Economics Bulletin

Volume 39, Issue 3

Estimating tax income elasticities using a group-averaged synthetic tax instrument

Allan Seuri
Tampere University

Abstract

A common approach to estimating the elasticity of income to taxation is to construct an instrumental variable using synthetic tax rates. Individual-level income dynamics threaten the validity of this instrument, but this problem can potentially be mitigated by group-averaging the instrument. In this article I show that rather than imposing an arbitrary minimum threshold for group sizes to avoid small-sample bias, researchers should use leave-one-out group averages. Using CPS data I show that this correction increases the estimate for broad income elasticity.

Citation: Allan Seuri, (2019) "Estimating tax income elasticities using a group-averaged synthetic tax instrument", *Economics Bulletin*, Volume 39, Issue 3, pages 2137-2141
Contact: Allan Seuri - allan.seuri@gmail.com.
Submitted: April 22, 2019. **Published:** September 30, 2019.

1 Introduction

The elasticity of taxable income and the elasticity of broad income are key parameters in determining the welfare effects of taxation (Feldstein 1999, Chetty 2008). A popular approach to estimating these elasticities is to regress the change in log income on the change in log of net-of-tax rate:

$$\triangle \ln\ y_{i,t} = \alpha + \beta \triangle \ln\ (1 - \tau_t(y_{i,t})) + \Gamma Q_{i,t-1} + \epsilon_{i,t}, \tag{1}$$

where $\triangle \ln\ y_{i,t} \equiv \ln\ (y_{i,t}/y_{i,t-1})$ and $\triangle \ln\ (1 - \tau_t(y_{i,t}))$ is defined similarly, Q is a vector of controls, and $\tau_t(y)$ gives the marginal tax rate given income y (and some other characteristics such as state and marital status, omitted for notational convenience). This equation was derived by Gruber & Saez (2002) from a static consumption-leisure optimization problem. Their original formulation also included income effects, but as they found that these are empirically quite small, they dropped income this term from most of their specifications. Much of the subsequent literature has made the same choice.

As the net-of-tax rate is a function of income, it is endogenous and must be instrumented. Gruber & Saez (2002), and many others in subsequent literature, use what is called the synthetic tax rate for identification. The instrument for $\ln\ [(1 - \tau_t(y_{i,t}))/(1 - \tau_{t-1}(y_{i,t-1}))]$ is $\ln\ [(1 - \tau_t(y_{i,t-1}))/(1 - \tau_{t-1}(y_{i,t-1}))]$. The counterfactual synthetic tax rate $\tau_t(y_{i,t-1})$ is calculated using a tax simulation model. Using the NBER panel of tax returns for 1979-1990, Gruber & Saez (2002) obtain a preferred estimate of 0.4 for the elasticity of taxable income and an estimate of 0.12 for the elasticity of broad income.

Ideally the generated instrument would only capture exogenous variations in tax policy. The instrument is a function of y_{t-1}, which is also on the left-hand side of the second-stage equation. If income growth is a function of income levels and tax reforms captured by the instrument correlate with income, the exclusion restriction is violated. One likely form of income dynamics and thus a source of bias is mean reversion, whereby transitory income shocks produce negative covariance between income levels and income growth.

As mean reversion is likely to be an individual-level phenomenon, it can potentially be mitigated or even eliminated by aggregating the instrument to group level, with group here referring to any larger observational unit. This was suggested in a recent contribution by Burns & Ziliak (2016), who define groups by birth cohort, education group, state, and year. To counter the problem of small-sample bias related to this aggregation, the authors impose a minimum group size of 50 observations.

In this article I show that the small-sample bias is better corrected by using leave-one-out group-averaging. Section 2 states the problem and the suggested solution in more detail, clarifying the potential small-sample biases. Section 3 presents results using the data and the specification of Burns & Ziliak (2016).

