1.9 Exponential and Power Function Data Fitting

This section will further explore the methodology behind modeling a data set with exponential and power functions. We will explore cases where we have no insight into the data as well as cases where the problem statements allow us to gain some predictive knowledge as to which model may fit the problem best.

Example 1 *Assume that you are a data analyst for a research company. Your company receives contracts from numerous outside agencies that require an unbiased evaluation of data sets that they have collected. In order to remain completely unbiased, you specify that you want to see their raw data – and only their raw data. You are totally unaware of what agency you are working with, unaware of the experiment that the data was collected from, and unaware of any preconceived ideas of the model with which the contracted agency would like the data modeled. The data you receive is represented in Table 1.15.*

x	*f(x)*
-2	0.02
-1	0.08
0	1.7
1	2.9
2	4.2
3	7.4
4	12.5
5	21.4
6	39.0
7	57.2
8	88.1

Table 1.15: Example 1 (Raw Data)

Step 1: *Transform the problem.* We are **given** a table of data, we must examine the table.

- Define the variables:
  - Independent variable (input): *x*
  - Dependent variable (output): *f(x)*

- Nature of the data: Apply the quantitative measures of the nature of the data to gain an idea of which type of function may be most appropriate. See Table 1.16.

x	*f(x)*	AVG RoC
-2	0.02	
-1	0.08	0.06
0	1.7	1.62
1	2.9	1.2
2	4.2	1.3
3	7.4	3.2
4	12.5	5.1
5	21.4	8.9
6	39.0	17.6
7	57.2	18.2
8	88.1	30.9

Table 1.16: Quantitative Measure of the Nature of the Data

Notice that the data are increasing, as is the rate. Therefore, we need a function that is increasing at an increasing rate. We may model these data by either an exponential or a power function. Let's graph the data to verify our conjecture. See Figure 1.57.

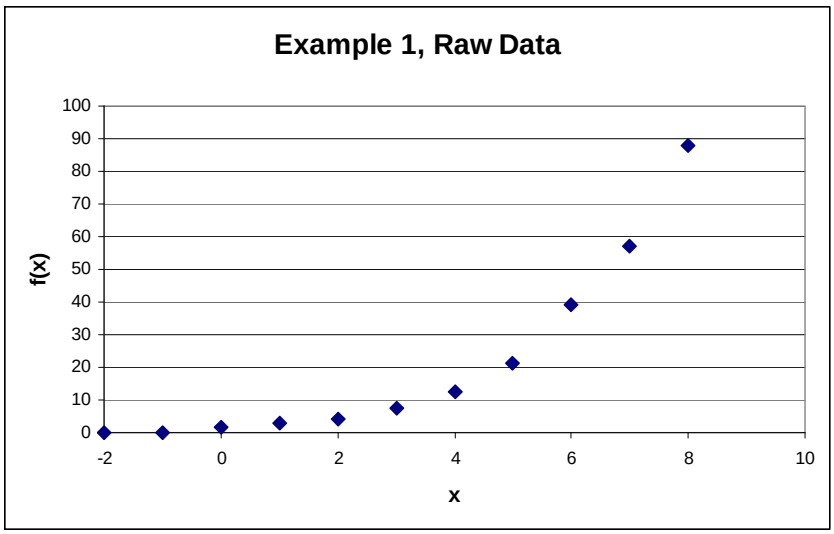

Figure 1.57: Scatterplot of Table 1.12 Data.

The data in Figure 1.57 seems to exhibit a pattern consistent with either an exponential or a power function. We know that we will have to **find** the model that fits through the data best. Because we have no knowledge as to what the data may do outside of the domain of the data in the given table, it would be wise to find a "good" exponential model **AND** a "good" power model. By completing both models, we can recommend to the agency that contracted our services which model fits the experiment from which the data came.

- Assumptions
  - Using an exponential function model of the form $f(x) = ab^x + d$, we can assume that the parameter $d = 0$ because there appears to be a horizontal asymptote at $f(x) = 0$.
  - Based on a power function model of the form $f(x) = a(x+c)^b + d$, a reasonable assumption seems to be that the vertex is at the first data point. Because the first data point is left of the origin by two units we will assume $c = 2$. Since the function value at the first data point is 0.02, we will assume that $d = 0.02$.

