

Volume 2

TAS RESEARCH AND RELATED STUDIES

Do Taxpayers Respond to the Substantial Understatement Penalty? Analysis of Bunching Below the Substantial Understatement Penalty Threshold

Do Taxpayers Respond to the Substantial Understatement Penalty? Analysis of Bunching Below the Substantial Understatement Penalty Threshold¹

EXECUTIVE SUMMARY

DISCUSSION

The Substantial Understatement Penalty Rules

Precedent for an Analysis of Bunching

Limitations

Methodology

Results

CONCLUSION

APPENDIX A

55

56

57

59

61

63

67

68

69

1

The principal authors of this study are Eric LoPresti, Senior Attorney Advisor, Jeff Wilson, Senior Research Advisor, Michael Nestor, Operation Research Analyst, and Kim Bloomquist, former Operation Research Analyst. For a prospectus for this research, see National Taxpayer Advocate 2017 Annual Report to Congress 196-208.

EXECUTIVE SUMMARY

The “economic deterrence” model of tax compliance suggests that higher or more certain penalties should produce more compliance.² This study aims to explore the extent to which taxpayers respond to the substantial understatement penalty.

An accuracy-related penalty applies to various understatements, including “substantial” understatements and those due to negligence.³ If the understatement exceeds the substantial understatement threshold, a penalty applies even if the IRS does not determine the taxpayer was negligent. For individuals, the threshold is generally the greater of \$5,000 or 10 percent of the tax required to be shown on the return.⁴ The substantial understatement penalty should increase the likelihood that substantial understatements will be subject to an accuracy-related penalty. If it does, the deterrence model suggests that it should deter taxpayers from understating their tax liabilities by more than the substantial understatement threshold.

If the substantial understatement penalty affects compliance behavior, some taxpayers whose understatements would otherwise be just over the threshold should adjust their reporting so that their understatements are just below it.⁵ If they do, we should see the density of understatements concentrated or “bunching” at or just below it, and fewer (*i.e.*, a crater) just above the threshold.⁶

To detect bunching at or below the threshold, we analyzed the distribution of individual examination assessments (*i.e.*, understatements) on returns selected at random as part of the National Research Program (NRP) for tax years (TY) 2006-2012 (excluding 2009).⁷ We reviewed histograms of the distribution of the understatements and applied statistical tests.

We did not detect statistically significant evidence of bunching at or just below the substantial understatement penalty threshold for taxpayers overall or for any taxpayer segment. In other words, we did not find evidence that taxpayers respond to the economic incentive provided by the substantial understatement penalty, as predicted by the economic deterrence model of tax compliance.

2 See, e.g., Maurice Allingham & Agnar Sandmo, *Income Tax Evasion: A Theoretical Analysis*, 1 J. Pub. Econ. 323-338 (1972).

3 Internal Revenue Code (IRC) § 6662.

4 IRC § 6662(d).

5 This threshold is analogous to a kink point or notch in the tax rate schedule. The point at which the marginal tax rate increases is called a kink point, whereas large jumps or stepped increases are generally called notches. For example, an income tax is notched if it requires a person to pay a higher average rate on all of his income when he or she reaches the next highest bracket, as is the case in Pakistan. See Henrik Kleven & Mazhar Waseem, *Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan*, 128 Qtrly J. Econ. 669, 670 (2013). Penalty thresholds are generally more similar to notches than to kink points. Notches may trigger a larger behavioral response than kink points, in part, because they are more salient. See James Sallee & Joel Slemrod, *Car Notches: Strategic Automaker Responses to Fuel Economy Policy* 3 (Nat'l Bureau Econ. Res. (NBER) Working Paper No. 16604, 2010), <http://www.nber.org/papers/w16604>.

6 Indeed, Congress recently lowered the 10 percent substantial understatement threshold to 5 percent for taxpayers claiming a deduction under IRC § 199A for qualified business income, presumably based on the assumption that doing so would reduce noncompliance. Pub. L. No. 115-97, § 11011, 131 Stat. 2054 (Dec. 22, 2017) (codified at IRC § 6662(d)(1)(C)). The 5 percent threshold did not apply to the years we studied.

7 We excluded 2009 because the IRS's data did not have weights for that year.

DISCUSSION

Under the deterrence model people pay taxes to avoid penalties.⁸ Tax compliance depends on the likelihood of being caught and the size of the penalty. This model is overly simplified.⁹

Insights from behavioral science (e.g., psychology and behavioral economics) suggest that other factors affect tax compliance. For example, people do what is easy, do what they think others are doing (i.e., follow social norms), and cheat only to the extent they can maintain a positive self-image (i.e., tax morale).¹⁰ These findings are consistent with other lines of research, which suggest that trust for the IRS, norms, fairness, reciprocity, tax morale, complexity, and similar factors drive compliance.¹¹

However, the deterrence model might suggest that a costless way for the government to increase tax compliance (and government revenue) is to increase the penalties for noncompliance.¹² Indeed, some lab

-
- 8 See, e.g., Maurice Allingham & Agnar Sandmo, *Income Tax Evasion: A Theoretical Analysis*, 1 J. Pub. Econ. 323-338 (1972).
- 9 Some have tried to tweak the model to solve the “compliance puzzle” of why, if deterrence is so important, we observe relatively high levels of compliance in the U.S. even though we have relatively few examinations and moderate penalty rates. See, e.g., Mark Phillips, *Reconsidering the Deterrence Paradigm of Tax Compliance*, IRS Research Conference (2011); Jack Manhire, *Toward A Perspective-Dependent Theory of Audit Probability for Tax Compliance*, 33 VA. TAX REV. 629 (2014). However, other factors appear to affect compliance. See, e.g., Nadja Dwenger et al., *Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany*, 8 AM. ECON. J. 203, 204-205 (2016) (finding that about 20 percent fully paid a church tax, even though they knew the tax was not enforced); Molly Freat, Jonathan Gruber & Benjamin Sommers, *Premium Subsidies, the Mandate, and Medicaid Expansion: Coverage Effects of the Affordable Care Act*, 5 (NBER Working Paper No. 22213, 2016), <http://www.nber.org/papers/w22213> (“the [Affordable Care Act coverage] mandate penalty had a negligible impact on coverage”). Similarly, the incidence of crime cannot be explained by the severity of the sanction, as predicted by the deterrence model. See, e.g., Anthony Doob & Cheryl Webster, *Sentence Severity and Crime: Accepting the Null Hypothesis*, 30 CRIME & JUST. 143 (2003).
- 10 See generally National Taxpayer Advocate 2016 Annual Report to Congress 50-63 (Most Serious Problem: *Voluntary Tax Compliance*); National Taxpayer Advocate 2016 Annual Report to Congress vol. 3 (Literature Review: *Behavioral Science Lessons for Taxpayer Compliance*); Richard Thaler, *Misbehaving: The Making of Behavioral Economics* (2015).
- 11 See, e.g., National Taxpayer Advocate 2007 Annual Report to Congress, vol. 2 138-150 (Marjorie E. Kornhauser, *Normative and Cognitive Aspects of Tax Compliance*); National Taxpayer Advocate 2012 Annual Report to Congress vol. 2 1-28; OECD, Forum on Tax Administration, Small/Medium Enterprise (SME) Compliance Subgroup, *Understanding and Influencing Taxpayers’ Compliance Behaviour* (Nov. 2010); OECD, Forum on Tax Administration Subgroup, *Right from the Start: Influencing the Compliance Environment for Small and Medium Enterprises* (Jan. 2012); Tom Tyler, *Why People Obey the Law* (2006); Tom Tyler, *Legitimacy and Criminal Justice: The Benefits of Self-Regulation*, 7 OHIO ST. J. CRIM. L. 307-359 (2009); Erich Kirchler, *The Economic Psychology of Tax Behaviour* (2007); Erich Kirchler & Erik Hoelzl, *Tax Behaviour in Economic Psychology*, ch. 16 (2017).
- 12 Under an extension of the deterrence model, some have suggested that tax agencies can maintain compliance when they reduce audit rates without increasing penalties by trusting taxpayers—rewarding compliant taxpayers with fewer audits. See Juan P. Mendoza & Jacco L. Wielhouwer, *Only the Carrot, Not the Stick: Incorporating Trust into the Enforcement of Regulation*, 10 PLOS ONE 1, 4 (2015), <https://www.ncbi.nlm.nih.gov/pubmed/25705898>. However, an increase in penalty rates makes the use of trust less feasible. *Id.* at 15.

experiments support this notion.¹³ However, there is relatively little evidence that marginal changes to penalty rates have a positive effect on tax compliance in the real world.¹⁴

This study examines the extent to which the substantial understatement penalty affects real-world tax reporting behavior by individuals. If some taxpayers whose understatements would otherwise be just over the substantial understatement threshold adjust their reporting so that their understatements are at or just below it, then we should see relatively more understatements bunching at or just below it, and fewer (*i.e.*, a crater) slightly above the threshold. This analysis is possible only because the substantial understatement penalty is triggered at a specific observable threshold.

The Substantial Understatement Penalty Rules

If a tax return is wrong and the taxpayer was negligent or disregarded a rule or regulation, the IRS may apply a 20-percent accuracy-related penalty to the underpayment.¹⁵ Even if the IRS cannot show that the taxpayer was negligent or disregarded a rule or regulation, it may apply a 20-percent accuracy-related penalty to any underpayment that is due to a “substantial understatement,”¹⁶ unless certain exceptions apply.¹⁷

