Model Dependence in Counterfactual Inference

Gary King

October 20, 2005

References

References

- Related Software: WhatIf, MatchIt, Zelig


Related Software: WhatIf, MatchIt, Zelig

http://GKing.Harvard.edu
Counterfactuals

Three types:

1. Forecasts: Will the U.S. be in Iraq in 2008?
2. What-if 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.
Counterfactuals

Three types:

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

1. **Forecasts** Will the U.S. be in Iraq in 2008?
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.
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?
Model Dependence in Practice

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

- collect the data over many months or years.
- finish recording and merging.

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?
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.

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?
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?
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 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?
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 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?
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?
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?
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.
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.
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?
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.
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.)

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.
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? $R^2$?
• 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.)
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 $x = 5$
- 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 $x = 5$
- 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

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.
Model Dependence Proof

Model Free Inference

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

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

Assumptions (Model-Based Inference)

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

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
**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 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 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.
Detecting Model Dependence
A (Hypothetical) 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.
Detecting Model Dependence
A (Hypothetical) Research Design

- Randomly select a large number of infants
Detecting Model Dependence
A (Hypothetical) Research Design

- Randomly select a large number of infants
- Randomly assign them to 0, 6, 8, 10, 12, 16 years of education
Detecting Model Dependence
A (Hypothetical) 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 $\hat{\beta} = 1,000,000$, big t-statistics, narrow confidence intervals, and pass every test for auto-correlation, fit, normality, linearity, homoskedasticity, etc.
Detecting Model Dependence
A (Hypothetical) 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
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?

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: \(\hat{Y} = X \hat{\beta} = 12 \times \$1,000 = \$12,000\)

Counterfactuals and Model Dependence

October 20, 2005 8 / 40
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 $X = 12$. 

Counterfactuals and Model Dependence
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: \(\hat{Y} = X\hat{\beta} = 12 \times $1,000 = $12,000\)
How much salary would someone receive with 14 years of education (an Associates Degree)? Model free estimates impossible.

\[ \hat{Y} = \hat{X} \hat{\beta} = 14 \times \$1,000 = \$14,000 \]
How much salary would someone receive with 14 years of education (an Associates Degree)?
How much salary would someone receive with 14 years of education (an Associates Degree)?

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

Model free estimates impossible.

$$\hat{Y} = X\hat{\beta} = 14 \times $1,000 = $14,000$$
How much salary would someone receive with 24 years of education (a Ph.D.)?

\[ Y = X \hat{\beta} = 24 \times \$1,000 = \$24,000 \]

Counterfactuals and Model Dependence

October 20, 2005 10 / 40
How much salary would someone receive with 24 years of education (a Ph.D.)?

\[ Y = X \hat{\beta} = 24 \times \$1,000 = \$24,000 \]
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

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

\[ \hat{Y} = X\hat{\beta} = 53 \times $1,000 = $53,000 \]
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.
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?
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 $\infty$ to 2, also by assumption.
Suppose $Y$ is starting salary; $X$ is education in 10 categories.
Suppose $Y$ is starting salary; $X$ is education in 10 categories.

To estimate $E(Y|X)$: we need 10 parameters, $E(Y|X = x_j)$, $j = 1, \ldots, 10$. 
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 $X$
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).
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.
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 $\infty$ to 2, also by assumption.
How many parameters do we now need to estimate? 20? No. It's $10 \times 10 = 100$. 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 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. It's $10 \times 10 = 100$. 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.
How many parameters do we now need to estimate?

\[ 10 \times 10 = 100. \]

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 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.
How many parameters do we now need to estimate? 20? Nope.
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.
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.
How many parameters do we now need to estimate? 20? Nope. It's $10 \times 10 = 100$. 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).
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.
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

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

- need to estimate $10^{15}$ (a quadrillion) parameters with how many observations?
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: 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.
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.
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.
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?