Our **plan** is to use the model development and evaluation techniques that we have used up to this point in the course to find the best models (exponential and power) for the data in Table 1.16.

Step 2. *Solve the problem using appropriate solution techniques.* To **solve**, we will develop a "good" exponential and a "good" power model. We will provide both options to our client.

- Exponential model development.

$$Model(x) = ab^x$$

- Exponential Model Parameter Estimation: Using modeling skills developed to date, we need to estimate the parameters (a and b) for an exponential model. To estimate two parameters, we need two data points that represent the data well. Let's try the fourth and tenth points: $(1, 2.9)$ and $(7, 57.2)$.

- Using Mathematica to solve for a and b yields,

```
Solve[{2.9 == a (b)^1, 57.2 == a (b)^7}, {a, b}]

{{a → -1.76427, b → -1.64374}, {a → -0.882135 - 1.5279 i, b → -0.82187 + 1.42352 i},
 {a → -0.882135 + 1.5279 i, b → -0.82187 - 1.42352 i},
 {a → 0.882135 - 1.5279 i, b → 0.82187 + 1.42352 i},
 {a → 0.882135 + 1.5279 i, b → 0.82187 - 1.42352 i}, {a → 1.76427, b → 1.64374}}
```

- To ensure exponential growth, we need to choose a and b to be positive. Therefore, our initial exponential model becomes:

$$Model(x) = 1.764(1.643)^x$$
$$Domain: \{x \mid x \in \Re\}$$
$$Range: \{Model(x) \mid 0 \le Model(x)\}$$

Figure 1.58 is a plot of the model, to see how well it appears to fit the data.

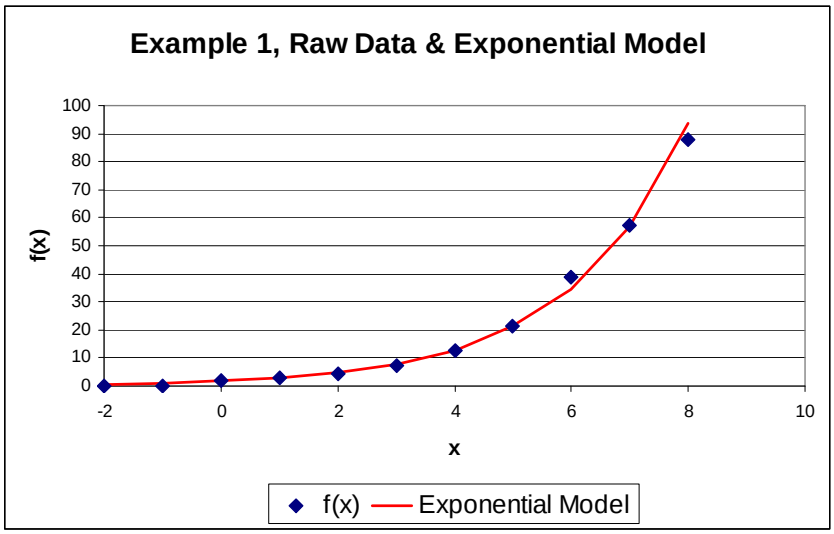

Figure 1.58: Table 1.16 Data Overlaid with Exponential Model

- Power model development.

$$Model2(x) = a\,(x+2)^b + 0.2$$

- Power Model Parameter Estimation: Using modeling skills developed to date, we need to estimate the parameters (a and b) for a power model. To estimate two parameters, we need two data points that represent the data well. Let's try the same points that we used for the exponential model: $(1, 2.9)$ and $(7, 57.2)$.

- Using Mathematica to solve for a and b yields,

```
Solve[{2.9 == a (1 + 2.9)^b + 0.02, 57.2 == a (7 + 2.9)^b + 0.02},
 {a, b}]

Solve::ifun :
  Inverse functions are being used by Solve, so some solutions may not be found;
     use Reduce for complete solution information. >>

Solve::svars : Equations may not give solutions for all "solve" variables. >>

{{a → 57.18 9.9^(-1. b)}}
```

- Note that this solution still shows parameter a in terms of parameter b, meaning we still must make an assumption for the b parameter.

