



ESRI Research Note No.53

ウェブスクレイピング等を活用した シェアリング・エコノミーの計測について

亀井 英人

June 2020



内閣府経済社会総合研究所
Economic and Social Research Institute
Cabinet Office
Tokyo, Japan

ESRI Research Note は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません（問い合わせ先：<https://form.cao.go.jp/esri/opinion-0002.html>）。

ESRI リサーチ・ノート・シリーズは、内閣府経済社会総合研究所内の議論の一端を公開するために取りまとめられた資料であり、学界、研究機関等の関係する方々から幅広くコメントを頂き、今後の研究に役立てることを意図して発表しております。

資料は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません。

The views expressed in “ESRI Research Note” are those of the authors and not those of the Economic and Social Research Institute, the Cabinet Office, or the Government of Japan.

ウェブスクレイピング等を活用した
シェアリング・エコノミーの計測について

亀井英人*

*内閣府経済社会総合研究所 研究官

[要旨]

シェアリング・エコノミーと呼ばれる経済活動の生産額の推計の精度向上、またデフレーター計測のため、ウェブスクレイピングを用いた手法により、民泊や大手クラウドソーシングサイトの市場規模、デフレターの試験的な推計を行った。あわせて、中間投入構造の把握のため、税務情報等の活用可能性について検討した。

[キーワード]

シェアリング・エコノミー、ウェブスクレイピング、デフレーター、民泊、クラウドソーシング

[目次]

1. 概要.....	2
2. 先行研究.....	2
3. 分析手法・結果.....	4
4. まとめ・今後の課題.....	16
参考文献.....	17

1. 概要

近年、IT 技術の進展により、個人間の需給がインターネット上のプラットフォームを介してマッチングされることによる財・サービスの売買、すなわちシェアリング・エコノミーと呼ばれる経済活動が盛んになってきている。日本でも、民泊やフリマアプリでの取引が社会的な現象となっている。内閣府においても、「統計改革の基本方針」 [1]においてシェアリング・エコノミーの経済規模に対して統計的捕捉を行う方針が示され、2017 年には生産額の推計、2018 年には付加価値額の推計が行われた [2]。

しかし、シェアリング・エコノミーの統計的な捕捉にむけて、依然としていくつかの課題が残っている。まずは、生産額の測定において、プラットフォーム事業者について網羅的な調査を行うことができず、特に海外のプラットフォーム事業者からの情報収集が困難な場合があるという問題がある。次に、付加価値額の測定において、財・サービスの提供者である個人事業主の中間投入構造の捕捉が必要となるが、個人事業主を対象としたアンケート調査を行う場合、精度の高い調査が比較的困難であることが確認されている [2]。このため、計測へ向けて、これらの実際的な課題を解決していくことが必要になる。最後に、得られた付加価値額を実質化する際に、財・サービスのデフレーターが必要となるが、現時点でシェアリング・エコノミーは小売物価統計調査の対象となっていないため、何らかの手法でそのデフレーターを計測する必要がある。本研究では、シェアリング・エコノミーの実装へむけて、以上3つの課題について検討した。

生産額の捕捉に関しては、海外の民泊仲介プラットフォーム事業者に直接聞き取りを行うことが困難であったため、プラットフォームのウェブスクレイピングを行うことによって生産額（取引規模）に関する情報収集が可能であるかの検証を試みた。その結果、試算に必要な取引情報が一部公開されていない場合もあるものの、取引規模の概算はプラットフォームによっては可能であることが判明した。

次に、付加価値額（提供者の中間投入構造）の把握に関しては、確定申告を行っている個人だけで取引額の多くの割合を占めていることに着目し、中間投入構造が納税書類から転記が可能であるかの確認を行った。その結果、一般的にはシェアリング・エコノミーは雑所得として申告されており、費用の明細を提出していないことから、納税書類からの転記によって中間投入構造を記入してもらうことは困難であることが判明した。

最後に、デフレーターの計測に関しては、価格がプラットフォームのウェブサイトから直接取得が比較的容易な情報であることから、ウェブスクレイピングによる取得を行った。価格分布が計測対象期間の異なるサンプル群について概ね安定的であることから、価格指数の値がある程度安定的であり¹、定義する事が妥当であることを確認した。

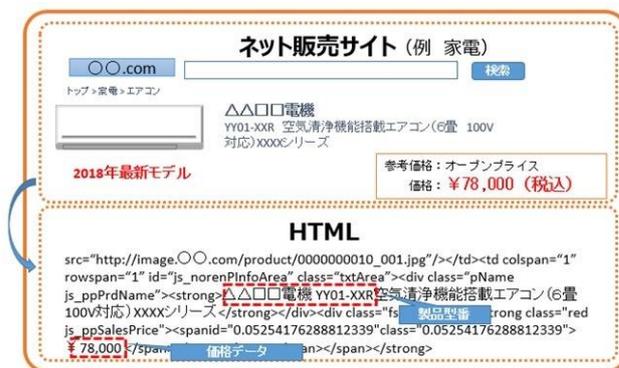
2. 先行研究

¹ 個別価格の値があまりに不安定であり、価格指数の変動に比して誤差が大きいような場合、価格指数を定義する正当性が低い。

シェアリング・エコノミーの生産額と付加価値額については、内閣府報告書 [2]にてその捕捉・推計の試みが行われている。プラットフォーム事業者へのヒアリングや旅行統計の活用等によって生産額を推計し、さらに財・サービス提供者へのアンケート調査を行いシェアリング・エコノミーの中間投入構造・付加価値額の把握が行われた。しかし、付加価値額推計の実装に向けて、上記の通りプラットフォーム事業者の経済活動捕捉の網羅性や提供者向けアンケートの精度向上という課題が指摘されている。

