Detecting Grouped Local Average Treatment Effects and Selecting True Instruments With an Application to the Effect of Imprisonment on Recidivism

Nicolas Apfel [1] Helmut Farbmacher [2] Rebecca Groh [2] Martin Huber [3] Henrika Langen [4]

[1]University of Southampton [2]TU Munich [3]University of Fribourg [4]University of Helsinki

ESEM 2023, Barcelona School of Economics

August 29, 2023

Motivation: The effect of imprisonment on recidivism

- Does incarceration prevent future crime (*recidivism*)?
- The results in the literature so far are mixed: crime-reducing, no effect or even a crime-inducing effect of prison
- Maybe there is heterogeneity of effects, maybe even inside states?
  - In some counties prisons are good at rehabilitating, in others they have reverse effect

This paper in a nutshell

1. Estimate the effect of incarceration on recidivism in the US based on judge instruments when allowing for effect heterogeneity

   Single estimate might mask coexistence of null, positive and negative effects

2. Proposition of a two-step method to allow for selection of valid IVs in presence of LATEs

   1. Find clubs of propensity scores (i.e. imprisonment rates) and apply clustering
   2. Inside each club find largest group of same reduced form estimates (i.e. recidivism rate)
   3. Estimate grouped heterogeneous treatment effects for all pairs of clubs using only the largest group of (valid) instruments (i.e. judges)

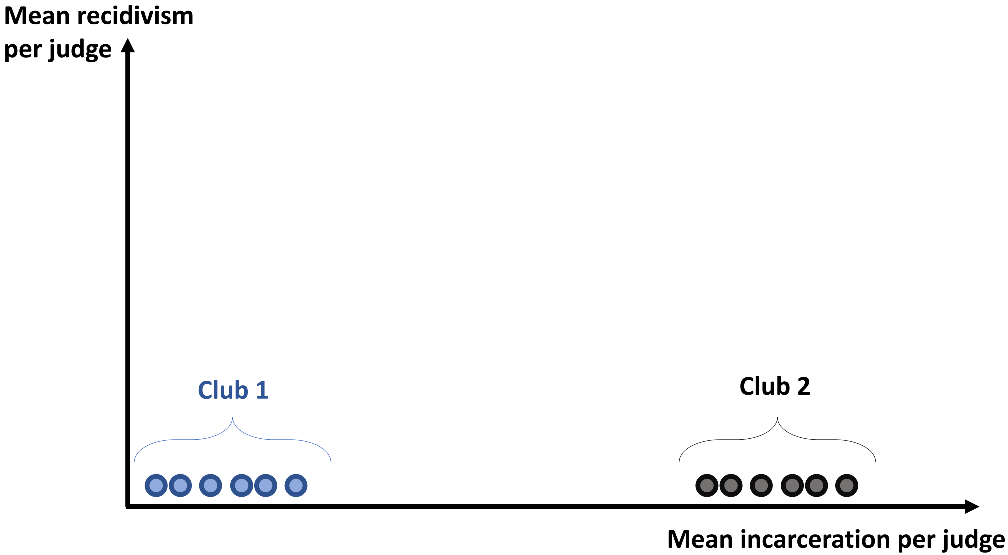
Assumptions

LATE Assumptions: For two distinct IV values $z \neq z'$:

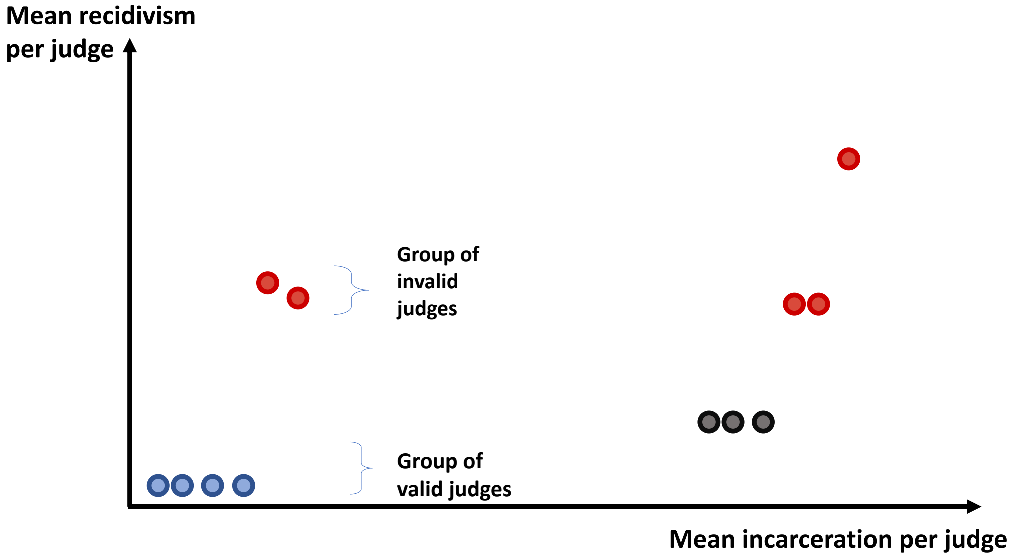
- **Assumption 1: Validity (Exogeneity and Exclusion restriction)**
  (i) Random assignment: $Z \perp (D(z), Y(z', d))$ and
  (ii) Exclusion: $Y(z, d) = Y(z', d) = Y(d)$
- **Assumption 2: Monotonicity** $\Pr(D(z) \geq D(z')) = 1$
- **Assumption 3: First Stage**: $E(D|Z = z) - E(D|Z = z') \neq 0$
  - *Clubs*: Sets of judges with comparable probability to incarcerate
  - *Group*: Set of judges inside club with same judge-specific mean of the outcome
- **Assumption 4: Plurality:** Largest group of judges in each club fulfils LATE assumptions

Model | Assumptions detail

Method: Visualization of Clubs and Groups



Method: Visualization of Clubs and Groups



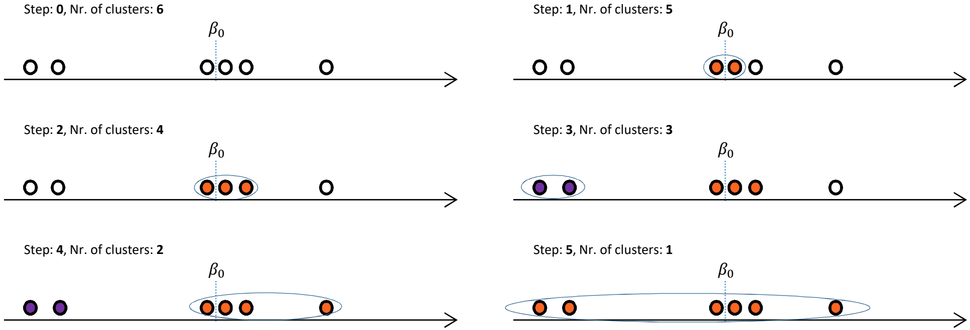
Method: Group-pair LATEs

$$LATE = \frac{R_{C1} - R_{C2}}{P_{C1} - P_{C2}}$$

- R_{C1}, R_{C2}: Recidivism rate for Group 1 and 2
- P_{C1}, P_{C2}: Imprisonment rate for Group 1 and 2
- LATE is the effect of imprisonment on recidivism for the sub-population of people who are sentenced to jail by a **judge in Group 1**, but **would not have been** sentenced by **a judge in Group 2**
- If we group judges into **4** clubs, we get only **6** estimates (instead of several thousands from a judge-wise comparison)

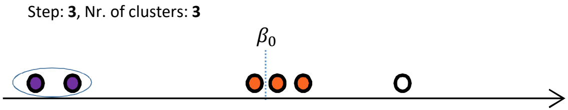
Method: AHC Visualization

Classification of mean imprisonment and recidivism via Agglomerative Hierarchical Clustering (AHC) (Ward, 1963; Apfel and Liang, 2021) Details