- In this form, it is difficult to intuitively make a good estimation of what value b should take on. However, if we consider the basic power function cases, we can make a solid *estimation* for b. We see that the data increases more steeply than we would expect in a quadratic function. The bottom of the "cup" is also more flat than we would expect in a quadratic, both characteristics of a function with a higher power. Let's assume that the parameter $b = 3$. Using the tenth point and $b = 3$ yields:

$$57.2 = a(7+2)^3 + 0.02$$
$$a = \frac{(57.2 - 0.02)}{9^3} \approx .0784$$

- Therefore, our initial exponential model becomes:

$$Model2(x) = .0784(x+2)^3 + 0.02$$
$$Domain: \{x \mid x \in \Re\}$$
$$Range: \{Model(x) \mid 0 \le Model(x)\}$$

Let's plot our models, Figure 1.59, to see how well they appear to fit the data.

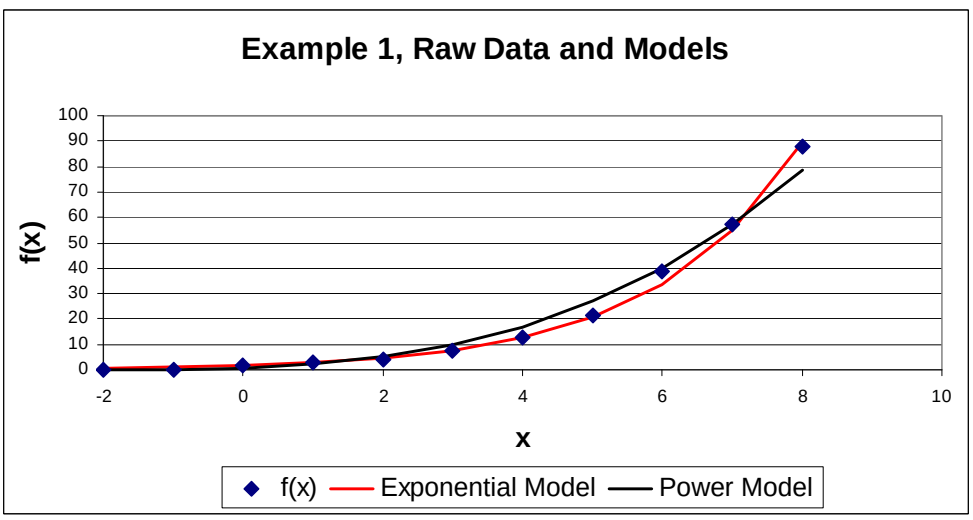

Figure 1.59: Table 1.12 Data Overlaid with Exponential & Power Models

Step 3. *Interpret the results of the solution.* Both models fit the trend of the data. The SSE for the exponential model is approximately 51.6 whereas the SSE for the power model is approximately 153.5. Currently, the exponential model seems to be the better model. Let's refine each model as best we can to provide the client with the best possible model for each type of function.

By manually adjusting the parameters for the exponential and power models, both SSEs can be reduced. With the exponential model, it seems that little can be done to lower SSE. In fact, we can only make minor adjustments to the a and b parameters to make the model less steep. The refined model is:

$$Model(x) = 1.76(1.635)^x .$$

The power model also can be transformed only by a little. If we make the curve steeper, it appears that it will fit the data a little better. To do this, we can change the exponent from a 3 to a 4. After doing that, we must decrease the a parameter because simply changing the b makes the curve too steep. Remember, refining models is an iterative process. The refined power model is:

$$Model2(x) = .00888(x+2)^4 + 0.02 .$$

You can see that, especially with power and exponential models, if good estimations are made at the beginning of the modeling process, then small parameter adjustments may be all that are necessary to improve models.

The refined exponential and power models have SSEs of approximately 50.04 and 23.46, respectively. The plot of both models follows.

