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

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

First draft: April 2019
Revised January 2022

Abstract

Low liquidity and a high marginal propensity to consume are tightly linked. This paper analyzes this link 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, and can explain the finding that households tend to spend tax refunds as if they valued liquidity, yet do not act to increase liquidity by reducing their withholding. The theory is supported by individual-level evidence based on financial account records, 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, marginal propensity to consume (MPC), account data.

*We thank William Boning, James Hines, Damon Jones, Joel Slemrod, Dmitry Taubinsky, 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 maintain low liquid assets balances and exhibit substantial sensitivity to transitory changes in income. Behavior surrounding income tax withholding and refunds represents an important example. Nearly a third of all personal income tax payments collected by the US government are later returned in the form of tax refunds. Households tend to spend substantial fractions of those refunds when they arrive. Hence, households 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, even those with substantial incomes, reduce their liquid assets buffer and yet respond to income fluctuations as if they would value a cushion against shocks? This paper offers a theory that rationalizes evidence both of large average tax refunds and of sensitivity of spending to the arrival of refunds. The theory features precautionary saving motivated by income uncertainty. It also incorporates the wedge between the return to saving in private liquid assets, and the return to "saving" in the illiquid form of income tax withholding. That wedge is positive when overwithheld and negative when under-withheld. The theory also makes the distinctive prediction, borne out empirically, that the propensity to spend out of refunds should be increasing with the refund size. This prediction emerges from the model because negative income shocks both raise the MPC out of transitory income and produce tax refunds. This link between MPC and refunds arises because lower than expected income means the household is overwithheld.

The theory maintains standard assumptions of modern models of consumption and saving. It provides a unified alternative to existing, often behavioral, models that separately explain either the tendency to receive tax refunds or the propensity to spend them. While we offer a mostly standard model consistent with these facts, we do not offer evidence to refute behavioral theories. Thus, we do not rule out the possibility that behavioral mechanisms, with different welfare implications, are playing important roles driving these behaviors.

The paper analyzes administrative data to describe the distribution and correlates of tax refunds and to estimate the propensity to spend them. Our sample consists of a 4-year panel of individual-level bank and credit card records for approximately 63,000 users of a financial aggregation application. These records offer an integrated source of high frequency, individual-level measures of income by source, of tax refunds, and of spending. They reveal several patterns: (1) On average, households receive nearly a third of their income from sources that are not subject to tax withholding (non-paycheck income). (2) This form of income is especially variable within households over time. While annual non-paycheck income averages \$38,764 compared to \$68,226 for paycheck income, annual non-paycheck income has an average, within-household standard deviation of \$19,879, compared to \$18,490 for paycheck income. (3) Average tax refunds are large, approximately \$3,200, and larger for households with higher fractions of non-paycheck income or greater annual variation in that source of income. (4) The marginal propensity to consume (MPC) out of tax refunds is 0.21 on average, and thus significantly above a certainty equivalent lifecycle model benchmark. (5) The MPC increases with the size of the refund in a way predicted by the theory.

Our theory accommodates these patterns with a parsimonious model based on standard assumptions. It assumes taxes on paycheck income are automatically withheld appropriately at the source, but that households need to make active withholding decisions to cover taxes on non-paycheck income. (Throughout the paper, "withholding" should be understood to include estimated tax payments.) It also assumes non-paycheck income is uncertain, and that withholding decisions are made before all non-paycheck income uncertainty is realized. This friction motivates households to maintain precautionary savings in part to cover their uncertain tax liability. Those savings could be held in private accounts, but we assume, as is the case in the U.S., that the penalties for being under-withheld are above the market return to saving. This wedge between the market return to saving and the return to saving in the form of income tax withholding motivates overwithholding. Importantly, when non-paycheck income represents a sufficiently large and unpredictable fraction of total income, the theory predicts large refunds.

The income uncertainty that rationalizes large tax refunds also rationalizes standard forms of precautionary saving and a positive MPC out of transitory income. In the model, households want to maintain cash on hand to smooth consumption as income fluctuates. They weigh the costs of accumulating a larger buffer against the benefits of less variable consumption. With a conventional calibration of the model parameters, optimization implies consumption smoothing is imperfect and there is a positive MPC out of transitory income, including tax refunds. The model can thus rationalize, in a unified way, both the deliberate decision to reduce liquidity by overwithholding income taxes and then the decision to spend a substantial fraction of those taxes once they are refunded. Our analysis does not rule out alternative theories of either behavior, but instead offers a parsimonious and mostly conventional theory of them both.

The analysis is carried out at annual frequency that maps directly into annual features of the tax code, e.g., that liabilities are calculated on an annual basis, that withholding is accrued on an annual basis, and that refunds are received once a year. Hence, the consequences of the illiquidity studied in this paper are also annual. In particular, the analysis addresses how the refund's relaxation of the liquidity constraint affects the MPC. The model is not designed, however, to study how this illiquidity affects the intra-year response to income volatility in the year when the overwithholding takes place, which is an important, but different phenomena.

