

## 住宅設備と賃料\*

賃貸住宅の詳細構造に対する推定

### Housing Facilities and Housing Rent: Estimation of microstructure of rental housing

西颯人†・浅見泰司‡・清水千弘§

Hayato Nishi, Yasushi Asami, Chihiro Shimizu

April 30, 2018

#### Abstract

This paper reveals the effect of housing facilities on housing rent prices in Tokyo, Japan. In previous research which applies hedonic models to housing market, theoretical and estimation problems are pointed out (Ekeland, Heckman, & Nesheim, 2004). But nowadays information technology (IT) drastically makes these problems' structure change, and will solve some of them, e.g. information asymmetry. We focus on housing facilities which did not get much importance from researchers in housing selection process or hedonic pricing in the previous research. It means that these variables might make omitted variable bias in previous research. Our analysis shows that housing facilities significantly influence housing rents. First, inclusion of variables on housing facilities (especially related to structure) improves hedonic model. Second, if vital variables in hedonic model are omitted, omitted variable bias occurs. Third, vital variables in hedonic model tend to be biased if other variables are omitted.

**Keywords:** hedonic model; omitted variable bias; housing facilities; IT (information technology); housing selection

**JEL Classification:** R31 - Housing Supply and Markets, R32 - Other Spatial Production and Pricing Analysis.

---

\*本研究の実施にあたり、スタートコーポレーション副会長 関戸博高氏はじめテックデザインラボの皆様からは、データの提供のみならず多くの示唆をいただいた。ここに記して御礼申し上げます。

† 東京大学大学院工学系研究科都市工学専攻 修士課程

‡ 東京大学大学院工学系研究科都市工学専攻 教授

§ 日本大学スポーツ科学部 教授, 東京大学空間情報科学研究センター客員研究員

## 1. 住宅設備とヘドニックアプローチ

住宅サービスに対する消費は、家計の支出のおおよそ四分の一を占めることから、その消費の選択によっては、それぞれの家計の効用水準を大きく変化させてしまうといっても過言ではない。家計は住宅選択の際に、住宅の品質と価格との対応を吟味しながら、その効用が最大化できるように行動する。

このような家計の選択行動と、住宅の機能と価値とを説明する経済理論として、ヘドニック理論がある。伝統的な価格理論では、一物一価の法則が市場分析を行う上での有効な仮定となるが、Lancaster (1966)が分析しているように、この仮定は差別化された商品を扱う上で理論的に（そして実証分析を行う上でも）きわめて不都合である。そこで、Rosen (1974)はこのような属性の束としての商品価格データが、どのような市場メカニズムで発生するのかを理論的に解明した Rosen の研究は、商品供給者のオファー関数 (offer function)、商品需要者の付け値関数 (bid function) およびヘドニック関数の構造ととの間の関係を厳密に検討し、商品の市場価格を消費者および生産者の行動から特徴づけている。

しかし、住宅市場を対象としたヘドニックモデルの適用においては、既存研究では様々な問題が指摘されてきた。この問題は、大きくは理論的な問題と推計論上の問題に大別される。理論的には、住宅市場は完全ではないために、ヘドニック理論が示す完全競争市場下でのオファー関数と付け根関数の交点から形成される市場価格関数が、完全市場価格から乖離しているのではないかという問題である。具体的には、売り手と買い手の情報の非対称性が大きく、供給者、所有者の市場行動において高いコストが存在しているということが明らかにされてきた (Shimizu, Nishimura, & Asami, 2004)。また、取引、つまり均衡状態に到達するまでに時間がかかることも知られている (Genesove & Mayer, 2001)。

実証分析においては、住宅価格関数を推計しようとしたときに、真のモデルがわからないため、また情報収集技術に強い制約があったことで、過少定式化バイアスの問題が存在するのではないかということが指摘されてきた (Ekeland, Heckman, Nesheim, 2004)。さらには、従来推計されてきたヘドニックモデルの多くは均質の家計を想定しているものの、多様な家計が存在することで、その関数形に非線形性が存在するために、家計の特性をも考慮したうえでヘドニック関数を推計していくことの必要性も指摘されてきた。この問題は、住宅ストックの多様性にも関係する。多様な選好を持つ家計が存在したとしても、市場で流通している住宅ストックにも多様性がなければ、消費者は強い制約の中で住宅選択をしなければならなくなる。また、市

場が流通している住宅の情報が多様に品質が掲載されていなければ、買い手はその品質を調べるために膨大なコストがかかったり、そのような選好を顕示できる機会を失ったりしてしまうという問題があった。

しかし、近年においては、IT, IoT の進化の下で、住宅の情報流通が飛躍的に変化する中で、これらの問題構造が大きく変化しようとしている。Shimizu et al. (2004) で測定された住宅市場の不完全性のコストは、住宅探索行動における情報選択は、新聞、情報誌またはチラシが中心の時代であった。新聞、情報誌、チラシといった情報媒体では、紙面に制約があり、その中には多様な情報を記載できないために、専有面積や建築後年数、最寄り駅とそこまでの距離、都市計画用途制限などの広報上の規制などと限定された情報だけを網羅的に一覧性を持つてみることに過ぎなかった。そのため、住宅の設備や周辺の状況などといった住宅選択においては潜在的に重要であると考えられていた情報であったとしても、消費者は現地に行き確認しなければならなかったのである。また、現在ほどに空き家率は高くなかったために、市場で流通している住宅ストックもより限定的であったと言えよう。

本稿では、首都圏の賃貸住宅市場を対象として、住宅の品質と家賃との対応関係を、ヘドニックモデルを推計することで明らかにしていく。ここで、東京の賃貸住宅市場を対象とすることには、次の意義がある。

第一に、首都圏には 4000 万人の人口を有する世界最大の都市圏であるとともに、持ち家率が 56.2% と自由な賃貸住宅市場が成立している世界でも珍しい市場となる。第二に、現在の賃貸住宅市場の空き家率は 19.2% と高く、買い手優位の市場である<sup>1</sup>。つまり、多様な住宅ストックの中から住宅選択が可能となっているのである。さらに、民間の住宅情報サイトが発達しており、従来のヘドニック関数にしばしば取り入れられていた、規模、ベッドルーム数、建築後年数や立地といった変数だけでなく、様々な設備に関する情報が掲載されるようになっていく。また、その情報の記載には、一定の公的な規制の下で配信されていることから、法で規定された主要な価格形成要因(専有面積、建築後年数、最寄り駅までの時間、都市計画用途制限)などの情報の信頼性は高く、また欠損率も極めて低いという優位性がある。

---

<sup>1</sup> 数値はいずれも平成 25 年住宅・土地統計調査の「関東大都市圏」の値から算出

このような市場の中では、消費者が様々な設備も含む住宅属性を考慮しながら、住宅選択が可能となることから、多様な選好を顕示させることができる市場へと成熟しているものと考えられる。

とりわけ本研究では、住宅の設備に注目する。住宅の設備は多様性を持ち、消費者にとっては一層個別性を有する住宅品質であるとともに、従来はこのような変数が取得できなかったという意味で、観察不可能な変数であった。そのために、ヘドニック関数の中に取り入れることができなかったという課題があったと考える。このような問題以前に、住宅の選択行動の中で、消費者はそのような情報までに到達することができなかったために、付け値関数の中に反映されていなかった可能性も高い。

しかし、住宅選択行動における住宅情報サイトの影響力が高まる中で、従来は考慮されなかった要因、例えば設備などを要素として条件付けして探すことができるようになったことで、専有面積、建築後年数、最寄り駅まで距離などといった従来の主要な住宅の属性情報とともに、賃料に及ぼす住宅設備の影響が相対的に大きくなってきているという仮説を持つに至った。日本の住宅情報サイトでは、検索の条件付けのために様々な住宅設備に関するチェックボックスが用意されており、利用者は住宅設備の有無によって候補を絞り込み、住宅を選択している。

