Working Paper Series
Congressional Budget Office
Washington, D.C.

Panel Data Techniques and the Elasticity of Taxable Income

Seth H. Giertz
Tax Analysis Division
Congressional Budget Office
Washington, D.C.

October 2008
CBO Working Paper 2008-11

Panel Data Techniques and the Elasticity of Taxable Income

Seth H. Giertz

Abstract

This paper examines the elasticity of taxable income with a focus on income controls designed to control for divergence in the income distribution and mean reversion. Additional emphasis is placed on the difference between short-run and longer-run responses to tax rate changes. Several panel techniques are applied to tax return data for years 1991 to 1997, followed by a cross-section analysis covering the same period. For each panel regression, an innovative inverted panel regression framework is employed to test the efficacy of the controls for mean reversion apart from controls for divergence in the income distribution. Finally, cross-section (and repeated cross-section) regressions are estimated for comparison. A major finding from comparing estimates from the standard and inverted panels is that even some of the more sophisticated techniques likely fail to adequately control for mean reversion at the top of the income distribution. Furthermore, the residual impact from mean reversion may still exert an enormous influence on elasticity estimates, which could help explain the lack of robustness reported in a number of papers in this literature. Analysis of cross-section data circumvents the problem of mean reversion and results in estimates that are robust with respect to sample income cutoffs. However, when vast differences likely exist between those experiencing a specific change in tax rates and other filers, estimates relying on either panel or cross-section data are likely to be poorly identified.

The author thanks Emmanuel Saez, Joel Slemrod, David Weiner, Ed Harris, and Tom Woodward.

I. Introduction

A series of CBO working papers has used panel data (primarily from tax returns) to measure the elasticity of taxable income (ETI).[1] This line of research has been motivated by an increasing recognition of the importance of the ETI in tax policy design. In particular, because the ETI measures the responsiveness of taxable income to changes in the net-of-tax rate (NTR),[2] it has emerged as a key parameter in assessing the efficiency and revenue implications of tax rate changes (Feldstein 1999). Attention in the CBO working papers has focused on the sensitivity of estimates to reasonable variants to the standard techniques employed in the literature. A major conclusion from this research is that isolating income responses to tax rate changes from the myriad other factors that influence income is difficult. And the range of estimates resulting from variations of what appear to be sound techniques is large, a conclusion supported by others working with panel data (Kopczuk 2005; Heim 2008). Where the true ETI falls within this wide range has tremendous implications for tax policy. For example, in examining the potential expiration of the Bush tax cuts, Giertz (2008b) reports that the deadweight loss per dollar of additional revenue from the federal income tax could range from $0.18 at an ETI of 0.2 to $1.25 at an ETI of 0.8.

It is widely recognized that mean reversion and diverging secular income trends across the income distribution are two of the key obstacles to identification in this literature.[3] Despite sophisticated techniques developed to address these phenomena, identification remains elusive. Furthermore, it is often unclear which of the two phenomena is dominant.

In this paper new elasticity estimates are produced with a focus on income controls designed to control for divergence in the distribution of income (between those with incomes in the highest 1 percent and the rest of filers) and mean reversion. Additional emphasis is placed on the difference between short-run and longer-run responses to tax rate changes. To this end, several different panel techniques are applied to tax return data for years 1991 to 1997, followed

---

[1] See Giertz (2005, 2006, 2008a, 2008c) and Eissa and Giertz (2006).

[2] More precisely, the ETI measures the percent change in taxable income associated with a 1 percent change in the NTR, where the NTR equal 1 minus the marginal tax rate – or the share of marginal income that taxpayer has after taxes.

[3] See Giertz (2004) and Slemrod (1998) for a discussion of these issues.

by a cross-section analysis covering the same period.[4] Cross-section analysis is not plagued by mean reversion, but must still confront divergence in the income distribution (as well as some potential obstacles not present with panel data analysis). The panel techniques include variations of a differencing approach employed by Gruber and Saez (2002) and others, as well as a regression-based tabulated differences-in-differences approach similar to that used by Moffitt and Wilhelm (2000). Additionally, for each panel regression, an innovative inverted panel regression framework is employed to test the efficacy of the controls for mean reversion apart from controls for divergence in the income distribution.[5] Finally, cross-section (and repeated cross-section) regressions are estimated for comparison.

A major finding from comparing estimates from the standard and inverted panels is that even some of the more sophisticated techniques likely fail to adequately control for mean reversion at the top of the income distribution. Furthermore, the residual impact from mean reversion may still exert an enormous influence on elasticity estimates, which could help explain the lack of robustness reported in a number of papers in this literature. Analysis of cross-section data circumvents the problem of mean reversion and results in estimates that are robust with respect to sample income cutoffs. However, in instances where vast differences likely exist between those experiencing a specific change in tax rates and other filers, estimates relying on either panel or cross-section data are likely to be poorly identified.

II. Data

Individual tax return data are from the Statistics of Income (SOI) and span years 1991 to 1997. The SOI heavily oversamples high-income filers, who, earlier research suggests, play a critical role in determining overall responses to changes in tax rates.[6] Those at the top of the income

[4] The key tax change over this period results from the Omnibus Budget Reconciliation Act of 1993 (OBRA 93), which created two new high-income tax brackets with marginal rates of 36 and 39.6 percent. OBRA 93 also lifted the cap on the Medicare portion of the FICA tax, which applies only to earned income.

[5] This technique effectively inverts the data. While the standard approach might compare changes in income from 1992 to 1993 to the increase in tax rates from 1992 to 1993, the inverted approach compares the changes in income from 1993 to 1992 to the *decrease* in tax rates from 1993 to 1992. For more detail, see Section III.

[6] The SOI's sampling strategy could cause regression estimates to be biased, but this possibility is averted by weighting observations by the inverse of their probability of appearing in the sample.

distribution are responsible for a very large and, for more than three decades, growing share of both overall income and tax payments (see CBO 2001 and Piketty and Saez 2003).

The focus of this paper is methodological (i.e., an examination of factors that may impede identification) rather than an attempt to produce a best estimate of the ETI, so adjusted gross income (AGI) may be a better measure than taxable income, since it eliminates some idiosyncratic variation that may be difficult to model. (And if current methods fail to adequately control for nontax factors influencing AGI, they will similarly fail with taxable income.) AGI is a simpler measure than taxable income and is likely less subject to fluctuation from factors such as changes to itemized deductions or exemptions. An accurate measure of the elasticity of AGI with respect to the net-of-tax rate (NTR) represents a lower bound for the ETI, since AGI responses are a subset of responses to taxable income.

