

TCER Working Paper Series

マイクロデータを用いた計量経済学的接近法 -ヘドニック法を用いた品質調整問題を題材として-

Econometric Approach using Microeconomic Data

清水千弘
Chihiro Shimizu

新井優太
Yuta Arai

2024年 1月

Working Paper J-28

<https://www.tcer.or.jp/wp/pdf/j28.pdf>

TCER 

公益財団法人東京経済研究センター
〒102-0072 東京都千代田区飯田橋1-7-10-703

©2024 by Chihiro Shimizu and Yuta Arai.

All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including ©notice, is given to the source.

概要

マイクロデータを用いた計量経済分析は、多様な大規模データの入手可能性が飛躍的に高まるとともに、大容量のデータを処理可能な計算機の普及、分析用プログラムの発達により、急速に進化してきている。しかし、マイクロデータが公開されていることは少ないため、データの希少性が研究の評価となることも依然として残っている。そのため、データ収集能力が研究推進力の大きな比重を占めることには変わりがない。また、実際にデータが取得できたとしても、データのノイズを取り除き、データを加工し、分析できるようになるまでには、高いデータ加工能力が要求される。また、モデルの推計においても、高度なプログラミング能力が要求される。本稿では、Diewert and Shimizu (2024) を題材として、一連の研究の中で行われた、a) データの整備・加工、b) 推計手順、そして推計された結果のc) 評価・検証などの具体的な手続きを公開し、汎用的なプログラム言語で再現し、網羅的に整理した。さらに、推計された結果の解釈を含む研究手法 (Research method)、研究リソースを公開することで、今後のマイクロデータを用いた計量分析、特にヘドニック・モデルに関する研究を進める大学院生や、これから本分野の研究を始める研究者に対して、一定の指針を示す。

清水千弘
東京経済研究センター (TCER) 及び
一橋大学
ソーシャル・データサイエンス研究科
東京都国立市中2-1
c.shimizu@r.hit-u.ac.jp

新井優太
麗澤大学
経済学部
千葉県柏市光ヶ丘団地2-1-1
yuarai20@reitaku-u.ac.jp

Abstract

Econometric analysis using microdata has evolved rapidly due to the dramatic increase in the availability of big data, the proliferation of computers capable of processing large volumes of data, and the development of econometric software programs. However, since microdata are rarely publicly available, the scarcity of data still poses a challenge for economic research. Therefore, the ability to collect data remains a major driving force in research. Additionally, even when data can be acquired, high data processing capabilities are required to remove data noise and effectively process and analyze the data. Advanced programming skills are also essential for estimating models. This paper focuses on Diewert and Shimizu (2024) as its subject matter and thoroughly organizes the specific procedures conducted in a series of studies, including: a) Data preparation and processing, b) Estimation procedures, and c) Evaluation and validation of the estimated results. We achieve this by reproducing these procedures in a general-purpose programming language R. Furthermore, by disclosing our research methods, including the interpretation of the estimated results, and sharing our research resources, we aim to provide guidelines for new researchers embarking on future studies in econometric analysis using microdata, especially hedonic models. These guidelines will also benefit new researchers entering this field in the future.

Chihiro Shimizu
TCER
and
Hitotsubashi University
School of Social Data Science
186-8601
c.shimizu@r.hit-u.ac.jp

Yuta Arai
Reitaku University
Department of Economics
2-1-1 Hikarigaoka, Kashiwa, Chiba 277-8686
yuarai20@reitaku-u.ac.jp

マイクロデータを用いた計量経済学的接近法 -ヘドニック法を用いた品質調整問題を題材として-

新井優太(麗澤大学)[†]・清水千弘(一橋大学)[‡]

要旨

マイクロデータを用いた計量経済分析は、多様な大規模データの入手可能性が飛躍的に高まるとともに、大容量のデータを処理可能な計算機の普及、分析用プログラムの発達により、急速に進化してきている。しかし、マイクロデータが公開されていることは少ないため、データの希少性が研究の評価となることも依然として残っている。そのため、データ収集能力が研究推進力の大きな比重を占めることには変わりがない。また、実際にデータが取得できたとしても、データのノイズを取り除き、データを加工し、分析できるようになるまでには、高いデータ加工能力が要求される。また、モデルの推計においても、高度なプログラミング能力が要求される。本稿では、Diewert and Shimizu(2024)を題材として、一連の研究の中で行われた、a)データの整備・加工、b)推計手順、そして推計された結果のc)評価・検証などの具体的な手続きを公開し、汎用的なプログラム言語で再現し、網羅的に整理した。さらに、推計された結果の解釈を含む研究手法(Research method)、研究リソースを公開することで、今後のマイクロデータを用いた計量分析、特にヘドニック・モデルに関する研究を進める大学院生や、これから本分野の研究を始める研究者に対して、一定の指針を示す。

キーワード

Quality adjustment, scanner data, hedonic regressions, predicted share similarity linking, expanding window approach to multilateral indexes, the chain drift problem, economic approach to index number theory.

Journal of Economic Literature Classification Code

C32, C43, D20, D57, E31

* 本稿は、Diewert, E and C. Shimizu (2024), “Product Churn and Quality Adjustment”. TECR Working paper, E-197 を日本語に翻訳するとともに、一連の分析手続きを、公開するものである。同論文では、分析手続きは Shazam によって推計されたが、それを汎用的な R プログラムに変換するとともに、一連の推計手続きの中で留意すべき点を解説した。本稿の執筆にあたり、一橋大学ソーシャル・データサイエンス研究科 馬場弘樹氏、西颯人氏には、原稿に目を通していただき、多くの示唆をいただいた。ここに記して御礼を申し上げます。

[†] 麗澤大学経済学部・助教

[‡] 一橋大学大学院ソーシャル・データサイエンス研究科・教授, e-mail: c.shimizu@r.hit-u.ac.jp.

1. はじめに

マイクロデータを用いた計量経済モデルは、多様なデータ資源の入手可能性が高まり、高い計算能力を持つ計算機の普及、汎用的な計算プログラムの開発などによって、年々研究者だけでなく、エコノミスト、学生に至るまで広く利用されるようになってきている。また、その推計手法の開発も目覚ましいものがある。計量経済モデルの背後には、経済モデルが存在することは言うまでもない。データサイエンスという研究領域が発達してきている現在においては、伝統的な経済モデルと融合されながら、経済学、データサイエンスの二つの領域が発展していく可能性が期待される場所である。

ミクロ経済分野において、経済理論と計量経済モデルが融合し、多くの研究論文が毎年出版されている。ミクロ経済学の研究の主要分野として、価格の解明が挙げられる。しかし、消費者が購入できる財やサービスは、それぞれに価格が設定されているが、その数は数百万にのぼる。生産側においては、企業は最終消費のための財やサービスを購入しているわけではなく、他の生産者が必要とする輸出財や中間財も生産しているため、企業は総計で数百万種類の輸入財・サービス、数千種類の労働サービス、数十万種類の特定の資本を使用している。さらに、物理的な商品を地理的に異なる場所や、生産または消費される季節や時間帯によっても区別すると、先進国において各年度内に取引される財・サービスの数は、何十億にまで膨れ上がり、それごとに価格が設定されている。このようなすべての財やサービスの価格をすべて捕捉することは困難であるために、物価指数や生産性を表す集計された統計量(GDP等)に情報を集約していくことになる。そのため価格分析は、その背後にあるマイクロな構造を理解するだけでなく、マクロ経済指標、集計指標を作成していく上で、極めて大きな貢献をしているといっても良い。

また、そのような膨大な数の価格を捕捉することが出来るデータ資源としては、POS(Point Of sales)データが登場し、価格を取り巻く研究が飛躍的に発達してきた。POS データでは、JAN(Japanese Article Number)コードと呼ばれる識別番号で事業者や商品とともに、商品が売れた段階でレジなどに記録されていく。そのため、多くの財の価格がリアルタイムで取引された場所も含めて捕捉が出来るようになってきたのである(付論A)。

本稿では、ミクロ経済分野でもヘドニック・モデルに注目する。その理由としては、価格分析において、ミクロ経済理論と計量経済理論とが融合して高度に発達してきた研究領域の一つであることが挙げられる。現実の市場に出回っている商品の多くは、使用目的が同じであったとしても性能や機能面で多くの差別化が図られていることが一般的である。性能や機能面での違いは、その消費を通じて家計が得ることのできる効用が変化していくために、その商品の市場価格に反映される。マクロ経済の政策立案や運営に必須の物価指数を計算する際にも、このような差別化された商品の価格と消費者行動の関係を明示化し、時間的な品質の異質性を制御したうえで、生計費指数を効用不変指数として計測していくことが要求される。

例えば、本稿で演習として取り扱うパソコンのように、日進月歩で技術革新が起こる財は、他にもテレビ、デジタル・カメラなどのデジタル家電製品が挙げられる。このような財は、たとえ価格が同じであったとしても、性能は向上し、機能も高度化されることが多い。そのため、実際の価格は低下していると考えべきである。このような技術革新、それに伴う商品の入れ替わりは、実質的に価格の低下をもたらすことになる。

近年における技術革新の例を挙げるのであれば、スマートフォンではないだろうか。例えば 2007 年に初代 iPhone が発売され、2009 年に発売された iPhone 3GS にはカメラにオートフォーカスとビデオ撮影機能が搭載された。2014 年に発売された iPhone6 では Apple Pay が搭載された。このように、既存製品にはなかった機能が、技術革新によって新製品には搭載されるようになるのである。高性能化した分、高価格化してしまうため、価格指数を作成する上で品質調整の果たす役割がどれだけ重要であるかがわかるであろう。

そのため、品質補正をしていなかったり、基準時点のマーケット・バスケットを固定したりした場合には、物価指数は商品の価格実態を正しく反映していないことになる。物価指数を測定する際に、新商品や既存商品の品質向上に対する調整を行わないと、価格を誤測定する要因になることは以前から知られていた。たとえば、ボスキン委員会の報告書 Boskin et al.(1996)¹では、米国の消費者物価指数が真のインフレ率を年率1.1%過大評価していると推定しており、このうち新商品と品質の変化を考慮しなかったことが半分以上の0.60%を占めているとされている。

わが国で作成されている総務省統計局による消費者物価指数においても、こうした問題点を踏まえた上で、一部の商品についてヘドニック・アプローチによる品質調整を行っている。ヘドニック・アプローチとは、ある商品の価格をさまざまな性能や機能の価値の集合体(属性の束)とみなし、広い意味での統計学における回帰分析のテクニックを利用して商品価格を推定する方法であると定義することができる。² 推計モデルでは、商品価格は属性の束からなる方程式で表現され、このような式はヘドニック価格関数とよばれる。

このようにヘドニック・モデルは、ミクロ経済領域におけるデータサイエンスを学ぶ上で、最適な教材の一つといっても良いであろう。しかし、大規模なマイクロデータに入手可能性が向上したからと言って、すぐに計量経済学的な接近法が適用できるわけではない。マイクロデータには、欠損値が存在するし、異常値・外れ値が含まれていることも多く、また、モデルの構築には、複数のデータベースを結合させることも要請される。さらに、マクロ計量分析と比較して、多次元のデータを扱うことが多いため、多重共線性などの問題にも直面しやすい。一方で、十分に変数が収集できないときには、過少定式化バイアスの問題に代表されるように、推定された値が一致性も不偏性も持たない問題にも直面することが多い。

データを用いた研究を実施していく中では、a)リサーチトピックを確定し、仮説を設定した上で、b)リサーチデザインを行い、c)理論モデルと推計モデルを開発し、モデルに応じて d)必要とされるデータを収集・加工または作成し、はじめて e)モデルの推計を行うことができる。このモデルの推計には、推計プログラムを作成しなければならない。推計プログラムには、さまざまな言語やソフトウェアがあるため、その特徴を理解したうえで、ふさわしい言語、ソフトウェアを選択する。³また、推計された結果を f)解釈・評価し、その結果と理論モデルや経験値と比較したうえで、適合するまで推計を繰り返し行う。この時にはデータを再取得、作成しなければならないことも多い。さらに、推計ができた後に、g)推計結果を議論し、h)論文を執筆していく。経済学分野において、研究論文を執筆していく中では、それぞれの過程で研究を進めるうえでの技術や専門性が要求されることになる。⁴

本稿では、Diewert and Shimizu(2024)を題材として、一連の推計の中で利用されてきたデータの入手方法からデータの加工方法、推計手順などを整理し、汎用的なプログラム言語で再現し、それを公開することで、今後のマイクロデータを用いた価格分析、指数理論、ヘドニック・モデルに関する研究、広い意味でのマイクロデータを用いた計量経済分析を行う大学院生やこれから当該野の研究を始める研究者に対して、その出発点となるように、研究のリソースを公開することを目的とする。

¹ <https://www.ssa.gov/history/reports/boskinrpt.html>

² 詳細は、清水・唐渡(2007)第2章および Diewert et al (2022) Chapter 2 参照されたい。

³ 近年では、Rなどの汎用的な言語が開発され、ライブラリなどが充実してくる中で、おおよその統計モデルの推計、統計量の計算はできるようになってきている。古くは、Fortran や C++、R の前身となる S 言語、汎用プログラムソフトとして、SAS、SPSS と限定されたものしか利用できなかったが、1990年代に入ると Shazam、TSP が普及し、さらに Gauss、Matlab、EViews、Stata といったソフトウェアが登場してきた。近年では Python なども利用されている。

⁴ このような一連の手続きは、研究方法(Research Methods)という分野が確立されており、大学・大学院の講義として提供されている場合もある。一橋大学大学院ソーシャル・データサイエンス研究科では、統計分析発展(実践)として講義が提供されている(2024年時点)。

2. 価格指数の推計と統計問題

物価指数 (Price Index) は、ある一定の期間における財・サービスの価格の変動を示す経済指標である。物価指数の測定においては、特定の基準年度と比較して、物価の変動を測定していくが、価格水準の変化を単純に調べているわけではない。物価指数の推計においては、価格指数(P)と数量指数(Q)を、経済学的・公理的・確率論的な条件と照らして、理想的な指数として集計していく(Diewert 2022)。この価格(P)と数量(Q)の間には、経済学の教科書が示すように、正常財では、価格が上昇すれば数量は減少するといった負の相関が存在している。さらに、複数の財を考えたときには、代替財や補完財が存在し、ある商品の価格の変化が、該当する商品だけでなく、他の商品の取引数量にも影響を及ぼすことになる。まずこのようなマイクロ経済学の基本的な貢献を踏まえて、データサイエンスの視点から、複数の財の価格(P)と数量(Q)をどのように集計していくのかという問題を整理する。

通常の Laspeyres 指数に代表される指数算式は、すべての価格(P)と数量(Q)が連続して市場で観察できるということを前提としている。しかし、実際の市場で観察される財の中には、特定の季節でしか生産または消費ができないような生鮮食料品や、正月やクリスマスなどの一部の時期でしか消費されないような財が存在し、このような財を「季節財」と呼ぶ。価格指数の推計問題においては、特定の季節に取引がなくなる中で、価格(P)も数量(Q)も観察が出来ない、つまり、周期的に、そして一時的にデータが欠損してしまうといった問題にどのように対応したらよいかといったことが課題として取り扱われてきた。そこで本節では、指数理論の基本算式と併せて、このような特定の季節で情報が欠損する場合での価格指数の推計方法を学習する。

また、パソコンなどのデジタル財は、製品によって品質が異なる。さらに、同じパソコンが継続して販売されているわけではない。新しい製品が技術進歩とともに市場に投入されてくる。古い製品は、製品の製造が中止され市場から撤退していく中で、二度と市場に出現してくることはない。このように品質が製品によって異なる場合や、時間とともに製品が入れ替わる場合には、品質を固定化して物価指数を測定していくことになる。その品質補正のための主要な方法の一つがヘドニック・モデルとなる。ヘドニック・モデルに代表される品質補正の方法は、3 節以降で整理する。

それでは、Diewert(2022)に基づき、基本的な指数算式に基づく計算方法と季節財の物価指数の計算方法をプログラミングと併せて学習しよう。まず、基本的な集計問題について簡単に整理する。ここでの実際の計算方法と計算式については、「付論 B1:micro01_basicindex」をご覧ください。

現在は多くの国において消費者物価指数(Consumer Price Index:CPI)を作成しているが、そのほとんどで Laspeyres 型の算式に基づいた指数が計算されている。これは上位指数(あるいは上位集計)と呼ばれるもので、一般的に物価指数は下位指数(あるいは下位集計)を上位指数に集計するという2段階で作成されている。例えばパンについて考えた場合、「食パン」「アンパン」「カレーパン」などの品目があり、これらの価格指数からある一定のウェイトとの加重平均によってパンの物価指数が作成される。

日本における物価指数の場合、最初に市町村別ごとに品目別の価格指数を算出し、次に全国及び地方・年階級別の品目別価格指数を算出した後、それぞれの地域ごとの上位類及び総合指標を算出している。最初の品目別の価格指数については、調査員が測定した品目の価格に基づいて計算される。そのため、下位集計としては価格のみを用いた物価指数が作成される。価格のみを用いた物価指数として、ここでは Carli 指数、Dutot 指数、Jevons 指数がある。最初に紹介するのは、歴史的に最も古い下位集計用の重み付けなし指数として Carli 指数である。これは N 個の財について、基準時点と比較時点との相対価格の平均値を計算したものである。次に、Dutot 指数は、基準時点における平均価格と比較時点における平均価格の比を計算したものである。重みなし指数の最後に、Jevons 指数は、比較時点と基準時点との価格比の幾何平均を計算するものである。

ここまで紹介してきた3つの指数は、いずれも価格情報のみを用いた指数であるが、それぞれ長所と短所を持っている。例えば Carli 指数は、式を理解しやすい反面、「時間反転性」という性質を満たさない。これは、基準時点と比較時点を入れ替えたときに、指数が元の指数の逆数になる性質をいう。一方で、Dutot 指数と Jevons 指数はこの性質を満たすが、Dutot 指数は商品価格を調査する際の単位に依存してしまう欠点がある。また Jevons 指数は掛け算を計算していることから、どれか1つでも比較時点の商品価格が0だと指数が0になってしまう欠点がある。日本における下位集計では Dutot 指数が採用されている。

また、上位集計を行う際に先ほどのパンについて考えると、例えば食パンの価格が高くなると消費者は効用を一定に保つために、食パンではなくお米やうどんを購入する可能性がある。代替材の存在である。前述したように CPI は、毎月同じ効用を得るために支払わなければいけない金額を測定するための指標であるから、指数を作成するためには価格だけでなく数量も考慮しなければならない。そこで上位集計を行う際には、価格だけでなく数量が重みとして用いられる。この代表的な指数が Laspeyres 指数や Paasche 指数であり、特に Laspeyres 指数は日本における消費者物価指数の基本算式として採用されている。重み付き指数はこれ以外にも Fisher 指数や Törnqvist 指数など、様々なものがある。

Laspeyres 指数は基準時点を設定し、比較時点の価格で基準時点と同じ商品を同じ数量購入した場合に、支出がどれだけ変化するかを計算している。一方で、Paasche 指数は比較時点での数量を重みとして用いて、比較時点での数量だけ同じ商品を購入した場合に、基準時点と比較時点とでどれくらい支出が変化するかを表した指数になっている。

Laspeyres 指数についてももう少し現実と照らし合わせながら考えてみると、世の中の消費構造は年々変化するため、基準年から時間が経過するに従って計算される指数は実態から乖離してしまう。例えば、景気が悪く価格の安い商品が好まれるような場合には、そのような財の購入数量が増えるはずである。ところが、指数計算に用いられる n 種の商品の各数量は基準時点のものに固定されているため、安い財の購入数量が増える効果が反映されずに過少評価されてしまうというバイアスが生じる。一方 Paasche 指数では逆のバイアスが生じてしまう。そこで、Laspeyres 指数と Paasche 指数の幾何平均をとる Fisher 指数は、この2つの指数の間の値をとることになる。この指数は今日でも、Laspeyres 指数や Paasche 指数よりも優れた指数算式だとされている。その理由として、Fisher 指数は Laspeyres 指数や Paasche 指数が満たさない「時間反転性」をはじめとした、理想的な指数が満たすべき性質の多くを満たすためである。理想の指数が満たすべき性質のことを指数の公理と呼び Fisher(1922)で整理されている。また、阿部(2023)でも詳しく解説されている。

Laspeyres 指数や Paasche 指数はバスケットの中に入れた商品の数量を基準時点と比較時点で固定し、価格の変化による支出総額の比を比較する固定バスケット方式と呼ばれる物価指数である。この2つは時間反転性を満たさないが、時間反転性を満たす指数となるためには、大きく分けて2つの方法がある。1つは、数量を基準時点とも比較時点とも異なる第3の時点にすること。2つ目は、Fisher 指数のように基準時点における数量と比較時点における数量を、時点を入れ替えても変化しないような形で取り入れることである。2つ目の方法を用いて提案された指数として Walsh 指数や、Törnqvist 指数がある。1つ目で挙げた、基準時点でも比較時点でもない第3の時点での数量を用いた指数としては Lowe 指数、Young 指数、Young 指数の幾何平均をとった Geo Young 指数がある。

日本における消費者物価指数の基本算式として Laspeyres 指数が採用されていると述べたが、実は2020年基準の日本の消費者物価指数は Young 指数になっている。その理由として、2020年は新型コロナウイルス感染症の影響を受け、多くのカテゴリで支出が大きく低下したことがある。そのため、2019年と2020年の平均をとったものを重みとして用いている⁵。

⁵ 総務省統計局内に設置されている「物価研究会」においては、このウェイトをどのように決定したらよいかを取り巻き、多くの議論が展開された。ウェイトを変更しない案、2019年と2020年の平均を使う案、さらには将来における消費構造の変化に確信が持てた場合に修正する案などである。もっとも大きな問題は、海外バック旅行など、消費が消滅してしまった財への対応であった。筆者の一人は、その議論に参加し、ブリティッシュ・コロンビア大学の Diewert 教授との議論の結果、2019年と2020年の平均をとることを指示した。『「消費者物価指数2020年基準改定計画」の一部改定』より。

ここまでは基本的に、ある基準となる時点を固定した上で指数の計算を実行してきた。しかし前述したように、基準時点と比較時点での消費構造がかけ離れてしまうと、計算された価格指数も実態から乖離してしまう。そこで連鎖指数という指数の計算方法もある。連鎖指数では、1時点前を基準として計算した指数を、1時点前の時点で求まっていた連鎖指数に掛けることによって計算される。そのため、消費構造の変化などに対応できるのである。

続いて季節財を扱う上での問題点と指数の作成方法を整理していく。ここでの実際の計算方法と計算式については、「付論 B2:micro02_Seasonal Product」をご覧ください。国際通貨基金から出版された CPI マニュアルの改訂版では、季節財のデータとしてイスラエルの生鮮果実消費に関するデータを用いて紹介されている。ここでは、同マニュアルの概要を紹介している。

前述したように季節財とは、価格や数量が特定の季節に増加・減少する傾向をもった財(商品)である。その中でも、ある特定の季節にのみ市場で入手が可能な財と、1年を通して市場で入手できるものの特定の時期に価格や数量が変動する性質を持った財の2種類がある。実際にデータの中身を確認してみるとわかるように、欠損値(価格や数量が入力されていない状態)の存在が確認できる。これが前者の「ある特定の季節にのみ市場で入手が可能な財」の特徴であり、季節性の強い商品と呼ばれる。また後者を季節性の弱い商品と呼ぶ。

価格指数を計算する上で、強い季節性のある商品の取り扱いは非常に厄介である。なぜなら、ある商品が y 年の1月には存在するが、翌年の1月には存在しないかもしれないのである。この場合、先程説明したような指数の計算方法をそのまま適用することは難しい。そこで、この問題を回避するために、大きく分けて2つの方法がある。1つ目は直近の異なる年における同じ月の価格で補完する方法、2つ目は比較時点と基準時点の両方に存在する財のみを用いる方法(Maximum Overlap 法)である。ただし、欠損価格の補完を行ってしまうと、指数に不当なバイアスを生じさせてしまう可能性がある。そのため、財の価格が補完元時点における価格と変わらないことが確実な場合にのみ使用すべきであることが推奨されている。そうでない場合には、overlap 法を用いるか品質調整を行うべきだとされている。

