Identification, Replication and the Effect on Employment of the TennCare Public Insurance Contraction

John C. Ham
National University of Singapore, IFAU, IRP and IZA

Ken Ueda
National University of Singapore

We would like to thank, without implicating, Thomas Buchmueller, Amy Finklestein, Robert Moffitt, Matthew Notowidigdo, and Lara D. Shore-Sheppard, as well as colleagues and seminar participants at the University of Maryland, the NBER, and the National University of Singapore, for very helpful comments. Ham thanks the National Institutes of Health for financial support. Any opinions, findings, and conclusions or recommendations in this material are those of the authors and do not necessarily reflect the views of the National Institutes of Health. We both thank the University of Maryland and the National University of Singapore for financial support. Of course, we take responsibility for all errors, and emphasize that our thanks to the individuals above in no way implies that they agree with the arguments in this paper.

ABSTRACT

In a recent paper, Garthwaite, Gross and Notowidigdo (2014) report large positive labor supply effects of a major contraction in public insurance coverage in Tennessee, announced at the end of 2004 and implemented in mid-2005, using data from the March CPS. These results are important given the expansions of Medicaid coverage under the Affordable Care Act and the potential for large Medicaid contractions under President Trump and the Republican Congress. Their results are surprising given the previous work on the employment effects of health insurance expansions, but the authors argue that these differences are due to the fact that the Tennessee program went much higher into the income distribution than the programs studied by other researchers.

In this paper we show, under reasonable parameter restrictions, that the framework used by Garthwaite, Gross and Notowidigdo (2014) only allows for estimating the lower bound on the labor supply response to the contraction, which makes their results all the more striking. However, we show next that their large estimates are the result of focusing on the March CPS in estimation. When we use their estimation strategy on a dataset based on all the months of the CPS, or a dataset based on the American Community Survey, we find much smaller, and sometimes negative, estimates of the lower bound on the labor supply response. Note that compared to the March CPS, these alternative datasets offer much larger sample sizes and are not affected by seasonal factors.

We attempt to distinguish between the estimates across databases using placebo tests. While these tests reject many estimates, there is still a very wide range in the surviving estimates. Hence we conclude that, at best, we do not have good estimates of the treatment effect of interest.

1. Introduction

There is a large body of literature on the labor supply effects of Medicaid eligibility, and it is fair to say that most of the literature argues against large effects for this zero-one dummy variable.[1] For example, Yelowitz (1995), from the March CPS, found large labor force participation effects of being able to obtain Medicaid coverage while working. However, Ham and Shore-Sheppard (2005), also using the March CPS and the Survey of Income and Program Participation (SIPP), showed that his results were an artifact of constraining welfare benefits and Medicaid availability to have the same coefficient. Once this constraint was relaxed, welfare benefits, but not Medicaid eligibility, continued to affect labor force participation. Further, Meyer and Rosenbaum (2001), using data from the Current Population Survey (CPS) Outgoing Rotation Group Files and from the March CPS, found an important role for welfare benefits, but not Medicaid provisions, in a static model of labor force participation. Recently Finkelstein et al. (2014) found that offering individuals Medicaid coverage in the Oregon Health Experiment had essentially no effect on participation or employment; since their result is based on a randomized trial, this evidence is perhaps the strongest to date.[2]

In a recent paper, Garthwaite, Gross and Notowidigdo (2014, hereafter referred to as GGN), reported large labor supply effects due to a major contraction in Medicaid coverage in the

[1] See Buchmueller, Ham and Shore-Sheppard (2016) for a summary of this research

[2] Note that one study that took a more sophisticated approach allowing for heterogeneous treatment effects found Medicaid effects. Specifically, Moffitt and Wolfe (1992) consider a reduced form model of employment and Medicaid participation, where an important independent variable of interest is the imputed value of Medicaid for a given family. They allow the Medicaid effect to vary across families and find that that the value of Medicaid matters for families with high expected medical expenses.

Tennessee program TennCare, using data from the March CPS. Specifically, in a program change announced at the end of 2004 and implemented in mid-2005, childless adults were no longer eligible for Medicaid coverage. GGN argued that their results were not necessarily inconsistent with the above studies, for example with Finklestein et al. 2014, since TennCare had been much more generous before the contraction relative to Oregon. Specifically, in Tennessee individuals could have earnings up to four times the poverty line and be eligible for Medicaid before the contraction, while eligibility in Oregon was restricted to individuals below the poverty line. Thus, GGN were considering a very different population than Finklestein et al. 2014, so there is little reason that the results from the two studies should be comparable.[3] While GGN's logic is correct, it is fair to say that their results imply by far the biggest labor supply response to Medicaid seen in the literature. A specific prediction of their paper was that the Affordable Care Act would have big labor supply effects once it was in place, since a major innovation of the ACA was to cover childless adults, albeit with incomes only up to 133% of the poverty line.[4] Furthermore, GGN's estimates also suggest large potential employment effects if President Trump and the Republican Congress repeal the large Medicaid expansion made possible by the Affordable Care Act.

In this paper we show that, in the language of modern econometrics, the treatment effect on labor supply in GGN's model is not point-identified, as they are estimating a reduced form employment equation. However, we also show that their framework does allow for partial identification of the labor supply effect of the Tennessee contraction (under reasonable parameter

[3] The TennCare population is closer to Finkelstein et al than to previous studies, since the latter tended to look at single mothers.
[4] See, e.g., Buchmueller (2016) for a discussion of the changes brought about by the ACA.

restrictions), as they can estimate a lower bound for the labor supply response to the contraction that they study. This of course makes the implication of their large estimated coefficients even more important. However, we show next that their large estimates are the result of focusing on the March CPS in estimation, and when we use their estimation strategy on a dataset based on all the months of the CPS, or a dataset based on the American Community Survey, we find much smaller, and sometimes negative, estimates of this lower bound. (We note that when compared to the March CPS, these two datasets offer much bigger sample sizes and are not affected by seasonal factors.)

We attempt to distinguish between the estimates from different datasets using placebo tests. While these tests reject many estimates, there is still a very wide range in the surviving estimates between, for example, the large triple difference (hereafter DDD) estimates from the March CPS and the much smaller, and sometimes significantly negative, DDD estimates based upon the American Community Survey. Hence, we conclude that, at best, the data are relatively uninformative about their parameter of interest.

The outline of the paper is as follows. In the next section we briefly review the TennCare contraction and the GGN approach to estimating its effect on labor supply using data from 2000 to 2007. In Section 3 we examine the identification of the TennCare effect on labor supply in GGN's framework and show that it is only possible, at best, to estimate a lower bound for this parameter. In Section 4 we replicate their estimation for not only the March CPS, but also the Basic March CPS, a dataset consisting of all months of the CPS, and a dataset based on the American Community Survey. In Section 5 we implement placebo tests with the aim of discriminating between the wide range of estimates that our replication exercise produces. However, while these tests do eliminate a number of estimates, we are still left with a wide range of estimates for the lower bound of the TennCare contraction on labor supply. Section 6 concludes the paper.

2. The TennCare Contraction and the GGN Approach to Estimating the Effect of its Contraction.

2.1 The TennCare Program [5]

TennCare started in 1994 with an enrollment cap of 1.775 million people and with 12 licensed managed care organizations, which included 8 HMOs and 4 PPOs. The goal was to provide insurance to people who were Medicaid ineligible but were either "uninsured" or "uninsurable." The eligible pool for TennCare started with people who were rejected by private insurance plans. TennCare eventually expanded to include uninsured children (without income restrictions) age 17 and older whose parents did not have access to workplace insurance, "dislocated workers" (i.e. displaced workers in the literature), and loosened income restriction levels. As a result of these and other expansions, TennCare had very generous coverage relative to other Medicaid beneficiaries. In 1995 roughly 40% of TennCare enrollees had incomes above 100% of the poverty line, and 1.3% had incomes above 400% of the poverty line (Wooldridge 1996).[6]

In 2001 TennCare encountered a 342 million dollar shortfall, resulting in their largest MCO, Blue Cross Blue Shield of Tennessee (BCBST), threatening to pull out of TennCare.[7] In

[5] Except when otherwise noted, in this section we draw on GGN, Wright (2001), Wooldridge (1966), and Chang and Steinberg (2009).
[6] This distinction is crucial for reconciling the GGN estimates with previous work looking at much poorer recipients.
[7] BCBST covered almost half of all TennCare patients at the beginning of 2001. They stated that rising costs could force them to withdraw.

response, TennCare required a medical review of whether enrollees were "insurable" and began a process called "reverification," which required enrollees to schedule appointments in order to determine eligibility benefits. The result was 100,000 individuals being removed from the Medicaid rolls, and roughly 20% of the TennCare enrollees being moved from the expansion program to traditional Medicaid.[8]

In November 2004 Governor Philip Bredesen announced that as of July 1, 2005, TennCare would no longer cover adults over 19 who did not qualify for traditional Medicaid. From December 2004 to June 2005, there was substantial discussion of the forthcoming July 2005 contraction in the press, and it is plausible that at least some of those who would be affected starting looking for a job with health insurance. The disenrollment process began in July 2005.

2.2 The GGN Approach to Estimating the Effect on Labor Supply of the TennCare Contraction

Using data aggregated by state from the March CPS files for 2000-2007, GGN first estimate a difference-in-difference (hereafter DD) version of the following equation

$L_{st} = \alpha_s + \gamma_t + \beta\, I[Tenn_s] * I[Year_t \geq 2006] + v_{st,}$ (1)

where L_{st} is employment for state s in year t, $I[Tenn_s]$ is an indicator function equal to 1 if the state is Tennessee and 0 otherwise, while $I[Year_t \geq 2006]$ is an indicator function equal to 1 if the year is 2006 or 2007 and 0 otherwise. Their sample consists of aggregated observations from Tennessee and thirteen other Southern States.

GGN note that estimating (1) will not provide a consistent estimate of β if the employment equation contains state specific trends:

[8] More details on the composition of enrollees can be found in GGN (2011).

$L_{st} = \alpha_s + \gamma_t + \beta\, I[Tenn_s] * I[Year_t \geq 2006] + t * \theta_s + v_{st},$ (2)

However, to address this possibility, they note that the TennCare contraction did not affect individuals with children, and only individuals without children (under 17 years old), so they use triple difference (hereafter DDD) estimation, where the dependent variables now contain aggregate employment for individuals with children and individuals without children, respectively, from each of the above states.[9] They go to some length to argue that β is the effect of the TennCare contraction on labor supply, as opposed to, for example, labor demand.

