Reexamining Income Tax Overwithholding as a Response to Uncertainty

Ashvin Gandhi
Pomona College

Michael Kuehlwein
Pomona College

Preliminary Version: November 13, 2012

Abstract

This paper reexamines the proposition of Highfill et al. (1998) that income tax overwithholding can be explained as taxpayers trying to avoid penalties for underwithholding when faced with uncertainty about tax liability. We first adjust the model presented in Highfill et al. to account for interest accumulated on underwithheld income and to enforce necessary boundary conditions. Both of these changes reduce the theoretically predicted probability of overwithholding. Further, we incorporate a relevant tax rule into the model and find that when this is taken into account, the model predicts only half of the true refund rate. Finally, we replace two seemingly-innocuous distributional assumptions with ones that allow for more realistic levels of taxpayer uncertainty about tax liability. We find that then, even under generous assumptions, the model predicts only about a fifth of the true refund rate on average. These results provide strong evidence that overwithholding cannot be fully explained by penalty-aversion.
I. Introduction

In the United States, income tax on salary is automatically withheld from one’s paycheck. While the default rate of automatic withholding tends to exceed one’s tax liability, informed employees can adjust their automatic withholding by filing a Form W-4 with their employer. Additionally, one must make estimated payments for any tax liability stemming from non-salary income.

Surprisingly, each year a significant majority of taxpayers choose to withhold more than they owe in taxes. And they overshoot their tax liability by a significant margin: the average taxpayer’s overpayment is 7% of their adjusted gross income (Jones 2010). While the government does eventually return this money as a tax refund, it amounts to 340 billion dollars, zero-interest loan made voluntarily by 83% of America’s 140 million income tax filers (Internal Revenue Service 2011). This high refund rate is not new, and the root of taxpayers’ persistent willingness to make this zero-interest loan rather than earn positive returns in the market is an area of some debate.

The significance of this phenomena is underscored by the fact that low income taxpayers are the most likely to withhold too much, and they do so by the largest relative margin: the average low income taxpayers’ overpayment is 13% of their adjusted gross income (Jones 2010). This group tends to accept high interest rates—such as from credit cards and payday loans—in order to smooth consumption over the year, so the implications of withholding too much extend beyond forgone interest. Jones (2012) estimates that the cost to consumption smoothing can be as high as 14% of income for taxpayers in the bottom quintile.

Highfill, Thorson, and Weber (1998), henceforth HTW, propose that taxpayers withhold more than necessary because they are uncertain about their tax liability and hope to avoid penalties for withholding too little. In the period we examine, 1983-1992, the Internal Revenue Service (IRS) did not impose a penalty if the taxpayer either withheld more than a preset percentage of his current-year tax liability or more than his previous year’s tax liability. A taxpayer who is penalized is said to be “underwithholding.” We consider a taxpayer whose withholding exceeds at least one of the above cutoffs, and thus will not be penalized, to be “overwithholding” because he could lower his withholdings without being penalized. It is important to observe the distinction between someone who is overwithholding and someone who receives a refund. One can be the former without being the latter. In fact, this arguably is the taxpayer’s ideal situation to be in: withholding less than he owes, but not so little as to be penalized.

HTW present a model in which risk-neutral taxpayers, who are uncertain about their tax liability, choose the level of withholding that minimizes the sum of their expected penalties from underwithholding and their expected forgone interest from overwithholding. Under this model, taxpayers withhold at rates which cause them to overwithhold with great frequency and receive refunds with only slightly less frequency. In fact, their model predicts refund rates that approach, and sometimes even exceed, the actual refund rate.

While HTW introduce a novel and informative way to model withholding, there are a few key oversights and tenuous assumptions in their model. This paper accounts for these oversights and adjusts their model to see whether their key results still hold. First, we expand the model to account for the interest that the taxpayer might earn on underwithheld income. Second, we impose important boundary conditions required by HTW’s assumptions. Each of these changes decreases the model’s predicted probability of overwithholding.

When we further account for the fact that taxpayers whose withholdings exceed last year’s liability are not penalized, the model’s estimates of the refund rate are about half of the actual refund rates. We independently replace two seemingly-innocuous assumptions about the PDF of tax liability with more realistic ones, the model’s estimates of the refund rate fall even further. The first of these changes is that we allow taxpayers to have some certain tax liability. We find that the more certain liability taxpayers have, the less inclined they are to overwithhold. Under reasonable assumptions about the ratio of certain to uncertain liability, the model predicts refund rates that average only a quarter of the actual ones. Second, we allow for tax liability to be normally distributed. Intuitively, this can be thought of as allowing for greater taxpayer precision in estimating their liability. The result is that the model’s predicted refund rate decreases as taxpayers are able

---

1 Since we can think of estimated tax payments as just a manual form of withholding, it should be understood that henceforth, we take “withholding” to include both automatic withholding and estimated payments.
2 Fennell (2006) aptly terms this decision a “hyperopic choice.”
3 The exact tax law on when penalties will be assessed is actually rather complicated, however these rules are a rather accurate distillation of the laws that were applied during the time period we analyze in this paper.
4 HTW impose a trapezoidal distribution on tax liability.
to more precisely estimate their liability. When liability is normally distributed under reasonable parameters, the model’s predicted refund rate averages 15.6% from 1983-1992, about a fifth of the real average of 74.8%. This demonstrates that penalty-aversion alone is not sufficient to fully explain overwithholding.

The remainder of this paper is organized as follows. Section II discusses the literature on income tax overwithholding. Section III summarizes the model and results from HTW. Section IV presents our own amendments and adjustments to the model and their implications. Section V concludes, and we include proofs of all non-immediate results in the Mathematical Appendix.

II. Literature

Outside of HTW, most of the literature views overwithholding as more of a behavioral phenomenon than the result of optimization under uncertainty. In particular, overwithholding with the intent of receiving a refund has been frequently viewed as a forced savings mechanism due to mental accounting or time inconsistency. For example, Neumark (1995) uses overwithholding as a metric for forced savings preferences and finds that overwithholding tends to correlate with steeper earnings profiles (with lower intercepts), which he takes as support for the hypothesis that rising earnings profiles are a forced savings mechanism.

Barr and Dokko (2007) use survey data to examine the relationship between portfolio liquidity and withholdings for low- to moderate-income taxpayers. They find evidence consistent with the theory that overwithholding is used as a commitment device to constrain consumption. Further, they find evidence against mental accounting of refunds as windfalls and loss aversion as the causes of overwithholding.

There is some evidence that overwithholding might not only be seen as a mechanism to force savings until refund day, but also as a mechanism for motivating future savings. Chambers and Spencer (2008) try to determine how refund timing affects consumption and savings patterns and finds that yearly refunds yield more savings and less consumption than do monthly refunds. This argument is also made explicitly in Thaler (1996) and evidenced in Shefrin and Thaler (1998).

Most recently, Jones (2012) proposes that inertia plays a large role in overwithholding. By analyzing taxpayer responses to changes in dependents, the expansion of the Earned Income Tax Credit, and the 1992 mandated change in default withholdings, Jones determines that taxpayer adjustments to changes in liabilities or the withholding rate are often slow or non-existent. This lines up with the kind of “default effects” seen in Madrian and Shay (2001).

Importantly, Jones provides strong empirical evidence against the claim that taxpayers are consciously setting withholding based on a desire to minimize expected penalties and opportunity costs. Thus, while it seems that empirical evidence strongly supports that overwithholding is a behavioral phenomenon, little has been done to challenge the theory presented by HTW. This is what we aim to do in this paper: reexamine HTW’s model and determine whether it robustly dictates that overwithholding is primarily a response to uncertainty.

III. Model and Results from Highfill et al.

HTW model a risk neutral taxpayer attempting to optimize his level of withholdings ($W$) under two different tax rules. The first is that the taxpayer must withhold at least a set percentage ($\alpha$) of his current year’s tax liability. HTW term this the “90% Rule” since they assume $\alpha = .9$ for all of years they examine (1983-1992). However, in reviewing historical IRS Form 1040 Instructions, we found that while $\alpha$ has been .9 since 1987, prior to 1987, $\alpha$ was .8. Therefore, we shall simply refer to this rule as Rule 1. The second tax rule more closely mirrors the tax law in the period we examine. Under the second rule, the taxpayer must withhold at least the minimum of what is required by Rule 1 and his previous year’s tax liability ($N$).

