Some Notes on Functional Form

Brad Jones

1Department of Political Science
University of California, Davis

March 2, 2010
Today: Lots of Different Things
Today

- “Functional form”
Today

- “Functional form”
- Issues w/functional form.
Today

- “Functional form”
- Issues w/functional form.
- Visualizing and transforming data.
Today

- “Functional form”
- Issues w/functional form.
- Visualizing and transforming data.
- Nonparametric regression
Visualizing Data

- Not wise to barge into analysis without knowing the nature of one’s data.
Not wise to barge into analysis without knowing the nature of one’s data.

Plots and density estimates are useful as pre-analysis.
Visualizing Data

- Not wise to barge into analysis without knowing the nature of one’s data.
- Plots and density estimates are useful as pre-analysis.
- The very interpretation of a model will be sensitive to placement of data points.
Visualizing Data

- Not wise to barge into analysis without knowing the nature of one's data.
- Plots and density estimates are useful as pre-analysis.
- The very interpretation of a model will be sensitive to placement of data points.
- Is $X$ truly continuous? Are there discontinuities in the data?
Visualizing Data

- Not wise to barge into analysis without knowing the nature of one’s data.
- Plots and density estimates are useful as pre-analysis.
- The very interpretation of a model will be sensitive to placement of data points.
- Is $X$ truly continuous? Are there discontinuities in the data?
- NOMINATE scores from the 109th Congress:
Visualizing Data

Pool-Rosenthal Scores [1st Dim]
Visualizing Data

- Think about “what a unit increase in $X$ means in this context.”
Visualizing Data

- Think about “what a unit increase in $X$ means in this context.”
- Start with some univariate quantities.
Think about “what a unit increase in $X$ means in this context.”

Start with some univariate quantities.

Density estimation using nonparametric smoothers.
Think about “what a unit increase in $X$ means in this context.”

Start with some univariate quantities.

Density estimation using nonparametric smoothers.

Useful to “smooth” out histograms sometimes.
Visualizing Data

- Histograms “chunk” up the data into distinct “bins.”
Visualizing Data

- Histograms “chunk” up the data into distinct “bins.”
- $n$ bins usually defined by software defaults.
Visualizing Data

- Histograms “chunk” up the data into distinct “bins.”
- $n$ bins usually defined by software defaults.
- We may be interested in considering a smooth histogram—one where the bin edges are smoothed and connected.

$\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^{n} K(x - X_i h)$

$K$ is the kernel function and $h$ is a bandwidth parameter. Think of $h$ as analogous to the width of a bin in a histogram.
Visualizing Data

- Histograms “chunk” up the data into distinct “bins.”
- $n$ bins usually defined by software defaults.
- We may be interested in considering a smooth histogram—one where the bin edges are smoothed and connected.
- Consider a _kernel density estimator_ (the basis for this estimator will be considered later):

\[
\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right)
\]

\(K\) is the kernel function and \(h\) is a bandwidth parameter. Think of \(h\) as analogous to the width of a bin in a histogram.
Visualizing Data

- Histograms “chunk” up the data into distinct “bins.”
- $n$ bins usually defined by software defaults.
- We may be interested in considering a smooth histogram—one where the bin edges are smoothed and connected.
- Consider a *kernel density estimator* (the basis for this estimator will be considered later):

$$
\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right)
$$

- $K$ is the kernel function and $h$ is a *bandwidth parameter*. 
Visualizing Data

- Histograms “chunk” up the data into distinct “bins.”
- $n$ bins usually defined by software defaults.
- We may be interested in considering a smooth histogram—one where the bin edges are smoothed and connected.
- Consider a *kernel density estimator* (the basis for this estimator will be considered later):

$$\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right)$$

- $K$ is the kernel function and $h$ is a *bandwidth parameter*.
- Think of $h$ as analogous to the width of a bin in a histogram.
Visualizing Data

- The function $K(z)$ may be one of a variety of density functions (including the normal).
Visualizing Data

- The function $K(z)$ may be one of a variety of density functions (including the normal).
- $z$ is given by $(x - X_i) / h$ where $X_i$ is a “focal” in a continuously moving window.
Visualizing Data

- The function $K(z)$ may be one of a variety of density functions (including the normal).
- $z$ is given by $(\frac{x-X_i}{h})$ where $X_i$ is a “focal” in a continuously moving window.
- Note the moving parts: if $h$ is large, the density estimator will produce smoother estimates; if $h$ is small, the estimator is more closely adapted to the observed data.
The function $K(z)$ may be one of a variety of density functions (including the normal).

$z$ is given by $(\frac{x-X_i}{h})$ where $X_i$ is a “focal” in a continuously moving window.

Note the moving parts: if $h$ is large, the density estimator will produce smoother estimates; if $h$ is small, the estimator is more closely adapted to the observed data.

Illustration in R
Visualizing Data

Denisty Estimate

Denisty Estimate (Default)

Denisty Estimate [bw=.1]

Denisty Estimate [bw=.5]

Denisty Estimate [bw=2]

Denisty Estimate [bw=4]
Visualizing Data

- Note how the smoother varies with selection of the bandwidth.
Visualizing Data

- Note how the smoother varies with selection of the bandwidth.
- Larger data sets can accommodate smaller $h$
Visualizing Data

- Note how the smoother varies with selection of the bandwidth.
- Larger data sets can accommodate smaller $h$
- We will return to nonparametric estimators a bit later . . .
Visualizing Data

- Note how the smoother varies with selection of the bandwidth.
- Larger data sets can accommodate smaller $h$
- We will return to nonparametric estimators a bit later . . .
- But this should give you a sense of things.
Visualizing Data

- Quantile comparison plots are useful sometimes.
Visualizing Data

- Quantile comparison plots are useful sometimes.
- Are my data “normally distributed?”
Visualizing Data

- Quantile comparison plots are useful sometimes.
- Are my data “normally distributed?”
- Derivation of a Q-Q Plot (quantile plot)
Visualizing Data

- Quantile comparison plots are useful sometimes.
- Are my data “normally distributed?”
- Derivation of a Q-Q Plot (quantile plot)
- Rank order the data from smallest to largest, $X(1), X(2), \ldots, X(n)$
Visualizing Data

- Quantile comparison plots are useful sometimes.
- Are my data “normally distributed?”
- Derivation of a Q-Q Plot (quantile plot)
- Rank order the data from smallest to largest, $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$
- The $X_{(k)}$ are the order statistics.
Visualizing Data

- Quantile comparison plots are useful sometimes.
- Are my data “normally distributed?”
- Derivation of a Q-Q Plot (quantile plot)
- Rank order the data from smallest to largest, $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$
- The $X_{(k)}$ are the order statistics.
- Identify the quantiles which are defined as points taken at regular intervals from the cumulative distribution function of a random variable.
Theoretically, we could use the order statistics directly to define proportions to construct the quantiles: $k/n$
Theoretically, we could use the order statistics directly to define proportions to construct the quantiles: \( k/n \)

Problem is at \( k = n \), the quantile is the 100th percentile, which may or may not be the upper limit in the population.
Visualizing Data

- Theoretically, we could use the order statistics directly to define proportions to construct the quantiles: $k/n$
- Problem is at $k = n$, the quantile is the 100th percentile, which may or may not be the upper limit in the population.
- Conventionally $\frac{k-.5}{n}$ is used to define the proportions.
Theoretically, we could use the order statistics directly to define proportions to construct the quantiles: $k/n$

Problem is at $k = n$, the quantile is the 100th percentile, which may or may not be the upper limit in the population.

Conventionally $\frac{k - .5}{n}$ is used to define the proportions.

