第9講　マッチング手法

先に述べたように、政策を実施した処理群のデータはあるが、対照群のデータは不在であり、外部情報から対処群をみつけてマッチングさせる必要が出てくる場合には次のような手法を用いる。

実際に外部データが十分にあり、処理群に含まれる個別サンプルの全ての変数、属性にぴったり一致するような対照群サンプルを選ぶことが出来れば、これを**完全一致マッチング（exact matching）**と呼ぶが、分析に用いる変数が増えるに従って、個別変数をマッチングさせることは難しくなる。このような場合、変数一つ一つをマッチングさせるのではなく、ある程度変数を集約して表現した条件付き確率（これを傾向スコア；**propensity score** と呼ぶ）を処理群と対照群でマッチングさせるという方法が考えられる[1]。

具体的な考え方は、処理群に選ばれる確率 $[\Pr[d_i = 1|\, x]]$ を全サンプルを用いたロジット推定によって求め、それを処理群と対照群に分け、さらに確率を均等な階層に分け、同じ階層に入るもの同士をマッチングさせ、処理効果の平均を求めるというものである。ここでマッチングをどうするかということが問題になる。すなわち、一度マッチングに使った対照サンプルを再び使うことを認めるかどうか、比較対照するために処理サンプルに対していくつの対照サンプルを割り当てるのか。最も propensity score が近いもの一つを選べばいいのか（**caliper matching** と言う）、それともその周辺の対照サンプルを複数割り当てるのがいいのか、また、具体的なマッチングの方法としてどのようなものを用いるのか、といった問題がある[2]。

マッチングの詳細な手法はかなり技術的に高度になるので、ここでは扱わないが、主要な手法の基本的な考え方を紹介しておきたい[3]。一般に処理効果は次のように表すことが出来る。

$$\triangle^M = \frac{1}{N_T} \sum_{i \in \{d=1\}} [y_{i1} - \sum_j w(i,j) y_{j0}]$$

[1] この方法は Rosenbaum and Rubin (1983) によって開発され、現在では以下で紹介するように沢山の拡張が行われている。

[2] 実際にこれらの問題にどう対処するかということは研究者の判断にゆだねられている。逆に、決定的に正しい方法が知られているわけではなく、試行錯誤するしかない。

[3] 最新のマッチング手法のアルゴリズムに関しては、Abadie et al (2004)、Becker and Ichino (2002)、Becker and Caliendo (2007) 等を参照されたい。

ここで $\sum w(i,j) = 1$、$0 < w(i,j) \leq 1$ となるマッチング・ウェイトである。N_T は処理群のサンプル数を表す。

対照群としてどのようなサンプルを処理群にマッチさせるかという事であるが次のようなマッチング手法が提案されている。(1) **最近傍マッチング（nearest-neighbor matching）**の考え方では、全ての処理サンプル i に対して、次のような条件を満たす集合 $A_i(x) = \{j | \min_j \|x_i - x_j\|\}$ を対照群として選択する[4]。(2) **カーネル・マッチング (kernel matching)** では、ウェイトを次のように定義する $w(i,j) = K(x_j - x_i)/\sum_{j=1}^{N_{ic}} K(x_j - x_i)$、ここで K はカーネル関数を表す。(3) **層化マッチング（stratification matching)** とは propensity score を均等に層化し、層内で処理群と対照群が同じスコアになるようにした後で、処理効果を推定する。同じスコアのペアが組めない場合には、その層内での処理効果は計算されない。(4) **半径マッチング（radius matching）**では対照群集合を次のように定義する。$A_i(p(x)) = \{p_j | \|p_i - p_j\| < r\}$ すなわち、propensity score の差が半径 r 以内であればペアとしてマッチングするという方法である。

マッチング手法[5]

I_0 the index for non-participant

I_1 the index for participant

The effect of treatment for each treated observation $i \in I$

$$\Delta^{MAT} = \frac{1}{N_1} \sum_{i \in Z_i} \left[Y_i^1 - \sum_{j \in Z_0} W_{N_0}(i,j) Y_j^0 \right]$$

$\Sigma_j W_{N_0}(i,j) = 1 \quad \nabla_i$ the total weight of all

controls sums up to one for each treated neighbourhood individual.