以上を踏まえ、本研究では住宅設備と賃料について以下の2点を明らかにする。

第1に、住宅設備を加えることがヘドニックモデルの向上に寄与することを示す。一般に説明変数の追加は回帰モデルのフィッティングを向上させるが、一方で多重共線性や過剰適合などの問題が生じさせる。本研究ではこれらの問題を回避するため、多重共線性のチェックを行うとともにAICによるモデル比較・モデル選択を行っている。

第2に、住宅設備がモデルから失われることが、他の変数の係数推定値に影響を及ぼすことを示す。過少定式化バイアスに対処する手法としてはRepeat Sales Methodや空間統計などの手法が考えられるが、住宅設備は時間的・空間的な定常性・類似性を必ずしも仮定できないため、こうした手法によって推定バイアスを回避することは困難であると考えられる。よって本研究では、これまで入手困難だった住宅設備に関する情報をモデルに直接投入することで、これによりバイアスを生じさせやすい変数や、過小定式化に対して頑健な変数を明らかにする。またバイアスの具体的な大きさや方向を実証的に提示することで、十分なデータ収集ができなかった際に、ヘドニックモデルの解釈において留意すべき点を示す。

得られた結論によると、第一に、住宅設備を考慮することによってヘドニックモデルの予測向上が見られる。また住宅設備の中では、住宅の構造に関する変数は賃料への影響が大きいことが明らかになった。第二に、モデルの予測向上や賃料への影響が大きい変数（例えば専有面積や築年）が失われた場合に、過少定式化バイアスが発生しやすいことが明らかになった。第三に、賃料への影響が大きい変数（例えば専有面積や東京距離）は過少定式化バイアスの影響を受けやすい傾向が見られるため、十分な変数のデータを集められなかった場合にはモデルの解釈に注意が必要である。

## 2. 先行研究

住宅市場の分析において、ヘドニック分析は広く用いられてきた(Chau & Chin, 2003)。住宅は様々な特性の複合物であり、ヘドニック分析ではこれらの特性の住宅価格に対する限界潜在価格を算出する(Sirmans, David, & Emily, 2005; Yoo & Wagner, 2016)。ヘドニック分析において考慮される住宅特性は、以下のように分類される。

Nelson (1972)は、住宅の位置的な特性はアクセシビリティにかかわるものを除き、「住宅設備」と「位置的有利性」の2つに分類できるとしている。このうち、「住宅設備」は土地所有者の行動によって決定するもので、「位置的有利性」は土地所有者の行動によって決定できないものである。一方、Chau & Chin (2003)はヘドニック価格モデルに通常使用される特性を、「位置的特性」「構造的特性」「近隣特性」の3つに分類している。Sirmans et al. (2005)は約125のヘドニックモデルをレビューし、そこに含まれる特性を8つのカテゴリ、「建築構造変数」「住宅内部特性」「住宅外部特性」「自然・環境特性」「環境・近隣・位置因子」「公的サービス」「マーケティング・占有・販売因子」「財務問題」に分類し、リストアップしている。また、Yoo & Wagner (2016)は住宅を構成する特性として、「構造変数」「近隣変数」「環境アメニティ」の3つを挙げている。

これらの研究をもとに、ヘドニック分析に用いられる住宅特性 (housing amenities) を以下のように分類した。本研究ではこのうち、住宅設備を主な研究対象として設定する。

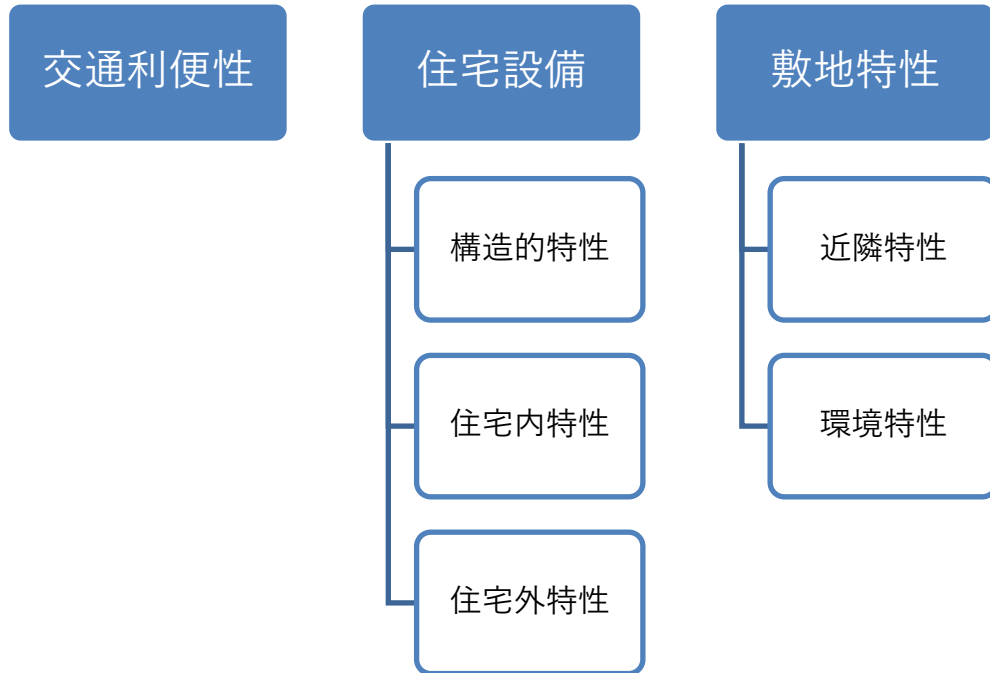


Figure 1：住宅特性の分類

本研究では、交通利便性(Accessibility)を「通勤通学に関わる主要な交通利便性」と定義する。都心までの距離に加え、東京では通勤通学手段として公共交通機関の利用が支配的であるため、これらに関わる特性が交通利便性として分類される。住宅設備(Housing Facilities)と敷地特性(Site Advantages)は Nelson (1972) に従い、交通利便性を除く特性のうち「土地所有者の決定によって変更可能なもの」と「そうでないもの」と定義する。これらの下位分類は (Chau & Chin, 2003; Sirmans et al., 2005; Yoo & Wagner, 2016) をもとに作成した。

交通利便性と住宅設備はヘドニックモデルの基本的な変数であり、多くの研究で使用されている。Sirmans et al. (2005) が示した「ヘドニック価格モデル研究で最も多く登場する 20 の特性」リストの多くは、交通利便性または住宅設備に該当する変数である。また住宅設備は時間経過に伴う劣化やリノベーションなどによって変化し、賃料に対して影響を及ぼすことが考えられる (Billings, 2015; McMillen & Thorsnes, 2006; Shimizu & Karato, 2015)。住宅設備は土地所有者が変更可能な住宅特性であることから、経営上・政策上の重要性が高いと考えられるが、知る限りにおいてこれらの特性を研究の主題として、幅広くモデルに組み込んだ研究は存在しない。

一方で、近年の研究では敷地特性に分類される特性の研究が盛んである。従来敷地特性に関する情報は取得するのが困難であったが、現在では GIS で特性を定量化するなどして直接的にモデルに組み込んだり、あるいは空間的特性を考慮したモデルによって間接的にモデルに組み込んだりすることが可能である(Shimizu, 2014)。

直接的にモデルに組み込んだ研究としては、Gao & Asami (2001), Kang (2017), Shimizu (2014)が挙げられる。Gao & Asami (2001)は戸建て住宅周辺の道路や公園の影響を分析している。また Shimizu (2014)は GIS を使用して近隣環境を示す変数を算出し、これらの変数がヘドニックモデルを向上させることを示した。Kang (2017)は徒歩圏の土地利用混合を表す指標をネットワーク解析によって複数算出し、住宅価格に対する影響を比較検討している。一方で Yoo & Wagner (2016)は、SLM, SEM, SAR, GWR などの空間的特性を考慮した統計モデルを使用した研究をレビューし、伝統的なモデルを使用した研究と比較している。また空間的特性を考慮したモデルではないが、Yoo, Im., & Wagner (2012)は機械学習手法の適用によって、住宅販売価格の予測残差の空間自己相関が低減することを報告している。