The quantile function is the inverse of the CDF:

$$z_i = P^{-1} \left( \frac{k - .5}{n} \right)$$
Theoretically, we could use the order statistics directly to define proportions to construct the quantiles: \( k/n \)

Problem is at \( k = n \), the quantile is the 100th percentile, which may or may not be the upper limit in the population.

Conventionally \( \frac{k-.5}{n} \) is used to define the proportions.

The quantile function is the inverse of the CDF:

\[
z_i = P^{-1} \left( \frac{k-.5}{n} \right)
\]

\( z_i \) now gives the cumulative probabilities.
Visualizing Data

- Plot $z_k$ against $X$. 

The idea: if $X$ is sampled from the proposed distribution that produced the quantiles, then the plot should be linear with slope 1 and $y$-intercept 0. (Why?)

A common comparison distribution is the normal (though in other instances, other distributions might be preferable).

In passing, if the normal CDF is used, the $z$ are sometimes called "rankits."
Visualizing Data

- Plot $z_k$ against $X$.
- The idea: if $X$ is sampled from the proposed distribution that produced the quantiles, then the plot should be linear with slope 1 and $y$-intercept 0. (Why?)
Visualizing Data

- Plot $z_k$ against $X$.
- The idea: if $X$ is sampled from the proposed distribution that produced the quantiles, then the plot should be linear with slope 1 and $y$-intercept 0. (Why?)
- A common comparison distribution is the normal (though in other instances, other distributions might be preferable).
Visualizing Data

- Plot $z_k$ against $X$.
- The idea: if $X$ is sampled from the proposed distribution that produced the quantiles, then the plot should be linear with slope 1 and $y$-intercept 0. (Why?)
- A common comparison distribution is the normal (though in other instances, other distributions might be preferable).
- In passing, if the normal CDF is used, the $z$ are sometimes called “rankits.”
Visualizing Data

- Plot $z_k$ against $X$.
- The idea: if $X$ is sampled from the proposed distribution that produced the quantiles, then the plot should be linear with slope 1 and $y$-intercept 0. (Why?)
- A common comparison distribution is the normal (though in other instances, other distributions might be preferable).
- In passing, if the normal CDF is used, the $z$ are sometimes called “rankits.”
- Illustration: Nominate and Income
Visualizing Data

QQ Plot of Nominate Scores
Visualizing Data

QQ Plot for Income

income

0 5000 10000 15000 20000 25000

-2 -1 0 1 2
Visualizing Data

- Basic utility of the plot is to assess deviations from normality.
Visualizing Data

- Basic utility of the plot is to assess deviations from normality.
- Of course what you choose to do about it is the next question!
Visualizing Data

- Basic utility of the plot is to assess deviations from normality.
- Of course what you choose to do about it is the next question!
- Continue with univariate plots.
Visualizing Data

- Basic utility of the plot is to assess deviations from normality.
- Of course what you choose to do about it is the next question!
- Continue with univariate plots.
- I think basic plots are essential first steps in analysis. This is particularly true wrt functional form of covariates.
Basic utility of the plot is to assess deviations from normality.
Of course what you choose to do about it is the next question!
Continue with univariate plots.
I think basic plots are essential first steps in analysis. This is particularly true wrt functional form of covariates.
Box plots
Visualizing Data

- The box gives information about the quartiles.
Visualizing Data

- The box gives information about the quartiles.
- The bottom line on the box denotes the $25^{th}$ percentile (the $1^{st}$ quartile); the middle line denotes the median; and the upper line denotes the $75^{th}$ percentile (the $3^{rd}$ quartile).
Visualizing Data

- The box gives information about the quartiles.
- The bottom line on the box denotes the 25\textsuperscript{th} percentile (the 1\textsuperscript{st} quartile); the middle line denotes the median; and the upper line denotes the 75\textsuperscript{th} percentile (the 3\textsuperscript{rd} quartile).
- The whiskers extend to the upper and lower adjacent values (that is, they approximately represent the upper and lower values in the distribution).
Visualizing Data

- The box gives information about the quartiles.
- The bottom line on the box denotes the $25^{th}$ percentile (the $1^{st}$ quartile); the middle line denotes the median; and the upper line denotes the $75^{th}$ percentile (the $3^{rd}$ quartile).
- The whiskers extend to the upper and lower adjacent values (that is, they approximately represent the upper and lower values in the distribution).
- Useful to visual distributions, outlying cases, coding errors.
Visualizing Data

- The box gives information about the quartiles.
- The bottom line on the box denotes the 25\textsuperscript{th} percentile (the 1\textsuperscript{st} quartile); the middle line denotes the median; and the upper line denotes the 75\textsuperscript{th} percentile (the 3\textsuperscript{rd} quartile).
- The whiskers extend to the upper and lower adjacent values (that is, they approximately represent the upper and lower values in the distribution).
- Useful to visual distributions, outlying cases, coding errors.
- Two examples:
Visualizing Data

Box Plot of Percent Hispanic

Percent Hispanic in C.D.

0 20 40 60 80
Visualizing Data
Categorical data and plotting is useful but sometimes uninformative.
Visualizing Data

- Categorical data and plotting is useful but sometimes uninformative.
- Coded scatterplots (a functionality in car) and jitter plots are useful.
Visualizing Data

- Categorical data and plotting is useful but sometimes uninformative.
- Coded scatterplots (a functionality in `car`) and jitter plots are useful.
- Coded scatterplot of percent Hispanic and DW-NOMINATE scores shows relationship between population characteristics and ideology of the MC controlling for party.
Visualizing Data
Visualizing Data

- Jitter plots were conceived of by Cleveland (1994).
Visualizing Data

- Jitter plots were conceived of by Cleveland (1994).
- Basic problem with categorical data in plots?
Visualizing Data

- Jitter plots were conceived of by Cleveland (1994).
- Basic problem with categorical data in plots?
- Consider example on next slide:
Visualizing Data

Scatterplot of Cat. Var.

- Vocabulary
- Education

- X-axis: Education
- Y-axis: Vocabulary

- Data points represented in red.
In this plot, education can assume 21 values and vocabulary can assume 11 values.
Visualizing Data

- In this plot, education can assume 21 values and vocabulary can assume 11 values.
- There are about 1000 observations so clearly many points are overplotted.
In this plot, education can assume 21 values and vocabulary can assume 11 values.

There are about 1000 observations so clearly many points are overplotted.

... but of course some are not.
In this plot, education can assume 21 values and vocabulary can assume 11 values.

There are about 1000 observations so clearly many points are overplotted.

... but of course some are not.

Of the 231 plot positions, some may be heavily overplotted; some may be sparse.
Visualizing Data

- In this plot, education can assume 21 values and vocabulary can assume 11 values.
- There are about 1000 observations so clearly many points are overplotted.
- ... but of course some are not.
- Of the 231 plot positions, some may be heavily overplotted; some may be sparse.
- Cleveland suggested adding a random quantity to each coordinate to create separation in the overplotting.
In this plot, education can assume 21 values and vocabulary can assume 11 values.

There are about 1000 observations so clearly many points are overplotted.

... but of course some are not.

Of the 231 plot positions, some may be heavily overplotted; some may be sparse.

Cleveland suggested adding a random quantity to each coordinate to create separation in the overplotting.

This is known as jittering.
Visualizing Data

Scatterplot of Cat. Var.

