Nonparametric regression–Estimation, inference, and effects

StataCorp LLC

May 09, 2018

Initial Thoughts

- Why is nonparametric regression relevant?
- Nonparametric regression is agnostic
  - Unlike parametric estimation, nonparametric regression assumes no functional form for the relationship between outcomes and covariates.
  - You do not need to know the functional form to answer important research questions
  - You are not subject to problems that arise from misspecification

Initial Thoughts

- Why is nonparametric regression relevant?
- Nonparametric regression is agnostic
- Unlike parametric estimation, nonparametric regression assumes no functional form for the relationship between outcomes and covariates.
- You do not need to know the functional form to answer important research questions
- You are not subject to problems that arise from misspecification

Initial Thoughts

- Why is nonparametric regression relevant?
- Nonparametric regression is agnostic
- Unlike parametric estimation, nonparametric regression assumes no functional form for the relationship between outcomes and covariates.
- You do not need to know the functional form to answer important research questions
- You are not subject to problems that arise from misspecification

Initial Thoughts

- Why is nonparametric regression relevant?
- Nonparametric regression is agnostic
- Unlike parametric estimation, nonparametric regression assumes no functional form for the relationship between outcomes and covariates.
- You do not need to know the functional form to answer important research questions
- You are not subject to problems that arise from misspecification

Mean Function

- Some parametric functional form assumptions.
  - regression: $E(Y|X) = X\beta$
  - probit: $E(Y|X) = \Phi(X\beta)$
  - Poisson: $E(Y|X) = \exp(X\beta)$
- The relationship of interest is also a conditional mean:

$$E(y|X) = g(X)$$

- Where the mean function $g(\cdot)$ is unknown

Mean Function

- Some parametric functional form assumptions.
  - regression: $E(Y|X) = X\beta$
  - probit: $E(Y|X) = \Phi(X\beta)$
  - Poisson: $E(Y|X) = \exp(X\beta)$
- The relationship of interest is also a conditional mean:

$$E(y|X) = g(X)$$

- Where the mean function $g(\cdot)$ is unknown

Mean Function

- Some parametric functional form assumptions.
  - regression: $E(Y|X) = X\beta$
  - probit: $E(Y|X) = \Phi(X\beta)$
  - Poisson: $E(Y|X) = \exp(X\beta)$
- The relationship of interest is also a conditional mean:

$$E(y|X) = g(X)$$

- Where the mean function $g(\cdot)$ is unknown

Mean Function

- Some parametric functional form assumptions.
  - regression: $E\left(Y|X\right) = X\beta$
  - probit: $E\left(Y|X\right) = \Phi\left(X\beta\right)$
  - Poisson: $E\left(Y|X\right) = \exp\left(X\beta\right)$
- The relationship of interest is also a conditional mean:

$$E\left(y|X\right) = g\left(X\right)$$

- Where the mean function $g(\cdot)$ is unknown

Parametric regression and coefficients

- In the parametric models β fully characterizes the mean function
- We work hard to look and understand β
- Most of the interesting questions and results are inferences about the relationship of interest, the mean function.
- The answers to these questions is not β_j
- Nonparametric regression invites us to think in terms of the questions of interest
- Nonparametric regression invites you to think about inference not about parameters

Parametric regression and coefficients

- In the parametric models β fully characterizes the mean function
- We work hard to look and understand β
- Most of the interesting questions and results are inferences about the relationship of interest, the mean function.
- The answers to these questions is not β_j
- Nonparametric regression invites us to think in terms of the questions of interest
- Nonparametric regression invites you to think about inference not about parameters

Parametric regression and coefficients

- In the parametric models β fully characterizes the mean function
- We work hard to look and understand β
- Most of the interesting questions and results are inferences about the relationship of interest, the mean function.
- The answers to these questions is not β_j
- Nonparametric regression invites us to think in terms of the questions of interest
- Nonparametric regression invites you to think about inference not about parameters

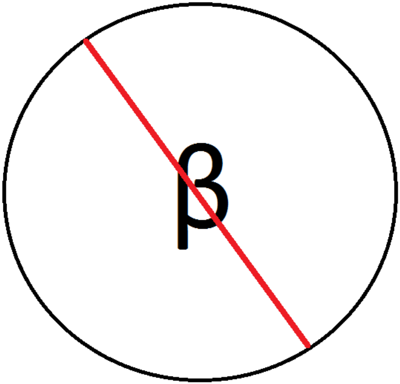
Parametric regression and coefficients

- In the parametric models β fully characterizes the mean function
- We work hard to look and understand β
- Most of the interesting questions and results are inferences about the relationship of interest, the mean function.
- The answers to these questions is not β_j
- Nonparametric regression invites us to think in terms of the questions of interest
- Nonparametric regression invites you to think about inference not about parameters

Parametric regression and coefficients

- In the parametric models β fully characterizes the mean function
- We work hard to look and understand β
- Most of the interesting questions and results are inferences about the relationship of interest, the mean function.
- The answers to these questions is not β_j
- Nonparametric regression invites us to think in terms of the questions of interest
- Nonparametric regression invites you to think about inference not about parameters

To Summarize the Discussion



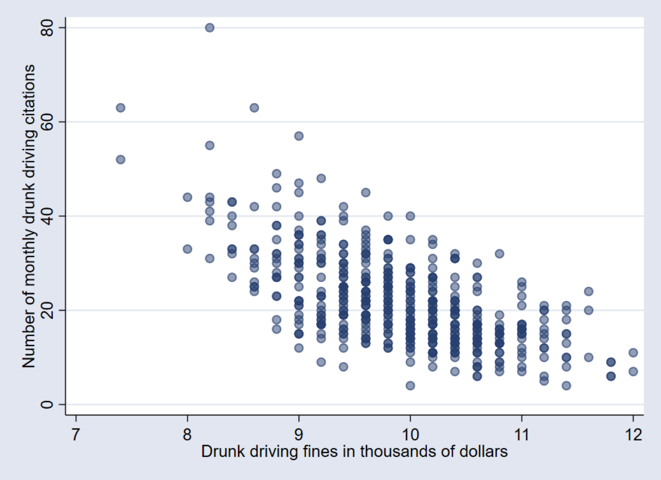
Traditional Approach to Nonparametric Estimation

- A cross section of counties
- `citations`: Number of monthly drunk driving citations
- `fines`: The value of fines imposed in a county in thousands of dollars if caught drinking and driving.

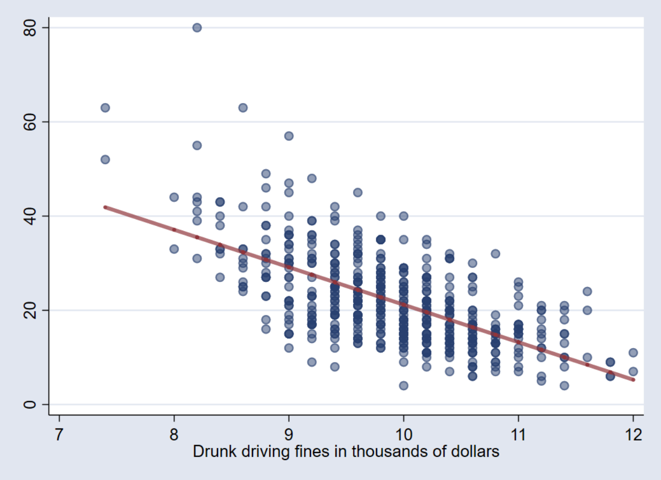
Traditional Approach to Nonparametric Estimation

- A cross section of counties
- `citations`: Number of monthly drunk driving citations
- `fines`: The value of fines imposed in a county in thousands of dollars if caught drinking and driving.

