

令和 3 年 6 月 11 日現在

機関番号：12102

研究種目：若手研究

研究期間：2019～2020

課題番号：19K13666

研究課題名(和文) A sample selection model with a monotone selection correction function

研究課題名(英文) A sample selection model with a monotone selection correction function

研究代表者

YU ZHENGFEI (YU, ZHENGFEI)

筑波大学・人文社会系・助教

研究者番号：40774758

交付決定額(研究期間全体)：(直接経費) 1,100,000円

研究成果の概要(和文)：本研究では、パラメトリックで単調なコントロール関数を仮定する有名なヘックマン選択モデルに触発され、コントロール関数の単調性を維持しつつ、潜在的な誤差にパラメトリックな分布の仮定を課さない選択モデルを研究する。コントロール関数の単調性を維持するために、潜在誤差のコピュラ関数に十分条件を見出す。この単調性を利用して、パラメータチューニングを必要としないセミパラメトリック推定法を提案し、推定値について、一貫性と漸近正規性を確立する。提案した手法の有用性を示すために、シミュレーションと実証分析を行う。この単調性制限付き推定法は、順序反応モデルなど、他のセミパラメトリックモデルにも適用可能である。

研究成果の学術的意義や社会的意義

This project shows that the monotonicity of the control function implied by the celebrated Heckman selection model is shared by a much larger family without parametric assumptions. It proposes a more convenient semiparametric estimation method for the sample selection model.

研究成果の概要(英文)：Motivated by the celebrated Heckman selection model which implies a parametric and monotone selection function, this project studies a sample selection model that does not impose parametric distributional assumptions on the latent errors, while maintaining the monotonicity of the control function. It shows that a positive dependence condition on the latent errors is sufficient for the monotonicity. The condition is equivalent to a restriction on the copula function of latent error terms. Using the monotonicity, this project proposes a tuning-parameter-free semiparametric estimation method and establishes root  $n$ -consistency and asymptotic normality for the estimates of finite-dimensional parameters. Simulations and an empirical application are conducted to illustrate the usefulness of the proposed methods. The shape-restricted estimation methods are also applicable to other semiparametric models including the ordered response model and accelerated failure time model.

研究分野：計量経済学

キーワード：Shape restriction Isotonic regression Isotonic regression Tuning parameter free Sample selection model Ordered response model Accelerated failure time

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

#### 1. 研究開始当初の背景

The sample selection problem arises frequently in economics when observations are not taken from a random sample of the population. Understanding the self-selection process and correcting selection bias is a central task in empirical studies of the labor supply behavior of females, the determinants of schooling choices, unionism status, and migration decisions, among others. The celebrated Heckman selection model yields a selection correction function (control function) that is parametric and monotone. When the joint distribution of error terms in the selection and outcome equations is unknown, the control function becomes nonparametric. This project studies a sample selection model that does not impose parametric distributional assumptions on the latent error terms, while maintaining the monotonicity of the control function. Using the monotonicity, this project develops a new semiparametric estimation method of the sample selection model that does not rely on parametric assumptions, and at the same time does not require tuning parameters such as the bandwidth.

#### 2. 研究の目的

- (1). To find a sufficient condition for the monotone control function, which is related to an intuitive dependence concept of two latent errors.
- (2). To propose a new semiparametric estimation method of the sample selection model that incorporates the monotonicity of the control function and thus frees practitioners from specifying tuning parameters.
- (3). To investigate the asymptotic properties of the proposed estimator.
- (4). To apply the monotonicity-restricted estimation method to other models including the ordered response model and the accelerated failure time model.

#### 3. 研究の方法

- (1). To characterize the sufficient conditions for the monotonicity of the control function, this project utilizes a concept of positive (negative) dependence in the probability literature: the right tail increasing (decreasing) (Esary and Proschan, 1972). Then using the copula theory, this project shows that the monotonicity of the control function in the Heckman model is shared by a much larger family without requiring any parametric assumptions.
- (2). To incorporate the monotonicity restriction into the estimation, this project uses the nonparametric maximum likelihood estimator (NPMLE, Groeneboom and Hendrick, 2018) in the first stage and then uses the isotonic regression (Huang, 2002) in the second stage.
- (3). To establish the convergence rate and the asymptotic normality of the proposed estimator, this project relies on the empirical process theory and the characterization of isotonic regression.

