Online Appendix for

Do State Earned Income Tax Credits Increase Participation in the Federal EITC?

David Neumark
UCI, NBER, and IZA

Katherine E. Williams
UCI

Potential Biases from Non-Representativeness of the Public-Use SOI Data

We use the SOI public-use files created by the IRS to measure EITC take-up. A potential issue is that these files are not constructed to be representative at the state level, but rather, with the sample weights, to be representative at the federal level. In this appendix, we document this issue, and explore whether our results are sensitive to this non-representativeness.

There are a couple of reasons this non-representativeness may not be a first-order problem for our regression estimates. First, a lack of representativeness need not bias regression estimates, if there is not heterogeneity of parameters across subgroups whose representation differs from the sampling universe. (This is essentially the same question that underlies the importance of weighting in regression estimates; see Solon et al., 2015). Thus, the critical question is how the regression estimates are affected by the sampling issue. Second, sampling errors in the public-use data are more common and larger for small states. But the estimates we report in the paper are weighted by the number of filers in the states, because we used grouped data, suggesting that any sampling errors will play a small role.

We first present some evidence on differences between the data sources, and then show that the results are insensitive to a number of ways to address the discrepancies. Note that we were not able to access the full confidential microdata, so that we cannot generate estimates avoiding the public-use SOI data, or estimate discrepancies by filing status.[1] However, we can study differences in the numbers of filers implied by the two data sources, and assess the sensitivity of the results to either trying to correct for the sampling error in the public-use SOI data, or dropping the states with large errors.

Appendix Figures A1 and A2 give some evidence on the sampling errors. In particular, for our first and last sample years (1997 and 2008) we plot the number of filers, by state, implied by the public-use SOI data, and in the full IRS data. (The plots are similar for all other years.) These figures clearly show some errors. And indeed if we zoom in on the smaller states to see more detail near the lower-left-hand corner, we see that these errors are larger in smaller states, which we show here just for the first year (Appendix Figure A3).

We take two different approaches to assessing whether the regression results are sensitive to these sampling errors. First, we simply drop states with large and persistent errors in the SOI public-use data. In particular, we drop states with 9 or more years of absolute errors in the SOI public-use data that exceed 10% (Delaware, the District of Columbia, Hawaii, Vermont, West Virginia, and Wyoming). Appendix Figure A4 shows these average absolute errors for each state, and the six states we drop based on this criterion (marked with x's). Appendix Table A1 shows the results in which we drop these states. In this analysis, we focus on the results for single filers with children, for the specification for 1 vs. 2+ kids. (Results for other runs were similar, as were results with different cutoffs for the states excluded, such as 5% and 20%.) The results in Appendix Table A1 are qualitatively similar to those in Table 4.B.

The second approach we take is to adjust the total number of filers implied by the SOI data for the percentage error implied by the difference between the public-use SOI and the full confidential SOI data. Note, again, that we do not have this error by filing status. The results again are very similar. (See Appendix Table A2.)

Overall, between these findings, and the fact the sampling errors are large for the small states that are downweighted, we conclude that our results are not materially influenced by the non-representative sampling in the public-use SOI data.

