



機械学習とモデル解釈手法のSHAPを用いた 光熱費の分析

電力中央研究所 社会経済研究所／グリッド
イノベーション研究本部
西尾 健一郎
nishio@criepi.denken.or.jp

「カーボンニュートラルに向けた家庭部門CO₂排出実態統計調査の活用」
シンポジウム（第13回ESIシンポジウム）

2021年7月1日

 電力中央研究所

内容

1. 家庭CO₂統計への期待

2. 機械学習と モデル解釈手法のSHAPを用いた 光熱費の分析

詳細

西尾：家庭CO₂統計の個票データと機械学習を用いた建築時期別光熱費の実態把握,
エネルギー・資源学会論文誌, 42(3), 175-184, 2021.5.

https://doi.org/10.24778/jjser.42.3_175

1. 家庭CO₂統計への期待

公開されている集計表や関連報告書はとても有用である

「家庭からの二酸化炭素排出量の推計に係る実態調査 試験調査」を業務に活用した例

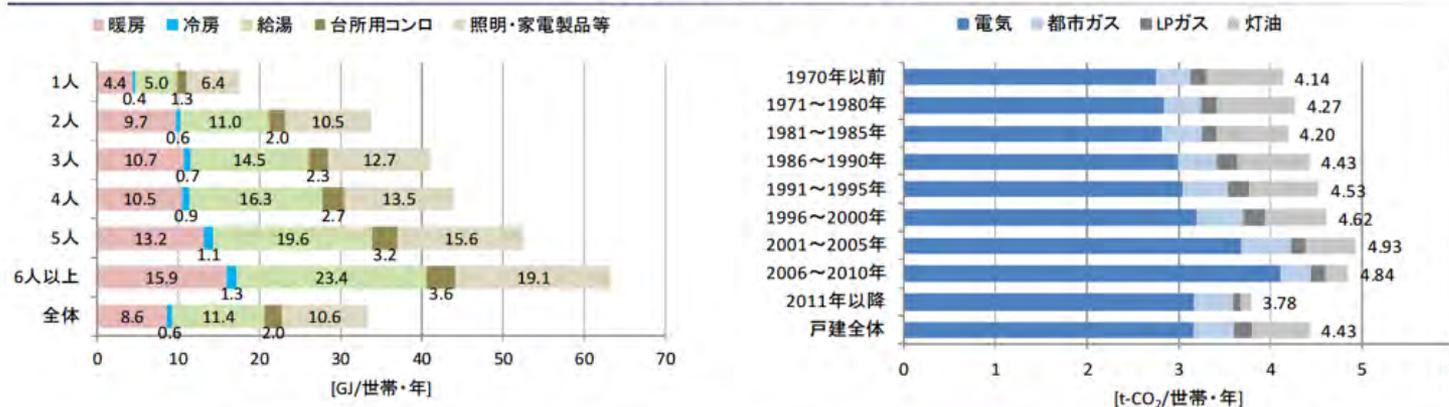


図 1-57 世帯人数別世帯当たり年間用途別エネルギー消費量 図 1-27 建て方別建築時期別世帯当たり年間エネルギー種別 CO₂ 排出量 (戸建)

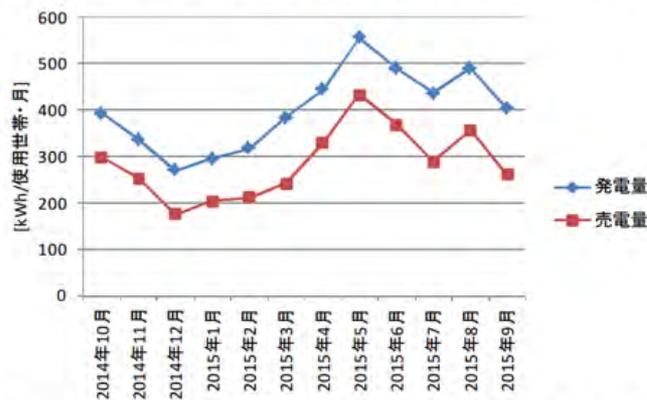
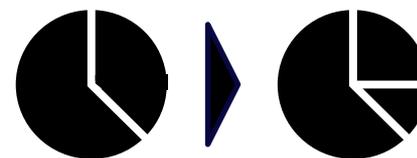


図 1-115 太陽光発電システムの月別発電量・売電量 (使用世帯当たり) 統合集計(参考値) (環境省、2016)

個票データにもポテンシャルがある

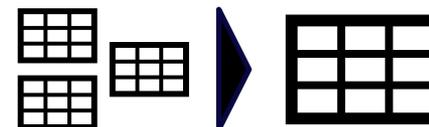
1. 統合・分割などの**柔軟なデータ処理**ができる

例
 ガス機器のエネルギー源について、
 都市ガスか液化石油（LP）ガスかを特定



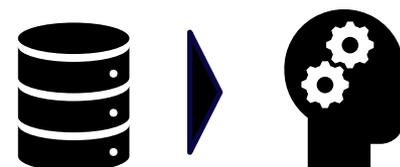
2. 公開情報にはない**詳細セグメント**での**実態把握**ができる

例
 建て方・建築時期・世帯人数別の
 機器シェアを直接特定



3. **統計解析**や**機械学習**ができる

例
 約1万世帯のデータを活かして
 エネルギー消費量や光熱費を推定



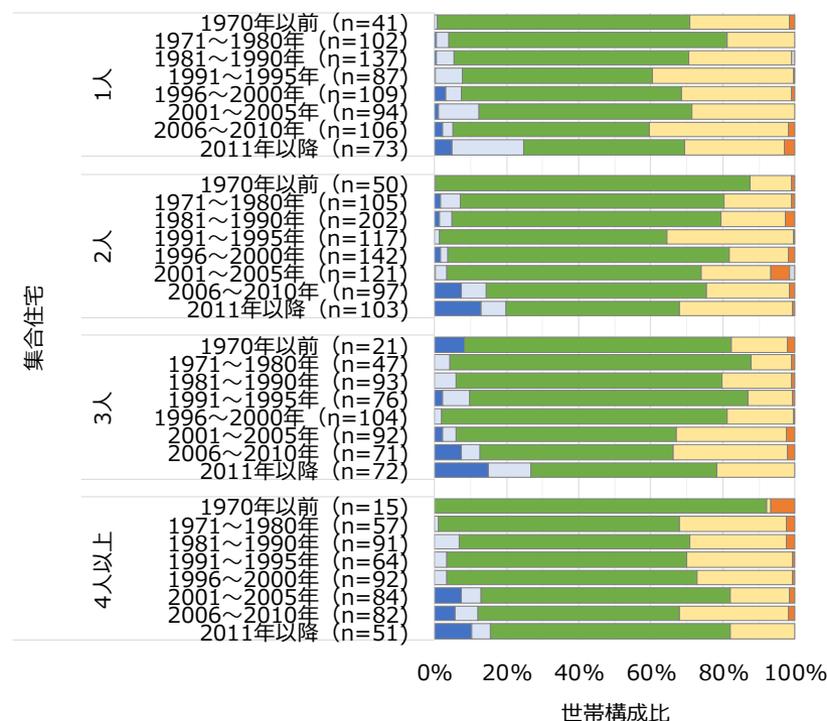
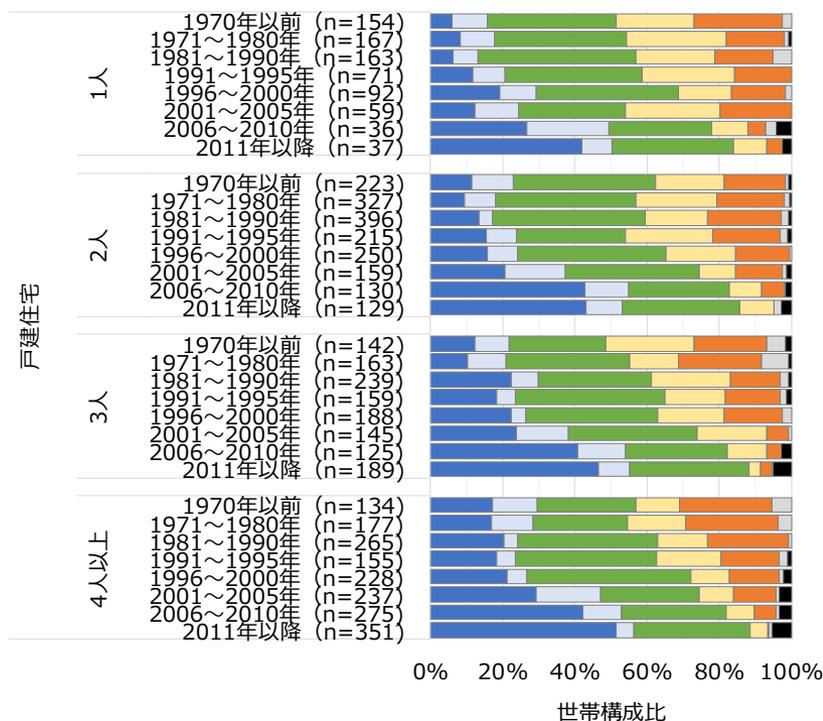
個票データの分析例 1 : 給湯機器の保有実態

■ カーボンニュートラルの実現に向けて、電化に期待が寄せられているが、住宅設備はロックインしがちであり早急な対応が求められる

戸建住宅：2011年以降に建築された住宅では、ヒートポンプ（HP）給湯機は5割程度に。発売前の2000年以前築の住宅を見ると、既築交換によりHP給湯機が設置済みなのは1~2割程度。受容性はあるだけに、いかにして普及を後押しできるか

集合住宅：2011年以降に建築された住宅でも、HP給湯機は伸び悩み。2000年以前築の住宅の既築交換率も極めて低調。交換が容易でないだけに、脱炭素化に向けては、新築時のHP給湯機の採用率を向上させていくことが不可欠

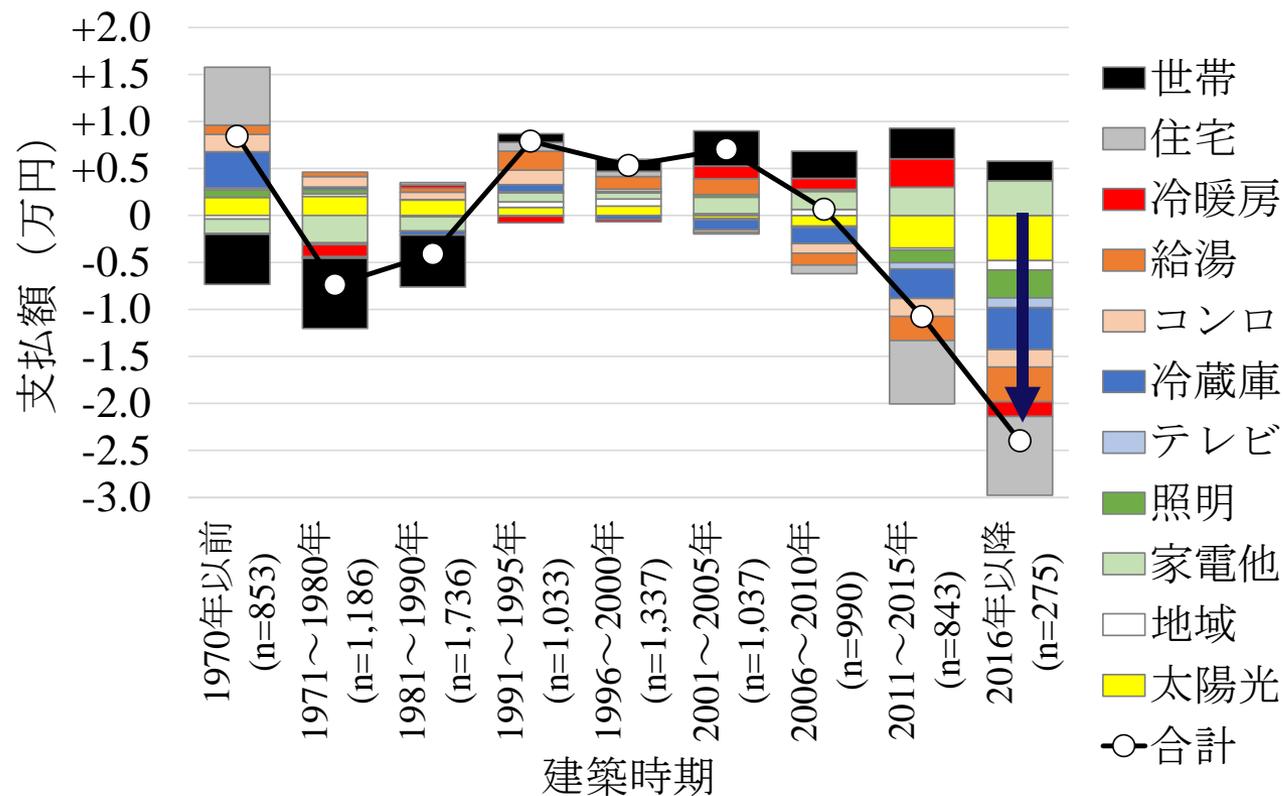
■ ヒートポンプ給湯機 ■ 電気温水器 ■ 都市ガス給湯器 ■ LPガス給湯器 ■ 灯油給湯器 ■ 太陽熱温水器 ■ コージェネレーション



西尾・中野：家庭CO₂統計の個票データを用いた給湯機器の保有実態などに関する基礎検討，第39回エネルギー・資源学会研究発表会 講演論文集，2020.