しかしながら研究に使用できるデータには、詳細な住宅の特性が含まれていないことが指摘されている(Billings, 2015)。このようなデータの制約がある場合、ヘドニック分析を行うと過少定式化バイアスが生じる恐れがある(Shimizu, 2014; Shimizu & Karato, 2015)。

この問題に対処する1つの方法として、Repeat Sales Method が挙げられる。Repeat Sales Method では、複数回売買されたデータのある物件を分析対象として、2時点における取引価格の差を算出する。これにより、交通利便性や住宅設備は時間経過で変化しないと仮定すれば、これらの特性をモデル式から除外することが可能である(McMillen & Thorsnes, 2006)。しかしリノベーションなどによって住宅設備が変化した場合、この手法では分析に不都合が生じる(Billings, 2015; McMillen & Thorsnes, 2006; Shimizu & Karato, 2015; Wong, Chau, Karato, & Shimizu, 2017)。

他の方法としては、空間的特性を考慮したモデルを用いること方法が挙げられる(Abbott & Klaiber, 2011; Shimizu, 2014)。手法は空間的に近い物件が近い特性を持っていることを仮定するため、敷地特性や交通利便性の統制には有効であると考えられる。しかし住宅設備は物件個別の特性であることから、必ずしもこれらの手法によって十分に統制できるとは言えない。

Web データを使用した研究としては、Boeing & Waddell (2016), Miura & Asami (2012)が挙げられる。Miura & Asami (2012)はブログから地域の情報を収集してヘドニックモデルに組み込んでいる。Boeing & Waddell (2016)は Craigslist から全米のデータを収集し、その有効性を示

している。これらの研究では、詳細・広域・最新の情報を Web データの利用によって分析している。

### 3. 住宅設備を考慮したヘドニックモデル

#### 3.1. 基本モデル

本節では、住宅設備を説明変数に加えることで、ヘドニックモデルの向上を試みる。一般的なヘドニックモデルは

$$y = \log p = \sum_i \beta_i x_i + \alpha \quad (1)$$

と表される。ここで、 $y$ は被説明変数、 $p$ は住宅賃料、 $x_i$ は説明変数（住宅特性）、 $\beta_i, \alpha$ は推定パラメータである。ただし、連続変数はその標準偏差で除することにより標準化を行った。これによりこのモデルは、説明変数 $x_i$ が1増加する<sup>2</sup>と賃料が $e^{\beta_i}$ 倍になるモデルであると解釈できる。

前章で述べた住宅変数の分類に従い説明変数を分けると、式(1)は以下のように書き直せる。

$$y = \sum_j \beta_j^{AC} AC_j + \sum_k \beta_k^{SF} SF_k + \sum_l \beta_l^{IH} IH_l + \sum_m \beta_m^{EH} EH_m + \sum_n \beta_n^{SA} SA_n + \alpha \quad (2)$$

各シンボルの意味は次節の表で示す。

本研究においては、投入する変数群を変えて以下のような6つのモデルを設定・推定し、比較を行う。

---

<sup>2</sup> これは説明変数が連続変数の場合は標準偏差の分だけ増大することを意味し、ダミー変数の場合は0から1になることを意味する。



Table 1: ヘドニックモデル

	Model 0	Model 1	Model 2	Model 3	Model 4	Full Model
交通利便性	○	○	○	○	○	○
住宅設備						
構造的特性	×	○	×	×	×	○
住宅内特性	×	×	○	×	×	○
住宅外特性	×	×	×	○	×	○
敷地特性	×	×	×	×	○	○

○：モデルに特性を含む    ×：モデルに特性を含まない

Model 0 は交通利便性のみを説明変数に含む基本モデルである。このモデルは、通勤通学における交通利便性のみで住宅賃料が決定されると仮定している。Model 1～Model 3 は交通利便性に加えて住宅設備の各下位分類をそれぞれ考慮したモデルである。これらは交通利便性と建物の特性から賃料を説明している。Model 4 は交通利便性と敷地特性を説明変数とするモデルであり、賃料が建物特性とは無関係に立地のみで決定されるとするモデルである。最後の Full Model は、すべての説明変数を含むモデルである。ただし、専有面積は最も基本的な変数の1つであるため、すべてのモデルに投入した。

一般に説明変数の増加はデータに対するモデルのフィッティングを良くするが、一方で多重共線性や過剰適合による汎化誤差の低減などの問題を引き起こす。よって、変数間の相関を確認するとともに、最も多くの説明変数を含むモデル(Full Model)について VIF を算出して多重共線性の発生を防いだ。また説明変数の数の増加による影響を差し引いて予測精度の比較を行うため、AIC によるモデル比較を行う。よって Full Model ではすべての変数を含むのではなく、AIC を用いた変数の逐次増減による変数選択を行った。

## 3.2. データ

本研究で扱うデータは東京距離<sup>3</sup>を除き、スタートが提供したピタットハウスに掲載された1都3県の賃貸住宅情報<sup>4,5</sup>である。ピタットハウスは日本国内で最もポピュラーな住宅情報サイトの1つであり、提供されたデータにはおよそ36万件の物件が含まれている。

管理費（共益費）と礼金の影響を加味するため、住宅の賃料に管理費（共益費）と礼金の24分の1を加えて補正を行った。住宅設備の賃料に対する一般的な影響を明らかにするため、分析に先立ちデータから外れ値を除外した<sup>6</sup>。これに加え欠損のあるデータの除去<sup>7</sup>を行ったことにより、データ数は236,411件に減少した。

---

<sup>3</sup> 東京距離は、各物件からの東京駅までの直線距離を、同データの住所情報からGISを用いて算出した。

<sup>4</sup> バスに乘車することへの心理的な負担を加味するため、バスダミーを作成して変数に追加した。

<sup>5</sup> 建物構造は「木造（2×4）」「木造（在来）」「軽量鉄骨造」「重量鉄骨造」「鉄筋コンクリート造」「鉄骨鉄筋コンクリート造」「プレキャストコンクリート造」「その他」に分類されている。これらすべてをダミー変数にすると変数が増えるため、RC/PCと鉄骨造に集約した。

<sup>6</sup> 竣工していない物件を除外した上で、HotellingのT<sup>2</sup>統計を用いて異常度の累積確率が99%以上のものを除外した。これに加え、賃料の上下0.1%・専有面積が15㎡未満ないし100㎡超・所在階が非正のいずれかに該当するデータを外れ値とした。

<sup>7</sup> 公的規制の対象となっていない項目に関しては、データの欠損が多く見られた。本研究では多様な住宅設備の影響を観察するため、欠損のあるデータに関しては除外した。

