Linear Regression: What do these numbers mean?

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

y

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

Ordinary Least Squares

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

m

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

Residual degrees of freedom

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

Residual degrees of freedom = number of observations - number of parameters (including intercept)

Model degrees of freedom

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

R^2

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

Best model minimizes

$$\sum_{i=1}^{m} \left(y_{\beta}(x^{(i)}) - y_{obs}^{(i)} \right)^2$$

Sum of Squared Error
SSE

Variance of observed points (times m) is

$$\sum_{i=1}^{m} \left(\overline{y}_{obs} - y_{obs}^{(i)} \right)^2$$

Total Sum of Squares
SST

$$R^2 = 1 - \frac{SSE}{SST}$$

Sum of Squared Error
SSE

Total Sum of Squares
SST

$$R^2 = 1 - \frac{SSE}{SST}$$

Randomness
left in the model

Variation in the data

$$R^2 = 1 - \frac{SSE}{SST}$$

Randomness left in the model

Variation in the data

SSE/SST is the portion of variation left unexplained by the model (handled by ε)

$$R^2 = 1 - \frac{SSE}{SST}$$

Randomness left in the model

Variation in the data

R^2 is the portion of variation explained by the model (R^2 is between 0 and 1)

(as long as the model has smaller residuals than the mean-only model)

OLS Regression Results

Dep. Variable:	DomesticTotalGross	$\boxed{\textbf{R-squared:} \quad 0.286}$	
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

R^2

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

F-test

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

Null hypothesis:

This data can be modeled by setting all β values to zero

(and the linear relationship we've found is purely due to chance)

Null hypothesis:

This data can be modeled by setting all β values to zero

(and the linear relationship we've found is purely due to chance)

If p-value <0.05, we can reject the null hypothesis.

Data is too extreme to fit this model just by chance.

Null hypothesis:
This data can be modeled by setting all β values to zero
(and the linear relationship we've found is purely due to chance)

If p-value <0.05, we can reject the null hypothesis.
Data is too extreme to fit this model just by chance.

Doesn't mean the model is "true"

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

Log L

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

Likelihood is a different cost function

$$L(\beta_0, \beta_1) = p(y_{obs} | \beta_0, \beta_1)$$

$$p(\beta_0, \beta_1 | y_{obs}) = \frac{p(y_{obs} | \beta_0, \beta_1) p(\beta_0, \beta_1)}{p(y_{obs})}$$

Likelihood is a different cost function

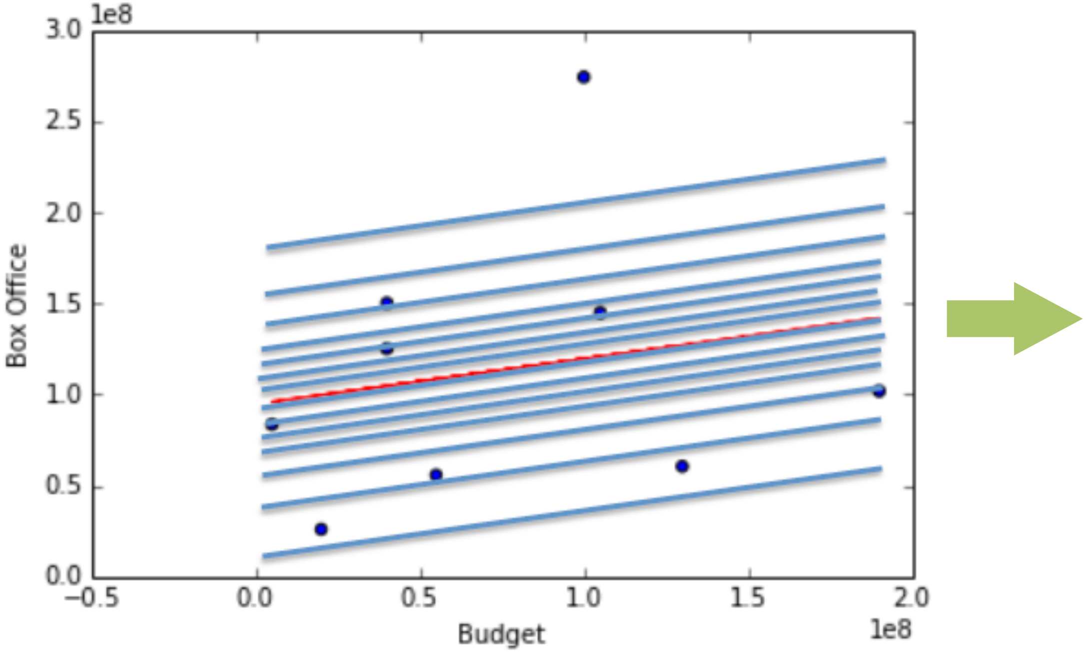
$$L(\beta_0, \beta_1) = p(y_{obs} | \beta_0, \beta_1)$$

For a given model (pair of β_0 And β_1 values), Likelihood is the prob. Of getting exactly this set of observed y values

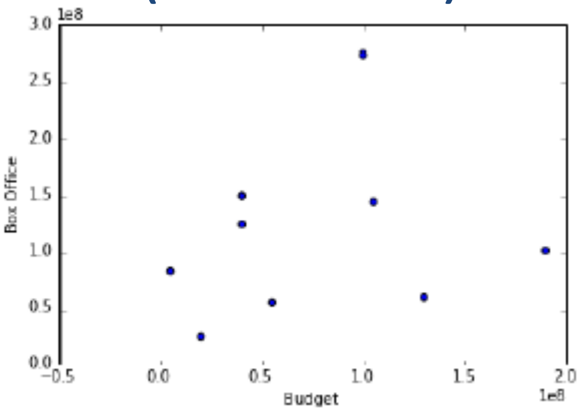
The model with maximum likelihood is the best fit.

Likelihood is a different cost function

$$L(\beta_0, \beta_1) = p(y_{obs} | \beta_0, \beta_1)$$



Our world
(observed)



OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

		coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
β_1	**Budget**	0.7846	0.133	5.901	0.000	0.520 1.049
β_0	**Ones**	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

Standard error of the coefficient

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

95% conf interval for coefficient's value

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

t-test

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

Null hypothesis:

This specific β value is zero

(and the data can be created by such a model (with the other β values intact)

Null hypothesis:

This specific β value is zero

(and the data can be created by such a model (with the other β values intact)

If p-value <0.05, we can reject the null hypothesis.

This variable DOES contribute to the model.

Null hypothesis:

This specific β value is zero

(and the data can be created by such a model (with the other β values intact)

If p-value <0.05, we can reject the null hypothesis.

This variable DOES contribute to the model.

Note: DOES or DOESN'T. Not how much.

