

Correcting Medicaid Enrollment Underreporting By The Current Population Survey: A Stochastic Frontier Analysis

2016

Brachel R. Champion
University of Central Florida

Find similar works at: <https://stars.library.ucf.edu/honorstheses>

University of Central Florida Libraries <http://library.ucf.edu>

 Part of the [Economics Commons](#)

Recommended Citation

Champion, Brachel R., "Correcting Medicaid Enrollment Underreporting By The Current Population Survey: A Stochastic Frontier Analysis" (2016). *Honors Undergraduate Theses*. 46.
<https://stars.library.ucf.edu/honorstheses/46>

This Open Access is brought to you for free and open access by the UCF Theses and Dissertations at STARS. It has been accepted for inclusion in Honors Undergraduate Theses by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.

CORRECTING MEDICAID ENROLLMENT UNDERREPORTING BY THE
CURRENT POPULATION SURVEY: A STOCHASTIC FRONTIER
ANALYSIS

by

BRACHEL R. CHAMPION

A thesis submitted in partial fulfillment of the requirements
for the Honors in the Major Program in Economics
in the College of Business Administration
and in the Burnett Honors College
at the University of Central Florida
Orlando, Florida

Spring Term, 2016

Thesis Chair: Richard Hofler, Ph.D.

Abstract

The Current Population Survey (CPS) is the most widely cited source for estimates on Medicaid enrollment. However, previous literature has shown the CPS underreports enrollment by 30-40% in comparison to state-level records. The question then is how to correct the Medicaid enrollment gap brought on by the CPS. Gross adjustments for the discrepancy may be made, but only if the true amount of enrollees is known. In years when administrative records are delayed or incomplete this is not possible. To date, the methods for correcting underreporting require access to the state-level data which is usually infeasible or unpublishable due to privacy issues. Redesigning the CPS questionnaire itself might alleviate a good part of the undercount but doing so is well beyond the scope of most researchers. A better correction would rely only on the CPS count of Medicaid enrollees so as to avoid privacy concerns and time delays. We propose using stochastic frontier analysis to shrink the gap between the CPS count of Medicaid enrollees and the state records by adjusting the CPS counts to be closer to the state records.

Acknowledgements

To my Thesis Chair, Dr. Richard Hofler, for his insight, advice, and direction on this work.

To Drs. Tirthatanmoy Das and Albert Liu for asking hard questions.

And to everyone else who supported me while I was studying, researching, and writing.

Table of Contents

Introduction.....	1
Background.....	3
Literature Review	7
Research Design	10
Data.....	10
Methodology.....	13
Results.....	16
Conclusion	22
References.....	24

List of Figures

Figure 1	12
Figure 2	19
Figure 3	20
Figure 4	21

List of Tables

Table 1. Results from Estimating Stochastic Frontier Model: Y is CPS Enrollment Counts. One-sided Error Term has Exponential Distribution.....	17
Table 2. Results from Estimating Stochastic Frontier Model: Y is CPS Enrollment Counts. One-sided Error Term has Half-Normal Distribution.....	17
Table 3. Results from Estimating Stochastic Frontier Model: Y is CPS Enrollment Counts. One-sided Error Term has Truncated Normal Distribution.....	17

Introduction

Every month the Current Population Survey (CPS) (produced jointly by the Census Bureau and the Bureau of Labor Statistics) collects and reports data on U.S. labor statistics. Every March a supplemental survey is conducted to gather information on healthcare coverage and, in particular, Medicaid enrollment. Using the CPS has many advantages to researchers who might be interested in data on people who are in various stages of Medicaid coverage. The survey is conducted regularly and carefully designed to be as representative as possible. The CPS also has a great breadth of information collected on each respondent, which allows for much more flexibility in analysis of the data. For these reasons the Current Population Survey (CPS) is the most widely cited source for estimates on Medicaid enrollment. Unfortunately the CPS undercounts the total number of people enrolled in Medicaid in comparison to state-level records. Previous literature has shown it falls as much as 30-40% below the true level of aggregate enrollment. Many different attempts have been made to improve the accuracy of reporting by the CPS but none have been both successful and feasible. An obvious solution is to make gross adjustments for the discrepancy, but this is only possible if the true amount of enrollees is known. It is not uncommon for administrative records to be delayed or incomplete for many years at a time. The most up-to-date administrative data that isn't missing any records is almost five years old at the time this was written. Other methods for correcting underreporting require access to the state-level data which is usually infeasible or unpublishable due to privacy issues. Redesigning the CPS questionnaire itself might eliminate much of the undercount but doing so is clearly impossible for anyone outside of the Census Bureau and the Bureau of Labor

Statistics. We show that using stochastic frontier analysis to shrink the gap between the CPS and state records is a viable alternative to these methods and requires only CPS enrollment data

Background

Medicaid provides health coverage for millions of Americans and is funded jointly by the federal government and the states. The program is targeted to low-income adults, children, pregnant women, elderly adults and people with disabilities. States who participate in Medicaid are required to cover population groups defined as mandatory eligibility groups but may choose to extend coverage to optional eligibility groups. Each state sets eligibility criteria within the federal minimum standards for determining low-income status, which is usually calculated as a percentage of the Federal Poverty Level. Non-financial criteria are also included such as the Supplemental Security Income (SSI) program, state residency requirements, immigration status, and documentation of U.S. citizenship.¹

States are required to cover certain services such as inpatient and outpatient hospital care, nursing facility services, laboratory and X-ray services, physician services and much more. Many states also cover optional benefits including prescription drugs, physical and occupational therapy, prosthetics, dental services, eyeglasses and hospice. States have the flexibility to charge premiums and establish cost-sharing programs such as copayments, co-insurance and deductibles. They also have the ability to charge higher premiums for groups whose family income is greater than 150% of the federal poverty level. However, maximum out-of-pocket payments are limited and at-risk groups like children and pregnant women are exempt from most of these charges. In addition, participants cannot be charged copayments and coinsurance for certain services. Medicaid gives states the ability to incentivize cost-effective behavior by

¹ The next two pages concerning the details of Medicaid have been excerpted from Medicaid.gov.

charging higher copayments for non-generic drugs or non-emergency visits to hospital emergency rooms.