- 13 See, e.g., Calvin Blackwell, *A Meta-analysis of Incentive Effects in Tax Compliance Experiments*, in *Developing Alternative Frameworks for Explaining Tax Compliance* 97, 109 (James Alm et al. eds., 2010); James Alm et al., *Estimating The Determinants Of Taxpayer Compliance With Experimental Data*, 45(1) NAT'L TAX J. 107, 110 (1992) (finding experiments generally show that the “response to an increase in the penalty rate is positive but small and not highly significant.”).
- 14 See, e.g., Ann D. Witte and Diane F. Woodbury, *The Effect of Tax Laws and Tax Administration on Tax Compliance: The Case of the U.S. Individual Income Tax*, 38 NAT'L TAX J. 1, 7-9 (1985) (analyzing IRS data from the Taxpayer Compliance Measurement Program (TCMP) and finding the probability of civil and criminal fraud penalties had no significant effect or a negative effect; and the severity of criminal sanctions had no significant effect, except for a small positive effect on high-income self-employed individuals); Joel Slemrod et al., *Cheating Ourselves: The Economics of Tax Evasion*, 21 J. ECON. PERSP. 25, 38 (2007) (“there has been no compelling empirical evidence addressing how noncompliance is affected by the penalty for detected evasion, as distinct from the probability that a given act of noncompliance will be subject to punishment.”); James Andreoni et al., *Tax Compliance*, 36 J. ECON. LIT. 818, 842 (1998) (finding only one real-world study (by Pommerehne and Frey) that suggested penalties may have a positive effect on compliance, but the effect was not statistically significant). See also Kimberly Varma & Anthony Doob, *Deterring Economic Crimes: The Case of Tax Evasion*, 40 CANADIAN J. CRIMINOLOGY 165, 175-176 (1998) (surveying Canadians and finding that “25.9 percent of those who thought that jail would be imposed for evasion... had evaded tax. In contrast, only 15.3% of those who thought nothing would happen had evaded tax.”). Even if raising penalties could increase compliance, there may be a point beyond which penalty increases decrease compliance, potentially due to a reduction in the perceived legitimacy of the penalties or in the agency's enforcement of them. See generally, Tom Tyler, *Why People Obey the Law* (2006) (discussing legitimacy).
- 15 IRC § 6662(a); Treas. Reg. § 1.6664-2. For this purpose, a refundable credit claim (e.g., the Earned Income Tax Credit (EITC)) can trigger an underpayment if it is paid. See Treas. Reg. §§ 1.6664-2(c), -(g) (Example 3). During tax year (TY) 2010-12, the IRS may have believed a refundable credit claim could trigger an underpayment, even if it was frozen and not paid. See, e.g., Program Manager Technical Advice (PMTA) 2010-01 (Nov. 20, 2009) and PMTA 2011-03 (Aug. 27, 2010). While it later revised this conclusion, taxpayers could not be sure if the IRS would freeze their claims or issue refunds. See PMTA 2012-16 (May 30, 2012). While there were subsequent developments in this area, they should not be relevant to the years under study. See, e.g., *Rand v. Comm'r*, 141 T.C. 376 (2013) (holding that refundable credit claim reduce the amount shown, but not below zero); Protecting Americans from Tax Hikes Act of 2015 (PATH Act), Pub. L. No. 114-113, Div. Q, § 209(d)(1), 129 Stat. 3040, 3084-85 (2015) (providing that effective for all returns filed after Dec. 18, 2015, and all returns filed on or before Dec. 18, 2015, for which the period of limitations specified in section 6501 had not expired as of that date, refundable credit claims reduce the amount shown and can reduce it below zero).
- 16 Treas. Reg. § 1.6662-4(a)(2)(ii); Treas. Reg. §§ 1.6662-4(b)(4), -4(b)(5). An understatement is “substantial” if it exceeds the greater of: (A) ten percent of the tax required to be shown on the return for the tax year, or (B) \$5,000. For corporations (other than an S-corporation or personal holding company), an understatement is “substantial” if it exceeds the lesser of: (A) ten percent of the tax required to be shown on the return for the tax year (or if greater, \$10,000), or (B) \$10,000,000. See IRC § 6662(d)(1).
- 17 If the IRS establishes that a taxpayer was both negligent and substantially understated the tax, the maximum accuracy-related penalty is capped at 20 percent of the understated tax. IRC § 6662(a).

Exceptions Do Not Eliminate the Incentive to Avoid the Threshold

The substantial understatement penalty does not apply if the taxpayer shows there is “substantial authority” for the tax treatment of the item.¹⁸ There is substantial authority for the tax treatment of an item only if the weight of the authorities supporting the treatment is substantial in relation to the weight of authorities supporting contrary treatment.¹⁹ Another way to avoid the penalty is to adequately disclose the position, typically on Form 8275, *Disclosure Statement*, Form 8275-R, *Regulation Disclosure Statement*, or on the return.²⁰ The adequate disclosure exception only applies if the taxpayer has a “reasonable basis” for the position and keeps adequate records.²¹ However, the substantial authority and adequate disclosure exceptions do not apply to understatements resulting from a “tax shelter.”²² Tax shelters are broadly defined to include any partnership, entity, investment plan, or arrangement having “a significant purpose” of tax avoidance (*i.e.*, potentially any type of tax planning).²³

Finally, a taxpayer may avoid an accuracy-related penalty (including the substantial understatement penalty), if he or she can show the error was made in good faith and due to “reasonable cause.”²⁴ However, this is a relatively narrow exception that is based on the facts and circumstances.²⁵

If any of these exceptions apply to an item (*i.e.*, there is substantial authority, adequate disclosure, or reasonable cause for the taxpayer’s treatment of an item), then it is treated as if it were properly shown on the return for purposes of computing the penalty.²⁶ Because of the relative ease with which the IRS can establish that a taxpayer made a substantial understatement (as compared to establishing negligence or disregard of a rule or regulation) and the significant uncertainty about whether a taxpayer will be able to show that an exception applies, taxpayers have an economic incentive to ensure that any understatement of tax does not exceed the substantial understatement threshold, even if there is some possibility that an exception might apply.

18 IRC § 6662(d)(2)(B)(i) (reduction for substantial authority).

19 Treas. Reg. § 1.6662-4(d)(3)(i); Treas. Reg. § 1.6662-4(d)(2) (describing the standard as “less stringent than the more likely than not standard ... but more stringent than the reasonable basis standard”).

20 IRC § 6662(d)(2); Treas. Reg. § 1.6662-3(b)(3) (defining reasonable basis); Treas. Reg. § 1.6662-3(c) and -4(f) (discussing the disclosure exception and Form 8275 or Form 8275-R); Rev. Proc. 2016-13, 2016-4 I.R.B. 290 (discussing alternative disclosure procedures for certain items).

21 Treas. Reg. § 1.6662-4(e)(2)(i) and (iii). The “reasonable basis” standard is “a relatively high standard of tax reporting, that is, significantly higher than not frivolous or not patently improper.” Treas. Reg. § 1.6662-3(b)(3) (explaining the reasonable basis “standard is not satisfied by a return position that is merely arguable or that is merely a colorable claim. If a return position is reasonably based on one or more of the authorities set forth in § 1.6662-4(d)(3)(iii) (taking into account the relevance and persuasiveness of the authorities, and subsequent developments), the return position will generally satisfy the reasonable basis standard even though it may not satisfy the substantial authority standard...”).

22 IRC § 6662(d)(2)(C) (reduction for substantial authority or adequate disclosure inapplicable to tax shelter items).

23 *Id.* For a discussion about problems with leaving “a significant purpose” undefined, see National Taxpayer Advocate 2008 Annual Report to Congress vol. 2, § 1 (*A Framework for Reforming the Penalty Regime*).

24 IRC § 6664(c) (reasonable cause exception).

25 Treas. Reg. § 1.6664-4(b)(1) (further explaining “[c]ircumstances that may indicate reasonable cause and good faith include an honest misunderstanding of fact or law that is reasonable in light of all of the facts and circumstances, including the experience, knowledge, and education of the taxpayer. An isolated computational or transcriptional error generally is not inconsistent with reasonable cause and good faith. Reliance on an information return or on the advice of a professional tax advisor or an appraiser does not necessarily demonstrate reasonable cause and good faith. Similarly, reasonable cause and good faith is not necessarily indicated by reliance on facts that, unknown to the taxpayer, are incorrect. Reliance on an information return, professional advice, or other facts, however, constitutes reasonable cause and good faith if, under all the circumstances, such reliance was reasonable and the taxpayer acted in good faith.”).

26 IRC § 6662(d)(2)(B) (understatement “reduced” by that “portion” for which there is substantial authority or adequate disclosure); IRC § 6664(c)(1) (reasonable cause exception for applicable “portion” of the understatement).

A Taxpayer Could Compute the Operative Threshold Before Filing

For individuals, an understatement is substantial if it exceeds the greater of \$5,000 or 10 percent of the tax required to be shown on the return.²⁷ For example, if the correct amount of tax is \$10,000 and an individual taxpayer reported \$6,000, then the penalty would not apply. Although the \$4,000 understatement is more than ten percent of the correct tax, it is less than the fixed \$5,000 threshold. Conversely, if the same individual reported a tax of \$4,000, the substantial understatement penalty would apply. The \$6,000 understatement is more than \$5,000, which is the greater of the two thresholds.

For relatively high income taxpayers who owe more than \$50,000 in tax, the operative threshold is 10 percent because 10 percent is more than \$5,000 ($10\% \times \$50,000 = \$5,000$). For the same reason, \$5,000 is the operative threshold for those who owe less than \$50,000 in tax (e.g., those with lower incomes or who are entitled to relatively large deductions or credits). Thus, because a taxpayer could calculate whether the understatement penalty might apply, the penalty lends itself to bunching analysis, as opposed to the more subjective penalty for negligence.

Precedent for an Analysis of Bunching

Because the substantial understatement penalty does not apply unless a person's understatement exceeds a threshold, the distribution of understatements around the threshold can reveal whether the penalty affects reporting behavior. Other studies have suggested that if we see a disproportionate number of taxpayers reporting income near a threshold, such as a notch or kink point in the tax rate schedule (called bunching), we can conclude that they are responding to the economic incentive created by the marginal rate.²⁸

To detect this type of bunching, researchers first divide the population into buckets or bins that represent fixed income ranges (e.g., \$1–\$500, \$501–\$1,000, etc.) and then create a histogram showing how many taxpayers fall into each bin. If there are a disproportionate number of taxpayers in the bins near the threshold, researchers may conclude that taxpayers are adjusting their behavior (e.g., earnings or reporting compliance) in response to it.

A number of studies have used bunching to analyze the behavioral response to the Earned Income Tax Credit (EITC) schedule. The EITC is a means-tested anti-poverty program that provides assistance

27 IRC § 6662(d)(1).

28 See, e.g., Emmanuel Saez, *Do Taxpayers Bunch at Kink Points?*, 2 AM. ECON. J. ECON. POL. 180, 182 n4. (2010) (finding evidence of bunching around the first marginal U.S. tax rate threshold); Raj Chetty et al., *Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records* (NBER Working Paper No. 15617, 2009), <http://www.nber.org/papers/w15617> (reviewing tax data from Denmark to find evidence of bunching at the top marginal rate thresholds); Henrik Kleven & Mazhar Waseem, *Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan*, 128 QTRLY J. ECON. 669, 672 (2013) (finding “bunching below every notch [in Pakistan’s income tax brackets] combined with missing mass (holes) above every notch...[and that these effects are] larger for self-employed individuals than for wage earners...”); Spencer Bastani & Håkan Selin, *Bunching and Non-Bunching at Kink Points of the Swedish Tax Schedule* (CESifo Working Paper No. 3865, 2012), <https://ssrn.com/abstract=2101038> (estimating the taxable income elasticity at a kink point in the Swedish tax schedule using the bunching method). For a technical discussion of a similar methodology called “regression discontinuity,” see, e.g., David S. Lee and Thomas Lemieux, *Regression Discontinuity Designs in Economics*, 48 J. ECON. LIT. 281-355 (2010) (describing methods for analyzing of the effects of treatments based on the insight that those immediately above and below a threshold that triggers a treatment can be compared as if they were selected at random if there is no reason to expect they are significantly different in other important respects); Justin McCrary, *Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test*, 142 J. ECONOMETRICS 698–714 (2008) (developing a statistical test to gauge whether the distribution of people above and below a threshold is random or subject to manipulation).

to the working poor.²⁹ It boosts the economic incentive to earn income from work within certain ranges. One study found a disproportionate number of self-employed taxpayers' income bunching near the first kink point in the EITC schedule.³⁰ Although the self-employed are generally able to adjust their earnings in response to incentives more easily than wage earners, the study concluded that tax evasion could best explain the results.³¹ Although EITC claimants are more likely to get tax preparation assistance from unregulated, unaffiliated preparers—the types of preparers who are most likely to make mistakes—these studies suggested that someone (perhaps a preparer) was responsive to the kinks in the rate schedule.³² A follow-up study suggested that low income taxpayers' incomes bunch around thresholds that maximize all of the refundable credits, rather than just the EITC.³³

