DEMS WORKING PAPER SERIES

Drivers of mean reversion bias in the estimation of elasticity of taxable income in an autoregressive framework

Bruno Bosco, Paolo Maranzano

No. 578 – June 2026

Department of Economics, Management and Statistics
University of Milano – Bicocca
Piazza Ateneo Nuovo 1 – 2016 Milan, Italy
http://dems.unimib.it/

DRIVERS OF MEAN REVERSION BIAS IN THE ESTIMATION OF ELASTICITY OF TAXABLE INCOME IN AN AUTOREGRESSIVE FRAMEWORK

Bruno Bosco[a] Paolo Maranzano[a,b,*]
bruno.bosco@unimib.it paolo.maranzano@unimib.it

a) Department of Economics, Management and Statistics (DEMS), University of Milan-Bicocca, Piazza Ateneo Nuovo, 1, 20127 Milan, Italy
b) Fondazione Eni Enrico Mattei (FEEM), Corso Magenta, 63, 20123 Milan, Italy
* Corresponding Author

ABSTRACT

Mean reversion, the tendency for taxpayers with unusually high or low income in one period to move towards their long-run average, is a key challenge for estimating the elasticity of taxable income (ETI), as it generates bias. We show that this bias can be characterised through the variance of the tax treatment and its covariance with windfall gains in an autoregressive model of taxable income dynamics. We further show that the bias decreases when the tax reaction of marginal treated taxpayers is weaker than that of average treated taxpayers. The paper provides analytical derivations and interpretation for these results.

Keywords: Elasticity of Taxable Income (ETI); Mean reversion and estimation bias; Variance-covariance between tax treatment assignment and windfall income shocks; Magnitude and persistence.

JEL Codes: H300; H24; C22;

Declarations

- Conflict of interest/Competing interests: The authors have no competing interests to declare that are relevant to the content of this article.
- Funding: The authors did not receive support from any organization for the submitted work.

1. Introduction: mean reversion in ETI estimation

Let z denote reported taxable income, whose elasticity (ETI) measures the responsiveness of z to changes in tax rates. Let the marginal tax rate (MTR) be $1 - \tau$; ETI can then be defined as:

$$ETI \ = \ \frac{\partial log\ z}{\partial log\ (1-\tau)} = \frac{\partial z}{\partial(1-\tau)}\left(\frac{1-\tau}{z}\right)$$

This elasticity is intended to capture all behavioural responses to income taxation (e.g., labour supply, avoidance, evasion and income shifting) and, once estimated, may be interpreted as a sufficient statistic for evaluating both the total allocative effects of taxation and optimal income tax rates under certain assumptions. (Feldstein, 1995; Saez, 2001; Saez, Slemrod, & Giertz, 2012).
To estimate ETI in a panel-data setting, the baseline specification is generally written as:

$$ln\ z_{it} = \alpha + \beta[log\ log\ (1-\tau_{it})\ ] + X_{it}'\omega + \mu_i + \lambda_t + u_{it} \tag{1}$$