A taxpayer who withholds less than the minimum required amount to satisfy the rule in effect is said to have “underwithheld.” Underwithholding taxpayers are made to pay a per-dollar penalty ($j$) on the deficit of their withholdings to the minimum level of withholding required by the rule in effect. On the other

---

5Shapiro and Slemrod (1995) implement surveys to determine the effect on consumption of President George H. W. Bush’s 1992 mandated decrease in default withholdings. Their results support that “myopic or rule-of-thumb decision-making” (as opposed to liquidity constraints) may be behind some household’s decisions to spend their newfound income, a result which they conclude is compelling in the context of the puzzle of overwithholding.
hand, if a taxpayer withholds more than necessary to avoid being penalized, we say that this taxpayer has “overwithheld.” There is a per-dollar opportunity cost \( k \) that applies to the taxpayer’s withholding surplus beyond the minimum required level of withholding. We can think of this as the return the taxpayer could have earned if the money he overwithheld were invested in the market.\(^6\) We expect that, in general, the IRS will set \( j \) so that \( j > k \) in order to discourage underwithholding.

In this model, the taxpayer aims to minimize his expected penalty or opportunity cost \( (PC) \) from withholding.\(^7\) Therefore, the taxpayer would like to set his withholding to exactly match the minimum rule-satisfying level. However, the taxpayer’s problem is complicated by uncertainty about his tax liability. HTW model this uncertainty by assuming that the taxpayer has some certain and uncertain income \((Y_c \) and \( Y \), respectively) and certain and uncertain allowances \((A_c \) and \( A \), respectively). They further assume that \( Y \) and \( A \) are bounded below by 0 and above by \( V \) and \( U \), respectively, and that the variables have the joint probability distribution function \( f(A, Y) \). Also, the taxpayer faces a fixed tax rate \( (t) \), so that his tax liability is \( t(Y_c + Y - A_c - A) = t(Y-A) + m \), where \( m \equiv t(Y_c - A_c) \). Finally, for mathematical convenience, it is assumed that \( \alpha m < W < \alpha m + \alpha t(V-U) \).\(^8\)

**Rule 1**

Under Rule 1, a per-dollar penalty of \( j \) percent is assessed on the deficit of one’s withholdings from \( \alpha (t(Y-A) + m) \). Similarly, there is also a per-dollar opportunity cost of \( k \) percent for any withholding in excess of \( \alpha (t(Y-A) + m) \). Therefore, HTW assert that the penalty or cost to a given level of withholding can be given as follows:

\[
PC = \begin{cases} 
  j \left[ \alpha (t(Y-A) + m) - W \right] & \text{if } W \leq \alpha [t(Y-A) + m], \\
  k \left[ W - \alpha (t(Y-A) + m) \right] & \text{if } W > \alpha [t(Y-A) + m]. 
\end{cases}
\]

(1)

It is clear that if \( Y \) and \( A \) were known, the taxpayer would set \( W = \alpha (t(Y-A) + m) \) and avoid any penalties or opportunity costs. But, since \( Y \) and \( A \) are random variables, the risk-neutral taxpayer sets \( W \) to minimize:

\[
E[PC] = \int_0^U \int_{\frac{W}{\alpha} - \frac{m}{\alpha} + A}^V j \left[ \alpha (t(Y-A) + m) - W \right] f(A,Y)dYdA \\
+ \int_0^U \int_{\frac{W}{\alpha} - \frac{m}{\alpha} + A}^V k \left[ W - \alpha (t(Y-A) + m) \right] f(A,Y)dYdA.
\]

(2)

The first term above gives the expected penalty from cases in which the taxpayer has underwithheld, and the second term gives the expected opportunity cost from cases in which the taxpayer has overwithheld. Solving the model yields that the taxpayer sets \( W \) so that:

\[
Pr(\text{underwithholding}) = \frac{k}{j+k}, \tag{3}
\]

\[
Pr(\text{overwithholding}) = \frac{j}{j+k}. \tag{4}
\]

Figure 1 gives a graphical interpretation of this result. The line \( Y = \frac{W}{\alpha} - \frac{m}{\alpha} + A \) is the “no penalty-cost” line: the line where the taxpayer has managed to neither overwithhold nor underwithhold. When \((A,Y)\) is above this line (i.e. \( W < m + t(Y-A) \)), the taxpayer has underwithheld. When \((A,Y)\) is below this line (i.e. \( W > m + t(Y-A) \)), the taxpayer has overwithheld. We can think of the taxpayer as setting \( W \) (and thus shifting the no penalty-cost line) so that the probability the areas above and below the no penalty-cost line satisfy equations (3) and (4).

The result is consistent with what one might expect: as the penalty to underwithholding \( (j) \) increases, the taxpayer withholds more and the probability of underwithholding decreases. And, as the opportunity cost cost

---

\(^6\) In some cases, however, this interpretation of \( k \) is insufficient. Many taxpayers accept high rates of interest to smooth consumption, so this interpretation of \( k \) may underestimate the true costs to overwithholding.

\(^7\) Note that the taxpayer is not directly concerned with withholding to meet his tax liability. This is because if he undershoots his liability, he will be expected to pay the balance, and if he overshoots his liability, the government will refund his overpayment. Therefore, the selected level of withholding only affects penalties and opportunity costs for under and overwithholding.

\(^8\) Note that implicit in this assumption is that \( V > U > 0 \).
to overwithholding \((k)\) increases, the taxpayer withholds less and probability of overwithholding decreases. Graphically, this can be interpreted as that when \(j\) increases, the taxpayer shifts the no penalty-cost line up, and when \(k\) increases, the taxpayer shifts the no penalty-cost line down. Finally, it is worth noting that, given realistic values for \(j\) and \(k\), the probability of overwithholding predicted is substantial. For example, if \(j = .07\) and \(k = .035\), then the taxpayer will select \(W\) such that his probability of overwithholding is \(\frac{2}{3}\).

**Rule 2**

Under Rule 2, a penalty of \(j\) percent is assessed on the deficit of one’s withholdings from \(M = \min\{\alpha(t(Y - A) + m), N\}\) the lower of the minimum withholding required under Rule 1 and one’s previous year’s liability. There is an opportunity cost of \(k\) percent to one’s withholdings in excess of \(M\). HTW therefore assert that the penalty-cost function \((PC)\) can be given as follows:

\[
PC = \begin{cases} 
  j(M - W) & \text{if } W \leq M, \\
  k(W - M) & \text{if } W > M.
\end{cases}
\]

By setting up an analogous minimization problem to that used in examining Rule 1, it can be shown that the taxpayer sets \(W\) according to the following rules:

\[
Pr(\text{overwithholding}) = Pr(\text{Area 3}) = \frac{j}{j + k} \text{ if } Pr(\text{Area 1}) \leq \frac{k}{j + k},
\]

\[
W = N \text{ if } Pr(\text{Area 1}) > \frac{k}{j + k},
\]

where Area 1 and Area 3 are as depicted in Figure 2.

Figure 2 shows the support of \(f(A,Y)\) partitioned into three areas by two lines. The bottom line is the no penalty-cost line, which represents the set of \((A,Y)\) for which the taxpayer has managed to neither underwithhold nor overwithhold. Area 3, the area under the no penalty-cost line, represents the set of possible \((A,Y)\) such that the taxpayer has overwithheld. Also as before, the taxpayer sets \(W\) to shift the no penalty-cost line in order to satisfy the conditions given in equation (6). In addition to the no penalty cost line, we now also have a “rules equal” line, \(Y = \frac{N}{\alpha}t - \frac{m}{T} + A\), which shifts upwards and downwards with \(N\). The rules equal line derives its name from the fact that if \(N\) is sufficiently large, then the rules equal line will be high enough that \(Pr(\text{Area 1}) \leq \frac{k}{j + k}\), where \(Pr(\text{Area 1})\) is the probability that the taxpayer’s current liability exceeds his previous year’s liability, and the taxpayer’s chosen withholding will be the same under both rules.

The result in equation (6) can actually be seen quite intuitively. Under Rule 2, \(W \geq N\) guarantees that a taxpayer will not be penalized. Therefore, a taxpayer should never set \(W > N\) because \(W = N\) is a strictly superior choice since it decreases overwithholding without any corresponding increase in the probability of underwithholding. In fact, we can think of the taxpayer as setting \(W\) as if Rule 1 applied, except in cases that this strategy would dictate \(W > N\), in which case the taxpayer sets \(W = N\). In other words, the only difference between how a taxpayer behaves under Rule 2 and Rule 1 is that under Rule 2, if a taxpayer is sufficiently confident that his current liability will exceed his previous year’s liability, he will simply withhold as much as he owed last year. Therefore, the taxpayer’s probability of overwithholding is weakly lower under Rule 2 than Rule 1.