$C(P_i)$ for each i in the participant sample and denote as neighbours for i those non-participants

$j \in A_i \quad$ where $\quad A_i = \{j \in I_0 \mid P_j \in C(P_i)\}$

Nearest-Neighbour-Matching 最近傍マッチング（NN マッチング）

$C^{NN}(p_i) = \min_j \| p_i - p_j \| . j \in N_0$

$$W_{N_0}^{NN}(i,j) = \begin{cases} 1 & if \ \| p_i - p_j \| = \min_j \| p_i - p_j \| \\ 0 & otherwise \end{cases}$$

NN matching with replacement and without replacement

If we allow replacement(再利用可） the average quality of matching will ***increase*** and the bias will ***decrease***.

[4] $\|\|$ はベクトル間のユークリッド距離を表す。

[5] 以下は Cameron and Trevedi (2005, pp.874-878）を参照。

When using oversampling, one has to decide how many matching partners m should be chosen for each individual i and which weight should be assigned to them. One possiblity is to use ***uniform weights***.(1/m)

All the m control individuals within set Ai receive the weight 1/m whereas all other individuals from the control group receive the weight zero

$$W^{NN^0}(i,j) = \begin{cases} 1 & \frac{1}{m}\ if\ j \in A_i \\ 0 & otherwise \end{cases}$$

Triangular weights, the m individuals with set A_i has to be ranked where $\rho = 1$ is the closest neighbour. $\rho = 2$ the next closest.

Caliper and Radius Matching

NN matching faces the risk of bad matcher, if the closest neighbour is far away. This can be avoided by imposing a *tolerance level* on the maximum distance $\| p_i - p_j \|$ allowed.

$\| p_i - p_j \| < \varepsilon\ j \in N_\sigma$

$$w^{CM}(i,j) = \begin{cases} 1 & if\ \ \| p_i - p_j \| = \min j \| p_i - p_j \| \curlywedge \| p_i - p_j \| < \varepsilon \\ 0 & otherwise \end{cases}$$

Stratification and Interval Matching

To implement STM, the propensity score is used to divide the full sample into M blocks of units of approximately equal probability of treatment.

Let J is be an indicator for unit i being in blocks. One way of implementing this is to divide the unit interval into ξ blocks with boundary values equal to $\frac{s}{S}$ for $s = 1, ..., s-1$.

$J_{is} = \left\{ \frac{\delta - 1}{S} < p(x_i) \leq \frac{\delta}{S} \right\}$

$N_{1s}(treated) = \sum_i | \{D_i = 1,\ J_{is} = 1\}$

$N_{0s}(untreated) = \sum_i | \{D_i = 0,\ J_{is} = 1\}$

Average Treatment effect within each block

$$\Delta_s^{STM} = \frac{1}{}N_{1s} \sum_{i=1}^{N} J_{is} D_i Y_i - \frac{1}{N_{0s}} \sum_{i=1}^{N} J_{is}(1 - D_)_i Y_i$$

overall average treatment effect

$$\Delta_{ATE}^{STM} = \sum_{s=1}^{S} \Delta_s^{STM} \frac{N_{1S} + N_{0S}}{N}$$

$$\Delta_{ATT}^{STM} = \sum_{s=1}^{S} \Delta_s^{STM} \frac{N_{1S}}{N_1}$$

The within block average treatment effects by the number of treated units.

Kernel and Local Polynomical Matching

Kernel Matching (KMA) and local linear matching (LLM) and non-parametric matching estimates that were all unitsin the control group to constructed a match for each programme participatnts.

And advantage is the lower variance is head for constructing counterfactual outcomes.

