Rational Illiquidity and Excess Sensitivity: Theory and Evidence from Income Tax Withholding and Refunds*

Michael Gelman, Shachar Kariv, Matthew D. Shapiro, and Dan Silverman †

January 16, 2019

Abstract

Nearly a third of all personal income tax collected by the U.S, government is later returned in the form of tax refunds; and households tend to spend disproportionately from those refunds. This paper develops a theory of liquid assets management that explains why households voluntarily reduce liquidity by overwitholding, but then spend in response to the liquidity provided by tax refunds. Liquidity constraints that arise endogenously when income is uncertain and when adjusting tax payments is not frictionless explain these behaviors. Tax refunds tend to arrive in circumstances where income is lower than expected, so liquidity is low and the MPC is endogenously high. The paper evaluates the empirical relevance of the theory with account data on income, refunds, and spending. The average amount of income not subject to withholding and the average annual fluctuations in that income are more than sufficient to explain the size of average tax refunds, and microevidence supports central mechanisms of the model.

JEL Classification Numbers: D14, H24, E21

Keywords: Income uncertainty, liquidity, tax withholding, account data

*We thank Damon Jones and participants at many seminars for helpful comments. This research is supported by a grant from the Alfred P. Sloan Foundation. Shapiro acknowledges additional support from the Michigan node of the NSF-Census Research Network (NSF SES 1131500).

†Gelman: Claremont McKenna College (mgelman@cmc.edu); Kariv: University of California, Berkeley (kariv@berkeley.edu); Shapiro: University of Michigan and NBER (shapiro@umich.edu); Silverman: Arizona State University and NBER (dsilver3@asu.edu).

1 Introduction

Nearly a third of all personal income tax collected by the US government is later returned in the form of tax refunds. Yet, households tend to spend disproportionately from those refunds when they arrive. Why do households make substantial, voluntary, interest-free loans to the government in the form of overwithholding, but then rapidly spend a high fraction of the loan repayments when they receive them in the form of tax refunds? These two phenomena represent puzzles for standard theory, which makes strong predictions about no-arbitrage and consumption smoothing. The refunds amount to repayment of a zero-interest loan from the taxpayer to the government, a loan the taxpayer could arbitrage with any interest-bearing account. Workhorse models predict that spending should respond to refunds as it would any other form of transitory income. But several studies estimate economically important "excess sensitivity" of spending to federal tax refunds (Souleles (1999), Baugh et al. (2018), Gelman (2018)).

This paper develops and evaluates a general theory of household liquid assets management with income shocks that both accounts for the prevelance of income tax refunds and the excess sensitivity of spending to the arrival of refunds. The theory preserves standard predictions of no-arbitrage and consumption smoothing but, different from prior work, takes account of the volatility of both paycheck and non-paycheck income, the latter not generally subject to tax withholding at the source. A central mechanism of the theory is the endogenous liquidity constraints that emerge as households manage annual fluctuations in non-paycheck income. The theory predicts large refunds, on average, when non-paycheck income represents a sufficiently large and unpredictable fraction of total income. The theory also predicts that these refunds will tend to arrive when cash-on-hand is low, and thus the marginal propensity to spend is high—thereby explaining what appears to be excess sensitivity of spending from tax refunds.

We evaluate the empirical relevance of the theory with administrative account data on income, spending, and refunds. These data show, with a standard calibration of the other parameters of the model, that the average amount of non-paycheck income and the average annual fluctuations in that income are sufficient to explain the size of average tax refunds. The micro data also provide evidence consistent with the central mechanisms of the theory. First, the micro evidence shows that the fraction of annual income that is not subject to withholding has an economically and statistically significant positive relationship with tax refunds. Second, as the model indicates, those whose non-paycheck income shares predict larger refunds also have larger marginal propensities to consume out tax refunds. The estimated magnitude of this excess sensitivity is quite similar to the predictions of the calibrated model.

The theory builds on insights from precautionary savings models. We assume that taxes on paycheck income are withheld at the source and, if annual fluctuations in this income were the only form of uncertainty, tax withholding would exactly match tax liability. In this special case, optimal behavior implies no tax refunds and no additional tax payments; as usual, saving would occur strictly in the household's private financial accounts. Non-paycheck income can, however, lead the household to elect additional tax withholding or estimated tax payments. We assume decisions about this additional withholding are made just once per year and before the realization of all income uncertainty. In principle, households need not withhold additional taxes, they could just save, and use that savings to pay any tax liability on non-paycheck income when taxes are due. Doing so is costly, however, because the IRS imposes penalties and interest for being under-withheld that drive a wedge between the private returns to saving and the returns to "saving" in the form of income tax withholding. When non-paycheck income is uncertain, precautionary motives combine with the return-on-savings wedge to produce over-withholding on average.

The endogenous liquidity constraints that emerge from precautionary savings motives also explain the excess sensitivity of spending to the arrival tax refunds. As is standard, in the model households choose to maintain cash on hand in order to smooth consumption as income fluctuates. Households balance the costs of accumulating a large buffer of cash on hand with the costs of more variable consumption. Optimal balancing implies that negative income shocks will meaningfully lower cash on hand and, due to the concavity of the consumption function (Zeldes (1989), Carroll and Kimball (1996)), raise the marginal propensity to consume out of income. These negative income shocks are, however, precisely the events that produce tax refunds: Lower than expected income means the household is overwithheld. The refund arrives when, due to a negative income shock in the previous period, cash on hand is low and the marginal propensity to consume is high. Optimal behavior thus will appear to display "excess" sensitivity.

The empirical relevance of this theory depends on several factors including the level and variability of non-paycheck income. Using a panel of individual-level, bank and credit card records for approximately 880,000 users over 4 years, we isolate non-paycheck income and find that it varies substantially at annual frequency. In this population, non-paycheck income averages $38,764 compared to $68,226 for paycheck income; but non-paycheck incomes has an average standard deviation of $19,879, compared to $18,490 for paycheck income. Simulating an otherwise standard calibration of the model with these income inputs, we find the average tax refund is $1,846—close to the average in the data. It is even higher when we account for a mechanical tendency to overwithhold caused by within-year variation in paycheck income and the convexity of the income tax schedule.

The same data allow us to present micro evidence consistent with the basic mechanisms of the model. Specifically, the data show that refunds decline with the share of total income that is received via paycheck; a 25 percentage point increase in the paycheck income share is associated with a 14% smaller refund. We also find evidence that non-paycheck income volatility is associated with higher refunds. The micro data show that those whose non-paycheck income shares predict larger refunds have larger marginal propensities to consume out tax refunds. The marginal propensity to consume out of refunds increases monotonically with refund size. As predicted by the model, the MPC out of tax refunds rises by approximately 30% from the bottom to the top quintile of the refund distribution.

The prevalence of income tax refunds and the evidence of excess sensitivity of spending from tax refunds may be explained by several mechanisms. Different from past research on the topic, this paper points to income uncertainty and precautionary savings motives as important and linked drivers behind these behaviors. The paper thus joins a growing body of evidence revealing the importance of liquid assets management for understanding how households respond to income and spending shocks.

1.1 Related Literature

1.1.1 On Over-withholding

Jones (2012) generalizes a theory of over-withholding based on the logic in Highfill, Thorson and Weber (1998). Both papers consider the "timing problem" we study: workers must choose their levels of withholding before knowing, for sure, what their incomes and tax liabilities will be. Highfill, Thorson and Weber (1998) explain over-withholding as the optimal response to the wedge between the opportunity costs of over-withholding and the IRS penalties of being under-withheld. Jones (2012) determines that this wedge is insufficient to justify the prevalence of over-withholding; based on tax liability uncertainty alone, he finds that the risk aversion necessary to justify large refunds is implausibly high. Jones (2012) adds adjustment costs to the model of uncertain tax liability and finds empirical support both for those adjustment costs and for their role in determining overwithholding. Our model predicts large refunds without inertia or defaults biased toward over-withholding because it takes explicit account of the volatility of income not subject to withholding.

Alternative explanations interpret over-withholding as a form of forced savings that helps workers deal with problems of self-control. Thaler (1994), Neumark (1995), and Fennell (2006) see over-withholding as an active choice to avoid the daily temptation to spend all that remains from a paycheck. Jones (2012) formalizes these ideas with a quasi-hyperbolic discounting model and finds that it, too, fails to account quantitatively for the observed level of over-withholding.[1] By incorporating volatile income not subject to withholding at the source, we find that a model of workers with time-consistent preferences can account quantitatively for the large average refunds observed in the data.

1.1.2 On Excess Sensitivity

The excess sensitivity of spending to the arrival of tax refunds has been documented by Souleles (1999), Gelman (2018), and Baugh et al. (2018). In these and related papers, the refund is (implicitly) understood as transitory but anticipated income. In the absence of liquidity constraints, standard models predict that the marginal propensity to consume from well-anticipated income is zero. If refunds are interpreted as fully anticipated, transitory income then the spending response is evidence of binding liquidity constraints, or worker myopia, or of a form of choice imperfection. This is consistent with the interpretation of excess sensitivity of spending to other forms of anticipated income such as the Alaska Fund payments (Hsieh (2003), Kueng (2018)) or predictable changes in payroll taxes (Parker (1999)).

Adopting the interpretation of tax refunds as fully anticipated, the model we develop below explains excess sensitivity as the result of endogenous liquidity constraints. It is optimal in the model for workers to have relatively little cash on hand when they receive a (large) tax refund. When the refund arrives, workers are effectively liquidity constrained and their marginal propensity to consume, even from anticipated income, is positive.

An alternative view is that, for purposes of evaluating standard models of consumption smoothing, tax refunds are not fully anticipated. In recent years, electronic tax returns could be filed no earlier than mid-January, and 90% of refunds arrive within 21 days of

[1] Rees-Jones (2018) provides evidence of tax liability bunching just to the right of zero and shows how a model of loss aversion can explain why taxpayers seek to avoid making additional payments at the time of tax filing.

filing.[2] Baugh et al. (2018) report that the average refund arrives 11 days after the tax return was filed. If, therefore, workers remain uncertain of the extent of their refund until the date of filing, the delay between learning about the size of the refund and receiving it is often so short that it can be interpreted as at least partially unanticipated.

If tax refunds are interpreted as transitory but unanticipated income, then the benchmark for defining "excess" sensitivity of the spending response could either be the certainty equivalent benchmark of Hall (1978) for the case of quadratic utility, or simply the average marginal propensity to consume from any transitory income shock. In this case, the puzzle is not why spending responds to the arrival of the tax refunds, but why it responds more than it does to other transitory and imperfectly anticipated forms of income.

1.1.3 On Liquidity and Consumption

The paper's focus on liquid asset management links it to a growing literature that studies, often using innovative data sources, the relationships between liquidity and consumption. Influential examples include Braxton et al. (2018), which links administrative employment and credit bureau data to study consumption smoothing during unemployment, and Herkenhoff (2018), which uses several data sources to show increasing access to credit led to an increased ability to smooth consumption during unemployment.

Given our focus on higher income workers, the paper also relates to Kaplan, Violante and Weidner (2014) and Kaplan and Violante (2014) who document the "wealthy hand-to-mouth;" households who are relatively high net worth but hold few liquid assets. Kaplan and Violante (2014) model this phenomenon by allowing for a higher yielding but less liquid asset and show how optimally low liquid asset holdings can induce a strong spending response to income changes even among higher income households. Our model does not include illiquid assets but focuses attention on the management of liquid cash on hand and its influence on the marginal propensity to consume from transitory income.

