

スパース推定を用いた 新しいヘドニック法について

おう ゆうすけ かわかみ あつし はたやまゆうだい ふるた さほ こ
王 悠介 / 川上淳史 / 畑山優大 / 古田早穂子

要 旨

物価指数の品質調整におけるヘドニック法の適用に当たっては、説明変数間の多重共線性や欠落変数によるバイアスなど、実務上の課題が存在する。本稿では、こうした課題を克服するため、「スパース推定 (sparse estimation)」を用いた新しいヘドニック法を提案する。新手法は、多重共線性に対する頑健性を高める「グループ効果」、および変数選択の適正性と係数の漸近的な不偏性を同時に満たす「オラクル性」を担保することで、ヘドニック法の推計上の課題に対処している。新手法をわが国の乗用車の価格指数に適用した実証分析の結果、従来の標準的な推計手法と比べて、(1) 回帰式に採用される変数の大幅な増加、(2) フィットの改善、(3) 欠落変数バイアスによる品質向上率の過大評価の緩和といった点で改善がみられた。こうした結果は、品質調整におけるヘドニック法の利便性を高めるとともに、物価指数の精度向上に寄与するものと考えられる。

キーワード： 物価指数、品質調整、ヘドニック法、多重共線性、欠落変数、スパース推定、アダプティブ・エラスティック・ネット

.....
本稿の作成過程では、舟岡史雄氏（信州大学）、北村行伸氏（立正大学）、阿部修人氏（一橋大学）、清水千弘氏（日本大学）、白塚重典氏（慶應義塾大学）から有益なご助言をいただいた。また、日本銀行の神山一成氏、肥後雅博氏、武藤一郎氏、桜健一氏、宇野洋輔氏、中島上智氏、倉知善行氏、増島綾子氏からも有益なコメントを頂いた。石黒雄人氏、久保太基氏、江口万里奈氏、新聞洗太郎氏からは、計数作成においてご協力を頂いた。記して感謝の意を表したい。ただし、本稿で示されている意見は、筆者たち個人に属し、日本銀行の公式見解を示すものではない。また、ありうべき誤りはすべて筆者たち個人に属する。

王 悠介 日本銀行調査統計局企画役補佐（現金融機構局企画役補佐、
E-mail: yuusuke.ou@boj.or.jp）
川上淳史 日本銀行調査統計局企画役（現国際局企画役、
E-mail: atsushi.kawakami@boj.or.jp）
畑山優大 日本銀行調査統計局（現金融機構局、
E-mail: yuudai.hatayama@boj.or.jp）
古田早穂子 日本銀行調査統計局（E-mail: sahoko.furuta@boj.or.jp）

1. はじめに

ヘドニック法は、物価指数の作成における品質調整方法の1つである¹。商品価格の時系列的な推移を表す物価統計において、市場における商品の入れ替わりに対応して、新旧商品の品質差を調整することが不可欠であり、ヘドニック法では、商品の品質がその商品の持つ個々の機能・性能に分解可能であるとの仮定に基づき、各機能・性能と価格の関係を統計的な回帰分析を通じて推計することで、表面価格の差を品質変化と純粋な価格変動に分解する。ヘドニック法は、品質評価を分析者の主観的な判断ではなく、データと統計的手法により客観的に行える点や、機能・性能の変化が多岐にわたる場合にも、それら複数の品質変化が価格に与える影響を総合的に評価できる点などにメリットがあり、各国における消費者物価指数や生産者物価指数の作成に適用されている。

もっとも、ヘドニック法を実際に適用する際に、直面する課題も少なくない²。まず、同手法では、商品の機能・性能に関する客観的なデータを用いて回帰分析を行う必要があるが、商品間での機能・性能の相関が強い場合には、回帰式における説明変数間の多重共線性の問題が生じやすくなり、パラメータが安定的でなくなる可能性がある。また、商品のすべての機能・性能をデータとして入手することが困難な場合には、欠落した変数の影響で、回帰式に含まれる変数のパラメータにバイアスが生じてしまう問題がある。さらに、ヘドニック法では、商品の機能・性能と価格の関係が必ずしも線形ではないと考え、非線形の回帰式を推計することが多いが、多重共線性や欠落変数の問題は、関数形が複雑なものになるほど深刻になることも知られている³。

こうしたヘドニック法の課題については以前から指摘されていたものの、近年に至るまで、必ずしも実務上満足できる解決法が与えられてこなかったのが実情である。そこで、本稿では、これらの問題を推計方法の改善により克服する試みを行う。具体的には、ヘドニック法の推計手法として、「スパース推定 (sparse estimation)」を導入することによって、変数間の多重共線性と欠落変数の問題に対処する手法を提案する。スパース推定は、統計学で従来から改良が加えられてきており、近年では、機械学習などの分野で用いられることが多い手法である。「まばら」の意を表すスパースの語のとおり、スパース推定は、多数の説明変数の中で、幾つかの意味のある変数だけを選択し、それ以外の変数にかかるパラメータを丁度ゼロとなるように推計する性質を持つ。この性質を利用することにより、従来のヘドニック法で

1 ヘドニック法に関する代表的な文献としては、例えば白塚 [1998] を参照。

2 ヘドニック法が直面する実務的な課題についての参考文献としては、例えば Triplett [2006] を参照。

3 この点については、Cropper, Deck, and McConnell [1988] を参照。

用いられてきた推計手法（例えば最小二乗法〈ordinary least squares: OLS〉）との対比では、変数選択を自動的に行わせることができる点に強みがある。特に、スパース推定の中でも、本稿で提案する新しい推計手法で用いる「アダプティブ・エラスティック・ネット（adaptive elastic net: AEN）」は、「グループ効果」と呼ばれる多重共線性に対する頑健性を高める性質に加え、「オラクル性」と呼ばれる変数選択の適正性と係数の漸近的な不偏性を同時に担保する性質を持つ点で非常に優れており（Zou and Zhang [2009]）、本稿では、AEN のこうした特性がヘドニック法の抱える課題を解消するうえで役立つことを示す。なお、AEN を用いた実証分析は、近年、さまざまな分野で存在するが、AEN をヘドニック法に適用した先行研究に関しては、筆者らの知る限り存在しない。

分析結果を先取りすると、以下のとおりである。本稿において、AEN による新手法を乗用車価格に適用する実証分析を行った結果、日本銀行調査統計局が作成するわが国の企業物価指数のヘドニック法において用いられてきた従来の推計手法と比べて、複数の点で改善がみられた。第1に、回帰式に採用される変数が大幅に増加し、品質調整を適用できる機能・性能の種類が拡大した。第2に、推計に用いたサンプル期間中のフィットが改善したほか、サンプル期間後についてもフィットの改善がみられた。また、推計期間の変更に対して従来手法よりもパラメータが安定的であることも確認された。第3に、推計方法の違いが価格指数に与える影響を確認したところ、新手法によって推計された価格指数の下落ペースは、従来手法と比べて緩やかとなった。これは、従来の手法では、欠落変数バイアスにより品質向上率を過大評価してしまう可能性がある一方、新手法ではその問題を解消しうることを示唆している。こうした結果は、品質調整におけるヘドニック法の利便性を高めるとともに、物価指数の精度の向上に寄与するものと考えられる。

本稿の構成は以下のとおりである。2節では、従来用いられてきたヘドニック法の概要とその課題について説明する。3節では、スパース推定を用いた新しいヘドニック法とその特性について説明する。4節では、新手法を乗用車の価格に適用した実証分析の結果を示す。5節では本稿のまとめを行う。

2. 従来のヘドニック法とその課題

(1) 従来のヘドニック法

ここでは、日本銀行調査統計局が作成する企業物価指数を例にとって、ヘドニック法で従来用いられてきた手法の概要を説明する。ヘドニック法では、商品の価格

を被説明変数、その商品の機能や性能を表すデータを説明変数とする回帰分析を行い、そこで得られたパラメータを用いて品質調整を行う。回帰分析を行うに当たっては、特定の関数形を想定する必要があるが、経済理論上、ヘドニック法の関数形には事前的な制約は存在しないことが知られている⁴。こうしたもとは、推計に用いることのできる関数形は無数に存在することになるため、実際の推計では、フィットのよさや推計パラメータの整合性（有意性や符号条件）などの観点から、純粋にパフォーマンスが高い関数形を選択する必要がある⁵。もっとも、商品の価格と機能・性能の関係は必ずしも線形とは限らないため、非線形性を考慮することも必要と考えられる。こうした観点から、先行研究では、非線形性を考慮するために、推計に用いる変数を以下のように変換する「Box-Cox 変換」と呼ばれる処理を施すことが提唱されている⁶。

【Box-Cox 変換】

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda} & (\lambda \neq 0), \\ \log x & (\lambda = 0). \end{cases} \quad (1)$$

(1) 式の λ は Box-Cox パラメータと呼ばれ、関数の非線形性の度合いを決める係数である。Box-Cox 変換を施した、従来のヘドニック法の回帰式は、以下のように表される。

【従来のヘドニック法の回帰式】

$$y_i^{(\lambda_0)} = \beta_0 + \sum_{j=1}^{p_c} \beta_{c_j} x_{c_j,i}^{(\lambda_j)} + \sum_{k=1}^{p_d} \beta_{dk} x_{dk,i} \quad (2)$$

y_i : 理論価格、 $x_{c_j,i}$: 連続変数、 $x_{dk,i}$: ダミー変数、

β_0 : 定数項、 β_{c_j} : 連続変数の係数、 β_{dk} : ダミー変数の係数、

λ_0 : 理論価格の Box-Cox パラメータ、 λ_j : 連続変数の Box-Cox パラメータ、

p_c : 連続変数の数、 p_d : ダミー変数の数

.....
4 ヘドニック法で用いる関数は、理論的には、すべての特性を連続的に選択可能な完全競争市場における、消費者の効用最大化を通じたある特性に対する買値関数 (bid function) や生産者の利潤最大化から導出される売値関数 (offer function) との包絡線として導かれ、関数形に関する事前的な制約は存在しない。詳細は、Rosen [1974] や白塚 [1998] の 5 章を参照。

5 白塚 [1997] は、ヘドニック法における関数選択の判断基準として、フィットのよさや推計パラメータの整合性のほか、解釈可能性や推計負担などにも注目すべきとしている。

6 Box-Cox 変換の詳細については、Box and Cox [1964] を参照。また、Halvorsen and Pollakowski [1981] は、ヘドニック法における一般的な関数形として Box-Cox 形を活用することや、尤度比検定を利用することを提唱している。

(2)式において、①被説明変数・説明変数の双方を対数線形とする場合 ($\lambda_0 = \lambda_j = 0$) を「両側対数形」、②被説明変数のみを対数線形とする場合 ($\lambda_0 = 0, \lambda_j = 1$) を「片側対数形」、③いずれの変数も線形とする場合 ($\lambda_0 = \lambda_j = 1$) を「線形」とし、上記に当てはまらないもののうち、④ $\lambda_j = 1$ の場合を被説明変数のみが Box-Cox 変換される「片側 Box-Cox 形」、それ以外の⑤被説明変数と説明変数（連続変数）のすべてが Box-Cox 変換されるものを「両側 Box-Cox 形」として 5 つの回帰式を推計し、検定を行ったうえで関数形を選択することが望ましい。なお、検定を行うと、実際には両側 Box-Cox 形が選択されるケースが多いことが知られている⁷。

(2) 従来手法の課題①：多重共線性

従来のヘドニック法が直面しうる課題の 1 つとして、多重共線性（multicollinearity）の問題が挙げられる。多重共線性とは、回帰式に含まれる説明変数間の相関が強い状態を指す。多重共線性がある場合、変数間での識別が困難となるため、パラメータを正確に推定することが難しくなる。その結果、本来重要な影響を果たしているはずの変数のパラメータが有意でなくなるといった事象が現れる。

ヘドニック法では、多重共線性の問題が発生しやすいことが知られている。乗用車の例で示すと、例えば車体の全長と重量は相関が強く、多重共線性の問題が生じる（図表 1）。ヘドニック法の適用品目のデータセットでは、全長と重量の例のように物理的な関係から必然的に発生するものに限らず、多くの変数間で相関関係がみられる。これは、ハイエンド、ミドルエンド、ローエンドといったように企業が価

図表 1 乗用車の連続変数間の相関係数

	乗車定員	全長	全幅	全高	重量	燃費	最大出力	最大トルク	リム径	ギア数
乗車定員	1.000									
全長	0.420	1.000								
全幅	0.223	0.849	1.000							
全高	0.787	0.317	0.251	1.000						
重量	0.521	0.885	0.844	0.553	1.000					
燃費	-0.252	-0.501	-0.618	-0.268	-0.563	1.000				
最大出力	0.017	0.628	0.696	0.018	0.665	-0.661	1.000			
最大トルク	0.008	0.599	0.711	-0.006	0.618	-0.535	0.812	1.000		
リム径	-0.051	0.625	0.785	0.004	0.584	-0.489	0.645	0.711	1.000	
ギア数	-0.058	0.198	0.152	-0.081	0.213	-0.080	0.409	0.319	0.159	1.000

.....
 7 Triplett [2006] は、統計的な検定を行うと、線形や対数形は Box-Cox 形に対し棄却されやすいことに言及している。実際、わが国の企業物価指数で、ヘドニック法による品質調整に用いている回帰式の多くは、両側 Box-Cox 形となっている。

格帯別に製品を揃える中で、ハイエンド製品にはさまざまな機能を同時に搭載し、ローエンド製品ではそれらの機能を必要最小限に絞り込むといった販売戦略がとられやすいためである。その結果、例えば最高出力とパワーシートの搭載有無など、物理的な関係性が必ずしも強くない変数間でも相関関係が発生しやすくなる⁸。

多重共線性を回避する手法としては、大きく2つのアプローチが考えられる。まず1つ目は、事前に主成分分析を行い、得られた主成分の一部を説明変数としてヘドニック関数を回帰する手法（主成分分析法）である。ただし、白塚 [1995] では、乗用車にヘドニック法を適用する際の回帰式に主成分を説明変数として加えた推計を行い、①回帰式の決定係数の改善は限界的であること、②主成分に対する個別機能の寄与が時間とともに変動するため、主成分の推計パラメータは解釈が困難であることを指摘したうえで、主成分分析は、重要な機能に見当をつける目的では有用なもの、得られた主成分を用いて多重共線性に対処するには適切な手法とは言い難い、と結論づけている。そのため、実際には、より簡便な第2の手法として、相関関係がある変数の片方を推計式から除外するアプローチ（ステップワイズ法）が広く用いられている。すなわち、推計結果から多重共線性の悪影響が疑われる場合には、原因と思われる変数を除外して推計をやり直すことで、多重共線性は一定程度回避することができる。もっとも、上述した乗用車の例のように、機能間の相関が強いもとは、取り除く変数を適切に選択することは必ずしも容易ではなく、妥当な結果が得られるまで推計を繰り返すことの負担は小さくない。