If weights from symmetric, non-negative, unimodel Karnel and Wed , then the average places highter weight on persons close in terms ofP_i and lower weight on more distanct obsevations. Kernel matching set $A_i = I_0$ *and* s uses the following weights:

$W_{N_D}^{KM}(i.j) = \frac{G_{ij}}{\Sigma_{K \in Z_0} G_{iK}}$

when $G_{iK} = G[(P_i - P_k)/GN_0]$ is a Kernel

that downweights distant observations from P_i and aN_0 is a bandwidth parameter.

Kernel Functions

Rectangular/uniform	$K(x) = \frac{1}{2}$	$\mid x \mid \leq 1$
Epanchnikov	$K(x) = \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}x^2)$	$\mid x \mid \leq \sqrt{5}$
	$\{-\frac{3}{4}(1 - x^2)$	$\mid x \mid \leq 1\}$
Quardrative (biweight)	$K(x) = \frac{15}{16}(1 - x^2)^2$	$\mid x \mid \leq 1$
Triangular	$K(x) = (1 - \mid x \mid)$	$\mid x \mid \leq 1$
Normal/Gaussian	$K(x) = \frac{1}{\sqrt{2\pi}} e^{-0.5x^2}$	$-\infty < x < \infty$

A generalized version of KM is local linear matching LLM has a faster rate of convergence near boundary points, and greater robustness to different data design densities.

$$W^{LLM}(i,j) = \frac{G_{ij}\Sigma_{k \in I_0} G_{ik}(P_k - P_i)^2 - [G_{ij}(P_j - P_i)]\left[\Sigma_{K \in I_0} G_{iK}(P_K - P_i)\right]}{\Sigma_{j \in I_0 G_{ij}} \Sigma_{K \in I_0} G_{ij}(P_K - P_i)^2 - (\Sigma_{K \in I_0} G_{iK})(P_K - P_i)}$$

Weighting on the Propensity Slone Hirano and Imbens (2002) suggest a straightforward way to implement this estimator by re-weighting treated and control observations to make them representative of the population of interest.

$$E\left[\frac{YD}{P(X)}\right] = E\left[\frac{Y^1 D}{P(X)}\right] = E\left[E\left[\frac{Y^1 D}{P(X)}|X\right]\right]$$
$$= E\left[\frac{P(X)E[Y^1|X]}{P(X)}\right] = E[Y^1]$$

Unconfoundedness

$$E\left[\frac{(1-D)Y}{1-P(X)}\right] = E[Y^0]$$

The Average Treatment Effect.

$$E\left[\frac{YD}{P(X)} - \frac{(1-D)Y}{1-P(X)}\right] = E[Y^1 - Y^0] = \triangle_{ATE}$$

If the propensity score is known, the estimator can directly be implemented as the differencebetween a weighted average of the outcomes for the treated individuals and a weighted

Trade-offs in Teams of bias and Efficiency

	Bias	*Variance*
Nearest neighbour matching:		
multiple neighbour	$+/-$	$-/+$
Single neibour with caliper/without caliper		$+/-$
Use of Control individuals:		
With replacement/without replacement	$-/+$	$(+)/(-)$
Use of Control individuals:		
With replacement/without replacement	$(-)/(+)$	$(+)/(-)$
NN-matching/Radical matching	$-/+$	$+/-$
KM/LLM/NN method	$+/-$	$-/+$
Band width choice with KM small/large	$-/+$	$+/-$

average of the outcomes for the non-participants.

The weights can be normalized to unity. The simple weighting estimator is given by

$$\triangle_{ATE}^{WG} = \sum_{i=1}^{N} \frac{P_i Y_i}{\hat{P}(X_i)} \Big/ \sum_{i=1}^{N} \frac{D_i}{\hat{P}(X_i)} - \sum_{i=1}^{N} \frac{(1-D_i)Y_i}{1-\hat{P}(X_i)} \Big/ \sum_{i=1}^{N} \frac{1-D_i}{1-\hat{P}(X_i)}$$

ATT

$$\triangle_{ATT}^{WG} = \left[\frac{1}{N_1} \sum_{i \in I_1} Y_1 \right] - \left[\sum_{i \in I_0} Y_i \cdot \frac{\hat{P}(X_i)}{1-\hat{P}(X_i)} \Big/ \sum_{i \in I_0} \frac{\hat{P}(X_i)}{1-\hat{P}(X_i)} \right]$$

参考文献

[1] 星野崇宏 (2009)『調査観察データの統計科学』、岩波書店

[2] 北村行伸 (2009) 『ミクロ計量経済学入門』、日本評論社

[3] Anderson, T.W. (1984) *Introduction to Multivariate Statistical Analysis*, Wiley.