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, it's 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. Results of one run apply to the class of all models, all estimators, and all 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?
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, it's too easy to fall into this trap.
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, it's 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.
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, it's 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 $X$. 
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, it's 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, it's 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.
  - Results of one run apply to the class of all models, all estimators, and all dependent variables.
Interpolation vs Extrapolation in one Dimension

\[ y_{\text{hat}} = X\beta + (X^2)\beta_2 \]

\[ E(\$|\text{Education}) \]

\[ y_{\text{hat}} = X\beta \]

Years of Education

\$
Interpolation or Extrapolation in One and Two Dimensions

Figure: The Convex Hull
Interpolation or Extrapolation in One and Two Dimensions

Figure: The Convex Hull

- **Interpolation**: Inside the convex hull

We show how to determine whether a point is in the hull without calculating the hull, so it's fast; see [http://GKing.harvard.edu/whatif](http://GKing.harvard.edu/whatif)
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

- **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 it's fast; see http://GKing.harvard.edu/whatif

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

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff</th>
<th>SE</th>
<th>P-val</th>
<th>Coeff</th>
<th>SE</th>
<th>P-val</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wartype</td>
<td>-1.742</td>
<td>.609</td>
<td>.004</td>
<td>-1.666</td>
<td>.606</td>
<td>.006</td>
</tr>
<tr>
<td>Logdead</td>
<td>-0.445</td>
<td>.126</td>
<td>.000</td>
<td>-0.437</td>
<td>.125</td>
<td>.000</td>
</tr>
<tr>
<td>Wardur</td>
<td>.006</td>
<td>.006</td>
<td>.258</td>
<td>.006</td>
<td>.006</td>
<td>.342</td>
</tr>
<tr>
<td>Factnum</td>
<td>-1.259</td>
<td>.703</td>
<td>.073</td>
<td>-1.045</td>
<td>.899</td>
<td>.245</td>
</tr>
<tr>
<td>Factnum2</td>
<td>.062</td>
<td>.065</td>
<td>.346</td>
<td>.032</td>
<td>.104</td>
<td>.756</td>
</tr>
<tr>
<td>Trnsfcap</td>
<td>.004</td>
<td>.002</td>
<td>.010</td>
<td>.004</td>
<td>.002</td>
<td>.017</td>
</tr>
<tr>
<td>Develop</td>
<td>.001</td>
<td>.000</td>
<td>.065</td>
<td>.001</td>
<td>.000</td>
<td>.068</td>
</tr>
<tr>
<td>Exp</td>
<td>-6.016</td>
<td>3.071</td>
<td>.050</td>
<td>-6.215</td>
<td>3.065</td>
<td>.043</td>
</tr>
<tr>
<td>Decade</td>
<td>-0.299</td>
<td>.169</td>
<td>.077</td>
<td>-0.284</td>
<td>.169</td>
<td>.093</td>
</tr>
<tr>
<td>Treaty</td>
<td>2.124</td>
<td>.821</td>
<td>.010</td>
<td>2.126</td>
<td>.802</td>
<td>.008</td>
</tr>
<tr>
<td>UNOP4</td>
<td>3.135</td>
<td>1.091</td>
<td>.004</td>
<td>.262</td>
<td>1.392</td>
<td>.851</td>
</tr>
<tr>
<td>Wardur*UNOP4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.037</td>
<td>.011</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>8.609</td>
<td>2.157</td>
<td>.000</td>
<td>7.978</td>
<td>2.350</td>
<td>.000</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>122</td>
<td></td>
<td></td>
<td>122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-45.649</td>
<td></td>
<td></td>
<td>-44.902</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>.423</td>
<td></td>
<td></td>
<td>.433</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
\[ d = \text{mean}(Y|D = 1) - \text{mean}(Y|D = 0) \]
$d = \text{mean}(Y|D = 1) - \text{mean}(Y|D = 0)$
\[ d = \text{mean}(Y|D = 1) - \text{mean}(Y|D = 0) \]

bias \equiv E(d) - \theta
\[ d = \text{mean}(Y|D = 1) - \text{mean}(Y|D = 0) \]

\[ \text{bias} \equiv E(d) - \theta = \Delta_o + \Delta_p + \Delta_i + \Delta_e \]
$$d = \text{mean}(Y|D = 1) - \text{mean}(Y|D = 0)$$

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

- $\Delta_o$ Omitted variable bias
Biases in Causal Inference: A New Decomposition

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

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

- \( \Delta_o \) Omitted variable bias
- \( \Delta_p \) Post-treatment bias
\[ d = \text{mean}(Y|D = 1) - \text{mean}(Y|D = 0) \]

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

- \( \Delta_o \) Omitted variable bias
- \( \Delta_p \) Post-treatment bias
- \( \Delta_i \) Interpolation bias
Biases in Causal Inference: A New Decomposition

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

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

- \( \Delta_o \) Omitted variable bias
- \( \Delta_p \) Post-treatment bias
- \( \Delta_i \) Interpolation bias
- \( \Delta_e \) Extrapolation bias
Interpolation vs Extrapolation Bias

