Why Propensity Scores Should Be Used for Matching

Ben Jann

University of Bern, ben.jann@soz.unibe.ch

2017 German Stata Users Group Meeting
Berlin, June 23, 2017
Contents

1. Potential Outcomes and Causal Inference
2. Matching
3. Propensity Score Matching
4. King and Nielsen’s “Why Propensity Scores Should Not Be Used for Matching”
5. Are King and Nielsen right?
6. Illustration using kmatch
7. Conclusions
Counterfactual Causality (see Neyman 1923, Rubin 1974, 1990)

- John Stuart Mill (1806–1873)

  Thus, if a person eats of a particular dish, and dies in consequence, that is, would not have died if he had not eaten of it, people would be apt to say that eating of that dish was the cause of his death.

  (Mill 2002[1843]:214)
Counterfactual Causality (see Neyman 1923, Rubin 1974, 1990)

- Treatment variable $D$
  \[ D = \begin{cases} 
  1 & \text{treatment (eats of a particular dish)} \\
  0 & \text{control (does not eat of a particular dish)} 
  \end{cases} \]

- Potential outcomes $Y^1$ and $Y^0$
  - $Y^1$: potential outcome with treatment ($D = 1$)
    - If person $i$ would eat of a particular dish, would she die or would she survive?
  - $Y^0$: potential outcome without treatment ($D = 0$)
    - If person $i$ would not eat of a particular dish, would she die or would she survive?

- Causal effect of the treatment for individual $i$:
  \[ \delta_i = Y^1_i - Y^0_i \]
Fundamental Problem of Causal Inference

- The causal effect of $D$ on $Y$ for individual $i$ is defined as the difference in potential outcomes: $\delta_i = Y^1_i - Y^0_i$

- However, the observed outcome variable is

$$Y_i = \begin{cases} 
Y^1_i & \text{if } D_i = 1 \\
Y^0_i & \text{if } D_i = 0 
\end{cases}$$

- That is, only one of the two potential outcomes will be realized and, hence, only $Y^1_i$ or $Y^0_i$ can be observed, but never both.

- Consequence:

  The individual treatment effect $\delta_i$ cannot be observed!
Average Treatment Effect

- Although individual causal effects cannot be observed, the average causal effect in a population (the so-called “Average Treatment Effect”) can be identified comparing the expected values of $Y^1$ and $Y^0$:

$$ATE = E[\delta] = E[Y^1 - Y^0] = E[Y^1] - E[Y^0]$$

- Some other quantities of interest:
  - Average Treatment Effect on the Treated (ATT)
    $$ATT = E[Y^1 - Y^0 | D = 1] = E[Y^1 | D = 1] - E[Y^0 | D = 1]$$
  - Average Treatment Effect on the Untreated (ATC)
    $$ATC = E[Y^1 - Y^0 | D = 0] = E[Y^1 | D = 0] - E[Y^0 | D = 0]$$
Average Treatment Effect

- To determine the average effect, unbiased estimates of $E[Y^0]$ and $E[Y^1]$ are required.
- If the independence assumption

$$ (Y^0, Y^1) \perp \perp D $$

applies, that is, if $D$ is independent from $Y^0$ and $Y^1$, then

$$ E[Y^0] = E[Y^0|D = 0] $$
$$ E[Y^1] = E[Y^1|D = 1] $$

- In this case the average causal effect can be measured by a simple group comparison (mean difference) of observations without treatment ($D = 0$) and observations with treatment ($D = 1$).
- Randomized experiments solve the problem: If the assignment of $D$ is randomized, $D$ is independent from $Y^0$ and $Y^1$ by design.
1 Potential Outcomes and Causal Inference

2 Matching

3 Propensity Score Matching

4 King and Nielsen’s “Why Propensity Scores Should Not Be Used for Matching”

5 Are King and Nielsen right?

6 Illustration using kmatch

7 Conclusions
Conditional Independence / Strong Ignorability

- Can causal effects also be identified from “observational” (i.e. non-experimental) data?
- Sometimes it can be argued that the independence assumption is valid *conditionally* (conditional independence, “unconfoundedness”):
  \[(Y^0, Y^1) \perp \! \! \! \perp D \mid X\]
- If, in addition, the overlap assumption
  \[0 < \Pr(D = 1 \mid X = x) < 1, \text{ for all } x\]
  is given, then the ATE (or ATT or ATC) can be identified by conditioning on \(X\).
- For example:
  \[ATE = \sum_x \Pr[X = x] \{E[Y \mid D = 1, X = x] - E[Y \mid D = 0, X = x]\}\]
Matching

- Matching is one approach to “condition on $X$” if strong ignorability holds.
- Basic idea:
  1. For each observation in the treatment group, find “statistical twins” in the control group with the same (or at least very similar) $X$ values (and vice versa).
  2. The $Y$ values of these matching observations are then used to compute the counterfactual outcome for the observation at hand.
  3. An estimate for the average causal effect can be obtained as the mean of the differences between the observed values and the “imputed” counterfactual values over all observations.
Matching

- Formally:

\[
\hat{ATT} = \frac{1}{N_{D=1}} \sum_{i \mid D=1} \left[ Y_i - \hat{Y}_i^0 \right] = \frac{1}{N_{D=1}} \sum_{i \mid D=1} \left[ Y_i - \sum_{j \mid D=0} w_{ij} Y_j \right]
\]

\[
\hat{ATC} = \frac{1}{N_{D=0}} \sum_{i \mid D=0} \left[ \hat{Y}_i^1 - Y_i \right] = \frac{1}{N_{D=0}} \sum_{i \mid D=0} \left[ \sum_{j \mid D=1} w_{ij} Y_j - Y_i \right]
\]

\[
\hat{ATE} = \frac{N_{D=1}}{N} \cdot \hat{ATT} + \frac{N_{D=0}}{N} \cdot \hat{ATC}
\]

- Different matching algorithms use different definitions of \( w_{ij} \).
Exact Matching

- Exact matching:
  \[ w_{ij} = \begin{cases} 
  1/k_i & \text{if } X_i = X_j \\
  0 & \text{else} 
  \end{cases} \]

  with \( k_i \) as the number of observations for which \( X_i = X_j \) applies.

- The result equivalent to “perfect stratification” or “subclassification” (see, e.g., Cochran 1968).

- Problem: If \( X \) contains several variables there is a large probability that no exact matches can be found for many observations (the “curse of dimensionality”).
An alternative is to match based on a distance metric that measures the proximity between observations in the multivariate space of $X$.