Although the states bear the responsibility of administering Medicaid, the program is funded by the federal government in conjunction with the states. The federal government pays a specified percentage of expenditures, called the Federal Medical Assistance Percentage (FMAP), which varies by state depending on criteria such as per capita income. The average FMAP is 57% but ranges from 50% in richer states to 72% in low-income per capita states. Each state's FMAP is updated every three years to allow for normal fluctuations in the economy. States must prove they are able to pay their share of expenditures using only recognized sources of funding before federal funding for the services in question are approved. This is especially important when states are applying to increase their FMAP or to cover new services. States are responsible for agreeing upon payment rates with providers which often come in the form of fee-for-service or managed care contracts. About 70% of Medicaid enrollees receive services through a managed care system and providers of said services are paid monthly capitation payments. These rates may be adjusted for trends based on Medicaid-specific factors or the Medicare Economic Index.

To track and evaluate the effectiveness of Medicaid, the Centers for Medicaid & Medicare Services (CMS) collect detailed records on Medicaid enrollees from the states. States report data on the application, eligibility and enrollment processes through the Medicaid and Children's Health Insurance Program (CHIP) Performance Indicator Project, child enrollment reporting through the Statistical Enrollment Data Systems and, more recently, total enrollment by group as well as overall enrollment through the Medicaid Budget and Expenditure System.

Beyond these, CMS also uses the Medicaid Statistical Information System (MSIS) to house data on eligibility, enrollment, utilization and expenditures for Medicaid and CHIP. MSIS contains records on participants in every state and territory which are sourced on a quarterly basis by the states. MSIS is used mainly for 1) health care research, 2) usage and expenditure forecasting, 3) analyzing proposed policy changes, 4) responding to questions from stakeholders and 5) matching to other health related databases.

While MSIS is commonly used for gathering data on Medicaid participants, extracting the data for analysis can be difficult and sometimes infeasible. CMS publishes certain datasets, but many of the records are personally identifiable and therefore require a linking process that protects individual privacy, which can be costly if not impossible. Further, CMS does not collect the data itself, rather depending on the states to submit the records to CMS. It is not uncommon for datasets to be incomplete or delayed (Davern et al. 2009). For these reasons, researchers often use survey data in lieu of administrative records from MSIS (or other sources). The most widely-cited survey for information on Medicaid coverage is the Current Population Survey Annual Social and Economic Supplement (hereafter CPS). CPS surveys about 70,000 households primarily during March but also in February and April, gathering data on Medicaid coverage during the previous year.

Because of its size and scope, state and national health policy research often uses the CPS (Blewett et al. 2004). State Children's Health Insurance Program (SCHIP) funding is allocated based on CPS estimates (Davern et al. 2003) and the Congressional Budget Office uses the data to calculate the cost of future legislation (Glied, Remler and Zivin 2002). CPS data is also used by states to analyze cost effectiveness of potential state health policy, to report to the federal

government the progress SCHIP makes in insuring low-income uninsured children and sending progress reports on other healthcare related efforts (Blewett and Davern 2006). The CPS is also used by academic researchers to analyze the effectiveness of state health policy reforms (Blewett et al. 2004). Determining the number of people who are eligible for Medicaid but not currently enrolled is an example of such a study.

Despite the widespread use of the CPS to estimate Medicaid enrollment, it has been shown the survey actually undercounts enrollment by approximately 30-40% on average when compared to the MSIS records (Pascale et al. 2009, Davern, Klerman, Baugh et al. 2009, Davern, Klerman, Ziegenfuss et al. 2009). In 2011, an undercount of 30-40% amounted to 20 to 27 million enrollees unaccounted for by the CPS. Even with such a large gap, the CPS is still a popular source of information on Medicaid enrollment. This is due in part to the large scale and long history of the CPS which, although inaccurate, is preferable to smaller, shorter running surveys that do not publish state identifiers such as the National Health Interview Survey or the Medical Expenditure Panel Survey Household Component (Klerman, et al. 2009). The CPS uses state identifiers and also has greater detailed income information.

Literature Review

There are many suspected causes of the gap in Medicaid enrollment reporting, the most prominent reason being misreporting of coverage status by respondents. Davern et al. (2009) found that approximately 41% of people in both the CPS and MSIS records failed to report being enrolled in Medicaid at least one day in the previous calendar year. They suggest the failure could be due to confusion between Medicaid and SCHIP, which often have similar sounding titles in many states, or the transition between the two can make it difficult for respondents to correctly report their enrollment status. The level of coverage may also cause confusion about current enrollment. People with partial Medicaid coverage may report being enrolled, though they are not considered to be by the CPS (Davern et al. 2009). However, Klerman, Ringel and Roth (2005) show that people with higher Medi-Cal (California's Medicaid program) coverage rates were more likely to report enrollment.

Other causes of enrollment underreporting include respondents failing to report temporary periods of Medicaid coverage (Klerman et al. 2009, Pascale et al. 2009), confusing Medicaid managed care with private coverage (Czjaka et al. 1998, Pascale et al. 2009), being unaware of their status (Pascale et al. 2009) or choosing not to admit enrollment in Medicaid due to social stigmas (Czjaka et al. 1998, Klerman et al. 2005). 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 (Davern et al. 2009).