(3) 従来手法の課題②：欠落変数によるバイアス

従来のヘドニック法が直面しうる課題の2つ目としては、欠落変数の存在によるバイアスが挙げられる。欠落変数とは、本来は被説明変数に対する説明力があるにもかかわらず、回帰式に含まれていない変数のことである。物価統計におけるヘドニック法の文脈では、本来は商品の価格に影響を与えているはずの機能・性能の情報が、回帰式に含まれていない場合がこれに当たる。

ヘドニック法で欠落変数が生じる状況には、①データセット構築の段階で発生するケースと、②変数選択の結果として発生するケースの2種類が存在する。①データセット構築の段階で発生するケースは、価格への影響を持つ機能であるにもかかわらず、それを観測できないことから生じる。例えば、製品のデザインやスタイル、ブランド価値といった定量化が難しい要素を回帰式に取り込むことは本質的に難しく、メーカーを識別するダミー変数を代理変数として用いるなど部分的な対応

8 Triplet [2006] では、全長と重量のような物理的な関係性などにより発生する多重共線性を「普遍的な多重共線性」と呼び、製品のグレードなどに応じて機能・性能の相関が高くなることにより生じる多重共線性を「サンプル上の多重共線性」として分類している。

しかできない。そのほか、技術革新により新しい特性が現れた場合、それを回帰式に取り込むには、その特性を有する製品が一定程度市場に浸透するまで待つ必要がある。②変数選択の結果として発生するケースは、回帰式で取り込める変数の数に限りがある中で、価格説明力の低い変数を誤って取り込んだ結果、真に価格説明力を持つ変数が除外されてしまうことから生じる。

欠落変数がある場合、回帰式に取り込んでいる変数（採用変数）のパラメータに歪みが生じる。その歪みによって物価指数にバイアスが生じる場合、そのバイアスが上下どちらの方向となるかは、欠落変数と採用変数の相対的な性能向上率の差によって定まる。例えば、欠落変数が採用変数より大幅に性能向上している場合には、品質向上が過小に評価される結果、物価指数には上方バイアスが生じる。逆に、欠落変数の性能向上が小幅にとどまる中で、パラメータに歪みを持つ採用変数が大きく性能向上した場合、品質向上を過大評価する結果として、物価指数に下方バイアスが生じる。Triplett [2006] は、パソコンの価格にヘドニック法を適用し、欠落変数が存在する場合、5ヵ月の期間で -0.2% ～ -1.0% 程度の下方バイアスが物価指数に生じるとの実証結果を示しており、その背景として、計算速度やメモリ容量といった採用変数が、欠落変数よりも速く性能向上していた可能性を指摘している。また、Sawyer and So [2018] では、ヘドニック法で算出される半導体の物価下落率が、説明変数の組合せによってどの程度変化するか分析を行っている。1種類の変数のみ採用した場合の物価下落率（4年平均）は最大 -45.11% と、全機能を採用した場合の -8.77% よりも、欠落変数バイアスにより大きくなることを示している。これらの先行研究は、ヘドニック法で欠落変数がある場合、物価指数に下方バイアス（品質向上率の過大評価）が生じる可能性を示唆している。

2節で述べたとおり、特にヘドニック法の推計では、多重共線性の問題が発生しやすい。このため、モデルに取り込む変数の選択に当たって、価格説明力の高い変数を適切に選択することは、必ずしも容易ではない。欠落変数が生じないように多くの変数をモデルに取り込んでいくと、多重共線性の問題が強く顕現し、安定的な推計パラメータを得にくくなる。他方で、使用する変数を少数に絞り込むと、上述した欠落変数の存在によるバイアスが深刻化する。この点、白塚 [1995] は、乗用車のヘドニック推計において多重共線性を除去することの難しさを指摘したうえで、変数選択に当たっては、さまざまな説明変数の組合せをテストしながらフィットが良好かつ安定的な推計結果が得られるものを可能な限り探索することで、多重共線性と欠落変数によるバイアスの問題にバランスよく対処していくことの重要性を指摘している。

なお、欠落変数によるバイアスは、関数形が複雑になるほど深刻になることが知られている。Cropper, Deck, and McConnell [1988] は、不動産価格に関するヘドニック法を通じて、欠落変数が存在する可能性が無視できない場合、単純な関数形を用

いる方が望ましいとの見解を示している。特に Box-Cox 項を含む関数形についてみると、説明変数の組合せ次第では Box-Cox パラメータが極端な値をとるリスクがある⁹。歪んだ関数形の推計式を採用すると、推計に用いたデータセットに対するフィットはよくても、その後新たに登場する製品に対するフィットが悪化する問題（これを「オーバーフィッティング」と呼ぶ）を引き起こすリスクが小さくない。そのため、Box-Cox パラメータが過度に高次の次数となることがないように、変数構成を変更しながら推計をやり直す必要がある。この点、わが国の企業物価指数で乗用車（ミニバン）の品質調整に用いられているヘドニック法の回帰式をみても、Box-Cox パラメータが 3.4 からほぼゼロ（対数形）になるなど、再推計の前後で結果が大きく変化する事例が時折みられている（図表 2）。こうした変化は、欠落変数の存在によるパラメータの不安定性を示唆している面もあると考えられる。

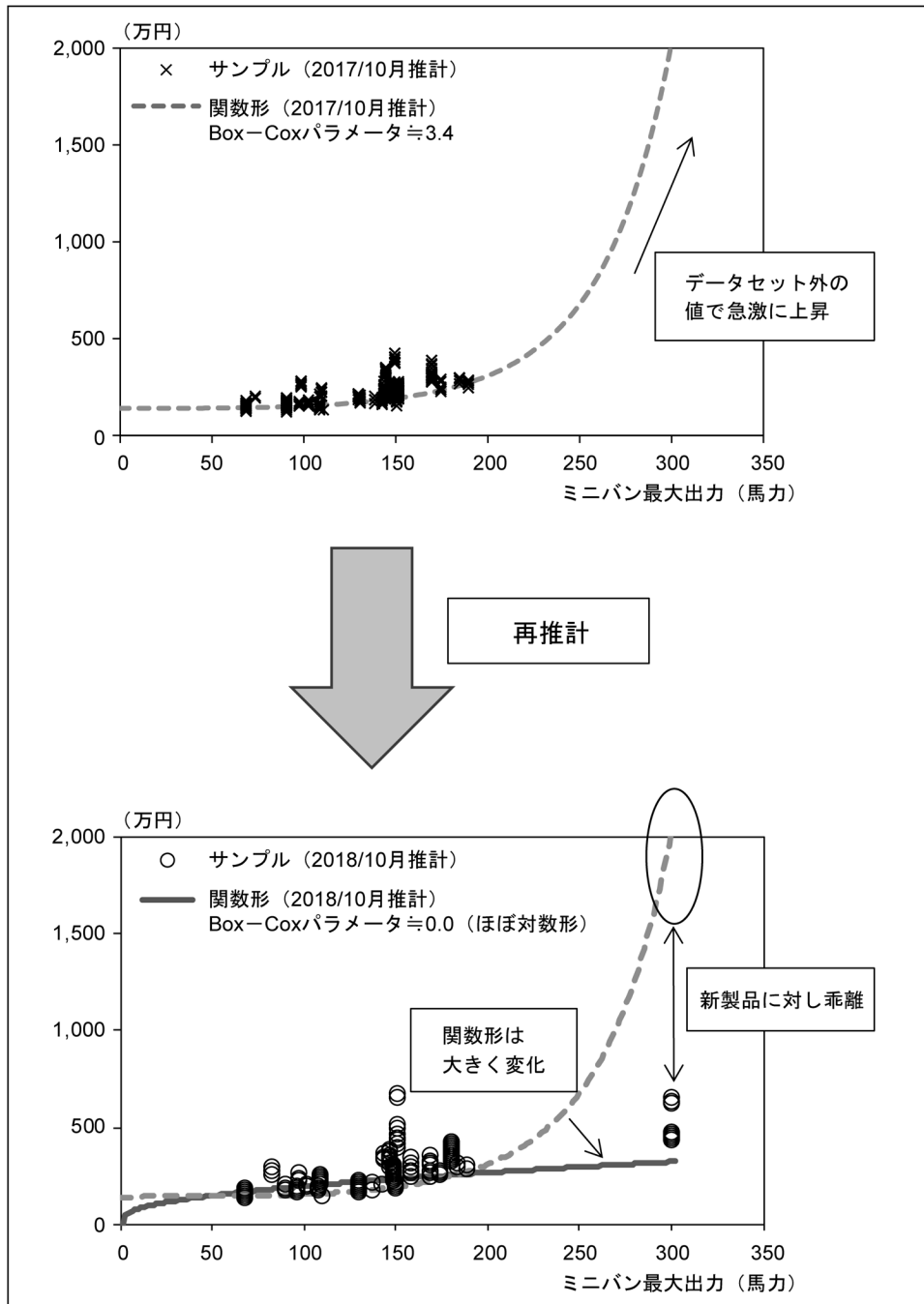
(4) 従来手法の課題③：交互作用の存在

ヘドニック法の適用に当たり直面する上記以外の課題としては、いわゆる「交互作用」の取込みに関する問題が挙げられる。ヘドニック法では、多くの場合、データセットに含まれる製品の間で、機能・性能にかかるパラメータが同一であると仮定して推計を行うが、実際には、ある機能が向上することにより別の機能の向上が価格に及ぼす影響力が強くなるといった交互作用が存在する場合や、同一の製品とみなして推計を実施しているが、実際には更に細かいカテゴリが存在する場合など、パラメータが全製品間で一定という仮定が成り立たない状況は一般に存在する。

こうした交互作用に対しては、変数の交差項を導入して推計を行うことが有用である。これにより、各機能が商品価格に与える影響が、別の機能の状態に依存する状況を捉えることが可能になる。もっとも、ヘドニック法を実際に適用する際、組合せが膨大に存在する交差項のすべてを導入することは、現実的には難しい。交差項を多く取り込めば、説明変数間での相関が高くなる結果、多重共線性の問題が生じやすくなり、パラメータが不安定化しやすくなってしまう。そのため、従来のヘドニック法では、交差項の取込みが限定的なものにとどまっていたというのが実情である。もっとも、上述の欠落変数のバイアスの問題を踏まえると、本来必要であるはずの交差項を回帰式に含めずに推計した場合には、回帰式に含めている変数や交差項のパラメータに歪みが生じることも考えられる。このように、交互作用の存在は、上述した多重共線性と欠落変数によるバイアスの双方の問題に直面するた

.....
⁹ Graves *et al.* [1988] も、複数の変数構成、複数の関数形の組合せで不動産価格に関するヘドニック法の回帰式を推計したうえで、Box-Cox 項を含む関数形の推計式が、変数構成の変化に対し極めて不安定な挙動を示すことに言及している。

図表 2 ミニバン最大出力の関数形の変化



め、これまでのヘドニック法では十分な解決が図られてこなかったものと考えられる。

3. スパース推定を用いた新しいヘドニック法

(1) スパース推定

本節では、「スパース推定」を用いたヘドニック法について解説する。スパース推定は、多数の説明変数の中で、幾つかの意味のある変数だけを選択し、それ以外の変数にかかるパラメータを丁度ゼロとなるように推計する性質（これを「スパース性」という）を持つ。この性質を利用することにより、従来のヘドニック法で用いられてきた OLS などの手法との対比では、変数選択と係数の推定を同時に行い、安定的かつフィットのよいモデルを自動的に構築することができる点に強みがある。スパース推定は、これまで経済分析に限らず、さまざまな分野の実証分析で利用されているが、本節では、この手法の一種が、物価統計のヘドニック法で直面する課題（多重共線性や欠落変数によるバイアス）を克服するうえで有効であることを説明する¹⁰。

スパース推定は、Tibshirani [1996] によって提案された「ラッソ (least absolute shrinkage and selection operator: Lasso)」を皮切りに、これまで多くの手法が提案されている。本稿で提案する新しいヘドニック法では、多重共線性に対する頑健性が得られる「グループ効果」と、変数選択と係数の適正性を保証する「オラクル性」という望ましい2つの性質を持つ「アダプティブ・エラスティック・ネット (adaptive elastic net: AEN)」を用いる¹¹。なお、AEN をヘドニック法に適用した研究は、これまでのところ、筆者らの知る限り存在しない¹²。以下では、順を追ってスパース推定の概要と、上記2つの性質について解説する。

10 スパース推定は、例えば、国際プロジェクト「イベント・ホライズン・テレスコープ」による世界初のブラックホールの撮影においても、観測データの解析で大きな役割を果たした (The Event Horizon Telescope Collaboration [2019])。また、地理情報科学の分野でも、経済・社会変数間における地域間の異質性などを定量化する中でスパース推定の活用がみられており、例えば Jin and Lee [2020] が、空間多変量自己回帰モデル (spatial vector auto regression) の推計においてスパース推定を用いて住宅価格の分析を行っているほか、Wheeler [2009] は、同分野で用いられる地理的加重回帰モデル (geographically weighted regression) においてスパース推定の採用を提案している。

11 推計手法および各種性質の詳細は、Zou and Zhang [2009] を参照。

12 スパース推定のうち Lasso をヘドニック法に適用した研究は、幾つか存在する。例えば、Zafar and Himpens [2019] は、パソコンの価格と機能に関するウェブ・スクレイピング・データに対して Lasso を適用し、非線形を考慮した他の推計手法などと比較した推計パフォーマンスの評価を行っている。

まず、代表的なスパース推定である Lasso を用いて、スパース性がどのように得られるか概説する。Lasso は、回帰式の誤差二乗和に正則化項として β の L_1 ノルム（絶対値の総和）を加えた関数を最小化する β を推定する手法である¹³。

【Lasso】

$$\hat{\beta}(Lasso) = \underset{\beta}{\operatorname{argmin}} \left(|Y - X\beta|^2 + \lambda \sum_{j=1}^p |\beta_j| \right), \quad (3)$$

$\lambda > 0$: 正則化パラメータ（ λ が大きい場合、より少ない変数を選択）

多重共線性の問題に対処しうる伝統的な手法としては、リッジ回帰（Ridge regression）が挙げられる¹⁴。回帰式の誤差二乗和に正則化項を加えた関数を最小化するという点では等しいが、リッジ回帰は正則化項として β の L_2 ノルム（二乗の総和）を用いる。その帰結として、Lasso はスパース性を有するが、リッジ回帰はスパース性を持たないという差異が生まれる。

【リッジ回帰】

$$\hat{\beta}(Ridge) = \underset{\beta}{\operatorname{argmin}} \left(|Y - X\beta|^2 + \lambda \sum_{j=1}^p \beta_j^2 \right), \quad (4)$$

