Population Project

Steven Turner*

January 10, 2017

Contents

1 Introduction

For our project, we will fit the populations of the United States (US) and Munich, Germany (Munich) to three different models for population to see which is best. The three models we will fit out data to are the exponential (1), logistic (2), and Gompertz models (2).

$$P(t) = Ae^{rt} \qquad \text{exponential} \tag{1}$$

$$P(t) = \frac{L}{1 + Ae^{-rt}} \qquad \text{logistic} \tag{2}$$

$$P(t) = Le^{-be^{-rt}} \qquad \text{Gompertz} \tag{3}$$

Where L is the carrying capacity, $P(t)$ is the population, and t is time. Carrying capacity is the amount of population that can be accommodated due to restrictions of resources. Resources such as food and water, but can also be things such as the area a population has to settle and climate of the region the population is settling. These equations are derived from the following differential equations

*Thanks to Sam (I think) for giving me the idea to separate my results section

$$\frac{dp}{dt} = rp \qquad \text{exponential} \tag{4}$$
$$\frac{dp}{dt} = r(1 - \frac{p}{L})p \qquad \text{logistic} \tag{5}$$
$$\frac{dp}{dt} = rp \log\left(\frac{L}{p}\right) \qquad \text{Gompertz} \tag{6}$$

To compare the differences in our three models, we will not be looking at the solutions ($P(t)$). Instead, we will be looking at the differential equations. Specifically we will be looking at the per unit population growth rate (PPGR); The PPGR is $\frac{1}{p}\frac{dp}{dt}$. PPGR is the amount of growth contributed by one unit of population. For the comparison we will be looking at the $\lim\limits_{p\to 0} \text{PPGR}$ and $\lim\limits_{p\to L} \text{PPGR}$ for all three models. What these limits will tell us is how fast the population (p) is growing per person as the population gets tiny and as the population gets to the limit of their resources (carrying capacity L).

We will start by analyzing the exponential model (1). For the exponential model we have

$$\text{PPGR} = \frac{1}{p}\frac{dp}{dt} = r.$$

Thus the two limits for the exponential model are

$$\lim_{p\to L} \text{PPGR} = r$$
$$\lim_{p\to 0} \text{PPGR} = r.$$

What these two limits mean is that no matter the size of the population, it will be growing at the same rate. The rate not changing is mostly a problem for the limit as p goes to L. Since as the population gets close to it carrying capacity its PPGR is not slowing which could lead to overpopulation.

For the logistic model, we have the following PPGR and limits

$$\text{PPGR} = r(1 - \frac{p}{L})$$
$$\lim_{p\to L} \text{PPGR} = 0$$
$$\lim_{p\to 0} \text{PPGR} = r.$$

Here our limit as the population goes to 0 is r just like the exponential model. However, as the population approaches its carrying capacity its PPGR slows to

0. This seems more realistic sense a population wouldn't want to overburden its resources.

Lastly, we will look at the PPGR and the limits of the Gompertz model. The PPGR and limits are

$$\begin{aligned} \text{PPGR} &= r \log\left(\frac{L}{p}\right) \\ \lim_{p \to L} \text{PPGR} &= 0 \\ \lim_{p \to 0} \text{PPGR} &= \infty. \end{aligned}$$

Like the logistic model as our population approaches L our PPGR slows to 0. However, unlike the other two models as our population gets tiny, our growth rate is infinite. In the physical world, this makes no sense if you say a population grows infinitely fast, but you can think instead of your pop having almost limitless resources to develop with so it can expand at an alarming rate.

Of the three models, I predict that the population of Munich and the US will be best modeled by the logistic model. I predict the logistic model will be the better model is because the logistic models PPGR does not have a limit of infinity. Despite being able to interpret what the infinite limit means, I don't think it is realistic for a population to be growing that fast.

2 Methods

We will be graphing the three models using matplotlib for Python. For our initial estimates for our parameters for the three models we used numpy's `polyfit` to estimate linear parameters by linearizing our models, and to make them better we used the `curve_fit` from the scipy library to get our final parameters. Linearizing means getting the three models into form $y = mx + k$ so that `polyfit` can make linear guess for our initial parameters.