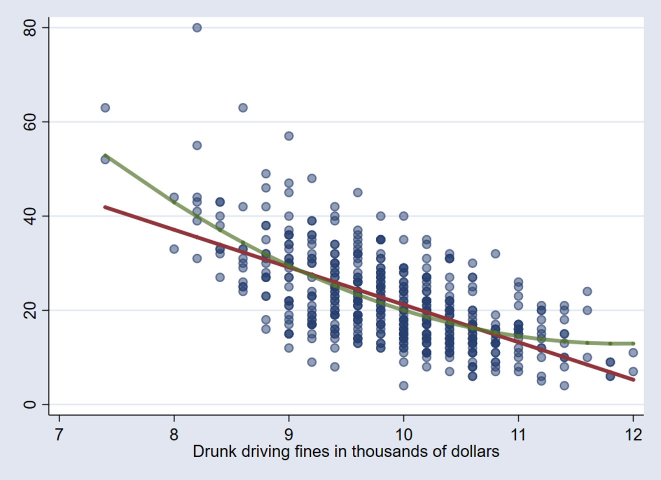
Implicit Relation



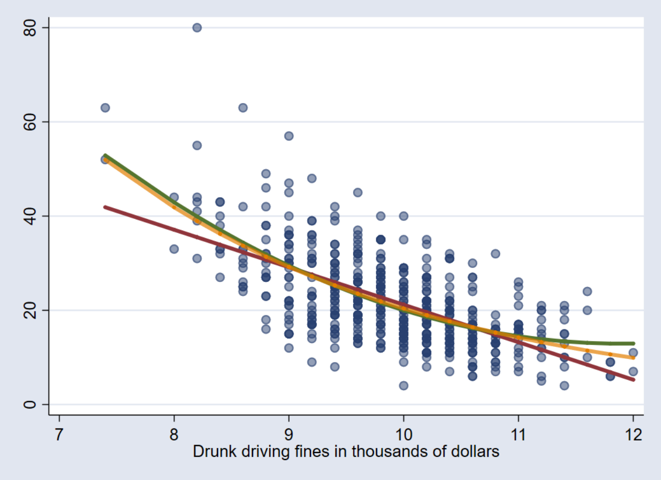
Simple linear regression



Regression with nonlinearities

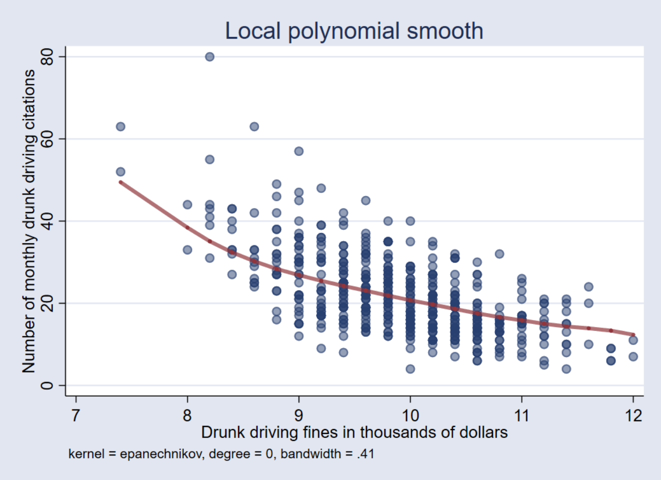


Poisson regression



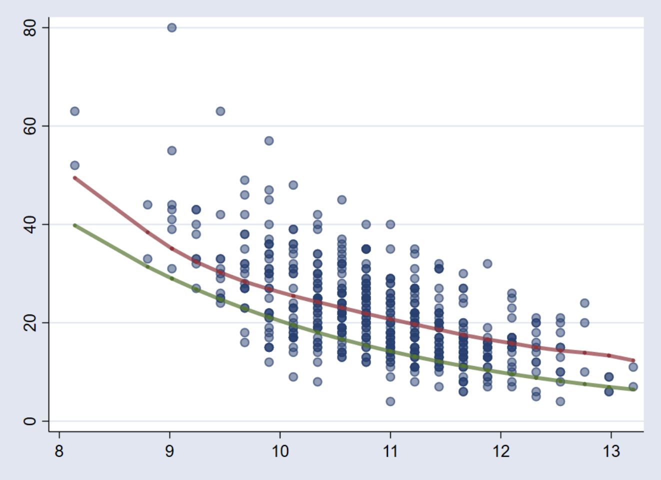
Nonparametric Estimation of Mean Function

```
. lpoly citations fines
```

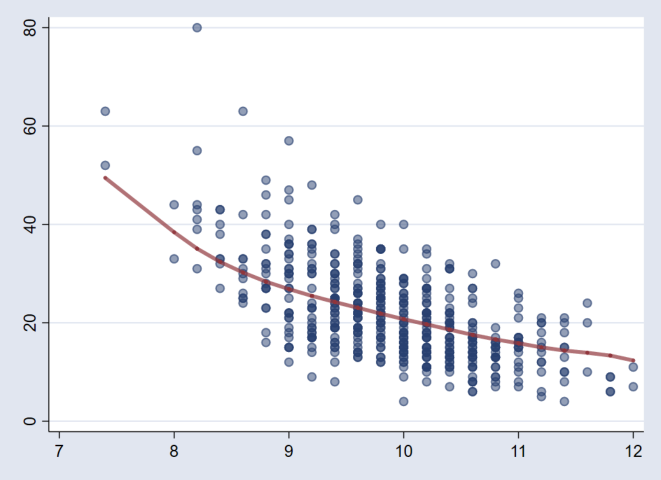


Now That We have the Mean Function

- What is the effect on the mean of `citations` of increasing `fines` by 10% ?



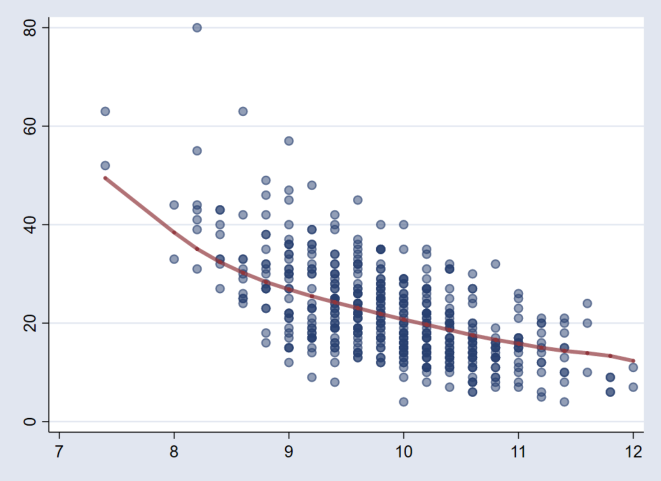
Traditional Approach Gives Us



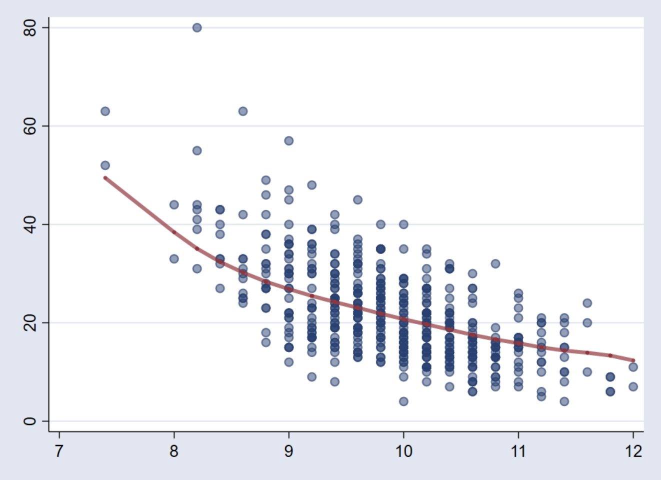
Additional Variables

- I would like to add controls
  - Whether county has a college town `college`
  - Number of highway patrol `patrols` units per capita in the county
- With those controls I can ask some new questions