where z_{it} is the taxable income of individual i at time t; τ_{it} is the MTR for individual i at time t; X_{it} is the vector of controls (demographics, geographic controls, economic sectors, etc.); μ_i denotes individual fixed effects; λ_t denotes time fixed effects; u_{it} is an error term and β is the parameter of interest. Equation (1) corresponds to a two-way fixed-effects (TWFE) model, which can be estimated by ordinary least squares (OLS) (see Section 3.2 of Baltagi, 2021; Wooldridge, 2025).
At least two statistical problems, namely *income mean reversion* and *income-tax endogeneity*, may affect the OLS estimation of Equation (1). Mean reversion is a central identification challenge in empirical studies of taxable income: transitory income shocks, which are particularly prevalent at the top of the income distribution, may be present in certain periods and may then mechanically dissipate over time. Mean reversion therefore reflects the empirical tendency for individuals with unusually high or low income in one period to move closer to their long-run average income in subsequent periods, even in the absence of any tax-policy change. This may be due to transitory income shocks (e.g., bonuses, capital gains and business income), measurement error in reported income, or life-cycle income dynamics. However, short-run movements in income may appear to be correlated with tax-rate changes: when taxpayers are grouped by base-year income or marginal tax rates, and when those shocks dissipate, income falls or rises can be spuriously interpreted as the effects of concomitant tax changes. Thus, some declines or increases in reported income may reflect the reversal of timing behaviour or reversion towards permanent income, rather than tax effects. Without correcting for mean reversion, these declines may be misinterpreted as tax responses, even though no causal relationship exists between income adjustments and tax changes.
In this paper, we investigate a case in which taxable income is affected by mean reversion and in which the variance of the tax treatment, together with its covariance with shocks, explains the bias in OLS estimation. We show that OLS estimates of ETI are upward biased, meaning that the regression mechanically, but incorrectly, attributes to behavioural tax responses an income adjustment associated with mean reversion (i.e., income falling after an unusually high year). As an original contribution to the literature, we analyse what drives this bias. First, we prove that the bias increases when the variance of transitory shocks is large and temporal persistence is low (strong mean reversion). In the extreme case of zero persistence, the bias is the largest possible. Second, we show that, under the standard mean-reversion assumption, the bias is strictly positive and is lower when the variance of the tax reform is higher. Indeed, the bias increases when the threshold of the tax reform lies far in the right tail of the income distribution (strong selection) and treatment assignment is sharp. This means that mean reversion causes upward bias in ETI estimates when treatment is assigned on the basis of high realizations of base-year taxable income and reforms raise top marginal tax rates only (the reform has smaller variance when it is concentrated on a small subset of the taxpayer population). Finally, we show that the bias decreases with the treatment if and only if average selection (global correlation) dominates marginal selection (local responsiveness), implying diminishing marginal sorting or concavity in the covariance-variance relationship. A convex relationship would instead generate a condition under which the bias increases with the treatment.

2. Mean reversion affecting ETI

Mean reversion in declared income (i.e., the tendency of income to move back towards its long-run average after extreme realisations) biases ETI estimates obtained from observed changes in reported income around tax reforms. With mean reversion, those income movements may reflect statistical reversion of the income process, rather than behavioural responses to changes in tax rates. For example, consider a tax reform that raises, at time t, the top marginal rates applied to high-income individuals identified in year $t-1$. As is customary, the new tax rate is applied to taxable income in $t+1$. In period $t+1$, those taxpayers may experience a decrease in income because of negative transitory shocks. In year $t+1$, the taxable income they declare may therefore fall relative to income in both t and $t-1$, irrespective of the tax change. However, that fall may be falsely attributed to the tax reform occurring at time t and referring to the top subsample of taxpayers identified at $t-1$. If the change in declared income is interpreted as a response to a tax reform, we will overestimate the responsiveness of income to taxation and obtain an upward-biased ETI estimate.

To focus on mean reversion only, we assume that the identification strategy exploits genuinely exogenous variation in marginal tax rates generated by truly unexpected tax reforms. Even in that case, however, mean reversion in declared income may exist and lead to identification bias. As shown by Saez et al. (2012), taxable income may be decomposed each year as the sum of a permanent (long-run) component z_i^P and a zero-mean transitory shock u_{it}:

$$ln\, ln\, z_{it} = ln\, ln\, z_i^P + v_{it} \quad (2)$$

where $ln\, ln\, (z_i^P)$ is the long-run deterministic component of income, which depends directly on exogenous factors and fixed individual characteristics. Transitory income shocks v_{it} are assumed to follow a first-order autoregressive stationary stochastic process (i.e., AR(1)), thus implying mean reversion by definition, such that

$$v_{i,t+1} = \rho v_{it} + \epsilon_{it} \quad \text{with} \quad 0 \leq \rho < 1 \quad E[\varepsilon_{it}] = 0 \quad Var(\varepsilon_{it}) = \sigma^2 \quad Cov(\varepsilon_{it+1}, \varepsilon_{it}) = 0$$

