Model Dependence in Counterfactual Inference

Gary King

October 20, 2005

References

- King, Gary and Langche Zeng. "The Dangers of Extreme Counterfactuals," *Political Analysis*, Vol. 14, No. 2, 2006, forthcoming.

References

- King, Gary and Langche Zeng. "The Dangers of Extreme Counterfactuals," *Political Analysis*, Vol. 14, No. 2, 2006, forthcoming.
- King, Gary and Langche Zeng. "When Can History be Our Guide? The Pitfalls of Counterfactual Inference," *International Studies Quarterly*, 2006, forthcoming.

References

- King, Gary and Langche Zeng. "The Dangers of Extreme Counterfactuals," *Political Analysis*, Vol. 14, No. 2, 2006, forthcoming.
- King, Gary and Langche Zeng. "When Can History be Our Guide? The Pitfalls of Counterfactual Inference," *International Studies Quarterly*, 2006, forthcoming.
- Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart, "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference" preprint.

References

- King, Gary and Langche Zeng. "The Dangers of Extreme Counterfactuals," *Political Analysis*, Vol. 14, No. 2, 2006, forthcoming.
- King, Gary and Langche Zeng. "When Can History be Our Guide? The Pitfalls of Counterfactual Inference," *International Studies Quarterly*, 2006, forthcoming.
- Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart, "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference" preprint.

- Related Software: WhatIf, MatchIt, Zelig

References

- King, Gary and Langche Zeng. "The Dangers of Extreme Counterfactuals," *Political Analysis*, Vol. 14, No. 2, 2006, forthcoming.
- King, Gary and Langche Zeng. "When Can History be Our Guide? The Pitfalls of Counterfactual Inference," *International Studies Quarterly*, 2006, forthcoming.
- Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart, "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference" preprint.

- Related Software: WhatIf, MatchIt, Zelig

`http://GKing.Harvard.edu`

Counterfactuals

Counterfactuals

- Three types:

Counterfactuals

- Three types:
  1. Forecasts Will the U.S. be in Iraq in 2008?

Counterfactuals

- Three types:
  1. Forecasts Will the U.S. be in Iraq in 2008?
  2. Whatif Questions What would have happened if the U.S. had not invaded Iraq?

Counterfactuals

- Three types:
  1. Forecasts Will the U.S. be in Iraq in 2008?
  2. Whatif Questions What would have happened if the U.S. had not invaded Iraq?
  3. Causal Effects What is the causal effect of the Iraq war on U.S. Supreme Court decision making? (a factual minus a counterfactual)

Counterfactuals

- Three types:
  1. Forecasts Will the U.S. be in Iraq in 2008?
  2. Whatif Questions What would have happened if the U.S. had not invaded Iraq?
  3. Causal Effects What is the causal effect of the Iraq war on U.S. Supreme Court decision making? (a factual minus a counterfactual)
- Counterfactuals are part of almost all research questions.

Model Dependence in Practice

Model Dependence in Practice

- How do you conduct empirical analyses?

Model Dependence in Practice

- How do you conduct empirical analyses?
  - collect the data over many months or years.

Model Dependence in Practice

- How do you conduct empirical analyses?
  - collect the data over many months or years.
  - finish recording and merging.

Model Dependence in Practice

- How do you conduct empirical analyses?
  - collect the data over many months or years.
  - finish recording and merging.
  - sit in front of your computer with nobody to bother you.

Model Dependence in Practice

- How do you conduct empirical analyses?
  - collect the data over many months or years.
  - finish recording and merging.
  - sit in front of your computer with nobody to bother you.
  - run one regression.

Model Dependence in Practice

- How do you conduct empirical analyses?
  - collect the data over many months or years.
  - finish recording and merging.
  - sit in front of your computer with nobody to bother you.
  - run one regression.
  - run another regression with different control variables.

Model Dependence in Practice

- How do you conduct empirical analyses?
    - collect the data over many months or years.
    - finish recording and merging.
    - sit in front of your computer with nobody to bother you.
    - run one regression.
    - run another regression with different control variables.
    - run another regression with different functional forms.

Model Dependence in Practice

- How do you conduct empirical analyses?
  - collect the data over many months or years.
  - finish recording and merging.
  - sit in front of your computer with nobody to bother you.
  - run one regression.
  - run another regression with different control variables.
  - run another regression with different functional forms.
  - run another regression with different measures.

Model Dependence in Practice

- How do you conduct empirical analyses?
  - collect the data over many months or years.
  - finish recording and merging.
  - sit in front of your computer with nobody to bother you.
  - run one regression.
  - run another regression with different control variables.
  - run another regression with different functional forms.
  - run another regression with different measures.
  - run yet another regression with a subset of the data.

Model Dependence in Practice

- How do you conduct empirical analyses?
  - collect the data over many months or years.
  - finish recording and merging.
  - sit in front of your computer with nobody to bother you.
  - run one regression.
  - run another regression with different control variables.
  - run another regression with different functional forms.
  - run another regression with different measures.
  - run yet another regression with a subset of the data.
  - end up with 100 or 1000 *different* estimates.

Model Dependence in Practice

- How do you conduct empirical analyses?
  - collect the data over many months or years.
  - finish recording and merging.
  - sit in front of your computer with nobody to bother you.
  - run one regression.
  - run another regression with different control variables.
  - run another regression with different functional forms.
  - run another regression with different measures.
  - run yet another regression with a subset of the data.
  - end up with 100 or 1000 *different* estimates.
  - put 1 or maybe 5 regression results in the paper.

Model Dependence in Practice

- How do you conduct empirical analyses?
  - collect the data over many months or years.
  - finish recording and merging.
  - sit in front of your computer with nobody to bother you.
  - run one regression.
  - run another regression with different control variables.
  - run another regression with different functional forms.
  - run another regression with different measures.
  - run yet another regression with a subset of the data.
  - end up with 100 or 1000 *different* estimates.
  - put 1 or maybe 5 regression results in the paper.
- What's the problem?

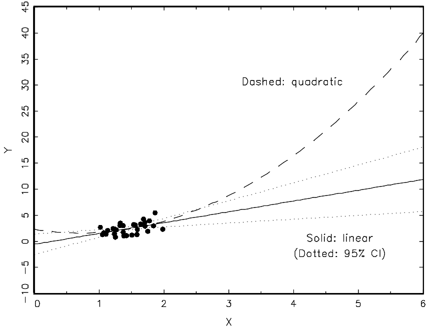
Model Dependence in Practice

- How do you conduct empirical analyses?
  - collect the data over many months or years.
  - finish recording and merging.
  - sit in front of your computer with nobody to bother you.
  - run one regression.
  - run another regression with different control variables.
  - run another regression with different functional forms.
  - run another regression with different measures.
  - run yet another regression with a subset of the data.
  - end up with 100 or 1000 *different* estimates.
  - put 1 or maybe 5 regression results in the paper.
- What's the problem?
  - Some specification is designated as the "correct" one, only after looking at the estimates.

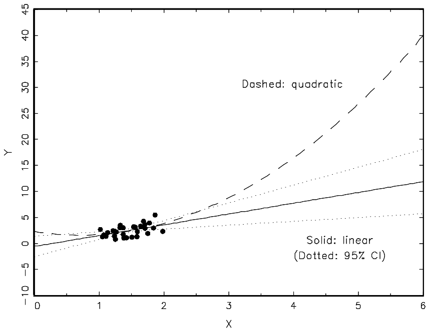
Model Dependence in Practice

- How do you conduct empirical analyses?
  - collect the data over many months or years.
  - finish recording and merging.
  - sit in front of your computer with nobody to bother you.
  - run one regression.
  - run another regression with different control variables.
  - run another regression with different functional forms.
  - run another regression with different measures.
  - run yet another regression with a subset of the data.
  - end up with 100 or 1000 *different* estimates.
  - put 1 or maybe 5 regression results in the paper.
- What's the problem?
  - Some specification is designated as the "correct" one, only after looking at the estimates.
  - Is this a true test of an ex ante hypothesis or merely a demonstration that it is *possible* to find results consistent with your favorite hypothesis?

Which model would you choose? (Both fit the data well.)

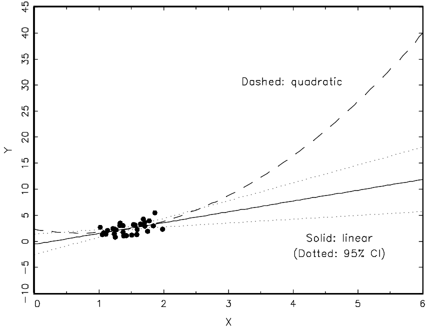


Which model would you choose? (Both fit the data well.)



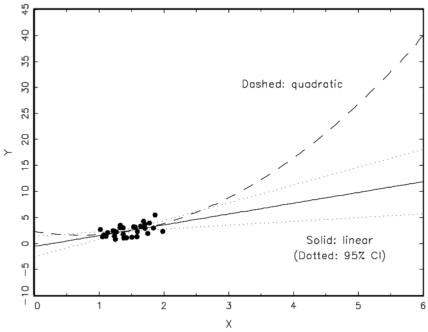
- Compare prediction at $x = 1.5$ to prediction at $x = 5$

Which model would you choose? (Both fit the data well.)