Dashed: quadratic
Solid: linear (dotted: 95% CI)

Treatment group data
Control group data
Causal Effect of Multidimensional UN Peacekeeping Operations

![Graph showing the marginal effects of UN peacekeeping operations over the duration of wars in months. The graph compares two models: the original model (dotted line) and a model with an interaction term (solid line).]
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.
Matching, a new statistics literature on causal inference:

- Nonparametric, non-model based methods promise 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:

1. Don't use matching as a substitute for parametric models.
2. Use matching to make parametric models work better.
3. Apply parametric analyses to preprocessed/matched data rather than raw data.

Resulting estimates are less model dependent.
Matching, a new statistics literature on causal inference:
- nonparametric, non-model based methods.
Matching, a new statistics literature on causal inference:
- nonparametric, non-model based methods.
- Promises to reduce or eliminate models and model dependence.
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...
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,
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.
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.
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.
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:
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
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.
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.
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.
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.
Key features of classical randomized experiments:
Key features of classical randomized experiments:

- random selection of units from a given population.
Key features of classical randomized experiments:
- random selection of units from a given population.
- random assignment of values of the treatment.
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.
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.
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:
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.
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 \textit{all} three requirements of classical randomized experiments
- Most research in every field is observational, and thus requires at least some assumptions.
Researchers typically assume a parametric model (up to unknown parameters): e.g., $Y_i \sim p(\mu_i, \theta)$ with $\mu_i \equiv E(Y_i|t_i, X_i) = g(\alpha + t_i \beta + X_i \gamma)$.

Estimate the causal effect:

$$\text{ATT} = \text{mean}[g(\hat{\alpha} + \hat{\beta} + \hat{X}_i \hat{\gamma}) - g(\hat{\alpha} + \hat{X}_i \hat{\gamma})]$$

But, the true model is unknown.

In experiments, $T$ and $X$ are independent; we can drop $X$.

$$\text{ATT} = g(\hat{\alpha} + \hat{\beta}) - g(\hat{\alpha})$$

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.
Researchers typically assume a parametric model (up to unknown parameters):

\[ Y_i \sim p(\mu_i, \theta) \]

with \( \mu_i \equiv \mathbb{E}(Y_i | t_i, X_i) = g(\alpha + t_i \beta + X_i \gamma) \)

Estimate the causal effect:

\[ \text{ATT} = \text{mean}\left[ g(\hat{\alpha} + \hat{\beta} + X_i \hat{\gamma}) - g(\hat{\alpha} + X_i \hat{\gamma}) \right] \]

But, the true model is unknown.

In experiments, \( T \) and \( X \) are independent; we can drop \( X \).

\[ \text{ATT} = g(\hat{\alpha} + \hat{\beta}) - g(\hat{\alpha}) \]

The ATT requires no calculation over \( i \).

MLE is invariant to reparametrization, so \( g(\cdot) \) is irrelevant!

In observational studies, results are dependent on choice of \( g(\cdot) \).

Curse of dimensionality looms large.
Researchers typically

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

- But, the true model is unknown.
- In experiments, \( T \) and \( X \) are independent; we can drop \( X \)
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: $ATT = \text{mean}[g(\hat{\alpha} + \hat{\beta} + X_i \hat{\gamma}) - g(\hat{\alpha} + X_i \hat{\gamma})]$ 

But, the true model is unknown.

In experiments, $T$ and $X$ are independent; we can drop $X$

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

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 \( i \).
Researchers typically assume a parametric model (up to unknown parameters):

\[ Y_i \sim p(\mu_i, \theta) \text{ with } \mu_i \equiv E(Y_i \mid t_i, X_i) = g(\alpha + t_i \beta + X_i \gamma) \]

Estimate the causal effect: \( \text{ATT} = \text{mean}[g(\hat{\alpha} + \hat{\beta} + X_i \hat{\gamma}) - g(\hat{\alpha} + X_i \hat{\gamma})] \)

But, the true model is unknown.

In experiments, \( T \) and \( X \) are independent; we can drop \( X \)

\( \text{ATT} = g(\hat{\alpha} + \hat{\beta}) - g(\hat{\alpha}) \)

The ATT requires no calculation over \( i \).