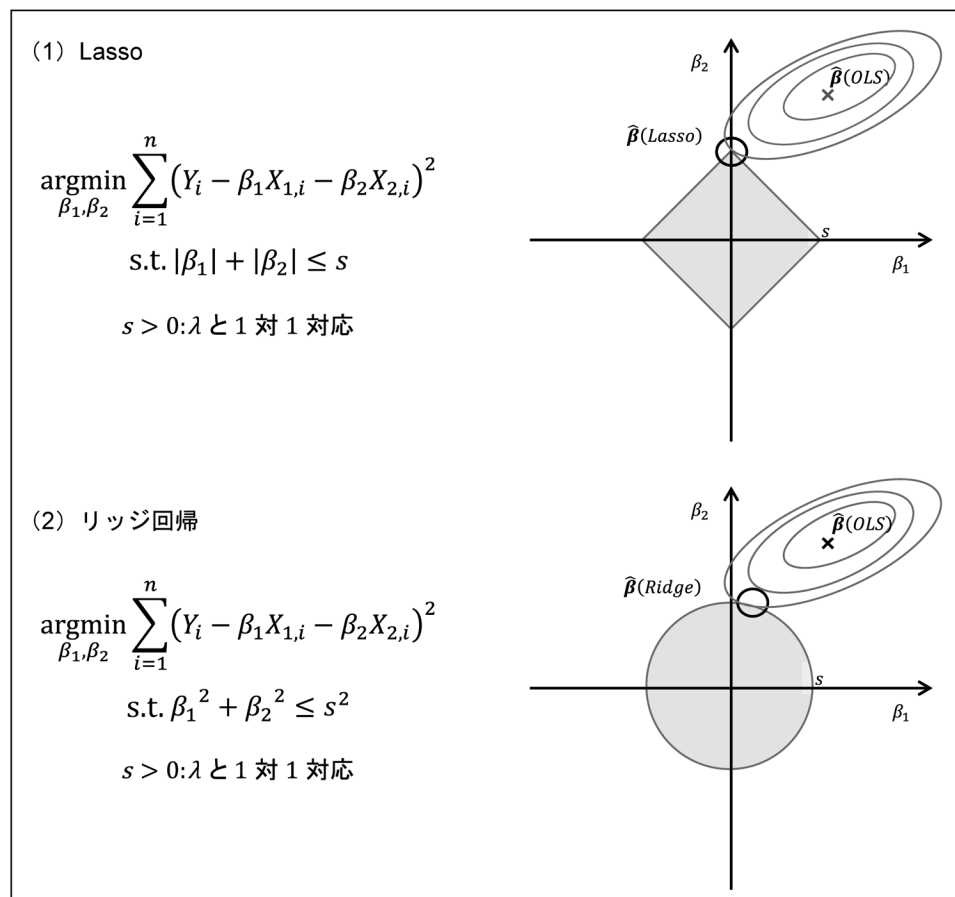
$\lambda > 0$: 正則化パラメータ（ λ が大きい場合、より小さい係数を推定）

正則化項の違いがスパース性の有無につながるメカニズムについて、Tibshirani [1996] の議論に沿って、図表 3 を用いて直観的に示す。変数が 2 種類の場合、Karush–Kuhn–Tucker 条件より、 $\hat{\beta}(Lasso)$ と $\hat{\beta}(Ridge)$ は図表 3 に記載の数式に変換することができる。 β_1 軸と β_2 軸から得られる平面を考えると、誤差二乗和は $\hat{\beta}(OLS)$ を中心とした楕円であり、正則化項と 1 対 1 対応する制約条件は、Lasso の場合は菱形、リッジ回帰の場合は円となる。こうした条件のもと、 β は誤差二乗和（楕円）と、制約条件（菱形か円）の接点から得られる。ここで、Lasso の場合は、制約条件が菱形であることから、両者が角で接しやすい、すなわち、 β の可動領域の中で端点解が選択されやすいことがわかる。このとき、片方のパラメータが丁度ゼロに推定され、変数選択が自動的に行われることとなる。その一方で、リッジ回帰は、制約条件が円であり、誤差二乗和と制約条件が特定の点で接しやすいわけではないことから、片方のパラメータが丁度ゼロに推定される可能性は低く、変数選択を自動的に行う性質を有していない。

13 スパース推定では、事前に被説明変数を中心化、説明変数を標準化したうえで計算を行う。すなわち、サンプル数 n に対して、 $\frac{1}{n} \sum_{i=1}^n y_i = 0$, $\frac{1}{n} \sum_{i=1}^n x_{ji} = 0$, $\frac{1}{n} \sum_{i=1}^n x_{ji}^2 = 1$ とする。

14 詳細については、Hoerl and Kennard [1970] を参照。

図表3 スパース推定のイメージ図



(2) グループ効果

他方、Lasso については、多重共線性の強いデータでは変数選択の結果が不安定になることが知られている。例えば、ある2つの変数にかかるパラメータの真の値が β_1^* 、 β_2^* であるとする。このとき、極端な例として、この2つの変数のデータセット上の値が全く同一であるとする、Lasso による最適化の解は、以下のとおり無数に存在し、一意に定まらない。

$$\hat{\beta}(Lasso) = \begin{pmatrix} s(\beta_1^* + \beta_2^*) \\ (1-s)(\beta_1^* + \beta_2^*) \end{pmatrix} \quad \text{for any } s \in [0, 1]. \quad (5)$$

これと同様に、相関が強い 2 変数が存在するとき、Lasso の変数選択はデータセットのわずかな変化に強く影響され、どちらの変数がモデルに取り込まれるかが安定しない。

ヘドニック法で用いるデータは相関が強いことが多いため、上記のような多重共線性の問題に対して頑健な性質を持つスパース推定を用いる必要がある。こうした性質を持つ代表的なスパース推定の 1 つが「エラスティック・ネット (elastic net: EN)」である。EN は、回帰式の誤差二乗和に正則化項として β の L_2 ノルムと L_1 ノルムの双方を加えた関数を最小化する β を推定する手法である¹⁵。その結果、変数選択できる Lasso の長所と、多重共線性に強いリッジ回帰の長所を併せ持った推計手法となっている。

【エラスティック・ネット (EN)】

$$\hat{\beta}(EN) = \left(1 + \frac{\lambda_2}{n}\right) \left\{ \operatorname{argmin}_{\beta} \left[|Y - X\beta|^2 + \lambda_2 \sum_{j=1}^p \beta_j^2 + \lambda_1 \sum_{j=1}^p |\beta_j| \right] \right\}, \quad (6)$$

$\lambda_2 > 0$: L_2 ノルムの正則化パラメータ、 $\lambda_1 > 0$: L_1 ノルムの正則化パラメータ、
 n : サンプル数

EN が有する多重共線性に対する頑健性は「グループ効果」と呼ばれる。グループ効果とは、説明変数間の相関が強い場合に、それらの変数にかかる係数の差が小さくなるような推計結果を与える性質である¹⁶。極端な例として、2 つの説明変数のデータセット上の値が全く同じである場合、EN は、その 2 つの変数にかかるパラメータを、以下のとおり、全く等しい値として推定する。そのため、多重共線性が強く、どの変数が真に説明力を有しているのかデータから識別するのが困難な状況でも、安定的な変数選択、パラメータの推定が可能となる。

$$\hat{\beta}(EN) = \begin{pmatrix} \frac{1}{2}(\beta_1^* + \beta_2^*) \\ \frac{1}{2}(\beta_1^* + \beta_2^*) \end{pmatrix}. \quad (7)$$

(3) オラクル性

スパース推定により得られる推定量が満たすべきもう 1 つの性質に「オラクル

.....
¹⁵ 詳細については、Zou and Hastie [2005] を参照。

¹⁶ より詳細に述べると、2 つの説明変数間の相関係数が $\rho > 0$ のとき、それらの説明変数 (標準化済み) のパラメータの差の絶対値の上限が $\sqrt{1-\rho}$ に比例する値になる制約条件を有している。

性」がある。具体的には、真の係数を β^* としたとき、スパース推定による推定量 $\hat{\beta}$ が以下の2つの条件を満たすことを、オラクル性を持つという。

【オラクル性】

①変数選択の一致性

$$\beta_j^* = 0 \text{ の場合} : \lim_{n \rightarrow \infty} P(\hat{\beta}_j = 0) = 1.$$

②非ゼロ係数の漸近正規性

$$\beta_j^* \neq 0 \text{ の場合} : \lim_{n \rightarrow \infty} \frac{(\hat{\beta}_j - \beta_j^*)}{\sigma(\hat{\beta}_j)} \sim N(0, 1),$$

$\sigma^2(\hat{\beta}_j)$: 推定量の漸近分散

上記2条件のうち、①の「変数選択の一致性」は、真の係数がゼロである変数について、その係数の推定量が一致性を満たすということである。②の「非ゼロ係数の漸近正規性」は、真の係数がゼロでない変数について、その係数に関する推定誤差が漸近的に正規分布に従うということである。

オラクル性は、スパース推定が同時に行う「変数選択」と「係数の推定」の両者の適正性を漸的に保証する重要な性質である。しかし、スパース推定の中でも、上述した Lasso や EN は、データセット次第では、どれほど適切に正則化パラメータを選択してもオラクル性が満たされることが知られている。そのため、本稿では、ヘドニック法の新手法として、スパース推定の中でもオラクル性を満たす、以下の「アダプティブ・エラスティック・ネット (AEN)」を用いる。

【アダプティブ・エラスティック・ネット (AEN)】

$$\hat{\beta}(AEN) = \left(1 + \frac{\lambda_2}{n}\right) \left\{ \underset{\beta}{\operatorname{argmin}} \left[|Y - X\beta|^2 + \lambda_2 \sum_{j=1}^p \beta_j^2 + \lambda_1^* \sum_{j=1}^p \hat{w}_j |\beta_j| \right] \right\}, \quad (8)$$

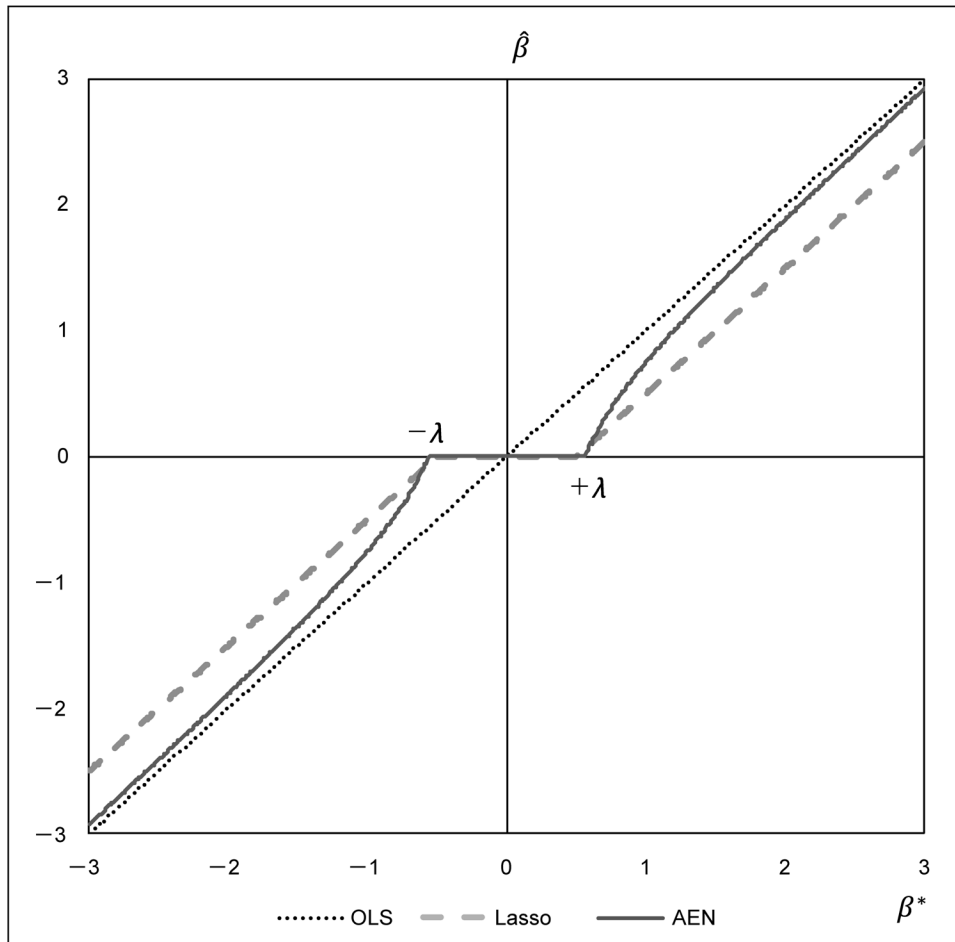
where $\hat{w}_j = (|\hat{\beta}_j(EN)|)^{-\gamma}$,

$\lambda_1^* > 0$: L_1 ノルムの正則化パラメータ (2段階目)、 $\hat{w}_j > 0$: 適応ウエイト、
 $\gamma > 0$: 適応パラメータ (大きいと係数の絶対値に応じて強い罰則を与える)

AEN の推計は、2段階で行われる。まず予備推計として EN で係数を推定する。そのうえで、係数の絶対値が小さい変数に対してより強い罰則を与えるよう L_1 ノルムの正則化項を変数ごとに調整したうえで、改めて EN を実施する¹⁷。このよう

.....
 17 1段階目の EN で係数がゼロと推定された変数は、2段階目の EN では用いない。

図表 4 AEN の持つ統計的特性



備考：1. X を 120,200 行 601 列の計画行列、 $\beta_i^* = -3 + 0.01i$ ($i = 0 \sim 600$)、 $Y = X\beta^* + \varepsilon$ ($\varepsilon \sim N(0, I)$) とした人工的なデータに対して各推定法から推定された $\hat{\beta}$ をプロット。なお、 X についてはすべての列で平均 0、標準偏差 1 とした。

2. Lasso について、 $\lambda = 0.5$ 、AEN について、 $\lambda_1 = \lambda_1^* = 0.2$ 、 $\lambda_2 = 0.001$ 、 $\gamma = 1$ としている。

に 2 段階推計を行うことで、データセットの性質にほとんど依存することなくオラクル性を得ることができるようになる。

図表 4 では、AEN がオラクル性を満たす理由について、Zou [2006] の議論を参考に、直観的に示している。ここでは、説明変数の行列 X と攪乱項のベクトル ε を人工的に生成し、真のモデル ($Y = X\beta^* + \varepsilon$) に基づき被説明変数のベクトル Y を算出した。そのうえで、 Y と X を観測値としたとき、OLS、Lasso、AEN が β をどのように推定するかを確認している。図中の横軸は真の係数 β^* を表しており、縦軸には