今回、生産額・デフレーター2の計測にあたりウェブスクレイピングによるウェブサイト上の取引情報の抽出を行った。スクレイピングとはプログラミングを用いてウェブ上の公開情報を取得する技術のことである。予約状況や価格情報など、人力でも原理的には可能な範囲の情報収集であるが、膨大な作業量を省くため、プログラミングを用いて人間の操作を模して検索やクリック等の操作を行うなどしてウェブ上に情報を表示したうえで、ウェブページのhtmlソースの構造や文字列における一定のパターンを検索して、ウェブページ上に表示されている必要情報を自動的に取得する。上記のプロセスのイメージとして、統計局の作成した [3]図表 1 が参考になると思われる。なお、本研究では Python の Selenium モジュール2を用いた。

〈図表 1 ウェブスクレイピングのイメージ〉



近年、国内外でウェブスクレイピングを経済統計に活用しようという試みは盛んとなっている。こうした試みは、プログラムによりウェブサイト上の公開取引情報を取得することにより、プラットフォーム上での取引規模や物価の計測を行おうというものである³。カナダ統計局ではスクレイピング業者から購入したデータを分析することによって地域別の民泊の取引規模を推計しており [4] [5]、価格情報の取得に関して、総務省統計局物価統計室

² モジュールとは新たな機能を追加するためのプログラムで、関数等を1つのファイルにまとめたもの。

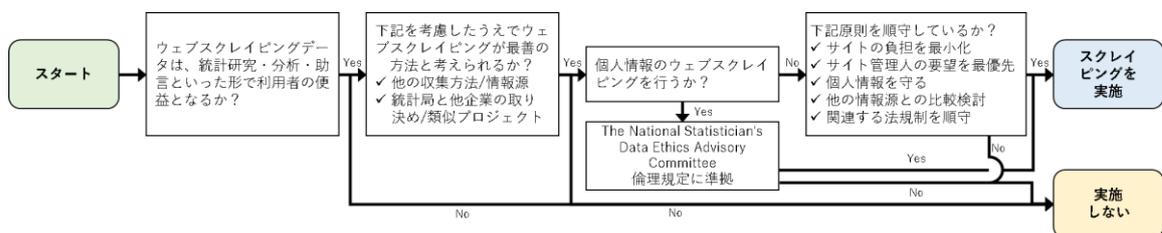
³ 後述するが、公開情報の範囲で取引規模が測定できるかどうかは、ひとえにウェブサイトの作りに依存する。民泊やクラウドソーシングサイトでは概ね捕捉が可能であるが、後述するように、民泊ならば宿泊客の数や泊数が不明であることから、またクラウドソーシングでは案件によって取引額が公開されていないことから、正確な取引額は不明である。

がスクレイピングにより宿泊サービスの物価指数を作成する [6]という試みや、日本銀行でも、家電製品の物価指数を作成し、機械学習手法を用いて品質調整を行う [7]という取り組みがなされている。そのほか各国でも通信販売価格の把握のため、スクレイピングが用いられている事例がみられる [8]。

取引情報がウェブ上の情報から取得できる場合、ウェブスクレイピングによりそれらを取得することで、例えば既存の代表的な商品・店舗・取得日において物価を調査する方法よりも報告者負担の少ない方法で、大規模で高精度な取引データが取得できることが期待されているし、他にもプラットフォームを経由した取引について網羅的に把握することができると考えられる。

一方、ウェブスクレイピングの経済統計への活用は比較的新しい試みであり、ウェブサイトへのトラフィック面での負担や個人情報に関する取扱いの注意など、慎重な実施が必要となる面もある。そこで、イギリス国家統計局（ONS）では、ウェブスクレイピングに関するポリシーを定め、これらの懸念点に関して問題が生じないように努めている [9]。具体的には、スクレイピングの実施に際して、図表2フローチャート上でのチェックを行う事や、データの管理はデータ保護法に則って行うことなどを定めている。

〈図表2 イギリス統計局 ウェブスクレイピングポリシーフローチャート〉



(備考) イギリス統計局ウェブスクレイピングポリシーフローチャート [9]より筆者作成

3. 分析手法・結果

まず、本研究は、仲介事業者・提供者へのヒアリング、またスクレイピングによる取引情報の抽出によりシェアリング・エコノミーに係るサービスの生産額・付加価値額・デフレーターを推計を行うものであるが、主に、取得の可能性、取得方法の検証を主眼に置き、分析を行った。また、図表3の通り、シェアリングのカテゴリを限定して分析を行った。対象とするカテゴリを、比較的規模が大きく、比較的財・サービスが均質的で捕捉が可能と思われる民泊、スキル・時間のシェアに絞った。

〈図表3 本分析の対象とするシェアリングのカテゴリ〉

シェアのカテゴリ	具体例	本分析の対象か
民泊	民泊新法、特区民泊に基づく宿泊施設	○（対象内）
スペース	店舗等・駐車場賃貸	比較的小規模であるため対象外
移動	ライドシェア	比較的小規模であるため対象外
モノ	フリマアプリ、バッグや衣服のレンタル	生産額は推計事例有 [10]、デフレターは取引品目・質が多様であり作成が困難
スキル・時間	クラウドソーシング（取引されるサービスは Web デザイン、ライティング、翻訳、入力作業等多岐にわたる）	○（対象内）
カネ	クラウドファンディング	主たる部分は金融取引であり付加価値に含まれないため対象外

(1) 生産額（市場規模）の推計

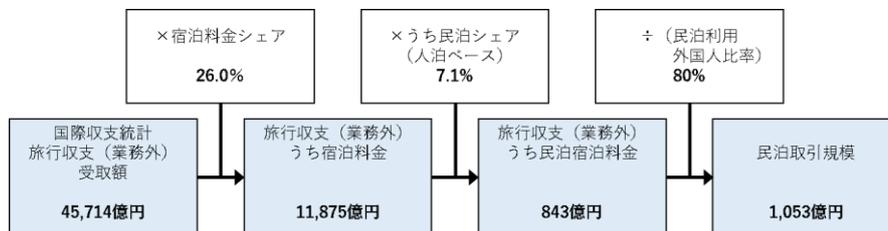
➤ 民泊

民泊の 2019 年の生産額について、内閣府報告書 [2]での推計方法を参考に推計し、加えて今回のスクレイピングによる取引データを用いて推計し、比較する。

【内閣府報告書を基にした推計（国際収支統計を用いた推計）】

内閣府報告書 [2]では国際収支統計の旅行収支受取額に対して、訪日外国人消費動向調査によって得られた外国人旅行者の民泊宿泊支出の割合を乗じて外国人旅行者の民泊支出額を算出し、さらに民泊利用者における外国人旅行者の比率で割り戻すことで全体の民泊取引規模（2017）を算出した。図表 4 は、内閣府の推計方法を参考に筆者が 2019 年の取引規模を試算した結果である。