Jitter Plot

jitter(vocabulary, factor = 2)
Visualizing Data

- Clearly many ways to plot and visualize data.
Visualizing Data

- Clearly many ways to plot and visualize data.
- Fox talks about more ways.
Visualizing Data

- Clearly many ways to plot and visualize data.
- Fox talks about more ways.
- WRT regression, what is the point of this?
Clearly many ways to plot and visualize data.
Fox talks about more ways.
WRT regression, what is the point of this?
Consideration of transformations of data and functional form.
Visualizing Data

- Clearly many ways to plot and visualize data.
- Fox talks about more ways.
- WRT regression, what is the point of this?
- Consideration of transformations of data and functional form.
- This is a major issue not only for OLS but any kind of modeling strategy.
Visualizing Data

- Clearly many ways to plot and visualize data.
- Fox talks about more ways.
- WRT regression, what is the point of this?
- Consideration of transformations of data and functional form.
- This is a major issue not only for OLS but any kind of modeling strategy.
- Entails the issue of “coding.”
Functional Form

▶ Recall interactions.
Recall interactions.

Through the use of multiplicative terms in interaction models, we can assess how slopes vary, conditional on the value of some other covariate.
Recall interactions.

Through the use of multiplicative terms in interaction models, we can assess how slopes vary, conditional on the value of some other covariate.

In this sense, we can estimate more sophisticated linear relationships among our variables.
Recall interactions.

Through the use of multiplicative terms in interaction models, we can assess how slopes vary, conditional on the value of some other covariate.

In this sense, we can estimate more sophisticated linear relationships among our variables.

Suppose we want to incorporate nonlinear relationships inside of a linear regression model?
Recall interactions.

Through the use of multiplicative terms in interaction models, we can assess how slopes vary, conditional on the value of some other covariate.

In this sense, we can estimate more sophisticated linear relationships among our variables.

Suppose we want to incorporate nonlinear relationships inside of a linear regression model?

Pedagogical motivation
To fix ideas, let us begin with a simple bivariate model of the form

$$\hat{Y} = \hat{a} + \hat{b}_1 X_1.$$  

(1)
To fix ideas, let us begin with a simple bivariate model of the form
\[ \hat{Y} = \hat{a} + \hat{b}_1 X_1. \] (1)

The interpretation of the slope coefficient is standard in that it gives us the change in \( E(Y) \) given a unit change in \( X \).
Functional Form

To fix ideas, let us begin with a simple bivariate model of the form

\[ \hat{Y} = \hat{a} + \hat{b}_1 X_1. \]  

The interpretation of the slope coefficient is standard in that it gives us the change in \( E(Y) \) given a unit change in \( X \).

However, this model is very restrictive in that this additive effect is constant over the range of \( X \).
To fix ideas, let us begin with a simple bivariate model of the form
\[
\hat{Y} = \hat{a} + \hat{b}_1 X_1. \tag{1}
\]

The interpretation of the slope coefficient is standard in that it gives us the change in $E(Y)$ given a unit change in $X$.

However, this model is very restrictive in that this additive effect is constant over the range of $X$.

Consider the simulated data on the next slide.
Functional Form

Table 1: Data for $Y$ and $X_1$.

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.46</td>
<td>.5</td>
</tr>
<tr>
<td>.47</td>
<td>1.5</td>
</tr>
<tr>
<td>.56</td>
<td>2.5</td>
</tr>
<tr>
<td>.61</td>
<td>3.5</td>
</tr>
<tr>
<td>.61</td>
<td>4.5</td>
</tr>
<tr>
<td>.67</td>
<td>5.5</td>
</tr>
<tr>
<td>.68</td>
<td>6.5</td>
</tr>
<tr>
<td>.78</td>
<td>7.5</td>
</tr>
<tr>
<td>.69</td>
<td>8.5</td>
</tr>
<tr>
<td>.74</td>
<td>9.5</td>
</tr>
<tr>
<td>.77</td>
<td>10.5</td>
</tr>
<tr>
<td>.78</td>
<td>11.5</td>
</tr>
<tr>
<td>.75</td>
<td>12.5</td>
</tr>
<tr>
<td>.8</td>
<td>13.5</td>
</tr>
<tr>
<td>.78</td>
<td>14.5</td>
</tr>
<tr>
<td>.82</td>
<td>15.5</td>
</tr>
<tr>
<td>.77</td>
<td>16.5</td>
</tr>
<tr>
<td>.8</td>
<td>17.5</td>
</tr>
<tr>
<td>.81</td>
<td>18.5</td>
</tr>
<tr>
<td>.78</td>
<td>19.5</td>
</tr>
</tbody>
</table>
Functional Form

- Using these data, we estimate the following model,

\[ \hat{Y} = 0.542 + 0.016X_1 \]

which has an $RMSE = 0.055$ and an $F = 60.05$.
Functional Form

- Using these data, we estimate the following model,

\[ \hat{Y} = 0.542 + 0.016X_1 \]

which has an $RMSE = 0.055$ and an $F = 60.05$.

- Plot of fitted values on next slide but main point is . . .
Functional Form

- Using these data, we estimate the following model,

\[
\hat{Y} = 0.542 + 0.016X_1
\]

which has an \(RMSE = 0.055\) and an \(F = 60.05\).

- Plot of fitted values on next slide but main point is . . .

- There is nothing noteworthy about this regression function.
Using these data, we estimate the following model,

\[ \hat{Y} = .542 + .016X_1 \]

which has an \( RMSE = .055 \) and an \( F = 60.05 \).

Plot of fitted values on next slide but main point is . . .

There is nothing noteworthy about this regression function.

We’ve known how to interpret this model since the first week of class.
Using these data, we estimate the following model,

\[ \hat{Y} = 0.542 + 0.016X_1 \]

which has an \( RMSE = 0.055 \) and an \( F = 60.05 \).

Plot of fitted values on next slide but main point is . . .

There is nothing noteworthy about this regression function.

We’ve known how to interpret this model since the first week of class.

BUT make sure you understand what is going on here: the slope is constant (and linear) across the range of \( X_1 \), this suggests that the change in \( E(Y) \) when \( X_1 \) increases from, say, .5 to 1.5, is exactly the same as when \( X_1 \) increases from 18.5 to 19.5.
Functional Form

Standard Model

$y = \beta_0 + \beta_1 x_1 + \epsilon$

$y$ is the dependent variable.

$x_1$ is the independent variable.

$\beta_0$ is the intercept.

$\beta_1$ is the slope.

$\epsilon$ is the error term.
Functional Form

- The $\hat{y}$ exhibit no marginality.
Functional Form

- The $\hat{y}$ exhibit no marginality.
- In the general case this may be unreasonable.
Functional Form

- The $\hat{y}$ exhibit no marginality.
- In the general case this may be unreasonable.
- The marginal change in $Y$ may decrease as $X$ continues to increase.
Functional Form

- The $\hat{y}$ exhibit no marginality.
- In the general case this may be unreasonable.
- The marginal change in $Y$ may decrease as $X$ continues to increase.
- Of course the obverse could be true: as $X$ increases to some point, the $E(Y)$ could begin to increase at an increasing rate.
Functional Form

- The $\hat{y}$ exhibit no marginality.
- In the general case this may be unreasonable.
- The marginal change in $Y$ may decrease as $X$ continues to increase.
- Of course the obverse could be true: as $X$ increases to some point, the $E(Y)$ could begin to increase at an increasing rate.
- Another feature of the plot is that of monotonicity.
Functional Form

- The $\hat{y}$ exhibit no marginality.
- In the general case this may be unreasonable.
- The marginal change in $Y$ may decrease as $X$ continues to increase.
- Of course the obverse could be true: as $X$ increases to some point, the $E(Y)$ could begin to increase at an increasing rate.
- Another feature of the plot is that of *monotonicity*.
- The modeled relationship between $X_1$ and $Y$ is monotonic.
The main point is that in some instances, the relationship between a covariate (or covariates) and the dependent variable may not possess the property of having a constant, monotonic, linear effect.
Functional Form

- The main point is that in some instances, the relationship between a covariate (or covariates) and the dependent variable may not possess the property of having a constant, monotonic, linear effect.
- Illustrations with simulated data.
Functional Form

Examples of Nonlinear Relationship between $x_1$ and $y$
Each of the relationships shown seem to suggest that the slope of $Y$ on $X_1$ may not be constant over the full range of $X_1$. 
Each of the relationships shown seem to suggest that the slope of $Y$ on $X_1$ may not be constant over the full range of $X_1$.