Another study found income bunching near a notch applicable to the saver's credit.³⁴ It concluded that some taxpayers who were claiming the saver's credit manipulated their incomes to qualify.³⁵ Another found that automakers responded to gas guzzler taxes by producing a disproportionate number of cars with fuel economy just above the notches in the tax rate schedule that would minimize the tax.³⁶ A study from the U.K. found bunching by small businesses with turnover just below the threshold at which they would be required to register for the value added tax (VAT).³⁷

-
- 29 To claim EITC for 2012, a person's income had to be between \$13,980 and \$50,270, depending on the number of children and whether the taxpayer was married. See IRS, Pub. 596, *Earned Income Credit (EIC)* 1 (2012). Although the thresholds have increased, we are citing the ones for 2012 because that was the last year of data that we studied. For a discussion of refundable credits, including the EITC, and related compliance challenges, see, e.g., *Improper Payments in the Administration of Refundable Tax Credits: Hearing Before the H. Subcomm. on Oversight of the Comm. on Ways and Means*, 112th Cong. (2011) (statement of Nina E. Olson, National Taxpayer Advocate).
- 30 See Emmanuel Saez, *Do Taxpayers Bunch at Kink Points?*, 2 AM. ECON. J. ECON. POL. 180–212 (2010).
- 31 *Id.* See also, Elira Kuka, *EITC and the Self-Employed: Real or Reporting Effects?* 42 PUB. FIN. REV. 691–719 (2013). Subsequent research used analysis of bunching to conclude that the EITC has a significant impact on reported earnings in areas of the country where knowledge about the EITC schedule is more widespread and preparers are readily available. Compare Raj Chetty & Emmanuel Saez, *Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients* (NBER Working Paper No. 14836, 2009), <http://www.nber.org/papers/w14836.pdf> (finding that having preparers educate taxpayers about EITC incentives did not increase bunching around the EITC kink points), with Raj Chetty, John N. Friedman, & Emmanuel Saez, *Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings*, 103 AM. ECON. REV. 2683–2721 (2013), <http://dx.doi.org/10.1257/aer.103.7.2683> (finding geographic and social proximity to peers (and preparers) with knowledge of the EITC kink points affected bunching around EITC kink points).
- 32 See Kara Leibel, IRS, Pub. 5161, *Taxpayer Compliance and Sources of Error for the Earned Income Tax Credit Claimed on 2006-2008 Returns* 41 (Aug. 2014) (finding that most EITC claimants use a preparer, that unenrolled return preparers are the most common type of preparer used by EITC claimants, and that these are also the most error prone). See also National Taxpayer Advocate 2002 Annual Report to Congress 216-230 (finding higher rates of error among unenrolled preparers). In some cases, preparer fraud is an issue. See, e.g., National Taxpayer Advocate 2012 Annual Report to Congress 68-94. For further discussion of the effect of preparers on compliance, see, e.g., National Taxpayer Advocate 2007 Annual Report to Congress vol. 2 2-74 (Leslie Book, *Study of the Role of Preparers in Relation to Taxpayer Compliance with Internal Revenue Laws*).
- 33 Jacob Mortenson & Andrew Whitten, *How Sensitive Are Taxpayers to Marginal Tax Rates? Evidence from Income Bunching in the United States* 3 (Jan. 2016), <https://www.aeaweb.org/conference/2016/retrieve.php?pdfid=1486>. This study did not analyze the effect of preparers.
- 34 Shanthi Ramnath, *Taxpayers' Responses to Tax-based Incentives for Retirement Savings: Evidence from the Saver's Credit Notch*, 101 J. PUB. ECON. 77–93 (2013). The saver's credit is a tax credit for using certain retirement savings vehicles, which provides larger credits to those with smaller adjusted gross incomes. IRC § 25B.
- 35 This study suggested that taxpayers with paid preparers were more likely to claim the credit, but did not otherwise analyze the effects of preparers. Shanthi Ramnath, *Taxpayers' Responses to Tax-based Incentives for Retirement Savings: Evidence from the Saver's Credit Notch*, 101 J. PUB. ECON. 77, 82 (2013).
- 36 James Sallee & Joel Slemrod, *Car Notches: Strategic Automaker Responses to Fuel Economy Policy* (NBER Working Paper No. 16604, 2010), <http://www.nber.org/papers/w16604>.
- 37 Li Liu & Ben Lockwood, *VAT Notches* (Nat'l Tax Assoc. 108th Annual Conf. Proc., Apr. 23, 2015), <https://www.ntanet.org/wp-content/uploads/proceedings/2015/023-lockwood-vat-notches.pdf>.

Another study found a small amount of income bunching near the threshold in the tax rate schedule that subjects Social Security benefits to tax, but only for self-employed individuals.³⁸ It concluded that the rules are so complex that most people (other than the self-employed) do not recognize this threshold. Although the Alternative Minimum Tax (AMT) is complicated, another study found bunching near the threshold at which people are subject to the AMT.³⁹ It observed that the behavioral response was largest for the self-employed.⁴⁰ It attributed the bunching to changes in both real economic activity and misreporting.⁴¹

Limitations

The bunching methodology has broad applicability. However, it is more challenging to analyze the distribution of understatements than it is to analyze the distribution of income. Income is reported by taxpayers on their returns, whereas the IRS can only detect understatements by examining a person's return. Moreover, the IRS only examines a subset of returns, making it more difficult to estimate the distribution of any understatements in the population as a whole or to determine whether taxpayers' understatements bunch at or below the substantial understatement threshold.

An Audit May Not Determine the Correct Tax Liability

One limitation is that an examination may not accurately reveal the taxpayer's understatement. An examiner may assert an understatement that is too high due to a misunderstanding of the facts or the rules. Alternatively, an examiner may not detect all of the understatements on a return, for example, because he or she is focused on a single issue.⁴²

To minimize the problem of inaccurate understatements, this study analyzed understatements detected as part of the National Research Program (NRP). NRP examinations are more likely to identify the true tax liability than other types of examinations.⁴³ However, even NRP exams do not identify all underreporting (or overreporting).⁴⁴

38 Leonard Burman *et al.*, *Older Taxpayers' Response to Taxation of Social Security Benefits*, IRS-Tax Policy Center Research Conference (June 20, 2013), <https://www.irs.gov/pub/irs-soi/14rptaxationofsocialsecuritybenefits.pdf>. Social Security benefits are only partially subject to tax. Over certain income ranges, taxpayers must include in their taxable income \$0.50-\$0.85 of benefit for every additional dollar of other taxable income. IRC § 86. These rules create kink points in the marginal rate schedule.

39 See Donald Bruce & Xiaowen Liu, *Tax Evasion and Self-Employment in the US: A Look at the Alternative Minimum Tax*, IRS Research Conference 165-179 (2015), <https://www.irs.gov/pub/irs-soi/14rescontaxevasion.pdf>. The alternative minimum tax (AMT) required taxpayers to compute their taxes twice—once under the regular tax rules and again under the AMT rules. If the “tentative minimum tax” liability exceeded the regular tax liability, the taxpayer paid the difference as AMT. See IRC § 55 (2017). The AMT was repealed for tax year 2018. See Pub. L. No. 115-97, 131 Stat. 2054 (Dec. 22, 2017).

40 *Id.*

41 *Id.*

42 Dispute resolution procedures may help to ensure that proposed assessments are accurate. However, the IRS may agree to settle for amounts that are less than the true tax liability based on the hazards of litigation, and taxpayers may agree to excessively high settlements to avoid the cost and risks of litigation, including reputational risks.

43 Compare Internal Revenue Manual (IRM) 4.10.2.7.1.1 (Feb. 11, 2016) (“The goal of an examination is to determine the ‘substantially correct’ tax liability”) and IRM 4.10.2.3.1 (Feb. 11, 2016) (noting that Office and Field “Examiners are expected to examine all large, unusual and questionable items (LUQ). However, it is not intended that examiners should consider every possible issue”) with IRM 4.22.1.3(4) (Sept. 6, 2017) (National Research Program (NRP) “examiners will verify information on the sampled returns and capture all adjustments, no matter how small or whether the adjustments favor the IRS or the taxpayers.”).

44 See, e.g., Mark J. Mazur & Alan H. Plumley, *Understanding the Tax Gap*, 60 NAT'L TAX J. 569, 573 (2007) (noting that because NRP audits do not detect all underreporting, the IRS applies a multiplier to the underreporting to estimate the total amount that should have been reported). Of course, IRS auditors may also fail to detect overreporting.

Some Bunching Could be Due to the Structure of the Tax Code, Rather than a Behavioral Response

Another limitation is that we would expect the structure of the tax code to cause understatements to bunch at specific dollar thresholds. This type of bunching does not necessarily reflect a behavioral response to a penalty. For example, we might expect the disallowance of a first-time homebuyer tax credit of up to \$6,500 to generate a disproportionate number of \$6,500 understatements, especially if an auditor is faced with a binary choice to either allow or disallow the credit.⁴⁵ Similarly, we might expect the disallowance of a dependent to cause a disproportionate number of understatements of specific amounts (e.g., child tax credits or earned income tax credits). By contrast, omitted income or overstated deductions can result in understatements of any amount, depending on the facts and circumstances.

In other words, we could observe understatements bunching in various ranges simply because of the structure of the tax code, even if taxpayers are oblivious to the penalty or penalty threshold. However, bunching due to the structure of the tax code should not interfere with our analysis of bunching below the substantial understatement threshold, unless the structural bunching occurs at or immediately below the substantial understatement threshold.

Bunching Could Reflect a Behavioral Response by Examiners, Rather than Taxpayers

If we observe understatements bunching below the threshold, it might not always reflect a behavioral response by the taxpayer. For example, when an understatement is otherwise close to the threshold, it is possible that some examiners might try to avoid proposing assessments that would trigger the substantial understatement penalty, perhaps to avoid controversy and extra paperwork.⁴⁶ Others might try to find understatements large enough to trigger the penalty, especially if the taxpayer seems negligent or uncooperative.⁴⁷ In other words, if understatements are bunching at a particular range, the bunching could reflect a behavioral response to the substantial understatement penalty threshold by IRS employees, rather than by taxpayers.

The Density or Distribution of the Understatements Detected by the IRS May Be Biased by the IRS's Examination Priorities

Another limitation is that if we only analyze the understatements on returns that have been selected for audit, the distribution of the understatements we detect could be due to the IRS's audit selection methods. It may not reflect the distribution of understatements in the population.⁴⁸ While some returns are selected for audit at random as part of the NRP,⁴⁹ others are selected because the IRS believes

45 The first-time homebuyer credit is a tax credit for the purchase of a new home after Apr. 8, 2008, and before May 1, 2010, with certain exceptions. See IRC § 36; IRS, *First-Time Homebuyer Credit Questions and Answers: Basic Information*, <https://www.irs.gov/newsroom/first-time-homebuyer-credit-questions-and-answers-basic-information> (revised Jan. 27, 2010). The maximum credit was generally \$7,500 or \$8,000, but beginning Nov. 7, 2009, a credit of up to \$6,500 was available to a new category of homebuyers. *Id.*

46 See IRM 20.1.5.1.4 (Dec. 13, 2016) (requiring exam employees to obtain managerial approval to either assert or not assert the substantial understatement penalty where there is a substantial understatement of tax).

