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

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

August 26, 2019

Abstract

Having low liquidity and a high marginal propensity to consume (MPC) are tightly linked. This paper analyzes this linkage in the context of income tax withholding and refunds. A theory of rational cash management with income uncertainty endogenizes the relationship between illiquidity and the MPC, which accounts for the finding that households tend to spend tax refunds as if they valued liquidity, yet do not act to increase liquidity by reducing their income tax withholding. The theory is supported by individual-level evidence, including a positive correlation between the size of tax refunds and the MPC out of those refunds.

JEL Classification Numbers: D14, H24, E21

Keywords: Income uncertainty, liquidity, account data, marginal propensity to consume

*We thank William Boning, James Hines, Damon Jones, Joel Slemrod, Basit Zafar, and participants at 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

Many households, even those with higher incomes, maintain low liquid assets balances and exhibit substantial sensitivity to transitory changes in income. The behavior around income tax withholding and refunds represents an important example. Nearly a third of all personal income tax collected by the US government is later returned in the form of tax refunds. Households tend to spend substantial fractions of those refunds when they arrive.[1] Hence, households choose to reduce their liquid assets by making interest-free loans to the government in the form of overwithholding, and then rapidly spend a high fraction of the loan repayments when they receive them in the form of tax refunds.

Why do households with substantial incomes choose to reduce their liquid assets buffer, yet respond to income fluctuations as if they would value a cushion against shocks? High degrees of impatience might explain both a lack of liquid savings and the sensitivity of consumption to changes in income, but are hard to reconcile with the foreward-looking decision to save through overwithholding.

This paper develops and evaluates a theory of household liquid assets management with income shocks that accounts for both the prevalence of income tax refunds and the sensitivity of spending to the arrival of refunds. The theory maintains the standard assumptions of modern, benchmark models of consumption and savings. It takes account of the volatility of both paycheck and non-paycheck income, with the latter generally not subject to tax withholding at the source. A central mechanism of the theory is that endogenous liquidity constraints 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, like most modern benchmark theories, that the marginal propensity to consume (MPC) out of transitory income shocks is positive. Distinctively, the model predicts the MPC will be especially high out of tax refunds, which tend to arrive when cash-on-hand is low.

We evaluate the empirical relevance of the theory with administrative account data on income, spending, and refunds. These data show, with a conventional calibration of the

[1] Several studies estimate economically important sensitivity of spending to federal tax refunds (Souleles (1999), Baugh et al. (2018), Gelman (2018)).

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 account data also provide evidence of the central mechanisms of the theory. First, individual-level 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 higher MPCs out of tax refunds.

To be more precise, the theory builds on precautionary savings models. We assume 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 made before the realization of all income uncertainty. In principle, households need not withhold additional taxes, they could just save, and use those savings to pay any tax liability on non-paycheck income when taxes are due. Doing so is costly, however, because the IRS charges interest and penalties 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 overwithholding on average.

The endogenous liquidity constraints that emerge from precautionary savings motives also explain the sensitivity of spending to the arrival tax refunds. As is standard, in the model households 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. This optimization implies, with a conventional calibration of the model parameters, positive MPCs out of transitory income shocks.

The model also shows how negative income shocks will lower cash on hand and, due to the concavity of the consumption function (Zeldes (1989a), Carroll and Kimball (1996)), raise the MPC 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 MPC is higher. Optimal behavior thus implies that the MPC is especially high when refunds arrive. The larger the refund the higher the MPC.

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. Non-paycheck income has an average standard deviation of \$19,879, compared to \$18,490 for paycheck income. With a conventional calibration of the rest of the model's parameters, we can calibrate the time discount factor to 0.982 to exactly match the average final tax settlement in the data, a \$1,704 refund.

The same data allow us to present indidivdual-level 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. In addition, the account data show that those whose non-paycheck income shares predict larger refunds also have larger MPCs out of tax refunds. As predicted by the model, the MPC out of tax refunds rises from the bottom to the top quintile of the refund distribution.

The prevalence of income tax refunds and the evidence of the 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 advances a growing body of evidence revealing the importance of liquid assets management for understanding how households respond to income and spending shocks.

2 Related Literature

The paper is related to three strands of literature. The first concerns the relationship between liquidity and consumption. The second and third relate to the overwithholding of income taxes and the response of spending to tax refunds. These last two strands of literature have been largely disconnected. One contribution of the paper is to provide a single model that explains both phenomena.

2.1 On Liquidity and Consumption

The analysis of liquid assets management presented here relates to a growing literature that studies, often using innovative data sources, the relationships between liquidity and consumption. Examples include Braxton, Phillips and Herkenhoff (2018), which links administrative employment and credit bureau data to study consumption smoothing during unemployment, and Herkenhoff (2019), which uses several data sources to show how increasing access to credit led to an increased ability to smooth consumption during unemployment.

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. They 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 to 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.

Our analysis develops a distinct 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 qualitatively consistent with the finding in Baugh et al. (2018) that spending reacts less to a tax payment than a tax refund, but provides a different 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.

2.2 On Overwithholding

Jones (2012) generalizes a theory of overwithholding based on the logic in Highfill, Thorson and Weber (1998). Both papers model a "timing problem" like the one we study: workers must choose their levels of withholding before knowing what their incomes and tax liabilities will be. Highfill, Thorson and Weber (1998) explain overwithholding as the optimal response to the wedge between the opportunity costs of overwithholding and the IRS penalties of being under-withheld. Jones (2012) determines that this wedge is insufficient to justify the prevalence of overwithholding; 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.[2]

Alternative (behavioral) explanations interpret overwithholding as a form of forced savings that helps workers deal with problems of self-control. Thaler (1994), Neumark (1995), and Fennell (2006) see overwithholding 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 overwithholding. Investigating another source of tax refunds, 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

[2] In recent work, Boning (2018) studies an unexpected shock that led to underwithholding and finds that some households, likely due to inattention, make late final settlements as a consequence.

at the time of tax filing.

By incorporating volatile income not subject to withholding at the source, we find that a model of workers with time-consistent and state-independent preferences can account quantitatively for the large average refunds observed in the data. Importantly, our model predicts large refunds without inertia or defaults biased toward overwithholding. [3]

2.3 On the Spending Response to Refunds

The tendency to spend large fractions of tax refunds around the time they arrive has been documented by Souleles (1999), Gelman (2018), and Baugh et al. (2018) and is qualitatively similar to the spending responses to related income changes such as in Hsieh (2003) and Kueng (2018) (Alaska Fund payments), Parker (1999) (changes in payroll taxes), and Wilcox (1989) (anticipated changes in Social Security payments). For purposes of analyzing the spending response to refunds, we will treat the income change they produce as partially unanticipated. This treatment is motivated by the fact that, in recent years, electronic tax returns could be filed no earlier than mid-January, and 90% of refunds arrive within 21 days of filing. 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.

When tax refunds are transitory but to some extent unanticipated income, then modern benchmark models (Zeldes (1989a), Carroll and Kimball (1996)) predict a positive MPC out of this income. Understood in this way, it is not puzzling that spending responds to the arrival of the tax refunds. Indeed, as the model developed below reveals, we should anticipate that spending responds more to larger refunds.

[3] There are many reasons to receive a tax refund that we do not model. Recipients of the Earned Income Tax Credit (EITC), for example, are almost certain to receive a tax refund because, since 2010, the credit cannot be paid out during the course of year. Our focus is on higher income households who are likely ineligible for the EITC.

3 Institutions and Model

In this section we describe and model key institutions that drive incentives for liquid assets management, income tax withholding, and consumption. To summarize the key elements: (1) Taxes on paycheck income are withheld at the source; (2) workers must remit taxes on non-paycheck income throughout the year or else owe interest; (3) interest on taxes owed exceeds the rate of return on low risk and liquid assets in the private market; (4) withheld taxes are illiquid, once remitted they cannot be accessed until taxes are filed; (5) overwithheld taxes earn no interest; and (6) we assume the withholding decision is made before all income uncertainty is resolved.

3.1 Background on Tax Withholding

Federal income tax liability is determined at annual frequency. Taxes on wage and salary income 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.[4,5]

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.[6]

Individuals with taxable income must file a tax return, or a request for an extension, by a mid-April deadline. There are costs assessed for underwithholding. Even if the tax bill is paid in full on the filing deadline, the IRS charges interest for underwithholding throughout the year. In particular, unpaid tax is subject to interest at the federal "short-term" interest rate plus 3%. So a taxpayer who is underwithheld by $10,000 and pays his tax bill on April 15, would face an interest rate of approximately 3.02% on the $10,000 he underwithheld (using the 0.2% short-term interest rate during the time period of this study). Assuming the withholding should have been done evenly throughout the year, this would amount to approximately $190 in interest.

There are also "failure to pay" penalties. If the tax payment is received after the filing

[4] If the tax liability on these other sources of income is more than $1,000, then estimated taxes must be paid quarterly. If those estimated taxes are not paid on time, then interest and late penalties may apply. We abstract from the late penalties and focus only on interest owed on underpayment. 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.

[5] 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.

[6] 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 5.4.1. We are grateful to Damon Jones for highlighting this effect for us.

deadline, then there is a penalty of 0.5% of the unpaid tax assessed every month that the remaining tax goes unpaid. (For purposes of calculating the penalty, filing for an extension does not extend the time to pay.) Thus, this same taxpayer who is underwithheld by $10,000 and remitted those unpaid taxes only on October 15 would owe approximately $303 in penalties. There are also relatively large penalties for late filing. Tax payers therefore face strong incentives to file on time, even if they have unpaid taxes.