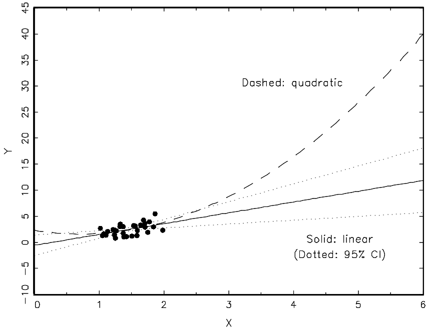
- Compare prediction at $x = 1.5$ to prediction at $x = 5$
- How do you choose a model?

Which model would you choose? (Both fit the data well.)



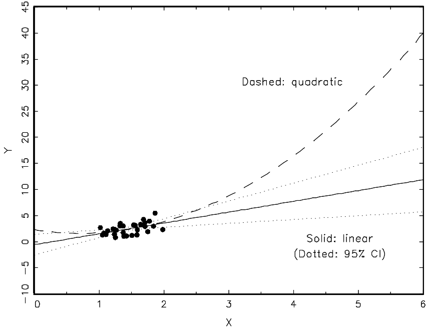
- Compare prediction at $x = 1.5$ to prediction at $x = 5$
- How do you choose a model? R^2?

Which model would you choose? (Both fit the data well.)



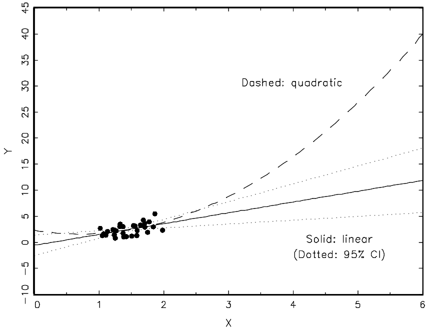
- Compare prediction at $x = 1.5$ to prediction at $x = 5$
- How do you choose a model? R^2? Some "test"?

Which model would you choose? (Both fit the data well.)



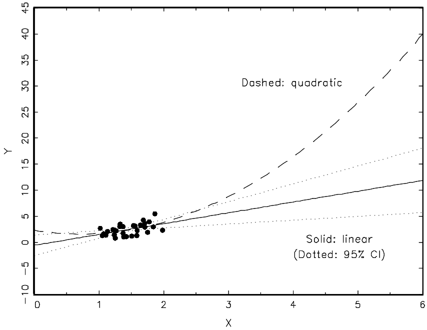
- Compare prediction at $x = 1.5$ to prediction at $x = 5$
- How do you choose a model? R^2? Some "test"? "Theory"?

Which model would you choose? (Both fit the data well.)



- Compare prediction at $x = 1.5$ to prediction at $x = 5$
- How do you choose a model? R^2? Some "test"? "Theory"?
- The bottom line: answers to some questions don't exist in the data.

Which model would you choose? (Both fit the data well.)



- Compare prediction at $x = 1.5$ to prediction at $x = 5$
- How do you choose a model? R^2? Some "test"? "Theory"?
- The bottom line: answers to some questions don't exist in the data.
- Same for what if questions, predictions, and causal inferences

Model Dependence Proof

Model Dependence Proof

Model Free Inference

Model Dependence Proof

Model Free Inference

To estimate $E(Y|X = x)$ at x, average many observed Y with value x

Model Dependence Proof

Model Free Inference

To estimate $E(Y|X = x)$ at x, average many observed Y with value x

Assumptions (Model-Based Inference)

Model Dependence Proof

Model Free Inference

To estimate $E(Y|X = x)$ at x, average many observed Y with value x

Assumptions (Model-Based Inference)

1. Definition: model dependence at x is the difference between predicted outcomes for any two models that fit about equally well.

Model Dependence Proof

Model Free Inference

To estimate $E(Y|X = x)$ at x, average many observed Y with value x

Assumptions (Model-Based Inference)

1. Definition: model dependence at x is the difference between predicted outcomes for any two models that fit about equally well.
2. The functional form follows strong continuity (think smoothness, although it is less restrictive)

Model Dependence Proof

Model Free Inference

To estimate $E(Y|X = x)$ at x, average many observed Y with value x

Assumptions (Model-Based Inference)

1. Definition: model dependence at x is the difference between predicted outcomes for any two models that fit about equally well.
2. The functional form follows strong continuity (think smoothness, although it is less restrictive)

Result

Model Dependence Proof

Model Free Inference

To estimate $E(Y|X = x)$ at x, average many observed Y with value x

Assumptions (Model-Based Inference)

1. Definition: model dependence at x is the difference between predicted outcomes for any two models that fit about equally well.
2. The functional form follows strong continuity (think smoothness, although it is less restrictive)

Result

The maximum degree of model dependence: solely a function of the distance from the counterfactual to the data

Detecting Model Dependence

- Randomly select a large number of infants

- Randomly select a large number of infants
- Randomly assign them to **0,6,8,10,12,16** years of education

Detecting Model Dependence

A (Hypothethical) Research Design

- Randomly select a large number of infants
- Randomly assign them to **0,6,8,10,12,16** years of education
- Assume 100% compliance, and no measurement error, omitted variables, or missing data

Detecting Model Dependence

A (Hypothethical) Research Design

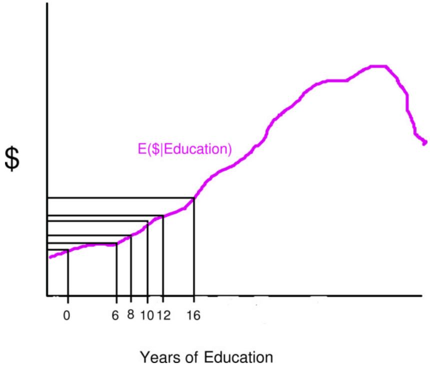
- Randomly select a large number of infants
- Randomly assign them to **0,6,8,10,12,16** years of education
- Assume 100% compliance, and no measurement error, omitted variables, or missing data
- Regress cumulative salary in year 17 on education

Detecting Model Dependence

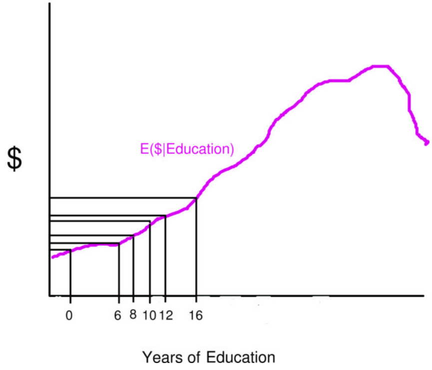
A (Hypothethical) Research Design

- Randomly select a large number of infants
- Randomly assign them to **0,6,8,10,12,16** years of education
- Assume 100% compliance, and no measurement error, omitted variables, or missing data
- Regress cumulative salary in year 17 on education
- We find a coefficient of $\hat{\beta} = \$1,000$, big t-statistics, narrow confidence intervals, and pass every test for auto-correlation, fit, normality, linearity, homoskedasticity, etc.

What Inferences Would You Be Willing to Make?

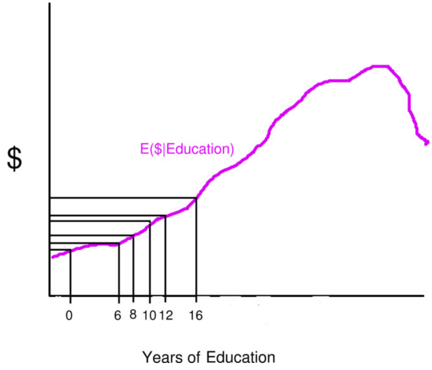


What Inferences Would You Be Willing to Make?



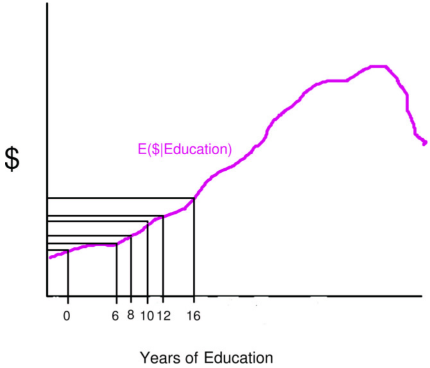
- A Factual Question: How much salary would someone receive with 12 years of education (a high school degree)?

What Inferences Would You Be Willing to Make?



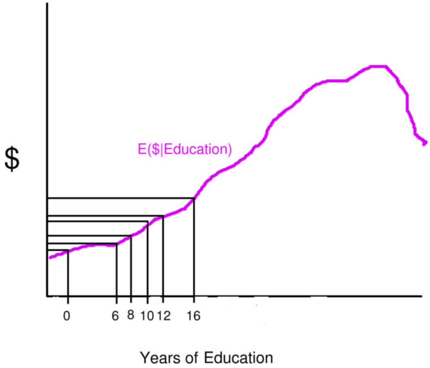
- A Factual Question: How much salary would someone receive with 12 years of education (a high school degree)?
- The model-free estimate: mean(Y) among those with $X = 12$.

What Inferences Would You Be Willing to Make?