〈図表4 国際収支統計を用いた年間民泊取引規模推計（2019）〉



（備考）IMF 国際収支統計、OECD 為替レート、訪日外国人消費動向調査（2019）を用いて推計した。

しかし、この推計には以下の問題点がある。

- 直接取引情報を用いることなく、いくつかの統計情報を組み合わせて間接

的に求められているため、推計の過程で誤差が生じる

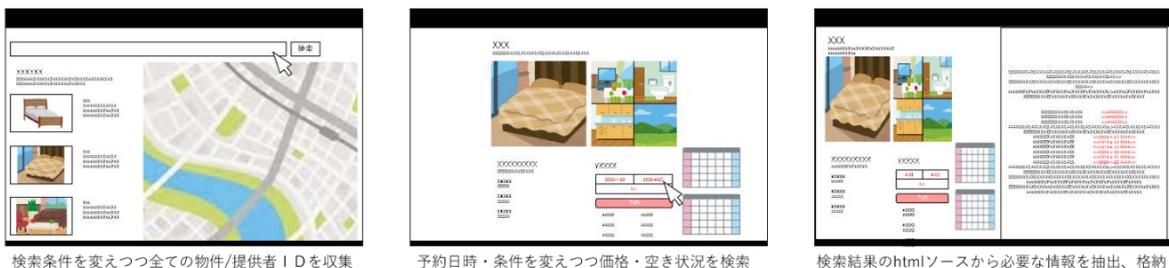
- 外国人消費動向調査における民泊利用率はアンケートによるものであり、旅行者の認識における「民泊」と法的な「民泊」（住宅宿泊事業法に基づく届出を行った民泊（本稿では新法民泊と呼ぶ）もしくは特区民泊）の間にずれが生じている可能性がある。⁴
- 民泊のシェアを計算する際に人泊ベースにより計算を行っており、民泊とホテルの単価の違いが反映されていない。しかも、複数種別の施設に宿泊した場合、内訳がわからないため泊数が不正確となる。

よって、今回は、民泊のプラットフォーム側の情報、また新法民泊の施行後に観光庁・内閣府によって集められた情報を元に、別の方法で本取引規模の推計を行う。

【本論における推計（ウェブスクレイピングデータを用いた推計）】

PythonのSeleniumモジュールを用いて大手民泊業者のプラットフォームにてスクレイピングを行った。物件住所・郵便番号・緯度経度や部屋タイプ、価格帯等の条件を変更して検索を行うことにより全物件リストを作成し、さらに全 리스팅の10%弱にあたる4000件を用いて、市場規模推計のため①単価② 리스팅のうち民泊（住宅宿泊事業者・特区民泊）の件数⁵③稼働率 をスクレイピングにより取得したデータから推計した⁶。プログラミングによる操作の模式図は図表5のとおりとなり、各物件について取得できたデータは図表6のようになる。

〈図表5 スクレイピング操作の模式図〉



⁴ 具体的には、住宅を用いた宿泊施設に見えて、旅館業法の枠内で営業している場合がある。

⁵ パッケージレンタルのサイトには、住宅宿泊事業法に基づく届出を行った民泊（本稿では新法民泊と呼ぶ）、特区民泊、旅館業法と主に3つの異なる法的根拠に基づく物件があり、前者二つを一般的に民泊と呼ぶ。

⁶ プラットフォームの全物件である約42000件を対象とし、価格は2019年6月13日（木曜日）の価格を用いた。実際の取引規模は、休日の前日等の単価が高くなることを考えると下記推計よりも大きくなると思われる。

〈図表 6 取得データ例〉

link	title	県名	地域	緯度	経度	評点	登録番号	最大 人数	寝室 数	ベッド 数の数	風呂 の数	基本料金	清掃費	サービス 料	取得日	空き状況 0日目	空き状況 1日目	...
https://w	【このは	Fukuoka-	Itoshima-	33.557	130.198	5	旅館業法 福岡県糸島市	6	1	6	5	10800	1504	12304	2019/6/13	TRUE	TRUE	FALSE
https://w	人気のビ	Chiba-ker	Shirako-m	35.459	140.4	4.5	M120006571	6	1	4	1	3800	6480	1432	2019/6/13	TRUE	FALSE	FALSE
https://w	JR Namba	Osaka,	Naniwa-ku, Osaka,			4	M270015824	5	1	3	1	6900	5000	1658	2019/6/13	FALSE	FALSE	FALSE

宿泊単価⁷は物件における 1 泊あたりの宿泊人数に依存するが、公開データからはあるリスティングに実際何人宿泊しているかが分からない（リスティングが空きか予約済かということしか分からない）ので、airbnb 社発表データ [11]による 1 泊あたりの平均宿泊人数が 2.8 人という数字に近づけるため、宿泊価格を取得する際に、宿泊人数を 3 人（ただし、物件の最大宿泊人数が 1 人または 2 人の場合は最大宿泊人数）と設定した。また、宿泊単価は、宿泊 1 日あたりに係る部分と滞在 1 回に係る部分に分けられる⁸が、1 日当たりの平均宿泊費を算出するため、1 回の滞在が平均的に 3.3 泊というデータ [11]を用いて、滞在 1 回あたりの費用が 1 泊につきその 1/3 ずつかかるものとした⁹。さらに、価格取得の際には、予約時点¹⁰から最も直近の予約可能日における価格を用いた。これは、もし宿泊日が近づくとつれ価格が一定の傾向を持って動く際には（取得した価格と取引価格が乖離するため）バイアスが発生する可能性があるが、 $\log(\text{価格})$ の時間平均からの乖離について

$$\text{各物件の平均をとったもの、すなわち } I_t = \frac{\sum_n (\log p_{nt} - \sum_t \log p_{nt} / T)}{N}$$

