<u>NEAS Time Series – Student Project</u>
By David Broomhead

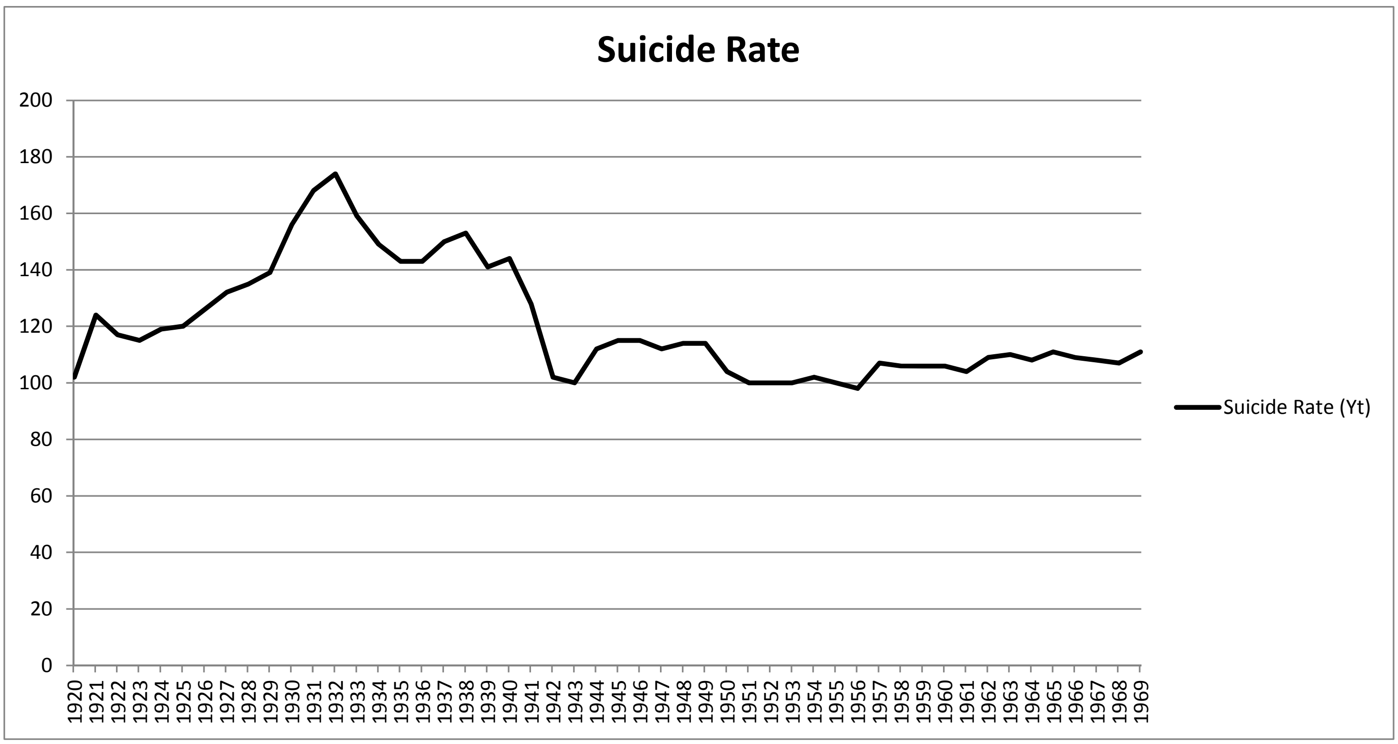
Session: Spring 2015

<u>US Suicide Rates</u>

Please refer to the excel file "Time Series Project – Suicide Rates – David Broomhead.xls" for details.