Correcting Medicaid enrollment may seem trivial at first glance, as one could adjust the total number of enrollees as reported by the CPS to match administrative records in the MSIS. But as Davern et al. (2009) point out, MSIS data is often lagged which means any adjusted

survey data would be at least a few years old. Others have suggested linking individual CPS records to corresponding entries in MSIS, but the same time delay problem arises. Furthermore, such data would be confidential and therefore only accessible by someone working at the Census Bureau. Davern et al. instead model measurement error to probabilistically impute Medicaid enrollment status based on linked records. Such regression models are publicly available since they serve only to predict the probability of being enrolled in Medicaid given a set of characteristics. They describe these models as “partial corrections” as they still lack up-to-date MSIS data to use for modeling the error. Davern et al. also admit the possibility of so-called “false positives” since records that cannot be matched are assumed not to be enrolled in Medicaid.

In a separate study by Davern et al. (2009), CPS records were again linked to MSIS records but not used to impute coverage probabilistically. Instead they make adjustments to MSIS in an effort to match the CPS. Duplicate records, SCHIP enrollees, partially covered enrollees and institutionalized groups are removed from the set. Doing so reduced the undercount by about 10% in 2000 and 2001 to 32% and 31% respectively. 10% is a significant reduction, but it still requires the process of linking survey records to the troublesome MSIS data.

Of course, if there were no misreporting of coverage by survey respondents, the undercount would be much smaller. Pascale et al. (2009) argue that perhaps the CPS itself is flawed and therefore the cause of response error. The questionnaire is unclear about which time period surveyors are discussing when asking about Medicaid coverage. The phrase “At any time during [the past 12 months/previous calendar year]” does not cause respondents to focus on the period in question, rather their current state of enrollment. Pascale et al. recommend redesigning

the survey to reflect cognitive tests that show more accurate response rates given certain rewordings of the question.

Research Design

Data

The CPS as we have defined it consists of two components: the base Current Population Survey and the March supplement, the Annual Social and Economic Supplement (ASEC). The supplement utilizes the base survey's sample to collect additional, more detailed information on Medicaid coverage. This is especially effective as the base survey is designed in a manner to ensure statistical consistency in an enormous and changing sample pool.

The CPS surveys households in a multistage, stratified manner so as to create a representative survey of the population. Approximately 72,000 housing units are chosen from 824 sampling areas, of which 12,000 housing units are sampled under the State Children's Health Insurance Program (SCHIP). These have been included in the CPS sample since July 2001. Potential households are identified from addresses obtained by the 2000 Decennial Census, but the sampling frame is continuously updated for households built after the Census.²

The U.S. is divided into primary sampling units (PSUs), each of which fall within a state boundary and comprise a metropolitan area, a large county or a group of small counties. PSUs are grouped into strata which are constructed to be as homogeneous as possible with respect to labor force and other social and economic characteristics correlated with unemployment. One PSU is sampled from each stratum with a probability of selection proportional to its population determined by the 2000 Census. Next, a sample of housing units are drawn from each PSU and grouped into small sets of housing units, called ultimate sampling units (USUs). Within a USU,

² The next two pages are excerpted from Census Bureau Technical Paper 66 (2006) and ASEC Technical Documentation 2013.

households from blocks with similar demographic and geographic characteristics are grouped together.

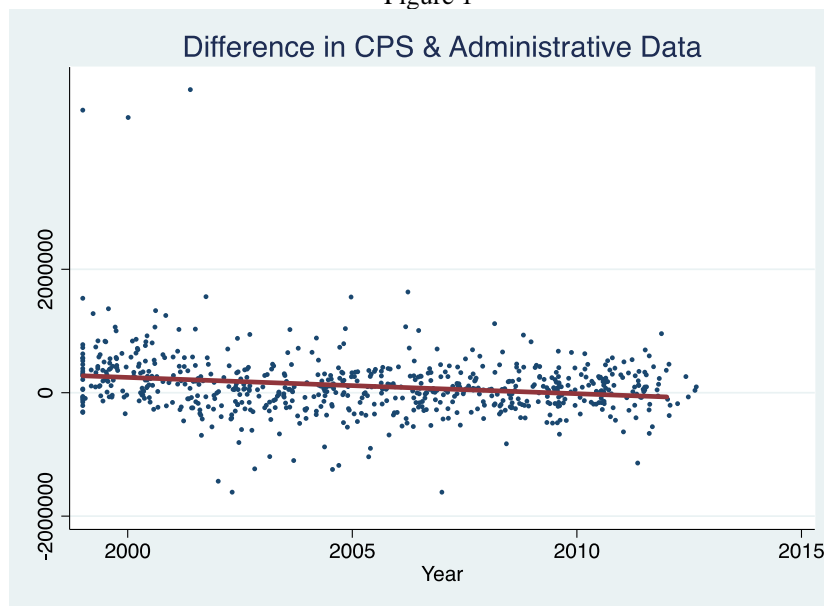
Interviewers collect data each month from the sample of housing units, with each housing unit being interviewed for four consecutive months. Although some 72,000 housing units are in the sample, only about 54,000 are actually interviewed typically because of nonresponse. After the first four months, the unit is then dropped from the sample for eight consecutive months, and then interviewed again for another four consecutive months. So each housing unit is interviewed a total of eight times. To improve accuracy of month-to-month and year-to-year estimates, households are rotated in and out of the sample such that in any given month an eighth of the housing units are being interviewed for the first time, another eighth are being interviewed a second time, and so on. Therefore, after the first month, there will be a 75 percent overlap between months. Similarly, after a year, there will be a 50 percent overlap between any given month and the same month the previous year. This rotation scheme minimizes four things: (i) variance of estimates of month-to-month change, (ii) variance of estimates of year-to-year change, (iii) variance of other estimates of change (outgoing housing units are replaced with housing units that have similar characteristics), and (iv) response burden.