We differ from the previous literature by raising the anticipation issue mentioned in Section 2.1. We believe this issue is important because it affects whether 2005 should be considered a comparison year or a treatment year. One could argue that this issue suggests that the March CPS used by GGN is superior, for 2005, to using annual data for 2005 (as we do below) since the disenrollment started after March 2005. However, this argument depends crucially on the assumption that individuals did not react until July 2005 to the scheduled contraction, i.e., they were unaffected by Governor Bredesen's announcement in November 2004 and the subsequent discussion in the press. In future work we will report on how the results discussed below are affected by dropping 2005 observations from the analysis.

3. Identification of the Effect of the TennCare Contraction on Labor Supply

In this section, we first show that DD estimation of β in (1), or DDD estimation of β in (2), does not provide a consistent estimate the effect of the TennCare contraction on labor supply. However,

[9] Individuals with children above 17 were still considered individuals without children.

we then show that this estimation produces a consistent estimate of a lower bound for the effect of the TennCare contraction on labor supply. To show this, we first write the structural equations for demand and supply respectively as

$$L_{st}^{\text{sup}} = \alpha_s' + \delta_t' + \pi_1 w_{st} + \pi_2 I[Tenn_s] * I[Year_t \geq 2006] + \varepsilon_{st}', \quad (3)$$

$$L_{st}^{\text{dem}} = \lambda_s' + \mu_t' + \phi_1 w_{st} + \phi_2 I[Tenn_s] * I[Year_t \geq 2006] + u_{st}'. \quad (4).$$

In (3) and (4) L_{st} is employment for state s in year t, w_{st} is the wage for state s in year t. α_s' and δ_t' represent the state and year effects within the labor supply equation, while λ_s' and μ_t' represent the same effects within the labor demand equation. π_1 and π_2 represent the effects of the wage and the TennCare contraction on labor supply respectively, while ϕ_1 and ϕ_2 represent the same effects for labor demand.[10] We expect $\pi_1 \geq 0$, $\pi_2 \geq 0$, $\phi_1 \leq 0$, and $\phi_2 \leq 0$. GGN do not take a stand on whether the TennCare contraction affects labor demand. We believe that it is plausible that this contraction will decrease the quantity of labor demanded at a given wage, since now workers will expect employers to provide health insurance, driving up the total cost of each unit of labor. However, our results are not qualitatively changed if $\phi_2 = 0$ and the contraction does not affect the demand curve (4).

To investigate the parameter that GGN do estimate, we solve for the model's reduced form by setting $L_{st}^{\text{sup}} = L_{st}^{\text{dem}} = L_{st}$. After solving for L_{st} we have

$$L_{st} = \Gamma_{st} + \beta\, I[Tenn_s] * I[Year_t \geq 2006] + v_{st}, \quad (5)$$

[10] Since L refers to employment, there is no possibility of a backward bending supply curve, and π_1 must be positive.

where $\beta = \left(1 - \frac{\pi_1}{\phi_1}\right)^{-1}\left(\pi_2 - \frac{\pi_1\phi_2}{\phi_1}\right)$, $v_{st} = \frac{\pi_1 u_{st} - \phi_1 \varepsilon_{st}}{\pi_1 - \phi_1}$, $\Gamma_{st} = (\alpha_s + \delta_t) + \frac{\pi_1}{\pi_1 - \phi_1}\big((\lambda_s - \alpha_s) + (\mu_t - \delta_t)\big)$. By comparing (1) with (5), one sees that GGN are estimating a reduced form employment equation. In particular, β is the reduced form effect of the TennCare contraction on employment. To compare the actual labor supply effect of the TennCare contraction, π_2, with the reduced form employment effect β, we note that (5) implies

$$\pi_2 = \left(1 - \frac{\pi_1}{\phi_1}\right)\beta + \frac{\pi_1\phi_2}{\phi_1}.$$

Since $\phi_1 < 0$ and $\pi_1 > 0$, we have $\left(1 - \frac{\pi_1}{\phi_1}\right)^{-1}\beta > \beta$. Further, since $\phi_1 < 0$, $\phi_2 < 0$, and $\pi_1 > 0$, $\frac{\pi_1\phi_2}{\phi_1} > 0$ and hence

$$\pi_2 = \left(1 - \frac{\pi_1}{\phi_1}\right)\beta + \frac{\pi_1\phi_2}{\phi_1} > \beta. \quad (6)$$

In other words, an estimate of β will, on average, underestimate π_2. In Figure 1, we show the relationship between π_2 and β in a simple labor demand and supply diagram.[11]

If the TennCare contraction does not affect the labor demand equation, $\phi_2 = 0$ and

$$\pi_2 = \left(1 - \frac{\pi_1}{\phi_1}\right)\beta = (1 - \frac{eta_1}{eta_1})\beta > \beta, \quad (7)$$

where eta_1 and eta_2 are the labor supply and demand elasticities respectively (assuming that they are constant).[12] Again, an estimate of β will, on average, provides a lower bound for π_2. In Figure 2, we show the relationship between π_2 and β for this case.

[11] Here we are assuming that $\beta > 0$ since this is implied by our model – see Figure 1.

[12] We thank Professor Notowidigdo for noting that (7) can be written in terms of the labor supply and demand elasticities.

To estimate the structural labor supply equation (3), one would need to add exogenous variables X_{st} and Z_{st} to (3) and (4) respectively. If Z_{st} contains at least one variable not in X_{st}, then the order condition for identification is satisfied. Assuming that the rank condition is also satisfied, then (3), and hence the labor supply effect of the TennCare contraction, is identified.[13] Unfortunately, we were unsuccessful in finding a variable that plausibly affected labor demand but not labor supply and was not a weak instrument.[14]

Hence, our analysis suggests that the labor supply effect of the TennCare contraction is not (point) identified. It is partially identified in the sense that an estimate of the reduced form TennCare coefficient provides an estimate of the lower bound for π_2. Since GGN's estimate of β is quite large, this implies that, in an expected value sense, π_2 also will be quite large (in contrast to the previous literature) and hence their approach is, in fact, quite informative. We now investigate the robustness of this result when we use data sets other than the March CPS.

4. DataSets Used for Replication

4. DataSets Used in Estimation

We will estimate GGN's equations (1) and (2) using their sample selection criterion on individuals from five different datasets. The first dataset is the CPS March Supplement (MCPS). Since this is also the data GGN use, we are essentially trying to replicate their results, and find that we can do this relatively well. (They post a data appendix but not the actual dataset they used.) When we use the MCPS from 2000-2007, we have approximately 250,000 observations before collapsing the microdata. Following their DD strategy, we aggregate the micro observations to the

[13] This discussion, of course, dates back to the Cowles Foundation, see, e.g., Hood and Koopmans (1953).
[14] If we have one excluded variable, but it is a weak instrument, the rank condition will not be satisfied asymptotically.

state level (for the states they used), while we aggregate to the state-child/state-childless level to replicate their DDD estimation. With this data and the other four datasets discussed below, we implement their sample selection strategy.

We use three other datasets, and aggregate the micro data as above for the DD and DDD approaches. Our first additional dataset is drawn from the March Basic Monthly File (BMCPS), which consists of 167,000 microdata observations from 2000 to 2007. Next, we pool all of the Basic CPS monthly surveys within a calendar year to form an "All CPS" dataset (AllCPS). This gives us 2.06 million micro observations over 2000 to 2007, which is (not surprisingly) a substantially larger sample than either of the two March datasets. Next, we draw from the 5% sample of the 2000 decennial Census data, the 2001-2004 pilot American Community Survey data, and the 2005-2007 American Community Survey (ACS) data. Hereafter, we refer to this combination as the ACS data; it contains 3.04 million micro observations, the largest dataset used in this analysis. The main difference between the 2005-2007 ACS and the 2001-2004 ACS data is that the 2005-2007 ACS has a representative sample from every county, while this is not true of the 2001-2004 data. Since we follow employment rates at the state level and not the county level, combining the ACS data may introduce problems in estimation. We therefore repeat our analysis using 2005-2007 instead of 2000-2007, and find that the results are very similar.

We specifically impose the GGN sample restrictions on each dataset as follows. In each of the datasets, education, age, and occupation are available, so we can ensure that the individual is between 21 and 64, and does not have an advanced degree. We use the occupation to determine the military status of the individual, and we exlucde those who were part of the military.We use the "age of the youngest own child in household" variable in the ACS, AllCPS, BMCPS and the

MCPS.[15] For all datasets, we use the state of residence to determine whether the respondent lived within Tennessee or one of the other southern states. We use the worker's employment status, which is defined as employment when the survey is administered, and usual hours worked per week for the five databases.

We note that there are potentially important differences between the databases. First, the ACS and AllCPS, while of course the MCPS and BMCPS survey people solely during March. Thus, the latter two datasets may be affected by seasonal factors, but this will be important only if these factors affect the treatment and comparison groups differently. Another difference is that in 2000 and 2005-2007, the ACS is more likely to draw from metropolitan areas relative to the AllCPS, BMCPS, and the MPS (2001-2004 ACS pilot data do not have urban/rural information). Below we investigate the likely sensitivity of these results to these differences, and conclude there is little evidence that these factors drive the differences in the estimation results in our analysis.

Finally, we consider differences in how the ACS and CPS ask about employment status. As Vroman (2003) and Kromer and Howard (2011) discuss, the CPS asks 4 questions: *'LAST WEEK, did you do ANY work for (pay/either pay or profit); LAST WEEK, did you do any unpaid work in the family business or farm?; LAST WEEK, did you have more than one job/job or business, including part time?; Altogether, how many jobs/jobs or businesses did you have?* In contrast, the ACS prior to 2008 (the years we use) asked only one question*: LAST WEEK, did this person do ANY work for either pay or profit?* (Kromer and Howard 2011). The CPS, as a result, captured more unpaid work and therefore had a slightly higher employment rate. Hence, if the TennCare contraction caused individuals to leave unpaid work for paid work in order to obtain

[15] If "age of the youngest own child in household" is missing or above 17 in the ACS/BMCPS/AllCPS/MCPS, then the adult is flagged as not having any children. Otherwise, the adult is flagged as having children. "Missing" is interpreted as not having any children at all.

health insurance, this will not be counted as an increase in employment in the CPS; note that this will not be an issue in the ACS.

