Matching to Reduce Model Dependence

Gary King

Institute for Quantitative Social Science
Harvard University, http://GKing.Harvard.edu

Talk at Washington University, St. Louis, 1/22/2010

References

- Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart."Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis*, 15 (2007): 199-236.

References

- Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart."Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis*, 15 (2007): 199-236.
- Stefano M. Iacus, Gary King, and Giuseppe Porro, "Causal Inference Without Balance Checking: Coarsened Exact Matching," 2010.

References

- Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart."Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis*, 15 (2007): 199-236.
- Stefano M. Iacus, Gary King, and Giuseppe Porro, "Causal Inference Without Balance Checking: Coarsened Exact Matching," 2010.
- Stefano M. Iacus, Gary King, and Giuseppe Porro, "Multivariate Matching Methods That are Monotonic Imbalance Bounding," 2010.

References

- Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart."Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis*, 15 (2007): 199-236.
- Stefano M. Iacus, Gary King, and Giuseppe Porro, "Causal Inference Without Balance Checking: Coarsened Exact Matching," 2010.
- Stefano M. Iacus, Gary King, and Giuseppe Porro, "Multivariate Matching Methods That are Monotonic Imbalance Bounding," 2010.

- Related Software: WhatIf, MatchIt, Zelig, CEM

References

- Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart."Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis*, 15 (2007): 199-236.
- Stefano M. Iacus, Gary King, and Giuseppe Porro, "Causal Inference Without Balance Checking: Coarsened Exact Matching," 2010.
- Stefano M. Iacus, Gary King, and Giuseppe Porro, "Multivariate Matching Methods That are Monotonic Imbalance Bounding," 2010.
- Related Software: WhatIf, MatchIt, Zelig, CEM

`http://GKing.Harvard.edu/projects/cause.shtml`

Matching for Causal Inference

Matching for Causal Inference

- Uses nonparametric, non-model based methods.

Matching for Causal Inference

- Uses nonparametric, non-model based methods.
- Goal: reduce model dependence

Matching for Causal Inference

- Uses nonparametric, non-model based methods.
- Goal: reduce model dependence
- Not a substitute for parametric models

Matching for Causal Inference

- Uses nonparametric, non-model based methods.
- Goal: reduce model dependence
- Not a substitute for parametric models
- makes parametric models work better (less hard)

Matching for Causal Inference

- Uses nonparametric, non-model based methods.
- Goal: reduce model dependence
- Not a substitute for parametric models
- makes parametric models work better (less hard)
- Matching is not an estimator; its a preprocessing method

Matching for Causal Inference

- Uses nonparametric, non-model based methods.
- Goal: reduce model dependence
- Not a substitute for parametric models
- makes parametric models work better (less hard)
- Matching is not an estimator; its a preprocessing method
- Apply model to preprocessed rather than raw data

Matching for Causal Inference

- Uses nonparametric, non-model based methods.
- Goal: reduce model dependence
- Not a substitute for parametric models
- makes parametric models work better (less hard)
- Matching is not an estimator; its a preprocessing method
- Apply model to preprocessed rather than raw data
- Valid standard errors use the same parametric procedures

Data Collection Mechanisms

Data Collection Mechanisms

- Key features of classical randomized experiments:

Data Collection Mechanisms

- Key features of classical randomized experiments:
  - random selection of units from a given population.

Data Collection Mechanisms

- Key features of classical randomized experiments:
  - random selection of units from a given population.
  - random assignment of values of the treatment.

Data Collection Mechanisms

- Key features of classical randomized experiments:
  - random selection of units from a given population.
  - random assignment of values of the treatment.
  - large *n*.

Data Collection Mechanisms

- Key features of classical randomized experiments:
  - random selection of units from a given population.
  - random assignment of values of the treatment.
  - large n.
  - (Implied: no missing data, measurement error, noncompliance, etc.)

Data Collection Mechanisms

- Key features of classical randomized experiments:
  - random selection of units from a given population.
  - random assignment of values of the treatment.
  - large n.
  - (Implied: no missing data, measurement error, noncompliance, etc.)
- Any study that meets all 3 can estimate causal inferences without modeling assumptions.

Data Collection Mechanisms

- Key features of classical randomized experiments:
  - random selection of units from a given population.
  - random assignment of values of the treatment.
  - large *n*.
  - (Implied: no missing data, measurement error, noncompliance, etc.)
- Any study that meets all 3 can estimate causal inferences without modeling assumptions.
- Observational data:

Data Collection Mechanisms

- Key features of classical randomized experiments:
  - random selection of units from a given population.
  - random assignment of values of the treatment.
  - large *n*.
  - (Implied: no missing data, measurement error, noncompliance, etc.)
- Any study that meets all 3 can estimate causal inferences without modeling assumptions.
- Observational data:
  - Any data that fails to meet all requirements of classical randomized experiments

Data Collection Mechanisms

- Key features of classical randomized experiments:
  - random selection of units from a given population.
  - random assignment of values of the treatment.
  - large *n*.
  - (Implied: no missing data, measurement error, noncompliance, etc.)
- Any study that meets all 3 can estimate causal inferences without modeling assumptions.
- Observational data:
  - Any data that fails to meet all requirements of classical randomized experiments
  - Encompasses most research in every field