Distinctively, the model additionally predicts that the MPC out of tax refunds is increasing in the size of the refund. In standard models of precautionary saving (Zeldes (1989a), Carroll and Kimball (1996)), the consumption function is concave; that is, the MPC for transitory income declines with cash on hand. In the model we develop, negative income shocks naturally lower cash on hand and thus raise the MPC out of transitory 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, so the MPC is higher. Optimal behavior thus implies that the MPC tends to be higher when refunds arrive. The larger the refund the higher the MPC.

The model's predictions are qualitatively consistent with the patterns we observe in the individual-level data. To perform a quantitative assessment, we make conventional assumptions about preference parameters and simulate the model based on the average level and variability of income, by source, in the data. Not all of that income variability represents relevant risk for the household. Some of those income fluctuations, perhaps a substantial fraction, may be anticipated. Without a long panel on income or detailed data on expectations, it is challenging to quantify the relevant risk. Instead, we estimate what fraction of income variability is relevant risk by matching the average size of tax refunds. Assuming log-utility and a discount factor of 0.985, the model matches the average tax refund of $3,184 when about 57% of the variability in non-paycheck income is unanticipated and therefore relevant to the precautionary motive for tax withholding.

Individual-level evidence from the administrative data is also consistent with the distinctive prediction of the model regarding the relationship between the size of tax refunds and the MPC out of those refunds. In particular, 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 substantially from the bottom to the top quintile of the refund distribution.

The rest of the paper is organized as follows. Section 2 relates the paper to the literature. Section 3 discusses institutional features that drive incentives for liquid assets management, income tax withholding, and consumption. Section 4 describes the data source, analysis sample, and the key measures. It also compares these data to other data sources. Section 5 presents the model and Section 6 shows how the calibrated version of the model matches the moments of the data. Section 7 contains some concluding remarks.

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 Liquidity and Consumption

The analysis of liquid assets management presented here relates to a growing literature that studies, often using innovative data sources, the relationship between liquidity and consumption. Examples include Gelman et al. (2014) that uses administrative account data to study the consumption response to regular paychecks, Gelman et al. (2020) that studies the consumption response to a liquidity shock, Braxton et al. (2018) that links administrative employment and credit bureau data to study consumption smoothing during unemployment, Herkenhoff (2019) that uses several data sources to show how increasing access to credit led to an increased ability to smooth consumption during unemployment, and Ganong and Noel (2019) that analyzes the response to unemployment shocks.

The paper also relates to Kaplan et al. (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. (2019), Bunn et al. (2018), and Fuster et al. (2020) examine how spending responds to hypothetical increases and decreases in income. Baugh et al. (2021) 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. (2021) 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 Overwithholding

Jones (2012) generalizes a theory of overwithholding based on the logic in Highfill et al. (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 et al. (1998) explain overwithholding as the optimal response to the wedge between the opportunity costs of overwithholding and the Internal Revenue Service (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.[1]

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 a related 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 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. Our model predicts large refunds without inertia or defaults biased toward overwithholding.[2]

[1] 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.

[2] 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 the year. Our focus is on middle to high income households who are likely ineligible for the EITC.

2.3 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 (2021), and Baugh et al. (2021) 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), Wilcox (1989) (anticipated changes in Social Security payments), Stephens (2003) (social security payments), and Shapiro and Slemrod (2003) (tax rebates). For purposes of analyzing the spending response to refunds, we will treat the income change tax refunds produce as partially unanticipated. This treatment is motivated by the fact that if individuals remain uncertain of the extent of their refund until the date of filing, then the delay between learning about the size of the refund and receiving it is typically so short that it can be interpreted as at least partially unanticipated.[3]

When tax refunds are transitory but to some extent unanticipated income, then 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 Institutions

Several institutional features drive incentives for liquid assets management, income tax withholding, and consumption. We describe here how (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, 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.

Federal income tax liability is determined annually. Taxes on wage and salary (paycheck) income are usually withheld at the source. The schedule for withholding at the source is

[3] 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. (2021) report that the average refund arrives 11 days after the tax return was filed.

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. On the W-4, the IRS provides guidelines for workers on how many allowances to take.[4]

Under some circumstances, following the IRS guidelines for allowances results in withholding that 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 a single source of earnings subject to withholding, and if those earnings exhibited little within-year variation, following the guidelines would result in withholding that closely matches tax liability. The worker would owe no additional income taxes and would receive no income tax refund.

Simple adherence to the W-4 allowance guidelines is, however, unlikely to result in accurate income tax withholding in several circumstances. In particular, workers who receive various forms of "non-paycheck income" will tend to be underwithheld if their only withholding results from following the W-4 guidelines for their wages and salary. Business income, including self-employment income, and income from partnerships, S-corporations, and rental properties, is typically not subject to withholding at the source. The same holds for most financial investment income and pension disbursements. To avoid underwithholding of taxes on these sources of income, additional taxes must be paid directly as estimated tax payments, or from income that is subject to withholding at the source.[5,6]

[4]Subsequent to the 2017 tax legislation, the IRS changed the withholding tables and the W-4 form. The discussion of tax rules as well as the tax parameters we use for modeling are based on the institutional arrangements around withholding that were in force during the 2013-2016 period for our data.

[5]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.

[6]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.