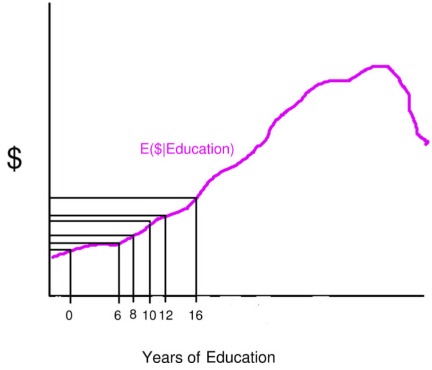


- A Factual Question: How much salary would someone receive with 12 years of education (a high school degree)?
- The model-free estimate: mean(Y) among those with $X = 12$.
- The model-based linear estimate: $\hat{Y} = X\hat{\beta} = 12 \times \$1,000 = \$12,000$

Counterfactual Inferences with Interpolation

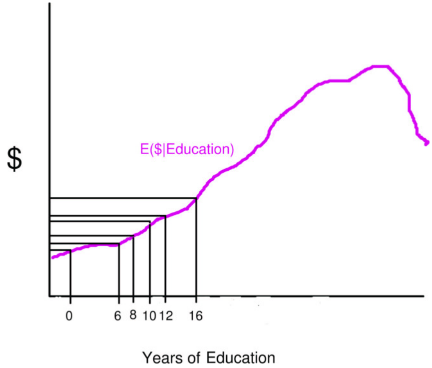


Counterfactual Inferences with Interpolation



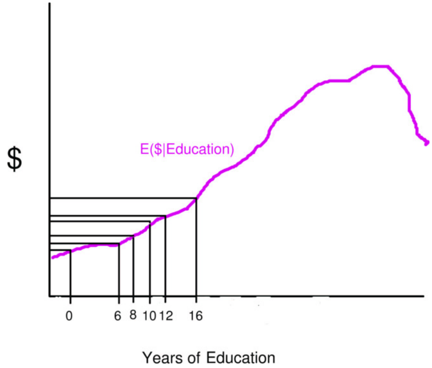
- How much salary would someone receive with 14 years of education (an Associates Degree)?

Counterfactual Inferences with Interpolation



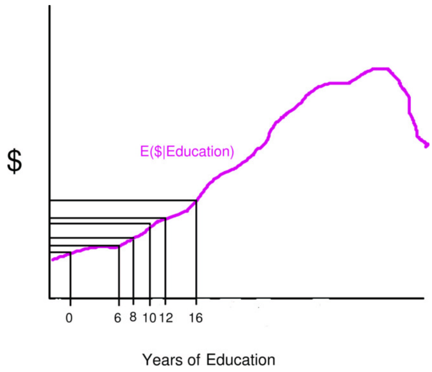
- How much salary would someone receive with 14 years of education (an Associates Degree)?
- Model free estimates impossible.

Counterfactual Inferences with Interpolation

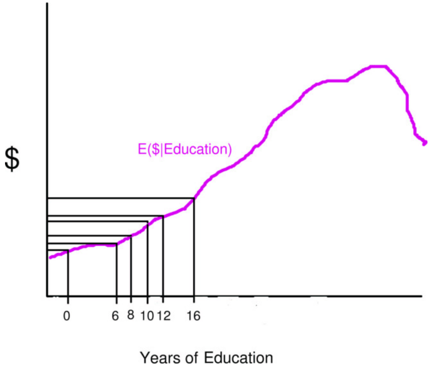


- How much salary would someone receive with 14 years of education (an Associates Degree)?
- Model free estimates impossible.
- $\hat{Y} = X\hat{\beta} = 14 \times \$1,000 = \$14,000$

Counterfactual Inference with Extrapolation

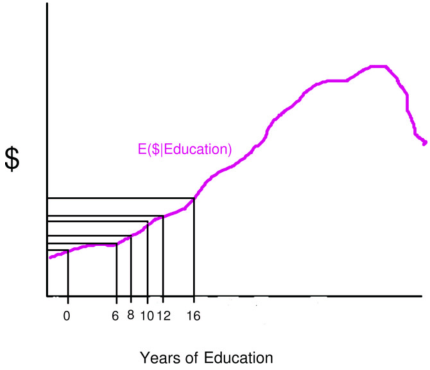


Counterfactual Inference with Extrapolation



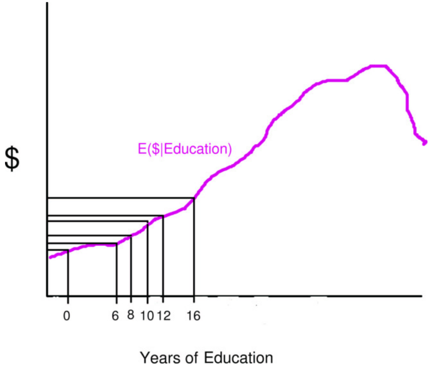
- How much salary would someone receive with 24 years of education (a Ph.D.)?

Counterfactual Inference with Extrapolation

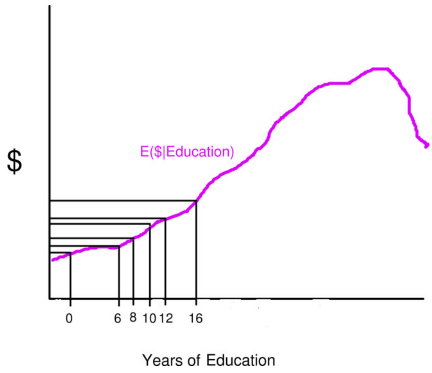


- How much salary would someone receive with 24 years of education (a Ph.D.)?
- $\hat{Y} = X\hat{\beta} = 24 \times \$1,000 = \$24,000$

Another Counterfactual Inference with Extrapolation

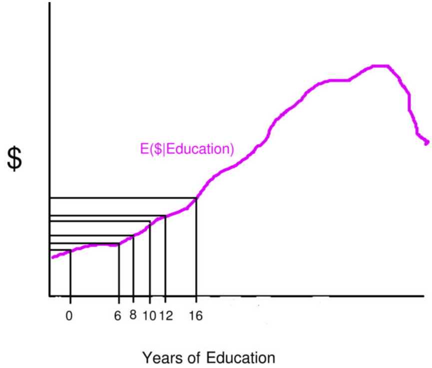


Another Counterfactual Inference with Extrapolation



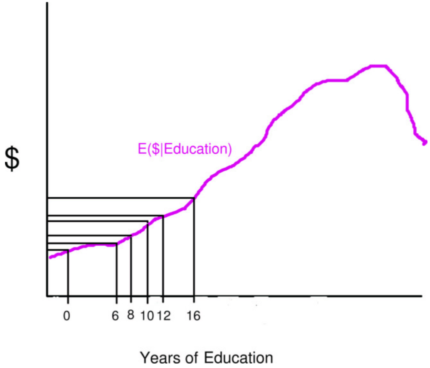
- How much salary would someone receive with 53 years of education?

Another Counterfactual Inference with Extrapolation



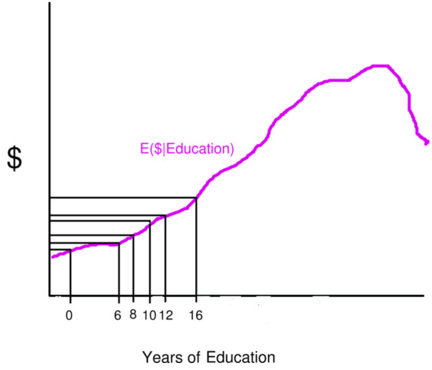
- How much salary would someone receive with 53 years of education?
- $\hat{Y} = X\hat{\beta} = 53 \times \$1,000 = \$53,000$

Another Counterfactual Inference with Extrapolation



- How much salary would someone receive with 53 years of education?
- $\hat{Y} = X\hat{\beta} = 53 \times \$1,000 = \$53,000$
- Recall: the regression passed every test and met every assumption; identical calculations worked for the other questions.

Another Counterfactual Inference with Extrapolation



- How much salary would someone receive with 53 years of education?
- $\hat{Y} = X\hat{\beta} = 53 \times \$1,000 = \$53,000$
- Recall: the regression passed every test and met every assumption; identical calculations worked for the other questions.
- What's changed? How would we recognize it when the example is less extreme or multidimensional?

Model Dependence with One Explanatory Variable

Model Dependence with One Explanatory Variable

- Suppose Y is starting salary; X is education in 10 categories.

Model Dependence with One Explanatory Variable

- Suppose Y is starting salary; X is education in 10 categories.
- To estimate $E(Y|X)$: we need 10 parameters, $E(Y|X = x_j)$, $j = 1, \ldots, 10$.

Model Dependence with One Explanatory Variable

- Suppose Y is starting salary; X is education in 10 categories.
- To estimate $E(Y|X)$: we need 10 parameters, $E(Y|X = x_j)$, $j = 1, \ldots, 10$.
- Model-free method: average 50 observations on Y for each value of X

Model Dependence with One Explanatory Variable

- Suppose Y is starting salary; X is education in 10 categories.
- To estimate $E(Y|X)$: we need 10 parameters, $E(Y|X = x_j)$, $j = 1, \ldots, 10$.
- Model-free method: average 50 observations on Y for each value of X
- Model-based method: regress Y on X, summarizing 10 parameters with 2 (intercept and slope).