We focus on tax payments and liabilities, not filing *per se*. We also assume throughout the analysis below that tax payments are made by the filing deadline and thus only penalty interest payments (not "failure to pay" penalties) obtain. Safe harbor provisions exempt some households from interest payments on underwithholding. In particular, no interest applies if the unpaid amount equals less than $1,000 total, or represents less than 10% of total taxes owed in current tax year, or withholding equals the total liability in the previous year (110% of the liability over certain income thresholds). In the model developed below, we account for the safe harbor provided by the 10% rule.[7]

3.2 Model

We model the interactions between income uncertainty, liquid assets management, and consumption in the context of income tax withholding and spending out of tax refunds. In the model 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 "non-paycheck" income, N_t, that is not subject to withholding at the source. To simplify the 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. Non-paycheck income is stochastic and is realized at the end of each period.

The key friction in the model is that withholding decisions must be made before the resolution of all income uncertainty. If this were not the case, then upon the resolution

[7]We find the $1,000 exception is not relevant for the average tax payer in the data. Modeling the safe harbor of the previous year's tax liability adds substantial complication with likely ambiguous effects on tax refunds given the 10% rule is already modeled.

of uncertainty workers would withhold the correct amount of taxes and it would never be optimal to overwithhold. In particular, we assume non-paycheck income is available to be spent only in the next period. As in the actual income tax system, we assume taxes on period t income are due at the beginning of period $t+1$. Note that a period is a year to correspond to annual calculation of tax liability and the time subscript t refers to the tax year.

Tax liability and withholding are important elements of the model. We specify the liability and tax withholding functions to capture key features of the tax system and, in the model simulations, calibrate these functions to match tax rules. The total tax liability in tax year t

$$\tau(Y_t + N_t)$$

is a function of annual total income. It is a nonlinear function that reflects the progressivity of the U.S. tax system.

The withholding function determines how much is withheld from paycheck income Y_t. The IRS sets the withholding table so that annual withholding equals annual tax liability if withheld income is the only source of income. The withholding schedule is therefore determined so that

$$W(Y_t) = \tau(Y_t),$$

meaning that withholding equals tax liability in the absence of non-paycheck income.[8] In that case, individuals would neither receive a refund nor owe any taxes. Given that individuals also earn non-paycheck income that is not subject to withholding, we allow individuals to make an additional withholding decision meant to offset some of the tax liability from non-paycheck income. $\widehat{W}_t \geq 0$ 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

[8]The model in this section is written with annual periodicity of paycheck. In the model, we impose that the annual withholding function and tax liability function are the same. In practice, the withholding table has two components, allowances and a piecewise linear function, and is implemented on a per-pay-period basis. In the empirical work with individual data, we incorporate these details into the analysis. In the theoretical model, without loss of generality, we elide withholding and estimated tax payments. In the text, when we refer to withholding, it should be understood to refer to both withholding from paycheck income and estimated tax payments made during the tax year.

of current income, and like withholding are presumed to be determined in advance of the realization on non-paycheck income.

The state variable for the worker is beginning-of-period "cash on hand," X_t, and consists of the current period's after-withholding paycheck income, plus the previous period's non-paycheck income (N_{t-1}), savings (S_{t-1}), and the final settlement (T_{t-1}) of the previous period's tax liability. The final settlement is positive if withholding W and other payments $\hat{W}$ are less than the tax liability for the previous years income. If it is negative, the taxpayer gets a refund. We assume the worker makes the final settlement from cash on hand. 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, non-paycheck income, N_t, is realized, and cash on hand for the subsequent period is determined.

Formally, cash on hand, X_{t+1}, evolves according to

$$X_{t+1} = S_t R + Y_{t+1} - W(Y_{t+1}) + N_t - T_t$$

where $S_t = X_t - C_t - \widehat{W}_t$, $R > 1$ is the rate of return on savings, T_t is the final settlement (refund if negative) based on year t income. Recall that T_t depends on non-paycheck income and is therefore realized only after period t consumption is completed. The final settlement also reflects a safe harbor provision where no interest is owed if withholding is at least 90 percent of the tax liability. Within the safe harbor, the final settlement is simply the tax liability net of withholding. Otherwise, the taxpayer has an additional liability equal to the interest rate ϕ times the shortfall between required withholding (90 percent of the tax liability) minus actual withholding.

The final settlement therefore satisfies:

$$T_t = \begin{cases} \tau(Y_t + N_t) - W(Y_t) - \widehat{W}_t & \text{if } 0.9\tau\left(Y_t + N_t\right) \leq W\left(Y_t\right) + \widehat{W}_t \\ \tau(Y_t + N_t) - W(Y_t) - \widehat{W}_t + \phi[0.9\ \tau(Y_t + N_t) - W(Y_t) - \widehat{W}_t], & \text{if } 0.9\tau\left(Y_t + N_t\right) > W\left(Y_t\right) + \widehat{W}_t \end{cases}$$

It is convenient to collapse the compound expression for the final settlement into the single equation

$$T_t = \tilde{\phi}[\tau(Y_t + N_t) - W(Y_t) - \widehat{W}_t] - [(\tilde{\phi} - 1)0.1\tau(Y_t + N_t)]$$

where $\tilde{\phi}$ is the penalty interest rate function that has a kink at the point where the individual is underwithheld,

$$\tilde{\phi} = \begin{cases} 1 & \text{if } 0.9\tau\left(Y_t + N_t\right) \leq W\left(Y_t\right) + \widehat{W}_t \\ 1 + \phi, & \text{if } 0.9\tau\left(Y_t + N_t\right) > W\left(Y_t\right) + \widehat{W}_t \end{cases}$$

This formulation captures the key asymmetry that drives the model. Individuals need to pay a penalty interest rate when they sufficiently underwithhold, but they do not receive any interest on the amount that they overwithhold.

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

$$\begin{aligned} V(X_t) &= \max_{S_t, \widehat{W}_t} u(C_t) + \beta \int_{N_t} V(X_{t+1}) \qquad (1) \\ s.t.\ C_t &= X_t - \widehat{W}_t - S_t \\ X_{t+1} &= S_t R + Y_{t+1} - W(Y_{t+1}) + N_t - T_t \\ T_t &= \tilde{\phi}[\tau(Y_t + N_t) - W(Y_t) - \widehat{W}_t] - [(\tilde{\phi} - 1)0.1\tau(Y_t + N_t)] \\ C_t, \widehat{W}_t, S_t &\geq 0 \end{aligned}$$

We do not consider the possibility of borrowing, so savings is constrained to be non-negative.[9]

3.3 Optimality

Systematic overwithholding of income taxes represents a deliberate reduction in liquidity. Overwithholding is a zero-interest loan to the government, a loan that can be arbitraged with any interest-bearing account. Given this intentional reduction in liquidity, the sensitivity of spending to the arrival of refunds is puzzling because it indicates the household values the

[9] Borrowing to make tax payments is very likely dominated by overwithholding, or even paying interest on underwithholding, because interest rates on (unsecured) credit are typically much higher than the short-term rate plus 3% charged by the IRS.

liquidity it has chosen to give away in the form of overwithholding. The optimality conditions for problem (1) reveal, however, the rationality of this illiquidity. They show the incentives to overwithhold and a simple logic driving both refunds and the sensitivity of spending to those refunds.

3.3.1 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 non-paycheck 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 sufficiently underwithheld.

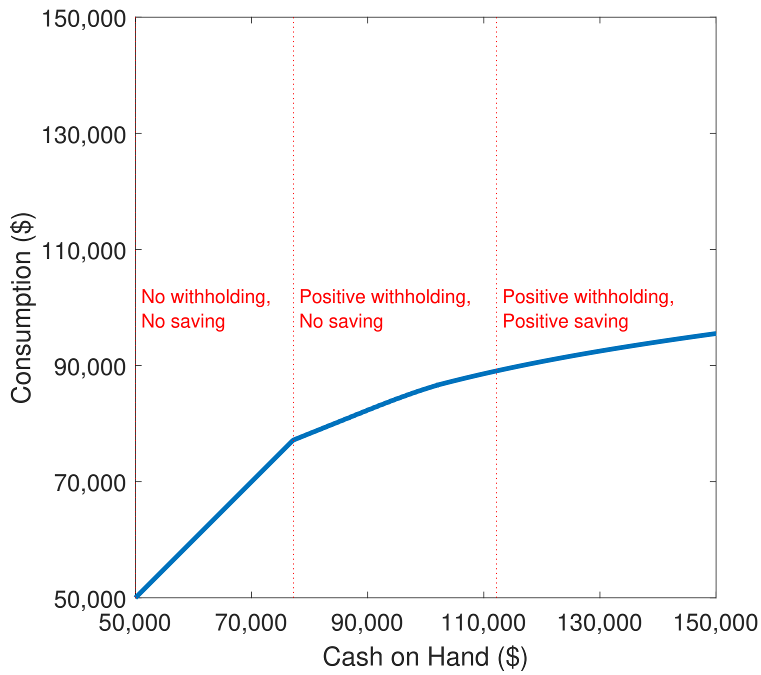
The two optimality conditions thus reveal the portfolio problem behind the withholding decision. Private saving and additional withholding are like two different assets; putting \$1 in additional withholding returns $\tilde{\phi}$ while saving returns R. It follows that, in the absence of income uncertainty, 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.