Figure A1: Total EITC filers: SOI vs IRS, 1997

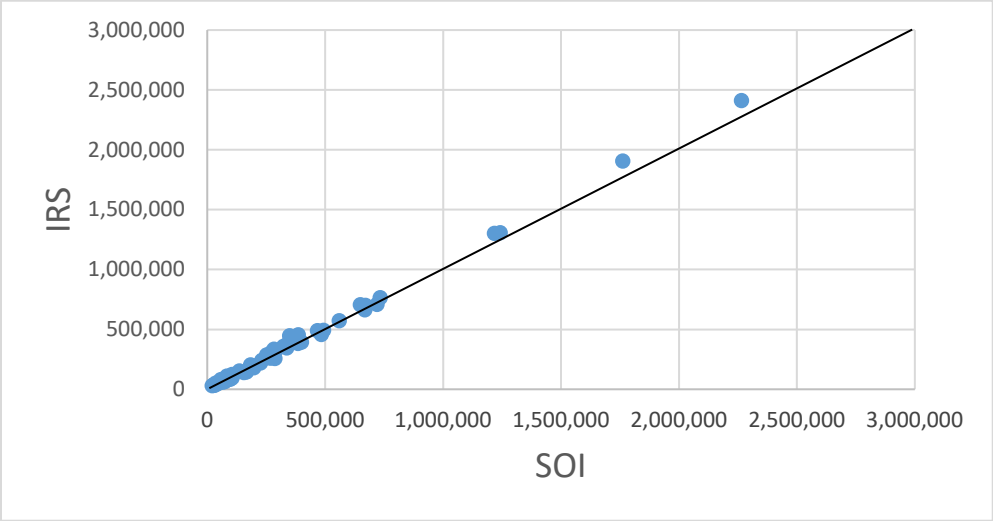


Figure A2: Total EITC filers: SOI vs IRS, 2008

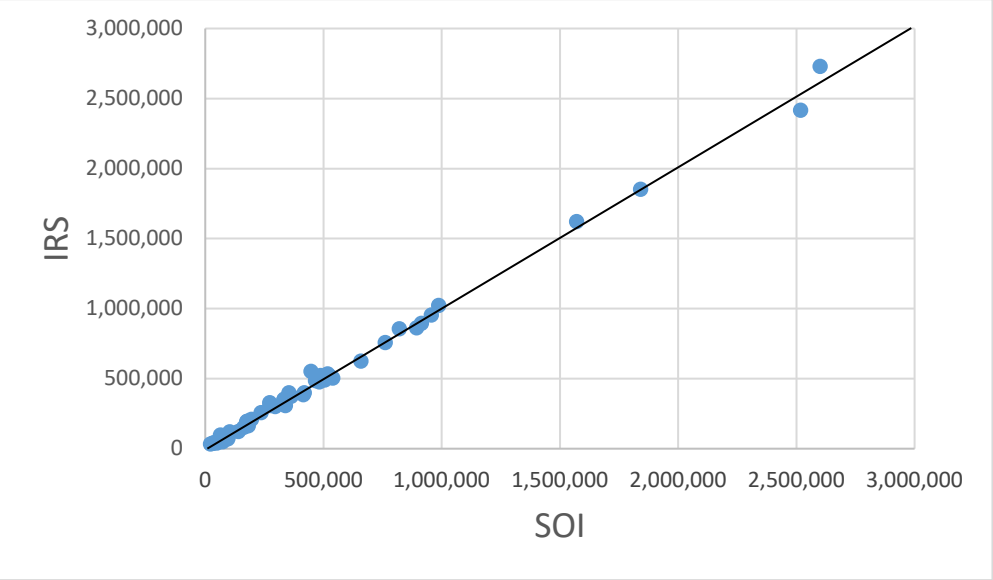


Figure A3: Total EITC filers: SOI vs IRS, 1997 (States with < 1 million filers)