Model Dependence with One Explanatory Variable

- Suppose Y is starting salary; X is education in 10 categories.
- To estimate $E(Y|X)$: we need 10 parameters, $E(Y|X = x_j)$, $j = 1, \ldots, 10$.
- Model-free method: average 50 observations on Y for each value of X
- Model-based method: regress Y on X, summarizing 10 parameters with 2 (intercept and slope).
- The difference between the 10 we need and the 2 we estimate with regression is pure assumption.

Model Dependence with One Explanatory Variable

- Suppose Y is starting salary; X is education in 10 categories.
- To estimate $E(Y|X)$: we need 10 parameters, $E(Y|X = x_j)$, $j = 1, \ldots, 10$.
- Model-free method: average 50 observations on Y for each value of X
- Model-based method: regress Y on X, summarizing 10 parameters with 2 (intercept and slope).
- The difference between the 10 we need and the 2 we estimate with regression is pure assumption.
- If X were continuous, we would be reducing ∞ to 2, also by assumption.

Model Dependence with Two Explanatory Variables

Model Dependence with Two Explanatory Variables

Variables: X (education) and Z, parent's income, both with 10 categories

Model Dependence with Two Explanatory Variables

Variables: X (education) and Z, parent's income, both with 10 categories

- How many parameters do we now need to estimate?

Model Dependence with Two Explanatory Variables

Variables: X (education) and Z, parent's income, both with 10 categories

- How many parameters do we now need to estimate? 20?

Model Dependence with Two Explanatory Variables

Variables: X (education) and Z, parent's income, both with 10 categories

- How many parameters do we now need to estimate? 20? Nope.

Model Dependence with Two Explanatory Variables

Variables: X (education) and Z, parent's income, both with 10 categories

- How many parameters do we now need to estimate? 20? Nope. Its $10 \times 10 = 100$.

Model Dependence with Two Explanatory Variables

Variables: X (education) and Z, parent's income, both with 10 categories

- How many parameters do we now need to estimate? 20? Nope. Its $10 \times 10 = 100$. This is the curse of dimensionality: the number of parameters goes up geometrically, not additively.

Model Dependence with Two Explanatory Variables

Variables: X (education) and Z, parent's income, both with 10 categories

- How many parameters do we now need to estimate? 20? Nope. Its $10 \times 10 = 100$. This is the curse of dimensionality: the number of parameters goes up geometrically, not additively.
- If we run a regression, we are summarizing 100 parameters with 3 (an intercept and two slopes).

Model Dependence with Two Explanatory Variables

Variables: X (education) and Z, parent's income, both with 10 categories

- How many parameters do we now need to estimate? 20? Nope. Its $10 \times 10 = 100$. This is the curse of dimensionality: the number of parameters goes up geometrically, not additively.
- If we run a regression, we are summarizing 100 parameters with 3 (an intercept and two slopes).
- But what about including an interaction? Right, so now we're summarizing 100 parameters with 4.

Model Dependence with Two Explanatory Variables

Variables: X (education) and Z, parent's income, both with 10 categories

- How many parameters do we now need to estimate? 20? Nope. Its $10 \times 10 = 100$. This is the curse of dimensionality: the number of parameters goes up geometrically, not additively.
- If we run a regression, we are summarizing 100 parameters with 3 (an intercept and two slopes).
- But what about including an interaction? Right, so now we're summarizing 100 parameters with 4.
- The difference is still one enormous assumption based on convenience, and neither evidence nor theory.

Model Dependence with Many Explanatory Variables

Model Dependence with Many Explanatory Variables

- Suppose: 15 explanatory variables, with 10 categories each.

Model Dependence with Many Explanatory Variables

- Suppose: 15 explanatory variables, with 10 categories each.
  - need to estimate 10^{15} (a quadrillion) parameters with how many observations?

Model Dependence with Many Explanatory Variables

- Suppose: 15 explanatory variables, with 10 categories each.
  - need to estimate 10^{15} (a quadrillion) parameters with how many observations?
  - Regression reduces this to 16 parameters, by assumption.

Model Dependence with Many Explanatory Variables

- Suppose: 15 explanatory variables, with 10 categories each.
  - need to estimate 10^{15} (a quadrillion) parameters with how many observations?
  - Regression reduces this to 16 parameters, by assumption.
- Suppose: 80 explanatory variables.

Model Dependence with Many Explanatory Variables

- Suppose: 15 explanatory variables, with 10 categories each.
  - need to estimate 10^{15} (a quadrillion) parameters with how many observations?
  - Regression reduces this to 16 parameters, by assumption.
- Suppose: 80 explanatory variables.
  - 10^{80} is more than the number of atoms in the universe.

Model Dependence with Many Explanatory Variables

- Suppose: 15 explanatory variables, with 10 categories each.
    - need to estimate 10^{15} (a quadrillion) parameters with how many observations?
    - Regression reduces this to 16 parameters, by assumption.
- Suppose: 80 explanatory variables.
    - 10^{80} is more than the number of atoms in the universe.
    - Yet, with a few simple assumptions, we can still run a regression and estimate only 81 parameters.

Model Dependence with Many Explanatory Variables

- Suppose: 15 explanatory variables, with 10 categories each.
  - need to estimate 10^{15} (a quadrillion) parameters with how many observations?
  - Regression reduces this to 16 parameters, by assumption.
- Suppose: 80 explanatory variables.
  - 10^{80} is more than the number of atoms in the universe.
  - Yet, with a few simple assumptions, we can still run a regression and estimate only 81 parameters.
- The curse of dimensionality introduces huge assumptions, often recognized.

We Ask: How Factual is your Counterfactual?

We Ask: How Factual is your Counterfactual?

- Readers have the right to know: is your counterfactual close enough to data so that statistical methods provide *empirical* answers?

We Ask: How Factual is your Counterfactual?

- Readers have the right to know: is your counterfactual close enough to data so that statistical methods provide *empirical* answers?
- If not, the same calculations will be based on indefensible model assumptions. With the curse of dimensionality, its too easy to fall into this trap.

We Ask: How Factual is your Counterfactual?

- Readers have the right to know: is your counterfactual close enough to data so that statistical methods provide *empirical* answers?
- If not, the same calculations will be based on indefensible model assumptions. With the curse of dimensionality, its too easy to fall into this trap.
- A good existing approach: *Sensitivity testing*, but this requires the user to specify a class of models and then to estimate them all and check how much inferences change

We Ask: How Factual is your Counterfactual?

- Readers have the right to know: is your counterfactual close enough to data so that statistical methods provide *empirical* answers?
- If not, the same calculations will be based on indefensible model assumptions. With the curse of dimensionality, its too easy to fall into this trap.
- A good existing approach: *Sensitivity testing*, but this requires the user to specify a class of models and then to estimate them all and check how much inferences change
- Our alternative approach:

We Ask: How Factual is your Counterfactual?

- Readers have the right to know: is your counterfactual close enough to data so that statistical methods provide *empirical* answers?
- If not, the same calculations will be based on indefensible model assumptions. With the curse of dimensionality, its too easy to fall into this trap.
- A good existing approach: *Sensitivity testing*, but this requires the user to specify a class of models and then to estimate them all and check how much inferences change
- Our alternative approach:
  - Specify your explanatory variables, X.

We Ask: How Factual is your Counterfactual?

- Readers have the right to know: is your counterfactual close enough to data so that statistical methods provide *empirical* answers?
- If not, the same calculations will be based on indefensible model assumptions. With the curse of dimensionality, its too easy to fall into this trap.
- A good existing approach: *Sensitivity testing*, but this requires the user to specify a class of models and then to estimate them all and check how much inferences change
- Our alternative approach:
  - Specify your explanatory variables, X.
  - Assume $\mathrm{E}(Y|X)$ is (minimally) smooth in X

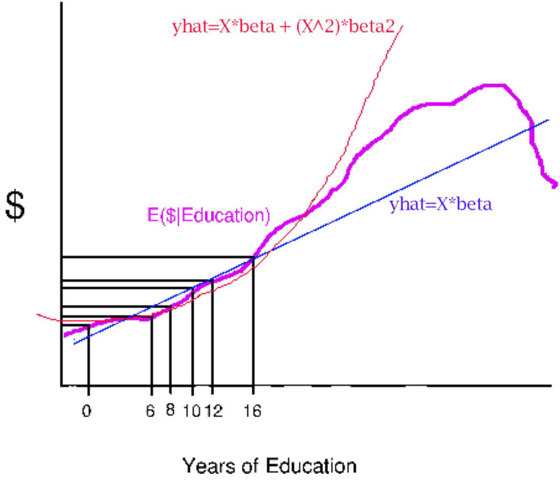
We Ask: How Factual is your Counterfactual?

- Readers have the right to know: is your counterfactual close enough to data so that statistical methods provide *empirical* answers?
- If not, the same calculations will be based on indefensible model assumptions. With the curse of dimensionality, its too easy to fall into this trap.
- A good existing approach: *Sensitivity testing*, but this requires the user to specify a class of models and then to estimate them all and check how much inferences change
- Our alternative approach:
  - Specify your explanatory variables, X.
  - Assume $E(Y|X)$ is (minimally) smooth in X
  - No need to specify models (or a class of models), estimators, or dependent variables.

We Ask: How Factual is your Counterfactual?

