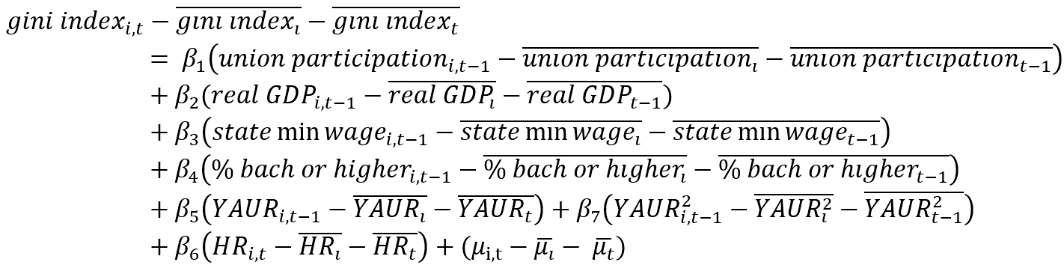
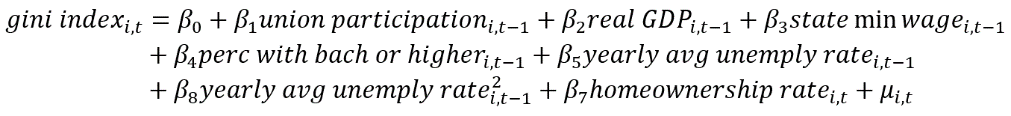
**Methodology**

Our empirical approach is summarized by follow two equations.

1:



2:



In both equations one and two we regress Gini index on the percent of the work force represented by a union for state `I` at time `T`, while controlling for one year lagged real GDP, state minimum wage, percent of population with a bachelor’s degree or higher, year averaged unemployment rate, year averaged unemployment rate squared, and non-lagged homeownership rate. We believe these control variables also influence a state’s income distribution, so it is important that we include them in our models. The difference between the two equations is that equation one contains state and time fixed effects, and equation two is a pooled OLS that does not account for state of time heterogeneity. Since we have panel data, there is a good chance that we have heterogeneity bias between states and years. If so, pooled OLS would not be a good estimation technique because the covariance between our independent variables and the unobserved heterogeneity will not be zero. That will make pooled OLS estimates biased and inconsistent.

To decide if we should include fixed effects in our model, we performed an F-test for individual and time effects. The null hypothesis is that we have homogeneous data and have no need for fixed effects. The F-test was statistically significant, with a p-value of less than 2.22e-16. We reject the null hypothesis and conclude that we have heterogeneity in our data, and fixed effects is the correct specification.

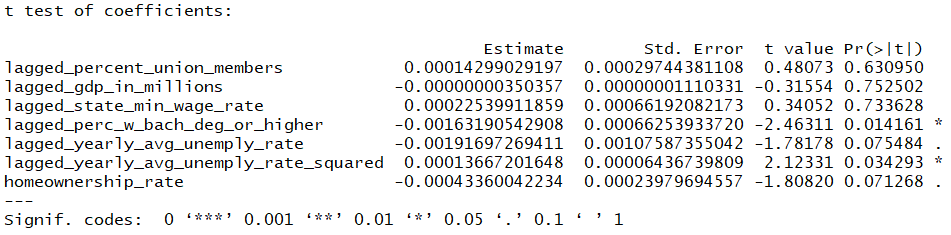
Next, we tested for heteroskedastic and serial correlation. We performed the Breusch-Pagan test for heteroskedastic and the Breusch-Godfrey test for serial correlation of the error term. The null hypothesis in these tests is that heteroskedastic and serial correlation are not present. We found statistically significant results in both tests with p-values less than 0.05. We reject the null hypothesis and conclude that heteroskedastic and serial correlation are present. We obtained Arellano robust standard errors for correct for these issues.

Finally, we checked the presence of multicollinearity in our model. We calculated the variance inflation factor for the fixed effects model. We did not find evidence of multicollinearity, as none of our variables had an VIF above 2, except for lagged average unemployment rate and lagged average unemployment rate squared, which can happen when including a squared term in a model. We tested the VIF of our model both including and excluding lagged average unemployment rate squared. We found the VIF of lagged average unemployment rate is below 2 when lagged average unemployment rate squared is not present in the model.

We were unable to find a suitable interment variable for union participation rate. Because of this, we are unable to address whether the exogeneity assumption holds for our model. Finding some suitable interment variables for union participation rate is a large area for future research for this project, as we believe that union participation rate may suffer from endogeneity.

**Results**

Our results with the robust standard errors are below.



Our overall model is statistically significant with a P-value of 0.0009. Our primary variables of interest, lagged union participation, is not statistically significant, holding all other variables constant. Based on this, our hypothesis that increases in union participation would lead to decreases in the Gini index is unfounded. However, we did find statistically significant evidence that lagged percent of the population with a bachelor’s degree or higher, lagged average yearly unemployment, and homeownership rate influences the distribution of income, holding all other variables constant. The coefficients on lagged percent with a bachelor’s degrees, lagged average yearly unemployment, and homeownership rate do have the expected signs, which is a positive signal about the specification of our model. However, none of these variables seem to have economic significance, as the variables have very small effect sizes.