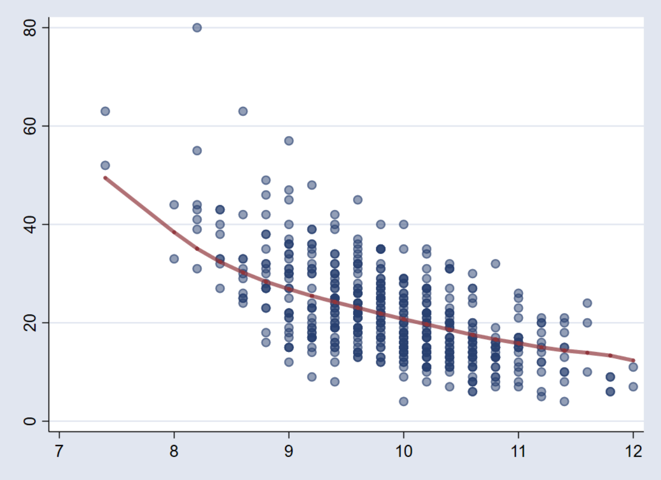
- What is the mean of `citations` if I increase `patrols` and `fines` ?



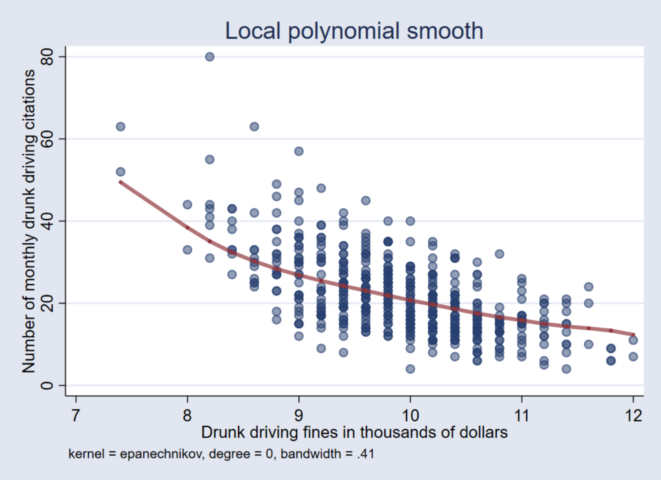
- How does the mean of `citations` differ for counties where there is a college town, averaging out the effect of `patrols` and `fines`?



- What policy has a bigger effect on the mean of `citations`, an increase in `fines`, an increase in `patrols`, or a combination of both?



What We Have Is



What We Have

- I have a mean function. That makes no functional form assumptions.
- I cannot answer the previous questions.
- My analysis was graphical not statistical
- My analysis is limited to one covariate
- This is true even if I give you the true mean function, $g(X)$

What We Have

- I have a mean function. That makes no functional form assumptions.
- I cannot answer the previous questions.
- My analysis was graphical not statistical
- My analysis is limited to one covariate
- This is true even if I give you the true mean function, $g(X)$

What We Have

- I have a mean function. That makes no functional form assumptions.
- I cannot answer the previous questions.
- My analysis was graphical not statistical
- My analysis is limited to one covariate
- This is true even if I give you the true mean function, $g(X)$

What We Have

- I have a mean function. That makes no functional form assumptions.
- I cannot answer the previous questions.
- My analysis was graphical not statistical
- My analysis is limited to one covariate
- This is true even if I give you the true mean function, $g(X)$

What We Have

- I have a mean function. That makes no functional form assumptions.
- I cannot answer the previous questions.
- My analysis was graphical not statistical
- My analysis is limited to one covariate
- This is true even if I give you the true mean function, $g(X)$

npregress

- New command in Stata 15 for nonparametric regression estimation
- We will be able to answer these question and make inferences using the mean function
- We will be able to include multiple continuous and discrete covariates
- `npregress` is an estimator not just a graphical tool
- It is a Stata estimator. You are going to be able to ask question and get inferences using your estimator.
- Stata is unique in being able to provide nonparametric graphics, estimation, and inference

npregress

- New command in Stata 15 for nonparametric regression estimation
- We will be able to answer these question and make inferences using the mean function
- We will be able to include multiple continuous and discrete covariates
- `npregress` is an estimator not just a graphical tool
- It is a Stata estimator. You are going to be able to ask question and get inferences using your estimator.
- Stata is unique in being able to provide nonparametric graphics, estimation, and inference

npregress

- New command in Stata 15 for nonparametric regression estimation
- We will be able to answer these question and make inferences using the mean function
- We will be able to include multiple continuous and discrete covariates
- npregress is an estimator not just a graphical tool
- It is a Stata estimator. You are going to be able to ask question and get inferences using your estimator.
- Stata is unique in being able to provide nonparametric graphics, estimation, and inference

`npregress`

- New command in Stata 15 for nonparametric regression estimation
- We will be able to answer these question and make inferences using the mean function
- We will be able to include multiple continuous and discrete covariates
- `npregress` is an estimator not just a graphical tool
- It is a Stata estimator. You are going to be able to ask question and get inferences using your estimator.
- Stata is unique in being able to provide nonparametric graphics, estimation, and inference

npregress

- New command in Stata 15 for nonparametric regression estimation
- We will be able to answer these question and make inferences using the mean function
- We will be able to include multiple continuous and discrete covariates
- `npregress` is an estimator not just a graphical tool
- It is a Stata estimator. You are going to be able to ask question and get inferences using your estimator.
- Stata is unique in being able to provide nonparametric graphics, estimation, and inference

`npregress`

- New command in Stata 15 for nonparametric regression estimation
- We will be able to answer these question and make inferences using the mean function
- We will be able to include multiple continuous and discrete covariates
- `npregress` is an estimator not just a graphical tool
- It is a Stata estimator. You are going to be able to ask question and get inferences using your estimator.
- Stata is unique in being able to provide nonparametric graphics, estimation, and inference

Nonparametric Estimation Intuition

Nonparametric regression: discrete covariates

Mean function for a discrete covariate

- Mean (probability) of low birthweight (`lbweight`) conditional on smoking 1 to 5 cigarettes (`msmoke`=1) during pregnancy

```
. mean lbweight if msmoke==1
Mean estimation                   Number of obs   =        480
--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
    lbweight |      .1125   .0144375      .0841313    .1408687
--------------------------------------------------------------
```

- `regress lbweight 1.msmoke, noconstant`
- $E(lbweigth|msmoke = 1)$, nonparametric estimate

Nonparametric regression: discrete covariates

Mean function for a discrete covariate

- Mean (probability) of low birthweight (`lbweight`) conditional on smoking 1 to 5 cigarettes (`msmoke`=1) during pregnancy

```
. mean lbweight if msmoke==1
Mean estimation                    Number of obs    =       480

             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
    lbweight |      .1125   .0144375      .0841313    .1408687
```

- `regress lbweight 1.msmoke, noconstant`
- $E(lbweigth|msmoke = 1)$, nonparametric estimate

Nonparametric regression: discrete covariates

Mean function for a discrete covariate

- Mean (probability) of low birthweight (`lbweight`) conditional on smoking 1 to 5 cigarettes (`msmoke`=1) during pregnancy

```
. mean lbweight if msmoke==1
Mean estimation                    Number of obs    =        480
```

	Mean	Std. Err.	[95% Conf.	Interval]
lbweight	.1125	.0144375	.0841313	.1408687

- `regress lbweight 1.msmoke, noconstant`
- $E(lbweigth|msmoke = 1)$, nonparametric estimate

Nonparametric regression: continuous covariates

Conditional mean for a continuous covariate

- low birthweight conditional on log of family income `fincome`
- $E(lbweight|fincome = 10.819)$
- Take observations **near** the value of 10.819 and then take an average
- $|fincome_i - 10.819| \leq h$
- h is a small number referred to as the bandwidth