The above conditions on the innovation sequence ε_{it} lead to the following results (see Section 3.1 of Shumway and Stoffer, 2025):

$$Var(v_{it}) = \frac{\sigma^2}{1-\rho^2} \quad Cov(v_{it}, v_{it-1}) = \sigma^2 \frac{\rho}{1-\rho^2}$$

As emphasised by Saez et al. (2012),, mean reversion is empirically important at the top of the income distribution, where capital income, bonuses and entrepreneurial income generate large temporary fluctuations. Saez, Slemrod, & Giertz (2012), provide a comprehensive survey of ETI estimation, highlight that mean reversion is a central identification challenge, and show how estimates are sensitive to income controls and sample definitions. Gruber and Saez (2002) use base-year income controls to mitigate mean-reversion bias. Giertz (2008) provides one of the clearest discussions of mean reversion in ETI. In particular, he shows that mean reversion is especially severe at the top of the income distribution and that even sophisticated panel estimators may fail to remove it. We follow a different approach, in which we incorporate an autoregressive structure in taxable income dynamics that explicitly accounts for mean reversion. This choice allows us not only to postulate the presence of mean reversion in general, but also to give mean reversion a proper functional form. Contrary to Saez, Slemrod & Giertz (2012, p. 18), that work under the hypothesis of time-independent random shocks, we allow log incomes to be correlated over time and use AR(1) as a simple and easily interpretable temporal structure: current income realisations are linearly related to previous occurrences because the transitory shock is correlated over time. The autoregressive approach was already taken by Weber (2014), whose contribution represents a fundamental reference point for understanding empirical identification failures and the inconsistency of standard ETI estimators in the presence of mean reversion. Weber's analysis is particularly important because it shows, in a rigorous empirical framework, that mean reversion is not merely a nuisance feature of taxable income data, but a structural source of inconsistency when income dynamics are not properly modelled. Specifically, she used an autoregressive model with p lags to characterise the temporal structure of the transitory income component and combined it with a new IV strategy to obtain consistent empirical estimates of tax effects. In this respect, her work provides one of the clearest and most influential demonstrations that the dynamic behaviour of taxable income must be explicitly incorporated into ETI estimation. Our paper is directly inspired by this insight, but differs from Weber's in both aim and emphasis. While Weber treats the autoregressive structure primarily as an econometric and empirical problem to be solved in order to recover consistent estimates, our purpose is interpretative. We use a parsimonious autoregressive framework to characterise the direction and magnitude of the mean-reversion bias and to show how it depends on the persistence of transitory shocks, the variance of the tax treatment, and the covariance between treatment assignment and windfall income gains. In the following sections, we use a two-period autoregressive model to obtain indications about the relationship between the variability of the tax treatment and the windfall shocks affecting income dynamics. We show that the OLS bias depends on the variability and temporal persistence of the shocks. Indeed, when transitory shocks are independent over time (Saez, Slemrod, & Giertz, S, 2012, p. 18), mean reversion is maximal and the bias is the largest possible.

3. A formal treatment of mean reversion in taxable income in a two-period case

Consider two periods $t = 0,1$ and call y_{it} the log of taxable income, that is, $y_{it} = ln\, ln\, (z_{it})$. Assume that y_{it} follows the AR(1) stationary process above, where $\mu_i = ln\, z_i^P$ is the permanent component and the idiosyncratic error term is v_{it}. The change in log income from period 0 to period 1 is given by

$$\Delta y_i = y_{i1} - y_{i0} = (v_{i1} - v_{i0})$$

Now suppose that, in period 1, a higher marginal tax rate is introduced for individuals with taxable income above a base-year threshold c. Let the tax treatment (e.g., an increase in the MTR for some taxpayers) be D and define it as