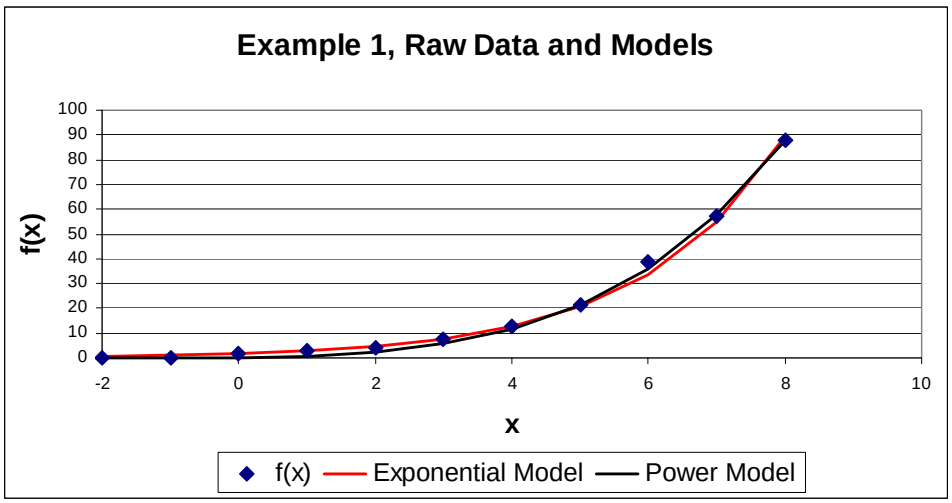

Figure 1.60: Comparison of Power and Exponential Models

We have determined the following: both models fit the trend of the data for the domain within the data set and both models increase at an increasing rate as the domain values increase past the values in the data set. The biggest difference in the models is for domain values that are less than -2. The exponential model will continue to decrease toward 0 as the domain values become more negative whereas the power model will start to "turn up" as the domain values become more negative. For domain values less than -2, the power model will become the reflection of the values greater than -2 about the line $x = -2$. Does this consideration affect your decision on which model to recommend? If you know where the data came from it might; if there were possibly negative values less than two then you would have to consider if the corresponding y values increased or decreased.

The two models that we developed appear to be good models, but are they the best possible models? Do they have the minimum possible SSE? The next subsection will help us to answer that very question!

The data for the following question is also linked on the course website.

Question 1 *Estimate your best possible exponential and power models for the following average salary data for professional basketball players in the years 1984-2004. Which model better represents the nature of the data? Is either model better for extrapolating outside of the domain of the data? Explain your conclusions.*

Year	AVG Salary
1984-85	$330,000
1985-86	$382,000
1986-87	$431,000
1987-88	$502,000
1988-89	$575,000
1989-90	$717,000
1990-91	$927,000
1991-92	$1,100,000
1992-93	$1,300,000
1993-94	$1,500,000
1994-95	$1,800,000
1995-96	$2,000,000
1996-97	$2,300,000
1997-98	$2,600,000
1998-99	$3,000,000
1999-2000	$3,600,000
2000-01	$4,200,000
2001-02	$4,500,000
2002-03	$4,546,000
2003-04	$4,917,000

Table 1.17: NBA Salary Data

1.9.1 Minimizing Sum of Squared Error

We have discovered a measure to *quantify* how far your model deviates from the actual data, the Sum of Squared Error. Since can quantify a model's "goodness," it is natural to want to develop the *best possible* model. Let's use technology to help us find better models, more quickly than we could find through trial and error.

Previously, we were intent on finding a model that predicted the value of the dependent variable of a situation, given the value of the independent variable. The models we have been exploring in this section resulted in the fitting of a data set with the following two models:

$$ExponentialModel: \quad Model(x) = 1.76(1.635)^{x}$$

$$PowerModel: \quad Model2(x) = .00888(x+2)^{4} + 0.02$$

Using the techniques mentioned earlier in this section, we determined the SSE of the exponential model to be about 50.04 and the power model to be about 23.46 (feel free to calculate this yourself, for additional practice). Since SSE is a relative measure, we also know that the power model we estimated appears to be a better model than the exponential; it certainly fits the data better over the domain of the data set. Are we done? Is this good enough? Can we do better?

