

# Propensity Score Matching in Accounting Research

Shipman, Jonathan E.      Swanquist, Quinn T.      Whited, Robert L.

*The Accounting Review* (2017), 92 (1), pp. 213–244

## 概要

会計研究において、傾向スコア・マッチング (propensity score matching, PSM) が平均処置効果 (ATEs) を推定するための一般的な手法として利用されるようになった。本論文では、PSM の有用性と限界について、伝統的な重回帰 (MR) 分析と比較して議論する。さまざまな PSM の設定 (design choice) について検討し、2008～2014 年に主たる会計雑誌に掲載された、PSM を採用している論文 86 本をレビューする。2008 年には、PSM を採用している論文数は 0 本であったのに対し、2014 年では 26 本と、顕著な増加が確認できる。しかしながら、論文によっては PSM の特性を過大評価しすぎてしまい、重要な設定およびまたは理論的に一貫した手続きに則って PSM を実施することを明記していない。そこで、はじめに、会計研究における 3 つの例から、PSM の複雑性について実証的に明らかにする。まず、処置群を表す変数 (treatment) がバイナリー変数でない場合、PSM を利用することで、効果量が最小限になるようなサブサンプルを対象とした分析になってしまうという例を示す。また、一見問題ないように思われる設定が、サンプルの構成 (sample composition) および ATE の推定に深刻な影響を及ぼすという例も提示する。さいごに、マッチング手法の利用を検討している将来の研究に対して、いくつかの示唆を提供する。

## 1 Introduction

### ■研究の背景: 内生性の問題に対する従来の手法の問題点

- 実証的会計研究において、因果処置効果 (causal treatment effects) を推定することが、主たる目的とされることがある (Gow, Larcker, and Reiss, 2016)。
- 非実験的データを利用した研究では、処置群が無作為割り当てではない (non-random treatment assignment) ために、内生性の問題が生じてしまう。
- 内生性を緩和させるための従来の手法
  - アーカイバル研究では、伝統的には重回帰 (multiple regression, MR) モデルを利用
  - ただし、MR を利用してバイアスのかかっていない推定値を得るためには、結果変数 (Y) と説明変数 (X) の関係に適切な仕様 (前提) を満たす必要がある。
  - Y と X の関係が前提を満たしていない場合、MR は“回帰式の特定ミス (functional form misspecification, FFM)”の影響を受けることになり、バイアスのかかった推定値を得ることになり得る。

- FFM による潜在的なバイアスは、処置群 (treatment groups) が似ていないほど、強くなってしまう。

#### ■Propensity score matching (PSM)

- 変数間の特定の仮定を少なくすることで、上記のような問題を緩和
- 処置群から選択された観測値と処置を受ける傾向の推定値を利用して、機械的に、様々な側面から判断したコントロール・グループをマッチ
- 変数間の関係 (functional relation) に関する緩い (relaxed) 仮定のもとで、処置効果を直接的かつ直感的に推定可能
- ただし、PSM は伝統的な MR の手法が持つ理論的有用性を損なう。

#### ■本論文の検討事項

- 内生性、MR の手法、FFM に関する問題、マッチング手法のメリット、PSM の設定のインプリケーションについて議論
- 以下の雑誌に掲載されている、PSM を利用した 86 本の論文に対する議論
  - *The Accounting Review, Contemporary Accounting Research, Journal of Accounting and Economics, Journal of Accountitng Reserach, Review of Accounting Studies*
  - PSM を利用した論文数は増加 (2008 年: 0 本、2014 年: 26 本)
  - 各論文について、PSM の利用および設定が正当なものであるのかを評価
- 財務報告に関する研究における 3 つの設定において、PSM を利用した場合に問題が生じる例
  1. 監査人の規模
  2. 内部統制の弱さ
  3. フォローしているアナリスト数
- マッチング手法の利用もしくは、FFM に基づく内生性問題に取り組む他の手法の利用を検討している研究に対するいくつかの提言