The idea then is to use observations that are “close”, but not necessarily equal, as matches.

A common approach is to use

$$MD(X_i, X_j) = \sqrt{(X_i - X_j)'\Sigma^{-1}(X_i - X_j)}$$

as distance metric, where $\Sigma$ is an appropriate scaling matrix.

- Mahalanobis matching: $\Sigma$ is the covariance matrix of $X$.
- Euclidean matching: $\Sigma$ is the identity matrix.
- Mahalanobis matching is equivalent to Euclidean matching based on standardized and orthogonalized $X$. 

Matching Algorithms

- Various matching algorithms can be employed to find potential matches based on $MD$, and determine the matching weights $w_{ij}$.

- Pair matching (one-to-one matching without replacement)
  - For each observation $i$ in the treatment group find observation $j$ in the control group for which $MD_{ij}$ is smallest. Once observation $j$ is used as a match, do not use it again.

- Nearest-neighbor matching
  - For each observation $i$ in the treatment group find the $k$ closest observations in the control group. A single control can be used multiple times as a match. In case of ties (multiple controls with identical $MD$), use all ties as matches. $k$ is set by the researcher.

- Caliper matching
  - Like nearest-neighbor matching, but only use controls for which $MD$ is smaller than some threshold $c$. 
Mahalanobis Matching

- **Radius matching**
  - Use *all* controls as matches for which $MD$ is smaller than some threshold $c$.

- **Kernel matching**
  - Like radius matching, but give larger weight to controls for which $MD$ is small (using some kernel function such as, e.g., the Epanechnikov kernel).

- In addition, since matching is no longer exact, it may make sense to refine the estimates by applying regression-adjustment to the matched data (also known as “bias-adjustment” in the context of nearest-neighbor matching).
1. Potential Outcomes and Causal Inference

2. Matching

3. Propensity Score Matching

4. King and Nielsen’s “Why Propensity Scores Should Not Be Used for Matching”

5. Are King and Nielsen right?

6. Illustration using kmatch

7. Conclusions
The Propensity Score Theorem (Rosenbaum and Rubin 1983)

- If the conditional independence assumption is true, then

\[
\Pr(D_i = 1 | Y_{i0}, Y_{i1}, X_i) = \Pr(D_i = 1 | X_i) = \pi(X_i)
\]

where \(\pi(X)\) is called the propensity score.

- That is,

\[
(Y^0, Y^1) \perp\!\!\!\!\!\!\!\!\!\!\!\!\perp D | X
\]

implies

\[
(Y^0, Y^1) \perp\!\!\!\!\!\!\!\!\!\!\!\!\perp D | \pi(X)
\]

so that under strong ignorability the average causal effect can be estimated by conditioning on the propensity score \(\pi(X)\) instead of \(X\).

- This is remarkable, because the information in \(X\), which may include many variables, can be reduced to just one dimension. This greatly simplifies the matching task.
Instead of computing multivariate distances, we can thus simply match on the (one-dimensional) propensity score.

Procedure
- Step 1: Estimate the propensity score, e.g. using a Logit model.
- Step 2: Apply a matching algorithm using differences in the propensity score, $|\hat{\pi}(X_i) - \hat{\pi}(X_j)|$, instead of multivariate distances.

PSM is tremendously popular
- https://scholar.google.ch/scholar?q="propensity+score"+AND+(matching+OR+matched+OR+match)
Potential Outcomes and Causal Inference

Matching

Propensity Score Matching

King and Nielsen’s “Why Propensity Scores Should Not Be Used for Matching”

Are King and Nielsen right?

Illustration using kmatch

Conclusions
In 2015/2016 Gary King and Richard Nielsen circulated a paper that created quite some concern among applied researchers.

The basic message of the paper is that PSM is really, really bad and should be discarded.

The paper
   ▶ http://j.mp/1sexgVw

Slides
   ▶ https://gking.harvard.edu/presentations/
     why-propensity-scores-should-not-be-used-matching-6

Watch it
   ▶ https://www.youtube.com/watch?v=rBv39pK1iEs
The story goes about as follows.

Argument 1

- Model dependence (i.e. dependence of results on modeling decisions made by the researcher) is bad because it leads to bias (people are selective in their decisions even if they try not to be).
- Matching is good because it reduces model dependence.
Matching to Reduce Model Dependence
(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)

- Education (years)
- Outcome

![Graph with T symbols indicating data points, showing the relationship between education and outcome.](image)
Matching to Reduce Model Dependence
(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)

![Graph showing the relationship between education (years) and outcome. The graph includes data points labeled T and C, indicating treatment and control groups respectively.](slides by King and Nielsen)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)

Education (years)

Outcome

C
C
C
C
C
C
C
C
C
C
C
C
C
C
C
C

T
T
T
T
T
T
T
T
T
T
T
T
T
T
T
T

3/23
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)

Education (years) vs Outcome

(slides by King and Nielsen)
Matching to Reduce Model Dependence
(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)

<table>
<thead>
<tr>
<th>Education (years)</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>C</td>
</tr>
<tr>
<td>14</td>
<td>T</td>
</tr>
<tr>
<td>16</td>
<td>C</td>
</tr>
<tr>
<td>18</td>
<td>T</td>
</tr>
<tr>
<td>20</td>
<td>C</td>
</tr>
<tr>
<td>22</td>
<td>T</td>
</tr>
<tr>
<td>24</td>
<td>C</td>
</tr>
<tr>
<td>26</td>
<td>T</td>
</tr>
<tr>
<td>28</td>
<td>C</td>
</tr>
</tbody>
</table>

(slides by King and Nielsen)
PSM approximates complete randomization.
Better are matching approaches that approximate fully blocked randomization, such as Mahalanobis matching, because complete randomization is less efficient than fully blocked randomization.

