

2025 年 7 月 31 日

切断回帰モデルによる  
専攻賃金プレミアムの再検証

源内雄貴

## 要旨

本研究は、東京大学「大学教育についての職業人調査」(2022) データに内在する**就業者のみの標本の偏り**が、大卒賃金プレミアムの推定に与える影響を分析する。この偏りにより、特に女性においてその影響が大きいと懸念される賃金プレミアムの過大推定を補正することが本研究の目的である。先行研究における区間回帰モデルの厳密性を踏まえ、本研究では「切断回帰」の概念を取り入れた**切断区間回帰モデル**を独自に構築し、その妥当性を OLS および既存の切断回帰モデルと比較検証した。分析は主に男性データを用いて行った。しかし、期待したほどの大きな補正効果は確認されず、推定値や有意性に目立った差異は見られなかった。この結果から、なぜ補正効果が限定的だったのかを考察し、今後の研究課題とする。

## 第1章

本稿は、大学教育が個人の賃金に与える影響、特に「大卒賃金プレミアム」の計測における標本選択バイアスの問題に取り組むものである。既存の調査データには、就業者のみが標本に含まれるという構造的な偏りが存在し、これが推定結果にバイアスをもたらす可能性がある。特に、労働市場における女性の参加特性を考慮すると、この偏りの影響は看過できない。本研究では、この問題に対し計量経済学的な手法を用いて補正を試み、その効果を検証する。

本稿の構成は以下の通りである。まず、第2章で先行研究を概観し、本研究との関係を説明する。第3章では、使用データと変数の作成方法について詳述する。第4章では、分析に用いる推定手法、特に独自に構築した切断区間回帰モデルについて説明する。第5章では分析結果を提示し、その解釈と直面した課題について考察する。最後に、第6章で結論を述べ、今後の研究の方向性について議論する。

## 第2章

大藤修史・荒井洋一(2022)「専攻および日本特有の属性変数による賃金プレミアムの分析」は、東京大学大学経営・政策研究センターによる「大学教育についての職業人調査」データを用いた研究である。彼らは、ミンサー型賃金関数を基礎とし、大学の学部・学校区分、学歴、経験年数などが個人の賃金に与える影響を分析した。その主要な主張は、高学歴であるほど賃金プレミアムが大きいことや、経験年数と賃金に非線形の関係があることなど、人的資本理論に整合的な知見を導き出した点にある。

彼らの研究は、分析手法の厳密性において特に注目される。被説明変数である年収が連続的な数値ではなく、**区間カテゴリ変数**として与えられているデータ特性に対し、通常の OLS

ではなく**区間回帰モデル**を適用した点である。これは、データの情報損失を最小限に抑え、より正確な推定を可能にする統計的に厳密なアプローチであると評価できる。賃金を対数変換する際に、賃金の最小カテゴリの下限が 0 万円であると負の無限大に発散するという問題に対し、独自に賃金の下限を 72 万円とすることで対処している。

本研究は、大藤らの研究のこの厳密なアプローチを高く評価し、その成果を土台としている。本研究はさらに、調査データに見られる「就業者のみが標本に含まれる」という**本質的な標本の偏り**に着目する。この偏りは、賃金が 0 の非就業者層がデータに含まれないことに起因し、先行研究の推定結果にバイアスをもたらす可能性がある。特に、女性の労働参加率の特性を考慮すると、このバイアスは重要である。したがって、本研究は、先行研究の区間回帰アプローチを踏襲しつつ、この「働かない選択」による標本の切断を明示的にモデルに組み込むことで、より厳密な賃金プレミアムの再評価を試みるものである。

### 第 3 章

本研究では、東京大学大学経営・政策研究センターが 2009 年に実施した「大学教育についての職業人調査」の個票データを使用する。この調査は、日本の大学を卒業し、現在も職業に就いている者を対象として実施されたものである。回答者は、日本各地の産業・企業に所属する職業人であり、多岐にわたる学歴や職務経験を持つ。