MLE is invariant to reparameterization, so \( g(\cdot) \) is irrelevant!
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: \( ATT = \text{mean}[g(\hat{\alpha} + \hat{\beta} + X_i \hat{\gamma}) - g(\hat{\alpha} + X_i \hat{\gamma})] \)

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 reparameterization, so \( g(\cdot) \) is irrelevant!

In observational studies,
Researchers typically assume a parametric model (up to unknown parameters):

\[ Y_i \sim p(\mu_i, \theta) \text{ with } \mu_i \equiv E(Y_i \mid t_i, X_i) = g(\alpha + t_i \beta + X_i \gamma) \]

Estimate the causal effect: \( \text{ATT} = \text{mean}[g(\hat{\alpha} + \hat{\beta} + X_i \hat{\gamma}) - g(\hat{\alpha} + X_i \hat{\gamma})] \)

But, the true model is unknown.

In experiments, \( T \) and \( X \) are independent; we can drop \( X \)

\[ \text{ATT} = g(\hat{\alpha} + \hat{\beta}) - g(\hat{\alpha}) \]

The ATT requires no calculation over \( i \).

MLE is invariant to reparameterization, so \( g(\cdot) \) is irrelevant!

In observational studies,

results are dependent on choice of \( g(\cdot) \).
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: $\text{ATT}=\text{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 $X$

- $\text{ATT} = g(\hat{\alpha} + \hat{\beta}) - g(\hat{\alpha})$
- The ATT requires no calculation over $i$.
- MLE is invariant to reparameterization, so $g(\cdot)$ is irrelevant!

In observational studies,

- results are dependent on choice of $g(\cdot)$.
- curse of dimensionality looms large
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. $p(X|t_i=1) = p(X|t_i=0)$ or $p(X|t_i=1) \approx p(X|t_i=0)$. 

Counterfactuals and Model Dependence

October 20, 2005 27 / 40
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.

Analogous to randomized blocks in experiments, stratified sampling in surveys.

With the preprocessed data set:

\[
p(X|t_i = 1) = p(X|t_i = 0) \quad \text{or} \quad p(X|t_i = 1) \approx p(X|t_i = 0).
\]
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.
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 $p(X \mid t_i = 1) \approx p(X \mid t_i = 0)$. 

Counterfactuals and Model Dependence
A Matching Example

Before Matching

After Matching

- Linear Model, Treated Group
- Linear Model, Control Group
- Quadratic Model, Treated Group
- Quadratic Model, Control Group
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.
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
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

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, ...
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

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).

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.)
Democratic senate majorities and FDA drug approval time (Carpenter 2002).
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).
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.
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:
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).
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.
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).
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.)
Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).

- omit post-treatment variables.
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.
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).
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.
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.
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

Covariate balance before and after matching

Absolute t−statistics before matching

Absolute t−statistics after matching

Estimated propensity score

Density Raw control group

Matched control group

Treatment group

Estimated average treatment effect

Counterfactuals and Model Dependence

October 20, 2005 33 / 40
Concluding Remarks

What can go wrong with the curse of dimensionality in balance diagnostics. Preprocessing data may increase variance while reducing bias.

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.
Concluding Remarks

- What can go wrong
Concluding Remarks

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

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.
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.
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.
http://GKing.Harvard.edu
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.
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:
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 $X$. 
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

- "Statistical insignificance" region
- QQ Plot Mean Deviation
- Difference in Means

Number of Controls Randomly Dropped

Number of Washington Post stories

Difference in Means

QQ Plot Mean Deviation
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 $P(X)$ and then compute hyper-volume near the counterfactual point:

$$\int_{x \in \mathbb{R}} P(X) \, dX.$$ 

A simple way to do this is to assume $P(X)$ is multivariate normal. Even with missing data, we can use Amelia for estimation. Could use Gower's nonparametric measure of distance:

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

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.
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 \( P(X) \) and then compute hyper-volume near the counterfactual point: \( \int_{x \in \mathbb{R}} P(X) \, dx \).

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

- Could estimate multivariate density $P(X)$ and then compute hyper-volume near the counterfactual point: $\int_{x \in R} P(X) dX$.
- A simple way to do this is to assume $P(X)$ 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 $P(X)$ and then compute hyper-volume near the counterfactual point: $\int_{x \in R} P(X) dX$.
- A simple way to do this is to assume $P(X)$ 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.
A useful question for counterfactuals just outside the hull or inside but far from the data.

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

- A simple way to do this is to assume $P(X)$ 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

- 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.
no sample selection bias.
no sample selection bias.

Biased inferences to some population can be valid for our sample.
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
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