OLS Regression Results

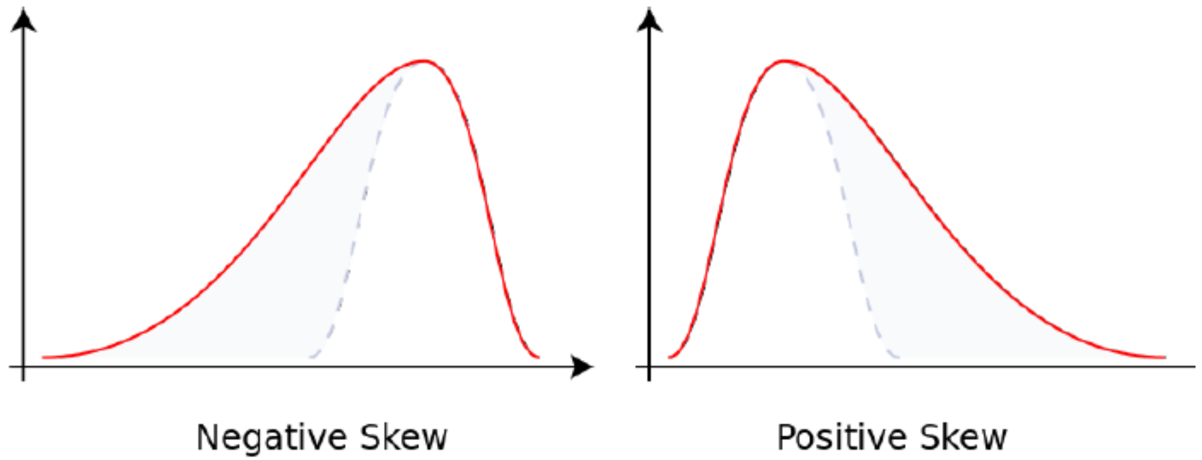
Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

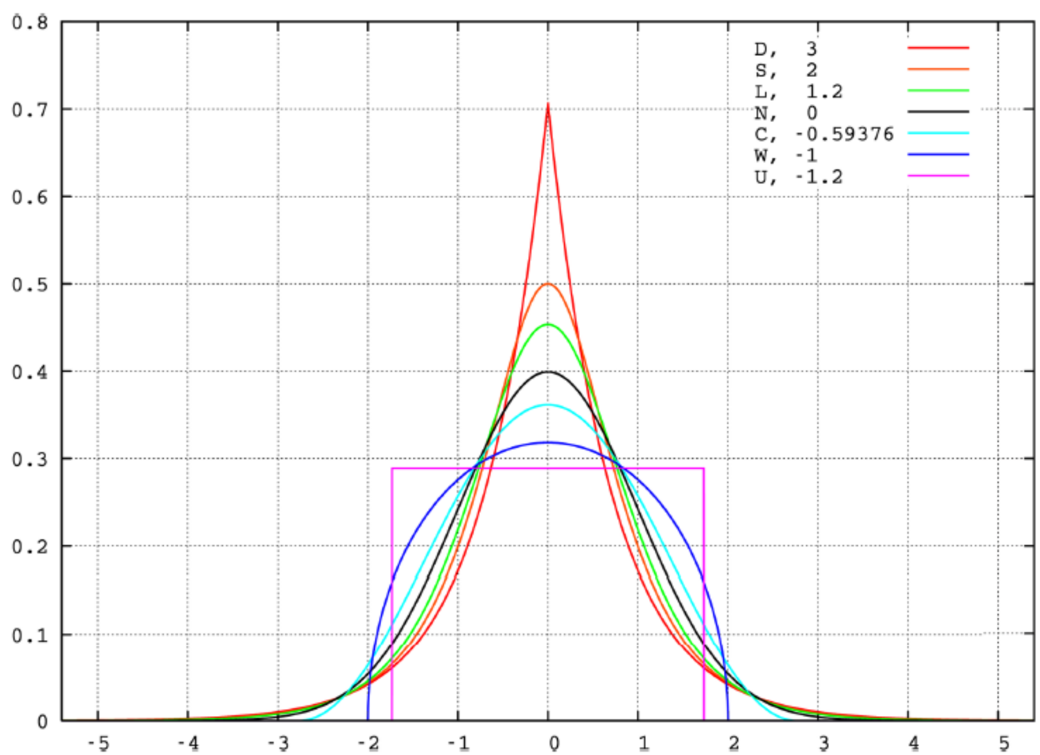
Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	$\boxed{1.587}$	**Prob(JB):**	2.55e-22
Kurtosis:	$\boxed{7.091}$	**Cond. No.**	1.54e+08

Skew & Kurtosis

Skew
(asymmetry)



Kurtosis
(peakness)



OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Normality test

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

Null hypothesis:

ε is normally distributed.

(no skew, no excess kurtosis)

If p-value <0.05, we can reject the null hypothesis.

ε does not exactly follow a normal distribution as we assumed

We may need to look closer.

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	$\boxed{99.441}$
Skew:	1.587	**Prob(JB):**	$\boxed{2.55e-22}$
Kurtosis:	7.091	**Cond. No.**	1.54e+08

Another normality test

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

Autocorrelation test

OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

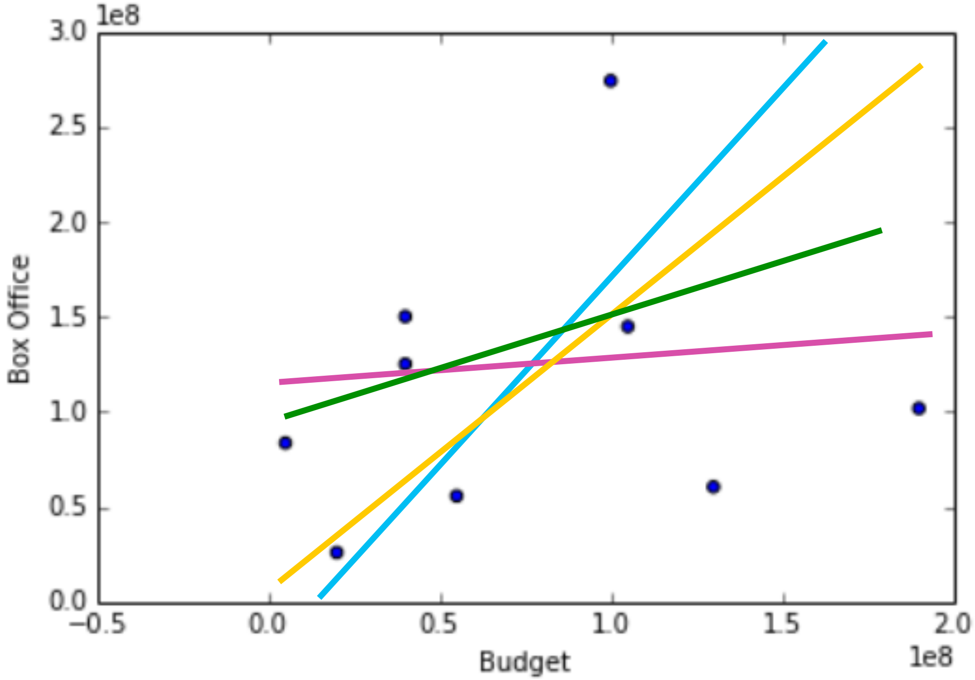
	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	$\boxed{1.54e+08}$

