

博士学位請求論文審査報告書

Essays on Unobserved Heterogeneity and Endogeneity in Health Econometrics

増原宏明氏の博士学位請求論文”Essays on unobserved heterogeneity and endogeneity in health econometrics”は医療経済学に特有のデータ問題への新たなマイクロ計量経済学の分析手法の構築を貢献とする。医療経済学で用いられるデータは、実験データではなく観察されたデータであるため、誤差が正規分布に従わず、サンプルセレクションの問題が常に存在する。さらに量的な変数よりも質的変数も多いために、計量分析の手法には工夫が必要とされる。

本論文は2部から構成される。第1部は観察不可能な異質性を取り上げる。最初にデータが観測値で異なる状況が生存期間分析やカウントデータ分析が非線形回帰分析においてもたらず問題を分析している。線形回帰モデルの場合、観察されない異質性が説明変数と独立であると最小二乗推定量は必ずしも効率的ではないが一致性を持つ。ところが非線形回帰モデルでこうした異質性を無視することで、例えば生存期間における負（正）の見せかけの時間依存性や、カウントデータにおける過大（過小）な分散をもたらす。しかし、生存期間分析では観察不可能な異質性と時間依存性を識別することが難しい。このため、見せかけの時間依存性という問題が生じて推定結果が誤る可能性を排除できない。この問題に対して、本論文ではエルミート多項式とボックスコックス変換によって一般化した対数線形生存期間モデルを提案している。対数線形の生存期間分析をベースに非説明変数をボックスコックス変換で一般化し、正規分布に従う誤差項をエルミートの直交多項式を用いて一般化したセミパラメトリックな分布で置き換えたものである。この分析手法の有効性を検証する。具体的には米国のGeneral Social Surveyを用い、第1子を産んだ年齢を被説明変数である生存期間とし、教育年数、兄弟姉妹の人数などを説明変数として回帰した。対数線形モデル、ガンマ分布に従う誤差項をもつ対数線形モデル、セミパラメトリックな分布に従う誤差項をもつ対数線形モデル、セミパラメトリックな分布に従う誤差項をもちボックスコックス変換された対数線形モデルの4つを比較している。推定結果としては、4つのモデルは互いによく似た値となり、有意水準もほぼ同じであるが、いくつかの変数では差異が生じた。情報量基準と非入れ子型の検定統計量の結果が良好であったセミパラメトリックな分布に従う誤差項をもつ対数線形モデルは絶対値が小さくなる。他のモデルで過大に推定されていることが示唆された。

続いて非連続型な異質性に関する分析を行っている。有限混合モデルにおいて被説明変

数が非連続変数のときに、要素密度の識別不可能性問題が発生するという問題が生じる。とりわけ、この状況は被説明変数が 2 値の横断面データの場合に発生して、3 期間以上のパネルデータでなければ、少なくとも要素密度が 2 個の有限混合モデルは推定できない。医療経済学で用いられるデータには、しばしば横断面データで、かつ 2 値データを非説明変数に使わざるをえないことも多い。この場合には、有限混合モデルを用いての推定が不可能となり、プロビットモデルやロジットモデルを使用することとなるが、これらが有限混合モデルを統計的に優越するかを確認できず、推定結果が頑健とはいえない。そこで 2 値の横断面データにいかなる条件を加えれば、有限混合モデルが識別可能となるのかを検証した。その結果、2 値の被説明変数以外に、ある連続の被説明変数が存在すれば、有限混合モデルが識別可能となることが、積率母関数を用いて証明された。医療経済学で用いられる横断面のデータであっても、2 値データ以外にある連続のデータが存在すれば、有限混合モデルの推定が可能となる。

第 2 部では、医療経済学の観点から内生性の問題をレビューした上で説明変数が内生変数である場合の推定方法を説明している。内生変数が存在と推定量が一致性を有しない。実験計画法のように、ランダムに割当られるか、割当に個人の意思がまったく介在しないならば、医療経済学において内生性の問題は生じない。しかし、これらの条件を満たすことはほとんどなく、内生性の問題を避けることはできない。本論文では、被説明変数が連続変数と非連続変数に分け、内生性をもつ説明変数が連続変数と非連続変数の場合の、一致性のある推定方法を述べている。最初に、被説明変数が連続変数の場合を取り上げ、内生変数があっても操作変数法が一致推定量を与え、応用分野で用いられる 2 段階推定法の説明している。内生変数が非連続である場合、中でも代表的なものとして 2 値の場合の処置効果であっても、操作変数法と 2 段階推定法が適用可能であることが説明された。次に、プロビットモデルとカウントデータモデルを例に挙げて、被説明変数が非連続変数であるケースにおける内生変数の議論を行った。内生変数が非連続（切断変数、打ち切り変数、離散変数）の場合には操作変数法と 2 段階推定法は一致推定料をもたらさず、完全情報最尤法のみが一致推定を可能とすることが明らかにされた。