For similar reasons, marginal tax rates are imputed (using CBO's federal tax model) only for the federal income tax, an approach that excludes changes in the payroll tax (which applies only to earned income) and in state tax rates. While this leaves an incomplete measure of the marginal tax rate, it has the virtue of simplicity, making it easier to focus on responses to OBRA 93 and exclude responses to changes in state tax rates. Additionally, for itemizers the effect of changes in state MTRs is mitigated since their overall NTR is $(1 - \tau_{fed})\cdot(1 - \tau_{state})$. Income, denoted z_{it}, is a constant-law measure of AGI for individual i at time t, excluding capital gains and including all Social Security benefits, such that

$$z = reported\_agi - realized\_CG + nontaxable\_SS\_benefits + deducted\_moving\_expenses. \quad (1)$$

Dollar values are adjusted by the IRS's inflation adjustment factors, using 1991 as the base.[7]

Summary Statistics

Average income for years 1991 to 1997 is about $39,000 (in 1991 dollars) and the average federal MTR is 16.2 percent.[8] Imputed MTRs are flat throughout the period, except at the very top of the income distribution (see Table 1). Data are examined in both cross-section and panel forms. For the cross section, individual identifiers are not used to link filers across years and thus fractiles

[7] See www.irs.gov/irb/2007-34_IRB/ar10.html.

[8] These averages are constructed to reflect the overall population of tax filers. This is done by weighting observations by the inverse of the probability of each filer being sampled.

Table 1. Average Federal Marginal Tax Rates by Income Fractile

Cross Section	1991	1992	1993	1994	1995	1996	1997
P50-80	18.5	18.3	18.2	18.4	18.5	18.8	19.1
P80-90	24.2	23.6	23.5	24.4	24.8	25.1	25.5
P90-95	27.8	27.7	27.8	28.1	28.0	28.2	28.2
P95-99	30.1	30.3	30.4	30.7	31.1	31.5	31.9
P99-100	32.4	32.3	39.3	39.5	39.6	39.6	39.5

Base Year:	1991		1992		1993		1994		1995		1996		1997	
Standard Panel	base year	year t+1	base year	year t+1	base year	year t+1	base year	year t+1	base year	year t+1	base year	year t+1		
P50-80	18.2	18.0	18.0	17.9	18.0	18.0	18.0	18.3	18.2	18.4	18.4	18.7		
P80-90	24.0	22.9	23.5	22.8	23.2	23.6	24.2	23.7	24.6	24.1	24.8	24.3		
P90-95	27.7	26.4	27.7	26.5	27.8	26.9	28.0	27.0	28.0	27.4	28.2	27.6		
P95-99	29.8	28.9	30.0	29.1	30.0	29.5	30.2	30.1	30.6	30.3	30.9	30.6		
P99-100	32.9	31.6	32.8	36.8	39.2	37.0	39.1	37.5	39.3	37.4	39.3	37.3		
Inverted Panel			base year	year t-1	base year	year t-1	base year	year t-1	base year	year t-1	base year	year t-1	base year	year t-1
P50-80			18.0	18.5	17.9	18.2	18.0	18.1	18.1	17.8	18.4	18.0	18.6	18.3
P80-90			23.4	23.3	23.2	22.8	24.1	22.6	24.6	23.4	24.9	23.5	25.3	23.9
P90-95			27.7	26.8	27.8	26.9	28.0	26.9	28.0	27.5	28.2	27.4	28.2	27.4
P95-99			29.8	29.2	29.9	29.3	30.1	29.6	30.5	29.9	30.8	30.2	31.2	30.6
P99-100			32.9	32.1	39.2	32.4	39.1	37.8	39.3	37.8	39.3	38.1	39.3	38.3

Source: Estimates are based on Statistics of Income data for years 1991 to 1997.

are recalculated each year based solely on reported income for the given year. Therefore, each year the filers composing a given fractile change based on the re-sorting of incomes. For the panel, individual identifiers are used and fractiles are based on the distribution of income in each of the base years. Thus, comparisons of income and MTRs between one year and either a subsequent (or preceding) year are based on the average of individual level changes. And filers in the subsequent (or preceding) year do not necessarily fall into the same fractile that they were in during the base year.

For fractiles in the top half of the income distribution, the change in average MTRs over consecutive years for the cross section is often well under 1 percent. However, the change in MTRs from 1992 to 1993 for the top 1 percent is 21.7 percent as a result of OBRA 93 (see Table 2). By comparison, the average MTR for those in the 95th to 99th percentile rose just 0.5 percent from 1992 to 1993, suggesting that OBRA 93 had little or no effect on those outside the top 1 percent.

For the panel, again the only substantial change in imputed MTRs is for the top 1 percent of the income distribution and occurs between 1992 and 1993. The average MTR for this top group increased 12.2 percent from 1992 to 1993—much less than the 21.7 percent increase observed in the cross section. This smaller growth is a result of mean reversion and rank reversal.[9] Many of those in the top 1 percent of the panel in 1992 did not experience an increase in tax rates as a result of OBRA 93 because their incomes fell in 1993, leaving them in a lower tax bracket. The average MTR for those in the 95th to 99th percentile actually fell by 2.7 percent from 1992 to 1993, suggesting that effects from mean reversion more than outweighed any increase in MTRs from OBRA 93.

Turning to income, AGI in the top 1 percent of the cross section drops by 6.5 percent from 1992 to 1993, as OBRA 93 takes effect (see Tables 3 and 4). In contrast, the average AGI for those

[9] Upper-income fractiles include many filers with unusually high incomes (compared to these filers' average incomes over a number of years). Examining the same filers over time (i.e., panel data analysis) implies that incomes in these top fractiles will fall (or grow at a slower rate) as a result of mean reversion (than they would were mean reversion not present). Put another way, filers are assigned to the top 1 percent fractile, for example, based on base (or initial) year income, independent of how much their income may drop between the base and subsequent year. Those with incomes below the top 1 percent in the base year are still assigned to their initial fractile, independent of how much their income may increase between the base and subsequent year. Repeated cross sections do not suffer from mean reversion because fractiles are recalculated each year. However, rank reversal is likely present with repeated cross sections. Rank reversal refers to changes in the composition of filers with reported incomes in a particular fractile over time. Rank reversal can make identification difficult if some sources of income are cross-sectionally correlated.