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td>On average</td>
</tr>
<tr>
<td>Unobserved</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

Goal of Each Matching Method (in Observational Data)
- PSM: complete randomization
- Other methods: fully blocked
- Other matching methods dominate PSM (wait, it gets worse)
Best Case: Mahalanobis Distance Matching
Best Case: Propensity Score Matching

Propensity Score

15/23

(slides by King and Nielsen)
Best Case: Propensity Score Matching is Suboptimal
Argument 3

- Random pruning (deleting observations at random) increases imbalance. This is because the sample size decreases so that variance increases (large differences become more likely).
- More imbalance/variance means more model dependence and researcher discretion.
- Because PSM approximates complete randomization, it engages in random pruning.
- PSM Paradox (“when you do ‘better,’ you do worse”)
  - When matching is made more strict (e.g., by decreasing the size of the caliper) PSM, like other matching methods, typically reduces imbalance. But soon the PSM Paradox kicks in, such that further pruning quickly increases imbalance.
  - If the data is such that there are no big differences between treated and untreated to begin with, the PSM Paradox kicks in almost immediately.
PSM Increases Model Dependence & Bias

Model Dependence

Bias

\[ Y_i = 2T_i + X_{1i} + X_{2i} + \epsilon_i \]
\[ \epsilon_i \sim N(0, 1) \]
The Propensity Score Paradox in Real Data

Finkel et al. (JOP, 2012) vs. Nielsen et al. (AJPS, 2011)

Similar pattern for > 20 other real data sets we checked

(slides by King and Nielsen)
Potential Outcomes and Causal Inference

Matching

Propensity Score Matching

King and Nielsen’s “Why Propensity Scores Should Not Be Used for Matching”

Are King and Nielsen right?

Illustration using kmatch

Conclusions
Are King and Nielsen right?

- Argument 1
  - Model dependence (i.e. dependence of results on modeling decisions made by the researcher) is bad because it leads to bias (people are selective in their decisions even if they try not to be).
  - Matching is good because it reduces model dependence.

- I fully agree!

- My view, however, may be somewhat less pessimistic. I believe that research results can be credible if researchers are well educated so that they know what they are doing and if modeling decisions are made transparent and robustness of results is evaluated (and documented).
Are King and Nielsen right?

Argument 2

- PSM approximates complete randomization.
- Better are matching approaches that approximate fully blocked randomization, such as Mahalanobis matching, because complete randomization is less efficient than fully blocked randomization.

That fully blocked randomization is more efficient than complete randomization – given the sample size – is of course true (how large the efficiency gains are depends on the strength of the relation between $X$ and $Y$).

However, if blocking reduces the sample size, it is not a priori clear whether estimates from the blocked sample are more efficient than estimates from the full sample (although often they will be).
Are King and Nielsen right?

- **Argument 2**
  - PSM approximates complete randomization.
  - Better are matching approaches that approximate fully blocked randomization, such as Mahalanobis matching, because complete randomization is less efficient than fully blocked randomization.

- That PSM approximates complete randomization is only partially true. PSM approximates complete randomization *within observations with the same propensity score*. Hence, PSM is somewhere between complete randomization and fully blocked randomization.
  - If the $X$ variables have no relation to $T$ (treatment), then all observations have the same propensity score. Hence we end up with complete randomization.
  - If the $X$ variables have a strong effect on $T$, there is lots of blocking.
Are King and Nielsen right?

- Argument 3
  - Random pruning $\Rightarrow$ imbalance $\Rightarrow$ more model dependence.
  - PSM $\Rightarrow$ complete randomization $\Rightarrow$ lots of random pruning.
  - PSM Paradox: “when you do ‘better,’ you do worse”

- That random pruning makes things worse is, of course, true because it unnecessarily reduces the sample size (without changing anything else).

- As argued above, that PSM applies random pruning is only true for $X$ variables unrelated to $T$ (so that we are in a “local” complete randomization situation; although something similar can probably also happen if effects from several $X$'s cancel each other out).

- Furthermore, it is only true if you employ a matching algorithm that throws away good matches! King and Nielsen’s results seem to be based on the worst possible algorithm: one-to-one matching without replacement.
Are King and Nielsen right?

- Argument 3
  - Random pruning $\Rightarrow$ imbalance $\Rightarrow$ more model dependence.
  - PSM $\Rightarrow$ complete randomization $\Rightarrow$ lots of random pruning.
  - PSM Paradox: “when you do ‘better,’ you do worse”

- If you use a matching algorithm that does not throw away good matches, such as radius or kernel matching (or also nearest-neighbor matching as long as all ties are kept and observations are matched with replacement), random pruning can be avoided.
  - Such algorithms block (and hence prune) where it is necessary to prevent bias, but they average where such pruning is not necessary.
  - Hence, efficiency differences between PSM and multivariate matching should only be minor for such algorithms.
Are King and Nielsen right?

- Argument 3
  - Random pruning $\Rightarrow$ imbalance $\Rightarrow$ more model dependence.
  - PSM $\Rightarrow$ complete randomization $\Rightarrow$ lots of random pruning.
  - PSM Paradox: “when you do ‘better,’ you do worse”

- True is that post-matching modeling can do more harm with PSM than with MDM (because PSM leaves more “free” variance in $X$ that can exploited by modeling decisions).

- In general, post-matching analyses are more limited for PSM than for MDM. For example, results from subgroup analyses will not be valid (you’d need to apply PSM stratified by subgroups in this case).
1 Potential Outcomes and Causal Inference

2 Matching

3 Propensity Score Matching

4 King and Nielsen’s “Why Propensity Scores Should Not Be Used for Matching”

5 Are King and Nielsen right?

6 Illustration using kmatch

7 Conclusions
The Command

- `kmatch`: new matching software for Stata that has been written over the last few months; available from SSC (`ssc install kmatch`).

Some key features:

- Multivariate Distance Matching (MDM) and Propensity Score Matching (PSM) (or MDM and PSM combined).
- Optional exact matching.
- Optional regression-adjustment bias-correction.
- Kernel matching, ridge matching, or nearest-neighbor matching.
- Automatic bandwidth selection for kernel/ridge matching.
- Flexible specification of scaling matrix for MDM.
- Joint analysis of multiple subgroups and multiple outcome variables.
- Various post-estimation commands for balancing and common-support diagnostics.
- Computationally efficient implementation.
Some Examples

. // Use the NLSW data to estimate the "effect" of union membership on
. // wages, controlling for some covariated such as education, labor market
. // experience, or industry
. sysuse nlsw88, clear
(NLSW, 1988 extract)
. drop if industry==2
(4 observations deleted)
. // Mahalanobis-distance kernel matching
. kmatch md union collgrad ttl_exp tenure i.industry i.race south ///
> (wage), nate att
(constructing bandwidth ... done)

Multivariate-distance kernel matching
Number of obs = 1,853
Kernel = epan
Treatment : union = 1
Metric : mahalanobis
Covariates: collgrad ttl_exp tenure i.industry i.race south
Matching statistics