続いて 2 値の内生変数をもつカウントデータモデルの頑健な推定方法を提案している。完全情報最尤法は分布を特定化するので、常に特定化の誤りという短所を抱えることになる。そこでセミパラメトリックでかつより簡便な、2 値内生変数をもつカウントデータモデルの推定方法を提案した。具体的には、カウントデータの誤差項と、2 値内生変数の誤差項が、2 変量の正規分布に従うモデルを考え、この同時確率密度関数を、エルミートの直交多項式で一般化した。すなわち、一般化された同時確率密度関数は、2 変量の正規分布をその特殊系として持つこととなる。この一般化された分布を、2 値内生変数をもつカウントデータモデルに適用した。具体的には米国の National Health Interview Survey を用いて、過去 2 週

間にアルコール飲料を飲んだ回数を被説明変数に、医師にアルコール習慣を減らすための診療の有無を 2 値内生変数にして、所得、年齢、教育、人種、日常生活活動指標等で回帰した。情報量基準、尤度比検定でもセミパラメトリックモデルが優越し、パラメトリックモデルは、平均処置効果を過大に推定したことが明らかとなった。本論文で議論した方法は、数値積分も比較的容易であり、単純ながらも頑健で、強力な推定法であった。

2014年3月20日に行われた口頭審問においては審査委員から次のような質問や示唆があった。第1にセミパラメトリックモデルで、たとえばエルミートの多項式で一般化された正規分布が、特定の分布を近似できるかも例が欲しいとの指摘があった。改訂版では第4章において Examples of Microdata in Health Economics の節を挿入してイギリスの Health and Lifestyle Survey と British Household Panel Survey, アメリカの Medical Expenditures Panel Survey を例に挙げ現実のデータ性質（問紙調査であるため離散変数が多く、連続少ないこと）内生性の可能について議論を行って離散変数が多く、連続少ないこと）内生性の可能について議論を行った。第2に本論文の貢献をより明確にして欲しいとの意見もあった。例えば手法の範囲が拡大することによる実証研究上メリット、例えば心臓血管系の入院日数が長い疾患だけでなく、白内障のように入院期間の短い疾患についても正確な分析が可能になった、ガウス分布を用いた分析の妥当性が高いことが傍証できたなどが指摘された。改訂版においては博士論文の貢献についてもより明確な整理がなされている。

本博士学位請求論文の主要な貢献は以下のとおりまとめられる。第1に観察不可能な異質性の真の分布は一般的に未知であるので、分布を一般化することは一致推定量をもたらす可能性を大幅に引き上げる。医療経済学で頻繁に用いられる非線形回帰モデルでは、観察不可能な異質性が存在するために、観測される、正もしくは負の関係は見せかけのものとなるかもしれない。一致推定量を得るための唯一の手段は、観察不可能な異質性を一般化するセミパラメトリックモデルを、係数の推定に用いることであり、本博士学位請求論文では一般化された方法を提案した。第2に一般化された観察不可能な異質性を用いることで、医療経済学における平均処置効果が正しく推定され、検定可能なものとなることである。一般的に、大規模の横断面もしくはパネルデータが、健康へのある処置効果を計測するさいに用いられる。医療経済学で用いられるこれらのデータは多くの非連続（切断変数、打ち切り変数、離散変数）が含まれ、連続変数は少ない。第3に、これらのデータにおいて、ある変数が外生変数であると仮定するのは困難であるので、潜在的な内生性の問題も孕む。したがって、医療経済学において平均処置効果を測定するためには、離散的な内生変数を持つ非線形回帰モデルを使うことは避けられない。本博士学位請求論文は、平均処置効果を推定し検定可能な、頑健でより正確な推定方法を示している。

以上から審査員一同は増原氏の論文に対して、一橋大学博士学位（経済学）を授与す

るに十分に値すると判断した。

2014年7月9日

審査員

佐藤主光

井伊雅子

小塩隆士

河口洋行

縄田和満