Table 2: データリスト

文字	変数名	内容	単位
<b>被説明変数</b>			
<i>p</i>	補正賃料	賃料+管理費（共益費）+礼金/24	円
<b>交通利便性</b>			
<i>AC</i> <sub>1</sub>	東京距離	東京駅までのユークリッド距離	m
<i>AC</i> <sub>2</sub>	徒歩分	バス分=0のとき：最寄り駅までの徒歩時間 バス分>0のとき：最寄りバス停までの徒歩時間	分
<i>AC</i> <sub>3</sub>	バス分	最寄りバス停から駅までバス移動したときの所要時間	分
<i>AC</i> <sub>4</sub>	バスダミー	バス分>0のとき1, そうでないとき0	(0,1)
<b>住宅設備</b>			
<i>構造的特性</i>			
<i>SF</i> <sub>1</sub>	所在階	物件がある階数	階
<i>SF</i> <sub>2</sub>	専有面積	専有面積	m <sup>2</sup>
<i>SF</i> <sub>3</sub>	築年	2017年時点での築年数	年
<i>SF</i> <sub>4</sub>	2階以上	2階以上	(0,1)
<i>SF</i> <sub>5</sub>	角部屋	角部屋	(0,1)
<i>SF</i> <sub>6</sub>	南向き	南向き	(0,1)
<i>SF</i> <sub>7</sub>	エレベータ	物件のある建物にエレベータがある	(0,1)
<i>SF</i> <sub>8</sub>	RC/PC	建物構造が「鉄筋コンクリート造」「鉄骨鉄筋コンクリート造」「プレキャストコンクリート造」のいずれかである	(0,1)
<i>SF</i> <sub>9</sub>	鉄骨造	建物構造が「軽量鉄骨造」「重量鉄骨造」のいずれかである	(0,1)
<i>住宅内特性</i>			
<i>IH</i> <sub>1</sub>	インターネット対応	インターネット対応	(0,1)
<i>IH</i> <sub>2</sub>	オートロック	オートロック	(0,1)
<i>IH</i> <sub>3</sub>	TVモニター付きインターホン	TVモニター付きインターホン	(0,1)
<i>IH</i> <sub>4</sub>	フローリング	フローリング	(0,1)
<i>IH</i> <sub>5</sub>	室内洗濯機置場	室内洗濯機置場	(0,1)
<i>IH</i> <sub>6</sub>	エアコン	エアコン	(0,1)
<i>IH</i> <sub>7</sub>	ガスコンロ	ガスコンロ	(0,1)
<i>IH</i> <sub>8</sub>	コンロ2口以上	コンロ2口以上	(0,1)
<i>IH</i> <sub>9</sub>	バス・トイレ別	バス・トイレ別	(0,1)
<i>IH</i> <sub>10</sub>	追い焚き風呂	追い焚き風呂	(0,1)
<i>IH</i> <sub>11</sub>	独立洗面台	独立洗面台	(0,1)
<i>IH</i> <sub>12</sub>	温水洗浄便座	温水洗浄便座	(0,1)
<i>IH</i> <sub>13</sub>	ペット相談可	ペット相談可	(0,1)
<i>住宅外特性</i>			
<i>EH</i> <sub>1</sub>	敷地内駐車場空きあり	敷地内駐車場空きあり	(0,1)
<i>EH</i> <sub>2</sub>	駐輪場	駐輪場	(0,1)
<b>敷地特性</b>			
<i>近隣特性</i>			
<i>SA</i> <sub>1</sub>	近隣駐車場空きあり	近隣駐車場空きあり	(0,1)
<i>SA</i> <sub>2</sub>	都市ガス	都市ガス	(0,1)

住宅情報サイトには、様々な住宅特性の情報が住宅設備をはじめとして掲載されており、これらを用いて物件の絞り込みを行うことができる。このとき、賃借人の関心が高い項目が優先的に記載されていると考えられ、また賃借人は実際にその情報をもとに物件の選択を行っているため、項目数の少ない敷地特性を除けば、賃料に影響する主要な要因は住宅情報サイトの情報によってカバーされていると考える。

変数間の相関などは Appendix に示した。連続変数の要約統計量は Table 3 にまとめた。なお、バス分の要約統計量はバス分が正であるときのみを対象として算出している。

Table 3: 要約統計量

	平均	標準偏差	最小値	最大値
<b>被説明変数</b>				
補正賃料	92950	39373	25958	384583
<b>交通利便性</b>				
東京距離	20749.89	15242.42	528.31	86464.61
徒歩分	8.52	5.17	0.00	29.00
バス分	6.49	1.42	1.00	9.00
<b>住宅設備</b>				
<b>構造的特性</b>				
所在階	3.02	2.46	1	20
専有面積	37.38	16.17	15.00	100.00
築年	14.51	11.38	0	52

### 3.3. 推計結果

Table 4 に各モデルの推定結果をまとめた。Full Model について VIF を算出すると、定数項とバス分・バスダミーを除けば最大値が 3.06 となっているため、重大な多重共線性は生じていないと考えられる。

Table 4: モデル比較

	Model 0	Model 1	Model 2	Model 3	Model 4	Full Model
const	0.608*	0.516*	0.312*	0.585*	0.579*	0.381*
<i>AC</i> <sub>1</sub>	-0.206*	-0.172*	-0.173*	-0.207*	-0.202*	-0.164*
<i>AC</i> <sub>2</sub>	-0.074*	-0.046*	-0.055*	-0.074*	-0.072*	-0.043*
<i>AC</i> <sub>3</sub>	-0.007*	-0.007*	-0.007*	-0.007*	-0.007*	-0.006*
<i>AC</i> <sub>4</sub>	-0.165*	-0.101*	-0.093*	-0.159*	-0.160*	-0.078*
<i>SF</i> <sub>1</sub>		0.037*				0.035*
<i>SF</i> <sub>2</sub>	0.251*	0.242*	0.227*	0.250*	0.249*	0.226*
<i>SF</i> <sub>3</sub>		-0.104*				-0.070*
<i>SF</i> <sub>4</sub>		0.010*				0.010*
<i>SF</i> <sub>5</sub>		0.018*				0.012*
<i>SF</i> <sub>6</sub>		0.005*				0.005*
<i>SF</i> <sub>7</sub>		0.055*				0.033*
<i>SF</i> <sub>8</sub>		0.123*				0.099*
<i>SF</i> <sub>9</sub>		0.024*				0.017*
<i>IH</i> <sub>1</sub>			0.016*			0.012*
<i>IH</i> <sub>2</sub>			0.162*			0.053*
<i>IH</i> <sub>3</sub>			0.027*			-0.002*
<i>IH</i> <sub>4</sub>			0.045*			0.041*
<i>IH</i> <sub>5</sub>			-0.006*			-0.021*
<i>IH</i> <sub>6</sub>			-0.014*			-0.006*
<i>IH</i> <sub>7</sub>			-0.037*			0.009
<i>IH</i> <sub>8</sub>			0.038*			0.030*
<i>IH</i> <sub>9</sub>			0.069*			0.058*
<i>IH</i> <sub>10</sub>			0.029*			0.024*
<i>IH</i> <sub>11</sub>			0.043*			0.024*
<i>IH</i> <sub>12</sub>			0.057*			0.030*
<i>IH</i> <sub>13</sub>			0.053*			0.023*
<i>EH</i> <sub>1</sub>				0.214*		0.118*
<i>EH</i> <sub>2</sub>				0.046*		-0.019*
<i>SA</i> <sub>1</sub>					0.188*	0.083*
<i>SA</i> <sub>2</sub>					0.038*	0.006*
R <sup>2</sup>	0.650	0.784	0.762	0.654	0.653	0.799
AIC	-31169	-145553	-122878	-34371	-33405	-162987

\*は5%有意な相関係数

モデルの推定結果を比較すると、モデルが含む説明変数の数が多いほど決定係数・AIC がともに大きい傾向にあることが分かる。しかし Model 1 と Model 2 を比較すると、Model 2 の方が説明変数の数が多いにもかかわらず、Model 1 の方が決定係数・AIC がともに良い。このことから、やみくもに変数を投入するのではなく、妥当な変数を投入することが予測精度の向上に寄与することが示唆される。

Model 0 と Model 1~Model 4 を比較すると、構造的特性及び住宅内特性を考慮した Model 1, Model 2 が他と比べて大きく AIC を改善している。また、立地のみを考慮した Model 0, Model 4 の説明力は決定係数で 0.65 程度であるのに対し、住宅設備を考慮したモデルでは説明力が大幅に改善されている (Full Model では 0.799)。これらから、住宅設備を考慮することでヘッドニックモデルの改善が見込めると考えられる。

Full Model に含まれた変数を見ると、すべての変数が選択されているものの、ガスコンロを除いてすべて 5%有意となっている。住宅設備に関する変数がすべて選択されていることは、これらが賃料や住宅選択行動に影響を及ぼしているという仮説と矛盾しない。

賃料に及ぼす影響が大きな変数は、専有面積・東京距離・敷地内駐車場空きありである。この結果は経験的な知見と矛盾しないが、敷地内駐車場空きありの影響が駅やバスの利便性よりも大きいという点は興味深い。住宅設備の中で比較的影響が大きい変数としては、専有面積・RC/PC・築年・バス・トイレ別・オートロックなどが挙げられる。RC/PC とオートロックに該当する物件は、全体のうちそれぞれ 5 割程度である。このため、これらの特性は住宅の付加価値として機能していることが予想される。またこれらはそれぞれ地震災害・住宅侵入盗に対する安全性を高めるための機能として解釈でき、これは上記の予想と合致する。一方でバス・トイレ別は 9 割以上の物件が該当する項目である。このためこれは付加価値というよりは、該当しない物件のウィークポイントとして解釈するべきである。