One surprising point of note is that taxpayers do not seem to adhere to the intuition that one should set \(W \leq N\). Using the Michigan Tax Panel of IRS Statistics of Income data, we calculated the percentage

\footnote{At this point in their exposition, HTW let \(m = 0\). Though it does not affect the probability of overwithholding, we continue to allow \(m \geq 0\) in this exposition for the sake of generalization.}

\footnote{This result contained a typo in Highfill et al. The published result was that the taxpayer sets \(W\) so that:}

\[
Pr(\text{Area 3}) = \frac{j}{j + k} \text{ if } Pr(\text{Area 1}) < \frac{k}{j + k},
\]

\[
W = N \text{ if } Pr(\text{Area 2}) > \frac{k}{j + k},
\]

\[
W = N \text{ if } Pr(\text{Area 1}) = \frac{k}{j + k}.
\]

\footnote{This implies that the rules equal line will always be weakly above the no penalty-cost line. The region between the two constitutes Area 2.}
of taxpayers who set $W > N$ and found that for almost all years form 1980-1990, between 63 and 72% of taxpayers set $W > N$! The sole exception was in 1987 when there were drastic changes in tax liabilities and withholding rates due to the Tax Reform Act of 1986. This is a strong indicator that taxpayers may not cognitively evaluate the problem of overwithholding in the way that is proposed by HTW.

The Refund Rate

The refund rate in the United States has historically been quite high: in the time period we examine, the average refund rate was 74.8%, and one of the major results propounded by HTW is that their model predicts similarly high rates. They solve for a theoretical range for the refund rate under the three key assumptions. First, they assume that $N$ is sufficiently small that Rule 1 and Rule 2 yield identical results. Second, they assume that $f(A,Y)$ is uniform. Third, they assume that $m = 0$, so the taxpayer has no certain tax liability.

Under these assumptions, they find that the optimal level of withholding is:

$$W = \frac{\alpha t (2jV - (j + k)U)}{2(j + k)}. \quad (7)$$

Since taxpayers receive a refund when they have withheld more than their liability, i.e. when $W > t(Y - A)$, the refund rate ($R$) is then:

$$R = \frac{\alpha j}{j + k} + \frac{(1 - \alpha)}{2} \frac{U}{V}. \quad (8)$$

HTW point out that since they assume $V > U > 0$, it must be that $0 < \frac{U}{V} < 1$, so:

$$\frac{\alpha j}{j + k} < R < \frac{\alpha j}{j + k} + \frac{1 - \alpha}{2}. \quad (9)$$

They compare the above theoretical interval for the refund rate with actual refund rates from 1983 to 1992 in Table 1. According to Table 1 the average range of refund rates predicted by the model for the time period is 64.2-69.4, only a few percentage points below the actual average of 74.4. It would seem from these results that uncertainty largely explains income tax overwithholding and the refund rate.

IV. Adjusted Model and Results

In this section, we extend and adjust the model presented by HTW. We include proofs for all non-immediate results in the Appendix.

Interest Earned While Underwithholding

One factor overlooked by HTW in their model is the interest that taxpayers can earn on their withholding deficit. While the IRS will penalize this deficit at the per-dollar rate $j$, this money can be invested to earn interest at the per-dollar rate $k$. Therefore, the effective penalty on any dollar underwithheld is actually $j - k$, not $j$. Accounting for this, the penalty-cost functions under Rule 1 and Rule 2 are respectively:

$$PC = \left\{ \begin{array}{ll} (j - k) [\alpha (t(Y - A) + m) - W] & \text{if } W \leq \alpha [t(Y - A) + m], \\ k [W - \alpha (t(Y - A) + m)] & \text{if } W > \alpha [t(Y - A) + m]. \end{array} \right. \quad (10)$$

$$PC = \left\{ \begin{array}{ll} (j - k) (M - W) & \text{if } W \leq M, \\ k (W - M) & \text{if } W > M. \end{array} \right. \quad (11)$$

By minimizing this function analogously to that in Section III, we find that under Rule 1, the taxpayer sets $W$ to satisfy:

12The theoretical refund rates in Table 1 assume that $\alpha = .9$ for all years, while in reality $\alpha = .8$ for 1983-1986. While we do not correct this in the replicated table, we apply the correct value for $\alpha$ henceforth in this paper.
such a case, there is no non-arbitrary reasoning behind why one cannot set $W < m$ and when the desired $W > \alpha t$ we enforce the new bounds and maintain the assumption that the effective penalty to overwithholding is some substantial probability of overwithholding! This probability is $Pr(\text{Area} T_j)$.

We can see from Figure 3 that even as $W$ approaches zero, the assumptions imply that there still exists some substantial probability of overwithholding! This probability is $Pr(\text{Area} T_2)$. If $Pr(\text{Area} T_2) \geq 1 - \frac{k}{j}$, then the taxpayer would aim to set $W \leq 0$, violating the assumption $W > 0$. Similarly, if $Pr(\text{Area} T_1) \geq \frac{k}{j}$, then the taxpayer would wish to set $W \geq \alpha t(V - U)$, violating $W < \alpha t(V - U)$.

Under the assumption that $A$ and $Y$ are uniformly distributed, $Pr[\text{Area} T_1] = Pr[\text{Area} T_2] = \frac{U}{V}$. Therefore, in order to insure that the taxpayer selects a $W$ that does not violate any assumptions, it must be that

$$0 < \frac{U}{V} < 2 \min\left\{ \frac{k}{j}, 1 - \frac{k}{j} \right\}. \quad (15)$$

These bounds for $\frac{U}{V}$ are tighter than those used by HTW in calculating the theoretical range of the refund rate. For example, when $j = .067$ and $k = .052$, as was the case in 1984, the bounds require $\frac{U}{V} < .45$. When we enforce the new bounds and maintain the assumption that the effective penalty to overwithholding is $j - k$, we find that:

$$W = \alpha t \left[ \left( 1 - \frac{k}{j} \right) V - \frac{U}{2} \right]. \quad (16)$$

Both the upper and lower bound for the refund rate here are weakly lower than those presented by HTW. The lower bound decreased because we employed the effective penalty, $j - k$, for underwitholding. The upper decreased both because we employed the effective penalty and because we enforced all of the required bounds on $\frac{U}{V}$.

Similarly, the Kuhn Tucker conditions for the optimization problem under Rule 2 dictate that the taxpayer sets $W$ to satisfy:

$$Pr(\text{overwithholding}) = Pr(\text{Area} 3) = 1 - \frac{k}{j} \quad \text{if} \quad Pr(\text{Area} 1) \leq \frac{k}{j},$$

$$W = N \quad \text{if} \quad Pr(\text{Area} 1) > \frac{k}{j}. \quad (14)$$

Just as in Section III, we find that taxpayers behave identically under Rule 2 as they did under Rule 1, unless their previous year’s liability is sufficiently low, i.e. unless $Pr(\text{Area} 1)$ is sufficiently high. In that case the taxpayer simply sets $W = N$.

In comparing these results to those in Section III, we see that the probability of overwithholding has weakly decreased: $1 - \frac{k}{j} \leq \frac{U}{V}$ for all positive $j$ and $k$. And, this difference can be substantial: for example, if $j = .07$ and $k = .035$, then $1 - \frac{k}{j} = \frac{1}{2} < \frac{U}{V}$. Therefore, by overstating the effective penalty for underwitholding, HTW overestimate the degree to which overwithholding is justified by penalty aversion under either Rule 1 or Rule 2. We continue to apply these new results in the rest of the paper.

### Enforcing Boundary Conditions

HTW calculate the theoretical range of refund rates by assuming $0 < \frac{U}{V} < 1$; however, their initial assumption that $am < W < am + \alpha t(V - U)$ dictates that even tighter bounds must be applied to $\frac{U}{V}$. To see this, note that when $m = 0$, then $0 < W < \alpha t(V - U)$, bounds which are depicted in Figure 3.

We can see from Figure 3 that even as $W$ approaches zero, the assumptions imply that there still exists some substantial probability of overwithholding! This probability is $Pr(\text{Area} T_2)$. If $Pr(\text{Area} T_2) \geq 1 - \frac{k}{j}$, then the taxpayer would aim to set $W \leq 0$, violating the assumption $W > 0$. Similarly, if $Pr(\text{Area} T_1) \geq \frac{k}{j}$, then the taxpayer would wish to set $W \geq \alpha t(V - U)$, violating $W < \alpha t(V - U)$.