Nonparametric regression: continuous covariates

Conditional mean for a continuous covariate

- low birthweight conditional on log of family income `fincome`
- $E(lbweight|fincome = 10.819)$
- Take observations **near** the value of 10.819 and then take an average
- $|fincome_i - 10.819| \leq h$
- h is a small number referred to as the bandwidth

Nonparametric regression: continuous covariates

Conditional mean for a continuous covariate

- low birthweight conditional on log of family income `fincome`
- $E(lbweight|fincome = 10.819)$
- Take observations **near** the value of 10.819 and then take an average
- $|fincome_i - 10.819| \leq h$
- h is a small number referred to as the bandwidth

Nonparametric regression: continuous covariates

Conditional mean for a continuous covariate

- low birthweight conditional on log of family income `fincome`
- $E(lbweight|fincome = 10.819)$
- Take observations **near** the value of 10.819 and then take an average
- $|fincome_i - 10.819| \leq h$
- h is a small number referred to as the bandwidth

Nonparametric regression: continuous covariates

Conditional mean for a continuous covariate

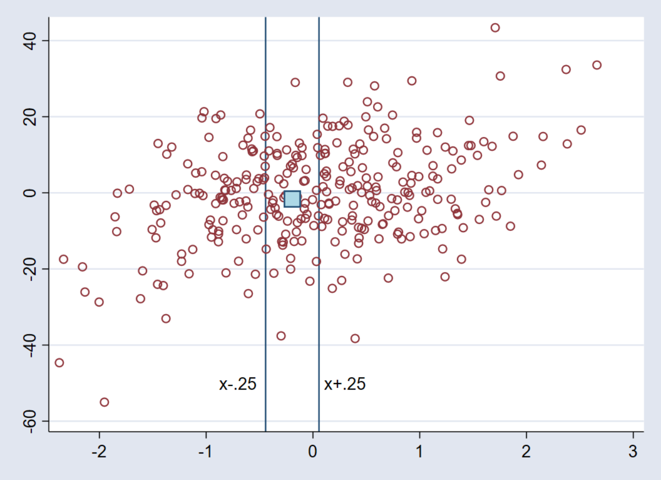
- low birthweight conditional on log of family income `fincome`
- $E(lbweight|fincome = 10.819)$
- Take observations **near** the value of 10.819 and then take an average
- $|fincome_i - 10.819| \leq h$
- h is a small number referred to as the bandwidth

Nonparametric regression: continuous covariates

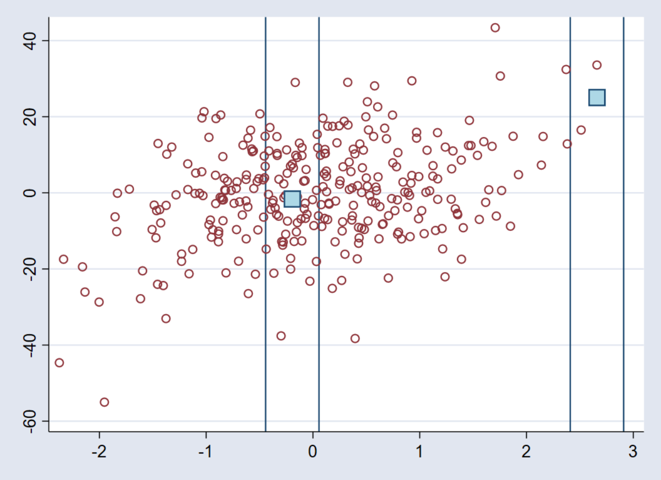
Conditional mean for a continuous covariate

- low birthweight conditional on log of family income `fincome`
- $E(lbweight|fincome = 10.819)$
- Take observations **near** the value of 10.819 and then take an average
- $|fincome_i - 10.819| \leq h$
- h is a small number referred to as the bandwidth

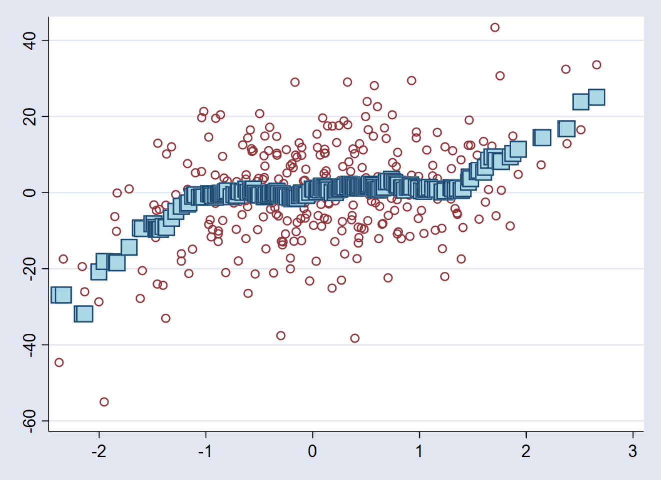
Graphical representation



Graphical example



Graphical example continued



Two concepts

1. h !!!!
2. Definition of distance between points, $|x_i - x| \leq h$

Kernel weights

- Epanechnikov
- Gaussian
- Epanechnikov2
- Rectangular(Uniform)
- Triangular
- Biweight
- Triweight
- Cosine
- Parzen

Kernel weights

- Epanechnikov
- Gaussian
- Epanechnikov2
- Rectangular(Uniform)
- Triangular
- Biweight
- Triweight
- Cosine
- Parzen

Discrete bandwidths

- Li–Racine Kernel

$$k(\cdot) = \begin{cases} 1 & \text{if} \quad x_i = x \\ h & \text{otherwise} \end{cases}$$

- Cell mean

$$k(\cdot) = \begin{cases} 1 & \text{if} \quad x_i = x \\ 0 & \text{otherwise} \end{cases}$$

- Cell mean was used in the example of discrete covariate estimate $E(lbweigth|msmoke = 1)$

Discrete bandwidths

- Li–Racine Kernel

$$k(\cdot) = \begin{cases} 1 & \text{if} \quad x_i = x \\ h & \text{otherwise} \end{cases}$$

- Cell mean

$$k(\cdot) = \begin{cases} 1 & \text{if} \quad x_i = x \\ 0 & \text{otherwise} \end{cases}$$

- Cell mean was used in the example of discrete covariate estimate $E(lbweigth|msmoke = 1)$

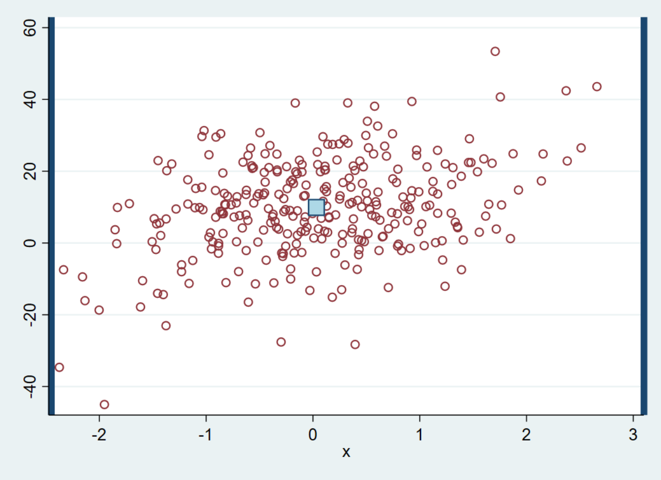
Selecting The Bandwidth

- A very large bandwidth will give you a biased estimate of the mean function with a small variance
- A very small bandwidth will give you an estimate with small bias and large variance