## 4. モデルの拡張:過少定式化バイアス

### 4.1. 過少定式化バイアスの特定

本節では、前節で推定した Full Model

$$y = \log p = \sum_i \beta_i^{full} x_i + \alpha \quad (3)$$

から住宅設備に含まれる変数 $SF_k, IH_l, EH_m$ のうちいずれか1つを除外して推定を行う。  $x_{omit}$ を説明変数に含めずに推定したモデルを

$$y = \log p = \sum_{i \neq omit} \beta_i^{x_{omit}} x_i + \alpha \quad (4)$$

とすると、その回帰係数の差を $x_{omit}$ の損失による $x_i$ の過少定式化バイアス

$$Bias_i^{x_{omit}} = \beta_i^{x_{omit}} - \beta_i^{full} \quad (5)$$

と定義する。バイアスが正である場合、Full Model に対して回帰係数が大きく推定されていることを示し、バイアスが負である場合、Full Model に対して回帰係数が小さく推定されていることを示す。なお、バイアスの有意性は、 $\beta_i^{x_{omit}}$ が $\beta_i^{full}$ の信頼区間に入っているか否かで判定し、信頼区間に入っていない場合に「バイアスが発生した」と判断する(Shimizu, 2014)。

## 4.2. 過少定式化バイアスに関する推計結果

Table 5 は推定されたバイアスと、各推定モデルの決定係数及び Full Model との AIC の差を示している。

結果を見ると、専有面積と築年の損失は AIC を大きく低下させ、また多くのバイアスを生じさせている。全体的には AIC の減少幅（表中の diff AIC）の大きい変数が失われると、バイアスの発生数（表中の sig #）も多くなる傾向にある（スピアマンの順位相関係数は 0.90）。このことから過少定式化バイアスの発生を抑えるためには、モデルの予測向上に寄与する変数をモデルに組み込んでいくことが効果的であると考えられる。また、Full Model の回帰係数の絶対値とバイアスの発生数の間にも正の相関があるため（スピアマンの順位相関係数 0.75）、賃料に対する影響が大きいと予想される変数を単純にモデルに組み込むことによっても、過少定式化バイアスの発生を抑えることができると考えられる。

バイアスを発生させやすい変数としては、構造的特性の中では所在階・専有面積・築年・RC/PC が挙げられる。また住宅内特性の中ではオートロックとコンロ 2 口以上が挙げられる。これらの変数には研究上の関心がなかったとしても、他の変数の回帰係数にバイアスを生じさせないためにモデルに含むことが望ましい。

Table 5: 過少定式化バイアス (100 倍した値)

Bias × 100		モデルから除外された変数																							sig #	
		SF <sub>1</sub>	SF <sub>2</sub>	SF <sub>3</sub>	SF <sub>4</sub>	SF <sub>5</sub>	SF <sub>6</sub>	SF <sub>7</sub>	SF <sub>8</sub>	SF <sub>9</sub>	IH <sub>1</sub>	IH <sub>2</sub>	IH <sub>3</sub>	IH <sub>4</sub>	IH <sub>5</sub>	IH <sub>6</sub>	IH <sub>7</sub>	IH <sub>8</sub>	IH <sub>9</sub>	IH <sub>10</sub>	IH <sub>11</sub>	IH <sub>12</sub>	IH <sub>13</sub>	EH <sub>1</sub>		EH <sub>2</sub>
モデルに残された変数	AC <sub>1</sub>	-0.247*	2.404*	0.167*	0.000	-0.010	0.015	-0.057	-0.145*	-0.007	-0.014	-0.301*	-0.005	-0.033	-0.028	-0.004	0.000	-0.219*	-0.002	0.026	0.006	0.008	-0.008	0.047	-0.082*	7
	AC <sub>2</sub>	-0.244*	2.848*	0.181*	-0.003	-0.022	0.004	-0.059	-0.230*	-0.001	-0.009	-0.130*	-0.001	-0.057	-0.004	0.002	0.000	-0.010	-0.023	0.021	-0.030	-0.003	0.044	-0.001	-0.046	5
	AC <sub>3</sub>	-0.028	0.816*	0.104	0.003	0.005	-0.004	0.003	-0.030	0.003	0.001	0.001	0.003	0.028	0.000	0.009	0.001	-0.005	0.004	-0.020	-0.014	0.019	0.022	-0.006	-0.018	1
	AC <sub>4</sub>	-0.224	3.589*	-0.337	-0.031	-0.088	0.036	-0.119	-0.121	-0.048	-0.114	-0.519	-0.042	-0.880	-0.140	-0.108	-0.004	-0.097	0.031	0.215	-0.110	-0.277	-0.055	0.022	0.065	1
	SF <sub>1</sub>	-	1.640*	0.158*	0.182*	-0.006	-0.009	0.350*	0.311*	0.004	0.006	0.088	-0.002	0.013	0.012	0.006	0.000	-0.007	-0.022	-0.011	-0.006	0.044	0.068	0.012	0.013	5
	SF <sub>2</sub>	0.247*	-	-1.366*	-0.022	0.027	0.028	0.060	0.382*	0.034	-0.034	-0.202*	0.003	-0.079	-0.024	0.024	0.000	0.027	0.338*	0.563*	0.302*	0.034	0.038	0.010	0.051	7
	SF <sub>3</sub>	-0.093	5.328*	-	0.023	0.012	0.013	-0.027	0.893*	0.074	-0.007	-0.913*	0.036	-0.045	0.081	0.006	0.002	-0.043	-0.293*	-0.160*	-0.190*	-0.485*	-0.051	0.012	-0.037	7
	SF <sub>4</sub>	2.545*	-2.070*	-0.551*	-	0.037	-0.003	0.039	0.934*	0.142	0.007	0.178	0.002	-0.043	0.015	-0.001	0.000	-0.013	0.008	-0.027	-0.004	-0.033	-0.053	-0.006	-0.023	4
	SF <sub>5</sub>	-0.040	1.204*	-0.135	0.018	-	0.003	-0.075	-0.495*	-0.011	0.014	-0.065	-0.001	0.043	-0.046	-0.010	-0.010	0.072	0.065	0.069	0.045	0.031	-0.012	0.006	0.010	2
	SF <sub>6</sub>	-0.153*	3.043*	-0.370*	-0.004	0.008	-	-0.086	-0.105	0.003	0.001	-0.072	0.002	0.055	0.002	0.011	0.000	-0.009	0.038	0.019	-0.006	-0.010	-0.032	-0.017	-0.014	3
	SF <sub>7</sub>	2.043*	2.338*	0.270*	0.016	-0.066	-0.030	-	3.154*	0.014	0.044	1.163*	-0.001	0.029	-0.038	0.016	-0.012	0.225*	0.040	-0.009	-0.001	-0.035	0.135	0.038	-0.089	6
	SF <sub>8</sub>	0.747*	6.080*	-3.648*	0.160	-0.177	-0.015	1.296*	-	-1.069*	0.045	1.867*	0.002	0.068	-0.032	0.027	0.000	-0.042	-0.158	-0.118	0.183	-0.014	0.155	-0.001	-0.153	6
	SF <sub>9</sub>	0.037	2.251*	-1.255*	0.101	-0.017	0.002	0.025	-4.417*	-	0.039	0.743*	-0.002	-0.068	-0.015	0.002	0.000	-0.154	0.089	-0.067	0.061	0.023	-0.001	-0.012	-0.142	4
	IH <sub>1</sub>	0.049	-1.803*	0.098	0.004	0.017	0.000	0.060	0.149	0.031	-	0.144	-0.008	0.512*	-0.052	-0.022	0.002	0.229*	-0.036	-0.053	0.094	-0.003	0.036	0.005	-0.341*	4
	IH <sub>2</sub>	0.263*	-4.016*	4.643*	0.038	-0.029	-0.013	0.595*	2.324*	0.224*	0.054	-	-0.006	0.134	-0.021	-0.016	0.000	0.099	-0.027	0.126	-0.010	0.019	-0.017	0.039	-0.064	6
	IH <sub>3</sub>	0.123	-1.291*	3.554*	-0.010	0.008	-0.007	0.015	-0.044	0.011	0.055	0.118	-	0.095	-0.045	-0.017	-0.011	0.256*	0.099	0.189*	0.055	0.657*	0.073	0.001	-0.130	5
	IH <sub>4</sub>	0.034	-1.387*	0.202*	-0.008	0.017	0.009	0.013	0.074	-0.018	0.169*	0.118	-0.004	-	-0.066	0.004	-0.018	0.279*	0.073	0.007	-0.017	0.058	0.032	0.017	-0.105	4
	IH <sub>5</sub>	-0.155	1.979*	1.749*	-0.013	0.087	-0.002	0.083	0.167	0.019	0.083	0.090	-0.010	0.318*	-	-0.037	0.000	-0.075	0.873*	-0.016	0.319*	-0.115	-0.077	0.006	-0.146	5
	IH <sub>6</sub>	-0.128	-3.215*	0.216*	0.001	0.030	-0.013	-0.053	-0.223*	-0.005	0.054	0.109	-0.006	-0.031	-0.059	-	0.000	-0.060	-0.063	0.060	0.011	0.105	-0.069	-0.014	-0.059	3
	IH <sub>7</sub>	0.046	-0.058	-1.356*	0.000	-0.641	0.012	-0.863	-0.014	-0.005	0.092	0.040	0.080	-2.966*	0.007	0.003	-	2.177*	0.215	0.120	0.056	-0.030	0.070	0.016	-0.022	3
	IH <sub>8</sub>	-0.024	0.610*	0.251*	-0.003	0.037	-0.002	0.131	-0.059	-0.053	0.098	0.113	-0.015	0.360*	0.020	0.010	0.017	-	0.194*	-0.009	0.204*	0.273*	0.067	0.011	-0.093	6
	IH <sub>9</sub>	-0.116	11.588*	2.569*	0.003	0.050	0.012	0.036	-0.339*	0.046	-0.023	-0.046	-0.009	0.143	-0.357*	0.016	0.003	0.294*	-	-0.080	0.731*	0.188	0.060	0.003	0.011	6
	IH <sub>10</sub>	-0.055	18.305*	1.330*	-0.010	0.051	0.006	-0.007	-0.241*	-0.033	-0.032	0.206*	-0.016	0.014	0.006	-0.015	0.001	-0.013	-0.076	-	0.157	0.278*	-0.091	-0.005	-0.057	5
	IH <sub>11</sub>	-0.032	10.625*	1.711*	-0.001	0.035	-0.002	-0.001	0.404*	0.033	0.063	-0.018	-0.005	-0.033	-0.134	-0.003	0.001	0.317*	0.750*	0.170	-	0.119	0.013	-0.009	-0.131	5
	IH <sub>12</sub>	0.173*	0.878*	3.237*	-0.009	0.018	-0.002	-0.023	-0.024	0.009	-0.001	0.026	-0.045	0.087	0.036	-0.020	0.000	0.316*	0.143	0.223*	0.088	-	0.079	0.005	-0.074	5
	IH <sub>13</sub>	0.485*	1.812*	0.619*	-0.027	-0.013	-0.014	0.165	0.459*	0.000	0.032	-0.041	-0.009	0.086	0.043	0.024	0.001	0.140	0.083	-0.132	0.017	0.144	-	0.006	-0.017	4
	EH <sub>1</sub>	0.572	3.206*	-1.006	-0.021	0.043	-0.049	0.323	-0.020	-0.060	0.028	0.644	-0.001	0.311	-0.024	0.035	0.002	0.153	0.028	-0.047	-0.082	0.066	0.043	-	-0.167	1
EH <sub>2</sub>	-0.062	-1.546*	-0.287*	0.007	-0.007	0.004	0.070	0.291*	0.066	0.195*	0.099	-0.010	0.183*	-0.053	-0.013	0.000	0.124	-0.010	0.054	0.113	0.086	0.011	0.016	-	5	
SA <sub>1</sub>	0.319	-6.520*	-0.875	0.017	0.184	-0.038	0.310	-0.145	-0.080	0.295	0.245	-0.009	0.549	0.007	-0.003	0.002	-0.016	0.109	-0.062	0.202	0.207	-0.045	-0.099	0.214	1	
SA <sub>2</sub>	-0.206*	0.769*	-0.578*	0.008	0.070	0.002	0.102	0.119	-0.013	0.078	0.006	0.004	0.150	-0.073	0.003	0.008	0.388*	0.005	0.055	0.034	0.117	-0.059	0.030	-0.135	4	
sig #	11	28	23	1	0	0	3	16	2	2	8	0	5	1	0	0	10	5	4	5	4	0	0	2		
R <sup>2</sup>	0.795	0.600	0.784	0.799	0.799	0.799	0.799	0.794	0.799	0.799	0.797	0.799	0.797	0.799	0.799	0.799	0.798	0.798	0.799	0.799	0.798	0.799	0.799	0.799	0.799	
dif AIC	5354.6	162955	18024	108.1	284.9	48.6	871	6375.5	266.7	233.7	2769.8	3.8	2427.6	259.8	46.6	1.1	1388.7	1741.6	767.6	705	1197.9	514.2	373.4	649.9		