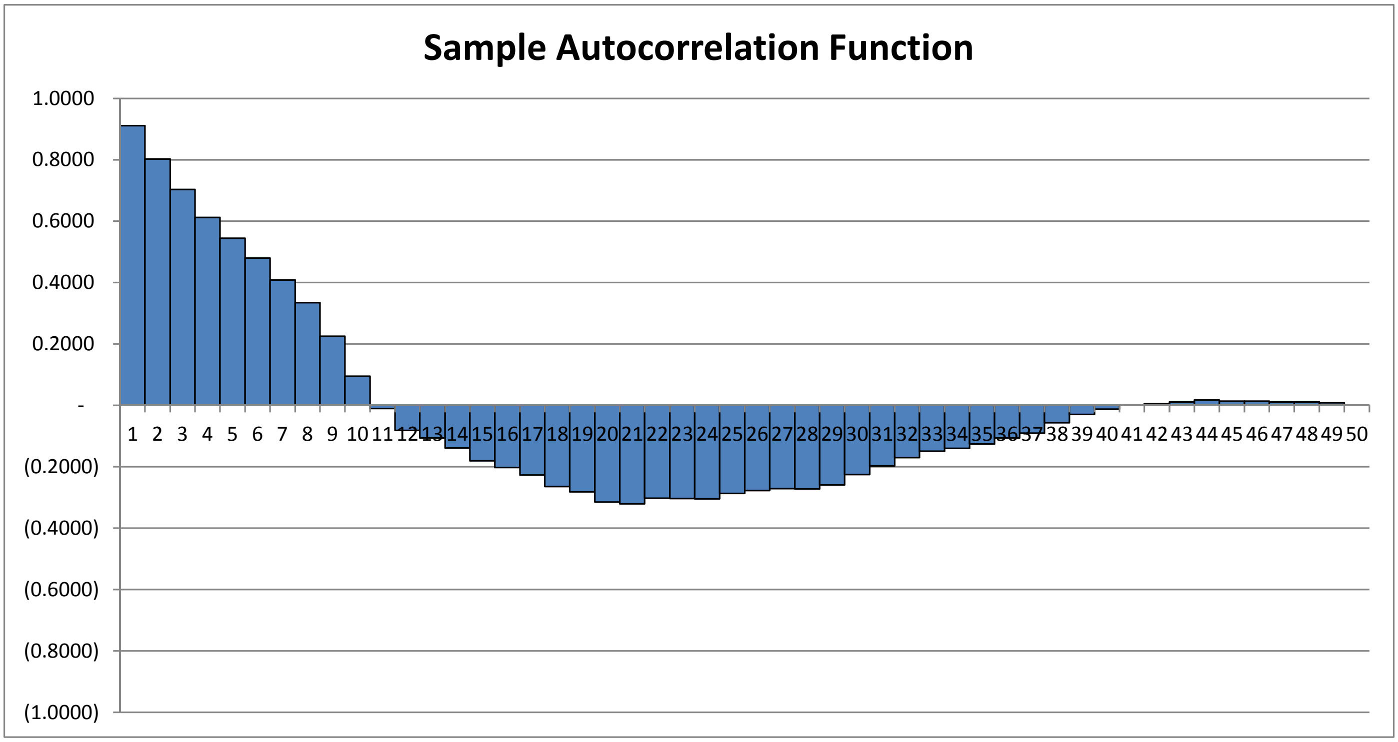
Intro:
I will analyze annual US suicide rate data for 1920-1969 pulled from datamarket.com and determine whether the historical data is reasonable to predict future suicide rates (as of 1969).

Data:
The data is shown in cells B14:B63 of the excel file, and represents the number of suicides per 100,000 people between 1920 and 1969 inclusive, in the United States. The excel file has a link to the source.