3.3.2 Optimal Responses of Spending to Income

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 non-paycheck income uncertainty motivates precautionary "savings" in the form of overwithholding. As shown in Zeldes (1989a) and Carroll and Kimball (1996), for a broad class of preferences, income uncertainty implies consumption is responsive to transitory income and the consumption function that maps existing financial assets into the optimal level consumption, is concave. The effect on consumption of transitory income depends 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 5. Note first that in both panels (a) and (b), the slope of the consumption function (the MPC) is everywhere positive. The positive MPC reflects the costs, in terms of foregone consumption, of accumulating a buffer sufficient to smooth consumption perfectly.

Figure 1: Consumption function

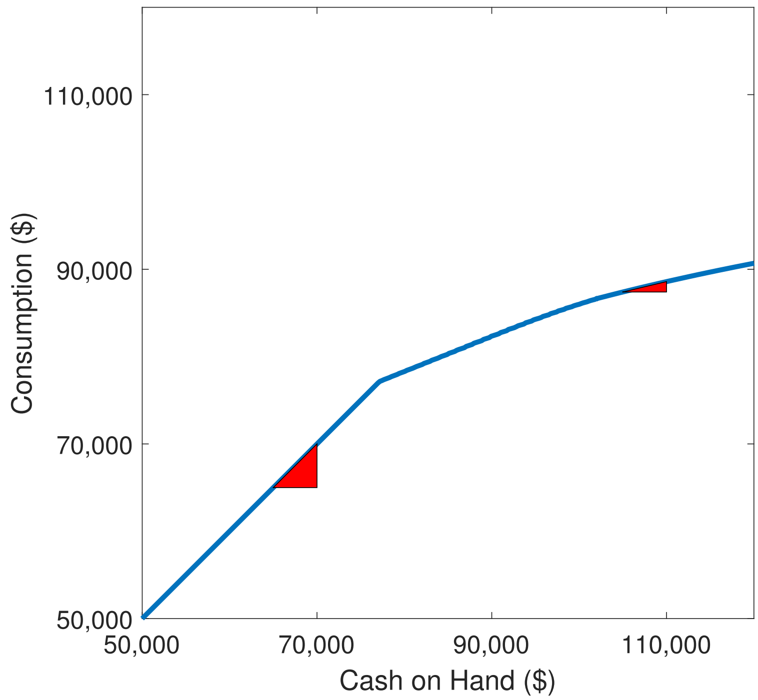


Notes: The consumption function shows optimal consumption as a function of cash on hand where cash on hand is the sum of current period after-tax paycheck income, saving, last period's non-paycheck income and tax liability.

Second, both panels (a) and (b) show the endogenous liquidity constraints that emerge from income uncertainty as consumption is a concave function of cash on hand. In particular, consumption rises 1-for-1 with after tax income when cash on hand is sufficiently low and households live hand to mouth. Once cash-on-hand exceeds a threshold, however, the MPC out of income declines. Panel (a) highlights the different regions of the consumption function. When cash on hand is very low (and the marginal utility of consumption is very high), the consumption function is linear because hand-to-mouth consumers do not save at all ($\hat{W} = 0, S = 0$) and consume all their resources despite the fact that the liquidity constraint does not literally bind. As cash on hand increases, the consumer has enough resources to devote to saving. As discussed in the previous section, consumers should always start with putting their extra resources in additional withholding ($\hat{W}$) instead of saving (S) because the the return is $\phi > R$ when an individual is underwithheld. In this second region of the consumption function $\hat{W} > 0$ and $S = 0$. As cash on hand increases further, the return to additional withholding is declining. Given the probability distribution of non-paycheck income (N_t), devoting an extra dollar to additional withholding will not reduce the chance of being underwithheld and the return on saving will dominate. This represents the last region of the consumption function where $\hat{W} > 0$ and $S > 0$.

Panel (b) of Figure 1 illustrates the quantitative importance of the concavity. In the figure, we consider deviations from an initial cash on hand level of \$85,000. We consider the marginal propensity to consume from a \$10,000 tax refund, which is larger than typical to make the figure readable. Panel (b) shows the effect of receiving the refund in two situations, one where cash in hand is \$20,000 below the initial level (the triangle to the left) and one where cash in hand is \$20,000 above the initial level (the triangle to the right). The base of each triangle is the \$10,000 refund and the height is the spending due to it. The MPC in the low-liquidity state is about one. In contrast, the MPC in the high-liquidity state less than a quarter as large.

Figure 2: Consumption function: concavity



Notes: The consumption function shows optimal consumption as a function of cash on hand where cash on hand is the sum of current period after-tax paycheck income, saving, last period's non-paycheck income and tax liability.

The higher MPC in the low-liquidity state illustrates an important mechanism of this paper's model. A negative surprise for non-paycheck income could lead to simultaneously having low liquidity and a big refund. The big refund arises as a consequence of the negative income shock because the previous-years tax payments were made in anticipation of higher non-paycheck income. On the other hand, there is no reason for the refund and income shocks to be correlated in the high liquidity state.

4 Data and Estimates of the Income Process

The quantitative relevance of the proposed mechanisms behind both tax refunds and the sensitivity of spending to refunds depends on 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 non-paycheck income processes.

4.1 Data Source

To estimate the paycheck and non-paycheck 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.[10] 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. We used these data previously to study the spending response to anticipated income, stratified by spending, income and liquidity (Gelman et al., 2014) and households' high-frequency responses to shocks such as the government shutdown (Gelman et al., 2018). Similar account data from other apps have been used in Baugh et al. (2018), Baker (2017), Baker and Yannelis (2017), Kuchler and Pagel (2018), Ganong and Noel (2019), and Kueng (2018).

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 from December 2012 to July 2016.

Because data we use are "naturally-occurring" or "non-designed" (aka "big 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 steps to assess whether the sample is 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

[10]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 project researchers. Analysis is carried out on data aggregated and normalized at the individual level. Only aggregated results are reported.

a diverse population.[11]

4.2 Sample and measurement

From the population of app users, 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 and income, we limit attention to app users who link all (or most) of their accounts to the app, and generate a long time series of observations. To study the importance of both paycheck and non-paycheck income, we also restrict attention to app users who receive regular paychecks throughout most of the time we observe them in our data. Regular paychecks are identified through textual analysis of the transaction description and regularity of amount and timing of the payments. 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. For ease of analysis of the individual data, we limit the sample to individuals on bi-weekly payrolls. (Approximately 61 percent of those with payroll income are paid bi-weekly.) There are 62,946 individuals in the panel with roughly 3.5 years of observations per individual on average.

4.3 Variable definitions

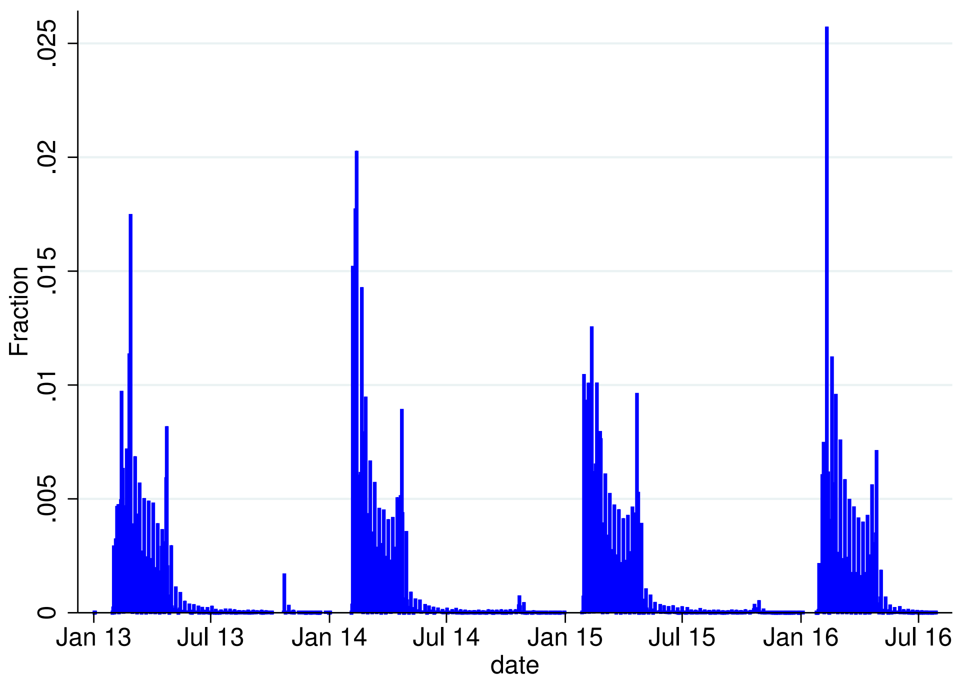
Key variables of the model are tax refunds, non-durable expenditure, and paycheck and non-paycheck income. We measure those variables from the transaction data as follows:

[11] To further evaluate the validity of the sample, after we define key variables in subsection 4.3, 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.

4.3.1 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"). Figure 3 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 3: Federal tax refund time series



Notes: The figure contains data from 142,315 tax refunds and 49,520 individuals who received at least one refund during the period observed.

4.3.2 Non-durable Expenditure

Another primary focus of our analysis is the spending response of individuals to the arrival of 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

H 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 ML algorithm uses a subset of the data where the MCC is recorded as a training dataset in order to create a mapping from transaction description to MCCs.[12] After training the ML algorithm on the data where the MCC is recorded, we apply the algorithm to the rest of the data set. We use spending on restaurants, groceries, gasoline, entertainment, and services to measure consumption C_t in the model.