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th></th>
<th></th>
<th>Controls</th>
<th></th>
<th></th>
<th>Band-</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Total</td>
<td>Used</td>
<td>Unused</td>
<td>Total</td>
<td>width</td>
</tr>
<tr>
<td>Treated</td>
<td>432</td>
<td>25</td>
<td>457</td>
<td>1105</td>
<td>291</td>
<td>1396</td>
<td>1.3394</td>
</tr>
</tbody>
</table>

Treatment-effects estimation

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>wage</td>
<td></td>
</tr>
<tr>
<td>ATT</td>
<td>0.6059013</td>
</tr>
<tr>
<td>NATE</td>
<td>1.432913</td>
</tr>
</tbody>
</table>
Some Examples

// some balancing statistics
kmatch summarize(refitting the model using the generate() option)

<table>
<thead>
<tr>
<th>Means</th>
<th>Raw Treated</th>
<th>Untrea-d</th>
<th>StdDiff</th>
<th>Matched(ATT) Treated</th>
<th>Untrea-d</th>
<th>StdDiff</th>
</tr>
</thead>
<tbody>
<tr>
<td>collgrad</td>
<td>.321663</td>
<td>.224212</td>
<td>.219912</td>
<td>.319444</td>
<td>.319444</td>
<td>0</td>
</tr>
<tr>
<td>ttl_exp</td>
<td>13.2685</td>
<td>12.7323</td>
<td>.117584</td>
<td>13.3205</td>
<td>13.1425</td>
<td>.039036</td>
</tr>
<tr>
<td>tenure</td>
<td>7.89205</td>
<td>6.17658</td>
<td>.29735</td>
<td>7.91744</td>
<td>7.58347</td>
<td>.057888</td>
</tr>
<tr>
<td>3.industry</td>
<td>.006565</td>
<td>.012178</td>
<td>-.058246</td>
<td>.00463</td>
<td>.00463</td>
<td>0</td>
</tr>
<tr>
<td>4.industry</td>
<td>.183807</td>
<td>.166905</td>
<td>.044425</td>
<td>.185185</td>
<td>.185185</td>
<td>0</td>
</tr>
<tr>
<td>5.industry</td>
<td>.105033</td>
<td>.027937</td>
<td>.312944</td>
<td>.085648</td>
<td>.085648</td>
<td>0</td>
</tr>
<tr>
<td>6.industry</td>
<td>.045952</td>
<td>.169771</td>
<td>-.407129</td>
<td>.048611</td>
<td>.048611</td>
<td>0</td>
</tr>
<tr>
<td>7.industry</td>
<td>.019694</td>
<td>.102436</td>
<td>-.35067</td>
<td>.020833</td>
<td>.020833</td>
<td>0</td>
</tr>
<tr>
<td>8.industry</td>
<td>.017505</td>
<td>.035817</td>
<td>-.113785</td>
<td>.009259</td>
<td>.009259</td>
<td>0</td>
</tr>
<tr>
<td>9.industry</td>
<td>.010941</td>
<td>.040115</td>
<td>-.18566</td>
<td>.011574</td>
<td>.011574</td>
<td>0</td>
</tr>
<tr>
<td>10.industry</td>
<td>.004376</td>
<td>.008596</td>
<td>-.052551</td>
<td>.002315</td>
<td>.002315</td>
<td>0</td>
</tr>
<tr>
<td>11.industry</td>
<td>.479212</td>
<td>.356734</td>
<td>.250073</td>
<td>.506944</td>
<td>.506944</td>
<td>0</td>
</tr>
<tr>
<td>12.industry</td>
<td>.122538</td>
<td>.07235</td>
<td>.169707</td>
<td>.12037</td>
<td>.12037</td>
<td>0</td>
</tr>
<tr>
<td>2.race</td>
<td>.330416</td>
<td>.244986</td>
<td>.189418</td>
<td>.3125</td>
<td>.3125</td>
<td>0</td>
</tr>
<tr>
<td>3.race</td>
<td>.017505</td>
<td>.011461</td>
<td>.050566</td>
<td>.006944</td>
<td>.006944</td>
<td>0</td>
</tr>
<tr>
<td>south</td>
<td>.297593</td>
<td>.466332</td>
<td>-.352408</td>
<td>.291667</td>
<td>.291667</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variances</th>
<th>Raw Treated</th>
<th>Untrea-d</th>
<th>Ratio</th>
<th>Matched(ATT) Treated</th>
<th>Untrea-d</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>collgrad</td>
<td>.218674</td>
<td>.174066</td>
<td>1.25628</td>
<td>.217904</td>
<td>.217904</td>
<td>1</td>
</tr>
<tr>
<td>ttl_exp</td>
<td>20.5898</td>
<td>21.0001</td>
<td>.980459</td>
<td>19.8177</td>
<td>18.2323</td>
<td>1.08696</td>
</tr>
<tr>
<td>tenure</td>
<td>37.2044</td>
<td>29.3629</td>
<td>1.26706</td>
<td>37.0399</td>
<td>34.9543</td>
<td>1.05966</td>
</tr>
<tr>
<td>3.industry</td>
<td>.006536</td>
<td>.012038</td>
<td>.542928</td>
<td>.004619</td>
<td>.004619</td>
<td>1</td>
</tr>
<tr>
<td>4.industry</td>
<td>.150351</td>
<td>.139148</td>
<td>1.08052</td>
<td>.151242</td>
<td>.151242</td>
<td>1</td>
</tr>
<tr>
<td>5.industry</td>
<td>.094207</td>
<td>.027176</td>
<td>3.46656</td>
<td>.078494</td>
<td>.078494</td>
<td>1</td>
</tr>
<tr>
<td>6.industry</td>
<td>.043936</td>
<td>.14105</td>
<td>.311496</td>
<td>.046355</td>
<td>.046355</td>
<td>1</td>
</tr>
<tr>
<td>7.industry</td>
<td>.019348</td>
<td>.092008</td>
<td>.210287</td>
<td>.020447</td>
<td>.020447</td>
<td>1</td>
</tr>
<tr>
<td>8.industry</td>
<td>.017237</td>
<td>.034559</td>
<td>.498769</td>
<td>.002315</td>
<td>.002315</td>
<td>1</td>
</tr>
</tbody>
</table>
Some Examples

. // make a graph of the balancing stats
. mat M = r(M)
. mat V = r(V)
. coefplot matrix(M[,3]) matrix(M[,6]) || matrix(V[,3]) matrix(V[,6]) || , ///
> bylabels("Std. mean difference" "Variance ratio") ///
> noci nolabels byopts(xrescale)
. addplot 1: , xline(0) norescaling legend(order(1 "Raw" 2 "Matched"))
. addplot 2: , xline(1) norescaling