You could probably continue to adjust the parameter values parameters to get smaller and smaller SSE values, but this would be tedious and we would not know if we truly had the best combination of parameter values to call it the "BEST" model. As it turns out, we have a powerful ally in our quest to find a low SSE in *Excel*. This ally is a tools add-in called Solver. What does Solver do you ask? Well... it solves! *(If you dig deep into the help files of Excel, you find that it uses something called the Simplex method to solve/optimize linear equations and a reduced gradient algorithm to solve/optimize non-linear equations – but that is a* touch *beyond the scope of this course).*

Before we begin to use Solver, we must first load it onto your laptop. To load Solver, you must first open up *Excel*. In your Tools pull down menu, select Add-Ins. Another screen will appear with several options as to which Add-In you wish to install. You must select the Solver Add-In as shown in Figure 1.61. Ensure you also select the first Add-in, the Analysis ToolPak.

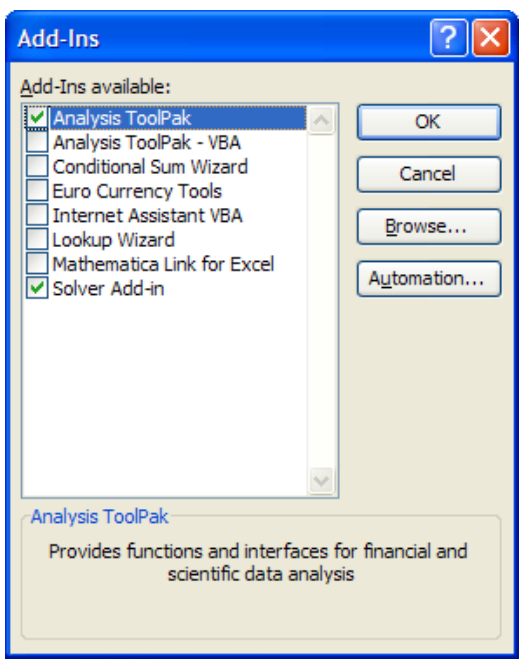

Figure 1.61: Add-In Dialog Box

Once we have Solver loaded, we can work on the problem at hand, that is, finding the "BEST" or minimum SSE.

In the Tools menu, you now have an option to run Solver. When you select this application, you get the Solver dialog box, shown in Figure 1.62.