In this way, the paper is also related is the literature testing the local concavity of the consumption function. Using surveys, Christelis et al. (2017), Bunn et al. (2018), and Fuster, Kaplan and Zafar (2018) examine how spending responds to hypothetical increases and decreases in income. Baugh et al. (2018) use transactions data to test the asymmetric spending responses to tax refunds and tax payments.

[2]There is an exception for refunds for households receiving the Earned Income Tax Credit (EITC). To combat fraud, the 2015 Protecting Americans from Tax Hikes Act precludes the IRS from issuing refunds that include an EITC before February 15.

Our analysis develops another important implication of the concavity of the consumption function. In our model, negative and positive income shocks move individuals along the consumption function while also influencing their tax refund. The refunds serve as both an indicator of the magnitude of the income shocks a worker faced and as an instrument with which to estimate the spending response. Our model is thus consistent with the finding in Baugh et al. (2018) that spending reacts less to a tax payment than a tax refund but provides a novel mechanism underlying this asymmetry. Tax payments result from positive shocks that increase liquidity and tax refunds result from negative shocks that decrease liquidity. The endogenous constraints that bind when tax refunds arrive lead to larger spending responses relative to when tax payments are made.

2 Institutions and Model

2.1 Background on Tax Withholding

Federal income tax liability is determined at annual frequency and taxes on salary and wages are usually withheld at the source. The schedule for withholding at the source is determined by the frequency of the paycycle, by the number of "allowances" a worker takes on the W-4 form, and by any additional withholding an individual elects to take on the W-4. The IRS provides guidelines to workers on how many allowances to take.

Under some circumstances, following the IRS guidelines for allowances results in withholding that very closely matches a worker's tax liability. The withholding schedule assumes, with exceptions for bonuses, that each paycheck is prorated annual income. On a bi-weekly pay schedule, for example, the withholding schedule for a paycheck of $2,000 assumes annual earnings of $52,000. Allowances on the W-4 are designed to mimic the effects of tax exemptions, deductions, and credits in the federal income tax code; they function to adjust the level of earnings in each paycycle subject to withholding. The IRS guidelines recommend allowances depending on family structure, employment and tax filing status, total income level, and other information. If taxable income were derived only from earnings subject to withholding, and if those earnings exhibited little within-year variation, following the allowances guidelines would result in withholding that very closely matches ultimate tax liability. The worker would owe no additional income taxes and would receive no income tax refund.

Simple adherence to the allowances guidelines is, however, unlikely to result in accurate withholding if the worker also receives income from any of several important sources. Independent contractor income, (self-employed) business or partnership income, capital income, and pension disbursements, are typically not subject to withholding at the source. To avoid underwithholding of taxes on these sources of income, additional taxes must be paid directly, or from income that is subject to withholding.[3,4]

In addition, even if a worker only receives income that is subject to withholding at the source, within-year variation in that income could also lead to overwithholding when a household adheres to the W-4 guidelines. This "mechanical effect" derives from the fact that the income tax schedule is convex and the withholding schedule treats each paycheck as prorated annual income.[5] Suppose, for example, that on alternating paydays a worker receives a small and then a large paycheck. Now suppose withholding from the small paychecks is appropriate for average tax rate τ while withholding from the large paycheck is appropriate for an average tax rate $\tau' > \tau$. In this case, if the average tax rate on annual income is strictly less than τ', adhering to the W-4 guidelines will leave the worker overwithheld. The details of this mechanical effect of high frequency variation are described in the Appendix and evaluated empirically in Section 4.

Individuals with taxable income must file a tax return, or a request for an extension, by a mid-April deadline. Assuming the return is filed on time, there are costs assessed for underwithholding.[6] The costs take two forms. First, there is a penalty of 0.5% of the unpaid tax, and this penalty is assessed again every month that the remaining tax goes unpaid. Thus, if a taxpayer underwithheld by $5,000 and remitted those unpaid taxes on April 15, he would owe and additional $25 in penalties.[7] If, instead, he remitted those

[3] If the tax liability on these other sources of income is sufficiently large relative to total tax liability, then estimated taxes must be paid quarterly. If those estimated taxes are not paid on time, then late penalties and interest apply. We abstract from these late penalties. Estimated taxes may also be paid by increasing withholding on paycheck income, in which case payments are deemed to be paid throughout the year regardless of the timing of the extra withholding.

[4] The influence of these other sources of income on tax liability may be simple and direct if they do not result in a change in the marginal tax bracket. If income not subject to withholding results in an increase in the individual's marginal tax bracket, then it also results in underwithholding on income that is withheld at the source.

[5] We are grateful to Damon Jones for highlighting this mechanical effect for us.

[6] There are relatively large penalties for late filing. Tax payers therefore face unambiguous incentives to file on time, even if they have substantial unpaid taxes. We will abstract from any material incentives to file late and focus on the behavior of timely income tax filers.

[7] Safe harbor provisions exempt households from penalties (but not interest) for underwithholding. Specifically, no penalty applies in the unpaid amount represents less than 10% of total taxes owed in current tax year, or withholding equals the liability in the previous year (110% of the liability over certain income thresholds). For simplicity the model developed below abstracts from these safe harbor provisions.

unpaid taxes only on September 15, he would owe \$150 in penalties. Second, the unpaid tax is also subject to interest at the federal "short-term" interest rate plus 3%. So this same taxpayer, if he remitted his unpaid tax on April 15, would owe about 3.5% interest (using the 0.5% short-term interest rate during the time period of this study) on \$5,000 or \$175. If he remitted those unpaid taxes only on September 15, he would owe approximately \$280 in interest. Thus, a taxpayer that underwithheld by \$5,000 and remitted those unpaid taxes on April 15 would owe approximately \$200 in penalities and interest. If he paid only on September 15, we would owe \$330 in penalties and interest.

2.2 Model

To evaluate the incentives to overwithhold and to spend differentially from tax refunds, we study a model in which time is discrete, the horizon is infinite, and a worker's preferences over period t consumption, C_t, are represented by $u(C_t)$. Income comes in two forms: "paycheck" income, Y_t, that is subject to withholding at the source, and "other" income, N_t, that is not subject to withholding at the source. To simplify analysis, we assume paycheck income is deterministic, arrives at the beginning of each period, and is available for the worker to spend in the current period. Other income is stochastic and is realized at the end of each period; it is available to be spent only in the next period. This timing captures, in a simple way, the primary friction in the model: that withholding decisions must be made before the resolution of all income uncertainty. If this were not the case, then upon the resolution of uncertainty workers would withhold the correct amount of taxes and it would never be optimal to overwithhold. As in the actual income tax system, we assume taxes on period t income are due at the beginning of period $t + 1$.

The state variable for the worker is "cash on hand," X_t, and consists of the current period's paycheck income, plus the previous period's other income, savings, and net tax liability. The net tax liability is negative if the worker is underwithheld, positive if overwithheld. We assume the worker pays a negative liability when it is due. Given cash on hand, the worker chooses how much to save, and how much (more) income to withhold to pay a tax liability that will come due next period. The remainder of the worker's cash on hand is consumed in period t. At the end of the period, other income, N_t, is realized, and cash on hand for the subsequent period is determined.

Formally, cash on hand, X_t, evolves according to

$$X_{t+1} = \left(X_t - C_t - W(Y_t - a) - \widehat{W}_t\right) \cdot R + Y_{t+1} + N_t - T_t$$

where $R > 1$ is the rate of return on savings, T_t is the net tax liability based on period t income. In this formulation a is a predetermined allowances level such that, $W\left(Y - a\right) = \tau \cdot Y$, where τ is the worker's average tax rate. We assume, that is, if all income were paycheck income then the withholding for allowances a would just match the tax liability. Finally, $\widehat{W}_t \geq 0$ is the additional income tax withholding chosen by the worker. Equivalently, $\widehat{W}_t$ can be estimated tax payments, which are also dollar amounts, not functions of current income, and like withholding are presumed to be determined in advance of the realization on other income.

Recall that T_t depends on other income and is therefore realized only after period t consumption is completed. That net tax liability satisfies

$$T_t = \tilde{\phi} \cdot [\tau \cdot (Y_t + N_t) - W(Y_t - a) - \widehat{W}_t]$$

where $\tilde{\phi}$ is a function that captures the costs of being underwithheld and satisfies

$$\tilde{\phi} = \begin{cases} 1 + \phi, & \text{if } \tau \cdot (Y_t + N_t) > W\left(Y_t - a\right) - \widehat{W}_t \\ 1 & \text{if } \tau \cdot (Y_t + N_t) \leq W\left(Y_t - a\right) - \widehat{W}_t \end{cases}$$

for $\phi > 1$. In words, if the worker owes more in taxes on period t income than what she had withheld in period t, then she pays a penalty in the next period of ϕ times the difference. If, instead, she owes less in taxes than what she had withheld then in the next period she receives the difference as a tax refund.

The value to the worker of state X_t is then given by

$$\begin{aligned} V(X_t) &= \max_{S_t, \widehat{W}_t} u(X_t - W(Y_t - a) - \widehat{W}_t - S_t) + \beta \int_N V(X_{t+1}) \qquad (1) \\ s.t.\ C_t &= X_t - W(Y_t - a) - \widehat{W}_t - S_t \\ X_{t+1} &= S_t R + Y_{t+1} + N_t - T_t \\ T_t &= \tilde{\phi} \cdot [\tau \cdot (Y_t + N_t) - W(Y_t - a) - \widehat{W}_t] \\ C_t, \widehat{W}_t, S_t &\geq 0 \end{aligned}$$

Note that period t savings is constrained to be non-negative. We do not consider the possibility of borrowing.

2.3 Optimality

The prevalence of tax refunds is puzzling because overwithholding income taxes amounts to a zero-interest loan to the government, a loan that can be arbitraged with any interest-bearing account. The "excess sensitivity" of spending to the arrival of refunds is puzzling because, in the absence of liquidity constraints, spending from it should either be smooth (if the income is well-anticipated) or like the response to any imperfectly anticipated and transitory income. The optimality conditions for problem (1) reveal the incentives to over-withhold and a common source driving both refunds and excess sensitivity of spending from those refunds.

2.3.1 Optimal Refunds

The incentive to overwithhold can be seen in the tradeoff between allocating the marginal dollar to private savings (S_t) or to additional withholding ($\hat{W}_t$). The first-order condition for saving is

$$u'(C_t) \geq \beta R \int_N V'(X_{t+1}) \tag{2}$$

with equality if $S_t > 0$. Written in terms of consumption, (2) becomes the standard consumption Euler equation, that is,

$$u'(C_t) \geq \beta R \int_N u'(C_{t+1}). \tag{3}$$

Deriving (3) from (2) makes use of the envelope theorem. At the optimum, the worker balances the cost of saving another dollar—the marginal utility of current consumption—against the discounted expected benefit of that saving—the rate of return times the expected marginal utility of next period's consumption. The expectation is with respect uncertain other income, N_t.

The optimality condition for additional withholding is like the one for saving. It is

$$u'(C_t) \geq \beta \int_N V'(X_{t+1}) \tilde{\phi} \tag{4}$$

or, parallel to the consumption Euler equation (3),

$$u'(C_t) \geq \beta \int_N u'(C_{t+1})\tilde{\phi}. \tag{5}$$

Again, both hold with equality if $\hat{W}_t > 0$.

Two features distinguish the optimality condition for additional withholding from that for saving. First, there is no rate of return R inflating the benefit side of the withholding equation (5); here $R = 1$. This distinction reflects the potential arbitrage opportunity; setting aside penalties for being underwithheld, "saving" in the form of income tax withholding is suboptimal for any $R > 1$. The reason this arbitrage opportunity does not always obtain is because of the second distinguishing feature of the optimality condition for additional withholding, the $\tilde{\phi}$ term on the marginal utility of next period's consumption. The $\tilde{\phi}$ term belongs inside the integration because it equals 1 when the realization of N_t is sufficiently low that the worker is overwithheld, and equals $\phi > R$ when the realization of N_t is sufficiently high that the worker is underwithheld.