個票データの分析例2：光熱費の差異をもたらす要因

- 2016年以降に建てられた住宅で光熱費が抑えられているのは、住宅・冷蔵庫・給湯・太陽光・照明・テレビ・コンロ要因などによる



詳細は
次ページ
以降

2. 機械学習と モデル解釈手法のSHAPを用いた 光熱費の分析

研究の背景・目的

光熱費の実態把握

- 温暖化対策を検討する上で、CO₂削減効果が重要であることは言うまでもないが、**需要家の視点に立つと光熱費も関心事項**である
- 検討に資するデータとして期待されるのが、**環境省が平成29年度より実施している「家庭部門のCO₂排出実態統計調査」（家庭CO₂統計）**である
 1. 政府統計ポータルサイト（e-Stat）や環境省Webサイトにおいて、集計表や関連報告書が多数公開されている



<https://www.e-stat.go.jp/>



<http://www.env.go.jp/earth/ondanka/ghg/kateiCO2tokei.html>

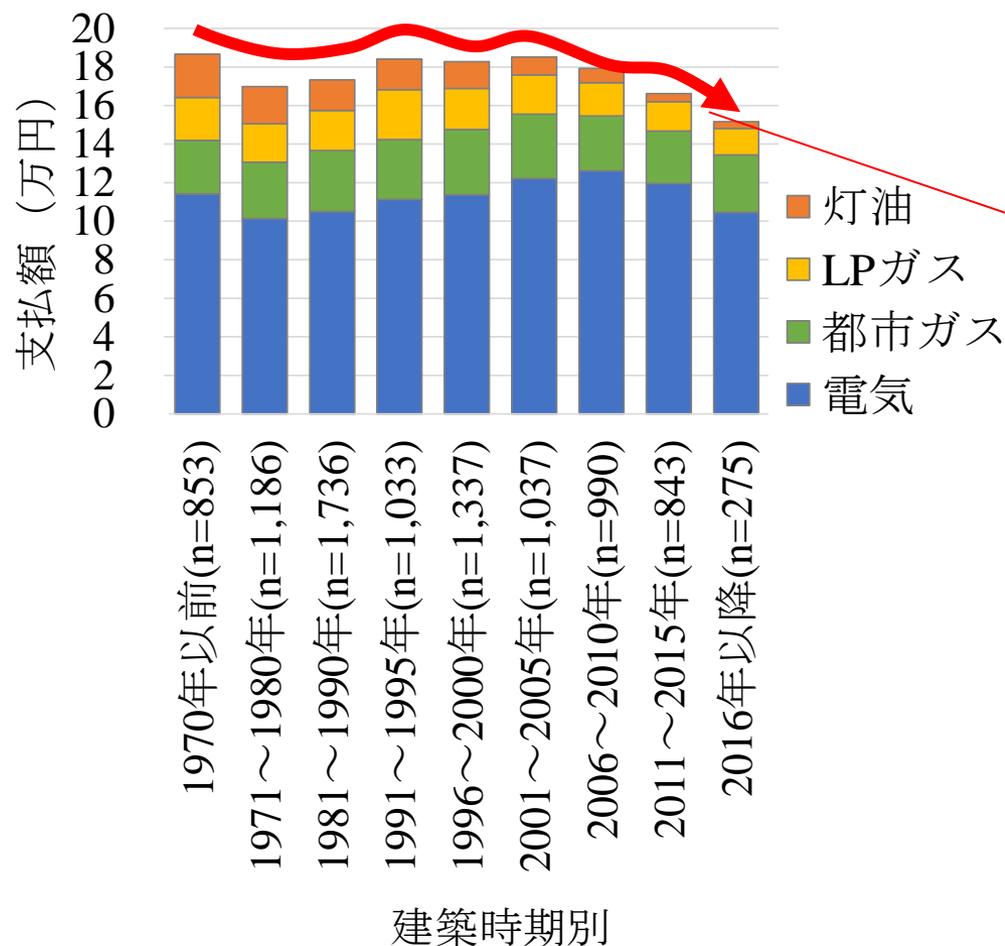
2. エネルギー・資源学会「家庭部門のCO₂排出実態統計調査利用研究会」では、**個票データを多様な視点で分析すること**としている

本研究

**平成30年度家庭CO₂統計の個票データを用いて、
建築時期による光熱費差異の要因を明らかにする**

建築時期別の光熱費

- 近年建てられた住宅では光熱費が抑えられているように見えるが、単純な分析（例：クロス集計）からは、その理由がよくわからない



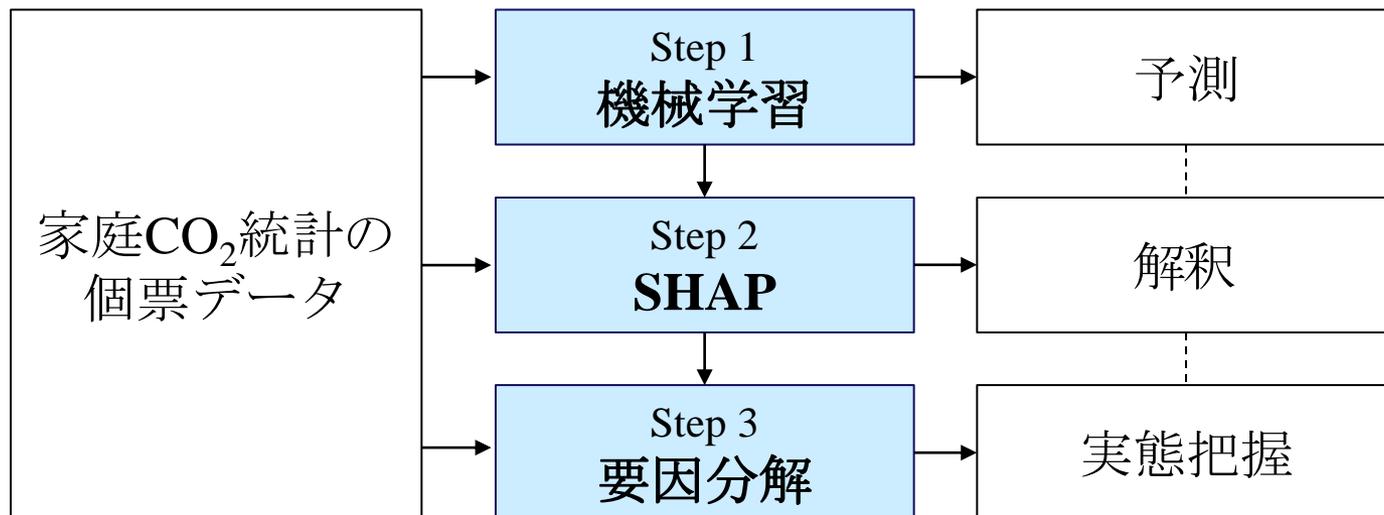
仮説の例
 住宅断熱性能が高い？
 省エネ製品が多い？
 太陽光発電が増えた？
 . . . ?

検証の方法
 積み上げ or
 重回帰分析 or
 .
 .
 .
 機械学習 (本研究)

手法

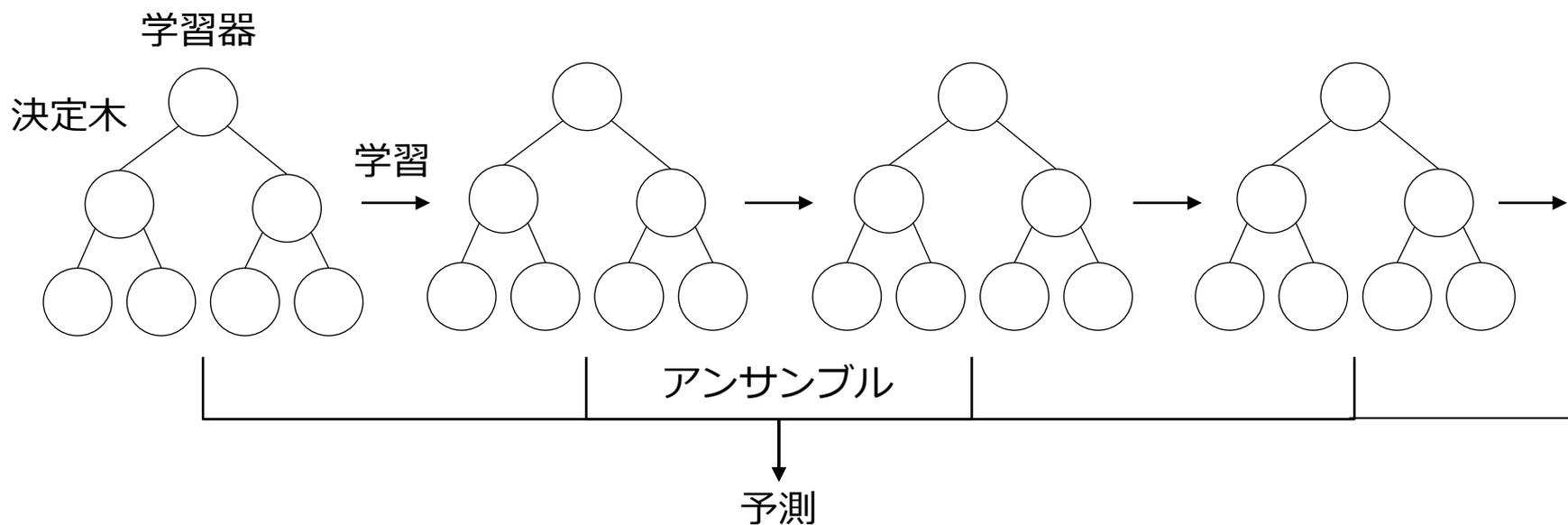
分析フロー

- データ分析の最新要素技術も活用して、光熱費の実態把握を行う



Step 1 勾配ブースティング木 (GBDT) とは

- **機械学習手法**の1つであり、**非線形性や相互作用を考慮しやすく**、**家庭CO₂統計と類似のデータでも良好な精度**が示されることが多い
 - ① **Gradient** : 勾配降下法により、誤差を最小化する
 - ② **Boosting** : 弱学習器を多く使うことで精度を上げるアンサンブル学習の1つで、目的変数と予測値の誤差を小さくしていく
 - ③ **Decision Tree** : 木構造により、分類や回帰をする



Step 1 GBDTモデルの構築

- 判断の加え過ぎで要因分解が部分的になってしまうのを避けるため、機械学習では**個票データの情報量を最大限活用**することにした

対象データ

平成30年度家庭CO₂統計の個票データのうち、建築時期不明の世帯を除く9,290サンプル

目的変数

光熱費支払額（万円／年）

- 2018年4月～2019年3月の電気・都市ガス・LPガス・灯油の合計（ガソリン・軽油は含まず）
- 太陽光発電の自家消費に伴う支払額の減少は反映されている（売電受取額は反映されない）
- 平均は17.7万円