Unless it is properly adjusted for, paycheck income from multiple employers can also produce underwithholding due to the convexity of the tax schedule. Tax units with two earners are an important instance. When tax liabilities are based on joint income, because the tax schedule is convex, both earners will typically be underwithheld, with the underwithholding greater the more unequal are the incomes. The W-4 was modified following the 2017 tax cuts to address the two-earner problem, but for the period covered by the data we study, the W-4 did not easily address the issue.

Additionally, if wage and salary income varies substantially within the year, then adherence to the W-4 guidelines can produce overwithholding, even if the only source of income is subject to withholding at the source. This "mechanical effect" of high frequency income variation also derives from the fact that the income tax schedule is convex and the withholding schedule treats each paycheck as prorated annual income.[7]

Underwithholding is penalized. Normally, individuals with taxable income must file a tax return, or a request for an extension, by a mid-April deadline. Even if the tax bill is paid in full by 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 3.2% 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 deadline, then there is a penalty of 0.5% of the unpaid tax assessed every month that the remaining tax goes unpaid. Filing for an extension does not extend the deadline for payment. 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

[7] 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 Appendix Section C.1. We are grateful to Damon Jones for highlighting this effect for us.

large penalties for late filing. Tax payers therefore face strong incentives to file on time, even if they have unpaid taxes. Fear of audit might also motivate taxpayers to file on time and avoid underwithholding.

Finally, safe harbor provisions exempt some households from interest payments on underwithholding. No interest applies if the unpaid amount equals less than $1,000 total, or represents less than 10% of total taxes owed in the 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.[8]

4 Data

This section describes the data we analyze. We first describe the data source, which is based on individual-level account records, and how we draw the analysis sample. We then explain how we construct the key measures for our analysis—tax refunds, expenditure, and income levels, sources, and volatility—and compare them to measures from other data sources. We conclude this section by examining the relationship between tax refunds and income measures.

4.1 Data Source and Analysis Sample

To conduct a unified study of the distribution and correlates of tax refunds, and of the propensity to spend them, we draw on administrative records 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.[9] The app offers financial aggregation and bill-paying services. Users can link almost any financial account to the app, including bank accounts, credit cards, utility bills, and more. We used these data previously to study the spending response to anticipated income, stratified

[8]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.

[9]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.

by spending, income, and liquidity (Gelman et al., 2014) and households' high-frequency responses to shocks such as the government shutdown (Gelman et al., 2020). Similar account data have been used in Baugh et al. (2021), Baker (2017), Baker and Yannelis (2017), Kuchler and Pagel (2021), Ganong and Noel (2019), Koustas (2018), 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 being supplied to the researchers, 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 these data are "naturally-occurring" or "non-designed," they reflect any non-random enrollment in the app. We have taken a number of steps to assess whether app users are 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 app users with representative samples. Although there are differences, notably that very low incomes and older individuals are underrepresented, the demographic and economic profile of the sample from the app captures a diverse population.

From the population of app users, we draw an analysis 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 who 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. Our methods for identifying regular paychecks are detailed in the next subsection. The specifics of all these filters are described in the appendix and the consequences for sample size are presented in Table A.1.

Our analysis is thus based on a sample of individuals with paycheck 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. We observe an individual level average of 88 transactions per month. Even the 1st percentile of the sample conducts an average of 16.5 transactions per month.

We next define the key measures for analysis. To further evaluate the validity of the sample, we will then compare the distributions of these variables in the sample with their distributions in other data sources. This analysis shows 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 income not subject to withholding.

4.2 Key Measures

The data from the app consist 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 keywords in the description field (all tax refunds include the keywords "TAX," "TREAS," and "REF"). Most refunds (96%) in these data are received in February, March, April, and May.

The spending response to the arrival of tax refunds is also key to our analysis. 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 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 is then combined to create the consumption variable.

The ML algorithm uses a subset of the data where the Merchant Category Code (MCC) is recorded as a training dataset in order to create a mapping from transaction description to MCCs.[10] 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 miscellaneous services to measure consumption expenditure.

We define an individual's income as the sum of all inflows to checking and saving accounts minus transfers between accounts.[11] From this measure of income, paycheck income is defined as the inflows from paychecks identified using an algorithm detailed in Appendix A.2. The raw measure of paycheck income from the app 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 is defined as non-paycheck income. In the analysis we use before-tax paycheck income.

Income volatility, and the precautionary saving it motivates, are central to our theoretical analysis. We are also interested in the potential for higher frequency income variation to have "mechanical" effects on tax refunds. We therefore exploit the panel structure of the app data to estimate both annual and bi-weekly income volatility, at the individual level.

Formally, we model the total income of individual i in bi-weekly interval b and year t, as arriving in the form of paycheck income $p_{i,t,b}$, and non-paycheck income $np_{i,t,b}$. These bi-weekly realizations derive from annual paycheck income $P_{i,t}$ and non-paycheck income $NP_{i,t}$ processes. Each of these four income variables contains a random component. Specifically, we assume these variables follow the processes:

$$p_{i,t,b} = \frac{P_{i,t}}{26} + \epsilon^{P}_{i,t,b} \tag{1}$$

$$np_{i,t,b} = \frac{NP_{i,t}}{26} + \epsilon^{NP}_{i,t,b} \tag{2}$$

$$P_{i,t} = \alpha_{i,P} + \nu^{P}_{i,t} \tag{3}$$

$$NP_{i,t} = \alpha_{i,NP} + \nu^{NP}_{i,t} \tag{4}$$

where the random components of bi-weekly and annual income variables are represented

[10] 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.