また、弱い季節性を持った商品の場合、特定の時期に価格や数量が変動する。そのため、同じ月のデータを抜き出して価格指数を作成した後、サンプル期間における月次の価格指数を構築する方法や、ある年の異なる月に同じ商品が存在しても別の商品だとみなして年次の指数を作成する方法などが存在する。

さらに、このようなデータに対して連鎖指数を計算すると、チェーン・ドリフトが発生することも知られている。そこで、この問題に対処するための Similarity Linked 指数という指数算式も提案されている。この方法を用いると、季節変動やチェーン・ドリフトが抑制されるため、最良の指数であると考えられている。Similarity Linked 指数については、後程詳しく解説する。

3. ヘドニック理論とヘドニック・モデル

3.1. ヘドニック理論

製品の品質変化による類似製品の価格調整の問題は、購入者にとって製品の有用性または効用に関係する。各製品は、購入者に対する製品の効用を決定するさまざまな特性の量を持っており、通常ヘドニック回帰では、製品価格を製品の様々な価格決定特性の量に回帰させる。もしくは代替的な方法として、製品価格を製品ダミー変数に回帰させる。これらのヘドニック回帰モデルは、購入者の効用関数として特定の関数形で表現できる。本節では、製品価格を製品ダミー変数に回帰させる方法について考察し、次節では、製品特性に回帰させるモデルについて考察する。

いま、 N 個の商品があり、 T の時点があると仮定し、 $p^t \equiv [p_{t1}, \dots, p_{tN}]$ および $q^t \equiv [q_{t1}, \dots, q_{tN}]$ は、時点 t における製品の価格ベクトルおよび数量ベクトルを表すとする。⁶ また、価格と数量はすべて正であり、

⁶ 本節の分析は、Diewert (2022; section 5)の分析に従う。

価格と数量に欠損がないと仮定し、 N 個の製品の各購入者は、各時点において、次の線形関数 $f(q)$ を最大化すると仮定する：

$$f(q) = f(q_1, q_2, \dots, q_N) \equiv \sum_{n=1}^N \alpha_n q_n \equiv \alpha \cdot q \dots (1)$$

このとき、 α_n は正のパラメータであり、品質調整係数と解釈できる。 N 個の商品の各購入者について、効用最大化行動が仮定され、時点 t における各購入者に対して価格ベクトル p^t が与えられたと仮定すると⁷、数量ベクトル q^t は、効用最大化問題 $\max_q \{\alpha \cdot q : p^t \cdot q = e^t ; q \geq 0_N\}$ の解であることが示される。ここで、 e^t は、 N 個の商品に対する時点 t における総支出に等しい。時点 t における効用最大化問題に対する解 q^t の一次条件は次の $N + 1$ 個の方程式となる。ただし、 λ_t はラグランジュ乗数である。

$$\alpha = \lambda_t p^t \dots (2)$$

$$p^t \cdot q^t = e^t \dots (3)$$

(2)式の両辺に数量ベクトル q^t を掛けて内積を計算し、その結果生じる式を λ_t について解き、(3)式を用いると、 λ_t について以下の式が得られる：

$$\lambda_t = \alpha \cdot q^t / e^t > 0 \dots (4)$$

$$\pi_t \equiv 1/\lambda_t \dots (5)$$

(2)式の両辺を λ^t で割り、定義式(5)を用いると、時点 t における時間製品ダミー推定式が得られる：⁸

$$p_{tn} = \pi_t \alpha_n \dots (6)$$

すると、このモデルの時点 t における集約価格 P^t と数量水準 Q^t は、以下のように定義される：

$$Q^t \equiv \alpha \cdot q^t \dots (7)$$

$$P^t \equiv e^t / Q^t = \pi_t \dots (8)$$

ここで、(8)式の2番目の式は(4)式と(5)式を用いて導かれる。したがって(6)式は、製品 n の時点 t における価格 p_{tn} が、時点 t の価格水準 π_t に製品 n の品質調整係数 α_n を掛けたものに等しいと解釈できる。⁹

この時点で、(6)式の消費者理論による導出がすべての経済学者に受け入れられているわけではないことに注意する必要がある。Rosen (1974)や Triplett (1987) (2004)は、需要条件だけでなく供給条件にも基づく、ヘドニック回帰モデルの導出に対するより一般的なアプローチを主張している。現在のアプローチは、明らかに消費者(または購入者)の選好のみに基づいている。この消費者嗜好のアプローチは、Griliches

⁷ これらは強い仮定であるが、ヘドニック回帰モデルを対象商品の効用に関連付けるためには、強い仮定が必要である。

⁸ このモデルは Court(1939;109-111)に遡る。式(6)の両辺の対数を取り、誤差項を加えることによって、これらの方程式を変換した。Diewert(2003b)(2023)は、基本方程式(6)の代替変換を行うことによる指数数の影響を検討し、最終的に Court の変換を支持した。

⁹ α_n は $n = 1, \dots, N$ の場合、製品 n の1単位の購入者の限界効用であることに注意されたい。 $c(p)$ は効用関数 $f(q)$ と双対的な単位費用関数であるとして、 π_t が時点 t における価格指数と等しくなることは Diewert (1974)を参照されたい。 $p = [p_1, \dots, p_N]$ を正の価格ベクトルとすると、線形効用関数 $f(q) = \sum_{n=1}^N \alpha_n q_n$ に対応する単位コスト関数は $c(p) = \min_n \{p_n/\alpha_n : n = 1, \dots, N\}$ である。効用を最大化する購買者がすべての N 個の製品を正の量だけ購入する場合、 $p_1/\alpha_1 = p_2/\alpha_2 = \dots = p_N/\alpha_N$ となり、購買者は各製品の1単位の購入から同じ量の効用を得なければならない。

(1971; 14-15), Muelbauer (1974; 988), Diewert (2003a) (2003b)によって支持された。もちろん、現在の消費者アプローチを正当化する関数形の仮定はかなり制限的であるが、それにもかかわらず、伝統的な消費者需要の設定にヘドニック回帰モデルを当てはめることは有用である。

経験的に、(6)式が正確に成立する可能性は低い。そこで Court (1939)に倣い、(6)式で定義される厳密なモデルはある程度の近似でのみ成立すると仮定し、(6)式の右辺に誤差項 e_{tn} を加える。すると未知のパラメータ $\pi \equiv [\pi_1, \dots, \pi_T]$ および $\alpha \equiv [\alpha_1, \dots, \alpha_N]$ は、次の(非線形)最小二乗問題の解として推定できる:

$$\min_{\alpha, \pi} \sum_{n=1}^N \sum_{t=1}^T [p_{tn} - \pi_t \alpha_n]^2 \dots (9)$$

しかし Diewert (2023)は、(9)式の最小化問題を解く推定価格水準 π_t^* は、満足のいく公理的性質を持たないことを示した。そこで Court(1939)に従い、厳密に成り立つ(6)式の両辺の対数を取り、得られた方程式に誤差項を加えることを考える。これによって、次の最小二乗問題が導かれる:¹⁰

$$\min_{\rho, \beta} \sum_{n=1}^N \sum_{t=1}^T [p_{tn} - \rho_t - \beta_n]^2 \dots (10)$$

ここで、新しいパラメータ ρ_t と β_n はそれぞれ、 π_t と α_n の対数として定義される:

$$\rho_t \equiv \ln \pi_t \dots (11)$$

$$\beta_n \equiv \ln \alpha_n \dots (12)$$

しかし、(10)式で定義される最小二乗法による最小化問題は、対数価格項 $[p_{tn} - \rho_t - \beta_n]^2$ をその経済的重要度によって重み付けしていないため、次のような重み付き最小二乗法による最小化問題を考える:¹¹

$$\min_{\rho, \beta} \sum_{n=1}^N \sum_{t=1}^T s_{tn} [\ln p_{tn} - \rho_t - \beta_n]^2 \dots (13)$$

ここで s_{tn} は、時点 t における製品 n の支出シェアである。 $\rho^* \equiv [\rho_1^*, \dots, \rho_T^*]$ と $\beta^* \equiv [\beta_1^*, \dots, \beta_N^*]$ が(13)式の解となるための一次条件は、次の(14)式と(15)式に単純化される:¹²

$$\rho_t^* = \sum_{n=1}^N s_{tn} [\ln p_{tn} - \beta_n^*] \dots (14)$$

¹⁰ このモデルは、Summer(1973)の国別製品ダミーモデルを時系列に適応させたものである。このモデルの時系列での初期の適用については、Aizcorbe, Corrado and Doms (2000)を参照されたい。

¹¹ Rao (1995) (2004) (2005; 574)は、支出シェアの重みを用いたこのモデルを初めて検討した。しかし、Balk (1980;70)はもっと早くから、やや異なる重みを用いてこのクラスのモデルを提案していた。2期間の場合については、Diewert (2004) (2005) および de Haan (2004a) を参照されたい。

¹² 支出や数量に関する情報が利用できない場合、重み付き最小二乗問題は、重み付けなし最小二乗問題(10)に置き換えられる。単純化された問題(10)式の一次条件は、(14)式と(15)式で与えられる。ここで、シェア s_{tn} は、すべての t と n について、 $1/N$ で置き換えられる。この重み付けのないケースでは、(17)式で定義される価格指数は、Jevons 指数に分解される。

$$\beta_n^* = \sum_{t=1}^T s_{tn} [\ln p_{tn} - \rho_n^*] / \left(\sum_{t=1}^T s_{tn} \right) \dots (15)$$

(14)式と(15)式の解は一意ではなく、 $\rho^* \equiv [\rho_1^*, \dots, \rho_T^*]$ と $\beta^* \equiv [\beta_1^*, \dots, \beta_N^*]$ が(14)式と(15)式の解となる場合、すべての λ について $[\rho_1^* + \lambda, \dots, \rho_T^* + \lambda]$ と $[\beta_1^* - \lambda, \dots, \beta_N^* - \lambda]$ も解となる。そこで(15)式において $\rho_1^* = 0$ と設定し、(14)式の最初の方程式を削除することで、方程式の一意解を見つけることができる。¹³解が見つかったら、推定によって得られた価格水準 π_t^* と品質調整係数 α_n^* を以下のように定義する：

$$\pi_t^* \equiv \exp[\rho_t^*], \alpha_n^* \equiv \exp[\beta_n^*] \dots (16)$$

なお、 $\rho_1^* = 0$ としたので、 $\pi_1^* = 1$ である。(16)式で定義される価格水準 π_t^* は、重み付き時間製品ダミー価格水準と呼ばれる。その結果、時点 t と τ の間の価格指数は、時点 t の価格水準の期間 τ の価格水準に対する比率として定義され、以下の式に等しくなる：

$$\pi_t^* / \pi_\tau^* = \prod_{n=1}^N \exp[s_{tn} \ln(p_{tn} / \alpha_n^*)] / \prod_{n=1}^N \exp[s_{\tau n} \ln(p_{\tau n} / \alpha_n^*)] \dots (17)$$

もし $s_{tn} = s_{\tau n}$ が $n = 1, \dots, N$ に対して成り立つならば、 π_t^* / π_τ^* は $p_{tn} / p_{\tau n}$ の重み付き幾何平均に等しくなる。したがって π_t^* / π_τ^* は、この場合 α_n^* に依存しなくなる。

π_t^* と α_n^* の推定値が計算されたら、 $t = 1, \dots, T$ について、時点ごとの価格と数量の水準である P^t と Q^t を構成するために2つの方法がある。1つ目は π_t^* の推計値を(18)式のように用いて構成する方法、2つ目は α_n^* の推計値を(19)式のように用いて時点 t の価格と数量の水準を構成する方法である：¹⁴

$$P_t^* \equiv \pi_t^* ; Q_t^* \equiv p^t \cdot q^t / \pi_t^* \dots (18)$$

$$Q_t^{**} \equiv \alpha^* \cdot q^t ; P_t^{**} \equiv p^t \cdot q^t / \alpha^* \cdot q^t \dots (19)$$

ここで誤差項 $e_{tn} \equiv \ln p_{tn} - \ln \pi_t^* - \ln \alpha_n^*$ を定義する。すべての誤差項が0であれば、 $t = 1, \dots, T$ について、 P_t^* は P_t^{**} と等しく、 Q_t^* は Q_t^{**} と等しくなる。¹⁵しかし、誤差項がすべてゼロに等しくない場合、統計機関は実務的な理由から、時点 t における価格水準と数量水準を形成するために、(18)式と(19)式のどちらを使用するかを決定しなければならない。

ここまでは価格と数量に欠損がないことを仮定していたが、(13)式の重み付き最小化問題を、価格と数量が欠損している場合に一般化するのは、そう難しくない。この場合、 N 個の商品と T 個の時点があるが、すべての期間ですべての商品が購入(または販売)されているわけではないと仮定する。各時点 t について、時

¹³ あるいは $(s_{tn})^{1/2} \ln p_{tn} = (s_{tn})^{1/2} \rho_t + (s_{tn})^{1/2} \beta_n + e_{tn}$ で定義される線形回帰モデルを設定することもできる。ここでは、厳密な多重共線性を避けるために $\rho_1 = 0$ とする。式(14)と式(15)の間を反復すると、これらの方程式の解も生成され、解は $\rho_1 = 0$ となるように正規化できる。

¹⁴ (19)式で定義される価格水準 P_t^{**} は、de Haan(2004b)で研究されたタイプの品質調整済み単価指数であることに留意されたい。

¹⁵ もしすべての誤差項が0であるならば(10)式で定義される重みなし(正確には等しく重み付けされた)最小化問題は、(13)式で定義される重み付き最小化問題によって生成されるのと同じ解を生成する。このことから、(10)式がデータに非常によく適合するならば、より複雑な重み付けされた問題である(13)式を扱う必要はない。

点 t に存在する製品 n の集合を $S(t) \equiv \{n: p_{tn} > 0\}$ と定義する。これらの集合は空ではない、つまり各期間に少なくとも1つの製品が購入されると仮定する。また、製品 n が存在する時点 t の集合を $S^*(n) \equiv \{t: p_{tn} > 0\}$ と定義する。ここでも、これらの集合は空ではないと仮定する。(13)式を欠品の場合に一般化すると、次の重み付き最小2乗最小化問題となる。¹⁶

$$\min_{\rho, \beta} \sum_{t=1}^T \sum_{n \in S(t)} s_{tn} [\ln p_{tn} - \rho_t - \beta_n]^2 = \min_{\rho, \beta} \sum_{t=1}^T \sum_{n \in S^*(t)} s_{tn} [\ln p_{tn} - \rho_t - \beta_n]^2 \cdots (20)$$

左辺では、時点 t に存在する製品 n の集合 $S(t)$ の定義を使用し、右辺では、製品 n が存在する期間 t の集合 $S^*(n)$ の定義を使用している。 ρ_1, \dots, ρ_T と β_1, \dots, β_N が(20)式を解くための一次の条件は、(14)式と(15)式に対応する以下のものである：¹⁷

$$\begin{aligned} \sum_{n \in S(t)} s_{tn} [\rho_t^* + \beta_n^*] &= \sum_{n \in S(t)} s_{tn} \ln p_{tn} \cdots (21) \\ \sum_{t \in S^*(n)} s_{tn} [\rho_t^* + \beta_n^*] &= \sum_{t \in S^*(n)} s_{tn} \ln p_{tn} \cdots (22) \end{aligned}$$

$\rho^* \equiv [\rho_1^*, \dots, \rho_T^*]$ と $\beta^* \equiv [\beta_1^*, \dots, \beta_N^*]$ が(21)式と(22)式の解となるのであれば、すべての λ について $[\rho_1^* + \lambda, \dots, \rho_T^* + \lambda]$ と $[\beta_1^* - \lambda, \dots, \beta_N^* - \lambda]$ も解となる。したがって、(22)式で $\rho_1^* = 0$ と設定し、(21)式の最初の式を削除することで一意解を見つけることができる。¹⁸

推定価格水準 π_t^* と品質調整係数 α_n^* を定義(11)式と(12)式により定義し、これらの定義を(21)式と(22)式に代入し整理すると次のようになる：

$$\begin{aligned} \pi_t^* &= \exp[\sum_{n \in S(t)} s_{tn} \ln(p_{tn} / \alpha_n^*)] \cdots (23) \\ \alpha_n^* &= \exp[\sum_{t \in S^*(n)} s_{tn} \ln(p_{tn} / \pi_t^*) / \sum_{t \in S^*(n)} s_{tn}] \cdots (24) \end{aligned}$$

一度、 π_t と α_n の推定値が計算されれば、先程と同様に時点ごとの価格水準と数量水準を構成するための方法が2つある。定義式(18)に対応する方法は以下の通りである：

$$\begin{aligned} P^{t*} &\equiv \pi_t^* = \exp[\sum_{n \in S(t)} s_{tn} \ln(p_{tn} / \alpha_n^*)] \cdots (25) \\ Q^{t*} &\equiv \sum_{n \in S(t)} p_{tn} q_{tn} / P^{t*} \cdots (26) \end{aligned}$$

P_t^* は、時点 t に存在する品質調整価格 p_{tn} / α_n^* の加重幾何平均であり、 p_{tn} / α_n^* の重みは、時点 t における製品 n の対応する支出(または売上)シェア s_{tn} である。一方で定義式(19)に対応するものは、以下の定義である：

$$\begin{aligned} Q^{t**} &\equiv \sum_{n \in S(t)} \alpha_n^* q_{tn} \cdots (27) \\ P^{t**} &\equiv \sum_{n \in S(t)} p_{tn} q_{tn} / Q^{t**} \\ &= \sum_{n \in S(t)} p_{tn} q_{tn} / \sum_{n \in S(t)} \alpha_n^* q_{tn} \\ &= \sum_{n \in S(t)} p_{tn} q_{tn} / \sum_{n \in S(t)} \alpha_n^* (q_{tn})^{-1} p_{tn} q_{tn} \end{aligned}$$

¹⁶ 価格情報しか入手できない場合は、(20)式の s_{tn} を $1/N(t)$ で置き換える。

¹⁷ (20)式の重み付けなし最小化問題は、 $n \in S(t)$ に対してすべての $s_{tn} = 1$ を設定する。結果として得られる一次条件は、正の s_{tn} を1に置き換えた(21)式と(22)式である。

¹⁸ 結果として得られる $T-1 + N$ 方程式系は、一意解を得るためにフルランクである必要がある。解は、線形回帰を実行することによっても得ることができる。

$$\begin{aligned}
&= [\sum_{n \in S(t)} s_{tn} (p_{tn}/\alpha_n^*)^{-1}]^{-1} \\
&\leq \exp[\sum_{n \in S(t)} s_{tn} \ln(p_{tn}/\alpha_n^*)] \\
&= P^{t*} \dots (28)
\end{aligned}$$

ここで、不等式は Schlömilch の不等式¹⁹から導かれる。価格水準に関する不等式 $P^{t**} \leq P^{t*}$ は、数量水準に関する不等式 $Q^{t**} \geq Q^{t*}$ を含意する。また(28)式は de Haan (2004b) (2010)と de Haan and Krsinich (2014) (2018; 763)によるものである。de Haan と Krsinich が用いたモデルは、より一般的なヘドニック回帰モデルであり、本節で用いた TPD モデルを特別なケースとして含んでいる。

重み付き最小化問題に現れる誤差項 $e_{tn}^* \equiv \ln p_{tn} - \rho_t^* - \beta_n^*$ が 0 に等しければ、方程式 $p_{tn} = \pi_t \alpha_n$ が成立し、ヘドニック回帰は現実への良い近似となる。さらに、これらの条件下において、 P^{t*} はすべての t について P^{t**} と等しくなる。モデルの適合が良好でない場合は、次節以降で検討されるような他のモデルを検討する必要がある。

(20)式で定義される重み付き最小 2 乗回帰問題の解は、欠損している製品の帰属価格を生成するためにも使用できる。したがって、時点 t の製品 n が欠損している場合、価格について $p_{tn} \equiv \pi_t^* \alpha_n^*$ と定義すれば、対応する数量は $q_{tn} \equiv 0$ と定義される。統計機関の中には、欠損価格に対する帰属価格を生成するためにヘドニック回帰モデルを使用するところもある。

重み付き時間製品ダミー価格水準 π_t^* における欠点の 1 つは T 期間のうち 1 時点しか利用できない製品は、集計価格水準 π_t^* に影響を与えないということである。²⁰ 商品の特性に関する情報を利用する次節のヘドニック回帰モデルは、本節で説明した重み付き時間製品ダミーヘドニック回帰モデルのこの欠点を回避することができる。

3.2. ヘドニック・モデルによる価格指数の推計

この節では、 T 期間にわたって入手可能な N 個の製品があり、前節と同様に、 N 個の製品の数量集計関数は線形関数であると仮定する。具体的には、 $f(q) \equiv \alpha \cdot q = \sum_{n=1}^N \alpha_n q_n$ であるとする。ここで、 q_n は検討期間に購入または販売された製品 n の数量であり、 α_n は製品 n の品質調整係数である。前節と異なるのは、品質調整係数が製品の K 個の特性のベクトルの関数であるという仮定である。したがって、製品 n は、 $n = 1, \dots, N$ に対して、特性 $z_n \equiv [z_{n1}, z_{n2}, \dots, z_{nK}]$ のベクトルを持つと仮定する。本節での新しい仮定は、品質調整係数 α_n が各製品の特性ベクトル z^n の関数であり、各製品の品質調整に同じ関数 $g(z)$ を使用できるとすることである。すなわち、以下が成り立つと仮定する：

$$\alpha_n \equiv g(z^n) = g(z_{n1}, z_{n2}, \dots, z_{nK}) \dots (29)$$

このように、製品 n はそれぞれ独自の特性の組合せ z^n を持つが、同じ関数 g を用いて、製品の購入者の相対効用を決定することができる。いま、 $t = 1, \dots, T$ において、時点 t の数量ベクトルを $q^t =$

$[q_{t1}, \dots, q_{tN}]$ と定義し、製品 n が第 t 期に欠損している場合、数量は $q_{tn} = 0$ であるとする。上記の仮定を用いると、時点 t における集約数量水準 Q^t は次式で定義される：

$$Q^t \equiv f(q^t) \equiv \sum_{n=1}^N \alpha_n q_{tn} = \sum_{n=1}^N g(z^n) q_{tn} \dots (30)$$

¹⁹ Hardy, Littlewood and Pólya (1934; 26)を 参照。

²⁰ Diewert (2004) はこの性質を証明した。

(30)式で定義される線形効用関数による効用最大化行動の仮定を用いると、(6)式は以下ようになる：

$$p_{tn} = \pi_t g(z^n) \cdots (31)$$

(6)式で定義される厳密なモデルはある程度の近似でのみ成立すると仮定するため、(31)式には誤差項を追加する必要がある。また、品質調整関数またはヘドニック評価関数 $g(z) = g(z_1, \dots, z_K)$ の関数形を選択する必要がある。一般的な評価関数のパラメータを推定することはできないので、 $g(z)$ は $g_k(z_k)$ という形の1変数の K 個の関数の積であると仮定する。すなわち、 $g(z)$ は以下のように定義されると仮定する：

$$g(z_1, \dots, z_K) \equiv g_1(z_1)g_2(z_2)\dots g_K(z_K) \cdots (32)$$

今回の例では、各特性は有限個の離散値しかとらないので、各 $g_k(z_k)$ はステップ関数、または各 z_k の範囲において有限個の離散値にて値がジャンプする「plateaux」関数であると仮定する。この仮定は、最終的にすべての独立変数がダミー変数である回帰モデルにつながる。