47 See IRM 20.1.5.1.1 (Dec. 13, 2016) ("An adjustment warranting a penalty may meet the threshold of a substantial understatement and also be attributable to the taxpayer's clearly negligent actions. The substantial understatement penalty under IRC § 6662(b)(2) should be developed as the primary position and the negligence penalty under IRC § 6662(b)(1) should be developed as the alternative position.").

48 John L. Czajka & Sonya Vartivarian, MPR Ref. No.: 6284-001, *An Evaluation of the Sample Design for the National Research Program Individual Income Tax Reporting Compliance Study* 4, n.1 (Mathematica, Policy Research, Inc., Apr. 30, 2007) ("The returns selected for operational audits do not comprise a probability sample of the full population of filed returns, and the audits themselves differ from research audits in their generally more narrow focus on reporting issues identified and flagged in the screening process [than NRP audits].").

49 IRM 4.10.2.7.1 (Feb. 11, 2016).

they contain significant understatements or because it wants to maintain a particular level of audit coverage for a given taxpayer segment.⁵⁰ To avoid the bias that might result from analyzing a non-random sample, this study analyzes the results of individual NRP examinations. We used NRP data for tax years (TYs) 2006-2012, excluding data for which no weights were available.⁵¹ However, working with NRP data presented another set of challenges.

Methodology

Bootstrapping Procedure

Instead of a simple random sample, the NRP uses a complex stratified random sample (*i.e.*, selects a minimum number of returns from various taxpayer segments).⁵² Stratified random samples must be weighted before the results can be projected to the population. The NRP weights reflect the relative frequency with which the IRS expects the audited returns to occur in the population. Because the statistical tests that we used to detect bunching could not use stratified weighted data, we expanded the data so that it would not require weighting. Then we used a “bootstrapping” method to improve the reliability of our statistical test results.

First, we duplicated each observation so that the chance of selecting it at random from the overall NRP dataset would match the likelihood of selecting it from the overall population. For example, observations from a segment of the population with a weight of five (*e.g.*, because they were relatively underrepresented in the NRP sample) would each appear five times.

Next, we drew 1,000 random samples of 10,000 observations (with replacement) from the expanded NRP dataset. For each of these 1,000 “bootstrap” samples, we estimated the probability that the understatements bunched at or just below the threshold. The results of these repeated estimates allowed us to come up with an average estimate of the results of these tests.

Separate Analysis of Various Taxpayer Segments

Because our literature review identified a number of taxpayer segments that were better able to adjust their income in response to monetary incentives, TAS analyzed these segments separately to determine if some are more responsive to the penalty threshold than others. Specifically, we analyzed segments claiming the EITC, sole proprietors, those who used a preparer, and those with income of at least \$100,000 for the following reasons:

- Studies (discussed above) suggest that self-employed taxpayers have a greater propensity than other taxpayers to adjust their tax reporting behavior in response to economic incentives such as thresholds applicable to marginal tax rates, the EITC, the taxation of Social Security benefits, and

50 For a general discussion of audit selection methods, see IRM 4.1.3.1 (Aug. 10, 2012); Government Accountability Office (GAO), GAO-16-103, *Certain Internal Controls for Audits in the Small Business and Self-Employed Division Should Be Strengthened* (2015). The extent to which the IRS can audit various types of taxpayers and issues may also depend on the staffing, training, and skill level of its workforce.

51 We analyzed the latest tax years for which the NRP weights were available. Weights were not available for TY 2009.

52 John L. Czajka & Sonya Vartivarian, MPR Ref. No.: 6284-001, *An Evaluation of the Sample Design for the National Research Program Individual Income Tax Reporting Compliance Study* (Mathematica, Policy Research, Inc., Apr. 30, 2007).

for triggering the AMT.⁵³ For this reason, they could also be more responsive to the economic incentives provided by the substantial understatement penalty threshold than other taxpayers.

- Studies suggest that some taxpayers adjust their behavior to claim the maximum EITC.⁵⁴ If EITC claimants (or their preparers) change their behavior in response to the economic incentive provided by the EITC thresholds, they could be similarly responsive to the substantial understatement penalty threshold.⁵⁵
- Preparers can have a wide range of effects on compliance, depending on the circumstances.⁵⁶ Preparers might educate taxpayers about the substantial understatement penalty threshold. Some taxpayers may use a preparer to avoid making an understatement, perhaps because the law is so complex they are not sure they can avoid one on their own. Others may use a preparer

- 53 See Emmanuel Saez, *Do Taxpayers Bunch at Kink Points?*, 2 AM. ECON. J. ECON. POL. 180 (2010) (“We find clear evidence of bunching around the first kink point of the Earned Income Tax Credit but concentrated solely among the self-employed.”); Elira Kuka, *EITC and the Self-Employed: Real or Reporting Effects?*, 42 PUB. FIN. REV. 691–719 (2013) (finding that real labor supply responses of the self-employed are similar to those of salaried workers, but that they exhibit greater bunching due to misreporting); Leonard Burman et al., *Older Taxpayers’ Response to Taxation of Social Security Benefits*, IRS-Tax Policy Center Research Conference 3 (June 20, 2013), <https://www.irs.gov/pub/irs-soi/14rptaxationofsocialsecuritybenefits.pdf> (“We find no evidence of bunching at or around the thresholds [at which Social Security benefits become taxable] for the population as a whole, and only a very small response for single self-employed taxpayers who have previously been found to be more sensitive to changes in tax rates...”); Donald Bruce & Xiaowen Liu, *Tax Evasion and Self-Employment in the US: A Look at the Alternative Minimum Tax*, IRS Research Conference 165, 165 (2015), <https://www.irs.gov/pub/irs-soi/14rescontaxevasion.pdf> (“We find the bunching created by self-employed individuals locates further away from the AMT threshold than the bunching created by wage earners, which suggests that the self-employed act more aggressively to avoid the AMT.”); Henrik Kleven & Mazhar Waseem, *Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan*, 128 QTRLY J. ECON. 669, 672 (2013), http://eml.berkeley.edu/~saez/course/kleven-waseem_qje2013.pdf (finding “bunching below every notch [in Pakistan’s income tax brackets] combined with missing mass (holes) above every notch...[and that these effects are] larger for self-employed individuals than for wage earners...”).
- 54 See, e.g., Raj Chetty, John N. Friedman, & Emmanuel Saez, *Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings*, 103 AM. ECON. REV. 2683–2721 (2013), <http://dx.doi.org/10.1257/aer.103.7.2683>.
- 55 The impact of the preparer on compliance probably depends on the issue and also on the combination of both the taxpayer and the preparer’s views toward compliance. See National Taxpayer Advocate 2010 Annual Report to Congress vol. 2, 72, 80 n.37 (Study: *Researching the Causes of Noncompliance: An Overview of Upcoming Studies*) (“Assume there are three types of preparers and taxpayers: (1) those who want to comply with the letter and spirit of the law, (2) those who are willing to be more aggressive, particularly in areas where the law is unclear, and (3) those who are willing to cheat. Type one preparers may increase compliance by type two and type three taxpayers. Alternatively, those taxpayers may seek out type two or type three preparers. However, type two and type three preparers may reduce compliance by type one taxpayers unless those taxpayers either seek out type one preparers or are particularly resistant to the preparer’s suggestions for tax savings. Similarly, type three taxpayers may pressure type one or type two preparers to be more aggressive than usual.”).
- 56 For a general discussion of the influence of preparers on sole proprietors and EITC claimants, see, e.g., National Taxpayer Advocate 2007 Annual Report to Congress vol. 2, 2-74 (Leslie Book, *Study of the Role of Preparers in Relation to Taxpayer Compliance with Internal Revenue Laws*). Some research suggests preparers enhance compliance with unambiguous rules, but reduce it with respect to ambiguous ones. See Steven Klepper, Mark Mazur, & Daniel Nagin, *Expert Intermediaries and Legal Compliance: The Case of Tax Preparers*, 34 J. L. AND ECON. 205 (1991). See also Kim M. Bloomquist, Michael F. Albert, & Ronald L. Edgerton, *Evaluating Preparation Accuracy of Tax Practitioners: A Bootstrap Approach*, Proceedings of the 2007 IRS Research Conference 77 (2007) (finding preparers reduce math errors, but increase the incidence of potential misreporting). Other research suggests preparers make frequent errors in a wide variety of areas. See, e.g., GAO-02-509, *Tax Deductions: Further Estimates of Taxpayers Who May Have Overpaid Federal Taxes by Not Itemizing* (2002) (finding in 1998 about two million taxpayers overpaid by failing to itemize even though about half used a preparer); Treasury Inspector General for Tax Administration (TIGTA), Ref. No. 2003-40-046, *Analysis of Statistical Information for Returns with Potentially Unclaimed Additional Child Tax Credit* (Jan. 31, 2003) (finding about 230,000 returns filed by paid preparers in 2002 where taxpayers appeared eligible for Additional Child Tax Credits they did not claim); Janet Holtzblatt & Janet McCubbin, *Issues Affecting Low-Income Filers*, in *The Crisis in Tax Administration* 148, 159 (Henry J. Aaron and Joel Slemrod eds., 2004) (observing that about two-thirds of EITC returns, which have high levels of noncompliance, were prepared by paid preparers); GAO, GAO-06-563T, *Paid Tax Return Preparers: In a Limited Study, Chain Prepares Made Serious Errors* 5, 23 (Apr. 4, 2006) (finding preparers made significant mistakes on 17 of the 19 returns prepared for GAO employees posing as taxpayers, including the omission of income on ten); TIGTA, Ref. No. 2008-40-171, *Most Tax Returns Prepared by a Limited Sample of Unenrolled Preparers Contained Significant Errors* 2 (Sept. 3, 2008) (finding preparers made mistakes on 17 of the 28 returns prepared for TIGTA employees posing as taxpayers, including six willful or reckless errors).

to minimize their taxes and to ensure that any understatement is not so severe as to trigger a penalty.⁵⁷

- Taxpayers with at least \$100,000 in income might be more likely (than lower income taxpayers) to be engaged in economic activities that allow for a wider range of reasonable reporting positions.⁵⁸ Some of these taxpayers might try to ensure that debatable positions would not push them over the substantial understatement threshold.

Analysis of the Gap Between the Understatement and the Threshold

To analyze the distribution of the understatements, we computed the gap between the assessed understatement and the substantial understatement penalty threshold using a methodology similar to the one employed by the authors of the AMT study (discussed above).⁵⁹ The AMT threshold is different for each person depending on what items are on the return, making it difficult to select an income threshold to study. The AMT study addressed this problem by analyzing the gap between each individual's AMT and regular tax to determine how close each person was to the threshold.⁶⁰ Because the substantial understatement threshold depends on how much tax is required to be shown on a person's return (*i.e.*, an amount that varies from person to person, just like the AMT), this study uses a similar gap analysis to determine how close taxpayers are to the substantial understatement threshold that would apply to their returns.