\*は Full Model の 95%信頼区間に入っていないもの。またバイアスは 100 倍した値を表中で示した。



一方で、残された変数のうちどんな変数に過少定式化バイアスが生じているかを確認すると、こちらはバイアスの発生数に大きな偏りは見られない。一方で、Full Model の回帰係数の絶対値とバイアスの発生数の間には、弱い正の相関が見られる（スピアマンの順位相関係数はスピアマン 0.42）。この点は、賃料に対する影響が大きい変数の回帰係数に対してバイアスが生じる可能性を示唆している。

この観点からは、専有面積と東京距離の回帰係数バイアスの解釈には特に注意が必要であると考えられる。これらについてのバイアスの発生数は7件で、全変数で最も多い。またこれらの偏回帰係数の絶対値は、専有面積が1番目に、東京距離は2番目に大きい。よってこれらは賃料に対する影響が大きく、かつバイアスも生じやすい変数であると考えられる。専有面積について見ると、所在階・築年・RC/PC・オートロック・バストイレ別・追い焚き風呂・独立洗面台が失われた場合にバイアスが生じることが分かる。このうち、築年とオートロックについては負の、それ以外については正のバイアスが生じていることから、これらの変数に関するデータが入手できなかった場合は、この分のバイアスを考慮して回帰係数の解釈を行う必要があるだろう。

一方で東京距離について見ると、所在階・専有面積・築年・RC/PC・オートロック・コンロ2口以上・駐輪場が失われた場合にバイアスが生じることが分かる。このうち、専有面積と築年については正の、それ以外については負のバイアスが生じていることから、これらの変数に関するデータが入手できなかった場合は、この分のバイアスを考慮して回帰係数の解釈を行う必要があるだろう。

## 5. 結論

本研究では住宅設備に着目して、変数の投入に伴うヘドニックモデルの予測向上への寄与と、変数の損失に伴う過少定式化バイアスの発生を実証的に明らかにした。

変数の投入に伴う予測向上については、住宅設備に分類される変数の投入によってAICの向上が見られることが確認できた。また変数選択ではすべての変数が選択されたが、これは本研究で使用したデータセットが極めて多くの物件を含んでいるために、説明変数の増加による汎化誤差の低減が抑えられたことが原因として考えられる。また、住宅設備の変数は1つも除外

されなかったことは、これらの賃料や住宅選択行動に対する影響が大きくなってきているという仮説を支持するものである。

過少定式化バイアスの発生については、モデルの予測や賃料への影響が大きいと考えられる変数が失われたとき、多くのバイアスが生じる傾向にあることが分かった。また、賃料への影響が大きい変数にバイアスが生じやすい傾向も見られ、こちらはヘドニックモデルの解釈において留意すべきであると考えられる。特に、オートロックのような個別の住宅事情に関する変数の損失によって、東京距離のような広域的な変数にバイアスが生じたのは興味深い。この点は、地域性を評価するためにヘドニック分析を行う際も、住宅設備をモデルに投入することが重要であるという示唆を与えている。

本研究では既往研究と比較して、数多くの住宅設備をモデルに組み込んだ。しかしながら敷地特性に関しては限られた情報しか扱えなかったため、敷地特性に対するバイアスなどは明らかにできていない。また、住宅設備が賃料に及ぼす影響が住宅情報サイトの普及に伴ってどのように増大したのか、という点は時系列的には明らかになっていない。よって、これらの点は今後の課題として指摘しておきたい。

## 引用

Abbott, J. K., & Klaiber, H. A. (2011). An Embarrassment of Riches: Confronting Omitted Variable Bias and Multi-Scale Capitalization in Hedonic Price Models. *Review of Economics and Statistics*, 93(4), pp. 1331-1342.