[11] Transfers are identified using the keywords "transfer," "xfer," "tfr," "xfr," or "trnsfr."

by $\epsilon^{P}_{i,t,b}$, $\epsilon^{NP}_{i,t,b}$, $\nu^{P}_{i,t}$, and $\nu^{NP}_{i,t}$. These components are normally distributed with mean zero and corresponding variances $\sigma^2_{i,\epsilon^P}, \sigma^2_{i,\epsilon^{NP}}, \sigma^2_{i,\nu^P}$, and $\sigma^2_{i,\nu^{NP}}$. The two components of annual income are modeled as independent processes.[12] In addition, bi-weekly income is subject to serially-uncorrelated noise.

Table 1 gives the average across individuals of the estimated parameters of the income processes. Appendix G describes in detail how we estimate the parameters. Table 1 shows an average annual paycheck income of \$68,226, with an average standard deviation of \$18,490. The bi-weekly component of volatility is also sizeable, with an average standard deviation of \$1,791. The analogous estimates for non-paycheck income are \$38,764, \$19,879, and \$3,182. In these data, non-paycheck income therefore is substantial, on average, and is especially volatile. We compare income moments with external data sources in the next section and use them for calibrating the model in Section 6.

Table 1: Income Moments: Average of Individual-Level Parameter Estimates (\$)

$\overline{\alpha}_P$	68,226
$\overline{\alpha}_{NP}$	38,764
$\overline{\sigma}_{\nu^P}$	18,490
$\overline{\sigma}_{\nu^{NP}}$	19,879
$\overline{\sigma}_{\epsilon^P}$	1,791
$\overline{\sigma}_{\epsilon^{NP}}$	3,182
NxT	251,784
N	62,946

Notes: The bar means averaged across individuals. NxT represents the number of individual-year observations. N represents the number of individual observations. All individual estimates winsorized at 1%.

4.3 Comparison with Other Data Sources

With these measures of tax refunds, expenditure, and income, we can compare statistics of the app sample to those from external data sources. Table 2 shows the average tax refund

[12] 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.

in the sample is \$1,704 where, because we cannot accurately observe most tax payments, the refund is set equal to zero for those who did not receive a refund. Restricting attention to positive refunds, the average size is \$3,184, slightly larger than the average reported by the IRS \$2,778. Segmenting the app sample according to the gross income levels used in publicly-available IRS tables, the average refund size in the app data matches very closely the IRS figures for those with incomes between \$0 and \$50,000, or between \$50,000 and \$100,000. (IRS numbers are conditional on receiving refunds and are comparable with the first column of app data.) For those with incomes exceeding \$100,000, the average refund in the app data is substantially smaller than that reported by the IRS. The app does not include the highest income households who also tend to receive the largest refunds.

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 higher than the \$58,410 average annual spending in the Consumer Expenditure Survey (CEX). The app undersamples those with low income. Segmenting by income somewhat reduces these discrepancies in average between the datasets. 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 is \$46,224 and average non-paycheck income is \$20,603. Stratifying by income, the IRS data appear to contain more of both the lowest and the highest income households. For example, average non-paycheck income in the app data is \$14,442 among those with incomes between \$0 and \$50,000, while in the IRS data it is \$4,556. At the other end, average non-paycheck income in the app data is \$83,408 among those with total incomes above \$100,000 while in the IRS data it is \$95,261.

While the average levels of income are somewhat higher in the app sample than in the population of tax filers, measures of income volatility are similar to those in other studies that use administrative records on income. In the app sample, the average standard deviation of the first difference of the log of annual income is 0.50.[13] The measure is very similar (0.48) if

[13]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.

we restrict attention to those who received at least one tax refund during the sample period. Income volatility is higher (0.57-0.60) for those with less than $50,000 of average annual income than for those earning more than $50,000 per year (0.45-0.48). Guvenen et al. (2014) use earnings histories from US Social Security Administration records covering 1978-2011. They show that the standard deviation of the first difference of men's log earnings varies from roughly 0.50 to 0.60 over their time period. Like, Guvenen et al. (2014), the app data show that income volatility is higher at the lower levels of income.[14]

[14] Debacker et al. (2013) use panel data from tax returns over the years 1997-2009. They can thus study the volatility of total income, not just earnings. They find that pre-tax household income shows very similar levels of volatility with an average standard deviation of the log change in income of approximately 0.4.