各特性 k について、 z_k の取りうる範囲を正確にカバーする $N(k)$ 個の離散区間に分割する。これを $k = 1, \dots, K$ および $j = 1, \dots, N(k)$ として、変数 z_k の j 番目の区間を $I(k, j)$ とする。そして時点 t における製品 n 、および各特性 k について、ダミー変数 $D_{tn,k,j}$ を次のように定義する：

$$D_{tn,k,j} \equiv \begin{cases} 1 & \text{時点 } t \text{ の製品 } n \text{ が、} j \text{ 番目の区間に属する特性 } k \text{ の量 } z_{nk} \text{ を持つ} \\ 0 & \text{時点 } t \text{ の製品 } n \text{ が、} j \text{ 番目の区間に属する特性 } k \text{ の量 } z_{nk} \text{ を持たない} \end{cases} \cdots (33)$$

定義(33)を用いて、製品 n が時点 t に購入された場合の $g(z^n) = g(z_{n1}, z_{n2}, \dots, z_{nK})$ を定義する：

$$g(z_{n1}, z_{n2}, \dots, z_{nK}) \equiv (\sum_{j=1}^{N(1)} a_{1j} D_{tn,1,j}) (\sum_{j=1}^{N(2)} a_{2j} D_{tn,2,j}) \cdots (\sum_{j=1}^{N(K)} a_{Kj} D_{tn,K,j}) \cdots (34)$$

(34)式を用いると、 π_t と $a_{1j}, a_{2j}, \dots, a_{Kj}$ が未知のパラメータであるような次式を得ることができる：

$$p_{tn} = \pi_t \left(\sum_{j=1}^{N(1)} a_{1j} D_{tn,1,j} \right) \left(\sum_{j=1}^{N(2)} a_{2j} D_{tn,2,j} \right) \cdots \left(\sum_{j=1}^{N(K)} a_{Kj} D_{tn,K,j} \right) \cdots (35)$$

さらに(35)式の両辺の対数をとると次式を得る：²¹

$$\ln p_{tn} = \ln \pi_t + \sum_{j=1}^{N(1)} (\ln a_{1j}) D_{tn,1,j} + \sum_{j=1}^{N(2)} (\ln a_{2j}) D_{tn,2,j} + \cdots + \sum_{j=1}^{N(K)} (\ln a_{Kj}) D_{tn,K,j} \cdots (36)$$

ここで、以下のパラメーターを定義する：

$$\begin{cases} \rho_t \equiv \ln \pi_t \\ b_{1j} \equiv \ln a_{1j}; j = 1, \dots, N(1) \\ b_{2j} \equiv \ln a_{2j}; j = 1, \dots, N(2) \cdots (37) \\ \vdots \\ b_{Kj} \equiv \ln a_{Kj}; j = 1, \dots, N(K) \end{cases}$$

²¹ (35)式で定義されるモデルによって生成されるヘドニック価格指数は、特性の測定単位の変化に対して不変ではないことに注意されたい。

(36)式に定義(37)を代入し、誤差項 e_{tn} を加えると、以下の線形回帰モデルが得られる:

$$\ln p_{tn} = \rho_t + \sum_{j=1}^{N(1)} b_{1j} D_{tn,1,j} + \sum_{j=1}^{N(2)} b_{2j} D_{tn,2,j} + \dots + \sum_{j=1}^{N(K)} b_{Kj} D_{tn,K,j} + e_{tn} \dots (38)$$

式(38)には、合計 $T + N(1) + N(2) + \dots + N(K)$ の未知パラメータがある。(38)式で定義される線形回帰モデルに対応する最小化問題は次のようになる。

$$\min_{\rho, b(1), b(2), \dots, b(K)} \sum_{t=1}^T \sum_{n \in S(t)} \left\{ \ln p_{tn} - \rho_t - \sum_{j=1}^{N(1)} b_{1j} D_{tn,1,j} - \sum_{j=1}^{N(2)} b_{2j} D_{tn,2,j} - \dots - \sum_{j=1}^{N(K)} b_{Kj} D_{tn,K,j} \right\}^2 \dots (39)$$

ここで、 ρ はベクトル $[\rho_1, \rho_2, \dots, \rho_T]$ 、 $b(k)$ は $k = 1, 2, \dots, K$ のベクトル $[b_{k1}, b_{k2}, \dots, b_{kN(k)}]$ である。最小二乗問題の解は存在するが、この場合も解は一意ではない。これまでと同様に、(39)式の有用な一意解は、 $\rho_1 = 0$ (これは $\pi_1 = 1$ に対応する)とし、 $k = 2, \dots, K$ に対して $b_{k1} = 0$ とすることで得られる。

しかし今回は、 $g(z)$ を少し変更し次のようにする:

$$p_{tn} = \pi_t a_0 \left(\sum_{j=1}^{N(1)} a_{1j} D_{tn,1,j} \right) \left(\sum_{j=1}^{N(2)} a_{2j} D_{tn,2,j} \right) \dots \left(\sum_{j=1}^{N(K)} a_{Kj} D_{tn,K,j} \right) \dots (40)$$

(35)式と異なる点は $\pi_1 = 1$ と $k = 1, \dots, K$ について $a_{k1} = 1$ とした点である。この変更は、 π_t には影響しないものの、回帰モデル中の定数項 a_0 の推定を可能にする。 $b_0 \equiv \ln a_0$ を定義し、(40)の両辺の対数をとると、次の線形回帰モデルが得られる:

$$\ln p_{tn} = \rho_t + b_0 + \sum_{j=2}^{N(1)} b_{1j} D_{tn,1,j} + \sum_{j=2}^{N(2)} b_{2j} D_{tn,2,j} + \dots + \sum_{j=2}^{N(K)} b_{Kj} D_{tn,K,j} + e_{tn} \dots (41)$$

$k = 1, \dots, K$ に対して $b_{k1} \equiv \ln a_{k1} = 0$ という制約が式(41)に課せられている。(39)で定義される最小二乗回帰問題の一意解を得るために必要な追加の制約は、 $\rho_1 = 0$ である。これによって、(41)式で定義される線形回帰を実行することで、未知パラメータ $[\rho_2^*, \dots, \rho_T^*]$ 、 b_0^* および $[b_{k2}^*, \dots, b_{kN(k)}^*]$ の推定値を得ることができる。これらの推定値を用いて、時点 t に購入された全製品 n の品質調整係数 α_n の対数を定義する:²²

$$\beta_{tn}^* \equiv b_0^* + \sum_{j=1}^{N(1)} b_{1j}^* D_{tn,1,j} + \sum_{j=1}^{N(2)} b_{2j}^* D_{tn,2,j} + \dots + \sum_{j=1}^{N(K)} b_{Kj}^* D_{tn,K,j} \dots (42)$$

対応する製品 n の品質調整係数 α_{tn}^* は、 β_{tn}^* を指数化することで得られる:

$$\alpha_{tn}^* \equiv \exp[\beta_{tn}^*] \dots (43)$$

上記の α_{tn}^* を用いれば、時点 t に購入者が得た数量水準または効用を次のように直接推定することができる:

$$Q^{t**} \equiv \sum_{n \in S(t)} \alpha_{tn}^* q_{tn} \dots (44)$$

²² 製品 n が複数の期間で入手可能な場合、品質調整係数は期間をまたいでも同じままである。

間接的に得られる時点 t の価格水準 P^{t**} は、時点 t の支出を時点 t の総量でデフレートすることによって定義される：

$$P^{t**} \equiv \sum_{n \in S(t)} p_{tn} q_{tn} / Q^{t**} = \sum_{n \in S(t)} p_{tn} q_{tn} / \sum_{n \in S(t)} \alpha_{tn}^* q_{tn} \dots (45)$$

時点 t の価格水準 π_t を直接推定にするために必要な式を得るための、 ρ_t に関して(39)式を最小化するための一次条件は次のとおりである：

$$\begin{aligned} 0 &= \sum_{n \in S(t)} \{ \ln p_{tn} - \rho_t^* - \sum_{j=1}^{N(1)} b_{1j}^* D_{tn,1,j} - \sum_{j=1}^{N(2)} b_{2j}^* D_{tn,2,j} - \dots - \sum_{j=1}^{N(K)} b_{Kj}^* D_{tn,K,j} \} \\ &= \sum_{n \in S(t)} \{ \ln p_{tn} - \rho_t^* - \beta_n^* \} \dots (46) \end{aligned}$$

ここで、2つ目の等式を導くために定義式(42)を用いた。 $N(t)$ を時点 t に購入された製品の数として、定義式(37)と(42)を用いると、(46)式から、時点 t の価格水準の直接推定値 π_t^* は次のようになる：

$$\pi_t^* = \prod_{n \in S(t)} (p_{tn} / \alpha_{tn}^*)^{1/N(t)} \equiv P^{t*} \dots (47)$$

したがって、時点 t の価格水準 P^{t*} の直接推定値は、時点 t に購入された製品の品質調整済み価格の幾何平均 (p_{tn} / α_{tn}^*) に等しい。この価格水準は価格情報だけで計算できるのに対して、間接的な価格水準 P^{t**} は、時点 t における製品の購入に関する価格と数量の情報を必要とする。

(41)式で定義される最小二乗問題の問題点は、製品の経済的重要性を考慮していないことである。そこで、対応する重み付き最小二乗問題を以下のように定義する：

$$\min_{\rho, b(1), b(2), \dots, b(K)} \sum_{t=1}^T \sum_{n \in S(t)} s_{tn} \left\{ \ln p_{tn} - \rho_t - b_0 - \sum_{j=2}^{N(1)} b_{1j} D_{tn,1,j} - \sum_{j=2}^{N(2)} b_{2j} D_{tn,2,j} - \dots - \sum_{j=2}^{N(K)} b_{Kj} D_{tn,K,j} \right\}^2$$

ここで、 $s_{tn} = p_{tn} q_{tn} / \sum_{j \in S(t)} p_{tj} q_{tj}$ であり、 $S(t)$ は(39)式の最小化問題で用いられたのと同じ定義を用いる。両辺の対数を取ることで、(41)式で定義された線形回帰モデルの新しい重み付き対応モデルを以下に示す：

$$(s_{tn})^{\frac{1}{2}} \ln p_{tn} = (s_{tn})^{\frac{1}{2}} \left(\rho_t + b_0 + \sum_{j=1}^{N(1)} b_{1j} D_{tn,1,j} + \sum_{j=1}^{N(2)} b_{2j} D_{tn,2,j} + \dots + \sum_{j=1}^{N(K)} b_{Kj} D_{tn,K,j} \right)$$

定義(42)～(45)を用いて、新しく β_{tn}^* , α_{tn}^* , Q^{t**} , P^{t**} を定義する。 P^{t**} を多少異なる方法で書き直すと以下のようになる：

$$\begin{aligned} P^{t**} &= \sum_{n \in S(t)} p_{tn} q_{tn} / \sum_{n \in S(t)} \alpha_{tn}^* q_{tn} \\ &= \sum_{n \in S(t)} p_{tn} q_{tn} / \sum_{n \in S(t)} (\alpha_{tn}^* / p_{tn}) p_{tn} q_{tn} \\ &= \left[\sum_{n \in S(t)} s_{tn} (p_{tn} / \alpha_{tn}^*)^{-1} \right]^{-1} \dots (50) \end{aligned}$$

時点 t の価格水準 π_t の直接推定に有用な式を得るため、 ρ_t に関して新しい重み付き最小化問題である(48)式の解を得るための一次条件は次のとおりである:

$$0 = \sum_{n \in S(t)} s_{tn} \left\{ \ln p_{tn} - \rho_t^* - \sum_{j=1}^{N(1)} b_{1j}^* D_{tn,1,j} - \sum_{j=1}^{N(2)} b_{2j}^* D_{tn,2,j} - \dots - \sum_{j=1}^{N(K)} b_{Kj}^* D_{tn,K,j} \right\} \\ = \sum_{n \in S(t)} s_{tn} \{ \ln p_{tn} - \rho_t^* - \beta_n^* \} \dots (51)$$

ここで、2番目の等式を導くために定義式(42)を用いた。定義式(37)と(43)を用いると、(51)式は時点 t の価格水準 π_t に対する直接推定が以下のようになることを意味する:²³

$$\pi_t^* = \prod_{n \in S(t)} (p_{tn} / \alpha_{tn}^*) s(t, n) \equiv P^{t*} \dots (52)$$

ここで、 $s(t, n) = s_{tn}$ である。続いて、時点 t における間接的な数量水準は、支出を P^{t*} で割ったものとして定義される:

$$Q^{t*} \equiv \sum_{n \in S(t)} p_{tn} q_{tn} / P^{t*} \dots (53)$$

(52)式で定義される時点 t における直接的な価格水準 P^{t*} は、品質調整価格 p_{tn} / α_{tn}^* の加重幾何平均である。その一方で、(50)式で定義される時点 t における間接的な価格水準 P^{t**} は、品質調整価格の加重調和平均である。したがって、de Haan の不等式($P^{t**} \leq P^{t*}$, $Q^{t**} \geq Q^{t*}$)が成り立つ。

次節では、代替のヘドニック回帰モデルを推定し、上記の分析を利用した実証例を紹介する。

4. ヘドニック・モデルの推計

4.1. ヘドニック回帰

ここでの実際の計算方法と計算式については、付論 B3 をご覧いただきたい。

全国のノートパソコンについて、日次で販売台数、価格、特性情報を収集している民間企業からデータを入手した。入手したデータは2015年1月～2022年11月までの期間であったが、本研究では、そのうちの2016年1月～2022年6月までの78ヶ月分のデータを用いる。また、第4節5項で市場に登場してからの期間に関する効果を分析するが、もともとのデータには市場に流通してからの期間に関する情報は含まれていない。そこで、データに初めて登場した月からの経過月数を市場に流通してからの期間としてNEW変数を作成し、第4節5項で用いる。²⁴このために、2015年1月から登場したデータは、新製品なのか、左側打ち切りのデータなのか区別がつけられないため、2015年2月以降にデータに登場した製品のみを用いた。ノートパソコン販売に関する日次データの件数は合計で706887件ある。

このデータには多くのノートパソコンに関する特性が含まれている。そのうち量的変数としてCLOCKはノートパソコンのクロックスピード、MEMはノートパソコンのメモリ容量(GB)、SIZEはノートパソコンの画面サイズ(インチ)、PIXはノートパソコンの画面に埋め込まれた画素数(万画素)、WEIGHTはノートパソコンの重量(キログラム)、Storageはノートパソコンに搭載されたHDDやSSDの合計容量(GB)などがある。

²³ 正規化すると、 $\pi_1^* = 1$ となる。

²⁴ JANコードから発売月を取得することも可能かもしれないが、商品数も多く発売月の取得が困難なため、今回はこのような方法で代替した

経験的に、購入者はノートパソコンのクロックスピード、メモリ容量、画面サイズ、画素数や HDD または SSD の容量を高く評価すると予想され、検討中の特性の値が高いほどダミー変数の推定係数が大きくなる。また購入者は、重いノートパソコンよりも軽いノートパソコンを重視すると予想される。

続いて、質的変数として CPU がある。これは、ノートパソコンが使用している中央処理装置の種類である。本データでは、代表的な 10 種類の CPU (Core i5 や AMD RYZEN 5 など) 以外はその他としてまとめた。²⁵ BRAND はノートパソコンのメーカー名であり、データファイルには、多くの BRAND が存在するものこちらにも CPU と同様に、Apple や NEC などの代表的な 10 種類の BRAND 以外はその他としてまとめた。それ以外にも GPU については搭載されていないか、CPU 内蔵か、搭載されているか、HDMI 接続端子が搭載されているか、HDD ではなく SSD が搭載されているか、などがある。

回帰分析および物価・数量指数の計算で系列の平均値をより単一に近づけるために、測定単位の一部を変換した。 p_{tn} は $p_{tn} / 100,000$ で置き換え、10 万円単位で価格を測定している。同様に MEM は MEM / 1000 に、SIZE は SIZE / 10 に、PIX は PIX / 10 に置き換えた。

日本においては、総務省でノートパソコンの価格指数が作成・公表されているが、その際には月次のデータに集計されたものを使用している。そこで本研究でも、商品ごと (JAN コードごと) に月次販売数と月次総売上から月次平均 (単価) 価格を算出するように月次で集約したものをを用いる。

これを R で集約するためには、dplyr というパッケージを利用するのが便利である。dplyr というのは、表型のデータ操作に特化したパッケージで、表型データに対して加工や集計を行うのに便利な関数が多数用意されている。例えばいま、df という変数に読み込んだデータが格納されているとして、その中に TDM という何年何月に販売されたかという年月の情報、JAN コード、(日次の)販売台数、(日次の)販売金額に関する項目があるとすると、このとき、商品ごとに月次平均 (単価) 価格を求めるためには、まず月ごとの総販売台数と総販売額を計算する必要がある。これは

```
df2 <- df %>% group_by(TDM,JAN コード) %>% summarise(Q=sum(販売台数),E=sum(販売金額))
```

のようにして計算することができる。上記のコードは、df の TDM と JAN コードの組み合わせが同じレコードについて、販売台数について足し合わせたものを Q という変数に、販売金額について足し合わせたものを E という変数に代入せよ、そしてその結果を df2 に代入せよ、という意味になっている。そのため、ここで作成された df2 は TDM、JAN コード、Q、E という 4 列のデータになる。そのうえで、

```
df2$P <- df2$E / df2$Q
```

とすれば月次平均 (単価) 価格を P として求めることができる。

月次で集約を行うと、データ件数は合計で 54647 件となった。この 54647 件に対する記述統計量を表 1 に示す。また、月ごとのデータ件数を表 2 にまとめた。

表 1: 記述統計量

	Min	Mean	Std.Dev	Max
NEW	0.0000	12.0352	9.6198	84.0000

²⁵ 本来は Core i5 などでも世代があるため細かく分けるべきであるが、今回入手したデータではそこまでの情報は得られなかったため、Core i5 や Core i7 などの分類までとした。また、種類でダミー変数化するのではなく、ベンチマークスコアを用いて量的変数として扱う事も考えられるが、ベンチマークスコアの取得も困難であるため、今回は採用しなかった。

Clock	0.9000	2.0141	0.5485	4.7000
MEM	2.0480	7.9590	4.6592	32.7680
SIZE	0.8000	1.4376	0.1599	1.7300
PIX	0.9830	2.1542	1.2036	9.2160
Weight	0.6500	1.7325	0.5965	2.9800
Q	1.0000	173.1561	604.4614	20301.0000
P	0.2000	1.1168	0.5648	4.0000

データ数：54647

表 2.月次データ件数

TDM	データ数	TDM	データ数
201601	399	201904	810
201602	426	201905	823
201603	435	201906	838
201604	460	201907	881
201605	464	201908	831
201606	476	201909	852
201607	488	201910	788
201608	497	201911	815
201609	491	201912	852
201610	526	202001	901
201611	520	202002	871
201612	561	202003	855
201701	578	202004	826
201702	599	202005	826
201703	641	202006	808
201704	622	202007	797
201705	601	202008	784
201706	634	202009	771
201707	649	202010	779
201708	651	202011	790
201709	656	202012	775
201710	685	202101	777
201711	700	202102	800

201712	716	202103	800
201801	736	202104	754
201802	721	202105	720
201803	754	202106	725
201804	719	202107	745
201805	688	202108	699
201806	745	202109	651
201807	790	202110	661
201808	780	202111	654
201809	766	202112	712
201810	788	202201	698
201811	803	202202	652
201812	807	202203	652
201901	816	202204	586
201902	810	202205	535
201903	835	202206	520

ここで、月次集計後のデータに対して時点 t における製品 n の価格と数量を p_{tn} と q_{tn} とすると、 p_{tn} は時点 t における製品 n の平均(単価)価格、 q_{tn} は販売台数になっている。サンプルの78ヵ月間で、2972の異なる商品が販売されたので、 $n = 1, \dots, 2972$ であり²⁶、 $t = 1, 2, \dots, 78$ とする。そして、製品 n が t 月に売れなかったなら、 p_{tn} と q_{tn} を0に設定する。もし各製品が各月に売れたなら、 $2972 \times 78 = 231816$ の価格 p_{tn} と数量 q_{tn} を持つことになるが、0でないデータの割合は $54647/231816 = 0.236$ なので、平均して月あたり24.0%の製品しか売れていないことがわかる。このように、日本のノートパソコン市場では、個々の製品が素早く市場に参入し、そして撤退するという、非常に大きな製品の入れ替わりがあるといえる。

時点 t に販売されたノートパソコンの平均価格 P_A^t を、サンプルの78ヶ月のデータそれぞれについて計算する:

$$P_A^t \equiv \sum_{n \in S(t)} p_{tn} / N(t) \dots (54)$$

$N(t)$ は時点 t に販売されたノートパソコンの数、 $S(t)$ は時点 t に販売された製品の集合である。

ノートパソコンの時点 t における平均価格 P_A^t は、時点 t の各ノートパソコン価格に等しく重み付けをしており、ノートパソコンの種類ごとの経済的重要性を考慮していない。時点 t におけるノートパソコンの平均価格について、より代表的な尺度は、時点 t の単価 P_{UV}^t であり、以下のように定義される:

$$P_{UV}^t \equiv \sum_{n \in S(t)} p_{tn} q_{tn} / \sum_{n \in S(t)} q_{tn} = \sum_{n \in S(t)} e_{tn} / \sum_{n \in S(t)} q_{tn} \dots (55)$$

ただし、 $e_{tn} \equiv p_{tn} q_{tn}$ は時点 t における製品 n の総売上である。

²⁶ 総販売台数が100台以上、市場での流通期間(生存時間)が7ヶ月以上の製品のみを利用した

(54)式と(55)式で定義された価格を、対応する系列値の最初の時点の値で割ることによって価格指数に変換する事ができる。つまり、時点 t における平均価格指数 P_A^t と単価価格指数 P_{UV}^t を以下のように定義する:

$$P_A^t \equiv P_A^t / P_A^1; P_{UV}^t \equiv P_{UV}^t / P_{UV}^1 \dots (56)$$

価格指数 P_A^t, P_{UV}^t については、ノートパソコンの平均的な品質の変化に対する調整は行われていない。そのためここでは、前節の(41)式で定義されたタイプのヘドニック回帰モデルを検討する。最初に、全 78 カ月のデータを使用して製品の様々な特性を用いた回帰を検討する。一般的に消費者物価指数は改定できないが、この方法で結果として得られる指数は、国の統計局が作成できるようリアルタイムの「実用的な」指数ではない。しかしこのタイプの回帰は、ヘドニック回帰に特定の特性を含めることがいかに重要であることを示すという点で有用である。また、サンプル期間中におけるチェーン・ドリフトの影響を受けない。

製品の「新しさ」を除く特性を使用した、ヘドニック回帰のモデルとして、ベクトル表記で次のようなものを用いる。

$$\begin{aligned} \ln P = & \sum_{t=2}^{78} \rho_t D_t + b_0 ONE + \sum_{j=2}^7 b_{Cj} D_{Cj} + \sum_{j=2}^5 b_{Mj} D_{Mj} + \sum_{j=2}^6 b_{Sj} D_{Sj} + \sum_{j=2}^6 b_{Pj} D_{Pj} \\ & + b_{H2} D_{H2} + \sum_{j=2}^{10} b_{Bj} D_{Bj} + \sum_{j=2}^{10} b_{Uj} D_{Uj} + \sum_{j=2}^5 b_{Wj} D_{Wj} + b_{SSD2} D_{SSD2} \\ & + \sum_{j=2}^3 b_{GPUj} D_{GPUj} + \sum_{j=2}^6 b_{STj} D_{STj} + e \dots (57) \end{aligned}$$

$\ln P$ は 54647 次元の対数価格のベクトル、 D_t は時間ダミー変数のベクトル、 ONE は 54647 次元の 1 のベクトルであり、 $D_{Cj}, D_{Mj}, D_{Sj}, D_{Pj}, D_{H2}, D_{Bj}, D_{Uj}, D_{Wj}, D_{SSD2}, D_{GPU}, D_{ST}$ はクロックスピード、メモリサイズ、画面サイズ、ピクセル数、HDMI の有無、ブランド、コンピュータチップの種類、ノートパソコンの重さ、SSD の有無、GPU の有無、ストレージ容量に関するダミー変数のベクトルである。