Method: AHC Visualization

Select the number of clusters via a **stopping rule**: Stop when the F-test for equality of all first-stage parameters in the cluster does not reject any more



Method

- **Idea:** Club pairs that fulfil LATE assumptions produce same LATE
- **Step I**: Classify Judges into Clubs
  - Find clubs of IVs with the same propensity score P_z
- **Step II**: Find largest groups of judges with same reduced form inside each club
- Estimate LATEs for group-pairs

Application: Incarceration and future crime

$$Recidivism_{ojt} = \beta_0 + Prison_{ojt}\beta + u_{ojt} \qquad \text{Equation of Interest}$$
$$Prison_{ojt} = \gamma_0 + \mathbf{Z}_{ojt}\gamma + \varepsilon_{ojt} \qquad \text{First Stage}$$

- o: Offender ID, j: Judge ID, t: Time period (2009-2014)
- *Recidivism*-Dummy: 1 if offender has reoffended within 3 years
- *Prison*-Dummy: 1 if offender has been convicted
- Adult offenders in US state Minnesota: Data obtained by linking Minnesota Judicial Branch case database with Minnesota Sentencing Guidelines dataset
- Judges assigned randomly according to the Minnesota Order for Assignment of Cases
- Controls: race-, gender-, offensetype-, year-dummies, severity of crime, age, the squares of the latter two and race-gender interaction dummies

Application: Incarceration and future crime

- Number of judges: 78
- Cases with presumptive sentences of up to three years (minor crimes), e.g. robbery, assault, theft, stalking, fleeing from the police, lottery fraud
- Number of cases per judge:
  - Min: 202
  - Mean: 307
  - Max: 935

