データ駆動型回帰分析 補足資料

前川 大空 *

2025 年 6 月 18 日

目次

1 回帰分析の課題

■データ駆動 従来の計量経済学では変数選択, ノンパラ, セミパラがこれにあたる. 何を指すのだろう $\cdots$

■一様妥当性 データ生成の母集団分布についての頑健性, といったことか. 頑健性と同じなのかしら $\cdots$

1.1 回帰分析

■p.1 観測可能性 観測可能性を前提としている. つまり, $\mathbf{W}$ はコントロール変数.

■p.1 構造モデル 構造モデルはデータ生成過程を表すのみ, 観測不可能な部分を誤差に全てまとめているため, 内生性などは排除されていない. つまり, 構造モデルは平均独立や条件付平均独立を満たすとは限らない.

■p.2 回帰モデル 回帰関数とは, 被説明変数の条件付期待値関数のことを指す. 回帰モデルとは:

$$Y = g(S, \mathbf{W}) + e := \underbrace{\mathbb{E}[\,Y \mid S, \mathbf{W}\,]}_{\text{回帰関数}} + e$$

と回帰関数を (1.1) 式に適応させた式である. LIE から条件付平均独立 $\mathbb{E}[\,e \mid S, \mathbf{W}\,] = 0$ が確認できる.

■p.2 回帰関数の識別 回帰関数 $\mathbb{E}[\,Y \mid S, \mathbf{W}\,]$ は $(Y, S, \mathbf{W})$ の同時分布から一意に定まる. これは, 一般の分布について, 母集団モーメントは分布が判明することによって一意に定まるためである. 条件付分布 $Y \mid_{S,\mathbf{W}}$ の平均はこの特殊ケースと見なせる. ここで, $(Y, S, \mathbf{W})$ は全て観測可能である.

* 一橋大学経済学部 4 年, 五年一貫専修コース公共経済プログラム

Def: 識別

観測されるデータの同時分布が既知の時, $\boldsymbol{\theta}$ の値が一意に定まるならば, $\boldsymbol{\theta}$ は識別されるという.

上の定義から分かるように, 回帰関数は識別される. 何故ならば, 観測可能なデータ $(Y, S, \mathbf{W})$ の分布が既知の時, 上記の議論から回帰関数 $g(S, \mathbf{W}) := \mathbb{E}[Y \mid S, \mathbf{W}]$ は一意に定まるためである.

■p.2 **構造的/記述的な分析** 以下のような区別が為されている.

構造的/記述的な分析

構造的な分析: 何かしらの決定メカニズムを背後に想定する分析

記述的な分析: 観測される情報のみから識別可能な変数間の関係の分析

つまり, 回帰モデルによる分析は記述的な分析といえる.

■p.2 **構造モデルの識別** 回帰モデルでない構造モデル, つまり, $\mathbb{E}[e \mid S, \mathbf{W}] \neq 0$ である場合, g は回帰関数ではない別の関数になる. 関数が特定できないため, このままでは識別できない. 構造モデルへの追加的仮定は, 識別のために, 経済学ならば経済理論に基づいた妥当性が実証データからは検証できない仮定を置く. [*1]

■p.3 **構造モデルにおける誤差項の加法分離性** (1.1) 式において, 大卒の因果効果を測るために $\mathbf{W}$ のみならず本来観測不可能な e も一定としていることに注意せよ. つまり, 因果効果の識別については, (1.1) 式に基づく構造的な分析の文脈においても述べられていない.

■p.3 **$\mathbf{W}$ 一定では高卒/大卒の賃金の差も一定という制約** 検証しておこう.

Proof 誤差項が加法分離可能な構造モデル:

$$Y_i = g(S_i, \mathbf{W}_i) + e_i \ ,$$

について, 個人 i の教育年数の賃金への因果効果は,

$$Y_i \mid_{S_i=16} - Y_i \mid_{S_i=12} = [g(16, \mathbf{w}_i) + e_i] - [g(12, \mathbf{w}_i) + e_i] = g(16, \mathbf{w}_i) - g(12, \mathbf{w}_i) \ ,$$

$\mathbf{w}_i = \mathbf{w}_j$ なる 2 個人の因果効果は $g(16, \mathbf{w}_i) - g(12, \mathbf{w}_i) = g(16, \mathbf{w}_j) - g(12, \mathbf{w}_j)$ で同一. □

■p.3 **回帰モデルと因果効果** 分かるのはあくまで平均で, 予測に過ぎない. 因果効果とは限らない.

■p.3 **因果推論は構造的な分析** 先述の例の通り, 因果効果は識別できるとは限らず, 単にメカニズムを記述したのみの, 即ち構造的な分析の範疇であった. しかし理論に基づく様々な仮定を置くことによって識別が可能となり, 因果効果の分析, 因果推論も記述的な分析に落とし込むことが出来る.

1.2 線形回帰モデル

1.3 本書の課題と構成

1.4 補論

*1 これが構造推定なる分野なのだろうか.

付録 A　記法

■不明点　分からない記述は赤文字を用いて記載する.

■ベクトル, 行列　共に `mathbf` を用いて記載する.

■条件付期待値　本文では $S = s,\ \mathbf{W} = \mathbf{w}$ である部分母集団について, $\mathbb{E}[Y \mid s, \mathbf{w}]$ と記述されている. 任意の $S, \mathbf{W}$ についてこの関係が成立する場合, $\mathbb{E}[Y \mid S, \mathbf{W}]$ と記載することにする.

■繰り返し期待値の法則　Law of Iterated Expectation, LIE と略す.

参考文献

[1] 末石 直也 (2024), データ駆動型回帰分析-計量経済学と機械学習の融合, 第 1 版, 日本評論社

[2] 末石 直也 (2015), 計量経済学 ミクロデータ分析へのいざない, 第 1 版, 日本評論社