Sensitivity of prediction to small errors in input

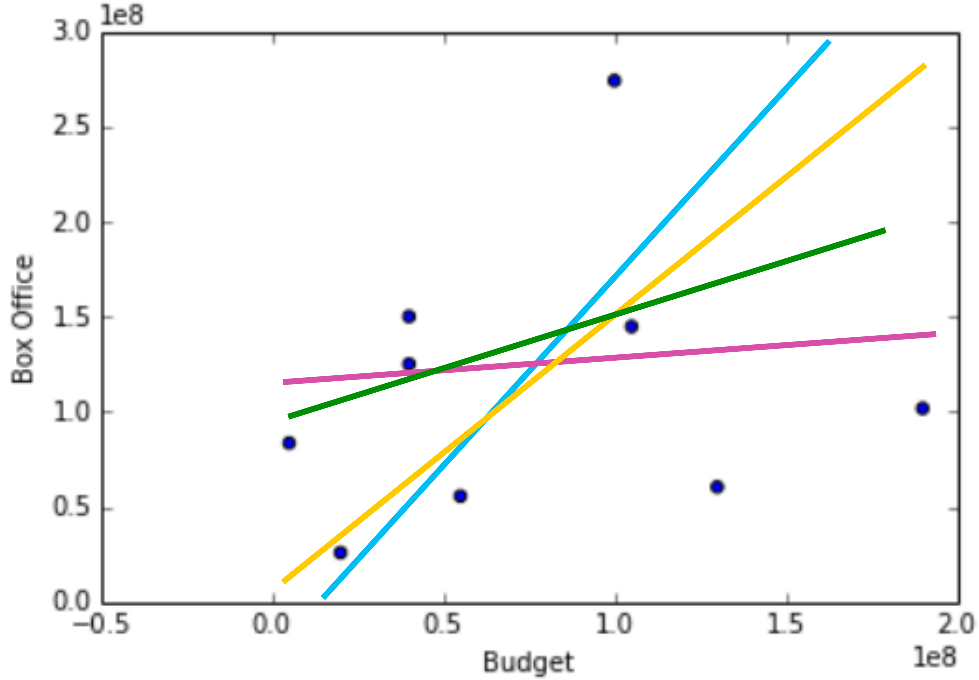
Model Selection I

$$y_\beta(x) = \beta_0 + \beta_1 x + \varepsilon$$



For models with the same amount of parameters, easy:

$$y_\beta(x) = \beta_0 + \beta_1 x + \varepsilon$$

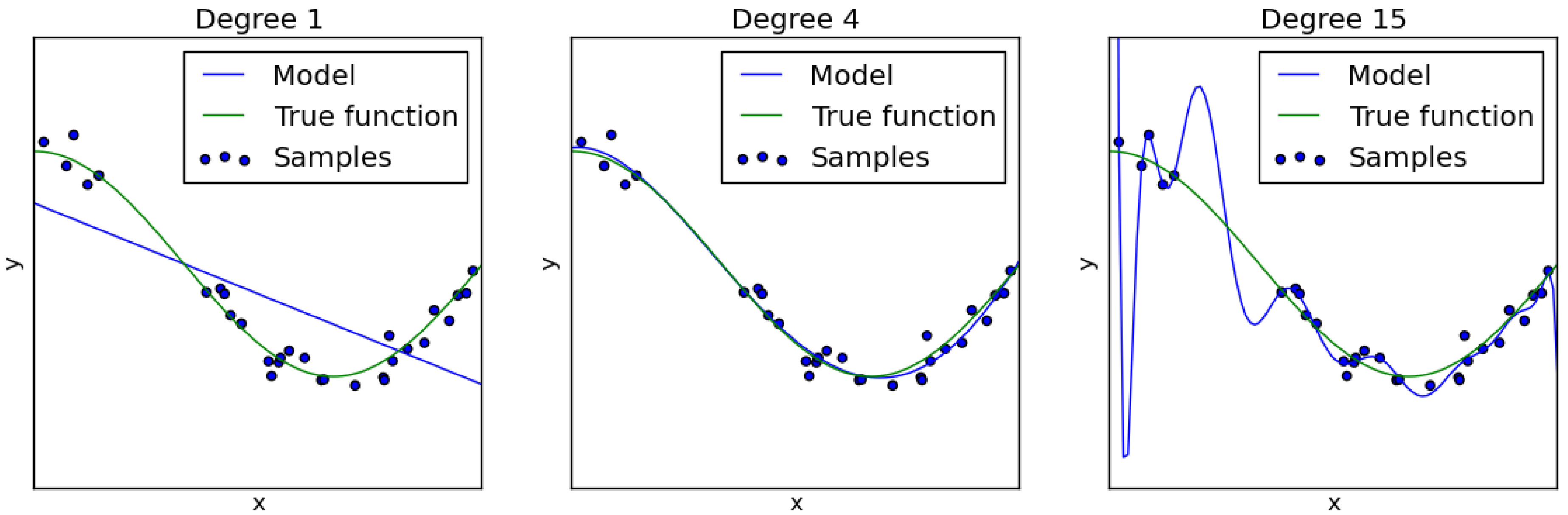


For models with the same amount of parameters, easy:

Take the one with the better cost function

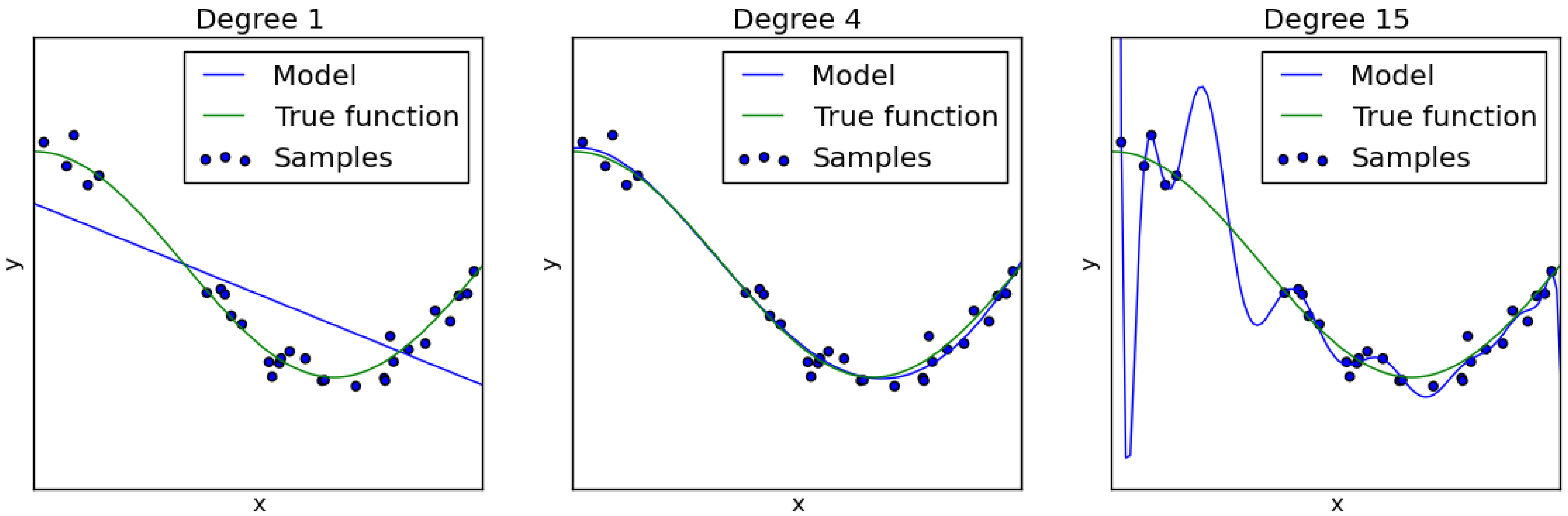
Log-Likelihood:	-1753.0

For models of different complexity: Beware under/overfitting



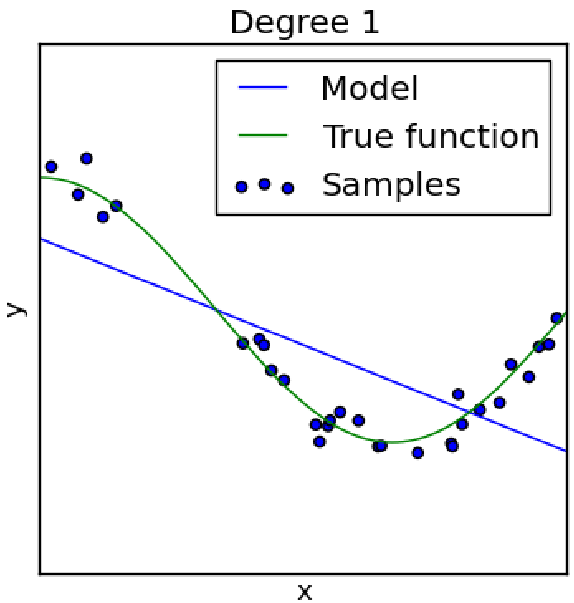
For models of different complexity: Beware under/overfitting