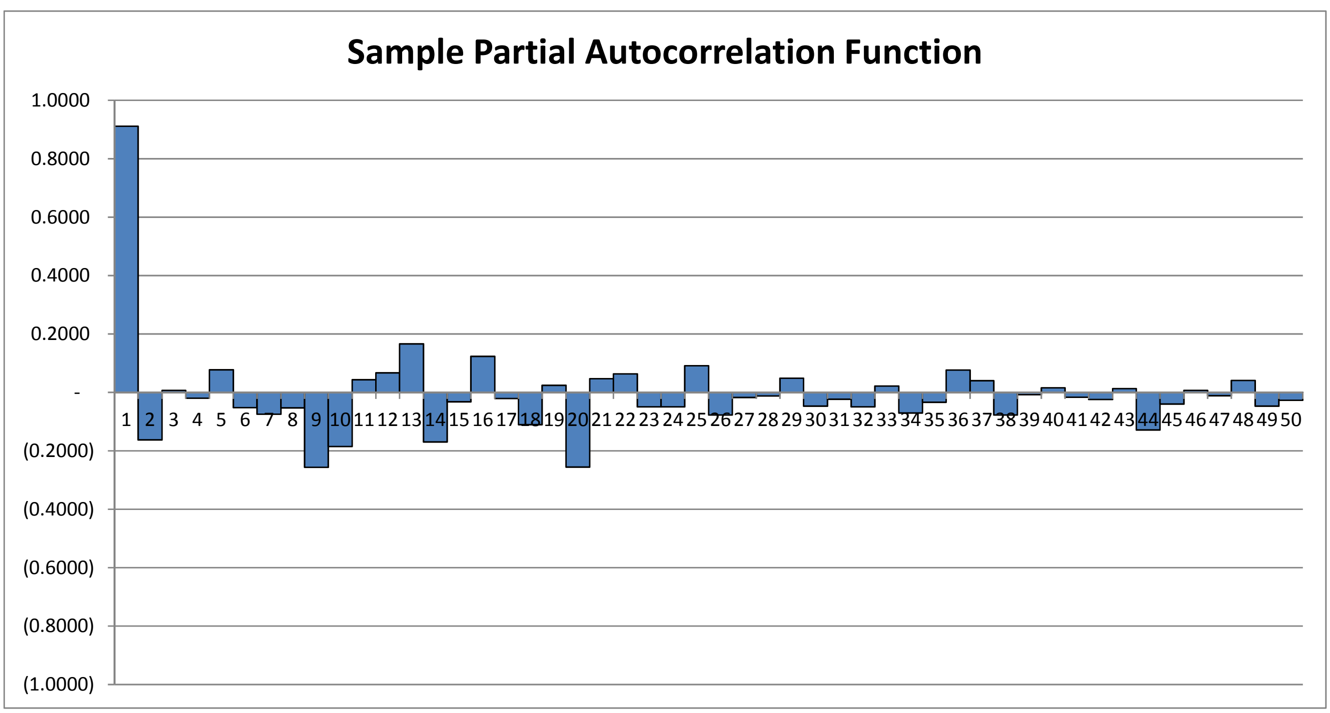
Analysis:
The data shows a period of increased suicide rate in the late 1920s to late 1930s, which coincided with the Great Depression.

The Sample Autocorrelation is calculated in the excel file and the graph is shown below:



There is significant autocorrelation at lag 1 of 0.91 which decays until lag 10 and then goes negative, increasing back to near 0 by lag 41. There is little to suggest that the data is not stationary, so I will assume that it is stationary. The lag pattern is consistent with an autoregressive (AR) model as the autocorrelation decays without cutting off (as would be seen in a moving average (MA) model). Accordingly, I will test AR(1) and AR(2) models.

The Partial Sample Autocorrelation Function is calculated in excel (using formula 6.2.9) and the graph is shown below.



The partial sample autocorrelation for lag 2 is insignificant, which suggests that an AR(2) model is likely not ideal, as an AR(p) model's PACF will cut off at lag p. This suggests that the AR(1) model is likely to be more appropriate.