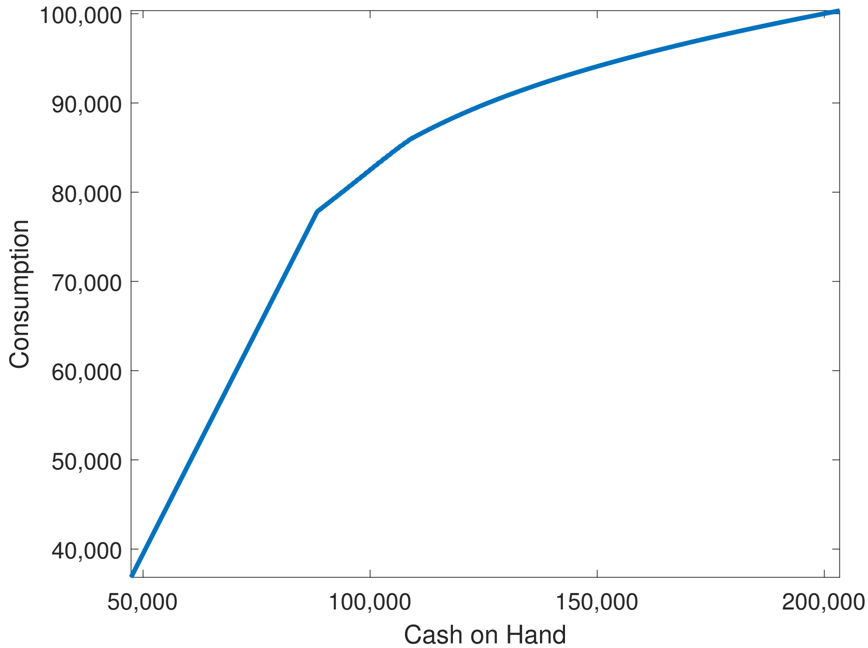
The two optimality conditions thus reveal the portfolio problem behind the withholding decision. Private saving and additional withholding are like two different assets; putting \$1 in additional withholding returns $\tilde{\phi}$ while saving returns R. It follows that, in the absence of uncertainty, overwithholding is never optimal because there is no additional return to allocating \$1 to withholding once the tax bill is paid in full. In the case of uncertainty, the optimal portfolio can result in overwithholding. Because $\phi > R$, how much an individual withholds depends on the distribution of uncertain income N_t, and there will be realizations of N_t that are low enough to produce income tax refunds.

2.3.2 Optimal "Excess Sensitivity"

Prior analyses of the demand for income tax refunds have been separated from analyses of how spending changes when the refunds arrive. In the model developed here, the two behaviors emerge from a common source. We saw that other income uncertainty motivates precautionary "savings" in the form of overwithholding. As shown in Zeldes (1989) and Carroll and Kimball (1996), for a broad class of preferences, income uncertainty implies consumption is responsive to transitory income and the "consumption function," which maps existing financial assets into the optimal level consumption, is concave. The effect on consumption of transitory income depends, that is, on the level of cash on hand.

Figure 1 presents the consumption function for a calibrated version of the model we study. A detailed description of the calibration is postponed until section 4. The figure shows the influence of the endogenous liquidity constraints that emerge from income uncertainty. Consumption responds to changes in cash on hand, and the function is concave. The marginal propensity to consume out of income is higher when cash on hand is low.

Figure 1: Consumption function



Notes: The consumption function shows the amount of optimal consumption as a function of cash on hand.

Recall from first order condition (5), optimal behavior implies that tax refunds occur when the realization of other income in the previous period was sufficiently low. When other income is lower in the previous period, current cash on hand is also lower. The concavity of the consumption function, confirmed in Figure 1, implies that this is precisely when the marginal propensity to consume is relatively high. It follows that tax refunds based on period t income and withholding arrive when the marginal propensity to consume from income is higher. Indeed, as we show via simulation of the model below, the bigger the refund the higher the marginal propensity to consume. In this model, refunds and “excess sensitivity” are therefore linked. The uncertain income that creates the demand for tax refunds also increases the marginal propensity to consume when those refunds arrive.

3 Data and Estimates of the Income Process

The quantitative relevance of the proposed mechanisms behind both tax refunds and excess sensitivity of spending to refunds depends on several factors. These factors include standard inputs such as preference parameters, interest rates, and tax schedules. More challenging, quantitative evaluation of the theory requires estimates of both the paycheck and other income processes.

3.1 Data Source

To estimate the paycheck and other income processes, and later to evaluate ancillary predictions of the model, we turn to administrative account information derived from de-identified transactions and balance data from individual-level, linked checking, saving, and credit card accounts. The data are captured in the course of business by a personal finance app.[8] The app offers financial aggregation and bill-paying services. Users can link almost any financial account to the app, including bank accounts, credit card accounts, utility bills, and more. Each day, the app logs into the web portals for these accounts and obtains central elements of the user's financial data including balances, transaction records and descriptions, the price of credit and the fraction of available credit used. Prior to analysis, the data are stripped of personally identifying information such as name, address, or account number. The data have scrambled identifiers to allow observations to be linked across time and accounts. We draw on the entire de-identified population of active users and data derived from their records from from December 2012 until July 2016.

3.2 Sample and measurement

From this population, we draw a sample that is filtered on several dimensions to reduce measurement error in key variables and to focus attention on workers with at least some regular paycheck income. In particular, to observe a sufficiently complete view of spending

[8] We gratefully acknowledge the partnership with the financial services application that makes this work possible. All data are de-identified prior to being made available to the project researchers. Analysis is carried out on data aggregated and normalized at the individual level. Only aggregated results are reported. We have used these data previously to study anticipated income, stratified by spending, income and liquidity (Gelman et al., 2014) and the high-frequency responses of households to shocks such as the government shutdown (Gelman et al., 2015). Similar account data from other apps have been used in Baugh et al. (2018), Baker (2017), Baker and Yannelis (2017), Kuchler (2015), Ganong and Noel (2016), and Kueng (2018).

and income, we limit attention to app users who link all (or most) of their accounts to the app, generate a long time series of observations, and have positive income in each month. To study the importance of both paycheck and other income, we also restrict attention to app users who receive regular paychecks throughout most of the time we observe them in our data. The specifics of these filters are provided in the Appendix and the consequences for sample size are presented in Table A.1.

Our analysis is therefore based on a sample of individuals with payroll income, with longitudinal observations that allow estimation of the variability of income, and with well-linked accounts. There are 62,946 individuals in the panel with 4 years of observations per individual on average. The account data we use are "non-designed" (aka "big data"). In contrast with designed, survey data, the sample is based on those who enroll in the app, which is selected non-randomly from the population. We have taken a number of number of steps to assess whether the sample is at least broadly representative of the population. In (Gelman et al., 2014), we conduct an external validation exercise that compares the distribution of demographic characteristics including age, education, and location and the distribution of income of our sample with representative samples. Although there are differences, notably that very low incomes and older individuals are underrepresented, the demographic and economic profile of the sample from the app captures a diverse population.

To further evaluate the validity of the sample, after we define key variables in subsection 3.3, we will compare the distributions of these variables in the filtered data with their distributions in other data sources. This analysis suggests that our sample is well-aligned with the population along key dimensions relevant for this analysis—propensity to receive tax refunds, size of refunds, and fraction of non-payroll income not subject to withholding.

The data have other features that make them particularly well-suited for addressing the questions in this paper. First, unlike most administrative account data, the data links credit card and bank accounts, so that they provide comprehensive measurement of income and spending rather than just income or spending from a particular account. Second, the data have both spending and income flows and stocks of liquid assets. Third, unlike survey sources such as the PSID or CEX, the data provide accurate, high-frequency data on a relatively large sample.

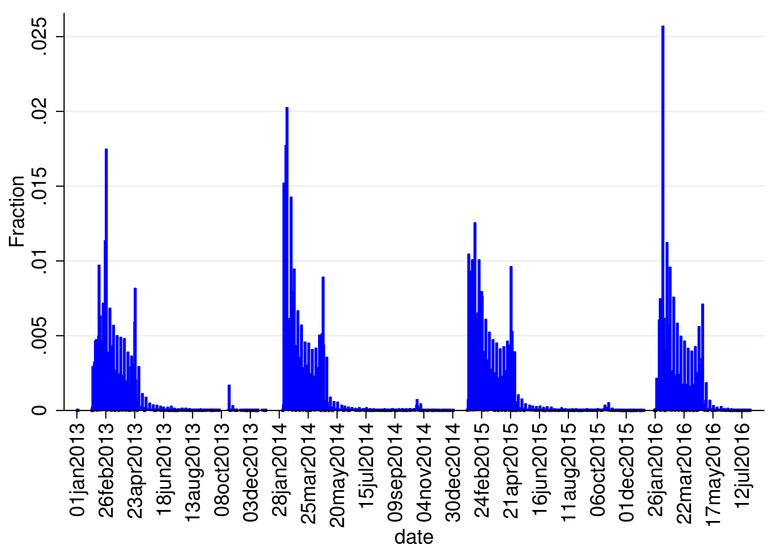
3.3 Variable definitions

The key variables for the analysis are tax refunds, non-durable expenditure, and paycheck and other income. Here we describe how those variables are measured from the transaction data.

3.3.1 Tax refunds

The data from the app consists of individual transactions and include information such as amount, transaction type (debit or credit), and a transaction description. We identify tax refunds by searching for identifying keywords in the description field (all tax refunds include the keywords "TAX", "TREAS", and "REF"). We exclude individuals that receive multiple tax refunds within the same year. Figure 2 shows the time series of the fraction of tax refunds observed in the data from December 2012 to July 2016. Most refunds in these data are received in February, March, April, and May.

Figure 2: Federal tax refund time series



Notes: The figure contains data from 142,315 tax refunds and 49,520 individuals.

3.3.2 Expenditure

Another primary focus of our analysis is the spending response individuals to the arrival of income tax refunds. Following the literature, we will calculate an empirical MPC out of refunds based on a measure of non-durable expenditure.

The transaction records do not indicate, directly, whether spending is on non-durable or durable goods. We therefore adopt a machine learning (ML) algorithm (see appendix section E.5 for more details) to aid in categorization. The goal of the ML algorithm is to provide a mapping from transaction descriptions to spending categories. For example, any transaction with the keyword "McDonald's" should map into "Fast Food." A subset of these categories are then combined to create the consumption variable.

The finest level of categorization is derived using merchant category codes (MCCs) which are indicated in each spending transaction record by two of the account providers in the data. MCCs are four digit codes used by credit and debit card companies to classify spending and are also recognized by the Internal Revenue Service for tax reporting purposes. The ML algorithm uses a subset of the data where the MCC is recorded in order to create a mapping from transaction description to MCCs. After training the ML algorithm on the data where the MCC is recorded, the algorithm is applied to the rest of the data set. We then define consumption as spending on restaurants, groceries, gasoline, entertainment, and services.

3.3.3 Income

We define income as the sum of all inflows to checking and saving accounts minus incoming transfers. From this measure of income, paycheck income is defined as the inflows from paychecks identified as such using the keywords from Appendix section E.4. This definition of paycheck income is net of deductions including income and payroll tax withholding. The proper input to the model is income before taxes. To obtain before-tax paycheck income, we adjust measured income to account for state and federal income taxes and for federal payroll taxes. See Appendix B for specifics. All income not classified as paycheck income is defined as other income.