特徴量

住宅・世帯属性や機器保有・利用状況など244個

- ガス機器は、都市ガス・LPガスで細分化
- 所在の10地域情報と共に、都道府県の代表地点の年間平均気温も投入
- 内容の重複度が極めて高いデータについては、情報量が多い方のデータを投入
- 欠損値の除外・補完処理はしない（それ自体が特徴として考慮される）

勾配Boosting木（GBDT）モデルの構築

予測精度の向上と過学習の抑制のため、交差検証によりパラメータチューニング

Step 1~3

特徴量とその要因分類

■ 世帯・住宅属性、機器保有・利用行動などとの関係を機械学習した

要因		特徴量
大分類	小分類	
世帯	世帯人数	世帯人数
	世代	世帯主の年齢, 世帯構成員の最低・最高年齢
	世帯年収	世帯年収
	在宅率他	平日昼間の在宅状況, 就業者人数比率
住宅	建て方	住宅の建て方
	所有関係	住宅の所有関係
	延床面積	延床面積
	窓状況	二重サッシまたは複層ガラスの窓の設置状況
冷暖房	建築時期	建築時期
	エアコン	エアコンの使用台数, 1・2・3・4・5目目エアコンの種類・製造時期
	暖房機器	主暖房機器, 各種暖房機器の使用台数, セントラル暖房・床暖房・太陽熱暖房の使用有無
	冷房行動	夏季平日の1台目エアコン使用時間, 1台目エアコンの冷房設定温度, ペットのための冷房使用状況
給湯	暖房行動	冬季平日の主暖房機器使用時間, 主暖房機器の暖房設定温度・強度, 暖房居室比率, 暖房の仕方, ペットのための暖房使用状況
	給湯機器	各種給湯機器の設置有無
	入浴行動	夏・冬季の湯はり・シャワー確率, 入浴の各種省エネ行動
	他行動	冬季の洗面・台所お湯使用確率, 入浴以外の各種省エネ行動
コンロ	コンロ機器	各種台所用コンロの設置有無
	調理行動	朝・昼・夜の調理確率, 調理の各種省エネ行動
冷蔵庫	台数	冷蔵庫の使用台数
	1台目仕様	1台目冷蔵庫の種類・内容積・製造時期
	2台目仕様	2台目冷蔵庫の種類・内容積・製造時期
	利用行動	冷蔵庫の各種省エネ行動

要因		特徴量
大分類	小分類	
テレビ	台数	テレビの使用台数
	1台目仕様	1台目テレビの種類・画面サイズ・製造時期
	2台目仕様	2台目テレビの種類・画面サイズ・製造時期
	3台目仕様	3台目テレビの種類・画面サイズ・製造時期
照明	利用行動	1台目テレビの平日使用時間, テレビの各種省エネ行動
	居間照明	居間における各種照明の使用有無, 居間主照明の種類
	他照明	食卓・台所・個室・その他場所における各種照明の使用有無
家電他	利用行動	居間照明の使用時間, 照明の各種省エネ行動
	洗濯乾燥	洗濯機・衣類乾燥機・浴室乾燥機の使用台数, 乾燥機能の使用確率
	食洗乾燥	食器乾燥機・食器洗い乾燥機の使用台数
	トイレ	温水洗浄便座・暖房便座の使用台数, トイレの各種省エネ行動
	水サーバー	ウォーターサーバーの使用台数
地域	その他	各種家電の使用台数, 家電の各種省エネ行動, ペット関連機器の使用有無, HEMSの導入有無, 電気自動車・プラグインハイブリッド車・電動オートバイ等の使用有無
	地域他	10地域区分, 契約電力会社, 都市階級, 都市ガス供給エリアの該当有無
	気温	都道府県代表地点の年間平均気温
太陽光	融雪機器	各種融雪機器の使用有無
	太陽光有無	太陽光発電の使用有無
	太陽光容量	太陽光発電の総容量

Step 3
要因に集約した上で
実態把握

Step 1&2
多数の特徴量を用いて
機械学習

Step 2

SHAP (SHapley Additive exPlanations) とは

- **複雑なモデルの予測の解釈手法**として2017年に提案されたもので、機械学習のブラックボックス問題への対応手段を与えてくれる

- **Shapley値は**、協力ゲーム理論で獲得利得を各プレイヤーに公平に分配する手法

例：

各条件下の利得	
A	3
B	4
C	5
A&B	6
A&C	7
B&C	9
A&B&C	10

➔

協力順 (全6通り) による限界利得			
	A	B	C
A⇒B⇒C	3	3	4
A⇒C⇒B	3	3	4
B⇒A⇒C	2	4	4
B⇒C⇒A	1	4	5
C⇒A⇒B	2	3	5
C⇒B⇒A	1	4	5

➔

Shapley値 (全6通りの平均)			
	A	B	C
	2	3.5	4.5

- これを応用したSHAPは、各特徴量がモデル予測値に及ぼす影響へと分解

加法的 (additive) な特徴帰属手法

特性例：
ローカルな正確性

$$f(x) = g(x') = \phi_0 + \sum_{i=1}^M \phi_i x'_i$$

元の予測モデル 元の入力値 単純化した入力値 出力値への影響

単純化した説明 (explanation) モデル 全てがオフの場合の出力値

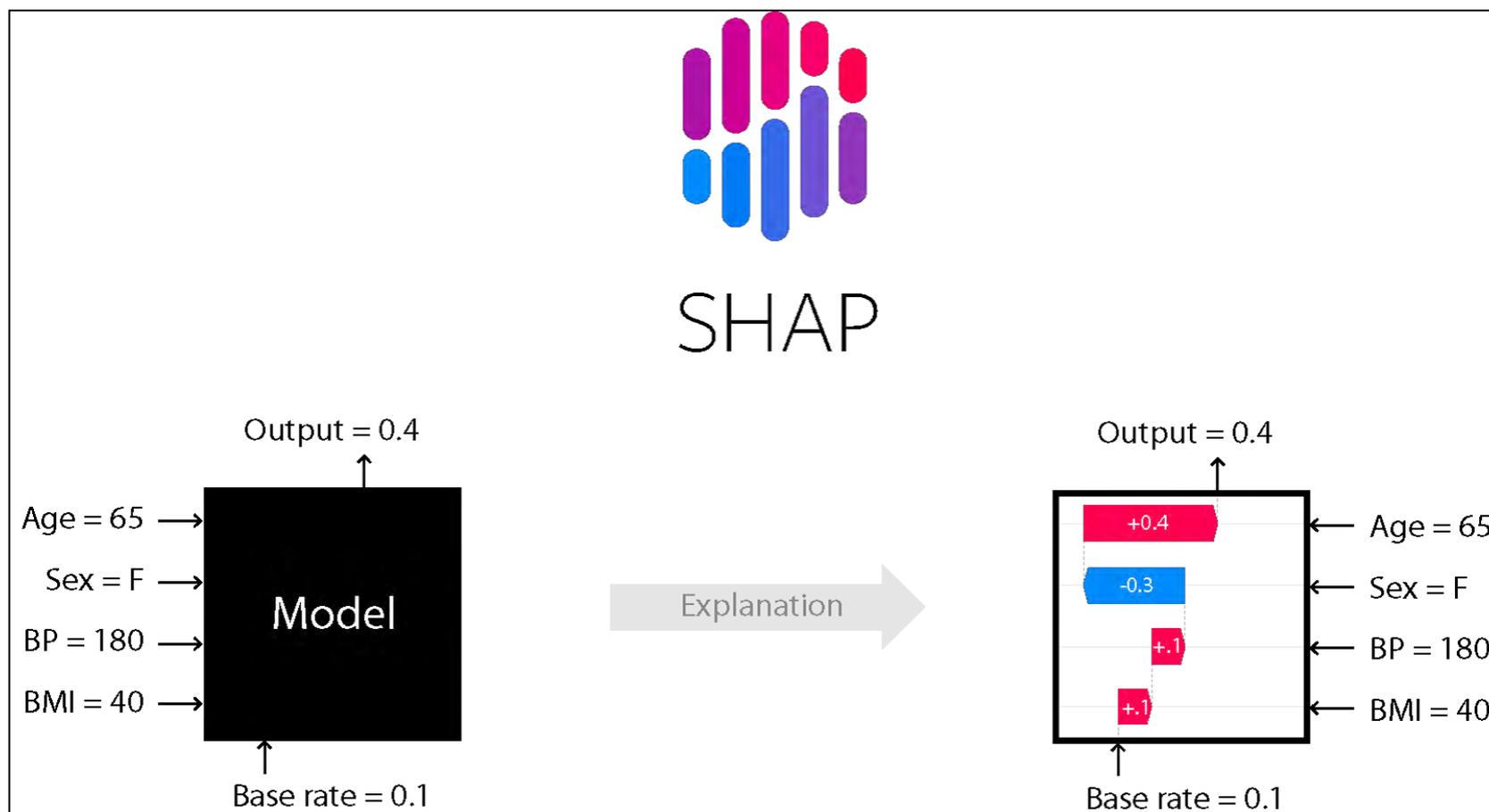
M ← 単純化した入力値の数

- ローカル (ミクロ) な解釈からグローバル (マクロ) な解釈までを一貫してできる
- 木構造モデルでは計算負荷を抑えられる

Step 2

SHAP (SHapley Additive exPlanations) とは

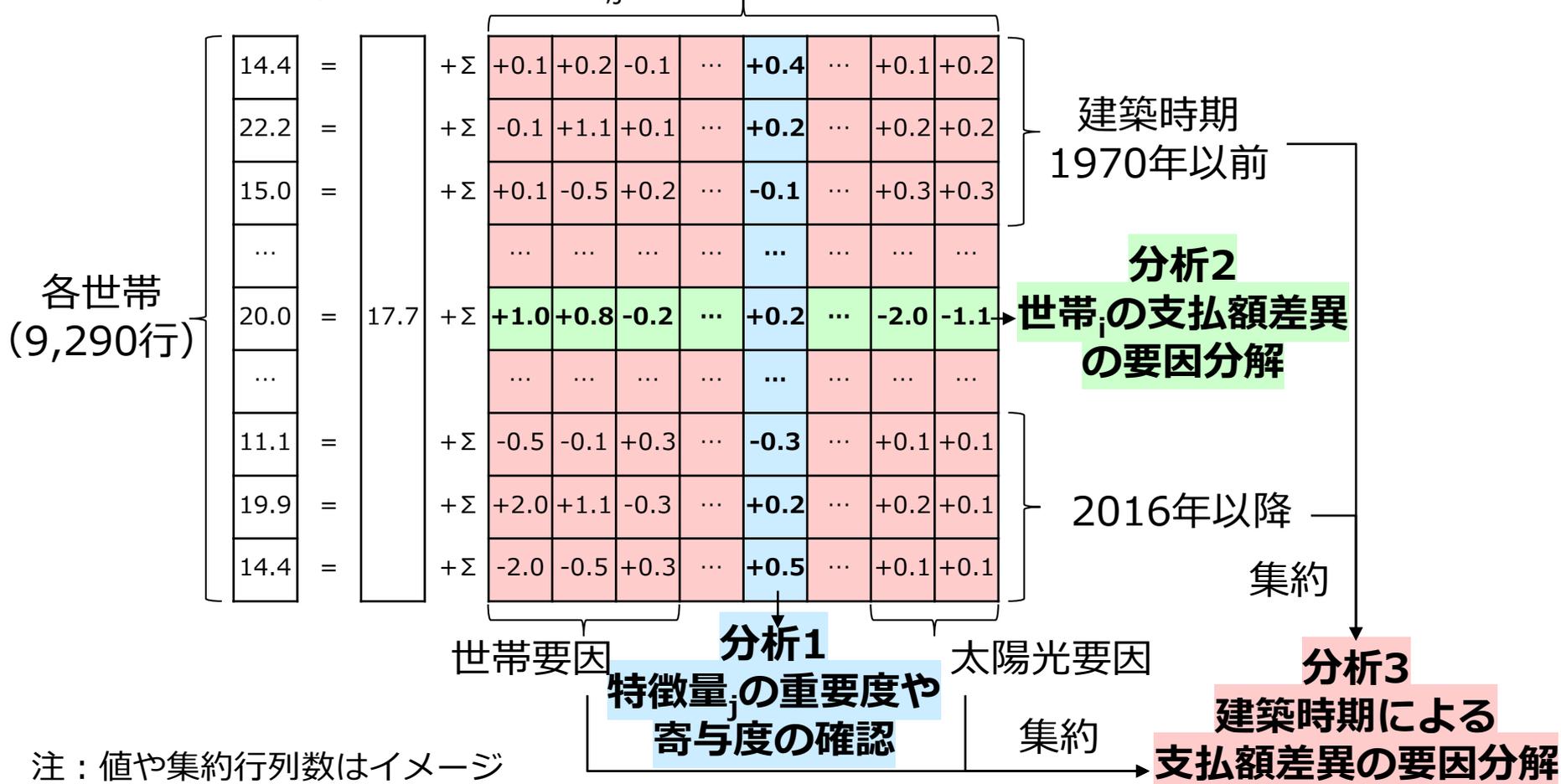
- **複雑なモデルの予測の解釈手法**として2017年に提案されたもので、機械学習のブラックボックス問題への対応手段を与えてくれる