#### ■本論文の貢献

- DeFond, Erkens and Zhang (2015) を補完
  - DeFond, Erkens and Zhang (2015)
    - \* マッチング手法を利用して、監査人の規模および監査の質の関係について分析
    - \* 設定 (design) をランダムに数千回実行し、総合的な結果を提示
    - \* 分析結果の多くは、4 大監査法人は、それ以外の監査法人よりも優れた監査の質を提供していることを示唆
    - \* Lawrence, Minutti-Meza and Zhang (2011) の結果と整合

– 本論文と DeFond et al. (2015) の違い

- \* DeFond et al. (2015) の主たる目的は、4 大監査法人の影響に関する実証的証拠を提供すること
  - \* 本研究も同様のテーマを共有
  - \* ただし、会計研究において PSM を利用している論文をレビューし、PSM の有用性および限界に関する議論を提供し、一般的な会計研究における設定 (setting) の、固有の設定 (design choices) に関する影響について明示する、という点で異なる。
  - \* 本論文におけるどの設定 (settings) においても、実証的証拠を提供するものではない。
- PSM の利用を検討している将来の研究に対する情報提供

■構成

1. Background on propensity score matching: PSM の有用性、誤解、適切なりサーチ・デザインの設計に関する議論
2. Propensity score matching in accounting research: 主な会計雑誌に掲載された PSM を利用した研究のサーベイ
3. Empirical examples of propensity score matching in accounting settings: 3 つの会計的な例を設定した場合における、PSM が引き起こす問題に関する事例
4. Suggestion and consideration for future research: マッチング手法の利用を検討している研究に対する提言
5. Conclusion: 本論文の発見事項とインプリケーション

### 3 PROPENSITY SCORE MATCHING IN ACCOUNTING RESEARCH

■会計研究における PSM の利用 (Table 1 Panel A)

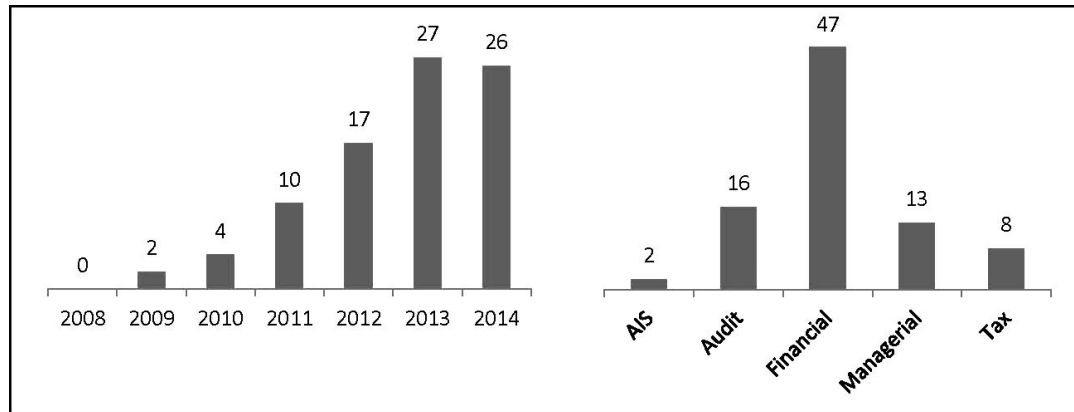
- 2008～2014 年における、*The Accounting Review*, *Contemporary Accounting Research*, *Journal of Accounting and Economics*, *Journal of Accounting Research*, and *Review of Accounting Studies* に掲載された論文延べ 86 件が対象。
- 会計研究で PSM が用いられはじめたのは最近 (86 件中 70 件は 2012～2014 の期間に刊行)。

■各研究の PSM の位置付け (Table 1 Panel B)

- 主要な分析 (primary analyses) として用いている研究が 37 件であるのに対し、ロバスト・チェック (sensitivity or robustness tests) として用いている研究は 49 件。
- PSM を採用する理由として FFM や重回帰分析の線形性の仮定を挙げている研究はわずか 20 件。

TABLE 1

## Descriptive Statistics for Accounting Studies Using PSM

Panel A: Number of Studies in Top Accounting Journals Using PSM Techniques by Year and Topic<sup>a</sup> (2008–2014)

<sup>a</sup> Topics are classified using the BYU Accounting Research taxonomy (Coyne et al. 2010). Twenty-three studies had more than one BYU topic classification (the majority of which include “financial”). Based on judgement, we placed each study into just one classification.