Specifically, TAS first computed the substantial understatement threshold applicable to each return (*i.e.*, the greater of \$5,000 and 10 percent of the tax required to be shown). Next, we subtracted the threshold from the understatement (including disallowed EITC) to compute the gap between the understatement and the threshold for each return. Positive gaps represent understatements above the threshold, and negative gaps represent understatements below the threshold.

As an example, if the amount of tax required to be shown is \$60,000 and the amount actually shown is \$52,000, the understatement is \$8,000 (\$60,000 minus \$52,000) and the threshold is \$6,000 (the greater of \$5,000 and 10 percent of \$60,000). The gap between the understatement and the threshold is \$2,000 (\$8,000 minus \$6,000). If, instead, the amount shown were \$56,000, then the understatement would be \$4,000 (\$60,000 minus \$56,000), the threshold would be \$6,000 (the greater of \$5,000 and 10 percent of \$60,000), and the gap would be -\$2,000 (\$4,000 minus \$6,000). Because our focus was the substantial understatement threshold (*i.e.*, a gap of \$0), we did not include outliers—understatement gaps of less than -\$4,000 or more than \$10,000 (*i.e.*, the “tails” of the distribution). If we had expanded our analysis to include gaps below -\$4,000, the significant bunching in this area would have made it difficult to detect any bunching closer to \$0—both visually and statistically.⁶¹ Although the density

57 For further discussion of various types of influences, see, e.g., National Taxpayer Advocate 2007 Annual Report to Congress vol. 2, 59-63 (Leslie Book, *Study of the Role of Preparers in Relation to Taxpayer Compliance with Internal Revenue Laws*).

58 Perhaps for this reason, high income taxpayers may even take more aggressive positions after being told they are likely to be audited. See Joel Slemrod et al., *Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Field Experiment in Minnesota*, 79 J. PUB. ECON. 455, 455 (2000) (finding that the reported tax liability of the high income taxpayers fell sharply relative to the control group, after being informed the returns they were about to file would be “closely examined”).

59 See Donald Bruce and Xiaowen Liu, *Tax Evasion and Self-Employment in the US: A Look at the Alternative Minimum Tax*, IRS Research Conference 165-179 (2015), <https://www.irs.gov/pub/irs-soi/14rescontaxevasion.pdf>.

60 *Id.*

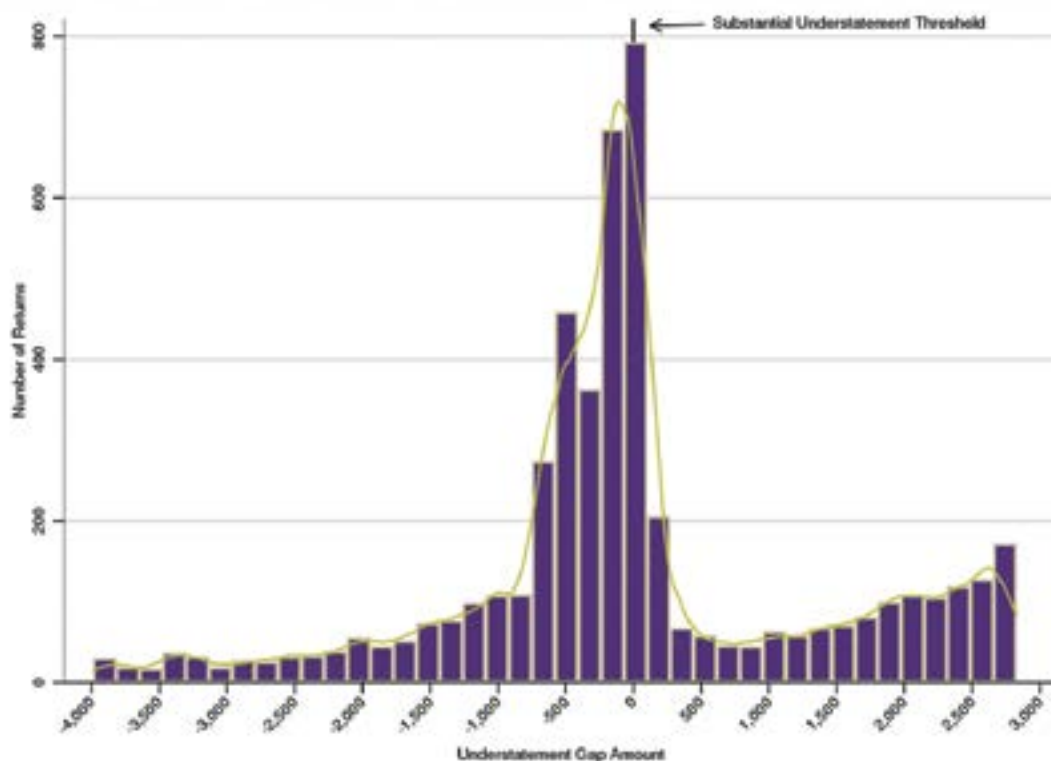
61 For taxpayers owing tax of less than \$50,000, gaps of less than -\$4,000 represented understatements of between \$0 and \$1,000. For about 84.2 percent of the returns in our sample, the understatement was in this range (including 48.9 percent that were not changed). As noted above, for purposes of this analysis, the understatement included the amount of any disallowed EITC claims.

of understatements of more than \$10,000 above the substantial understatement threshold trailed off gradually, we cut off this tail for the same reason—to make it easier to see any bunching near \$0.

An Analysis of Histograms to Detect Bunching

To see the distribution of the understatement gaps, the gap amounts can be sorted into bins and plotted on a histogram.⁶² If taxpayers are manipulating their understatements so that they are at or below the threshold (with a few overshooting by a little), we would expect to see bunching on a histogram that looks something like Figure 3.1.

FIGURE 3.1, Hypothetical Example of Bunching at or Below the Substantial Understatement Threshold



The appearance of the histogram depends to some extent on bin width selection.⁶³ Large bins look more choppy than small ones. However, kernel smoothing (the yellow-green line) can help to give a better representation of the density. In Figure 3.1, the number of returns clearly bunches near the substantial

62 See Justin McCrary, *Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test*, 142 J. ECONOMETRICS 698–714 (2008). We used the R software, executing package “rdd,” as documented by Drew Dimmery, *Regression Discontinuity Estimation* (Mar. 14, 2016), <https://cran.r-project.org/web/packages/rdd/rdd.pdf>.

63 Some researchers suggest choosing a bin size based on the amount of variation in the data, using larger sizes for groups with more variation. For example, one recommendation is to compute the bin sizes as $2 \cdot SD \cdot n^{-1/5}$, where SD is the standard deviation of the understatement gap for the group and n is the number of returns in the group. See Justin McCrary, *Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test*, 142 J. ECONOMETRICS 698–714 (2008). We chose the bin sizes for purposes of the histograms, but allowed the software we were using to choose them for purposes of the statistical tests, as discussed below.

understatement threshold (*i.e.*, near an understatement gap of \$0). If there is no bunching at or below the threshold then we would expect the kernel density line to go nearly straight through the threshold, or to show seemingly random hills and valleys around the threshold, perhaps corresponding to the structure of the tax code.

An Analysis of Statistical Tests to Detect Bunching

A visual analysis of a histogram is sometimes sufficient to detect bunching or the lack thereof. However, TAS confirmed the results reflected on the histograms with a statistical test. Justin McCrary developed one of the first statistical tests for the “manipulation” of a “running variable.”⁶⁴ In this case, the understatement gap is the running variable. The McCrary test first creates a histogram where no one bin contains points both to the left and to the right of the threshold point. Then it uses local linear regression—trend lines on each side of the threshold—to provide an estimate of the density and slope of the running variable (*i.e.*, the understatement gap) on each side. It also computes the bandwidth to use for these regressions (*i.e.*, how long the regression lines should be) using the method described by Guido and Kalyanaraman.⁶⁵ However, the McCrary test requires researchers to select a bin size and the amount of data to analyze on each side of the threshold (*i.e.*, “bandwidth”) to determine whether any bunching that appears on the histogram is likely to have occurred by chance (*i.e.*, is statistically significant). These choices introduce an element of subjectivity into the analysis. Accordingly, we used a similar test that was subsequently developed by Cattaneo, Jansson, and Ma, which does not require researchers to select a starting bin size or bandwidth.⁶⁶

Results

TAS reviewed the histograms of the gap between taxpayers’ assessed understatement and the applicable substantial understatement threshold to see if understatements bunched in the bin immediately below the substantial understatement penalty threshold. Our significance tests generally validated our visual observations.

Overall, understatements by taxpayers selected at random and examined as part of the NRP program for TYs 2006-2012 (excluding 2009 because the data did not include weights) did not bunch immediately below the substantial understatement threshold, as shown on Figure A.1 (in the Appendix). Moreover, we did not detect any significant evidence of bunching at or below the substantial understatement threshold for any taxpayer segment that we analyzed.

However, we see some visual evidence of what could be interpreted as bunching among those who claimed the EITC, particularly if they used a preparer, as shown on Figures A.4 and A.5 (in the Appendix), though it is not significant at a 95 percent level of confidence. It is possible that the structure of the EITC resulted in some bunching that happened to coincide with the threshold in these

64 See, e.g., Justin McCrary, *Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test*, 142 J. ECONOMETRICS 698–714 (2008).

65 *Id.* See Imbens, Guido & Karthik Kalyanaraman, *Optimal Bandwidth Choice for the Regression Discontinuity Estimator* (NBER Working Paper No. 14726, 2009). Finally, the test measures whether the differences in the density of the substantial understatement gaps on the left and right-hand sides of the threshold are statistically significant.

66 See Matias D. Cattaneo, Michael Jansson, & Xinwei Ma, *Simple Local Polynomial Density Estimators*, (Sept. 27, 2017), https://eml.berkeley.edu/~mjansson/Papers/CattaneoJanssonMa_LocPolDensity.pdf. We applied these tests by using commands that have been incorporated into the “R” and “Stata” software (called “rddensity” and “rdbwdensity”). These packages are available in R from the CRAN-R Project and in Stata from Google site RD Packages. We used these commands in R to test for manipulation of the understatement gap amount at zero, as described below. If the density is discontinuous at the threshold (*i.e.*, there is bunching at or below the threshold, which suggests manipulation of the running variable) then the test statistic will be less than 0.05 if the discontinuity is significant at the 95 percent confidence level.

two graphs. Indeed, we see several bunches on each of the histograms of understatements by taxpayers claiming the EITC (*i.e.*, Figures A.4, A.5, and A.6), which correlate roughly with the maximum amount of EITC that could be disallowed to taxpayers with various numbers of qualifying children. Figure 3.2 shows the maximum amount of EITC that could be claimed for the years we studied based on the number of qualifying children (QC).