2 The small-sample biases of grouping approaches

Suppose we have a system of three equations with outcome variable y, endogenous regressor x, and instrument z, along with a vector of covariates Q. Each variable is observed for individual i belonging to group g.[1]

$$y_{i,g} = \beta x_{i,g} + \Gamma Q_{i,g} + e_{i,g} \tag{2}$$

$$x_{i,g} = \gamma z_{i,g} + \Pi Q_{i,g} + v_{i,g} \tag{3}$$

$$z_{i,g} = z_g + \Psi Q_{i,g} + u_{i,g} \tag{4}$$

Suppose we have reason to suspect our exclusion restriction may be violated at the individual level with $\text{cov}(e_{i,g}, u_{i,g} | Q_{i,g}) \neq 0$. [2] Suppose further that we are willing to assume that the exclusion restriction holds

[1] Group here refers to the level to which the instrument is aggregated. In the application in section 3 groups are defined by birth cohort, education, state, and year. For a reader familiar with grouping estimation this may sound unusual as groups often denote groups defined by demographic variables (here birth cohort, education, and state), while their time interaction is called a cell. To avoid confusion of general-interest readers between cells and groups, I only use the term group.

[2] See Weber (2014) for a discussion of how individual-level income dynamics violate the exclusion restriction of the synthetic tax instrument.

at group level: $\text{cov}(e_{i,g}, z_g|Q_{i,g}) = 0$. The population group average z_g is not observed, so we use a sample average $\overline{z_g} \equiv 1/n_g \sum_{j=1}^{n_g} z_{i,g}$, where n_g is the number of observations in the group.

In small samples, this sample average contains a non-negligible individual component from $u_{i,g}$ and thus $\text{cov}(e_{i,g}, \overline{z_g}|Q_{i,g}) \neq 0$. Burns & Ziliak (2016) approach this problem by discarding small groups from the analysis.

The authors motivate their approach by noting that a minimum group size criterion is also used in grouping estimation (Blundell et al. 1998). Grouping estimation is, however, different in its identification and possible biases. Grouping estimation does not use exogenous instruments, but instead relies on group-time interaction in the endogenous variable of interest for identifying variation. The exclusion restriction is the same as for grouped instruments: residual variation in the outcome variable should not have a group component.

Grouping estimation produces biased estimates in finite samples (Deaton 1985). One way of understanding this is to note that grouping estimation can be implemented by estimating the equation of interest at group level (see e.g. Angrist & Pischke 2008, section 4.1.3). Group averages estimated from a sample are error-ridden measures of true population group averages. Assuming random sampling, this imparts classical measurement error to the regressors, resulting in attenuation bias. In grouping estimation it is common to solve this problem by setting a minimum group size, commonly the aforementioned 50. This threshold is *ad hoc*, and an analytical solution for correcting the small-sample bias in grouping estimation has been derived by Devereux (2007).

Grouped instruments are similarly error-ridden measures of true population group averages, but classical measurement error in instrumental variables does not produce bias. The problem is that the sampling error is non-classical due to the non-validity of the instrument at the individual level.

Because the nature of the small-sample problem is different in grouping estimation and in estimation using grouped instruments, the proper solution is also different. For grouped instruments it is better to calculate leave-one-out group averages $\overline{z}_{i,g}^{LOO} \equiv 1/(n_g - 1) \sum_{j=1, j \neq i}^{n_g} z_{j,g}$.[3] As $\text{cov}(\overline{z}_{i,g}^{LOO}, u_{i,g}) = 0$ the instrument should be valid, unless instrument validity is compromised at the group level as well with $\text{cov}(z_g, e_{i,g}|Q_{i,g}) \neq 0$. An added benefit to leave-one-out averaging is that discarding observations is not necessary, cases with $n_g = 1$ notwithstanding.

$\overline{z}_{i,g}^{LOO}$ is still an error-ridden measure of z_g but with the sample mean being calculated without observation i, the measurement error is classical. Even though classical measurement error in an instrumental variable does not produce bias it does, however, reduce instrument strength, which can be assessed using conventional criteria.