Data Collection Mechanisms

- Key features of classical randomized experiments:
  - random selection of units from a given population.
  - random assignment of values of the treatment.
  - large *n*.
  - (Implied: no missing data, measurement error, noncompliance, etc.)
- Any study that meets all 3 can estimate causal inferences without modeling assumptions.
- Observational data:
  - Any data that fails to meet all requirements of classical randomized experiments
  - Encompasses most research in every field
- Most knowledge learned is from observational data — even in experimental work (where most treatments fail)

Characteristics of Observational Data

Characteristics of Observational Data

- Lots of data

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment:

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random,

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random, not controlled by investigator,

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random, not controlled by investigator, not known

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random, not controlled by investigator, not known
- Bias-Variance Tradeoff

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random, not controlled by investigator, not known
- Bias-Variance Tradeoff

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random, not controlled by investigator, not known
- Bias-Variance Tradeoff
- The idea of matching: sacrifice some data to avoid bias

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random, not controlled by investigator, not known
- Bias-Variance Tradeoff
- The idea of matching: sacrifice some data to avoid bias
- Removing heterogeneous data will often reduce variance too

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random, not controlled by investigator, not known
- Bias-Variance Tradeoff
- The idea of matching: sacrifice some data to avoid bias
- Removing heterogeneous data will often reduce variance too
- (Medical experiments are the reverse: small-n with random treatment assignment; don't match unless something goes wrong)

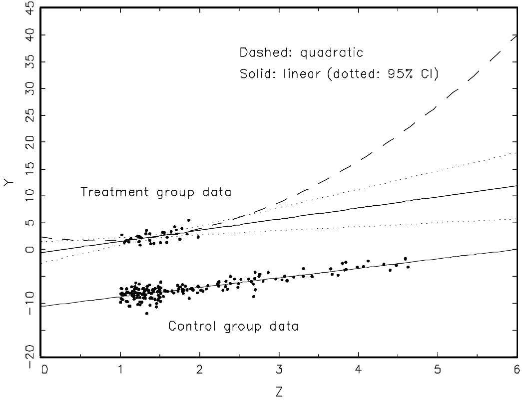
Model Dependence

Model Dependence

(King and Zeng, 2006: fig.4 *Political Analysis*)

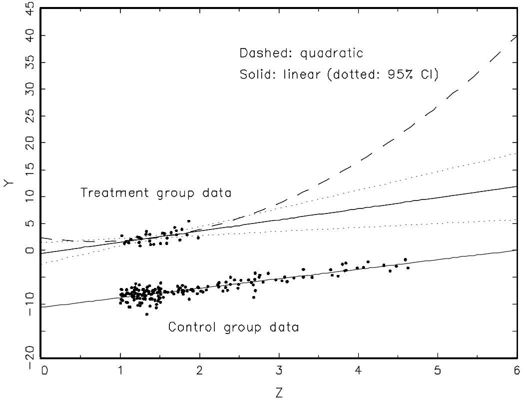
Model Dependence

(King and Zeng, 2006: fig.4 *Political Analysis*)



Model Dependence

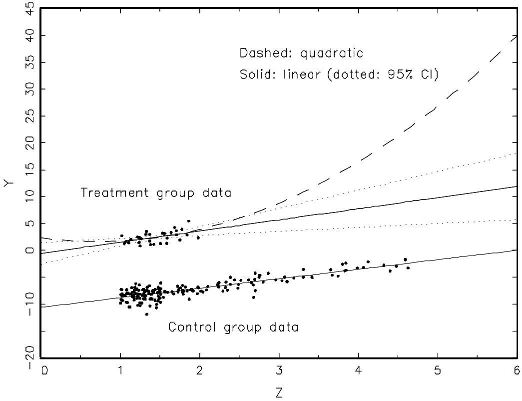
(King and Zeng, 2006: fig.4 *Political Analysis*)



What to do?

Model Dependence

(King and Zeng, 2006: fig.4 *Political Analysis*)

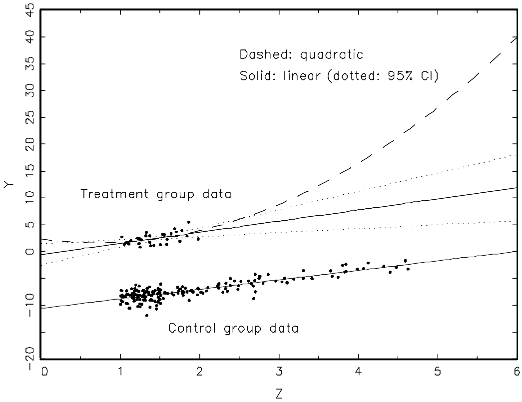


What to do?