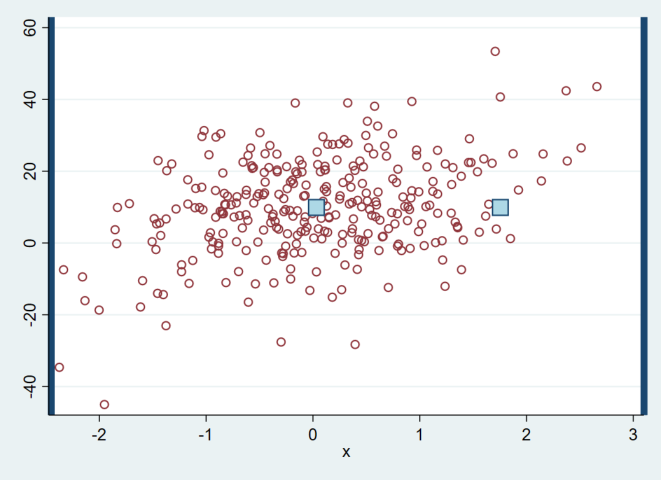
Selecting The Bandwidth

- A very large bandwidth will give you a biased estimate of the mean function with a small variance
- A very small bandwidth will give you an estimate with small bias and large variance

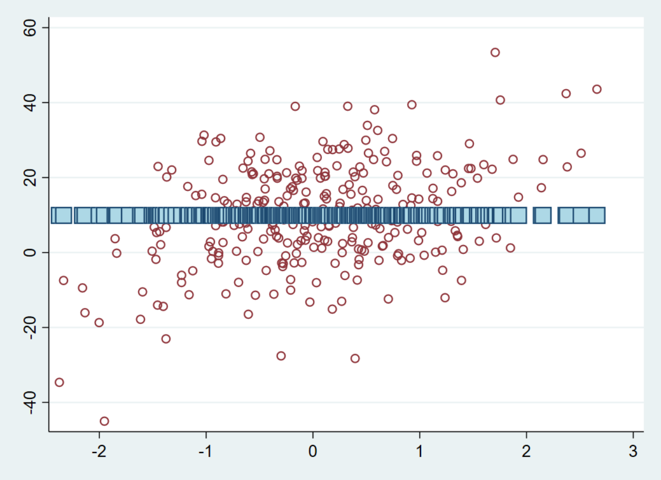
A Large Bandwidth At One Point



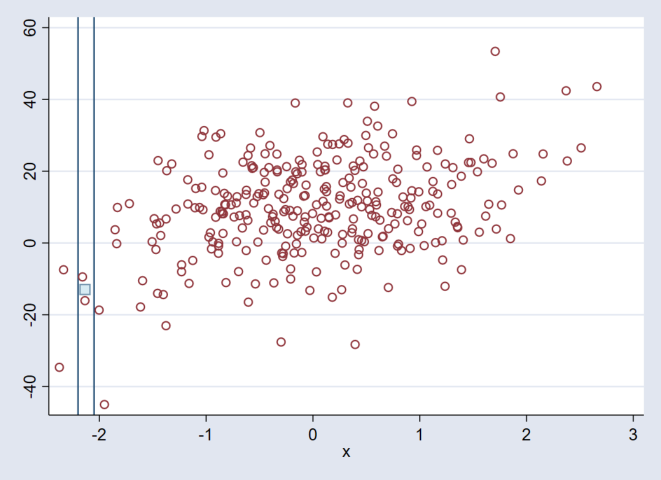
A Large Bandwidth At Two Points



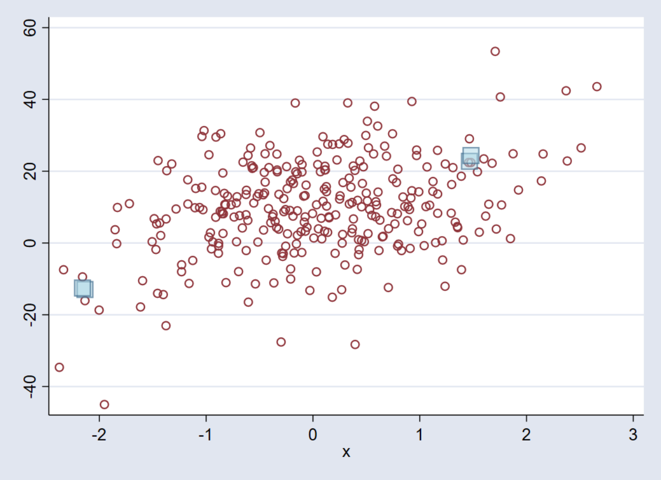
No Variance but Huge Bias



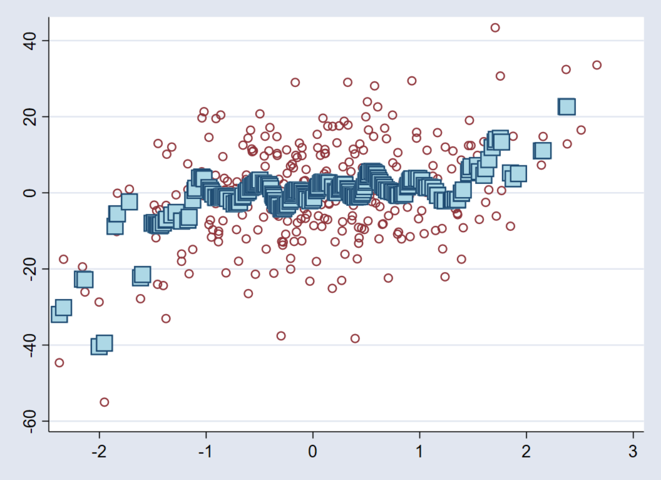
A Very Small Bandwidth at a Point



A Very Small Bandwidth at 4 Points



Small Bias Large Variance



Estimation

- Choose bandwidth optimally. Minimize bias–variance trade–off
  - Cross-validation (default)
  - Improved AIC (IMAIC)
- Compute a mean for every point in data (local-constant)
- Compute a regression for every point in data (local linear)
  - Computes constant (mean) and slope (effects)
  - Mean function and derivatives and effects of mean function
  - There is a bandwidth for the mean computation and another for the effects.
- Local-linear regression is the default

Estimation

- Choose bandwidth optimally. Minimize bias–variance trade–off
  - Cross-validation (default)
  - Improved AIC (IMAIC)
- Compute a mean for every point in data (local-constant)
- Compute a regression for every point in data (local linear)
  - Computes constant (mean) and slope (effects)
  - Mean function and derivatives and effects of mean function
  - There is a bandwidth for the mean computation and another for the effects.
- Local-linear regression is the default

Estimation

- Choose bandwidth optimally. Minimize bias–variance trade–off
  - Cross-validation (default)
  - Improved AIC (IMAIC)
- Compute a mean for every point in data (local-constant)
- Compute a regression for every point in data (local linear)
  - Computes constant (mean) and slope (effects)
  - Mean function and derivatives and effects of mean function
  - There is a bandwidth for the mean computation and another for the effects.
- Local-linear regression is the default

Estimation

- Choose bandwidth optimally. Minimize bias–variance trade–off
  - Cross-validation (default)
  - Improved AIC (IMAIC)
- Compute a mean for every point in data (local-constant)
- Compute a regression for every point in data (local linear)
  - Computes constant (mean) and slope (effects)
  - Mean function and derivatives and effects of mean function
  - There is a bandwidth for the mean computation and another for the effects.
- Local-linear regression is the default

Estimation

- Choose bandwidth optimally. Minimize bias–variance trade–off
  - Cross-validation (default)
  - Improved AIC (IMAIC)
- Compute a mean for every point in data (local-constant)
- Compute a regression for every point in data (local linear)
  - Computes constant (mean) and slope (effects)
  - Mean function and derivatives and effects of mean function
  - There is a bandwidth for the mean computation and another for the effects.
- Local-linear regression is the default

Estimation

- Choose bandwidth optimally. Minimize bias–variance trade–off
  - Cross-validation (default)
  - Improved AIC (IMAIC)
- Compute a mean for every point in data (local-constant)
- Compute a regression for every point in data (local linear)
  - Computes constant (mean) and slope (effects)
  - Mean function and derivatives and effects of mean function
  - There is a bandwidth for the mean computation and another for the effects.
- Local-linear regression is the default