3 Results

I illustrate the issue by replicating the analysis of Burns & Ziliak (2016), who use data from the March Supplement of the Current Population Survey (CPS) for 1979-2008. The authors present numerous elasticity estimates using different income definitions, income controls, income truncation specifications, and demographic controls. Here I will present results following their baseline specification for broad income, presented in Table 1 in their paper, using income splines as income control. I use broad income as the data does not include information on use of exemptions and deductions required for defining taxable income. Groups are defined by 5-year birth cohorts, a 3-category education variable (less than high school, high school, more than high school), state, and year. The vector of controls includes marital status, year, state, and education group-birth cohort interaction. The sample consists of 25-60 year-old household heads with minimum income of 10,000 real USD (2008 levels). Both observed and counterfactual net-of-tax rates are calculated using TAXSIM.

[3] For other applications of leave-one-out estimators, see Angrist (2014, section 3), and Goldsmith-Pinkham et al. (2018).

Table I presents results using Burns & Ziliak's (2016) replication data, provided as an attachement to their article. All specifications exclude observations in groups with less than 50 observations. The first two columns exactly replicate estimates from Table I in Burns & Ziliak (2016). They illustrate the authors' main result: grouping the Gruber-Saez instrumental variable (ln $[(1-\tau_t(y_{i,t-1}))/(1-\tau_{t-1}(y_{i,t-1}))]$) increases the elasticity estimate. The third column estimates broad income elasticity using the leave-one-out group-averaged instrumental variable. Excluding the individual observation in question from the group-averaged instrument reduces the F-value of the instrument from 153 to 8, below common thresholds for weak instruments. This reduction in instrument strength indicates that identifying variation in the second column comes from individual level, which is in contrast to the motivation for using group averages in the first place.

Table I: Broad income elasticities estimated using the Burns & Ziliak (2016) replication data

Instrument	Gruber-Saez	Group	Leave-one-out group
Elasticity	.12	.29	2.31
(s.e.)	(.09)	(.19)	(1.96)
First-stage F-statistic	119.6	153.1	8.3
# obs	198,285	198,285	194,248

The large reduction in instrument strength when using leave-one-out averaging is due to small group sizes, a result of an unconventional way Burns & Ziliak (2016) impose the minimum group size criterion. The authors exclude observations with less than 50 observations in groups defined only by birth cohort, education, and year, even though the instrument is also grouped by state. Looking at groups defined by the level of instrument aggregation (birth cohort, education, year, and state) I find that 97 % of their observations are in groups with less than 50 observations, with median group size being 11.

Strength of the instrument can potentially be improved by decreasing its sampling error. This can be achieved by constructing larger groups, which can be done by redefining the grouping or by using a larger sample. For better comparability to Burns & Ziliak's (2016) results, I will follow the latter approach.

Burns & Ziliak (2016) generate their group-averaged instrument from the same set of observations they use in the regressions, which are the first-year observations of longitudinally linked eligible households. I will now construct the instrument using an expanded dataset, which includes all eligible individuals in the CPS data. The expanded dataset includes 878,297 observations, compared to 198,285 observations used by Burns & Ziliak.[4] Group sizes are correspondingly larger: median group size is 43. Note that I rely on the longitudinal linking of Burns & Ziliak (2016). Thus the number of observations used in the regressions remains the almost[5] the same, while the number of observations used to construct the instrument increases.

The first two columns present results using the Gruber-Saez instrument and the group-averaged instrument estimated from longitudinally linked observations. The differences between the first two columns in Table II and the first two columns in Table I are due to differences in how income and tax variables are defined.[6] The elasticity estimates are similar in magnitude to those of Burns & Ziliak (2016), although the Gruber-Saez instrument and the group instrument produce almost the same estimate for the income elasticity. The estimate obtained using the Gruber-Saez instrument is slightly higher than those obtained by Gruber & Saez (2002) themselves and by Burns & Ziliak (2016).

The third and the fourth column use a larger set of observations to construct the instrument. Comparing columns 2 and 3 we find that calculating the group average of the instrument using larger groups increases the elasticity estimate. Increasing group sizes reduces the weight of any individual in the group-averaged

[4]CPS is a rotating panel with each household followed for two consecutive years. The reduction in sample size when moving from the expanded dataset to a longitudinally linked dataset beyond 50 % is due to imperfect matching of households across years. See Madrian & Lefgren (1999) for an analysis of longitudinal matching in CPS.