Under the assumption that $A$ and $Y$ are uniformly distributed, $Pr[\text{Area} T_1] = Pr[\text{Area} T_2] = \frac{U}{V}$. Therefore, in order to insure that the taxpayer selects a $W$ that does not violate any assumptions, it must be that

$$0 < \frac{U}{V} < 2 \min\left\{ \frac{k}{j}, 1 - \frac{k}{j} \right\}. \quad (15)$$

These bounds for $\frac{U}{V}$ are tighter than those used by HTW in calculating the theoretical range of the refund rate. For example, when $j = .067$ and $k = .052$, as was the case in 1984, the bounds require $\frac{U}{V} < .45$. When we enforce the new bounds and maintain the assumption that the effective penalty to overwitholding is $j - k$, we find that:

$$W = \alpha t \left[ \left( 1 - \frac{k}{j} \right) V - \frac{U}{2} \right]. \quad (16)$$

Both the upper and lower bound for the refund rate here are weakly lower than those presented by HTW. The lower bound decreased because we employed the effective penalty, $j - k$, for underwitholding. The upper decreased both because we employed the effective penalty and because we enforced all of the required bounds on $\frac{U}{V}$.

**Note:** Alternatively, we could enforce these boundary conditions by assuming that if the taxpayer’s desired $W < m$, then $W = m$, and when the desired $W > \alpha t(V - U)$, then $W = \alpha t(V - U)$. This method, however, seems unrealistic when $m > 0$ because in such a case, there is no non-arbitrary reasoning behind why one cannot set $W < m$ and $W > \alpha t(V - U)$. Further, the method would violate the strictness of the inequalities in the bounds.

**Note:** It is worth noting that we found that properly calculating the refund rate when the bounds are violated actually yields very similar results.
Rephrasing the Problem

Until this point, we have maintained consistency with HTW by phrasing the withholding problem in terms of income \((Y_c + Y)\) and allowances \((A_c + A)\). However, one can rephrase the withholding problem in a way that simplifies the mathematics and elucidates some of the underlying assumptions. To do so, we observe that the taxpayer is not individually concerned with either income or allowances. Rather, the taxpayer is actually concerned with tax liability: \(t(Y_c - A_c + Y - A) = m + t(Y - A)\), where \(m\) is liability stemming from certain income and allowances, and \(t(Y - A)\) is liability stemming from uncertain income and allowances. By defining \(L = t(Y - A)\), so \(L\) is the stochastic part of liability, the taxpayer’s problem can be distilled into one of trying to set \(W\) given uncertainty about his current tax liability \((m + L)\). Under Rule 1, the taxpayer is penalized for withholding less than \(\alpha(m + L)\), and under Rule 2, the taxpayer is penalized for withholding less than \(M = \min\{\alpha(m + L), N\}\).

The assumption that \(f(A, Y)\) is uniform over \([0, U] \times [0, V]\) translates to that the PDF of \(L\), which we denote \(f(L)\), is:

\[
f(L) = \begin{cases} 
\frac{1}{V} + \frac{L}{V^2} & \text{if } -tU \leq L \leq 0, \\
\frac{1}{V} & \text{if } 0 < L < t(V - U), \\
\frac{1}{V} - \frac{L}{V^2} & \text{if } t(V - U) \leq L \leq tV, \\
0 & \text{otherwise.}
\end{cases}
\]

(18)

In other words, \(f(L)\) has the trapezoidal distribution depicted in Figure 4. The tails of the trapezoidal distribution correspond to T1 and T2 in the previous section, which we have also marked in Figure 4. Additionally, we have marked the range of \(W\) that the assumptions will allow with \(W_{\text{min}}\) and \(W_{\text{max}}\). Area 1 is now represented by the interval \([-tU, \frac{W}{\alpha}]\), and Area 2 and 3 are represented by the intervals \([\frac{W}{\alpha}, \frac{N}{\alpha}]\) and \([\frac{N}{\alpha}, tV]\), respectively. Therefore, the point \(L = \frac{W}{\alpha}\) is analogous to the no penalty-cost line because it represents the only case in which the taxpayer has neither under nor overwithheld. Moreover, if \(L < \frac{W}{\alpha}\), then the taxpayer has overwithheld, and if \(L > \frac{W}{\alpha}\), then the taxpayer has underwithheld.\(^{15}\) Similarly, the point \(L = \frac{N}{\alpha}\) is analogous to the rules equal line.

This form for \(f(L)\) acts as a significant impediment to our ability to examine the refund rate under Rule 2. The piecewise form of \(f(L)\) requires either for the analysis to be broken down into many cases or for variables, such as \(W\), to be highly restricted as they were in HTW. The limited range of \(W\) is problematic because it allows for area \(T_1\) to be an area in which the taxpayer always receives a refund, no matter what the taxpayer’s preferences. This is actually a very significant problem: under the given assumptions, \(\Pr[\text{Area } T_1] = \frac{W}{V}\), so as \(\frac{V}{U} \to 1\), \(\Pr[\text{Area } T_1] \to \frac{1}{2}\). In other words, as \(U\) approaches \(V\), the model presupposes that the lower-bound for the refund rate is at least 50%.

Fortunately, we can circumvent these problems by simply assuming that \(L\) is uniformly distributed between 0 and \(T = t(V - U) > 0\)\(^{16}\) where \(T\) is the taxpayer’s maximum possible tax liability stemming from uncertain income and allowances. This new \(f(L)\) does not presuppose either a minimum rate of overwithholding or a minimum rate of underwithholding, and it does not require that we place any bounds on \(W\). We depict \(f(L)\) under this assumption in Figure 5. Area 1 is now the interval \([0, \frac{W}{\alpha}]\) and Area 3 is now the interval \([\frac{N}{\alpha}, T]\), while Area 2 remains the same.

Importantly, this distributional assumption does not differ too greatly from that made by HTW, either in form or in result. In fact, we can think of it as truncating the previous distribution so \(\Pr[\text{Area } T_1] = \Pr[\text{Area } T_2] = 0\). In other words, we have simply removed the problematic cases. And, in doing so, we have not greatly changed the results: under the new distributional assumptions, the refund rate calculated under Rule 1 is equivalent to the lower-bound calculated under Rule 1 given the previous assumptions. Since the difference between the upper and lower bound of the refund rate in the previous analysis was only \((1 - \alpha)\min\left\{\frac{k}{\alpha}, 1 - \frac{k}{\alpha}\right\} \leq \frac{1 - \alpha}{2} = .05\) when \(\alpha = .9\), we conclude that the new assumed distribution for \(L\) is a rather innocuous change.

\(^{15}\)As before, this relies on that the taxpayer will always set \(W \leq N\).

\(^{16}\)In fact, the selection of \(T = t(V - U)\) is actually not necessary. Any \(T > 0\) will yield the same results in our following analysis.
Refund Rate Under Rule 2

In their calculation of the refund rate, HTW restrict their analysis to the case in which Rule 1 and Rule 2 yield equivalent withholding. In other words, they examine the refund rate only under Rule 1. They argue that one needn’t examine the refund rate under Rule 2 because Rule 2 will only be binding infrequently as it requires the taxpayer to anticipate a high probability of a substantial increase in liability.

The exact condition for Rule 2 to be binding—so Rule 2 yields different levels of withholding than Rule 1—is that \( \Pr(\text{Area 1}) > 1 - \frac{k}{j} \). Since

\[
\Pr(\text{Area 1}) = \Pr(L > \frac{N}{\alpha}) = \Pr(\alpha L > N),
\]

we can think of \( \Pr(\text{Area 1}) \) as the probability that \( \alpha \) of this year’s tax liability exceeds last year’s tax liability. Given salary’s correlation with factors such as experience and inflation, it does not seem unlikely that a taxpayer will experience such an increase in nominal liability. In fact, using the Michigan Tax Panel, we found that from in each of the years from 1980 to 1990 (excluding 1987), between 27 and 42% of taxpayers saw such an increase in their liability. Moreover, Neumark (1995) finds that refunds correlate with steeper earnings profiles, so we would expect the case in which tax liability increases (or is very likely to increase) to be highly relevant to any refund-related analysis. So, there may be more merit to examining the refund rate under Rule 2 than was given by HTW. Therefore, in this section, we aim to calculate the refund rate under Rule 2.

For our analysis, we assume that \( N \) has the same PDF as \( m + L \). In other words, the taxpayer faces the same range of tax liabilities this year as he did last year. Because our assumption implies an expected stagnant nominal earnings profile, while most people face increasing nominal earnings profiles, we likely underestimate the frequency with which Rule 2 is binding. Therefore, the chosen PDF of \( N \) is quite charitable to the argument of overwithholding as a response to uncertainty.