- Preprocess I: Eliminate extrapolation region (a separate step)

Model Dependence

(King and Zeng, 2006: fig.4 *Political Analysis*)

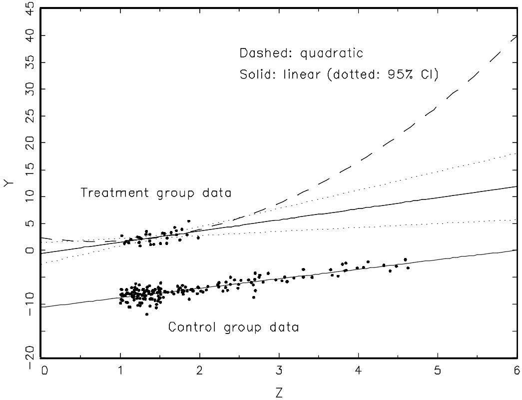


What to do?

- Preprocess I: Eliminate extrapolation region (a separate step)
- Preprocess II: Match (prune bad matches) within interpolation region

Model Dependence

(King and Zeng, 2006: fig.4 *Political Analysis*)



What to do?

- Preprocess I: Eliminate extrapolation region (a separate step)
- Preprocess II: Match (prune bad matches) within interpolation region
- Model remaining imbalance

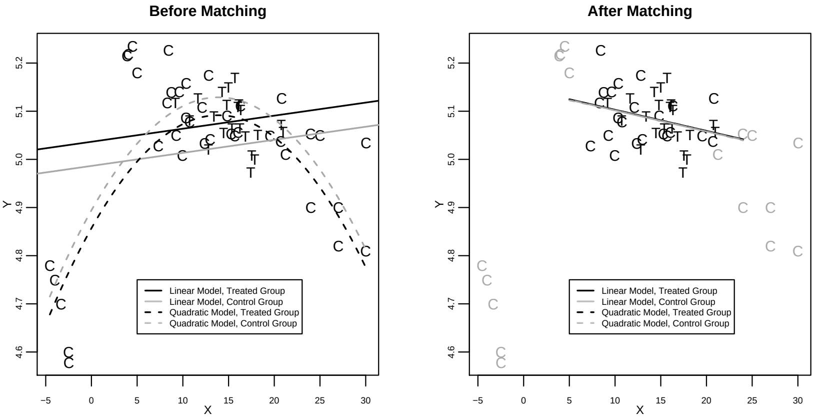
Matching within the Interpolation Region

Matching within the Interpolation Region

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

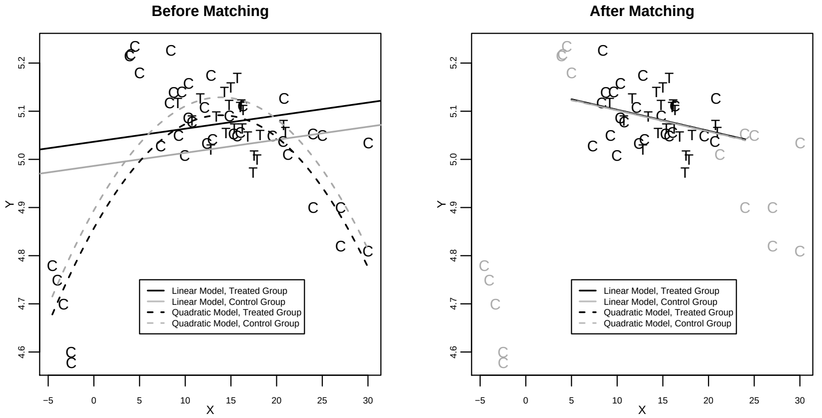
Matching within the Interpolation Region

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



Matching within the Interpolation Region

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



Matching reduces model dependence, bias, and variance

Parametric Analysis Methods

Parametric Analysis Methods

- Researchers typically

Parametric Analysis Methods

- Researchers typically
  - assume a parametric model (up to unknown parameters):
    e.g., $Y_i \sim p(\mu_i, \theta)$ with $\mu_i \equiv E(Y_i \mid t_i, X_i) = g(\alpha + t_i\beta + X_i\gamma)$

Parametric Analysis Methods

- Researchers typically
  - assume a parametric model (up to unknown parameters): e.g., $Y_i \sim p(\mu_i, \theta)$ with $\mu_i \equiv E(Y_i \mid t_i, X_i) = g(\alpha + t_i\beta + X_i\gamma)$
  - Estimate the causal effect: ATT=mean$[g(\hat{\alpha} + \hat{\beta} + X_i\hat{\gamma}) - g(\hat{\alpha} + X_i\hat{\gamma})]$

Parametric Analysis Methods

- Researchers typically
  - assume a parametric model (up to unknown parameters): e.g., $Y_i \sim p(\mu_i, \theta)$ with $\mu_i \equiv E(Y_i \mid t_i, X_i) = g(\alpha + t_i\beta + X_i\gamma)$
  - Estimate the causal effect: ATT=mean$[g(\hat{\alpha} + \hat{\beta} + X_i\hat{\gamma}) - g(\hat{\alpha} + X_i\hat{\gamma})]$
- But, the true model is unknown.