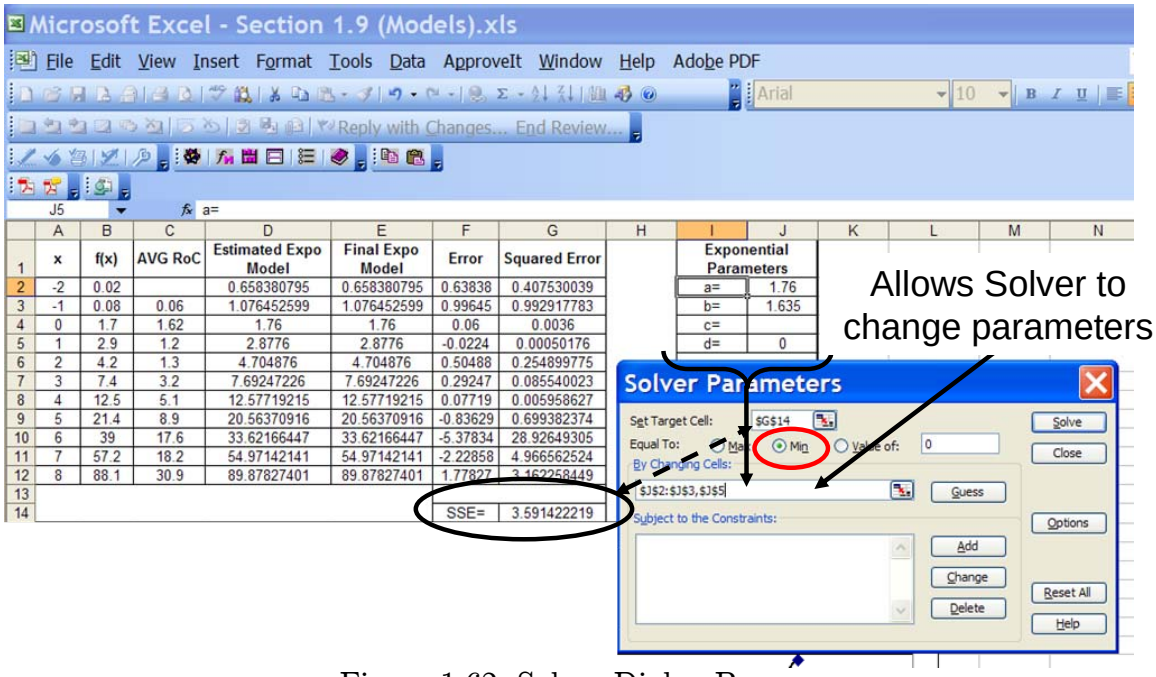

Figure 1.62: Solver Dialog Box

When you were adjusting the parameter values, what was your goal? To minimize the SSE, right? Solver's first input is to set the target cell. The target cell is the cell that you want to minimize, maximize, or go to a specific value. In this case, we wish to minimize the cell that contains the SSE. As a matter of practice, do not have Solver attempt to achieve a value of 0, this is most likely an unachievable goal.

When you were adjusting the parameter values to get a low SSE, you were changing specific cells to help you achieve that goal. In order for Solver to find a "best" solution, we must tell it which cells on the spreadsheet it can change in order to find the minimum SSE. In this case, we want to allow Solver to change the parameter values. Now we just need to tell it to Solve. You may notice that in Figure 1.64, there is a $ before each reference in the Solver parameter box. Take some time to look up **absolute referencing** in the *Excel* help files to Figure out what this means. *Your use of absolute referencing is vitally important to your modeling efforts.*

Solver will usually find its best solution quite quickly. Before Solver writes the solution it obtained onto your spreadsheet, you will be asked if you want to change to the new solution that was just obtained, or restore the original values with which you started. For this example, we will accept the Solver values. See Figure 1.63 for the *better* exponential solution. Our eyeball test shows that that the final model appears to be a closer fit to the data. Note that the already small SSE was cut by more than two-thirds. Both of our model evaluation tests agree: the final model fits the data better than the estimated model.

M	N	O	P	Q	R	S	T	U	V
x	f(x)	Final Exponential Model	Squared Mean Error	Model Error	Squared Model Error				
-2	0.02	-0.69681766	453.6125	0.716818	0.513828				
-1	0.08	-0.05649564	451.0604	0.136496	0.018631		a	2.805963	
0	1.7	0.93219453	384.8731	0.767805	0.589525		b	1.544051	
1	2.9	2.45878307	339.2294	0.441217	0.194672		c	0	
2	4.2	4.81591438	293.0321	-0.61591	0.379351		d	-1.87377	
3	7.4	8.45544652	193.7158	-1.05545	1.113967				
4	12.5	14.0750716	77.76033	-1.57507	2.48085				
5	21.4	22.752062	0.006694	-1.35206	1.828072				
6	39	36.1497821	312.6467	2.850218	8.123742				
7	57.2	56.8365519	1287.505	0.363448	0.132095				
8	88.1	88.7779897	4459.811	-0.67799	0.45967				
	21.31818	SST:	8253.253	SSE:	15.8344				
				r^2:	0.998081				