Suppose we fit a regression model?
Each of the relationships shown seem to suggest that the slope of $Y$ on $X_1$ may not be constant over the full range of $X_1$.

Suppose we fit a regression model?

Visualizing the data suggests the garden variety model may not be optimal for us.
## Functional Form

### Regression Models

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Marginality 1</th>
<th>Marginality 2</th>
<th>Nonmonotonicity 1</th>
<th>Nonmonotonicity 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.754 (.019)</td>
<td>.616 (.025)</td>
<td>.617 (.04)</td>
<td>1.66 (.12)</td>
</tr>
<tr>
<td>Slope</td>
<td>-.193 (.038)</td>
<td>.0007 (.0001)</td>
<td>-.0003 (.0035)</td>
<td>-.00009 (.01)</td>
</tr>
<tr>
<td>( RMSE )</td>
<td>.074</td>
<td>.076</td>
<td>.091</td>
<td>.27</td>
</tr>
<tr>
<td>( r^2 )</td>
<td>.58</td>
<td>.56</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>
Ignoring the graphical displays of data, the interpretation of these models is simple.

- For a unit increase in $X_1$, the expected value of $Y$ decreases by about -.193 units.
- This effect is constant over the full range of $X_1$, although the graphical display of the data seems to suggest that the slope decreases (i.e. tends to 0) as $X_1$ gets large.
- The extreme cases are found for the two regression models estimated for the nonmonotonic data.
- The slope coefficient tells us that a unit increase in $X_1$ is associated with no change in $E(Y)$. 

Functional Form
Functional Form

- Ignoring the graphical displays of data, the interpretation of these models is simple.
- Looking at the first column, we see that for a unit increase in $X_1$, the expected value of $Y$ decreases by about -.193 units.
Functional Form

- Ignoring the graphical displays of data, the interpretation of these models is simple.
- Looking at the first column, we see that for a unit increase in $X_1$, the expected value of $Y$ decreases by about -.193 units.
- This effect is constant over the full range of $X_1$, although the graphical display of the data seems to suggest that the slope decreases (i.e. tends to 0) as $X_1$ gets large.
Ignoring the graphical displays of data, the interpretation of these models is simple. Looking at the first column, we see that for a unit increase in $X_1$, the expected value of $Y$ decreases by about -.193 units. This effect is constant over the full range of $X_1$, although the graphical display of the data seems to suggest that the slope decreases (i.e. tends to 0) as $X_1$ gets large. The extreme cases are found for the two regression models estimated for the nonmonotonic data.
Ignoring the graphical displays of data, the interpretation of these models is simple.

Looking at the first column, we see that for a unit increase in $X_1$, the expected value of $Y$ decreases by about -0.193 units.

This effect is constant over the full range of $X_1$, although the graphical display of the data seems to suggest that the slope decreases (i.e. tends to 0) as $X_1$ gets large.

The extreme cases are found for the two regression models estimated for the nonmonotonic data.

The slope coefficient tells us that a unit increase in $X_1$ is associated with no change in $E(Y)$. 
Functional Form

Regression Functions with Nonlinear Data
This figure illustrates an obvious, but important point: a linear model will produce a straight line.
Functional Form

- This figure illustrates an obvious, but important point: a linear model will produce a straight line.
- The actual data may belie this.
Functional Form

- This figure illustrates an obvious, but important point: a linear model will produce a straight line.
- The actual data may belie this.
- Consider the regression residuals.
Functional Form

Residual vs. Fitted Values Plots
The distribution of the residuals is far from randomly distributed about 0.
The distribution of the residuals is far from randomly distributed about 0.

Residual plots like these are useful in diagnosing potential problems with the standard regression model.
Functional Form

- The distribution of the residuals is far from randomly distributed about 0.
- Residual plots like these are useful in diagnosing potential problems with the standard regression model.
- What to do?
The distribution of the residuals is far from randomly distributed about 0.

Residual plots like these are useful in diagnosing potential problems with the standard regression model.

What to do?

Consideration of functional form and transformation of data seem natural at this point.
The idea of segmented slopes or *piecewise* slopes gives rise to the idea that the slope coefficient may vary across different ranges of the data.
The idea of segmented slopes or *piecewise* slopes gives rise to the idea that the slope coefficient may vary across different ranges of the data.

This leads to the consideration of a model where different slope coefficients are estimated, conditional on a given range of the data.
Functional Form

- The idea of segmented slopes or *piecewise* slopes gives rise to the idea that the slope coefficient may vary across different ranges of the data.
- This leads to the consideration of model where different slope coefficients are estimated, conditional on a given range of the data.
- We’ve already seen the idea of conditional slope coefficients in practice. Interaction terms allow the slope to conditionally vary as a function of some other covariate.
The idea of segmented slopes or *piecewise* slopes gives rise to the idea that the slope coefficient may vary across different ranges of the data.

This leads to the consideration of a model where different slope coefficients are estimated, conditional on a given range of the data.

We’ve already seen the idea of conditional slope coefficients in practice. Interaction terms allow the slope to conditionally vary as a function of some other covariate.

The basic ideas motivating interaction terms help lead us to the idea of segmenting the slopes.
The idea of segmented slopes or *piecewise* slopes gives rise to the idea that the slope coefficient may vary across different ranges of the data.

This leads to the consideration of model where different slope coefficients are estimated, conditional on a given range of the data.

We’ve already seen the idea of conditional slope coefficients in practice. Interaction terms allow the slope to conditionally vary as a function of some other covariate.

The basic ideas motivating interaction terms helps lead us to the idea of segmenting the slopes.

Consider these data again.
Segmented Slopes
Segmented Slopes

The figure seems to suggest that the relationship between $Y$ and $X_1$ is curvilinear such that the “slope” seems to become less steep at higher levels of $X_1$. 
Segmented Slopes

- The figure seems to suggest that the relationship between $Y$ and $X_1$ is curvilinear such that the “slope” seems to become less steep at higher levels of $X_1$.
- Plot the residuals against the values of $X_1$ from a bivariate model.
Segmented Slopes

- The figure seems to suggest that the relationship between $Y$ and $X_1$ is curvilinear such that the “slope” seems to become less steep at higher levels of $X_1$.
- Plot the residuals against the values of $X_1$ from a bivariate model.
- In the figure, it appears that the slope changes pitch around the value $X_1 = 7$. 
Segmented Slopes

Residual Plot

Residuals

x1
Segmented Slopes

Suppose we account for the apparent nonlinearity in the data?
Segmented Slopes

- Suppose we account for the apparent nonlinearity in the data?
- A segmented slopes approach would allow us to do this.
Suppose we account for the apparent nonlinearity in the data? A segmented slopes approach would allow us to do this. The logic is to estimate a separate slope for the data above and below $X_1 = 7$. 
Segmented Slopes

- Suppose we account for the apparent nonlinearity in the data?
- A segmented slopes approach would allow us to do this.
- The logic is to estimate a separate slope for the data above and below $X_1 = 7$.
- To do this, we first create a dummy variable $z$ such that

$$z = \begin{cases} 
1 & \text{if } X_1 \geq 7 \\
0 & \text{if } X_1 < 7 
\end{cases}$$
Segmented Slopes

To define the separate slopes, create an interaction term between \( z \) and \( X_1 \).
Segmented Slopes

To define the separate slopes, create an interaction term between $z$ and $X_1$.