（ある同一の予約実施日に対し、宿泊予定日の時点 t を変えながらデータを取得した）プロットすると¹¹、図表 7 のように、週末や連休等によって影響をうけるものの、予約時点から宿泊日までの日数について一定の傾向はみられず、直近の予約可能日における価格を用いることによる大きな誤差は無いものと考えられる。

⁷ 物件あたり 1 泊あたりの宿泊単価であり、1 人あたり 1 泊あたりの宿泊単価ではない。

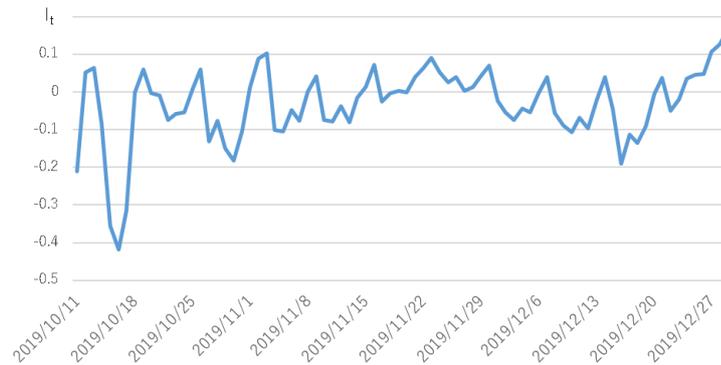
⁸ 対象とした民泊プラットフォームでは、「基本料金」が 1 日あたりに係る部分で、「サービス料」と「清掃費」と呼ばれる金額が滞在 1 回に係る部分となる。

⁹ 宿泊単価は、予約可能日として提示されている部分の価格を用いた。収集対象日が予約不可能な場合、予約可能な日のうち対象日から最も近い日程を指定した。

¹⁰ ウェブスクレイピングの実施日のことを指す。取引規模の推計にあたっては 2019 年 6 月 13 日（木曜日）のデータを用いた。

¹¹ 42000 件のうちサンプル 100 件について、2019 年 10 月 10 日時点で取得した価格を用いた。 p_{nt} は時点 t におけるサンプル n の価格であり、 N はサンプル数合計、 T は時点の数を表す。

〈図表7 民泊宿泊価格 宿泊日に関する価格変動の傾向〉



新法民泊・特区民泊のリスティング件数の推計には、観光庁・内閣府の把握している件数をベースとし、一つの届出（認定）に対してプラットフォーム上で複数のリスティングがある場合を考慮して算出した。サンプルの4000件から得られた届出あたりのリスティング数分布（図表8）を基に¹²、母集団の届出あたりのリスティング数分布がべき分布であると仮定し、分布形状のパラメーターを仮定する。さらに、ランダムに抽出したサンプルが図表8の分布に最も近い分布を与えるパラメーターをKL情報量¹³の最小化により求めて母集団の分布を特定し¹⁴、新法民泊につき1届出あたりのリスティングが1.3件、特区民泊につき1認定あたりのリスティングが2.0件¹⁵と推計した。新法民泊の全届出件数が16528件、特区民泊の全許可件数が6539件であることとあわせて全リスティング件数が推計される。

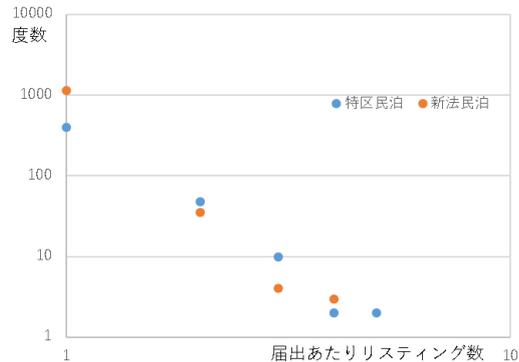
¹² なお、新法民泊あるいは特区民泊という法律上の分類は、ウェブサイト上の「登録番号」欄に記載されている。ちなみに、サンプル4,000件のうち新法民泊が1,231件、特区民泊が545件、それ以外は、旅館業法が2012件、その他が30件、リンク切れ等で不明な物件が182件であった。

¹³ 届出あたりリスティング数を*i*とし、母分布の仮定から計算したサンプルにおける*i*の分布を*Q(i)*、実際のサンプルにおける分布を*P(i)*とし、 $\sum_i P(i) \log(P(i)/Q(i))$ によって計算される量である。*Q(i)*を真の分布、*P(i)*を観測される分布と考えた時、この観測される分布が実現される確率にlogをとったものと考えてよい。

¹⁴ これらはきわめて強い仮定であり、本来はプラットフォームの全リスティングに対する届出（認定）の数を求めるべきであるが、データが限られていることから推計を行った。結果として、民泊全体におけるリスティング数に大きい誤差が生じている可能性がある。

¹⁵ 実際の特区民泊の物件について同じ数字を公表されている情報で把握できる範囲で（部屋番号が記載されていない認可については1件として数えたため、過少に見積もっていると思われる）集計したところ、大田区では4.3、大阪市では2.1となったため、2.0という数字はやや幅を持って考えた方が良いと思われる。