3.4 Comparison with Other Data Sources

With these measures of tax refunds, expenditure, and income, we can compare statistics of the app analytic sample to those from external data sources. Table 1 shows that the average tax refund in the sample is $1,704, setting the refund or tax payment equal to zero for those who did not receive a refund. Restricting attention to those who received a refund, the average size is $3,184, slightly larger than the average reported by the IRS $2,778.

Average annual spending in the full sample is $83,253, and $77,854 among those who received a refund in the relevant year. Consistent with positive selection on income, these average spending numbers are, respectively, 43% and 33% higher than those in the Consumer Expenditure Survey.

Comparing both average paycheck and other income to analogous statistics from the IRS also shows the analytic sample is higher income than the population at large. Average paycheck income in the whole sample is $68,226 and average other income is $38,764. Among all income tax filers, the IRS reports average paycheck income to be $46,224 and average other income is $20,603.

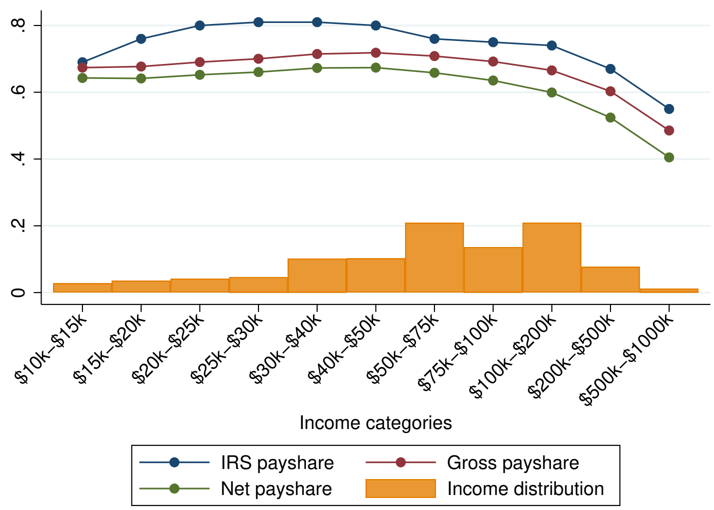
Table 1: Summary statistics

	Full sample mean	Only refunds mean	External sources mean
Tax refund	1,704	3,184	2,778
Spending	83,253	77,854	58,410
Paycheck income	68,226	67,415	46,224
Other income	38,764	36,607	20,603
Paycheck share	.68	.68	.69
NxT	251,784	134,752	
N	62,946	49,520	

Notes: NxT represents the number of individual-year observations. N represents the number of individual observations. External tax refund data from IRS databook. Results are based on individual taxes not including the child tax credit or the EITC. External spending data is calculated from the Consumer Expenditure Survey. External income data is calculated from IRS, Statistics of Income Division Publication 1304.

While the average levels of paycheck and other income are higher in the app sample than in the population of tax filers, the ratio of paycheck income to total income is similar, about 0.68. More generally, Figure 3 plots the paycheck share from the IRS along with two of our measures for different segments of the income distribution. Net payshare is the paycheck share calculated before adjusting the data to take account of various withholding as explained in Appendix B. Gross payshare computes the paycheck share after the adjustment for various tax withholding is taken into account. This adjustment tends to make a bigger difference as income increases due to the progressive nature of the federal tax code. The income distribution box plots show how total income is distributed in our sample. The bulk of the data falls within the $30k to $200k range where the gap between our measure and the IRS data is at its smallest.

Figure 3: Paycheck share comparison across income groups



Notes: IRS payshare is calculated from IRS, Statistics of Income Division, Publication 1304.

3.5 Estimating Income Processes

In the model developed above, a time period is a year; simulating the model's prediction therefore requires estimates, at annual frequency, of the expected level and volatility both paycheck and other income. Moreover, to evaluate the "mechanical" effects of higher frequency income variation on tax refunds, we allow a bi-weekly component to the income variables introduced in Section 2.2.

In the estimating equations, t indexes the year and b indexes the bi-week interval. Total bi-weekly income is a combination of paycheck income in that two-week interval, $y_{t,b}$, and non-paycheck income in that two-week interval $n_{t,b}$. Bi-weekly income results from an annual income process and a higher frequency, bi-weekly, process. We model these variables as

$$y_{t,b} = \frac{Y_t}{26} + \epsilon^Y_{t,b} \tag{6}$$

$$n_{t,b} = \frac{N_t}{26} + \epsilon^N_{t,b} \tag{7}$$

$$Y_t = \alpha_Y + \nu^Y_t \tag{8}$$

$$N_t = \alpha_N + \nu^N_t \tag{9}$$

$$\nu^Y_t \sim F(0, \sigma^2_{\nu^Y}), \epsilon^Y_{t,b} \sim F(0, \sigma^2_{\epsilon^Y}), \nu^N_t \sim F(0, \sigma^2_{\nu^N}), \epsilon^N_{t,b} \sim F(0, \sigma^2_{\epsilon^N}) \tag{10}$$

where lower case variables represent bi-weekly frequencies and upper case variables represent annual frequencies. The random components of bi-weekly and annual variables for the two components of income are represented by $\epsilon^Y_{t,b}$, $\epsilon^N_{t,b}$, ν^Y_t, and ν^N_t, respectively. Annual income for the two components are modeled as independent processes.[9] On top of these annual components, bi-weekly income is subject to serially-uncorrelated noise. Section E.3 in the Appendix describes how we compute the parameters from moments in the data.

Table 2 gives the estimated parameter values. Recall that σ_ϵ is measured at the bi-weekly level while σ_ν is measured at the yearly level.

[9] We considered more elaborate time series processes but ultimately decided on a parsimonious specification. Because our time-series sample only consists of four years, there is little hope of estimating more elaborate annual income processes with sufficient precision.

Table 2: Parameter estimates

α_N	38,764
α_Y	68,226
σ_{ν^N}	19,879
σ_{ν^Y}	18,490
σ_{ϵ^N}	3,182
σ_{ϵ^Y}	1,791
NxT	251,784
N	62,946

Notes: NxT represents the number of individual-year observations. N represents the number of individual observations. All values winsorized at the top 1%.

4 Results

4.1 Calibration

With estimated parameters of the income process, we have the necessary inputs to calibrate the model. Adopting a standard, constant relative risk aversion form for utility, Table 3 presents the levels of the parameters used to simulate the model. To ensure the results do not depend on unusually high degrees of risk aversion or time discounting, we assume a coefficient of risk aversion equal to one (log utility) and an annual time discount factor of 0.98. We approximate the tax schedule as a function of income with a third degree polynomial.

Table 3: Model calibration

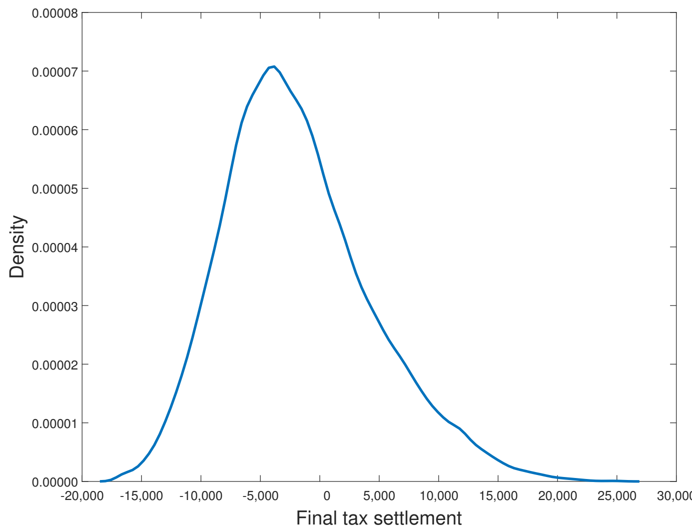
Parameter	Value	Notes
$u(x)$	$\frac{x^{1-\theta}}{1-\theta}$	utility function
θ	1	risk aversion
β	0.98	discount factor
α_Y	68,226	mean income subject to withholding
α_N	38,764	mean income not subject to withholding
σ_{ν^N}	19,879	standard deviation of income not subject to withholding
$\tau(y)$		tax schedule (polynomial approx)
R	1.007	1-yr treasury yield
ϕ	0.042	0.5% penalty + (R-1) + 3%

4.2 Explaining the level of refunds

To compare the distribution of tax refunds predicted by the model to that in the data, we simulate the calibrated model for 100,000 periods[10] and record the final tax payment in each period. As discussed in section 2.3.1, we expect workers to over-withhold on average in order to avoid the under-withholding penalty. Figure 4 shows the distribution of final tax settlements after simulating the model. A negative number represents a tax refund while a positive number represents a tax payment.

[10] We discard the first 1,000 periods.

Figure 4: Distribution of final tax settlement



Notes: This figure shows the density of final tax settlement for 100,000 simulated observations. A negative settlement represents a refund while a positive settlement represents a payment.

As expected, workers overwithhold on average. Less anticipated, the calibration produces an average refund quite close to that observed in the data. Table 4 shows the average tax settlement is a refund of \$1,846. In the data, the average is a refund of \$1,704. Thus, with standard calibration of other key parameters, the estimated level and variation in non-paycheck income is sufficient to justify the average size of refunds in the data.

The specific realization of the refund/tax payment depends on the income shocks faced by the worker. In cases where a worker receives a large negative income shock, they will tend to receive a refund because their tax liability will be lower than expected. Conversely, a large positive income shock is associated with a tax payment.

Table 4: Final tax payment statistics

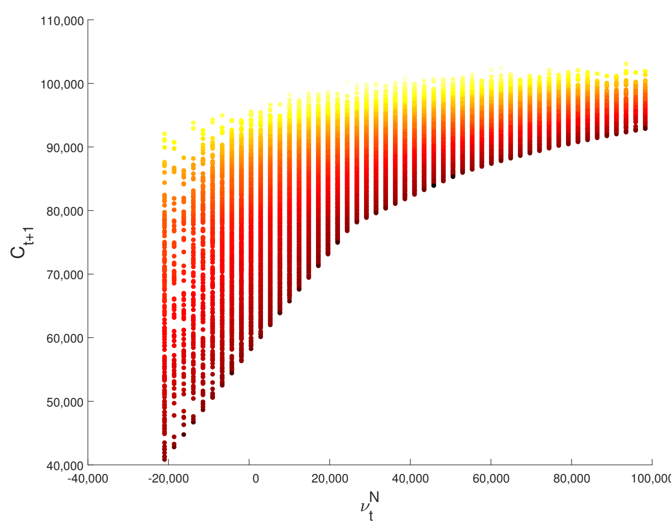
$\mathbb{E}$[Tax]	25%	50%	75%
-1,846	-6,360	-2,378	1,892

4.3 Explaining the high MPC

The previous section showed how the model explains large refunds on average. The money costs of under-withholding combined with uncertainty in non-paycheck income makes it optimal to over-withhold. This section uses the model to understand the second puzzle: why workers spend disproportionately out of their refunds. Most important, the model reveals links between a negative non-paycheck income shock, cash on hand, consumption, and the final tax settlement.

To better understand the mechanism linking shocks, refunds, and expenditure, Figure 5 plots the relationship between the non-paycheck income shock (ν_t^N) and next period consumption (C_{t+1}) in the calibrated model. The different colors of the circles represent different levels of cash on hand (X_t). Darker colors represent lower values. On average, a negative ν_t^N shock results in lower levels of consumption next period. Except for those with very high levels of cash on hand, these periods when consumption is low tend to be periods when the MPC is very high because the marginal utility of consumption is high. Because negative non-paycheck income shocks tend to lead to larger refunds, the MPC tends to be higher when individuals receive tax refunds.