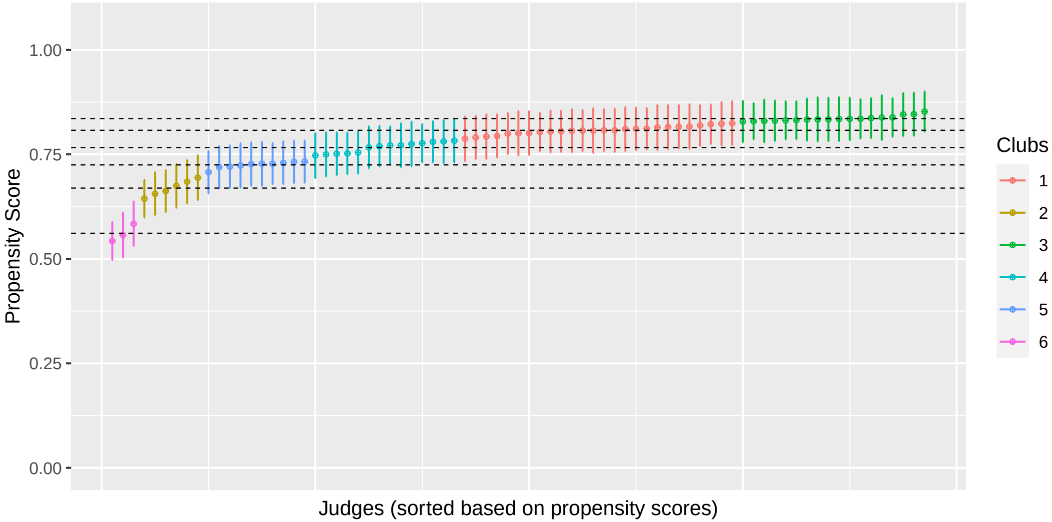
Descriptives

Application: Clusters based on imprisonment rate

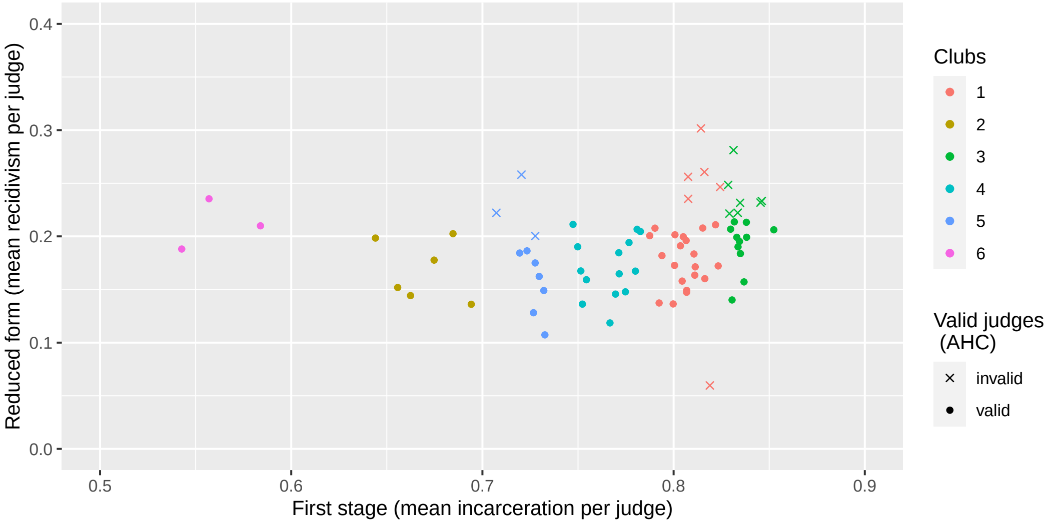
Club	Mean	Nr
1	0.81	26
2	0.67	6
3	0.84	18
4	0.77	14
5	0.72	10
6	0.56	3
7	0.15	1

Table: First step: Club allocation

Application: Judges sorted by imprisonment rate



Application: Imprisonment & Recidivism rate



Application: Group-pair LATEs

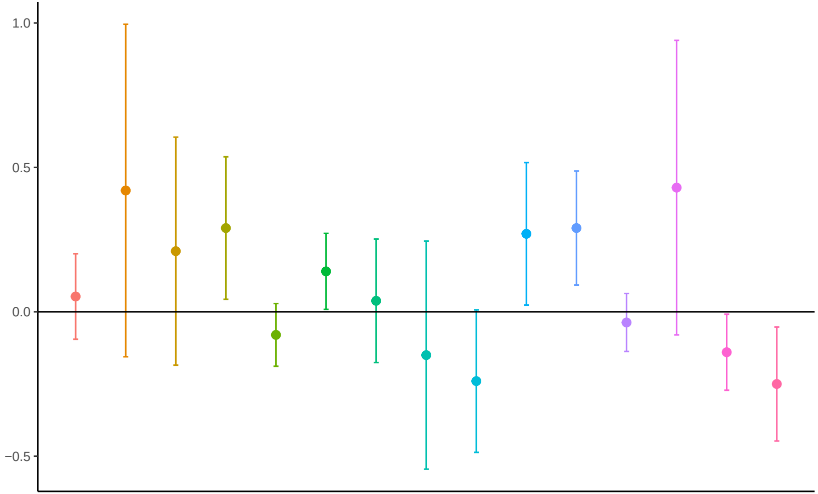


Figure: Effect of Imprisonment on Recidvism - AHC (group-pair LATEs)

Application: Group-pair LATEs (significant only)

Table: Effect of Imprisonment on Recidivism - AHC

	OLS	2SLS	1-5	2-3	3-5	5-6
Prison	0.061	0.095	0.294	0.135	0.286	-0.248
	(0.0095)	(0.022)	(0.154)	(0.078)	(0.121)	(0.114)
J			27	17	18	10
N	23958	23958	7156	5787	5758	2703
Diff			-0.083	0.166	-0.111	-0.163

Cluster-robust standard errors in parentheses.

Conclusion

This paper...

- suggests a cluster method to detect grouped LATEs
- suggests a cluster method to select valid instruments
- applies the method to estimate heterogeneous effects on incarceration on recidivism

Thank you for your attention!