[4] Anderson, T.W. and Rubin, H.(1949) “Estimators of the Paramters of a Single Equation in a Complete Set of Stochastic Equations”, *Annals of Mathematical Statistics*, 21, pp.570-82.

[5] Andrew, Donald W.K. and Stock, James H.(2005) “Inference with Weal Instruments”, NBER Technical Working Paper 313.

[6] Angrist, J.D.and Krueger, A.B.(1991) “Does Compulsory School Attendance Affect Schooling and Earnings?”, *Quarterly Journal of Economics*, 106, pp.979-1014.

[7] Angrist, Joshua, D., Imbens, Guid, W. and Rubin, Donald B.(1996) “Identification of Causal Effects Using Instrumental Variables”, *Journal of the American Statistical Association*, 91(434), pp.444-455.

[8] Angrist, Joshua D. and Krueger, Alan B.(2001) “Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments”, *Journal of Economic Perspectives*, 15(4), pp.69-85.

[9] Angrist, Joshua, D.,Bettinger, Eric., Bloom, Erik., King, Elizabeth., and Kremer, Michael. (2002) “Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment”, *American Economic Review*, 92(5), pp.1535-1558.

[10] Basmann, R.L. (1960) “On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics”, *Journal of the American Statistical Association*, 55(292), pp.650-59.

[11] Baum, Christopher. (2006) *An Introduction to Modern Econometrics Using Stata*, Stata Press.

[12] Blackburn, McKinley and Neumark, David.(1992) “Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials”, *Quaterly Journal of Economics*, 107(4), pp.1421-1436.

[13] Bound, John., Jaeger, David.A. and Baker, Regina. M.(1995) “Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak”, *Journal of the American Statistical Association*, 90(430), pp.443-50.

[14] Bowden, R.J. and Turkington, D.A.(1984) *Instrumental Variables*, Cambridge University Press.

[15] Breusch, Trevor., Qian, Hailong., Schmidt, Peter., and Wyhowski, Donald.(1999) "Redundancy of Moment Conditions", *Journal of Econometrics*, 91, pp.89-111.

[16] Cameron, A.C.and Trivedi, P.K.(1998) *Regression Analysis of Count Data*, Cambridge University Press.

[17] Cameron, A.C. and Trivedi, P.K.(2005) *Microeconometrics: Methods and Applications*, Cambridge University Press.

[18] Chao, John.C. and Swanson, Norman R.(2005) "Consistent Estimation with a Large Number of Weak Instruments", *Econometrica*, 73(5), PP.1673-1692.

[19] Cragg, John G.and Donald, Stephen G.(1993) "Testing Identifiability and Specification in Instrumental Varaible Models", *Econometric Theory*, 9, pp.222-40.

[20] Davidson, Russell and MacKinnon, James G.(2004) *Econometric Theory and Methods*, Oxford University Press.

[21] Durbin, J.(1954) "Errors in variables", *Review of the Internatinal Statistical Institute*, 22, pp.23-32.

[22] Griliches, Zvi.(1976) "Wages of Very Young Men", *Journal of Political Economy*, 84(4. Part 2), pp. S69-S85.

[23] Griliches, Zvi.(1977) "Estimating the Returns to Schooling: Some Econometric Problems", *Econometrica*, 45(1), pp.1-22.

[24] Griliches, Zvi., Hall, Bronwyn., and Hausoman, Jerry.(1978) "Missing Data and self-Selection in Large Panels", *Annales de L'INSEE*, XXX-XXXI, pp.137-76.

[25] Hahn, Jinyoung and Hausman, Jerry. (2002a) "A New Specification Test for the Validity of Instrumental variables", *Econometrica*, 70(1), pp.163-189.