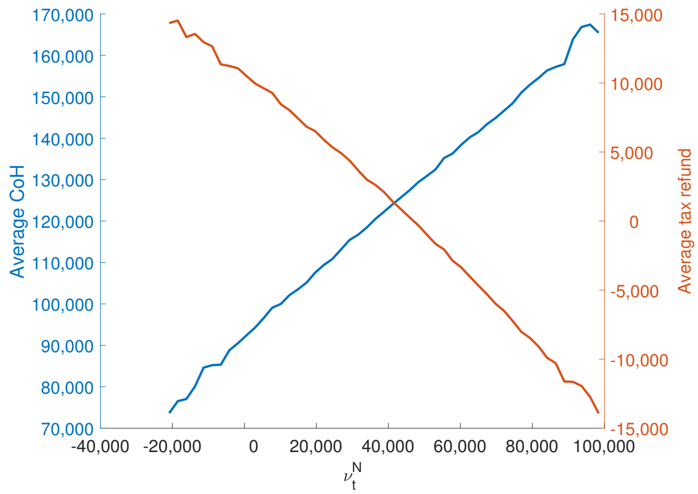
Figure 5: Non-paycheck Income Shock vs Next Period Consumption



Notes: This figure plots the relationship between the non-paycheck income shock (ν_t^N) and next period consumption (C_{t+1}). For any given ν_t^N, the value of C_{t+1} may vary depending on what cash on hand (X_t) is. Low values of X_t are represented by darker colors. It represents data from 100,000 simulated observations.

The link between tax refunds and cash on hand is seen in Figure 6. When a worker experiences a negative non-paycheck income shock (ν_t^N), this directly reduces cash on hand because the worker's wealth falls. At the same time, a negative ν_t^N will tend to lead to a tax refund because tax liability will be lower than expected. In cases of a positive ν_t^N shock, the argument is reversed and individuals tend to have an increase in cash on hand and will owe the government a tax payment (negative tax refund).

Figure 6: Average CoH and Tax Refund Conditional on ν_t^N

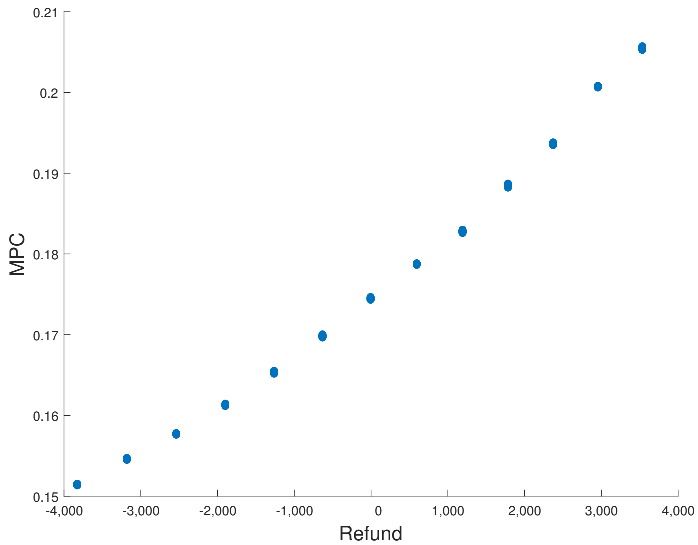


Notes: This figure plots the relationship between the non-paycheck income shock (ν_t^N) and average cash on hand on the left Y-axis and average tax refund on the right Y-axis. It represents data from 100,000 simulated observations.

We can combine the mechanisms described in Figures 5 and 6 to characterize the relationship between the average MPC and the average tax refund in the calibrated model. Figure 7 shows the positive relationship between the MPC and the level of tax refunds for observations close to average cash on hand values. When individuals are near their average cash on hand, a negative non-paycheck income shock will lead to a large refund because tax liability will be lower than expected. At the same time, the negative shock results in lower cash on hand levels and, because the consumption function is concave, lower cash on hand leads to a higher MPC.

In this way, the model predicts "excess sensitivity" of spending to the arrival of refunds like that observed in (Souleles (1999), Baugh et al. (2018), Gelman (2018). Because cash on hand tends to be low refunds arrive, workers will spend disproportionately out of this transitory income.

Figure 7: Predicted Relationship Between the MPC and Tax Refunds



Notes: This figure shows the positive relationship between the MPC and the level of tax refunds for observations close to average cash on hand values. It represents data from 100,000 simulated observations.

4.4 Micro Evidence

The balance and transaction records made available in the app data provide key inputs to the calibrated model analyzed above. Specifically, the app data provide the first, to our knowledge, individual-level measures of the level of and the variation in both paycheck and non-paycheck income. In addition, the individual-level data allow direct evaluation of some of the key mechanisms in the model. To that end, this section estimates the empirical relationship between individual-level tax refunds and individual-level variation in income processes.

4.4.1 Excess withholding due to high frequency paycheck volatility

The model highlights the relationship between yearly fluctuations in income and tax refunds. As explained in section 2.1, however, there is also a potential "mechanical" effect of within-year paycheck volatility on refunds. Within-year paycheck income changes can lead to excess withholding because the withholding schedule assumes biweekly paychecks are pro-rated annual income. Other things equal, therefore, the convex income tax schedule implies withholding will increase, weakly, as within-year paycheck income rises.

To quantify the potential magnitude of the mechanical effect and to isolate its influence from that of annual fluctuations, we define potential excess withholding from high frequency paycheck volatility as:

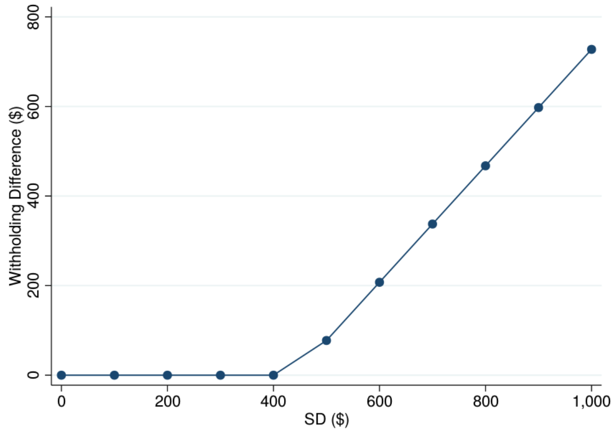
$$ExcessW_y = \sum_{b=1}^{26} (w(p_{by}; s, e) - w(\bar{p}_y; s, e)) \tag{11}$$

where $w(\cdot; s, e)$ is a withholding function that takes in paycheck income as its argument and is influenced by filing status s and number of exemptions e, p_{bt} is the bi-weekly pre-withholding paycheck in bi-week b of year y, and $\bar{p}_y$ is the average bi-weekly pre-withholding paycheck in year y.[11]

Figure 8 illustrates the relationship between this measure of potential excess withholding and within-year paycheck volatility. The example in the figure assumes paychecks are one standard deviation above average half the time and one standard deviation below average the other half of the time. As expected, the measure of potential excess withholding, $ExcessW_y$, increases as within year paycheck variation increases. The relationship is not linear, however, because potential excess withholding is positive only if annualized paycheck income crosses marginal tax rates. Because the tax schedule is a piece-wise linear function of income, there are regions where modest within-year variation doesn't lead to any excess withholding.

[11] We do not observe pre-withholding income p_{by} directly. Instead we observe post-withholding income $\tilde{p}_{by} = p_{by} - w(p_{by}; s, e)$. Therefore, we estimate p_{by} from $\tilde{p}_{by}$ conditional on s and e. Because observed post-withholding paycheck income is a function of pre-withholding income and other tax parameters, $\tilde{p}_{by} = f(p_{by}; s, e)$ and we can simply take the inverse of this function to estimate p_{by} by $p^*_{by} = f^{-1}(\tilde{p}_{by}; s, e)$.

Figure 8: $ExcessW_y$ as a function of within year paycheck variation



Notes: $ExcessW_y$ is calculated based on a single filer with two exemptions. The paycheck fluctuates one standard deviation above the average half of time and one standard deviation below the average the rest of time.

In the analytic sample of the app data, the average potential excess withholding is large, $1,340 in the full sample and $1,151 among those who receive refunds. If workers did not account for this mechanical effect of within-year income variation, predicted refunds would be even larger. Workers may, however, internalize the effects of this within-year income variation and adjust their withholding accordingly. In what follows, we evaluate the extent of this internalization as we correlate tax refunds with this measure of potential excess withholding.

4.4.2 The Empirical Relationship Between Tax Refunds and the Sources and Variation of Income.

Table 5 provides summary statistics from the app data of both tax refunds and measures of the sources and variation in income. These latter measures include the share of income from paychecks, the annual volatility of non-paycheck income and of paycheck income, and the measure of potential excess withholding described above.

Table 5: Summary statistics

	(1) Full sample	(2) Only refunds
$Refund_{it}$	1,704	3,184
	(2,542)	(2,712)
$payshare_i$	.679	.682
	(.197)	(.186)
$ExcessW_{it-1}$	1,340	1,151
	(2,358)	(2,006)
$\sigma^2_{\nu_i^Y}/1000000$	810	649
	(2,480)	(2,047)
$\sigma^2_{\nu_i^N}/1000000$	1,765	1,302
	(8,577)	(7,018)
NxT	251,784	134,752
N	62,946	49,520

Notes: Mean values reported with standard deviation in parenthesis. NxT represents the number of individual-year observations. N represents the number of individual observations. For $ExcessW_{it-1}$, NxT is 167,644 and N is 62,813.

Table 6 presents OLS estimates of the relationship between refunds and these sources and variation of income. In each specification, the dependent variable is the log of an individual's refund in year t. There are 49,520 individuals in the analysis sample that receive at least one refund during the period. On average each of these individuals receives 2.7 (out of a max of 4) refunds during the period.

Specification (1) estimates the relationship between tax refunds and the individual's average share of annual income that comes from a paycheck. That average is calculated over the four years of observation. The basic mechanisms of the model indicate that refunds emerge only when substantial fractions of taxable income is not subject to withholding at the source (non-paycheck). Consistent with these mechanisms, we find that the relationship between the paycheck share and refunds is negative, and both economically and statistically significant. The point estimate indicates that a worker who earns 90% of her income from paycheck would have a refund that is about half the size of a worker who earned just 20% of her income from a paycheck.

The results of Table 6 are also consistent with the model's emphasis on income uncertainty as a driving force behind tax refunds. Column (2) provides estimates of the correlation between the log of refunds and the log of the individual's average annual variance of both other and paycheck income. Consistent with the model, the annual variance of other income is associated with higher refunds, and this relationship is statistically significant at standard levels of confidence. The results also point to a role for annual variation in paycheck income, suggesting as in Jones (2012) that the default levels of withholding paycheck income may be biased toward refunds.

Column (3) of Table 6 evaluates the simple correlation between the log of tax refunds and the measure of excess withholding due to high frequency variation in paycheck income described in 4.4.1. The sample size declines because the estimate is based on the prior year's income variation and so only three years are available. These results are consistent with a significant, though economically modest, mechanical effect of this high frequency variation on tax refunds. The modest size of the point estimate indicates that workers internalize much of the "mechanical effect" and adjust withholding accordingly.