Prior to 2001, the ASEC supplement occurred in April and was known as the Annual Demographic Supplement (the name was later changed to ASEC in 2003). After 2001, the sample was increased to allow for more time to interview units. Thus interviews are now also conducted in February and April, though the majority are still conducted in March. The ASEC differs from the CPS in that it includes certain members of the Armed forces and therefore minor sample practices and weighting procedures are changed slightly. The total sample consists of the

entire March CPS sample, additional Hispanic households, Non-Hispanic non-White households, and non-Hispanic White households with children 18 years or younger. These additional cases are in addition to those in the CPS sample. Thus, the effective sample size of the ASEC increases from 72,000 to 98,000. At an interview rate of about 75% (54,000/72,000), one could expect 73,500 households to be interviewed out of the possible 98,000.

We selected CPS enrollment data on 51 states (including District of Columbia) from 1999 to 2012 to match MSIS enrollment records available as of 2016. On average the CPS counts of enrollment are less than the corresponding MSIS count as shown in Figure 1. However the CPS does not always undercount enrollment for every state in every year. In some instances it either matches or exceeds the MSIS. Thus we choose a subset of 15 states that undercount only over the entire time period. This simplification creates the most ideal situation for a stochastic frontier model to accurately predict the true value of enrollment.

Figure 1



Along with enrollment data we use state-level aggregate data on demographic and socioeconomic characteristics as predictors of CPS counts of enrollment. Keeping with previous literature (Davern et al. 2009, Pascale et al. 2009) we use measures of state GDP, GDP per capita, average household income, population above the federal poverty limit, race, and age brackets similar to Davern et al. (2009). GDP data was downloaded from the Bureau of Economic Analysis along with total state population estimates. Population race, age, and poverty estimates were found by collecting CPS microdata from Integrated Public Use Microdata Series (IPUMS), calculating the proportions of the population possessing the characteristic in question and then multiplying the percentage by the number of people living in the state that year. Finally, average household income was estimated by calculating the mean income of each state by year using the same microdata.

While these measures are important in predicting CPS counts of enrollment we are more interested in calculating the gap between CPS reporting and MSIS than in estimating coefficients of regressors. Stochastic frontier analysis allows us to do so by separating random error of observations from the error between CPS and MSIS.

Methodology

Stochastic frontier analysis (SFA) is a unique econometric model developed by Aigner, Lovell and Schmidt (1977), ALS hereafter, that estimates the difference between some observed outcome and an unobserved frontier outcome. In some cases, that estimated difference is the shortfall of an observed outcome *below* the unobserved frontier. The usual SFA model includes the normal matrix of variables and their estimated coefficients that is typical of linear regression, however the approach to estimating the error is different in SFA. It decomposes the regression

error term (ε) into two parts: a two sided error term and a nonnegative error term (or nonpositive if the frontier is below the observed output, as is the case in minimization problems). We can write the model formally as,

$$y = \mathbf{X}\boldsymbol{\beta} + v - u$$

where v is the stochastic component of the composite error term and $u \geq 0$ is the “inefficiency” component. We assume that v is independent and identically distributed as normal with mean zero and variance σ_v^2 . u is assumed to be distributed independently of v , and follows either a truncated normal or an exponential, or a half-normal distribution (three of the four standard one-sided distribution used in SFA models.)

The economic interpretation of u and v is quite simple if we consider Aigner, Lovell and Schmidt’s example of a firm and its production function. The firm’s production depends on their own efficiency and purely random disturbances. Since the firm must produce at or below its production possibility frontier, u measures the distance between an observed output level and the frontier. If the firm produces at maximum efficiency, we have $u = 0$. On the other hand, v represents the random variation in production caused by factors outside the firm’s control. Examples include unusually good or bad weather, luck or breakdowns, which are all factors that affect production but are exogenous to the firm.

In our case, MSIS is assumed to be the true count of Medicaid enrollees (Pascale et al. 2009) so we will treat it as the frontier to which we are comparing observations from the CPS. The CPS count of enrollees will be the dependent variable in our model and the socioeconomic and demographic variables will be the independent variables. Then v will account for the random error of the model and u will measure the distance between the CPS and MSIS. Once a

distribution for u has been selected and a model has been specified we can predict the value of u for each year and create a new variable that is the sum of the error prediction and CPS. If the stochastic frontier is accurately predicting the distance between CPS and MSIS, this new variable should be close or equal to MSIS.

Because our sample size is so small for an individual state, we chose to use a panel data stochastic frontier instead of a cross-sectional model. Defining states as the entities and year as the unit of time, we are able to increase our sample size by a factor of 15. It also allows us to use fixed effects to filter out any unobserved state-specific variation. In reality the data is not a true panel because of the household rotation scheme used in sampling for the CPS, but we make the assumption of state homogeneity to simplify our analysis. This is not unwarranted as the Census Bureau selects replacement households that are as similar as possible to those exiting the sample. Furthermore, Medicaid legislation is state-specific and so enrollment within a state should be fairly stable over time.