本調査は、卒業から 10 年・20 年が経過した時点での大学教育の成果を追跡するために、全国の主要大学から卒業生名簿を得て、無作為抽出した者を対象としている。

分析に用いるデータについて、川口 (2011) の指摘に基づき、サンプルを 59 歳以下に絞っている。これは、日本では定年を迎える 60 歳時点で平均賃金に大幅な下落があり、この事実を無視した賃金関数の推定にはゆがみが生じる可能性があるためである。また、大藤ら (2022) に倣い、標本を新卒者飲みに制限した。また、専攻区分・大学区分・学歴について、ダミー変数を用いた。

最終的な分析対象は男性データのみとし、有効サンプルサイズは 8599 である。

### 第 4 章

本研究では、以下の 3 種類のモデルを推定した。

#### 4.1. 通常の最小二乗法 (OLS)

基準モデルとして、通常の最小二乗法 (OLS) を用いる。OLS は線形回帰モデルの標準的な推定手法であり、他モデルとの結果比較の基準とする。推定式は以下で与えられる。

$$\ln(Y_i) = \beta_0 + \sum_{k=1}^K \beta_k X_{ik} + \varepsilon_i$$

ここで、 $Y_i$ はi番目の個人の年収、 $X_{ik}$ はk番目の説明変数、 $\beta_0$ は切片、 $\beta_k$ は回帰係数、 $\varepsilon_i$ は誤差項である。

#### 4.2. 切断回帰モデル (Truncated Regression Model)

本研究で取り組む主要な課題である標本の偏り、特に年収が特定の下限值以下の個人が標本に存在しないという「切断 (Truncation)」の問題を補正するために切断回帰モデルを用いる。このモデルは、潜在的な年収 $Y_i^*$ が正規分布に従うと仮定し、観測されるのは $Y_i^* > c$ となる場合のみである。推定は最尤法によって行われる

$$\ln(Y_i) = \beta^0 + \sum_{k=1}^K \beta_k X_{ik} + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$$

尤度関数は以下の通りである。

$$L(\beta, \sigma | \ln(Y), X, \ln(c)) = \prod_{i=1}^N \frac{\phi((\ln(Y_i) - X_i\beta)/\sigma)}{\sigma \times (1 - \Phi((\ln(c) - X_i\beta)/\sigma))}$$

ここで、 $\phi(\cdot)$ は標準正規分布の確率密度関数、 $\Phi(\cdot)$ は標準正規分布の累積分布関数である。切断点  $c$  は、年収の最下位階級 (72 万円～300 万円未満) の中央値である 186 万円 (対数スケール) とした。

#### 4.3. 切断区間回帰モデル (Truncated Interval Regression Model)

本研究で独自に構築したモデルであり、年収が区間データとして与えられていることと、標本が切断されていることの両方を同時に考慮する。潜在的な年収 $Y_i^*$ は正規分布に従うと仮定し、観測されるのはそれが区間 $[L_i, U_i]$ 内に存在し、かつ切断点  $c$  を超える場合のみである。推定は最尤法によって行われる。各観測  $i$  の尤度への寄与は、以下で与えられる。

$$P(L_i \leq Y_i^* \leq U_i | Y_i^* > c) = \frac{\Phi(\frac{U_i - X_i\beta}{\sigma}) - \Phi(\frac{L_i - X_i\beta}{\sigma})}{1 - \Phi(\frac{c - X_i\beta}{\sigma})}$$

全体としての対数尤度関数は以下の通りである。

$$\begin{aligned} \log L(\beta, \sigma | \{(\ln(L_i), \ln(U_i), X_i)\}_{i=1}^N, \ln(c)) \\ = \sum_{i=1}^N \log \left( \Phi\left(\frac{U_i - X_i\beta}{\sigma}\right) - \Phi\left(\frac{L_i - X_i\beta}{\sigma}\right) \right) - \log \left( 1 - \Phi\left(\frac{c - X_i\beta}{\sigma}\right) \right) \end{aligned}$$

この尤度関数は統計ソフトウェア R を用いて独自にコーディングし、optim 関数により最尤推定を行った。切断点  $c$  は、年収の最下位階級 (72 万円～300 万円未満) の中央値である 186 万円 (対数スケール) とした。

