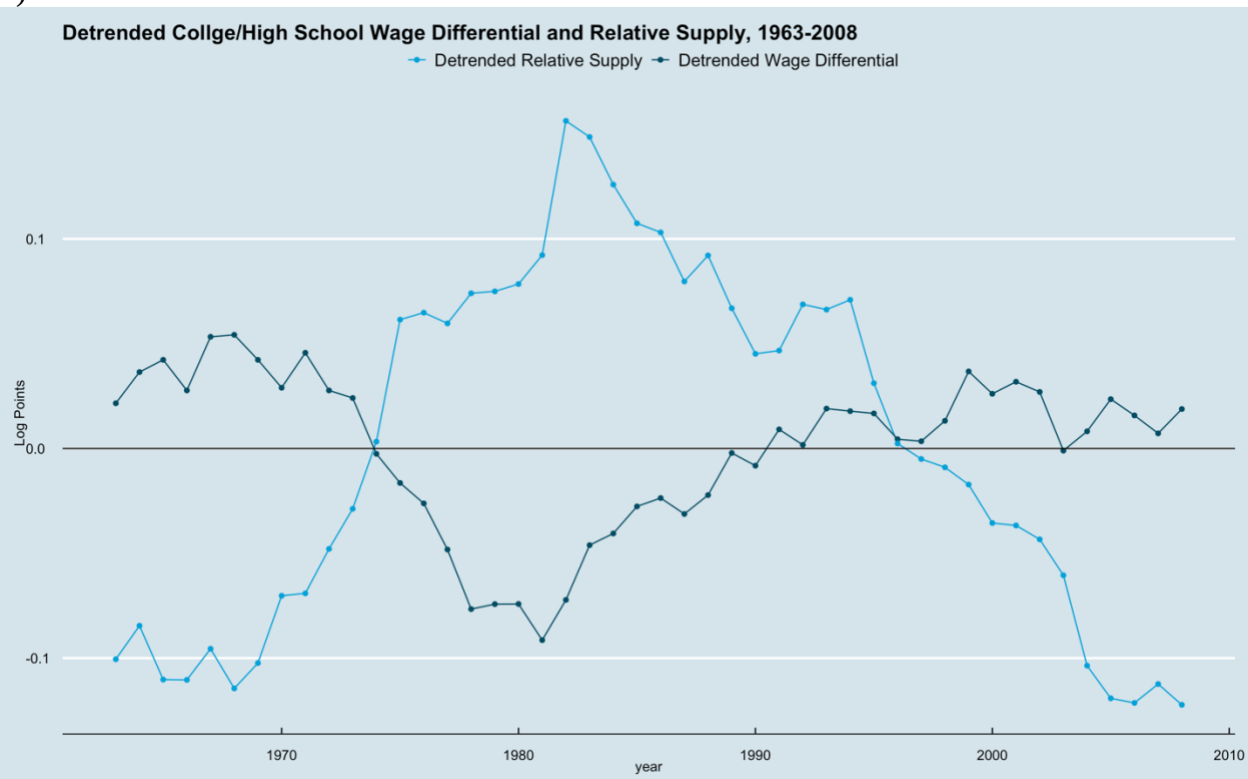
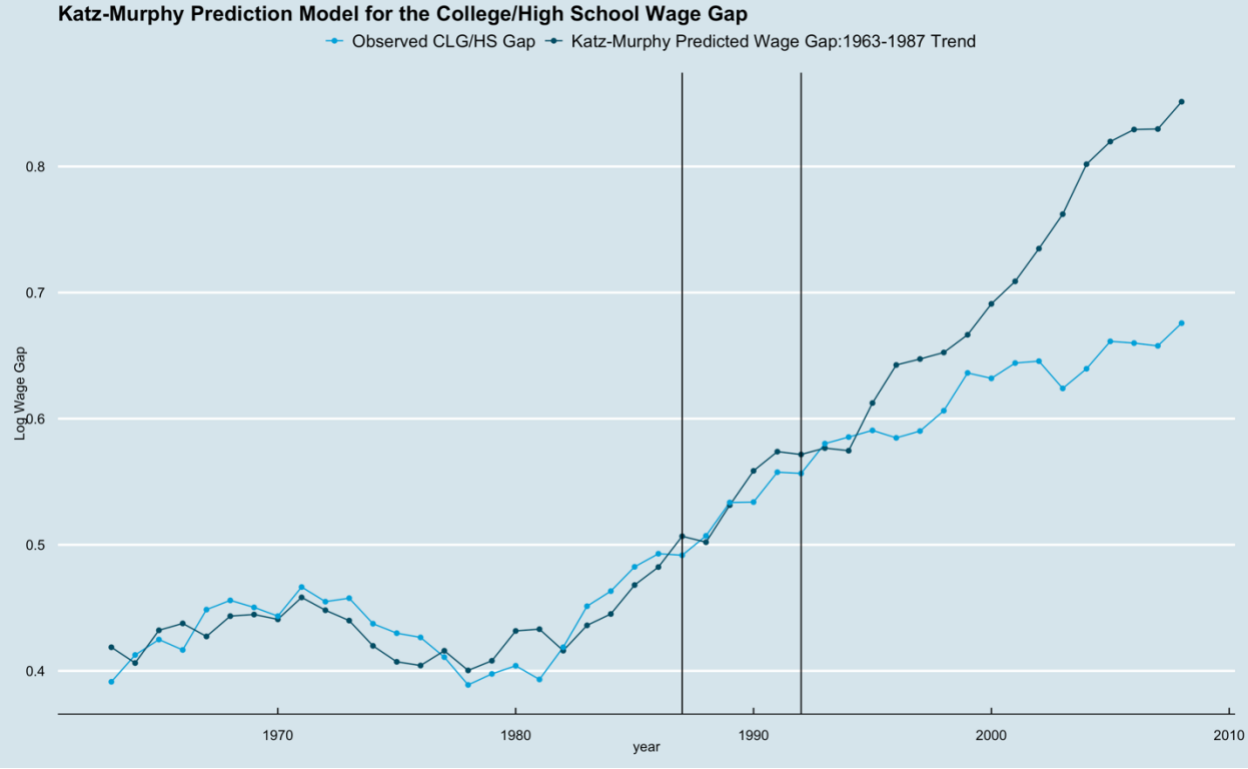
This assignment was to reproduce figure 4 (with colors encouraged) and table 2 from [Trends in U.S. Wage Inequality, Autor-Katz-Kearney] using provided data, extending through 2008.