(57)式では、大量のダミー変数を用いており、たとえば時点に関するダミー変数だけで 78 個もある。そのため当然、R でこの大量のダミー変数を作成することになる。R でダミー変数を作成するためには、例えば `df` という変数に月次集計されたデータが格納されており、年月の情報を表す変数として `TDM` という項目があるとすれば、次のように作成する事ができる。

```
D1 <- ifelse(df$TDM==201601,1,0)
```

このようにすることで、`TDM` の値が 201601(2016 年 1 月)だったら 1、そうでなかったら 0 というダミー変数が作成される。ダミー変数のイメージについては付論 A を参照していただきたい。しかし、このように手打ちで 78 個のダミー変数を作成するのは非常に手間である。そこで今回は次のように、繰り返し文を用いることで大量のダミー変数化を行った。実際のコードは次の通りである。

```
TDN <- length(unique(sort(df$TDM)))
TDlist <- unique(sort(df$TDM))
df_D <- data.frame(D01=ifelse(df$TDM==TDlist[1],1,0))
for(ii in 2:length(TDlist)){
  df_D <- data.frame(df_D,ifelse(df$TDM==TDlist[ii],1,0))
  names(df_D)[dim(df_D)[2]] <- paste("D",formatC(ii,width=2,flag=0),sep="")
}
```

処理の手順は次のとおりである

- (1). まず TDM の値を小さい順に並び替え、ユニーク化したものの長さを取得することで、月次のサンプル期間長を取得。(1 行目)
- (2). TDM の中身の一覧を取得。(2 行目)
- (3). まずは TDlist の最初に格納されている 201601 について D01 という列名でダミー変数化したのち、df_D というデータフレームに格納。(3 行目)
- (4). TDM が 201602 以降も順次ダミー変数化、ダミー変数化したものを df_D に追加していく。(5 行目) (data.frame(df_D, ...) とすることで、... に記述したデータが df_D の後ろ側の列に追加される())
- (5). 新たに追加されたダミー変数に列名を付ける。(6 行目)(formatC(ii,width,flag) で ii に代入されている数値の桁数が width 未満であれば flag の値で埋める関数である。今の場合、ii が 2 であれば、02 になる。また、paste 関数は文字列を結合する関数なので D02 という列名を付けることができる)

このように、時点ダミーに関するデータフレームを作成することができる。²⁷他の変数についても同様に作成できる。

こうして作成したダミー変数を説明変数として、回帰分析を実行する。R では lm 関数によって実行でき、使い方は次の通りである。

```
lm("推計を行う式",data)
```

例えば、y を目的変数として、x と z を説明変数として推計を行いたい場合は

```
lm("y~x+z",data)
```

のように、「~」の左側に目的変数を、右側に説明変数を書く。説明変数が複数ある場合には、「~」の右側に足し算の形で記述する。また、data として与えたデータフレームに x、y、z 列しかない場合、上記のコードは

```
lm("y~.",data)
```

のように書くこともできる。「y~.」のように書くと、目的変数として用いる y 以外の変数を目的変数として利用することになる。

さらに、lm 関数では特に指定しない場合定数項を含んだ推計を行うが

```
lm("y~.+0",data)
```

のように「+0」を付け加えると定数項を使用しない推計になる。今回の場合、(57)式を見るとわかるように b_0ONE という定数項に対応する項をデータフレームに含んでいるため、「+0」を付け加えて推計を実行している。

その結果、観測価格ベクトルと予測価格ベクトルの間の R^2 は 0.7675 であった。この回帰における時間ダミー変数の推定係数は $\rho_2^*, \rho_3^*, \dots, \rho_{78}^*$ であり、 $\rho_1^* \equiv 0$ と価格水準 $\pi_t^* \equiv \exp[\rho_t^*]$ を定義する。この製品特性を用いた価格水準から作成した価格指数のことを、時間ダミー特性価格指数(Time Dummy Characteristic; TDC)と呼ぶことにする。時点 t における時間ダミー特性価格指数 $P_{TDC}^t \equiv \pi_t^*$ を定義し、図 1 にプロットした。

²⁷ ダミー変数の作成についても fastDummies などの有用なパッケージがある

時点 t に購入された製品 n に対応する支出シェアは、 $n \in S(t)$ に対して $s_{tn} = p_{tn}q_{tn} / \sum_{j \in S(t)} p_{tj}q_{tj}$ と計算されるので、(57)式で定義されたヘドニック回帰モデルの重み付きに対応するものを得るには、各 s_{tn} の正の平方根からなる 54647 次元の新しいベクトル SS を求め、(57)式の両辺に SS を掛ければ良い。

この新しい重み付き回帰モデルの R^2 は 0.8810 で、対応する重みなしモデルの R^2 0.7675 よりもかなり高いことがわかる。

この回帰における時間ダミー変数の推定係数は、 $\rho_2^*, \rho_3^*, \dots, \rho_{78}^*$ である。先程と同様に、 $\rho_1^* \equiv 0$ とし、価格

水準 $\pi_t^* \equiv \exp[\rho_t^*]$ を定義する。これから求まる価格指数を、重み付き時間ダミー特性(Weighted Time Dummy Characteristic; WTDC)価格指数と呼ぶことにする。時点 t におけるこの価格指数 $P_{WTDC}^t \equiv \pi_t^*$ を定義し、こちらも図 1 にプロットした。

また、(57)式では 54647 件の月次集計したデータに対するヘドニック回帰モデルを想定しているが、参考として 706887 件の日次集計されたデータに対するヘドニック回帰モデルで月次指数を作成した $P_{TDC_D}^t$ と $P_{WTDC_D}^t$ も図1にプロットした。これら各指数の値を表 3 に掲載する。また、WTDC の結果得られた回帰係数の値を表 4 に示す。

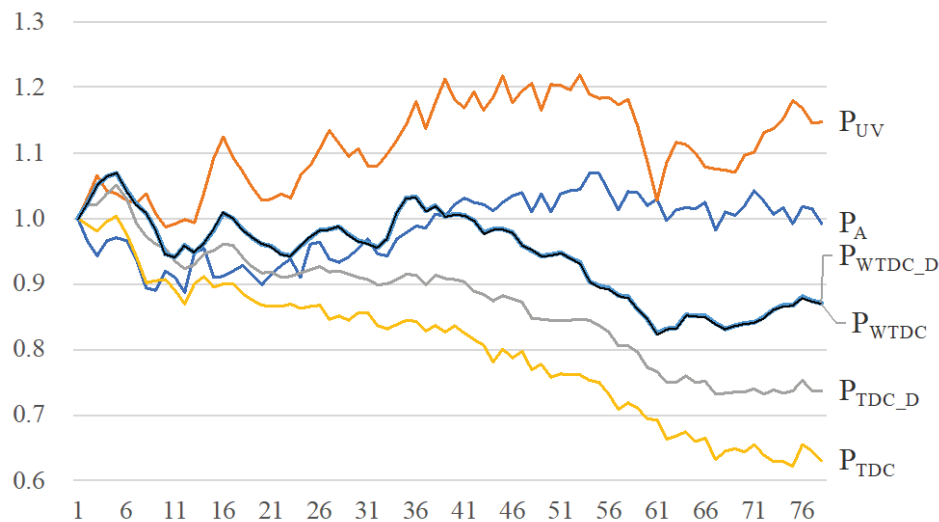


図1. 平均価格指数と製品特性を用いたヘドニック回帰モデルによる指数

表 3.製品特性を用いて作成した指数

TDM	P_A^t	P_{UV}^t	$P_{TDC_D}^t$	P_{TDC}^t	$P_{WTDC_D}^t$	P_{WTDC}^t
201601	1.000	1.000	1.000	1.000	1.000	1.000
201602	0.966	1.031	1.022	0.991	1.024	1.024
201603	0.943	1.066	1.022	0.980	1.052	1.052
201604	0.967	1.043	1.038	0.995	1.065	1.065
201605	0.971	1.038	1.051	1.003	1.070	1.070
201606	0.965	1.029	1.030	0.975	1.043	1.042
201607	0.936	1.023	0.991	0.941	1.021	1.021
201608	0.893	1.037	0.972	0.903	1.009	1.009

201609	0.890	1.006	0.962	0.906	0.983	0.983
201610	0.920	0.987	0.955	0.908	0.946	0.946
201611	0.910	0.993	0.935	0.891	0.941	0.941
201612	0.887	0.998	0.923	0.870	0.959	0.959
201701	0.947	0.994	0.930	0.900	0.948	0.948
201702	0.955	1.039	0.946	0.912	0.962	0.962
201703	0.910	1.091	0.952	0.895	0.982	0.982
201704	0.912	1.124	0.961	0.901	1.009	1.009
201705	0.921	1.094	0.959	0.900	1.001	1.001
201706	0.927	1.072	0.942	0.886	0.983	0.983
201707	0.913	1.048	0.927	0.877	0.971	0.971
201708	0.899	1.028	0.917	0.868	0.962	0.961
201709	0.915	1.030	0.918	0.867	0.957	0.957
201710	0.929	1.038	0.911	0.866	0.947	0.946
201711	0.938	1.032	0.913	0.869	0.943	0.943
201712	0.910	1.065	0.916	0.862	0.957	0.956
201801	0.962	1.081	0.922	0.866	0.972	0.971
201802	0.964	1.107	0.927	0.868	0.982	0.981
201803	0.938	1.134	0.919	0.846	0.984	0.984
201804	0.933	1.116	0.921	0.851	0.988	0.987
201805	0.941	1.094	0.915	0.845	0.976	0.976
201806	0.955	1.107	0.911	0.856	0.966	0.965
201807	0.969	1.081	0.908	0.856	0.963	0.962
201808	0.946	1.081	0.899	0.836	0.956	0.955
201809	0.942	1.098	0.900	0.833	0.969	0.968
201810	0.970	1.120	0.908	0.838	1.008	1.007
201811	0.978	1.145	0.916	0.844	1.031	1.030
201812	0.988	1.178	0.914	0.844	1.033	1.032
201901	0.986	1.137	0.899	0.829	1.011	1.011
201902	1.007	1.177	0.913	0.836	1.020	1.019
201903	1.004	1.213	0.909	0.827	1.003	1.002
201904	1.022	1.182	0.908	0.837	1.007	1.006
201905	1.032	1.169	0.903	0.826	1.006	1.005
201906	1.026	1.193	0.889	0.815	0.998	0.996

201907	1.022	1.166	0.884	0.807	0.979	0.977
201908	1.011	1.185	0.874	0.782	0.984	0.983
201909	1.024	1.217	0.882	0.801	0.984	0.983
201910	1.034	1.176	0.878	0.788	0.979	0.978
201911	1.039	1.195	0.873	0.798	0.959	0.959
201912	1.011	1.206	0.849	0.770	0.952	0.951
202001	1.038	1.165	0.847	0.777	0.944	0.943
202002	1.010	1.205	0.845	0.757	0.945	0.944
202003	1.037	1.203	0.845	0.763	0.949	0.947
202004	1.043	1.197	0.845	0.761	0.939	0.938
202005	1.045	1.219	0.847	0.761	0.933	0.931
202006	1.068	1.190	0.844	0.754	0.904	0.903
202007	1.069	1.184	0.837	0.750	0.897	0.895
202008	1.042	1.185	0.827	0.732	0.893	0.892
202009	1.013	1.174	0.805	0.709	0.882	0.881
202010	1.041	1.183	0.806	0.719	0.880	0.879
202011	1.039	1.140	0.796	0.710	0.862	0.860
202012	1.019	1.087	0.772	0.695	0.847	0.846
202101	1.030	1.028	0.766	0.693	0.825	0.824
202102	0.996	1.085	0.750	0.664	0.832	0.831
202103	1.014	1.117	0.751	0.668	0.834	0.833
202104	1.017	1.113	0.760	0.675	0.853	0.852
202105	1.015	1.100	0.750	0.660	0.852	0.850
202106	1.025	1.079	0.752	0.666	0.851	0.850
202107	0.983	1.075	0.733	0.633	0.840	0.839
202108	1.009	1.074	0.733	0.645	0.832	0.831
202109	1.005	1.071	0.735	0.649	0.837	0.836
202110	1.020	1.097	0.736	0.644	0.840	0.839
202111	1.043	1.101	0.740	0.655	0.842	0.840
202112	1.026	1.131	0.732	0.639	0.849	0.848
202201	1.007	1.138	0.739	0.629	0.862	0.860
202202	1.017	1.152	0.734	0.629	0.868	0.866
202203	0.992	1.181	0.738	0.623	0.868	0.867
202204	1.018	1.169	0.753	0.655	0.881	0.879

202205	1.015	1.146	0.737	0.646	0.875	0.873
202206	0.992	1.148	0.737	0.630	0.872	0.869

注)この出力は、付論 B3 の 116 ページの推計結果として出力されたものである

$P_{TDC_D}^t$ と P_{TDC}^t を比較すると、 P_{TDC}^t のほうが $P_{TDC_D}^t$ を下回っている。一方で重み付けされた P_{WTDC}^t と $P_{WTDC_D}^t$ はほぼ一致しており、区別することができない。我々の推奨するヘドニック指数である P_{WTDC}^t については、日次集計されたものも、月次集計されたものもほぼ一致しているため、以降はこれまで通り月次集計された 54647 件のデータに対して議論を進めていく。

表 4.重み付き時間ダミー特性回帰の推定結果

Coef.	Estimate	Coef.	Estimate	Coef.	Estimate	Coef.	Estimate
b ₀	-1.5141 ***	ρ ₃₄	0.0073	ρ ₆₇	-0.1761 ***	b _{B02}	0.1388 ***
ρ ₂	0.0236 **	ρ ₃₅	0.0298 ***	ρ ₆₈	-0.1854 ***	b _{B03}	0.1419 ***
ρ ₃	0.0505 ***	ρ ₃₆	0.0318 ***	ρ ₆₉	-0.1794 ***	b _{B04}	0.0350 ***
ρ ₄	0.0631 ***	ρ ₃₇	0.0105	ρ ₇₀	-0.1757 ***	b _{B05}	0.4205 ***
ρ ₅	0.0672 ***	ρ ₃₈	0.0193 *	ρ ₇₁	-0.1742 ***	b _{B06}	0.4580 ***
ρ ₆	0.0415 ***	ρ ₃₉	0.0021	ρ ₇₂	-0.1653 ***	b _{B07}	0.4461 ***
ρ ₇	0.0205 **	ρ ₄₀	0.0059	ρ ₇₃	-0.1506 ***	b _{B08}	0.2805 ***
ρ ₈	0.0089	ρ ₄₁	0.0045	ρ ₇₄	-0.1437 ***	b _{B09}	0.4538 ***
ρ ₉	-0.0175 *	ρ ₄₂	-0.0037	ρ ₇₅	-0.1429 ***	b _{B10}	0.8450 ***
ρ ₁₀	-0.0558 ***	ρ ₄₃	-0.0231 **	ρ ₇₆	-0.1288 ***	b _{CPU2}	0.0480 ***
ρ ₁₁	-0.0611 ***	ρ ₄₄	-0.0174 *	ρ ₇₇	-0.1361 ***	b _{CPU3}	0.2753 ***
ρ ₁₂	-0.0421 ***	ρ ₄₅	-0.0172 *	ρ ₇₈	-0.1399 ***	b _{CPU4}	-0.0301 *
ρ ₁₃	-0.0534 ***	ρ ₄₆	-0.0226 **	b _{C2}	0.0365 ***	b _{CPU5}	0.1247 ***
ρ ₁₄	-0.0388 ***	ρ ₄₇	-0.0423 ***	b _{C3}	0.1277 ***	b _{CPU6}	0.1528 ***
ρ ₁₅	-0.0178 *	ρ ₄₈	-0.0499 ***	b _{C4}	0.1555 ***	b _{CPU7}	0.2936 ***
ρ ₁₆	0.0088	ρ ₄₉	-0.0587 ***	b _{C5}	0.1546 ***	b _{CPU8}	0.2866 ***
ρ ₁₇	0.0009	ρ ₅₀	-0.0573 ***	b _{C6}	0.1645 ***	b _{CPU9}	0.2417 ***
ρ ₁₈	-0.0173 *	ρ ₅₁	-0.0539 ***	b _{C7}	0.1881 ***	b _{CPU10}	0.1691 ***
ρ ₁₉	-0.0292 ***	ρ ₅₂	-0.0637 ***	b _{M2}	0.1824 ***	b _{GPU2}	-0.0566 *
ρ ₂₀	-0.0394 ***	ρ ₅₃	-0.0711 ***	b _{M3}	0.2547 ***	b _{GPU3}	-0.1061 ***
ρ ₂₁	-0.0444 ***	ρ ₅₄	-0.1017 ***	b _{M4}	0.4656	b _{SSD}	0.3340 ***
ρ ₂₂	-0.0552 ***	ρ ₅₅	-0.1107 ***	b _{M5}	1.0092 ***	b _{ST2}	0.4098 ***
ρ ₂₃	-0.0590 ***	ρ ₅₆	-0.1140 ***	b _{S2}	-0.0350 ***	b _{ST3}	0.5957 ***
ρ ₂₄	-0.0445 ***	ρ ₅₇	-0.1268 ***	b _{S3}	0.0636 ***	b _{ST4}	0.6814 ***
ρ ₂₅	-0.0293 ***	ρ ₅₈	-0.1289 ***	b _{S4}	-0.0386 ***	b _{ST5}	0.8212 ***

ρ_{26}	-0.0189	*	ρ_{59}	-0.1504	***	b_{S5}	-0.0920	***	b_{ST6}	0.6619	***
ρ_{27}	-0.0163	*	ρ_{60}	-0.1672	***	b_{S6}	0.0125		b_{W2}	-0.0343	***
ρ_{28}	-0.0127	.	ρ_{61}	-0.1942	***	b_{PI2}	-0.2616	***	b_{W3}	-0.0589	***
ρ_{29}	-0.0248	**	ρ_{62}	-0.1853	***	b_{PI3}	0.0884	***	b_{W4}	-0.0086	
ρ_{30}	-0.0351	***	ρ_{63}	-0.1825	***	b_{PI4}	0.0897	***	b_{W5}	0.0257	***
ρ_{31}	-0.0383	***	ρ_{64}	-0.1600	***	b_{PI5}	0.0137	*			
ρ_{32}	-0.0458	***	ρ_{65}	-0.1624	***	b_{PI6}	0.2769	***			
ρ_{33}	-0.0320	***	ρ_{66}	-0.1629	***	b_{H2}	-0.0271	***			

Adj.R.Squared:0.8818

$p < 0.001$:***, $p < 0.005$:**, $p < 0.01$:*, $p < 0.1$:.

また、図より P_{WTDC} と P_{TDC} の間にはかなりの差があることから、重み付けには必要な意味があり、価格と数量の情報が入手可能なのであれば、統計機関によって作成される指数は重み付けを行って作成されるべきであることが示唆される。

ただし、製品の特徴を用いた現時点で「最良」だと考えている重み付きヘドニック価格指数 P_{WTDC}^t には、

- 6年半分のデータを使って算出される遡及指数である;²⁸
- 購入者の徐々に変化する選好を反映させることができない

という2つの問題がある。そこで、この課題を解決するために隣接期間のヘドニック回帰を考える。

隣接期間の方法を開始するために、1ヶ月目と2ヶ月目に販売された商品 n の価格データを使用する。また、前述の最後の回帰で使用された商品の特性に関するデータも使用する。一番最初の隣接期間における(57)式で定義された重み付けされていない時間ダミー特性ヘドニック回帰は、次の回帰モデルで表現することができる:

$$\begin{aligned} \ln P = & \rho_2 D_2 + b_0 ONE + \sum_{j=2}^7 b_{Cj} D_{Cj} + \sum_{j=2}^5 b_{Mj} D_{Mj} + \sum_{j=2}^6 b_{Sj} D_{Sj} + \sum_{j=2}^6 b_{Pj} D_{Pj} \\ & + b_{H2} D_{H2} + \sum_{j=2}^{10} b_{Bj} D_{Bj} + \sum_{j=2}^{10} b_{Uj} D_{Uj} + \sum_{j=2}^5 b_{Wj} D_{Wj} + b_{SSD2} D_{SSD2} \\ & + \sum_{j=2}^3 b_{GPUj} D_{GPUj} + \sum_{j=2}^6 b_{STj} D_{STj} + e \dots \quad (58) \end{aligned}$$

ここで $\ln P$ は、1ヶ月目と2ヶ月目にのみ販売された製品の対数価格ベクトルである。同様に、(58)式右辺における独立変数のベクトルも54647次元ではなく、1ヶ月目と2ヶ月目に販売された製品の数に等しい次元のみである。(58)式の右側の他の特性ダミー変数の定義は、(57)式における定義と同様であるが、いまはこれらの特性ダミー変数は、1ヶ月目と2ヶ月目に販売された製品についてのみ定義される。

これをRで実行するにはどうすればよいのかと言うと、最初の2時点のデータが存在する行番号のみを取得し、この行番号に該当するデータのみを用いれば良い。例えば、最初の2時点のデータが存在する行番号は

```
IDX <- which(df$TDM<=201602)
```

²⁸ ローリング・ウィンドウのヘドニック回帰を用いることで克服することができる。新しいパネル・データの結果を以前のパネルの結果とリンクさせることをめぐる問題については、Diewert and Fox (2022; 360-361)を参照されたい。

または

```
IDX <- which(df$TDM==201601| df$TDM==201602)
```

のようにすればよい。which 関数は条件に当てはまっている要素の番号を返すので、df のうち、TDM が 201601 もしくは()201602 である df の行番号が返ってくる。あとは、先程の全サンプル期間での回帰において指定した data においてこの行番号のみを指定すればよい。

このとき $P^{1*} \equiv 1$ を 1 ヶ月目の価格指数水準とすると、 ρ_2^* は(58)式で定義される時間ダミー係数であり、これの対数を取った $\pi_2^* \equiv \exp[\rho_2^*]$ を 2 ヶ月目の価格水準 $P_2^* \equiv \pi_2^*$ として用いる。

次に、 $\ln P$ の定義を 2 ヶ月目と 3 ヶ月目にのみ販売された製品に限定する。新しい隣接期間ヘドニック回帰は、(58)式右辺における時間ダミー項 $\rho_2 D_2$ を、 $\rho_3 D_3$ に置き換えた以外は(58)式で定義されたものと同様となる。 ρ_3^* が推定されると、 $\pi_3^* \equiv \exp[\rho_3^*]$ として、3 ヶ月目の価格水準を $P^{3*} \equiv \pi_3^* P^{2*}$ のように求める。付論 B3 に掲載したプログラムでは、2 ヶ月目と 3 ヶ月目の隣接期間における回帰以降を for 文で自動的に処理している。ii を時点に関する変数だとして

```
IDX <- which(df$TDM==TDlist[ii]|df$TDM==TDlist[ii+1])
```

とすることによって、ii 回目の隣接期間のデータが含まれる行番号を取得し、その行番号に該当するデータのみを用いて推計を行う。そして得られた時点ダミーを exp 関数で真数に戻し、1つ前の時点の指数に掛け合わせて連鎖指数を作成する。たとえば、いま $X[1]$ に時点ダミーの係数が格納されているとして、PCH に ii-1 時点までの価格水準が格納されている場合、ii 時点での価格水準は

$$PCH[ii-1]*\exp(X[1])$$

のように計算すれば良い。

上記の手順を、77 ヶ月目と 78 ヶ月目に販売された製品についてまで繰り返すことで、最終的なヘドニック価格指数が求まる。これを隣接期間時間ダミー特性(Adjacent period Time Dummy Characteristic; ATDC) 価格指数 P_{ATDC}^t として以下のように定義した:

$$P_{ATDC}^t \equiv P^{t*} \dots (59)$$

ただし、(59)式で定義される価格指数は、各製品の経済的重要性を考慮していないので、満足できるものではない。時点 t に販売された製品 n の経済的重要性は、対数価格 $\ln p_{tn}$ に対応する支出シェアの平方根 $(s_{tn})^{1/2}$ を乗じることによって考慮することができる。この項 $(s_{tn})^{1/2}$ は(58)式の右辺に現れる様々なダミー変数ベクトルにも適用される。これによって、時間ダミー係数 $\rho_2^*, \rho_3^*, \dots, \rho_{78}^*$ の対数の推定値が再び得られる。これらの新しい推定値が得られたら、その指数をとって $t = 2, 3, \dots, 78$ の一連の物価水準 π_t^* を求める。時点 t における重み付き隣接期間時間ダミー特性(Weighted Adjacent period Time Dummy Characteristic; WATDC) 価格指数 P_{WATDC}^t を(59)式と同様の手順で求めたものを図 2 に示す。また、指数の値を表 5 に示す。