画像出所 <https://github.com/slundberg/shap>

Step 2&3 モデル説明から要因分解へ

- 予測支払額の規定要因を、様々な角度から考察できる
 予測値_i 平均値 SHAP_{i,j} 各特徴量 (244列)



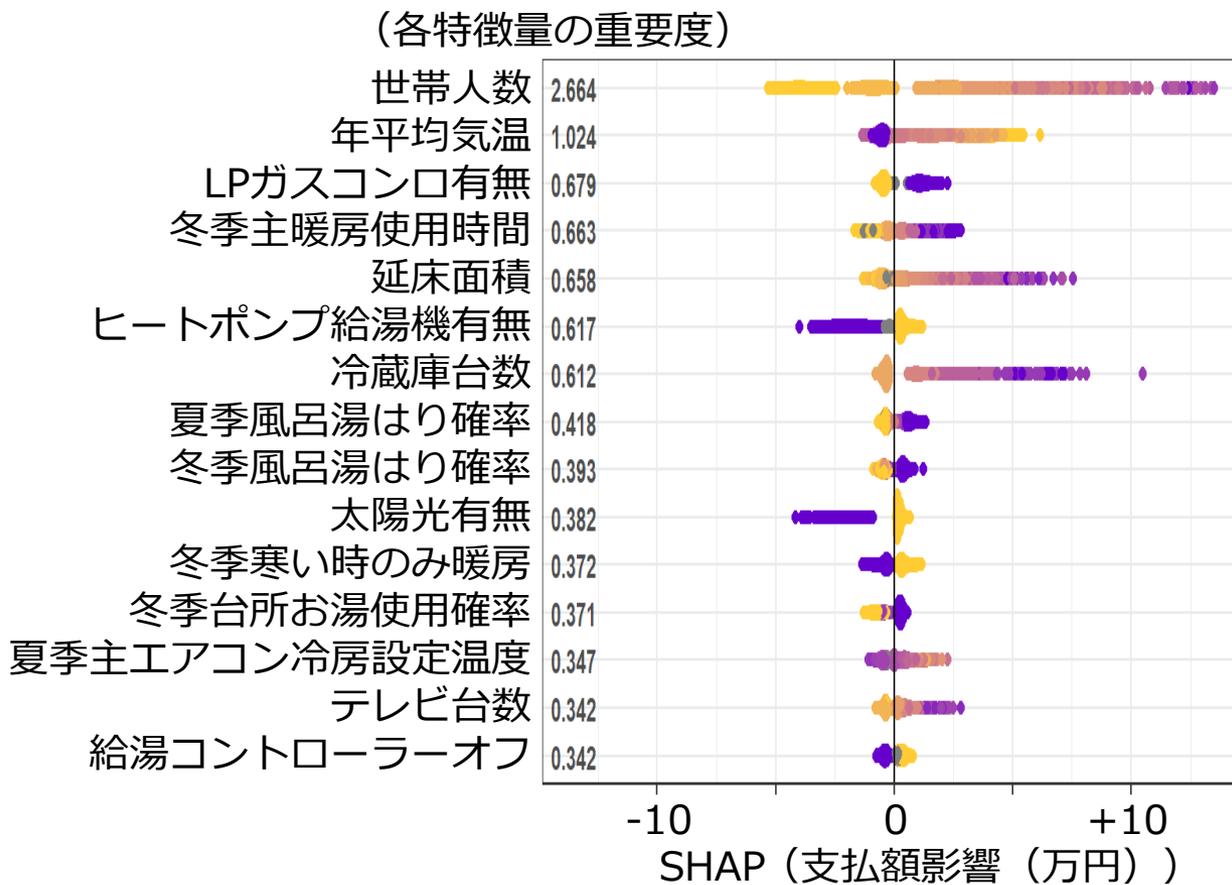
注：値や集約行列数はイメージ

分析①

特徴量の重要度や寄与度の確認

特徴量（上位15個）の重要度

- 支払額を予測する上で、どの特徴量が重要かを確認できる

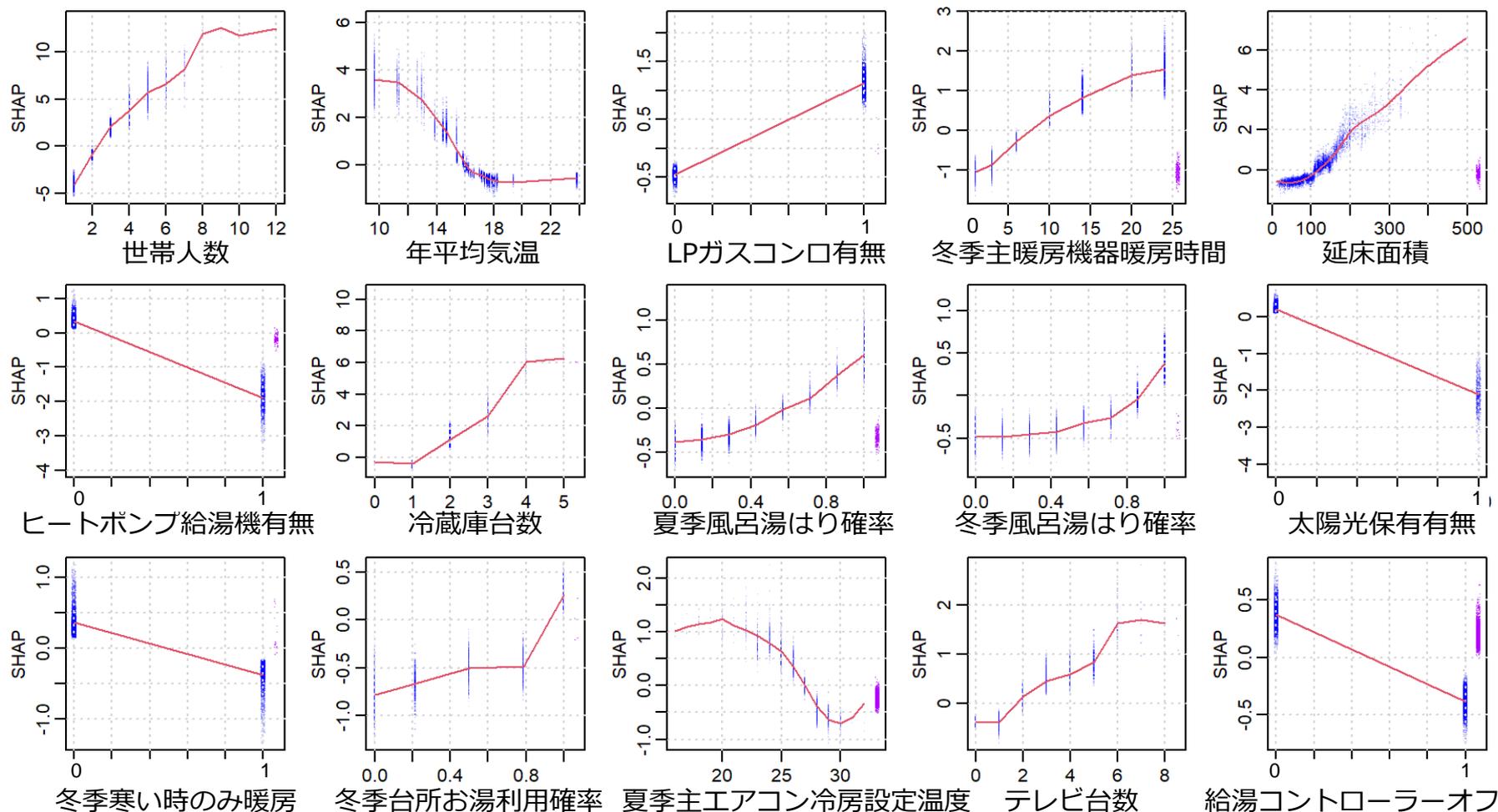


例
 世帯人数が多いほど支払増
 年平均気温が低いほど支払増

各特徴量の水準：低 高

特徴量（上位15個）の寄与度

■ 特徴量がどのように寄与していたかを確認できる（例：非線形性）



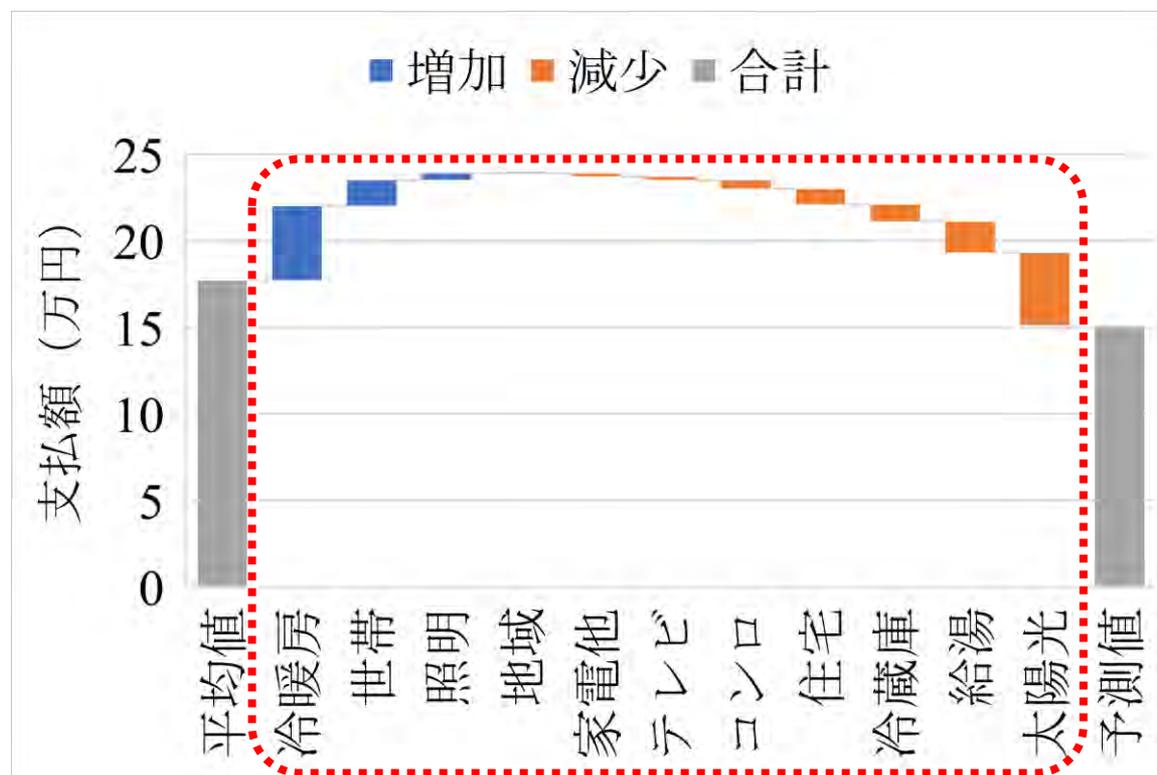
注：縦軸は支払額影響（万円）、横軸は各特徴量の値、右端は欠損値、青点は各サンプル、赤線は平滑曲線