1)





To reproduce panel A, I used the "detrend" function in R to generate a vector of the residuals from both the wage differential and the relative supply variable. I then used ggplot2 with the Economist theme to graph it.

For panel B, I observed from section III of the paper that the Katz-Murphy model used data from 1963-1987 and included real minimum wage (deflated using PCE) and prime-age male unemployment rate. Because the given data provided only the nominal minimum wage, I downloaded monthly PCE data from the FRED database. I generated annual PCE averages, which I used to convert the nominal minimum wage to real minimum wage (in 2008 dollars). I subsetted the data to 1963-1987 and ran an OLS model as specified in the paper. I then defined a function in R using the coefficients generated in this regression. I plugged in the 1963-2008 data into this function, which allowed me to generate a vector of the predicted values. I then graphed this with the actual CPS data using ggplot2.

(2)

REGRESSION MODELS FOR THE COLLEGE/HIGH SCHOOL LOG WAGE GAP, 1963–2005

	(1)	(2)	(3)	(4)	(5)
CLG/HS relative supply	-0.612	-0.339	-0.591	-0.562	-0.556
	(0.128)	(0.043)	(0.068)	(0.112)	(0.094)
Time	0.027	0.016	0.026	0.029	0.020
	(0.005)	(0.001)	(0.003)	(0.006)	(0.006)
$Time^2/100$				-0.013	0.036
				(0.006)	(0.012)
$Time^3/1000$					-0.007
					(0.002)
Time × post-1992			-0.010		
			(0.002)		
Constant	-0.217	0.059	-0.198	-0.189	-0.145
	(0.134)	(0.039)	(0.068)	(0.122)	(0.103)
Elasticity Estimates:	-1.635	-2.95	-1.691	-1.778	-1.797
p-Values of Elasticity	0	0	0	0	0
Observations	25	46	46	46	46
R^2	0.558	0.935	0.965	0.941	0.960

Note: $^{*}p$ $^{**}p$ $^{***}p<0.01$

Standard errors in parentheses

I also noticed the constants in my models originally differed dramatically from those in the paper, so I changed the "year" variable to be 1-46 instead of 1963-2008. For model 3, I made a dummy variable "post 1992" and interacted it with the "year" variable.

(3) The third question asked to add a trend-break to each year to find the highest R^2. I iterated through this regression with different years by making the dummy variables for each year and the interaction term. I found that the highest R^2 was with the break in 1993, with $R^2 = 0.965$