4. Empirical Results

4.1 Basic Results

In Panel A of Table 1 we present the baseline DD estimates for the treatment effect of the Medicaid contraction on employment based on equation (1). In the first column, we copy the results from GGN; they estimated a coefficient (standard error) of 2.5 (1.1) percentage point (hereafter PP) increase in Tennessee's employment rate relative to other southern states as a result of the TennCare contraction. In column (2), we present our estimates from the MCPS, and find that for all practical purposes we replicate their results, since we find a 2.2 (1.0) PP increase.[16] In column (3) we report the results for the BMCPS, and find a treatment effect of 2.0 (1.1) PP.

We start to diverge from the GGN results when we use the AllCPS data, since now we estimate a significant but smaller treatment effect of 1.3 (0.4) PP – see column (4). This estimate is about half of GGN's estimated treatment effect. We diverge even further from the GGN results when we use the ACS, since we estimate a significantly *negative* treatment effect of -1.1 (0.4) PP.

Panel B of Table 1 presents the triple difference (hereafter DDD) estimates for the different datasets. The first column contains GGN's estimate of 4.6 (2.0) PP. Columns (2) and (3) show our estimated treatment effects for the MCPS and BMCPS as 5.5 (2.0) PP and 7.0 (2.4) PP, respectively. As in GGN, the estimated DDD treatment effect is at least 100% larger than the DD estimated effect when we use any variant of the March CPS. For the AllCPS, the DD and DDD

[16] Recall that their actual data are not available online.

estimated effects are very similar at 1.3 (0.4) PP and 1.4 (0.8) PP respectively, but now the AllCPS DDD estimate is between one-third and one-quarter of the DDD estimates from the March CPS datasets. The estimated effect from the ACS of 0.2 (0.6) is small and insignificant. It is well known that the DD and DDD estimates will differ if there are different linear trends for the comparison and treatment groups. In Table 1 the main differences are between the DD and DDD estimates for the March CPS datasets, but it is not clear whether these are significantly different.[17]

The upshot of these results is that the GGN results do replicate for versions of the March CPS, but not for the other datasets we consider. The DD and DDD employment effects, or the lower bound on the labor supply effect, are much smaller for the AllCPS data, and significantly negative or insignificant (with a relatively small confidence interval), respectively for the ACS. Since the AllCPS and ACS datasets are much larger than either of the March CPS datasets, and will not be affected by seasonality, they arguably should be preferred to either of the March CPS datasets. However, we must also investigate whether the identifying assumptions underlying the DD and DDD estimation are satisfied for the March CPS but not the AllCPS or ACS datasets, and we do this below.

4.2 Allowing Treatment Effects to Vary by Demographic Characteristics

4.2.1 Overview

GGN also estimate treatment effects (separately) across hours worked, age, and education groups. Since TennCare was much more generous than the versions of public insurance studied in other papers, and GGN is the only paper that finds large effects (especially with DDD estimation), one

[17] If the DD estimates were efficient, we could use a Hausman test here. Given the DD estimates are not efficient, we would have use the bootstrap to test the null hypothesis that they are equal.

would expect that the GGN effects come from workers who have higher earnings and thus would have not been covered by the public insurance plans analyzed in previous work. GGN find significant treatment effects for two (not mutually exclusive) groups likely to comprise higher earners in the TennCare eligible sample: i) those with more hours of work per week; ii) those with more education. We also compare the effects for younger and older workers.

Given that we are using a number of datasets, a full discussion of these results would be lengthy. We summarize the results as follows. For individuals outside subsets i)-iii) above, all datasets produce insignificant results. For subsets i)-iii), the results generally mimic Table 1 by dataset. Specifically, we see large positive effects of the TennCare contraction from the March CPS, with the DD estimates being about half of the DDD estimates. Further, when we use the AllCPS, the effects are generally relatively small, but positive and statistically significant. Moreover, for the ACS, the DD estimated effects are negative and significant, whereas the DDD estimates are insignificant. Therefore, below we discuss only results that deviate from the above for the respective subsets of the data.

4.2.2 Estimating Treatment Effects That Vary by Hours of Work

Table 2A contains the DD results by hours worked; note that we have copied the DD estimates from Table 1 to make comparisons easy. We analyze four different outcomes: the percent working more than 0 but less than 20 hours per week; the percent working more than 20 hours per week; the percent working between 20 and 35 hours per week; the percent working more than 35 hours per week. These are the same outcomes GGN analyzed. We see insignificant DD effects for working between 0 and 20 hours per week and between 20 to 35 hours per week. The only qualitative deviation from the summary of the results in Section 4.2.1 for working more than 35 hours per week is that the DD estimate for the BMCPS is insignificant at 1.8 (1.2) PP. Table 2B contains the DDD results by hours worked; we have included the DDD estimates from Table 1 at the top of this table. There are three deviations from the general summary in Section 4.2.1. First, the estimated DDD coefficient from GGN is insignificant at 2.6 (2.1) PP and is half the size of the effect we estimate from the MCPS and BMCPS. Second, the DDD estimate using the AllCPS is significant for the 20-35 hours per week, as is the ACS DD estimate. However, the latter two estimated coefficients are relatively small and do not have much overlap with the significant estimates for the March CPS. Therefore, again we argue that the AllCPS and ACS datasets produce qualitatively different results than the March CPS.

4.2.3 Estimating Treatment Effects That Vary by Education Level

We next follow GGN and examine how education levels affect the estimated effects. Unlike the coefficients by hours worked, overall we do not find a clear-cut relationship between size of treatment effects and higher education (and hence those likely to have higher earnings). The DD results for those with a high school degree or more follow the pattern described in Section 4.2.1. (Recall that those with more than a college degree are not included here.) But the DD results for those with less than a high school degree deviate from the summary in Section 4.2.1 in that we find significant results from the AllCPS and ACS of 2.9 (1.1) PP and -2..2 (0.9) PP, respectively, which are actually larger in absolute value than the DD coefficients in Table 1 for these datasets.

The DDD results for those with more schooling follow the narrative in Section 4.2.1, except that the estimated coefficient of 3.4 (2.3) PP reported by GGN for those with more than high school education is not statistically significant. Also, for those with less schooling, the MCPS estimates are not only significant but are more than twice the size of the respective estimates in Table 1. The AllCPS estimate for those with less than a high school education is significant, unlike the respective element in Table 1, and is now much larger at 4.4 (2.2) PP.

4.2.4 Estimating Treatment Effects That Vary by Age

Following GGN, we separated those between 21-39 years from those between 40-64 years. To interpret different (unconditional) treatment effect by age group, we need to determine which group will be the higher earners among those previously eligible for TennCare, since earnings tend to be higher for older workers conditional on schooling level, but the average level of schooling will be lower for the group of older workers. To investigate this, we calculated the fraction of those earning between 100%-400% of the poverty line in 2005 for each age group. The results are shown in Table 4. For the younger workers, none of the DD or DDD estimates across the datasets, are significant; hence, they follow the pattern described in Section 4.2.1. The DDD estimates also follow this pattern, except that the coefficient for the AllCPS is now insignificant.

5. Which Estimates Should One Use for Policy

5.1 Overview

The March CPS, the dataset based on all the CPS months, and the ACS produced substantially different coefficients, so the question is: which estimates are appropriate for policy? One answer to this question is that none of these estimates is particularly useful, since we cannot replicate them with another dataset. However, before drawing such a negative conclusion, it is useful to explore whether we can explain the differences in the results across datasets by focusing on the characteristics of the datasets. We explore this in Section 5.2. Finally, we use specification/placebo tests to help choose between models in Section 5.3.

5.2 Investigating Differences in the Estimates Across Datasets

For each dataset, the DD and DDD estimates will differ if there are different linear trends for the comparison and treatment groups. Hence, the DD estimates from the MCPS and BMCPS could differ from the AllCPS and ACS if the linear trends differ by season, while the DDD estimates could differ if higher order trends differ by season. If so, we may be able to reduce the differences between the March datasets and the annual datasets if we remove highly seasonal employment in manufacturing and construction from each dataset. We do so in Table 5. The results duplicate the results from Table 1, except that the ACS DD estimate, while still negative, is no longer significant.

Another source of the differences could be differing proportions of urban and rural workers across the datasets, combined with parameter heterogeneity across these groups. For example, the ACS and MCPS have 16.5% and 22%, respectively, of their respondents explicitly stating they do not live within a metropolitan area. We investigate whether this difference is important in Table 6 by restricting the data to include only those residing within a metropolitan area. However, the only major difference from Table 1 is that now the AllCPS DD and DDD estimates are no longer statistically significant.

Finally, we investigate whether the 2000 decennial and the 2001-2004 ACS pilot datasets are inconsistent with the 2005-2007 ACS. While we do not expect the combination of these datasets this to be an issue, we re-ran our analysis for 2005-2007 in Table 7: the only difference from Table 1 is that the ACS DD estimate is no longer statistically significant. We note that the MCPS, BMCPS, and AllCPS coefficients, for both the DD and DDD estimation, are now larger than in Table 1, with the MCPS and BMCPS estimates being considerably larger now. This difference could be attributable to differences in the employment trends prior to TennCare repeal across the treatment and comparison groups for the March CPS data. This raises the issue of model misspecification in the different datasets, and we now turn to this issue.

5.3 Specification/Placebo Tests for the GGN model Across Datasets

Carrying out specification/placebo tests across datasets is a natural way to distinguish between the estimates. To carry out our specification tests, for each dataset, we run the GGN model for the years 2000-2005 while allowing for a placebo treatment that is coded 1 for 2003-2005, and 0 otherwise; as is well known, if the model is properly specified for the dataset, the placebo treatment variable should not have a significant estimated coefficient. If the placebo coefficient is significantly different from 0, its size is interesting since it gives a measure of how badly the assumption underlying the DD and DDD estimates are violated. We also repeat the analysis letting the placebo variable start to take on the value one in 2002-2005, and then the value one in only 2004-2005, respectively.