[5]The difference is due to the fact that a small number of households in the replication data have duplicate identifiers, and these household are dropped in Table II.

[6]I am unable to replicate Burns & Ziliak's (2016) income and tax variables, as the programs used to construct their replication dataset exist only partially (I thank the authors for patiently responding to my questions regarding this issue). The correlation coefficient between log change in income in their data and in my own is .92, and for log change in net-of-tax rate it is .77.

instrument, and consequently decreases the associated bias. Finally column 4 uses the leave-one-out group-averaged instrument, which reduces the weight of any individual i in the instrument assigned to him/her to zero. The elasticity estimate increases further, and importantly the instrument remains relatively strong with an F-value of 24.

Table II: Broad income elasticities estimated using an expanded sample

Instrument	Gruber-Saez	Group	Group	Leave-one-out group
Elasticity	.22	.22	.60	.83
(s.e.)	(.08)	(.14)	(.21)	(.39)
First-stage F-statistic	304.2	136.5	130.0	23.8
# obs used in the regression	198,197	198,197	198,197	198,197
# obs used in constructing the instrument	198,197	198,197	878,297	878,297

The increase in the elasticity estimate is economically significant. Using the parametrization of the Saez (2001) formula for the revenue-maximizing top marginal tax rate used by Burns & Ziliak (2016), the increase in the broad income elasticity from .22 to .83 would imply a reduction of the revenue-maximizing top marginal tax rate on broad income by 30 percentage points.

4 Conclusions

When estimating income tax elasticities such as the elasticity of taxable income or the elasticity of broad income, group-averaging the synthetic tax instrument is potentially a good approach to reducing bias due to individual-level income dynamics. More generally, grouping an instrumental variable is a valid approach in any setting where instrument validity is threatened at the individual level but less so at group level. Researchers applying this method should, however, use leave-one-out group averages to properly purge individual-level variation from the grouped instrument. The potential small-sample bias associated with this approach is a conventional weak instrument bias due to measurement error in the group-averaged instrument. The magnitude of the weak instrument problem can be assessed using conventional criteria, such as the first-stage F-test on the excluded instrument. The instrument can potentially be strengthened by increasing group sizes, which can be done by redefining the grouping or by using additional observations to estimate the instrument.

References

Angrist, J. & Pischke, J-S. (2008) ”Mostly Harmless Econometrics: An Empiricist's Companion”, Princeton University Press

Angrist, J. (2014) ”The perils of peer effects” *Labour Economics*, 30: 98-108

Blundell, R., Duncan, A. & Meghir, C. (1998) ”Estimating labour supply responses using tax reforms” *Econometrica*, 66(4): 827-861

Burns, S. K., & Ziliak, J. P. (2016) Identifying the elasticity of taxable income. *The Economic Journal*, 127(600), 297-329.

Chetty, R. (2009) ”Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance.” *American Economic Journal: Economic Policy*, 1(2): 31–52.

Deaton, A. (1985) ”Panel data from a time series of cross-sections” *Journal of Econometrics*, 30, 109–126

Devereux, P. (2007) ”Improved errors-in-variables estimators for grouped data” *Journal of Business & Economic Statistics*, 278-287

Feldstein, M. (1999) “Tax avoidance and the deadweight loss of the income tax” *Review of Economics and Statistics*, 81(4): 674–80.

Goldsmith-Pinkham, P., Sorkin, I. & Swift, H. (2018) ”Bartik instruments: what, when, why, and how” *NBER Working Paper 24408*

Gruber, J. & Saez, E. (2002) ”The elasticity of taxable income: evidence and implications” *Journal of Public Economics* 84: 1-32

Madrian, B. & Lefgren, L. (1999) ”A note on longitudinally matching current population survey (CPS) respondents” *NBER Technical Working Paper 247*

Saez, E. (2001). ”Using elasticities to derive optimal income tax rates” *Review of Economic Studies*, 68(1): 205–29.

Weber, C. E. (2014). Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences. *Journal of Public Economics*, 117, 90-103.