Table 2. Average Percent Change in Federal Marginal Tax Rates by Income Fractile

Cross Section	1991 to 1992	1992 to 1993	1993 to 1994	1994 to 1995	1995 to 1996	1996 to 1997
P50-80	-1.17	-0.16	0.86	0.88	1.57	1.14
P80-90	-2.50	-0.66	4.08	1.59	1.12	1.56
P90-95	-0.03	0.36	0.79	-0.17	0.71	0.08
P95-99	0.45	0.51	0.75	1.50	1.21	1.42
P99-100	-0.15	21.66	-0.64	0.27	-0.13	-0.05

Standard Panel	1991 to 1992	1992 to 1993	1993 to 1994	1994 to 1995	1995 to 1996	1996 to 1997
P50-80	-1.26	-0.17	0.44	1.30	1.28	1.49
P80-90	-4.68	-2.76	1.76	-2.20	-1.91	-2.00
P90-95	-4.85	-4.55	-3.33	-3.53	-2.05	-2.10
P95-99	-2.84	-2.74	-1.73	-0.45	-1.02	-1.15
P99-100	-3.86	12.20	-5.56	-4.16	-4.65	-5.02

Inverted Panel	1992 to 1991	1993 to 1992	1994 to 1993	1995 to 1994	1996 to 1995	1997 to 1996
P50-80	2.74	1.56	0.40	-1.85	-2.05	-1.95
P80-90	-0.17	-1.57	-6.30	-5.07	-5.37	-5.51
P90-95	-3.17	-3.23	-3.96	-1.79	-2.98	-2.91
P95-99	-1.88	-2.09	-1.42	-2.11	-2.07	-1.90
P99-100	-2.49	-17.32	-3.29	-3.66	-3.08	-2.66

Source: Estimates are based on Statistics of Income data for years 1991 to 1997.

in the 95th to 99th percentile fell by just 0.7 percent from 1992 to 1993. And over all other two-year comparisons, AGI rises for all income groups, although the increases are greatest for the top 1 percent, often more than 6 percent, and from 1991 to 1992, 11.7 percent.

For the panel, the top 1 percent's drop in income from 1992 to 1993 is over twice as large (compared to the cross section), at 13.3 percent. This larger drop in AGI is again driven by mean reversion (within the panel and rank reversal with the repeated cross sections) at the top of the income distribution. The cross-section numbers show that reported AGI for the top 1 percent falls and for the panel it falls even more because some of those in the top 1 percent in 1992 are further down in the distribution in 1993. Those in the 95th to 99th percentile saw their incomes fall by 4.6 percent from 1992 to 1993. For the top 1 percent, incomes fell by 5.8 percent from 1993 to 1994 (when the cap on the Medicare portion of the FICA tax was lifted). For the other

Table 3. Average Incomes by Income Fractile

Cross Section	1991	1992	1993	1994	1995	1996	1997
P50-80	36,677	36,815	36,148	36,255	36,918	37,191	38,272
P80-90	61,201	61,404	60,668	60,980	62,549	62,987	64,724
P90-95	81,421	82,021	80,941	81,690	84,667	85,608	88,353
P95-99	121,036	123,414	122,565	123,897	129,432	132,610	138,572
P99-100	411,733	460,016	429,921	434,023	466,321	496,219	538,084

Base Year:	1991		1992		1993		1994		1995		1996		1997	
Standard Panel	base year	year t+1	base year	year t+1	base year	year t+1	base year	year t+1	base year	year t+1	base year	year t+1		
P50-80	36,804	36,776	36,858	36,417	36,207	36,770	36,350	37,686	37,033	37,784	37,254	38,785		
P80-90	61,241	60,107	61,419	59,755	60,679	60,629	60,984	61,915	62,522	62,215	62,952	64,094		
P90-95	81,482	79,909	81,987	78,743	80,963	80,072	81,740	83,302	84,659	85,157	85,651	87,930		
P95-99	120,934	120,227	123,675	117,917	122,845	120,182	123,934	128,088	129,672	130,155	132,918	136,711		
P99-100	413,888	426,366	460,503	399,066	432,308	407,074	433,717	448,600	468,296	463,556	495,598	498,917		
Inverted Panel			base year	year t-1	base year	year t-1	base year	year t-1	base year	year t-1	base year	year t-1	base year	year t-1
P50-80			37,056	36,975	36,213	36,393	36,337	35,528	37,016	35,570	37,206	36,285	38,307	36,628
P80-90			61,311	60,349	60,543	60,435	60,860	59,654	62,636	59,772	63,025	61,037	64,717	61,794
P90-95			82,061	80,020	80,970	80,327	81,731	79,418	84,679	80,909	85,628	82,385	88,354	84,054
P95-99			122,192	119,682	121,271	121,774	122,674	119,467	128,613	120,981	131,780	127,807	137,598	133,599
P99-100			439,292	407,559	422,050	453,989	424,318	423,415	456,327	431,619	485,969	466,860	523,686	500,346

Source: Estimates are based on Statistics of Income data for years 1991 to 1997.