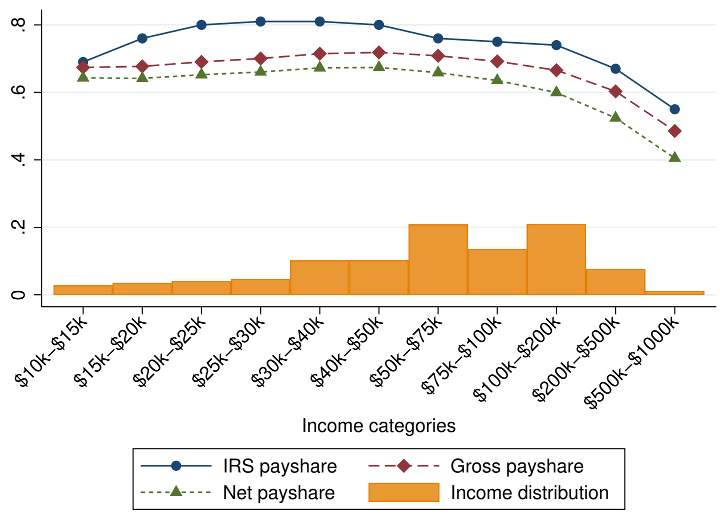
Table 2: Comparing App and External Sources

	App		
	Received refunds	Full sample	External sources
Panel A: Means			
Tax refund ($)	3,184	1,704	2,778
low income (0-50k)	2,402	1,152	2,323
medium income (50-100k)	3,051	1,796	3,006
high income (>100k)	4,258	2,335	5,869
Spending ($)	77,855	83,254	58,410
low income (0-50k)	41,763	45,069	32,913
medium income (50-100k)	64,676	66,946	58,775
high income (>100k)	135,986	152,653	107,799
Paycheck income ($)	67,416	68,226	46,224
low income (0-50k)	32,361	33,187	17,461
medium income (50-100k)	57,791	58,732	54,588
high income (>100k)	118,767	124,719	145,624
Non-paycheck income ($)	36,607	38,764	20,603
low income (0-50k)	13,844	14,442	4,556
medium income (50-100k)	27,456	27,891	17,680
high income (>100k)	73,808	83,408	95,261
Paycheck share	.68	.68	.69
low income (0-50k)	.70	.70	.79
medium income (50-100k)	.69	.69	.76
high income (>100k)	.65	.64	.60
Panel B: Standard deviations			
First difference of total income	.48	.50	.52
low income (0-50k)	.57	.60	.64
medium income (50-100k)	.45	.48	.47
high income (>100k)	.45	.45	.45
NxT	134,752	251,784	
N	49,520	62,946	

Notes: NxT represents the number of individual-year observations. N represents the number of individual observations. For panel B, the observations are at the individual level so only N is relevant. For panel A, received refunds refers to person-years where a refund is received. For panel B, it refers to individuals who ever received a refund. External tax refund data are from the 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 data on the level of income is calculated from IRS, Statistics of Income Division Publication 1304. External measures of the standard deviation of income are from are from Guvenen et al. (2014). Because they do not break down their results by income, we approximate low, medium, and high income individuals using terciles.

The average ratio of paycheck income to total income in the app sample is similar to that in the IRS data, about 0.68. Compared with the IRS data, the average payshares are somewhat lower in the app data for lower income households, and slightly higher for the higher income households. Figure 1 plots the paycheck share from the IRS along with two alternative measures of the payshare in the app data for different segments of the income distribution. The raw paycheck income in the app is net of withholding. Net payshare is the paycheck share calculated based on the raw paycheck as recorded by the app. Gross payshare uses the gross paycheck income as described above and used throughout the analysis. 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 1: Paycheck share comparison across income groups



Notes: Payshare is the 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.

4.4 Correlates of Tax Refunds

We next describe the relationship between income levels, income sources, income volatility, and tax refunds in the app data. Table 3 presents OLS estimates of the individual-level relationship between income tax refunds and the 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 1 refund during the period. On average each of these individuals receives 2.7 (out of a maximum of 4) refunds during the period. The sample is conditional on a positive refund.

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 4 years of observation. Consistent with a link between tax refunds and income not subject to withholding at the source, we find 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.

Table 3 also reveals a statistically significant relationship between non-paycheck income volatility and 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. The results indicate that a 1% increase in the volatility of non-paycheck income is associated with a 0.09% increase in the tax refund amount. Column (3) adds the log of the variance of individuals' annual paycheck income. The results point to a role for annual variation in paycheck income in determining tax refunds, though a substantially smaller one than for volatility of non-paycheck income. Inclusion of the paycheck income volatility does not much affect the estimated effect of non-paycheck volatility, which is a key mechanism in the model.

Table 3: Tax refunds and income volatility: $Log(Refund)_{it}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$payshare_i$	-0.903				-0.362	-0.907	
	(0.0246)				(0.0339)	(0.0279)	
$Log(\sigma^2_{\nu_i^{NP}})$		0.104	0.0921		0.0814		0.105
		(0.00224)	(0.00235)		(0.00305)		(0.00246)
$Log(\sigma^2_{\nu_i^{P}})$			0.0382				
			(0.00256)				
$Log(ExcessW_{it-1})$				0.0458	0.0336		
				(0.00186)	(0.00182)		
Constant	8.195	5.679	5.179	7.346	6.176	8.247	5.706
	(0.0180)	(0.0413)	(0.0544)	(0.0132)	(0.0704)	(0.0202)	(0.0451)
NxT	134,752	134,752	134,752	87,712	87,712	87,712	87,712
N	49,520	49,520	49,520	44,324	44,324	44,324	44,324
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 parentheses. NxT represents the number of individual-year observations. N represents the number of individual observations. Columns (4) and (5) are based on one fewer year's observations to allow for the lagged variable. Columns (6) and (7) repeat the estimates of columns (1) and (2) with this sample.