This gives:

$$\hat{Y} = \hat{a} + \hat{b}_1 X_1 + \hat{b}_2 z + \hat{b}_3 (zX_1),$$

which returns:

$$\hat{Y} = \hat{a} + \hat{b}_1 X_1,$$

for $X_1 < 7$, and

$$\hat{Y} = \hat{a} + \hat{b}_2 z + (\hat{b}_1 + \hat{b}_3) X_1,$$

for $X_1 \geq 7$. 

Estimate this model:

$$\hat{Y} = 441 + 0.404 X_1 + 0.263 z - 0.034 (zX_1),$$

where the RMSE = 0.027 and $F = 102.09$. 

**Segmented Slopes**

- To define the separate slopes, create an interaction term between \( z \) and \( X_1 \).
- This gives:
  \[
  \hat{Y} = \hat{a} + \hat{b}_1 X_1 + \hat{b}_2 z + \hat{b}_3 (zX_1),
  \]
  which returns:
  \[
  \hat{Y} = \hat{a} + \hat{b}_1 X_1,
  \]
  for \( X_1 < 7 \), and
  \[
  \hat{Y} = \hat{a} + \hat{b}_2 z + (\hat{b}_1 + \hat{b}_3) X_1,
  \]
  for \( X_1 \geq 7 \).
- Estimate this model:
  \[
  \hat{Y} = .441 + .040 X_1 + .263 z + -.034 (zX_1),
  \]
  where the \( RMSE = .027 \) and \( F = 102.09 \).
Segmented Slopes

Segmented Slopes Model

0 5 10 15 20
x1
Segmented Slopes

- This model explicitly accounts for the curvilinear nature of the data.
Segmented Slopes

- This model explicitly accounts for the curvilinear nature of the data.
- Further, it should be clear how the use of an interaction term can be used to capture segmented slopes.
Segmented Slopes

- This model explicitly accounts for the curvilinear nature of the data.
- Further, it should be clear how the use of an interaction term can be used to capture segmented slopes.
- In passing note the discontinuity at $X = 7$.
Segmented Slopes

- This model explicitly accounts for the curvilinear nature of the data.
- Further, it should be clear how the use of an interaction term can be used to capture segmented slopes.
- In passing note the discontinuity at $X = 7$.
- More formally, the difference in the ordinates at $X_1 = 7$ represent a discontinuity in the data.
Segmented Slopes

- This model explicitly accounts for the curvilinear nature of the data.
- Further, it should be clear how the use of an interaction term can be used to capture segmented slopes.
- In passing note the discontinuity at $X = 7$.
- More formally, the difference in the ordinates at $X_1 = 7$ represent a discontinuity in the data.
- This continuity is given by
  
  \[ \text{discontinuity at } X_1 = 7 = \hat{b}_2 - 7(\hat{b}_3). \]
Segmented Slopes

- This result holds because $\hat{b}_3$ is only defined for $X_1 = 7$ and above.
Segmented Slopes

- This result holds because $\hat{b}_3$ is only defined for $X_1 = 7$ and above.
- We could constrain the fit to make this difference 0 by imposing the restriction that
  
  $$\hat{b}_2 = -7(\hat{b}_3).$$
Segmented Slopes

- This result holds because $\hat{b}_3$ is only defined for $X_1 = 7$ and above.
- We could constrain the fit to make this difference 0 by imposing the restriction that
  \[ \hat{b}_2 = -7(\hat{b}_3). \]
- Constraining the model to satisfy this constraint is achieved by imposing this condition on the model, which gives rise to
  \[ \hat{Y} = \hat{a} + \hat{b}_1 X_1 + \hat{b}_3 (z \ast (X_1 - 7)). \]
Segmented Slopes

- This result holds because $\hat{b}_3$ is only defined for $X_1 = 7$ and above.
- We could constrain the fit to make this difference 0 by imposing the restriction that
  $$\hat{b}_2 = -7(\hat{b}_3).$$
- Constraining the model to satisfy this constraint is achieved by imposing this condition on the model, which gives rise to
  $$\hat{Y} = \hat{a} + \hat{b}_1 X_1 + \hat{b}_3 (z \ast (X_1 - 7)).$$
- The last term in this model is zero for $X_1 < 7$ (why?) and is $X_1 - 7$ if $X_1 > 7$. (Why?)
Segmented Slopes

- This result holds because $\hat{b}_3$ is only defined for $X_1 = 7$ and above.
- We could constrain the fit to make this difference 0 by imposing the restriction that $\hat{b}_2 = -7(\hat{b}_3)$.
- Constraining the model to satisfy this constraint is achieved by imposing this condition on the model, which gives rise to

$$\hat{Y} = \hat{a} + \hat{b}_1 X_1 + \hat{b}_3 \left( z \ast (X_1 - 7) \right).$$

- The last term in this model is zero for $X_1 < 7$ (why?) and is $X_1 - 7$ if $X_1 > 7$. (Why?)
- Fitting this model would force the two slopes to meet at $X_1 = 7$.

$$\hat{Y} = .434 + .043X_1 + -.037 \left( z \ast (X_1 - 7) \right)$$
Segmented Slopes

Segmented Slopes Model with Continuity at $x_1=7$
Transformations on $X$

- Is what we just did “curve fitting”?
Transformations on $X$

- Is what we just did “curve fitting”?
- Alternative approaches?
Transformations on $X$

- Is what we just did “curve fitting”?  
- Alternative approaches?  
- Suppose we have monotonicity but one that exhibits marginality.
Transformations on $X$

- Is what we just did “curve fitting”? 
- Alternative approaches? 
- Suppose we have monotonicity but one that exhibits marginality. 
- The natural log transformation on $x$ may be one “solution”.
Transformations on $X$

- Is what we just did “curve fitting”?
- Alternative approaches?
- Suppose we have monotonicity but one that exhibits marginality.
- The natural log transformation on $x$ may be one “solution”.
- The log transformation works because, in a sense, it compresses the $x$ axis.
Transformations on $X$

- When taking the log of a variable, the “distance” between adjacent values of the logged variable decreases as the values of the unlogged variable increase.
Transformations on $X$

- When taking the log of a variable, the “distance” between adjacent values of the logged variable decreases as the values of the unlogged variable increase.

- Estimating a regression model with a logged covariate presents no challenges:

$$\hat{Y} = \hat{a} + \hat{b}_1 \log(X_1),$$

where the independent variable is transformed by the natural log.
Transformations on $X$

- When taking the log of a variable, the “distance” between adjacent values of the logged variable decreases as the values of the unlogged variable increase.

- Estimating a regression model with a logged covariate presents no challenges:

$$\hat{Y} = \hat{a} + \hat{b}_1 \log(X_1),$$

where the independent variable is transformed by the natural log.

- It is easy to see that the “compression” in $X_1$ is a desirable property when saturation effects are present because by logging $X_1$, we are multiplying by $\hat{b}_1$, a number that is increasing in values at a decreasing rate.
Transformations on $X$

- When taking the log of a variable, the “distance” between adjacent values of the logged variable decreases as the values of the unlogged variable increase.

- Estimating a regression model with a logged covariate presents no challenges:

  \[ \hat{Y} = \hat{a} + \hat{b}_1 \log(X_1), \]

  where the independent variable is transformed by the natural log.

- It is easy to see that the “compression” in $X_1$ is a desirable property when saturation effects are present because by logging $X_1$, we are multiplying by $\hat{b}_1$, a number that is increasing in values at a decreasing rate.