$$D_i = I\{y_{i0} \geq c\}$$

where $I(\cdot)$ is an indicator function and c is the level of taxable income above which treatment (i.e., higher tax rates) occurs. Let the log net-of-tax-rate change be

$$x_i = \Delta log\, (1 - \tau_i) = -\delta D_i$$

where $\delta > 0$, such that $x_i < 0$ for treated units after a tax increase. The true period-to-period behavioural response is

$$\Delta y_i = \beta x_i + (v_{i1} - v_{i0}) \tag{3}$$

Notice that the canonical ETI regression (Gruber & Saez, 2002), usually estimated by OLS, is given by

$$\Delta y_i = \beta x_i + \varepsilon_i \tag{4}$$

where ε_i is assumed to be a white-noise error term with zero mean, constant variance and uncorrelated components, (then, stationary and mean reverting) in line with the classical OLS assumptions. It follows that the OLS estimator of the coefficient is given by

$$\hat{\beta} = \frac{Cov(x_i, \Delta y_i)}{Var(x_i)}$$

However, in our autoregressive setting, the true error term is the first difference of an AR(1) process, with

$$Var(v_{i1} - v_{i0}) = \frac{2\sigma^2}{1+\rho} \quad \text{and} \quad Cov(v_{it}, v_{it-1}) = \sigma^2 \frac{\rho-1}{1+\rho}$$

It follows that the model in Equation (4) is mis-specified, with consequent estimation bias for $\hat{\beta}$. Indeed, substituting for the true model, we obtain

$$Cov(x_i, \Delta y_i) = Cov(x_i, \beta x_i + v_{i1} - v_{i0}) = \beta Var(x_i) + Cov(x_i, v_{i1} - v_{i0})$$

and thus

$$\hat{\beta} = \beta + \underbrace{\frac{Cov(x_i, \nu_{i1} - \nu_{i0})}{Var(x_i)}}_{Bias}$$

Since the variance is always positive, to investigate the sign of the bias we must examine the sign of the covariance term. Since $x_i = -\delta D_i$ then:

$$Cov(x_i, \nu_{i1} - \nu_{i0}) = -\delta\, Cov(D_i, \nu_{i1} - \nu_{i0})$$

Therefore

$$Bias = -\delta \frac{Cov(D_i, \nu_{i1} - \nu_{i0})}{Var(x_i)}$$

Since $Var(x_i) = \delta^2 Var(D_i)$, we get

$$\hat{\beta} = \beta + \underbrace{\frac{-Cov(D_i, \nu_{i1} - \nu_{i0})}{\delta Var(D_i)}}_{Bias > 0} \tag{5}$$

which is an expression for the bias whose sign depends on the sign of the new numerator $Cov(D_i, \nu_{i1} - \nu_{i0})$, that is,

$$Bias = -\frac{1}{\delta}\left[\frac{Cov(D_i, \nu_{i1} - \nu_{i0})}{Var(D_i)}\right] = -\left[(+)\frac{?}{+}\right] \tag{6}$$

We rewrite the tax treatment as a function of permanent income and the initial transitory shock, that is,

$$D_i = I\{\mu_i + \nu_{i0} > c\}$$

For positive initial shocks ν_{i0}, we have that

$$E[\nu_{i0} \mid D_i = 1] > 0$$

High-base-year-income individuals are selected partly because of positive transitory shocks. Now compute:

$$E[\nu_{i1} \mid D_i = 1] = E[E[\nu_{i1} \mid \nu_{i0}] \mid D_i = 1]$$

Since $E[\nu_{i0}] = \rho\nu_{i0}$, we get that

$$E[\nu_{i1} \mid D_i = 1] = \rho E[\nu_{i0} \mid D_i = 1]$$

Therefore

$$E[\nu_{i1} - \nu_{i0} \mid D_i = 1] = (\rho - 1)E[\nu_{i0} \mid D_i = 1]$$

Because we assumed that $[\nu_{i0} \mid D_i = 1] > 0$ and $\rho < 1$, we have that