AR(1):
$\mu = 120.54$
$r_1 = 0.9113$

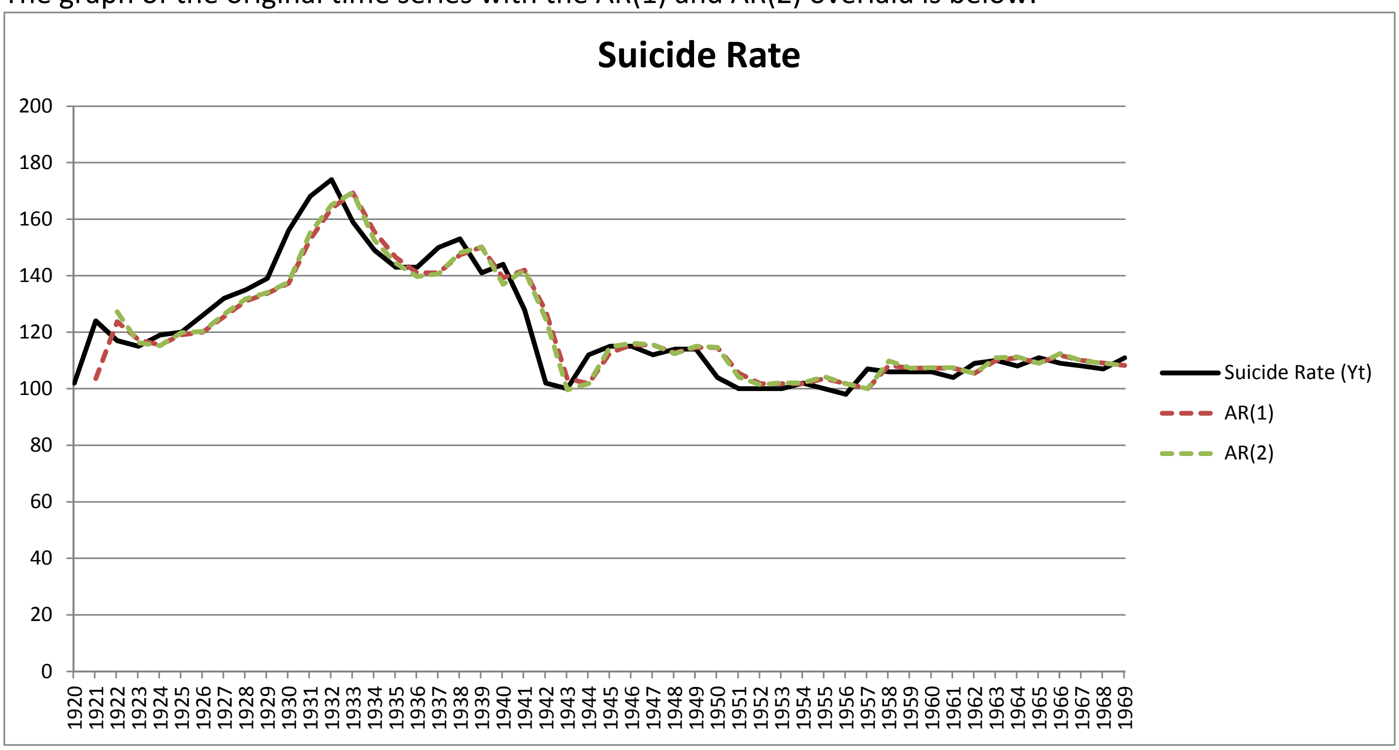
Using the method of moments gives $\phi = r_1 = 0.9113$. Our AR(1) model is thus given as:
$Y_t = \mu(1-\phi) + \phi Y_{t-1} = 10.70 + 0.9113Y_{t-1} + e_t$

AR(2):
$\mu = 120.54$
$r_1 = 0.9113$
$r_2 = 0.8028$

Using the method of moments gives $\phi_1 = r_1 (1-r_2)/(1-r_1^2) = 1.0598$, and $\phi_2 = (r_2 - r_1^2)/(1-r_1^2) = -0.1630$. Our AR(2) model is thus given as:
$Y_t = \mu(1-\phi_1 - \phi_2) + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} = 12.44 + 1.0598Y_{t-1} - 0.1630Y_{t-2} + e_t$