Row 1 of Table 8 contains the DD estimates of the placebo effects for the case where the placebo variable equals one for 2003 and later, for the full datasets. These estimates imply that the model is misspecified for all datasets except the ACS data, and the biases for the MCPS, BMCPS and AllCPS are considerable at -0.030, -0.025, and -0.015 respectively. Row 2 of Table 8 contains the DDD estimates of the placebo effects for the full datasets. Now only the AllCPS data produces a significant placebo effect. The corresponding figures showing the trends for the treatment and comparison group relevant for DD estimation across the datasets are in Figures 3A-3C, while the corresponding figures relevant for DDD estimation across the datasets are in Figures 4A-4C. We view these as bearing out the more formal results above, with the possible exception of the DDD figures for the March CPS, which would seem to indicate different high order trends..

We repeat the analysis for the case where the placebo variable takes on the value of one in 2002 and later in Table 9; the corresponding estimates for the case where the placebo variable takes on the value of one in 2004 and later are in Table 10. The 2002 placebo effects are significant only in the MCPS and the AllCPS data when we use DD estimation, and are insignificant for all the datasets when we use DDD estimation. When we let the placebo treatment take place in 2004 and later, they are significant with DD estimation for all datasets. However, when we use DDD estimation, the placebo effects are significant for the AllCPS data. If we consider only the estimates that never have a significant placebo effect, we are left with the DDD estimates from the MCPS (or BMCPS) data and the ACS data. Unfortunately, the DDD estimates for the MCPS and the BMCPS at 5.5 (2.0) PP and 7.0 (2.5) PP, respectively, are very different than the DDD estimate for the ACS at -0.2 (0.6), since the 95% confidence intervals for the DDD estimates from the MCPS, BMCPS and ACS are]1.5, 9.5[PP, [2.0, 12.0[PP and [-1.4, 1.0] PP. respectively. Thus, at least for the full datasets, the placebo tests do not reduce our uncertainty about the overall TennCare treatment effects.

Moreover, we consider the estimates of the 2003 placebo effect for when we let the treatment effect vary by hours worked, education, and age, respectively, in Tables 11-13. For hours, we focus on the DD and DDD placebo effects in Tables 11A and 11B (11A has the DD, 11B has the DDD estimates) for working more than 20 hours per week and working more than 35 hours per week, where we found large treatment effects. For the working more than 20 hours results, the placebo tests mimic those for the full data in Table 8: for the MCPS and BMCPS, the placebo effects are significant for DD estimation but not DDD estimation; for the AllCPS, the placebo effect is significant for both DD and DDD estimation, and the ACS placebo effect is not significant for either DD or DDD estimation. Returning to Tables 2A and 2B, wWe are left choosing between estimates of 4.6 (2.2) PP, 5.2 (2.6) PP, 0.4(0.7) PP from the DDD estimation and -0.3 (0.7) PP. from DD estimation. When we consider the results for working more than 35 hours, the major difference is that the MCPS and BMCPS placebo effects are significant for both DD and DDD estimation, and we are left with only the DD ACS estimate of -1.3 (0.4) PP and the DDD ACS estimate of -0.8 (0.8) PP. Hence these tests lead us away from the GGN estimates.

Table 12 contains the estimated placebo effects by education group. Here we focus on the results for those with a high school degree or more, as that group that had the larger treatment effects in Table 3. Now the placebo tests mirror those for the full data in Table 8: for the MCPS and BMCPS, the placebo effects are significant for DD estimation but not DDD estimation; for the AllCPS, the placebo effect is significant for both DD and DDD estimation, and the ACS placebo effect is not significant for either DD or DDD estimation. Thus, returning to Table 3, we are left with estimates of 3.8 (2.1) PP, 6.9 (2.5) PP, and -0.01(0.06) PP from the DDD estimation and -0.08 (0.04) PP and -0.07 (0.09) from the DD estimation.

Finally, Table 14 contains the estimated placebo effects by age group. Again, we focus on older workers, the group that had the larger treatment effects in Table 4. Now the placebo effects are significant for the MCPS, BMCPS, and AllCPS when we use DD estimation, but none of the datasets produces a significant placebo effect when we use DDD estimation. As a result, returning to Table 4, we are left with estimates of 6.8 (2.7), 5.8 (3.3) 0.9 (1.0) PP and 0.7 (0.8) PP from the DDD estimation and -0.014 (0.005) PP from the DD estimation.

Conclusion

In this paper we first show that the GGN approach does not allow for point identification of the labor supply response to the TennCare contraction. However, we also show that, given reasonable assumptions, the GGN approach, in fact estimates a lower bound for the treatment effect.

As a result, GGN's large estimated effects become even more striking when compared to the previous estimates of the effect of public health insurance expansions and contractions. We reestimate the model across different datasets and find that the results are not robust to moving to a dataset based on all of the CPS months, or one based on the American Community Survey. However, we also find important differences between the AllCPS results and the ACS results. We find that the differences between the estimates based on the March dataset and annual datasets are not the result of the March CPS being substantially affected by seasonality in manufacturing and construction. We also find that the differences between the AllCPS and ACS are not due to the ACS sample containing more urban households than the AllCPS, or to an improper merging of the 2001-2004 ACS data with the 2000 and 2005-2007 data.

We then turn to specification tests to distinguish between the different estimates. While these tests eliminate many specifications, they do not reduce the range of the estimates since the very large DDD March CPS estimates, and the small ACS DDD estimates are not rejected. Thus, readers may differ in their estimate of the TennCare contraction treatment effect.

We argue that the ACS data dominate the March CPS data by virtue of being much larger and not subject to seasonality. Therefore, we conclude that the employment effect of TennCare contraction are close to zero, and we learn nothing about the labor supply effect. Alternatively, one could put the results from the ACS data and the March CPS data on equal footing, which would imply that the TennCare contraction, despite its large size, is essentially uninformative with regard to its employment and labor supply effect. We find it difficult to come up with convincing arguments that imply that the March CPS data dominate the ACS data.

References

Buchmueller, Thomas, John C. Ham and Lara Shore-Sheppard. (2016). "The Medicaid Program." In The Economics of Means-Tested Transfer Programs II, edited by Robert Moffitt, NBER volume, Chicago: University of Chicago Press.

Chang, Cyril and Stephanie Steinberg, ''TennCare Timeline: Major Events and Milestones from 1992 to 2016,'' Methodist LeBonheur Center for Healthcare Economics, University of Memphis, 2009.

Finkelstein, Amy, Katherine Baicker, Jae Song, and Taubman, Sarah (2014). "The Impact of Medicaid on Labor Market Activity and Program Participation: Evidence from the Oregon Health Insurance Experiment." *American Economic Review*, Volume 104, Issue 5 (May): 322-28.

Garthwaite, Craig, Tal Gross, and Matthew Notowidigdo, (2014). "Public Health Insurance, Labor Supply, and Employment Lock." *The Quarterly Journal of Economics*, Volume 129, Issue 2: 653–696.

Ham, John and Lara Shore-Sheppard (2005). "Did Expanding Medicaid Affect Welfare Participation?" *Industrial and Labor Relations Review*, Volume 58, Issue 3: 3452-470.

Hood, W. C. and T. C. Koopmans (editors) (1953) Studies in Econometric Method. New York: John Wiley and Sons.

Meyer, Bruce and Daniel Rosembaum (2001). "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." *The Quarterly Journal of Economics*, Volume 116, Issue 3: 1063-1114.

Moffitt, R. and Wolfe, B.L., 1993. Medicaid, welfare dependency, and work: Is there a causal link?. *Health care financing review*, *15*(1), p.123.

Wooldridge, Judith, Leighton Ku, Teresa Coughlin, Lisa Dubay, Marilyn Ellwood, Shruti Rajan, and Sheila Hoag, ''Implementing State Health Care Reform: What Have We Learned from the First Year? The First Annual Report of the Evaluation of Health Reform in Five States,'' *Mathmatica Policy Research* (1996).

Wright, Douglas, ''TennCare: A Closer Look,'' (Tennessee Comptroller of the Treasury Legislative Briefing Paper, 2001).

Yelowitz, Aaron (1995) "The Medicaid Notch, Labor Supply, and Welfare Participation: Evidence from Eligibility Expansions." *The Quarterly Journal of Economics*, Volume 110, Issue 4: 909-939.

Table 1: TennCare Effect on Employment by Database

	GGN	MCPS	BMCPS	AllCPS	ACS
	(1)	(2)	(3)	(4)	(5)
Panel A: Difference-in-Difference					
Point Estimate	0.025**	0.022**	0.020*	0.013***	-0.011***
Standard Error	(0.011)	(0.010)	(0.011)	(0.004)	(0.004)
N	136	136	136	136	136
Panel B: Triple Difference					
Point Estimate	0.046**	0.055***	0.070***	0.014*	-0.002
Standard Error	(0.020)	(0.020)	(0.024)	(0.008)	(0.006)
N	272	272	272	272	272
Unconditional Average	0.705	0.705	0.707	0.705	0.709
Microdata Sample		249,559	167,368	2,057,701	3,036,337

Notes: The years used are every year from 2000 to 2007 for the MCPS, ACPS, and ACS. We use all people between 21 to 64 years old (inclusive) who were not part of the active military, with at most a bachelor's degree for the MCPS, ACPS, and ACS. We use the same set of southern states used within Garthwaite et al. (Alabama, Arkansas, Delaware, the District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, Tennessee, Texas, Virginia, South Carolina, and West Virginia). We compute the share of employment within each state-year combination. Column (1) shows the same set of statistics displayed within Garthwaite et al. (2014). Column (2) shows our version of their empirical specification using the MCPS. Column (3) shows the equivalent statistics using only March from the ACPS. Column (4) shows the equivalent statistics using all months from the ACPS. Column (5) shows the equivalent statistics using the ACS. The standard errors are calculated in the same fashion as Garthwaite et al.

Table 2A: TennCare Effect on Employment (Hours) by Database: Difference in Difference

	GGN	MCPS	BMCPS	AllCPS	ACS
	(1)	(2)	(3)	(4)	(5)
			Panel A: Working		
Point Estimate	0.025**	0.022**	0.020*	0.013***	-0.011***
Standard Error	(0.011)	(0.010)	(0.011)	(0.004)	(0.004)
			Panel B: 0<Working < 20 hours		
Point Estimate	-0.001	-0.001	-0.004	0.000	0.001
Standard Error	(0.004)	(0.004)	(0.005)	(0.001)	(0.001)
			Panel C: Working ≥ 20 hours		
Point Estimate	0.026***	0.024**	0.024**	0.014***	-0.012***
Standard Error	(0.010)	(0.011)	(0.012)	(0.004)	(0.004)
			Panel D: Working ≥ 20 hours, < 35 hours		
Point Estimate	0.001	0.000	0.007	0.001	0.001
Standard Error	(0.007)	(0.007)	(0.008)	(0.002)	(0.002)
			Panel E: Working ≥ 35 hours		
Point Estimate	0.025**	0.023**	0.018	0.014***	-0.013***
Standard Error	(0.011)	(0.011)	(0.012)	(0.004)	(0.004)

Notes: We use the same years and states as Table 1. For the MCPS, ACPS, and ACS, within each state-year combination, we compute the share of employment, those who work less than 20 hours a week, more than 20 hours a week, more than 20 hours a week and below 35 hours week, and more than 35 hours. Each row represents a different difference in difference regression where the corresponding outcome was used as the dependent variable. The standard errors are calculated in the same fashion as Garthwaite et al.