分析②

世帯の支払額差異の要因分解

平均値との差異の要因分解例

- 予測支払額が平均を上（下）回る理由を、世帯ごとに説明できる

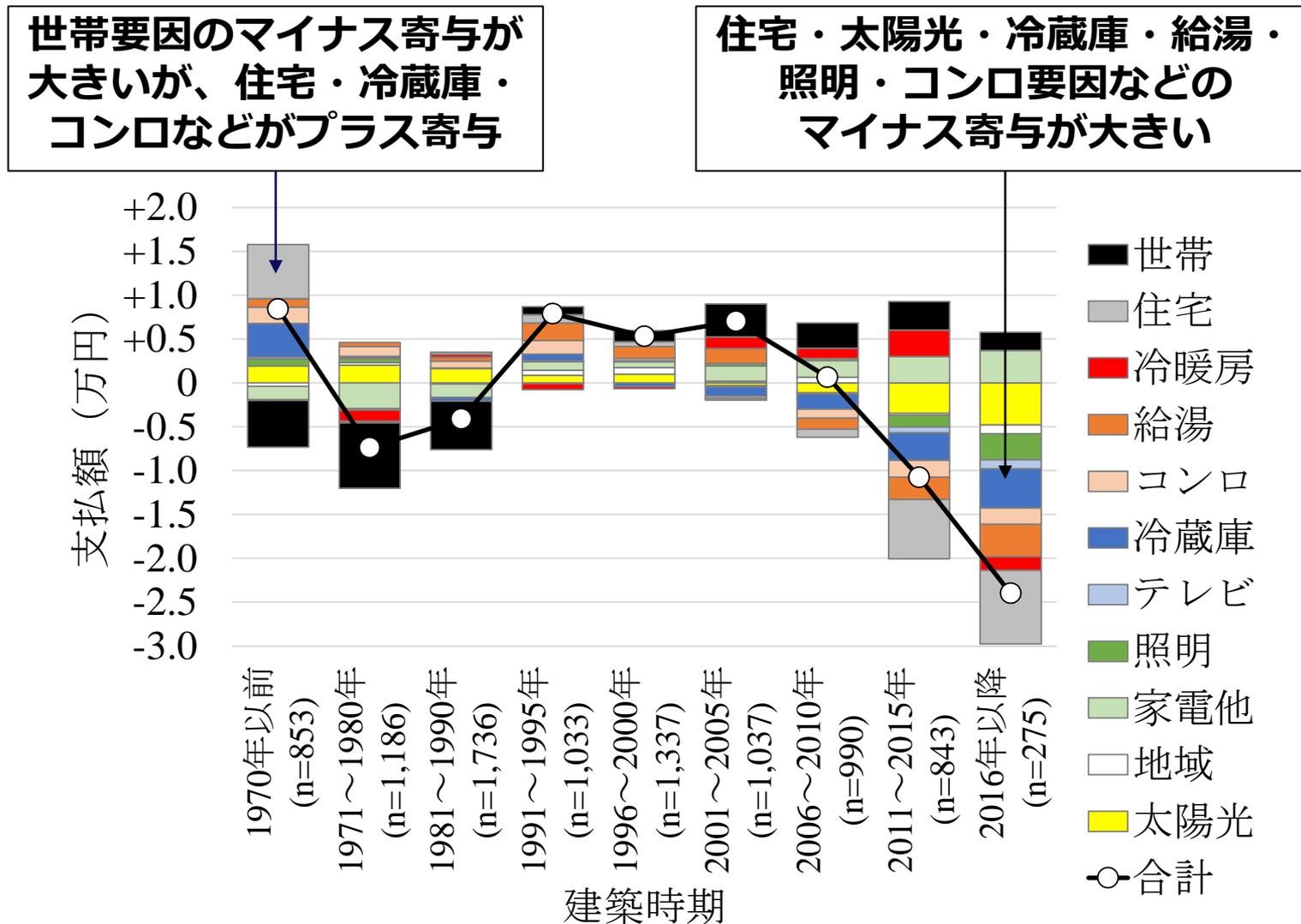


省エネ情報提供に活用できれば、周囲比較にとどまらず、省エネ余地の気づきへ

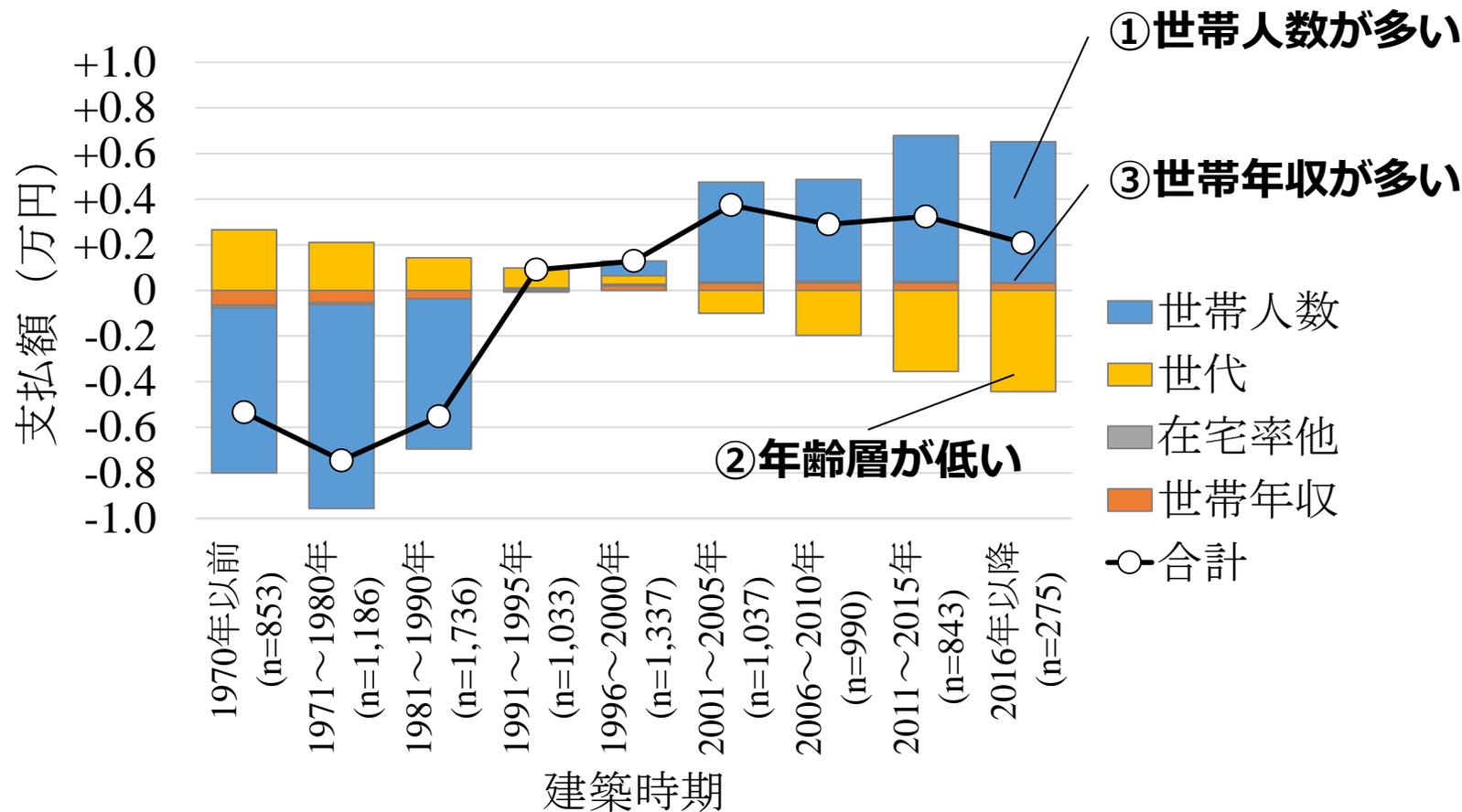
分析③

建築時期による支払額差異の要因分解

光熱費支払額の差異の要因分解（全体像）

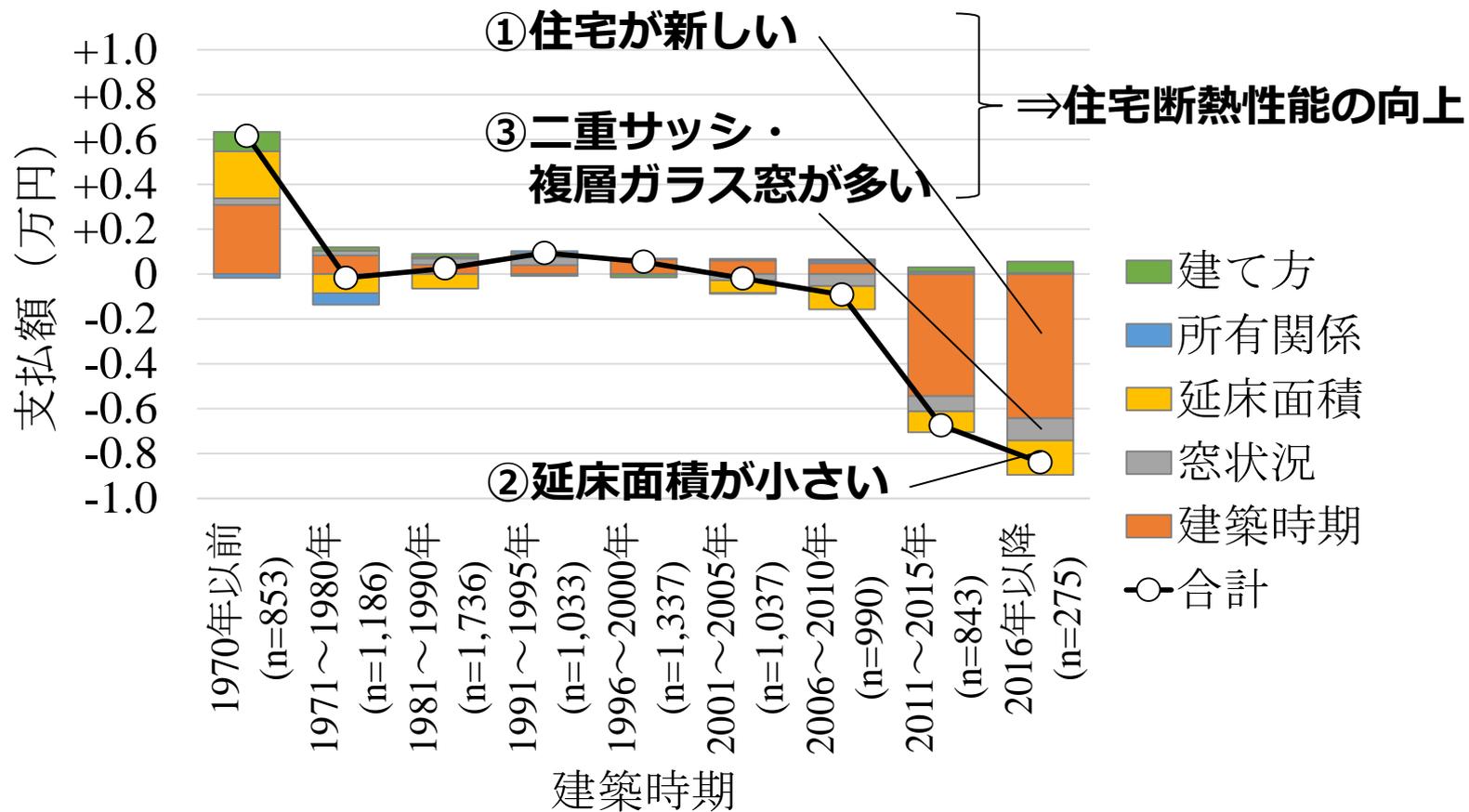


世帯要因



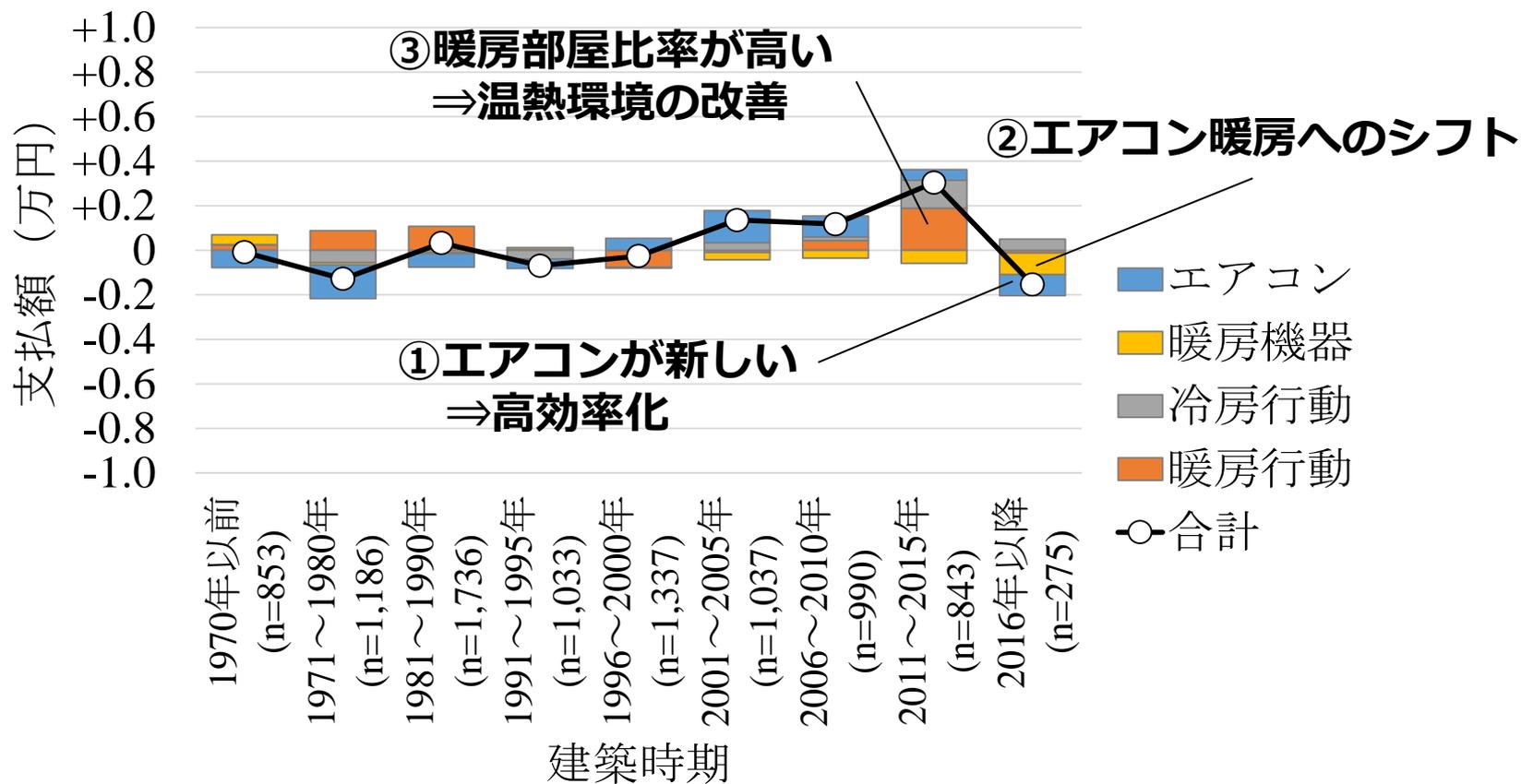
注 図中の説明は、近年の住宅における主な傾向を示したもの

住宅要因



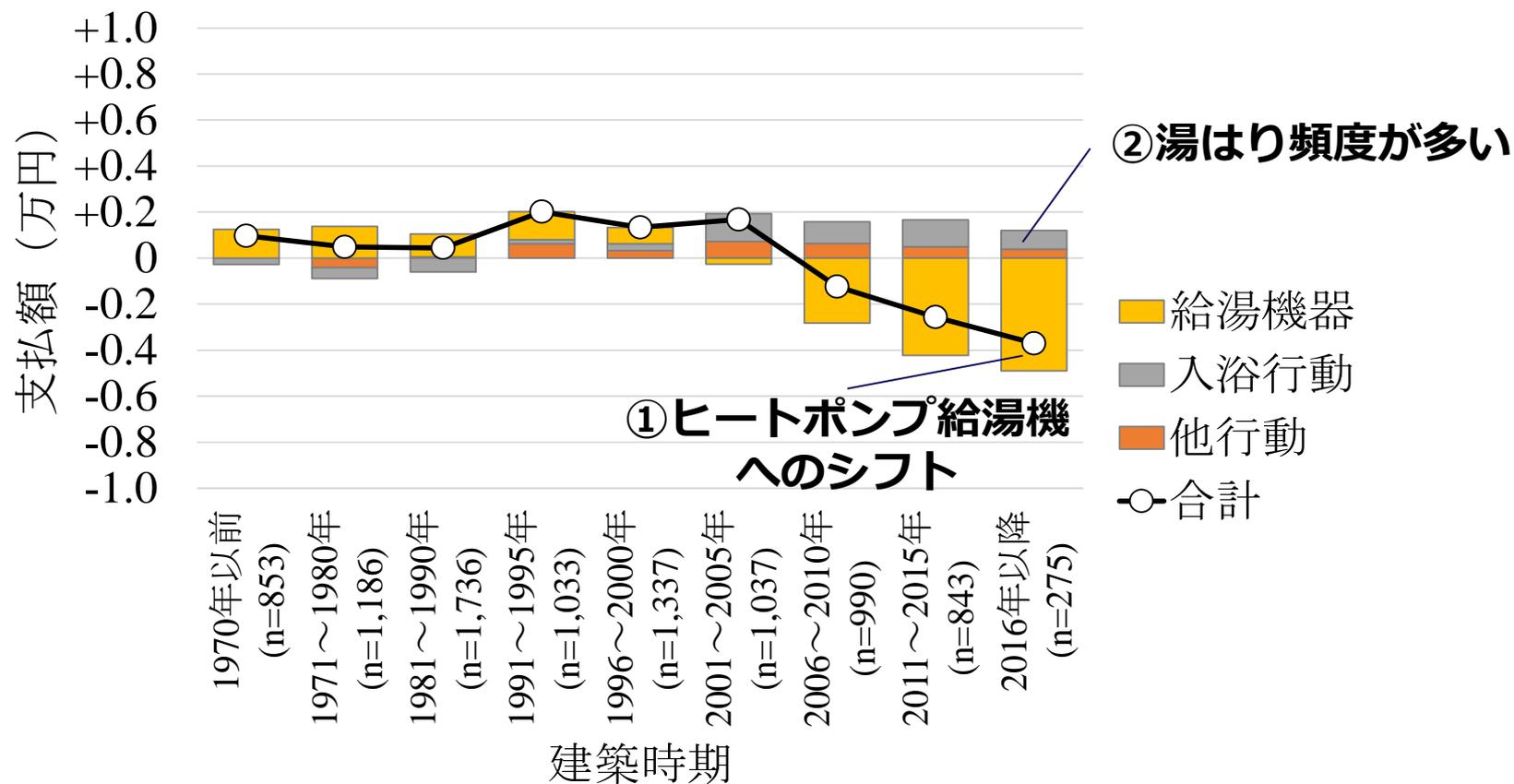