As noted earlier, the rules governing paycheck tax withholding schedules, and the convexity of the tax schedule, may induce a "mechanical" relationship between within-year paycheck volatility and refunds. The withholding schedule assumes periodic paychecks are pro-rated annual income. Other things equal, therefore, the convex income tax schedule implies the withholding rate will increase, weakly, as within-year paycheck income volatility rises. To quantify the magnitude of the mechanical effect, and isolate its influence from that of annual fluctuations, we define potential excess withholding from high frequency paycheck volatility:

$$ExcessW_{it} = \sum_{b=1}^{26} (w(p_{i,b,t}; s, e, t) - w(\bar{p}_{i,t}; s, e, t)) \tag{5}$$

where $w(\cdot; s, e, t)$ 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 t.[15] Let $p_{i,b,t}$ denote the bi-weekly paycheck (before withholding) for individual i in bi-week b of year t, and $\bar{p}_{i,t}$ the average bi-weekly pre-withholding paycheck in year t.[16] We assume single filing status and two exemptions in our calculations of excess withholding.

The average potential excess withholding is large, \$1,340 in the full sample and \$1,151 among those who receive refunds. The distribution of excess withholding is positively skewed with a 25^{th} percentile of \$170 (resp. \$157) and a 75^{th} percentile of \$1,386 (resp. \$1,229) for the full sample (resp. refund receivers). If workers do not account for this mechanical effect of within-year income variation, predicted refunds should be larger. Alternatively, workers may internalize the effects of this within-year income variation and adjust their withholding accordingly.

We evaluate the extent of this internalization as we relate tax refunds with this measure of potential excess withholding. In particular, Column (4) of Table 3 presents the correlation between the log of tax refunds and the measure of excess withholding due to high frequency variation in paycheck income.[17] These results show a statistically significant, though economically modest, mechanical effect of high frequency income variation on tax refunds. The modest size of the point estimate indicates that workers internalize much of the "mechanical

[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 C.

[16] We do not observe pre-withholding income $p_{i,b,t}$. Instead we observe post-withholding income $\tilde{p}_{i,b,t} = p_{i,b,t} - w(p_{i,b,t}; s, e, t)$ and estimate $p_{i,b,t}$ from $\tilde{p}_{i,b,t}$ conditional on s, e, and t. Because post-withholding paycheck income is a function of pre-withholding income and other tax parameters, $\tilde{p}_{i,b,t} = f(p_{i,b,t}; s, e, t)$, we can take the inverse of this function to estimate $p_{i,b,t}$ by $p^*_{i,b,t} = f^{-1}(\tilde{p}_{i,b,t}; s, e, t)$. See Appendix C for details of the calculation of bi-weekly withholding.

[17] The sample size declines because the estimate is based on the prior year's income variation and therefore only three years are available.

effect" and adjust withholding accordingly. Conditioning on the share of paychecks in total income, the volatility of non-paycheck income, and excess withholding from within-year volatility of paycheck income in Column (5), the qualitative results are unchanged. 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.

The qualitative relationships estimated in Table 3 are not driven by households with higher income who, on average, have both lower paycheck income shares and larger refunds. In the appendix, Table E.1 shows that the focal relationships largely hold, qualitatively, in separate analyses of the top, middle, and bottom terciles of income.[18]

5 The Model

5.1 Specification

To understand how the source and variation of income creates incentives to overwithhold, and also why households tend to spend large fractions of tax refunds, we adapt a standard model of consumption and saving with income uncertainty. The model incorporates key institutional features of the tax system related to withholding and 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. Y_t is income where the tax liability is withheld at the source or where the taxpayer knows the amount in advance and so can make estimated tax payments that exactly match the tax liability. N_t is stochastic income where estimated taxes must be paid in advance of the realization of income. Y_t includes paycheck income plus the pre-determined part of non-paycheck income (e.g., certain self-employment income or retirement distributions). N_t is the uncertain non-paycheck income (e.g., end of year bonuses or investment returns). In Section 6.1, we will

[18]The exceptions are that, in specification (3), the conditional correlation between the volatility of paycheck income refunds is positive only for the lowest income tercile and, in specification (5) for the middle tercile, the point estimate of the relationship between the volatility of non-paycheck income and refunds is not statistically distinguishable from zero and has a negative sign.

relate the measured income in the data to these modeled income concepts.

To simplify the analysis, and consistent with only a modest correlation between paycheck income volatility and refunds, we assume income Y_t is pre-determined. We also assume income Y_t arrives at the beginning of each period, and is available for the worker to spend in the current period. Stochastic non-paycheck income N_t is realized at the end of each period.