Table 4. Average Percent Change in Income by Income Fractile

Cross Section	1991	1992	1993	1994	1995	1996	1997
P50-80		0.38	-1.81	0.30	1.83	0.74	2.91
P80-90		0.33	-1.20	0.51	2.57	0.70	2.76
P90-95		0.74	-1.32	0.93	3.64	1.11	3.21
P95-99		1.96	-0.69	1.09	4.47	2.46	4.50
P99-100		11.73	-6.54	0.95	7.44	6.41	8.44

Standard Panel	1991 to 1992	1992 to 1993	1993 to 1994	1994 to 1995	1995 to 1996	1996 to 1997
P50-80	-0.08	-1.20	1.56	3.67	2.03	4.11
P80-90	-1.85	-2.71	-0.08	1.53	-0.49	1.81
P90-95	-1.93	-3.96	-1.10	1.91	0.59	2.66
P95-99	-0.58	-4.66	-2.17	3.35	0.37	2.85
P99-100	3.01	-13.34	-5.84	3.43	-1.01	0.67

Inverted Panel	1992 to 1991	1993 to 1992	1994 to 1993	1995 to 1994	1996 to 1995	1997 to 1996
P50-80	-0.22	0.50	-2.23	-3.90	-2.47	-4.38
P80-90	-1.57	-0.18	-1.98	-4.57	-3.16	-4.52
P90-95	-2.49	-0.79	-2.83	-4.45	-3.79	-4.87
P95-99	-2.05	0.41	-2.61	-5.93	-3.01	-2.91
P99-100	-7.22	7.57	-0.21	-5.41	-3.93	-4.46

Source: Estimates are based on Statistics of Income data for years 1991 to 1997.

two-year comparisons, AGI reported by top 1 percent of the panel either increases or marginally declines.

III. Methodology

Both panel and cross-section regressions are run on data for pairs of years in order to capture the short-, medium-, and long-run responses to the increase in top MTRs in 1993. The short run comprises responses from 1992 to 1993; the medium run, from 1991 to 1994; and the long run, 1991 to 1997. Additionally, each of the panel specifications is followed by a pooled panel, which focuses on short-run responses and includes all years 1991 to 1997. Likewise, the cross-section analysis is followed by a repeated cross section that includes all years in the sample.

Standard Panel

As defined in the previous section, z_{it} is the income reported by filer i in year t; τ_{it} is the marginal tax rate (on ordinary income); and k is the interval (one, three, or six years) over which behavior is examined.

The estimating equation for the standard panel is

$$\log[z_{i,t+K}/z_{it}] = \alpha + \psi \cdot \log[(1 - \tau_{i,t+K})/(1 - \tau_{it})] + X_{it}\beta + \varepsilon_{it}. \quad (2)$$

The key coefficient (ψ) is the elasticity, which measures the percent change in income associated with a 1 percent change in the NTR. The equation is estimated via two-stage least squares (2SLS) and all regressions are weighted by z_{it}.

All of the panel regressions are estimated using two different instruments for the tax variable. Instrumenting is necessary in order to focus on the effect of the NTR on income. The NTR is also a function of income; e.g., with a progressive rate structure, increases in income lead to lower NTRs. The first instrument is an imputed tax rate of the form $\log[(1-\tau_{i,t}^{t+K})/(1-\tau_{it})]$; i.e., the log change in the NTR applying to the last dollar of income from year t to $t+K$, where imputed income in year $t+K$ equals income in year t inflated (by the IRS's inflation adjustment factors) to year $t+K$; thus, differences across the two years arise from statute and not from behavioral changes. Regressions are rerun using a second instrument that targets the top 1 percent of the income distribution (i.e., those affected by the tax change). This instrument is simply a dummy for those in the top 1 percent of the reported distribution of z_{it}; the instrument equals 1 where $p_{it} > 0.99$ and zero otherwise, where p is the tax filer's place in the income distribution, as measured in percentiles. The latter approach (using a dummy variable as the instrument) is similar in spirit to Feldstein's (1995) tabulated differences-in-differences approach, but put into a regression framework (Moffitt and Wilhelm 2000). The regression framework allows for the inclusion of control variables.

There are tradeoffs between the two approaches. The downside of the dummy variable approach is that the top 1 percent (those for whom $p_{it} > 0.99$) is an imperfect approximation of the group experiencing a tax change, because it will include some filers who do not experience a tax change and exclude others who do. Additionally, the instrument does not account for variation in NTR decreases for those in the top 1 percent. However, the virtue of this method is not in the precision of the instrument but in its simplicity: although the instrument is cruder (than the imputed rate instrument) it is also more transparent and easier to interpret.[10] Its source of identification is the disproportionate tax change experienced by the top 1 percent. With just two years of data and no controls, the regression coefficient (i.e., estimated elasticity) equals the difference in the average income growth rate between the top 1 percent and the rest of the distribution divided by the difference in average NTR changes for the two groups.[11]

Each equation is estimated first with no controls and then with two different sets of income controls, where X_{it} equals either (1) log(z_{it}) or (2) log(z_{it}) plus additional splines of log(z_{it}). Each equation is also estimated using three different income cutoffs: (1) $p_{it} > 0.5$; (2) $p_{it} > 0.9$; and (3) $z_{it} > \$20,000$. Because the effects from OBRA 93 are concentrated within the top 1 percent of tax filers, the income cutoffs will effectively change the composition of only the "counterfactual" group of filers who did not experience a tax change.

Inverted Panel

One of the problems in identifying an ETI is that changes in tax rates can be spuriously correlated with mean reversion at the top of the income distribution. For tax rate increases, the NTR will be positively correlated with mean-reverting changes; for rate decreases, the reverse is true. Researchers often include control variables to account for mean reversion at the top, but the effectiveness of these controls is not known. A potential solution to this problem is to focus on samples that include both tax increases and decreases (see Giertz 2007), in the hope that mean reversion from the two tax changes will cancel each other out (i.e., will have a net effect of zero). However, mean reversion at the top of the income distribution may vary across the two time periods because of different macroeconomic conditions or other unobserved factors.

Another alternative is to "construct" a tax change by inverting the panel. For example, instead of starting in 1991 and continuing to 1997 (with a tax increase in 1993), suppose that one starts in 1997 and continues to 1991 (with a tax *decrease* in 1993). Thus, OBRA 93 would

[10] The imputed rate instrument ($\log[(1-\tau_{i,t}^{t+K})/(1-\tau_{it})]$) incorporates many factors that affect the MTR simultaneously to estimate the elasticity. It is less transparent in the sense that is difficult to grasp what all those changes are. For example, small changes at the lower end of the income distribution due to a quirk in the law (such as the rounding error for adjustment for inflation) could potentially have a substantial effect on the estimated elasticities.