注 図中の説明は、近年の住宅における主な傾向を示したもの

冷暖房要因



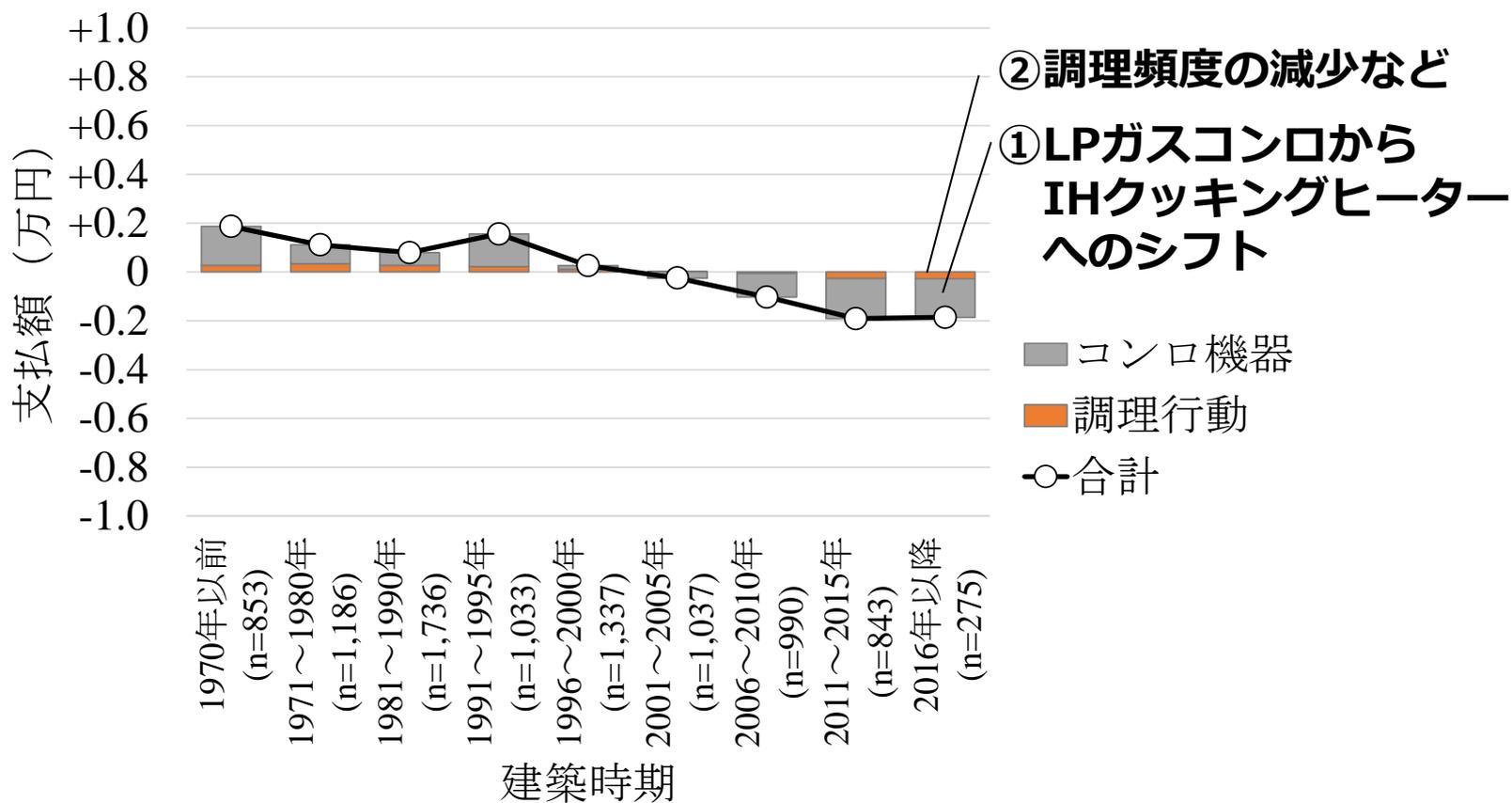
注 図中の説明は、近年の住宅における主な傾向を示したもの

給湯要因



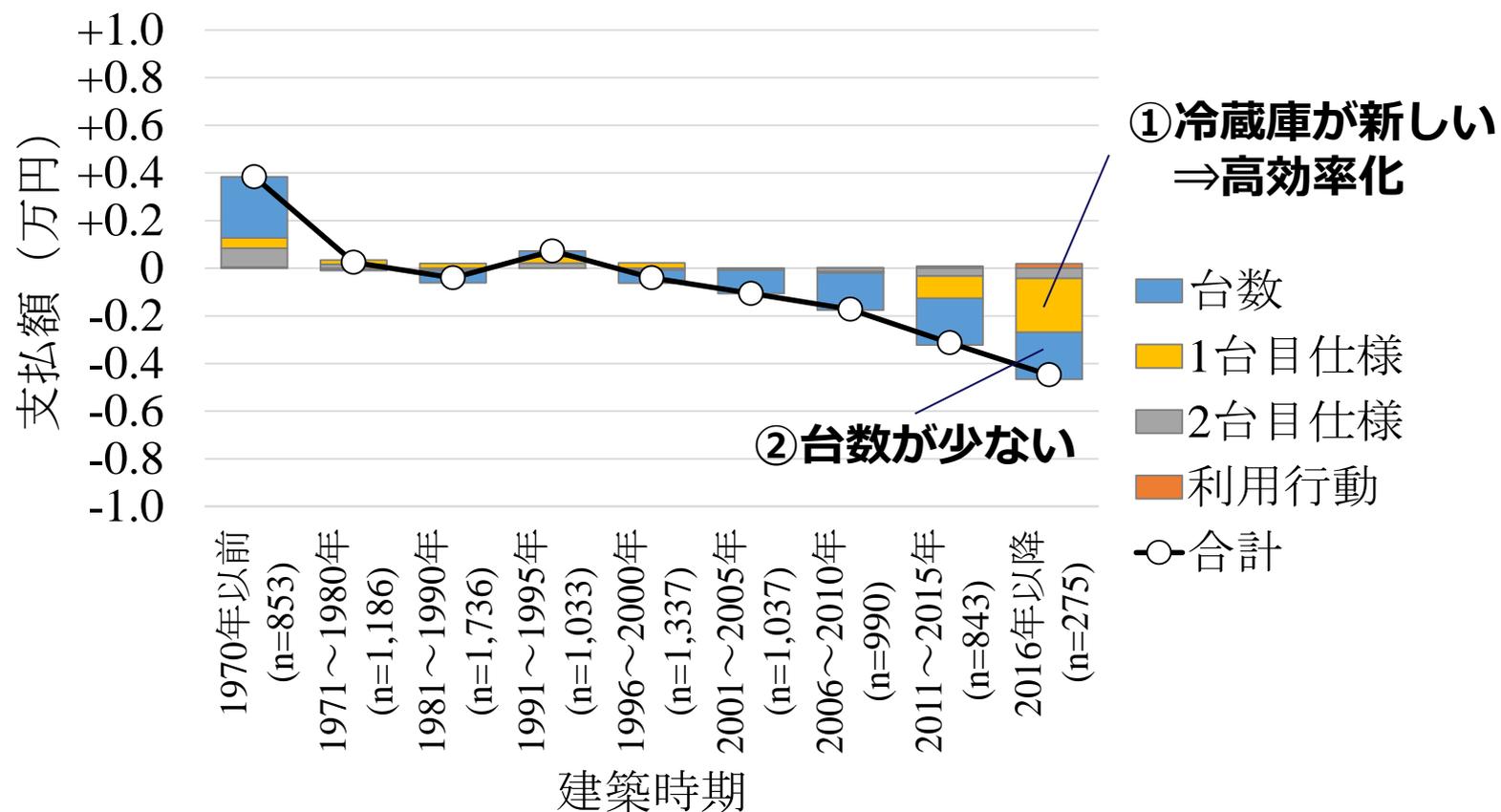
注 図中の説明は、近年の住宅における主な傾向を示したものの

コンロ要因



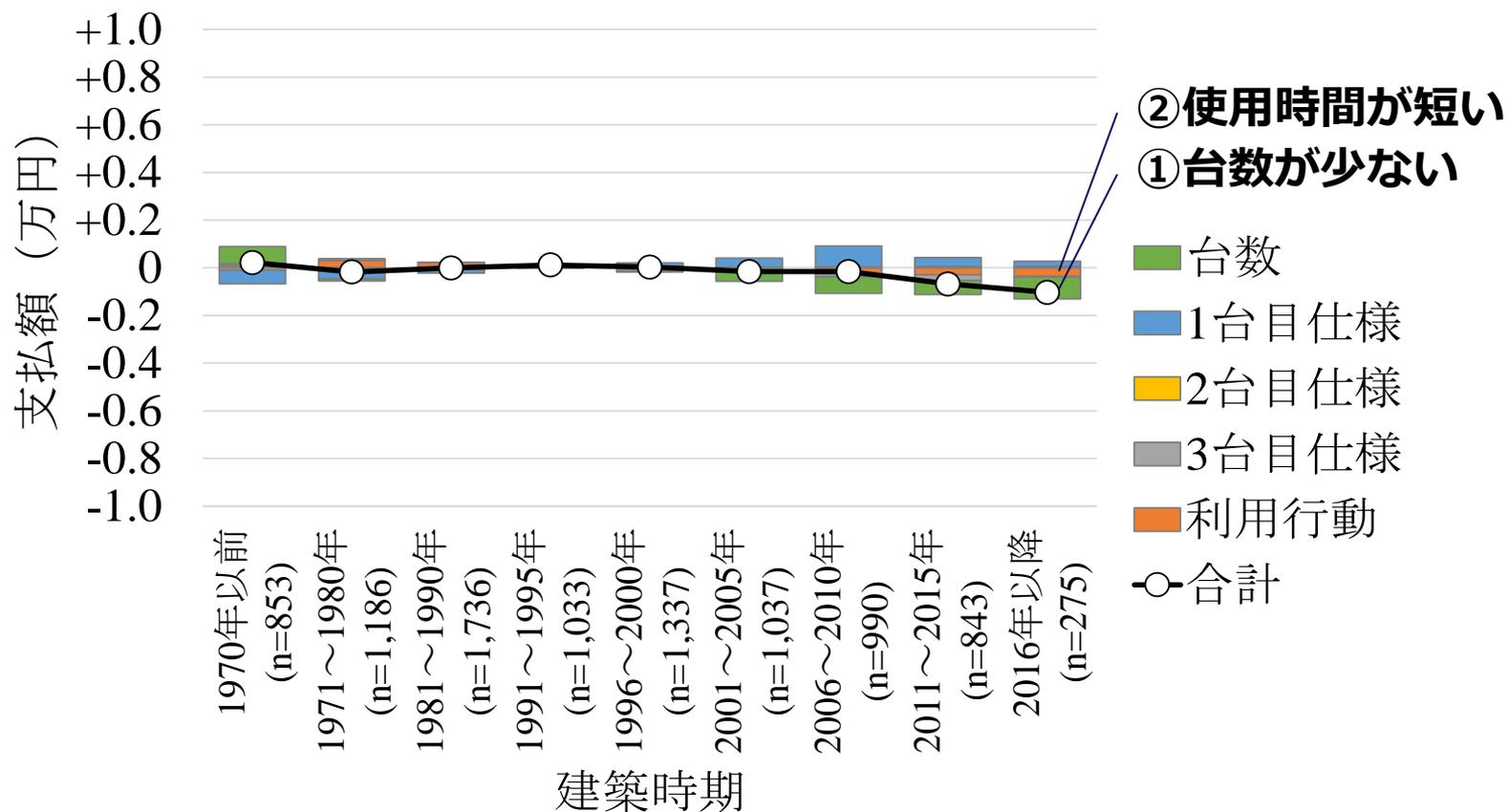
注 図中の説明は、近年の住宅における主な傾向を示したもの

冷蔵庫要因



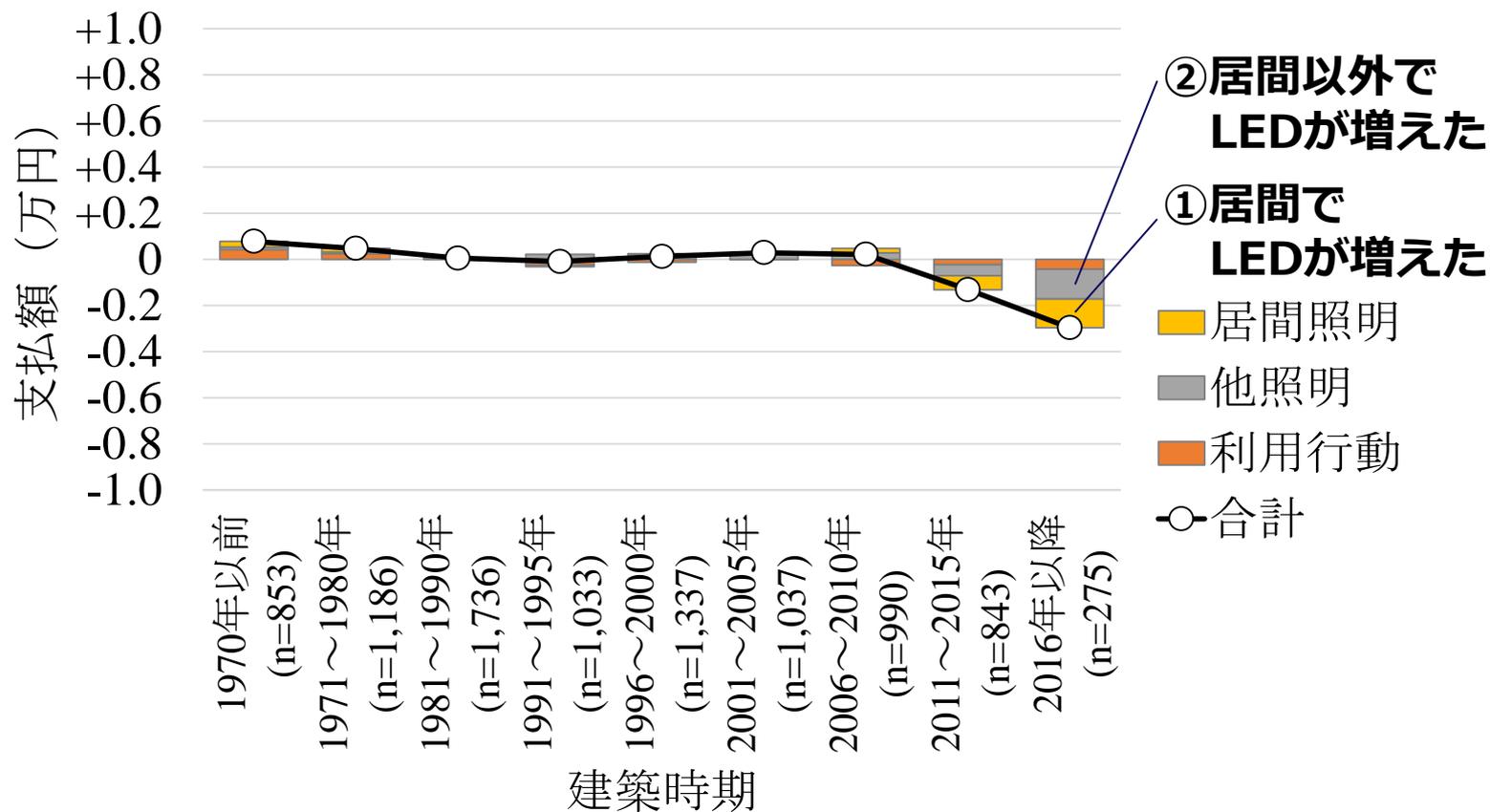
注 図中の説明は、近年の住宅における主な傾向を示したものの

テレビ要因



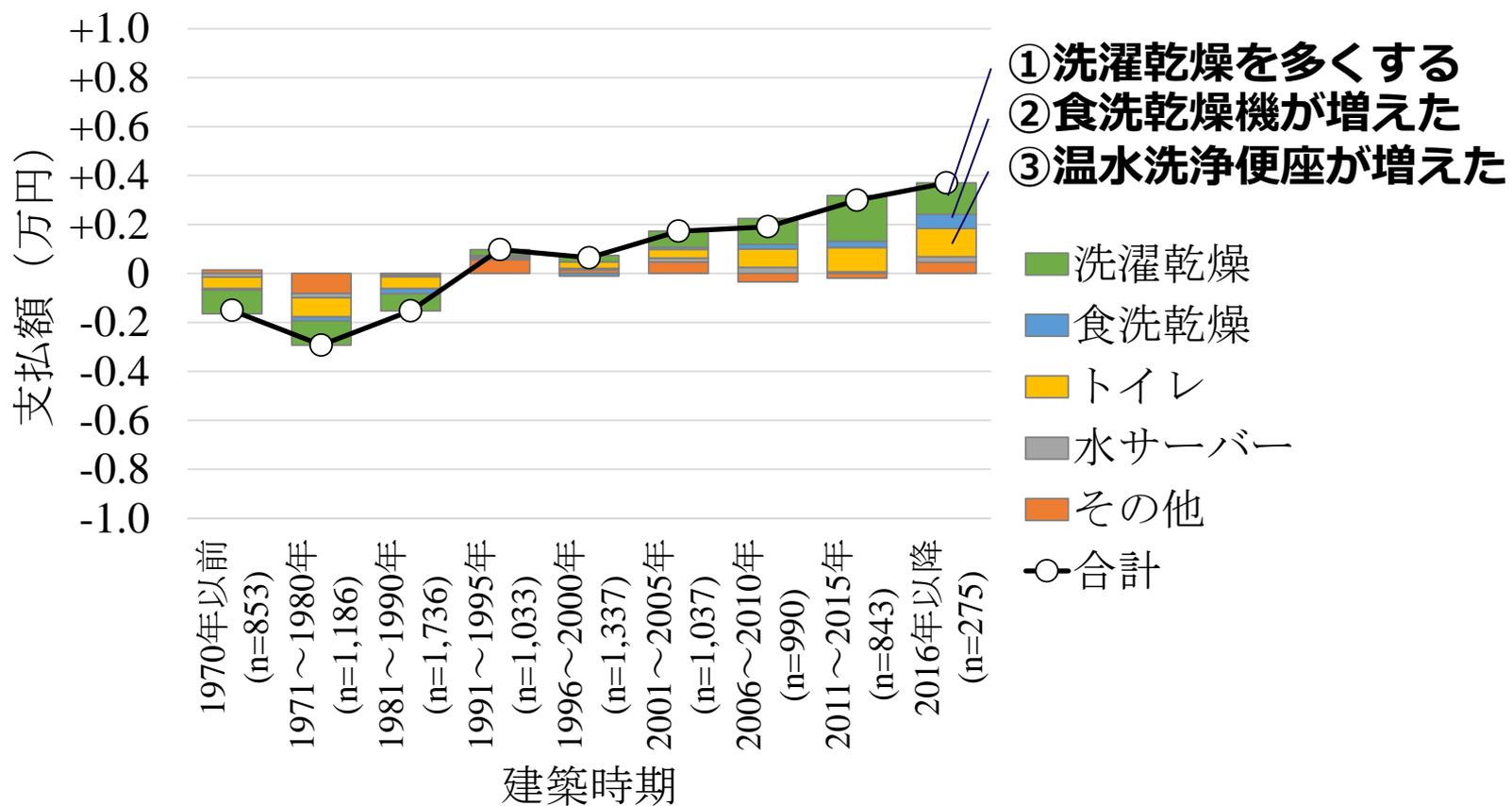