![Graph of balancing stats](image-url)
. // Propensity-score kernel matching
. kmatch ps union collgrad ttl_exp tenure i.industry i.race south ///
> (wage), nate att
(computing bandwidth ... done)

Propensity-score kernel matching
Number of obs = 1,853
Kernel = epan

Treatment : union = 1
Covariates: collgrad ttl_exp tenure i.industry i.race south
PS model : logit (pr)

Matching statistics

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated</td>
<td>431</td>
<td>26</td>
<td>457</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Used</th>
<th>Unused</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated</td>
<td>1214</td>
<td>182</td>
<td>1396</td>
</tr>
</tbody>
</table>

Bandwidth = .00188

Treatment-effects estimation

<table>
<thead>
<tr>
<th>wage</th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATT</td>
<td>.3887224</td>
</tr>
<tr>
<td>NATE</td>
<td>1.432913</td>
</tr>
</tbody>
</table>
Some Examples

. // Kernel density balancing plot
. kmatch density, lw(*6 *2) lc(*.5 *1)
(refitting the model using the generate() option)
(applying 0-1 boundary correction to density estimation of propensity score)
(bandwidth for propensity score = .06803989)
Some Examples

. // Cumulative distribution balancing plot
. kmatch cumul, lw(*6 *2) lc(*.5 *1)
(refitting the model using the generate() option)
Some Examples

.// Balancing box plot
.kmatch box
(refitting the model using the generate() option)
Some Examples

. // Standard errors
. kmatch md union collgrad ttl_exp tenure i.industry i.race south ///
> (wage), nate ate att atc vce(bootstrap)
(computing bandwidth for treated ... done)
(computing bandwidth for untreated ... done)
(running kmatch on estimation sample)

Bootstrap replications (50)

1 2 3 4 5 .................................................. 50

Multivariate-distance kernel matching

Number of obs = 1,853
Replications = 50
Kernel = epan

Treatment : union = 1
Metric : mahalanobis
Covariates: collgrad ttl_exp tenure i.industry i.race south

Matching statistics

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th>Controls</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Total</td>
</tr>
<tr>
<td>Treated</td>
<td>432</td>
<td>25</td>
<td>457</td>
</tr>
<tr>
<td>Untreated</td>
<td>1386</td>
<td>10</td>
<td>1396</td>
</tr>
<tr>
<td>Combined</td>
<td>1818</td>
<td>35</td>
<td>1853</td>
</tr>
</tbody>
</table>

Treatment-effects estimation

|         | Observed Coef. | Bootstrap Std. Err. | z     | P>|z| | Normal-based [95% Conf. Interval] |
|---------|----------------|---------------------|-------|-------|----------------------------------|
| wage    |                |                     |       |       |                                  |
| ATE     | .4095729       | .1920853            | 2.13  | 0.033 | .0330928                         | .7860531 |
| ATT     | .6059013       | .2472069            | 2.45  | 0.014 | .1213846                         | 1.090418 |
| ATC     | .3483797       | .1893653            | 1.84  | 0.006 | -.0227695                        | .7195289 |
| NATE    | 1.432913       | .2333282            | 6.14  | 0.000 | .9755981                         | 1.890228 |
Some Examples

. // Do some tests
. lincom ATT-NATE
( 1) ATT - NATE = 0

| wage | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|------|--------|-----------|-------|------|----------------------|
| (1)  | -.8270117 | .1810415 | -4.57 | 0.000 | -1.181847 -.4721768 |

. test ATT = ATC
( 1) ATT - ATC = 0

    chi2( 1) =  2.42
    Prob > chi2 =  0.1200
Some Examples

. // Nearest-neighbor matching (1 neighbor)
. kmatch md union collgrad ttl_exp tenure i.industry i.race south (wage), att nn

Multivariate-distance nearest-neighbor matching

Number of obs = 1,853
Neighbors: min = 1
max = 1
Treatment : union = 1
Metric : mahalanobis
Covariates: collgrad ttl_exp tenure i.industry i.race south

Matching statistics

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Treated</td>
<td>457</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>328</td>
<td>1068</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>

Treatment-effects estimation

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>wage</td>
<td>.7246969</td>
</tr>
</tbody>
</table>

. teffects nnmatch (wage collgrad ttl_exp tenure i.industry i.race south) (union), atet

Treatment-effects estimation

Number of obs = 1,853
Estimator : nearest-neighbor matching
Matches: requested = 1
Outcome model : matching
Distance metric: Mahalanobis

|       | AI Robust Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-------|----------------|-----------|-------|------|---------------------|
| wage  | .7246969       | .2942952  | 2.46  | 0.014 | .147889  1.301505   |
| ATET  | union          |           |       |      |                     |
| (union vs nonunion) | .7246969 | .2942952 | 2.46  | 0.014 | .147889  1.301505   |
Some Examples

. // Nearest-neighbor matching (5 neighbors)
. kmatch md union collgrad ttl_exp tenure i.industry i.race south (wage), att nn(5)
Multivariate-distance nearest-neighbor matching

Number of obs = 1,853
Neighbors: min = 5
max = 5
Treatment : union = 1
Metric : mahalanobis
Covariates: collgrad ttl_exp tenure i.industry i.race south

Matching statistics

<table>
<thead>
<tr>
<th>Matched Yes</th>
<th>Matched No</th>
<th>Matched Total</th>
<th>Controls Used</th>
<th>Controls Unused</th>
<th>Controls Total</th>
<th>Band-width</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated</td>
<td>457</td>
<td>0</td>
<td>457</td>
<td>870</td>
<td>526</td>
<td>1396</td>
</tr>
</tbody>
</table>

Treatment-effects estimation

<table>
<thead>
<tr>
<th>wage</th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATT</td>
<td>.5590823</td>
</tr>
</tbody>
</table>

. tefects nnmatch (wage collgrad ttl_exp tenure i.industry i.race south) (union), atet nn(5)

Treatment-effects estimation

Number of obs = 1,853
Estimator : nearest-neighbor matching
Matches: requested = 5
Outcome model : matching
Distance metric: Mahalanobis

| wage | Coef. | Std. Err. | z    | P>|z|   | [95% Conf. Interval] |
|------|-------|-----------|------|-------|----------------------|
| ATET union (union vs nonunion) | .5590823 | .2381752 | 2.35 | 0.019 | .0922675 1.025897 |
Some Examples