#### 4. 研究成果

- (1). This project shows that a specific positive (negative) dependence condition between the two latent error terms (in the selection and outcome equations, respectively), i.e., the right tail increasing (RTI) (or right tail decreasing), is sufficient to induce a monotone control function. This condition only depends on the copula function without restricting the marginal distributions of the latent errors. This project presents several parameterized joint distributions

of latent errors that yield monotone control functions, and it also discusses the economic meaning of RTI by comparing it to another popular measure of positive dependence: stochastic increasing.

(2). This project proposes a new semiparametric estimator (MCF hereafter) for the sample selection model that incorporates the monotonicity of the control function and circumvents the choice of user-specified tuning parameters. The proposed estimation method contains two stages: in Stage 1, the estimator first computes the NPMLLE for the distribution function of the error term in the selection equation for any fixed value of the coefficients in the selection equation, and then estimates these coefficients via moment conditions; in Stage 2, the coefficients in the outcome equation are estimated via a partial linear model where the nonparametric component part is estimated by the isotonic regression. The algorithm can be implemented conveniently in R.

(3). Although the NPMLLE and the isotonic regression for the nonparametric components are known to have a slow convergence rate, this project shows that the proposed estimator for the coefficients in the outcome and selection equations keeps the desirable root-n convergence rate. In addition, the asymptotic normality and bootstrap consistency are also established.

(4). Free from the choice of bandwidth, the Monte Carlo simulation study shows that the finite sample performance of proposed MCF estimator is competitive relative to the kernel estimator using different bandwidth selectors. Table 1, which presents a part of the simulation results, compares the root mean square error (RMSE) of the proposed estimator (MCF) to the parametric Heckman's estimator and kernel estimators for the outcome equation's coefficient.

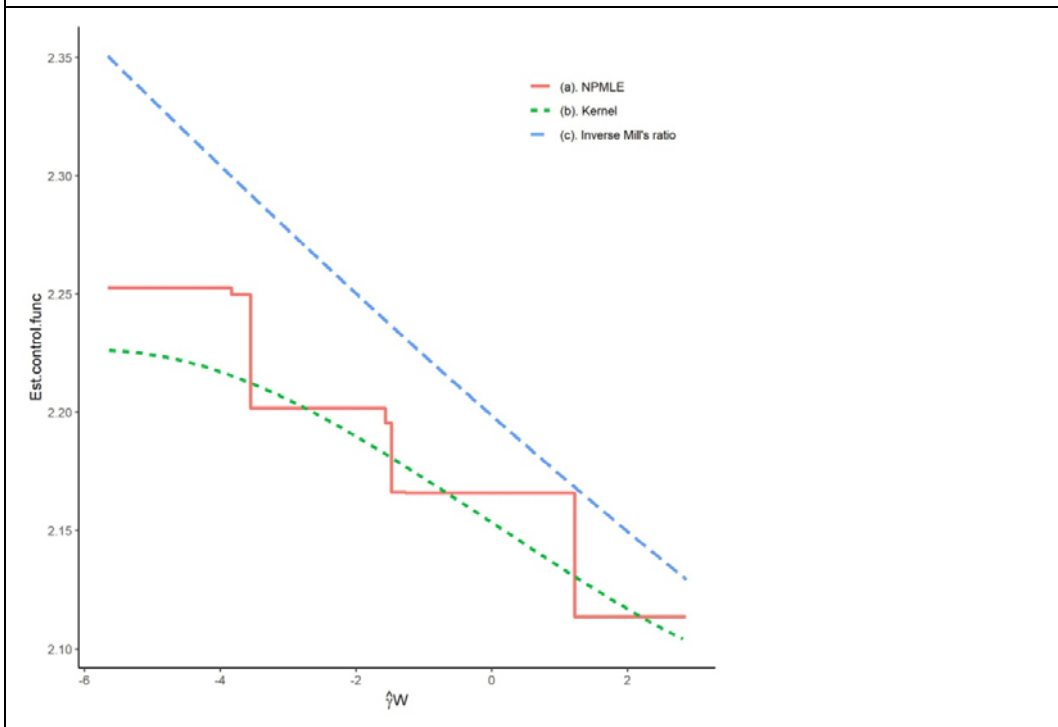
Table 1: Root mean square error of the Heckman's estimator, the proposed estimator (MCF) and the kernel estimator (with different bandwidth selectors) for the single coefficient in the outcome equation. CV=cross-validation bandwidth selector, PI= plug-in bandwidth, 2ord=Linton (1995)'s optimal bandwidth based on the second order approximation, ROT=rule of thumb bandwidth, N= sample size.