The key departure from an otherwise standard model is a friction: withholding decisions must be made before the resolution of income uncertainty. If this were not the case, then upon the resolution of uncertainty workers would withhold the correct amount of taxes and it would never be optimal to overwithhold. In particular, we assume stochastic non-paycheck income N_t 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 $t+1$. 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 central to the analysis. 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 tax liability $\tau(Y_t + N_t)$ in tax year 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 income Y_t.[19] 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 $W(Y_t)$. In the model, we impose that the annual withholding function and tax liability function are the same, so that liabilities equal withholding absent stochastic non-paycheck income. Hence,

$$W(Y_t) = \tau(Y_t). \tag{6}$$

[19] Recall that by "withheld" we mean the withholding from payroll income plus any extra withholding or estimated tax related to the pre-determined component of non-paycheck income. Institutionally, this assumption means that individuals do not adjust either their W-4 or estimated tax payments within the year. The assumption of non-adjustment of estimated taxes is consistent with the practice of accountants instructing taxpayers to make equal tax payments. There is good reason for this practice because if estimated tax payments are equal, taxpayers do not need to account for the timing of income during the year in calculating any penalty and interest. Making an adjustment within year in estimated tax payments triggers a requirement for a complicated calculation based on the within-year timing of income.

Given that workers also can receive stochastic non-paycheck income that is not subject to withholding, we allow workers to make an additional withholding decision meant to offset some of the tax liability from income N_t. $\widehat{W}_t \geq 0$ is the additional income tax withholding chosen by the worker. Equivalently, $\widehat{W}_t$ can be estimated tax payments, which are also dollar amounts 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 income Y_t, plus the previous period's stochastic 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 year's 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, stochastic 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

$$X_{t+1} = S_t R + Y_{t+1} - W(Y_{t+1}) + N_t - T_t \tag{7}$$

where saving is

$$S_t = X_t - C_t - \widehat{W}_t \tag{8}$$

$R > 1$ is the gross 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 stochastic 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 penalty interest rate ϕ times the shortfall between required withholding (90 percent of the tax liability) minus actual withholding.

The final settlement therefore satisfies:

$$T_t = \begin{cases} \tau(Y_t + N_t) - W(Y_t) - \widehat{W}_t & \text{if } 0.9\tau\left(Y_t + N_t\right) \leq W\left(Y_t\right) + \widehat{W}_t \\ \tau(Y_t + N_t) - W(Y_t) - \widehat{W}_t + \phi[0.9\ \tau(Y_t + N_t) - W(Y_t) - \widehat{W}_t], & \text{if } 0.9\tau\left(Y_t + N_t\right) > W\left(Y_t\right) + \widehat{W}_t \end{cases} \tag{9}$$

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

$$T_t = \tilde{\Phi}[\tau(Y_t + N_t) - W(Y_t) - \widehat{W}_t] - [(\tilde{\Phi} - 1)0.1\tau(Y_t + N_t)] \tag{10}$$

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

$$\tilde{\Phi} = \begin{cases} 1 & \text{if } 0.9\tau\left(Y_t + N_t\right) \leq W\left(Y_t\right) + \widehat{W}_t \\ 1 + \phi, & \text{if } 0.9\tau\left(Y_t + N_t\right) > W\left(Y_t\right) + \widehat{W}_t \end{cases} \tag{11}$$

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

$$\begin{aligned} V(X_t) &= \max_{S_t, \widehat{W}_t} u(C_t) + \beta \int_{N_t} V(X_{t+1})\ dN_t \qquad (12) \\ s.t.\ C_t &= X_t - \widehat{W}_t - S_t \\ X_{t+1} &= S_t R + Y_{t+1} - W(Y_{t+1}) + N_t - T_t \\ T_t &= \tilde{\Phi}[\tau(Y_t + N_t) - W(Y_t) - \widehat{W}_t] - [(\tilde{\Phi} - 1)0.1\tau(Y_t + N_t)] \\ \widehat{W}_t, S_t &\geq 0 \end{aligned}$$

We do not consider the possibility of borrowing, so savings is constrained to be non-negative.[20]

[20]Borrowing to make tax payments is very likely dominated by overwithholding, or even paying interest on

5.2 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 liquidity it has chosen to give away in the form of overwithholding. The optimality conditions for problem (12) 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.

5.2.1 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

$$u'(C_t) \geq \beta R \int_{N_t} V'(X_{t+1})\, dN_t \tag{13}$$

with equality if $S_t > 0$. Written in terms of consumption, (13) becomes the standard consumption Euler equation, that is,

$$u'(C_t) \geq \beta R \int_{N_t} u'(C_{t+1})\, dN_t. \tag{14}$$

Deriving (14) from (13) 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 to stochastic non-paycheck income, N_t.

underwithholding, because interest rates on (unsecured) credit are typically much higher than the short-term rate plus 3% charged by the IRS.

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

$$u'(C_t) \geq \beta \int_{N_t} V'(X_{t+1})\tilde{\Phi} \, dN_t \tag{15}$$

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

$$u'(C_t) \geq \beta \int_{N_t} u'(C_{t+1})\tilde{\Phi} \, dN_t. \tag{16}$$

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 (16), 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 $R > 1$.[21] 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 $1 + \phi > R$ 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 $1 + \phi > R$, 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.

