Yidong Wang
Time Series Fall 2015
Student Project

GINI Index in the United States

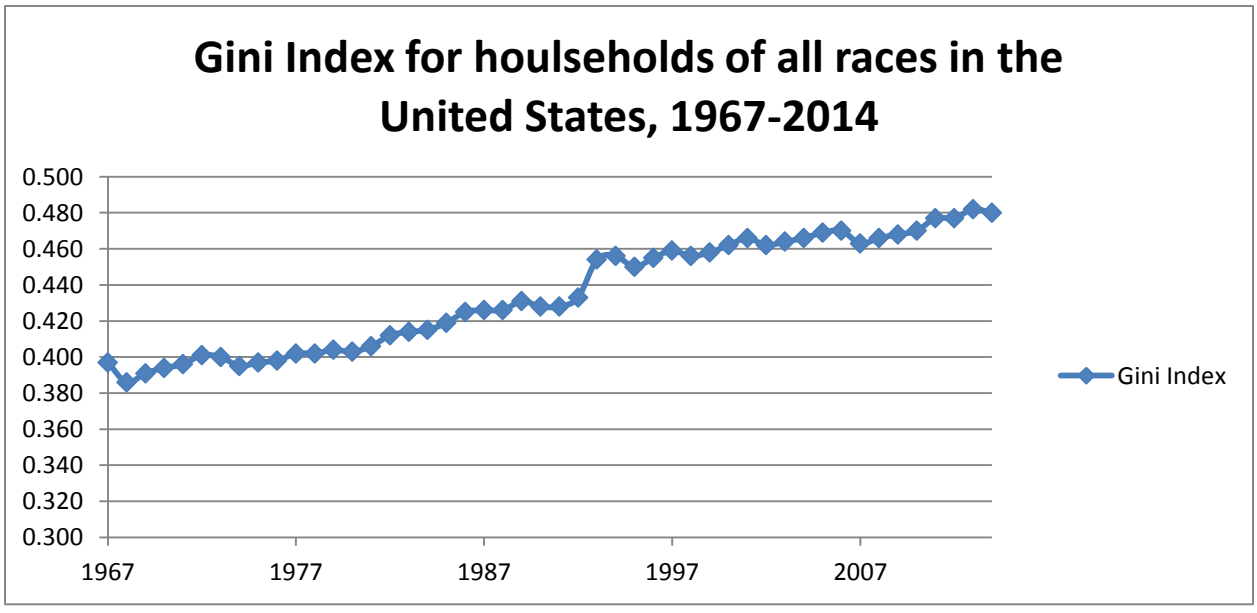
Introduction

Gini index is a measure of statistical dispersion developed by the Italian statistician and sociologist Corrado Gini and published in his 1912 paper "Variability and Mutability".[1] A Gini index of zero indicates perfect equality where every group of population has an exactly equal income. A Gini index of one means extreme gap among different group of population that one group owns all the wealth in the society.

Worldwide, Gini indexes for income range from about 0.23 (Sweden) to 0.71 (Namibia) in 140 countries. Wealth inequality has been an increasing issue in the United States and wealth ownership has long been concentrated in the hands of a small minority of the population. Recent trends in wealth inequality have been particularly startling. The Gini-coefficient of the U.S. in 2010 is approximately 0.47, indicating a relatively large income gap. This study will examine different time series model for Gini index trend in the U.S. since 1967.

Gini indexes for households, 1967-2014

Our data comes from the U.S. Census Bureau and has the Gini index of the United States from 1967 to 2014, as shown below.

