Social Security Disability and the Affordable Care Act

Background

The Patient Protection and Affordable Care Act (ACA) was signed into law by President Barack Obama on March 23, 2010 with the intent of improving access and affordability of health insurance within the United States. Through a number of employer and individual mandates, tax credits, subsidies for lower income Americans, and the creation of regulated health insurance marketplaces, the ACA aimed to expand coverage and control the rate of increase of health care costs nationwide.

Among other provisions, the ACA created health insurance exchange markets where individuals could purchase private health insurance without exclusions for pre-existing conditions. Participation in these exchanges by health insurers varies by State, and in certain States the exchange itself is State-run rather than using the Federal heathcare.gov platform. In addition, the ACA contained a provision for the expansion of Medicaid eligibility to individuals earning up to 128% of the Federal Poverty Line. This expansion, while intended to be mandatory for all States, was made optional by a later Supreme Court decision. Nineteen States did not adopt the expansion. Thus, the precise implementation of the ACA varies by State.

The ACA was an incremental reform that leveraged existing regulatory structures and programs to expand coverage. As such, the Congressional Budget Office's (CBO) task in estimating the interacting budgetary effects of the reform needed to take into account the how both covered and uncovered individuals would move between non-coverage, employer provided coverage, existing individual policies, the new health insurance marketplaces, Medicaid and other government programs. Given this complex problem, the CBOs initial estimates must be continually re-evaluated based on its observed results as the program matures and changes. While most ACA provisions went into effect in 2014 or prior, there are provisions that are in flux through 2020 under current law and changes to the law are, of course, possible at any time.

Problem

One potential effect that went un-estimated by the CBO was movements within the Social Security Disability program. While the program is focused on income rather than health insurance, there is a provision that provides for Medicaid eligibility after one year of being enrolled in the disability program (or Medicare eligibility after two years, depending on which program(s) the enrollee is eligible for). With this provision in place, prior to the ACA, the disability program was a viable way in which for individuals to eventually acquire health insurance coverage that was otherwise not available, even though that is not the primary goal of the program. Given that the ACA marketplaces provide an alternative means to acquire health insurance to all Americans, it could be the case that the ACA would reduce enrollments in the program and save the U.S. Government money on the Social Security side of the ledger. However, it is also possible that people who would otherwise go on disability are holding on to jobs that provide health insurance, and will go on disability now that they have access to health insurance independent of being employed. As CBO Senior Analyst Joyce Manchester testified to the House Ways and Means Committee (Manchester, 2013):

> *Looking ahead, the Affordable Care Act is likely to influence application rates for the DI [disability insurance] program, but whether it will result in more or fewer beneficiaries is difficult to predict. Among other changes, that legislation will make it easier for people who have health problems to buy their own insurance; it will also provide new subsidies for individually purchased coverage and expand eligibility for Medicaid in states that choose to do so. On the one hand, people who do not have employment-based health insurance will find it easier to obtain subsidized coverage as well as to gain access to health care without applying for DI benefits. That change will tend to reduce applications to the DI program. On the other hand, some people who would lose employment-based health coverage if they left their jobs to apply for DI benefits will have access to insurance during the two-year waiting period for Medicare benefits, with no exclusions for preexisting conditions, through the health insurance exchanges that will be established under the law. Moreover, that insurance might be subsidized, depending on an individual's income. Those considerations will tend to increase applications to the DI program.*

This project will investigate whether or not the ACA has measurably reduced or increased enrollment in Social Security Disability in the early implementation of the law. If the ACA did influence enrollments in the disability program, that would indicate that as the CBO makes its ongoing budget estimates for the ACA, it should take into account the costs or savings as a result of these effects.