〈図表 8 サンプルにおける届出当たりのリスティング数〉



稼働率については全リスティングの平均をとった。稼働率としては、6月13日時点での一日後の予約率が約75%であった一方（図表9）、住宅宿泊事業者（新法民泊）については法律上の年間営業可能日数が180日であることから齟齬があり、得られた数字は全物件の稼働率とは乖離しているものと考えられる¹⁶。そのため、観光庁により公表 [12]されている2019年6月－7月稼働率、33%を用いて推計する。特区民泊については75%を採用した。上記のように、2019年6月13日の価格・件数・稼働率から大手民泊事業者の年間売上を推計した¹⁷。

単価、件数、稼働率を乗じて売上を求めた結果、新法民泊の推計取引規模は約400億円、特区民泊の推計取引規模は約400億円となり、あわせて2019年の民泊市場規模の推計値（用いたデータがかなり限定的であるうえ、一定の仮定に基づいて推計されたため、参考値として扱うことが妥当である）は約800億円となった。

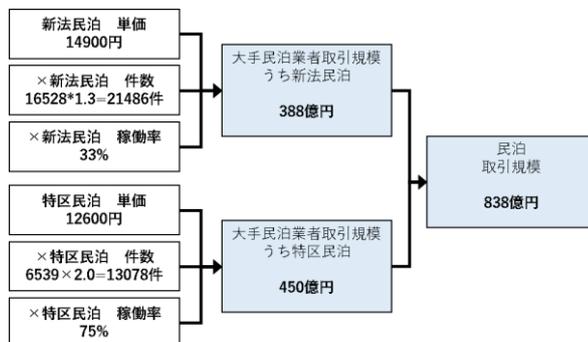
¹⁶住宅宿泊事業者についてスクレイピングによる予約率と法律上可能な予約率が乖離する理由は不明だが、年間営業日数の上限に達したリスティングはそもそもシステム上表示されず稼働率の母数から除かれてしまっている可能性や、ウェブ上の予約不可能日は予約済日とは限らず、元々宿泊を受け付けていない日であり、予約不可能日の割合を稼働率とすることで過大評価が生じた、等の可能性が考えられる。また、上記の観光庁による稼働率は届出ごとに対するものであり、リスティングごとの稼働率とは異なる可能性がある。

¹⁷6月の価格を用いて推計していることで、図表17の民泊の価格やホテルの消費者物価指数を見ると6月の価格と全期間の平均価格に10%弱の差があることから、同程度の誤差が生じうる。また、平日の価格を用いていることで、図表7の価格やホテルの消費者物価指数を見ると平日と休日の間に20%程度の差があることから、休日の日数に応じて誤差が発生しうる。さらに、新法民泊の件数自体も、観光庁のデータ [17]によれば2019年初頭で12166件、2020年初頭で20525件であるので、6月時点での件数16528件を用いることでずれが生じている可能性がある。

〈図表9 大手民泊 予約から宿泊までの日数と予約率(2019年6月13日時点)〉



〈図表10 プラットフォーム・観光庁情報¹⁸を用いた年間民泊取引規模推計・参考値(2019)〉



(備考) ウェブスクレイピングによる大手民泊事業者サイト、民泊制度ポータルサイト「住宅宿泊事業法の届出状況」、内閣府地方創生推進事務局「国家戦略特区 特区民泊について」の情報を用いて推計した。

➤ クラウドソーシング

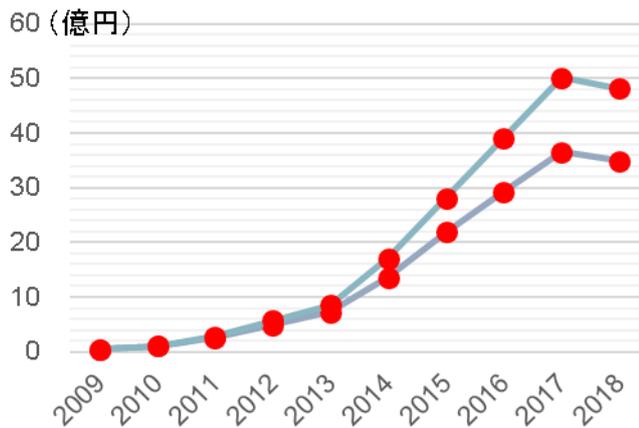
クラウドソーシングウェブサイトに対して、ウェブスクレイピングの手法を用いて取引規模を推計した(民泊同様、PythonのSeleniumモジュールを用いてスクレイピングを行った)。まずは、サービス提供者のリストを各都道府県・性別・年齢ごとに作成した。さらに、各提供者のページ上には過去の取引履歴が大まかな金額とともに記載されており、この取引金額を合計することにより、取引規模が推計できる。ここでは大手クラウドソーシングウェブサイトの各年の取引規模を推計した。2009年～2018年の全仕事件数約910000件のうち90000件をランダムに抽出し、サンプルの取引規模から全体の取引規模を推計した。(ただし、90000件のうち150件程度は時間単価しかわからないため、稼働時間を1時間のみとして推

¹⁸ 新法民泊・特区民泊の全体の件数のみ観光庁情報、他はスクレイピングによる。

計した。このため、取引規模に 0.1%程度の誤差が生じる可能性がある。）

その結果、取引規模の推移は図表 11 のようになった。なお、参考までに、クラウドソーシング市場は内閣府報告書 [2]でのプラットフォームへのヒアリング結果により寡占状態であることがわかっており、競合のうちの 1 社であるクラウドワークス社では 2017 年売上が 73 億円、2018 年売上が 123 億円となっている。