To linearize the exponential model we do the following

$$\begin{aligned} \ln(P(t)) &= \ln(Ae^{rt}) \\ &= \ln(A) + rt. \end{aligned}$$

Where $\ln(P(t))$ is our y and t (or time) is our x. `polyfit` estimates the parameters $m = r$ and $k = \ln(A)$.

To linearize the logistic model we do the following

$$
\begin{aligned}
P &= \frac{L}{1+Ae^{-rt}}\\
\frac{L}{P} &= 1+Ae^{-rt}\\
\frac{L}{P}-1 &= Ae^{-rt}\\
\ln\left(\frac{L}{P}-1\right) &= \ln(A)-rt.
\end{aligned}
$$

Here our $y = \ln\left(\frac{L}{P}-1\right)$ and $x = t$, and `polyfit` estimates our $m = r$ and $k = \ln(A)$.

Last but not least is the linearization of the Gompertz model which is

$$
\begin{aligned}
P &= Le^{-be^{-rt}}\\
\frac{L}{P} &= e^{be^{-rt}}\\
\ln\left(\frac{L}{P}\right) &= be^{-rt}\\
\ln\left(\ln\left(\frac{L}{P}\right)\right) &= \ln(b)-rt.
\end{aligned}
$$

Where $y = \ln\left(\ln\left(\frac{L}{P}\right)\right)$, $x = t$, and from `polyfit` $m = r$ and $k = \ln(b)$.

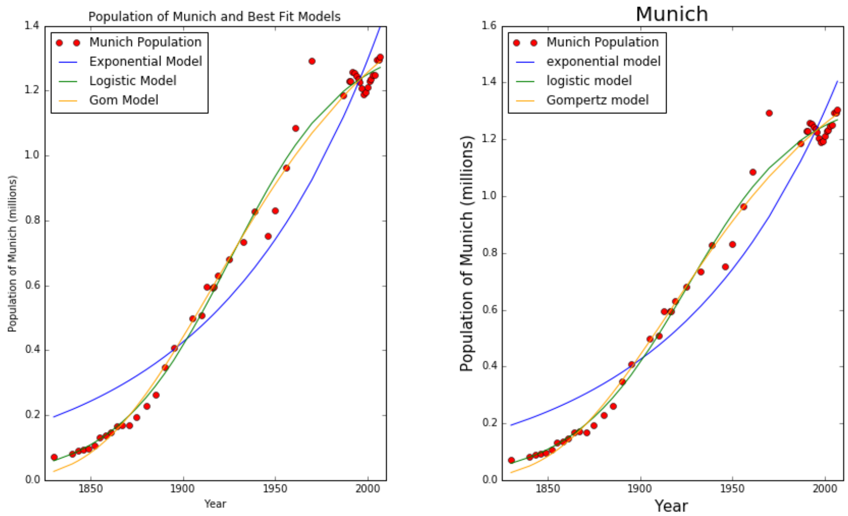
For the carrying capacity of Munich, I used 1.5 million. This was a guess based on the population in the data wik never going above 1.4 million. For the carrying capacity of the US, we used 350 Million appears to be tapering off below that. It should also be noted that once `curve_fit` is used it will take the initial L's and adjust them.

To evaluate the models, we are going to use the Leave One Out (LOO) method. We will be using the R^2 error (from LOO) and the error from the prediction of the last point to judge how good the models are. R^2 error is a measure of the goodness of fit of a model with numbers close to one being "good" and close to 0 being "bad."

3 Results: Munich

Fit	Munich data; Population in Millions (from LOO)		
Exponential	$A = 0.1935$	$r = 0.0111$	
Logistic	$L = 1.359$	$A = 21.96$	$r = -0.0323$
Gompertz	$L = 1.612$	$b = 4.100$	$r = -0.0164$
L is in millions			

Munich Population Data; Pop in Millions (from LOO)				
Fit	R^2 error	Population 2007	Prediction 2007	Percent Error
Exponential	0.9383	1.3055	1.4120	7.56%
Logistic	0.9909	1.3055	1.2662	-2.83%
Gompertz	0.9898	1.3055	1.2881	-1.26%



(a) This is graph the three original models and the data that was used to generate them.

(b) This is graph the three models using the LOO method and the data minus the last point.