## Panel B: Purpose and Reliance on PSM in Empirical Tests (2008–2014)

Is PSM used as a primary or sensitivity analysis?	Primary	Sensitivity
	37	49
If used as a primary analysis, is PSM the only method for at least one conclusion?	Yes	No
	22	15
Is PSM motivated by concerns about FFM or nonlinearities?	Yes	No
	20	66
Did the paper test for FFM or nonlinearities?	Yes	No
	2	84
Is PSM motivated by generic concerns about “self-selection,” “endogeneity,” or “omitted variable” bias?	Yes	No
	33	53

## Panel C: Implementation of PSM (2008–2014)

Was the underlying treatment construct dichotomous?	Yes	No	
	52	34	
For the 64 studies that used other non-PSM tests, were the matching/control variables consistent with other tests?	Yes	No	Unknown
	13	43	8
Does the study match with replacement or without replacement?	With	Without	Unknown
	5	26	55
Did the paper impose a caliper distance?	Yes	No/Unknown	
	29	57	
Was the matching procedure 1:1 or 1:m?	1:1	1:m	Unknown
	68	11	7
Did the paper discuss covariate balance?	Yes	No	
	51	35	
Does the study use MR or a t-test for the second stage?	MR	t-test	Unknown
	58	22	6

Table 1 presents descriptive statistics on the use of PSM in the leading accounting journals from 2008–2014. Studies were identified by searching all publications in *The Accounting Review* (28 studies), *Contemporary Accounting Research* (20 studies), *Journal of Accounting and Economics* (13 studies), *Journal of Accounting Research* (16 studies), and *Review of Accounting Studies* (nine studies) for PSM-related key words (e.g., “propensity,” “PSM”) and manually determining whether a PSM technique was used. Panel A categorizes the studies by year and topic. Panels B and C classify the studies by motivation and methodology. All studies identified are listed in Appendix A.

- PSM が対処しうる内生性の問題を提示することなく、広く“自己選択 (self-selection),” “内生性 (endogeneity),” および“欠落変数バイアス (omitted variable bias)” への対応として PSM を用いている研究が 33 件ある。
- Heckman (1979) の代わりとして誤用してしまっている研究も存在する。

#### ■処置群の選択の方法 (Table 1 Panel C)

- 問題の所在
  - 処置が 2 値変数 (dichotomous) であるならば、PSM の実施は単純である。
  - しかしながら、多様な状況でマッチングを実施するため、連続 (あるいは順序) 変数に閾値を設けて変換することがある<sup>\*1</sup>。
- 問題点
  - このような場合、閾値付近の観測値が over-represent される傾向があり、それによって、効果の大きさ (および平均処置効果) が消失し、第 II 種の過誤が生じる可能性が増大する。
- 連続変数を用いている研究の数
  - 34 件。また、この影響により、効果の大きさのみならずサンプル・サイズも低下する。
  - 59 (12) 件の研究において、MR のサンプル・サイズの大きさは PSM の 3 (10) 倍である。
  - サンプルサイズが小さいほど、サブサンプルは母集団を代表しなくなる。

#### ■コントロール変数の選択 (Table 1 Panel C)

- MR と PSM のいずれを用いるにせよ、同様のコントロール変数を用いるべきであるにもかかわらず、しばしば異なるコントロール変数が用いられていることがわかった。
- MR からマッチングに用いた変数を除外することは、その変数が処置変数 (treatment) にも結果変数 (outcome) にも影響を与えないこと、ひいては、その変数によるマッチングが不要であることを意味するに他ならない。
- 分析においては、*post hoc* なモデルの特定 (model specification) をおこなっているという疑念 (appearance; 外観) を避けるため、PSM と他のテストとの説明変数の不一致を検討すべきである。

---

<sup>\*1</sup> 2 値変数を用いた処置群の選択について、例えば修正再表示のアナウンスメントや IFRS のアドプションがあげられる。一方で、非 2 値変数を用いた処置群の選択について、企業の所有構造や監査人の産業特殊性 (auditor industry specialization) があげられる

■傾向スコア推定後のマッチング・プロシージャ (Table 1 Panel C)