FIGURE 3.2, Maximum EITC by Tax Year and Number of Qualifying Children (QC)⁶⁷

Tax Year	3 + QC	2 QC	1 QC	0 QC
2012	\$5,891	\$5,236	\$3,169	\$475
2011	\$5,751	\$5,112	\$3,094	\$464
2010	\$5,666	\$5,036	\$3,050	\$457
2008	\$4,824	\$4,824	\$2,917	\$438
2007	\$4,716	\$4,716	\$2,853	\$428
2006	\$4,536	\$4,536	\$2,747	\$414

For several of the tax years we reviewed, the maximum EITC was just below \$5,000—the operative substantial understatement threshold. Any taxpayer who claimed the maximum EITC for the years in question and whose EITC was denied would be faced with an understatement of a similar amount. Therefore, we cannot be sure that what may appear to be slight bunching under the substantial understatement threshold for EITC claimants is actually a behavioral response to the threshold.

CONCLUSION

Although it may be convenient for economists and policymakers to assume that people are motivated primarily by monetary penalties, there is relatively little real-world evidence that marginal changes to accuracy-related penalty rates affect tax reporting compliance. However, if people know about the substantial understatement penalty and try to avoid it, as assumed by the economic deterrence model of tax compliance, then some who might otherwise have an understatement just above the threshold should reduce their underreporting so that they are at or just below the threshold.

A wide range of studies have found that if people respond to an economic incentive that includes a threshold, we should observe bunching at or near that threshold (*e.g.*, the EITC, the AMT, a gas guzzler tax, and a VAT). This study reveals no such bunching below the substantial understatement threshold—except for slight bunching that is not statistically significant and that is most likely due to the structure of the EITC—rather than a behavioral response to the penalty. Therefore, this study does not support the hypothesis that we can improve reporting compliance by increasing penalties or reducing the substantial understatement threshold. One explanation could be that other (nonmonetary) factors drive tax compliance decisions for most taxpayers. Another could be that the IRS cannot reliably detect understatements even when using NRP examinations, which are the most comprehensive types of examinations. Alternatively, it is possible that the substantial understatement threshold is too uncertain or complicated to affect behavior.

67 This table is based on the EITC tables printed on IRS Pub. 596, *Earned Income Credit (EIC)*, for each of the tax years in question.

Appendix A

FIGURE A.1, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for All Taxpayers

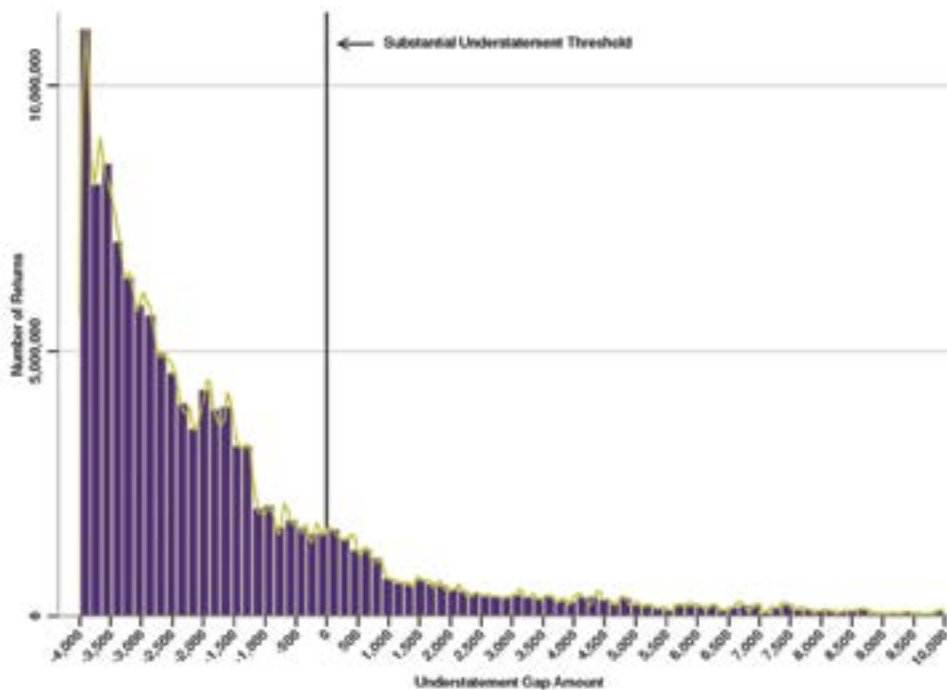


FIGURE A.2, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for Those with Income of at Least \$100,000 Who Used a Paid Preparer

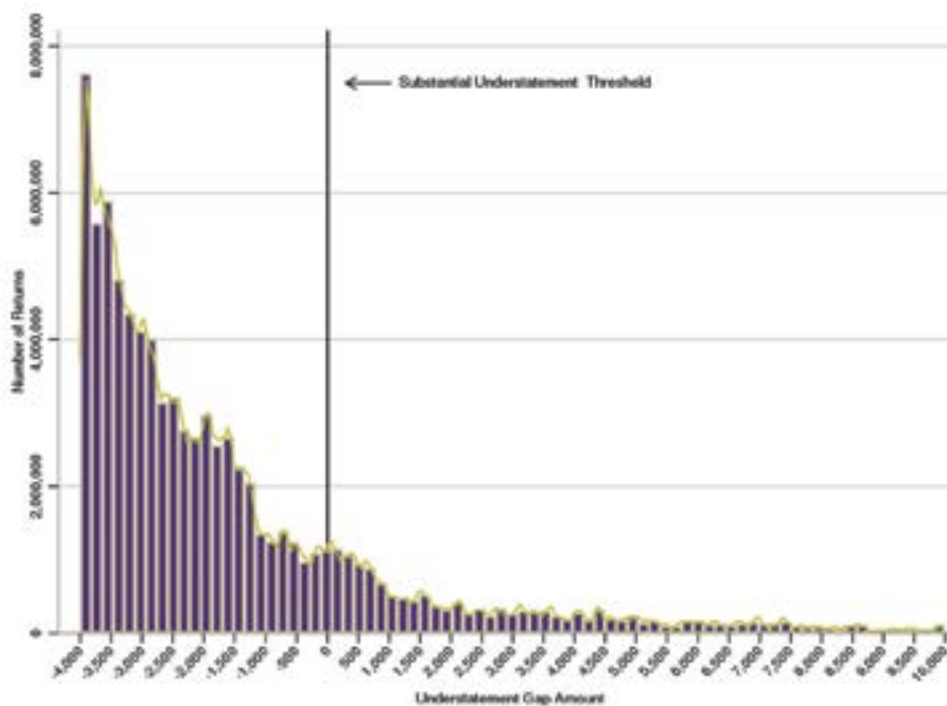


FIGURE A.3, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for Those with Income of at Least \$100,000 Who Did Not Use a Paid Preparer

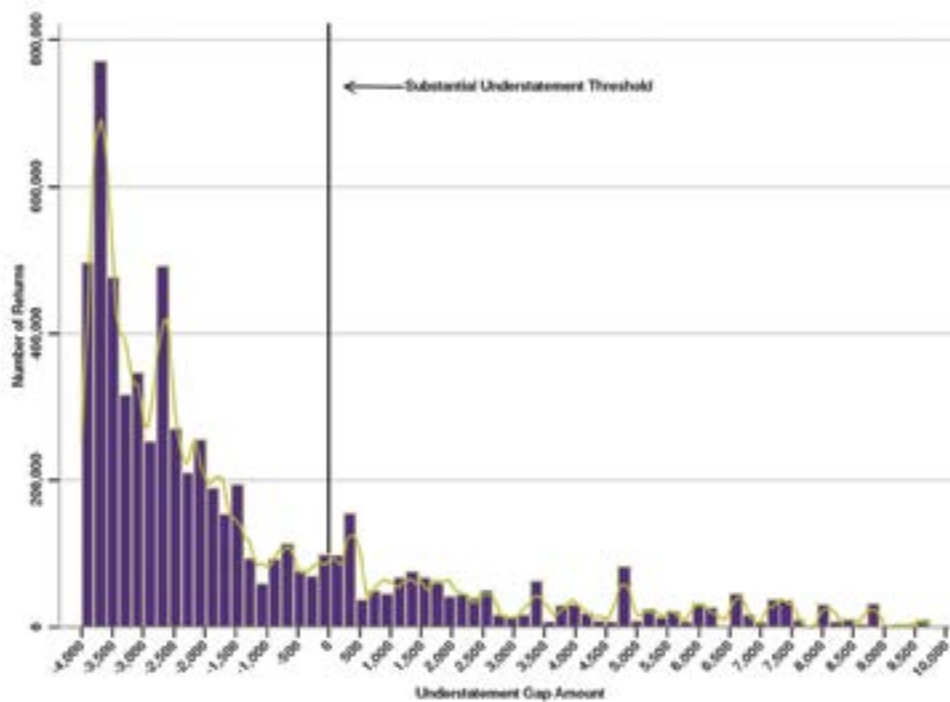


FIGURE A.4, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for Those Who Claimed EITC

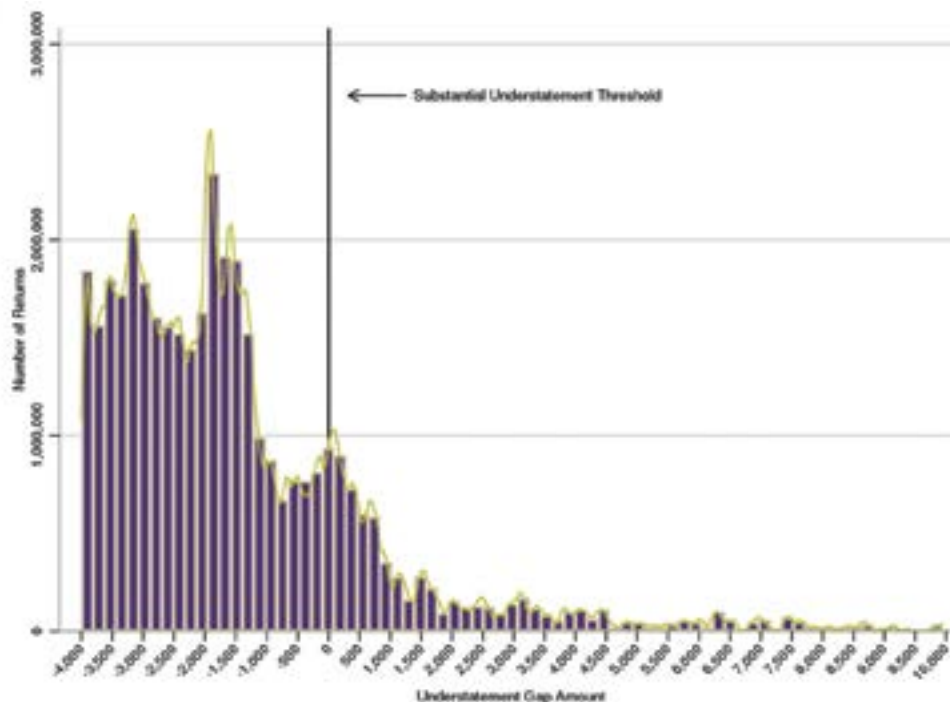


FIGURE A.5, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for Those Who Claimed EITC and Used a Preparer