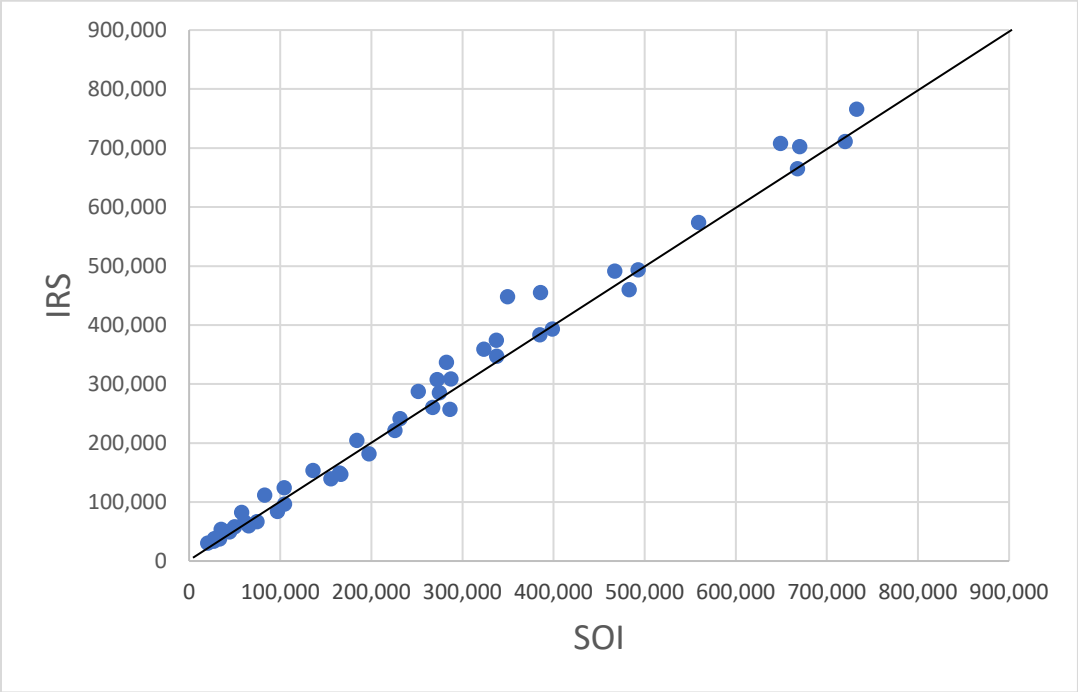
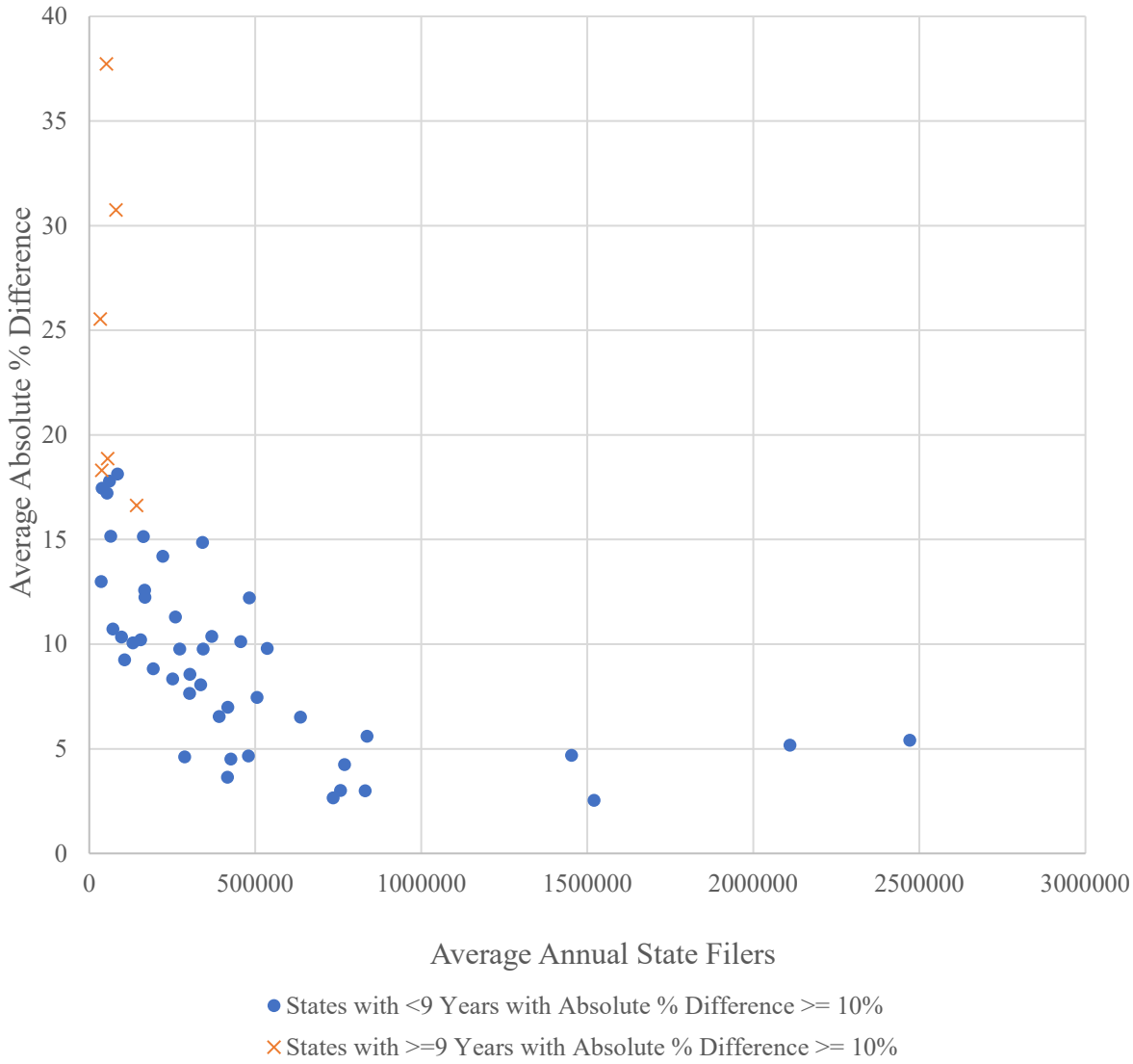


Figure A4: Average Absolute Percent Difference Between Public-Use SOI and Confidential IRS SOI EITC State Filers



Appendix Table A1: Estimated State EITC Effects on EITC Participation Per Potentially Eligible Filer, Dropping States with Persistently Large Errors (DE, DC, HI, VT, WV, and WY): *Single Filers with Children, 2+ Kids vs. 1 Kid, 1997-2008*