Conditioning on each of these measures of an individual worker's sources and variation in income, in Column (4), the qualitative results are unchanged. Even conditional on the mechanical effect of high-frequency variation in income, both the payshare and the measures of annual income volatility have relationships with tax refunds of the expected sign. In each case, the relationship is statistically significant and economically substantial.[12]

[12]Columns (5) and (6) repeat the analysis in columns (1) and (2) restricting the sample to those years for which we can calculate the excess withholding measure. These results indicate that the changes in the coefficients on payshare and the log of the variance in income is not due to the change in sample

Table 6: Tax refunds and income volatility

	(1)	(2)	(3)	(4)	(5)	(6)
	$Log(Refund)_{it}$	$Log(Refund)_{it}$	$Log(Refund)_{it}$	$Log(Refund)_{it}$	$Log(Refund)_{it}$	$Log(Refund)_{it}$
$payshare_i$	-0.903			-0.578	-0.907	
	(0.0246)			(0.0361)	(0.0279)	
$Log(\sigma^2_{\nu_i^N})$		0.0921		0.0573		0.0930
		(0.00235)		(0.00335)		(0.00257)
$Log(\sigma^2_{\nu_i^Y})$		0.0382		0.0507		0.0394
		(0.00256)		(0.00314)		(0.00287)
$Log(ExcessW_{it-1})$			0.0456	0.0255		
			(0.00185)	(0.00181)		
Constant	8.195	5.179	7.348	5.866	8.247	5.186
	(0.0180)	(0.0544)	(0.0131)	(0.0734)	(0.0202)	(0.0599)
NxT	134,752	134,752	87,736	87,736	87,736	87,736
N	49,520	49,520	44,328	44,328	44,328	44,328

Notes: NxT represents the number of individual-year observations. N represents the number of individual observations.

4.4.3 MPC analysis

Another key mechanism of the model is the link between the arrival of tax refunds and the marginal propensity to consume (MPC). The endogenous liquidity constraints that emerge in the model imply that, when larger refunds arrive due to lower than expected non-paycheck income, cash on hand is lower and the MPC is higher.

To evaluate this link in the micro data we estimate the MPC as a function of tax refund size using the following specification

$$C_{it} = \sum_{j=1}^{5} MPC_j \times Refund_{it} \times Q_i^j + \sum_{j=2}^{5} Q_i^j + month_t + \varepsilon_{it} \tag{12}$$

where C_{it} is our measure of non-durable spending, $Refund_{it}$ is the level of tax refund received, Q_i represents quintiles of tax refunds, and $month_t$ are month fixed effects. MPC_j captures the average MPC out of refunds for each quintile of estimated tax refunds.

To isolate changes in refunds attributable to non-paycheck income, we adopt a two-stage estimation strategy that, in the second stage, replaces the worker's quintile of the refund distribution with its prediction from a regression of refund quintile on payshare. The results of the second stage estimation are presented in Table 7 and summary statistics of the predicted quintiles are provided in Appendix Table D.2.

Table 7: MPC estimates

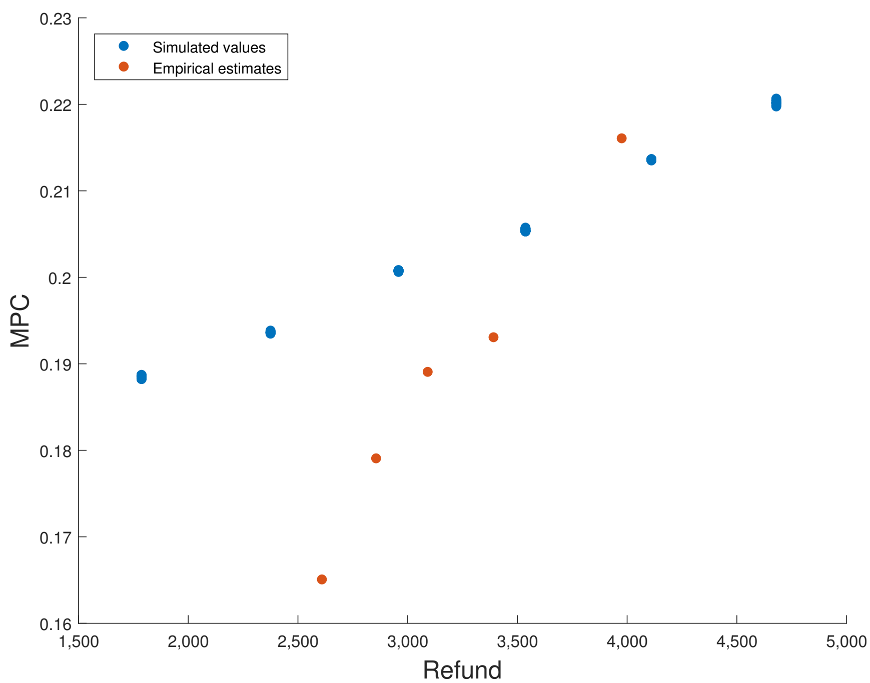
	(1)
VARIABLES	C_{it}
$Q2_i$	-0.007
	(0.002)
$Q3_i$	-0.018
	(0.002)
$Q4_i$	-0.032
	(0.002)
$Q5_i$	-0.058
	(0.002)
$Refund_{it} \times Q1_i$	0.165
	(0.006)
$Refund_{it} \times Q2_i$	0.179
	(0.005)
$Refund_{it} \times Q3_i$	0.189
	(0.005)
$Refund_{it} \times Q4_i$	0.193
	(0.005)
$Refund_{it} \times Q5_i$	0.216
	(0.005)
Constant	0.286
	(0.001)
Observations	1,615,572
R-squared	0.047

Robust standard errors in parentheses

The results of Table 7 are consistent with the link between the MPC and tax refunds derived from the model. Those predicted to be in higher refund quintiles because of their lower paycheck income shares, have higher MPCs out of refunds, and the estimated relationship is monotonic. The highest quintile of predicted tax refunds distribution has an MPC that is 30% higher than that of the lowest quintile.

These micro estimates of the relationship between refunds and the MPC out of refunds are qualitatively consistent with the link proposed in the model. A quantitative comparison of these point estimates to the predictions of the calibrated model is presented in Figure 7. Comparing the estimated relationship to the predicted indicates that, in the micro data, the MPCs out of refunds rise more quickly with refunds than would be predicted by the model. But the estimated level of MPCs are remarkably close to those predicted by the simple calibrated model.

Figure 9: Caption here



Notes: This blue dots plot the relationship between the MPC and the level of tax refunds for simulated observations close to average cash on hand values. It represents data from 100,000 simulated observations.

5 Conclusion

This paper presents and evaluates a simple theory of household liquid assets management with income shocks that can explain both the prevalence of income tax refunds and the tendency of households to spend disproportionately from those refunds when they arrive. The theory maintains standard assumptions but, different from prior analyses of tax refunds or excess sensitivity, takes account of the volatility of income not subject to tax withholding at the source.

A central mechanism of the theory is the endogenous liquidity constraints that emerge as households manage annual fluctuations in income not subject to withholding. The theory predicts large refunds, on average, when non-paycheck income represents a sufficiently large and unpredictable fraction of total income. The theory also predicts that these refunds will tend to arrive when cash-on-hand is low, and thus the marginal propensity to spend is high—thereby explaining what appears to be excess sensitivity of spending from tax refunds.

Administrative account data on income, spending, and refunds show, that the average level of and variation in non-paycheck income is more than sufficient to explain the size of average tax refunds. The micro data also provide evidence consistent with the basic mechanisms of the theory: The fraction of annual income that is not subject to withholding has an economically and statistically significant positive relationship with tax refunds, and those whose non-paycheck income shares predict larger refunds also have larger marginal propensities to consume out tax refunds.

The model and evidence presented here further underscore the importance of income uncertainty and precautionary savings motives for household behavior and well-being. The preceding analysis shows, in particular, the importance of different sources of income uncertainty for understanding how households manage liquid assets and respond to tax policy.

References

Baker, Scott (2017) "Debt and the response to household income shocks: Validation and application of linked financial account data," *Journal of Political Economy*.

Baker, Scott R and Constantine Yannelis (2017) "Income changes and consumption: Evidence from the 2013 federal government shutdown," *Review of Economic Dynamics*, Vol. 23, pp. 99–124.

Baugh, Brian, Itzhak Ben-David, Hoonsuk Park, and Jonathan A Parker (2018) "Asymmetric Consumption Response of Households to Positive and Negative Anticipated Cash Flows," Working Paper 25086, National Bureau of Economic Research.

Braxton, J Carter, Gordon Phillips, Kyle Herkenhoff et al. (2018) "Can the Unemployed Borrow? Implications for Public Insurance," in *2018 Meeting Papers*, No. 564, Society for Economic Dynamics.

Bunn, Philip, Jeanne Le Roux, Kate Reinold, and Paolo Surico (2018) "The consumption response to positive and negative income shocks," *Journal of Monetary Economics*, Vol. 96, pp. 1–15.

Carroll, Christopher D. and Miles S. Kimball (1996) "On the Concavity of the Consumption Function," *Econometrica*, Vol. 64, No. 4, pp. 981–992.

Christelis, Dimitris, Dimitris Georgarakos, Tullio Jappelli, Luigi Pistaferri, and Maarten van Rooij (2017) "Asymmetric Consumption Effects of Transitory Income Shocks," SSRN Scholarly Paper ID 2914247, Social Science Research Network, Rochester, NY.

Fennell, Lee Anne (2006) "Hyperopia in Public Finance," in Edward J. McCaffery and Joel Slemrod eds. *Behavioral Public Finance*. Russell Sage Foundation, pp. 141–172.

Fuster, Andreas, Greg Kaplan, and Basit Zafar (2018) "What Would You Do With $500? Spending Responses to Gains, Losses, News and Loans," Working Paper 24386, National Bureau of Economic Research.

Ganong, Peter and Pascal Noel (2016) "How Does Unemployment Affect Consumer Spending?"

Gelman, Michael (2018) "What Drives Heterogeneity in the Marginal Propensity to Consume? Temporary Shocks vs Persistent Characteristics."

Gelman, Michael, Shachar Kariv, Matthew D. Shapiro, Dan Silverman, and Steven Tadelis (2014) "Harnessing naturally occurring data to measure the response of spending to income," *Science*, Vol. 345, No. 6193, pp. 212–215.

——— (2015) "How Individuals Smooth Spending: Evidence from the 2013 Government Shutdown Using Account Data," Working Paper 21025, National Bureau of Economic Research.

Hall, Robert E (1978) "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, Vol. 86, No. 6, pp. 971–87.

Herkenhoff, Kyle F (2018) "The Impact of Consumer Credit Access on Unemployment," Working Paper 25187, National Bureau of Economic Research.

Highfill, Jannett, Douglas Thorson, and William V. Weber (1998) "Tax Overwithholding as a Response To Uncertainty," *Public Finance Review*, Vol. 26, No. 4, pp. 376–391.

Hsieh, Chang-Tai (2003) "Do Consumers React to Anticipated Income Changes? Evidence from the Alaska Permanent Fund," *American Economic Review*, Vol. 93, No. 1, pp. 397–405.

Jones, Damon (2012) "Inertia and Overwithholding: Explaining the Prevalence of Income Tax Refunds," *American Economic Journal: Economic Policy*, Vol. 4, No. 1, pp. 158–85.

Kaplan, Greg and Giovanni L Violante (2014) "A model of the consumption response to fiscal stimulus payments," *Econometrica*, Vol. 82, No. 4, pp. 1199–1239.

Kaplan, Greg, Giovanni L. Violante, and Justin Weidner (2014) "The Wealthy Hand-to-Mouth," *Brookings Papers on Economic Activity*, Vol. 45, No. 1 (Spring), pp. 77–153.

Kuchler, Theresa (2015) "Sticking to your plan: Hyperbolic discounting and credit card debt paydown," *Available at SSRN 2629158*.