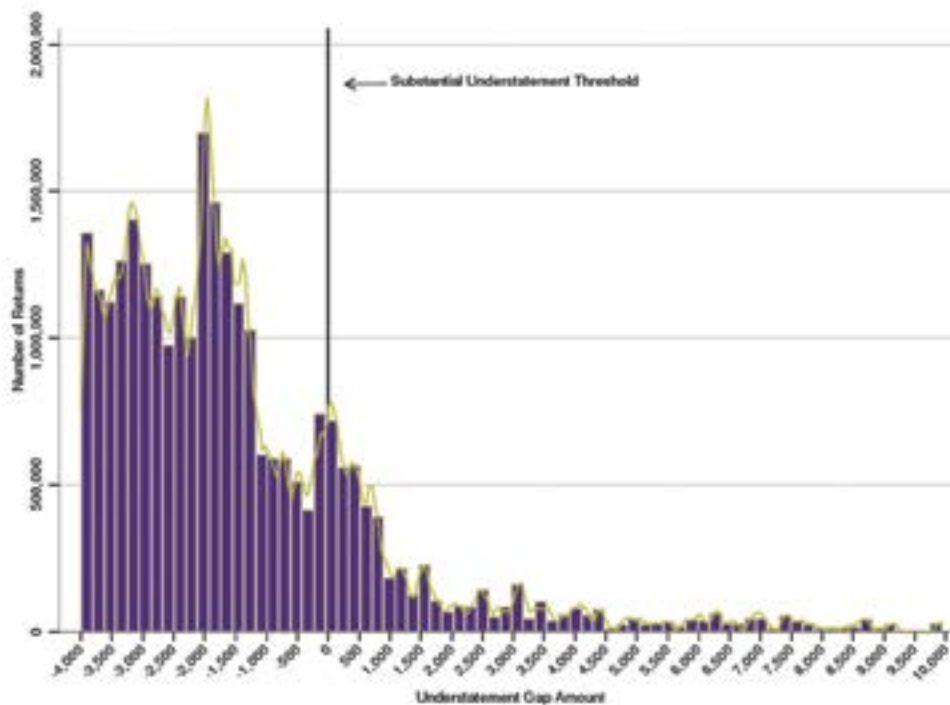


FIGURE A.6, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for Those Who Claimed EITC and Did Not Use a Preparer

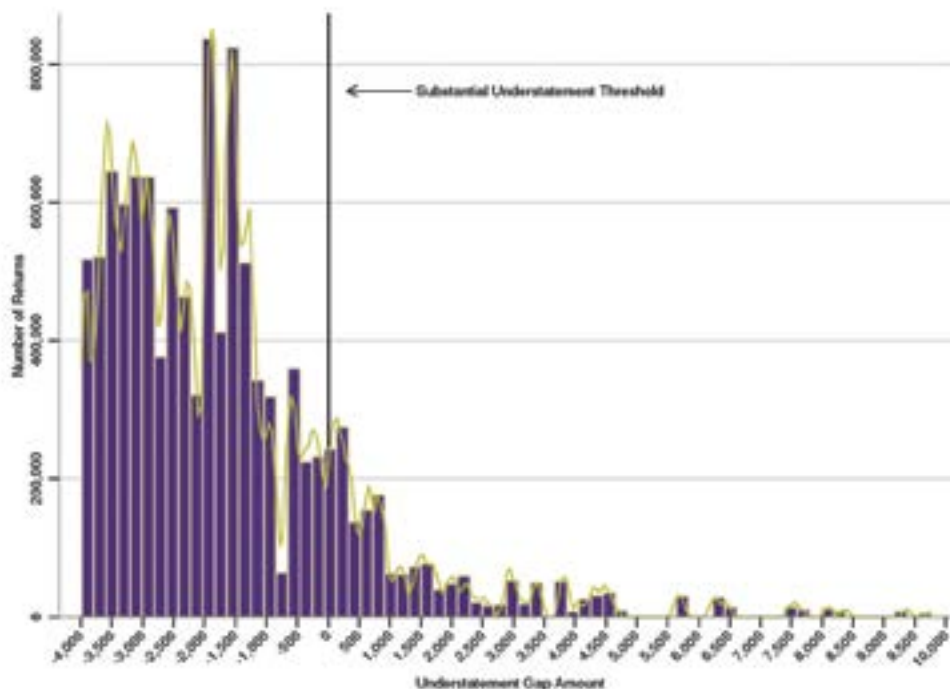


FIGURE A.7, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for Those Who Did Not Claim EITC

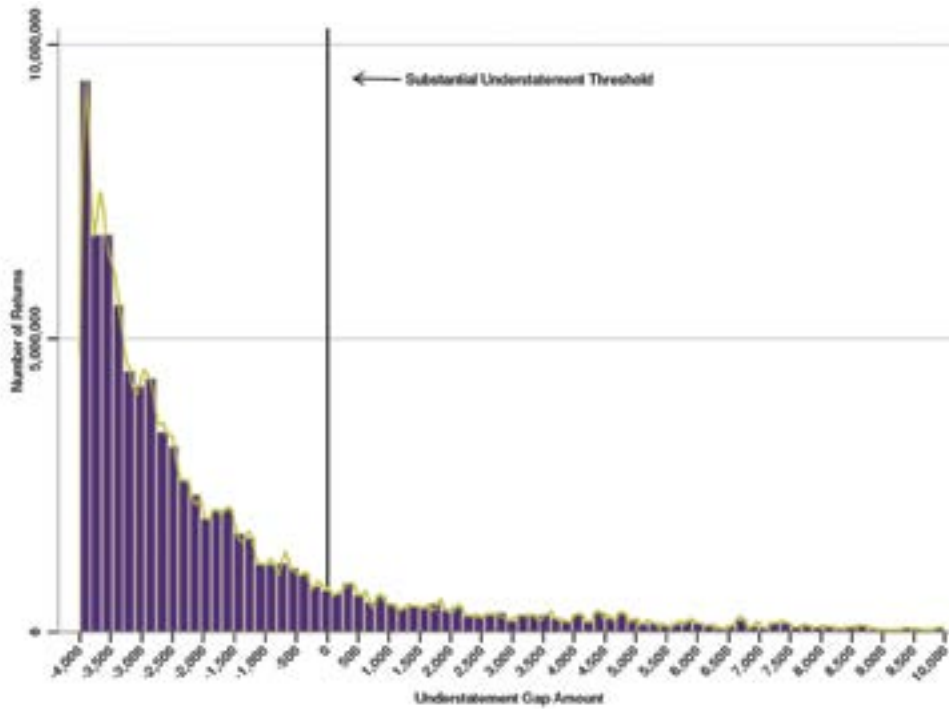


FIGURE A.8, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for Those with Income Below \$100,000 Who Did Not Claim EITC

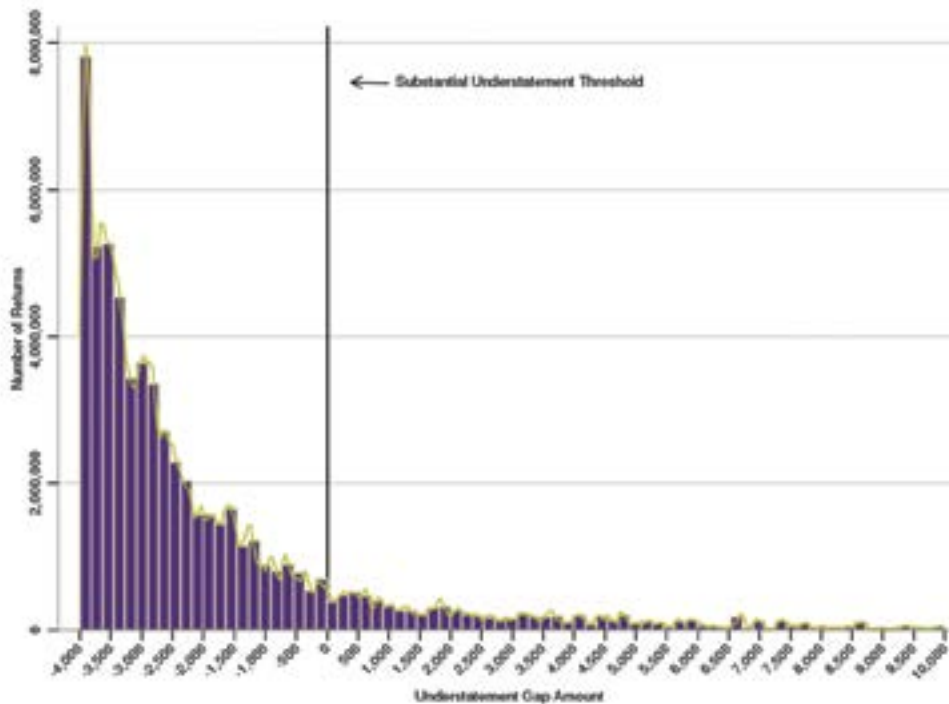


FIGURE A.9, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for Those Who Used a Paid Preparer and Did Not Claim EITC

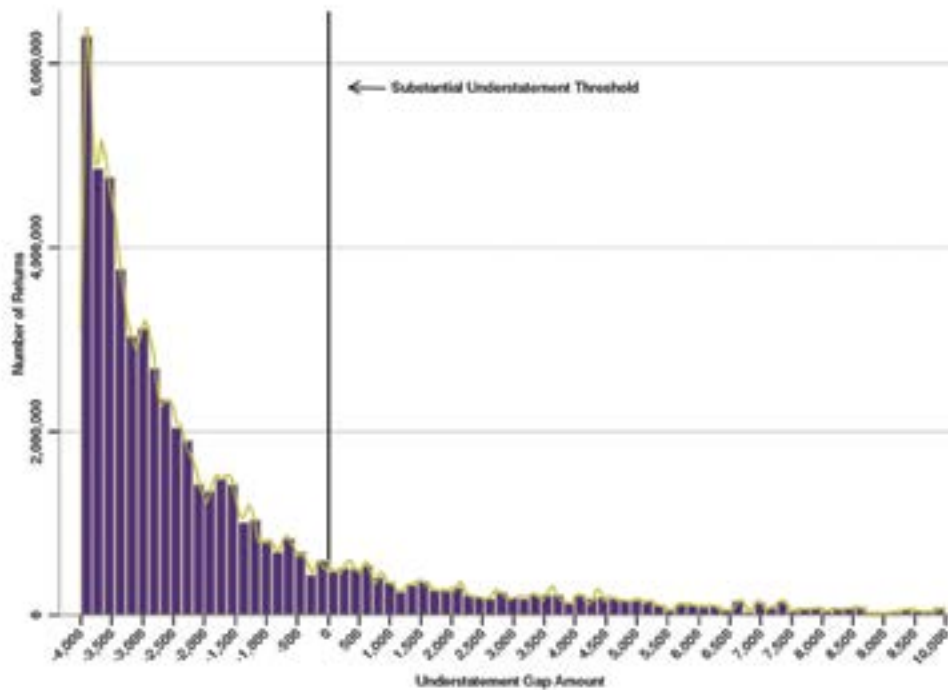


FIGURE A.10, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for Those Who Did Not Use a Paid Preparer or Claim EITC