Data

Data was obtained through numerous, mostly US Government, sources in the course of this project. Links to the data sources are provided in a bibliography, and the project folder in github contains the actual files used/generated for this project. The data was assembled from numerous source files into a dataframe containing time series from 2006-15 for each State as well as a national series:

- For ACA Enrollment data, I used the annual (final) enrollment reports published by the US Department of Health and Human Services. This data had to be extracted from internal tables in the pdf documents to plain text, then read into pandas frames and cleaned up via python code.
- For Economic Data (GDP, Personal Income, Employment, Compensation, Wages), I used the US Department of Commerce, Bureau of Economic Analysis to obtain series by State and Nationally
- For Demographic Data (Age of Population, Poverty Rates) and necessary reference data (State/Region mapping and State Name/State Code), I used data from the US Census Bureau.
- For Data on the Medicaid Expansion (a simple table of which States did and did not implement the expansion), I found structured data on the Kaiser Family Foundation website.

Data Exploration

Nationally, enrollments in the disability program have risen over time, both in levels and per capital terms. Figures 1 and 2 illustrate the growth in program utilization.

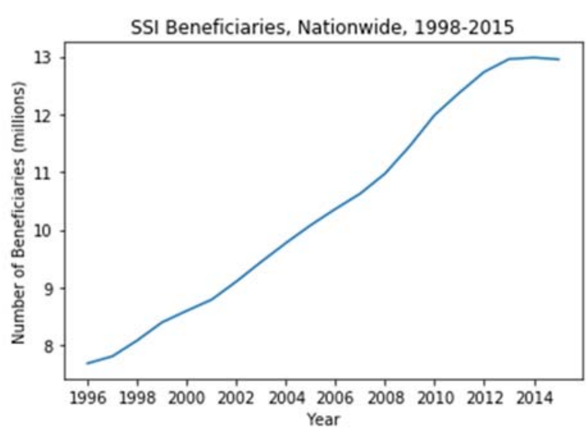

Figure 1

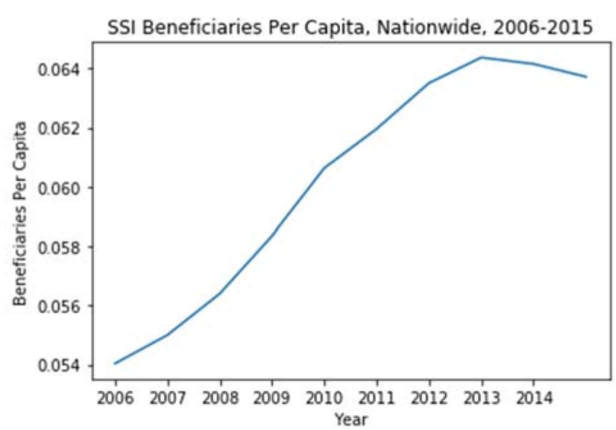

Figure 2

In both plots, there is a distinct reduction in the growth rate (and, in per capita terms a visible reduction) in enrollments corresponding with the timeframe of interest (2013-2015). However, the plots also appear to show an acceleration in disability enrollments beginning in around 2008. This corresponds with the major recession/depression resulting from the financial crisis. It is possible that at least some of the correction to trend disability enrollment growth is simply related to the overall economic recovery of the last few years. This can be seen most clearly in Figure 2, which plots the series from 2006-2015 only.

Disaggregated to the State level, the picture is somewhat messier. Figures 3-6 plot the percent change in per capita disability recipients by State, separated by Census region for simplicity of display. There seems to be more variability between States earlier in the series, with a compression later in the series but no picture emerges that is not simply consistent with the overall economic recovery. Adjustment for economic conditions is required.