表 5.隣接期間時間ダミー特性回帰による指数

TDM	P_A^t	P_{UV}^t	P_{TDC}^t	P_{WTDC}^t	P_{ATDC}^t	P_{WATDC}^t
201601	1.000	1.000	1.000	1.000	1.000	1.000
201602	0.966	1.031	0.991	1.024	0.992	1.015
201603	0.943	1.066	0.980	1.052	0.980	1.035
201604	0.967	1.043	0.995	1.065	0.985	1.046
201605	0.971	1.038	1.003	1.070	0.992	1.044
201606	0.965	1.029	0.975	1.042	0.964	1.017
201607	0.936	1.023	0.941	1.021	0.930	0.990
201608	0.893	1.037	0.903	1.009	0.889	0.969
201609	0.890	1.006	0.906	0.983	0.884	0.939
201610	0.920	0.987	0.908	0.946	0.883	0.904
201611	0.910	0.993	0.891	0.941	0.860	0.889
201612	0.887	0.998	0.870	0.959	0.839	0.904
201701	0.947	0.994	0.900	0.948	0.862	0.899
201702	0.955	1.039	0.912	0.962	0.860	0.896
201703	0.910	1.091	0.895	0.982	0.845	0.898
201704	0.912	1.124	0.901	1.009	0.849	0.910
201705	0.921	1.094	0.900	1.001	0.843	0.898
201706	0.927	1.072	0.886	0.983	0.827	0.879
201707	0.913	1.048	0.877	0.971	0.813	0.866
201708	0.899	1.028	0.868	0.961	0.802	0.855
201709	0.915	1.030	0.867	0.957	0.798	0.850
201710	0.929	1.038	0.866	0.946	0.793	0.838
201711	0.938	1.032	0.869	0.943	0.791	0.830
201712	0.910	1.065	0.862	0.956	0.782	0.833
201801	0.962	1.081	0.866	0.971	0.784	0.837
201802	0.964	1.107	0.868	0.981	0.779	0.836
201803	0.938	1.134	0.846	0.984	0.760	0.822
201804	0.933	1.116	0.851	0.987	0.763	0.821
201805	0.941	1.094	0.845	0.976	0.756	0.812
201806	0.955	1.107	0.856	0.965	0.764	0.801
201807	0.969	1.081	0.856	0.962	0.766	0.800
201808	0.946	1.081	0.836	0.955	0.744	0.790

201809	0.942	1.098	0.833	0.968	0.739	0.794
201810	0.970	1.120	0.838	1.007	0.735	0.805
201811	0.978	1.145	0.844	1.030	0.739	0.803
201812	0.988	1.178	0.844	1.032	0.736	0.799
201901	0.986	1.137	0.829	1.011	0.719	0.792
201902	1.007	1.177	0.836	1.019	0.724	0.793
201903	1.004	1.213	0.827	1.002	0.716	0.780
201904	1.022	1.182	0.837	1.006	0.724	0.779
201905	1.032	1.169	0.826	1.005	0.715	0.777
201906	1.026	1.193	0.815	0.996	0.705	0.773
201907	1.022	1.166	0.807	0.977	0.696	0.757
201908	1.011	1.185	0.782	0.983	0.673	0.758
201909	1.024	1.217	0.801	0.983	0.688	0.755
201910	1.034	1.176	0.788	0.978	0.676	0.744
201911	1.039	1.195	0.798	0.959	0.685	0.730
201912	1.011	1.206	0.770	0.951	0.664	0.724
202001	1.038	1.165	0.777	0.943	0.669	0.714
202002	1.010	1.205	0.757	0.944	0.653	0.721
202003	1.037	1.203	0.763	0.947	0.655	0.724
202004	1.043	1.197	0.761	0.938	0.655	0.726
202005	1.045	1.219	0.761	0.931	0.653	0.723
202006	1.068	1.190	0.754	0.903	0.645	0.705
202007	1.069	1.184	0.750	0.895	0.640	0.693
202008	1.042	1.185	0.732	0.892	0.625	0.685
202009	1.013	1.174	0.709	0.881	0.603	0.674
202010	1.041	1.183	0.719	0.879	0.610	0.672
202011	1.039	1.140	0.710	0.860	0.599	0.658
202012	1.019	1.087	0.695	0.846	0.584	0.643
202101	1.030	1.028	0.693	0.824	0.583	0.631
202102	0.996	1.085	0.664	0.831	0.559	0.630
202103	1.014	1.117	0.668	0.833	0.562	0.624
202104	1.017	1.113	0.675	0.852	0.565	0.630
202105	1.015	1.100	0.660	0.850	0.550	0.623
202106	1.025	1.079	0.666	0.850	0.551	0.624

202107	0.983	1.075	0.633	0.839	0.526	0.612
202108	1.009	1.074	0.645	0.831	0.531	0.604
202109	1.005	1.071	0.649	0.836	0.534	0.607
202110	1.020	1.097	0.644	0.839	0.529	0.606
202111	1.043	1.101	0.655	0.840	0.536	0.598
202112	1.026	1.131	0.639	0.848	0.527	0.595
202201	1.007	1.138	0.629	0.860	0.517	0.595
202202	1.017	1.152	0.629	0.866	0.514	0.595
202203	0.992	1.181	0.623	0.867	0.508	0.593
202204	1.018	1.169	0.655	0.879	0.526	0.596
202205	1.015	1.146	0.646	0.873	0.516	0.584
202206	0.992	1.148	0.630	0.869	0.509	0.580

注)この出力は、付論 B3 の 121 ページの推計結果として出力されたものである

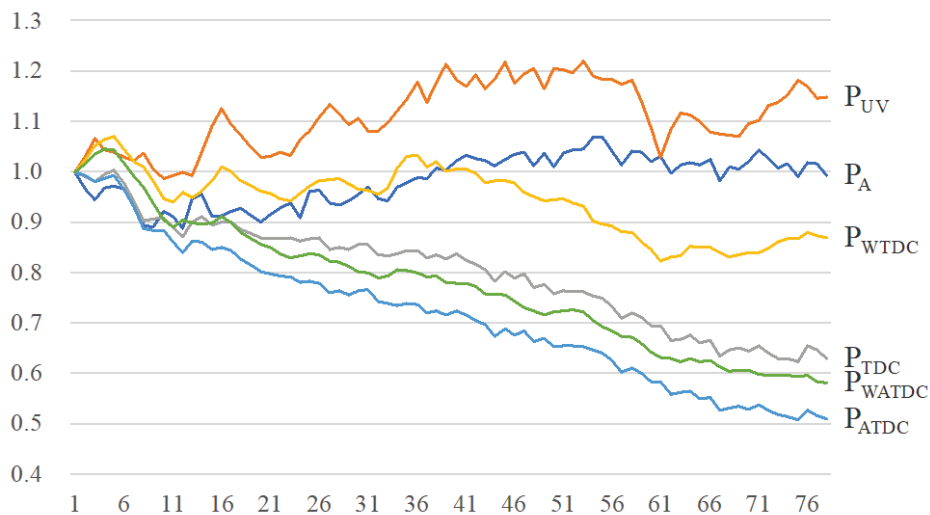


図 2. 製品特性を用いた隣接期間ダミー価格指数

ここでは、 P_{WATDC}^t と P_{WTDC}^t に対する長所と短所をいくつか挙げる：

- 全期間における回帰では品質調整パラメータが固定されるのに対し、隣接期間での回帰ではそれぞれ新しい品質調整パラメータを推定するため、隣接期間指数はデータにはるかによく適合する。
- 品質調整パラメータを毎月変更できるようにする隣接期間方式は、購入者が長期にわたって安定して一貫した選好を持たない可能性があることを意味し、経済学者の中には、その結果生じる指数の矛盾に異議を唱える人もいるだろう。
- 隣接期間指数は連鎖指数である。月次の商品価格と数量に大きな変動がある場合、これらの指数がチェーン・ドリフトの問題にさらされる危険性がある。今回のデータでは、月次の価格と数量に大きな変動があるため、隣接期間指数が下向きの連鎖ドリフトを起こす危険性がある。この点については、後述の第 4 項で再検討する。

新たに求めた P_{WATDC}^t と P_{ATDC}^t は、 $t = 78$ において、 P_{WTDC}^t と P_{TDC}^t を大きく下回る。また重み付けの効果を見ると $t = 78$ において、重み付けされていない P_{ATDC}^t は、重み付けされている P_{WATDC}^t を10%近く下回っている。重み付けをしない指数は、代表的でない商品に重みを与えすぎるため、 P_{WATDC}^t の方が好ましいと考える。

ここまで代表的と考えられる商品の特性を用いて推計を行ってきたが、欠落した特性が価格指数に重大な影響を与えることはよく知られている。すべての可能な製品特性²⁹を含むモデルは、第3節で提示された時間製品ダミー(Time Product Dummy; TPD)モデルである。したがって次項では、重み付きおよび重み付けなしのTPDヘドニック回帰モデルを検討する。

4.2. 代理変数を用いたヘドニック回帰モデル

重み付き時間製品ダミーの最小二乗問題は、(20)式で定義され、この問題の一意解を得るために、正規化 $\rho_1 = 0$ を加えた。一方で、重み付けのないTPDモデルにおける最小二乗問題は、すべての支出シェア s_{tn} を1に等しく設定した(20)式で定義される。

ここでは、時間ダミー変数のベクトル D_1, \dots, D_{78} と、54647次元の2972個の製品ダミー変数のベクトル D_{J1}, \dots, D_{J2972} を用い、前節と同様に、観察されたノートパソコン価格の対数ベクトルを $\ln P$ として定義する。(60)式において、未知のパラメータは対数価格水準 $\rho_2, \rho_3, \dots, \rho_{78}$ と2972製品の対数品質調整係数 $\beta_1, \beta_2, \dots, \beta_{2972}$ である：

$$\ln P = \sum_{t=2}^{78} \rho_t D_t + \sum_{k=1}^{2972} \beta_k D_{Jk} + e \dots (60)$$

特性についてダミー変数を作成したように、JANコードについてダミー変数を作成することによって(60)式の D_{Jk} を求めることができる。そのため、前項と比較してRで実行するための新しい知識は必要ない。ただし、JANコードダミーによって作成される 54647×2972 のデータは非常に大きく、通常のパソコンでは計算できない可能性が高い。

上記の回帰を実行すると、 R^2 は0.9259であった。したがって、(60)式で定義されたモデルの適合度は、前節の(57)式で定義された重みなし回帰モデルの R^2 (0.7675)よりもかなり改善されている。推定された ρ_t^* を指数化し、 $\pi_t^* \equiv \exp[\rho_t^*]$ として得られた指数を P_{TPD}^t とする。

続けて、重み付き時間製品ダミー(Weighted Time Product Dummy; WTPD)価格指数は(60)式の両辺の誤差項を除くベクトルに、各時点で購入された製品の月別支出シェア s_{tn} の正の平方根を掛けたモデルを用いることで求めることができる。同じパラメータ $\rho_2, \rho_3, \dots, \rho_{78}, \beta_1, \beta_2, \dots, \beta_{2972}$ で線形回帰を行った結果、 R^2 は0.9777であった。ここでも $\rho_1^* = 0$ とし、推定された ρ_t^* を指数化することで、 $\pi_t^* \equiv \exp[\rho_t^*]$ によって求まる新しい指数を P_{WTPD}^t とする。

前節の製品特性を用いた指数である P_{TDC}^t, P_{WTDC}^t と同様に、 P_{WTPD}^t はリアルタイム指数ではない。そこで、リアルタイム指数を求めるために前節同様、隣接期間時間製品ダミー(Adjacent period Time Product Dummy; ATPD)回帰を実行する。

先程と同様に、隣接期間法を開始するために、1ヶ月目と2ヶ月目に販売された商品 n の価格データを使用する。 $S(1,2)$ を1ヶ月目と2ヶ月目に購入された商品の集合と定義すると、(60)式で定義される重み付け

²⁹ 対象製品の購入者の効用に影響を与える外部環境要因(時間とともに変化する)があるかもしれない。対象製品の購入者が他の製品から分離可能な嗜好を持っていると仮定しているが、これは現実への大まかな近似に過ぎない。また、製品の「新しさ」や「古さ」も、購入者の効用に影響を与えるかもしれない。

されていない時間製品ダミーヘドニック回帰と対になる。1ヶ月目と2ヶ月目に販売された製品のみを用いた回帰は、次のモデルで表現される:

$$\ln P^* = \rho_2 D_2^* + \sum_{k=1}^{2972} \beta_k D_{jk}^* + e^{t^*} = \rho_2 D_2^* + \sum_{k \in S(1,2)} \beta_k D_{jk}^* + e^{t^*} \dots (61)$$

ここで、対数価格ベクトル $\ln P^*$ 、時間ダミー・ベクトル D_2^* 、および製品ダミーベクトル $D_{j1}^*, \dots, D_{j2972}^*$ は、1ヶ月目と2ヶ月目に販売された製品に制限される。この回帰モデルの結果を用いて、隣接期間時間製品ダミー (Adjacent period Time Product Dummy; ATPD) 物価指数 $P_{ATPD}^1 = 1$ と $P_{ATPD}^2 = \exp[\rho_2^*]$ (ρ_2^* は時間ダミーのパラメータ) を定義した。

続けて実行される回帰モデルでは、 $\ln P^{**} = \rho_3 D_3^{**} + \sum_{k \in S(2,3)} \beta_k D_{jk}^{**} + e^{t^{**}}$ という形式となる。ここで、 $\ln P^{**}, \rho_3 D_3^{**}, D_{jk}^{**}, e^{t^{**}}$ は、2ヶ月目と3ヶ月目に販売された製品に限定される。推定された ρ_3^* を指数化することで、3ヶ月目の隣接期間時間製品ダミー物価指数 P_{ATPD}^3 は $P_{ATPD}^2 \times \pi_3^*$ として求められる。ただし $\pi_3^* \equiv \exp[\rho_3^*]$ である。これを77ヶ月目と78ヶ月目に販売された製品に限定されるまで繰り返し、最終的な隣接時間製品ダミー価格指数 P_{ATPD}^t を作成する。また前節と同様、重み付き隣接期間時間製品ダミー (Weighted Adjacent period Time Product Dummy; WATPD) 価格指数 P_{WATPD}^t を得るために、支出シェアの平方根を乗じた77回の回帰も実行した。

これらについて、Rでの計算方法において特筆すべきことはないが、隣接期間にデータを制限したことによって、こちらであれば通常のパソコンでも計算ができる可能性が高い。

そしてこれら $P_{WATPD}^t, P_{ATPD}^t, P_{TPD}^t, P_{WTPD}^t$ と P_A^t, P_{UV}^t を図3にプロットした。なお、指数の値を表6に示す。

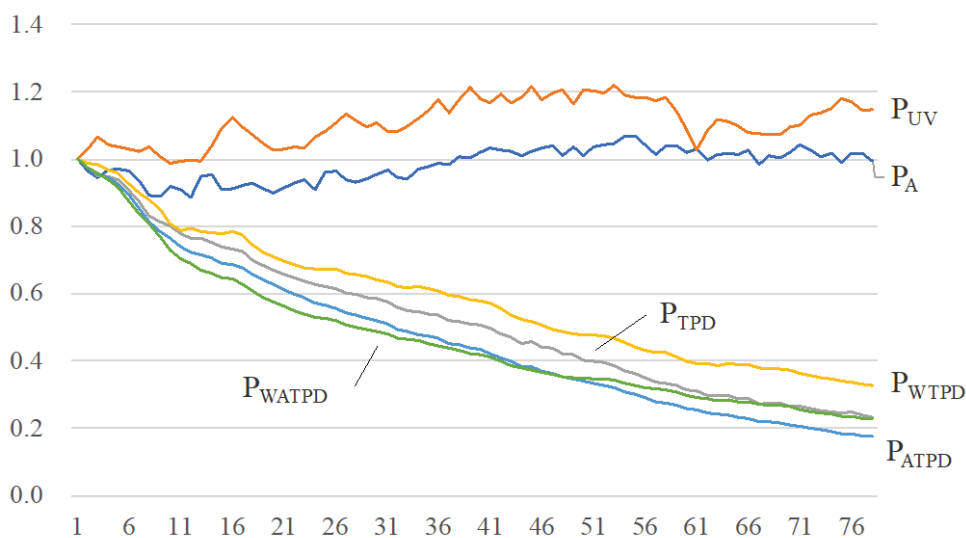


図3. 時間製品ダミーヘドニック回帰モデル価格指数

表6. 時間製品ダミーヘドニック回帰モデルによる価格指数の値

TDM	P_A^t	P_{UV}^t	P_{TPD}^t	P_{WTPD}^t	P_{ATPD}^t	P_{WATPD}^t
201601	1.000	1.000	1.000	1.000	1.000	1.000
201602	0.966	1.031	0.973	0.988	0.975	0.976
201603	0.943	1.066	0.952	0.983	0.955	0.958

201604	0.967	1.043	0.949	0.969	0.944	0.938
201605	0.971	1.038	0.940	0.958	0.926	0.916
201606	0.965	1.029	0.907	0.926	0.892	0.874
201607	0.936	1.023	0.871	0.900	0.852	0.838
201608	0.893	1.037	0.830	0.879	0.814	0.809
201609	0.890	1.006	0.815	0.849	0.786	0.770
201610	0.920	0.987	0.802	0.807	0.766	0.729
201611	0.910	0.993	0.779	0.789	0.740	0.702
201612	0.887	0.998	0.764	0.796	0.723	0.690
201701	0.947	0.994	0.765	0.783	0.718	0.672
201702	0.955	1.039	0.753	0.781	0.706	0.662
201703	0.910	1.091	0.738	0.778	0.691	0.649
201704	0.912	1.124	0.731	0.784	0.685	0.643
201705	0.921	1.094	0.725	0.774	0.678	0.628
201706	0.927	1.072	0.701	0.746	0.657	0.607
201707	0.913	1.048	0.683	0.722	0.641	0.589
201708	0.899	1.028	0.670	0.708	0.626	0.574
201709	0.915	1.030	0.658	0.697	0.611	0.562
201710	0.929	1.038	0.647	0.687	0.600	0.550
201711	0.938	1.032	0.638	0.677	0.589	0.539
201712	0.910	1.065	0.628	0.674	0.573	0.531
201801	0.962	1.081	0.622	0.674	0.565	0.526
201802	0.964	1.107	0.614	0.672	0.555	0.520
201803	0.938	1.134	0.602	0.661	0.542	0.507
201804	0.933	1.116	0.598	0.658	0.535	0.502
201805	0.941	1.094	0.587	0.651	0.525	0.495
201806	0.955	1.107	0.585	0.642	0.519	0.487
201807	0.969	1.081	0.576	0.634	0.510	0.479
201808	0.946	1.081	0.560	0.620	0.495	0.468
201809	0.942	1.098	0.550	0.618	0.487	0.464
201810	0.970	1.120	0.545	0.623	0.479	0.461
201811	0.978	1.145	0.541	0.616	0.473	0.453
201812	0.988	1.178	0.537	0.607	0.469	0.446
201901	0.986	1.137	0.519	0.596	0.452	0.438

201902	1.007	1.177	0.516	0.593	0.448	0.432
201903	1.004	1.213	0.510	0.582	0.437	0.422
201904	1.022	1.182	0.506	0.581	0.434	0.418
201905	1.032	1.169	0.496	0.572	0.423	0.412
201906	1.026	1.193	0.479	0.555	0.409	0.400
201907	1.022	1.166	0.470	0.535	0.400	0.386
201908	1.011	1.185	0.452	0.524	0.384	0.378
201909	1.024	1.217	0.457	0.515	0.382	0.373
201910	1.034	1.176	0.441	0.507	0.368	0.365
201911	1.039	1.195	0.439	0.495	0.365	0.359
201912	1.011	1.206	0.422	0.488	0.352	0.355
202001	1.038	1.165	0.419	0.481	0.348	0.350
202002	1.010	1.205	0.402	0.478	0.339	0.350
202003	1.037	1.203	0.399	0.476	0.335	0.347
202004	1.043	1.197	0.395	0.474	0.329	0.347
202005	1.045	1.219	0.386	0.467	0.322	0.344
202006	1.068	1.190	0.370	0.453	0.309	0.335
202007	1.069	1.184	0.363	0.442	0.302	0.328
202008	1.042	1.185	0.351	0.432	0.292	0.322
202009	1.013	1.174	0.337	0.426	0.280	0.317
202010	1.041	1.183	0.334	0.424	0.276	0.313
202011	1.039	1.140	0.327	0.411	0.267	0.307
202012	1.019	1.087	0.315	0.400	0.260	0.298
202101	1.030	1.028	0.312	0.392	0.254	0.291
202102	0.996	1.085	0.299	0.392	0.244	0.289
202103	1.014	1.117	0.298	0.386	0.243	0.283
202104	1.017	1.113	0.297	0.391	0.241	0.283
202105	1.015	1.100	0.288	0.388	0.233	0.279
202106	1.025	1.079	0.289	0.388	0.230	0.278
202107	0.983	1.075	0.273	0.381	0.221	0.272
202108	1.009	1.074	0.275	0.376	0.218	0.267
202109	1.005	1.071	0.275	0.376	0.215	0.267
202110	1.020	1.097	0.266	0.372	0.209	0.264
202111	1.043	1.101	0.266	0.364	0.205	0.256

202112	1.026	1.131	0.257	0.355	0.200	0.250
202201	1.007	1.138	0.254	0.351	0.196	0.246
202202	1.017	1.152	0.248	0.345	0.190	0.242
202203	0.992	1.181	0.244	0.339	0.185	0.237
202204	1.018	1.169	0.248	0.338	0.184	0.236
202205	1.015	1.146	0.238	0.331	0.178	0.231
202206	0.992	1.148	0.233	0.327	0.175	0.228

注)この出力は、付論 B3 の 126 ページの推計結果として出力されたものである

P_{TPD}^t と P_{WTPD}^t については、重み付けされた指数 P_{WTPD}^t の方が前節同様に減衰率が小さく、 P_{TPD}^t よりも大きな値をとっている。これまで同様、重み付けされていない指数よりも重み付けされた指数の方がより好ましい指数であると考えられる。一方で、隣接期間指数である P_{ATPD}^t と P_{WATPD}^t は P_{TPD}^t と P_{WTPD}^t を大きく下回っており、隣接期間指数はチェーン・ドリフトの問題に悩まされている可能性が高いと考える。

次項では、伝統的な指数と、連鎖ドリフトを減少させる可能性のある Similarity Linked 指数を計算し、これまでに求めた指数との比較を行う。

4.3. 品質調整の代替的手法

時点 t において構築された指数は、時点 t の価格と時点 $(t - 1)$ の価格を比較している。しかし、2 時点間の価格の比較がすべて同じように正確であるというわけではない。時点 r と時点 t のマッチした商品の相対価格が非常に似ている場合、Laspeyres 価格指数と Paasche 価格指数は互いに非常に近くなり、これら 2 つの間に存在する「真の」価格指数は(指数理論への経済的アプローチを用いて³⁰)、Fisher 指数に非常に近くなると考えられる。特に、2 つの価格ベクトルが正確に比例している場合、この 2 時点間の価格指数は(関連する数量ベクトルが比例していなくても)比例係数に等しくなることが望ましく、この 2 時点間の Fisher 価格指数はこの比例テストを満たす。これは、相対的な価格構造が最も類似している 2 時点間で価格の両側比較が行われた場合、より正確な価格指数のセットが構築される可能性があることを示唆している。³¹相対価格の構造が最も似ている時点同士をリンクさせる Predicted Share 法は、リアルタイムで価格指数 P^t を構築する必要があるという仮定の下で説明される。³²