Figure 1: The Munich models and data

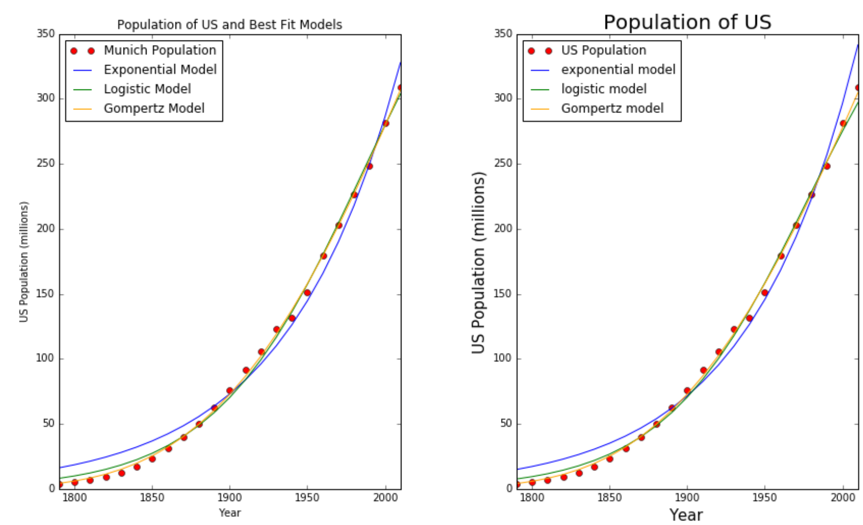
Our initial results for fitting the models to the Munich data is shown in figure 1a. What we can see is that the exponential overshoots our guess for carrying capacity which is a problem that was discussed earlier. Both the logistics model and Gompertz model appear to fit the data quite well, with both flattening out as we get closer to the carrying capacity as shown with the PPGR and figure 1a. By observation, it is interesting that the logistic model fits the data quite well up until WW2.

Figure 1b is the graph of out LOO method. With the LOO method, we still see the exponential model still overshooting our data. As well as, the flatting of the other two models much like figure 1a. As shown in the table above we can see the R^2 error for the exponential model is much lower than the other two R^2 errors. The reason for this is because the exponential has one fewer parameter to fit the data with it. By both the combination of R^2 error and prediction error I would say that for the population of Munich that the Gompertz model is the better of the three models.

4 Results: US

Fit	US data; Population in millions (from LOO)		
Exponential	$A = 15.05$	$r = 0.01418$	
Logistic	$L = 444.19$	$A = 56.42$	$r = -0.0215$
Gompertz	$L = 1274.68$	$b = 5.718$	$r = -0.006297$

United States data (from LOO)				
Fit	R^2 error	Population 2010	Prediction 2010	Percent error
Exponential	0.9837	308.7	341.4	10.58%
Logistic	0.9971	308.7	296.9	-3.81%
Gompertz	0.9989	308.7	304.7	-1.28%



(a) This is graph the three original models and the data that was used to generate them.

(b) This is graph the three models using the LOO method and the data minus the last point.

Figure 2: The US models and data

As we can see in figure 2a much like figure 1a the exponential model overshoots our data for the US towards the end. Again much like figure 1a for the US models in figure 2a we can see that the logistic and Gompertz models both follow the data very closely.

Much like the data for Munich if we look at the LOO models and the original model for the US the graphs are very similar. Also like the Munich population model based on the R^2 error and prediction error the Gompertz model is the best fit for the US population.

5 Conclusion

As we can clearly see from the combination of R^2 and closeness of prediction the Gompertz model is the best model for predicting both populations. The Gompertz being the better model is not what I predicted when starting to do the population modeling. Gompertz being the better model is interesting to me because it has the infinite limit which makes doesn't make a lot of physical sense.

That being said I don't think any of these models would be good at predicting the population of a region over a longer period due to immigration and emigration. This is especially relevant during years with volatile political climates. Think the US now and the Middle East and the refugee crisis. Also as global warming causing sea levels to rise is another way that emigration could change the population. The region that comes to mind first is Florida.

Another thing that makes these models inferior is the fact that they always increase. This makes it hard for them to account for a decrease in the population that happens for whatever reason.

References

Munich. URL `https://en.wikipedia.org/wiki/Munich`.