Underfitting Just Right Overfitting

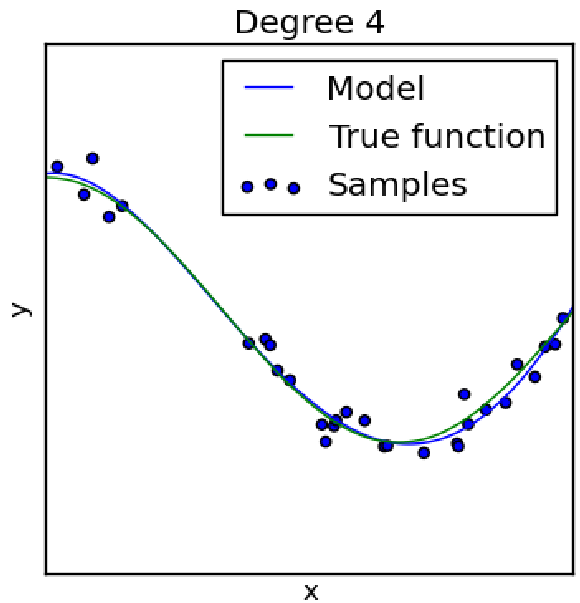


In machine learning, this is also called Bias/variance tradeoff

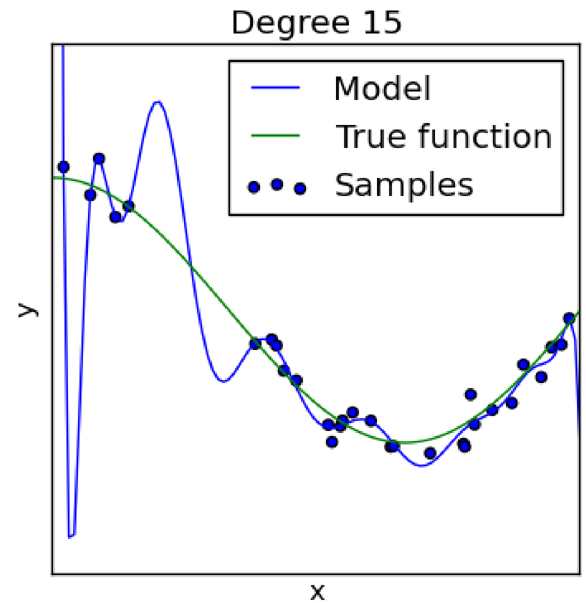
High bias
Low varience



Just Right



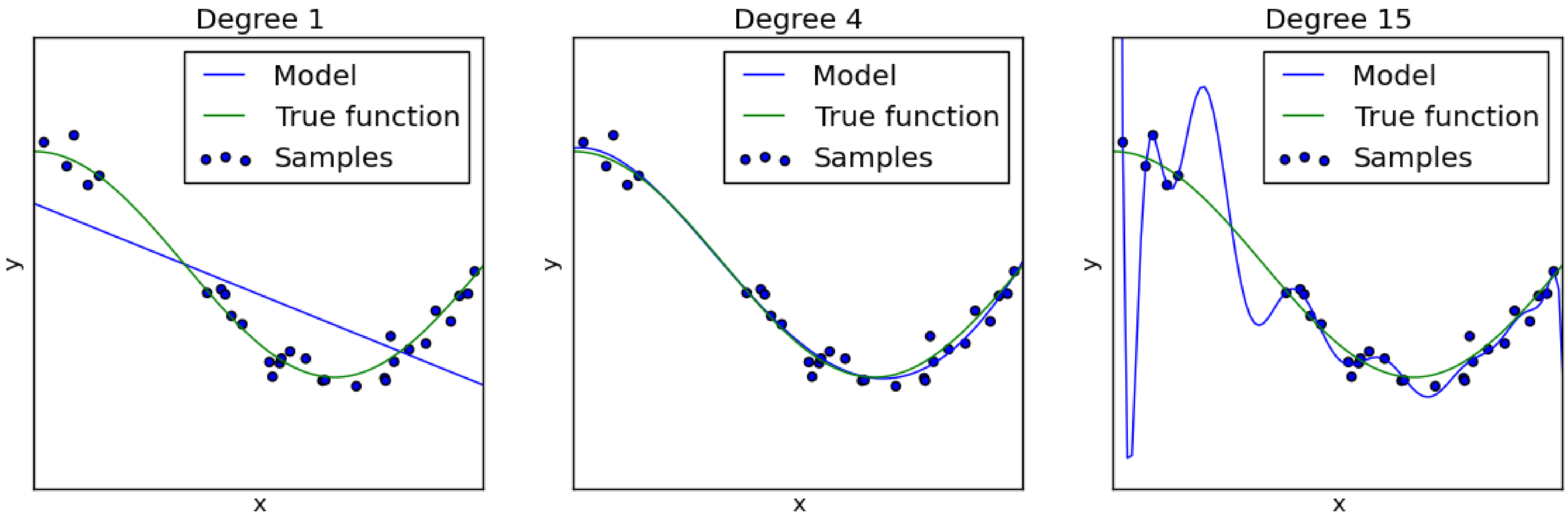
Low bias
High varience



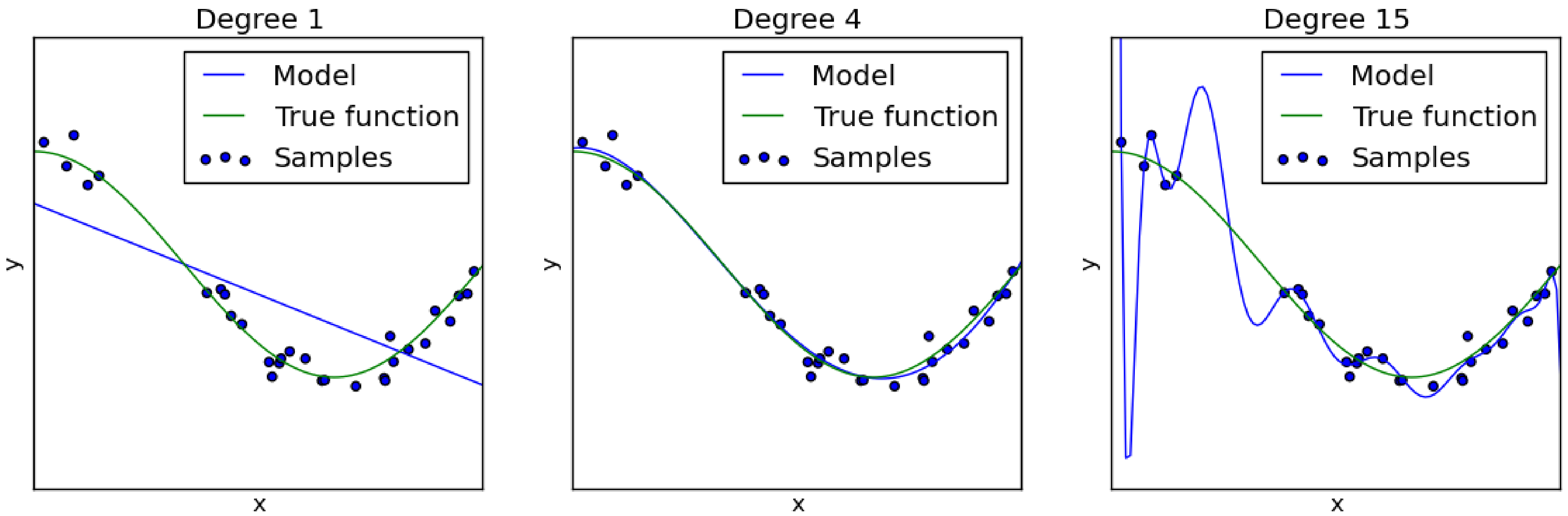
First training poorly, predictions bad

Just Right

First training very well, can't generalize

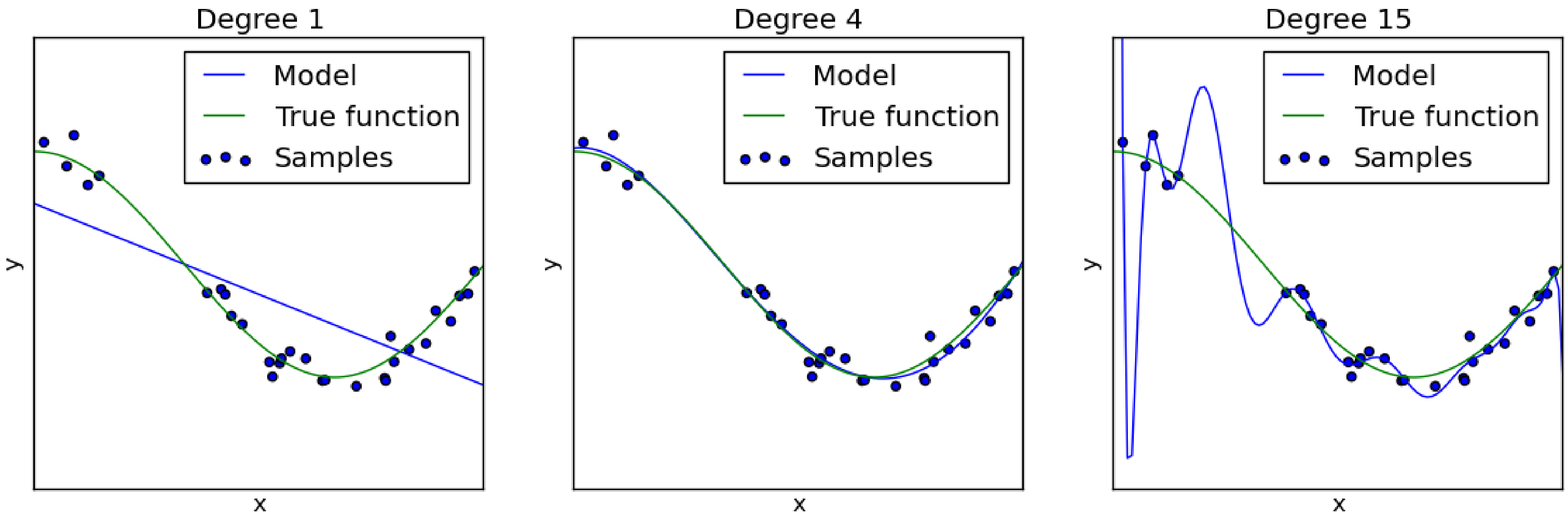


First and third will do poorly in the test set



Challenge: Fit a training set, calculate mean squared error on your test set (scikit learn)

There are a few metrics that try to measure this (without even looking at a test set yet)



OLS Regression Results

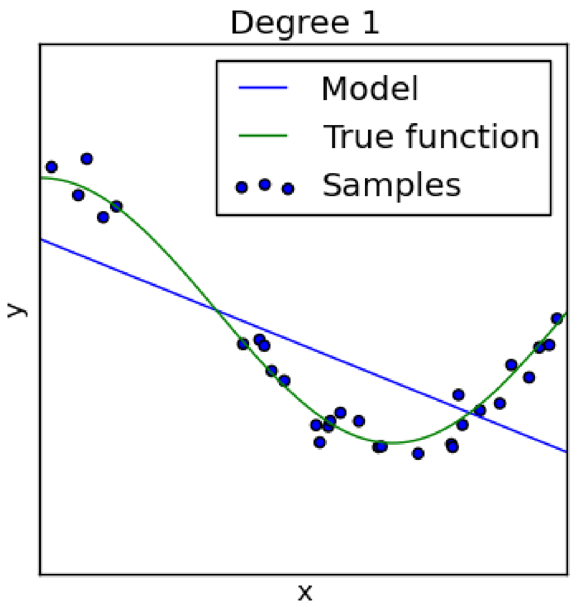
Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

Adjusted R^2

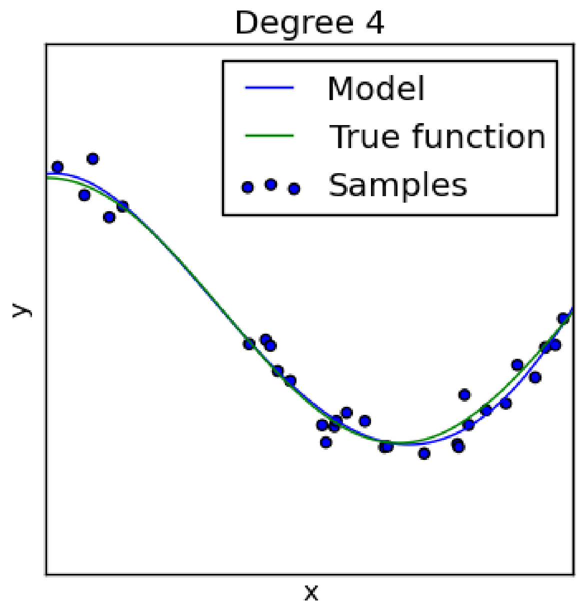
	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