予備段階として、価格と数量のデータを $t = 1, \dots, 78$ について、2972 次元の 78 個の価格と数量のベクトル $p^t \equiv [p_1^t, p_2^t, \dots, p_{2972}^t]$ 、 $q^t \equiv [q_1^t, q_2^t, \dots, q_{2972}^t]$ に再編成する必要がある。商品 k が時点 t に購入されなかった場合、 $p_k^t = q_k^t = 0$ とし、時点 r と時点 t において、両時点に存在する商品 k の集合を $S(r, t)$ と定義する。すると、時点 t と時点 r の価格を関連付ける Laspeyres 指数 $P_L(r, t)$ と Paasche 指数 $P_P(r, t)$ は以下のように定義される:

$$P_L(r, t) \equiv \sum_{k \in S(r, t)} p_k^t q_k^t / \sum_{k \in S(r, t)} p_k^r q_k^r \quad \dots (62)$$

³⁰ Fisher 指数と指数理論の経済学的アプローチとの関係については、Diewert (1976)を参照されたい。

³¹ 国間の物価比較を行う文脈では、相対価格の構造が最も類似している国を結びつける手法が Hill (1997) (1999a) (1999b) (2009)、Hill and Timmer (2006)、Diewert (2009) (2013) (2023)、Hill, Rao, Shankar and Hajargasht (2017)によって追求されてきた。Hill (2001) (2004) も時系列の文脈でこの相対価格の類似性アプローチを追求している。

³² この方法については、Diewert (2023)で詳しく説明されている。

$$P_P(r, t) \equiv \sum_{k \in S(r,t)} p_k^t q_k^t / \sum_{k \in S(r,t)} p_k^r q_k^t \dots (63)$$

時点 t における価格は(62)式と(63)式の分子にあり、時点 r の対応する価格は(62)式と(63)式の分母にある。時点 t の価格と時点 r の価格を関係付けた Fisher 指数は、 $P_L(r, t)$ と $P_P(r, t)$ の幾何平均として次のように定義される:³³

$$P_F(r, t) \equiv [P_L(r, t)P_P(r, t)]^{1/2} \dots (64)$$

また、2972 商品の時点 t における支出シェア s_k^t のベクトル $s_k^t \equiv [s_1^t, s_2^t, \dots, s_{2972}^t]$ を以下のように定義する:

$$s_k^t \equiv p_k^t q_k^t / p^t \cdot q^t \dots (65)$$

ここで、 $p^t \cdot q^t$ はベクトル p^t と q^t の内積を表す。

相対価格の類似性の尺度の選択は、類似した時点をリンクさせる手法において重要な役割を果たす。相対価格構造の類似性または非類似性の様々な尺度が、Allen and Diewert (1981), Kravis, Heston and Summers (1982; 104-106), Hill (1997) (2009), Sergeev (2001) (2009), Hill and Timmer (2006), Aten and Heston (2009) and Diewert (2009) (2023) によって提案されている。相対価格の類似性のほとんどの尺度の問題点は、いくつかの製品が欠落している場合、それらがうまく定義されないことである。そこでここでは、比較される2時点におけるいくつかの製品価格がゼロに等しくても定義されるよう、次のような非類似度尺度 (Dissimilarity)を用いる:³⁴

$$\Delta(p^r, p^t, q^r, q^t) \equiv \sum_{k=1}^{2972} [s_k^t - (p_k^r q_k^t / p^r \cdot q^t)]^2 + \sum_{k=1}^{2972} [s_k^r - (p_k^t q_k^r / p^t \cdot q^r)]^2 \dots (66)$$

$\Delta(p^r, p^t, q^r, q^t)$ が r と t の任意の組の期間に対して定義されるためには、 $r = 1, \dots, 78$ と $t = 1, \dots, 78$ に対して $p^r \cdot q^t > 0$ であることが必要である。(66)式の右辺の2つの和は2乗項の和なので、 $\Delta(p^r, p^t, q^r, q^t) \geq 0$ であり、 $\Delta(p^r, p^t, q^r, q^t) = 0$ であれば、時点 r と t の価格ベクトルは比例する。もし物価が2時点で比例しているなら、2時点における物価指数は比例係数に等しいはずである。つまり、ある正の比例係数 λ に対して $p^t = \lambda p^r$ とすれば、(64)式で定義される Fisher 指数 $P_F(r, t)$ は λ と等しくなる。(66)式で定義される類似性の尺度におけるもう1つの非常に重要な特性は、2つの時点 r と t にまたがる製品のマッチングの欠損に対してペナルティを与えることである。したがって、時点 r と t のマッチング価格が等しくても、2時点のどちらかできしか入手できない製品がある場合、 $\Delta(p^r, p^t, q^r, q^t)$ は0より大きくなる。

また、 $\Delta(p^r, p^t, q^r, q^t)$ を用いて $t = 1, \dots, 78$ の類似連動指数 P_S^t を構築することができる。構築方法としては、たとえば時点 $t = 4$ において計算した尺度 $\Delta(p^r, p^4, q^r, q^4)$ の値が、 $r = 3$ よりも $r = 2$ のほうが0に近い場合、 $P_S^4 = P_S^2 \times P_F(2,4)$ のようにリンクさせることによって、指数を計算すれば良い。

それでは、この連鎖をRでどのように実行すればよいだろうか。いま時点 r と時点 t の Dissimilarity が DPS2[t,r]に格納されており、時点 r を基準とした時点 t の Fisher 指数が PF[t,r]に格納されているとする。ただし、実際に Dissimilarity を計算すると共通して存在する商品がなく、DPS2にはNAとなっている要素も存在する。そこで、DPS2における欠損値は100などの大きな値で置換しておく。すると、時点iiにおける Similarity Linked Index は次のように計算できる。

³³ 6984=78×78 の Fisher(1922)指数 $P_F(r, t)$ があることに注意。

³⁴ この尺度の公理的性質については Diewert (2023)を参照。

```
tt <- which(DPS2[ii,1:(ii-1)]==min(DPS2[ii,1:(ii-1)]))
PS[ii] <- PS[tt]*PF[ii,tt]
```

まず1行目であるが、時点 ii に比較しうる Dissimilarity の値は DPS2 の ii 行目において1列目から $(ii-1)$ 列目に存在する。そのため

$$\min(\text{DPS2}[\text{ii},1:(\text{ii}-1)])$$

とすることで最小の Dissimilarity の値を取得できる。すると which 関数は条件にあった要素の番号を返してくれるので、DPS2 の ii 行目の1列目から $(ii-1)$ 列目において最小値と一致するのが何列目なのかがわかる。このようにして得られた tt がリンクに使用する基準時点になるので、2行目のように指数をリンクさせれば良い。

このようにして計算した P_S^t を $P_L(r, t), P_P(r, t), P_F(r, t)$ に基づいて作成した連鎖指数 $P_{LCH}(r, t), P_{PCH}(r, t), P_{FCH}(r, t)$ とともに図4に示す。また、指数の値を表7に掲載する。

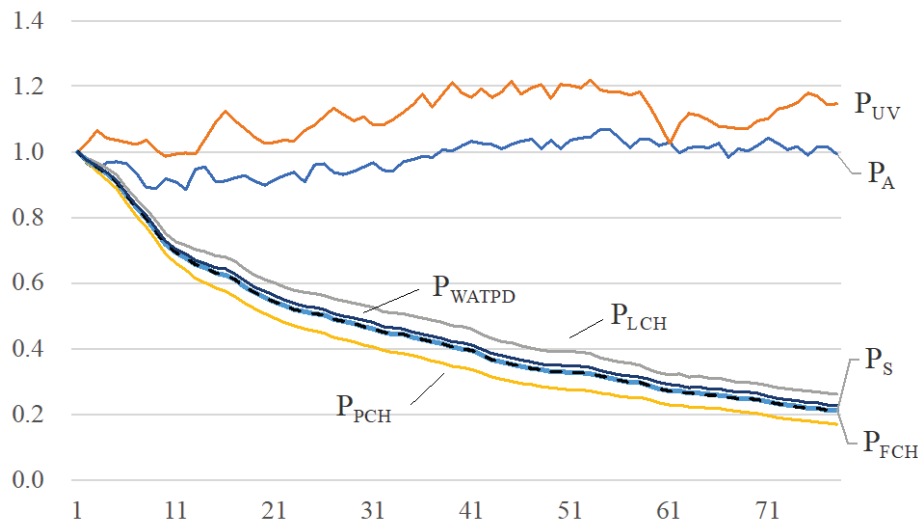


図4. 類似連動指数とその他の指数比較

連鎖 Fisher 指数には、複雑なヘドニック回帰法を必要としないという利点があり、一般消費者への説明も比較的容易である。しかし、セール品や季節性の強い商品など多くの用途では、連鎖フィッシャー指数は連鎖ドリフトを起こしやすいという欠点もある。類似連動指数は Fisher 連鎖指数に非常に近いので、連鎖ドリフトを受ける可能性が高い。³⁵したがって、 $P_S^t, P_{FCH}^t, P_{LCH}^t, P_{PCH}^t, P_{WATPD}^t$ はすべて何らかの連鎖ドリフトに悩まされている可能性が高い。この潜在的な連鎖ドリフト問題については、次の節で取り上げる。

表7. 類似連動指数とその他の指数値

TDM	P_A^t	P_{UV}^t	P_{LCH}^t	P_{PCH}^t	P_{FCH}^t	P_S^t	P_{WATPD}^t
201601	1.000	1.000	1.000	1.000	1.000	1.000	1.000
201602	0.966	1.031	0.981	0.968	0.974	0.974	0.976

³⁵ したがって、類似連動指数が Walsh の多期間同一性テストを満たすだけでは、連鎖的ドリフトの可能性を排除するには不十分である。

201603	0.943	1.066	0.967	0.941	0.954	0.954	0.958
201604	0.967	1.043	0.952	0.916	0.934	0.934	0.938
201605	0.971	1.038	0.930	0.889	0.909	0.909	0.916
201606	0.965	1.029	0.890	0.844	0.867	0.867	0.874
201607	0.936	1.023	0.855	0.805	0.830	0.830	0.838
201608	0.893	1.037	0.826	0.773	0.799	0.799	0.809
201609	0.890	1.006	0.790	0.731	0.760	0.760	0.770
201610	0.920	0.987	0.752	0.689	0.720	0.720	0.729
201611	0.910	0.993	0.724	0.662	0.692	0.692	0.702
201612	0.887	0.998	0.715	0.641	0.677	0.677	0.690
201701	0.947	0.994	0.702	0.616	0.658	0.658	0.672
201702	0.955	1.039	0.696	0.602	0.648	0.648	0.662
201703	0.910	1.091	0.684	0.585	0.632	0.632	0.649
201704	0.912	1.124	0.681	0.575	0.626	0.626	0.643
201705	0.921	1.094	0.666	0.560	0.611	0.611	0.628
201706	0.927	1.072	0.645	0.539	0.590	0.590	0.607
201707	0.913	1.048	0.626	0.522	0.572	0.572	0.589
201708	0.899	1.028	0.612	0.506	0.556	0.556	0.574
201709	0.915	1.030	0.600	0.493	0.544	0.544	0.562
201710	0.929	1.038	0.589	0.481	0.533	0.533	0.550
201711	0.938	1.032	0.578	0.470	0.521	0.521	0.539
201712	0.910	1.065	0.572	0.461	0.513	0.513	0.531
201801	0.962	1.081	0.568	0.455	0.508	0.508	0.526
201802	0.964	1.107	0.564	0.448	0.502	0.502	0.520
201803	0.938	1.134	0.551	0.436	0.490	0.490	0.507
201804	0.933	1.116	0.547	0.429	0.484	0.484	0.502
201805	0.941	1.094	0.541	0.421	0.477	0.477	0.495
201806	0.955	1.107	0.533	0.413	0.469	0.469	0.487
201807	0.969	1.081	0.526	0.406	0.462	0.462	0.479
201808	0.946	1.081	0.514	0.395	0.451	0.451	0.468
201809	0.942	1.098	0.510	0.391	0.446	0.446	0.464
201810	0.970	1.120	0.508	0.387	0.444	0.444	0.461
201811	0.978	1.145	0.501	0.378	0.435	0.435	0.453
201812	0.988	1.178	0.495	0.372	0.429	0.429	0.446

201901	0.986	1.137	0.487	0.364	0.421	0.421	0.438
201902	1.007	1.177	0.482	0.357	0.415	0.415	0.432
201903	1.004	1.213	0.471	0.347	0.404	0.404	0.422
201904	1.022	1.182	0.467	0.343	0.400	0.400	0.418
201905	1.032	1.169	0.460	0.338	0.394	0.394	0.412
201906	1.026	1.193	0.446	0.326	0.381	0.381	0.400
201907	1.022	1.166	0.431	0.314	0.368	0.368	0.386
201908	1.011	1.185	0.422	0.307	0.360	0.360	0.378
201909	1.024	1.217	0.417	0.302	0.355	0.355	0.373
201910	1.034	1.176	0.409	0.295	0.348	0.348	0.365
201911	1.039	1.195	0.403	0.290	0.342	0.342	0.359
201912	1.011	1.206	0.397	0.285	0.336	0.336	0.355
202001	1.038	1.165	0.393	0.281	0.332	0.332	0.350
202002	1.010	1.205	0.394	0.278	0.331	0.331	0.350
202003	1.037	1.203	0.391	0.276	0.329	0.329	0.347
202004	1.043	1.197	0.389	0.275	0.327	0.327	0.347
202005	1.045	1.219	0.385	0.273	0.324	0.324	0.344
202006	1.068	1.190	0.374	0.267	0.316	0.316	0.335
202007	1.069	1.184	0.367	0.261	0.309	0.309	0.328
202008	1.042	1.185	0.361	0.254	0.303	0.303	0.322
202009	1.013	1.174	0.355	0.251	0.298	0.298	0.317
202010	1.041	1.183	0.351	0.250	0.296	0.296	0.313
202011	1.039	1.140	0.338	0.245	0.288	0.287	0.307
202012	1.019	1.087	0.328	0.237	0.279	0.278	0.298
202101	1.030	1.028	0.322	0.230	0.272	0.271	0.291
202102	0.996	1.085	0.323	0.228	0.271	0.270	0.289
202103	1.014	1.117	0.316	0.223	0.265	0.264	0.283
202104	1.017	1.113	0.316	0.223	0.265	0.264	0.283
202105	1.015	1.100	0.311	0.219	0.261	0.260	0.279
202106	1.025	1.079	0.309	0.218	0.260	0.259	0.278
202107	0.983	1.075	0.303	0.213	0.254	0.253	0.272
202108	1.009	1.074	0.298	0.209	0.249	0.249	0.267
202109	1.005	1.071	0.298	0.208	0.249	0.248	0.267
202110	1.020	1.097	0.295	0.204	0.246	0.245	0.264

202111	1.043	1.101	0.288	0.198	0.239	0.238	0.256
202112	1.026	1.131	0.282	0.191	0.232	0.231	0.250
202201	1.007	1.138	0.279	0.187	0.229	0.228	0.246
202202	1.017	1.152	0.276	0.183	0.225	0.224	0.242
202203	0.992	1.181	0.271	0.179	0.221	0.220	0.237
202204	1.018	1.169	0.270	0.178	0.219	0.218	0.236
202205	1.015	1.146	0.264	0.174	0.214	0.213	0.231
202206	0.992	1.148	0.261	0.171	0.211	0.211	0.228

注)この出力は、付論 B3 の 129 ページの推計結果として出力されたものである

4.4. ヘドニック・モデルの拡張

ある価格指数が連鎖ドリフトの問題に苦しんでいるかどうかは、連鎖ドリフトに苦しんでいない「合理的な」指数と比較することで判断できる。しかし、連鎖ドリフトを起こさない「合理的な」指数を見つけるにはどうすればよいのだろうか。

この問いに答えるための出発点となるのは、第 4 節 2 項で定義した全データを用いた重み付き時間製品ダミー価格指数 P_{WTPD}^t である。まず、製品ごとの品質調整係数である β_n^* を用いて、製品ごとの線形効用関数パラメータ α_n^* を求める。いま、 $\alpha^* \equiv [\alpha_1^*, \dots, \alpha_{2972}^*]$ とし、(19)式を使って時点 t における数量水準 Q^{t**} と間接的
重み付き時間製品ダミー価格水準 P^{t**} を定義する：

$$Q^{t**} \equiv \alpha^* \cdot q^t ; P^{t**} \equiv p^t \cdot q^t / \alpha^* \cdot q^t = e^t / Q^{t**} \dots (67)$$

(67)式からわかるように $Q^{3**} / Q^{1**} = (Q^{2**} / Q^{1**}) (Q^{3**} / Q^{2**})$ であり、時点 t における総支出 $e^t \equiv p^t \cdot q^t$ も推移的であるため、 $e^{3**} / e^{1**} = (e^{2**} / e^{1**}) (e^{3**} / e^{2**})$ となる。このことから、間接的に決定された価格水準 P^{t**} も推移的である。いま、時点 t における間接的な重み付き時間製品ダミー (Indirect Weighted Time Product Dummy; IWTPD) 価格指数 P_{IWTPD}^t を、価格水準 P^{t**} の正規化として以下のように定義する：

$$P_{IWTPD}^t \equiv P^{t**} / P^{1**} \dots (68)$$

すると、価格指数 P_{IWTPD}^t も推移的であることから、循環性テストを満たすので、連鎖ドリフトが存在しないことになる。これは常に「合理的な」目標指数というわけではないが、基礎となるヘドニック回帰の R^2 が 0.9777 だったため、おそらく合理的なインデックスであると考えられる。重み付き時間製品ダミー価格指数 P_{WTPD}^t と間接的に得られた価格指数 P_{IWTPD}^t を、図 5 に示す。これら 2 つの指数はほぼ一致している。しかし、 P_{IWTPD}^t はリアルタイムの指標ではないため、リアルタイムで計算できるこの目標指数の適切な近似値を見つけなければならない。

1 つ目の近似法は、 $t = 1, 2, \dots, 78$ に対して、拡張ウィンドウ重み付き時間製品ダミー (Expanding Window weighted time product dummy; EW) 価格指数 P_{EW}^t を構築するものである。この方法を示す。

ステップ 1: $P_{EW}^1 \equiv 1$ を定義する。

ステップ 2: 第 4 節 2 項と同様に、1 ヶ月目と 2 ヶ月目に販売された製品を用いる重み付き時間製品ダミー回帰を実行する。したがって、(61)式の重み付きバージョンを実行して、製品に対応する β_k^* についての推定を得る。そして α_k^* を得るために、これらの β_k^* を指数化する。ここで、 $t = 1, 2$ に対する定義(79)式を用いて、 $P_{EW}^2 \equiv P^{2**} / P^{1**}$ を定義する。

ここで、この計算を実際に R で実行するにはどうすればよいか説明する。いま、最初の 2 時点のデータが格納されているデータフレームの行数が rowidx という変数に格納されているとする。また、重み付き対数価格が SLP、重み付き時点ダミーが df_SD、重み付き JAN ダミーが df_SDJ という変数に格納されている。このとき、この 2 時点における重み付き時間製品ダミー回帰は、次のように実行できる。

```
res <- lm("SLP~.+0",data=data.frame(SLP=SLP[rowidx],df_SD[df_SD[rowidx,-1]],df_SDJ[df_SDJ[rowidx,]]))
```

すると、得られた回帰係数の 1 つ目に時点ダミーの係数が、2 つ目以降に JAN コードダミーの係数が求まる。そこで

```
summary(res)$coef[2:dim(summary(res)$coef)[1],1]
```

とすることで、2 番目から回帰係数の最後の要素までを取得することができる。³⁶そのため、この exp をとることで α_k^* が求まる。次に、 $\alpha^* \cdot q^t$ を計算する。そのためにまず、各 JAN ダミーに α_k^* をかけたものを次のように計算する。ただし df_DJ は重み付けなしの JAN コードダミー変数である。

```
DAJ <- data.frame(DAJ0001=df_DJ[rowidx,1]*A[1])
for(jj in 2:length(JANlist)){
  DAJ <- data.frame(DAJ,df_DJ[rowidx,jj]*A[jj])
  names(DAJ)[dim(DAJ)[2]] <- paste("DAJ",formatC(jj,width=4,flag=0),sep="")
}
```

これを行方向に足し合わせるとデータごとに対応する α_k^* をもつ列が作成できた。つぎに、(77)式の Q^{t**} を計算するために tapply 関数を用いれば良い。具体的には次のとおりである。

```
NTU <- AJ*df$Q[rowidx]
U <- tapply(NTU,df$TDM[rowidx],sum)
```

tapply 関数では上記の場合、NTU を TDM ごとに足し合わせる。さらに、 $p^t \cdot q^t = e^t$ は、同様に次のように計算すれば良い。

```
V <- tapply(df$E[rowidx,],df$TDM[rowidx],sum)
```

これによって、価格水準は $P^{t**} \equiv e^t / Q^{t**}$ として計算できるため、

```
PLEV <- V / U
```

³⁶ summary(res)\$coef は、1 列目に係数、2 列目に標準誤差、3 列目に・・・という形の行列形式になっている。

によって価格水準を求める事ができた。さらに、 $P_{EW}^2 \equiv P^{2**} / P^{1**}$ なので、 $t = 2$ における価格指数は $PLEV[2]/PLEV[1]$ として求まる。あとはこれを次のステップ以降も繰り返せば良い。この際にこれまで同様、for 文によって繰り返し処理を行うのが良い。

ステップ 3: 1ヶ月目、2ヶ月目、3ヶ月目のデータ³⁷ を使って重み付き時間製品ダミー回帰を実行し、1ヶ月目、2ヶ月目、3ヶ月目に購入された製品に対応する β_k の推定値を得る。1,2 または3ヶ月目に購入されなかった製品 k については、 $\beta_k^* = 0$ とする。 $k = 1, \dots, 2972$ についての α_k^* を得るために、これら β_k^* を指数化する。ここで、 $t = 1$ および3に関する(77)式を使用して、 $P_{EW}^3 \equiv P^{3**} / P^{1**}$ を定義する。 P_{IWTPD}^3 は78ヶ月目の終わりに入手可能な、より完全なデータに基づくので、 P_{EW}^3 は一般的に P_{IWTPD}^3 と等しくはならない。

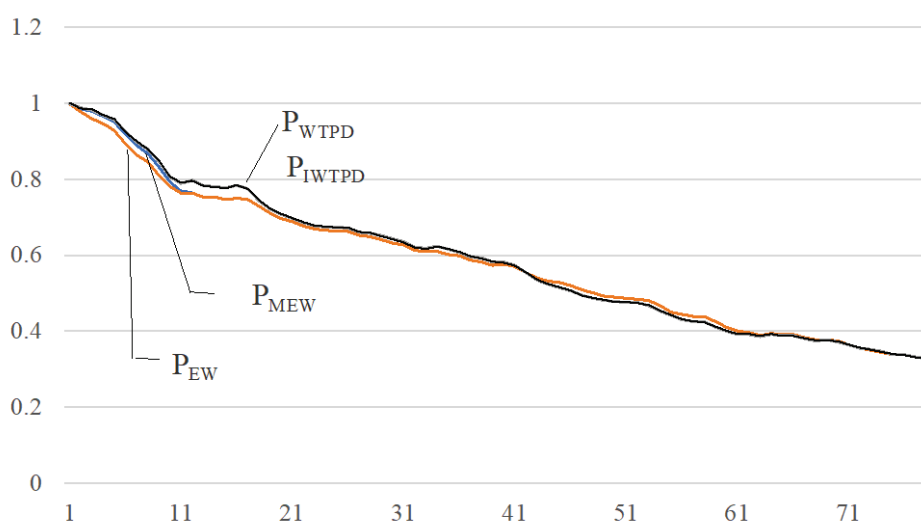
...