## 第 5 章

本章では、前章で述べた 3 つのモデルを用いて男性データに対する推定結果を示す。各モデルの推定値を表 1 にまとめた。

変数名	推定値 (OLS)	推定値 (切断)	推定値 (切断区間)
切片	5.504***	5.504***	5.397***
心理・社会	-0.068**	-0.068**	-0.083**
法・政治	0.055**	0.055**	0.056**
経済	-0.002	-0.002	-0.004
経営	0.006	0.006	0.005
教育	-0.053*	-0.053*	-0.064
理学	0.004	0.004	0.000
工学	-0.001	-0.001	-0.004
農学	-0.066**	-0.066**	-0.071**
薬学	0.181***	0.181***	0.196**
保健・福祉	-0.120**	-0.120**	-0.158*
その他学部	-0.025	-0.025	-0.035
国立	0.104***	0.104***	0.110***
公立	0.064**	0.064**	0.068*
修士	0.181***	0.181***	0.205***
博士	0.427***	0.427***	0.448***
その他学歴	-0.066**	-0.066**	-0.067*
経験年数	0.070***	0.070***	0.080***
経験年数 <sup>2</sup>	-0.001***	-0.001***	-0.001***
Sigma	---	0.323***	0.368***

表 1

表 1 に示されるように、各モデルの推定結果を比較したところ、主要な独立変数の推定値およびその統計的有意性に、細かな変化が確認された。OLS と比較して、切断回帰モデルや、独自に構築した切断区間回帰モデルにおいて、係数推定値にごくわずかながらも変化が見られ、特定の変数では有意性の傾向にも変化が生じるケースがあった。

修士や博士といった学歴は、どのモデルにおいても統計的に有意な正の効果を示しており、ミンサー型賃金関数に整合的な結果である。切断区間回帰では、OLS と比較してわずかに係数が上昇する傾向が見られた。

経験年数もどのモデルでも有意な正の効果を示し、経験年数二乗項の負の符号により、経験年数が進むにつれて賃金の伸びが鈍化するという非線形な関係が確認された。これもミソサー型賃金関数に整合的である。

一方で、当初、より顕著な変化を期待していたのに対し、その度合いは予想ほど大きくなかった。特に、女性データについては、ヘシアン行列の収束問題により信頼できる推定値が得られず、他モデルとの比較はできていない。

## 第6章

本研究は、東京大学「大学教育についての職業人調査」データにおける「就業者のみ」という標本の偏りに着目し、大卒賃金プレミアムの推定に対するその影響を検証した。先行研究の区間回帰モデルの厳密性を踏まえ、本研究では新たに切断の概念を導入した切断区間回帰モデルを独自に構築し、最尤法で推定を行った。

分析の結果、OLS、既存の切断回帰モデルと比較して、主要な独立変数の推定値に微細な変化が確認されたものの、当初期待したほどの大きな補正効果は得られなかった。この、わずかな変化がなぜ限定的だったのか、そしてなぜ女性データの切断区間回帰モデルにおける推定が行えなかったことが、今後の大きな課題として残された。

働いていない人に関する偏りの補正には、今回用いた切断という操作だけでは限定的であった可能性がある。より複雑な「選択メカニズム」をモデル化できる標本選択モデルの適用を検討し、より厳密な推定を目指す必要がある。

また、女性データで直面しているヘシアン行列の収束問題については、尤度関数の数値的な安定性の再検証、初期値設定のさらなる最適化、あるいは他の最適化アルゴリズムや専門パッケージの検討が必要である。

これらの課題に取り組み、本研究で得られた「わずかな変化」の意味を深掘りし、標本の偏りが賃金プレミアムの推定に与える影響について、より強固な知見を確立することが今後の研究の方向性となる。

## 参考文献

- 大藤・荒井 (2022)「専攻および日本特有の属性変数による 賃金プレミアムの分析」『日本労働研究雑誌』 64(9), 79-98.
- 安井健悟. (2019). 大学と大学院の専攻の賃金プレミアム. 『経済分析/内閣府経済社会総合研究所 編』, (199), 42-67.
- 川口大司. (2011). ミンサー型賃金関数の日本の労働市場への適用. 『現代経済学の潮流』, 67-98.