Parametric Analysis Methods

- Researchers typically
  - assume a parametric model (up to unknown parameters):
    e.g., $Y_i \sim p(\mu_i, \theta)$ with $\mu_i \equiv E(Y_i \mid t_i, X_i) = g(\alpha + t_i\beta + X_i\gamma)$
  - Estimate the causal effect: ATT=mean$[g(\hat{\alpha} + \hat{\beta} + X_i\hat{\gamma}) - g(\hat{\alpha} + X_i\hat{\gamma})]$
- But, the true model is unknown.
- In experiments, T and X are independent; we can drop X

Parametric Analysis Methods

- Researchers typically
  - assume a parametric model (up to unknown parameters): e.g., $Y_i \sim p(\mu_i, \theta)$ with $\mu_i \equiv E(Y_i \mid t_i, X_i) = g(\alpha + t_i\beta + X_i\gamma)$
  - Estimate the causal effect: ATT=mean$[g(\hat{\alpha} + \hat{\beta} + X_i\hat{\gamma}) - g(\hat{\alpha} + X_i\hat{\gamma})]$
- But, the true model is unknown.
- In experiments, T and X are independent; we can drop X
  - $ATT = g(\hat{\alpha} + \hat{\beta}) - g(\hat{\alpha})$

Parametric Analysis Methods

- Researchers typically
  - assume a parametric model (up to unknown parameters): e.g., $Y_i \sim p(\mu_i, \theta)$ with $\mu_i \equiv E(Y_i \mid t_i, X_i) = g(\alpha + t_i\beta + X_i\gamma)$
  - Estimate the causal effect: ATT=mean$[g(\hat{\alpha} + \hat{\beta} + X_i\hat{\gamma}) - g(\hat{\alpha} + X_i\hat{\gamma})]$
- But, the true model is unknown.
- In experiments, T and X are independent; we can drop X
  - $ATT = g(\hat{\alpha} + \hat{\beta}) - g(\hat{\alpha})$
  - The ATT requires no calculation over i.

Parametric Analysis Methods

- Researchers typically
  - assume a parametric model (up to unknown parameters): e.g., $Y_i \sim p(\mu_i, \theta)$ with $\mu_i \equiv E(Y_i \mid t_i, X_i) = g(\alpha + t_i\beta + X_i\gamma)$
  - Estimate the causal effect: ATT=mean$[g(\hat{\alpha} + \hat{\beta} + X_i\hat{\gamma}) - g(\hat{\alpha} + X_i\hat{\gamma})]$
- But, the true model is unknown.
- In experiments, T and X are independent; we can drop X
  - $ATT = g(\hat{\alpha} + \hat{\beta}) - g(\hat{\alpha})$
  - The ATT requires no calculation over i.
  - MLE is invariant to reparamerization, so $g(\cdot)$ is irrelevant!

Parametric Analysis Methods

- Researchers typically
  - assume a parametric model (up to unknown parameters): e.g., $Y_i \sim p(\mu_i, \theta)$ with $\mu_i \equiv E(Y_i \mid t_i, X_i) = g(\alpha + t_i\beta + X_i\gamma)$
  - Estimate the causal effect: ATT=mean$[g(\hat{\alpha} + \hat{\beta} + X_i\hat{\gamma}) - g(\hat{\alpha} + X_i\hat{\gamma})]$
- But, the true model is unknown.
- In experiments, T and X are independent; we can drop X
  - $ATT = g(\hat{\alpha} + \hat{\beta}) - g(\hat{\alpha})$
  - The ATT requires no calculation over i.
  - MLE is invariant to reparamerization, so $g(\cdot)$ is irrelevant!
- In observational studies,

Parametric Analysis Methods

- Researchers typically
  - assume a parametric model (up to unknown parameters): e.g., $Y_i \sim p(\mu_i, \theta)$ with $\mu_i \equiv E(Y_i \mid t_i, X_i) = g(\alpha + t_i\beta + X_i\gamma)$
  - Estimate the causal effect: ATT=mean$[g(\hat{\alpha} + \hat{\beta} + X_i\hat{\gamma}) - g(\hat{\alpha} + X_i\hat{\gamma})]$
- But, the true model is unknown.
- In experiments, T and X are independent; we can drop X
  - $ATT = g(\hat{\alpha} + \hat{\beta}) - g(\hat{\alpha})$
  - The ATT requires no calculation over i.
  - MLE is invariant to reparamerization, so $g(\cdot)$ is irrelevant!
- In observational studies,
  - results are dependent on choice of $g(\cdot)$.