Billings, S. B. (2015). Hedonic Amenity Valuation and Housing Renovations. *Real Estate Economics*, 43(3), pp. 652-682.

Boeing, G., & Waddell, P. (2016). New Insights into Rental Housing Markets across the United States. *Journal of Planning Education and Research*, 37(4), pp. 457-476.

Chau, K. W., & Chin, T. L. (2003). A Critical Review of Literature on the Hedonic Price Model. *International Journal for Housing Science and Its Applications*, 2(27), pp. 145-165.

Ekeland, I., Heckman, J. J., & Nesheim, L. (2004). Identification and Estimation of Hedonic Models. *Journal of Political Economy*, 112(S1), pp. S60-S109.

Gao, X., & Asami, Y. (2001). The External Effects of Local Attributes on Living Environment in Detached Residential Blocks in Tokyo. *Urban Studies*, 38(3), pp. 487-505.

Genesove, D., & Mayer, C. (2001). Loss Aversion and Seller Behavior: Evidence from the Housing Market. *The Quarterly Journal of Economics*, 116(4), pp. 1233-1260.

Kang, C.-D. (2017). Effects of spatial access to neighborhood land-use density on housing prices: Evidence from a multilevel hedonic analysis in Seoul, South Korea. *Environment and Planning B: Urban Analytics and City Science*, 0(0), pp. 1-23.

- Lancaster, K. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), pp. 132–157.
- McMillen, D. P., & Thorsnes, P. (2006). Housing renovations and the quantile repeat-sales price index. *Real Estate Economics*, 34(4), pp. 567–584.
- Miura, T., & Asami, Y. (2012). Hedonic analysis for the estimation of condominium rent utilizing web information. *Environment and Planning B: Planning and Design*, 39(6), pp. 1049–1068.
- Nelson, R. H. (1972). HOUSING FACILITIES, SITE ADVANTAGES, AND RENT. *Journal of Regional Science*, 12(2), pp. 249–259.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), pp. 34–55.
- Shimizu, C. (2014). Estimation of hedonic single-family house price function considering neighborhood effect variables. *Sustainability (Switzerland)*, 6(5), pp. 2946–2960.
- Shimizu, C., & Karato, K. (2015). Property Price Index Theory and Estimation : A Survey. *IRES Working Paper Series*, (May).
- Shimizu, C., Nishimura, K. G., & Asami, Y. (2004). Search and Vacancy Costs in the Tokyo Housing Market: Attempt To Measure Social Costs of Imperfect Information. *Review of Urban & Regional Development Studies*, 16(3).
- Sirmans, G. S., David, A., & Emily, N. (2005). The Composition of Hedonic Pricing Models. *Journal of real estate literature*, 13(1), pp. 1–44.
- Wong, S. K., Chau, K. W., Karato, K., & Shimizu, C. (2017). Separating the Age Effect from a Repeat Sales Index: Land and Structure Decomposition. *Journal of Real Estate Finance and Economics*, pp. 1–16.
- Yoo, S., Im, J., & Wagner, J. E. (2012). Variable selection for hedonic model using machine learning approaches: A case study in Onondaga County, NY. *Landscape and Urban Planning*, 107(3), pp. 293–306.
- Yoo, S., & Wagner, J. E. (2016). A review of the hedonic literatures in environmental amenities from open space: a traditional econometric vs. spatial econometric model. *International Journal of Urban Sciences*, 20(1), pp. 141–166.

## Appendix: Prior Analysis

ヘドニック分析を行う前に、データの特徴について事前に分析を行う。まず連続変数について、散布図行列 (Figure A.1) を表示する。行列の対角成分はヒストグラムとなっている。

ヒストグラムを見てみると、バス分・所在階はゼロ近辺にピークがあり、それ以降徐々に頻度が低下していく分布になっていることが分かる。築年の分布からは、このサイトには近年作られた物件ほど多く登録されている傾向が見られる。その他の項目は、ゼロより少し大きい値にピークを持つ分布となっている。専有面積は複数のピークを持つ分布となっており、借主層によって求められる物件の大きさが離散的に変動していることが予想される。また東京距離の分布から、東京にある程度近い地域では高密度に賃貸住宅が集積していることが観察される。補正賃料は正の値を取る特に滑らかな分布となっている。

次に散布図を観察する。バス分に関する散布図を見ると、バス分がゼロの場合とそれより大きい場合で、点の分布状況が異なることが見て取れる。この点から、バスダミーの投入は妥当であると考えられる。その他の散布図については点の分布の極端な偏りが見られないため、異常値の除外は正常に行われたと考えられる。