- Readers have the right to know: is your counterfactual close enough to data so that statistical methods provide *empirical* answers?
- If not, the same calculations will be based on indefensible model assumptions. With the curse of dimensionality, its too easy to fall into this trap.
- A good existing approach: *Sensitivity testing*, but this requires the user to specify a class of models and then to estimate them all and check how much inferences change
- Our alternative approach:
  - Specify your explanatory variables, X.
  - Assume $\mathrm{E}(Y|X)$ is (minimally) smooth in X
  - No need to specify models (or a class of models), estimators, or dependent variables.
  - Results of one run apply to the class of all models, all estimators, and all dependent variables.

Interpolation vs Extrapolation in one Dimension



Interpolation or Extrapolation in One and Two Dimensions

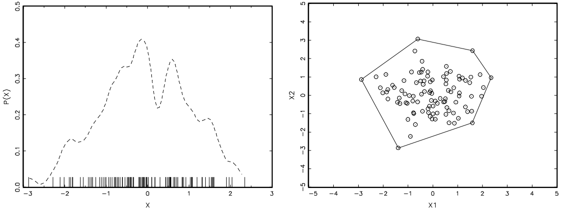


Figure: The Convex Hull

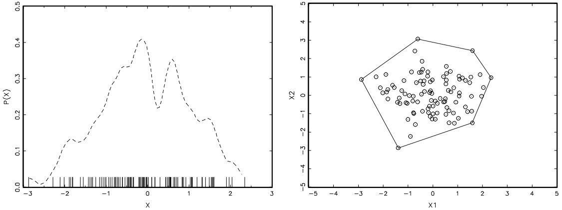


Figure: The Convex Hull

- Interpolation: Inside the convex hull

Interpolation or Extrapolation in One and Two Dimensions

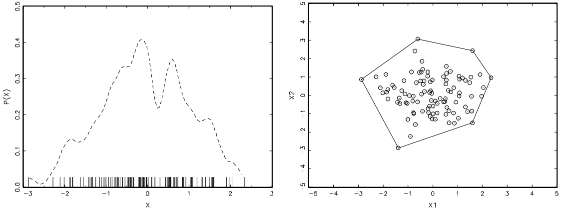


Figure: The Convex Hull

- Interpolation: Inside the convex hull
- Extrapolation: Outside the convex hull

Interpolation or Extrapolation in One and Two Dimensions

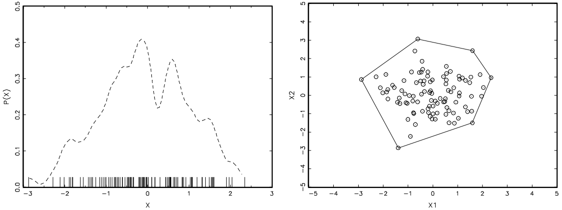


Figure: The Convex Hull

- Interpolation: Inside the convex hull
- Extrapolation: Outside the convex hull
- Works mathematically for any number of X variables

Interpolation or Extrapolation in One and Two Dimensions

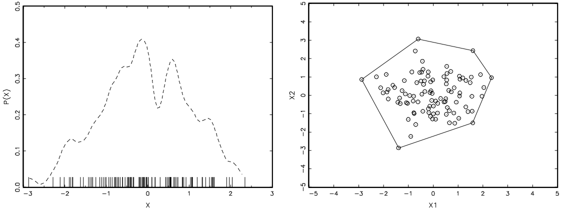


Figure: The Convex Hull

- Interpolation: Inside the convex hull
- Extrapolation: Outside the convex hull
- Works mathematically for any number of X variables
- We show how to determine whether a point is in the hull without calculating the hull, so its fast; see `http://GKing.harvard.edu/whatif`

Replication: Doyle and Sambanis, APSR 2000

Replication: Doyle and Sambanis, APSR 2000

- Data: 124 Post-World War II civil wars

Replication: Doyle and Sambanis, APSR 2000

- Data: 124 Post-World War II civil wars
- Dependent variable: peacebuilding success

Replication: Doyle and Sambanis, APSR 2000

- Data: 124 Post-World War II civil wars
- Dependent variable: peacebuilding success
- Treatment variable: multilateral UN peacekeeping intervention (0/1)

Replication: Doyle and Sambanis, APSR 2000

- Data: 124 Post-World War II civil wars
- Dependent variable: peacebuilding success
- Treatment variable: multilateral UN peacekeeping intervention (0/1)
- Control variables: war type, severity, and duration; development status; etc...

Replication: Doyle and Sambanis, APSR 2000

- Data: 124 Post-World War II civil wars
- Dependent variable: peacebuilding success
- Treatment variable: multilateral UN peacekeeping intervention (0/1)
- Control variables: war type, severity, and duration; development status; etc...
- Counterfactuals: UN intervention switched (0/1 to 1/0) for each observation

Replication: Doyle and Sambanis, APSR 2000

- Data: 124 Post-World War II civil wars
- Dependent variable: peacebuilding success
- Treatment variable: multilateral UN peacekeeping intervention (0/1)
- Control variables: war type, severity, and duration; development status; etc...
- Counterfactuals: UN intervention switched (0/1 to 1/0) for each observation
- Percent of counterfactuals in the convex hull:

Replication: Doyle and Sambanis, APSR 2000

- Data: 124 Post-World War II civil wars
- Dependent variable: peacebuilding success
- Treatment variable: multilateral UN peacekeeping intervention (0/1)
- Control variables: war type, severity, and duration; development status; etc...
- Counterfactuals: UN intervention switched (0/1 to 1/0) for each observation
- Percent of counterfactuals in the convex hull: 0%

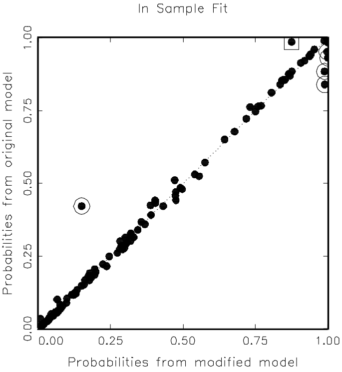
Replication: Doyle and Sambanis, APSR 2000

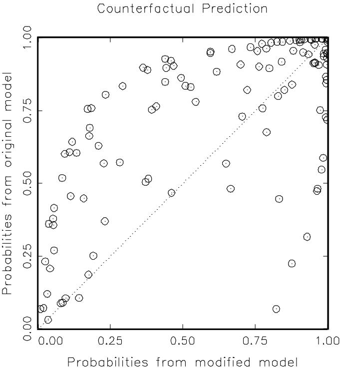
- Data: 124 Post-World War II civil wars
- Dependent variable: peacebuilding success
- Treatment variable: multilateral UN peacekeeping intervention (0/1)
- Control variables: war type, severity, and duration; development status; etc...
- Counterfactuals: UN intervention switched (0/1 to 1/0) for each observation
- Percent of counterfactuals in the convex hull: 0%
- Thus, without estimating any models, we know inferences will be model dependent; for illustration, let's find an example. . . .

Doyle and Sambanis, Logit Model

	Original Model			Modified Model		
Variables	Coeff	SE	P-val	Coeff	SE	P-val
Wartype	−1.742	.609	.004	−1.666	.606	.006
Logdead	−.445	.126	.000	−.437	.125	.000
Wardur	.006	.006	.258	.006	.006	.342
Factnum	−1.259	.703	.073	−1.045	.899	.245
Factnum2	.062	.065	.346	.032	.104	.756
Trnsfcap	.004	.002	.010	.004	.002	.017
Develop	.001	.000	.065	.001	.000	.068
Exp	−6.016	3.071	.050	−6.215	3.065	.043
Decade	−.299	.169	.077	−0.284	.169	.093
Treaty	2.124	.821	.010	2.126	.802	.008
UNOP4	3.135	1.091	.004	.262	1.392	.851
Wardur*UNOP4	—	—	—	.037	.011	.001
Constant	8.609	2.157	0.000	7.978	2.350	.000
N		122			122	
Log-likelihood		-45.649			-44.902	
Pseudo R^2		.423			.433	