Nonparametric Estimation With `npregress`

The Data

- `lbweight`: 1 if low birthweight baby
- `msmoke`: cigarettes smoked during pregnancy (3 categories)
- `mage`: mother's age
- `medu`: mother's educational attainment
- `alcohol`: 1 if alcohol is consumed during pregnancy

`npregress` Syntax

```
. npregress kernel lbweight mage medu i.msmoke i.alcohol
```

- `kernel` refers to the kind of nonparametric estimation
- By default Stata assumes variables in my model are continuous
- `i.msmoke` States the variable is categorical
- Interactions between continuous variables and between continuous and discrete variables are implicit

`npregress` Syntax

```
. npregress kernel lbweight mage medu i.msmoke i.alcohol
```

- `kernel` refers to the kind of nonparametric estimation
- By default Stata assumes variables in my model are continuous
- `i.msmoke` States the variable is categorical
- Interactions between continuous variables and between continuous and discrete variables are implicit

`npregress` Syntax

```
. npregress kernel lbweight mage medu i.msmoke i.alcohol
```

- `kernel` refers to the kind of nonparametric estimation
- By default Stata assumes variables in my model are continuous
- `i.msmoke` States the variable is categorical
- Interactions between continuous variables and between continuous and discrete variables are implicit

`npregress` Syntax

```
. npregress kernel lbweight mage medu i.msmoke i.alcohol
```

- `kernel` refers to the kind of nonparametric estimation
- By default Stata assumes variables in my model are continuous
- `i.msmoke` States the variable is categorical
- Interactions between continuous variables and between continuous and discrete variables are implicit

`npregress` Syntax

```
. npregress kernel lbweight mage medu i.msmoke i.alcohol
```

- `kernel` refers to the kind of nonparametric estimation
- By default Stata assumes variables in my model are continuous
- `i.msmoke` States the variable is categorical
- Interactions between continuous variables and between continuous and discrete variables are implicit

npregress Bandwidth

```
 . npregress kernel lbweight mage medu i.msmoke i.alcohol
Computing mean function

Minimizing cross-validation function:

Iteration 0:   Cross-validation criterion = -1.7960703
Iteration 1:   Cross-validation criterion = -1.8051048
Iteration 2:   Cross-validation criterion = -1.8051048
Iteration 3:   Cross-validation criterion = -1.8097678
Iteration 4:   Cross-validation criterion = -1.8161976
Iteration 5:   Cross-validation criterion = -1.8295231
Iteration 6:   Cross-validation criterion = -1.8295231
Iteration 7:   Cross-validation criterion = -1.8327629
Iteration 8:   Cross-validation criterion = -1.8327629
Iteration 9:   Cross-validation criterion = -1.8344806
Iteration 10:  Cross-validation criterion = -1.8348909
Iteration 11:  Cross-validation criterion = -1.8348909
Iteration 12:  Cross-validation criterion = -1.8348909

Computing optimal derivative bandwidth

Iteration 0:   Cross-validation criterion =  1.0020523
Iteration 1:   Cross-validation criterion =    .997563
Iteration 2:   Cross-validation criterion =  .99756116
```

`npregress` Output

```
 Bandwidth
─────────────┬──────────────────────
             │      Mean     Effect
─────────────┼──────────────────────
        mage │  3.149233   36.95622
        medu │  1.092557   12.82115
      msmoke │  .4397903   .4397903
     alcohol │  .0369884   .0369884
─────────────┴──────────────────────

Local-linear regression                    Number of obs      =      1,000
Continuous kernel : epanechnikov           E(Kernel obs)      =      1,000
Discrete kernel   : liracine               R-squared          =     0.4215
Bandwidth         : cross validation
─────────────┬──────────────────────────────────────────────────────────────
    lbweight │   Estimate
─────────────┼──────────────────────────────────────────────────────────────
Mean         │
    lbweight │   .0964155
─────────────┼──────────────────────────────────────────────────────────────
Effect       │
        mage │   -.002998
        medu │   -.023344
             │
      msmoke │
  (1-5 vs 0) │   .0969135
   (6+ vs 0) │   .2136147
             │
     alcohol │
 (yes vs no) │   .2147543
─────────────┴──────────────────────────────────────────────────────────────
Note: Effect estimates are averages of derivatives for continuous covariates
      and averages of contrasts for factor covariates.
Note: You may compute standard errors using vce(bootstrap) or reps().
```

npregress Standard Errors

- https://www.stata.com/manuals/rnpregress.pdf
- `npregress kernel y x i.a, vce(bootstrap, reps(1000) seed(111))`
- `npregress kernel y x i.a, reps(1000) seed(111)`

`npregress` Standard Errors

- https://www.stata.com/manuals/rnpregress.pdf
- `npregress kernel y x i.a, vce(bootstrap, reps(1000) seed(111))`
- `npregress kernel y x i.a, reps(1000) seed(111)`

`npregress` Standard Errors

- https://www.stata.com/manuals/rnpregress.pdf
- `npregress kernel y x i.a, vce(bootstrap, reps(1000) seed(111))`
- `npregress kernel y x i.a, reps(1000) seed(111)`

npregress Confidence Intervals

```
. npregress kernel lbweight mage medu i.msmoke i.alcohol, reps(1000) seed(111)
(running npregress on estimation sample)
Bootstrap replications (1000)
(output omitted)
Bandwidth
------------------------------------
             |      Mean     Effect
-------------+----------------------
        mage |  3.149233   36.95622
        medu |  1.092557   12.82115
      msmoke |  .4397903   .4397903
     alcohol |  .0369884   .0369884
------------------------------------

Local-linear regression                    Number of obs      =      1,000
Continuous kernel : epanechnikov           E(Kernel obs)      =      1,000
Discrete kernel   : liracine               R-squared          =     0.4215
Bandwidth         : cross validation
------------------------------------------------------------------------------
             |   Observed   Bootstrap                          Percentile
    lbweight |   Estimate   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Mean         |
    lbweight |   .0964155   .0092934    10.37   0.000     .0784985    .1146061
-------------+----------------------------------------------------------------
Effect       |
        mage |   -.002998   .0012575    -2.38   0.017    -.0055092   -.0006704
        medu |   -.023344   .0033661    -6.94   0.000    -.0298985   -.0167461
             |
      msmoke |
  (1-5 vs 0) |   .0969135   .0110657     8.76   0.000     .0758351    .1183473
   (6+ vs 0) |   .2136147   .0243908     8.76   0.000     .1621227    .2603727
             |
     alcohol |
 (yes vs no) |   .2147543   .0219014     9.81   0.000     .1712824    .2571516
------------------------------------------------------------------------------
Note: Effect estimates are averages of derivatives for continuous covariates
      and averages of contrasts for factor covariates.
```

Inference

- What is the population–average probability of low birthweight?
- Average of the mean function (conditional probability)

- What is the population–average probability of low birthweight?
- Average of the mean function (conditional probability)

- margins

```
. margins
Predictive margins                              Number of obs     =      1,000
Expression   : mean function, predict()
```

	Margin
_cons	.0964155

Note: You may compute standard errors using **vce(bootstrap)** or **reps()**.