VARIABLES	(1)	(2)	(3)	(4)
	Y = EITC Filers/Potentially Eligible Population			
	Single Individuals with Children, Aged 21-64			
State EITC	-0.25	0.08	-0.08	0.26
	(0.25)	(0.17)	(0.30)	(0.43)
2+ Kids	0.04	0.05	0.06	-0.68
	(0.07)	(0.06)	(0.07)	(1.04)
Share low-skilled	1.21***	1.39***	1.28***	2.26***
	(0.40)	(0.39)	(0.40)	(0.78)
State EITC×2+ Kids	0.63		0.33	0.14
	(0.40)		(0.53)	(0.57)
State EITC×Share low-skilled		6.53***	4.72	13.88**
		(2.33)	(3.28)	(6.88)
2+ Kids×Share low-skilled			0.09	-1.73
			(0.24)	(1.04)
State EITC×2+ Kids×Share low-skilled			0.15	-13.45
			(1.67)	(8.93)
State minimum wage	-0.24	-0.25	-0.25	-0.26
	(0.18)	(0.18)	(0.18)	(0.20)
Real maximum welfare benefit/1000, family size = 3	-0.09	-0.08	-0.08	-0.10
	(0.35)	(0.34)	(0.34)	(0.36)
State unemployment rate	-1.12	-1.14	-1.13	-1.02
	(1.36)	(1.32)	(1.34)	(1.32)
Fixed Effects				
State and Year FE	Y	Y	Y	Y
Kids-by-State FE and Kids-by-Year FE	Y	Y	Y	Y
Low-skilled-by-State FE and Low-skilled-by-Year FE	Y	Y	Y	Y
Kids-by-Skill-by-State FE and Kids-by-Skill-by-Year FE				Y
Observations	1,220	1,220	1,220	1,220
R-squared	0.75	0.75	0.75	0.77

Appendix Table A2: Estimated State EITC Effects on EITC Participation Per Potentially Eligible Filer, SOI Data Error Correction: *Single Filers with Children, 2+ Kids vs. 1 Kid, 1997-2008*

	(1)	(2)	(3)	(4)
	Y= [EITC Filers in Public-Use Data/Potentially Eligible Population] - [(EITC Filers Public-Use Data - EITC Filers IRS SOI)/Total Filers IRS SOI])			
VARIABLES	*Single Individuals with Children, Aged 21-64*			
State EITC	-0.30	0.04	-0.01	0.59
	(0.27)	(0.19)	(0.39)	(0.53)
2+ Kids	0.24***	0.25***	0.25***	-2.64**
	(0.07)	(0.07)	(0.08)	(0.99)
Share low-skilled	1.37***	1.46***	1.37***	2.69***
	(0.45)	(0.44)	(0.48)	(0.89)
State EITC×2+ Kids	0.52		0.09	-0.36
	(0.44)		(0.66)	(0.62)
State EITC×Share low-skilled		7.29**	6.76	20.25**
		(2.96)	(4.85)	(8.58)
2+ Kids×Share low-skilled			0.15	-2.15*
			(0.25)	(1.19)
State EITC×2+ Kids×Share low-skilled			-0.11	-19.37
			(1.88)	(11.77)
State minimum wage	-0.25	-0.27	-0.27	-0.29
	(0.20)	(0.20)	(0.20)	(0.21)
Real maximum welfare benefit/1000, family size = 3	-0.07	-0.06	-0.06	-0.10
	(0.39)	(0.37)	(0.37)	(0.39)
State unemployment rate	-0.98	-1.04	-1.04	-0.95
	(1.59)	(1.54)	(1.54)	(1.51)
Fixed Effects				
State and Year FE	Y	Y	Y	Y
Kids-by-State FE and Kids-by-Year FE	Y	Y	Y	Y
Low-skilled-by-State FE and Low-skilled-by-Year FE	Y	Y	Y	Y
Kids-by-Skill-by-State FE and Kids-by-Skill-by-Year FE				Y
Observations	1,080	1,080	1,080	1,080
R-squared	0.75	0.75	0.75	0.77

[1] The IRS's SOI division takes a sample of federal tax returns filed each year. These data are not publicly available, but from this sample, the SOI division produces a smaller publicly available file known as the Public Use File (PUF). We are using the PUF data. These publicly available data contain state identifiers, designed to be nationally representative but not necessarily representative at the state level. Both the Office of Tax Analysis (OTA) and the Urban-Brookings Tax Policy Center (TPC) have worked on creating weights to make the data representative at the state level for their tax models (Fisher and Lin, 2015; Khitatrakun et al., 2016). However, these weights are not publicly available. Additionally, the OTA weights are based on the non-publicly available microdata from which the PUF are derived, and the TPC data do not cover most of our sample years.

Additional References Not in Main Document

Fisher, Robin, and Emily Lin. 2015. *Re-weighting to Produce State-Level Tax Microsimulation Estimates*. Office of Tax Analysis Technical Paper 6, Washington, D.C.

Khitatrakun, Surachai, Gordon Mermin, and Norton Francis. 2016. Incorporating state analysis into the Tax Policy Center's microsimulation model: Documentation and methodology. Urban Institute & Brookings Institution Tax Policy Center Working Paper, Washington, D.C.

Solon, Gary, Steven Haider, and Jeffrey Wooldridge. 2015. What are we weighting for? *Journal of Human Resources*, 50(2): 301-316.