4.3.3 Income

We define income as the sum of all inflows to checking and saving accounts minus transfers between accounts. From this measure of income, paycheck income is defined as the inflows from paychecks identified using an algorithm detailed in Appendix section A.2. This measure of paycheck income is net of deductions including income and payroll tax withholding. To obtain before-tax paycheck income, we add estimates of state and federal income taxes and federal payroll taxes. See Appendix B for specifics. All income not classified as paycheck income Y_t is defined as non-paycheck income N_t.

4.4 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 where the refund is set 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. These average spending numbers are, respectively, 43% and 33% higher than average annual spending in the Consumer Expenditure Survey.

[12]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.

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

Table 1: Comparing App and External Sources: Means

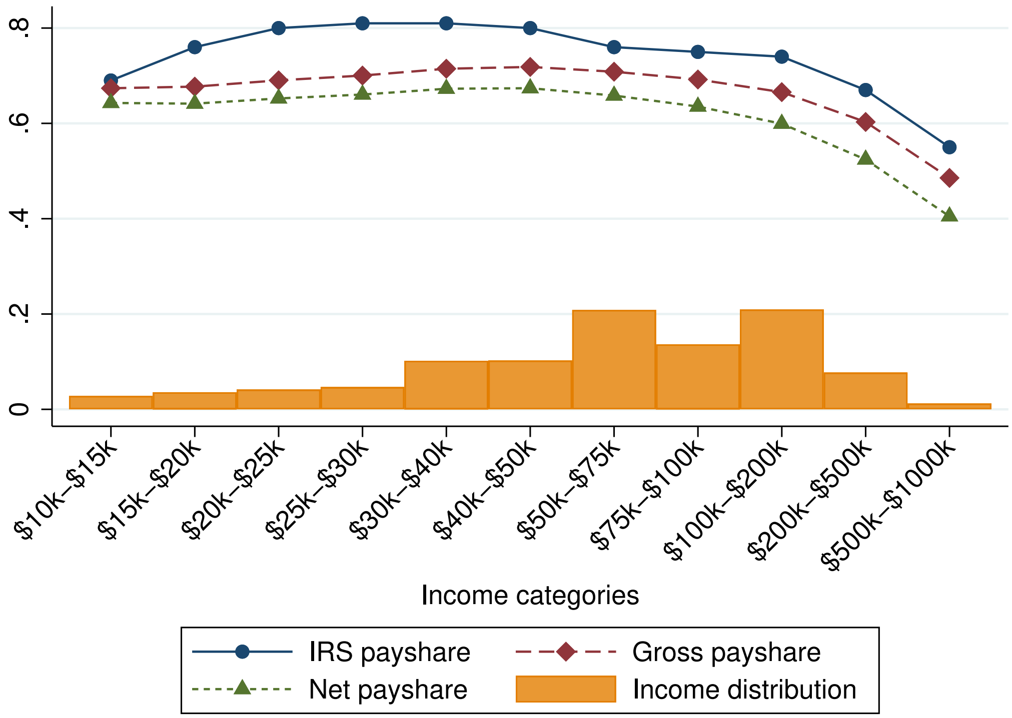
	App		
	Full sample	Received refunds	External sources
Tax refund (\$)	1,704	3,184	2,778
Spending (\$)	83,253	77,854	58,410
Paycheck income (\$)	68,226	67,415	46,224
Non-paycheck income (\$)	38,764	36,607	20,603
Paycheck share	.68	.68	.69
NxT	251,784	134,752	
N	62,946	49,520	

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 non-paycheck 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. Figure 4 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 the app 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 4: Paycheck share comparison across income groups



Notes: Payshare is fraction of wage and salary income in total income. IRS payshare is calculated from IRS, Statistics of Income Division, Publication 1304. Net payshare is calculated using paycheck income (net of withholding and other deductions) as reported in the App. Gross payshare is our calculation grossing up net paycheck income as described in Appendix B.

4.5 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 non-paycheck 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 3.2.

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.[13] On top of these annual components, bi-weekly income is subject to serially-uncorrelated noise. Section G 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 annual level.

[13] We considered more general time series processes but ultimately decided on a parsimonious specification. Because our time-series sample only consists of less than four years, there is little hope of estimating more elaborate annual income processes with sufficient precision. In particular, with only four years of data, it is not possible to estimate the persistence of annual income. Serially correlated income would complicate the solution of the model, but not change its main message.

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%.

To assess the validity of these estimates of income volatility, we would like to compare them to existing estimates from other studies. The literature uses a variety of statistical models and methods to measure income volatility, but many use the standard deviation of the first difference in log total income. In the app data, the standard deviation of the first difference of total income is 0.50. Because we have incomplete data in 2016, we extrapolate the missing months based on month fixed effects estimated from 2013-2015. If, instead, we drop data from 2016 then the estimated standard deviation is 0.46.

Our estimates of income volatility are similar to those in other studies that use administrative records on income. Two prominent examples are Debacker et al. (2013) and Guvenen, Ozkan and Song (2014). Debacker et al. (2013) use panel income data from tax returns over the years 1997-2009. In their Figure 2, they show that the standard deviation of the first difference of log male earnings has varied from roughly 0.40 to 0.44 over their sample. Their Figure 6 analyzes pre-tax household income and shows very similar levels of volatility. Guvenen, Ozkan and Song (2014) use earnings histories from US Social Security Administration records covering 1978-2011. Their Figure 5 shows that the standard deviation of the first difference of log earnings varies from roughly 0.50 to 0.60 over their time period.

5 Results

We next present the empirical results. We first calibrate the model and show how it explains large refunds on average. We then show that the model predicts workers spending a large fraction of their refunds, on average, and that the larger the refund the higher the MPC. Using our individual-level data, we then estimate the empirical relationship between tax refunds and variation in income at the individual level.

5.1 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, we assume a coefficient of risk aversion equal to one (log utility). We calibrate the annual time discount factor to 0.98212 to match the average final tax settlement observed in the data. We approximate the tax liability schedule as a function of income with a fifth degree polynomial.

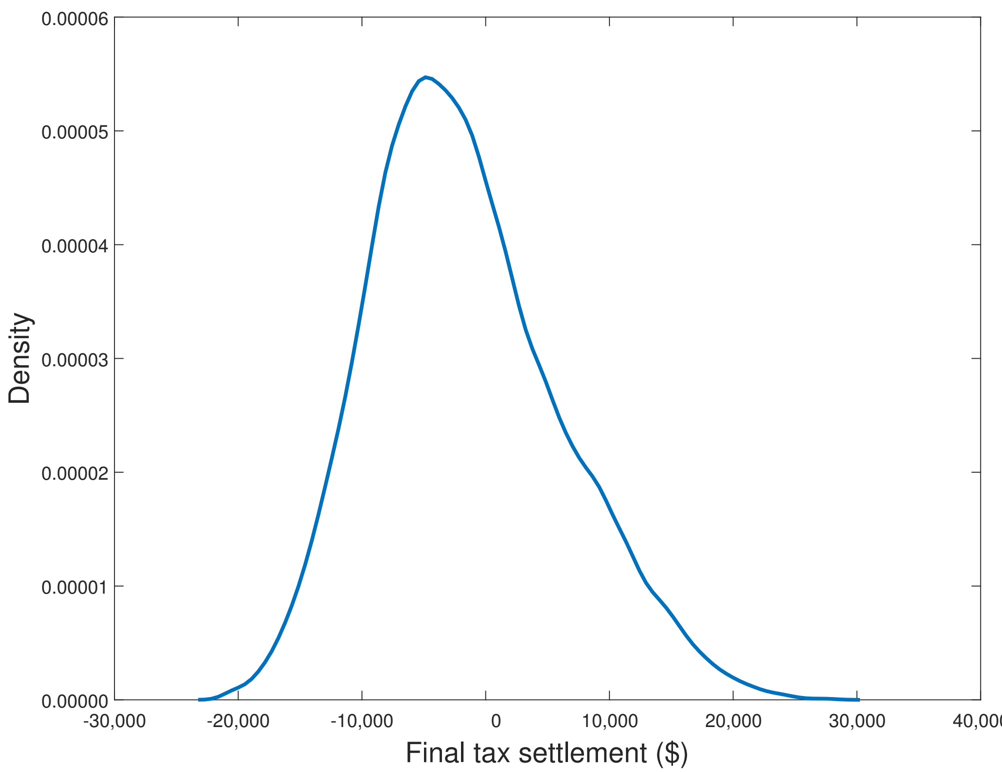
Table 3: Model calibration

Parameter	Value	Notes
$u(C)$	$\frac{C^{1-\theta}}{1-\theta}$	utility function
θ	1	risk aversion
β	0.98212	discount factor (calibrated to match empirical final tax settlement)
T_t	\$1,704	mean final tax settlement (targeted)
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
$\tau(Y+N)$		tax liability schedule approx (see Appendix A.6)
R	1.002	1-year treasury yield
ϕ	0.032	(R-1) + 3%

5.2 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, discarding the first 1,000 periods, and record the final tax settlement in each period. Figure 5 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. The specific realization of the tax refund/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.

Figure 5: 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.