- margins

```
. margins
Predictive margins                              Number of obs     =      1,000
Expression   : mean function, predict()
```

	Margin
_cons	.0964155

Note: You may compute standard errors using **vce(bootstrap)** or **reps()**.

```
. margins, reps(1000) seed(111)
(running margins on estimation sample)
Bootstrap replications (1000)
(output omitted)
Predictive margins                              Number of obs     =      1,000
                                                Replications      =      1,000

Expression   : mean function, predict()

             |   Observed   Bootstrap                         Percentile
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .0964155   .0092934    10.37   0.000     .0784985    .1146061
```

- What is the population–average probability of low birthweight evaluated at different values of `msmoke`
- What is the population–average probability of low birthweight for different levels of the treatment `msmoke`
- Counterfactual

- What is the population–average probability of low birthweight evaluated at different values of `msmoke`
- What is the population–average probability of low birthweight for different levels of the treatment `msmoke`
- Counterfactual

- What is the population–average probability of low birthweight evaluated at different values of `msmoke`
- What is the population–average probability of low birthweight for different levels of the treatment `msmoke`
- Counterfactual

Probabilities at Different Values of `msmoke`

```
. margins msmoke, reps(1000) seed(111)
(running margins on estimation sample)
Bootstrap replications (1000)
(output omitted)
Predictive margins                              Number of obs     =      1,000
                                                Replications      =      1,000

Expression   : mean function, predict()
```

	Observed Margin	Bootstrap Std. Err.	z	P>|z|	Percentile [95% Conf.	Interval]
msmoke						
0	.0220855	.0074394	2.97	0.003	.0081794	.0377342
1-5	.1189991	.0110734	10.75	0.000	.0986169	.140994
6+	.2357003	.0224171	10.51	0.000	.1913104	.276417

- What is the average treatment effect of smoking
- The population–average of the difference of mean function estimates at each level with respect to the base level

- What is the average treatment effect of smoking
- The population–average of the difference of mean function estimates at each level with respect to the base level

Treatment Effects

```
. margins r.msmoke, contrast(nowald) reps(1000) seed(111)
(running margins on estimation sample)
Bootstrap replications (1000)
(output omitted)
Contrasts of predictive margins
                                         Number of obs     =      1,000
                                         Replications      =      1,000

Expression   : mean function, predict()
```

	Observed Contrast	Bootstrap Std. Err.	Percentile [95% Conf.	Interval]
msmoke				
(1-5 vs 0)	.0969135	.0110657	.0758351	.1183473
(6+ vs 0)	.2136147	.0243908	.1621227	.2603727

```
Expression   : mean function, predict()
```

	Observed Margin	Bootstrap Std. Err.	z	P>|z|	Percentile [95% Conf.	Interval]
msmoke						
0	.0220855	.0074394	2.97	0.003	.0081794	.0377342
1-5	.1189991	.0110734	10.75	0.000	.0986169	.140994
6+	.2357003	.0224171	10.51	0.000	.1913104	.276417

Treatment Effects

```
. margins r.msmoke, contrast(nowald) reps(1000) seed(111)
(running margins on estimation sample)
Bootstrap replications (1000)
(output omitted)
Contrasts of predictive margins
                                         Number of obs     =      1,000
                                         Replications      =      1,000
Expression    : mean function, predict()
```

	Observed Contrast	Bootstrap Std. Err.	Percentile [95% Conf.	Interval]
msmoke				
(1–5 vs 0)	.0969135	.0110657	.0758351	.1183473
(6+ vs 0)	.2136147	.0243908	.1621227	.2603727

```
Expression    : mean function, predict()
```

	Observed Margin	Bootstrap Std. Err.	z	P>|z|	Percentile [95% Conf.	Interval]
msmoke						
0	.0220855	.0074394	2.97	0.003	.0081794	.0377342
1–5	.1189991	.0110734	10.75	0.000	.0986169	.140994
6+	.2357003	.0224171	10.51	0.000	.1913104	.276417

- What would the effects be if I compared them to adjacent levels instead of the base level.
- Incremental effect

- What would the effects be if I compared them to adjacent levels instead of the base level.
- Incremental effect

```
. margins ar.msmoke, contrast(nowald) reps(1000) seed(111)
(running margins on estimation sample)
Bootstrap replications (1000)
(output omitted)
Contrasts of predictive margins
                                               Number of obs     =      1,000
                                               Replications      =      1,000

Expression   : mean function, predict()
```

	Observed Contrast	Bootstrap Std. Err.	Percentile [95% Conf.	Interval]
msmoke				
(1–5 vs 0)	.0969135	.0110657	.0758351	.1183473
(6+ vs 1–5)	.1167012	.0169506	.0815164	.1484021

- What would the population–averaged probability of low birthweight be if all mothers in the population increased their education by 4 years.

Counterfactual Education Levels

```
. margins, at(medu=generate(medu)) at(medu=generate(medu+4))      ///
>         reps(1000) seed(111)
(running margins on estimation sample)
Bootstrap replications (1000)
(output omitted)
Predictive margins                              Number of obs     =      1,000
                                                Replications      =      1,000

Expression   : mean function, predict()
1._at        : medu            = medu
2._at        : medu            = medu+4

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Percentile
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .0964155   .0092934    10.37   0.000     .0784985    .1146061
          2  |   .0321729   .0114623     2.81   0.005     .0140669    .0575301
------------------------------------------------------------------------------
```

Contrasting

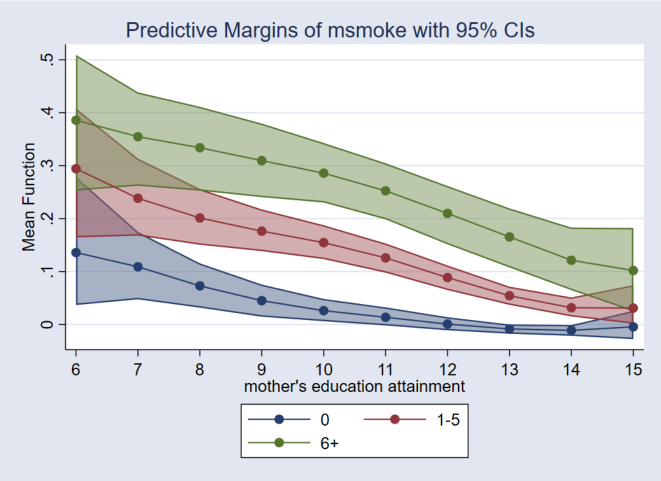
```
. margins, at(medu=generate(medu)) at(medu=generate(medu+4)) ///
>                   contrast(at(r)) reps(1000) seed(111)
(running margins on estimation sample)
Bootstrap replications (1000)
(output omitted)
Contrasts of predictive margins
                                                Number of obs     =      1,000
                                                Replications      =      1,000

Expression    : mean function, predict()
1._at         : medu            = medu
2._at         : medu            = medu+4

-------------------------------------------------------------
             |   Observed   Bootstrap         Percentile
             |   Contrast   Std. Err.     [95% Conf. Interval]
-------------+-----------------------------------------------
         _at |
   (2 vs 1)  |  -.0642426    .0121771     -.0856404   -.0376858
-------------------------------------------------------------
```

- What is the population–average of low birthweight for each counterfactual level of `msmoke` if everyone in the population received a different level of education
- population–average of low birthweight for fixed level of `msmoke` and `medu`

- What is the population–average of low birthweight for each counterfactual level of `msmoke` if everyone in the population received a different level of education
- population–average of low birthweight for fixed level of `msmoke` and `medu`

Counterfactual Education Level

- `margins msmoke, at(medu=(6(1)15))`
- `marginsplot, recastci(rarea) ciopts(fcolor(%50))`



Predictive Margins of msmoke with 95% CIs
Mean Function
0
.1
.2
.3
.4
.5
6
7
8
9
10
11
12
13
14
15
mother's education attainment
0
1-5
6+

Interpretation of Confidence Bands and `contrast`

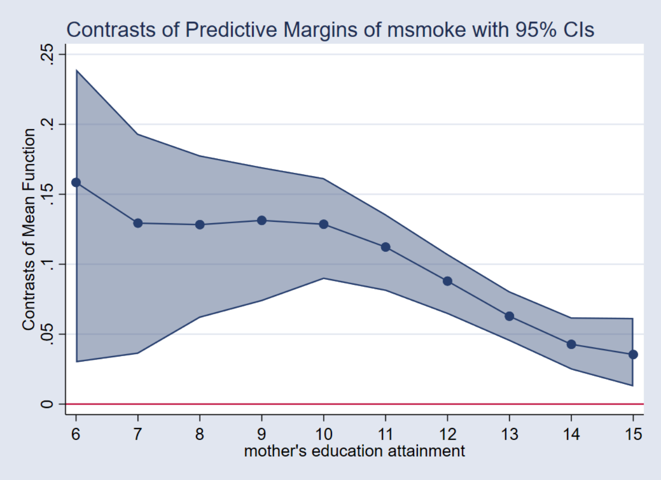
- Overlapping confidence intervals do not mean there is no difference between (0 and 1-5)
- Confidence intervals are for the point estimates not the difference
- The way to test this is testing for the differences, the effects.