[26] Hahn, Jinyoung and Hausman, Jerry. (2002b) "Notes on Bias in Estimators for Simultaneous Equation Models", *Economics Letters*, 75. pp.237-241.

[27] Hahn, Jinyoung and Hausman, Jerry. (2003) "Weak Instruments: Diagnosis and Cures in Empirical Econometrics", *American Economic Review*, 93(2), pp.118-125.

[28] Hall, Alastair R., Rudebusch, Glenn D. and Wilcox, David W.(1996) "JUdging Instrument Relevance in Instrumental Variables Estimation", *International Economic Review*, 37(2), pp.283-298.

[29] Hall, Alastair R. and Peixe, Fernanda P.M.(2000) "A Consistent Method for the Selection of Relevant Instruments", A paper presented at Econometric Society World Congress 2000.

[30] Hansen, Lars.P (1982) "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50(4), pp.1029-1054.

[31] Hausman, Jerry. (1978) "Specification tests in econometrics", *Econometrica*, 46, pp.1251-72.

[32] Hausman, Jerry., Stock, James H. and Yogo, Motohiro.(2005) "Asymptotic Properties of the Hahn-Hausman Test for Weak-Istruments", *Economics Letters,* 89, pp.333-42.

[33] Hayashi, Fumio.(2000) *Econometrics*, Princeton University Press.

[34] Imbens, Guido W. and Angrist, Joshua D.(1994) "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, 62(2). pp.467-475.

[35] Koenker, Roger.(1981) "A Note on Studentizing a test for Heteroscedasticity", *Journal of Econometrics*, 17., pp.107-112.

[36] Koenker, Roger. (2005) *Quantile Regression*, Cambridge University Press.

[37] Nelson, Charles R. and Startz, Richard.(1990a) "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One", *Journal of Business*, 63(1, Part.2), pp. S125-S140.

[38] Nelson, Charles R.and Startz, Richard.(1990) "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator", *Economerica*, 58(4), pp.967-76.

[39] Pagan, A.R. and Hall, D. (1983) "Diagnostic Tests as Residual Analysis", *Econometric Reviews*, 2(2), pp.159-218.

[40] Ruud, P.A. (2000) *An Introduction to Classical Econometric Theory*, Oxford University Press.

[41] Sargan, J.D. (1958) "The Estimation of Economic Relationships Using Instrumental Variables", *Econometrica*, 26(3), pp.393-415.

[42] Shea, John.(1997) “Instrument Relevance in Multivariate Linear Models: A Simple Measure”, *Review of Economics and Statistics*, 79(2), pp.348-352.

[43] Staiger, Douglas. and Stock, James.H. (1997) “Instrumental Variables Regression with Weak Instrumetns”, *Econometrica*, 65(3), pp.557-86.

[44] Stock, James H. and Wright Jonathan H. (2000) “GMM with Weak Identification”, *Econometrica*, 68(5), pp.1055-96.

[45] Stock, James H., Wright, Jonathan H. and Yogo, Motohiro. (2002) “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments”, *Journal of Business and Economic Statistics*, 20(4), pp.518-29.

[46] Stock, James H. and Yogo, Motohiro. (2005) “Testing for Weak Instruments in Linear IV Regression”, in Andrews, D.W.K. and Stock, J.H.(eds) *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press. pp.80-108.

[47] Winklemann, Rainer and Boes, Stefan. (2005) *Analysis of Microdata*, Springer.

[48] White, Halbert. (1980) “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity”, *Econometrica*, 48(4), pp.817-838.

[49] White, Halbert. (1982) “Instrumental Variables Regression with Independent Observations”, *Econometrica*, 50(2), pp.483-499.

[50] Wooldridge, Jeffrey. M. (2002) *Econometric Analysis of Cross Section and Panel Data*, The MIT Press

[51] Wu, D-M. (1973) “Alternative tests of independence between stochastic regressors and disturbances”, *Econometrica*, 41, pp.733-50.