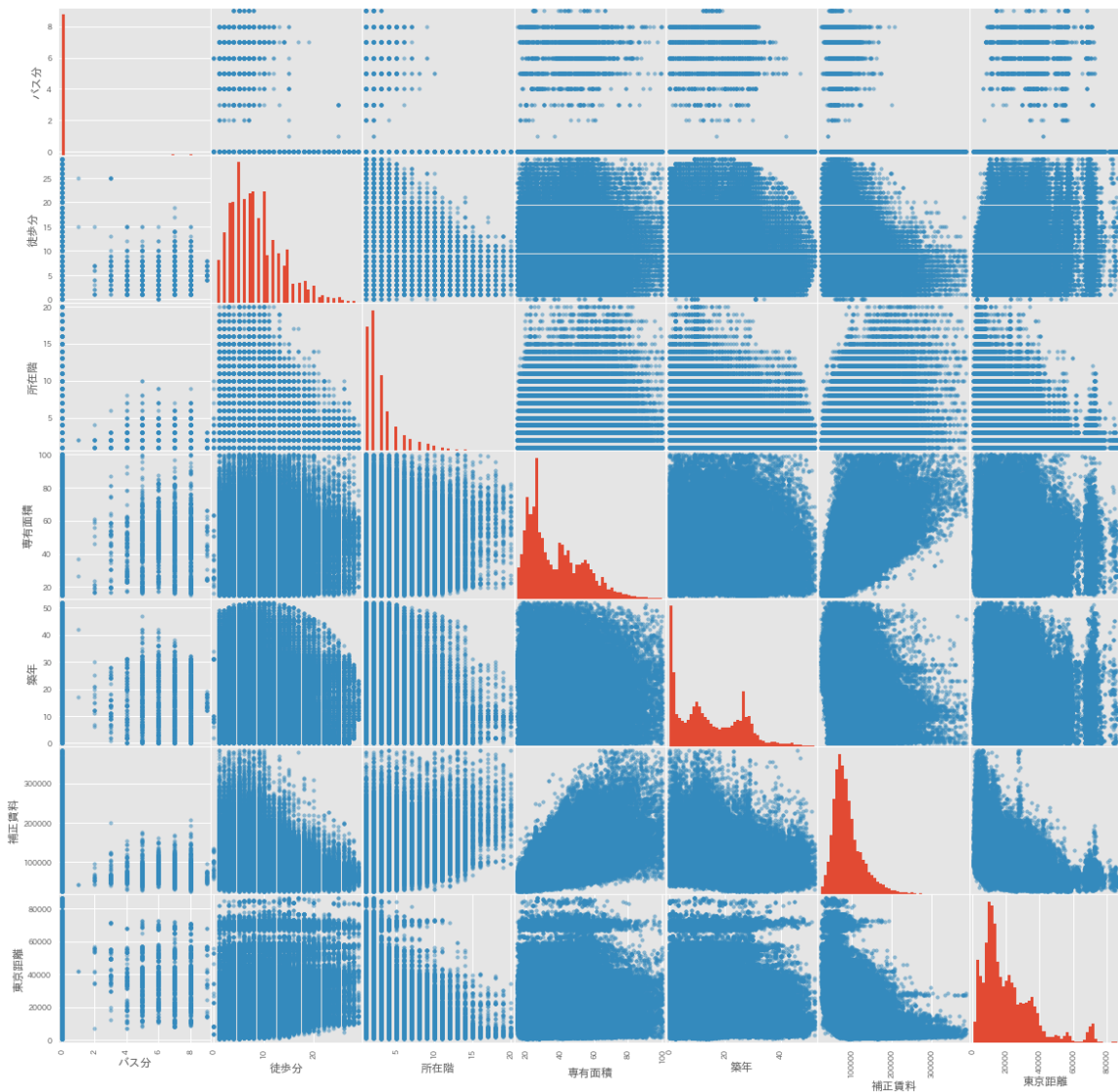


Figure A.1：散布図行列

次にダミー変数について、該当率（ダミー=1となる物件の割合）を図示した（Figure A.2）。該当率はダミー変数の平均値と一致する。これを見ると、バスダミー・ガスコンロ・近隣駐車場空きあり・敷地内駐車場空きありについては該当する物件が極めて限られている一方で、バス・トイレ別や室内洗濯機置場などについてはほとんどの物件が該当していることが分かる。一方で、オートロックや温水洗浄便座などの設備は物件により有無がまちまちであるので、このような設備が物件の差別化に寄与していることが予想される。

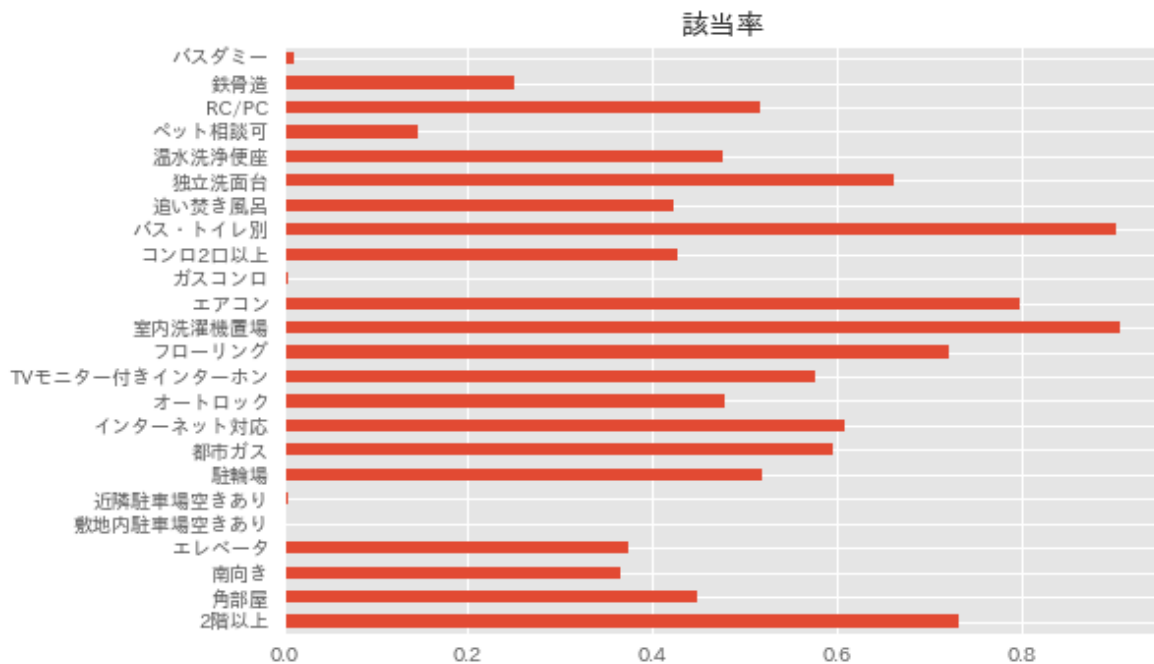


Figure A.2 : ダミー変数該当率

項目間の関係性を見るため、すべての項目についてピアソンの積率相関係数を算出し、ヒートマップにして表示した (Figure A.3)。これを見ると、バス分とバスダミーの相関が極めて強いことを除けば、項目間の相関係数の絶対値は 0.7 未満である。比較的相関が強いものとしては、築年や建物構造・建物高さに関連する項目と、一部の設備の有無との相関が挙げられる。また、補正賃料と専有面積も比較的相関が強い。

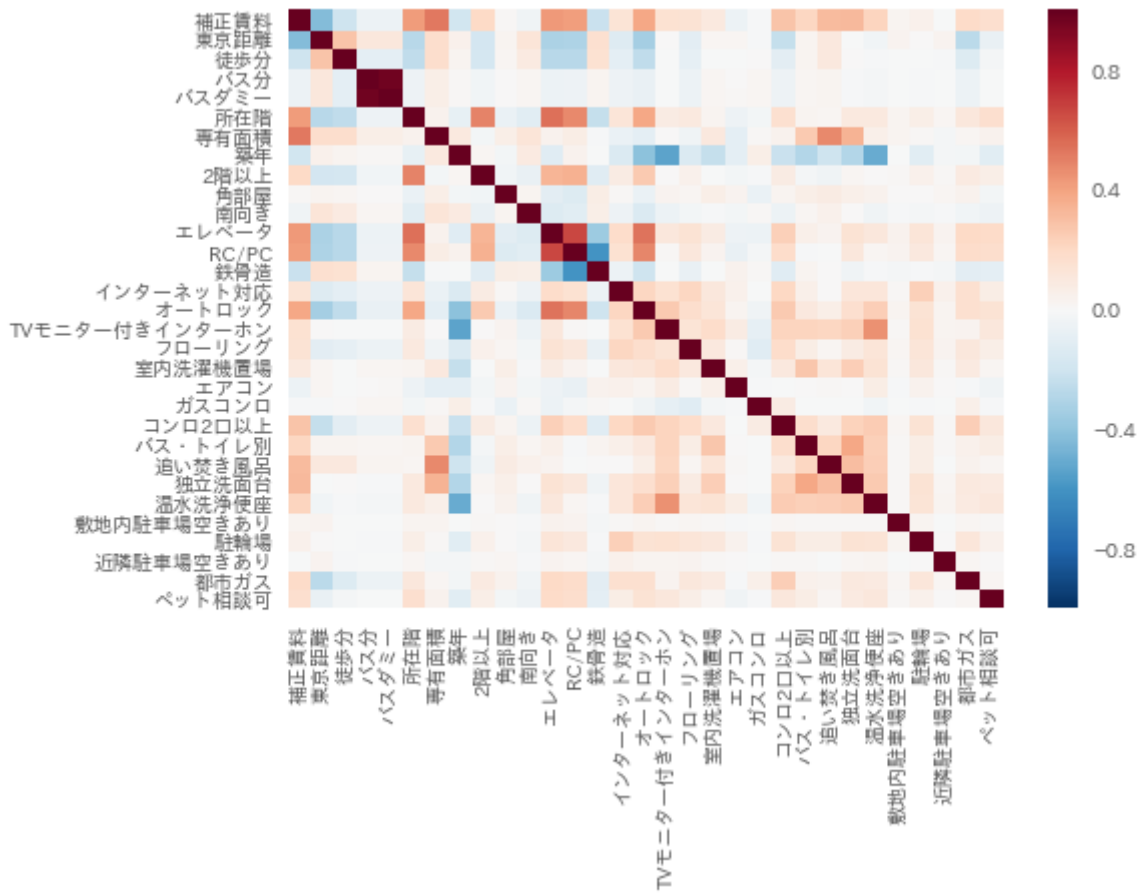


Figure A.3 : 相関係数