We calibrated only the discount factor to match the average final tax settlement observed in the data. Despite the simple structure of the model and just one degree of freedom, the model is successful at fitting that moment, $1,704, precisely with a standard discount factor of 0.982. These results show that, with a conventional calibration of other model parameters, the estimated level and variation in non-paycheck income is sufficient to justify the average final tax settlement in the data. Table 4 shows this predicted average and some of the other statistics of the final tax settlement distribution.

Table 4: Simulated Distribution of Final Tax Settlement ($)

Mean	25^{th} percentile	50^{th} percentile	75^{th} percentile
-1,704	-7,084	-2,483	3,015

Notes: Table shows statistics of final tax settlements of the distribution shown in Figure 5. Refunds are negative.

Appendix F illustrates how the key to matching the average final tax settlement is the curvature of the consumption function which, itself, depends on a combination of the coefficient of relative risk aversion θ and the discount factor β. For a any level of risk aversion above 1, Table F.1 indicates there is always a discount factor less than 1 that is sufficiently high to rationalize average tax refunds at least as large as observed in the data. The role of the consumption function is revealed in Figure F.1 which compares the function for two different sets of parameters that produce approximately the same average final tax settlement. The two consumption functions are practically identical.

5.3 Explaining the level and slope of the MPC

The previous section showed how the model explains large refunds on average. The monetary costs of underwithholding combined with uncertainty in non-paycheck income makes it optimal to overwithhold. This section uses the model to understand why workers spend a large fraction of their refunds when they arrive, and why that fraction tends to rise with the size of the refund.

5.3.1 The average MPC out of refunds

Textbook models of life-cycle/permanent-income consumption and saving imply a very small MPC in the aggregate. Such models include those without income uncertainty or models of certainty-equivalence. Models with precautionary saving (Zeldes (1989b), Deaton (1991), Carroll (1997)) moved the benchmark for the MPC to higher than the annuity value of lifetime resources. Recent models such as Kaplan and Violante (2014) and Carroll et al. (2017) predict larger MPCs (around 0.25) which are more consistent with the empirical literature.[14] Kaplan and Violante (2014) generate this larger MPC by introducing "wealthy hand-to-mouth" individuals who hold large amounts of wealth but do not smooth transitory shocks because they invest much of that wealth in illiquid assets. Carroll et al. (2017) generate large MPCs using a combination of impatience and transitory shocks. Our approach is closer to Carroll et al. (2017) in the sense that we are also able to generate an aggregate MPC that is more in line with the empirical literature using a combination of modest impatience and large transitory shocks.

Empirical estimates of the MPC out of a tax refund suggest a wide range of results depending on the data used and the definition of consumption. For example, Parker (1999) estimates an MPC out of tax refunds that ranges from 0.05-0.09 for nondurables and 0.34-0.64 for total spending using the Consumer Expenditure Survey. More recent studies using administrative data estimate larger MPCs out of nondurable spending. For example, Baugh et al. (2018) estimates an MPC of roughly 0.4 while this study estimates an MPC of roughly 0.2. While the data used in Baugh et al. (2018) is similar to our study, different approaches to defining nondurable spending can lead to large differences in reported MPC estimates. Because the MPC is defined as the additional spending out of an extra dollar of income, more comprehensive measures of spending will lead to higher estimates of the MPC.

While previous consumption models have explicitly targeted the MPC, we calibrate our model to match the average level of refunds. Nevertheless, the average MPC in our model of 0.3 is within range of the empirical estimates in the literature.

5.3.2 The slope of the MPC with respect to refunds

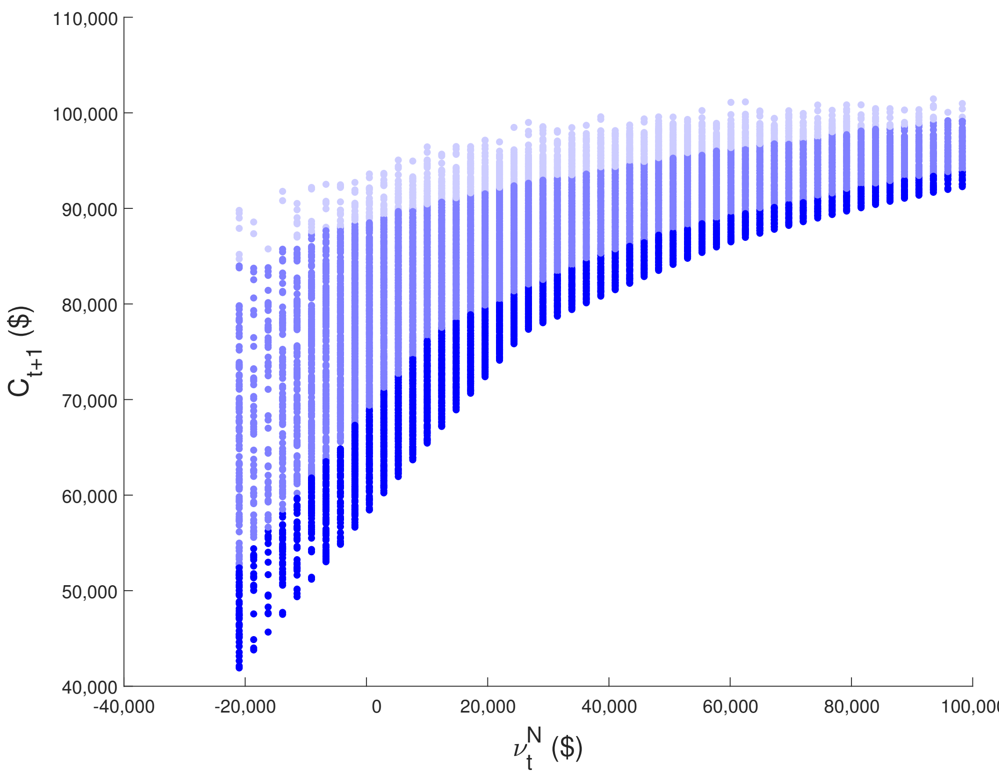
The average MPC out of refunds predicted by the model reflects average levels of cash on hand and average levels of refunds. The model predicts, however, heterogeneity in MPCs

[14] Other recent models that incorporate heterogeneity (e.g., Krusell and Smith (1998)) predict a low aggregate MPC because they locate most households in the flat portion of the consumption function.

out of refunds depending on prior non-paycheck income shocks and, thus, the size of the refund.

To better understand the mechanism linking shocks, refunds, and expenditure, Figure 6 plots the relationship between the non-paycheck income shock (ν_t^N) and next period consumption (C_{t+1}) in the simulation of the model. The different colors 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.

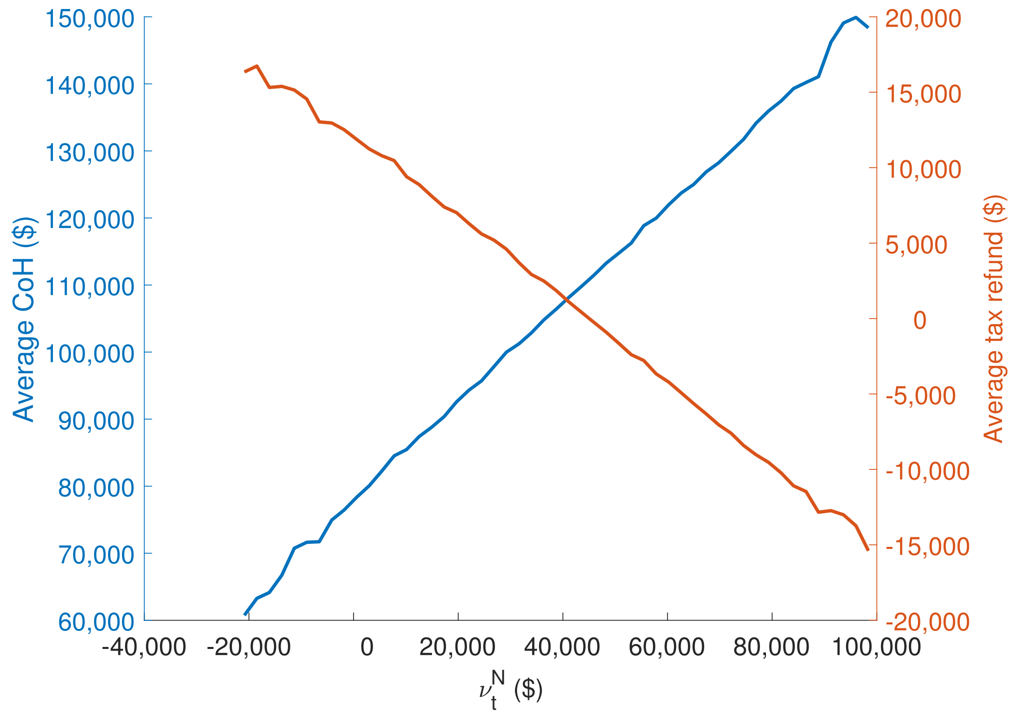
Figure 6: Non-paycheck Income Shock vs Next Period Consumption: Simulation



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. 100,000 simulated observations.

The link between tax refunds and cash on hand is seen in Figure 7. 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).