$\hat{\beta}(OLS)$ 、 $\hat{\beta}(Lasso)$ 、 $\hat{\beta}(AEN)$ をプロットしている。まず Lasso をみると、 $|\beta^*| < \lambda$ のとき、 $\hat{\beta} = 0$ となっており、スパース性を有していることがわかる。一方で、 $|\beta^*| \geq \lambda$ のとき、 $\hat{\beta}$ は真の値である β^* と比較して、絶対値でみて λ だけ小さく推定されていることがわかる。すなわち、正則化パラメータ λ とオラクル性の条件との関係は、 λ が大きくなるほど、①ゼロ係数を推定しやすくなり変数選択の一致性を満たしやすくなる一方、②推定値が絶対値でみて λ だけ小さくなってしまい非ゼロ係数の漸近正規性を満たしにくくなる、というトレードオフの関係となっていることがわかる。

これに対して AEN をみると、 $|\beta^*|$ が小さい場合、1 段階目の推計における係数の小ささを映じて罰則が強くと与えられることで $\hat{\beta} = 0$ が導かれる一方で、 $|\beta^*|$ が大きい場合には、罰則があまり与えられない結果として、 $\hat{\beta}$ が β^* に漸近していく関係となっている。このように、罰則の大きさを 1 段階目の推計値に応じて調整することで、係数が小さいときはゼロが推定されやすくなると同時に、係数が大きいときは、絶対値でみた推計値の縮小幅が最小限に抑えられ、オラクル性の 2 条件が同時に満たされやすくなっていることがわかる。

(4) 関数形の選択

本稿で提案する新しい手法では、ヘドニック関数を 2 次多項式として立式する。2 次多項式のどの項を回帰式に含めるかという点を AEN により判断することで、変数選択と関数形選択を同時に行う。そこではすべての交差項が勘案されるため、従来の手法とは異なって、交互作用を明示的に回帰式に取り込むことが可能となる¹⁸。なお、次数をあらかじめ 2 次までに制限しているのは、Box-Cox アプローチで時折みられていた、高次項によるオーバーフィッティングを未然に防ぐためである。

以上を踏まえると、ヘドニック法の回帰式は、新手法では以下のとおり推計されることとなる。

【AEN を用いたヘドニック法の回帰式】

$$Y_i \equiv \log y_i,$$

$$Y_i = \hat{\beta}_{00} + \sum_{j=1}^p \hat{\beta}_{0j} x_{j,i} + \sum_{j=1}^p \hat{\beta}_{jj} x_{j,i}^2 + \sum_{k>j \geq 1} \hat{\beta}_{jk} x_{j,i} x_{k,i}, \quad (9)$$

18 なお、全変数の組合せについて交差項を計算すると、ほとんどの場合において変数間に完全な多重共線性的関係が生じてしまう。このため、本稿では、他の変数の 1 次従属となっている幾つかの変数を事前に除去する処理を行っている。

$$\text{where } \hat{\beta} = \left(1 + \frac{\lambda_2}{n}\right) \left\{ \underset{\beta}{\operatorname{argmin}} \left(|Y - X\beta|^2 + \lambda_2 \sum_{k \geq j \geq 0} \beta_{jk}^2 + \lambda_1^* \sum_{k \geq j \geq 0} \hat{w}_{jk} |\beta_{jk}| \right) \right\},$$

$$\hat{w}_{jk} = (\hat{\beta}_{jk}^{1st})^{-\gamma},$$

$$\hat{\beta}^{1st} = \left(1 + \frac{\lambda_2}{n}\right) \left\{ \underset{\beta}{\operatorname{argmin}} \left(|Y - X\beta|^2 + \lambda_2 \sum_{k \geq j \geq 0} \beta_{jk}^2 + \lambda_1 \sum_{k \geq j \geq 0} |\beta_{jk}| \right) \right\},$$

y_i : 理論価格、 $x_{j,i}$: 説明変数、 $\hat{\beta}_{jk}$: $x_{j,i}x_{k,i}$ の係数、
 p : 候補となる説明変数の数、 n : データセットにおけるサンプル数、
 $\lambda_1 > 0$: L_1 ノルムの正則化パラメータ (1 段階目)、
 $\lambda_1^* > 0$: L_1 ノルムの正則化パラメータ (2 段階目)、
 $\lambda_2 > 0$: L_2 ノルムの正則化パラメータ、
 $\gamma > 0$: 適応パラメータ、 $\hat{w}_{jk} > 0$: 適応ウエイト

4. 新しいヘドニック法を用いた実証分析

(1) 推計に用いたデータ

本節では、前節で説明した、AEN を用いた新たなヘドニック法を、わが国の乗用車の価格に適用し、その特徴について考察する¹⁹。

分析に際しては、日本銀行調査統計局が作成するわが国の企業物価指数のヘドニック法におけるデータセットと同一のものを用いる。具体的には、株式会社プロトコーポレーション作成のデータベース「ゲーネット」に掲載された新車の希望小売価格に、株式会社内外出版社「月刊自家用車」に掲載された該当車種の平均値引率を掛け合わせることで価格データを作成した。対象期間は 2016 年第 3 四半期～2018 年第 2 四半期であり、サンプル数は 940 である。

機能・性能に関する指標についても基本的には「ゲーネット」に掲載されたデータを使用した。同データベースに記載されていない特性値のうち、重要であると考えられるものは、個別車種の仕様書などから取得した。具体的に用いた機能・性能の指標は図表 5 に示したとおりである。データセットには、乗用車の定量的

19 わが国の乗用車の価格指数にヘドニック法を適用した分析としては、前述した白塚 [1995] のほか、太田 [1978]、Ohta [1987] などが先駆的な研究である。太田 [1978] では、わが国の消費者物価指数 (CPI) における乗用車の価格指数と、ヘドニック法によって推計した価格指数を比較したうえで、両者の品質調整手法の違いが指数精度に及ぼす影響について検討している。また、Ohta [1987] では、中古車市場での取引価格にもヘドニック法を適用し、新車価格との比較を行っている。

図表 5 乗用車のヘッドニック推計における候補変数の一覧

候補変数一覧					
連続変数	乗車定員 (人)	性能 ダメージ 変数	ETC	性能 ダメージ 変数	バックカメラ
	全長 (mm)		ナビゲーション		サイドカメラ
性能 ダメージ 変数	全幅 (mm)	DVD プレイヤー	性能 ダメージ 変数	フロントカメラ	
	全高 (mm)	ブルーレイプレイヤー		全周囲カメラ	
	重量 (kg)	AM/FM ラジオ		AFS	
	ホイールベース (mm)	USB 入力端子		ヒルスタートアシスト	
	最小回転半径 (m)	アイドリングストップ		寒冷地仕様	
	燃費 (JC08 モード、km/l)	フルオートエアコン		レインセンサー	
	燃料タンク容量 (l)	デュアルエアコン		盗難防止システム	
	最大出力 (馬力)	前席左右独立温度調整式エアコン		車 型 ダ ミ ー 変 数	セダン
	最大トルク (kg × m)	運転席シートヒーター			ワゴン
	気筒数 (気筒)	シートポジションメモリー機能		クーペ	
	総排気量 (cc)	分割可倒式リアシート		オープンカー	
	リム径 (インチ)	前席パワーシート		ミニバン	
	タイヤ幅 (mm)	助手席パワーシート		SUV	
	タイヤ扁平率 (%)	後席パワーシート		ハッチバック	
	ギア数 (段)	本革シート		メ ー カ ー ダ ミ ー 変 数	国産車 A
	室内空間 (m ³)	本革巻きステアリング			国産車 B
	ディーゼル	ステアリングテレスコピック			国産車 C
	ハイブリッド	ステアリングコントローラー			国産車 D
	プラグインハイブリッド	ウッドパネル			国産車 E
	ハイオク仕様	アルミニウムホイール			国産車 F
ターボ	LED ヘッドランプ	国産車 G			
スーパーチャージャー	LED フォグランプ	国産車 H			
ツインターボ	フロントフォグランプ	国産車 I			
水平対向エンジン	リアフォグランプ	輸入車 A			
FF	キセノンヘッドランプ	輸入車 B			
FR	プロジェクターヘッドランプ	輸入車 C			
フルタイム式 4WD	LSD	輸入車 D			
パートタイム式 4WD	クルーズコントロール	発 売 時 期 ダ ミ ー 変 数	2016/3Q		
AT	ACC		2016/4Q		
MT	ACC (全車速対応)		2017/1Q		
CVT	クリアランスソナー		2017/2Q		
フロントスポイラー	LDWS		2017/3Q		
リアスポイラー	LKAS		2017/4Q		
リアワイパー	トラクションコントロール		2018/1Q		
サンルーフ	誤発進抑制装置		2018/2Q		
ガラスストップ	AEBS				
プライバシーガラス	ブレーキアシスト				
サイドエアバッグ	パーキングアシスト				

な品質情報を表す連続変数が約 20 種類、定性的な品質情報を表すダミー変数が約 100 種類含まれている²⁰。品質情報の多さは乗用車の複雑な製品特性に起因しているが、こうした複雑な製品では特に、変数選択をどのように行うかという点が課題になる。3 節で説明したように、本稿で提案するスパース推定を用いた手法では、変数選択を自動的に行う点に優れているが、そのメリットは、乗用車のように、品質情報が多岐にわたる製品の品質調整を行う際に、特に大きくなると考えられる。

(2) 推計上のパラメータの設定

実際に AEN で推計を行うに当たっては、(9) 式におけるハイパーパラメータ $(\lambda_1, \lambda_1^*, \lambda_2, \gamma)$ の設定方法が重要となる。本稿では、Zou and Zhang [2009] において挙げられている幾つかの設定基準の中から、機械学習の分野で広く用いられている「 K -分割交差検証 (K -fold cross validation)」を用いて、これらのハイパーパラメータを設定している²¹。具体的には、まず、第 1 段階の推計では、 $\lambda_2 = 10^a$: $a = -5, -4, \dots, 0, \dots, 5$ の計 11 通りの設定に対して交差検証を行い、 λ_1 を推定する²²。次に、第 2 段階の推計では、第 1 段階で得た (λ_1, λ_2) の組合せに対して、 L_1 ノルムに与える罰則の大きさを規定する適応パラメータ γ についてもいくつかのパターンを試行し、交差検証によって λ_1^* を推定する。 γ については、過度に小さい値に設定した場合に罰則が不十分となってオラクル性が満たされなくなることが知られている。本稿では、Zou [2006] に倣い γ の下限を 0.5 に設定したうえで、3 以下の範囲で 0.5 刻みのグリッドによって計 6 通りを設定している。以上のようなプロセスを経て、ハイパーパラメータ $(\lambda_1, \lambda_1^*, \lambda_2, \gamma)$ の組合せについて計 66 通りのパターンを試し、その中で交差検証による誤差が最も小さくなるケースの推計結果を採用することとしている。

また、AEN による推計では、最大で自由度と等しい個数の変数がモデルに採用されうる。もっとも、十分なサンプル数があって自由度が大きい場合は、モデルに採用される変数の個数が一定以下となるように別途の制約を与えることで、モデルが過度に複雑化することを避けつつ、計算負荷や品質調整における実務負荷を抑制

.....
20 ダミー変数には、性能のほか、車型やメーカー、発売時期に関するものを含む。

21 K -分割交差検証とは、対象データを K 個に分割し、そのうちの 1 個を検証データ、残りの $K-1$ 個を訓練データとして用い、検証データを変更しながら学習と検証を計 K 回繰り返すことで偏りなく性能を評価する手法である。 K を決定する際には、推計精度に影響を及ぼす偏り（モデルで推定したパラメータと真のパラメータの差）と分散（訓練データの違いに由来するランダムさ）のトレードオフに留意する必要がある。本稿の分析では、一般的に用いられることの多い $K=10$ を採用した。

22 実際には、 λ_1 の探索範囲として 100 通りのグリッドを定め（本研究では、R のパッケージ“gcdnet”が、 λ_2 の値に応じて自動生成する離散数列を使用）、その中で交差検証エラーが最小となる λ_1 を探索している。なお、第 2 段階の推計における交差検証においても、同様のプロセスを経る。

することも可能である。実際、本稿の推計では、自由度 939 に対し、採用変数の個数の上限を 140 に設定している。そのうえで、上限を 140 から更に引き上げたケースについても試行したが、回帰式のフィットの改善は、ごくわずかなものにとどまった。本稿における採用変数の個数上限の設定が、推計パフォーマンスに及ぼす影響は、限定的とみられる。

(3) 新旧手法による推計結果の比較

ここでは、ヘドニック法に関する従来の推計手法を適用した結果と、AEN を用いた新しい手法を適用した結果を示す。まず、従来のヘドニック法による推計結果は、図表 6 に掲載している。2 節で説明したように、従来の回帰式は変数の Box-Cox 変換を行ったうえで推計しており、尤度比検定の結果、被説明変数と説明変数（連続変数）をともに Box-Cox 変換した「両側 Box-Cox 形」が望ましいとの結果が得られた。説明変数のうち、定量的な品質情報を表す連続変数については、室内空間、燃費×等価慣性重量、最大出力のみが選択された。なお、ここでは車型別のダミー変数を利用して、機能・性能の価格に与える影響が車型によって異なることを考慮しており、室内空間については、例えばセダン・ワゴンでは有意とならず、ミニバンのみで有意となった。定性的な品質情報を表すダミー変数については、駆動系ダミー（4WD、FR）、オプションダミー（本革シート、LED ヘッドランプなど）、メーカーダミー（自動車メーカー各社に関するダミー）で、おのおの統計的に有意となるものがみられた。

次に、本稿で提案した AEN を用いた新しいヘドニック法による結果は、図表 7 に示している。前述したように、AEN などのスパース推定では、多数の説明変数を同時に推計式に含め、変数選択と係数の推定を同時に行うことができる。また、本稿では、非線形性の次数を 2 次までに限定したうえで、変数間での交互作用の存在を考慮し、交差項を多数取り込んでいる。その結果、従来の手法と比べて、多数の変数が採用されており、とりわけ多くの交差項が回帰式の中で捕捉されるようになっている²³。

もっとも、従来の手法と新しい手法における回帰式は関数形が大きく異なり、推計により得られたパラメータを単純に比較することができない。このため、以下のとおり理論価格への寄与率を変数ごとに算出し、その結果を比較する。

.....
23 なお、AEN における推定値の有意性の検証方法について、例えば残差ブートストラップ法の利用を提唱する論文もあるが（Chatterjee and Lahiri [2013]）、現時点ではコンセンサスは得られていない（Bonaldi, Hortaçsu, and Kastl [2015]）。