Table 2B: TennCare Effect on Employment (Hours) by Database: Triple Difference

	GGN	MCPS	BMCPS	AllCPS	ACS
	(1)	(2)	(3)	(4)	(5)
		Panel A: Working			
Point Estimate	0.046**	0.055***	0.070***	0.014*	-0.002
Standard Error	(0.020)	(0.020)	(0.024)	(0.008)	(0.006)
		Panel B: 0<Working < 20 hours			
Point Estimate	0.002	-0.002	0.003	-0.001	0.001
Standard Error	(0.009)	(0.009)	(0.010)	(0.003)	(0.002)
		Panel C: Working ≥ 20 hours			
Point Estimate	0.044**	0.057***	0.067***	0.015*	-0.003
Standard Error	(0.020)	(0.021)	(0.025)	(0.008)	(0.006)
		Panel D: Working ≥ 20 hours, < 35 hours			
Point Estimate	0.018	0.008	0.022	0.014***	-0.008
Standard Error	(0.013)	(0.014)	(0.016)	(0.005)	(0.004)
		Panel E: Working ≥ 35 hours			
Point Estimate	0.026	0.046**	0.052**	0.004	0.004
Standard Error	(0.021)	(0.022)	(0.026)	(0.009)	(0.007)

Notes: We use the same years and states as Table 1. For the MCPS, ACPS, and ACS, within each state-year combination, we compute the share of employment, those who work less than 20 hours a week, more than 20 hours a week, more than 20 hours a week and below 35 hours week, and more than 35 hours. Each row represents a different triple difference regression where the corresponding outcome was used as the dependent variable. The standard errors are calculated in the same fashion as Garthwaite et al.

Table 3: TennCare Effect on Employment by Education Attainment

	GGN	MCPS	BMCPS	AllCPS	ACS
	(1)	(2)	(3)	(4)	(5)
Difference-in-Difference					
Table 1 Estimates					
Point Estimate	0.025**	0.022**	0.020*	0.013***	-0.011***
Std. Error	(0.011)	(0.010)	(0.011)	(0.004)	(0.004)
Panel A: < High School					
Point Estimate	--	0.009	0.016	0.029***	-0.022**
Std. Error	--	(0.028)	(0.033)	(0.011)	(0.009)
Panel B: High School or more					
Point Estimate	--	0.025**	0.022*	0.012***	-0.008**
Std. Error	--	(0.011)	(0.012)	(0.004)	(0.004)
N	136	136	136	136	136
Triple Difference					
Table 1 Estimates					
Point Estimate	0.046**	0.055***	0.070***	0.014*	-0.002
Std. Error	(0.020)	(0.020)	(0.024)	(0.008)	(0.006)
Panel C: < High School					
Point Estimate	0.125**	0.193***	0.092	0.044**	0.002
Std. Error	(0.054)	(0.059)	(0.072)	(0.022)	(0.019)
Panel D: High School or more					
Point Estimate	0.034	0.038*	0.069***	0.011	-0.001
Std. Error	(0.023)	(0.021)	(0.025)	(0.008)	(0.006)
N	272	272	272	272	272
Microdata Sample: < HS	--	40,843	25,635	313,351	415,638
Microdata Sample: HS or More	--	208,716	141,733	1,744,350	2,620,699

Notes: We use the same years and states as Table 1. We first compute the employment share within each state-year combination for those who had less than a high school education and those who had either a high school degree or higher separately. We then calculate the difference in difference/triple difference estimates for each cohort separately. The standard errors are calculated in the same fashion as Garthwaite et al.

Table 4: TennCare Effect on Employment by Age

	GGN	MCPS	BMCPS	AllCPS	ACS
	(1)	(2)	(3)	(4)	(5)
			Difference-in-Difference		
			Table 1 Estimates		
Point Estimate	0.025**	0.022**	0.020*	0.013***	-0.011***
Std. Error	(0.011)	(0.010)	(0.011)	(0.004)	(0.004)
			Panel A: Age 21-39, Inclusive		
Point Estimate	--	-0.008	-0.019	-0.001	-0.006
Std. Error	--	(0.015)	(0.018)	(0.006)	(0.005)
			Panel B: Age 40-64, Inclusive		
Point Estimate	--	0.046***	0.050***	0.024***	-0.014***
Std. Error	--	(0.013)	(0.015)	(0.006)	(0.005)
N		136	136	136	136
			Triple Difference		
			Table 1 Estimates		
Point Estimate	0.046**	0.055***	0.070***	0.014*	-0.002
Std. Error	(0.020)	(0.020)	(0.024)	(0.008)	(0.006)
			Panel C: Age 21-39, Inclusive		
Point Estimate	0.01	0.02	0.054	0.013	-0.007
Std. Error	(0.031)	(0.029)	(0.036)	(0.011)	(0.009)
			Panel D: Age 40-64, Inclusive		
Point Estimate	0.060**	0.068**	0.058*	0.009	0.007
Std. Error	(0.028)	(0.027)	(0.033)	(0.010)	(0.008)
N		272	272	272	272
Microdata Sample: Age 21-39		40,843	25,635	313,351	415,638
Microdata Sample: Age 40-64		208,716	141,733	1,744,350	2,620,699

Notes: We use the same years and states as Table 1. We first compute the employment share within each state-year combination for those in the following age brackets separately: 21-39, 40-64. We then calculate the difference in difference/triple difference estimates for each cohort separately. The standard errors are calculated in the same fashion as Garthwaite et al.

Table5: TennCare Effect on Employment by Database - Omit Construction, Manufacturing

	GGN	MCPS	BMCPS	AllCPS	ACS
	(1)	(2)	(3)	(4)	(5)
Panel A: Difference-in-Difference					
Point Estimate	--	0.026***	0.024*	0.013***	-0.005
Standard Error	--	(0.010)	(0.014)	(0.005)	(0.005)
N		136	136	136	136
Panel B: Triple Difference					
Point Estimate	--	0.054**	0.074***	0.019**	-0.007
Standard Error	--	(0.024)	(0.028)	(0.009)	(0.007)
Unconditional Average	--	0.663	0.664	0.663	0.685
N		272	272	272	272
Microdata Sample		207,036	139,105	1,712,681	2,468,114

Notes: We use the same years and states as Table 1 and we impose an additional restriction: We further restrict the sample of people to those who were not part of the construction or manufacturing sectors. All other calculations are performed analogously to those in Table 1.

Table6: TennCare Effect on Employment by Database - Only Metropolitan Areas

	GGN	MCPS	BMCPS	AllCPS	ACS
	(1)	(2)	(3)	(4)	(5)
Panel A: Difference-in-Difference					
Point Estimate	--	0.021*	0.024*	0.002	-0.010**
Standard Error	--	(0.012)	(0.014)	(0.004)	(0.004)
N		136	136	136	136
Panel B: Triple Difference					
Point Estimate	--	0.044*	0.061**	0.008	-0.003
Standard Error	--	(0.023)	(0.029)	(0.009)	(0.007)
Unconditional Average	--	0.715	0.718	0.716	0.714
N		272	272	272	272
Microdata Sample		193,290	128,292	1,584,874	2,655,685

Notes: We use the same years and states as Table 1 and we impose an additional restriction: We further restrict the sample of people to those who were part of a metropolitan area. All other calculations are performed analogously to those in Table 1.

Table7: TennCare Effect on Employment by Database - 2005 to 2007

	GGN	MCPS	BMCPS	AllCPS	ACS
	(1)	(2)	(3)	(4)	(5)
Panel A: Difference-in-Difference					
Point Estimate	--	0.040***	0.042**	0.017***	0.002
Standard Error	--	(0.015)	(0.017)	(0.005)	(0.004)
N		51	51	51	51
Panel B: Triple Difference					
Point Estimate	--	0.071**	0.087***	0.022**	0.006
Standard Error	--	(0.030)	(0.034)	(0.010)	(0.009)
Unconditional Average	--	0.701	0.703	0.705	0.709
N		102	102	102	102
Microdata Sample		99,381	65,594	802,771	1,680,411

Notes: We use the same states as Table 1. We use the years 2005 to 2007 instead of 2000 to 2007. All other restrictions and calculations are done in the same manner as in Table 1.

Table 8: Placebo Tests Assuming a 2003 Treatment Year

	GGN	MCPS	BMCPS	ACPS	ACS
	(1)	(2)	(3)	(4)	(5)
		Difference-in-Difference			
Point Estimate	--	-0.030***	-0.025**	-0.015***	-0.001
Standard Error	--	(0.011)	(0.013)	(0.004)	(0.004)
N	102	102	102	102	102
		Triple Difference			
Point Estimate	--	-0.012	-0.036	-0.022***	-0.006
Standard Error	--	(0.021)	(0.024)	(0.008)	(0.008)
N	204	204	204	204	204

Notes: We use the same states as Table 1. We use the years 2000 to 2005 instead of 2000 to 2007. All restrictions and calculations are done in the same manner as in Table 1. We treat 2003, 2004, and 2005 as the treatment years as the placebo test.

Table 9: Placebo Tests Assuming a 2002 Treatment Year

	GGN	MCPS	BMCPS	ACPS	ACS
	(1)	(2)	(3)	(4)	(5)
		Difference-in-Difference			
Point Estimate	--	-0.023**	-0.018	-0.007*	-0.001
Standard Error	--	(0.012)	(0.013)	(0.004)	(0.004)
N	102	102	102	102	102
		Triple Difference			
Point Estimate	--	0.001	-0.014	-0.010	-0.003
Standard Error	--	(0.022)	(0.026)	(0.008)	(0.007)
N	204	204	204	204	204

Notes: We use the same states as Table 1. We use the years 2000 to 2005 instead of 2000 to 2007. All restrictions and calculations are done in the same manner as in Table 1. We treat 2002, 2003, 2004, and 2005 as the treatment years as the placebo test.