[11] Because the equations are estimated in log form, the estimated coefficients are in percent terms.

represent a *cut* in top MTRs. If the model is correctly identified, then the estimated elasticity should be the same for both the standard and the inverted panel.[12]

For the inverted panel approach, the estimating equation is identical to equation (3) with t and $t+K$ inverted throughout, such that

$$\log[z_{i,t}/z_{i,t+K}] = \alpha + \psi\cdot\log[(1-\tau_{it})/(1-\tau_{i,t}+K)] + X_{i,t+k}\,\beta + \varepsilon_{it}. \quad (4)$$

The imputed rate instrument is now $\log[(1-\tau^{t}_{i,t+K})/(1-\tau_{i,t+K})]$—i.e., the log change in the MTR from year $t+K$ to t, where year t income equals income in year $t+K$ deflated to year t. For the approach that targets the top 1 percent of income distribution, the instrument equals 1 where $p_{it+1} > 0.99$ (i.e., for the top 1 percent of z_{it+k}).

Pooled Panel

In addition to estimating equation (2) for three different pairs of years (using the different income cutoffs and control variables), the equation is estimated using one-year differences and including all years, 1991 to 1997. The estimating equation can be expressed such that

$$\log[z_{i,t+1}/z_{it}] = \alpha + \zeta_t + \psi\cdot\log[(1-\tau_{i,t+1})/(1-\tau_{it})] + X_{it}\beta + \varepsilon_{it}\,, \quad (3)$$

where ζ_t are year dummies. The same income cutoffs and income controls (X_{it}) used for the panel analysis that relied on a single set of paired years are also used here. The imputed rate instrument for the pooled regressions is constructed identically to that for the two-year panel regressions. The instrument for the approach that targets the top 1 percent of the income distribution is again a dummy variable that equals 1 for the top 1 percent of filers for paired observations where t equals 1992 (and $t + 1$ equals 1993) and equals zero for all other paired observations; thus, the instrument is turned on only for the reform year at the top 1 percent of the distribution.

Paired Cross Sections

A solution to the problem of mean reversion is to use pooled cross-section data instead of a panel. With a cross section, the data are not paired across years, so individual indentifiers are

[12] This inverted methodology was suggested to me by Joel Slemrod, who, to my knowledge, was the first to come up with the idea (at least in the context of the tax responsiveness literature). The approach was used as a sensitivity check by Auten and Carroll (1999).

not needed (except possibly to adjust for heteroskedasticity) and the sample is not restricted to filers present in both years. The cross-section sample is constructed such that years t and $t+k$ are stacked. The estimating equation takes the form

$$\log z_{i,s} = \alpha + \psi \cdot \log(1 - \tau_{i,s}) + \gamma_1 \cdot \textit{post-obra\_dummy}_{is} + \gamma_2 \cdot \textit{high-income\_dummy}_{is} + \varepsilon_{is}, \qquad (5)$$

where s is the year. The post-OBRA dummy equals 1 for observations for years 1993 to 1997 and zero otherwise. The treatment dummy equals 1 for taxpayers in the top 1 percent of the distribution (i.e., those affected by the reform) and zero otherwise. The instrument for the NTR equals the product of the NTR and a dummy that equals 1 only for the top 1 percent of the income distribution in the end year ($s = t+k$). As with the panel, regressions are run using two years of data (to yield short-, medium-, and long-run estimates). For each set of years, regressions are weighted by z_{is} and estimated for the following segments of the income distribution: (1) $p_{is} > 0.5$; (2) $p_{is} > 0.8$; (3) $p_{is} > 0.9$; (4) $p_{is} > 0.95$; and (5) $z_{is} > \$20{,}000$.

Pooled Cross Sections

The cross-section analysis is repeated when including all years 1991 to 1997. The estimating equation for the repeated cross section is

$$\log z_{i,s} = \alpha + \psi \cdot \log(1 - \tau_{i,s}) + \gamma_1 \cdot \textit{post-obra\_dummy}_{is} + \gamma_2 \cdot \textit{high-income\_dummy}_{is} + \varepsilon_{is}. \qquad (6)$$

The postreform dummy equals 1 for years 1993 to 1997 and zero otherwise. The treatment dummy equals 1 for observations in the top 1 percent of the distribution in years 1993 to 1997 and zero otherwise. This model is then reestimated after adding two time trend variables, one of the form s and one in which s is interacted with a dummy for those in the top 1 percent of the income distribution (i.e., $p_{is} > 0.99$). The time trends are intended to control for differing secular trends in income between those affected and unaffected by the tax change.

Sample Weighting

In addition to weighting regressions by income, regressions are also weighted to adjust for the SOI's nonrandom sampling properties. Inclusion in the SOI is conditional on several factors, including income. Sampling probabilities reach 100 percent for very high income filers. The SOI is also constructed such that once filers are sampled, they continue to be sampled in all subsequent years, so long as their income increases (and other characteristics, such as filing status, do not change). In fact, the probability that one is observed in two different years is simply the minimum of the sampling probabilities for the two years. Without weighting, that sampling strategy raises the potential for spurious correlation between the dependent variable and the independent variables, including the tax variable. To avoid that possibility, observations (or paired observations) from the full SOI are weighted by the reciprocal of their probability of appearing in the sample. That strategy is discussed in Auten and Carroll (1999), who also employ the strategy using SOI data.

IV. Results

Findings from the Panel and Inverted Panel Analysis

Without income controls, mean reversion (at the top) appears to dominate. In the standard panel, mean reversion appears to be positively correlated with a falling NTR, leading to large positive elasticity estimates.[13] With the inverted panel, a rising NTR is negatively correlated with mean reversion, resulting in negative estimated elasticities (see the panel estimates for specification *a* for each of the income cutoffs in Table 5).[14] These results are consistent with those reported by Auten and Carroll (1999) for the Tax Reform Act of 1986. With their richest set of controls, Auten and Carroll report an estimated elasticity (for gross income) of 0.57. With the inverted approach, this estimate nearly doubles to 1.13. Note that Auten and Carroll's result is also consistent with mean reversion, since their estimate is much larger when the data ordering simulates a tax increase (or a rising NTR).

Sample income cutoffs, which determine the comparison group of taxpayers not experiencing a tax change, often have a big impact on my estimated elasticities.[15] Restricting the sample to those in the top 10 percent of z_{it} generally results in estimates from the standard panel

[13] Without control variables, the drop in top incomes by filers above their longer-run average incomes will likely be spuriously correlated with a falling NTR. In addition to income controls, including multiple pairs of years (some of which do not span a tax change) might mitigate this problem.