Under these conditions, the taxpayer sets \( W \) as follows:

\[
W = \begin{cases} 
\alpha T (1 - \frac{k}{j}) & \text{if } N \geq \left(1 - \frac{k}{j}\right) \alpha T, \\
N & \text{if } N < \left(1 - \frac{k}{j}\right) \alpha T.
\end{cases}
\] (19)

Therefore, the probability of receiving a refund given \( N = n \), which we shall denote \( R_n \), is:

\[
R_n = \begin{cases} 
\alpha \left(1 - \frac{k}{j}\right) & \text{if } n \geq \left(1 - \frac{k}{j}\right) \alpha T, \\
n & \text{if } n < \left(1 - \frac{k}{j}\right) \alpha T.
\end{cases}
\] (20)

So, the refund rate can thus be calculated as:

\[
R = \int_0^T R_n \Pr[N = n] dn.
\] (21)

We solve and find:

\[
R = \alpha \left(1 - \frac{k}{j}\right) - \frac{\alpha^2}{2} \left(1 - \frac{k}{j}\right)^2.
\] (22)

This refund rate is lower than under Rule 1 whenever \( 0 < k < j \). Table 2 is analogous to Table 1, and it demonstrates that when we apply Rule 2 and the effective penalty \( (j - k) \), there is a substantial gap between the theoretical and actual refund rates. In fact, on average, the model explains only half of the actual refund rate.

Nonzero Certain Liability

One seemingly innocuous assumption made in both our and HTW’s previous analyses of the refund rate is that taxpayers have no certain tax liability, i.e. \( m = 0 \). Intuitively, this is unrealistic because many taxpayers with stable employment may anticipate having at least some minimum tax liability. Mathematically, the
assumption implies a great deal of uncertainty in tax liability: it implies that the coefficient of variation of
$L$, \( \frac{\sigma_L}{\mu L} \), is almost 58%! In this section, we relax the assumption of zero certain liability, letting \( m \geq 0 \), and
find that it is not innocuous.

We maintain the assumption that the taxpayer’s current and previous year’s liabilities are drawn from the
same distribution. Therefore, \( N \) is uniformly distributed between \( m \) and \( m + T \). Under these assumptions:

\[
\Pr(\text{Area 3}) = \frac{W - m}{T},
\]

\[
\Pr(\text{Area 1}) = \frac{m + T - N}{T}.
\]

By equation (14), the taxpayer sets \( W \) in the following manner:

\[
W = \begin{cases} 
\alpha \left( (1 - \frac{k}{j}) T + \alpha m \right) & \text{if } N \geq \alpha \left( (1 - \frac{k}{j}) T + \alpha m \right), \\
N & \text{if } N < \alpha \left( (1 - \frac{k}{j}) T + \alpha m \right).
\end{cases}
\] (23)

Therefore, the refund rate is:

\[
R = \begin{cases} 
R^o - (1 - \alpha)(1 - \alpha D) \frac{m}{T} - \frac{(1-\alpha)^2}{2} \left( \frac{m}{T} \right)^2 & \text{if } D \geq \frac{(1-\alpha)m}{\alpha T}, \\
0 & \text{if } D < \frac{(1-\alpha)m}{\alpha T}.
\end{cases}
\] (24)

where \( D = \left( 1 - \frac{k}{j} \right) \) and \( R^o \) denotes the refund rate given in equation (22).

We can think of \( \frac{m}{T} \) as the ratio of certain liability to uncertain liability. Figure 6 depicts \( R \) as a function
of \( \frac{m}{T} \) for different values of \( \frac{m}{T} \) when \( \alpha = .9 \). Figure 7 shows the same functions when \( \alpha = .8 \). It is clear from
these figures that as \( \frac{m}{T} \) increases, the probability of receiving a refund decreases significantly. Furthermore,
when \( \alpha = .8 \), the refund rates are considerably lower than when \( \alpha = .9 \).

Table 3 compares actual refund rates to theoretical ones when \( \frac{m}{T} = 2 \) for the period 1983-1992. When \( \frac{m}{T} = 2 \), there is only a 5% chance that the taxpayer’s liability will be more than 19% off of his expected
liability. For most taxpayers, such a large changes in liability are likely uncommon. One possible source for
such a large, unexpected change in liability is involuntary job loss. Letting \( \frac{m}{T} = 2 \) actually aligns reasonably
well with real rates of involuntary job loss: Boisjoly et al. (1998) found that the rate of such job loss averaged
only 3.2% annually for men in the period 1980-1992. Furthermore, the possibility of ending up with much
less liability than expected should provide an incentive for taxpayers to withhold less. Therefore, even if we
understate the probability of a large loss in liability, we only stand to overstate the refund rate. On the
other hand, a taxpayer’s liability may also see a large unexpected increase in liability if he changes or gains
employment unexpectedly. However, neither of these cases pose a significant threat to our analysis: both
are relatively uncommon, and when taxpayers are hired by a new firm, they are asked to file a new Form
W-4 with their employer, allowing them to adjust their withholdings rate appropriately.

The theoretical estimates in Table 3 are considerably lower than those in either Table 1 or Table 2. This
is especially true for the years 1983-1986 in which \( \alpha = .8 \). In fact, for 1983, 1984, and 1986, the theoretical
refund rate is zero. For 1983-1992, the theoretical refund rate is on average only about a quarter of the
actual refund rate, averaging only 19.3%.

### Applying a Normal Distribution

In the previous sections, we assumed a uniform distribution for \( L \) and \( N \). Notably, a uniform distribution
gives no probabilistic preference to liabilities [within the support of \( L \)] that are closer to the taxpayer’s
expected liability over those that are farther away. Or, in other words, the taxpayer’s specific point of
expectation has no meaning: all liabilities in the support, no matter how far or close to the taxpayer’s
expectation are equally probable.

One might instead expect that tax liabilities closer to the taxpayer’s expectation are more probable than
those farther away. In this section, we apply what we believe to be a more realistic assumption, one that
satisfies the aforementioned expectation. Namely, we let \( m = 0, L \sim N(\mu_L, \sigma_L), \) and \( N \sim N(\mu_N, \sigma_N) \) so that taxable income and previous year’s tax liability are normally distributed.\(^{17}\)

Given these distributions, the taxpayer sets \( W \) in the following manner under Rule 2:

\[
W = \begin{cases}
\alpha \left[ \sigma_L \Phi^{-1} \left( 1 - \frac{k}{T} \right) + \mu_L \right] & \text{if } N \geq \alpha \left[ \sigma_L \Phi^{-1} (1 - \frac{k}{T}) + \mu_L \right], \\
N & \text{if } N < \alpha \left[ \sigma_L \Phi^{-1} (1 - \frac{k}{T}) + \mu_L \right],
\end{cases}
\]

(25)

where \( \Phi \) is the CDF of the standard normal distribution. Therefore, the probability of receiving a refund given \( N = n \), which we shall denote \( R_n \), is:

\[
R_n = \begin{cases}
\Phi \left( \frac{\alpha \sigma_L \Phi^{-1} (1 - \frac{k}{T}) - (1 - \alpha) \mu_L}{\sigma_L} \right) & \text{if } n \geq \alpha \left[ \sigma_L \Phi^{-1} (1 - \frac{k}{T}) + \mu_L \right], \\
\Phi \left( \frac{n - \mu_L}{\sigma_L} \right) & \text{if } n < \alpha \left[ \sigma_L \Phi^{-1} (1 - \frac{k}{T}) + \mu_L \right],
\end{cases}
\]

(26)

The refund rate can thus be calculated as:

\[
R = \int_{-\infty}^{\infty} R_n \Pr[N = n] dn.
\]

(27)

We can substitute in \( \Pr[N = n] = \phi \left( \frac{n - \mu_N}{\sigma_N} \right) \), where \( \phi \) is the PDF of a standard normal distribution, and \( R_n \) from equation (26) to find:

\[
R = \int_{-\infty}^{\alpha \sigma_L \Phi^{-1} (1 - \frac{k}{T}) + \mu_L} \Phi \left( \frac{n - \mu_L}{\sigma_L} \right) \phi \left( \frac{n - \mu_N}{\sigma_N} \right) dn 
+ \Phi \left( \frac{\alpha \sigma_L \Phi^{-1} (1 - \frac{k}{T}) - (1 - \alpha) \mu_L}{\sigma_L} \right) \left( 1 - \Phi \left( \frac{\alpha \sigma_L \Phi^{-1} (1 - \frac{k}{T}) + \mu_L}{\sigma_L} - \mu_N \right) \right).
\]

(28)

In order to meaningfully compare the refund rate when liability is normally distributed with that when liability is uniformly distributed, we assume that \( \mu_L = \frac{T}{2} \) and define \( s = \frac{\mu_L}{\sigma_L} \) to be the inverse of the coefficient of variation of \( L \). Thus, \( L \) is normally distributed with the same mean as under the assumption of a uniform distribution, such that \( 2s \) standard deviations fit in \([0, T] \), the support under the assumption of a uniform distribution.