- This is precisely the transformation we are after.
Transformations on $X$

- Return to simulated data from before.
Transformations on $X$:

- Return to simulated data from before.
- Using data from top two panels of previous figure estimate regression with logged $X$. 
Transformations on $X$

- Return to simulated data from before.
- Using data from top two panels of previous figure estimate regression with logged $X$.
- Table next slide.
Transformations on $X$

Table 3: Regression Models Based on Data from Figure 32 with Transformations.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Standard Model 1</th>
<th>With log $X_1$</th>
<th>Standard Model 2</th>
<th>With log $X_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.754 (.019)</td>
<td>.481 (.018)</td>
<td>.616 (.025)</td>
<td>.482 (.018)</td>
</tr>
<tr>
<td>Slope</td>
<td>-.193 (.038)</td>
<td>-.112 (.008)</td>
<td>.0007 (.0001)</td>
<td>.056 (.004)</td>
</tr>
<tr>
<td>$RMSE$</td>
<td>.074</td>
<td>.033</td>
<td>.076</td>
<td>.033</td>
</tr>
<tr>
<td>$r^2$</td>
<td>.58</td>
<td>.91</td>
<td>.56</td>
<td>.91</td>
</tr>
</tbody>
</table>
Transformations on X

Regression with x1 Untransformed

Regression with x1 Logged

Regression with x1 Untransformed

Regression with x1 Logged
Transformations on $X$

- Take caution in interpreting these results!
Transformations on $X$

- Take caution in interpreting these results!
- While the least squares solution works with no special problems for these models, you need to remember that the parameter estimate for the transformed model is based on $\log X_1$, not the untransformed $X_1$. When computing predicted values, you are multiplying the coefficient by the transformed variable. In order to make substantive sense of the coefficient, you have to be careful to "backtransform" the data into units that are interpretable to your reader. The correct interpretation of the regression model with transformed $X_1$ is that for a unit increase in $\log X_1$, the expected change is $\hat{Y}$ is given by the parameter estimate $\hat{b}_1$. 
Transformations on $X$

- Take caution in interpreting these results!
- While the least squares solution works with no special problems for these models, you need to remember that the parameter estimate for the transformed model is based on $\log X_1$, not the untransformed $X_1$.
- When computing predicted values, you are multiplying the coefficient by the transformed variable.
Transformations on $X$

- Take caution in interpreting these results!
- While the least squares solution works with no special problems for these models, you need to remember that the parameter estimate for the transformed model is based on $\log X_1$, *not* the untransformed $X_1$.
- When computing predicted values, you are multiplying the coefficient by the transformed variable.
- In order to make substantive sense of the coefficient, you have to be careful to “backtransform” the data into units that are interpretable to your reader.
Transformations on $X$

- Take caution in interpreting these results!
- While the least squares solution works with no special problems for these models, you need to remember that the parameter estimate for the transformed model is based on $\log X_1$, not the untransformed $X_1$.
- When computing predicted values, you are multiplying the coefficient by the transformed variable.
- In order to make substantive sense of the coefficient, you have to be careful to “backtransform” the data into units that are interpretable to your reader.
- The correct interpretation of the regression model with transformed $X_1$ is that for a unit increase in $\log X_1$, the expected change is $Y$ is given by the parameter estimate $\hat{b}_1$. 
Transformations on $X$

Regression with $x_1$ Logged

![Graph showing regression with $x_1$ logged](image-url)
Transformations on $X$

- Apart from the log transformation, another transformation to account for marginality is the square root transformation.
Transformations on $X$:

- Apart from the log transformation, another transformation to account for marginality is the square root transformation.
- This transformation also compresses the $x$-axis and so the logic of the transformation is similar to that of the log transformation.

Additionally, one thing to remember is that the log $X$ is undefined for $X \leq 0$ and the square root of negative numbers is also undefined.

If you have many 0s or if you have negative numbers, one or both of these transformations will be unsuitable and could result in a staggering loss of data.
Transformations on $X$

- Apart from the log transformation, another transformation to account for marginality is the square root transformation.
- This transformation also compresses the $x$-axis and so the logic of the transformation is similar to that of the log transformation.
- Either are suitable, and so one may want to compare $F$ statistics or $r^2$ measures to compare models using the different transformations.

One thing to remember is that the log $X \leq 0$ is undefined and the square root of negative numbers is also undefined. If you have many 0s or if you have negative numbers, one or both of these transformations will be unsuitable and could result in a staggering loss of data.
Transformations on $X$

- Apart from the log transformation, another transformation to account for marginality is the square root transformation.
- This transformation also compresses the $x$-axis and so the logic of the transformation is similar to that of the log transformation.
- Either are suitable, and so one may want to compare $F$ statistics or $r^2$ measures to compare models using the different transformations.
- One thing to remember is that the log $X_1 \leq 0$) is undefined and the square root of negative numbers is also undefined.
Transformations on $X$

- Apart from the log transformation, another transformation to account for marginality is the square root transformation.
- This transformation also compresses the $x$-axis and so the logic of the transformation is similar to that of the log transformation.
- Either are suitable, and so one may want to compare $F$ statistics or $r^2$ measures to compare models using the different transformations.
- One thing to remember is that the log $X_1 \leq 0$) is undefined and the square root of negative numbers is also undefined.
- If you have many 0s or if you have negative numbers, one or both of these transformations will be unsuitable and could result in a staggering loss of data.
Mosteller and Tukey’s Blueling Rule

Figure 4.6. Tukey and Mosteller’s “bulging rule”: The direction of the “bulge” indicates the direction of the power transformation of $Y$ and/or $X$ to straighten the relationship between them.

Simple monotone nonlinearity can often be corrected by a power transformation of $X$, of $Y$, or of both variables. Mosteller and Tukey’s “bulging rule” assists in the selection of a transformation.
Transformations on \( X \)

- Suppose your data exhibit nonmonotonicity.
Transformations on $X$

- Suppose your data exhibit nonmonotonicity.
- For example, imagine the relationship between $Y$ and $X_1$ is u-shaped (or inverted u-shaped).
Suppose your data exhibit nonmonotonicity.

For example, imagine the relationship between $Y$ and $X_1$ is u-shaped (or inverted u-shaped).

The question naturally arises as to how to account for nonlinearity of this form in the context of a regression model.
Transformations on $X$

- Suppose your data exhibit nonmonotonicity.
- For example, imagine the relationship between $Y$ and $X_1$ is u-shaped (or inverted u-shaped).
- The question naturally arises as to how to account for nonlinearity of this form in the context of a regression model.
- A common approach to handling this is through polynomial regression.
Transformations on $X$

- Polynomials are an algebraic expression of the form
  
  $$a_nx^n + a_{n-1}x^{n-1} + \ldots + a_3x^3 + a_2x^2 + a_1x + a_0,$$

  where $a_0, a_1, \ldots, a_n$ are constants that are the coefficient of the polynomial, and $n$ is a positive integer.
Transformations on $X$

- Polynomials are an algebraic expression of the form
  
  \[ a_n x^n + a_{n-1} x^{n-1} + \ldots + a_3 x^3 + a_2 x^2 + a_1 x + a_0, \]

  where $a_0, a_1, \ldots, a_n$ are constants that are the coefficient of the polynomial, and $n$ is a positive integer.

- The degree of the polynomial is the highest power of the variable that appears.
Transformations on \( X \)

- Polynomials are an algebraic expression of the form
  \[
  a_n x^n + a_{n-1} x^{n-1} + \ldots + a_3 x^3 + a_2 x^2 + a_1 x + a_0,
  \]
  where \( a_0, a_1, \ldots, a_n \) are constants that are the coefficient of the polynomial, and \( n \) is a positive integer.

- The degree of the polynomial is the highest power of the variable that appears.