〈図表 11 大手クラウドソーシングサイトの取引規模推移 (2019)〉



(備考) 赤い点は各年での推計の上限、下限を表す。

(2) 付加価値額の推計のための財／サービス提供者へのアンケート調査に関して

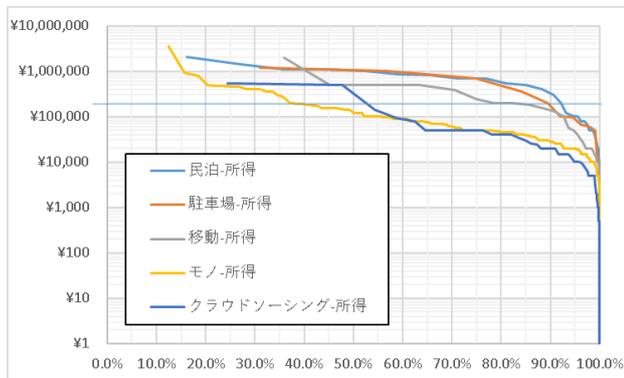
シェアリング・エコノミーの付加価値額計測のためには、財／サービスの提供者である個人事業主の中間投入構造の把握が必要である。内閣府調査 [2]では提供者に対してアンケート調査を行うことで中間投入構造の把握を図った。これは、試験的な推計という位置づけであったが、本調査では、今後の付加価値額推計の実装へ向け、提供者に納税情報からの転記により調査票を記載してもらうことで、精度の向上が可能であるかどうかを確認した。

経済センサス活動調査の個人経営調査票は費用の内訳（給料賃金、地代家賃、減価償却費、租税公課について）を確定申告書類から転記する前提として設計されており、シェアリング・エコノミーの費用内訳も確定申告書類からの転記が可能かを検討する事は意味があると考ええる。以下では、シェアリング・エコノミーの提供者が確定申告を行っているか、また費用内訳を申告書に記載しているかという観点から確認を行った。

まずは、売上高基準で、シェアリング・エコノミー提供者のどれだけの割合が確定申告を行っているかを確認した。内閣府調査におけるシェアリング・エコノミーのカテゴリ別の所得累積分布（図表 12）は、縦軸の金額以上の所得を持つシェアリング・エコノミーの提供者だけで、各カテゴリの全所得のどれくらいの割合を占め

るかを横軸に示したものとなっている。ここで、確定申告が必須となる 20 万円以上の所得の提供者（図表 12 の青線部分）だけで、民泊、駐車場（スペースのシェア）、移動のシェアについては 80%以上の所得を占めることがわかる。モノのシェア、クラウドソーシングについても 40%以上の所得を占め、納税情報からの転記を前提とした調査は大部分の所得を捕捉出来るため有効である可能性がある。ここでは、その有効性の確認のため、実際に納税書類からの転記が可能であるかの確認を行った。

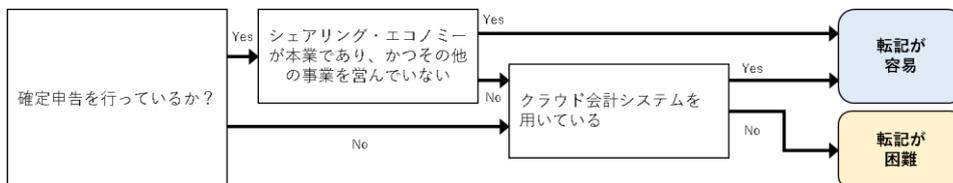
〈図表 12 シェアのカテゴリ別 個人事業主所得累積分布¹⁹⁾〉



（備考） [2]で収集したデータを ESRI の許可を得て使用、筆者作成

確定申告書類・申告手続きを確認した結果、図表 13 のように、確定申告を行っていて、かつシェアリング・エコノミーが本業とみなされる場合、あるいは（クラウド）会計システムを使用している場合は、既に粒度が十分なデータが作成されていることになるため、調査票情報の記入が比較的容易であると考えられる。

〈図表 13 納税書類/会計システムからの転記の可能性〉



ただし、確定申告を行っている場合でも、シェアリング・エコノミーを営んでいる個人事業主はあくまで副業として行っている場合が多いと想定され、その場合、

¹⁹ 一定以上の所得の合計だけで全体の所得の何%を占めるかをグラフにした。縦軸が所得の下限、横軸が全体の所得に占める割合を示す。累積分布のサンプル数は、民泊のシェア：38 件、駐車場のシェア：14 件、移動のシェア：8 件、モノのシェア：34 件、クラウドソーシング：71 件とやや少ないため、注意が必要である。

得られた所得は雑所得として申告され、収支内訳書を提出する義務は無い²⁰ [13]。

(図表 14 参照) よって、個人事業主について、確定申告書からの転記は、確定申告の際に費用内訳も記載する必要がある法人と違って困難である可能性が高い。

〈図表 14 確定申告書における雑所得の記入〉

○ 雑所得 (公的年金等以外)、組合費等の配当所得・譲渡所得、一時所得に関する事項				
所得の種類	種目・所得の生ずる場所	収入金額 円	必要経費等 円	差引金額 円

(備考) 確定申告書 B (個人事業主等向け書式) より抜粋

一方、近年急速に普及している(クラウド)会計システムのユーザーであれば、システム上のデータを用いて、中間投入構造に関するデータを出力することが可能ではないかと考えられる。

よって、すべてのシェアリング・エコノミー提供者について有効な捕捉方法は現時点ではないものの、シェアリング・エコノミーを本業として営んでいる者、あるいは(クラウド)会計システムを利用している者については、中間投入構造について詳細な情報が得られる可能性がある。

(3) デフレーター推計に関して

➤ 民泊

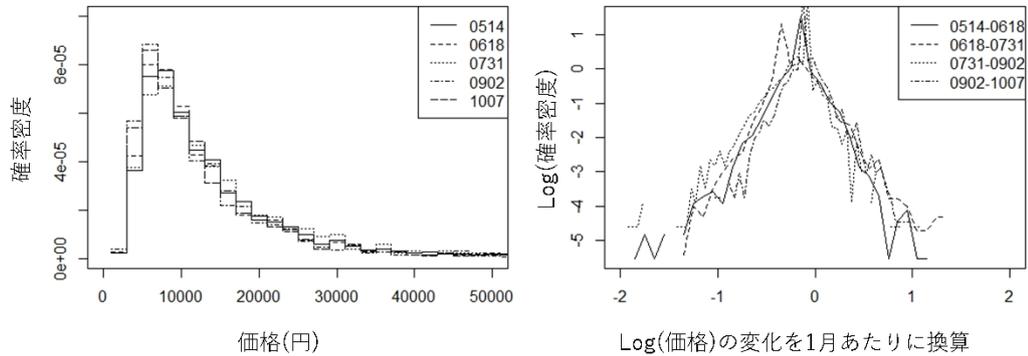
(取引規模の試算の際と同様に) 大手民泊業者のリスティングを全国から 4000 件ランダムに抽出し²¹、個々の物件に対して、5月～10月にわたる価格の分布(図表 15)と価格比分布(図表 16)を確認したうえで5月～10月のデフレーターを計測した。その際、5月の価格をベースとし、5月と対象月の両方にデータがある物件のみを用いてデフレーターのインデックスを計算した。その上、2018年における同期間の消費者物価指数のホテル宿泊価格との比較を行った。