Parametric Analysis Methods

- Researchers typically
  - assume a parametric model (up to unknown parameters): e.g., $Y_i \sim p(\mu_i, \theta)$ with $\mu_i \equiv E(Y_i \mid t_i, X_i) = g(\alpha + t_i\beta + X_i\gamma)$
  - Estimate the causal effect: ATT=mean$[g(\hat{\alpha} + \hat{\beta} + X_i\hat{\gamma}) - g(\hat{\alpha} + X_i\hat{\gamma})]$
- But, the true model is unknown.
- In experiments, T and X are independent; we can drop X
  - $ATT = g(\hat{\alpha} + \hat{\beta}) - g(\hat{\alpha})$
  - The ATT requires no calculation over i.
  - MLE is invariant to reparamerization, so $g(\cdot)$ is irrelevant!
- In observational studies,
  - results are dependent on choice of $g(\cdot)$.
  - curse of dimensionality looms large

What Matching Does

What Matching Does

- Notation:

What Matching Does

- Notation:
  - Y_i Dependent variable

What Matching Does

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1)

What Matching Does

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1)
  X_i pre-treatment covariates

What Matching Does

- Notation:
  - Y_i Dependent variable
  - T_i Treatment variable (0/1)
  - X_i pre-treatment covariates
- Treatment Effect for treated ($T_i = 1$) observation i:

What Matching Does

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1)
  X_i pre-treatment covariates

- Treatment Effect for treated ($T_i = 1$) observation i:

$$\text{TE}_i = Y_i(T_i = 1) - Y_i(T_i = 0)$$

What Matching Does

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1)
  X_i pre-treatment covariates

- Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

What Matching Does

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1)
  X_i pre-treatment covariates

- Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with Y_j from matched ($X_i \approx X_j$) controls
  $\hat{Y}_i(0) = Y_j(0)$ or a model $\hat{Y}_i(0) = \hat{g}_0(X_j)$

What Matching Does

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1)
  X_i pre-treatment covariates

- Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with Y_j from matched ($X_i \approx X_j$) controls
  $\hat{Y}_i(0) = Y_j(0)$ or a model $\hat{Y}_i(0) = \hat{g}_0(X_j)$
- Prune unmatched units to improve balance (so X is unimportant)

What Matching Does

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1)
  X_i pre-treatment covariates

- Treatment Effect for treated ($T_i = 1$) observation i:

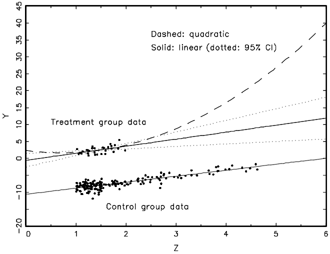
$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with Y_j from matched ($X_i \approx X_j$) controls $\hat{Y}_i(0) = Y_j(0)$ or a model $\hat{Y}_i(0) = \hat{g}_0(X_j)$
- Prune unmatched units to improve balance (so X is unimportant)
- Sample Average Treatment effect on the Treated:

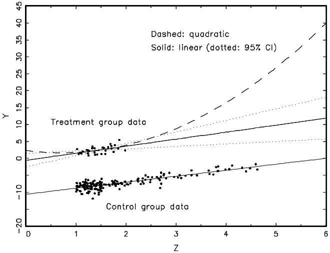
$$\text{SATT} = \frac{1}{n_T} \sum_{i \in \{T_i = 1\}} \text{TE}_i$$

Remove Extrapolation Region, then Match

Remove Extrapolation Region, then Match

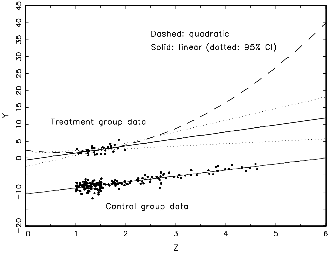


Remove Extrapolation Region, then Match



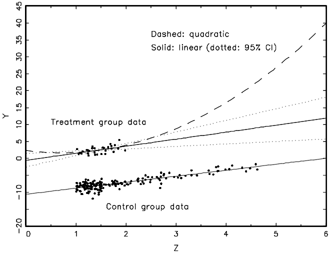
- Must remove data (selecting on X) to avoid extrapolation.

Remove Extrapolation Region, then Match



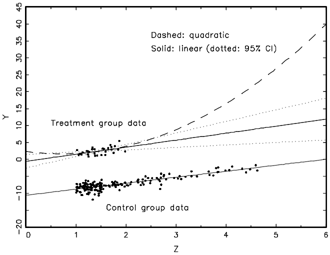
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|t = 1)$ and $P(X|t = 0)$

Remove Extrapolation Region, then Match



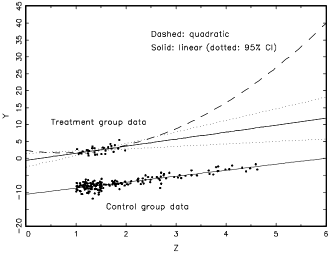
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|t=1)$ and $P(X|t=0)$
  1. exact match, so support is defined only at data points

Remove Extrapolation Region, then Match



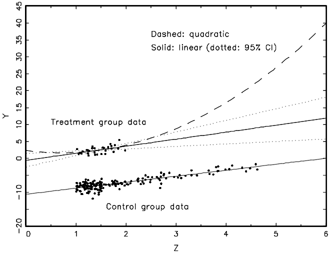
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|t=1)$ and $P(X|t=0)$
  1. exact match, so support is defined only at data points
  2. less, but still conservative: convex hull approach