Results

Using Stata's *sfp* command (when Stata is open, type *ssc describe sfp*) we experimented with different combinations of independent variables to find a model that would most accurately predict the gap between CPS and MSIS. Any model that failed to converge, was missing estimates, or produced poor estimates (as measured by large p-values) was rejected. Because our sample size is small, Stata produced different estimates each time the command was executed, with some iterations performing better than others. Of the models deemed satisfactory, each was ran 10 times to find the best set of estimates as determined by the highest log likelihood value for each distribution of the one-sided error term. Predicted frontiers were then plotted and compared with MSIS to determine the accuracy of each model's gap estimates. Models with gap estimates that were either too high, too low, or a combination of both were discarded. The most successful model was

$$ASEC_{it} = \beta_0 + \beta_1 GDPPERCAPITA_{it} + \beta_2 POPNPOV_{it} + \beta_3 MEANHHI_{it} + \alpha_i + v_{it} - u_{it}$$

where subscripts i and t denote each state and year, respectively. The dependent variable, $ASEC$, was the CPS count of enrollees in state i during year t . $GDPPERCAPITA$ represented GDP per capita, $POPNPOV$ was the count of people living above the federal poverty limit, and $MEANHHI$ was the average household income for state i during year t . α_i was a true fixed-effects parameter to capture any state-specific effects (as specified by Greene, 2005) and the two-sided error term for state i in year t was given by v_{it} . Finally, the one-sided error term was u_{it} for which we used exponential, half-normal and truncated normal distributions. Below are the regression results for this model under different one-sided error distribution assumptions.

Table 1. Results from Estimating Stochastic Frontier Model: Y is CPS Enrollment Counts. One-sided Error Term has Exponential Distribution.

VARIABLES	Frontier	Usigma	Vsigma
GDP per capita	-0.066*** (0.002)		
Population Above Federal Poverty	0.677*** (0.000)		
Average Household Income	3.881*** (0.000)		
Constant		25.458*** (0.138)	-9.984 (158.257)

Table 2. Results from Estimating Stochastic Frontier Model: Y is CPS Enrollment Counts. One-sided Error Term has Half-Normal Distribution.

VARIABLES	Frontier	Usigma	Vsigma
GDP per capita	-17.159*** (0.643)		
Population Above Federal Poverty	1.160*** (0.000)		
Average Household Income	11.572*** (0.424)		
Constant		26.815*** (0.098)	-10.725 (432.939)

Table 3. Results from Estimating Stochastic Frontier Model: Y is CPS Enrollment Counts. One-sided Error Term has Truncated Normal Distribution.

VARIABLES	Frontier	Mu	Usigma	Vsigma
GDP per capita	-0.066*** (0.005)			
Population Above Federal Poverty	0.677*** (0.000)			
Average Household Income	3.881*** (0.001)			
Constant		-4.711e+08 3.71E+08	32.708*** (0.789)	-3.456 (16.698)
Observations (all models)	210	210	210	210
Number of States (all models)	15	15	15	15

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: Fixed effects estimates have been omitted, however all tested significant at the 99% level.

Although the SFA estimates are not perfect, they are an improvement in most cases compared to the CPS and particularly accurate for a few states. For example, predicted enrollments for Arizona, California, Colorado, and Kentucky were very close to the true values of enrollment. Figures 2-4 show plots of MSIS, CPS, and estimates of MSIS by state for the aforementioned cases under different distributional assumptions. The poorest gap estimates were produced by the half-normal while the exponential and truncated normal fared much better. The estimates under the latter two were almost identical, with the exponential only slightly outperforming the truncated normal distribution in a few cases. Despite some variation in estimates (both above and below MSIS) our predictions have relatively little error for these states when using the exponential or truncated normal distribution. Even during 1999-2001 when the CPS undercounts by a much larger degree than subsequent years the predicted value of enrollment is still fairly close to the MSIS.

Figure 2

True Fixed-Effects with Exponential Distribution

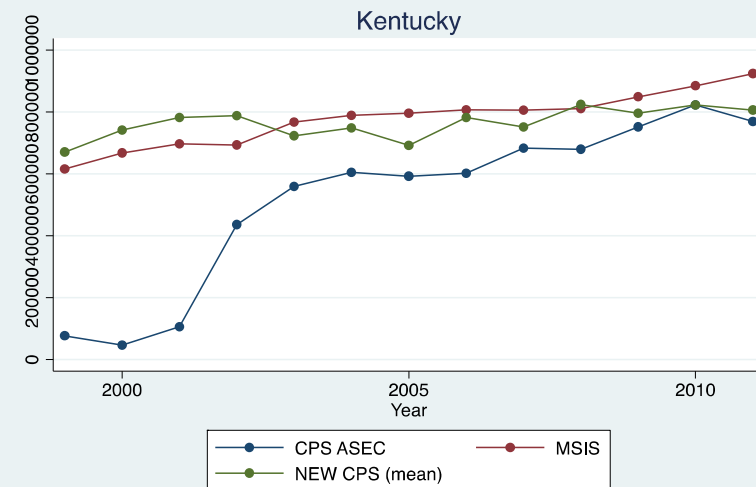
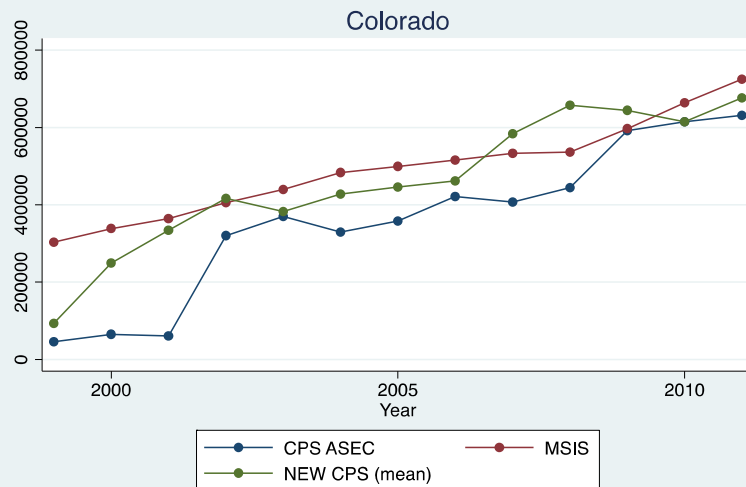
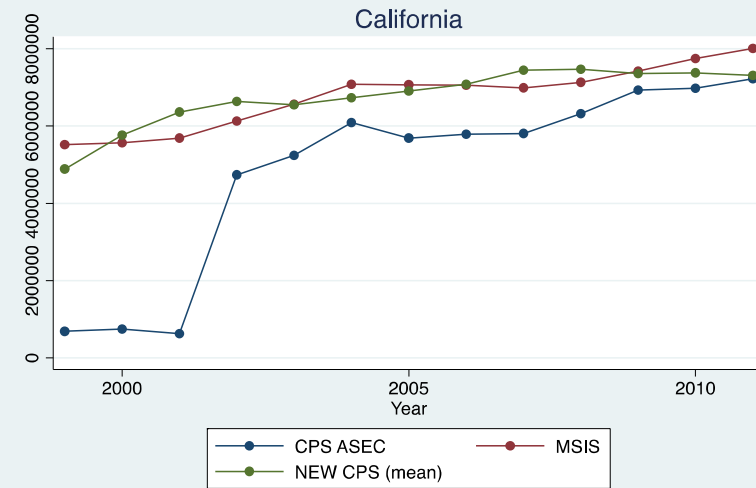
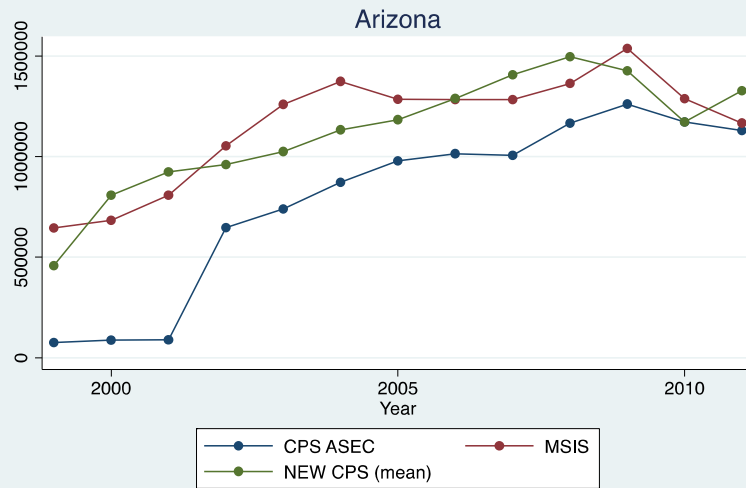