5/14 から 10/7 まで五時点でのリスティングの価格分布は図表 15 のように似通った分布になり、概ね価格分布は時間的に安定していると考えられる。

²⁰ 上記事実については税務署に確認を行った。

²¹ ただし、全てのリスティングを含むため、民泊以外の「旅館業法に則った物件」や「その他の物件」が含まれてしまっていることに注意

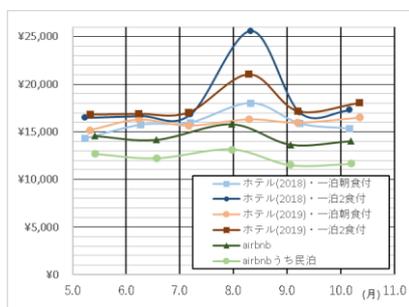
〈図表 15 リスティングの価格分布〉 〈図表 16 リスティングの価格比分布〉



一方、二期の価格比の分布は図表 16 のようにこちらも概ね安定的な分布になった。データの取得日が等間隔ではないため、1 月あたりの $\log(\text{価格比})$ に換算した。2 期の価格をそれぞれ p_1 、 p_2 とし、取得日の間隔を Δt 日とすると、横軸を $(31/\Delta t) \log(p_2/p_1)$ とした。

最後に、実際に 5 月～10 月のデフレーターを計算した。大手民泊業者サイトの 4000 件のサンプルについて、価格取得日から直近の宿泊日について 1 名・1 泊分の宿泊価格を取得し、ラスパイレス式で指数を求め、ホテル宿泊料との比較のため 5 月 14 日時点での宿泊価格を乗じて求めた。その結果、ホテル宿泊料の物価指数（2018 年、2019 年）とスクレイピングにより得られたデフレーター（2019 年）の推移は概ね同じような同程度の水準になったが、同様の季節性を持つとは明確に判断できない結果となった（図表 17）。民泊サイトのリスティングのうち新法民泊・特区民泊の物件（図表凡例の「airbnb うち民泊」に該当）に絞って価格指数を計算した場合、airbnb 全体の水準より幾分低い水準となった。

〈図表 17 ホテル宿泊料・大手民泊サイトリスティングのデフレーターと比較〉



（備考）ホテル宿泊料は総務省統計局「小売物価統計調査」（2018）のデータを用いた。民泊のデータはここまでと同様、大手民泊業者サイトから取得した。

今回は、比較する 2 時点間で価格データが入手できる物件を対象としてデフレ

ーターを計算した²²が、厳密には、2 時点間で新規参入・退出を行った物件に対して、その価格が相場よりも低いか高いかによってデフレーターへの影響を与えると考えられる。その場合、厳密にはヘドニックアプローチを用いて、物件が存在しない時点での価格も推計し、デフレーターの計算に含めるべきであると考えられる²³ [14]。ただし、民泊の品質変化は IT・家電製品等に比べ小さいと思われるので、ヘドニックアプローチを使う必要性はそこまで高くないと思われる。²⁴

▶ クラウドソーシング

民泊と違って提供される財・サービスが毎回の契約ごとに異なり、さらに財の中での異質性も高いため、そもそもデータから物価指数というものを定義することが妥当かという問いがある²⁵。提供される業務内容が種類別に分類されているため、そのうちの「ウェブサイト制作・デザイン」に絞って価格分布を見た（大手クラウドソーシングサイトからデータをスクレイピングにより取得したうえで、取引成立の時点について各年ごとに分けて集計した）（図表 18）。その結果、価格は大きなばらつきを見せ、タスクの質・量も大きな差異があることが予想される²⁶。しかし、価格分布は下記の通り、年次が異なってもある程度似通ったパターンを見せた。これは、クラウドソースされるタスクの量・質の違いを無視したうえで、同一カテゴリに含まれるタスクをひとつの財・サービス（のバスケット）としてみなすことに一定の合理性があることを示唆している。²⁷

²² 対象日が予約済で価格情報が取得できない場合、先述のように対象日から直近の予約可能日の価格を用いた。また、宿泊料が 100 万円を越える物件は除外した。

²³ 例えば、新たに市場に参入した物件が、その品質に比して市場価格より割高であれば、デフレーターを上昇させるものと思われる。そのような場合に「市場価格」の水準を求めるためヘドニックアプローチが必要となる。

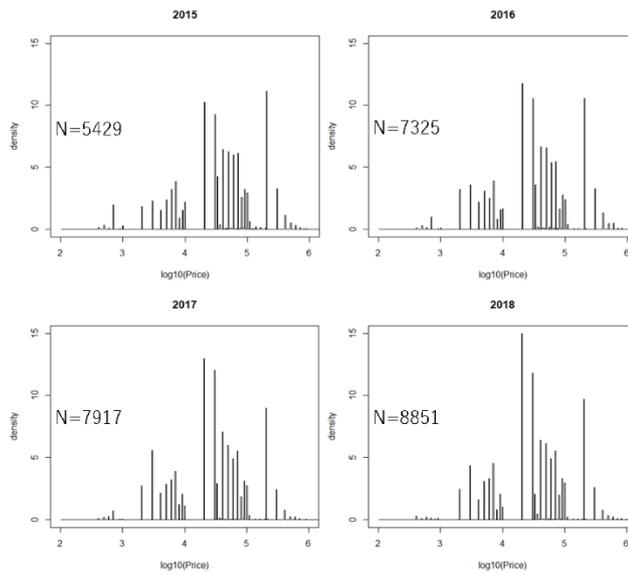
²⁴ IT・家電製品は 1987 年から 2015 年にかけて -90%も減少しているが [16]、ホテルの消費者物価指数は 1993 年から 2018 年にかけて高々 +5%程度しか変化していない [15]。よって、IT・家電製品に比して、重大な品質の変化はないものと考えられる。さらに、民泊の場合同一物件に関する比較が異なる時点で行えるため、ヘドニックアプローチを用いる必要性は高くないと思われる。