図表 6 従来手法によるヘドニック回帰式

関数形 被説明変数の Box-Cox パラメータ		両側 Box-Cox -0.280
定数項		3472.763***
室内空間 (m ³)	ミニバン Box-Cox パラメータ	1.360E-05*** 3.400
	セダン・ワゴン Box-Cox パラメータ	2.543E-09*** 1.372
燃費 (JC08 モード、km/l) × 等価慣性重量 (kg)	ミニバン Box-Cox パラメータ	1.606E-09*** 1.455
	SUV Box-Cox パラメータ	6.841E-09*** 1.330
	ハッチバック Box-Cox パラメータ	7.152E-18*** 3.351
	セダン・ワゴン Box-Cox パラメータ	2.846E-04*** 0.647
	ミニバン Box-Cox パラメータ	0.007*** 6.240E-06
最大出力 (馬力)	SUV Box-Cox パラメータ	5.880E-06*** 1.337
	ハッチバック Box-Cox パラメータ	0.008*** 3.621E-06
	ミニバン	-1,162.565***
	SUV	0.006***
ハッチバック	-2,306.764***	
駆動系ダミー		
4WD		0.002***
FR		0.002***
オプションダミー (標準搭載の場合のみ)		
本革シート		0.001***
サイドエアバッグ		4.504E-04**
パワーシート		0.002***
アルミニウムホイール		0.002***
LED ヘッドランプ		0.001***
差動制限装置 (LSD)		0.002***
定速走行・車間距離制御装置 (ACC) (全車速対応)		0.001***
車線逸脱警報システム (LDWS)		0.001***
配光可変型前照灯システム (AFS)		0.001***
パーキングアシスト		0.001***
メーカーダミー		
国産車 A		0.003***
国産車 B		-0.002***
国産車 C		-0.003***
国産車 D		-0.001***
国産車 E		0.004***
輸入車 A		0.006***
輸入車 B		0.006***
輸入車 C		0.008***
決定係数		0.957
自由度修正済み決定係数		0.956
回帰の標準誤差		0.002
被説明変数の平均値		3.509
両側 Box-Cox に対する検定 (H ₁ : 両側 Box-Cox)		
H ₀ : 片側 Box-Cox ($\lambda_j = 1$)		85.560***
H ₀ : 両側対数 ($\lambda_0 = \lambda_j = 0$)		273.705***
H ₀ : 片側対数 ($\lambda_0 = 0, \lambda_j = 1$)		130.257***
H ₀ : 線形 ($\lambda_0 = \lambda_j = 1$)		1,905.192***

備考：1. 等価慣性重量は、燃費消費率試験 (JC08 モード) において、試験自動車重量に応じてシャシダイナモメーターに設定する重量。

2. 駆動系ダミー「4WD」は、フルタイム式・パートタイム式を含む。

3. 説明変数には、上記のほか発売時期ダミーも含む。

資料：日本銀行

図表7 AENを用いた新手法によるヘドニック回帰式

説明変数		説明変数	
定数項	12.939	2016Q4 × 本革シート	-0.031
輸入車 C	0.244	2017Q3 × 全高 (mm)	-1.943E-05
輸入車 A	0.204	2017Q4 × CVT	-0.038
スーパーチャージャー	0.119	2018Q1 × フロントフォグランプ	0.005
ナビゲーション	0.142	2018Q2 × FF	-0.037
後席パワーシート	0.053	クーペ × 最大出力 (馬力) ÷ 重量 (kg)	0.223
アルミニウムホイール	0.032	ハッチバック × 最大出力 (馬力) ÷ 重量 (kg)	-0.315
LDWS	0.007	全高 (mm) × ハイブリッド	4.580E-05
ブルーレイプレイヤー	0.057	全高 (mm) × フロントフォグランプ	1.029E-05
人口密度 (人/m ²)	-0.055	全高 (mm) × デュアルエアコン	3.248E-05
重量 (kg) ÷ ボディサイズ (m ³)	0.001	全高 (mm) × サルーフ	1.827E-06
リム径 (インチ)	0.001	全高 (mm) × 運転席シートヒーター	4.776E-06
FF	-0.025	全高 (mm) × シートポジションメモリー機能	2.691E-05
重量 (kg) ÷ ボディサイズ (m ³) : 2次項	2.550E-05	全高 (mm) × ACC (全車速対応)	1.741E-05
全長 (m) × 全幅 (m) : 2次項	0.009	全高 (mm) × 重量 (kg) ÷ ボディサイズ (m ³)	2.412E-06
最大出力 (馬力) ÷ 重量 (kg) : 2次項	4.854	燃費 (JC08 モード, km/l) × 最大トルク (kg × m)	1.906E-05
国産車 G × 2017Q4	-0.003	燃費 (JC08 モード, km/l) × フロントスポイラー	0.001
国産車 G × フロントフォグランプ	-0.034	燃費 (JC08 モード, km/l) × ナビゲーション	2.649E-04
国産車 G × フロントスポイラー	-0.056	燃費 (JC08 モード, km/l) × デュアルエアコン	0.001
国産車 E × 2018Q2	0.248	燃費 (JC08 モード, km/l) × 本革巻きステアリング	0.001
国産車 E × 本革シート	0.018	燃費 (JC08 モード, km/l) × サルーフ	9.915E-05
国産車 E × AFS	0.032	燃費 (JC08 モード, km/l) × 運転席シートヒーター	0.001
国産車 E × 最大出力 (馬力) ÷ 重量 (kg)	0.865	最大トルク (kg × m) × リアスポイラー	0.001
国産車 D × 2016Q4	0.350	最大トルク (kg × m) × 最大出力 (馬力) ÷ 重量 (kg)	2.818E-04
国産車 D × 2017Q3	-0.012	ハイオク仕様 × フロントスポイラー	0.024
国産車 D × ハッチバック	-0.003	ハイオク仕様 × リアスポイラー	0.002
国産車 D × ミニバン	-0.135	MT × LED ヘッドランプ	0.019
国産車 D × 全高 (mm)	-4.036E-05	ギア数 (段) × リム径 (インチ)	0.001
輸入車 B × 最大トルク (kg × m)	0.003	LSD × 本革巻きステアリング	0.015
輸入車 B × フルオートエアコン	0.117	LSD × 本革シート	0.049
輸入車 B × リム径 (インチ)	0.005	クルーズコントロール × 重量 (kg) ÷ ボディサイズ (m ³)	3.507E-04
国産車 F × 2017Q3	-0.130	本革シート × 全長 (m) × 全幅 (m)	0.004
国産車 F × ハイブリッド	-0.061	LED ヘッドランプ × 重量 (kg) ÷ ボディサイズ (m ³)	4.970E-04
国産車 C × CVT	-0.135	ハイパーパラメータ	
国産車 B × CVT	-0.053	λ ₁	0.013
国産車 B × キセノンヘッドランプ	0.088	λ _{1*}	1.970E-05
輸入車 A × 2017Q3	0.073	λ ₂	1.000E-05
2016Q4 × 最大トルク (kg × m)	-0.003	λ	0.5

備考：2016年第3四半期から2018年第2四半期までの間に発売された製品のデータを用いて推計。ボディサイズ = 全長 × 全幅 × 全高。人口密度 = 乗車定員 ÷ (全長 × 全幅)。

$$\pi_l^{func} = \frac{y^{func}(\bar{x}_l + \Delta x_l, \bar{x}_{-l}) - y^{func}(\bar{x}_l, \bar{x}_{-l})}{y^{func}(\bar{x}_l, \bar{x}_{-l})} \times 100, \quad (10)$$

$$\log y^{AEN}(\bar{x}_l, \bar{x}_{-l}) = \hat{\beta}_{00} + \sum_{j=1}^p \hat{\beta}_{0j} \bar{x}_j + \sum_{j=1}^p \hat{\beta}_{jj} \bar{x}_j^2 + \sum_{k>j \geq 1} \hat{\beta}_{jk} \bar{x}_j \bar{x}_k, \quad (11)$$

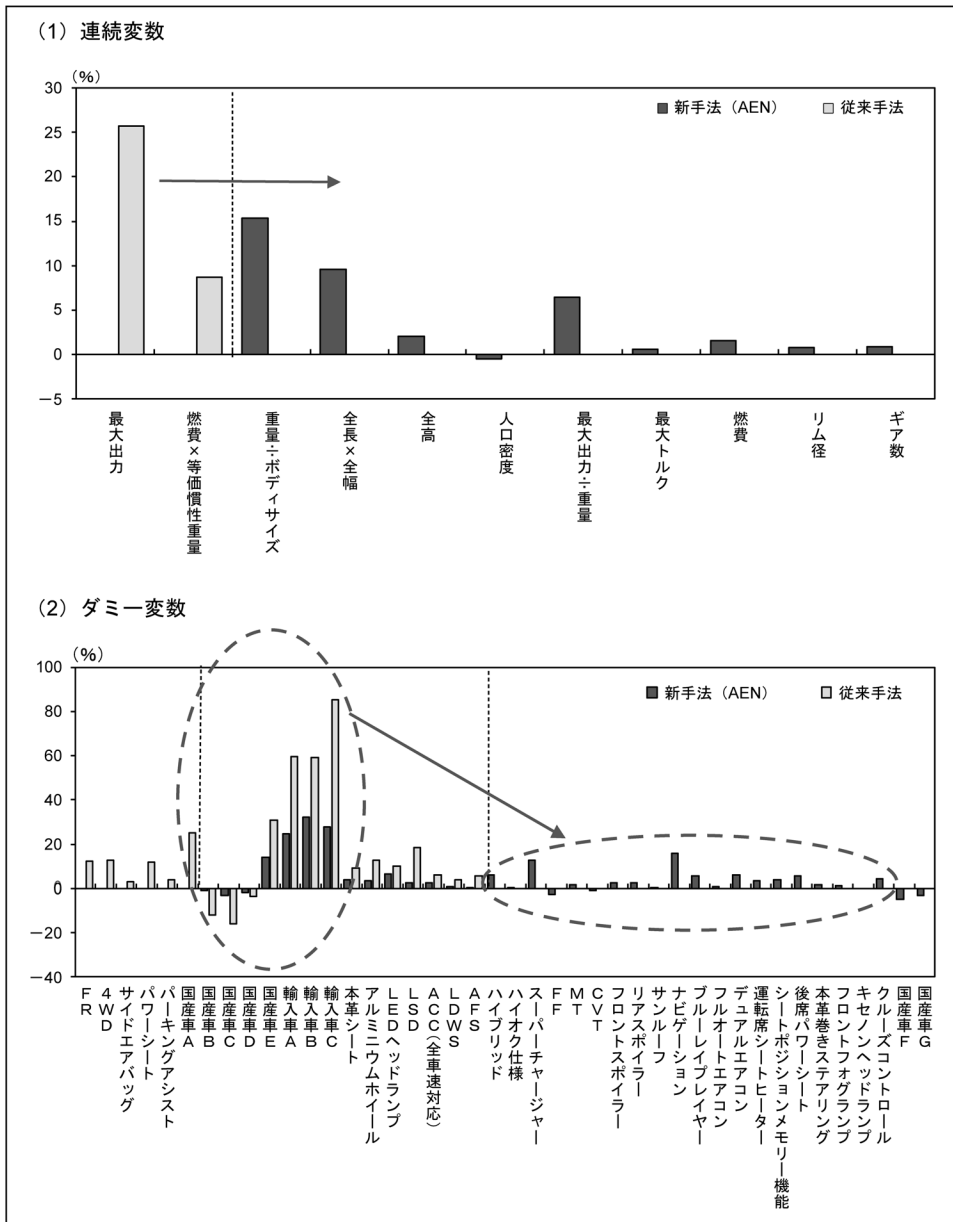
$$y^{Box-Cox}(\bar{x}_l, \bar{x}_{-l})^{(\lambda_0)} = \hat{\beta}_0 + \sum_{j=1}^{p_c} \hat{\beta}_{cj} \bar{x}_j^{(\lambda_j)} + \sum_{k=1}^{p_d} \hat{\beta}_{dk} \bar{x}_{dk}, \quad (12)$$

π_l^{func} : x_l の寄与率 ($func = AEN$ or $Box-Cox$),

\bar{x}_l : 説明変数 l の平均、 \bar{x}_{-l} : l 以外の説明変数の平均、

Δx_l : 説明変数 l の標準偏差 (連続変数) または 1 (ダミー変数)

図表 8 各変数の寄与率



備考：回帰式には、このほか車型ダミーと発売時期ダミーも存在。

図表 8 は、図表 6、7 で採用された連続変数およびダミー変数（交差項で採用されたものも含む）の、乗用車価格に対する寄与率を計算したものである。具体的に

は、すべての変数をサンプル期間中の平均値に設定した仮想的なデータセットを想定し、連続変数の場合 1 標準偏差、ダミー変数の場合 1 だけ増加させたときの理論価格の変化率 π_l^{func} を示している。

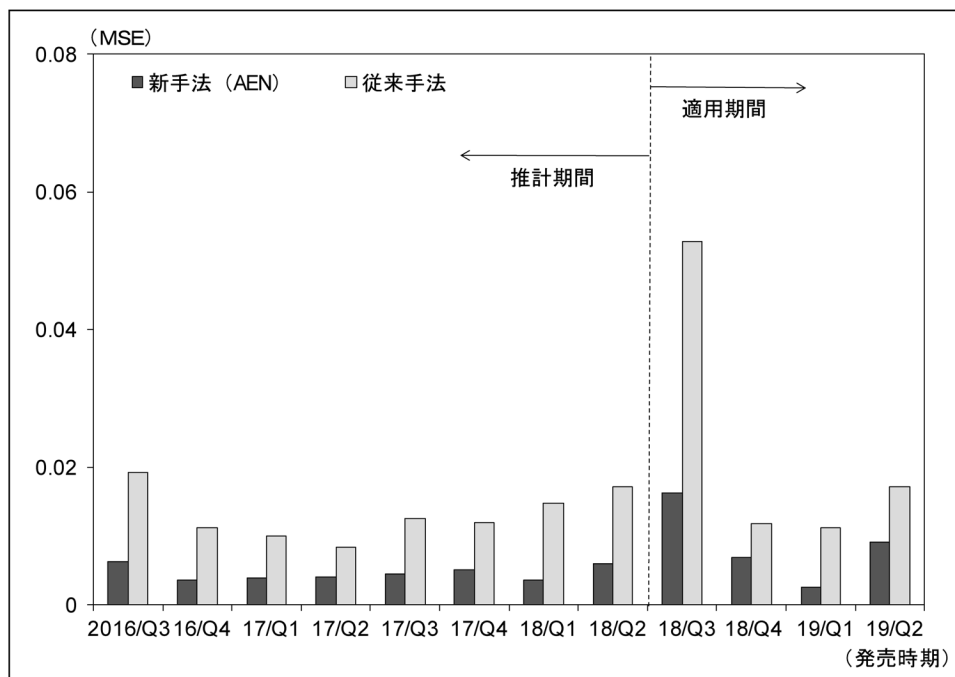
まず連続変数の結果をみると、新手法では、従来手法と比べて多くの変数が採用されている。すなわち、従来手法では 2 つのみであった採用変数が、新手法では 9 個に増えており、新手法によりヘドニック法の適用可能性が広がったものとみることができる。次に、ダミー変数の結果をみると、ここでも、新手法で採用された変数は、従来手法と比べて大幅に増加している。特に注目されるのが、新手法では、従来手法と比べ、機能・性能を表す多くのダミー変数が捕捉されるようになった反面、メーカーダミーの寄与率が低下しているということである。これは、これまでメーカー固有の要素として捕捉していた品質を、より具体的な機能に関する変数で捕捉できるようになったことを意味する。実際、乗用車の品質調整を行う場合には、同一メーカーによりモデルチェンジが行われるケースを対象とすることが多いため、回帰式がメーカーダミーに依存していると、実際にはモデルチェンジにより機能・性能が向上していると思われる場合でも、メーカーが変わらない以上、品質調整が行われないことになる。このため、回帰式の説明力がメーカーダミーに依存しなくなることは、品質調整の実務において、大きなメリットをもたらすと考えられる。