Figure 3

True Fixed-Effects with Half Normal Distribution

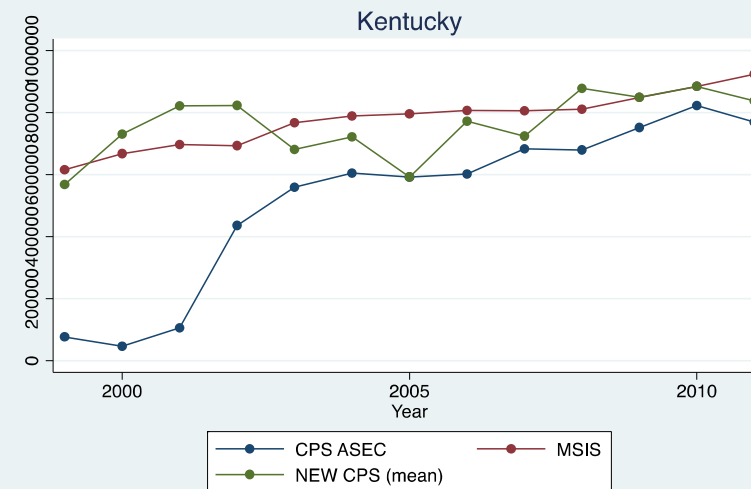
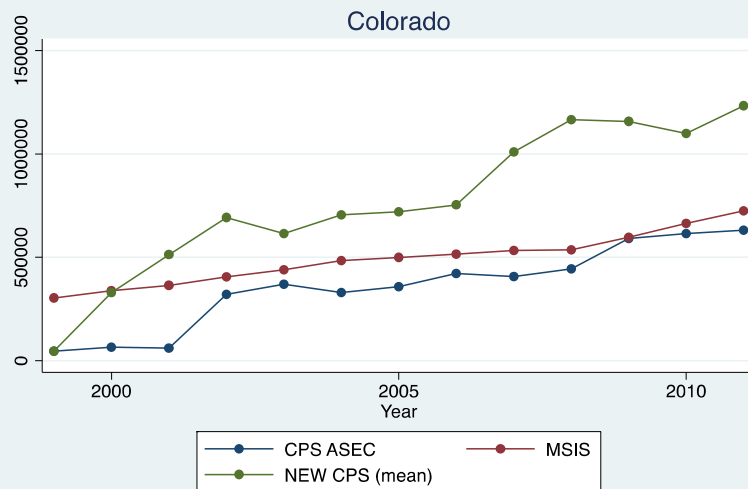
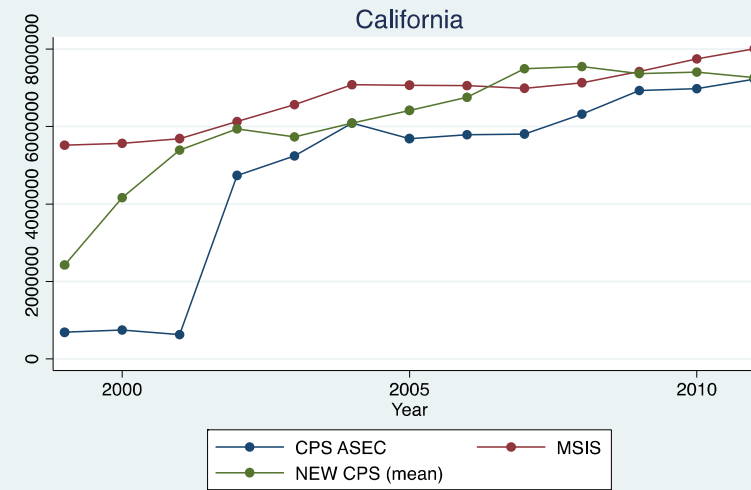