ステップ 78: 最後のステップでは、単純に P_{EW}^{78} を78ヶ月目のIWTPD 価格指数 P_{IWTPD}^{78} と等しく設定する。

これによって得られた価格指数 P_{EW}^t も、図5にプロットされている。統計機関が遡及的に対象製品の価格と数量のデータを収集できる場合は、修正拡張ウィンドウ重み付き時間製品ダミー(Modified Expanding Window weighted time product dummy;MEW) 価格指数 P_{MEW}^t を使用することができる。これらの指数を構築するためには、まず12ヶ月のデータウィンドウから開始し、この12ヶ月のウィンドウに対して間接的な重み付き時間製品ダミー価格指数を構築する。その後、13~78ヶ月目の拡張ウィンドウ価格指数に切り替えるだけである。そのため、Rでの計算手続きもさきほどと同様に行えば良い。

こうして出来上がった指数は、13~78ヶ月のリアルタイム指数となる。修正拡張ウィンドウ重み付き時間製品ダミー価格指数 P_{MEW}^t も図5にプロットしている。また、これらの指数の値を表8に掲載する。

直接的および間接的な重み付き時間製品ダミー価格指数である P_{WTPD}^t と P_{IWTPD}^t は、図5では区別できないほど近接している。また、拡張ウィンドウ法を用いた価格指数である P_{EW}^t と P_{MEW}^t は、 $t = 12 \sim 17$ で P_{IWTPD}^t を下回るものの、その後は P_{IWTPD}^t にかなり近接する。そのため、 P_{EW}^t と P_{MEW}^t は今回の例において、連鎖ドリフト問題に満足のいく方法で対処していると言える。



³⁷ 重み付き回帰の対応する重み付けなしの式は、 $\ln P^* = D\rho_{22}^* + D\rho_{33}^* + \sum_{k \in S(1,2,3)} \beta_k D_{jk}^* + e^{t*}$ という形式を持つ。ここで、 $S(1,2,3)$ は、1ヶ月目~3ヶ月目で販売された製品の集合である。

図 5. 拡張ウィンドウ価格指数

表 8. 拡張ウィンドウ法による指数の値

TDM	P_{MEW}^t	P_{EW}^t	P_{IWTPD}^t	P_{WTPD}^t	TDM	P_{MEW}^t	P_{EW}^t	P_{IWTPD}^t	P_{WTPD}^t
201601	1.000	1.000	1.000	1.000	201904	0.576	0.576	0.581	0.581
201602	0.986	0.976	0.988	0.988	201905	0.570	0.570	0.573	0.572
201603	0.980	0.960	0.983	0.983	201906	0.555	0.555	0.554	0.555
201604	0.965	0.946	0.970	0.969	201907	0.539	0.539	0.535	0.535
201605	0.951	0.928	0.959	0.958	201908	0.532	0.532	0.524	0.524
201606	0.917	0.892	0.927	0.926	201909	0.528	0.528	0.515	0.515
201607	0.889	0.865	0.900	0.900	201910	0.519	0.519	0.507	0.507
201608	0.868	0.845	0.879	0.879	201911	0.508	0.508	0.495	0.495
201609	0.835	0.813	0.849	0.849	201912	0.500	0.500	0.488	0.488
201610	0.793	0.781	0.807	0.807	202001	0.492	0.492	0.482	0.481
201611	0.771	0.763	0.789	0.790	202002	0.489	0.489	0.478	0.478
201612	0.763	0.763	0.796	0.796	202003	0.486	0.486	0.476	0.476
201701	0.754	0.754	0.783	0.783	202004	0.483	0.483	0.474	0.474
201702	0.753	0.753	0.781	0.781	202005	0.480	0.480	0.467	0.467
201703	0.747	0.747	0.778	0.778	202006	0.466	0.466	0.454	0.453
201704	0.751	0.751	0.784	0.784	202007	0.451	0.451	0.442	0.442
201705	0.749	0.749	0.774	0.774	202008	0.446	0.446	0.432	0.432
201706	0.731	0.731	0.745	0.746	202009	0.439	0.439	0.426	0.426
201707	0.711	0.711	0.722	0.723	202010	0.438	0.438	0.424	0.424
201708	0.695	0.695	0.708	0.708	202011	0.425	0.425	0.411	0.411
201709	0.687	0.687	0.698	0.698	202012	0.412	0.412	0.400	0.400
201710	0.677	0.677	0.687	0.687	202101	0.401	0.401	0.392	0.392
201711	0.669	0.669	0.677	0.677	202102	0.397	0.397	0.391	0.392
201712	0.665	0.665	0.675	0.674	202103	0.390	0.390	0.386	0.386
201801	0.664	0.664	0.674	0.674	202104	0.395	0.395	0.391	0.391
201802	0.663	0.663	0.673	0.672	202105	0.391	0.391	0.388	0.388
201803	0.652	0.652	0.661	0.661	202106	0.390	0.390	0.388	0.388
201804	0.649	0.649	0.658	0.658	202107	0.383	0.383	0.381	0.381
201805	0.640	0.640	0.651	0.651	202108	0.378	0.378	0.376	0.376
201806	0.633	0.633	0.643	0.642	202109	0.378	0.378	0.376	0.377

201807	0.626	0.626	0.634	0.634	202110	0.373	0.373	0.372	0.372
201808	0.612	0.612	0.620	0.620	202111	0.363	0.363	0.364	0.364
201809	0.609	0.609	0.618	0.618	202112	0.354	0.354	0.355	0.355
201810	0.610	0.610	0.623	0.623	202201	0.350	0.350	0.351	0.351
201811	0.603	0.603	0.616	0.616	202202	0.345	0.345	0.345	0.345
201812	0.599	0.599	0.607	0.607	202203	0.339	0.339	0.340	0.339
201901	0.588	0.588	0.597	0.596	202204	0.338	0.338	0.338	0.338
201902	0.583	0.583	0.593	0.593	202205	0.331	0.331	0.331	0.331
201903	0.574	0.574	0.582	0.582	202206	0.327	0.327	0.327	0.327

注)この出力は、付論 B3 の 135 ページの推計結果として出力されたものである

4.5. 新しい変数の追加:製品の入れ替わり効果

消費者は時に、市場に登場したばかりの製品にプレミアムを支払うことを厭わない。つまり、単にその「新しさ」ゆえに、この種の製品に高い価格を支払う消費者もいる。このような商品をファッション商品と呼び、ある種の衣料品、携帯電話、自動車などが代表的なものである。そこで本節では、ノートパソコンがファッション製品であるかどうかを調べるが、この場合にもヘドニック回帰モデルを用いることができる。

第3節で説明した時間製品ダミーヘドニック回帰モデルの特別なケースとして、製品それ自体と製品が市場に出てからの月数という2つの特性を考える。つまり、 p_{tn} が時点 t における商品 n の価格である場合、基本モデルは次のようになる:

$$p_{tn} \approx \pi_t \alpha_n \delta_a \dots (69)$$

ここで、 π_t は時点 t の価格水準、 α_n は製品 n の製品別品質調整パラメータ、 δ_a は時点 t に販売された製品 n の年齢「 a 」に応じて価格を調整する新たに追加した品質調整のパラメータである。購入者の嗜好がデータの全期間にわたって一定であると仮定すると、 t は1~78、 n は1~2972、「 a 」は1~85に相当する。

冒頭で述べたように、今回用いたデータでは78ヶ月の各ノートパソコンの「新しさ」または「年齢」に関する情報(変数NEW)を追加で作成した。変数NEWは、78ヶ月のサンプルで1から85の範囲であり、1は現在の時点 t が製品が販売された最初の月であることを示す。そのため、NEW変数について85個のダミー変数 $D_{A1} - D_{A85}$ を作成することができるが、これまでと同様にNEW変数をいくつかのグループに集約して、NEWダミー変数を作成する。

そのためにまず、本研究で対象としている2972製品について、市場に初めて登場してから最後に登場するまでの期間を生存時間として、製品の生存時間について分析を行う。まず、2972製品についての生存時間に関するヒストグラムと製品の累積比率を図6に示す。

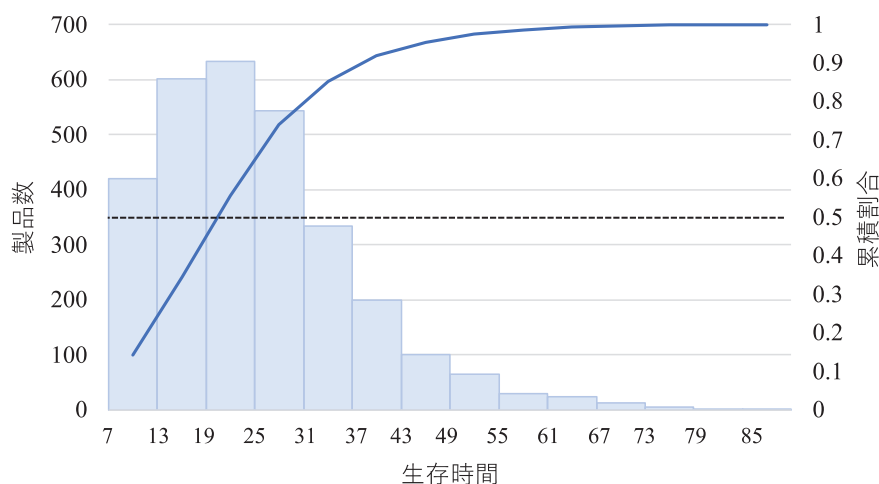


図 6. 製品の生存時間と累積製品数の割合

図 6 からわかるように、おおよそ 24 ヶ月までで半数の製品は市場から撤退もしくは購入されなくなることがわかる。一方で 84 ヶ月以上(7 年以上)も市場に存在する製品がわずかながら、存在することもわかる。³⁸また、表 2 から生存時間の平均値は 24 ヶ月程度であることが示されている。このデータには右側打ち切りのデータが含まれているため、生存時間を正しく見積もるために Kaplan-Meier 法による生存時間分析を行った。

R で生存時間分析を行うには、survival パッケージの Surv 関数を用いれば Kaplan-Meier 法による生存時間分析を行うことができる。この分析を行うには、JAN コードごとにデータとして市場に存在していた時間と、右側打ち切りであるかどうかという 2 つのデータがあれば良い。実際には次のようにして使うことができる。

Surv(time=市場に存在していた時間が格納された列名,event=打ち切りかどうかの情報が格納された列名)

その結果、2972 製品のうち右側打ち切りであるような製品は 348 製品であり、推定された 50%生存時間は 24 ヶ月であった(表 9)。このことから、NEW 変数ダミーを作成するに当たり、25 ヶ月以上は同一のダミー変数にまとめることとした。

続けて、25 ヶ月未満の NEW 変数について、どのような区分幅を設定するかが問題となる。そこで、2972 製品のうち右側打ち切りとなる 348 製品を除いた 2624 製品について、販売開始月からの累積販売額比率について月毎の平均を計算した(図 7)。12 ヶ月を超えると上昇が緩やかになり、24 ヶ月立つ頃には平均的に総売上の 99%近くに到達している。

表 9. ノートパソコンの生存時間に関する記述統計量:

	Min	Mean	Std. Dev	Max
生存時間	7.00	24.70	11.98	91.00

製品数 : 2972

³⁸ 実際 Amazon などでは、7 年以上前のモデルの Macbook も販売されている

図7から、平均して全売上の20%に到達するまでは消費者も新製品だと考えると仮定し、3ヶ月という区間幅でダミー変数を作成することとした。これによって最終的には、NEW変数について $D_{A1} - D_{A9}$ で表す9つのダミー変数ベクトルに集約した。

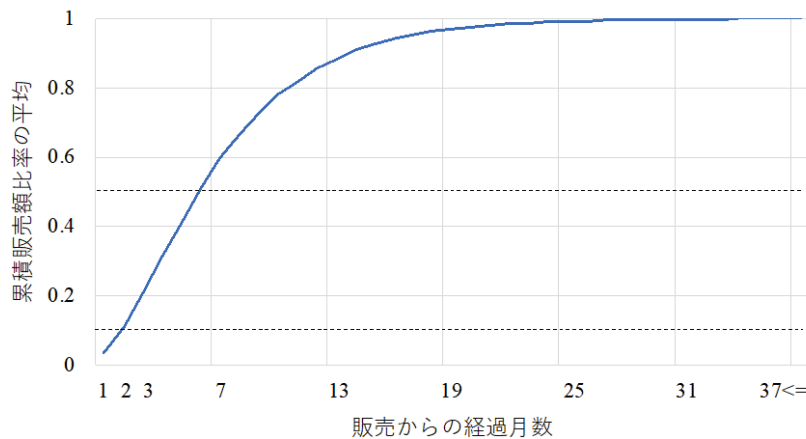


図7.販売月からの累積販売額比率の月次平均

(69)式に話を戻して、この式の両辺を対数化すると、次のような近似モデルが得られる：

$$\ln p_{tn} \approx \rho_t + \beta_n + \gamma_a \dots (70)$$

ここで、 $\rho_t \equiv \ln \pi_t$; $\beta_n \equiv \ln \alpha_n$, $\gamma_a \equiv \ln \delta_a$ である。"a"は時点 t の製品 n の適切なNEWのダミー変数を示す。(70)式において右辺のパラメータを一度にすべて推定することはできないため、少なくとも2つの正規化(たとえば、 $\rho_1 = 0$ や $\gamma_1 = 0$)が必要となる。

いま、 $S(1,2) \equiv S(1) \cup S(2)$ を、1ヶ月目と2ヶ月目に購入された商品の集合と定義すると、(70)式に基づいて1ヶ月目と2ヶ月目の価格だけを使った隣接期間ヘドニック回帰モデルとして次の回帰モデルを考えることができる：

$$\ln P^* = \rho_2 D_2^* + \sum_{k \in S(1,2)} \beta_k D_{jk}^* + \sum_{i=2}^9 \gamma_i D_{Ai}^* + e^{t^*} \dots (71)$$

(71)式において対数価格ベクトル $\ln P^*$ 、時間ダミー変数 D_2^* 、製品ダミー変数 $D_{j1}^*, \dots, D_{j2972}^*$ 、新しさを表すダミー変数ベクトル D_{Ai}^* は、1ヶ月目と2ヶ月目に販売された製品 n に制限される。この回帰の結果は、第4節2項で用いた隣接期間時間製品ダミーヘドニック回帰モデルと比較することができる。NEW変数に関するパラメータ $\gamma_1^* - \gamma_5^*$ ³⁹を追加することで、AICが5.107減少し、新しい回帰の R^2 は0.9919であった。対応する第4節2項で実行された回帰の R^2 は0.9918だった。 γ_i^* の推定値は、 $\gamma_1^* = 0$ と正規化しているため、 $\gamma_2^* = 0.0092, \gamma_3^* = 0.0105, \gamma_4^* = 0.0447, \gamma_5^* = 0.0425$ であった。これらのパラメータ推定は、ノートパソコンの購入者は、市場で入手可能な非常に古いノートパソコンに価格プレミアムを支払うことを望んでおり、したがってノートパソコンはファッション製品ではない可能性を示している。古いモデルに対する製品プレミアムが存在した。

³⁹ 1ヶ月目と2ヶ月目のデータにおいては、 $\gamma_6^* - \gamma_9^*$ に対応するNEW変数は存在しなかったため、これらのパラメータは0としている。以下同様に、隣接期間について考えた場合に、存在しないNEW変数に対応するダミー変数にかかるパラメータについては0として扱う。

このヘドニック回帰モデルの推定結果を使って、1ヶ月目と2ヶ月目に販売された商品コードと商品の「新しさ」を価格決定特性として使用した、重み付けなし隣接期間時間製品ダミーNEW(Adjacent period time Product dummy Newness; APN)価格指数 P_{APN}^t を定義する。 $P_{APN}^{1*} \equiv 1$ を1ヶ月目の指数水準として、(71)式で定義される回帰の月次ダミーの推定係数 ρ_2^* を用いて $P_{APN}^{2*} \equiv \pi_2^*$ と計算する。ただし、 $\pi_2^* \equiv \exp[\rho_2^*]$ である。

続いて3ヶ月目の価格指数を定義するために、 $\ln P$ の定義を2ヶ月目と3ヶ月目にのみ販売された商品に限定した。新しい隣接期間ヘドニック回帰は、様々なベクトルが2ヶ月目と3ヶ月目に売れた商品に制限され、(71)式の右辺の時間ダミー項 $\rho_2 D_2^*$ を項 $\rho_3 D_3^*$ に置き換えることによって、次のようになる：

$$\ln P^{**} = \rho_3 D_3^{**} + \sum_{k \in S(2,3)} \beta_k D_{jk}^{**} + \sum_{i=1}^9 \gamma_i D_{Ai}^{**} + e^{t**} \dots (72)$$

ρ_3^* が回帰の結果によって推定されると、 $\pi_3^* \equiv \exp[\rho_3^*]$ を用いて3ヶ月目の価格指数を $P_{APN}^{3*} \equiv \pi_3^* P_{APN}^{2*}$ のように求める。この回帰の結果は、第4節2項で実行したNEWダミー変数を除いた対応する重み付けなし隣接期間時間製品ダミー回帰と比較できる。NEWダミーを追加した回帰モデルではAICが、9.461減少した。 R^2 は0.9873で、対応する第4節2項の回帰では R^2 が0.9872だった。 γ_i^* の推定値は、 $\gamma_1^* = 0$ と正規化しているため、 $\gamma_2^* = -0.0301$, $\gamma_3^* = -0.0352$, $\gamma_4^* = -0.0256$, $\gamma_5^* = -0.0026$ であった。これらのパラメータ推定値は、ノートパソコンの購入者が、市場で入手可能な新しい製品に対して価格プレミアムを支払うだけでなく、古い製品に対しても若干の価格プレミアムを支払うことを示している。

この回帰モデルを、77ヶ月目と78ヶ月目に販売された商品の対数商品価格のみを使用した推定が行われるまで続ける。最終的なヘドニック回帰では、 ρ_{78}^* の推定値を得ることができ、 $\pi_{78}^* \equiv \exp[\rho_{78}^*]$ によって、78時点目のNEW変数を追加した重み付けのない隣接期間時間製品ダミー価格指数 $P_{APN}^{78*} \equiv \pi_{77}^* P_{APN}^{77*}$ を求める。この指数 P_{APN}^t を図8にプロットした。

重み付けされていない価格指数 P_{APN}^t は、各製品の経済的重要性を考慮していないので、満足できるものではない。いつものように、時点 t で販売された製品 n の経済的重要性は、このヘドニック回帰に現れる対数価格 $\ln p_{tn}$ に、対応する支出シェアの平方根 $s_{tn}^{1/2}$ を乗じることによって、(71)式で定義される最初の回帰モデルで考慮することができる。 $s_{tn}^{1/2}$ の項は(71)式で与えられる形式の推定方程式の右辺に現れるさまざまなダミー変数ベクトルにも適用される。

(71)式で定義される重み付けなしの回帰に対応する、重み付き回帰の結果は、NEWダミー変数を除いた、対応する隣接期間時間製品ダミー回帰の結果と比較できる。新しい重み付き回帰のAICは、5つの新しいNEWダミー変数を追加することで0.508減少した。 R^2 は0.9958で、対応する5節における回帰の R^2 も0.9958だった。 γ_i^* の推定値は次のとおりであった(括弧内は対応する重みなし推定)：0.0190 (0.0092), 0.0256 (0.0105), 0.0313 (0.0447), 0.0401 (0.0425)。このように重み付けを行った場合でも、消費者は市場で入手可能な古い製品にプレミアムを支払うことを示している。また決定係数は、NEWダミー変数を追加した場合でもほとんど改善されなかった。

$t = 1, \dots, 78$ について、時点 t における重み付き隣接期間時間製品ダミーNEW(Weighted Adjacent period time Product dummy NEWNESS)価格指数を P_{WAPN}^t として、 $P_{WAPN}^{1*} \equiv 1$ を1ヶ月目の指数水準と定義する。 ρ_2^* を(81)式で定義される回帰の2ヶ月目の時間ダミー係数とし、 $P_{WAPN}^{2*} \equiv \pi_2^*$ と定義する。ただし、 $\pi_2^* \equiv \exp[\rho_2^*]$ である。

続けて(82)式で定義された2ヶ月目と3ヶ月目に販売された製品のみを用いた重みなし回帰に対応する重み付き回帰を実行する。推定された ρ_3^* から $\pi_3^* \equiv \exp[\rho_3^*]$ を求め、3ヶ月目の価格指数として $P_{WAPN}^{3*} \equiv \pi_3^* P_{WAPN}^{2*}$ を計算する。この回帰の結果は、NEWダミー変数を除外した対応する重み付き隣接期間時間製品ダミー回帰と比較できる。新しい回帰を実行した結果、 γ_i^* を追加することによってAICが283.335減少し

た。R²は0.997で、対応する第4節2項でのR²は0.9960だった。γ_i^{*}の新しい推定値は、γ₂^{*} = -0.0442, γ₃^{*} = -0.0593, γ₄^{*} = -0.0506, γ₅^{*} = -0.0738であった。これらのパラメータ推定値は、これまでとは逆にノートパソコンの購入者が、市場で入手可能な新しいノートパソコンに対して価格プレミアムを支払うことを示している。ここでも、新しい回帰の決定係数は対応する第4節2項の回帰よりも大きくは改善されなかった。

上記の重み付き回帰の手順は、77ヶ月目と78ヶ月目に販売された商品に対する回帰が実行されるまで続けられる。最終的なヘドニック回帰から、ρ₇₈^{*}が推定されると、π₇₈^{*} ≡ exp[ρ₇₈^{*}]によって78時点目の重み付き隣接期間価格指数 $P_{WAPN}^{78*} \equiv \pi_{78}^* P_{WAPN}^{77*}$ を求める。得られた指数 P_{WAPN}^t も図8にプロットした。また、これら指数の値は表10に示す。

表 10. NEW ダミーを追加した時間製品ダミー回帰の結果得られた指数の値

TDM	P_A^t	P_{UV}^t	P_{ATPD}^t	P_{WATPD}^t	P_{APN}^t	P_{WAPN}^t
201601	1.000	1.000	1.000	1.000	1.000	1.000
201602	0.966	1.031	0.975	0.976	0.970	0.968
201603	0.943	1.066	0.955	0.958	0.947	0.947
201604	0.967	1.043	0.944	0.938	0.940	0.944
201605	0.971	1.038	0.926	0.916	0.921	0.920
201606	0.965	1.029	0.892	0.874	0.889	0.875
201607	0.936	1.023	0.852	0.838	0.855	0.845
201608	0.893	1.037	0.814	0.809	0.818	0.816
201609	0.890	1.006	0.786	0.770	0.789	0.775
201610	0.920	0.987	0.766	0.729	0.766	0.725
201611	0.910	0.993	0.740	0.702	0.740	0.692
201612	0.887	0.998	0.723	0.690	0.726	0.682
201701	0.947	0.994	0.718	0.672	0.720	0.663
201702	0.955	1.039	0.706	0.662	0.708	0.652
201703	0.910	1.091	0.691	0.649	0.696	0.641
201704	0.912	1.124	0.685	0.643	0.693	0.642
201705	0.921	1.094	0.678	0.628	0.687	0.626
201706	0.927	1.072	0.657	0.607	0.664	0.605
201707	0.913	1.048	0.641	0.589	0.648	0.588
201708	0.899	1.028	0.626	0.574	0.634	0.574
201709	0.915	1.030	0.611	0.562	0.618	0.561
201710	0.929	1.038	0.600	0.550	0.606	0.549
201711	0.938	1.032	0.589	0.539	0.596	0.539
201712	0.910	1.065	0.573	0.531	0.581	0.529
201801	0.962	1.081	0.565	0.526	0.571	0.526

