

クロスセクションデータからの「差分の差」による因果効果推定：背景情報が利用出来る場合

名古屋大学 大学院経済学研究科 星野 崇宏

経済学における政策評価においては、介入の対象となるグループ（処置群）と対象とならないグループ（対照群）は質的に大きく異なることが多い。そこで単に政策介入の後に得られた結果変数の差（介入後の差）を利用するだけでなく、介入前の差も測定しておき、「介入後の差 - 介入前の差」によって純粋な政策効果の推定値とする場合がある。これが差分の差推定量 (Difference in Differences) である。この方法は一般には複数時点で同一対象を繰り返して調査する、いわゆるパネル調査での処置群と対照群の事後の差と事前の差を利用することが多い。しかし一般にパネル調査では対象を追跡するのが難しく、追跡を正確に行うためには様々なコストがかかり、また二時点目で調査を受けない”ドロップアウト”が多数起こることがある。そこで二時点で別々の調査対象に調査を行ういわゆるクロスセクションデータの繰り返しによって政策効果を推定することが望まれる場合がある。しかしクロスセクションデータでの”差分の差”は、実際には様々な前提条件を仮定しないと適切な推定量とはならない。このことを理解するためには、潜在的結果変数を用いたモデルの記述を行う。まず b 時点での調査対象であれば $\delta = 1$ 、 $a(< b)$ 地点での調査対象であれば $\delta = 0$ とするインディケータ δ 、さらに a 時点、および b 時点で処遇群か対照群かを示すインディケータ z_a 、 z_b を考える（とりあえずは共変量 x は無視して議論する）。 y_a は政策を実施する前での結果変数の値、 y_{1b} はもし政策の対象となる集団に所属した場合における政策を実施した後の結果変数の値、 y_{0b} はもし政策の対象となる集団に所属しない場合における政策を実施した後の結果変数の値である。さて、ある介入を行った後の $t = b$ 時点において、政策の対象となった集団 $z = 1$ における「政策を実行した場合の結果変数と行わなかった場合の結果変数の期待値」は経済学等の社会科学において政策効果を計る量の一つであり、これは $t = b$ における処遇群での介入効果 (TET: treatment effect for the treated) $TET = E(y_{1b} - y_{0b} | z = 1)$ である。一方、単純な差分の差

$$\begin{aligned} DID &= E(y | \delta = 1, z_b = 1) - E(y | \delta = 0, z_a = 1) - \left\{ E(y | \delta = 1, z_b = 0) - E(y | \delta = 0, z_a = 0) \right\} \\ &= E(y_{1b} | \delta = 1, z_b = 1) - E(y_a | \delta = 0, z_a = 1) - \left\{ E(y_{0b} | \delta = 1, z_b = 0) - E(y_a | \delta = 0, z_a = 0) \right\} \quad (1) \end{aligned}$$

の単純な観測平均は、何も仮定を置かない場合には、TET の不偏な推定量ではない。このとき、(1) 二時点間で調査対象者は等質である、つまり $E(y_{1b} | z_b, \delta = 1) = E(y_{1b} | z_b, \delta = 0) = E(y_{1b} | z_b)$ 、 $E(y_{0b} | z_b, \delta = 1) = E(y_{0b} | z_b, \delta = 0) = E(y_{0b} | z_b)$ 及び $E(y_a | z_a, \delta = 1) = E(y_a | z_a, \delta = 0) = E(y_a | z_a)$ 、(2) 介入しなかった場合の従属変数の変化が b 時点における処遇群と対照群で等しい、つまり $E(y_{0b} - y_a | z_b = 1) = E(y_{0b} - y_a | z_b = 0)$ 、(3) 2つのデータ間での処遇群と対照群での a 時点での従属変数の平均の違いは等しい、つまり $E(y_a | z_b) = E(y_a | z_a)$ 、という3つの条件が成立していれば、これを式 (1) に代入することにより、 $DID = TET$ となることを示すことができる。逆に言えば、クロスセクションデータから差分の差推定量を得るとき、これを TET の推定量として利用している場合には、上記の3つの仮定を置いているということである。また、調査対象者について様々な共変量を利用出来る場合には、上記の3つの条件で期待値をすべて「共変量 x を所与とする」場合に変更した条件のもとで DID が $TET = E(y_{1b} - y_{0b} | z_b = 1)$