Remove Extrapolation Region, then Match



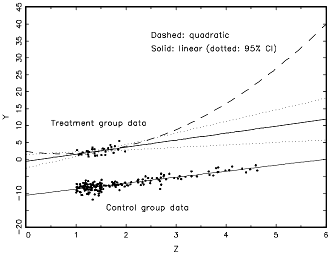
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|t=1)$ and $P(X|t=0)$
  1. exact match, so support is defined only at data points
  2. less, but still conservative: convex hull approach
     - let t^* and X^* denote subsets of t and X s.t. $\{1-t^*, X^*\}$ falls within the convex hull of $\{t, X\}$

Remove Extrapolation Region, then Match



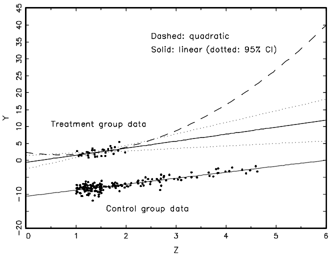
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|t=1)$ and $P(X|t=0)$
  1. exact match, so support is defined only at data points
  2. less, but still conservative: convex hull approach
     - let t^* and X^* denote subsets of t and X s.t. $\{1-t^*, X^*\}$ falls within the convex hull of $\{t, X\}$
     - use X^* as estimate of common support (deleting remaining observations)

Remove Extrapolation Region, then Match



- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|t=1)$ and $P(X|t=0)$
  1. exact match, so support is defined only at data points
  2. less, but still conservative: convex hull approach
     - let t^* and X^* denote subsets of t and X s.t. $\{1-t^*, X^*\}$ falls within the convex hull of $\{t, X\}$
     - use X^* as estimate of common support (deleting remaining observations)
  3. Other approaches, based on distance metrics, pscores, etc.

Remove Extrapolation Region, then Match



- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|t=1)$ and $P(X|t=0)$
  1. exact match, so support is defined only at data points
  2. less, but still conservative: convex hull approach
     - let t^* and X^* denote subsets of t and X s.t. $\{1-t^*, X^*\}$ falls within the convex hull of $\{t, X\}$
     - use X^* as estimate of common support (deleting remaining observations)
  3. Other approaches, based on distance metrics, pscores, etc.
- Then match within interpolation (common support) region

Choosing a Matching Procedure

Choosing a Matching Procedure

- The goal: improve balance (bias) without losing too many observations (efficiency)

Choosing a Matching Procedure

- The goal: improve balance (bias) without losing too many observations (efficiency)
- Try many matching procedures until better balance is achieved. (to avoid selection bias, do not examine Y during preprocessing)

Choosing a Matching Procedure

- The goal: improve balance (bias) without losing too many observations (efficiency)
- Try many matching procedures until better balance is achieved. (to avoid selection bias, do not examine Y during preprocessing)
- Select Covariates: include all variables that would have been included in the parametric model, but avoid posttreatment bias.

Choosing a Matching Procedure

- The goal: improve balance (bias) without losing too many observations (efficiency)
- Try many matching procedures until better balance is achieved. (to avoid selection bias, do not examine Y during preprocessing)
- Select Covariates: include all variables that would have been included in the parametric model, but avoid posttreatment bias.
- Try Exact Matching: if a large number of units are matched, begin parametric analysis.

Choosing a Matching Procedure

- The goal: improve balance (bias) without losing too many observations (efficiency)
- Try many matching procedures until better balance is achieved. (to avoid selection bias, do not examine Y during preprocessing)
- Select Covariates: include all variables that would have been included in the parametric model, but avoid posttreatment bias.
- Try Exact Matching: if a large number of units are matched, begin parametric analysis.
- Use approximate matching (many options!)

Choosing a Matching Procedure

- The goal: improve balance (bias) without losing too many observations (efficiency)
- Try many matching procedures until better balance is achieved. (to avoid selection bias, do not examine Y during preprocessing)
- Select Covariates: include all variables that would have been included in the parametric model, but avoid posttreatment bias.
- Try Exact Matching: if a large number of units are matched, begin parametric analysis.
- Use approximate matching (many options!)
- Evaluate the Matching Procedure: look at low-dimensional summaries of X (no hypothesis tests!)

Choosing a Matching Procedure

- The goal: improve balance (bias) without losing too many observations (efficiency)
- Try many matching procedures until better balance is achieved. (to avoid selection bias, do not examine Y during preprocessing)

- Select Covariates: include all variables that would have been included in the parametric model, but avoid posttreatment bias.
- Try Exact Matching: if a large number of units are matched, begin parametric analysis.
- Use approximate matching (many options!)
- Evaluate the Matching Procedure: look at low-dimensional summaries of X (no hypothesis tests!)
- Parametric Outcome Analysis: same method, same algorithm, same software, same model checking procedures, ...

Nearest Neighbor Approximate Matching

Nearest Neighbor Approximate Matching

- Choose metric for measuring distances between observation vectors (Euclidean, Mahalanobis, etc.)