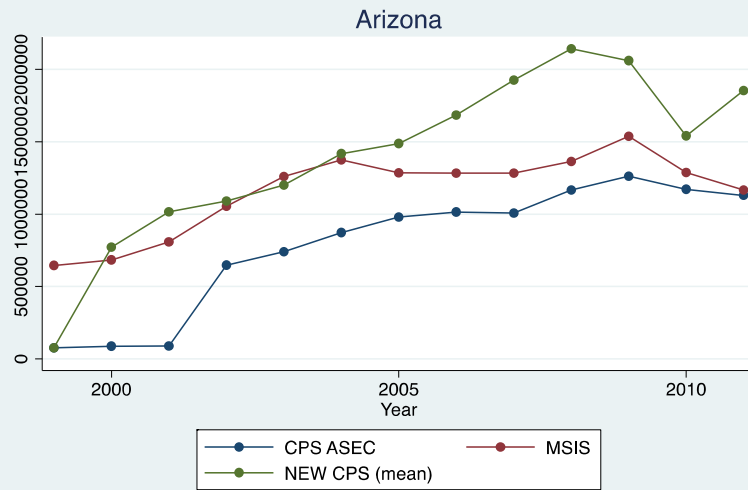
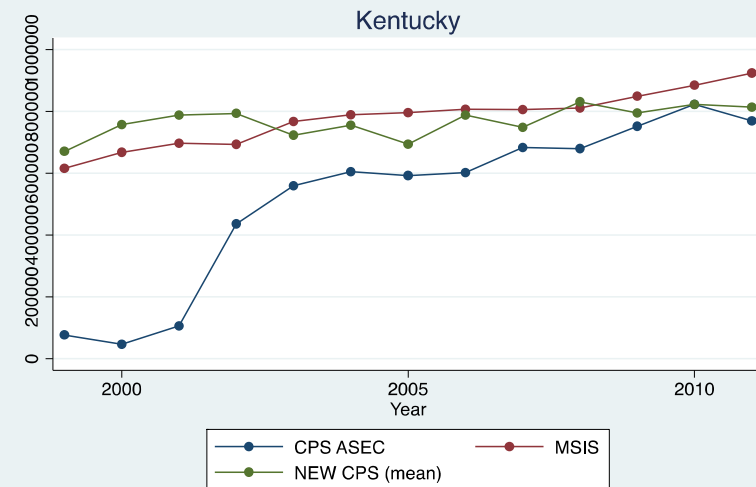
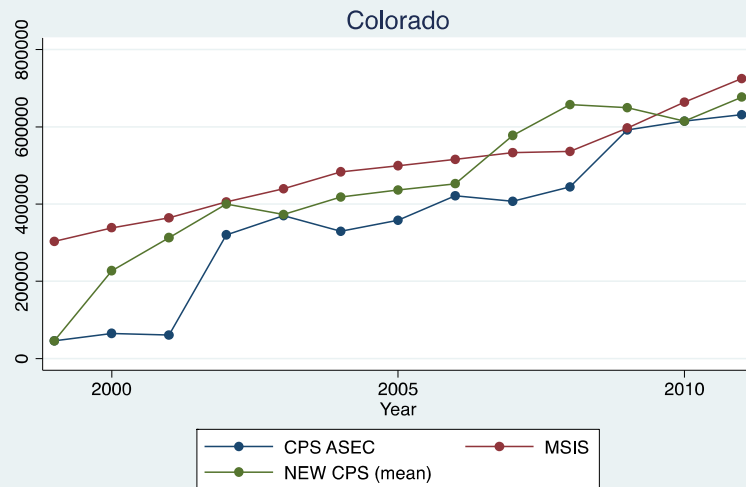
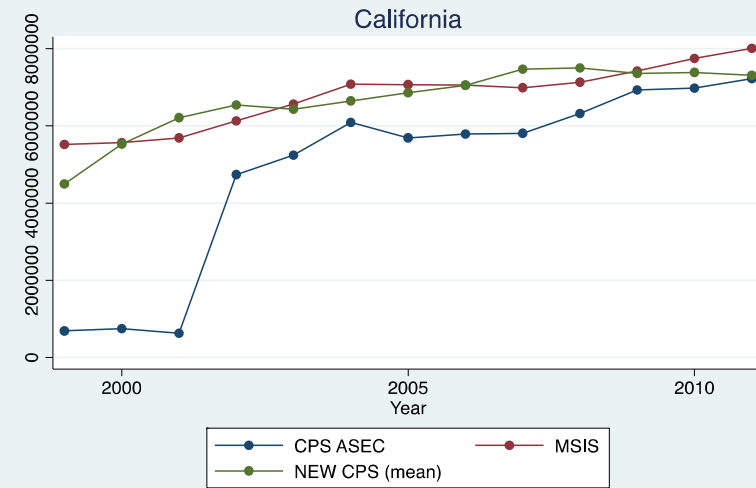
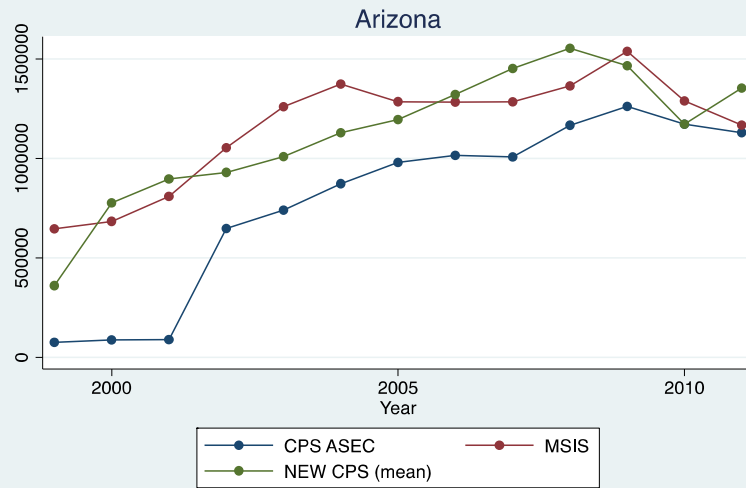