As in the previous section, we assume that \( N \) and \( L \) are identically distributed (i.e. \( \mu_N = \mu_L \) and \( \sigma_N = \sigma_L \)). Also as before, by assuming a stagnant expected earnings profile, we likely understate the frequency with which Rule 2 binds and thus overstate the refund rate. Substituting into equation (28) and changing the variable of integration to \( x = \frac{n - \mu_L}{\sigma_L} = \frac{n - \mu_N}{\sigma_N} \), we find:

\[
R = \int_{-\infty}^{\alpha \Phi^{-1} (1 - \frac{k}{T}) - (1 - \alpha) s} \Phi (x) \phi (x) dx 
+ \Phi \left( \alpha \Phi^{-1} (1 - \frac{k}{T}) - (1 - \alpha) s \right) \left( 1 - \Phi \left( \alpha \Phi^{-1} (1 - \frac{k}{T}) - (1 - \alpha) s \right) \right),
\]

(29)

Figure 8 depicts how the refund rate is affected by different values of \( s \) when \( \alpha = .9 \). Figure 9 shows the same function when \( \alpha = .8 \). As in our analysis of \( R \) when \( m \geq 0 \), we see that the refund rate is considerably lower when \( \alpha = .8 \) than when \( \alpha = .9 \).

We can think of \( s \) as a measure of one’s confidence that his income will be close to its expected value. So, as one’s confidence in the accuracy of one’s prediction increases, the probability of receiving a refund decreases. Figures 8 and 9 show that the normal distribution assumption yields much lower estimates of the refund rate than does the uniform distribution assumption, especially when \( s \) is large. Table 4 gives the theoretical and actual refund rates for 1983-1990 assuming that \( s = 10 \). When \( s = 10 \), the taxpayer perceives only a 5% chance that his liability will be more than 20% off of his expected liability—a very similar case to when \( \frac{m}{T} = 2 \).

\(^{17}\)While it may seem problematic that normal distributions allow for cases in which \( L < 0 \) or \( N < 0 \), these cases are highly marginal in our analysis. The assumptions we make later insure that such cases occur with at most a few percent probability. Furthermore, truncating these distributions only strengthens the result that applying a normal distribution decreases the theoretical refund rate.

\(^{18}\)We thank Christopher Overstreet for his advice on computational methods.
As in the previous section, the theoretical refund rates are only a small fraction of the actual ones. Also as before, this is especially true for the years 1983-1986 in which $\alpha = .8$. For 1983-1992, the theoretical refund rate is on average only about a fifth of the actual refund rate, averaging only 15.6%.

V. Conclusion

In our analysis, we have identified various weaknesses in the proposition of HTW that income tax overwithholding can be explained as a risk neutral taxpayer’s response to uncertain tax liability and the possibility of being penalized for underwithholding. We found that accounting for interest earned on underwithheld income and enforcing boundary conditions decreases the model’s predictions of the taxpayer’s probability of overwithholding and receiving a refund. We also showed that incorporating the previous year’s liability rule further decreases the taxpayer’s probability of receiving a refund. This result is possibly understated because our assumptions about the taxpayer’s previous year’s liability are rather generous, especially given that refunds have been shown to correlate with steeper earning profiles. Additionally, we discovered that relaxing the assumption of zero certain tax liability significantly widens the gap between the theoretical refund rate due to uncertainty and the real refund rate. Finally, we demonstrated that the more realistic approach of using normal distributions also weakens the argument for overwithholding as a response to uncertainty. If anything, our results are likely to be understated, for our methodology was highly conservative.

We must qualify our results, however, because the models presented assume risk neutrality, and a risk averse setting may prove more favorable to the theory of overwithholding as a response to uncertainty. Jones (2012) finds that when the cost to overwithholding from consumption smoothing is taken into account, a very high degree of risk aversion is required to justify withholding patterns seen in the United States. Our findings complement Jones’ result and provide important traction for behavioral analyses of overwithholding. There is still work to be done isolating and determining the degree to which different behavioral phenomena contribute to overwithholding. However, this research may now be performed with the greater confidence that a rational paradigm examination of uncertainty does not adequately explain income tax overwithholding.
References


Mathematical Appendix

Proof of Equations 12 and 13
The taxpayer sets $W$ to minimize:

$$E[PC] = \int_0^U \int_0^V \left((j - k) \left[\alpha \left(t (Y - A) + m\right) - W\right] f(Y, A) dYdAight)$$

$$+ \int_0^U \int_0^{\frac{N}{\alpha t} - \frac{m}{t} + A} k \left[W - \alpha \left(t (Y - A) + m\right)\right] f(Y, A) dYdA.$$ 

The results follow from taking the first order conditions with respect to $W$. A more detailed proof is given of equation \((14)\). The result for equations \((12)\) and \((13)\) also be derived from selecting an arbitrary large $N$.

Proof of Equation 14
The taxpayer sets $W$ to minimize:

$$E[PC] = \int_0^U \int_0^{\frac{N}{\alpha t} - \frac{m}{t} + A} k(W - \alpha t(Y - A) - \alpha m) f(A, Y) dYdA$$

$$+ \int_0^U \int_0^{\frac{N}{\alpha t} - \frac{m}{t} + A} (j - k)(at(Y - A) + am - W) f(A, Y) dYdA$$

$$+ \int_0^U \int_0^{\frac{N}{\alpha t} - \frac{m}{t} + A} (j - k)(N - W) f(A, Y) dYdA,$$

subject to $N \geq W$.

Our Lagrangian is as follows:

$$\mathcal{L} = \int_0^U \int_0^{\frac{N}{\alpha t} - \frac{m}{t} + A} k(W - \alpha t(Y - A) - \alpha m) f(A, Y) dYdA$$

$$+ \int_0^U \int_0^{\frac{N}{\alpha t} - \frac{m}{t} + A} (j - k)(at(Y - A) + am - W) f(A, Y) dYdA$$

$$+ \int_0^U \int_0^{\frac{N}{\alpha t} - \frac{m}{t} + A} (j - k)(N - W) f(A, Y) dYdA - \lambda(N - W).$$

The first order condition with respect to $W$ is:

$$\int_0^U \left[\frac{1}{\alpha t} \left(k[\alpha t - \frac{W}{\alpha t} - t - A] - \alpha m\right) f(A, \frac{W}{\alpha t} - \frac{m}{t} + A) + \int_0^{\frac{N}{\alpha t} - \frac{m}{t} + A} k f(A, Y) dY\right] dA$$

$$+ \int_0^U \left[-\frac{1}{\alpha t} \left(j - k\right)\left[\alpha t - \frac{W}{\alpha t} - t - A\right] + am - W\right] f(A, \frac{W}{\alpha t} - \frac{m}{t} + A)$$

$$- (j - k) \int_0^{\frac{N}{\alpha t} - \frac{m}{t} + A} f(A, Y) dY + (j - k) \int_0^U \int_0^{\frac{N}{\alpha t} - \frac{m}{t} + A} f(A, Y) dYdA + \lambda = 0.$$ 

We note:

$$\frac{1}{\alpha t} \left(k[\alpha t - \frac{W}{\alpha t} - t - A] - \alpha m\right) f(A, \frac{W}{\alpha t} + A) = 0.$$ 

$$-\frac{1}{\alpha t} \left(j[\alpha t - \frac{W}{\alpha t} - t - A] + am - W\right) f(A, \frac{W}{\alpha t} + A) = 0.$$

13
Therefore, the first order condition is:
\[ k \int_0^U \int_0^{W - \frac{m}{A} + A} f(A,Y)dYdA - (j - k) \int_0^U \int_\frac{N - m}{A}^{W - \frac{m}{A} + A} f(A,Y)dYdA \]
\[ - (j - k) \int_0^U \int_{\frac{N}{A}}^V f(A,Y)dYdA + \lambda = 0. \]

This is equivalent to the following:
\[ k \Pr[\text{Case 3}] - (j - k) \Pr[\text{Case 2}] - (j - k) \Pr[\text{Case 1}] = -\lambda. \]

We find consistency when \( N - W > 0 \) and \( \lambda = 0 \). In this case, our first order condition is equivalent to:
\[ k \Pr[\text{Case 3}] - (j - k) \Pr[\text{Case 2}] - (j - k) \Pr[\text{Case 1}] = 0 \]
\[ \iff k \Pr[\text{Case 3}] - (j - k) (\Pr[\text{Case 2}] + \Pr[\text{Case 1}]) = 0 \]
\[ \iff k \Pr[\text{Case 3}] = j - k - (j - k) \Pr[\text{Case 3}] \]
\[ \iff j \Pr[\text{Case 3}] = j - k \]
\[ \iff \Pr[\text{Case 3}] = 1 - \frac{k}{j} \]
\[ \iff \Pr[\text{Case 1}] \leq \frac{k}{j} \]

Additionally, we find the opposite consistency that holds when \( W = N \) and \( \Pr[\text{Case 1}] > \frac{k}{j} \).