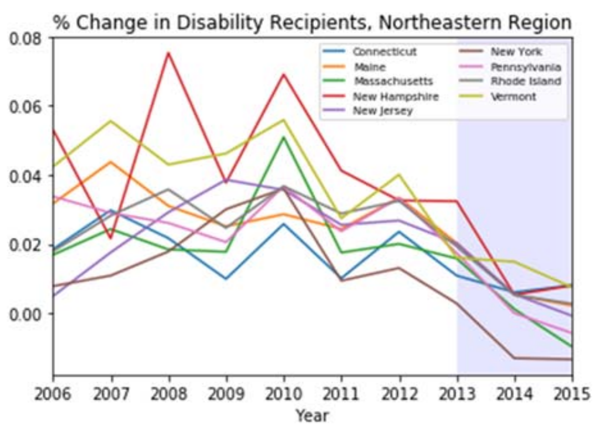

Figure 3

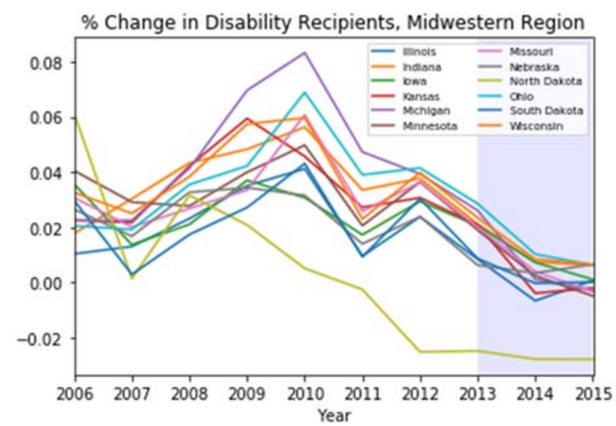

Figure 4

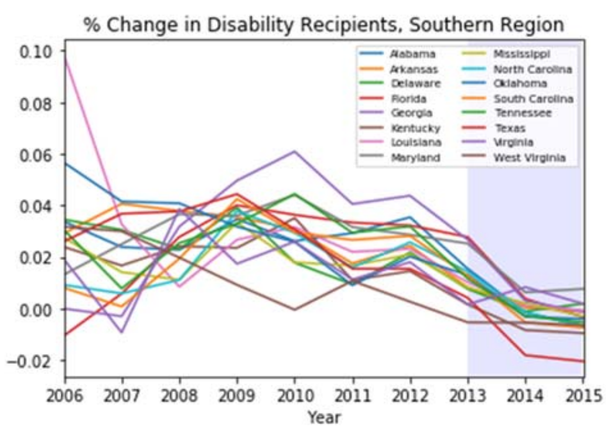

Figure 5

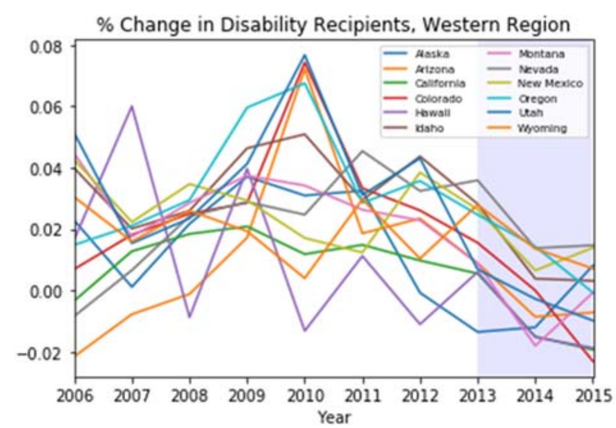

Figure 6

Data for Modeling

In order to account for changes in economic conditions, several annual data series related to GDP, Employment and personal income were obtained and built into a pandas dataframe. In addition, datapoints related to population age and poverty rates were obtained from a separate source and added to the dataframe. These datapoints are maintained in the frame in terms of levels and percentage changes, and where appropriate adjusted for state/national population levels. Single period lags of these variables were also considered as part of the model. The datapoints are:

- SSI Beneficiaries Per Capita (dependent variable in our model)
- Wages Per Capita
- Compensation Per Capita
- Compensation Per Job
- Employment to Population Ratio
- GDP Per Capita
- Personal Income Per Capita
- Average Age of Population
- Poverty Rate

The correlation matrix in Figure 7 was produced in R (GGally/ggpairs) rather than python (matplotlib) due to some useful extra built-in features included in the redundant cells of the table. It is clear (and expected) that many of our economic indicators are highly correlated with each other and will likely not be included in our final model due to the inefficiency caused by multicollinearity. Each datapoint was chosen from the available Bureau of Economic Analysis data due to it being plausibly related to disability enrollments (which are known to depend on overall economic conditions).