. // Bias-correction / regression adjustment
. kmatch md union collgrad ttl_exp tenure i.industry i.race south ///
> (wage = collgrad ttl_exp tenure i.industry i.race south), att nn(5)

Multivariate-distance nearest-neighbor matching

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Treated</td>
<td>457</td>
<td>0</td>
</tr>
</tbody>
</table>

Treatment-effects estimation

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>wage</td>
<td>.5288023</td>
</tr>
<tr>
<td>ATT</td>
<td></td>
</tr>
</tbody>
</table>

adjusted for collgrad ttl_exp tenure i.industry i.race south

. teffects nnmatch (wage collgrad ttl_exp tenure i.industry i.race south) ///
> (union), atet nn(5) biasadj(collgrad ttl_exp tenure i.industry i.race south)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AI</td>
<td>Robust</td>
<td></td>
<td></td>
<td>95% Conf. Interval</td>
</tr>
<tr>
<td></td>
<td>Coef.</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>wage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATET</td>
<td>union</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(union vs nonunion)</td>
<td>.5288023</td>
<td>.2420635</td>
<td>2.18</td>
<td>0.029</td>
<td>.0543666</td>
</tr>
</tbody>
</table>
Some Examples

. // Mahalanobis-distance and propensity-score matching combined
. kmatch md union collgrad ttl_exp tenure (wage), att ///
>    psvars(i.industry i.race south) psweight(3)
(computing bandwidth ... done)

Multivariate-distance kernel matching

Number of obs = 1,853
Kernel = epan

Treatment : union = 1
Metric : mahalanobis (modified)
Covariates: collgrad ttl_exp tenure
PS model : logit (pr)
PS covars : i.industry i.race south

Matching statistics

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Total</td>
<td>Used</td>
</tr>
<tr>
<td>Treated</td>
<td></td>
<td>439</td>
<td>18</td>
<td>457</td>
<td>1258</td>
</tr>
</tbody>
</table>

Treatment-effects estimation

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>wage</td>
<td>.6408443</td>
</tr>
</tbody>
</table>

Ben Jann (University of Bern)
. // Exact matching
. kmatch md union collgrad ttl_exp tenure (wage), att ematch(industry race south)
(computing bandwidth ... done)

Multivariate-distance kernel matching

<table>
<thead>
<tr>
<th>Matched</th>
<th></th>
<th>Total</th>
<th>Controls</th>
<th>Controls</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated</td>
<td>Yes</td>
<td>No</td>
<td>Total</td>
<td>Used</td>
<td>Unused</td>
</tr>
</tbody>
</table>

Number of obs = 1,853
Kernel = epan

Treatment : union = 1
Metric : mahalanobis
Covariates: collgrad ttl_exp tenure
Exact : industry race south

Matching statistics

Treated: 432 25 457 1103 293 1396 1.3013

Treatment-effects estimation

<table>
<thead>
<tr>
<th>wage</th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATT</td>
<td>.6047374</td>
</tr>
</tbody>
</table>
Some Examples

. // Bandwidth selection: the default (based on distribution of distances in one-nearest-neighbor matching)
. kmatch md union collgrad ttl_exp tenure i.industry i.race south (wage), att (computing bandwidth ... done)

Multivariate-distance kernel matching

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th>Controls</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Total</td>
</tr>
<tr>
<td>Treated</td>
<td>432</td>
<td>25</td>
<td>457</td>
</tr>
</tbody>
</table>

Treatment-effects estimation

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>wage</td>
<td></td>
</tr>
<tr>
<td>ATT</td>
<td>.6059013</td>
</tr>
</tbody>
</table>
Some Examples

. // Bandwidth selection: cross validation with respect to X
. kmatch md union collgrad ttl_exp tenure i.industry i.race south (wage), ///
>       att bwidth(cv)
(computing bandwidth ................. done)

Multivariate-distance kernel matching

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated</td>
<td>Used</td>
<td>Unused</td>
</tr>
<tr>
<td>Yes</td>
<td>448</td>
<td>1184</td>
</tr>
<tr>
<td>No</td>
<td>9</td>
<td>212</td>
</tr>
<tr>
<td>Total</td>
<td>457</td>
<td>1396</td>
</tr>
</tbody>
</table>

Treatment-effects estimation

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>wage</td>
<td>.6651578</td>
</tr>
</tbody>
</table>

. kmatch cvplot, ms(o) index mlabposition(1) sort

![Graph showing MSE vs. Bandwidth](attachment:image.png)
Some Examples

. // Bandwidth selection: cross validation with respect to Y
. kmatch md union collgrad ttl_exp tenure i.industry i.race south (wage), ///
>    att bwidth(cv wage)
(computing bandwidth ................. done)

Multivariate-distance kernel matching

Number of obs = 1,853
Kernel = epan

Treatment : union = 1
Metric : mahalanobis
Covariates: collgrad ttl_exp tenure i.industry i.race south

Matching statistics

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Treated</td>
<td>453</td>
<td>4</td>
</tr>
</tbody>
</table>

Bandwidth = 2.433

Treatment-effects estimation

<table>
<thead>
<tr>
<th>wage</th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATT</td>
<td>0.6928956</td>
</tr>
</tbody>
</table>

. kmatch cvplot, ms(o) index mlabposition(1) sort
Some Examples

. // Bandwidth selection: weighted cross validation with respect to Y
. kmatch md union collgrad ttl_exp tenure i.industry i.race south (wage), //>
>      att bwidth(cv wage, weighted)
(computing bandwidth ............... done)

Multivariate-distance kernel matching
Number of obs = 1,853
Kernel = epan

Treatment : union = 1
Metric : mahalanobis
Covariates: collgrad ttl_exp tenure i.industry i.race south

Matching statistics

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th>No</th>
<th>Total</th>
<th>Controls</th>
<th>Used</th>
<th>Unused</th>
<th>Total</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated</td>
<td>455</td>
<td>2</td>
<td>457</td>
<td>1356</td>
<td>40</td>
<td>1396</td>
<td></td>
<td>2.7626</td>
</tr>
</tbody>
</table>

Treatment-effects estimation

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>wage</td>
<td>.7308166</td>
</tr>
<tr>
<td>ATT</td>
<td>.7308166</td>
</tr>
</tbody>
</table>