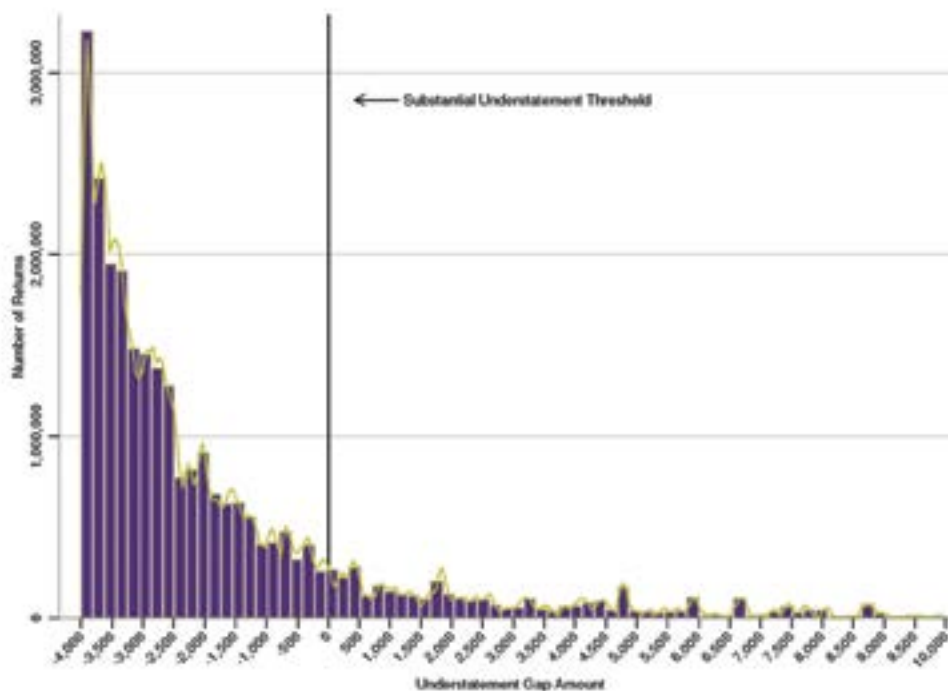


FIGURE A.11, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for Sole Proprietors

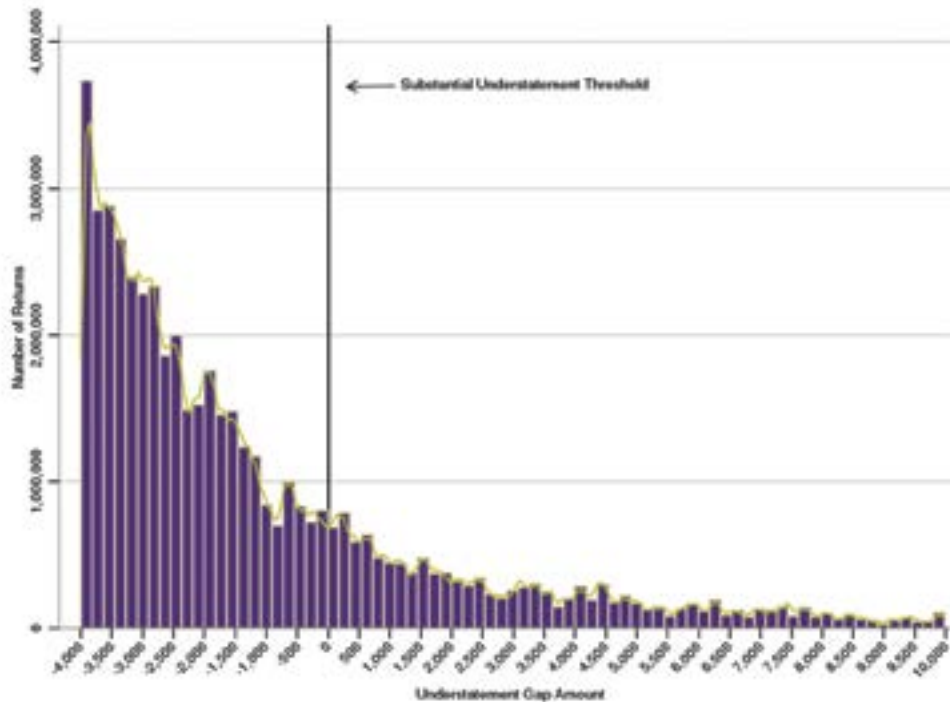


FIGURE A.12, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for Sole Proprietors Who Claimed the EITC

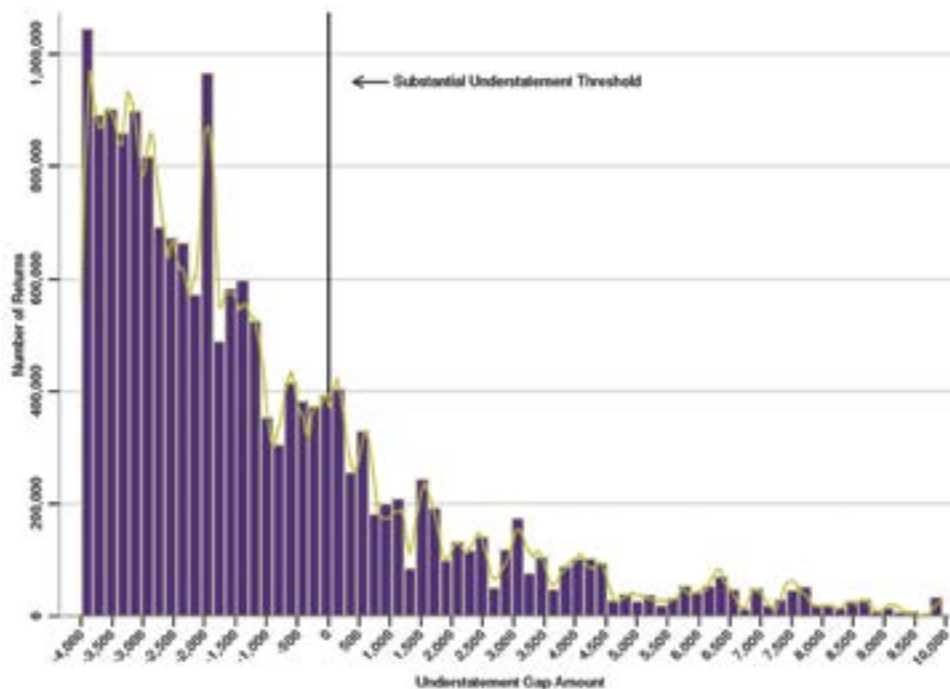


FIGURE A.13, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for Sole Proprietors Who Did Not Claim the EITC

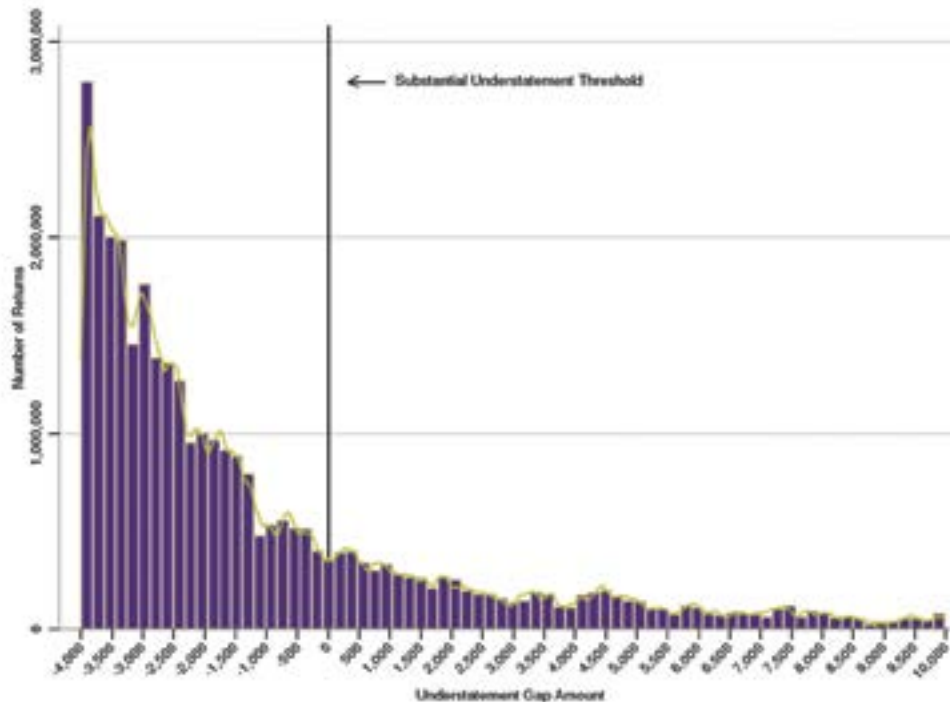
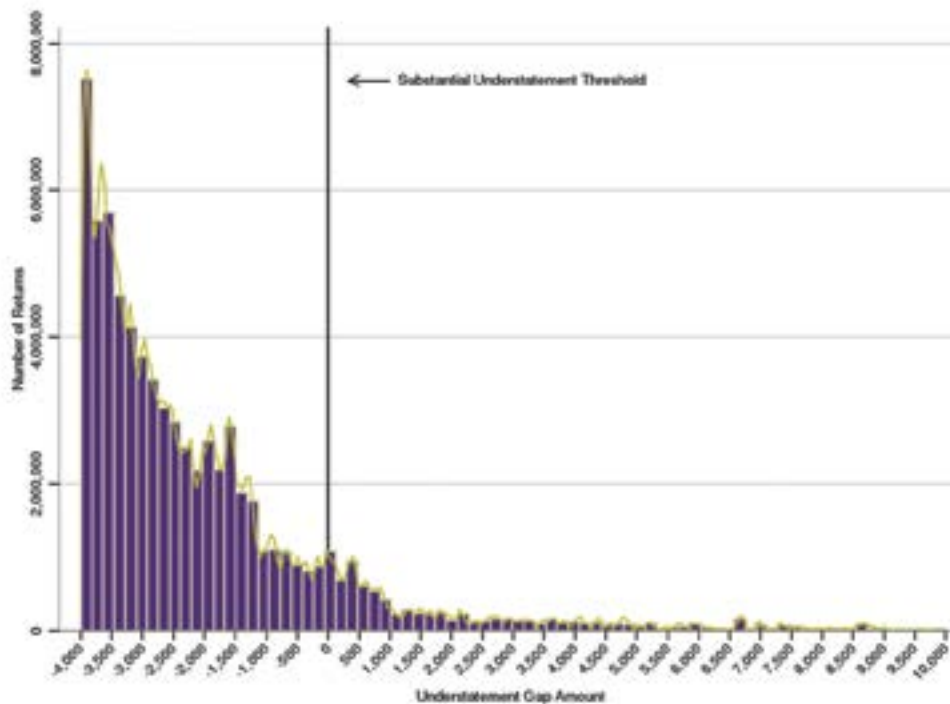


FIGURE A.14, The Gap Between Substantial Understatement Thresholds and NRP-Detected Understatements for Those Who Were Not Sole Proprietors



This page intentionally left blank.