Kueng, Lorenz (2018) "Excess Sensitivity of High-Income Consumers," *The Quarterly Journal of Economics*, Vol. 133, No. 4, pp. 1693–1751.

Neumark, David (1995) "Are Rising Earnings Profiles a Forced-Saving Mechanism?" *The Economic Journal*, Vol. 105, No. 428, pp. 95–106.

Parker, Jonathan A. (1999) "The Reaction of Household Consumption to Predictable Changes in Social Security Taxes," *American Economic Review*, Vol. 89, No. 4, pp. 959–973.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011) "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830.

Rees-Jones, Alex (2018) "Quantifying Loss-Averse Tax Manipulation," *The Review of Economic Studies*, Vol. 85, No. 2, pp. 1251–1278.

Souleles, Nicholas S. (1999) "The Response of Household Consumption to Income Tax Refunds," *The American Economic Review*, Vol. 89, No. 4, pp. 947–958.

Thaler, Richard H. (1994) "Psychology and Savings Policies," *The American Economic Review*, Vol. 84, No. 2, pp. 186–192.

Zeldes, Stephen P (1989) "Consumption and liquidity constraints: An empirical investigation," *The Journal of Political Economy*, Vol. 97, No. 2, pp. 305–346.

A Appendix – Data Filters, Definitions

The main analysis sample is drawn from the full s dimensions to reduce measurement error in key variables and to focus attention on workers with at least some regular paycheck income. In particular, to observe a sufficiently complete view of spending and income, we limit attention to app users who link all (or most) of their accounts to the app, generate a long time series of observations, and have positive income in each month. To study the importance of paycheck vs other income, we also restrict attention to app users who receive regular paychecks throughout most of the time we observe them in our data. The specifics of these filters are provided in the Appendix and the consequences for sample size are presented in Table 1, below.

A.0.1 Defining account linkage

The analysis may be biased if all accounts that are used for receiving income and making expenditures are not observed. For example, an individual may have a checking account that is used to pay most bills and a credit card that it used when income is low. If credit card expenditures are not properly observed the MPC will be biased downwards.

In order to identify linked accounts, we use a method that calculates how many credit card balance payments are also observed in a checking account. We define the variable *linked* as the ratio of the number of credit card balance payments observed in all checking accounts that matches a particular payment that originated from all credit card accounts. For example, a typical individual will pay their credit card bill once a month. If they existed in the data for the whole year, they will have 12 credit card balance payments. If 10 of those credit card payments can be linked to a checking account the variable $linked = \frac{10}{12} \approx 0.83$.

One drawback to this approach is that it requires individuals to have a credit card account. To ensure that those without credit cards are still likely to have linked accounts, we also condition on individuals who have three or more accounts.

A.0.2 Defining regular paycheck

In order to identify regular paychecks, we start by using keywords that are commonly associated with these transactions (see Appendix section E.4 for more details). We condition on four statistics to ensure that these transactions represent regular paychecks.

1. Number of paychecks ≥ 5
2. Median paycheck amount $>$ \$200
3. Median absolute deviation of days between paychecks is ≤ 5
4. Coefficient of variation of the paycheck amount ≤ 1

A.0.3 Defining stable paycheck

The ratio of paycheck and other income is an essential ingredient in our model. To ensure we are estimating the ratio correctly, we restrict attention to users who have received a paycheck at least 2/3 of the time we observe them in the sample.

A.0.4 Sample size

Table A.1 shows the evolution of the sample size from all users in the sample to those that survive the selection criteria. The criteria selects users who have a long time series ($\geq$ 40 months), a high linked account ratio (≥ 0.8), a reasonable number of accounts linked ([3,15]), receive a regular paycheck, receive positive income in each month, and receive more than 1 tax refund. We choose to drop users that have over 15 accounts linked because these accounts typically represent business users. Table 1 shows that this final sample compares well with external data for the variables that are important in our analysis.

Table A.1: Effect of sample filters

	Individuals	%
Full sample as of December 2012	883,529	100
Long time series (N $\geq$ 40)	341,841	39
Linked ratio $\geq$ 0.8	264,043	30
Linked accounts $\in$ [3,15]	197,530	22
Has regular paycheck	146,112	17
Has no months with zero income	77,052	9
Has stable paycheck	62,946	7

B Appendix – Estimating gross paycheck income

In our model, an individual makes withholding and saving decisions based on gross (pre-withheld) paycheck income and non-withheld income. In our data, we only observe net (post-withheld) income so we estimate gross paycheck income based on which taxes are withheld from an individuals' paycheck income.

The various types of withholding are

1. Federal income tax withholding (based on the yearly withholding schedule published by the IRS under Publication 15 or "Circular E")
2. Social security payroll tax (6.2%)
3. Medicare tax (1.45%)

4. State and local tax (based on yearly average state and local taxes collected)[13]

The observed net paycheck income is a function of gross paycheck income

$$Y_t^{net} = f(Y_t; s, e) \tag{13}$$

where s represents filing status and e represents the number of exemptions . We assume single filing status with two exemptions. We then invert this function to recover gross paycheck income.

C Appendix – High Frequency Income Variation and Overwithholding

High frequency variation in paycheck income can lead "mechanically" to overwithholding for households that adhere to the W-4 guidelines. This mechanical form of overwithholding results from the fact that the income tax schedule is convex and the withholding schedule treats each paycheck as prorated annual income.

To better understand why, Figure C.1 displays a table used to calculate withholding for individuals who receive a biweekly paycheck. Imagine two individuals who both earn $52,000 a year. One individual has no within year variation in their paycheck so they receive 26 bi-weekly paychecks of $2,000. The other individual has an extreme (and unrealistic) within year variation so that they receive 13 paychecks of $0 and 13 paychecks of $4,000. While this example is extreme, it easily highlights how within year variation will lead to overwithholding. The withholding table shows that the individual with smooth within year income has 25% of their paycheck withheld while the individual with extreme volatility has 28% of their paycheck withheld.

[13] We take total state and local income tax collected from "U.S. Census Bureau, Quarterly Summary of State and Local Government Tax Revenue" and divide it by total payroll tax reported in "IRS, Statistics of Income Division, Publication 1304" to arrive at an average state and local tax rate. The rates are 5.320%, 5.154%, 4.921%, and 5.291% for 2013,2014,2015, and 2016 respectively.

Figure C.1: Withholding table example

TABLE 2—BIWEEKLY Payroll Period

(a) SINGLE person (including head of household)—

If the amount of wages (after subtracting withholding allowances) is: Not over $88 . . . . The amount of income tax to withhold is: $0

Over—	But not over—		of excess over—
$88	—$447 . .	$0.00 plus 10%	—$88
$447	—$1,548 . .	$35.90 plus 15%	—$447
$1,548	—$3,623 . .	$201.05 plus 25%	—$1,548
$3,623	—$7,460 . .	$719.80 plus 28%	—$3,623
$7,460	—$16,115 . .	$1,794.16 plus 33%	—$7,460
$16,115	—$16,181 . .	$4,650.31 plus 35%	—$16,115
$16,181		$4,673.41 plus 39.6%	—$16,181

(b) MARRIED person—

If the amount of wages (after subtracting withholding allowances) is: Not over $333 . . . . The amount of income tax to withhold is: $0

Over—	But not over—		of excess over—
$333	—$1,050 . .	$0.00 plus 10%	—$333
$1,050	—$3,252 . .	$71.70 plus 15%	—$1,050
$3,252	—$6,221 . .	$402.00 plus 25%	—$3,252
$6,221	—$9,308 . .	$1,144.25 plus 28%	—$6,221
$9,308	—$16,360 . .	$2,008.61 plus 33%	—$9,308
$16,360	—$18,437 . .	$4,335.77 plus 35%	—$16,360
$18,437		$5,062.72 plus 39.6%	—$18,437

Source: IRS publication 15 (aka circular E) https://www.irs.gov/pub/irs-pdf/p15.pdf.

D Appendix – Predicted refund quintiles

Table D.2: Summary statistics for each predicted refund quintile

Q_i^j	Mean	p25	p50	p75
1	2,612	2,548	2,624	2,686
2	2,859	2,803	2,859	2,913
3	3,094	3,030	3,092	3,155
4	3,394	3,304	3,388	3,482
5	3,978	3,731	3,897	4,160
Total	3,167	2,793	3,071	3,449

E Online Appendix

E.1 Solution method

We use a combination of traditional value function iteration and the endogenous grid method to solve the maximization problem in three steps.

1. Step 1: Solve for optimal s and $\hat{w}$ when both are positive

   (a) Assume a grid of values for the control variable s_t

   (b) Conditional on s_t, use the FOC for $\widehat{w}_t$ to solve for $\widehat{w}_t$. $u'(c_t(\widehat{w}_t)) = \beta \int_\nu u'(c_{t+1}(\widehat{w}_t))\tilde{\phi}$

   (c) Calculate $(X_{t+1} = sR + \nu_t + Y_{t+1} - \tilde{\phi}\,[\tau(\nu_t + Y_t) - w(Y_t - a) - \widehat{w}_t])$ using the optimal $\widehat{w}_t$

   (d) Use the current iteration of the consumption function to solve for $c_{t+1}(X_{t+1})$

   (e) Use the EE to backout current period $c_t = u'^{-1}(\beta R \int_\nu u'(c_{t+1}))$

   (f) Use CoH LOM to calculate $X_t = c_t + s_t + \widehat{w}_t$

2. Step 2: Solve for $\hat{w}$ when $s = 0$

   (a) Specify a grid for X_t from 0 up until the minimum X_t solved in Step 1

   (b) Use the FOC for $\widehat{w}_t$ to solve for the optimal $\widehat{w}_t$ assuming $s = 0$

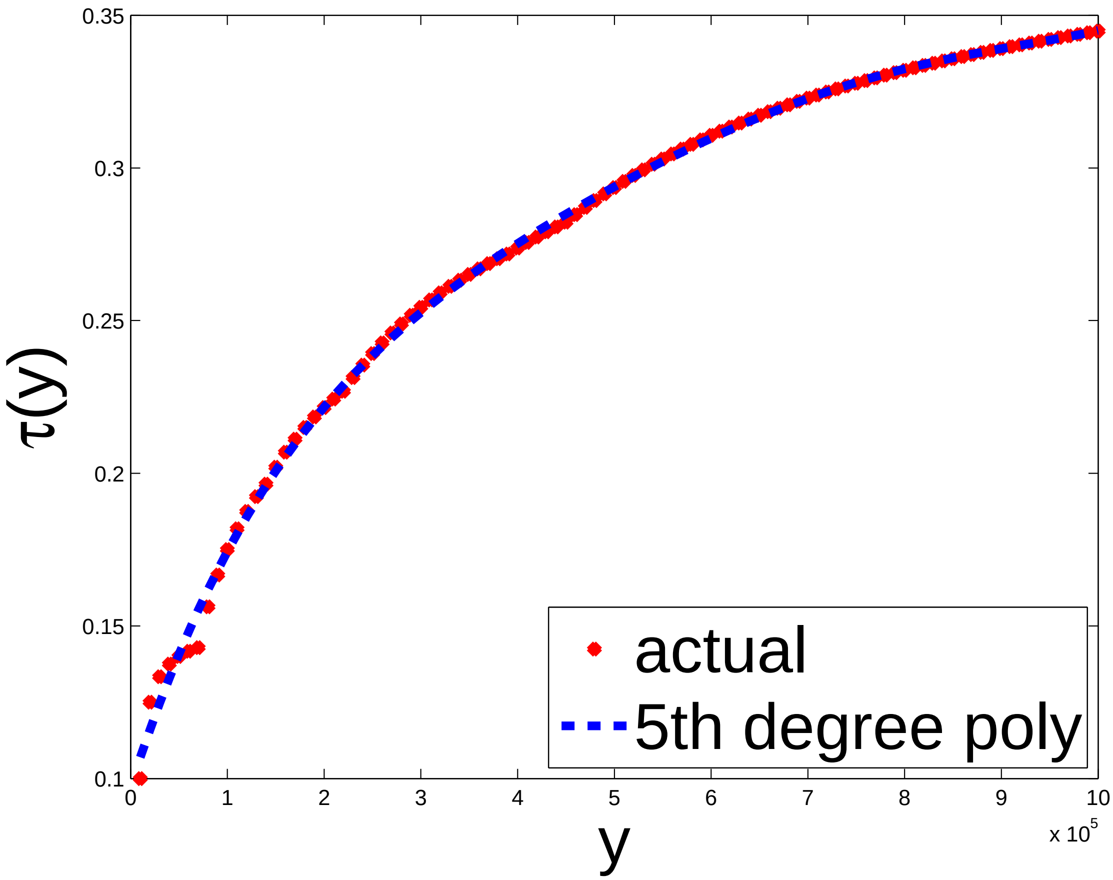
   (c) Conditional on X_t and $\widehat{w}_t$ we can back out what consumption will be

3. Step 3: Iterate until the consumption function $c(X_t)$ converges.

E.2 2013 tax schedule

We use the 2013 marginal tax schedule and calculate the average tax rates. We then approximate the ATR schedule with a 5th degree polynomial. The actual and smoothed schedule is shown in Figure 6.