以上がパラメータに関する推計結果の比較であったが、新手法と従来手法のパフォーマンスを比較するには、推計式全体のフィットを比べる必要がある。このため、最新のデータセットを用いて、2 つの回帰式の平均二乗誤差を各四半期に発売された製品に対して算出したものが図表 9 である。新手法を用いることにより、平均二乗誤差は、従来手法と比べて全期間で縮小しており、推計精度が向上したことが確認される²⁴。特に、新手法では、推計期間中の誤差だけでなく、推計期間の後におけるサンプルに対しても、誤差が縮小している。実際の品質調整は、推計のサンプル期間の後に登場する製品に対して適用するため、こうしたアウト・オブ・サンプルに対するフィットの改善の意義は大きいと考えられる。

また、ヘドニック法を実務で運用する際は、回帰式を定期的に再推計する必要があるが、新手法では、データセットのサンプル期間を変更したときの推計結果の変化が小幅にとどまることも確認された（詳細は補論を参照）。このような、時系列安定性の改善も、ヘドニック法の適用可能性を向上させるものと考えられる。

.....
24 図表 9 を仔細にみると、2018 年第 3 四半期において、従来手法の誤差が極端に大きくなっている。これは、同時期にフルモデルチェンジされた特定の高級車が新たにサンプルに加わり、新たな安全性の導入や、デザインや内装面の大幅な刷新など、当該サンプルに固有の機能向上が数多くみられたことによるものと考えられる。

図表 9 新旧手法のフィット比較

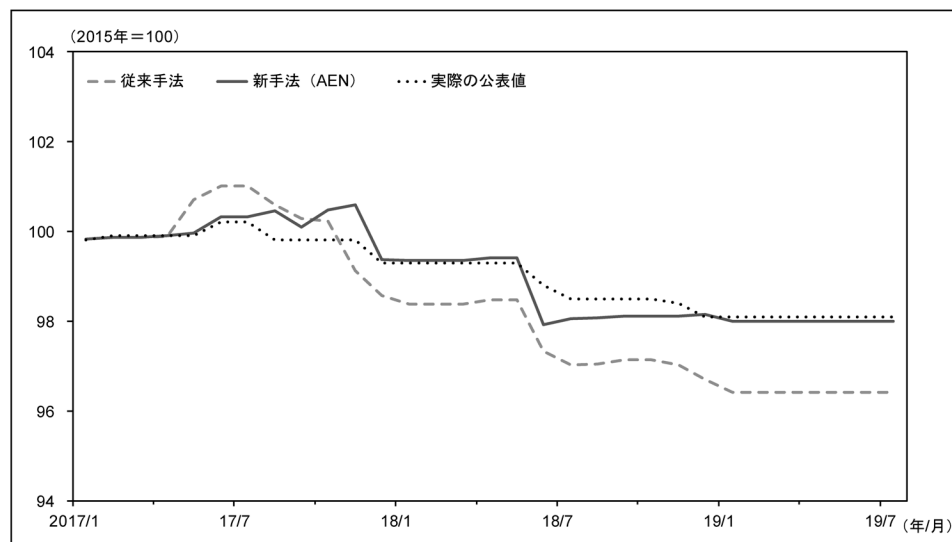


(4) 新手法導入の価格指数への影響

ここでは、AEN を用いた新たなヘドニック法の導入が、価格指数にどのような影響をもたらすかを確認する。具体的には、国内企業物価指数の品目「普通乗用車（ガソリン車）」について、2017 年以降に生じた調査価格の変更（旧製品の生産停止や市場での出回り製品の変化などに伴う調査対象商品の入替え）の際に、新旧のヘドニック法を機械的に適用していた場合に、価格指数がどのような推移を辿ったかという点を試算する。

そのうえで、新旧それぞれの手法による結果の妥当性を判断するため、国内企業物価指数の実際の公表値との比較を行う。なお、日本銀行調査統計局による企業物価指数の作成においては、ヘドニック法の適用対象品目であっても、同手法を必ずしも機械的に適用するのではなく、調査先企業への妥当性の確認や、他の品質調整方法（例えばコスト評価法など）との比較考量などを踏まえて、価格指数の作成・公表を行っている。言い換えれば、従来のヘドニック法による推計結果の機械的な適用が、実務面からみて妥当性を有しないと判断されれば、ヘドニック法の適用が

図表 10 PPI「普通乗用車（ガソリン車）」品質調整手法別推移



見送られているということである²⁵。このため、新手法を機械的に適用した場合の価格指数が、実務判断も踏まえて作成・公表された国内企業物価指数の公表値と類似しているかをみることにより、新しいヘドニック法の適用によって、品質向上率を適切に推計できるかどうかを評価することができる。

図表 10 が試算結果を表す。図中の破線は、過去に生じた調査価格の変更のすべてについて、従来のヘドニック法を機械的に適用した場合の価格指数の試算値である。これをみると、従来型のヘドニック法では、価格指数の下落幅がやや大きくなっている。一方、図中の太実線は、AEN を用いた新たなヘドニック法を、過去の調査価格変更にも機械的に適用した場合であるが、この場合、価格の下落ペースが従来手法と比べ、マイルドになっている。このように、新旧手法による推計結果の相違は、乗用車の価格指数に対して、定量的に無視できない影響を与えている。

また、図表 10 では、国内企業物価指数の公表値を点線で示しており、その推移は、従来のヘドニック法による試算値よりも、AEN による新しいヘドニック法によるものに近くなっている。この結果は、従来の手法を機械的に適用していれば、品質向上率を過大に見積もる結果、価格指数の下落幅が過大になるリスクがあった一方、新しい手法では、たとえそれを機械的に用いたとしても、品質向上率を概ね適切に評価できる可能性があることを示している。

この結果は、2 節で説明した欠落変数のバイアスに関する先行研究の結果と整合

25 企業物価指数の作成における品質調整方法（ヘドニック法やその他の品質調整方法）の利用に関する考え方については、日本銀行調査統計局 [2017] の 7 章を参照。

的である。すなわち、従来手法では、多重共線性の問題などにより、取り込める説明変数の数が限られる結果、欠落変数が生じやすく、それにより、回帰式に採択される変数のパラメータに歪みが生じ、このことが品質向上率を過大に評価させていたとの解釈が成り立つ。一方で、AENによる推計では、捕捉される説明変数が増加した結果、欠落変数によるバイアスが生じにくくなり、採択された変数にかかる推計パラメータの歪みが小さい結果、品質向上率の評価が適切に行われやすく、その結果が価格指数の違いに表れていると考えられる。

5. おわりに

本稿では、ヘドニック法で直面しうる課題について整理したうえで、スパース推定を用いた新しい推計手法の詳細やその推計結果についてまとめた。本稿で導入したスパース推定の1つ「アダプティブ・エラスティック・ネット (adaptive elastic net: AEN)」は、説明変数間の多重共線性に対する頑健性を高める「グループ効果」と、変数選択の適正性と係数の漸近的な不偏性を同時に満たす「オラクル性」を有し、ヘドニック法が直面する実務上の課題を克服できる可能性を持つ。実際、本稿において、わが国の乗用車の価格に関する実証分析を行ったところ、AENを用いた新手法は、従来の推計手法と比べて、①回帰式に採用される変数の大幅な増加、②フィットの改善、③欠落変数によるバイアスの解消の観点で改善がみられた。とりわけ、新手法を適用することにより、乗用車の価格指数の下落率が、従来手法と比べてマイルドなものになった点は、従来手法が抱えていた、欠落変数のバイアスに伴う品質向上率の過大評価のリスクを緩和するものと考えられる。そのことは、ヘドニック法による品質調整の精度改善につながるるとともに、実務上、個別の調査対象商品の変更を行う際に、ヘドニック法の適用可能性を向上させるものと期待される。1節で述べたように、ヘドニック法は、データと統計的手法に基づいて品質評価を客観的に行える点や、新旧商品間の機能・性能の変化が多岐にわたる場合にも対応可能な点に強みを有する。こうした特長のあるヘドニック法の利便性が増すことは、物価指数の精度向上にも寄与するものと考えられる。

本稿では、乗用車を例にとって推計を行ったが、本稿で提案した手法は、「スパース推定」と「多項式回帰」という汎用的なアプローチを軸として用いており、より幅広い製品についても適用可能である。ヘドニック法では、個別の製品の特徴を十分に踏まえてデータセットの構築やモデル化を行う必要があるが、本稿で指摘した課題は、どの製品においても概ね共通して抱えるものであり、そうした課題への対応を企図した新手法は、さまざまな製品においてヘドニック法のパフォーマンスを改善しうるものと考えられる。また、汎用的なアプローチであることは、統計的手

法の研究の進展や、実務上の要請に応じてカスタマイズ可能な柔軟性の高い推計手法であることも意味している。例えば、より高度なスパース推定を適用することで、推計精度やパラメータの安定性は改善可能か、ハイパーパラメータ設定の際、より高度な交差検証手法を用いることで汎化性能を更に高めることは可能かなどの論点は、研究の余地として考えられる。他方で、実務上求められる推計精度が必ずしも高くない場合、採用変数の上限を低めに設定してより単純な関数形、変数構成を選択したり、交差項を作成する範囲を初めから限定するなど、推計手法のフレームワークは維持しつつ、解釈可能性をより重視するアプローチも十分に想定される。

本稿の手法は、ヘドニック法の回帰モデルから直接的に物価指数を作成するアプローチにおいても有用である。ヘドニック法では、複数期間のクロスセクション・データをプールしたうえで推計式に時間ダミーを加え、その時間ダミーのパラメータの変化から品質調整済物価指数を作成する手法がとられることもある²⁶。こうした手法は、推計サンプル全体の価格情報を考慮することで、物価指数の代表性を確保しやすい一方、品質調整の結果がダミー変数の推計精度に依存するため、多重共線性や欠落変数の問題が顕在するもとは精度に課題がある。特に、本稿で取り上げた乗用車においては、近年、ハイブリッド車や電気自動車などの環境対応、安全性能の追加といった機能面の多様化が一段と進展している中、調査価格の代表性確保と品質変化の適切な捕捉を両立して行っていくことの重要性が増している。本稿で提案したスパース推定による手法は、こうした課題に対処していくうえでも一定の効果が期待される。

本稿は、ヘドニック法にスパース推定を適用することで、多重共線性と欠落変数の問題に対処することに焦点を当てたものであるが、同推計を巡っては、それ以外にも数多くの論点が存在する。例えば、データセットの作成方法は、欠落変数バイアスにも通ずる重要な論点である。「Garbage in Garbage out」という格言が示すとおり、技術革新に伴う製品機能の高度化を的確に捉えながら、必要に応じて新しい機能に関する変数を取り込んでいくといった対応により、ヘドニック法で用いるデータセットの質を維持していくことが重要である。ヘドニック法分野では、ビックデータ解析など近年の情報処理技術の進歩をデータセット構築にどのように活かしていくか、という論点も、現在研究が進められているところである²⁷。大型

26 他方で、日本銀行調査統計局が作成するわが国の企業物価指数においては、ヘドニック法の回帰モデルに基づき、製品の機能・特性を金額換算した「理論価格」を推計したうえで、新旧製品間の理論価格の差を品質変化とみなし、表面価格の差から控除することで純粋な価格変動を特定するアプローチをとっている。こうした手法は、品質調整の精度が、推計されたダミー変数の精度のみに依存されずに済む一方、サンプル・バイアスを避ける観点からは、市場シェアに応じた調査価格の代表性確保が必要となる。

27 例えば、ウェブ・スクレイピングで入手したデータを基にヘドニック法を実施する研究として Zafar and Himpens [2019] や Efthymiou and Antoniou [2013] が挙げられる。

データセットの利用は、今後一層容易となっていくことが見込まれる。こうした状況下、本稿で提案した推計手法は、従来の推計手法では扱いきれない程潤沢なデータであっても、必要な情報を漏らさず取捨選択したうえで、パフォーマンスのよいモデルの構築を自動的に行うことができるという意味において効率性の高いものとなっており、今後の実証研究、統計実務における更なる活用が期待される。

参考文献

- 太田 誠、「ヘドニック・アプローチの理論的基礎，方法および日本の乗用車価格への応用」、『季刊理論経済学』第 29 巻第 1 号、日本経済学会、1978 年、31～55 頁
- 白塚重典、「乗用車価格の変動と品質変化—ヘドニック・アプローチによる品質変化の計測と CPI への影響—」、『金融研究』第 14 巻第 3 号、日本銀行金融研究所、1995 年、77～120 頁
- 、「ヘドニック・アプローチによる品質変化の捕捉—理論的枠組みと実証研究への適用—」、日本銀行金融研究所ディスカッションペーパー No. 97-J-6、日本銀行金融研究所、1997 年
- 、『物価の経済分析』、東京大学出版会、1998 年
- 日本銀行調査統計局、「2015 年基準 企業物価指数の解説」、日本銀行、2017 年 (https://www.boj.or.jp/statistics/outline/exp/pi/cgpi_2015/data/excgpi15a.pdf)
- 平形尚久、「ヘドニック関数の時系列変化と価格指数への影響について—デスクトップパソコンのケース—」、日本銀行ワーキングペーパー No. 05-J-1、日本銀行、2005 年
- Bonaldi, Pietro, Ali Hortaçsu, and Jakub Kastl, “An Empirical Analysis of Funding Costs Spillovers in the Euro-Zone with Application to Systemic Risk,” NBER Working Paper No. 21462, National Bureau of Economic Research, 2015.
- Box, George E. P., and David R. Cox, “An Analysis of Transformations,” *Journal of the Royal Statistical Society: Series B*, 26(2), 1964, pp. 211–252.
- Chatterjee, Anjan, and Soumendra N. Lahiri, “Rates of Convergence of the Adaptive LASSO Estimators to the Oracle Distribution and Higher Order Refinements by the Bootstrap,” *The Annals of Statistics*, 41(3), 2013, pp. 1232–1259.
- Cropper, Maureen L., Leland B. Deck, and Kenneth E. McConnell, “On the Choice of Functional Form for Hedonic Price Functions,” *The Review of Economics and Statistics*, 70(4), 1988, pp. 668–675.
- Efthymiou, Dimitrios, and Constantinos Antoniou, “How Do Transport Infrastructure and Policies Affect House Prices and Rents? Evidence from Athens, Greece,” *Transportation Research Part A*, 52, 2013, pp. 1–22.
- The Event Horizon Telescope Collaboration, “First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole,” *The Astrophysical Journal Letters*, 875(1), 2019.
- Graves, Philip E., James C. Murdoch, Mark A. Thayer, and Donald M. Waldman, “The Robustness of Hedonic Price Estimation: Urban Air Quality,” *Land Economics*, 64(3), 1988, pp. 220–233.
- Halvorsen, Robert, and Henry O. Pollakowski, “Choice of Functional Form for Hedonic

