TITLE: Correcting Medicaid Enrollment Underreporting by the Current Population Survey: A Stochastic Frontier Analysis

ABSTRACT: Medicaid expenditures represent a large proportion of the U.S. budget and are increasing. In 2013, about 15% of the National Health Expenditure or approximately $450 billion was devoted to Medicaid and its recipients. Medicaid spending is projected to increase by 5.9% per year between 2015 and 2024 as 7.6 million new beneficiaries are expected to enroll. Naturally, the number of people enrolled in Medicaid is of particular interest and the Current Population Survey (CPS) is the most widely cited source for estimates on enrollment. However, previous literature has shown the CPS underreports enrollment in comparison to state-level records by 30-40% which was between 20 and 27 million enrollees in 2011. We attempt to correct this inaccuracy using stochastic frontier analysis (SFA) to estimate the true value of Medicaid enrollment given the CPS data.

BACKGROUND: Medicaid is a federal assistance program that provides healthcare to eligible low-income adults, children, pregnant women, elderly adults (though recipients are typically younger than age 65), and people with disabilities. Minimum requirements for eligibility are set by the federal government, but each state independently administers coverage. Funding for the program is provided by both the states and the federal government. Together, Medicaid spending represents about 15% of the National Health Expenditure, which amounted to approximately $450 billion in 2013. Consequently, detailed data on Medicaid is valuable to policymakers and researchers. Although administrative records are considered the gold standard, they are often incomplete, lagged or lacking the level of detail necessary. For these reasons, researchers tend to prefer using survey data instead. The most widely-cited survey on Medicaid coverage is the Current Population Survey Annual Social and Economic Supplement (CPS ASEC; hereafter CPS), particularly its measurement of Medicaid enrollment. Unfortunately, it has been shown the CPS undercounts enrollment by 30 to 40% on average when compared to state-level administrative records. This underreporting is generally thought to be caused by survey respondents misreporting coverage status. Individuals may not report temporary periods of Medicaid coverage, may confuse Medicaid managed care with private coverage, or may not be aware of their status. The Census Bureau itself may also be somewhat responsible for the undercount, as Medicaid enrollment status is sometimes imputed probabilistically based on other factors about the individual. There are different approaches to correcting the underreporting by the CPS, such as survey redesign to reduce confusion among respondents, logistic regression to estimate the probability an individual is enrolled in Medicaid and linking CPS data to administrative records. Each of these focuses on improving the accuracy of individual enrollment reporting, which requires access to household level data. We avoid the need for costly or inaccessible data by using SFA to estimate aggregate enrollment by state.

METHODS: To begin our analysis, we will first collect data on annual enrollment from both the CPS and the administrative records which are housed by the Medicaid Statistical Information System (MSIS), as well as demographic and socioeconomic data correlated with Medicaid enrollment. It may seem trivial to estimate enrollment when the true data is available, but administrative records often lag behind CPS data by multiple years. Thus a good measure of present enrollment would be useful. Once the data are collected, we will use stochastic frontier analysis (SFA) to estimate the actual enrollment data that is higher than what CPS reports. In general, SFA uses observed data on an outcome to estimate the value of the unobserved outcome (the “frontier”). In our case, enrollment reporting by MSIS is assumed to be correct so we will treat it as the frontier. Therefore, the CPS data will be the dependent variable and demographic and socioeconomic variables that control for natural variation in Medicaid enrollment will act as the observed set of covariates that explain enrollments. Finally, we will compare the estimated enrollment figures from the SFA with the correct figures from administrative records telling us how effectively the SFA estimates are closing the gaps between the CPS numbers and the actual enrollments.
EXPECTED OUTCOME: SFA will allow us to extend beyond the available MSIS data and adjust CPS enrollment figures much closer to the true amounts of enrollment using only the CPS and socioeconomic/demographic data. The ability to accurately estimate Medicaid enrollment multiple years in advance will be extremely useful in budgeting, reforming healthcare policy, creating new policy, and more. Furthermore, with the recent push to expand Medicaid coverage through the Affordable Care Act, accurate, up-to-date enrollment data will be crucial in states’ decision to move forward on expansion.

TIMELINE/PLAN OF WORK:

November 2015
- Write Honors in the Major proposal

December 2015
- Finish editing proposal and submit for review
- Finish collecting demographic and socioeconomic data

January 2016
- Run SFA using Stata/SE
- Troubleshoot any data issues (e.g. missing values)
- Finalize regression model and compare to administrative data
- Apply to present research at the Showcase of Undergraduate Research Excellence

February 2016
- Write research thesis and edit
- Submit to HIM committee for review

March 2016
- Finalize thesis
- Prepare to defend thesis

April 2016
- Defend thesis before committee
- Present research at SURE
- Make changes to thesis as recommended by committee

May 2016
- Submit completed thesis to Undergraduate Research Journal and economics-specific journals

ITEMIZED BUDGET AND BUDGET JUSTIFICATION:

Stata/SE $395
Vendor: StataCorp
Total: $395

Will be used to run stochastic frontier analysis on the large dataset of Medicaid enrollment.