Nearest Neighbor Approximate Matching

- Choose metric for measuring distances between observation vectors (Euclidean, Mahalanobis, etc.)
- For each treated unit, choose the "closest" control unit

Nearest Neighbor Approximate Matching

- Choose metric for measuring distances between observation vectors (Euclidean, Mahalanobis, etc.)
- For each treated unit, choose the “closest” control unit
- Alternatively: use “optimal matching” by choosing the set of controls as close as possible to the set of treated units

Propensity Score Approximate Matching

Propensity Score Approximate Matching

- Define the (true) propensity score: $\pi_i = P(T_i = 1|X_i)$

Propensity Score Approximate Matching

- Define the (true) propensity score: $\pi_i = P(T_i = 1|X_i)$
- Apparently solve the curse of dimensionality problem: match on (one-dimensional) π_i instead of (multidimensional) X_i

Propensity Score Approximate Matching

- Define the (true) propensity score: $\pi_i = P(T_i = 1|X_i)$
- Apparently solve the curse of dimensionality problem: match on (one-dimensional) π_i instead of (multidimensional) X_i
- Since: $p(X|t = 1) = p(X|t = 0) \quad \Leftrightarrow \quad p(\pi|t = 1) = p(\pi|t = 0)$

Propensity Score Approximate Matching

- Define the (true) propensity score: $\pi_i = P(T_i = 1|X_i)$
- Apparently solve the curse of dimensionality problem: match on (one-dimensional) π_i instead of (multidimensional) X_i
- Since: $p(X|t=1) = p(X|t=0) \quad \Leftrightarrow \quad p(\pi|t=1) = p(\pi|t=0)$
- Problem: π is unobserved

Propensity Score Approximate Matching

- Define the (true) propensity score: $\pi_i = P(T_i = 1|X_i)$
- Apparently solve the curse of dimensionality problem: match on (one-dimensional) π_i instead of (multidimensional) X_i
- Since: $p(X|t = 1) = p(X|t = 0) \quad \Leftrightarrow \quad p(\pi|t = 1) = p(\pi|t = 0)$
- Problem: π is unobserved
- Usual practice: estimate with a logit of t_i on X_i

Propensity Score Approximate Matching

- Define the (true) propensity score: $\pi_i = P(T_i = 1|X_i)$
- Apparently solve the curse of dimensionality problem: match on (one-dimensional) π_i instead of (multidimensional) X_i
- Since: $p(X|t = 1) = p(X|t = 0) \quad \Leftrightarrow \quad p(\pi|t = 1) = p(\pi|t = 0)$
- Problem: π is unobserved
- Usual practice: estimate with a logit of t_i on X_i
- Problem with the real (estimated) version of the pscore

Propensity Score Approximate Matching

- Define the (true) propensity score: $\pi_i = P(T_i = 1|X_i)$
- Apparently solve the curse of dimensionality problem: match on (one-dimensional) π_i instead of (multidimensional) X_i
- Since: $p(X|t = 1) = p(X|t = 0) \quad \Leftrightarrow \quad p(\pi|t = 1) = p(\pi|t = 0)$
- Problem: π is unobserved
- Usual practice: estimate with a logit of t_i on X_i
- Problem with the real (estimated) version of the pscore
  - the Propensity Score Tautology: check for correct pscore specification $\Leftrightarrow$ check for balance

Propensity Score Approximate Matching

- Define the (true) propensity score: $\pi_i = P(T_i = 1|X_i)$
- Apparently solve the curse of dimensionality problem: match on (one-dimensional) π_i instead of (multidimensional) X_i
- Since: $p(X|t = 1) = p(X|t = 0) \quad \Leftrightarrow \quad p(\pi|t = 1) = p(\pi|t = 0)$
- Problem: π is unobserved
- Usual practice: estimate with a logit of t_i on X_i
- Problem with the real (estimated) version of the pscore
  - the Propensity Score Tautology: check for correct pscore specification $\Leftrightarrow$ check for balance
  - (If it works, it works; if it doesn't work, it doesn't work)

Propensity Score Approximate Matching

- Define the (true) propensity score: $\pi_i = P(T_i = 1|X_i)$
- Apparently solve the curse of dimensionality problem: match on (one-dimensional) π_i instead of (multidimensional) X_i
- Since: $p(X|t = 1) = p(X|t = 0) \quad \Leftrightarrow \quad p(\pi|t = 1) = p(\pi|t = 0)$
- Problem: π is unobserved
- Usual practice: estimate with a logit of t_i on X_i
- Problem with the real (estimated) version of the pscore
  - the Propensity Score Tautology: check for correct pscore specification $\Leftrightarrow$ check for balance
  - (If it works, it works; if it doesn't work, it doesn't work)
  - Infinite regress: can't use it to identify regions of extrapolation unless pscore specification is correct; can't check pscore validity (via checking balance) until removing regions of extrapolation!