Table 10: Placebo Tests Assuming a 2004 Treatment Year

	GGN	MCPS	BMCPS	ACPS	ACS
	(1)	(2)	(3)	(4)	(5)
			Difference-in-Difference		
Point Estimate	--	-0.023*	-0.027**	-0.012***	-0.007*
Standard Error	--	(0.012)	(0.013)	(0.004)	(0.004)
N	102	102	102	102	102
			Triple Difference		
Point Estimate	--	-0.010	-0.016	-0.017**	-0.007
Standard Error	--	(0.023)	(0.025)	(0.008)	(0.008)
N	204	204	204	204	204

Notes: We use the same states as Table 1. We use the years 2000 to 2005 instead of 2000 to 2007. All restrictions and calculations are done in the same manner as in Table 1. We treat 2004 and 2005 as the treatment years as the placebo test.

Table 11A: Placebo Tests Assuming a 2003 Treatment Year – Treatment Effect Varies with Hours Worked; Difference in Difference

	GGN	MCPS	BMCPS	AllCPS	ACS
	(1)	(2)	(3)	(4)	(5)
			Panel A: Working		
Point Estimate	--	-0.030***	-0.025**	-0.015***	-0.001
Standard Error	--	(0.011)	(0.012)	(0.004)	(0.004)
			Panel B: 0<Working < 20 hours		
Point Estimate	--	-0.001	0.001	0.000	0.001
Standard Error	--	(0.005)	(0.005)	(0.001)	(0.001)
			Panel C: Working ≥ 20 hours		
Point Estimate	--	-0.029***	-0.026**	-0.014***	-0.002
Standard Error	--	(0.011)	(0.013)	(0.004)	(0.004)
			Panel D: Working ≥ 20 hours, < 35 hours		
Point Estimate	--	-0.013*	-0.012	-0.006**	-0.001
Standard Error	--	(0.007)	(0.008)	(0.003)	(0.003)
			Panel E: Working ≥ 35 hours		
Point Estimate	--	-0.018	-0.020	-0.008**	-0.002
Standard Error	--	(0.012)	(0.013)	(0.004)	(0.004)

Notes: We use the same states as Table 1. The outcomes are the same as Table 2A. We use 2000 to 2005 instead of 2000 to 2007. We use 2003, 2004, and 2005 as the treatment years.

Table 11B: Placebo Tests Assuming a 2003 Treatment Year – Treatment Effect Varies with Hours Worked; Triple Difference

	GGN	MCPS	BMCPS	AllCPS	ACS
	(1)	(2)	(3)	(4)	(5)
		Panel A: Working			
Point Estimate	--	-0.012	-0.036	-0.022***	-0.006
Standard Error	--	(0.021)	(0.023)	(0.007)	(0.007)
		Panel B: 0<Working < 20 hours			
Point Estimate	--	0.002	0.002	-0.008***	-0.001
Standard Error	--	(0.009)	(0.010)	(0.003)	(0.002)
		Panel C: Working ≥ 20 hours			
Point Estimate	--	-0.013	-0.038	-0.014*	-0.005
Standard Error	--	(0.021)	(0.025)	(0.008)	(0.008)
		Panel D: Working ≥ 20 hours, < 35 hours			
Point Estimate	--	0.020	0.023	0.005	0.002
Standard Error	--	(0.014)	(0.017)	(0.005)	(0.005)
		Panel E: Working ≥ 35 hours			
Point Estimate	--	-0.039*	-0.057**	-0.020**	-0.008
Standard Error	--	(0.022)	(0.026)	(0.008)	(0.008)

Notes: We use the same states as Table 1. The outcomes are the same as Table 2B. We use 2000 to 2005 instead of 2000 to 2007. We use 2003, 2004, and 2005 as the treatment years in the difference in difference analysis.

Table 12: Placebo Tests Assuming a 2003 Treatment Year – Treatment Effect Varies with Education

	GGN	MCPS	BMCPS	AllCPS	ACS
	(1)	(2)	(3)	(4)	(5)
			Difference-in-Difference		
			Panel A: < High School		
Point Estimate	--	-0.029	-0.070**	-0.025**	0.002
Std. Error	--	(0.029)	(0.033)	(0.010)	(0.011)
			Panel B: High School or more		
Point Estimate	--	-0.033***	-0.021*	-0.014***	-0.003
Std. Error	--	(0.012)	(0.013)	(0.004)	(0.004)
			Triple Difference		
			Panel C: < High School		
Point Estimate	--	-0.015	-0.059	0.025	-0.002
Std. Error	--	(0.054)	(0.062)	(0.020)	(0.025)
			Panel D: High School or more		
Point Estimate	--	-0.013	-0.026	-0.027***	-0.008
Std. Error	--	(0.022)	(0.025)	(0.008)	(0.007)

Notes: We use the same states as Table 1. The outcomes and comparisons are the same as Table 3. We use 2000 to 2005 instead of 2000 to 2007. We use 2003, 2004, and 2005 as the treatment years in the difference in difference analysis.

Table 13: Placebo Tests Assuming a 2003 Treatment Year – Treatment Effect Varies with Age

	GGN	MCPS	BMCPS	AllCPS	ACS
	(1)	(2)	(3)	(4)	(5)
Difference-in-Difference					
Panel A: Age 21-39, Inclusive					
Point Estimate	--	-0.022	-0.006	0.001	-0.002
Std. Error	--	(0.015)	(0.017)	(0.005)	(0.005)
Panel B: Age 40-64, Inclusive					
Point Estimate	--	-0.033**	-0.038**	-0.028***	-0.001
Std. Error	--	(0.015)	(0.017)	(0.006)	(0.005)
Triple Difference					
Panel C: Age 21-39, Inclusive					
Point Estimate	--	-0.009	-0.041	-0.018*	-0.001
Std. Error	--	(0.031)	(0.034)	(0.010)	(0.011)
Panel D: Age 40-64, Inclusive					
Point Estimate	--	0.000	-0.008	-0.009	-0.012
Std. Error	--	(0.028)	(0.034)	(0.011)	(0.010)

Notes: We use the same states as Table 1. The outcomes and comparisons are the same as Table 4. We use 2000 to 2005 instead of 2000 to 2007. We use 2003, 2004, and 2005 as the treatment years in the difference in difference analysis.