Comments & Questions?

rebecca.groh@tum.de

References

Apfel, N. and X. Liang (2021). Agglomerative Hierarchical Clustering for Selecting Valid Instrumental Variables. *arXiv preprint arXiv:2101.05774*.

Autor, D., A. Kostøl, M. Mogstad, and B. Setzler (2019). Disability benefits, consumption insurance, and household labor supply. *American Economic Review 109*(7), 2613–54.

Bhuller, M., G. B. Dahl, K. V. Løken, and M. Mogstad (2020). Incarceration, recidivism, and employment. *Journal of Political Economy 128*(4), 1269–1324.

Bhuller, M., L. Khoury, and K. V. Løken (2021). Prison, mental health and family spillovers. *Mental Health and Family Spillovers (November 23, 2021). NHH Dept. of Economics Discussion Paper* (19).

Frandsen, B. R., L. J. Lefgren, and E. C. Leslie (2019). Judging judge fixed effects. *NBER working paper 25528, National Bureau of Economic Research*.

Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2022). Contamination bias in linear regressions. Technical report, National Bureau of Economic Research.

Ke, Z. T., J. Fan, and Y. Wu (2015). Homogeneity pursuit. *Journal of the American Statistical Association 110*(509), 175–194.

Loeffler, C. E. and D. S. Nagin (2021). The impact of incarceration on recidivism. *Annual Review of Criminology 5*.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology 66*(5), 688–701.

Ward, J. H. J. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association 58*(301), 236–244.

Appendix: Simulation: Setting w/o Invalidity

- $V \sim Unif(0,1)$
- $U = 0.5V + Unif(0,1)$
- $D = 1(\mathbf{Zp} > V)$
- $\mathbf{Z}$ matrix of judge dummy
- Three clubs: $\mathbf{p} = (0.9\boldsymbol{\iota}_4, 0.7\boldsymbol{\iota}_4, 0.5\boldsymbol{\iota}_2)$
- $Y = (D \cdot 0.5 + D \cdot U + U)^4$
- $J = 10$
- Number of cases: $Unif(3,5)$ multiplied by 10 (few cases setting) or by 100 (many cases setting).
- 1000 repetitions

Simulation: Setting with Invalidity

- $Y = (D \cdot 0.5 + D \cdot U + \mathbf{Z}\gamma + U)^4$
- $$\gamma_{inv} = \begin{matrix} 0.1, 0\boldsymbol{\iota}_3 \\ 0\boldsymbol{\iota}_2, -0.1, -0.2 \\ 0\boldsymbol{\iota}_2 \end{matrix}$$
- Club 1: Three judges are valid and one is invalid (majority and plurality holds)
  Club 2: Two judges are valid and two are invalid but invalidity has different magnitudes (that is, majority is violated but plurality holds)
  Club 3: All judges are valid

Simulation: Extent of effect heterogeneity

Setting	Invalidity	OLS	Overall 2SLS	Oracle Clubs 2SLS 1-2	1-3	2-3
few cases	no	18.94	20.39	30.78	26.47	22.08
	yes	22.75	24.52	29.13	32.02	43.59
many cases	no	18.90	20.39	30.29	26.19	20.10
	yes	22.60	24.40	28.67	31.62	36.26

- Overall 2SLS using judge dummies as IVs would ignore effect heterogeneity.
- Hansen test rejects due to effect heterogeneity and invalidity.

Simulation: 1st clustering based on first stage parameters

- #clubs: Mean number of clusters selected by AHC
- #corr: fraction of times the correct number of clubs has been selected

Setting	Invalidity	Method	#clubs	#corr	Hansen p
few cases	no	Oracle	3	1	0.51
		AHC	2.36	0.36	0.48
	yes	Oracle	3	1	0.00
		AHC	2.35	0.35	0.15
many cases	no	Oracle	3	1	0.51
		AHC	3.02	0.98	0.49
	yes	Oracle	3	1	0.00
		AHC	3.02	0.99	0.00

Simulation: 1st clustering based on first stage parameters

- #clubs: Mean number of clusters selected by AHC
- #corr: fraction of times the correct number of clubs has been selected