注 図中の説明は、近年の住宅における主な傾向を示したもの

照明要因



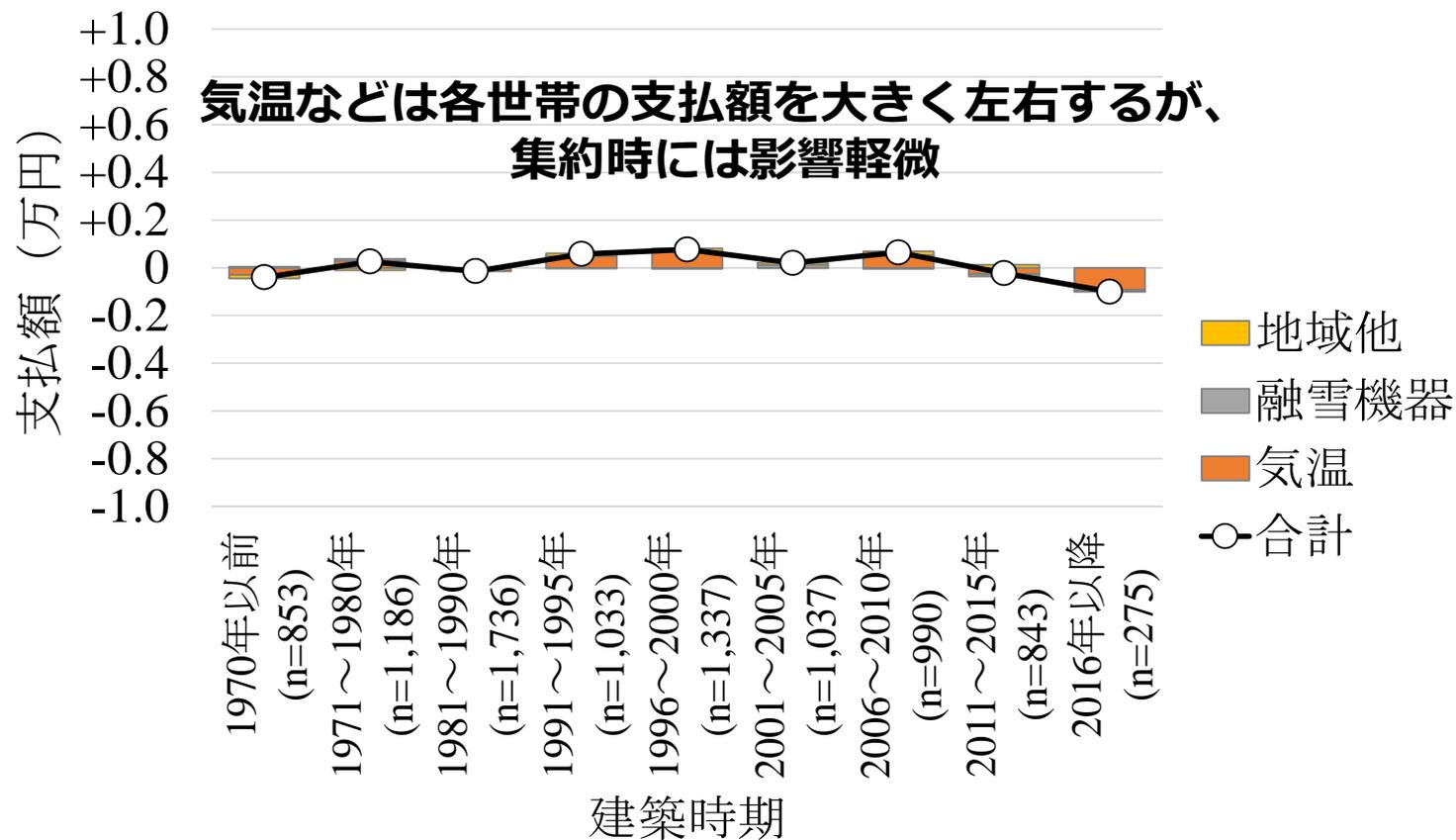
注 図中の説明は、近年の住宅における主な傾向を示したもの

家電他要因



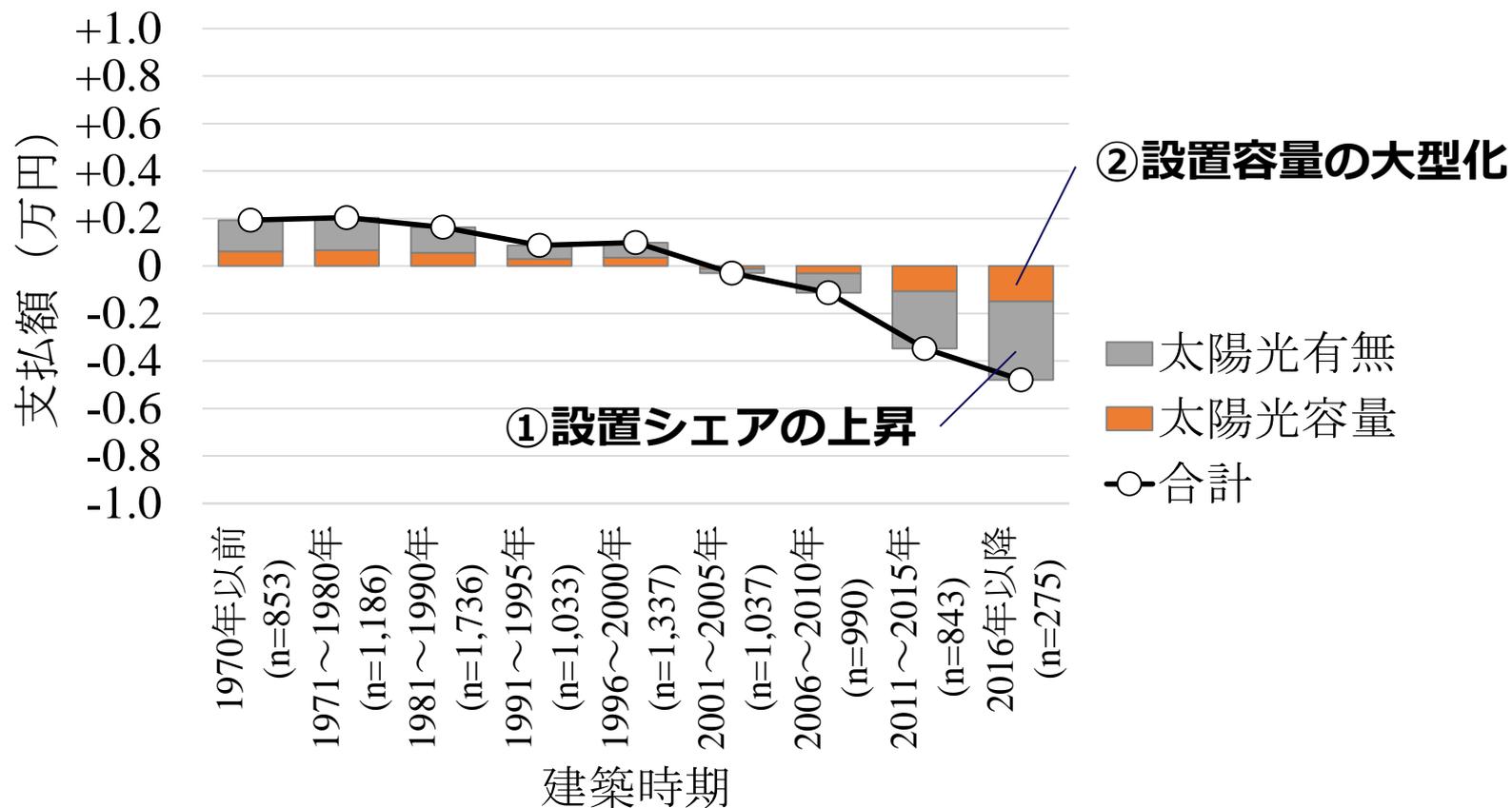
注 図中の説明は、近年の住宅における主な傾向を示したもの

地域要因



注 図中の説明は、近年の住宅における主な傾向を示したもの

太陽光要因



注 図中の説明は、近年の住宅における主な傾向を示したもの

まとめ

まとめ

- 平成30年度家庭CO₂統計の個票データ約1万件を用いて、**近年建てられた住宅で光熱費が抑えられている要因**を明らかにした

手法

- Step 1 機械学習手法の**勾配ブースティング木**により、精度の高いモデルを構築
- Step 2 モデル解釈手法の**SHAP**により、特徴量の寄与度をマイクロレベルで同定
- Step 3 SHAPを活用した**要因分解**により、マクロレベルで実態把握

分析に一貫性や柔軟性をもたせやすく、大規模データや最新要素技術との相性もよい

分析結果

- ✓ **住宅断熱性能の向上**に加えて、**ヒートポンプ機器の普及**や、**LPガスコンロからIHクッキングヒーターのシフト**などにより光熱費が抑えられていること、**冷蔵庫・照明の高効率化**などの効果も着実に現れていることが確認できた
- ✓ 暖房行動や給湯行動の変化により**サービス水準が向上**していることが示唆され、省エネの進展と同時に、**リバウンド効果**や**ライフスタイルの変化**などについても今後注視が必要であることが示唆された

本研究では光熱費や建築時期に注目したが、これに限らず実態把握のニーズに応じていく