Figure 1

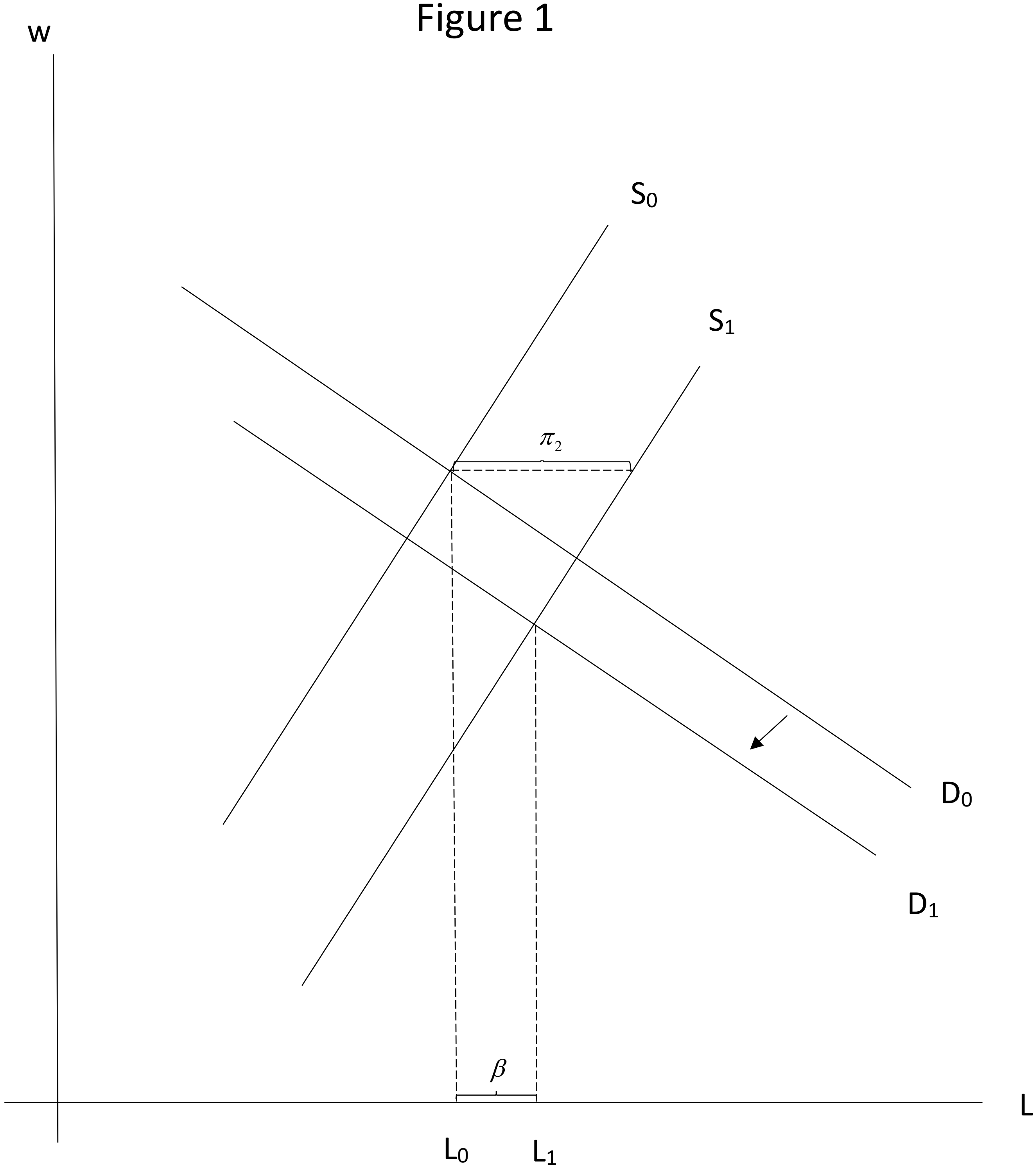


Figure 2

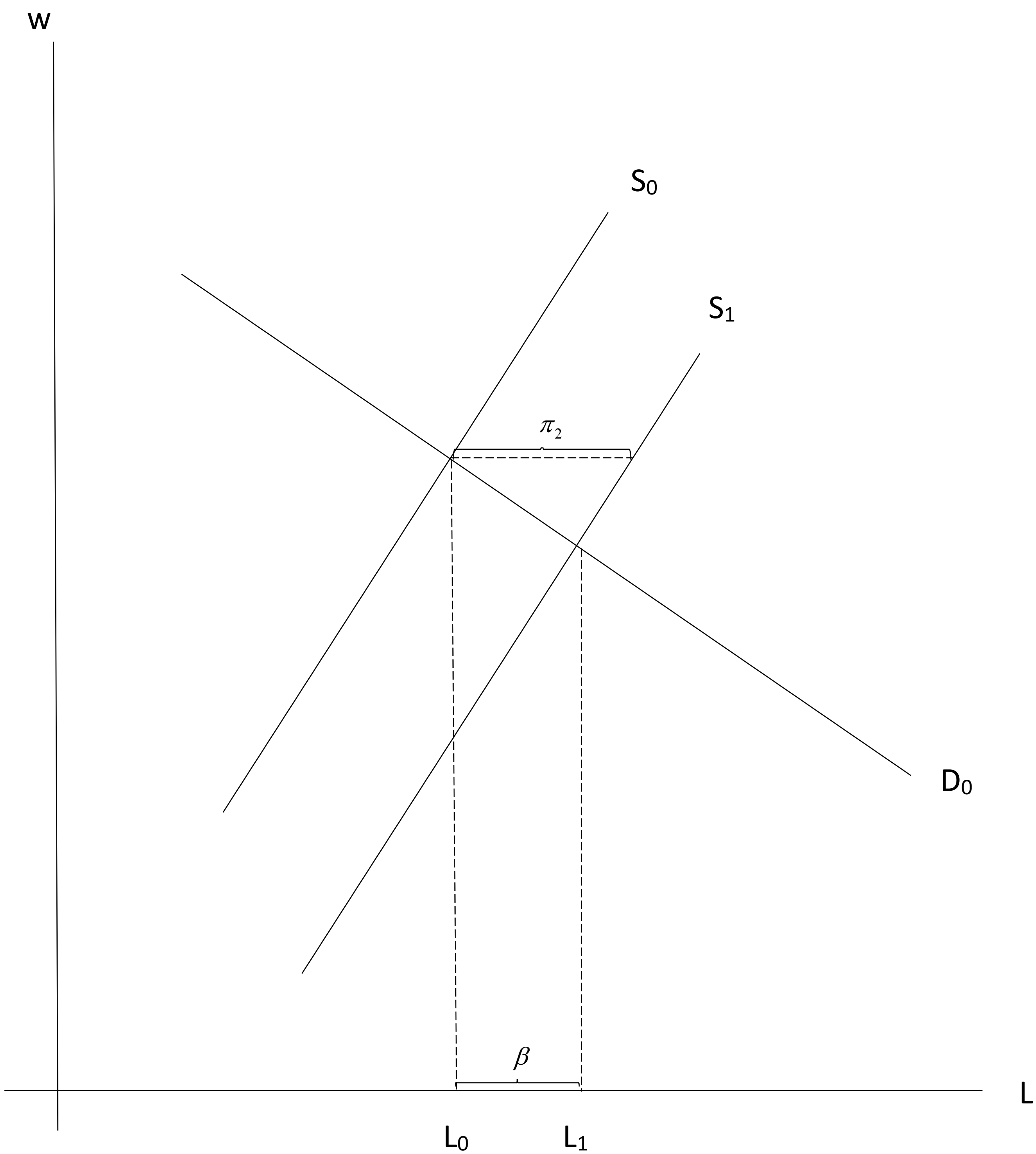
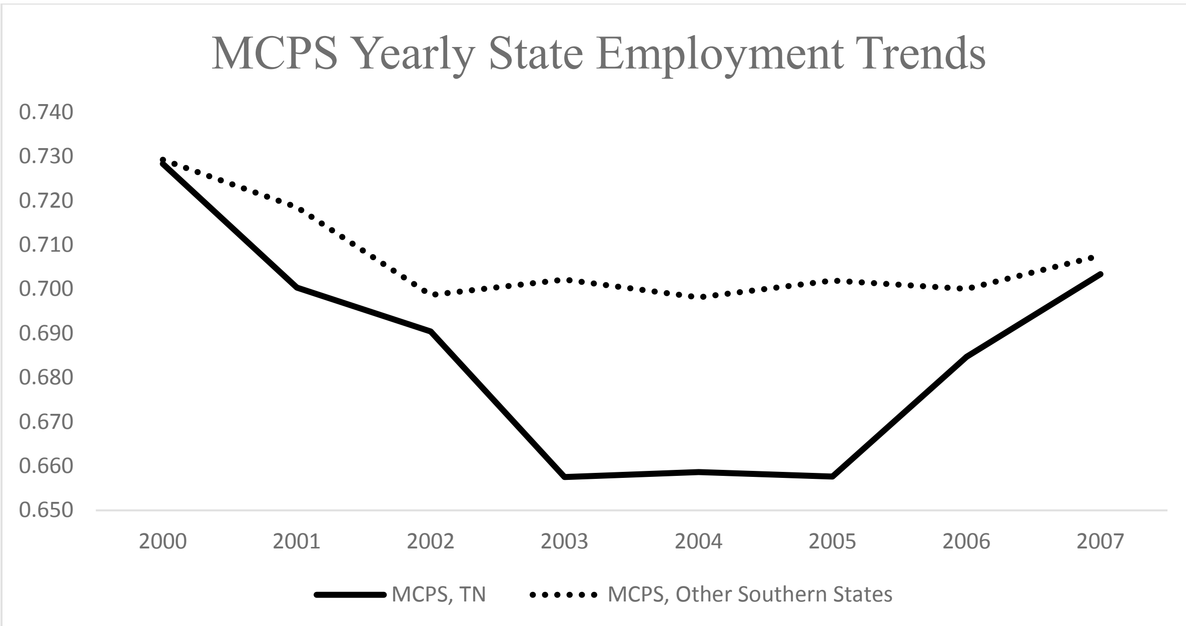
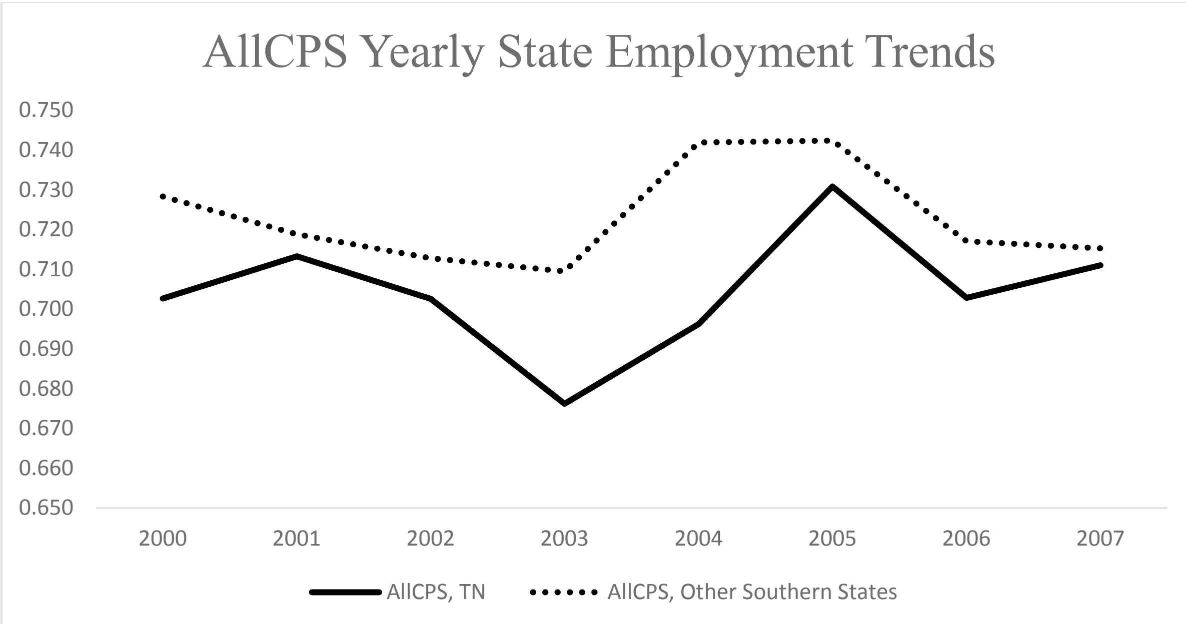


Figure 3A: MCPS Yearly State Employment Trends



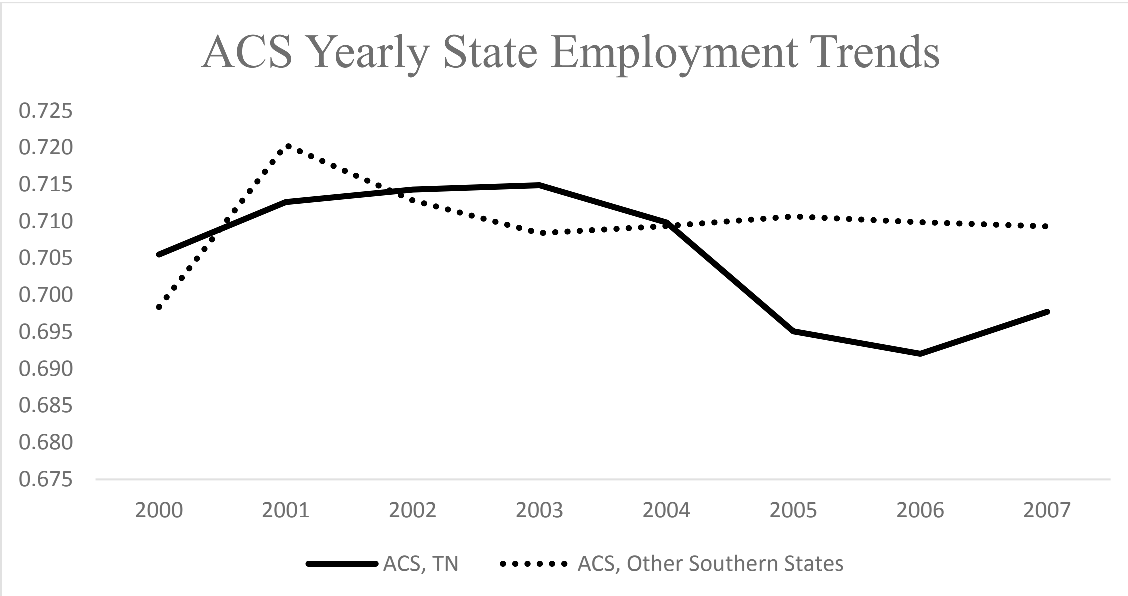
Notes: The years used are every year from 2000 to 2007. The sample of workers is the same sample as the one used in Table 1. The Y-axis is the employment share; the X-axis is the year.