Setting	Invalidity	Method	#clubs	#corr	Hansen p
few cases	no	Oracle	3	1	0.51
		AHC	2.36	0.36	0.48
	yes	Oracle	3	1	0.00
		AHC	2.35	0.35	0.15
many cases	no	Oracle	3	1	0.51
		AHC	3.02	0.98	0.49
	yes	Oracle	3	1	0.00
		AHC	3.02	0.99	0.00

- If number of cases is sufficiently large, clustering works well (even if some judge violate LATE assumptions)
- Invalid IV setting: Hansen test still rejects due to invalidity

Simulation: Estimation after 1st clustering

			2SLS		
Setting	Invalidity	Method	1-2	1-3	2-3
many cases	no	Oracle	30.23	26.15	20.01
		AHC Clubs	30.29	26.19	20.09
	yes	Oracle	30.16	26.08	19.94
		AHC Clubs	28.67	31.62	36.25

- Invalid IV setting: LATEs are biased

Simulation: 2nd clustering based on reduced form parameters

- Separately for each club (i.e. constant imprisonment rate) we perform a 2nd clustering based on RF parameters (Hansen test as stopping rule)
- If plurality holds (i.e. largest group of judges are valid), we can identify the valid judges.

Setting	Method	ValDet	InvDet	Hansen p
few cases	Oracle	1	1	0.50
	AHC	0.96	0.33	0.34
many cases	Oracle	1	1	0.51
	AHC	0.99	0.86	0.44

- Again clustering works well if number of cases is large enough
- Hansen test does not reject any longer.

Simulation: Estimation after 1st and 2nd clustering

- Average LATE estimates for the three possible comparisons

Method	1-2	1-3	2-3
Oracle	30.28	26.13	20.12
AHC Clubs & Groups	28.74	26.12	22.37

- Estimation of the LATEs work very well if we drop the invalid judges

Appendix: LATE assumptions

There exist three compliance types under Assumption 3:

Type	Potential treatment variable	Interpretation in application
Compliers (C)	$D^1 = 1, D^0 = 0$	$>$ 2 children only if same sex
~~Defiers (F)~~	~~$D^1 = 0, D^0 = 1$~~	~~$>$ 2 children only if opposite sex~~
Always-takers (A)	$D^1 = 1, D^0 = 1$	$>$ 2 children in any case
Never-takers (N)	$D^1 = 0, D^0 = 0$	2 children in any case

Which are distributed in the data as

	Z=1	Z=0
D=1	C,A	A
D=0	N	C,N

Back

Model

Potential outcomes model (Rubin, 1974)

$$D = 1(\psi(Z) > V) \qquad \text{First Stage relationship}$$

$$Y = \eta(D, U) \qquad \text{Outcome}$$

- V and U are unobservable
- $\psi(\cdot)$ is a nonparametric function of Z
- Binary instruments Z
- Potential treatment: $D(z)$
- Potential outcome: $Y(d, z)$
- Propensity score: $Pr(D = 1|Z = z) = P_z$

Model II: Key implication

Under the LATE Assumptions (Exogeneity and Exclusion, Monotonicity and First Stage), the Wald estimator identifies the LATE $\Delta_{z,z'}$

Central implication

Two IV pairs with values (z, z') and (z'', z^*) identify the same LATE *iff* their propensity scores are equal $p(z) = p(z'')$ and $p(z') = p(z^*)$:

$$\Delta_{z,z'} = E[Y(1) - Y(0)|p(z) > V \geq p(z')] = \Delta_{z'',z*}$$

Intuition: We are looking at the same section of the population in terms of V

Model III: LATE

$$\Delta_{z,z'} = \frac{E(Y|Z=z) - E(Y|Z=z')}{\Pr(D=1|Z=z) - \Pr(D=1|Z=z')}$$
$$= E[Y(1) - Y(0)|D(z)=1, D(z')=0]$$
$$= E[Y(1) - Y(0)|p(z) > V \geq p(z')]$$