- Price Equations,” *Journal of Urban Economics*, 10(1), 1981, pp. 37–49.
- Hoerl, Arthur E., and Robert W. Kennard, “Ridge Regression: Biased Estimation for Nonorthogonal Problems,” *Technometrics*, 12(1), 1970, pp. 55–67.
- Jin, Chanwoo, and Gunhak Lee, “Exploring Spatiotemporal Dynamics in a Housing Market using the Spatial Vector Autoregressive Lasso: A Case Study of Seoul, Korea,” *Transactions in GIS*, 24(1), 2020, pp. 27–43.
- Ohta, Makoto, “Hedonic Price Indexes of Japanese Passenger Cars over 1970–83: A Note,” *The Economic Studies Quarterly*, 38(3), 1987, pp. 264–274.
- Pakes, Ariel, “A Reconsideration of Hedonic Price Indexes with an Application to PC’s,” *American Economic Review*, 93(5), 2003, pp. 1578–1596.
- Rosen, Sherwin, “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition,” *Journal of Political Economy*, 82(1), 1974, pp. 34–55.
- Sawyer, Steven D., and Alvin So, “A New Approach for Quality-Adjusting PPI Microprocessors,” *Monthly Labor Review*, Bureau of Labor Statistics, 2018.
- Tibshirani, Robert, “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society: Series B*, 58(1), 1996, pp. 267–288.
- Triplett, Jack E., *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, OECD Publishing, 2006.
- Wheeler, David C., “Simultaneous Coefficient Penalization and Model Selection in Geographically Weighted Regression: The Geographically Weighted LASSO,” *Environment and Planning A*, 41(3), 2009, pp. 722–742.
- Zafar, Jean-Denis, and Stéphanie Himpens, “Web scraping Laptop Prices to Estimate Hedonic Models and Extensions to Other Predictive Methods,” paper presented at the 16th meeting of the Ottawa Group on Price Indices, Rio de Janeiro, 2019.
- Zou, Hui, “The Adaptive Lasso and Its Oracle Properties,” *Journal of the American Statistical Association*, 101(476), 2006, pp. 1418–1429.
- , and Trevor Hastie, “Regularization and Variable Selection via the Elastic Net,” *Journal of the Royal Statistical Society: Series B*, 67(2), 2005, pp. 301–320.
- , and Hao H. Zhang, “On the Adaptive Elastic-Net with a Diverging Number of Parameters,” *The Annals of Statistics*, 37(4), 2009, pp. 1733–1751.

補論. 回帰式の時系列安定性

ヘドニック法の回帰式は時系列的に不安定であることが広く知られている。これは、技術進歩や消費者の嗜好の変化などの影響を受けて、機能・性能と価格の関係が時間とともに変化しうるためである。例えば、Pakes [2003] は、パソコンの回帰式の推計に関し、技術革新によって半導体の価格が大きく下落すると、この半導体を搭載するパソコンの回帰式も変化しうることを指摘したうえで、実際に推定されるパラメータが大きく変わりうることを実証している。こうした変化を適切に捉えていくためには、関数形や説明変数の構成の変化に柔軟に対応しながら、回帰式を定期的に再推計していく必要がある。

ここでは、本論と同様、乗用車を例にとり、推計期間の異なる回帰式を比較することで、回帰式の時系列的な安定性について分析する。具体的には、4 節 (3) で用いたものより 1 年古いデータセット (2015 年第 3 四半期～2017 年第 2 四半期、サンプル数 1,188) を用いて、従来手法により推計した結果が図表 A-1、AEN を用いた新手法による推計結果が図表 A-2 である。以下ではデータセットの変化が推計結果をどの程度変化させるかにつき、新旧手法間の比較を行う。

初めに、サンプルの発売時期が 1 年古くなることで、各変数のパラメータがどの程度変化するかについて整理する。(A-1)、(A-2) 式のとおり、新手法、従来手法それぞれについて、図表 8 と同様の手順で算出した各変数の寄与率が旧サンプルへの変更によってどの程度変化したか箱ひげ図で示したものが図表 A-3 である²⁸。

$$\text{変数 } l \text{ の寄与率の変化} = \pi_l^{func,OLD} - \pi_l^{func,NEW}, \quad (\text{A-1})$$

$$\pi_l^{func,smpl} = \frac{y^{func,smpl}(\bar{x}_l + \Delta x_l, \bar{x}_{-l}) - y^{func,smpl}(\bar{x}_l, \bar{x}_{-l})}{y^{func,smpl}(\bar{x}_l, \bar{x}_{-l})} \times 100, \quad (\text{A-2})$$

$\pi_l^{func,smpl}$: x_l の寄与率 ($func = AEN \text{ or Box-Cox, } smpl = NEW \text{ or OLD}$),

$y^{func,smpl}$: 理論価格 ($func = AEN \text{ or Box-Cox, } smpl = NEW \text{ or OLD}$).

図表 A-3 をみると、データセットが 1 年古くなることによるパラメータの変化は、AEN を用いた新手法では $\pm 5\%$ ポイント程度の範囲に概ね収まっているのに対し、従来手法では $\pm 20\%$ ポイント程度の範囲で生じていることがわかる。これをみる限り、新手法は従来手法に比べて、変数ごとのパラメータが時系列方向に安定的であることが示唆される。

もっとも、変数ごとのパラメータの安定性が、直ちに品質調整結果の安定性につながるとは限らない。個々のパラメータの変化は小さくても、その変化の方向が同

.....
28 片方の回帰式にしか存在しない変数については、存在しない側の回帰式の寄与率を 0 として差分を計算している。

図表 A-1 従来手法によるヘドニック回帰式（旧データ）

関数形		両側 Box-Cox
被説明変数の Box-Cox パラメータ		0.150
定数項		1,664.131***
室内空間 (m ³)	セダン・ワゴン	2.433***
	Box-Cox パラメータ	1.066
	ミニバン	0.039***
	Box-Cox パラメータ	2.770
燃費 (JC08 モード、km/l) × 等価慣性重量 (kg)	セダン・ワゴン	9.512E-08***
	Box-Cox パラメータ	1.637
	ミニバン	1.754E-09***
	Box-Cox パラメータ	2.097
	SUV	3.147***
	Box-Cox パラメータ	0.002
最大出力 (馬力)	ハッチバック	1.951E-26***
	Box-Cox パラメータ	5.773
	セダン・ワゴン	5.993***
	Box-Cox パラメータ	0.003
	ミニバン	3.825E-07***
	Box-Cox パラメータ	3.384
SUV	Box-Cox パラメータ	3.243***
	Box-Cox パラメータ	0.040
	ハッチバック	4.408***
	Box-Cox パラメータ	0.018
車型ダミー		
ミニバン		2,275.621***
SUV		819.239***
ハッチバック		2,019.161***
原動機ダミー		
ハイブリッド		0.393***
プラグインハイブリッド		2.137***
駆動系ダミー		
4WD		0.846***
オプションダミー (標準搭載の場合のみ)		
本革シート		1.003***
サイドエアバッグ		0.559***
パワーシート		0.869***
プライバシーガラス		0.782***
差動制限装置 (LSD)		0.630***
衝突被害軽減ブレーキ (AEBS)		0.358***
定速走行・車間距離制御装置 (ACC)		0.405***
車線逸脱警報システム (LDWS)		0.184**
配光可変型前照灯システム (AFS)		0.624***
メーカーダミー		
国産車 A		-0.896***
国産車 B		-1.557***
国産車 C		-1.353***
国産車 D		-1.237***
国産車 E		2.648***
国産車 F		-0.611***
国産車 G		-1.803***
国産車 H		-0.523***
輸入車 A		2.987***
輸入車 B		3.825***
輸入車 C		4.723***
決定係数		0.962
自由度修正済み決定係数		0.961
回帰の標準誤差		0.792
被説明変数の平均値		52.810
両側 Box-Cox に対する検定 (H ₁ : 両側 Box-Cox)		
H ₀ : 片側 Box-Cox ($\lambda_j = 1$)		220.310***
H ₀ : 両側対数 ($\lambda_0 = \lambda_j = 0$)		158.589***
H ₀ : 片側対数 ($\lambda_0 = 0, \lambda_j = 1$)		238.038***
H ₀ : 線形 ($\lambda_0 = \lambda_j = 1$)		641.781***

備考：1. 等価慣性重量は、燃費消費率試験 (JC08 モード) において、試験自動車重量に応じてシャシダイナモメーターに設定する重量。

2. 駆動系ダミー「4WD」は、フルタイム式・パートタイム式を含む。

3. 説明変数には、上記のほかに発売時期ダミーも含む。

資料：日本銀行

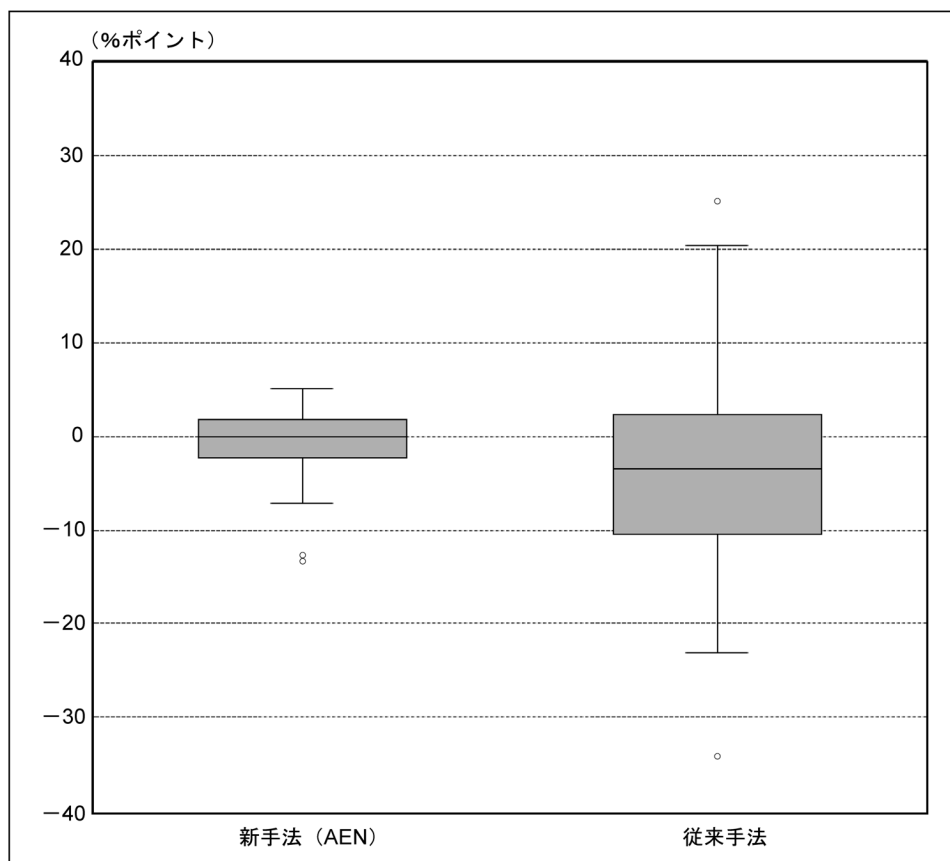
図表 A-2 AEN を用いた新手法によるヘドニック回帰式（旧データ）

説明変数		説明変数	
定数項	12.879	重量 (kg) ÷ ボディサイズ (m ³) × 全高 (mm)	2.436E-06
人口密度 (人/m ²)	-0.192	重量 (kg) ÷ ボディサイズ (m ³) × リム径 (インチ)	9.914E-05
重量 (kg) ÷ ボディサイズ (m ³)	0.002	重量 (kg) ÷ ボディサイズ (m ³) × ステアリングテレスコピック	1.396E-04
全長 (m) × 全幅 (m)	0.032	重量 (kg) ÷ ボディサイズ (m ³) × 盗難防止システム	3.299E-04
ナビゲーション	0.039	重量 (kg) ÷ ボディサイズ (m ³) × ブルーレイプレイヤー	3.530E-04
本革シート	1.432E-05	全長 (m) × 全幅 (m) × リム径 (インチ)	0.002
後席パワーシート	0.091	全高 (mm) × 本革シート	3.468E-05
全長 (m) × 全幅 (m) : 2 次項	0.002	全高 (mm) × LED ヘッドランプ	3.315E-05
輸入車 C × 全高 (mm)	1.123E-05	全高 (mm) × LDWS	4.677E-06
輸入車 C × 最大出力 (馬力) ÷ 重量 (kg)	2.093	燃費 (JC08 モード, km/l) × ハイブリッド	0.001
国産車 G × CVT	-0.025	燃費 (JC08 モード, km/l) × ナビゲーション	0.002
国産車 G × アルミニウムホイール	-0.067	燃費 (JC08 モード, km/l) × デュアルエアコン	0.002
国産車 E × 最大出力 (馬力) ÷ 重量 (kg)	1.211	燃費 (JC08 モード, km/l) × クルーズコントロール	0.001
国産車 E × 運転席シートヒーター	0.014	燃費 (JC08 モード, km/l) × 本革巻きステアリング	0.001
国産車 D × 2016Q3	-0.034	燃費 (JC08 モード, km/l) × 前席パワーシート	0.001
国産車 D × 2016Q4	0.203	燃費 (JC08 モード, km/l) × 運転席シートヒーター	0.001
国産車 D × ミニバン	-0.019	ハイブリッド × レインセンサー	0.013
国産車 D × 人口密度 (人/m ²)	-0.114	最大出力 (馬力) ÷ 重量 (kg) × 最大トルク (kg × m)	0.007
輸入車 B × 最大出力 (馬力) ÷ 重量 (kg)	1.016	最大出力 (馬力) ÷ 重量 (kg) × ハイオク仕様	0.167
輸入車 B × 最大トルク (kg × m)	0.002	最大出力 (馬力) ÷ 重量 (kg) × FF	-0.312
輸入車 B × フルオートエアコン	0.137	最大出力 (馬力) ÷ 重量 (kg) × ナビゲーション	0.065
輸入車 B × ウッドパネル	0.018	最大出力 (馬力) ÷ 重量 (kg) × デュアルエアコン	0.143
輸入車 B × プライバシーガラス	0.011	最大出力 (馬力) ÷ 重量 (kg) × ウッドパネル	0.268
国産車 F × 2016Q1	-0.068	最大出力 (馬力) ÷ 重量 (kg) × アルミニウムホイール	0.591
国産車 F × 燃費 (JC08 モード, km/l)	-0.001	最大出力 (馬力) ÷ 重量 (kg) × サイドエアバッグ	0.193
国産車 C × CVT	-0.146	最大出力 (馬力) ÷ 重量 (kg) × 盗難防止システム	0.009
国産車 B × 2015Q4	-0.124	最大出力 (馬力) ÷ 重量 (kg) × ブルーレイプレイヤー	0.010
国産車 B × ハッチバック	-0.074	フロントフォグランプ × AEB	0.014
輸入車 A × 本革巻きステアリング	0.145	アルミニウムホイール × AEB	0.006
2016Q3 × クルーズコントロール	0.009	ハイパーパラメータ	
2016Q3 × ウッドパネル	0.012	λ ₁	0.017
2016Q3 × ETC	0.056	λ _{1*}	2.068E-05
2016Q3 × 運転席シートヒーター	0.019	λ ₂	0.010
ハッチバック × 人口密度 (人/m ²)	-0.047	λ	0.5
ミニバン × ハイブリッド	0.043		
ミニバン × 運転席シートヒーター	0.029		
人口密度 (人/m ²) × 前席左右独立温度調整式エアコン	0.027		
重量 (kg) ÷ ボディサイズ (m ³) × 全長 (m) × 全幅 (m)	2.605E-04		

