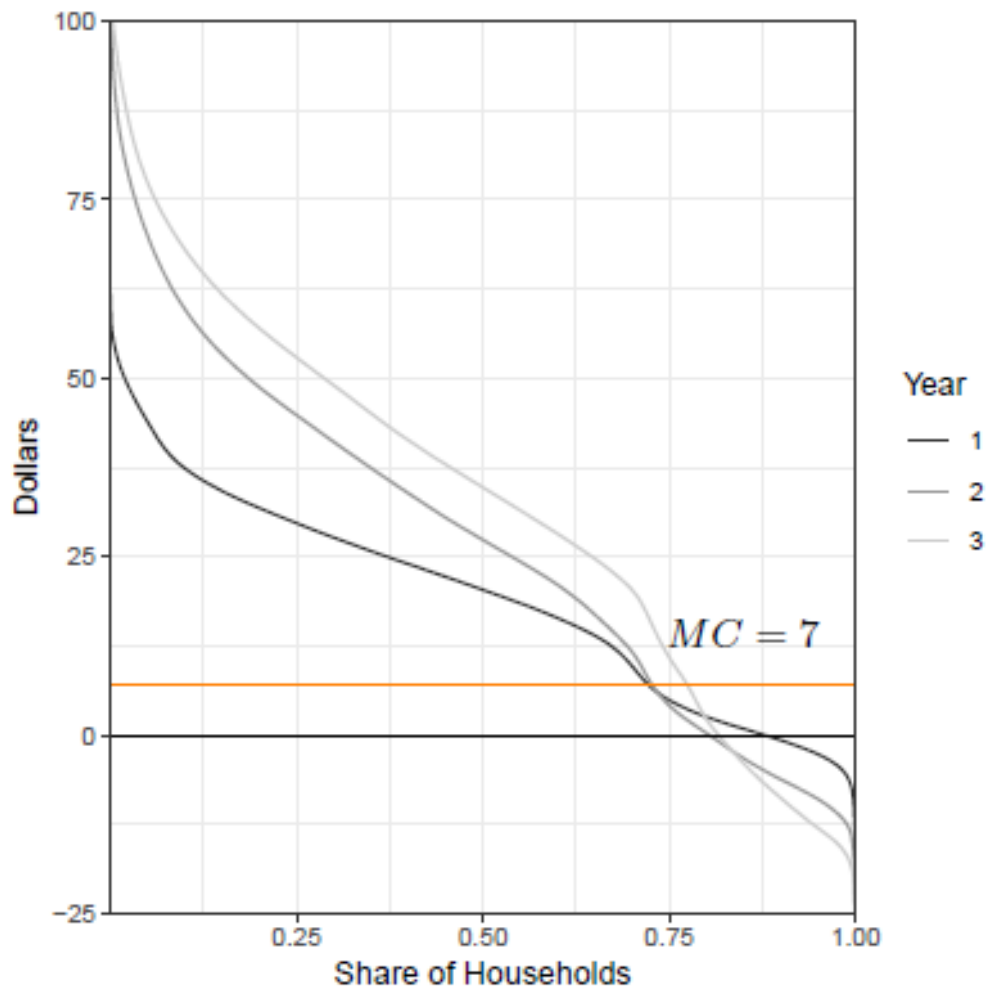


図 4:HER による純便益の分布

Knittel & Stolper. (2019) Figure 10 の Panel A より引用.

表 2: CF を用いた実証研究

	内容	HTE の分布	変数重要度	$\hat{\tau}(x)$ と X の関係性	$\hat{\tau}(x)$ による厚生評価 (ターゲティング)
Carter, Tjernström, & Toledo (2019)	農家の所得増加プログラムが所得に与える効果の検証.	×	×	○	×
Bertrand et al. (2017)	臨時雇用が雇用状態に与える効果の検証.	○	×	○	○
Davis & Heller (2017)	若者のフルタイム雇用と犯罪率に対して就業プログラムが与える効果の検証.	×	×	○	○
Farbmacher, Kögel & Spindler (2019)	貧困状態が認知力に与える効果の検証.	×	○	○	×
Hoffman & Mast (2019)	犯罪発生率に政府支出が与える効果の検証.	○	×	○	×
Miller (2020)	漁業の衰退に漁獲量個別割当制度が与える効果の検証.	○	×	○	×
O'Neill & Weeks (2018)	時間帯別電気料金が節電行動に与える効果の検証.	○	○	○	×
Knittel & Stolper (2019)	HER が節電行動に与える効果の検証.	○	○	○	○
Allcott & Kessler (2019)	HER が節電行動に与える効果の検証.	×	×	×	○
De Neve et al. (2019)	納税行動にナッジが与える効果の検証.	○	×	○	×

※ 表中の○は論文または appendix に報告があることを, ×は論文または appendix に報告がないことを示す.