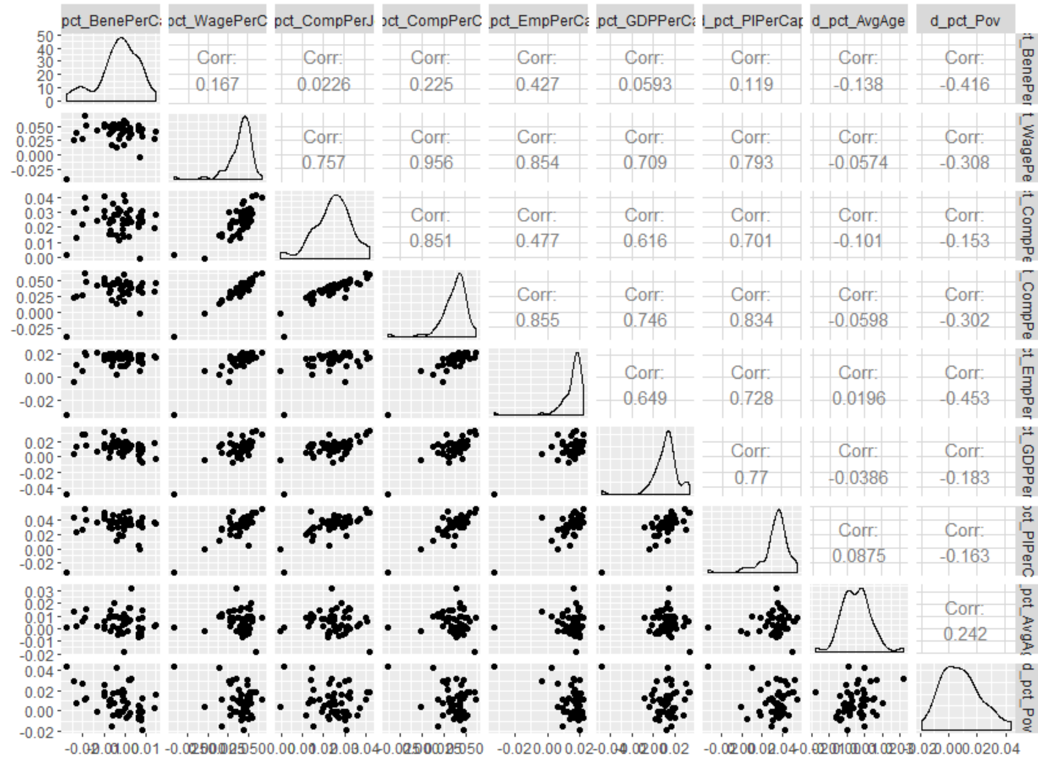

Figure 7

Testing for Structural Break in Panel Regression

As a first attempt to establish an “ACA effect” on disability enrollments, I tried to determine where structural breaks in the disability enrollment time series may occur. If there was a structural break corresponding to ACA implementation (roughly 2013, given delays in approving disability claims), it would provide some evidence that the ACA has changed disability enrollments, controlling for the economic and demographic factors in the dataset.

The dataset was structured as a longitudinal panel, with a 10 year time series of the percentage changes in each variable for each State. I assumed each State would have unobserved effects related to disability administration (the States administer these programs in conjunction with the Federal Government and the State agencies have some influence in how benefits are allocated). So I estimated a fixed effects panel data model, where each grouping (State) can have its own intercept.

To select which features to include in the model, I used a step forward feature selection algorithm scored using the F-statistic. All features were included in the selection process (in percentage change form) as were all their corresponding single period lags. In addition to the entity effects (group intercepts) endemic to the model, the selection process chose just one of the variables in the data set: percentage change in employment to population ratio, at a one period lag. This model unfortunately explains only 22% of the within-group variation in the panel, but the coefficient is statistically and practically significant. Other specifications resulted in statistically insignificant coefficients and a reduction in the significance of the model. A straightforward interpretation of this model is that a decline in the employment to population ratio in one period leads to an increase in disability enrollments in the next period as some of the unemployed apply for disability after being unable to obtain a suitable job.