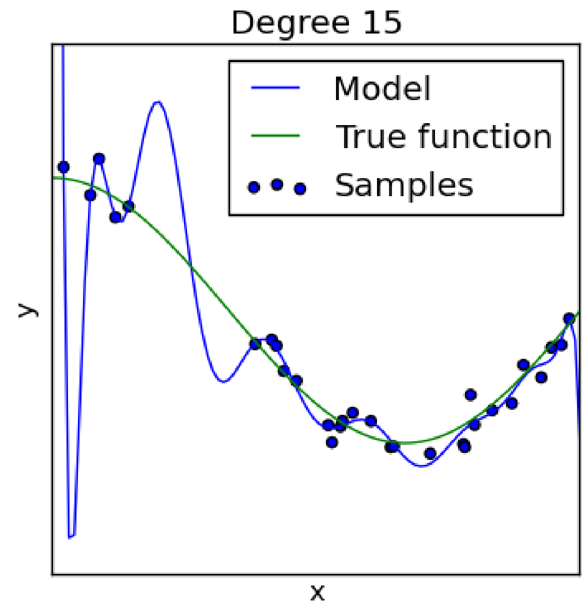
Low R^2



Higher R^2

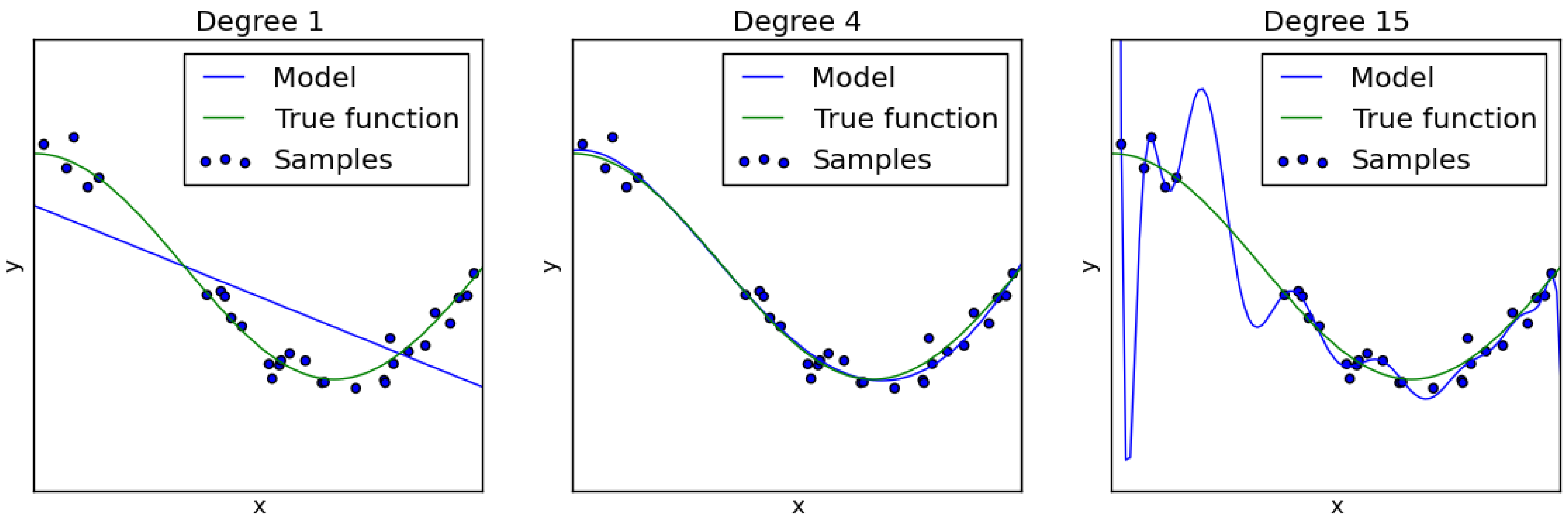


Highest R^2



$$\overline{R}^2 = 1 - \frac{SSE / df_e}{SST / df_t}$$

Low R^2 | Higher R^2 | Highest R^2



$$\bar{R}^2 = 1 - \frac{SSE / df_e}{SST / df_t}$$

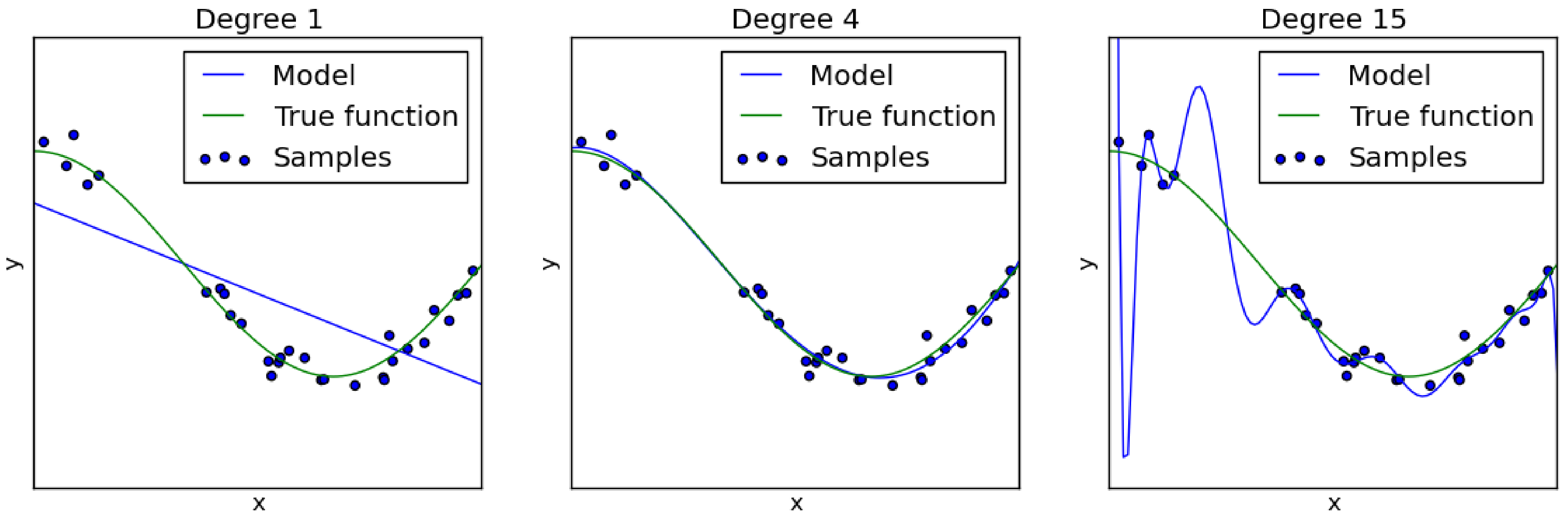
$m - k - 1$

$m - 1$

m= # points

k = # parameters

Low R^2 | Higher R^2 | Highest R^2



$$\overline{R}^2 = 1 - \frac{SSE / df_e}{SST / df_t}$$

$df_e \rightarrow m - k - 1$

$df_t \rightarrow m - 1$

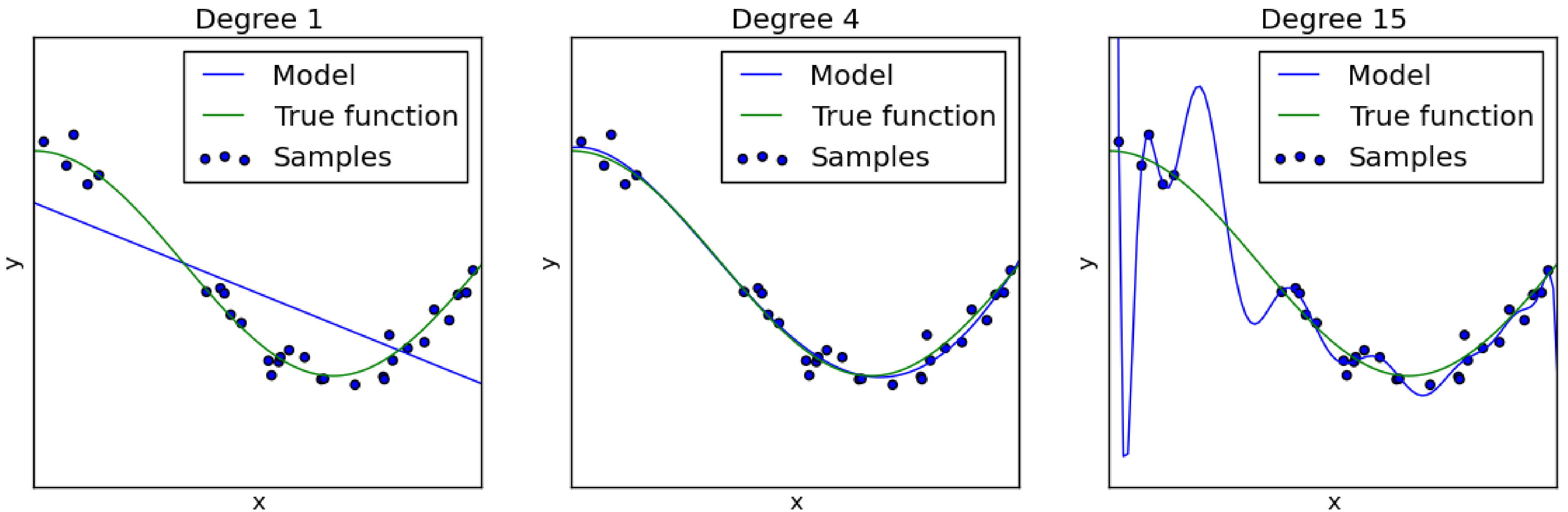
m= # points

k = # parameters

Low adj. R^2

Max. adj R^2

Low adj. R^2



OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