備考：2015 年第 3 四半期から 2017 年第 2 四半期までの間に発売された製品のデータを用いて推計。ボディサイズ = 全長 × 全幅 × 全高。人口密度 = 乗車定員 ÷ (全長 × 全幅)。

じであれば、全体としての品質変化率（理論価格の変動率）は大きく変化しうる。逆に、ある変数のパラメータが上昇する一方で、その変数と相関が強い別の変数のパラメータが低下しており、品質調整全体としてみればパラメータ変化の影響が相殺されている、といった状況も想像に難くない。このため、品質調整結果の安定性を評価するに当たっては、回帰式全体の示唆する品質変化率がどの程度変化するかにも注目する必要がある。こうした点を踏まえ、平形 [2005] は、デスクトップパソコンの回帰式を複数の関数形で推計したうえで、推計期間の異なる回帰式を用いて品質調整を行った場合の結果を比較し、物価指数がどの程度変化しうるかについて分析している。以下では、平形 [2005] の手法に倣い、乗用車におけるモデルチェンジを想定し、ヘドニック法の示唆する品質変化率の時系列安定性について、新旧手法間の比較を行う。

図表 A-3 1年古いデータセットで推計したときの変数ごとの寄与率の変化



- 備考：1. 「セダン・ワゴン」について、すべての変数を平均値とした仮想的なサンプルを想定し、連続変数の場合1標準偏差、ダミー変数の場合1だけ増加させたときの理論価格の変化率を、現サンプル、旧サンプル双方から推計された回帰式を用いて算出。各変数の寄与率の現旧差（旧－現）を掲載。ただし、車型ダミーと発売時期ダミーは寄与率の計算対象から除外している。
2. 寄与率の変化を算出した変数の数は、新手法（AEN）は55個、従来手法は31個。片方の回帰式にしか存在しない変数については、存在しない側の回帰式の寄与率を0として差分を計算している。
3. 「第一四分位数 $-1.5 \times$ 四分位範囲」を下回る値、「第三四分位数 $+1.5 \times$ 四分位範囲」を上回る値については、外れ値として表示している。

まず、図表9でフィットのよさを確認するために用いたテスト用のデータセットについて、「セダン・ワゴン」に分類されるサンプルを抽出し、発売四半期別にグループ分けする²⁹。その後、各四半期について、全スペックが各グループの平均値

.....
²⁹ 2016年第3四半期～2019年第2四半期。ただし、2017年第2四半期、2018年第1四半期、2019年第1四半期の3期については、「セダン・ワゴン」に該当するサンプルが存在しなかったため、分析の対象外とした。

である仮想的な製品を構築する。そのうえで、これらの仮想製品間でのモデルチェンジによる品質向上率を、新手法、従来手法のそれぞれで算出する^{30,31}。こうして得られた品質向上率を、現サンプルによるものと旧サンプルによるものとの間で比較することで、回帰式の時系列変化が物価指数に与える影響を、変数間の相関も加味して包括的に考察することができると考えられる。

$$\Pi^{func,smpl} = \frac{y^{func,smpl}(\bar{x}') - y^{func,smpl}(\bar{x})}{y^{func,smpl}(\bar{x})} \times 100, \quad (A-3)$$

$\Pi^{func,smpl}$: 品質変化率 ($func = AEN \text{ or } Box-Cox, smpl = NEW \text{ or } OLD$),

$y^{func,smpl}$: 理論価格 ($func = AEN \text{ or } Box-Cox, smpl = NEW \text{ or } OLD$),

\bar{x} : 仮想的な製品 (モデルチェンジ前) のスペック、

\bar{x}' : 仮想的な製品 (モデルチェンジ後) のスペック

図表 A-4 は、新手法および従来手法について、横軸を現サンプルによる算出結果 $\Pi^{func,NEW}$ 、縦軸を旧サンプルによる算出結果 $\Pi^{func,OLD}$ として、仮想的なモデルチェンジの品質変化率を比較したものである。新手法の散布図は概ね対角線の周りに分布している一方で、従来手法の散布図は対角線からの乖離が大きくなっている。これは、回帰式の推計期間の変化によって生じる品質変化率の乖離は、新手法を適用したときの方が小さいことを示唆している。この点を定量的に評価するため、以下のとおり算出される推計期間の変化に伴う品質変化率の乖離 (絶対値) を、新手法と従来手法のそれぞれについて表で示したものが、図表 A-5 である。

$$\text{品質変化率の乖離} = |\Pi^{func,OLD} - \Pi^{func,NEW}|, \quad (A-4)$$

$$\text{where } \Pi^{func,smpl} = \frac{y^{func,smpl}(\bar{x}') - y^{func,smpl}(\bar{x})}{y^{func,smpl}(\bar{x})} \times 100,$$

$\Pi^{func,smpl}$: 品質変化率 ($func = AEN \text{ or } Box-Cox, smpl = NEW \text{ or } OLD$),

$y^{func,smpl}$: 理論価格 ($func = AEN \text{ or } Box-Cox, smpl = NEW \text{ or } OLD$),

\bar{x} : 仮想的な製品 (モデルチェンジ前) のスペック、

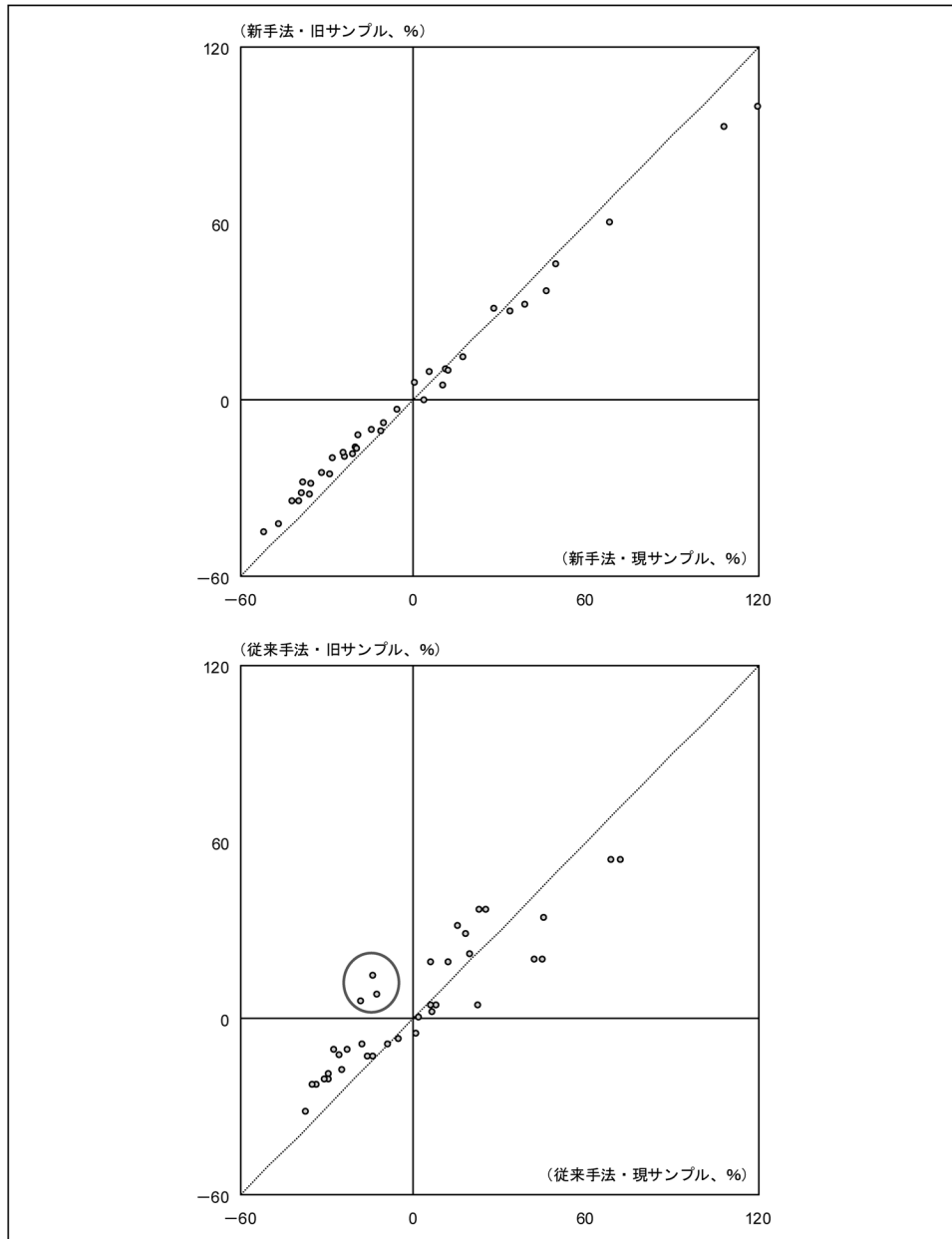
\bar{x}' : 仮想的な製品 (モデルチェンジ後) のスペック

AEN を用いた新手法の示唆する品質向上率の新旧乖離は、全期間平均で、従来手法の概ね半分程度となっており、AEN の導入によって推計結果の時系列安定性が増したことが確認できる。これは、本論で確認したとおり、新手法が、イン・サンプルだけでなくアウト・オブ・サンプルに対しても比較的良好なフィットを示すことと整合的であり、時間の経過とともに製品の価格と機能の関係が変化した場

30 品質変化率を算出する際、発売時期ダミーとメーカーダミーは平均値ではなく 1 つの値に固定している。

31 すなわち、合計で $36 (= {}_9C_2) \times 2$ (AEN or Box-Cox) $\times 2$ (現サンプル or 旧サンプル) = 144 通りの品質向上率が算出される。

図表 A-4 1年古いデータセットで推計したときの品質変化率の変化



備考：ある四半期に発売開始された製品の平均的なスペックを有する仮想的な製品から、別の四半期に対して同様に構築した仮想的な製品へのモデルチェンジが発生したと仮定。その際の品質変化率（理論価格の変動率）を現サンプル・旧サンプルから推計された回帰式で算出し、その組合せを掲載。

図表 A-5 1年古いデータセットで推計したときの品質変化率の乖離

モデルチェンジ 前\後	新手法 (AEN)							
	16/Q4	17/Q1	17/Q3	17/Q4	18/Q2	18/Q3	18/Q4	19/Q2
16/Q3	4.6	1.8	3.3	4.3	0.1	3.8	6.6	2.9
16/Q4	—	5.3	6.2	7.0	3.9	2.1	9.5	6.1
17/Q1	—	—	2.3	3.6	2.6	8.4	6.3	1.6
17/Q3	—	—	—	1.7	6.9	15.8	4.9	1.2
17/Q4	—	—	—	—	9.8	20.3	3.1	3.3
18/Q2	—	—	—	—	—	4.1	7.5	3.3
18/Q3	—	—	—	—	—	—	6.5	3.8
18/Q4	—	—	—	—	—	—	—	6.2
全期間平均	5.3							

モデルチェンジ 前\後	従来手法							
	16/Q4	17/Q1	17/Q3	17/Q4	18/Q2	18/Q3	18/Q4	19/Q2
16/Q3	4.6	8.3	9.2	7.8	6.8	1.2	6.8	20.4
16/Q4	—	11.6	11.9	10.7	2.5	6.2	9.8	23.6
17/Q1	—	—	2.4	0.7	18.8	12.0	1.0	12.7
17/Q3	—	—	—	1.9	25.6	18.5	4.1	11.2
17/Q4	—	—	—	—	22.8	15.2	2.0	13.6
18/Q2	—	—	—	—	—	9.9	12.6	27.8
18/Q3	—	—	—	—	—	—	4.9	16.1
18/Q4	—	—	—	—	—	—	—	15.3
全期間平均	10.8							

備考：ある四半期に発売開始された製品の平均的なスペックを有する仮想的な製品から、別の四半期に対して同様に構築した仮想的な製品へのモデルチェンジが発生したと仮定。その際の品質変化率（理論価格の変動率）を現サンプル・旧サンプルから推計された回帰式で算出し、その差分の絶対値を掲載。

合であっても、モデルの示唆する品質向上率の誤差が、比較的小幅にとどまる傾向を示唆している。

また、品質調整の実施に当たって、品質向上ないし劣化の定性的な方向感を正しく捉えられることは、何よりも優先される。この点に関し、図表 A-4 をみると、従来の手法では、丸で囲んでいるように推計期間の変化によって品質変化率の正負がはっきりと異なるモデルチェンジ事例が確認される。このことは、時間の経過とともに回帰式が陳腐化すると、実際には品質向上（ないし劣化）しているにもかかわらず、品質劣化（ないし向上）と誤って評価した品質調整を行ってしまう可能性を示唆している。他方で、新手法では、品質変化率の正負が新旧回帰式間で逆転する事例は、ごくわずかにとどまっている。AEN の導入によるこうした面での改善も、ヘドニック法の適用可能性の向上に繋がりをうる。