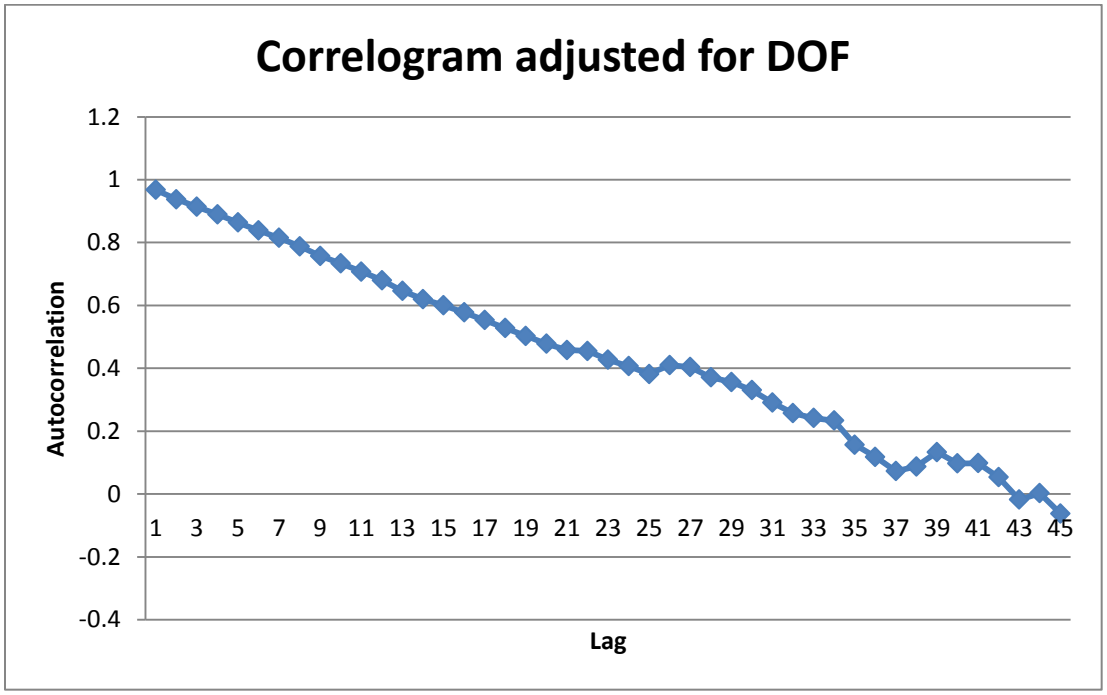


Source: U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplements

[1] Wikipedia contributors. "Gini coefficient." *Wikipedia, The Free Encyclopedia,* http://en.wikipedia.org/w/index.php?title=Gini_coefficient&oldid=489292482 (accessed May 2, 2012).

Methodology and Analysis

Autocorrelation



An autocorrelation plot of Gini index shows that the sample autocorrelations are very strong and positive, decaying steadily from lag 1 to lag 25 and decaying towards zero after that.

My study will examine three orders of Autoregressive models, AR(1), AR(2) and AR(3).

Autoregressive (AR) models

Using the regression tool in MS Excel Office, the study tests the three models mentioned above.

AR(1) Model

Regression Statistics	
Multiple R	0.988280067
R Square	0.976697491
Adjusted R Square	0.976179657
Standard Error	0.004704094
Observations	47

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.041737068	0.041737068	1886.122505	2.20995E-38
Residual	45	0.000995783	2.21285E-05		
Total	46	0.042732851			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.003461022	0.009968509	0.34719555	0.730062997	-0.016616587	0.023538631	-0.016616587	0.023538631
X Variable 1	0.99609068	0.022935802	43.42951191	2.20995E-38	0.949895603	1.042285756	0.949895603	1.042285756

The R Square is about 0.98, showing great explaining power of this model. Although the intercept is not significant, the coefficient of y(t-1) is very significant judging from the p-values.

This regression results into the following equation

$$Y_t = 0.9961Y_{t-1} + 0.0035$$

It shows the positive correlation between the Y(t) and Y(t-1) and suggests an increasing trend of Gini index since 1968, which proves our assumption earlier.

AR(2) Model

Regression Statistics	
Multiple R	0.989777186
R Square	0.979658877
Adjusted R Square	0.978712779
Standard Error	0.004363149
Observations	46

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	0.03942471	0.019712355	1035.472148	4.26639E-37
Residual	43	0.000818594	1.90371E-05		
Total	45	0.040243304			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.008580047	0.009499716	0.90318985	0.371457204	-0.010577956	0.027738049	-0.010577956	0.027738049
X Variable 1	0.949720867	0.139239154	6.820788844	2.3402E-08	0.668918352	1.230523382	0.668918352	1.230523382
X Variable 2	0.035381992	0.141027402	0.250887353	0.803096216	-0.24902687	0.319790854	-0.24902687	0.319790854

This model has similar R Square, also showing great explaining power of the whole model. However, the p-value of Y(t-2) is about 0.80, suggesting that Y(t-2) does not correlate with Y(t) significantly. Similarly, the p-value of the intercept is also not small enough to be significant.

Therefore, AR(1) is a better model so far.