201802	0.964	1.107	0.555	0.520	0.561	0.518
201803	0.938	1.134	0.542	0.507	0.545	0.503
201804	0.933	1.116	0.535	0.502	0.540	0.499
201805	0.941	1.094	0.525	0.495	0.532	0.496
201806	0.955	1.107	0.519	0.487	0.527	0.485
201807	0.969	1.081	0.510	0.479	0.517	0.479
201808	0.946	1.081	0.495	0.468	0.501	0.469
201809	0.942	1.098	0.487	0.464	0.494	0.465
201810	0.970	1.120	0.479	0.461	0.485	0.463
201811	0.978	1.145	0.473	0.453	0.479	0.454
201812	0.988	1.178	0.469	0.446	0.475	0.448
201901	0.986	1.137	0.452	0.438	0.457	0.442
201902	1.007	1.177	0.448	0.432	0.451	0.436
201903	1.004	1.213	0.437	0.422	0.438	0.425
201904	1.022	1.182	0.434	0.418	0.434	0.422
201905	1.032	1.169	0.423	0.412	0.423	0.416
201906	1.026	1.193	0.409	0.400	0.408	0.405
201907	1.022	1.166	0.400	0.386	0.399	0.396
201908	1.011	1.185	0.384	0.378	0.381	0.387
201909	1.024	1.217	0.382	0.373	0.380	0.382
201910	1.034	1.176	0.368	0.365	0.367	0.374
201911	1.039	1.195	0.365	0.359	0.363	0.369
201912	1.011	1.206	0.352	0.355	0.349	0.364
202001	1.038	1.165	0.348	0.350	0.346	0.360
202002	1.010	1.205	0.339	0.350	0.336	0.358
202003	1.037	1.203	0.335	0.347	0.330	0.355
202004	1.043	1.197	0.329	0.347	0.324	0.353
202005	1.045	1.219	0.322	0.344	0.318	0.349
202006	1.068	1.190	0.309	0.335	0.306	0.339
202007	1.069	1.184	0.302	0.328	0.298	0.333
202008	1.042	1.185	0.292	0.322	0.288	0.325
202009	1.013	1.174	0.280	0.317	0.277	0.320
202010	1.041	1.183	0.276	0.313	0.272	0.316
202011	1.039	1.140	0.267	0.307	0.264	0.310

202102	1.019	1.087	0.260	0.298	0.257	0.302
202101	1.030	1.028	0.254	0.291	0.252	0.292
202102	0.996	1.085	0.244	0.289	0.241	0.290
202103	1.014	1.117	0.243	0.283	0.241	0.284
202104	1.017	1.113	0.241	0.283	0.240	0.284
202105	1.015	1.100	0.233	0.279	0.231	0.280
202106	1.025	1.079	0.230	0.278	0.228	0.279
202107	0.983	1.075	0.221	0.272	0.221	0.273
202108	1.009	1.074	0.218	0.267	0.220	0.267
202109	1.005	1.071	0.215	0.267	0.217	0.268
202110	1.020	1.097	0.209	0.264	0.210	0.265
202111	1.043	1.101	0.205	0.256	0.208	0.258
202112	1.026	1.131	0.200	0.250	0.202	0.250
202201	1.007	1.138	0.196	0.246	0.198	0.247
202202	1.017	1.152	0.190	0.242	0.192	0.244
202203	0.992	1.181	0.185	0.237	0.187	0.240
202204	1.018	1.169	0.184	0.236	0.188	0.238
202205	1.015	1.146	0.178	0.231	0.180	0.233
202206	0.992	1.148	0.175	0.228	0.179	0.230

注)この出力は、付論 B3 の 138 ページの推計結果として出力されたものである

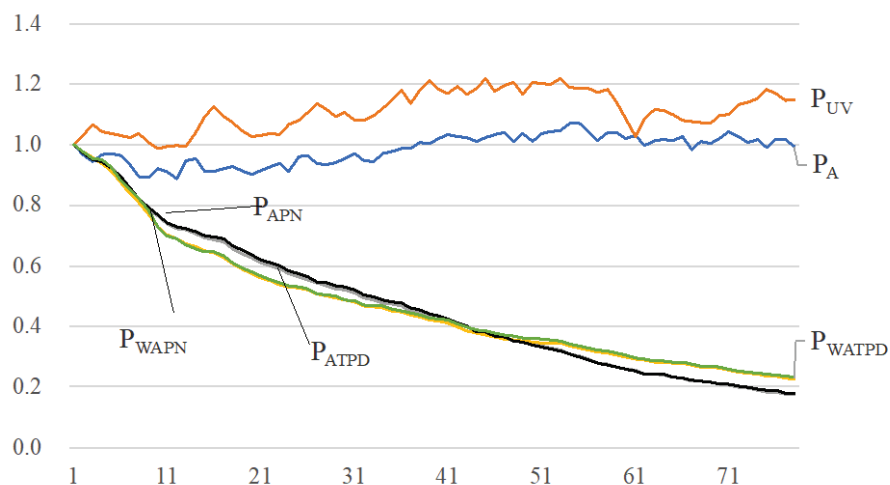


図 8. 時間製品ダミー回帰と NEW ダミー変数を追加した時間製品ダミー回帰による指数比較

ここまでは、NEW ダミーの区分幅を 3 ヶ月としてヘドニック回帰モデルを作成したが、より詳細な NEW 変数の効果を見るために、区分幅を 2 ヶ月と 6 ヶ月としたダミー変数を作成し、NEW ダミー変数の推定値の変化をみる。区分幅を 2 ヶ月と 6 ヶ月に設定した理由は、今回用いたノートパソコンのデータにおいて、製品

の累積販売額比率が平均して10%になるのが2ヶ月、50%になるのが6ヶ月であったためである。区分幅が2ヶ月の場合、1ヶ月目と2ヶ月目については次のようなモデルで表現できる。

$$(73) \ln P^* = \rho_2 D_2^* + \sum_{k \in S(1,2)} \beta_k D_{Jk}^* + \sum_{i=2}^{13} \gamma_{2,i} D_{Ai}^* + e^{t^*}$$

先程と同じように、1ヶ月と2ヶ月、2ヶ月と3ヶ月について推定を行ったところ、得られたNEWダミー変数の値は区分幅を2ヶ月とした場合、1ヶ月目と2ヶ月目の隣接期間では、 $\gamma_{2,2}^* = -0.0232(-0.0529)$, $\gamma_{2,3}^* = -0.0145(-0.0484)$, $\gamma_{2,4}^* = -0.0175(-0.0551)$, $\gamma_{2,5}^* = -0.0244(-0.0556)$, $\gamma_{2,6}^* = -0.0134(-0.0561)$, $\gamma_{2,7}^* = -0.0198(-0.0607)$ であった。ただし、()内は重み付きの場合に推定された係数である。同様に、2ヶ月目と3ヶ月目については $\gamma_{2,2}^* = -0.0248(0.0090)$, $\gamma_{2,3}^* = 0.0226(0.0504)$, $\gamma_{2,4}^* = 0.0224(0.0543)$, $\gamma_{2,5}^* = 0.1542(0.2254)$, $\gamma_{2,6}^* = 0.1216(0.2168)$, $\gamma_{2,7}^* = 0.1496(0.2127)$, $\gamma_{2,8}^* = 0.1693(0.2058)$ が得られた。

続いて、区分幅を6ヶ月にした場合の、1ヶ月目と2ヶ月目に販売された製品のモデルは次式のようになる。

$$(74) \ln P^* = \rho_2 D_2^* + \sum_{k \in S(1,2)} \beta_k D_{Jk}^* + \sum_{i=2}^5 \gamma_{6,i} D_{Ai}^* + e^{t^*}$$

この結果得られたNEW変数ダミーに関する推定値は、1ヶ月目と2ヶ月目の隣接期間では、 $\gamma_{6,2}^* = -0.0017(0.0064)$, $\gamma_{6,3}^* = -0.0069(0.0148)$ であった。同様に、2ヶ月目と3ヶ月目に売れた製品でのNEW変数ダミーに関する推定値は $\gamma_{6,2}^* = -0.0002(0.0034)$, $\gamma_{6,3}^* = 0.0278(-0.0013)$ であった。

このことから、NEW変数ダミーを作成する区分によっても推定されたパラメータの傾向が異なり、隣接する販売期間によってもその傾向が異なっているとわかる。以上のことから、ノートパソコンがファッション製品であるような明確な結果は得られなかった。また、NEW変数については、ダミー変数化する際の区分や時点によっても推定されたパラメータの傾向が大きく異なることから、モデルに追加する際には慎重に扱う必要があると考えられる。

5. 結論

経済分野において、とりわけ過去四半世紀ではマイクロ計量経済分析の発展は目覚ましいものがある。かつては、データの蓄積は紙によって行われていたことが多く、そのデータを分析用データへと変換するために、付論Aでも整理しているが、文字データをカテゴリーデータに変換したり、ダミー変数として分析できるように加工したりするために、コーディング表を作成し、それに基づき、データをコーディングし、データの入力作業を行って初めてデータ分析が可能となった。この際、データ入力には人為的なミスが伴うために、例えば二人体制で入力し、不一致のものは、再度確認して入力しなおすなど、データの生産には多大な労力を費やす必要があった。このようなデータ生成そのものにも高い研究技能が要請されていたのである。そのため、利用することができるデータ量に大きな制限があったといってもよい。または、データを発掘し、分析用データへと変換していく能力が、マイクロデータを扱う研究者に要求されていたといってもよい。

このような作業は、データの発生プロセス(Data generation process)や研究倫理を学ぶ上で極めて有効な機会でもあった。データがどのように生まれ、データ資源として蓄積され、生成されているのかを知らなければいけなかったし、データ利用においては、データの所有者と交渉する中で、一定のルールを順守しなければ、その利用ができないために、研究における倫理感が醸成されていたと考えている。

近年において、ビッグデータという言葉に象徴されるように、大量かつ多様な属性を持ったデータの利用可能性が一気に高まってきた。本稿で解説した POS データもその一つである。しかし、そのようなデータが入手できたからといって、すぐに研究を行うことができるわけではない。本稿は、マイクロデータを用いた推計を事例として、大学院生をはじめとする新しい研究者が、当該分野に参入していく中でぶつかる可能性のある障壁を乗り越えるための一助となるように、Diewert and Shimizu(2024)を題材として、一連の推計の中で利用された、データの整備、加工、推計手順などを整理し、汎用的なプログラム言語で再現し、できる限りの研究資源と情報を公開することに努めた。

筆者の一人が1990年代初頭に計量経済分析を開始した際には、利用できるプログラム言語がFortranやC++が中心である中で、Rの前身となるS言語が登場するとともに、汎用プログラムソフトとしては、SAS、SPSSと限定されたものしか利用できなかった。その後、Shazam, TSP, Gauss, Matlab, EViews, Stataといったソフトウェアが登場してくる中で、マイクロ計量分析に関する研究・教育環境は一変した。マイクロデータ、とりわけ大規模データを利用する際には、計算機だけでなく、それぞれのプログラム言語が処理可能なデータ容量も制限が存在するために、プログラム言語の特性を熟知し、研究対象に応じて使用するプログラム言語を変更しながら、研究を進めていた。現在でも、この問題は残っている。RやPhytonといった汎用的なプログラム言語も登場してきたが、大規模データになると計算時間が長かったり、計算そのものができなかったりするなどの問題にも、依然として直面することがある。

Diewert and Shimizu(2024)は、ブリテッシュ・コロンビア大学が中心となって開発したShazamによって分析がされている。そして、同論文では、2年間といった限定された期間を分析対象としている。これは、この期間だけで十分にモデルが検証できること以外にも、Shazamが持つ計算能力の限界も存在していた。Shazamは、価格指数に関するコマンドが充実しているものの、大規模データを扱うことができないという欠点も持ち合わせている。さらに、利用者が限定されているために、一般に推計プログラムを公開していくといった意味でも問題は多い。そのような意味で、本稿では、分析データの期間を5年に拡大し、汎用的なRで教材として作成することとした。そして、一つの学術論文を書き上げる中で、研究ノートに記載されていた分析上での判断や工夫を公開することで、初学者に対して公開することにした。

Diewert and Shimizu(2024)、または分析対象期間を5年に拡大した本稿では、次の点について結論と残された課題として整理することができる。

- 対象製品の価格と数量のデータが入手可能な場合、商品の経済的重要性を考慮した重み付きヘドニック回帰を用いるのが最善である。重み付きヘドニック回帰と重み付けなしヘドニック回帰(正確には等しい重み付けを行ったヘドニック回帰)の間には大きな違いがあることがわかった。
- ヘドニック回帰の時間ダミー特性アプローチは、今回用いたデータに対してはうまく機能しなかった。この方法は、より詳細な特性に関するデータを必要とし、特性データが欠損している問題の影響を受ける。
- 隣接期間に対してヘドニック回帰を実行し、連鎖指数を作成したすべてのモデルには、チェーン・ドリフトの問題があった。このチェーン・ドリフトの問題は、Similarity Linked 指数を用いた方法でも解決されなかった。この理由としては、連結指数を作成する際に、2時点の連結指数を使用している事が考えられる。⁴⁰
- この例のチェーン・ドリフト問題に対する満足のいく解決策は、拡張ウィンドウの手法によってもたらされた。この方法は、競合製品間の代替性が高い多くの製品でうまく機能するはずである。

また、本稿で追加した分析では、次の点が明らかになった。

⁴⁰ Fox, Levell and O'Connell (2023)による最近の論文でも、数百の製品クラスについて同じ結果が得られている。

- 製品の生存時間の分布は右側に裾野の長い分布になっており、生存時間の平均値は 24.7 ヶ月であった。また、生存時間分析を行ったところ 50%生存時間は約 2 年であり、半数の製品は市場に登場してから 2 年以内に撤退している。
- 製品の新鮮さを表す変数をヘドニック回帰モデルに追加することで、ノートパソコンがファッション製品であるかどうかを調べた結果、ファッション製品であることを示すような結果は得られず、逆に、消費者は古い製品に対してプレミアムを支払う可能性が示唆された。また、NEW 変数の区分幅の取り方を変えることによっても、得られた結果の傾向が異なることから、正しく推計ができていない可能性がある。

さらに、マイクロデータを用いた計量経済学的接近法として、また価格指数・データサイエンスとして残された課題としては、ランダムに発生するマイクロデータの扱いである。本節では、a)推計期間を通じて価格(P)と数量(Q)が完全に取得可能なケース、から出発し、b)特定の季節などに取引が集中したり、データが欠損したりするケース、の問題への対応方法を示した。いわゆる一時的に欠損値が存在する場合での統計分析の取り合う使い方となる。さらに、技術進歩などにより商品が入れ替わり、c)特定の時期から価格(P)と数量(Q)が永遠に欠損してしまうケース、への価格指数の推計方法としてのヘドニック・モデルを学んだ。

さらに残されている問題としては、ランダムに取引が発生する場合でのデータ分析の方法である。例えば、住宅価格は、同じ性能の建物であったとしても立地が異なれば、価格水準が異なる。同じ立地の住宅が繰り返し取引されたとしても、時間の経過とともに建物が劣化したりするなど性能が変化してしまうために、ただ一つとして同じ財は存在しないということになる。さらに、住宅は空間的な特性をも持ち合わせている。このような問題に対応した計量経済学的接近法については、別稿として整理したい。

計量経済学的な接近法は、今後も進化を続けていくことは確実である。経済理論・モデルもまた追時的に改善が続けられていく。分析者は、そのような進化と併せて利用可能なデータの集計手法、データの発生プロセスにも注意を払いながら、データ分析をしていくことが求められるのである。また、ここで整理された残された課題は、今後の研究対象にもなりうるものと考えている。多くの方が、この研究領域に参加し、国際的にも競争力を持った新しい研究成果が公表されていくことを期待したい。

本稿が、マイクロ計量分析、ヘドニック分析に新規参入される若い研究者の一助になれば幸いである。

参考文献

- Allen, R.C., Diewert, W.E., 1981. Direct versus Implicit Superlative Index Number Formulae, *Review of Economics and Statistics*. 63, 430-435.
- Aten, B., Heston, A., 2009. Chaining Methods for International Real Product and Purchasing Power Comparisons: Issues and Alternatives. *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S.P. Rao (ed.), Cheltenham UK: Edward Elgar. 245-273.
- Balk, B.M., 1980. A Method for Constructing Price Indices for Seasonal Commodities. *Journal of the Royal Statistical Society. Series A* 143, 68-75.
- Court, A.T., 1939. Hedonic Price Indexes with Automotive Examples. *The Dynamics of Automobile Demand*, New York: General Motors Corporation. 99-117.
- de Haan, J., 2004a. The Time Dummy Index as a Special Case of the Imputation Törnqvist Index. paper presented at The Eighth Meeting of the International Working Group on Price Indices (the Ottawa Group), Helsinki, Finland.
- de Haan, J., 2004b. Estimating Quality-Adjusted Unit Value Indices: Evidence from Scanner Data. Paper presented at the Seventh EMG Workshop, Sydney, Australia, December 12-14.

- de Haan, J., 2010. Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Re-pricing Methods. *Jahrbücher für Nationökonomie und Statistik* 230, 772-791.
- de Haan, J., Krsinich, F., 2014. Scanner Data and the Treatment of Quality Change in Nonrevisable Price Indexes. *Journal of Business and Economic Statistics* 32, 341–358.
- de Haan, J., Krsinich, F., 2018. Time Dummy Hedonic and Quality-Adjusted Unit Value Indexes: Do They Really Differ?. *Review of Income and Wealth* 64:4, 757-776.
- Diewert, W.E., 1976. Exact and Superlative Index Numbers. *Journal of Econometrics* 4, 114-145.
- Diewert, W.E., 2002. Weighted Country Product Dummy Variable Regressions and Index Number Formulae. Department of Economics, Discussion Paper 02-15, University of British Columbia, Vancouver, B.C., Canada, V6T 1Z1.
- Diewert, W.E., 2003a. Hedonic Regressions: A Consumer Theory Approach. in *Scanner Data and Price Indexes, Studies in Income and Wealth (Vol. 61)*, eds. R.C. Feenstra and M.D. Shapiro, Chicago: University of Chicago Press. 317-348.
- Diewert, W.E., 2003b. Hedonic Regressions: A Review of Some Unresolved Issues. Paper presented at the Seventh Meeting of the Ottawa Group, Paris, 27–29 May.
- Diewert, W.E., 2004. On the Stochastic Approach to Linking the Regions in the ICP. Discussion Paper no. 04-16, Department of Economics, The University of British Columbia, Vancouver, Canada.
- Diewert, W.E., 2005. Weighted Country Product Dummy Variable Regressions and Index Number Formulae. *Review of Income and Wealth* 51, 561-570.
- Diewert, W.E., 2009. Similarity Indexes and Criteria for Spatial Linking. *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S.P. Rao (ed.), Cheltenham, UK: Edward Elgar. 183-216.
- Diewert, W.E., 2013. Methods of Aggregation above the Basic Heading Level within Regions. *Measuring the Real Size of the World Economy: The Framework, Methodology and Results of the International Comparison Program—ICP*, Washington D.C.: The World Bank. 121-167
- Diewert, W.E., Fox, K.J., 2022. “Substitution Bias in Multilateral Methods for CPI Construction,” *Journal of Business and Economic Statistics* 40:1, 355-369.
- Diewert, E., 2022, *Consumer Price Index Theory*, International Monetary Fund, Empirical Examples <https://www.imf.org/en/Data/Statistics/cpi-manual>
- Diewert, E and C. Shimizu 2024, “Product Churn and Quality Adjustment”. TCER Working Paper Series, E-197.
- Fisher, I., 1922. *The Making of Index Numbers*, Boston: Houghton-Mifflin.
- Fox, K.J., Levell, P., O’Connell, M., 2023. Inflation Measurement with High Frequency Data. Working Paper 23/29, Institute for Fiscal Studies, London, U.K. <https://ifs.org.uk/publications/inflation->
- Griliches, Z., 1971. Introduction: Hedonic Price Indexes Revisited. *Price Indexes and Quality Change*, Z. Griliches (ed.), Cambridge MA: Harvard University Press. 3-15.
- Hardy, G.H., Littlewood, J.E., Pólya, G., 1934. *Inequalities*, Cambridge: Cambridge University Press.
- Hill, R.J., 1997. A Taxonomy of Multilateral Methods for Making International Comparisons of Prices and Quantities. *Review of Income and Wealth* 43(1), 49-69.
- Hill, R.J., 1999a. Comparing Price Levels across Countries Using Minimum Spanning Trees. *The Review of Economics and Statistics* 81, 135-142.
- Hill, R.J., 1999b. International Comparisons using Spanning Trees. *International and Interarea Comparisons of Income, Output and Prices*. A. Heston and R.E. Lipsey (eds.), *Studies in Income and Wealth Volume 61*, NBER, Chicago: The University of Chicago Press. 109-120.
- Hill, R.J., 2001. Measuring Inflation and Growth Using Spanning Trees. *International Economic Review* 42, 167-185.

- Hill, R.J., 2004. Constructing Price Indexes Across Space and Time: The Case of the European Union. *American Economic Review* 94, 1379-1410.
- Hill, R.J., 2009. Comparing Per Capita Income Levels Across Countries Using Spanning Trees: Robustness, Prior Restrictions, Hybrids and Hierarchies. *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S.P. Rao (ed.), Cheltenham UK: Edward Elgar. 217-244.
- Hill, R.J., Rao, D.S.P., Shankar, S., Hajargasht, R., 2017. Spatial Chaining as a Way of Improving International Comparisons of Prices and Real Incomes. paper presented at the Meeting on the International Comparisons of Income, Prices and Production, Princeton University, May 25-26.
- Hill, R.J., Timmer, M.P., 2006. Standard Errors as Weights in Multilateral Price Indexes. *Journal of Business and Economic Statistics* 24:3, 366-377.
- Kravis, I.B., Heston, A., Summers, R., 1982. *World Product and Income: International Comparisons of Real Gross Product*, Statistical Office of the United Nations and the World Bank, Baltimore: The Johns Hopkins University Press.
- Muellbauer, J., 1974. Household Production Theory, Quality and the Hedonic Technique. *American Economic Review* 64:6, 977-994.
- Rao, D.S.P., 1995. On the Equivalence of the Generalized Country-Product-Dummy (CPD) Method and the Rao-System for Multilateral Comparisons. Working Paper No. 5, Centre for International Comparisons, University of Pennsylvania, Philadelphia.
- Rao, D.S.P., 2004. The Country-Product-Dummy Method: A Stochastic Approach to the Computation of Purchasing Power parities in the ICP”, paper presented at the SSHRC Conference on Index Numbers and Productivity Measurement, June 30-July 3, 2004, Vancouver, Canada.
- Rao, D.S. P., 2005. On the Equivalence of the Weighted Country Product Dummy (CPD) Method and the Rao System for Multilateral Price Comparisons. *Review of Income and Wealth* 51:4, 571-580.
- Rosen, S, 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy* 82, 34-55.
- Sergeev, S., 2001. Measures of the Similarity of the Country’s Price Structures and their Practical Application. Conference on the European Comparison Program, U. N. Statistical Commission. Economic Commission for Europe, Geneva, November 12-14, 2001.
- Sergeev, S., 2009. Aggregation Methods Based on Structural International Prices. *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S.P. Rao (ed.), Cheltenham UK: Edward Elgar. 274-297.
- Summers, R., 1973. International Comparisons with Incomplete Data. *Review of Income and Wealth* 29:1, 1-16.
- Triplet, J., 1987. Hedonic Functions and Hedonic Indexes. John Eatwell, Murray Milgate, and Peter Newman (eds.), *The New Palgrave: A Dictionary of Economics*, Volume 2. New York, NY: Stockton Press. 630-634.
- Triplet, J., 2004. *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*, Directorate for Science, Technology and Industry, DSTI/DOC(2004)9, Paris: OECD.
- 阿部修人, 2023「物価指数理論」日本評論社
- 清水千弘, 唐渡広志, 2007「不動産市場の計量経済分析」, 中央評論社

「付論 A.laptop データについて」として POS データの構造を説明するとともに、分析用データを構築する手続きを、「付論 B. 推計プログラム」とし Diewert and Shimizu (2024)で作成された推計プログラムを汎用のプログラム言語である R により再現された推計用のプログラムを、それぞれ整理した。データおよびプログラムが必要な場合は、下記をご覧ください。また、筆者まで連絡いただきたい。
<https://www.shmzlab.jp/Paper/240118MicroDataScience.pdf>