Back

Method IV: First-step selection of clubs

1. Cluster the mean incarceration per judge via Agglomerative Hierarchical Clustering (AHC) (Ward, 1963; Apfel and Liang, 2021) Details
2. Select the number of clusters via a stopping rule: Stop when the F-test for equality of all first-stage parameters in the cluster does not reject any more

Method V: Dealing with invalidity

1. Create a set of instrumental variables from two clubs
2. Test overidentifying restrictions (Hansen-Sargan test)
3. Select valid IVs
   - Apply Agglomerative Hierarchical Clustering on the judge-specific means of recidivism

     Details
   - Search for largest cluster (group) inside each club found in the preceding step

Method VI: Agglomerative Hierarchical Clustering

1. **Input:** Calculate all propensity scores $\hat{p}_j$ with a first-stage regression
2. **Initialization:** Each $\hat{p}_j$ has its own cluster. The total number of clusters in the beginning hence is J.
3. **Joining:** The two clusters k and l which are closest in terms of their weighted Euclidean distance $\frac{J_k \cdot J_l}{J_k + J_l} ||\bar{\boldsymbol{p}}_{\boldsymbol{k}} - \bar{\boldsymbol{p}}_{\boldsymbol{l}}||^2$ are joined to a new cluster.
4. **Merging:** Recalculate the cluster means. Recalculate the pair-wise Euclidean distances with the new cluster.
5. **Iteration:** The joining and merging steps are repeated until all just-identified point-estimates are in one cluster. For each joining step, the number of clusters decreases by 1.

Method VI: Agglomerative Hierarchical Clustering

1. **Input:** Calculate all propensity scores $\hat{p}_j$ with a first-stage regression
2. **Initialization:** Each $\hat{p}_j$ has its own cluster. The total number of clusters in the beginning hence is J.
3. **Joining:** The two clusters k and l which are closest in terms of their weighted Euclidean distance $\frac{J_k \cdot J_l}{J_k + J_l} ||\bar{\boldsymbol{p}}_{\boldsymbol{k}} - \bar{\boldsymbol{p}}_{\boldsymbol{l}}||^2$ are joined to a new cluster.
4. **Merging:** Recalculate the cluster means. Recalculate the pair-wise Euclidean distances with the new cluster.
5. **Iteration:** The joining and merging steps are repeated until all just-identified point-estimates are in one cluster. For each joining step, the number of clusters decreases by 1.

Back

Method VII: Consistent classification

Consistent classification

As $(J, T) \to \infty$, classification is individually consistent if the probability of wrongly assigning judges goes to zero for all judges, for all clubs.
Possible wrong assignments:

- **Not** assigning a judge from a certain club C_k^0 to an estimated club $\hat{C}_k$
- Assign a judge from club C_k^0 to an estimated club $\hat{C}_k$ to which it **doesn't** belong

Back

Descriptives

Variable	Overall	$D = 1$	$D = 0$	Diff	pval
Recidivism	0.292	0.3	0.219	0.08	4.4e-37
Female	0.183	0.173	0.276	-0.1	4.3e-53
Age at Sentence	33.04	33.02	33.2	-0.18	0.28
Race					
White	0.59	0.585	0.638	-0.05	6.9e-13
Black	0.261	0.267	0.204	0.06	6.3e-24
Amerindian	0.073	0.074	0.065	0.01	0.021
Hispanic	0.05	0.047	0.074	-0.03	5.6e-12
Asian	0.026	0.027	0.018	0.01	5e-05
Unknown	0	0	0	0	0.73
Crime Type					
Property Crime	0.332	0.324	0.412	-0.09	9.7e-32
Crime against a Person	0.299	0.306	0.235	0.07	8e-28
Drug Crime	0.238	0.242	0.2	0.04	7.5e-12
Sex Offenses	0.056	0.056	0.054	0	0.49
Weapons Offense	0.007	0.007	0.009	0	0.19
Other	0.068	0.065	0.091	-0.03	1.3e-09
Severity	3.44	3.46	3.25	0.21	6.1e-09

Back