This regression results into the following equation

$$Y_t = 0.9497Y_{t-1} + 0.0354Y_{t-2} + 0.0086$$

AR(3) Model

Regression Statistics	
Multiple R	0.990353495
R Square	0.980800046
Adjusted R Square	0.979395171
Standard Error	0.004225803
Observations	45

ANOVA

	df	SS	MS	F	Significance F
Regression	3	0.037400957	0.012466986	698.1405785	3.30972E-35
Residual	41	0.000732154	1.78574E-05		
Total	44	0.038133111			

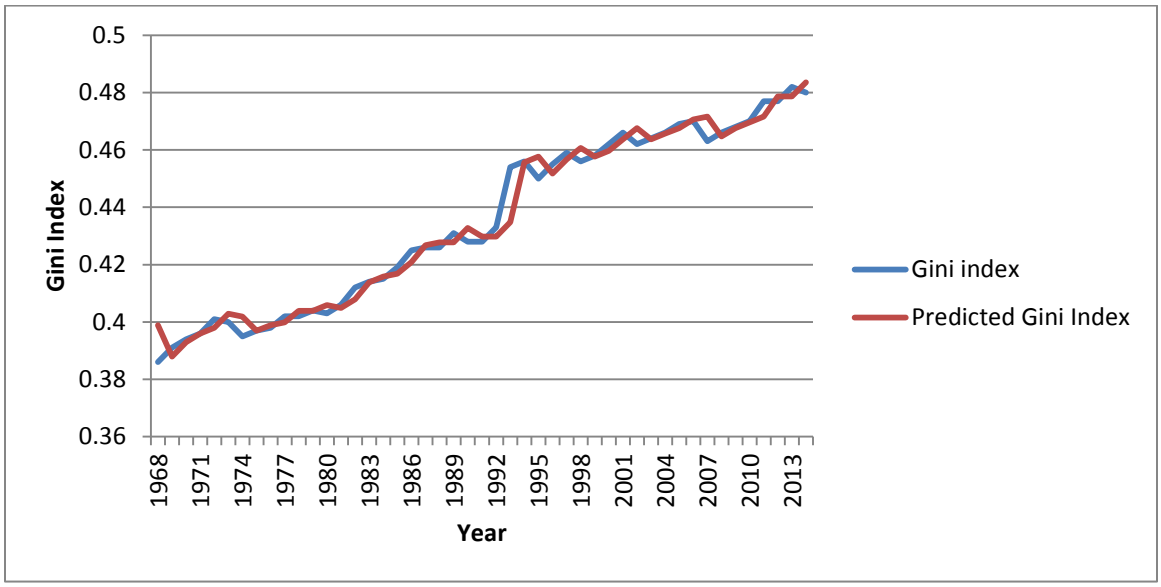
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.005364818	0.009485642	0.565572483	0.574766763	-0.013791825	0.024521461	-0.013791825	0.024521461
X Variable 1	0.969839162	0.148728947	6.520850051	7.8602E-08	0.66947496	1.270203364	0.66947496	1.270203364
X Variable 2	-0.270435435	0.196996536	-1.372792847	0.177279432	-0.66827801	0.12740714	-0.66827801	0.12740714
X Variable 3	0.294135219	0.136955569	2.147668917	0.037700577	0.017547836	0.570722603	0.017547836	0.570722603

AR(3) model is very similar to AR(2) model in that the p-value for Y(t-2) and the intercept are both not significant. Y(t-1) shows strong correlation with Y(t), as expected. Additionally, Y(t-3) also shows a positive correlation with Y(t). The equation is shown as below.