Figure 3B: AllCPS Yearly State Employment Trends



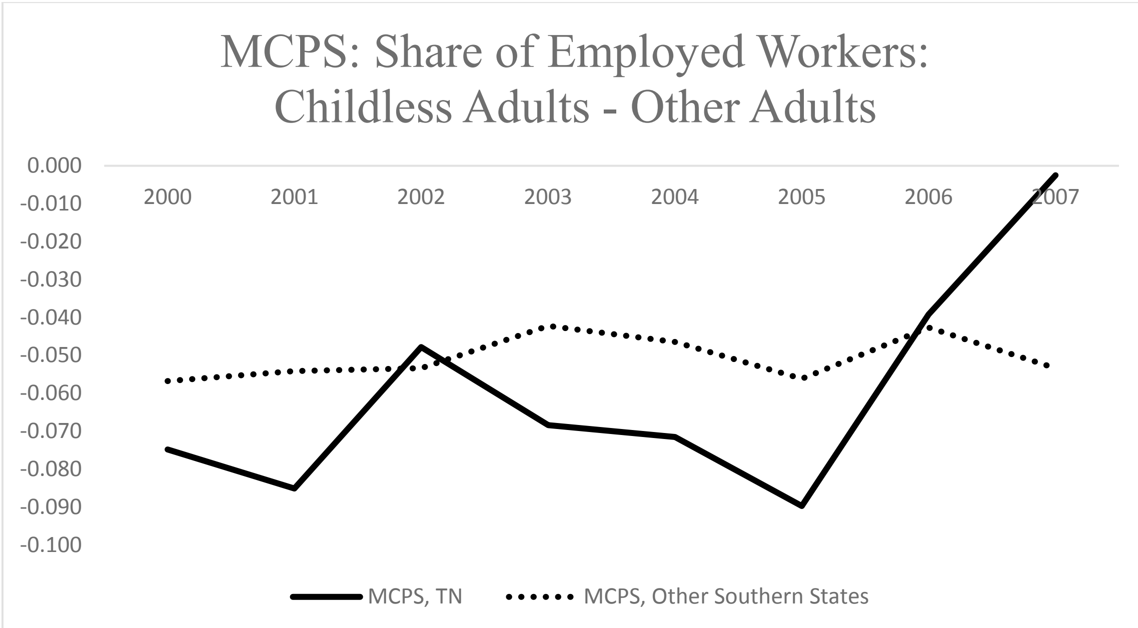
Notes: The years used are every year from 2000 to 2007. The sample of workers is the same sample as the one used in Table 1. The Y-axis is the employment share; the X-axis is the year.

Figure 3C: ACS Yearly State Employment Trends



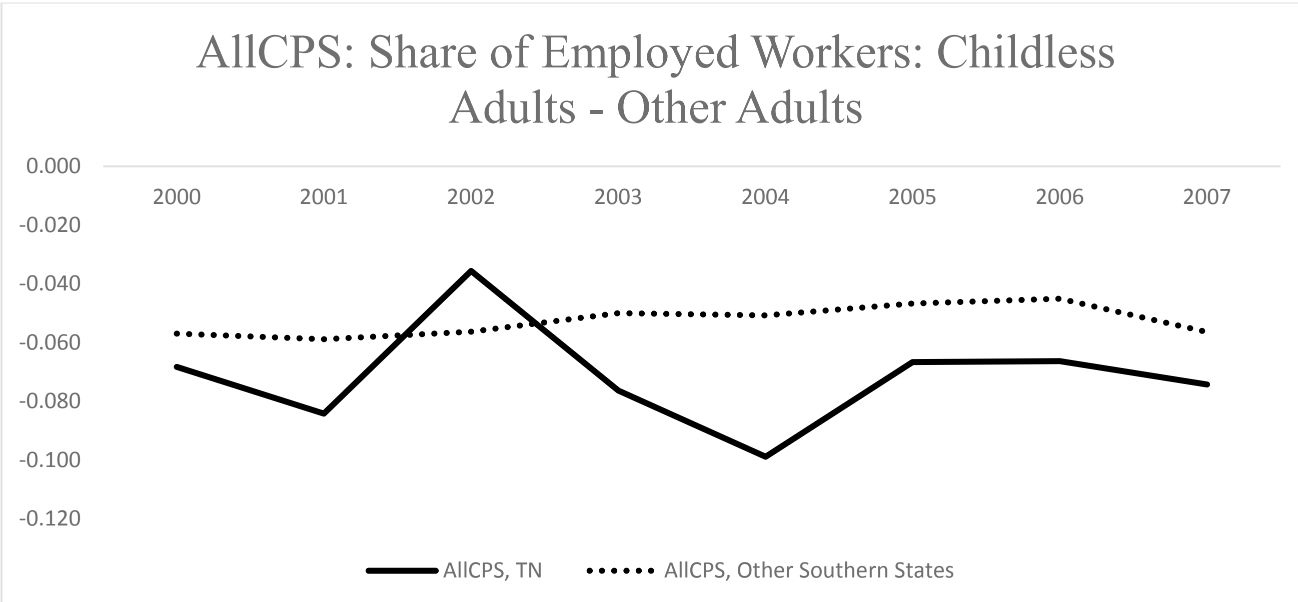
Notes: The years used are every year from 2000 to 2007. The sample of workers is the same sample as the one used in Table 1. The Y-axis is the employment share; the X-axis is the year.

Figure 4A: MCPS: Share of Employed Workers; Childless Adults – Other Adults



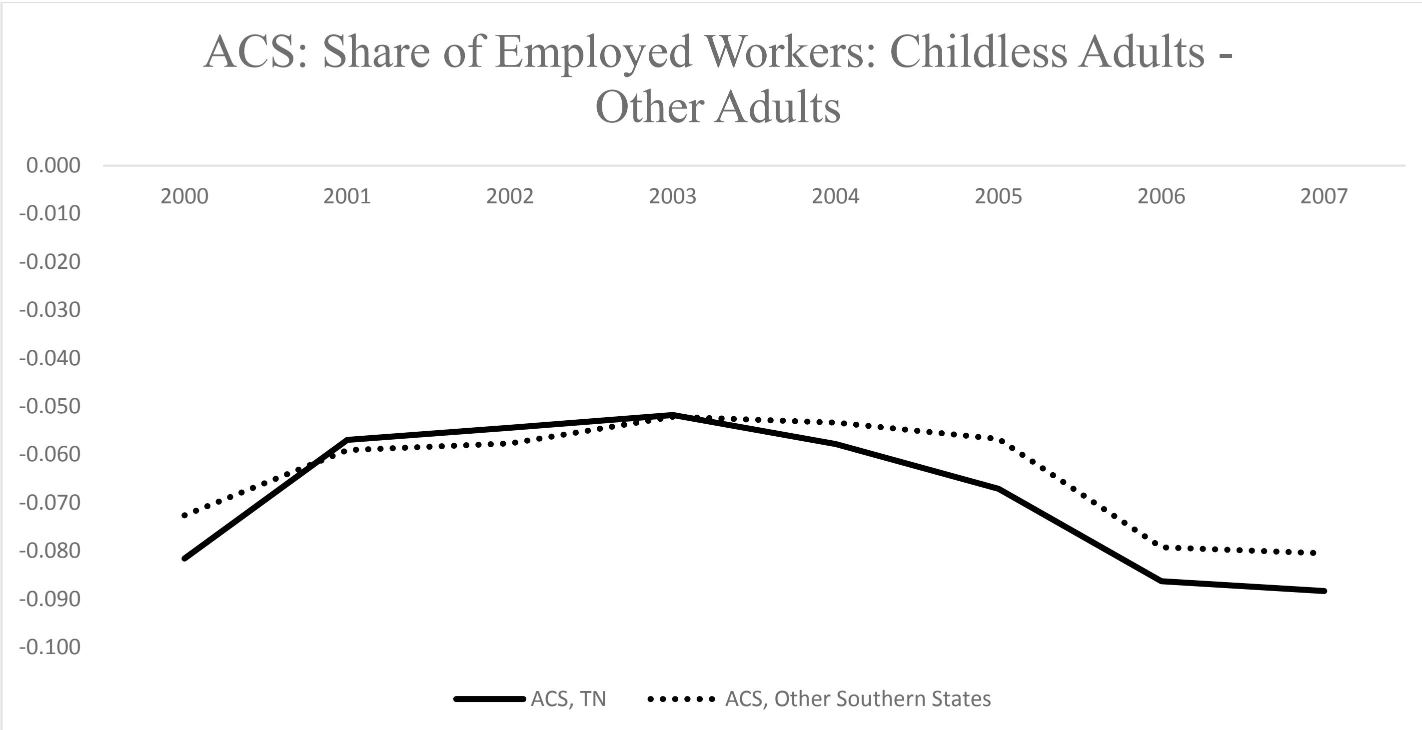
Notes: The years used are every year from 2000 to 2007. The sample of workers is the same sample as the one used in Table 1. The Y-axis is the difference in the employment share between those without children under 18 years old vs. those with children under 18 years old. The X-axis is the year.

Figure 4B: AllCPS: Share of Employed Workers; Childless Adults – Other Adults



Notes: The years used are every year from 2000 to 2007. The sample of workers is the same sample as the one used in Table 1. The Y-axis is the difference in the employment share between those without children under 18 years old vs. those with children under 18 years old. The X-axis is the year.

Figure 4C: ACS: Share of Employed Workers; Childless Adults – Other Adults



Notes: The years used are every year from 2000 to 2007. The sample of workers is the same sample as the one used in Table 1. The Y-axis is the difference in the employment share between those without children under 18 years old vs. those with children under 18 years old. The X-axis is the year.