Akaike Information Criterion

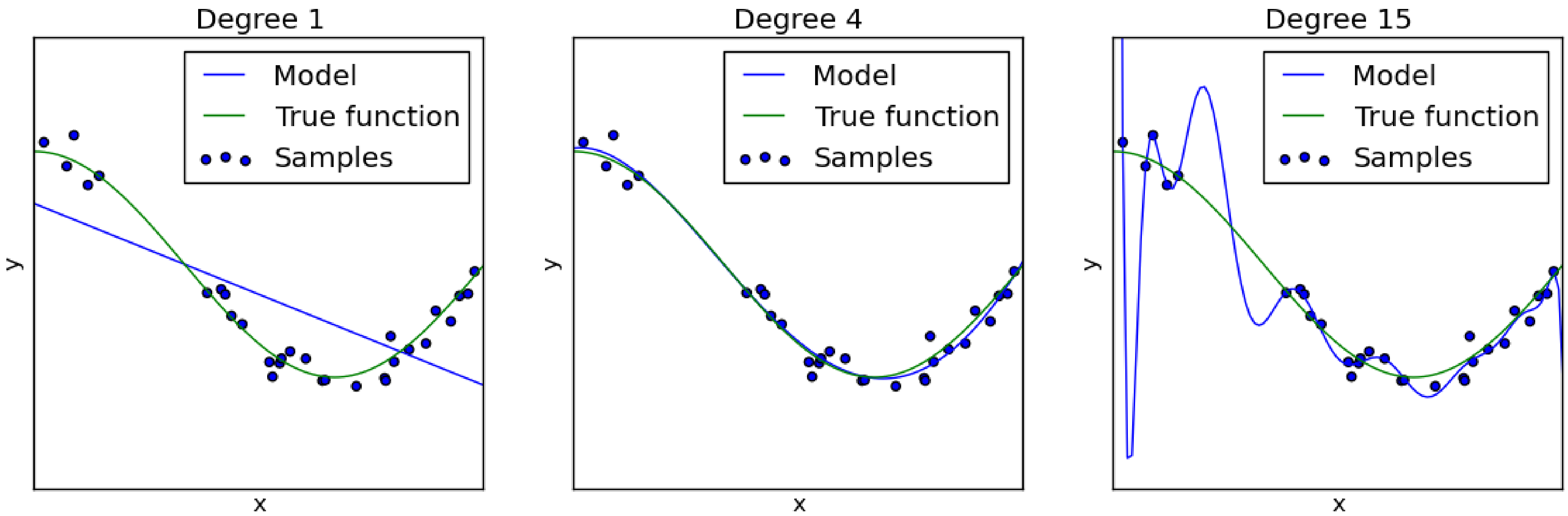
	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

$$AIC = 2k - 2\ln(L)$$

parameters

Log likelihood



$$AIC = 2k - 2\ln(L)$$

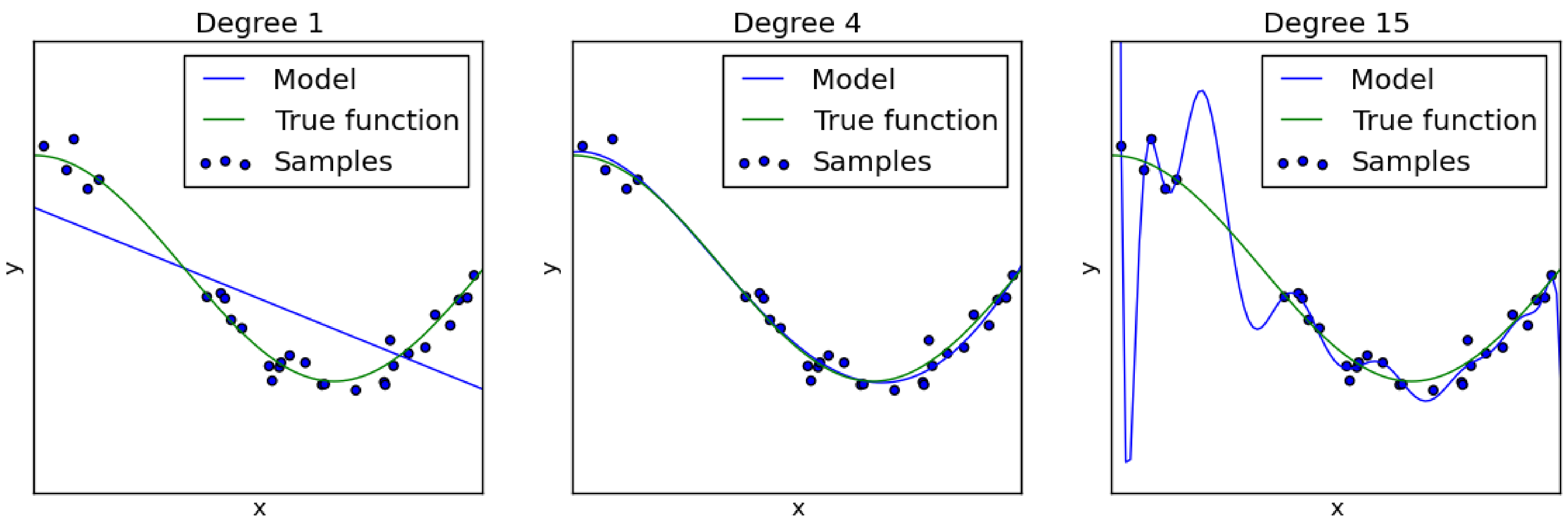
parameters

Log likelihood

Higher AIC

Min. AIC

Higher AIC



OLS Regression Results

Dep. Variable:	DomesticTotalGross	**R-squared:**	0.286
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	34.82
Date:	Sun, 14 Sep 2014	**Prob (F-statistic):**	6.80e-08
Time:	21:59:46	**Log-Likelihood:**	-1738.1
No. Observations:	89	**AIC:**	3480.
Df Residuals:	87	**BIC:**	3485.
Df Model:	1		