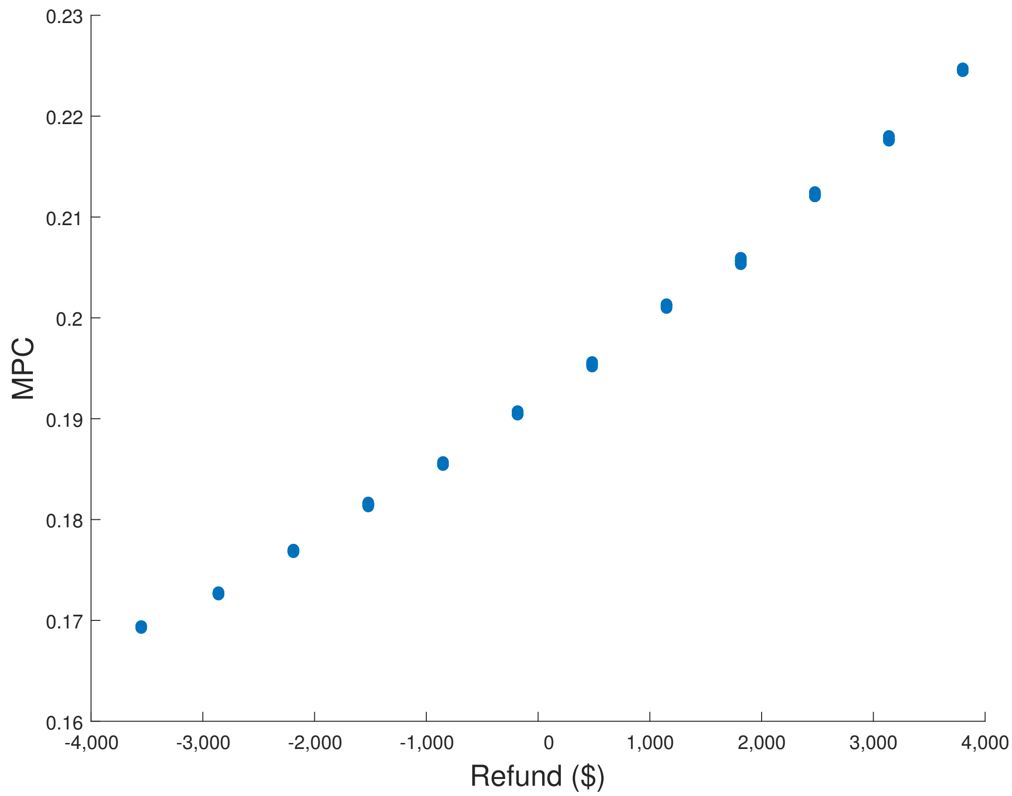
Figure 7: Average Cash on Hand and Tax Refund Conditional on ν_t^N: Simulation



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. 100,000 simulated observations.

We can combine the mechanisms described in Figures 6 and 7 to characterize the relationship between the MPC and tax refunds in the calibrated model. Figure 8 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.

Figure 8: 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. 100,000 simulated observations. Points are binned means of simulated observations.

In this way, the model predicts both that workers will spend substantial fractions of tax refunds, on average, and that the larger the refund the higher the MPC. This latter effect emerges because cash on hand tends to be especially low when especially large refunds arrive, and thus workers' have unusually high MPCs in those circumstances.

The upward sloping relationship between the level of tax refunds and the MPC is qualitatively consistent with the results in Baugh et al. (2018) who find that the MPC out of tax refunds is much higher than the MPC out of tax payments.

5.4 Individual-level 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 yield individual-level measures of the level of and the variation in both paycheck and non-paycheck income that we can link to tax refunds and spennding. The individual-level data allow direct evaluation of some of the key mechanisms in the model. This section estimates the empirical relationship between individual-level tax refunds and individual-level variation in income processes.

5.4.1 Excess withholding due to high frequency paycheck volatility

The model highlights the relationship between annual fluctuations in income and tax refunds. As explained in section 3.1, 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 periodic paychecks are prorated annual income. Other things equal, therefore, the convex income tax schedule implies withholding will increase, weakly, as within-year paycheck income rises.

Recall that we limited the sample to individuals on bi-weekly pay periods, so they are paid 26 weeks per year. 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:

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

where $w(\cdot; s, e, y)$ is a periodic withholding function that takes paycheck income as its argument and is influenced by filing status s, number of exemptions e, and year y.[15] 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.[16] We assume single filing status and two exceptions in our calculations of excess withholding.

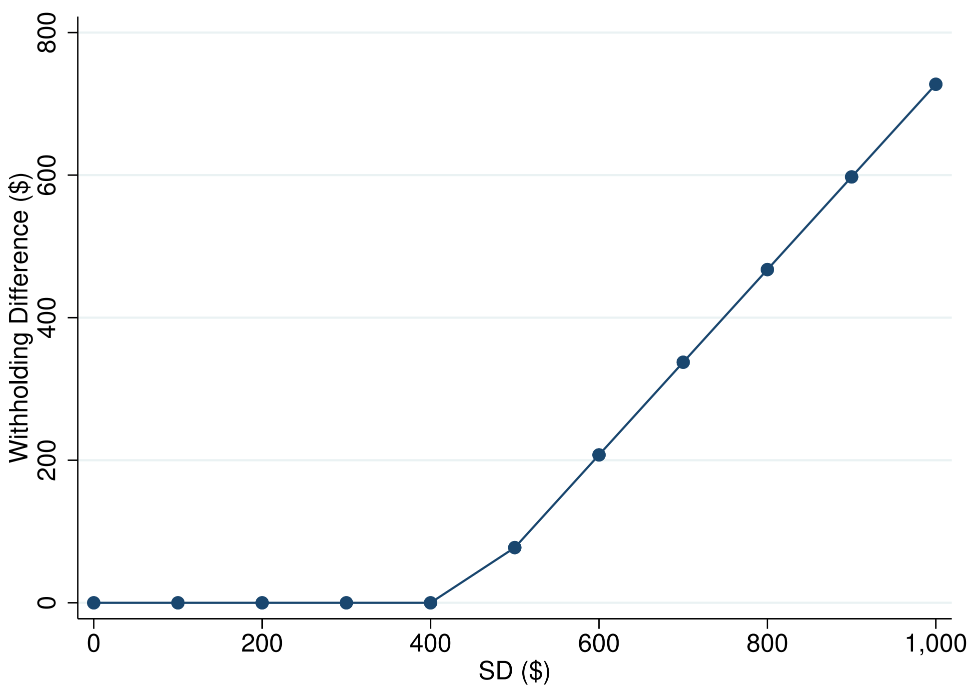
Figure 9 illustrates the relationship between this measure of potential excess withholding and within-year paycheck volatility. The example in the figure assumes paychecks are one standard deviation above average half the time and one standard deviation below average the

[15]The withholding function is based on the actual withholding schedule in form IRS publication 15 (aka circular E) https://www.irs.gov/pub/irs-pdf/p15.pdf. For more details see Appendix section C.

[16]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 $p^*_{by} = f^{-1}(\tilde{p}_{by}; s, e)$. See Appendix C for details of the calculation of bi-weekly withholding.

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.

Figure 9: $ExcessW_y$ 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 relate tax refunds with this measure of potential excess withholding.

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

Table 5 provides summary statistics of tax refunds, measures of the sources and variation in income, and the share of total spending that is captured as non-durable spending. 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 for Estimation: Means and Standard Deviations

	(1) Full sample	(2) Only refunds
Tax refund	1,704	3,184
	(2,542)	(2,712)
Paycheck share	.679	.682
	(.197)	(.186)
Excess withholding	1,340	1,151
	(2,358)	(2,006)
Paycheck variance	810	649
	(2,480)	(2,047)
Non-paycheck income variance	1,765	1,302
	(8,577)	(7,018)
Non-durable spending share	.198	.211
	(.218)	(.216)
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. In the table, paycheck and non-paycheck income variance is scaled down by a factor of 1,000,000 to increase readability. Non-durable spending share is the average of the share of total spending captured by non-durable spending at the individual level.

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 maximum 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 strongly negative, and both economically and statistically significant. The point estimate indicates that a worker who earns 90% of her income from a paycheck would have a refund that is less then 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 about non-paycheck oncome 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 variance of annual non-paycheck income. Consistent with the model, the variance of annual non-paycheck income is strongly associated with higher refunds, and this relationship is statistically significant.

The IRS withholding tables do a good job of capturing liabilities from paycheck income, so its variability should matter less. There is, however, a potential interaction between volatility of paycheck and non-paycheck income in our framework because of increasing marginal tax rates. Though we do not model it, uncertain paycheck income would create tax liability risk if its realizations cause total income to cross marginal tax rates. Hence, volatility of paycheck income can increase the precautionary amount of withholding against non-paycheck income. Column (3) adds the log of the variance of individuals annual paycheck income. The results also point to a role for annual variation in paycheck income, though a substantially smaller one than for volatility of non-paycheck income.

Column (4) 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 5.4.1. The sample size declines because the estimate is based on the prior year's income variation and therefore 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 the key cross-sectional determinants of excess withholding in our model, payshare, volatility of non-paycheck income, and excess withholding from within-year volatility of paycheck income Column (5), 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 and substantial magnitudes. The coefficients of payshare and volatility of non-paycheck income fall relative to the univariate specification because they are positively correlated.