Doyle and Sambanis: Model Dependence





Biases in Causal Inference: A New Decomposition

$$d = \text{mean}(Y|D = 1) - \text{mean}(Y|D = 0)$$

Biases in Causal Inference: A New Decomposition

$$d = \text{mean}(Y|D = 1) - \text{mean}(Y|D = 0)$$

Biases in Causal Inference: A New Decomposition

$$d = \text{mean}(Y|D = 1) - \text{mean}(Y|D = 0)$$

$$\text{bias} \equiv E(d) - \theta$$

Biases in Causal Inference: A New Decomposition

$$d = \text{mean}(Y|D=1) - \text{mean}(Y|D=0)$$

$$\text{bias} \equiv E(d) - \theta = \Delta_o + \Delta_p + \Delta_i + \Delta_e$$

Biases in Causal Inference: A New Decomposition

$$d = \text{mean}(Y|D = 1) - \text{mean}(Y|D = 0)$$

$$\text{bias} \equiv E(d) - \theta = \Delta_o + \Delta_p + \Delta_i + \Delta_e$$

- Δ_o Omitted variable bias

Biases in Causal Inference: A New Decomposition

$$d = \text{mean}(Y|D = 1) - \text{mean}(Y|D = 0)$$

$$\text{bias} \equiv E(d) - \theta = \Delta_o + \Delta_p + \Delta_i + \Delta_e$$

- Δ_o Omitted variable bias
- Δ_p Post-treatment bias

Biases in Causal Inference: A New Decomposition

$$d = \text{mean}(Y|D=1) - \text{mean}(Y|D=0)$$

$$\text{bias} \equiv E(d) - \theta = \Delta_o + \Delta_p + \Delta_i + \Delta_e$$

- Δ_o Omitted variable bias
- Δ_p Post-treatment bias
- Δ_i Interpolation bias

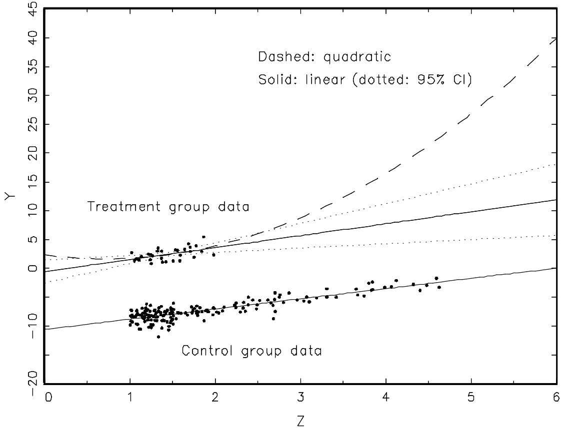
Biases in Causal Inference: A New Decomposition

$$d = \text{mean}(Y|D=1) - \text{mean}(Y|D=0)$$

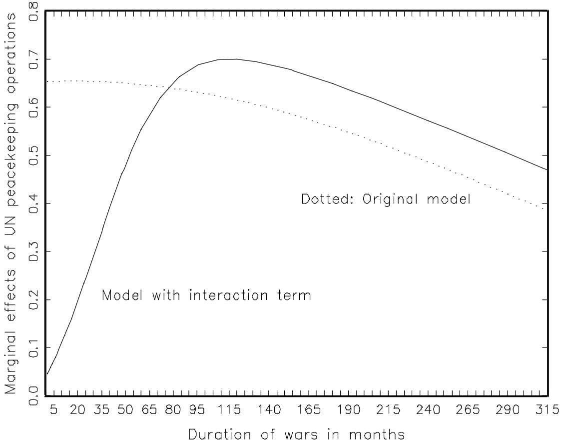
$$\text{bias} \equiv E(d) - \theta = \Delta_o + \Delta_p + \Delta_i + \Delta_e$$

- Δ_o Omitted variable bias
- Δ_p Post-treatment bias
- Δ_i Interpolation bias
- Δ_e Extrapolation bias

Interpolation vs Extrapolation Bias



Causal Effect of Multidimensional UN Peacekeeping Operations



The Matching Literature

The Matching Literature

- Matching, a new statistics literature on causal inference:

The Matching Literature

- Matching, a new statistics literature on causal inference:
  - nonparametric, non-model based methods.

The Matching Literature

- Matching, a new statistics literature on causal inference:
  - nonparametric, non-model based methods.
  - Promises to reduce or eliminate models and model dependence

The Matching Literature

- Matching, a new statistics literature on causal inference:
  - nonparametric, non-model based methods.
  - Promises to reduce or eliminate models and model dependence
  - Theory is sophisticated, but...

The Matching Literature

- Matching, a new statistics literature on causal inference:
  - nonparametric, non-model based methods.
  - Promises to reduce or eliminate models and model dependence
  - Theory is sophisticated, but...
- From the point of view of practical researchers,

The Matching Literature

- Matching, a new statistics literature on causal inference:
  - nonparametric, non-model based methods.
  - Promises to reduce or eliminate models and model dependence
  - Theory is sophisticated, but...
- From the point of view of practical researchers,
  - conflicting techniques, practices, guidelines, and rules of thumbs.

The Matching Literature

- Matching, a new statistics literature on causal inference:
  - nonparametric, non-model based methods.
  - Promises to reduce or eliminate models and model dependence
  - Theory is sophisticated, but...
- From the point of view of practical researchers,
  - conflicting techniques, practices, guidelines, and rules of thumbs.
  - calculation of valid standard errors is complicated or unavailable.

The Matching Literature

- Matching, a new statistics literature on causal inference:
  - nonparametric, non-model based methods.
  - Promises to reduce or eliminate models and model dependence
  - Theory is sophisticated, but...
- From the point of view of practical researchers,
  - conflicting techniques, practices, guidelines, and rules of thumbs.
  - calculation of valid standard errors is complicated or unavailable.
  - few relevant theoretical results exist.

The Matching Literature

- Matching, a new statistics literature on causal inference:
  - nonparametric, non-model based methods.
  - Promises to reduce or eliminate models and model dependence
  - Theory is sophisticated, but...
- From the point of view of practical researchers,
  - conflicting techniques, practices, guidelines, and rules of thumbs.
  - calculation of valid standard errors is complicated or unavailable.
  - few relevant theoretical results exist.
- Our unifying idea and proposed framework:

The Matching Literature

- Matching, a new statistics literature on causal inference:
  - nonparametric, non-model based methods.
  - Promises to reduce or eliminate models and model dependence
  - Theory is sophisticated, but...
- From the point of view of practical researchers,
  - conflicting techniques, practices, guidelines, and rules of thumbs.
  - calculation of valid standard errors is complicated or unavailable.
  - few relevant theoretical results exist.
- Our unifying idea and proposed framework:
  - Don't use matching as a substitute for parametric models

The Matching Literature

- Matching, a new statistics literature on causal inference:
  - nonparametric, non-model based methods.
  - Promises to reduce or eliminate models and model dependence
  - Theory is sophisticated, but...
- From the point of view of practical researchers,
  - conflicting techniques, practices, guidelines, and rules of thumbs.
  - calculation of valid standard errors is complicated or unavailable.
  - few relevant theoretical results exist.
- Our unifying idea and proposed framework:
  - Don't use matching as a substitute for parametric models
  - use matching to make parametric models work better.

The Matching Literature

- Matching, a new statistics literature on causal inference:
  - nonparametric, non-model based methods.
  - Promises to reduce or eliminate models and model dependence
  - Theory is sophisticated, but...
- From the point of view of practical researchers,
  - conflicting techniques, practices, guidelines, and rules of thumbs.
  - calculation of valid standard errors is complicated or unavailable.
  - few relevant theoretical results exist.
- Our unifying idea and proposed framework:
  - Don't use matching as a substitute for parametric models
  - use matching to make parametric models work better.
  - apply parametric analyses to preprocessed/matched data rather than raw data.

The Matching Literature

- Matching, a new statistics literature on causal inference:
  - nonparametric, non-model based methods.
  - Promises to reduce or eliminate models and model dependence
  - Theory is sophisticated, but...
- From the point of view of practical researchers,
  - conflicting techniques, practices, guidelines, and rules of thumbs.
  - calculation of valid standard errors is complicated or unavailable.
  - few relevant theoretical results exist.
- Our unifying idea and proposed framework:
  - Don't use matching as a substitute for parametric models
  - use matching to make parametric models work better.
  - apply parametric analyses to preprocessed/matched data rather than raw data.
  - can calculate valid standard errors using the same procedures.

The Matching Literature

- Matching, a new statistics literature on causal inference:
    - nonparametric, non-model based methods.
    - Promises to reduce or eliminate models and model dependence
    - Theory is sophisticated, but...
- From the point of view of practical researchers,
    - conflicting techniques, practices, guidelines, and rules of thumbs.
    - calculation of valid standard errors is complicated or unavailable.
    - few relevant theoretical results exist.
- Our unifying idea and proposed framework:
    - Don't use matching as a substitute for parametric models
    - use matching to make parametric models work better.
    - apply parametric analyses to preprocessed/matched data rather than raw data.
    - can calculate valid standard errors using the same procedures.
    - resulting estimates are less model dependent.

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.
- Any study that meets all three 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*.
- Any study that meets all three can estimate causal inferences without modeling assumptions.
- Observational studies:

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.
- Any study that meets all three can estimate causal inferences without modeling assumptions.
- Observational studies:
  - Any study that fails to meet *all* three 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.
- Any study that meets all three can estimate causal inferences without modeling assumptions.
- Observational studies:
  - Any study that fails to meet *all* three requirements of classical randomized experiments
  - Most research in every field is observational, and thus requires at least some assumptions.

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

Nonparametric Preprocessing

Nonparametric Preprocessing

- Adjust the data prior to the parametric analysis so that the relationship between t_i and X_i is eliminated or reduced.

Nonparametric Preprocessing

- Adjust the data prior to the parametric analysis so that the relationship between t_i and X_i is eliminated or reduced.
- Fundamental rule for avoiding selection bias: do not select on dependent variable.

Nonparametric Preprocessing

- Adjust the data prior to the parametric analysis so that the relationship between t_i and X_i is eliminated or reduced.
- Fundamental rule for avoiding selection bias: do not select on dependent variable.
  - Use a valid selection rule – a function of t_i and X_i only.