It is worth noting that if \( 1 - \frac{k}{j} < \int_0^U \int_0^A f(A,Y)dYdA \), the taxpayer sets \( W = 0 \). This is actually intuitive: if one would optimally like to set \( W < 0 \), one will set \( W = 0 \). This case, however, is ignored in Highfill et al. and this section of our analysis.

\[ \square \]

**Proof of Equation 16:**
Under the assumption of uniformity, we know that the probability of overwithholding is the area of the bottom trapezoid of our graph normalized by the area of the entire rectangular region of positive probability. Thus, using equations for the area of a trapezoid and the area of a rectangle, we see that the optimal level of withholding should be set such that:
\[ U \left( \frac{W}{2} + \frac{W + U}{2} \right) = 1 - \frac{k}{j} \]
This can be rearranged to yield:
\[ W = \alpha \left[ (1 - \frac{k}{j})V - \frac{U}{2} \right]. \]

\[ \square \]

**Proof of Equation 17:**
Since \( W > t(Y - A) \), the condition for receiving a refund, is equivalent to \( Y < \frac{W}{t} - A \), we see that the probability of getting a refund is equivalent to the probability of falling in the region located beneath the line \( Y = \frac{W}{t} - A \). Therefore, by normalizing the area of aforementioned region by the area of the entire rectangular region of positive probability, we find that the probability of receiving a refund is:
\[ R = \frac{U \left( \frac{W}{2} + \frac{W + U}{2} \right)}{UV}. \]
This can be rearranged algebraically to yield:

\[ = \alpha \left( 1 - \frac{k}{j} \right) + \frac{1 - \alpha U}{2} V. \]

By substituting the conditions on \( \frac{U}{V} \) in equation (15), we have:

\[ \alpha(1 - \frac{k}{j}) < R < \alpha(1 - \frac{k}{j}) + (1 - \alpha) \min\left(\frac{k}{j}, 1 - \frac{k}{j}\right). \]

□

**Proof of Equation 22:**

For notational simplicity, we denote \( D = 1 - \frac{k}{j} \). Then, from equation (21):

\[
R = \int_0^T R_n \Pr[N = n]dn \\
= \int_0^D \alpha T \int_0^T R_n \Pr[N = n]dn + \int_D^T \alpha D \int_0^T \Pr[N = n]dn.
\]

We note that \( \Pr[N = n] = \frac{1}{T} \) and substitute from equation (20) and find that:

\[
R = \int_0^D \frac{n}{T^2} \frac{1}{T} dn + \int_D^T \alpha D \frac{1}{T} dn \\
= \frac{D^2 \alpha^2}{2} + \alpha D(1 - \alpha D) = \alpha D - \frac{\alpha^2 D^2}{2}.
\]

By substituting \( D = 1 - \frac{k}{j} \), we have:

\[
R = \alpha \left( 1 - \frac{k}{j} \right) - \frac{\alpha^2}{2} \left( 1 - \frac{k}{j} \right)^2.
\]

□

**Proof of Equation 25:**

Rule 2 does not bind only when

\[
P(L < \frac{N}{\alpha} | N = n) = \Phi \left( \frac{n - \mu_L}{\sigma_L} \right) \geq 1 - \frac{k}{j}
\]

\[ \iff N \geq \alpha \left[ \sigma_L \Phi^{-1} (1 - \frac{k}{j}) + \mu_L \right]. \]

In such a case, the taxpayer sets \( W \) so that:

\[
P(L < \frac{W}{\alpha}) = \Phi \left( \frac{W - \mu_L}{\sigma_L} \right) = 1 - \frac{k}{j}
\]

\[ \iff W = \alpha \left[ \sigma_L \Phi^{-1} (1 - \frac{k}{j}) + \mu_L \right]. \]

□
Proof of Equation 28: We see that:

\[
R = \int_{-\infty}^{\alpha \{\sigma L \Phi^{-1}(1-\frac{k}{2})+\mu L\}^}\Phi\left(\frac{n-\mu L}{\sigma L}\right)\phi\left(\frac{n-\mu N}{\sigma N}\right)dn
\]

\[
+ \int_{\alpha \{\sigma L \Phi^{-1}(1-\frac{k}{2})+\mu L\}^\infty} \Phi\left(\frac{\alpha \sigma L \Phi^{-1}\left(1-\frac{k}{T}\right) - (1-\alpha)\mu L}{\sigma L}\right)\Phi\left(\frac{n-\mu L}{\sigma L}\right)dn.
\]

Equation (28) then follows by pulling the CDF out of the second integral.

Proof of Equation 24:

Note that if \( W < m \), \( R = 0 \). This can only happen if \( \alpha (DT + m) < m \), which is when \( D < \frac{(1-\alpha) D T}{m} \). For the following analysis, we assume that \( D \) is sufficiently large when compared to \( \frac{m}{T} \) that \( W \geq m \).

Then, when \( N = n \), we have that the refund rate \( R_n \) can be given as:

\[
R_n = \begin{cases} 
\alpha D - (1-\alpha) \frac{m}{T} & \text{if } n \geq \alpha (DT + m), \\
\frac{n-m}{T} & \text{if } n < \alpha (DT + m).
\end{cases}
\]

Substituting in \( n' = n - m \), we have that:

\[
R = \int_{m}^{m+T} R_n \Pr[N = n]dn = \int_{0}^{\alpha D T - (1-\alpha)m} \frac{n'}{T^2}dn' + \int_{\alpha D T - (1-\alpha)m}^{T} \left(\frac{\alpha D}{T} - (1-\alpha) \frac{m}{T^2}\right)dn'
\]

\[
= \frac{1}{2T^2} (\alpha DT - (1-\alpha)m)^2 + \left(1 - \alpha D + (1-\alpha) \frac{m}{T}\right) \left(\alpha D - (1-\alpha) \frac{m}{T}\right)
\]

\[
= \frac{1}{2} \left(\alpha D - (1-\alpha) \frac{m}{T}\right)^2 + \alpha D - (1-\alpha) \frac{m}{T} - \left(\alpha D - (1-\alpha) \frac{m}{T}\right)^2
\]

\[
= \alpha D - (1-\alpha) \frac{m}{T} - \frac{1}{2} \left(\alpha^2 D^2 - 2\alpha(1-\alpha) D \frac{m}{T} + (1-\alpha)^2 \left(\frac{m}{T}\right)^2\right)
\]

\[
= R - (1-\alpha)(1-\alpha) \frac{m}{T} = \frac{(1-\alpha)^2}{2} \left(\frac{m}{T}\right)^2.
\]

\[\square\]
## Tables and Figures

### Table 1: HTW’s Actual and Theoretical Refund Rates, 1983-1992

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual Refund Rate</th>
<th>Underwithholding Penalty</th>
<th>Opportunity Cost</th>
<th>Theoretical Refund Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>77.9</td>
<td>10.3</td>
<td>5.4</td>
<td>59.0-64.0</td>
</tr>
<tr>
<td>1984</td>
<td>76.1</td>
<td>6.7</td>
<td>5.2</td>
<td>50.7-55.7</td>
</tr>
<tr>
<td>1985</td>
<td>76.1</td>
<td>8.4</td>
<td>3.9</td>
<td>61.5-66.5</td>
</tr>
<tr>
<td>1986</td>
<td>75.8</td>
<td>7.6</td>
<td>4.1</td>
<td>58.5-63.5</td>
</tr>
<tr>
<td>1987</td>
<td>74.6</td>
<td>5.6</td>
<td>2.2</td>
<td>64.6-69.6</td>
</tr>
<tr>
<td>1988</td>
<td>72.1</td>
<td>6.4</td>
<td>2.6</td>
<td>64.0-69.0</td>
</tr>
<tr>
<td>1989</td>
<td>72.2</td>
<td>6.7</td>
<td>3.3</td>
<td>60.3-65.3</td>
</tr>
<tr>
<td>1990</td>
<td>73.4</td>
<td>5.6</td>
<td>2.1</td>
<td>65.5-70.5</td>
</tr>
<tr>
<td>1991</td>
<td>74.6</td>
<td>6</td>
<td>1.2</td>
<td>75.0-80.0</td>
</tr>
<tr>
<td>1992</td>
<td>71.3</td>
<td>5</td>
<td>0.4</td>
<td>83.3-88.3</td>
</tr>
<tr>
<td>Mean</td>
<td>74.4</td>
<td>6.8</td>
<td>3</td>
<td>64.2-69.2</td>
</tr>
<tr>
<td>SD</td>
<td>2.02</td>
<td>1.5</td>
<td>1.56</td>
<td></td>
</tr>
</tbody>
</table>