- Graphs of polynomial functions are useful to understand because they illustrate the point that the curve of the function can change directions.
Transformations on $X$

- Polynomials are an algebraic expression of the form
  
  \[ a_n x^n + a_{n-1} x^{n-1} + \ldots + a_3 x^3 + a_2 x^2 + a_1 x + a_0, \]

  where $a_0, a_1, \ldots, a_n$ are constants that are the coefficient of the polynomial, and $n$ is a positive integer.

- The degree of the polynomial is the highest power of the variable that appears.

- Graphs of polynomial functions are useful to understand because they illustrate the point that the curve of the function can change directions.

- The number of “turning points” in a polynomial is odd if the degree of the polynomial is even, and vice versa.
Transformations on $X$

- Polynomials are an algebraic expression of the form
  \[ a_n x^n + a_{n-1} x^{n-1} + \ldots + a_3 x^3 + a_2 x^2 + a_1 x + a_0, \]
  where $a_0, a_1, \ldots, a_n$ are constants that are the coefficient of
  the polynomial, and $n$ is a positive integer.
- The degree of the polynomial is the highest power of the
  variable that appears.
- Graphs of polynomial functions are useful to understand
  because they illustrate the point that the curve of the function
  can change directions.
- The number of “turning points” in a polynomial is odd if the
  degree of the polynomial is even, and vice versa.
- A polynomial of degree 1 has no turning points and so it
  produces a straight line. A polynomial of degree 2 has 1
  turning point and is known as a quadratic. A polynomial of
  degree 3 has 0 or 2 turning points and is known as a cubic.
Transformations on $X$

- I’m going to focus on the quadratic model, which is widely applied in political science settings, though it is often incorrectly interpreted.
Transformations on $X$

- I’m going to focus on the quadratic model, which is widely applied in political science settings, though it is often incorrectly interpreted.
- To fix ideas, suppose we have data that look roughly like that shown in earlier.
Transformations on $X$

- I’m going to focus on the quadratic model, which is widely applied in political science settings, though it is often incorrectly interpreted.

- To fix ideas, suppose we have data that look roughly like that shown in earlier.

- Estimation of

  \[ Y = \hat{a} + \hat{b}X_1 \]

  will be unsatisfactory.
Transformations on $X$

- I’m going to focus on the quadratic model, which is widely applied in political science settings, though it is often incorrectly interpreted.
- To fix ideas, suppose we have data that look roughly like that shown in earlier.
- Estimation of
  \[ Y = \hat{a} + \hat{b}X_1 \]
  will be unsatisfactory.
- Suppose we estimate
  \[ Y = \hat{a} + \hat{b}X_1 + \hat{b}X_1^2, \]
  where $X_1^2$ is $X_1$ squared
Transformations on $X$

- I’m going to focus on the quadratic model, which is widely applied in political science settings, though it is often incorrectly interpreted.
- To fix ideas, suppose we have data that look roughly like that shown in earlier.
- Estimation of
  \[ Y = \hat{a} + \hat{b}X_1 \]
  will be unsatisfactory.
- Suppose we estimate
  \[ Y = \hat{a} + \hat{b}X_1 + \hat{b}X_1^2, \]
  where $X_1^2$ is $X_1$ squared.
- This regression model is a polynomial model with degree 2.
Transformations on $X$

- It is also known as a quadratic model.
Transformations on $X$

- It is also known as a quadratic model.
- Estimating the model in this form will produce a response function (or a curve) that will have a single bend in it.
Transformations on $X$

- It is also known as a quadratic model.
- Estimating the model in this form will produce a response function (or a curve) that will have a single bend in it.
- Understand that the bend *may or may not* be observed in the range of data with which you are working.
Transformations on $X$

- It is also known as a quadratic model.
- Estimating the model in this form will produce a response function (or a curve) that will have a single bend in it.
- Understand that the bend *may or may not* be observed in the range of data with which you are working.
- Nevertheless, at some point, the response function will inflect, or change directions.
Transformations on $X$

- It is also known as a quadratic model.
- Estimating the model in this form will produce a response function (or a curve) that will have a single bend in it.
- Understand that the bend *may or may not* be observed in the range of data with which you are working.
- Nevertheless, at some point, the response function will inflect, or change directions.
- At the point at which the response function changes direction, the curve has a horizontal tangent.
Transformations on $X$

- In regression terms, at the point where the curve changes direction, the slope (partial or otherwise) between $Y$ and $X_1$ is exactly 0.
Transformations on $X$

- In regression terms, at the point where the curve changes direction, the slope (partial or otherwise) between $Y$ and $X_1$ is exactly 0.

- This point is critical to understand because it demonstrates quite clearly that the slope of the relationship is not constant with respect to $X_1$; rather, it is *conditional* on $X_1$. 
Transformations on X

- In regression terms, at the point where the curve changes direction, the slope (partial or otherwise) between $Y$ and $X_1$ is exactly 0.
- This point is critical to understand because it demonstrates quite clearly that the slope of the relationship is not constant with respect to $X_1$; rather, it is conditional on $X_1$.
- We have seen conditional slopes before:

  $$Y = \hat{a} + \hat{b_1}X_1 + \hat{b_2}X_2 + \hat{b_3}X_1X_2.$$
Transformations on $X$

- In regression terms, at the point where the curve changes direction, the slope (partial or otherwise) between $Y$ and $X_1$ is exactly 0.
- This point is critical to understand because it demonstrates quite clearly that the slope of the relationship is not constant with respect to $X_1$; rather, it is *conditional* on $X_1$.
- We have seen conditional slopes before:
  \[ Y = \hat{a} + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \hat{b}_3 X_1 X_2. \]
- It should be obvious that a polynomial model is equivalent to an “interactive” model. This is easy to see if we rewrite the quadratic model as
  \[ Y = \hat{a} + \hat{b} X_1 + \hat{b} X_1 X_1, \]
  where the last term is (obviously) equivalent to $X_1^2$. 
Transformations on \( X \)

- In a quadratic model, we’re essentially interacting \( X_1 \) with itself.
Transformations on $X$

- In a quadratic model, we’re essentially interacting $X_1$ with itself.
- The slope of $Y$ on $X_1$ will be conditional dependent upon the value of $X_1$. 

\[
\frac{\partial Y}{\partial X_1} = \hat{b}_1 + 2\hat{b}_2 X_1.
\]

This partial derivative illustrates that the rate of change is conditional on $X_1$ and that the rate will vary as $X_1$ changes. The slope is not constant.
Transformations on $X$

- In a quadratic model, we’re essentially interacting $X_1$ with itself.
- The slope of $Y$ on $X_1$ will be conditional dependent upon the value of $X_1$.
- In general, the rate of change of $Y$ with respect to $X$ is given by

$$\frac{\partial Y}{\partial X_1} = \hat{b}_1 + 2\hat{b}_2 X_1.$$
Transformations on $X$

- In a quadratic model, we’re essentially interacting $X_1$ with itself.
- The slope of $Y$ on $X_1$ will be conditional dependent upon the value of $X_1$.
- In general, the rate of change of $Y$ with respect to $X$ is given by
  \[
  \frac{\partial Y}{\partial X_1} = \hat{b}_1 + 2\hat{b}_2X_1.
  \]
- This partial derivative illustrates that the rate of change is conditional on $X_1$ and that the rate will vary as $X_1$ changes.
Transformations on $X$

- In a quadratic model, we’re essentially interacting $X_1$ with itself.
- The slope of $Y$ on $X_1$ will be conditional dependent upon the value of $X_1$.
- In general, the rate of change of $Y$ with respect to $X$ is given by

$$\frac{\partial Y}{\partial X_1} = \hat{b}_1 + 2\hat{b}_2 X_1.$$ 

- This partial derivative illustrates that the rate of change is conditional on $X_1$ and that the rate will vary as $X_1$ changes.
- The slope is not constant.
Transformations on $X$

- To illustrate, suppose we reexamine the data from before that was shown in the lower left panel.
Transformations on $X$

- To illustrate, suppose we reexamine the data from before that was shown in the lower left panel.