I tested the model for structural breaks for all years between 2009 and 2013 using the Wald method (e.g. dummy variables to distinguish the period before and after the year tested). For all years tested, the coefficient on the dummy variable was statistically significant, indicating the possibility of a structural break. While only in 2011 and 2013 do the dummy variables improve the model fit significantly, there is a potential structural break at all time periods, so there is no evidence here that there is a regime change affecting disability enrollments, but rather that there are factors unaccounted for by the model affecting year to year changes in disability enrollments.

<u>Prediction Using Enrollment Data</u>

Two additional datapoints are available for a single year within the panel: State by State enrollments in the ACA Exchanges, and a list of states that expanded Medicaid in 2014. Having estimated fixed effects for each State in the previous section, I estimated an OLS model on this single year of data using the entity effects for each State (e.g. the estimated intercept for each State group in the fixed effects model) in lieu of a constant, the lagged change in Employment to Population ratio that was found to be significant in the larger panel estimate, and the change in ACA exchange enrollment from 2014-2015 and a dummy variable indicating States that expanded Medicaid as part of the ACA.

I split the 50 States into a testing and training set using the sklearn test_train_split() function and estimated the model on the training set. 40 randomly chosen States were in the training set and 10 States were in the test set. In order to compare this model to the simpler model estimated in the previous section, I also trained a version of the model excluding the ACA enrollment and Medicaid variables.

Unfortunately, the coefficients on the ACA enrollment and Medicaid variables in the estimate were both practically and statistically insignificant, indicating there was likely little to no relationship between ACA enrollment and Medicaid expansion on the level of disability enrollments in the model. It is very likely that there are factors unaccounted for, as the explanatory power of all models estimated so far using the available economic and demographic data is quite low.

Non-linear Prediction Model

As an alternative specification, and simply to exercise some of the other concepts covered in the course (such as hyperparameter selection via k-fold validation), I estimated a non-linear Kernel Ridge Regression model. I used the same training set to estimate the model, and chose the alpha parameter using k-fold cross validation. I split the training set (again) into 5 parts and for each potential value of alpha I used each of the 5 parts as a test set after estimating the model on the other 4 and evaluated the average r-squared for each of the 5 tests. The alpha with the highest average r-squared was chosen and the model was re-estimated on the entire training set with that alpha hyperparameter.

Results of Prediction Models

The prediction models were then used to predict the values in the test set of ten States and were scored according to the r-squared. As expected, they did not predict disability enrollments very accurately. The model without ACA and Medicare data predicted 7 percent of the variation, while the model with those data points predicted 5 percent of the varation.

Conclusion and Lessons Learned

I was unable to establish any relationship at all between the implementation of the ACA and Medicaid expansion and the level of enrollment in the Social Security disability program. Because my data gathering focused primarily on economic factors, I ended up gathering a number of time series that were moderately to highly correlated and were dropped from the model entirely due to statistical insignificance when included in the same model as each other. There does seem to be a statistically and practically significant lagged effect of employment on disability enrollments but this is well known and its inclusion in the model was a control variable and it was not the treatment of interest.

From a project standpoint, I picked a topic without a well known, readily available dataset and spent a lot of time searching for economic and demographic data that had the same longitudinal, State-grouped structure as the disability data I was able to find. This limited the available regressors and resulted in some very weak models. When I first started estimating the models and had underwhelming results I went back and attempted to find additional demographic data (particularly epidemiological data) but was unable to find State by State time series for rates of cancer incidence, drug usage, etc. that may be factors in a State's level of disability enrollment. As such, the estimations in this project were of little use. However the act of coding them was a learning experience, as was the very significant amount of data wrangling I did in order to put together an ultimately not-very-useful panel data set. I would advise future students in this course to pick a more purely machine-learning based project with a readily available, relevant dataset and not bring time-series or panel data models into the equation.