[14] For both the standard and inverted panels, more rapid secular income growth at the top of the income distribution will tend to bias elasticity estimates downwards—unless controls adequately account for divergence in the income distribution.

[15] Income in the subsequent year of each paired observation is not subject to the same income restriction, but is restricted to be greater than zero.

Table 5. Panel Elasticity Estimates Using the Imputed Rate Instrument

Income Cutoff:	$p_{it} > 0.5$			$p_{it} > 0.9$			$z_{it} > 20,000$		
specification:	1	2	3	1	2	3	1	2	3
1992v1993	1.068	0.638	0.425	1.108	-0.097	-0.019	1.268	0.917	0.739
std. err.	(0.129)	(0.122)	(0.175)	(0.092)	(0.137)	(0.186)	(0.152)	(0.156)	(0.216)
obs.	52,236	52,236	52,236	32,315	32,315	32,315	55,313	55,313	55,313
1991v1994	2.248	1.819	1.941	1.323	0.403	0.133	1.284	1.028	1.090
std. err.	(0.224)	(0.213)	(0.257)	(0.154)	(0.213)	(0.230)	(0.101)	(0.102)	(0.115)
obs.	45,603	45,603	45,603	28,055	28,055	28,055	47,936	47,936	47,936
1991v1997	1.956	1.285	1.370	1.030	0.305	-0.101	1.174	0.659	0.617
std. err.	(0.254)	(0.224)	(0.261)	(0.184)	(0.241)	(0.225)	(0.120)	(0.113)	(0.124)
obs.	41,524	41,524	41,524	25,697	25,697	25,697	43,572	43,572	43,572
INVERTED PANEL	dependent var: ln(z[it]/z[it+k])								
1992v1993	-0.651	-0.694	-0.703	-0.661	-0.579	-0.857	-0.738	-0.780	-0.827
std. err.	(0.079)	(0.098)	(0.122)	(0.048)	(0.073)	(0.090)	(0.079)	(0.087)	(0.112)
obs.	52,211	52,211	52,211	32,248	32,248	32,248	54,935	54,935	54,935
1991v1994	-0.144	-0.079	0.099	-0.828	-0.126	-0.086	-0.289	-0.273	-0.194
std. err.	(0.082)	(0.095)	(0.099)	(0.064)	(0.084)	(0.092)	(0.085)	(0.105)	(0.106)
obs.	45,593	45,593	45,593	27,813	27,813	27,813	47,859	47,859	47,859
1991v1997	-0.137	0.106	0.423	-2.207	-1.040	-1.178	-0.259	0.005	0.223
std. err.	(0.099)	(0.114)	(0.117)	(0.077)	(0.104)	(0.110)	(0.105)	(0.131)	(0.132)
obs.	42,350	42,350	42,350	25,752	25,752	25,752	44,895	44,895	44,895
POOLED REGULAR PANEL with YEAR DUMMIES									
	dependent var: ln(z[it+1]/z[it])								
ln(z[it+1]/z[it])	0.289	0.255	0.237	0.330	0.145	0.143	0.288	0.241	0.238
std. err.	(0.077)	(0.076)	(0.077)	(0.201)	(0.201)	(0.200)	(0.063)	(0.062)	(0.063)
obs.	331,782	331,782	331,782	208,745	208,745	208,745	349,208	349,208	349,208
POOLED INVERTED PANEL with YEAR DUMMIES									
	dependent var: ln(z[it]/z[it-1])								
ln(z[it]/z[it-1])	-0.211	-0.683	-0.606	-0.388	-1.008	-0.707	-0.180	-0.787	-0.729
std. err.	(0.058)	(0.046)	(0.035)	(0.106)	(0.083)	(0.054)	(0.056)	(0.043)	(0.034)
obs.	332,720	332,720	332,720	208,711	208,711	208,711	350,781	350,781	350,781

1) No controls. 2) log z[it] control. 3) log z[it] and additional splines in log z[it].
Regressions are weigted by income.

Source: Estimates are based on Statistics of Income data for years 1991 to 1997.

that are much smaller (and not statistically different from zero) than estimates from identical methods applied to samples with lower income cutoffs. Even modestly raising the income cutoff for inclusion in the sample, from $z_{it} < \$20{,}000$ to $p_{it} < 0.5$, leads to very different elasticity estimates: Estimated elasticities based on the standard panel rise sharply over the medium and long term, but drop over the short term.

Including splines of log(z_{it}) has a substantial impact on the estimates, yielding positive, but smaller, elasticity estimates for the standard panel that, when imposing a $20,000 income cutoff, range from 0.62 to 1.1 depending on the two years compared. For the inverted panel, the controls raise the estimated elasticity in some cases, but in many instances estimates are still negative (see Table 5, specification *c*). When imposing a $20,000 income cutoff, estimates from the inverted panel range from −0.83 to (a statistically insignificant) 0.22 depending on the two years compared. Including the richest set of controls brings the estimates for the inverted and standard panels closer together (when comparing results over the same time interval), but the differences that remain are still substantial. This suggests that while including the richest set of income controls may improve identification, mean reversion still appears to severely contaminate elasticity estimates.

One might expect estimates for 1992 versus 1993 to be larger than those covering longer intervals because the literature suggests substantial intertemporal income shifting, even when longer-run estimated elasticities are small. A comparison of panel estimates using the richest controls is not consistent with this hypothesis—for both the standard and inverted panels.

Using the cruder instrument, a dummy variable turned on for those in the top 1 percent of the income distribution when the reform occurs (i.e., the instrument equals 1 when both $p_{it} > 0.99$ and $t = 1992$, and zero otherwise) results in elasticity estimates from the model without income controls that are qualitatively similar to those found from the standard panel analysis. In other words, mean reversion appears to dominate (see Table 6).

In contrast to the analysis with the imputed rate instrument, estimates when including the richest set of income controls appear to severely *overcorrect* for mean reversion, resulting in large negative estimates for the regular panel and positive estimates for the inverted panel. Standard errors are also very large, so even large estimates are rarely statistically different from zero.