Finally, Columns (6) and (7) 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: $Log(Refund)_{it}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$payshare_i$	-0.903				-0.360	-0.907	
	(0.0246)				(0.0339)	(0.0279)	
$Log(\sigma^2_{\nu_i^N})$		0.104	0.0921		0.0816		0.105
		(0.00224)	(0.00235)		(0.00305)		(0.00246)
$Log(\sigma^2_{\nu_i^Y})$			0.0382				
			(0.00256)				
$Log(ExcessW_{it-1})$				0.0456	0.0334		
				(0.00185)	(0.00181)		
Constant	8.195	5.679	5.179	7.348	6.173	8.247	5.705
	(0.0180)	(0.0413)	(0.0544)	(0.0131)	(0.0704)	(0.0202)	(0.0451)
NxT	134,752	134,752	134,752	87,736	87,736	87,736	87,736
N	49,520	49,520	49,520	44,328	44,328	44,328	44,328
R^2	0.022	0.041	0.045	0.009	0.047	0.021	0.041

Notes: Dependent variable is $Log(Refund)_{it}$. Robust standard errors in parenthesis. NxT represents the number of individual-year observations. N represents the number of individual observations. Columns (3) and (4) are based on one fewer year's observations to allow for the lagged variable. Columns (5) and (6) repeat the estimates of columns (1) and (2) with this sample.

5.4.3 MPC heterogeneity

Another key mechanism of the model is the positive association between tax refunds and the 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 relationship in the individual-level data we estimate the MPC as a function of tax refund size using the following specification

$$\tilde{C}_{it} = \sum_{j=1}^{5} MPC_j \times \widetilde{Refund}_{it} \times Q_i^j + \sum_{j=2}^{5} Q_i^j + month_t + \varepsilon_{it} \tag{12}$$

where $\tilde{C}_{it}$ is our measure of non-durable spending normalized by total spending, $\widetilde{Refund}_{it}$ is the tax refund normalized by total spending, 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.1.

Table 7: MPC estimates

	(1)
	Consumption
$MPC_{quantile1}$	0.165
	(0.006)
$MPC_{quantile2}$	0.179
	(0.005)
$MPC_{quantile3}$	0.189
	(0.005)
$MPC_{quantile4}$	0.193
	(0.005)
$MPC_{quantile5}$	0.216
	(0.005)
Constant	0.286
	(0.001)
Observations	1,615,572
R^2	0.047

Notes: Robust standard errors in parenthesis. Estimates include monthly and quantile fixed effects (not reported).

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 simulated MPCs rise similarly with the level of refunds. The correlation between the MPCs estimated across quintiles in Table 7 and the analogous simulated the values is 0.90.

The levels of the MPC are difficult to compare across simulation and estimates because the model is somewhat stylized, in particular, it presumes all spending is on a single non-durable good. The levels of the simulated MPCs and estimated MPCs are quite close, though they are not necessarily directly comparable. If much of spending beyond the strictly non-durable spending we use in the econometric analysis is pre-committed and hard to adjust (e.g., housing, utilities, vehicles, etc.), then our empirical MPC aligns closely with the theoretical value (see also Kaplan and Violante (2014)). On the other hand, if spending on durables is highly responsive to cash-on-hand, then our empirical MPCs understate the response of total spending. Since a full analysis of durable goods consumption is beyond the scope of the paper, we focus on the correlation between simulated and estimated MPCs by level of refund as the more robust support for the predictions of the model.

6 Conclusion

This paper presents and evaluates a simple theory of household liquid assets management with income shocks that can explain the prevalence of income tax refunds, the tendency of households to spend large fractions of those refunds, and the positive relationship between the propensity to spend and the size of the refunds. The theory maintains standard assumptions but, different from prior analyses of tax refunds or the responses of spending to income changes, 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 a positive marginal propensity to consume and that refunds will tend to arrive when cash on hand is relatively low, and thus the marginal propensity to consume is relatively high. The model thus explains a positive relationship between the size of tax refunds and the propensity to spend them.

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.

This analysis also has broader implications for understanding recent models of consumer behavior. Kaplan, Violante and Weidner (2014) show that there are many households that have substantial resources, yet act as if they were liquidity constrained. This wealthy hand-to-mouth behavior must therefore depend on costs of converting illiquid assets to liquid assets. This paper presents an important example of costly liquidity. Tax withholding and estimated payments are completely illiquid for a period of time. They cannot be withdrawn until the taxpayer files the annual income tax return. Individuals choose to save in the form of illiquid excess tax payments despite the zero nominal return on this saving because of the asymmetric cost of being under- and overwithheld.

The paper thus shows how wealthy hand-to-mouth behavior arises rationally because of features of the tax system and their interaction with income uncertainty. More generally, the paper reveals how a small adjustment friction, modest wedges between the returns to assets in different categories, and empirically relevant income volatility can generate substantial rational illiquidity. This illiquidity has consequential implications for spending from transitory income.

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.

Boning, William C (2018) "Paying Taxes Automatically: Behavioral Effects of Withholding Income Tax," Unpublished manuscript, University of Michigan.

Braxton, J. Carter, Gordon Phillips, and Kyle Herkenhoff (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. (1997) "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis," *The Quarterly Journal of Economics*, Vol. 112, No. 1, pp. 1–55.

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

Carroll, Christopher, Jiri Slacalek, Kiichi Tokuoka, and Matthew N. White (2017) "The distribution of wealth and the marginal propensity to consume," *Quantitative Economics*, Vol. 8, No. 3, pp. 977–1020.

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.

Deaton, Angus (1991) "Saving and Liquidity Constraints," *Econometrica*, Vol. 59, No. 5, pp. 1221–1248.

Debacker, Jason, Bradley Heim, Vasia Panousi, Shanthi Ramnath, and Ivan Vidangos (2013) "Rising Inequality: Transitory or Persistent? New Evidence from a Panel of U.S. Tax Returns," *Brookings Papers on Economic Activity*, Vol. 2013, No. 1, pp. 67–142.

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 J Noel (2019) "Consumer spending during unemployment: Positive and normative implications," *American Economic Review*. (forthcoming).

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.

——— (2018) "How individuals respond to a liquidity shock: Evidence from the 2013 government shutdown," *Journal of Public Economics*.

Guvenen, Fatih, Serdar Ozkan, and Jae Song (2014) "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, Vol. 122, No. 3, pp. 621–660.

Herkenhoff, Kyle F (2019) "The Impact of Consumer Credit Access on Unemployment," *The Review of Economic Studies*. (forthcoming).

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.

Krusell, Per and Jr. Smith, Anthony A. (1998) "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, Vol. 106, No. 5, pp. 867–896.

Kuchler, Theresa and Michaela Pagel (2018) "Sticking to Your Plan: The Role of Present Bias for Credit Card Paydown," Working Paper 24881, National Bureau of Economic Research.

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.

Wilcox, David W. (1989) "Social Security Benefits, Consumption Expenditure, and the Life Cycle Hypothesis," *Journal of Political Economy*, Vol. 97, No. 2, pp. 288–304.

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

Zeldes, Stephen P. (1989b) "Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence," *The Quarterly Journal of Economics*, Vol. 104, No. 2, pp. 275–298.

A Appendix – Data Filters, Definitions

The main analysis sample is drawn from the full 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 non-paycheck income, we also restrict attention to app users who receive regular bi-weekly 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.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.2 Defining regular paycheck

In order to identify regular paychecks, we start by using keywords that are commonly associated with these transactions.[17] 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

[17] 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","dfas-in" and DON'T include the keywords "ing direct","refund","direct deposit advance","dir dep adv."

A.3 Defining stable paycheck

The ratio of paycheck and non-paycheck 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.4 Payroll periodicity

We limit the sample to individuals with bi-weekly payroll. Bi-weekly paychecks are identified as a series of paychecks with the median number of days between each paycheck equalling 14 days.

A.5 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]), and receive a regular bi-weekly paycheck. 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.