Bayesian Information Criterion

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Budget	0.7846	0.133	5.901	0.000	0.520 1.049
Ones	4.44e+07	1.27e+07	3.504	0.001	1.92e+07 6.96e+07

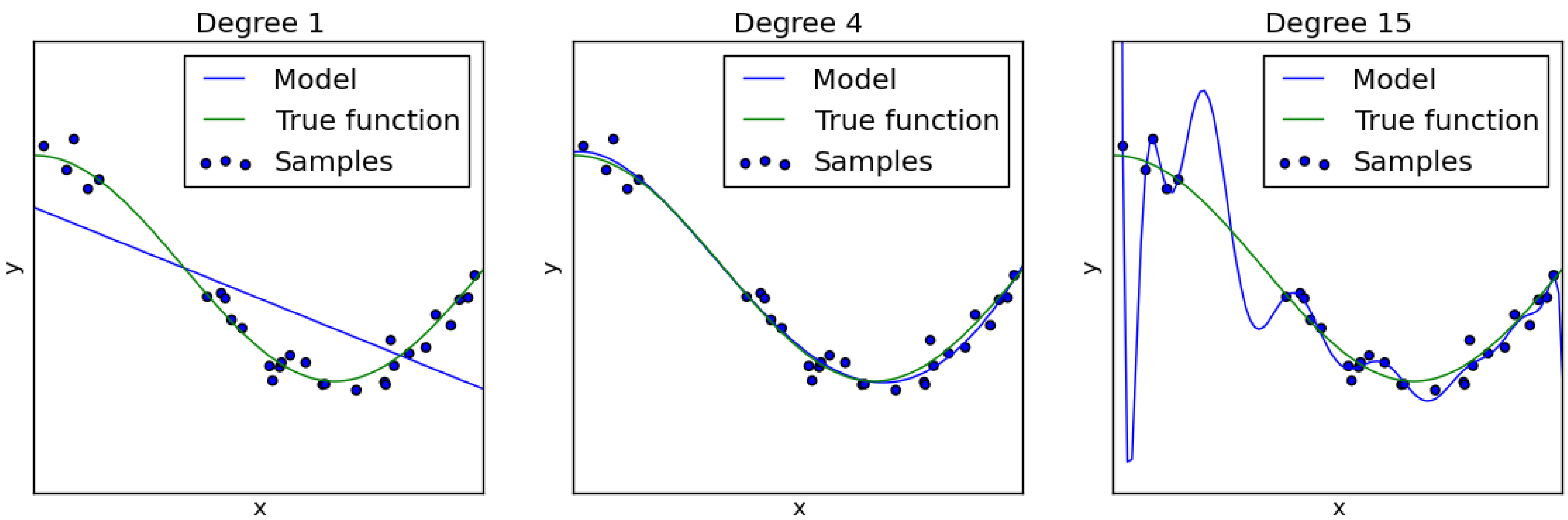
Omnibus:	39.749	**Durbin-Watson:**	0.674
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	99.441
Skew:	1.587	**Prob(JB):**	2.55e-22
Kurtosis:	7.091	**Cond. No.**	1.54e+08

$$BIC = k\ln(m) - 2\ln(L)$$

parameters

points

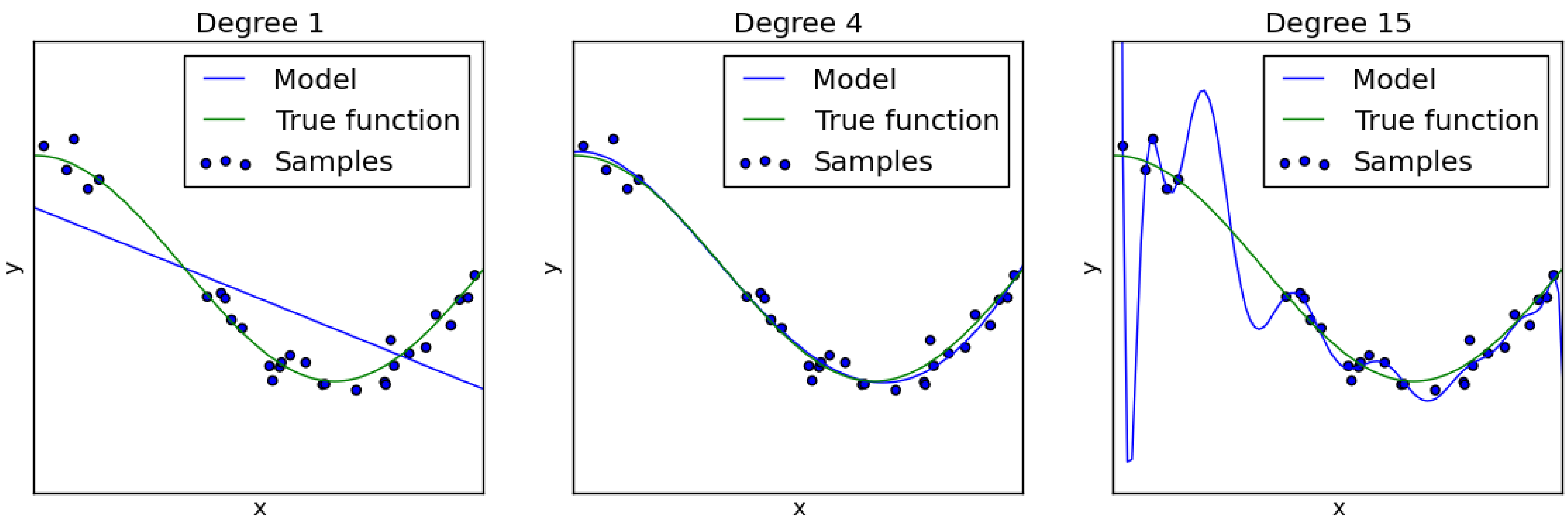
Log likelihood



$$BIC = k\ln(m) - 2\ln(L)$$

parameters   # points   Log likelihood

Higher BIC   Min. BIC   Higher BIC



My model is not awesome enough.

What do I do?

Try these and check test error
(and AIC,BIC,etc.) again:

Use a smaller set of features
Try adding polynomials
Check functional forms for each feature
Try including other features
Use more data (bigger training set)
Regularization (tomorrow)
Try some other model (later)