Figure 4

True Fixed-Effects with Truncated Normal Distribution



Conclusion

We have shown that stochastic frontier analysis can be used to shrink the gap between the observed value and the true value of Medicaid enrollment in certain states using only the CPS measurements of enrollment and data on GDP per capita, poverty, and average household income. Some states' enrollment levels were approximated more accurately than others, but on average our prediction is closer to MSIS records than the corresponding CPS counts.

A major limitation of this study is the short time period over which we are predicting. We could improve our estimates if we extended our sample to the full reporting period of the CPS (1980-present), however MSIS data is not available outside 1999-2012. Although our overall goal is to estimate the true value of Medicaid enrollment without any administrative data, this study seeks to first prove the veracity of SFA estimates. We cannot claim our model is a good predictor of MSIS if we extend to years where there is no available MSIS data.

Another more obvious limitation is choosing a subset of states for our sample. The most common SFA models are only applicable if the observed data falls below (or above) the frontier in every case. We are limited to only states in which the CPS undercounts enrollment in every year, reducing our sample size to 30% (15 states out of 51) of what is available. Furthermore, our goal was to predict enrollment for every state in the U.S., which we were not able to do. However this may be achieved by using two-tier stochastic frontier analysis instead. Two-tier SFA, developed by Polachek and Yoon (1987, 1996), decomposes the error term into three parts: a two-sided component, a one-sided nonnegative component, and a one-sided nonpositive component. This model is equipped to handle data which is a mixture of observations that may lie above or below the frontier and would allow us to use data from any state, regardless of

whether or not the CPS counts are above the MSIS frontier. A future study using two-tier SFA may produce even better results than we have shown here.

References

- Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and Estimation Of Stochastic Frontier Production Function Models. *Journal Of Econometrics*, 6(1), 21-37.
- Blewett, L., Brown, M. E., Call, K. T., & Davern, M. (2004). Monitoring the Uninsured: A State Policy Perspective. *Journal Of Health Politics, Policy And Law*, (1), 107.
- Blewett, L. A., & Davern, M. (2006). Meeting the Need for State-Level Estimates of Health Insurance Coverage: Use of State and Federal Survey Data. *Health Services Research*, 41(3), 946-975. doi:10.1111/j.1475-6773.2006.00543.x
- Davern, M., Blewett, L., Bershadsky, B., Call, K., & Rockwood, T. (2003). State variation in SCHIP allocations: how much is there, what are its sources, and can it be reduced?. *Inquiry (00469580)*, 40(2), 184-197 14p.
- Davern, M., Klerman, J. A., Baugh, D. K., Call, K. T., & Greenberg, G. D. (2009). An Examination of the Medicaid Undercount in the Current Population Survey: Preliminary Results from Record Linking. *Health Services Research*, 44(3), 965-987. doi:10.1111/j.1475-6773.2008.00941.x
- Davern, M., Klerman, J. A., Ziegenfuss, J., Lynch, V., & Greenberg, G. (2009). A partially corrected estimate of medicaid enrollment and uninsurance: Results from an imputational model developed off linked survey and administrative data. *Journal Of Economic & Social Measurement*, 34(4), 219-240. doi:10.3233/JEM-2009-0324
- Flood, S., King, M., Ruggles, S., & Warren, J. R. Integrated Public Use Microdata Series, Current Population Survey: Version 4.0. [Machine-readable database]. Minneapolis: University of Minnesota, 2015.

- Glied, S., Remler, D. K., Zivin, J. G., Ferry, D. H., Gil, I., & Neufeld, A. (2002). Inside the Sausage Factory: Improving Estimates of the Effects of Health Insurance Expansion Proposals. *The Milbank Quarterly*, (4). 603
- Greene, William. "Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model." *Journal of Econometrics* 126.2 (2005): 269-303. Web. 31 Mar. 2016.
- Klerman, J. A., Davern, M., Call, K. T., Lynch, V., & Ringel, J. D. (2009). Understanding The Current Population Survey's Insurance Estimates And The Medicaid 'Undercount'. *Health Affairs*, 28(6), w991-w1001. doi:10.1377/hlthaff.28.6.w991
- Klerman, J., Ringel, J., & Roth, B. (2005). Under-Reporting of Medicaid and Welfare in the Current Population Survey. 84 pages.
- Lewis, Kimball, Marilyn Ellwood, and John Czajka. 1998. "Counting the Uninsured: A Review of the Literature." The Urban Institute, Washington, DC.
- Pascale, J., Roemer, M. I., & Resnick, D. M. (2009). Medicaid Underreporting in the CPS: Results from a Record Check Study. *Public Opinion Quarterly*, 73(3), 497-520.
- Polachek, Solomon W., and Bong Joon Yoon. "A Two-Tiered Earnings Frontier Estimation of Employer and Employee Information in the Labor Market." *The Review of Economics and Statistics* 69.2 (1987): 296. Web. 31 Mar. 2016.
- Polachek, Solomon W., and Bong Joon Yoon. "Panel Estimates of a Two-tiered Earnings Frontier." *J. Appl. Econ. Journal of Applied Econometrics* 11.2 (1996): 169-78. Web. 31 Mar. 2016.