With the richest controls (and using the cruder instrument), estimates are generally much larger for 1992 versus 1993 than when comparing longer time intervals. This, in contrast to the

Table 6. More Panel Elasticity Estimates: Tabulated Differences-in-Differences within a Regression Framework

Income Cutoff:	$p_{it} > 0.5$			$p_{it} > 0.9$			$z_{it} > 20{,}000$		
specification:	1	2	3	1	2	3	1	2	3
1992v1993	1.878	0.814	-1.669	1.395	-0.721	-1.866	1.858	1.002	-1.701
std. err.	(0.184)	(0.149)	(1.052)	(0.107)	(0.213)	(0.711)	(0.189)	(0.190)	(1.139)
obs.	52,236	52,236	52,236	32,315	32,315	32,315	55,313	55,313	55,313
1991v1994	3.352	1.347	-2.602	2.420	2.201	-12.494	3.137	2.233	-2.795
std. err.	(0.446)	(0.332)	(2.966)	(0.221)	(1.115)	(18.013)	(0.421)	(0.615)	(3.414)
obs.	45,603	45,603	45,603	28,055	28,055	28,055	47,936	47,936	47,936
1991v1997	4.955	0.498	-3.742	3.022	-26.805	-6.482	4.720	0.362	-3.818
std. err.	(0.963)	(0.432)	(1.528)	(0.382)	(22.439)	(1.809)	(0.885)	(0.751)	(1.687)
obs.	41,524	41,524	41,524	25,697	25,697	25,697	43,572	43,572	43,572
INVERTED PANEL	dependent var: ln(z[it]/z[it+k])								
1992v1993	-0.488	-0.468	0.948	-0.430	0.314	0.948	-0.484	-0.504	0.949
std. err.	(0.093)	(0.155)	(0.707)	(0.058)	(0.129)	(0.375)	(0.103)	(0.133)	(0.766)
obs.	52,211	52,211	52,211	32,248	32,248	32,248	54,935	54,935	54,935
1991v1994	1.317	13.156	0.517	2.215	0.107	0.562	1.139	10.940	0.503
std. err.	(0.545)	(8.331)	(0.910)	(0.549)	(0.244)	(0.497)	(0.556)	(5.239)	(0.967)
obs.	48,764	48,764	48,764	30,573	30,573	30,573	51,129	51,129	51,129
1991v1997	-3.272	-9.561	0.154	-3.713	0.299	0.465	-3.135	-10.440	0.165
std. err.	(0.192)	(1.101)	(1.202)	(0.121)	(0.511)	(2.194)	(0.195)	(1.238)	(1.302)
obs.	42,350	42,350	42,350	25,752	25,752	25,752	44,895	44,895	44,895
POOLED REGULAR PANEL with YEAR DUMMIES									
	dependent var: ln(z[it+1]/z[it])								
ln(z[it+1]/z[it])	1.878	0.955	0.723	1.395	0.537	0.564	1.858	1.115	0.736
std. err.	(0.338)	(0.247)	(0.260)	(0.296)	(0.264)	(0.259)	(0.325)	(0.276)	(0.257)
obs.	331,784	331,784	331,784	208,747	208,747	208,747	349,210	349,210	349,210
POOLED INVERTED PANEL with YEAR DUMMIES									
	dependent var: ln(z[it]/z[it-1])								
ln(z[it]/z[it-1])	-0.488	-3.996	-1.874	-0.430	-3.003	-1.897	-0.484	-4.150	-1.913
std. err.	(0.099)	(0.064)	(0.083)	(0.115)	(0.097)	(0.079)	(0.101)	(0.063)	(0.085)
obs.	332,720	332,720	332,720	208,711	208,711	208,711	350,781	350,781	350,781

1) No controls. 2) log z[it] control. 3) log z[it] and additional splines in log z[it].
Regressions are weigted by income.

Source: Estimates are based on Statistics of Income data for years 1991 to 1997.

standard panel analysis, is consistent with the larger short-term responses and smaller long-term responses often reported in the literature.

Findings from the Pooled Panel Analysis

The pooled analysis confirms the results from the separate panels. Using either of the two instruments, mean reversion still seems to dominate regression estimates when excluding income controls (see Tables 5 and 6). With the richest set of income controls, comparing estimates from the standard and inverted panels suggests that mean reversion is still present and likely severe. Standard errors are much smaller than for the separate panels, so that estimates from applying the cruder instrument are now statistically significant. In contrast to the two-year panels, the pooled panel analysis suggests that the approach using the cruder instrument still undercorrects for mean reversion. Estimated elasticities when using the standard panel are positive and estimates from the inverted panel are negative. The disparity between the standard and inverted panel estimates is smaller (but still very large in terms of its economic importance) when using the imputed rate instrument.

Findings from the Cross-Section Analysis

In contrast to the panel analysis, estimates from the cross-section analysis are robust to changes to the sample income cutoff. Estimates are large (and positive) when including just years 1992 and 1993. For the other pairs of years, estimated elasticities are negative. This suggests that short-term shifting is much larger than longer-term behavioral responses. However, the large negative estimates for the other pairs of years suggest that the model is not identified.

Findings from the Repeated Cross-Section Analysis

The simple repeated cross-section analysis, which excludes time trends, yields negative elasticity estimates (see Table 7). This result may suggest that differential secular income trends are driving the estimates. Those at the top of the income distribution, whose NTR falls in 1993,

Table 7. Cross-Section-Based Elasticity Estimates

	Income Cutoff:	$p_{it} > 0.5$	$p_{it} > 0.8$	$p_{it} > 0.9$	$p_{it} > 0.95$	$z_{it} > \$20k$
	dependent var: $\ln(z_{it})$					
1	1992v1993	0.958	0.977	0.998	1.024	1.026
		(0.107)	(0.108)	(0.109)	(0.111)	(0.108)
		133,825	100,446	86,869	78,193	140,634
2	1991v1994	-0.576	-0.536	-0.481	-0.409	-0.520
		(0.111)	(0.113)	(0.115)	(0.117)	(0.109)
		145,751	111,310	96,930	86,550	152,349
3	1991v1997	-2.830	-2.768	-2.565	-2.587	-2.801
		(0.139)	(0.144)	(0.143)	(0.159)	(0.134)
		159,046	123,170	108,482	98,169	166,708
4	Repeated CS	-0.570	-0.499	-0.404	-0.315	-0.542
		(0.087)	(0.089)	(0.089)	(0.093)	(0.086)
		519,265	395,489	345,990	313,126	544,360
5	Repeated CS	1.342	1.326	1.329	1.280	1.460
	w/ 2 time-trend vars	(0.104)	(0.104)	(0.107)	(0.106)	(0.105)
		519,265	395,489	345,990	313,126	544,360

Regressions are weigted by income.
Repeated cross-section regressions include year fixed effects.