$$E[\nu_{i1} - \nu_{i0} \mid D_i = 1] < 0$$

and thus

$$Cov(D_i, \nu_{i1} - \nu_{i0}) < 0$$

As a result, the bias is positive. Moreover, since the true elasticity is β and $\hat{\beta} = \beta + Bias$, we obtain that

$$\hat{\beta} > \beta$$

that is, if the tax reform consists of a tax increase levied on high earners, it follows that $x_i = \Delta log\ (1 - \tau_i) < 0$ for treated taxpayers. Recalling that $Cov(v_{i1}, v_{i0}) = \sigma^2 \frac{\rho}{1-\rho^2}$, the bias increases when the transitory variance σ^2 is large and the persistence coefficient ρ is low (strong mean reversion). Indeed, in the extreme case of $\rho = 0$, mean reversion is maximal and the bias is the largest possible.

4. A discussion of the variance of the treatment and the bias

The way in which the treatment affects the bias can also be appreciated through the following derivative:

$$\frac{\partial(\hat{\beta} - \beta)}{\partial D_i} = -\frac{\left[\frac{\partial Cov(D_i, v_{i1} - v_{i0})}{\partial D_i}\right] Var(D_i) - \left[Cov(D_i, v_{i1} - v_{i0}) \frac{\partial Var(D_i)}{\partial D_i}\right]}{\delta Var(D_i)^2}$$

Note that, since $\delta > 0$, the derivative is negative if the numerator is positive, that is, if

$$\left[\frac{Cov(D_i, v_{i1} - v_{i0})}{Var(D_i)}\right] > \frac{\left[\frac{\partial Cov(D_i, v_{i1} - v_{i0})}{\partial D_i}\right]}{\frac{\partial Var(D_i)}{\partial D_i}}$$

The above inequality shows that the bias diminishes with the treatment if the mean relationship between $Cov(D_i, v_{i1} - v_{i0})$ and $Var(D_i)$, which is the OLS estimator of how the treatment affects the shock, is greater than the ratio of the derivatives of the two quantities with respect to the treatment[1]. In other words, the bias is reduced by the treatment if the OLS estimator (average linear association) of the shock-treatment relationship is greater than the ratio of the two marginal variations with respect to D_i, which is a ratio of two marginal sensitivities. The latter indicates how fast the covariance grows relative to how fast the variance grows as D_i changes (a marginal local analogue of the regression coefficient). In other terms, if the estimated mean reaction of the transitory shock to the treatment between two time points is greater than the ratio of the two marginal variations, the treatment reduces the bias. The condition implies that, to obtain a reduction in the bias, the global association must exceed the local sensitivity. In practical terms, the overall correlation structure between D_i and $v_{i1} - v_{i0_i}$ must be stronger than what one would infer from infinitesimal or local changes in D_i. This suggests concavity (or diminishing returns) in the relationship between covariance and variance as D_i changes.

To make the interpretation easier, we may assume that if D_i is a treatment and $\Delta y_i = v_{i1} - v_{i0}$ is an individual-specific, unexpected windfall taxable-income gain, then the left-hand side of the inequality captures a form of selection on gains (similar to standard Heckman-style selection), whereas the right-hand side captures how that selection changes at the margin. Hence, the inequality states that taxpayers selected into treatment exhibit stronger selection on gains on average than marginal changes in treatment would suggest. This often arises when high-gain individuals (i.e., those who have enjoyed high windfall gains) are disproportionately represented in treated groups (sorting), as would be the case when the treatment consists of increasing the tax rates of high-income taxpayers. By contrast, expanding the treatment (e.g., lowering the threshold c) reaches lower-gain individuals (a lower selection threshold increases the sample of treated units) and reduces the bias. The inverse relationship in the inequality condition would imply that the bias increases with the treatment because the inequality encodes the following message: the bias decreases if and only if average selection (global correlation) dominates marginal selection (local responsiveness), implying diminishing marginal sorting or concavity in the covariance-variance relationship. The presence of a convex relationship would generate a condition under which the bias increases with the treatment.