Figure E.1: ATR schedule and approximation



E.3 Estimating the parameters of the income process

The following equations derive expressions for each of our income parameters as functions of the data.

α_Y

$$\mathbb{E}[y_{t,b}] = \mathbb{E}[\frac{Y_t}{26}] + \mathbb{E}[\epsilon^Y_{t,b}] \tag{14}$$

$$\alpha_Y = \mathbb{E}[y_{t,b}]26(1 - \rho_Y) \tag{15}$$

$\sigma^2_{\epsilon^Y}$

$$y_{t,b} - \bar{y}_t = \epsilon^Y_{t,b} \tag{16}$$
$$\sigma^2_{\epsilon^Y} = \mathbb{V}[y_{t,b} - \bar{y}_t] \tag{17}$$

$\sigma^2_{\nu^Y}$

$$\mathbb{V}[y_{t,b}] = \mathbb{V}\left[\frac{Y_t}{26}\right] + \mathbb{V}[\epsilon^Y_{t,b}] \tag{18}$$
$$\mathbb{V}[y_{t,b}] = \frac{\sigma^2_{\nu^Y}}{(26(1-\rho_Y))^2} + \sigma^2_{\epsilon^Y} \tag{19}$$
$$\sigma^2_{\nu^Y} = (\mathbb{V}[y_{t,b}] - \sigma^2_{\epsilon^Y})(26(1-\rho_Y))^2 \tag{20}$$

α_N

$$\mathbb{E}[n_{t,b}] = \mathbb{E}\left[\frac{N_t}{26}\right] \tag{21}$$
$$\alpha_N = \mathbb{E}[n_{t,b}]26(1-\rho_N) \tag{22}$$

$\sigma^2_{\epsilon^N}$

$$n_{t,b} - \bar{n}_t = \epsilon^N_{t,b} \tag{23}$$
$$\sigma^2_{\epsilon^N} = \mathbb{V}[n_{t,b} - \bar{n}_t] \tag{24}$$

$\sigma^2_{\nu^N}$

$$\mathbb{V}[n_{t,b}] = \mathbb{V}\left[\frac{N_t}{26}\right] + \mathbb{V}[\epsilon^N_{t,b}] \tag{25}$$
$$\mathbb{V}[n_{t,b}] = \frac{\sigma^2_{\nu^N}}{(26(1-\rho_N))^2} + \sigma^2_{\epsilon^N} \tag{26}$$
$$\sigma^2_{\nu^N} = (\mathbb{V}[n_{t,b}] - \sigma^2_{\epsilon^N})(26(1-\rho_N))^2 \tag{27}$$

E.4 Identifying paychecks

Keywords used to identify paychecks are "dir dep", "dirde p", "salary", "treas xxx fed", "fed sal", "payroll", "ayroll", "payrll", "payrl", "payrol", "pr payment", "adp", "dfas-cleveland", "dfasin" and DON'T include the keywords "ing direct","refund","direct deposit advance","dir dep adv."

E.5 Machine learning algorithm

Most transactions in the data do not contain direct information on spending category types. However, category types can be inferred from existing transaction data. In general, the mapping is not easy to construct. If a transaction is made at "McDonalds," it's easy to surmise that the category is "Fast Food Restaurants." However, it is much harder to identify smaller establishments such as "Bob's store." "Bob's store" may not uniquely identify an establishment in the data and it would take many hours of work to look up exactly what types of goods these smaller establishments sell. Luckily, the merchant category code (MCC) is observed for two account providers in the data. MCCs are four digit codes used by credit card companies to classify spending and are also recognized by the U.S. Internal Revenue Service for tax reporting purposes. If an individual uses an account provider that provides MCC information "Bob's store" will map into a spending category type.

The mapping from transaction data to MCC can be represented as $Y = f(X)$ where Y represents a vector of MCC codes and X represents a vector of transactions data. The data is partitioned into two sets based on whether Y is known or not.[14] The sets are also commonly referred to as training and prediction sets. The strategy is to then estimate the mapping $\hat{f}(\cdot)$ from (Y_1, X_1) and predict $\hat{Y}_0 = \hat{f}(X_0)$.

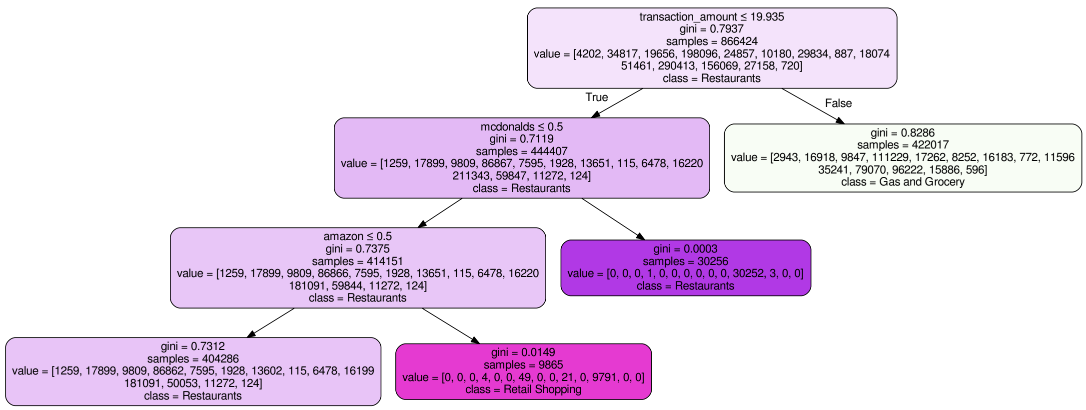
One option for the mapping is to use the multinomial logit model since the dependent variable is a categorical variable with no cardinal meaning. However, this approach is not well suited to textual data because each word would need its own dummy variable. Furthermore, interactions may be important for classifying spending categories. For example "jack in the box" refers to a fast food chain while "jack s surf shop" refers to a retail store. Including a dummy for each word can lead to about 300,000 variables. Including interaction terms will cause the number of variables to grow exponentially and will typically be unfeasible to estimate.

In order to handle the textual nature of the data I use a machine learning algorithm called random forest. A random forest model is composed of many decision trees that map transaction data to MCCs. This mapping is created by splitting the sample up into nodes depending on the features of the data. For example, for transactions that have the keyword

[14] Y_0 represents the set where Y is not known and Y_1 represents the set where Y is known.

"McDonalds" and transaction amounts less that \$20, the majority of the transactions are associated with a MCC that represents fast food. To better understand how the decision tree works, Figure E.2 shows an example. The top node represents the state of the data before any splits have been made. The first row "transaction_amount $\leq$ 19.935" represents the splitting criteria of the first node. The second row is the Gini measure which is explained below. The third row show that there are 866,424 total transactions to be classified in the sample. The fourth row "value=[4202,34817,...,27158,720]" shows the number of transactions in each spending category. The last row represents the majority class in this node. Because "Restaurants" has the highest number of transactions, assigning a random transaction to this category minimizes the categorization error without knowing any information about the transaction. At each node in the tree, the sample is split based on a feature. For example, the first split will be based on whether the transaction amount is $\leq$ 19.935. The left node represents all the transactions for which the statement is true and vice versa. Transactions $\leq$ 19.935 are more likely to be "Restraunts" spending while transactions $>$ 19.934 are more likely to be "Gas and Grocery." In our example, the sample is split further to the left of the tree. Transactions with the string "mcdonalds" are virtually guaranteed to be "Restaurant" spending. A further split shows that the string "amazon" is almost perfectly correlated with the category "Retail Shopping." How does the algorithm decide which features to split the sample on? The basic intuition is that the algorithm should split the sample based on features that lead to the largest disparities in the different groups. For example, transactions that have the word "mcdonalds" will tend to split the sample into fast food and non-fast food transactions so it is a good feature to split on. Conversely, "bob" is not a very good feature to split on because it can represent a multitude of different types of spending depending on what the other features are.

Figure E.2: Decision tree example



I state the procedure more formally by adapting the notation used in (Pedregosa et al., 2011). Define the possible features as vectors $X_i \in R^n$ and the spending categories as vector $y \in R^l$. Let the data at node m be presented by Q. For each candidate split $\theta = (j, t_m)$ consisting of a feature j and threshold t_m, partition the data into $Q_{left}(\theta)$ and $Q_{right}(\theta)$ subsets so that

$$Q_{left}(\theta) = (X, y)|x_j \leq t_m \tag{28}$$

$$Q_{right}(\theta) = Q \setminus Q_{left}(\theta) \tag{29}$$

The goal is then to split the data at each node in the starkest way possible. A popular quantitative measure of this idea is called the Gini criteria and is represented by

$$H(X_m) = \sum_k p_{mk}(1 - p_{mk}) \tag{30}$$

where $p_{mk} = 1/N_m \sum_{x_i \in R_m} \mathbb{I}(y_i = k)$ represents the proportion of category k observations in node m.

If there are only two categories, the function is is minimized at 0 when the transactions are perfectly split into the two categories[15] and maximized when the transactions are evenly

[15]because 0*1 + 1*0 = 0.

split between the two categories.[16]

Therefore, the algorithm should choose the feature to split on that minimizes the Gini measure at node m

$$\theta^* = argmin_\theta \frac{n_{left}}{N_m} H(Q_{left}(\theta)) + \frac{n_{right}}{N_m} H(Q_{right}(\theta)) \tag{31}$$

The algorithm acts recursively so the same procedure is performed on $Q_{left}(\theta^*)$ and $Q_{right}(\theta^*)$ until a user-provided stopping criteria is reached. The final outcome is a decision rule $\hat{f}(\cdot)$ that maps features in the transaction data to spending categories.

This example shows that decision trees are much more effective in mapping high dimensional data that includes text to spending categories. However, fitting just one tree might lead to over-fitting. Therefore, a random forest fits many trees by bootstrapping the samples of the original data and also randomly selecting the features used in the decision tree. With the proliferation of processing power, each tree can be fit in parallel and the final decision rule is based on all the decision trees. The most common rule is take the majority decision of all the trees that are fit.

[16] because 0.5*0.5 + 0.5*0.5 = 0.5.