Interpretation of Confidence Bands and `contrast`

- Overlapping confidence intervals do not mean there is no difference between (0 and 1-5)
- Confidence intervals are for the point estimates not the difference
- The way to test this is testing for the differences, the effects.

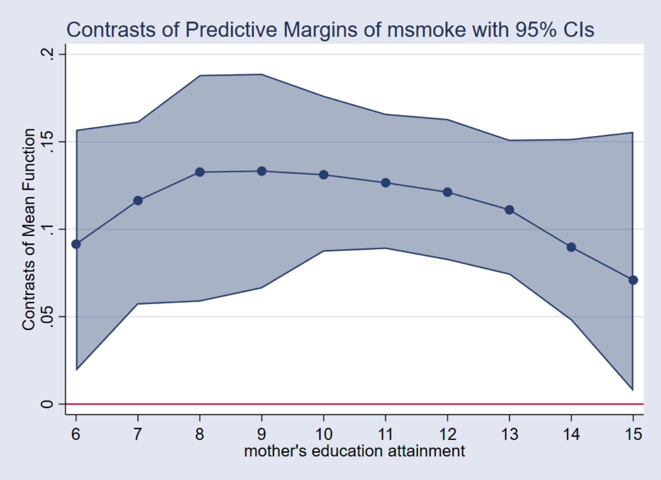
0 vs 1-5

```
margins r(0 1).msmoke , at(medu=(6(1)15))
```



1-5 vs 6+

```
margins r(1 2).msmoke , at(medu=(6(1)15))
```



Continue Exploring Function

- `margins msmoke, at(medu=(6(1)15))`
- `margins msmoke, at(medu=(6(1)15) alcohol=0) at(medu=(6(1)15) alcohol=1)`
- `margins msmoke, at(medu=(6(1)15) alcohol=0 mage=(16(1)34)) at(medu=(6(1)15) alcohol=1 mage=(16(1)34))`

Continue Exploring Function

- `margins msmoke, at(medu=(6(1)15))`
- `margins msmoke, at(medu=(6(1)15) alcohol=0) at(medu=(6(1)15) alcohol=1)`
- `margins msmoke, at(medu=(6(1)15) alcohol=0 mage=(16(1)34)) at(medu=(6(1)15) alcohol=1 mage=(16(1)34))`

Continue Exploring Function

- `margins msmoke, at(medu=(6(1)15))`
- `margins msmoke, at(medu=(6(1)15) alcohol=0) at(medu=(6(1)15) alcohol=1)`
- `margins msmoke, at(medu=(6(1)15) alcohol=0 mage=(16(1)34)) at(medu=(6(1)15) alcohol=1 mage=(16(1)34))`

Parting Words

How Good Is Your Approximation?

- If you have a lot of data per covariate
- Data per region is important (identification)
- Time consuming
- Benefits

How Good Is Your Approximation?

- If you have a lot of data per covariate
- Data per region is important (identification)
- Time consuming
- Benefits

How Good Is Your Approximation?

- If you have a lot of data per covariate
- Data per region is important (identification)
- Time consuming
- Benefits

How Good Is Your Approximation?

- If you have a lot of data per covariate
- Data per region is important (identification)
- Time consuming
- Benefits

`margins` After Correctly Specified Functional Form

```
probit lbweight c.mage#c.mage c.medu##c.mage i.msmoke##i.alcohol
```

```
. margins, dydx(*)
Average marginal effects                        Number of obs     =      1,000
Model VCE    : OIM
Expression   : Pr(lbweight), predict()
dy/dx w.r.t. : mage medu 1.msmoke 2.msmoke 1.alcohol
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
mage	-.0030745	.0012678	-2.43	0.015	-.0055593	-.0005898
medu	-.025059	.0035449	-7.07	0.000	-.0320069	-.0181112
msmoke						
1-5	.088816	.0146566	6.06	0.000	.0600897	.1175424
6+	.2155829	.0242201	8.90	0.000	.1681125	.2630533
alcohol						
yes	.2103893	.0198121	10.62	0.000	.1715584	.2492203

Note: dy/dx for factor levels is the discrete change from the base level.

`margins` After Correctly Specified Functional Form

```
probit lbweight c.mage#c.mage c.medu##c.mage i.msmoke##i.alcohol
```

```
. margins, dydx(*)
Average marginal effects                      Number of obs     =      1,000
Model VCE     : OIM
Expression    : Pr(lbweight), predict()
dy/dx w.r.t.  : mage medu 1.msmoke 2.msmoke 1.alcohol
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
mage	-.0030745	.0012678	-2.43	0.015	-.0055593	-.0005898
medu	-.025059	.0035449	-7.07	0.000	-.0320069	-.0181112
msmoke						
1-5	.088816	.0146566	6.06	0.000	.0600897	.1175424
6+	.2155829	.0242201	8.90	0.000	.1681125	.2630533
alcohol						
yes	.2103893	.0198121	10.62	0.000	.1715584	.2492203

Note: dy/dx for factor levels is the discrete change from the base level.

npregress

```
. npregress
```

lbweight	Observed Estimate	Bootstrap Std. Err.	z	P>|z|	Percentile [95% Conf.	Interval]
Mean						
lbweight	.0964155	.0101926	9.46	0.000	.0754131	.1146312
Effect						
mage	-.002998	.001135	-2.64	0.008	-.0047728	-.0007828
medu	-.023344	.003595	-6.49	0.000	-.030541	-.0164684
msmoke						
(1-5 vs 0)	.0969135	.0125798	7.70	0.000	.0660747	.1185829
(6+ vs 0)	.2136147	.0255175	8.37	0.000	.1568306	.2565002
alcohol						
(yes vs no)	.2147543	.021038	10.21	0.000	.1752073	.2571517

Note: Effect estimates are averages of derivatives for continuous covariates and averages of contrasts for factor covariates.

Conclusion

- Obtain understanding and intuition about nonparametric regression
- Think about nonparametric regression as a tool to make inferences about the mean function
  - `npregress` as a tool to obtain the mean function, effects, and average derivatives
  - `margins` as a tool to explore the mean function and ask interesting questions
- Unique to Stata

Conclusion

- Obtain understanding and intuition about nonparametric regression
- Think about nonparametric regression as a tool to make inferences about the mean function
    - `npregress` as a tool to obtain the mean function, effects, and average derivatives
    - `margins` as a tool to explore the mean function and ask interesting questions
- Unique to Stata

Conclusion

- Obtain understanding and intuition about nonparametric regression
- Think about nonparametric regression as a tool to make inferences about the mean function
  - `npregress` as a tool to obtain the mean function, effects, and average derivatives
  - `margins` as a tool to explore the mean function and ask interesting questions
- Unique to Stata

Conclusion

- Obtain understanding and intuition about nonparametric regression
- Think about nonparametric regression as a tool to make inferences about the mean function
  - `npregress` as a tool to obtain the mean function, effects, and average derivatives
  - `margins` as a tool to explore the mean function and ask interesting questions
- Unique to Stata