---

[1] Notice that the inequality can be rewritten as $\left[\frac{Cov(D_i, v_{i1} - v_{i0})}{Var(D_i)}\right] > \frac{\partial Cov(D_i, v_{i1} - v_{i0})}{\partial Var(D_i)}$, that is, the bias reduces if the expected average variation of the random transitory shock as a function of the tax reform is greater than the marginal variation.

5. Conclusions

The paper provides an analytical interpretation of mean reversion bias in ETI estimation by linking its magnitude to shock persistence, treatment variance, and the covariance between treatment assignment and windfall income gains. We show that, when shocks are correlated over time and subject to autoregressive mean reversion, the OLS estimates of ETI are upward biased, meaning that the regression mechanically, but incorrectly, attributes to behavioural tax responses an income adjustment associated with mean reversion. The analysis clarifies the mechanisms behind this bias. First, the bias is larger when transitory shocks are more volatile and less persistent (strong mean reversion), reaching its maximum when persistence is zero. Second, the bias is strictly positive and is lower when the variance of the tax reform is higher. Indeed, the bias increases when the threshold of the tax reform lies far in the right tail of the income distribution (strong selection) and treatment assignment is sharp. This means that mean reversion causes upward bias in ETI estimates when treatment is assigned on the basis of high base-year taxable income and the reform raises top marginal tax rates only (the reform has smaller variance when it is concentrated on a small subset of the taxpayer population). Finally, we show that the bias decreases when the response of marginal treated taxpayers is weaker than that of average treated taxpayers. For example, when the marginal tax rate applying to the highest tax bracket increases (i.e., only the richest taxpayers are treated), the bias increases or decreases depending on whether the reaction of the average taxpayer within the bracket is higher or lower than the reaction of the richest taxpayer in the bracket.

References

[1] Baltagi, B. H. (2021). Econometric Analysis of Panel Data (6th ed.). Springer Cham. https://doi.org/10.1007/978-3-030-53953-5https://doi.org/https://doi.org/10.1007/978-3-030-53953-5

[2] Feldstein, M. (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), pp. 551–572.

[3] Giertz, S.H. (2008), Panel data techniques and the elasticity of taxable income. Economics Department Faculty Publications. 50. https://digitalcommons.unl.edu/econfacpub/50

[4] Gruber, J., & Saez, E. (2002). The elasticity of taxable income: evidence and implications. Journal of Public Economics, 84(1), 1-32. https://doi.org/https://doi.org/10.1016/S0047-2727(01)00085-8

[5] Saez, E. (2001), Using Elasticities to Derive Optimal Income Tax Rates. Review of Economic Studies, 68(1), pp. 205–229

[6] Saez, E., Slemrod, J., & Giertz, S. H. (2012). The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. Journal of Economic Literature, 50(1), 3–50. https://doi.org/10.1257/jel.50.1.3

[7] Shumway, R. H., & Stoffer, D. S. (2025). Time Series Analysis and Its Applications (5th ed.). Springer Cham. https://doi.org/10.1007/978-3-031-70584-7https://doi.org/10.1007/978-3-031-70584-7

[8] 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. https://doi.org/https://doi.org/10.1016/j.jpubeco.2014.05.004

[9] Wooldridge, J. M. (2025). Two-way fixed effects, the two-way Mundlak regression, and difference-in-differences estimators. Empirical Economics. https://doi.org/10.1007/s00181-025-02807-z https://doi.org/10.1007/s00181-025-02807 -z