Nonparametric Preprocessing

- Adjust the data prior to the parametric analysis so that the relationship between t_i and X_i is eliminated or reduced.
- Fundamental rule for avoiding selection bias: do not select on dependent variable.
  - Use a valid selection rule – a function of t_i and X_i only.
  - Analogous to randomized blocks in experiments, stratified sampling in surveys.

Nonparametric Preprocessing

- Adjust the data prior to the parametric analysis so that the relationship between t_i and X_i is eliminated or reduced.
- Fundamental rule for avoiding selection bias: do not select on dependent variable.
  - Use a valid selection rule – a function of t_i and X_i only.
  - Analogous to randomized blocks in experiments, stratified sampling in surveys.
- With the preprocessed data set:

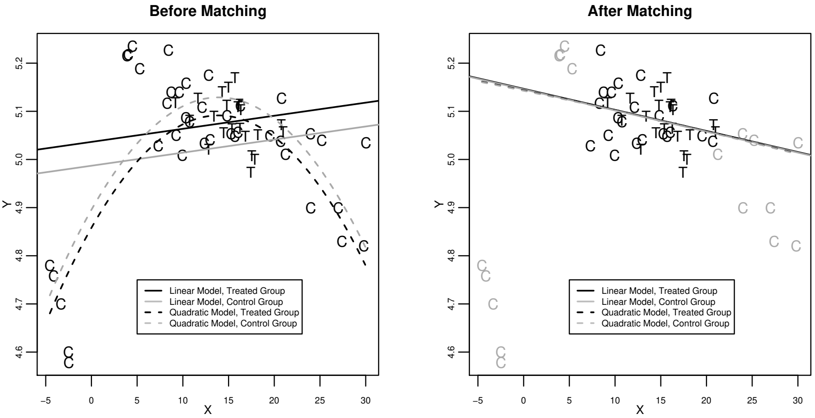
Nonparametric Preprocessing

- Adjust the data prior to the parametric analysis so that the relationship between t_i and X_i is eliminated or reduced.
- Fundamental rule for avoiding selection bias: do not select on dependent variable.
  - Use a valid selection rule – a function of t_i and X_i only.
  - Analogous to randomized blocks in experiments, stratified sampling in surveys.
- With the preprocessed data set:
  - model-dependence is reduced.

Nonparametric Preprocessing

- Adjust the data prior to the parametric analysis so that the relationship between t_i and X_i is eliminated or reduced.
- Fundamental rule for avoiding selection bias: do not select on dependent variable.
  - Use a valid selection rule – a function of t_i and X_i only.
  - Analogous to randomized blocks in experiments, stratified sampling in surveys.
- With the preprocessed data set:
  - model-dependence is reduced.
  - $p(X \mid t_i = 1) = p(X \mid t_i = 0)$ or $p(X \mid t_i = 1) \approx p(X \mid t_i = 0)$.

A Matching Example



Why Exact Matching Helps

Why Exact Matching Helps

- The goal, balance: $p(X|t = 1) = p(X|t = 0)$

Why Exact Matching Helps

- The goal, balance: $p(X|t=1) = p(X|t=0)$
- Exact matching: for every value of $X = x$ and $t = 0$, we have another for which $X = x$ and $t = 1$. Then by definition, $p(X|t=1) = p(X|t=0)$ holds

Why Exact Matching Helps

- The goal, balance: $p(X|t=1) = p(X|t=0)$
- Exact matching: for every value of $X = x$ and $t = 0$, we have another for which $X = x$ and $t = 1$. Then by definition, $p(X|t=1) = p(X|t=0)$ holds
- Normally, we will only approximate this goal, and will sacrifice some bias reduction (due to lack of balance) for more observations.

Choosing a Matching Procedure

Choosing a Matching Procedure

- The goal: improve balance without losing too many observations.

Choosing a Matching Procedure

- The goal: improve balance without losing too many observations.
- Try many matching procedures until better balance is achieved.

Choosing a Matching Procedure

- The goal: improve balance without losing too many observations.
- Try many matching procedures until better balance is achieved.
- But, do not examine the outcome variable during preprocessing.

Choosing a Matching Procedure

- The goal: improve balance without losing too many observations.
- Try many matching procedures until better balance is achieved.
- But, do not examine the outcome variable 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 without losing too many observations.
- Try many matching procedures until better balance is achieved.
- But, do not examine the outcome variable 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 without losing too many observations.
- Try many matching procedures until better balance is achieved.
- But, do not examine the outcome variable 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

Choosing a Matching Procedure

- The goal: improve balance without losing too many observations.
- Try many matching procedures until better balance is achieved.
- But, do not examine the outcome variable 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
- Evaluate the Matching Procedure: look at low-dimensional summaries of X (no hypothesis tests!)

Choosing a Matching Procedure

- The goal: improve balance without losing too many observations.
- Try many matching procedures until better balance is achieved.
- But, do not examine the outcome variable 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
- 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, ...

Empirical Illustration

Empirical Illustration

- Democratic senate majorities and FDA drug approval time (Carpenter 2002).

Empirical Illustration

- Democratic senate majorities and FDA drug approval time (Carpenter 2002).
  - Hypothesis: "expected approval times are greater when Democrats control the White House, when the agency's oversight committees are more liberal, and when the House and Senate are more liberal" (p.495).

Empirical Illustration

- Democratic senate majorities and FDA drug approval time (Carpenter 2002).
  - Hypothesis: "expected approval times are greater when Democrats control the White House, when the agency's oversight committees are more liberal, and when the House and Senate are more liberal" (p.495).
  - liberal FDA oversight should lead to slower approval of new drugs.

Empirical Illustration

- Democratic senate majorities and FDA drug approval time (Carpenter 2002).
  - Hypothesis: "expected approval times are greater when Democrats control the White House, when the agency's oversight committees are more liberal, and when the House and Senate are more liberal" (p.495).
  - liberal FDA oversight should lead to slower approval of new drugs.
- Original analysis:

Empirical Illustration

- Democratic senate majorities and FDA drug approval time (Carpenter 2002).
  - Hypothesis: "expected approval times are greater when Democrats control the White House, when the agency's oversight committees are more liberal, and when the House and Senate are more liberal" (p.495).
  - liberal FDA oversight should lead to slower approval of new drugs.
- Original analysis:
  - 408 new drugs (262 approved, 146 pending).

Empirical Illustration

- Democratic senate majorities and FDA drug approval time (Carpenter 2002).
  - Hypothesis: "expected approval times are greater when Democrats control the White House, when the agency's oversight committees are more liberal, and when the House and Senate are more liberal" (p.495).
  - liberal FDA oversight should lead to slower approval of new drugs.
- Original analysis:
  - 408 new drugs (262 approved, 146 pending).
  - lognormal survival model.

Empirical Illustration

- Democratic senate majorities and FDA drug approval time (Carpenter 2002).
  - Hypothesis: "expected approval times are greater when Democrats control the White House, when the agency's oversight committees are more liberal, and when the House and Senate are more liberal" (p.495).
  - liberal FDA oversight should lead to slower approval of new drugs.
- Original analysis:
  - 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

- Democratic senate majorities and FDA drug approval time (Carpenter 2002).
    - Hypothesis: "expected approval times are greater when Democrats control the White House, when the agency's oversight committees are more liberal, and when the House and Senate are more liberal" (p.495).
    - liberal FDA oversight should lead to slower approval of new drugs.
- Original analysis:
    - 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

- 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.

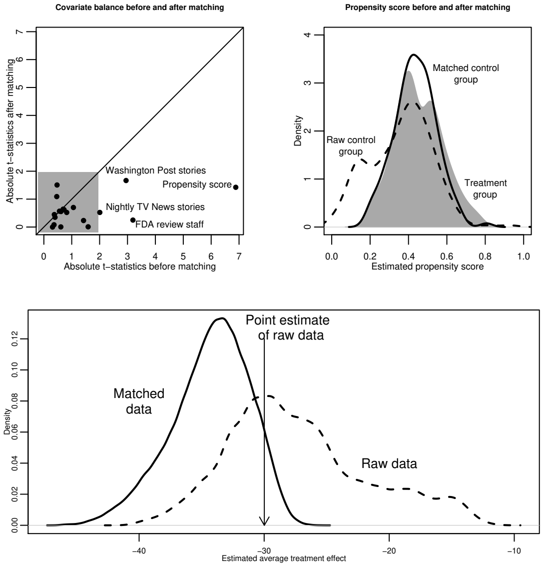
Evaluating Reduction in Model Dependence

- 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.)

Improved Balance and Reduced Model Dependence



Concluding Remarks

Concluding Remarks

- What can go wrong

Concluding Remarks

- What can go wrong
  - curse of dimensionality in balance diagnostics.

Concluding Remarks

- What can go wrong
  - curse of dimensionality in balance diagnostics.
  - preprocessing data may increase variance while reducing bias.

Concluding Remarks

- What can go wrong
  - curse of dimensionality in balance diagnostics.
  - preprocessing data may increase variance while reducing bias.
  - change in quantities of interest.

Concluding Remarks

- What can go wrong
  - curse of dimensionality in balance diagnostics.
  - preprocessing data may increase variance while reducing bias.
  - change in quantities of interest.
- Matching provides a way to get around ethical and methodological problems of choosing a model specification to present.