Propensity Score Approximate Matching

- Define the (true) propensity score: $\pi_i = P(T_i = 1|X_i)$
- Apparently solve the curse of dimensionality problem: match on (one-dimensional) π_i instead of (multidimensional) X_i
- Since: $p(X|t=1) = p(X|t=0) \quad \Leftrightarrow \quad p(\pi|t=1) = p(\pi|t=0)$
- Problem: π is unobserved
- Usual practice: estimate with a logit of t_i on X_i
- Problem with the real (estimated) version of the pscore
  - the Propensity Score Tautology: check for correct pscore specification $\Leftrightarrow$ check for balance
  - (If it works, it works; if it doesn't work, it doesn't work)
  - Infinite regress: can't use it to identify regions of extrapolation unless pscore specification is correct; can't check pscore validity (via checking balance) until removing regions of extrapolation!
- Pscore is one practical way to start, but better alternatives exist

Empirical Illustration: Carpenter, AJPS, 2002

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).
- lognormal survival model.

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).
- lognormal survival model.
- seven oversight variables (median adjusted ADA scores for House and Senate Committees as well as for House and Senate floors, Democratic Majority in House and Senate, and Democratic Presidency).

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).
- lognormal survival model.
- seven oversight variables (median adjusted ADA scores for House and Senate Committees as well as for House and Senate floors, Democratic Majority in House and Senate, and Democratic Presidency).
- 18 control variables (clinical factors, firm characteristics, media variables, etc.)

Evaluating Reduction in Model Dependence

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.
- discard 49 units (2 treated and 17 control units).

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.
- discard 49 units (2 treated and 17 control units).
- run 262,143 possible specifications and calculates ATE for each.

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.
- discard 49 units (2 treated and 17 control units).
- run 262,143 possible specifications and calculates ATE for each.
- Look at *variability* in ATE estimate across specifications.

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.
- discard 49 units (2 treated and 17 control units).
- run 262,143 possible specifications and calculates ATE for each.
- Look at *variability* in ATE estimate across specifications.
- (Normal applications would only do one or a small number of specifications.)

Example of Balance Assessments

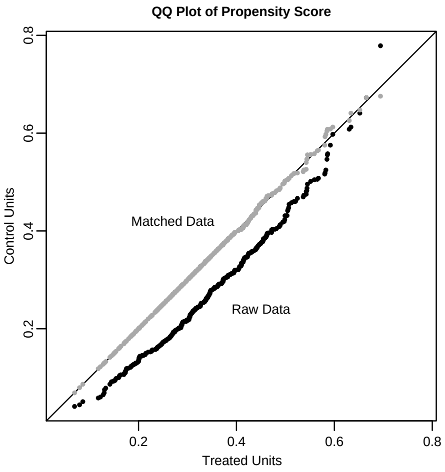


Figure: QQ plot of propensity score

Reducing Model Dependence

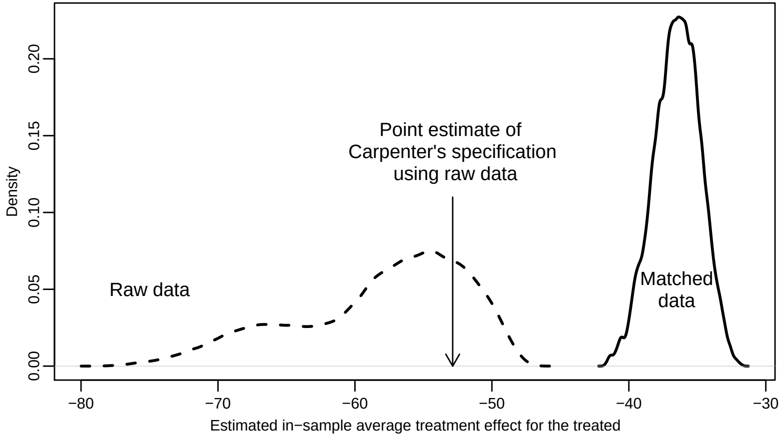


Figure: Histogram of estimated in-sample average treatment effect for the treated (ATT) of the Democratic Senate majority on FDA drug approval time across 262,143 specifications.

Another Example: Jeffrey Koch, AJPS, 2002

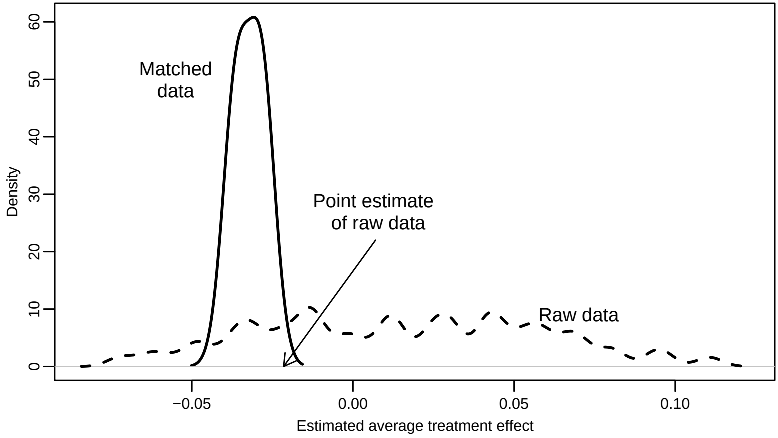


Figure: Estimated effects of being a highly visible female Republican candidate across 63 possible specifications with the Koch data.