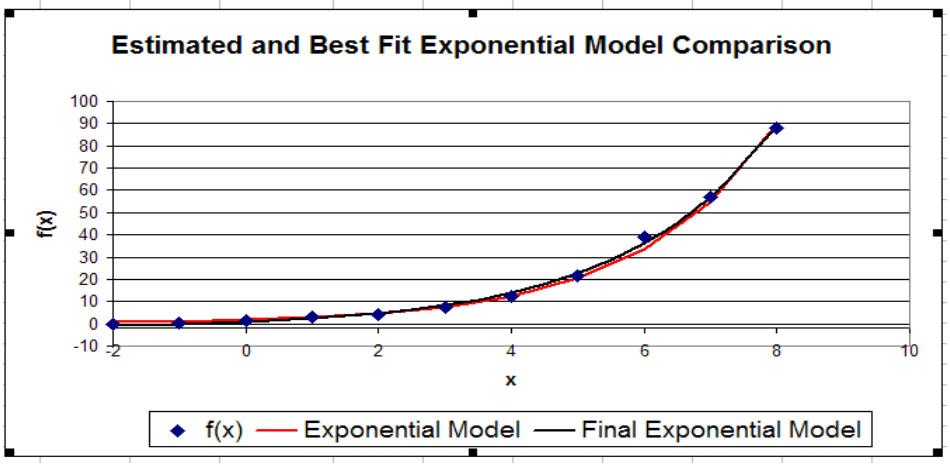

Figure 1.63: Model Post-Solver

After using Solver, the exponential model of *best* fit is:

$$Model(x) = 2.806(1.544)^x - 1.874$$
$$Domain: \quad \{x \mid -2 \le x \le 8\}$$
$$Range: \quad \{y \mid 0 \le y \le 88.1\}$$

Are we done? Can we do better with our model? Run Solver again to see if it results in an even more refined answer. Occasionally, with models that are a little more complex, Solver will not always find an optimal solution. It is best to give Solver a couple of chances to see if it "converges" to an answer within its specified tolerance. Another important point is that often, Solver needs initial parameter values that are reasonable before it can optimize.

Since starting with different parameter values may result in different SSEs for complex models, care should be taken in selecting the initial start or "guess" values. You should select start values that allow a reasonably good fit between your model and the data which is why the parameter estimation techniques you've been taught thus far are of vital importance. In other words, if you put "garbage" in, you get "garbage" out.

The data for all of the following problems are linked on the course website.

Question 2 *Develop a best fit model for the power function modeled in Example 1. **HINT** Do not allow Solver to optimize the b parameter, the exponent. The result will be a decimal (root) function which will provide undefined answers for negative values of (x+c). You may manually change the exponent, say, from a 3 to a 4. Is the power model or exponential better? Why?*

Question 3 *The following table represents the average life expectancy for females in the United States[1] since 1900. Model the data with an exponential and power model. Which model best reflects the trend of the data for the 20th century? Which model is better for extrapolating female life expectancies into the 21st century? Explain.*

Year	1900	1909	1919	1929	1939	1949	1959	1969	1979	1989	2002
Life Expectancy (US Females)	50.7	53.2	57.4	60.9	65.9	71	73.2	74.6	77.6	78.8	79.9

***Question 4** *LoggerPro, a computer-based data collection tool used by the Mathematics Department at USMA, collected the following potential measurements within a capacitor. If you are interested in what a capacitor is and how it works, check out http://electronics.howstuffworks.com/capacitor.htm.*

Model the data with an exponential and power model. Which model best fits the data within the specified domain? Which model do you think would be best to extrapolate data for domain values that are greater than those provided? Explain.

Time (s)	Potential (V)	Time (s)	Potential (V)	Time (s)	Potential (V)
0	1.5995	30	0.2271	60	0.0464
2	1.4286	32	0.2027	62	0.0415
4	1.2527	34	0.1783	64	0.0317
6	1.0867	36	0.1587	66	0.0317
8	0.9499	38	0.1441	68	0.0269
10	0.8327	40	0.1197	70	0.0269
12	0.7350	42	0.1148	72	0.0366
14	0.6374	44	0.1001	74	0.0220
16	0.5592	46	0.0904	76	0.0220
18	0.4957	48	0.0855	78	0.0317
20	0.4322	50	0.0708	80	0.0269
22	0.3785	52	0.0659	82	0.0269
24	0.3346	54	0.0562	84	0.0171
26	0.2906	56	0.0611	86	0.0269
28	0.2662	58	0.0513	88	0.0220

[1] Source: <http://www.ncseonline.org/nle/crsreports/05mar/RL32792.pdf>