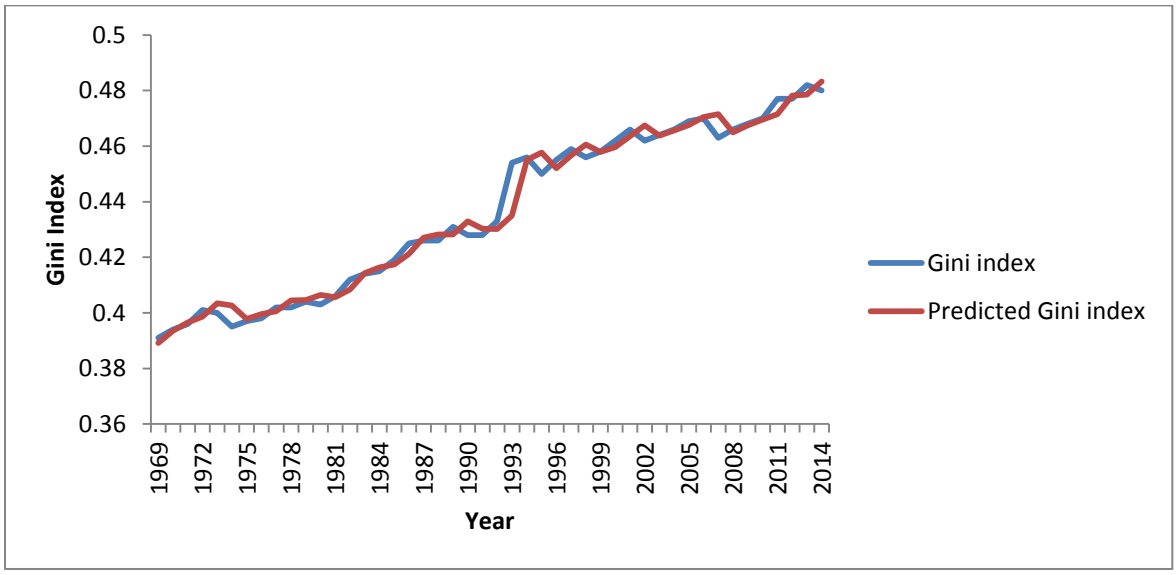
$$Y_t = 0.9698Y_{t-1} - 0.2704Y_{t-2} + 0.2941Y_{t-3} + 0.0054$$

Actual vs Predicted Gini Index for AR Models

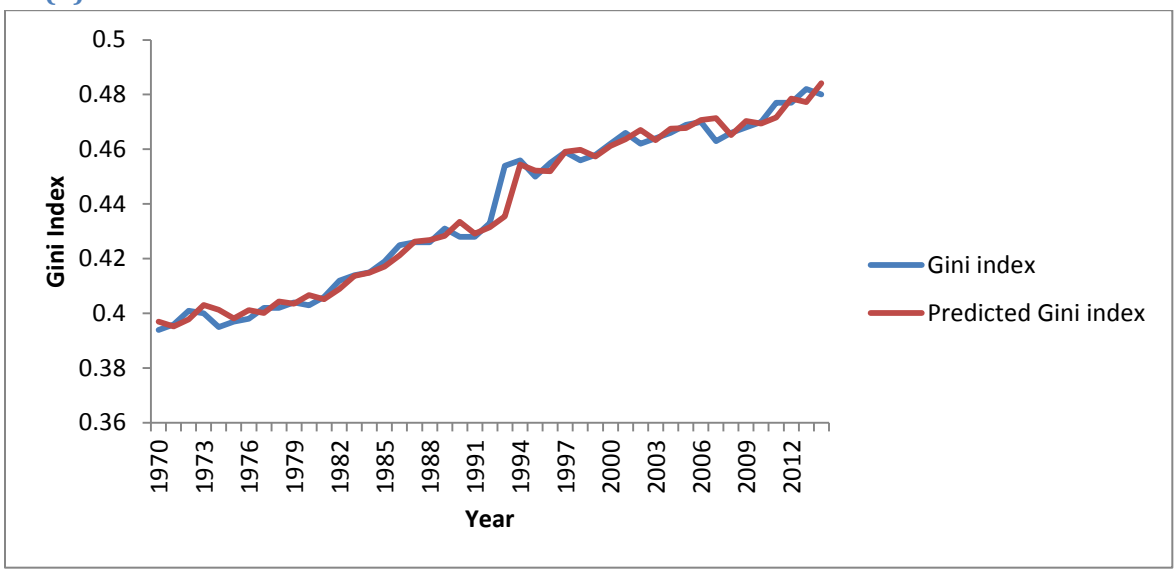
AR(1)



AR(2)



AR(3)



Durbin Watson Test

This test will be used to examine the correlation of the residuals for each of the AR processes. A number close to 0 represents perfect positive correlation. A number close to 2 represents no correlation. A number close to 4 represents perfect negative correlation. For AR(1) model, the result of DW test is shown below, suggesting no correlation between the residuals.

Durbin-Waston D	1.91315
Obs.	46
1st order Autocorrelation	0.98828

Conclusion

According to the fitted values of Gini index, all three models seem to fit pretty well with the actual data. But since Y(t-2) does not explain the model in a significant fashion, we therefore don't want to include this variable into our AR model. Additionally, DW test also suggests that AR(1) model has not much correlation between the residuals. Hence, AR(1) is the best fit model for Gini index trend in the United States since 1967.