Concluding Remarks

- What can go wrong
  - curse of dimensionality in balance diagnostics.
  - preprocessing data may increase variance while reducing bias.
  - change in quantities of interest.
- Matching provides a way to get around ethical and methodological problems of choosing a model specification to present.
- Preprocessing the raw data with matching procedures makes familiar parametric models a much more reliable tool.

Concluding Remarks

- What can go wrong
  - curse of dimensionality in balance diagnostics.
  - preprocessing data may increase variance while reducing bias.
  - change in quantities of interest.
- Matching provides a way to get around ethical and methodological problems of choosing a model specification to present.
- Preprocessing the raw data with matching procedures makes familiar parametric models a much more reliable tool.
- Readers (and authors) need not worry that slightly different specifications alter the empirical conclusions.

Papers, software, etc.

http://GKing.Harvard.edu

Propensity Score Matching

Summarize all the variables in X with a single variable, $e_i(X_i) = \Pr(t_i = 1 \mid X_i)$.

- Estimate $e_i(X_i)$ with logistic regression, GAM, CART, neural network, etc.

Propensity Score Matching

Summarize all the variables in X with a single variable,
$e_i(X_i) = \Pr(t_i = 1 \mid X_i)$.

- Estimate $e_i(X_i)$ with logistic regression, GAM, CART, neural network, etc.
- Propensity score theory: if you have the true pscore specification, then the curse of dimensionality is solved and you can match on the one-dimensional pscore

Propensity Score Matching

Summarize all the variables in X with a single variable, $e_i(X_i) = \Pr(t_i = 1 \mid X_i)$.

- Estimate $e_i(X_i)$ with logistic regression, GAM, CART, neural network, etc.
- Propensity score theory: if you have the true pscore specification, then the curse of dimensionality is solved and you can match on the one-dimensional pscore
- Propensity score tautology:

Propensity Score Matching

Summarize all the variables in X with a single variable, $e_i(X_i) = \Pr(t_i = 1 \mid X_i)$.

- Estimate $e_i(X_i)$ with logistic regression, GAM, CART, neural network, etc.
- Propensity score theory: if you have the true pscore specification, then the curse of dimensionality is solved and you can match on the one-dimensional pscore
- Propensity score tautology:
    - if the pscore is correct, it balances X.

Propensity Score Matching

Summarize all the variables in X with a single variable, $e_i(X_i) = \Pr(t_i = 1 \mid X_i)$.

- Estimate $e_i(X_i)$ with logistic regression, GAM, CART, neural network, etc.
- Propensity score theory: if you have the true pscore specification, then the curse of dimensionality is solved and you can match on the one-dimensional pscore
- Propensity score tautology:
    - if the pscore is correct, it balances X. How do you know if it is correct?

Propensity Score Matching

Summarize all the variables in X with a single variable,
$e_i(X_i) = \Pr(t_i = 1 \mid X_i)$.

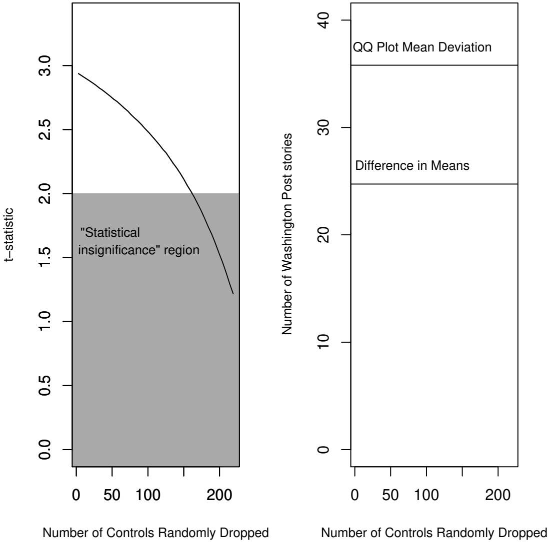
- Estimate $e_i(X_i)$ with logistic regression, GAM, CART, neural network, etc.
- Propensity score theory: if you have the true pscore specification, then the curse of dimensionality is solved and you can match on the one-dimensional pscore
- Propensity score tautology:
  - if the pscore is correct, it balances X. How do you know if it is correct? If it balances X, its correct.

Propensity Score Matching

Summarize all the variables in X with a single variable,
$e_i(X_i) = \Pr(t_i = 1 \mid X_i)$.

- Estimate $e_i(X_i)$ with logistic regression, GAM, CART, neural network, etc.
- Propensity score theory: if you have the true pscore specification, then the curse of dimensionality is solved and you can match on the one-dimensional pscore
- Propensity score tautology:
  - if the pscore is correct, it balances X. How do you know if it is correct? If it balances X, its correct.
  - I.e., it works when it works, and when it doesn't work, it doesn't work.

Hypothesis Tests for Balance Make No Sense



How Far Away Are the Data?

How Far Away Are the Data?

A useful question for counterfactuals just outside the hull or inside but far from the data.

How Far Away Are the Data?

A useful question for counterfactuals just outside the hull or inside but far from the data.

- Could estimate multivariate density $\mathrm{P}(X)$ and then compute hyper-volume near the counterfactual point: $\int_{x \in R} \mathrm{P}(X) dX$.

How Far Away Are the Data?

A useful question for counterfactuals just outside the hull or inside but far from the data.

- Could estimate multivariate density $\mathrm{P}(X)$ and then compute hyper-volume near the counterfactual point: $\int_{x \in R} \mathrm{P}(X) dX$.
- A simple way to do this is to assume $\mathrm{P}(X)$ is multivariate normal. Even with missing data, we can use Amelia for estimation.

How Far Away Are the Data?

A useful question for counterfactuals just outside the hull or inside but far from the data.

- Could estimate multivariate density $\mathrm{P}(X)$ and then compute hyper-volume near the counterfactual point: $\int_{x \in R} \mathrm{P}(X) dX$.
- A simple way to do this is to assume $\mathrm{P}(X)$ is multivariate normal. Even with missing data, we can use Amelia for estimation.
- Could use Gower's nonparametric measure of distance:

$$G_{ij} = \frac{1}{K} \sum_{k=1}^{K} \frac{|x_{ik} - x_{jk}|}{r_k}$$

where r_k is the range of variable k.

How Far Away Are the Data?

A useful question for counterfactuals just outside the hull or inside but far from the data.

- Could estimate multivariate density $\mathrm{P}(X)$ and then compute hyper-volume near the counterfactual point: $\int_{x \in R} \mathrm{P}(X) dX$.
- A simple way to do this is to assume $\mathrm{P}(X)$ is multivariate normal. Even with missing data, we can use Amelia for estimation.
- Could use Gower's nonparametric measure of distance:

$$G_{ij} = \frac{1}{K} \sum_{k=1}^{K} \frac{|x_{ik} - x_{jk}|}{r_k}$$

where r_k is the range of variable k.

- Regression confidence intervals widen as $\hat{y}$'s are farther from the data. This does not include model uncertainty, but we could use it as an index of how far we are from the data.

Assumptions for Causal Inference in Observational Data

Assumptions for Causal Inference in Observational Data

- no sample selection bias.

Assumptions for Causal Inference in Observational Data

- no sample selection bias.
  - Biased inferences to some population can be valid for our sample.

Assumptions for Causal Inference in Observational Data

- no sample selection bias.
    - Biased inferences to some population can be valid for our sample.
    - We could change the population to sample ATE or ATT

Assumptions for Causal Inference in Observational Data

- no sample selection bias.
  - Biased inferences to some population can be valid for our sample.
  - We could change the population to sample ATE or ATT
  - in-sample inferences are useful, and sometimes preferable, but generalization remains an issue.

Assumptions for Causal Inference in Observational Data

- no sample selection bias.
  - Biased inferences to some population can be valid for our sample.
  - We could change the population to sample ATE or ATT
  - in-sample inferences are useful, and sometimes preferable, but generalization remains an issue.
- no omitted variable bias: an unprovable issue for observational studies

Assumptions for Causal Inference in Observational Data

- no sample selection bias.
  - Biased inferences to some population can be valid for our sample.
  - We could change the population to sample ATE or ATT
  - in-sample inferences are useful, and sometimes preferable, but generalization remains an issue.
- no omitted variable bias: an unprovable issue for observational studies
- no posttreatment bias: possibly the most important overlooked problem in comparative politics and international relations

Assumptions for Causal Inference in Observational Data

- no sample selection bias.
  - Biased inferences to some population can be valid for our sample.
  - We could change the population to sample ATE or ATT
  - in-sample inferences are useful, and sometimes preferable, but generalization remains an issue.
- no omitted variable bias: an unprovable issue for observational studies
- no posttreatment bias: possibly the most important overlooked problem in comparative politics and international relations
- independent units (no interference between units), after taking into account X

Assumptions for Causal Inference in Observational Data

- no sample selection bias.
  - Biased inferences to some population can be valid for our sample.
  - We could change the population to sample ATE or ATT
  - in-sample inferences are useful, and sometimes preferable, but generalization remains an issue.
- no omitted variable bias: an unprovable issue for observational studies
- no posttreatment bias: possibly the most important overlooked problem in comparative politics and international relations
- independent units (no interference between units), after taking into account X
- same treatment within each treatment group

Omitted Variable Bias