- In applying the quadratic model from above to these data, we obtain the following estimates:

$$\hat{Y} = 0.428 + 0.056X_1 - 0.003X_1^2.$$
Transformations on $X$

- To illustrate, suppose we reexamine the data from before that was shown in the lower left panel.
- In applying the quadratic model from above to these data, we obtain the following estimates:
  \[ \hat{Y} = 0.428 + 0.056X_1 - 0.003X_1^2. \]
- The $r^2$ for these data is .95 and the RMSE is .02. (Compare these results to the linear regression results).
Transformations on $X$

- To illustrate, suppose we reexamine the data from before that was shown in the lower left panel.

- In applying the quadratic model from above to these data, we obtain the following estimates:

\[
\hat{Y} = 0.428 + 0.056X_1 - 0.003X_1^2.
\]

- The $r^2$ for these data is 0.95 and the RMSE is 0.02. (Compare these results to the linear regression results).

- Note the signs on the coefficients: the positive sign on the constituent term (i.e. $\hat{b}_1$) and the negative sign on the squared term (i.e. $\hat{b}_2$) indicates the curve is first increasing to some point and then decreasing.
Transformations on $X$

- To illustrate, suppose we reexamine the data from before that was shown in the lower left panel.
- In applying the quadratic model from above to these data, we obtain the following estimates:

$$\hat{Y} = .428 + .056X_1 - .003X_1^2$$

- The $r^2$ for these data is .95 and the $RMSE$ is .02. (Compare these results to the linear regression results).
- Note the signs on the coefficients: the positive sign on the constituent term (i.e. $\hat{b}_1$) and the negative sign on the squared term (i.e. $\hat{b}_2$) indicates the curve is first increasing to some point and then decreasing.
- That is, this combination of signed coefficients will produce an inverted u-shaped function.
Transformations on $X$
Transformations on $X$

- Just like the interactive models, the slope is conditional on $X_1$. 
Transformations on $X$

- Just like the interactive models, the slope is conditional on $X_1$.
- ...so too will be the standard errors.
Transformations on $X$

- Just like the interactive models, the slope is conditional on $X_1$.
- ...so too will be the standard errors.
- This implies that some points on the quadratic response function may be no different from 0.
Transformations on $X$

- Just like the interactive models, the slope is conditional on $X_1$. 
- ...so too will be the standard errors. 
- This implies that some points on the quadratic response function may be no different from 0. 
- So even if the quadratic “holds” based on inspection of the standard errors given by the default output, it need not be the case that entirety of the quadratic function is statistically significant (this is just like the interaction model setting).
Transformations on $X$

- We need to compute the standard error of the conditional slope

\[ \hat{b}X_1 + \hat{b}X_1^2. \]
Transformations on $X$

- We need to compute the standard error of the conditional slope
  \[ \hat{b}X_1 + \hat{b}X_1^2. \]

- Using standard results from the variance and covariance of two random variables, the variance of the conditional slope is given by
  \[ \text{var}(\hat{b}X_1 + \hat{b}X_1^2) = \text{var}(\hat{b}_1) + 4X_1^2\text{var}(\hat{b}_2) + 4X_1\text{cov}(\hat{b}_1\hat{b}_2), \]
  and the standard error is given by the square root of this term.
Transformations on $X$

- We need to compute the standard error of the conditional slope
  $$\hat{b}X_1 + \hat{b}X_1^2.$$  

- Using standard results from the variance and covariance of two random variables, the variance of the conditional slope is given by
  $$\text{var}(\hat{b}X_1 + \hat{b}X_1^2) = \text{var}(\hat{b}_1) + 4X_1^2\text{var}(\hat{b}_2) + 4X_1\text{cov}(\hat{b}_1\hat{b}_2),$$
  and the standard error is given by the square root of this term.

- With the standard errors in hand, we can compute the $t$ ratios for each conditional slope and then determine which points, if any, on the quadratic term are no different from 0.
Transformations on $X$

- We need to compute the standard error of the conditional slope

$$\hat{b}X_1 + \hat{b}X_1^2.$$

- Using standard results from the variance and covariance of two random variables, the variance of the conditional slope is given by

$$\text{var}(\hat{b}X_1 + \hat{b}X_1^2) = \text{var}(\hat{b}_1) + 4X_1^2\text{var}(\hat{b}_2) + 4X_1\text{cov}(\hat{b}_1 \hat{b}_2),$$

and the standard error is given by the square root of this term.

- With the standard errors in hand, we can compute the $t$ ratios for each conditional slope and then determine which points, if any, on the quadratic term are no different from 0.

- As in the case of the interaction model, sometimes there will be many dozens (or even hundreds) of possible conditional slopes.
Transformations on $X$

- We need to compute the standard error of the conditional slope
  $$\hat{b}X_1 + \hat{b}X_1^2.$$  

- Using standard results from the variance and covariance of two random variables, the variance of the conditional slope is given by
  $$\text{var}(\hat{b}X_1 + \hat{b}X_1^2) = \text{var}(\hat{b}_1) + 4X_1^2\text{var}(\hat{b}_2) + 4X_1\text{cov}(\hat{b}_1\hat{b}_2),$$
  and the standard error is given by the square root of this term.

- With the standard errors in hand, we can compute the $t$ ratios for each conditional slope and then determine which points, if any, on the quadratic term are no different from 0.

- As in the case of the interaction model, sometimes there will be many dozens (or even hundreds) of possible conditional slopes.
Transformations on $X$
Transformations on $X$

- What about the inflection point?
Transformations on $X$

- What about the inflection point?
- As noted above, the inflection point may or may not be found in the observed data.
Transformations on $X$

- What about the inflection point?
- As noted above, the inflection point may or may not be found in the observed data.
- Formally, the inflection point from a quadratic model is given by
  
  $$\frac{\partial Y}{\partial X_1} = 0.$$
Transformations on $X$

- What about the inflection point?
- As noted above, the inflection point may or may not be found in the observed data.
- Formally, the inflection point from a quadratic model is given by

$$\frac{\partial Y}{\partial X_1} = 0.$$ 

- To compute this directly, note that this expression is equivalent to

$$\frac{-\hat{b}_1}{2\hat{b}_2}.$$
Transformations on $X$

- What about the inflection point?
- As noted above, the inflection point may or may not be found in the observed data.
- Formally, the inflection point from a quadratic model is given by
  \[ \frac{\partial Y}{\partial X_1} = 0. \]
- To compute this directly, note that this expression is equivalent to
  \[ \frac{-\hat{b}_1}{2\hat{b}_2}. \]
- Substituting the parameter estimates into this expression, the quotient is 9.94.
Transformations on $X$

Quadratic Model with Inflection Point Noted
Transformations on $X$

- Observations:
  
  - Take seriously functional form and think about what a linear, additive model actually means.
  
  - Interpret the model correctly once transformed.
  
  - Compute the right quantities for uncertainty.
Transformations on $X$

- Observations:
  - Take seriously functional form and think about what a linear, additive model actually means.
Transformations on X

- Observations:
  - Take seriously functional form and think about what a linear, additive model actually means.
  - Interpret the model correctly once transformed.
Transformations on $X$

- Observations:
- Take seriously functional form and think about what a linear, additive model actually means.
- Interpret the model correctly once transformed.
- Compute the right quantities for uncertainty.