²⁵ クラウドソースされるタスクの分量や品質のようなものは、公開情報としては入手不可能であるため、尚更財の間での価格の比較が難しい。

²⁶ クラウドソーシングのような質・量が取引ごとに異なる取引ではヘドニックアプローチを使う意義があるが、プラットフォーム上に質・量に関する情報が比較可能な形で掲載されていないため、同手法を用いることが困難である。

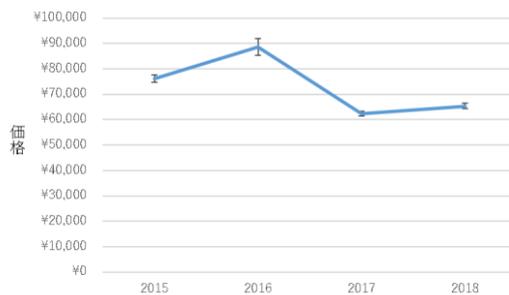
²⁷ ただし、価格指数の変動が質・量の変化を反映しないものになってしまうという意味で、妥当性に疑問が生じる部分もある。クラウドソーシングの場合、民泊と異なり一定の質の財・サービスが継続的に生じるような状況とは異なることに注意。

〈図表 18 Web 制作サービスの価格分布〉



続いて、上記分布から 2015 年～2018 年の 4 期の平均価格を計算した。各期に観測された分布が母分布と等しいものとして、そこから同数サンプルを非復元抽出にて観測した際の指数の標準偏差を図表 19 のエラーバーの幅とした²⁸。

〈図表 19 Web 制作サービスの価格指数変動〉



4. まとめ・今後の課題

本ノートでは、ウェブスクレイピングを用いたシェアリング・エコノミーの生産額・デフレーターを試算を行い、また中間投入構造の調査において税務調査の活用が困難であるという結論を得た。特に、民泊とクラウドソーシングについて、ウェブスクレイピングを用いることにより、市場規模とデフレーターの推計が可能であることを示した。プラットフォーム事業者へのヒアリングやその他の推計手法が困難である場合、スクレイピングを用いた

²⁸ 誤差は、観測された分布を母分布と同じとみなして観測された標準偏差/ $\sqrt{\text{標本数}}$ によって求めたため、誤差を過小評価している可能性が大きい。

推計により代替することに一定の合理性はあるものと考えられる。

課題としては、生産額の推計には消費される財・サービスの種類等（民泊の場合宿泊人数等）に依然いくつかの仮定を置かざるを得なかったこと、取得に相応の手間がかかることから、実推計を行う際は事業者から直接データを取得することが望ましいこと、また、デフレーター計測に関して、スキル・時間のシェアについては、取引の特性上財・サービスの（ウェブ情報からは判断できない）量・品質のばらつきが極めて大きいため、デフレーターの定義に注意が必要であることが挙げられる。

参考文献

- [1] 経済財政諮問会議, “統計改革の基本方針,” 2016.
- [2] 内閣府 経済社会総合研究所, “シェアリング・エコノミー等新分野の経済活動の計測に関する調査研究 報告書,” 2018,2019.
- [3] 総務省統計局統計調査部消費統計課物価統計室長 中村英昭, “統計 Today No.130 消費者物価指数へのネット販売価格の反映,” [オンライン]. Available: <https://www.stat.go.jp/info/today/130.html>.
- [4] Statistics Canada, “Measuring the Canadian economy in increasingly digitalized world,” 2018.
- [5] Statistics Canada, “Measuring private short-term accomodation in Canada,” 2019.
- [6] 総務省統計局, 著: *物価指数研究会 (第10回)*, 2018.
- [7] 安部展弘, 篠崎公昭, “価格比較サイトのビッグデータと機械学習手法を用いた物価指数の試算: 品質調整方法の比較分析と妥当性の検証,” 日本銀行ワーキングペーパーシリーズ no.18-j-6, 2018.
- [8] 総務省統計局, 著: *物価指数研究会(第5回)*, 2015.
- [9] Office for National Statistics, “Web-scraping policy,” [オンライン].
- [10] 経済産業省, “我が国におけるデータ駆動型社会に係る基盤整備(電子商取引に関する市場調査),” 経済産業省 商務情報政策局 情報経済課, 2018.
- [11] airbnb 株式会社, “民泊新法に向けた記者説明会,” 2018.
- [12] 観光庁, 住宅宿泊事業の宿泊実績について(令和元年6月-7月分: 住宅宿泊事業者からの定期報告の集計), 2019.
- [13] 国税庁, “タックスアンサー No.1906,” 2018. [オンライン]. Available: <https://www.nta.go.jp/m/taxanswer/1906.htm>.
- [14] J. Triplett, “Handbook on Hedonic Indexes and Quality,” OECD, 2004.

- [15] みずほ総合研究所, “みずほりポート 2018 年 8 月 29 日 ホテル市場の変調の兆しをどうみるか”.
- [16] 熊倉正修, “デフレと消費者物価指数の品質調整,” 『経済統計研究』 第 43 巻 I 号, 2015.
- [17] 観光庁, 民泊制度ポータルサイト minpaku 住宅宿泊事業法の施行状況, 2020.