Source: Estimates are based on Statistics of Income data for years 1991 to 1997.

experience a sharp increase in their share of total income (due to nontax factors). This could lead to a spurious negative correlation between the falling NTR and their rising incomes. Adding two time trends to the model (i.e., a standard time trend and a second time trend that applies only to those in the top 1 percent of the income distribution) fundamentally changes the elasticity estimates. They are now very large, ranging from 1.28 to 1.46, depending on the income cutoff for inclusion in the sample. The time trends, by allowing for separate secular income trends for those experiencing the tax change, counteracts the negative bias present in the first set of repeated cross-section estimates. While the time trends have an enormous impact, the model may still not be identified. Income trends differ within the top 1 percent of the income distribution, which are not accounted for here. Furthermore, there are a host of unobservables between those experiencing the tax increase and the rest of the income distribution that likely influence income growth in complex ways that cannot be accurately modeled with simple time trends.

V. Conclusion

This paper examines behavioral responses to the 1993 tax increase that primarily affected the top 1 percent of the income distribution. Several different panel techniques are applied to tax return data for years 1991 to 1997, followed by a cross-section analysis covering the same period. For each panel regression, an inverted regression is also run to test the efficacy of the controls for mean reversion. Estimates from the panel analysis are sensitive to sample income cutoffs, income controls and choice of instrument. Additionally, comparing similarly estimated elasticities from the standard and inverted panels suggests that even the richest income controls imperfectly control for mean reversion—and the residual impact from mean reversion may still be of enormous economic importance. Analysis of cross-section data circumvents the problem of mean reversion and results in estimates that are robust with respect to sample income cutoffs. However, when vast differences likely exist between those who experience a specific change in tax rates and other filers, estimates relying on panel or cross-section data are likely to be poorly identified.

References

Auten, Gerald, and Robert Carroll. 1999. The Effect of Income Taxes on Household Behavior. *Review of Economics and Statistics* 81(4):681-693.

Congressional Budget Office. 2001. *Effective Federal Tax Rates, 1979-1997.* Washington: Congressional Budget Office.

Eissa, Nada and Seth Giertz. 2006. "Trends in High-Income and Behavioral Responses to Taxation: Evidence from Executive Compensation and Statistics of Income Data," Congressional Budget Office, Washington D.C., Working Paper 2006-14, (www.cbo.gov/ftpdocs/77xx/doc7711/2006-14.pdf)

Feldstein, Martin. 1995. The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act. *Journal of Political Economy* 103(3):551-572.

——. 1999. Tax Avoidance and the Deadweight Loss of the Income Tax. *Review of Economics and Statistics* 4(81):674-680.

Giertz, Seth. 2008a . Taxable Income Responses to 1990s Tax Acts: Further Explorations. Working Paper 2008-08. Washington: Congressional Budget Office. www.cbo.gov/ftpdocs/89xx/doc8936/2008-01.pdf.

——. 2008b (forthcoming). How Does the Elasticity of Taxable Income Affect Economic Efficiency and Tax Revenues and What Implications Does This Have for Tax Policy Moving Forward. In *Tax Policy Lessons from the 2000s*. Alan Viard, ed. Washington: AEI Press.

——. 2008c. A Sensitivity Analysis of the Elasticity of Taxable Income. Working Paper 2008-01. Washington: Congressional Budget Office. (www.cbo.gov/ftpdocs/89xx/doc8936/2008-01.pdf).

——. 2007. The Elasticity of Taxable Income over the 1980s and 1990s. *National Tax Journal* 60(4):743-768.

——. 2006. The Elasticity of Taxable Income During the 1990s: A Sensitivity Analysis. Working Paper 2006-3. Washington: Congressional Budget Office. (www.cbo.gov/doc.cfm?index=7037).

——. 2004. Recent Literature on Taxable-Income Elasticities. Technical Paper 2004-16. Washington: Congressional Budget Office. (www.cbo.gov/ftpdocs/60xx/doc6028/2004-16.pdf).

Goolsbee, Austan. 2000. What Happens When You Tax the Rich? Evidence from Executive Compensation. *Journal of Political Economy* 108(2):352-378.

Gruber, Jonathan, and Emmanuel Saez. 2002. The Elasticity of Taxable Income: Evidence and Implications. *Journal of Public Economics* 84(1):1-32.

Heim, Bradley. 2008. The Elasticity of Taxable Income: Evidence from a New Panel of Tax Returns. Working paper.

Kopczuk, Wojciech. 2005. Tax Bases, Tax Rates and the Elasticity of Reported Income. *Journal of Public Economics* 89(11-12):2093-2119.

Moffitt, Robert, and Mark Wilhelm. 2000. "axation and the Labor Supply Decisions of the Affluent. In Slemrod, J., ed., *Does Atlas Shrug? The Economic Consequences of Taxing the Rich.* New York: Harvard University Press and Russell Sage Foundation.

Piketty, Thomas, and Emmanuel Saez. 2003. Income Inequality in the United States, 1913–1998. *Quarterly Journal of Economics* 118(1):1-39.

Saez, Emmanuel. 2004. Reported Incomes and Marginal Tax Rates, 1960–2000: Evidence and Policy Implications. In Poterba, J., ed., *Tax Policy and the Economy 18*. Cambridge: MIT Press, 117-173.

Sammartino, Frank, and David Weiner. 1997. Recent Evidence on Taxpayers' Response to the Rate Increases in the 1990s. *National Tax Journal* 50(3):683-705.

Slemrod, Joel. 1996. High Income Families and the Tax Changes of the 1980s: The Anatomy of Behavioral Response. In Feldstein, M., Poterba, J., eds., *Empirical Foundations of Household Taxation.* Chicago: University of Chicago Press, 169-189.

——. 1998. Methodological Issues in Measuring and Interpreting Taxable Income Elasticities. *National Tax Journal* 51(4):773-788.

Steurle, C. Eugene. 2004. Contemporary U.S. Tax Policy. Washington: Urban Institute Press.