| N    | Heckman | MCF   | Kernel-based (Klein-Spady + Robinson) |        |          |         |
|------|---------|-------|---------------------------------------|--------|----------|---------|
|      |         |       | CV, CV                                | CV, PI | CV, 2ord | ROT,ROT |
| 1000 | .1892   | .1531 | .1706                                 | .1646  | .1642    | .1696   |
| 2000 | .1467   | .1073 | .1131                                 | .1120  | .1132    | .1215   |
| 5000 | .1002   | .0633 | .0658                                 | .0669  | .0652    | .0675   |

(5). The proposed estimation method is applied to estimate female wage equation, using the Merged Outgoing Rotation Groups (MORG) of the US's current population survey for the year 2013. In this example, the control function will be decreasing if there is a strong enough positive dependence between the unobserved productivity in the wage equation and the working incentive in the labor market participation equation. In addition, such a monotonicity restriction is also supported by the kernel estimate of the control function in Figure 1.#

Fig-1. The estimated control functions for the wage equation of married women in US, MORG 2013. Solid line = NPMLLE estimate of the control function, dotted line = kernel

estimate, dashed line = Heckman's model



(6). This project also proposes monotonicity-restricted estimation methods for other semiparametric models including the ordered response model and accelerated failure time model. In the former case, there is an inherent monotonicity restriction on the distribution function of the error term. In the latter case, the project shows that the monotonicity of the hazard function of the error term is implied by the log-concavity of its survival function, which is in turn satisfied by a large family of commonly used parametric duration models. By incorporating the monotonicity restriction into the estimation method, the proposed estimator for the ordered response model is free from any tuning parameters, and the estimating equations for weighted rank estimators for the accelerated failure time model is guaranteed to have a unique global solution.

<引用文献>

- ① Esary, J. D., & Proschan, F. (1972). Relationships among some concepts of bivariate dependence. *The Annals of Mathematical Statistics*, 43, 651-655.
- ② Groeneboom, P., & Hendrickx, K. (2018). Current status linear regression. *Annals of Statistics*, 46, 1415-1444.
- ③ Huang, J. (2002). A note on estimating a partly linear model under monotonicity constraints. *Journal of Statistical Planning and Inference*, 107, 343-351.
- ④ Linton, O. (1995). Second order approximation in the partially linear regression model. *Econometrica*, 63, 1079-1112.

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 2件 / うちオープンアクセス 0件）

|  |                         |
|--|-------------------------|
| 1. 著者名<br>Liu Ruixuan, Yu Zhengfei                                 | 4. 巻<br>23              |
| 2. 論文標題<br>Accelerated failure time models with log-concave errors | 5. 発行年<br>2020年         |
| 3. 雑誌名<br>The Econometrics Journal                                 | 6. 最初と最後の頁<br>251 ~ 268 |
| 掲載論文のDOI (デジタルオブジェクト識別子)<br>10.1093/ectj/utz024                    | 査読の有無<br>有              |
| オープンアクセス<br>オープンアクセスではない、又はオープンアクセスが困難                             | 国際共著<br>該当する            |

|  |                 |
|--|-----------------|
| 1. 著者名<br>Liu Ruixuan, Yu Zhengfei                                 | 4. 巻<br>-       |
| 2. 論文標題<br>Sample selection models with monotone control functions | 5. 発行年<br>2021年 |
| 3. 雑誌名<br>Journal of Econometrics                                  | 6. 最初と最後の頁<br>- |
| 掲載論文のDOI (デジタルオブジェクト識別子)<br>10.1016/j.jeconom.2021.01.010          | 査読の有無<br>有      |
| オープンアクセス<br>オープンアクセスではない、又はオープンアクセスが困難                             | 国際共著<br>該当する    |

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 1件）

|   |
|---|
| 1. 発表者名<br>Zhengfei Yu  |
| 2. 発表標題<br>Simple Semiparametric Estimation of Ordered Response Models                    |
| 3. 学会等名<br>The 15th International Symposium on Econometric Theory and Applications (国際学会) |
| 4. 発表年<br>2019年   |

|  |
|--|
| 1. 発表者名<br>Zhengfei Yu   |
| 2. 発表標題<br>A Nonparametric Test for Index Sufficiency in the Selection Model |
| 3. 学会等名<br>Workshop on Advances in Econometrics 2019                         |
| 4. 発表年<br>2019年  |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

|  | 氏名<br>(ローマ字氏名)<br>(研究者番号) | 所属研究機関・部局・職<br>(機関番号) | 備考 |
|--|---------------------------|-----------------------|----|
|--|---------------------------|-----------------------|----|

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

| 共同研究相手国 | 相手方研究機関 |
|---------|---------|
|---------|---------|