と一致する。調査対象者についての様々な背景情報を所与とする場合には3つの条件はより成立しやすくなるが、一方差分の差推定量を構成する4つの項それぞれが共変量についての回帰関数を含んでいるため、ここでも回帰関数を指定しないセミパラメトリックな推定量があれば望ましい。そこで Abadie (2005) は2時点のクロスセクションデータを利用した TET の推定量として

$$\frac{\sum_{i=1}^N \left[e_i \frac{\delta_i - \lambda}{\lambda(1-\lambda)} \frac{z_i - e_i}{e_i(1-e_i)} y_i \right]}{\sum_{i=1}^N e_i} \quad (2)$$

が漸近正規性を有する一致推定量であることを示した。

2 時点のクロスセクションデータが利用出来る場合であっても、“処遇群での因果効果” $TET = E(y_{1b} - y_{0b} | z = 1)$ を推定することのみが目的であれば、“強く無視可能な割り当て”条件を仮定した上で「二時点目 ($\delta = 1$) のデータのみ利用した解析を行う」ことも可能である。しかし一般には一時点目のデータを利用することによって推定効率が向上することが予想される。また、クロスセクションデータに対しては、パネル調査のような経験尤度法は開発されておらず、共変量情報をフルに利用した推定法を開発することが有用であると考えられる。このような観点から、以下の5つの解析方法をクロスセクションデータに適用し、手法間の比較を行う。具体的には (1) $\delta = 1$ のデータのみ利用した二重にロバストな推定、(2) クロスセクションデータにおける傾向スコアを用いたセミパラメトリック推定 (Abadie, 2005)、(3) 傾向スコアを用いたマッチングによる疑似パネルデータに対する経験尤度法を利用、(4) 傾向スコアを用いたマッチングによる疑似パネルデータに対する二重にロバストな推定法、(5) ディリクレ過程混合モデルを用いたセミパラメトリックベイズ回帰 (Dunson et al., 2007)、である。このうち、方法 (3) としては解析法として Qin and Zhang (2008) を利用しているが、疑似パネルには利用されたことはない。また (4) についてはそもそもパネル調査における“二重にロバストな推定法”が提案されていないが、これは Abadie (2005) の方法を素直にロバストにすればよい。結果として

$$\sum_{i=1}^N \left\{ (y_{bi} - y_{ai}) \frac{z_i - e_i}{1 - e_i} + \frac{e_i^2 - z}{1 - e_i} (y_{bi} - g_{a1}(x_i) - g_{b0}(x_i) + g_{a0}(x_i)) \right\} / \sum_{i=1}^N e_i \quad (3)$$

を推定量とすればよい。但し g_{a1}, g_{b0}, g_{a0} はそれぞれ $E(y_a | z = 1, x), E(y_{0b} | z = 0, x), E(y_a | z = 0, x)$ の線形な回帰関数である。上記の5つの解析法の比較のためのシミュレーションを Qin and Zhang (2008) の設定に従って実施したところ、平均平方二乗誤差の観点からは (1) よりもそれ以外の方法、つまり2時点のデータを利用した方が誤差が減少し、誤差の大きさは (2) > (3) > (4) = (5) という関係になった。

参考文献

- Abadie, A. (2005) “Semiparametric Difference-in-Differences Estimators,” *Review of Economics and Statistics*, Vol. 72, pp. 1-19.
- Dunson, D.B., Pillai, N., and Park, J-H. (2007) “Bayesian Density Regression,” *Journal of the Royal Statistical Society, series B*, Vol. 69, pp. 163-183.
- Qin, J. and Zhang, B. (2008) “Empirical-likelihood-based Difference-in-differences Estimators,” *Journal of the Royal Statistical Society, series B*, Vol. 70, pp. 329-349.