. kmatch cvplot, ms(o) index mlabposition(1) sort
Some Examples

. // Common-support diagnostics
. kmatch md union collgrad ttl_exp tenure i.industry i.race south (wage), ///
> att bwidth(0.5)

Multivariate-distance kernel matching
Number of obs = 1,853
Kernel = epan

Treatment : union = 1
Metric : mahalanobis
Covariates: collgrad ttl_exp tenure i.industry i.race south

Matching statistics

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Treated</td>
<td>366</td>
<td>91</td>
</tr>
</tbody>
</table>

Treatment-effects estimation

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>wage</td>
<td>ATT</td>
</tr>
<tr>
<td>.3303161</td>
<td></td>
</tr>
</tbody>
</table>

. kmatch csummarize
(refitting the model using the generate() option)

<table>
<thead>
<tr>
<th>Means</th>
<th>Common support (treated)</th>
<th>Standardized difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Matched</td>
<td>Unmatched</td>
</tr>
<tr>
<td>collgrad</td>
<td>.322404</td>
<td>.318681</td>
</tr>
<tr>
<td>ttl_exp</td>
<td>13.3929</td>
<td>12.7682</td>
</tr>
<tr>
<td>tenure</td>
<td>8.12614</td>
<td>6.95055</td>
</tr>
<tr>
<td>3.industry</td>
<td>.002732</td>
<td>.021978</td>
</tr>
<tr>
<td>4.industry</td>
<td>.191257</td>
<td>.153846</td>
</tr>
<tr>
<td>5.industry</td>
<td>.062842</td>
<td>.274725</td>
</tr>
<tr>
<td>6.industry</td>
<td>.057377</td>
<td>0</td>
</tr>
<tr>
<td>7.industry</td>
<td>.019126</td>
<td>.021978</td>
</tr>
<tr>
<td>8.industry</td>
<td>.005464</td>
<td>.065934</td>
</tr>
<tr>
<td>9.industry</td>
<td>.010929</td>
<td>.010989</td>
</tr>
<tr>
<td>10.industry</td>
<td>0</td>
<td>.021978</td>
</tr>
<tr>
<td>11.industry</td>
<td>.554645</td>
<td>.175824</td>
</tr>
<tr>
<td>12.industry</td>
<td>.092896</td>
<td>.241758</td>
</tr>
<tr>
<td>2.race</td>
<td>.243169</td>
<td>.681319</td>
</tr>
<tr>
<td>3.race</td>
<td>.002732</td>
<td>.076923</td>
</tr>
</tbody>
</table>
Some Examples

. // make a graph of the common-support stats
. mat M = r(M)
. coefplot matrix(M[,4]), title("Std. difference") noci nolabels xline(0)
Some Examples

.kmatch md union collgrad ttl_exp tenure i.industry i.race south ///
> (wage hours), nate att
(choosing bandwidth ... done)

Multivariate-distance kernel matching

Number of obs = 1,852
Kernel = epan

Treatment : union = 1
Metric : mahalanobis
Covariates: collgrad ttl_exp tenure i.industry i.race south

Matching statistics

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Treated</td>
<td>432</td>
<td>25</td>
</tr>
</tbody>
</table>

Treatment-effects estimation

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>wage</td>
<td></td>
</tr>
<tr>
<td>ATT</td>
<td>.6021049</td>
</tr>
<tr>
<td>NATE</td>
<td>1.430823</td>
</tr>
<tr>
<td>hours</td>
<td></td>
</tr>
<tr>
<td>ATT</td>
<td>1.263759</td>
</tr>
<tr>
<td>NATE</td>
<td>1.450303</td>
</tr>
</tbody>
</table>
Some Examples

. // Multiple outcome variables with different regression-adjustment
. // equations
. kmatch md union collgrad ttl_exp tenure i.industry i.race south ///
>   (wage = collgrad ttl_exp tenure) ///
>   (hours = i.industry i.race), nate att
(computing bandwidth ... done)

Multivariate-distance kernel matching                      Number of obs  =   1,852
                Kernel       =       epan

Treatment : union = 1
Metric : mahalanobis
Covariates: collgrad ttl_exp tenure i.industry i.race south

Matching statistics

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th></th>
<th>Controls</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Total</td>
<td>Used</td>
<td>Unused</td>
</tr>
<tr>
<td>Treated</td>
<td>432</td>
<td>25</td>
<td>457</td>
<td>1104</td>
<td>291</td>
</tr>
<tr>
<td></td>
<td>1.3392</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Treatment-effects estimation

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>wage</td>
<td></td>
</tr>
<tr>
<td>ATT</td>
<td>.5152752</td>
</tr>
<tr>
<td>NATE</td>
<td>1.430823</td>
</tr>
<tr>
<td>hours</td>
<td>1.263759</td>
</tr>
<tr>
<td>ATT</td>
<td>1.450303</td>
</tr>
</tbody>
</table>

wage: adjusted for collgrad ttl_exp tenure
hours: adjusted for i.industry i.race
Some Examples

. // Treatment effects by subpopulation
. kmatch md union collgrad ttl_exp tenure i.industry i.race (wage), ///
   att vce(boot) over(south)
   (south=0: computing bandwidth ... done)
   (south=1: computing bandwidth ... done)
   (running kmatch on estimation sample)

Bootstrap replications (50)
    1  2  3  4  5
    1  2  3  4  5
---------------------------------------------------------------------
Multivariate-distance kernel matching
   Number of obs = 1,853
   Replications  = 50
   Kernel       = epan

Treatment : union = 1
Metric : mahalanobis
Covariates: collgrad ttl_exp tenure i.industry i.race
   0: south = 0
   1: south = 1

Matching statistics

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th></th>
<th>Controls</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Total</td>
<td>Used</td>
<td>Unused</td>
</tr>
<tr>
<td>0</td>
<td>306</td>
<td>15</td>
<td>321</td>
<td>625</td>
<td>120</td>
</tr>
<tr>
<td>1</td>
<td>126</td>
<td>10</td>
<td>136</td>
<td>473</td>
<td>178</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bandwidth

Treatment-effects estimation

|        | wage | Observed Coef. | Bootstrap Coef. | Std. Err. | z     | P>|z|  | Normal-based Coef. | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|--------|------|----------------|-----------------|-----------|------|-----|-------------------|-----------|------|-----|---------------------|
| 0      | ATT  | .4586332       | .276358         | 1.66      | 0.097|     | -.082975          | 1.000241  |
| 1      | ATT  | .9518705       | .406903         | 2.34      | 0.019|     | .1543553          | 1.749386  |