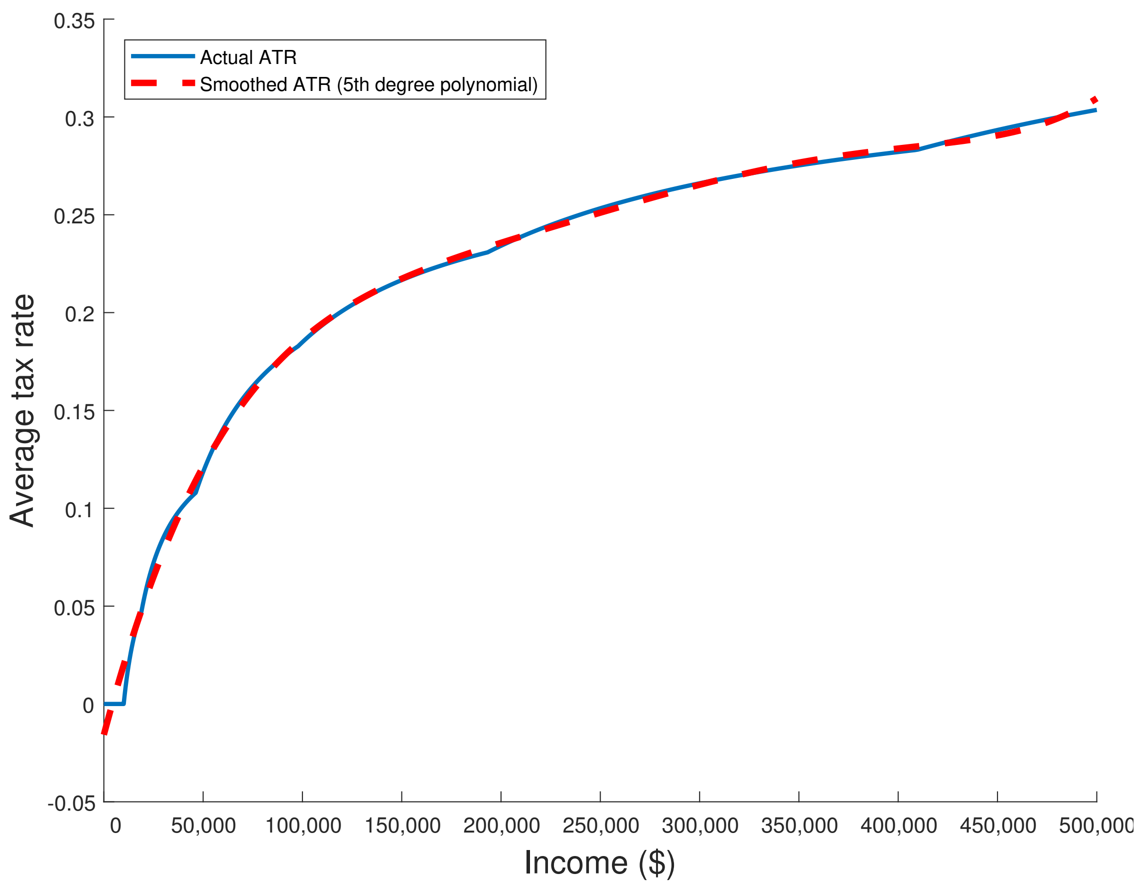
Table A.1: Effect of sample filters

	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 bi-weekly paycheck	92,883	11
Has stable paycheck	62,946	7

A.6 2013 tax schedule

We use the 2013 marginal tax rate schedule and calculate the average tax rate (ATR) schedule. We assume that individuals claim one personal exemption ($3,900) and the standard deduction ($6,100).[18] We then approximate the ATR schedule with a 5th degree polynomial. The actual and smoothed schedule is shown in Figure A.1. Note that while the smoothed function is negative for very low levels of income, income in the model is never this low.

Figure A.1: Actual and smoothed average tax rate function



Notes: This table plots the actual and smoothed average tax rate function. The smoothed average tax rate are calculated using a 5th degree polynomial.

The tax liability function is then defined as

$$\tau(Y) = ATR(Y) \times Y$$

[18] These values are taken from IRS publication 501 (https://www.irs.gov/pub/irs-prior/p501–2013.pdf).

where Y is income and $ATR(\cdot)$ represents the smoothed average tax rate function plotted above.

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)[19]

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.

[19] 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.

C Appendix – Withholding function calibration

The withholding function is calibrated using IRS publication 15 (aka circular E).[20] Figure C.1 displays an example of a table used to calibrate the withholding for individuals who receive a bi-weekly paycheck. We calibrate a withholding function for each year to account for the yearly changes in the schedules.

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.1: 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

[20] https://www.irs.gov/pub/irs-pdf/p15.pdf

E Appendix – 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 $(X_{t+1} = sR + N_t + Y_{t+1} - W(Y_{t+1}) - \tilde{\phi}\left[\tau(N_t + Y_t) - w(Y_t) - \widehat{W_t}\right])$ 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 $\widehat{W_t}$, back out what C_t will be

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

F Appendix – Sensitivity analysis

This section describes the sensitivity of the simulated average final settlement in the model to different utility parameters. Table F.1 shows the average final tax settlement under different values for θ and β. The table shows that other than our prefered parameterization of $\beta = 0.98212$ and $\theta = 1$, the values $\beta = 0.97$ and $\theta = 2$ also come close to matching the data.

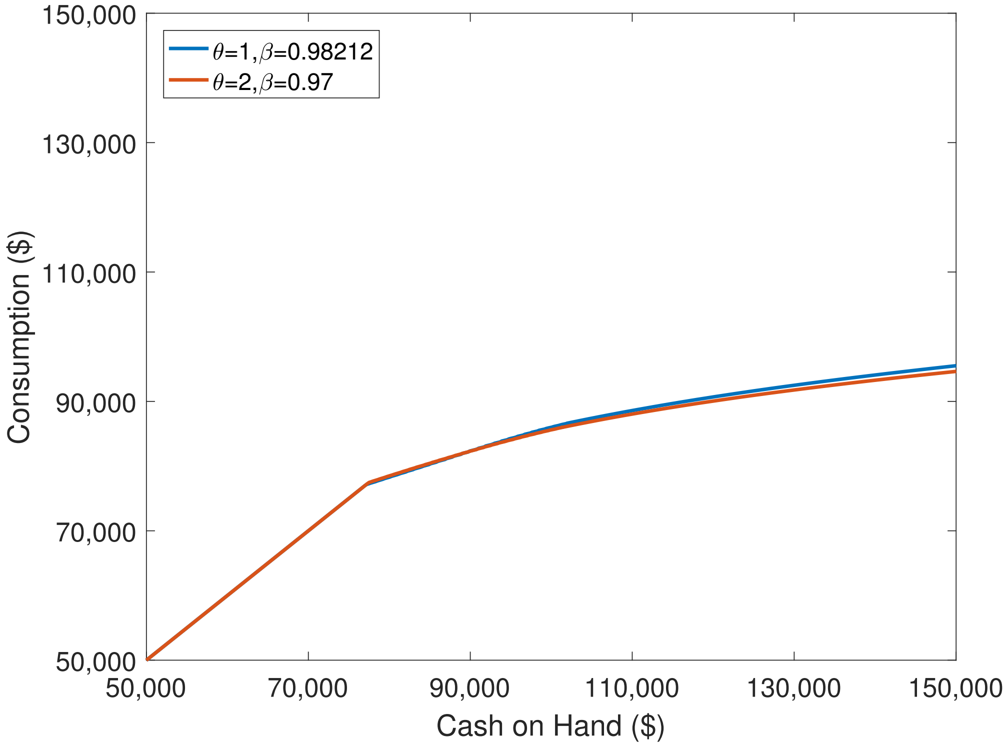
Table F.1: Average final tax settlement under different parameter values

	$\theta = 1$	$\theta = 2$	$\theta = 3$	$\theta = 4$
$\beta = 0.96$	2160	-769	-2268	-3173
$\beta = 0.97$	633	-1760	-2951	-3655
$\beta = 0.98$	-1249	-2893	-3689	-4149
$\beta = 0.99$	-3564	-4342	-4671	-4835

Notes: This table calculates the average final tax settlement for 100,000 simulated observations under different parameter values. A negative settlement represents a refund while a positive settlement represents a payment.

Figure F.1 compares the consumption function for two different sets of parameters that achieve roughly the same average final tax settlement. The curves are very similar for most values of cash on hand. The main difference is that the consumption function with a higher risk aversion parameter curves slightly more than the consumption function with a lower risk aversion parameter.

Figure F.1: Consumption function comparison



Notes: This figure shows the consumption function of two different sets of parameter values.

G Appendix – 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

$$\mathbb{E}[n_{t,b}] = \mathbb{E}\left[\frac{N_t}{26}\right] \tag{21}$$

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

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

$$n_{t,b} - \bar{n}_t = \epsilon^N_{t,b} \tag{23}$$

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

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

$$\mathbb{V}[n_{t,b}] = \mathbb{V}\left[\frac{N_t}{26}\right] + \mathbb{V}[\epsilon^N_{t,b}] \tag{25}$$

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

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

H Appendix – 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.[21] 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 $\hat{Y}_0 = \hat{f}(X_0)$.

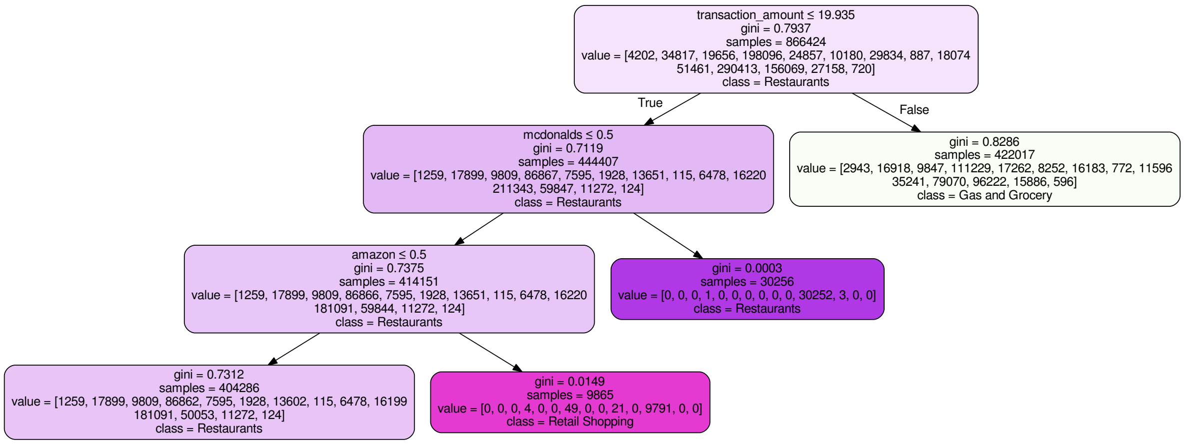
One option for the mapping is to use the multinomial logit model since the dependent variable is a categorical variable with no cardinal meaning. However, this approach is not well suited to textual data because each word would need its own dummy variable. Furthermore, interactions may be important for 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

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

depending on the features of the data. For example, for transactions that have the keyword "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 H.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 H.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[22] and maximized when the transactions are evenly split between the two categories.[23]

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.

---

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

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