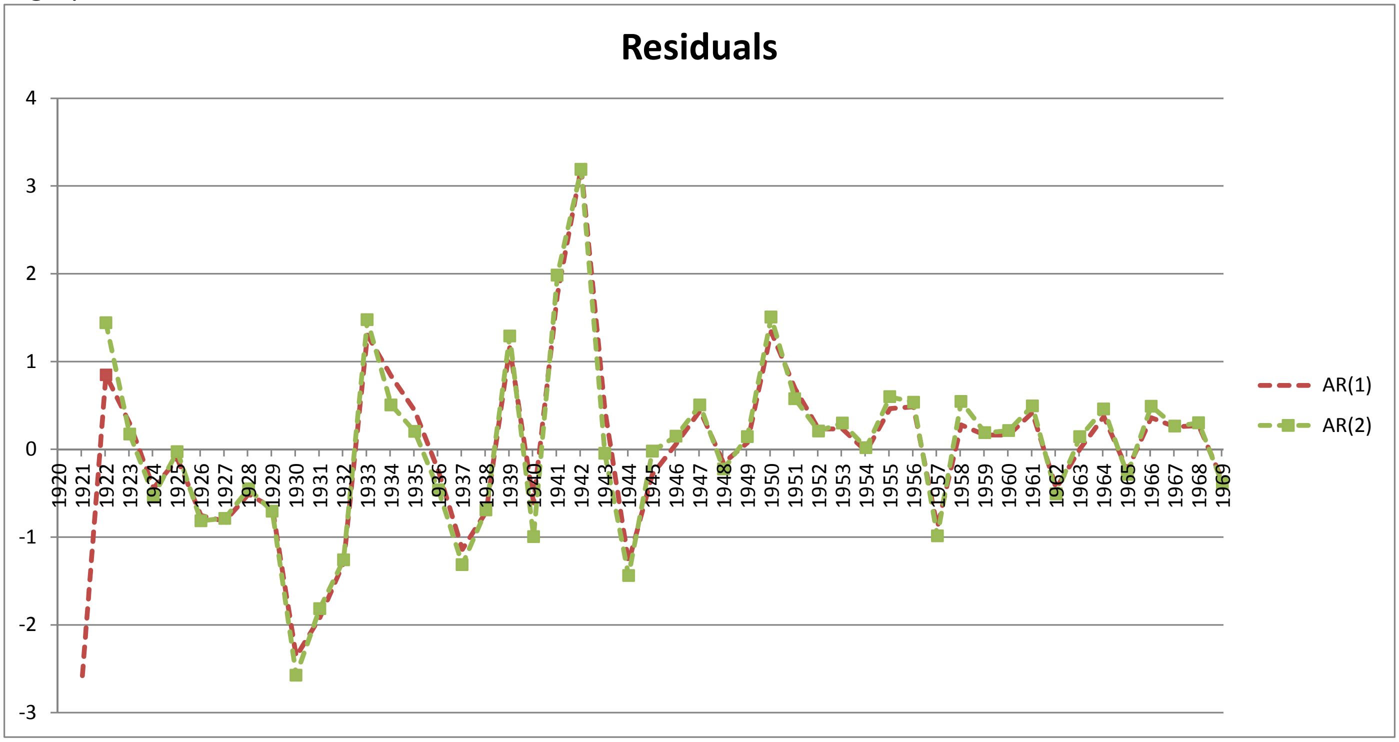
We can see that ϕ_2 is close to 0, which is consistent with the observation that the partial sample autocorrelation was insignificant at lag 2. This suggests that the AR(2) model does not represent a significant refinement over the AR(1) model.

The graph of the original time series with the AR(1) and AR(2) overlaid is below:



We can see that both the AR(1) and AR(2) models provide a reasonable fit to the historical data, but that the AR(2) model is not a significant improvement over the AR(1) model.

A graph of the residuals is below:



Again, we see that the residuals for both the AR(1) and AR(2) models are similar. Principle of parsimony suggests we should use the AR(1) model since is it less complex and the AR(2) model is not a significant improvement.

Conclusion:

I would use an AR(1) to fit this data and to predict future suicide rates.

The final selected model was: $Y_t = 10.70 + 0.9113Y_{t-1} + e_t$