. test [0]ATT = [1]ATT
   ( 1)  [0]ATT - [1]ATT = 0
   chi2( 1) = 1.23
   Prob > chi2 = 0.2679

. lincom [1]ATT - [0]ATT
   ( 1)  [0]ATT + [1]ATT = 0

|        | wage | Coef. | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|--------|------|-------|-----------|------|-----|---------------------|
| (1)    |      | .4932373 | .4452343 | 1.11 | 0.268| -.379406            | 1.365981  |
Simulation

- Population data from Swiss census of 2000.
- Outcome: Treiman occupational prestige (recoded from ISCO codes of the current job using command iskotrei by Hendrickx 2002) (values from 6 to 78; mean 44).
- Estimand: ATT of nationality on occupational prestige, with resident aliens as the treatment group and Swiss nationals as the control group.
- Control variables: gender, age, and highest educational degree.
- Population restricted to people between 24 to 60 years old who are working.
- 2'308'006 individuals, of which 17.5% belong to the treatment group.
- Draw random samples ($N = 500, 1000, \text{ or } 5000$) from population and compute various matching estimators.
Simulation

- Substantial differences between resident aliens and Swiss nationals on all three covariates.
- Propensity score in population (computed from fully stratified data)

McFadden $R^2 = 0.121$
Simulation

- Raw mean difference in occupational prestige (NATE): $-4.79$
- Population ATT (computed from fully stratified data): $-3.96$
- There is some treatment effect heterogeneity (ATE $= -3.51$, ATC $= -3.41$)
Results: Variance

- Nearest-neighbor matching
  - 1 neighbor
  - 5 neighbors
- Kernel matching
  - fixed bandwidth
  - pair-matching bandwidth
  - cross-validation with respect to X
  - cross-validation with respect to Y
  - weighted CV with respect to Y

- N = 500
- N = 5000

- MDM with bias correction
- PSM with bias correction
In this slide we can see that for the same algorithm PSM typically is somewhat less efficient than MDM, but that across algorithms PSM can also be much more efficient than MDM. For example, kernel matching PSM has a much smaller variance than 1-nearest-neighbor MDM. That is, the choice of algorithm matters much more than the choice between PSM and MDM.

For kernel matching the efficiency differences between PSM and MDM are only small; additional post-matching regression adjustment further reduces the differences.
Results: Bias reduction (in percent)

- Nearest-neighbor matching
  - 1 neighbor
  - 5 neighbors

- Kernel matching
  - fixed bandwidth
  - pair-matching bandwidth
  - cross-validation with respect to X
  - cross-validation with respect to Y
  - weighted CV with respect to Y

- N = 500
- N = 5000

- MDM with bias correction
- PSM with bias correction

Ben Jann (University of Bern) Propensity Scores Matching Berlin, 23.06.2017
Propensity Scores Matching

Illustration using kmatch

Here we see that PSM has a bias that does not vanish as the sample size increases. The reason is that the same propensity-score model specification is used for both sample sizes. The model is rather simple (linear effect of age, no interactions) and due to the specific pattern of the data (in particular, the sharp drop in the outcome variable after propensity score 0.3) small imprecisions can have substantial effects on the results. In practice, one would probably use a more refined specification in the large-sample situation, which would reduce bias.

The bias also vanishes once post-matching regression adjustment is applied.
Results: Mean squared error

<table>
<thead>
<tr>
<th>Method</th>
<th>N = 500</th>
<th>N = 5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nearest-neighbor matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 neighbor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 neighbors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kernel matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fixed bandwidth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pair-matching bandwidth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cross-validation with respect to X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cross-validation with respect to Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weighted CV with respect to Y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- MDM with bias correction
- PSM with bias correction

- Results: Mean squared error

Ben Jann (University of Bern)
Results: Relative standard error

<table>
<thead>
<tr>
<th>Method</th>
<th>N = 500</th>
<th>N = 5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nearest-neighbor matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 neighbor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 neighbors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nearest-neighbor matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 neighbor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 neighbors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kernel matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fixed bandwidth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pair-matching bandwidth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cross-validation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>with respect to X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cross-validation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>with respect to Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weighted CV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>with respect to Y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MDM with bias correction
PSM with bias correction

Ben Jann (University of Bern)
Results: Coverage of 95% CIs

- Nearest-neighbor matching (teffects)
  - 1 neighbor
  - 5 neighbors

- Nearest-neighbor matching (bootstrap)
  - 1 neighbor
  - 5 neighbors

- Kernel matching (bootstrap)
  - fixed bandwidth
  - pair-matching bandwidth
  - cross-validation with respect to X
  - cross-validation with respect to Y
  - weighted CV with respect to Y

<table>
<thead>
<tr>
<th>N</th>
<th>MDM with bias correction</th>
<th>PSM with bias correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Potential Outcomes and Causal Inference

2. Matching

3. Propensity Score Matching

4. King and Nielsen’s “Why Propensity Scores Should Not Be Used for Matching”

5. Are King and Nielsen right?

6. Illustration using kmatch

7. Conclusions
Conclusions

- The arguments brought forward by King and Nielsen against Propensity Score Matching are valid, but they mostly apply to one specific form of PSM: pair matching (one-to-one matching without replacement).

- Other PSM matching algorithms perform much better because they are less affected by the random pruning problem.
Conclusions

- Overall, I agree that MDM has advantages over PSM, but it also has some disadvantages. In applied research the choice may not be that clear.
  - MDM leaves less scope for post-matching modeling decision biases.
  - Theoretical results (see, e.g., Frölich 2007) suggest that MDM will generally tend to outperform PSM in terms of efficiency (but differences are likely to be small).
  - Less restrictions in terms of possible post-matching analyses.
  - Choice of scaling matrix largely arbitrary; various suggestions in the literature (somewhat unclear, e.g., how categorical variables should be treated).
  - Computational complexity.

- One clear conclusion we can draw, however, is:
  
  Do not use propensity scores for pair matching!
  
  (But don’t use pair matching anyhow.)
Conclusions

Some conclusions from the simulation

- For PSM, application of regression-adjustment seems like a great idea (reduction of bias and variance); for MDM the advantages of regression-adjustment are less clear.
- Bootstrap standard error/confidence interval estimation seems to be mostly ok for kernel/ridge matching; this is in contrast to nearest-neighbor matching, where bootstrap standard errors are clearly biased.

To do

- Run some simulations comparable to the ones by King and Nielsen using various matching algorithms.
References I