1 This table is replicated from HTW. Refer to Table 1 of HTW for details.

### Table 2: Actual and Theoretical Refund Rates Recalculated, 1983-1992

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual Refund Rate</th>
<th>Underwithholding Penalty</th>
<th>Opportunity Cost</th>
<th>Theoretical Refund Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>77.9</td>
<td>10.3</td>
<td>5.4</td>
<td>30.8</td>
</tr>
<tr>
<td>1984</td>
<td>76.1</td>
<td>6.7</td>
<td>5.2</td>
<td>16.3</td>
</tr>
<tr>
<td>1985</td>
<td>76.1</td>
<td>8.4</td>
<td>3.9</td>
<td>33.7</td>
</tr>
<tr>
<td>1986</td>
<td>75.8</td>
<td>7.6</td>
<td>4.1</td>
<td>30.1</td>
</tr>
<tr>
<td>1987</td>
<td>74.6</td>
<td>5.6</td>
<td>2.2</td>
<td>39.7</td>
</tr>
<tr>
<td>1988</td>
<td>72.1</td>
<td>6.4</td>
<td>2.6</td>
<td>39.2</td>
</tr>
<tr>
<td>1989</td>
<td>72.2</td>
<td>6.7</td>
<td>3.3</td>
<td>35.2</td>
</tr>
<tr>
<td>1990</td>
<td>73.4</td>
<td>5.6</td>
<td>2.1</td>
<td>40.4</td>
</tr>
<tr>
<td>1991</td>
<td>74.6</td>
<td>6.0</td>
<td>1.2</td>
<td>46.1</td>
</tr>
<tr>
<td>1992</td>
<td>71.3</td>
<td>5.0</td>
<td>0.4</td>
<td>48.5</td>
</tr>
<tr>
<td>Mean</td>
<td>74.4</td>
<td>6.8</td>
<td>3.0</td>
<td>36.0</td>
</tr>
<tr>
<td>SD</td>
<td>2.02</td>
<td>1.5</td>
<td>1.56</td>
<td>9.18</td>
</tr>
</tbody>
</table>
### Table 3: Actual and Theoretical Refund Rates When \( \pi = 2 \), 1983-1990

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual Refund Rate</th>
<th>Underwithholding Penalty</th>
<th>Overwithholding Opportunity Cost</th>
<th>Theoretical Refund Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>77.9</td>
<td>10.3</td>
<td>5.4</td>
<td>0</td>
</tr>
<tr>
<td>1984</td>
<td>76.1</td>
<td>6.7</td>
<td>5.2</td>
<td>0</td>
</tr>
<tr>
<td>1985</td>
<td>76.1</td>
<td>8.4</td>
<td>3.9</td>
<td>2.8</td>
</tr>
<tr>
<td>1986</td>
<td>75.8</td>
<td>7.6</td>
<td>4.1</td>
<td>0</td>
</tr>
<tr>
<td>1987</td>
<td>74.6</td>
<td>5.6</td>
<td>2.2</td>
<td>28.6</td>
</tr>
<tr>
<td>1988</td>
<td>72.1</td>
<td>6.4</td>
<td>2.6</td>
<td>27.9</td>
</tr>
<tr>
<td>1989</td>
<td>72.2</td>
<td>6.7</td>
<td>3.3</td>
<td>22.4</td>
</tr>
<tr>
<td>1990</td>
<td>73.4</td>
<td>5.6</td>
<td>2.1</td>
<td>29.7</td>
</tr>
<tr>
<td>1991</td>
<td>74.6</td>
<td>6.0</td>
<td>1.2</td>
<td>38.5</td>
</tr>
<tr>
<td>1992</td>
<td>71.3</td>
<td>5.0</td>
<td>0.4</td>
<td>43.1</td>
</tr>
<tr>
<td>Mean</td>
<td>74.4</td>
<td>6.8</td>
<td>3.0</td>
<td>19.3</td>
</tr>
<tr>
<td>SD</td>
<td>2.02</td>
<td>1.5</td>
<td>1.56</td>
<td>17.00</td>
</tr>
</tbody>
</table>

### Table 4: Actual and Theoretical Refund Rates When \( s = 10 \), 1983-1990

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual Refund Rate</th>
<th>Underwithholding Penalty</th>
<th>Overwithholding Opportunity Cost</th>
<th>Theoretical Refund Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>77.9</td>
<td>10.3</td>
<td>5.4</td>
<td>2.0</td>
</tr>
<tr>
<td>1984</td>
<td>76.1</td>
<td>6.7</td>
<td>5.2</td>
<td>0.05</td>
</tr>
<tr>
<td>1985</td>
<td>76.1</td>
<td>8.4</td>
<td>3.9</td>
<td>2.6</td>
</tr>
<tr>
<td>1986</td>
<td>75.8</td>
<td>7.6</td>
<td>4.1</td>
<td>1.8</td>
</tr>
<tr>
<td>1987</td>
<td>74.6</td>
<td>5.6</td>
<td>2.2</td>
<td>20.0</td>
</tr>
<tr>
<td>1988</td>
<td>72.1</td>
<td>6.4</td>
<td>2.6</td>
<td>19.3</td>
</tr>
<tr>
<td>1989</td>
<td>72.2</td>
<td>6.7</td>
<td>3.3</td>
<td>15.0</td>
</tr>
<tr>
<td>1990</td>
<td>73.4</td>
<td>5.6</td>
<td>2.1</td>
<td>21.0</td>
</tr>
<tr>
<td>1991</td>
<td>74.6</td>
<td>6.0</td>
<td>1.2</td>
<td>32.3</td>
</tr>
<tr>
<td>1992</td>
<td>71.3</td>
<td>5.0</td>
<td>0.4</td>
<td>42.2</td>
</tr>
<tr>
<td>Mean</td>
<td>74.4</td>
<td>6.8</td>
<td>3.0</td>
<td>15.6</td>
</tr>
<tr>
<td>SD</td>
<td>2.02</td>
<td>1.5</td>
<td>1.56</td>
<td>14.28</td>
</tr>
</tbody>
</table>
Figure 1: HTW’s Taxpayer’s Decision Under Rule 1 (Replicated from HTW)

\[
\begin{align*}
\text{Underwithholding} & \quad \text{Probability} = \frac{1}{1 + \alpha} \\
\text{Overwithholding} & \quad \text{Probability} = \frac{1}{1 + \beta}
\end{align*}
\]
Figure 2: HTW’s Taxpayer’s Decision Under Rule 2 (Replicated from HTW)

Area 1:
Penalty = \( j(N - W) \)

Area 2:
Penalty = \( j(\alpha t(Y - A) - W) \)

Area 3:
Cost = \( j(\alpha t(Y - A) - W) \)

---

“No Penalty-Cost Line”

“Rules Equal Line”
Figure 3: Boundary Conditions ($m = 0$)

Figure 4: $f(L)$ Under HTW’s Assumptions

\[ f(L) \]

\[ \frac{W_{\min}}{\alpha t} = 0 \quad \frac{W}{\alpha} = 0 \quad \frac{W_{\max}}{\alpha} = t(V - U) \]
Figure 5: Uniformly Distributed Liability \((m = 0)\)

\[ f(L) \]

\[ T = t(V - U)^L \]

Figure 6: Refund Rates When \(m \geq 0 \quad (\alpha = .9)\)

\[
\begin{array}{c c c c c}
\text{Refund Rate} & \text{Refund Rate} & \text{Refund Rate} & \text{Refund Rate} & \text{Refund Rate} \\
0 & .2 & .4 & .6 & .8 \\
.1 & .3 & .5 & .7 & .9 \\
\end{array}
\]

\[
\begin{array}{c c c c c}
\text{Refund Rate} & \text{Refund Rate} & \text{Refund Rate} & \text{Refund Rate} & \text{Refund Rate} \\
0 & .2 & .4 & .6 & .8 \\
.1 & .3 & .5 & .7 & .9 \\
\end{array}
\]
Figure 7: Refund Rates When $m \geq 0$ ($\alpha = .8$)

Figure 8: Uniform Distribution Versus Normal Distribution ($\alpha = .9$)
Figure 9: Uniform Distribution Versus Normal Distribution ($\alpha = .8$)