Code Files

- data_wrangling_basic_plots.ipynb contains the code used to import, transform and clean the data from HHS, BEA, Census and the Social Security Administration
- panel.ipynb contains the code used to estimate the PanelOLS model and choose parameters
- enrollment_estimation.ipynb contains the attempt to use ACA enrollment data to predict disability enrollment and the nonlinear model done as an exercise
- natl_data_plots.ipynb and state_region_data_plots.ipynb contain the matplotlib code used for the charts
- ggpairsplot.R contains the R code to generate the correlation matrix

References

Manchester, J. (2013). Testimony: The Social Security Disability Insurance Program. *House Ways and Means Committee.* https://waysandmeans.house.gov/UploadedFiles/Manchester_Testimony.pdf

Data Sources:

Medicaid Expansion by State (Kaiser Family Foundation): http://www.kff.org/health-reform/state-indicator/state-activity-around-expanding-medicaid-under-the-affordable-care-act/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

Mean Age by State, Poverty Rates (US Dept of Commerce, Census Bureau): https://www.census.gov/cps/data/cpstablecreator.html

Census Regions and Divisions (Census): https://www2.census.gov/programs-surveys/popest/geographies/2011/state-geocodes-v2011.xls

Per Capita GDP (US Dept of Commerce, Bureau of Economic Analysis): https://www.bea.gov/itable/iTable.cfm?ReqID=70&step=1#reqid=70&step=10&isuri=1&7003=1000&70

35=-1&7004=naics&7005=1&7006=xx&7036=-1&7001=11000&7002=1&7090=70&7007=-1&7093=levels

Per Capital Personal Income (BEA): https://www.bea.gov/itable/iTable.cfm?ReqID=70&step=1#reqid=70&step=30&isuri=1&7003=1000&7035=-1&7004=naics&7022=21&7005=1&7023=0&7033=-1&7024=non-industry&7006=xx&7025=0&7026=xx&7027=-1&7036=-1&7001=421&7028=-1&7002=1&7031=0&7040=-1&7083=levels&7029=21&7090=70&7093=levels&7007=-1

Wages, Compensation, Employment (BEA): https://www.bea.gov/regional/downloadzip.cfm (Annual State Personal Income and Employment, all areas)

ACA Enrollment Data (US Department of Health and Human Services), extracted from internal pdf tables:

- 2014 https://aspe.hhs.gov/system/files/pdf/76806/ib_2014mar_enrollment.pdf
- 2015 https://aspe.hhs.gov/system/files/pdf/83656/ib_2015mar_enrollment.pdf
- 2016 (unused) https://aspe.hhs.gov/system/files/pdf/187866/Finalenrollment2016.pdf, https://aspe.hhs.gov/system/files/aspe-files/187871/marketplacestatefinal2016.xlsx

Disability Enrollment data by State and total (Social Security Administration):

- 2005 https://www.ssa.gov/policy/docs/statcomps/di_asr/2005/sect05.xlsx
- 2006 https://www.ssa.gov/policy/docs/statcomps/di_asr/2006/sect05.xlsx
- 2007 https://www.ssa.gov/policy/docs/statcomps/di_asr/2007/sect05.xlsx
- 2008 https://www.ssa.gov/policy/docs/statcomps/di_asr/2008/sect05.xlsx
- 2009 https://www.ssa.gov/policy/docs/statcomps/di_asr/2009/sect05.xlsx
- 2010 https://www.ssa.gov/policy/docs/statcomps/di_asr/2010/sect05.xlsx
- 2011 https://www.ssa.gov/policy/docs/statcomps/di_asr/2011/sect05.xlsx
- 2012 https://www.ssa.gov/policy/docs/statcomps/di_asr/2012/sect05.xlsx
- 2013 https://www.ssa.gov/policy/docs/statcomps/di_asr/2013/sect05.xlsx
- 2014 https://www.ssa.gov/policy/docs/statcomps/di_asr/2014/sect05.xlsx
- 2015 https://www.ssa.gov/policy/docs/statcomps/di_asr/2015/sect05.xlsx