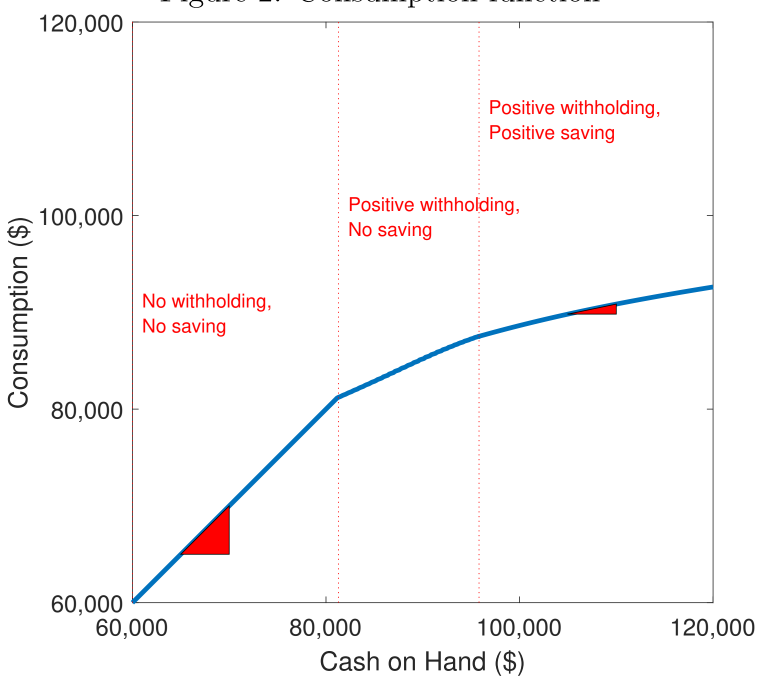
[21] The wedge is even greater for those borrowing on credit cards. The model does not address that behavior.

5.2.2 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 of consumption is concave. The effect on consumption of transitory income depends on the level of cash on hand.

Figure 2 presents the consumption function for a calibrated version of the model we study. A detailed description of the calibration is postponed until Section 6. The figure shows the endogenous liquidity constraints that emerge from income uncertainty as consumption is a concave function of cash on hand.

Figure 2: Consumption function



Notes: The consumption function shows optimal consumption as a 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. Figure 2 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 has a slope equal to one because hand-to-mouth consumers do not save at all ($\hat{W} = 0, S = 0$) despite the fact that the liquidity constraint does not literally bind. As cash on hand increases, the consumer has enough resources to make optimal saving positive. As discussed in the previous section, consumers should always start with putting their extra resources in additional withholding ($\hat{W}$) instead of private saving (S) because the return is $1 + \phi > R$ 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 stochastic 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 $S > 0$.

Figure 2 also illustrates the quantitative importance of the concavity. The figure shows the effect of receiving a \$5,000 refund in two situations, one where cash on hand is relatively low (the triangle to the left) and one where cash on hand is relatively high (the triangle to the right). The base of each triangle is the \$5,000 refund and the height is the spending due to it. The low-liquidity state is in the no withholding and no saving region for cash on hand and the MPC is about one. The high-liquidity state is in the positive saving and positive withholding region for cash on hand and the MPC less than a quarter the size.

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-year's 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.

6 Model: Calibration and Empirical Implications

We next calibrate the model and present its empirical implications. In this section, we first provide the details of the calibration including functional forms, parameters, and moments. We then evaluate the model on its ability to explain both the average size and correlates of tax refunds, and the response of spending to the arrival of refunds.

6.1 Relating Measured Income to Modeled Moments

The estimated mean and variance of the different sources of income are key inputs for calibrating the model. In particular, the variability of non-paycheck income is central to the mechanism driving the size and response to tax refunds. That mechanism, however, depends on the additional withholding of taxes on non-paycheck income being pre-determined with respect to the realization of income not subject to withholding. If, instead, the observed fluctuations in such income were fully anticipated, the taxpayer could adjust withholding on paycheck income, or make estimated tax payments, and refunds would never be optimal. The data provide an estimate of the variability of observed income but no feasible way to estimate directly the extent to which the variability corresponds to anticipated versus unanticipated realizations. This section therefore introduces the parameter γ, defined to be the fraction of the observed variability of non-paycheck income that is not fully anticipated by the taxpayer. We will treat anticipated fraction, $(1-\gamma)$, as if it were pre-determined, paycheck income. This additional parameter thus provides a degree of freedom for targeting the mean level of tax refunds.

We assume that the taxpayer can observe separately the predetermined, anticipated paycheck income and the stochastic component of non-paycheck income, but that the econometrician cannot. We can, however, infer the fraction γ given the structure of the model. At the annual frequency of the model, Y_t paycheck income plus anticipated non-paycheck income are non-stochastic. The mean of Y_t, denoted α_Y, is

$$\alpha_Y = \alpha_P + (1-\gamma)\alpha_{NP} \tag{17}$$

where α_P and α_{NP} are the mean of observed paycheck and non-paycheck income. When we simulate the model, since Y_t is non-stochastic, its value is constant at this mean.

The only source of uncertainty at annual frequency is the stochastic component of non-paycheck income N_t. The fraction of non-paycheck income that is unanticipated is γ, so the relation of the mean α_N and standard deviation σ_N of modeled income N_t are

$$\alpha_N = \gamma \alpha_{NP} \tag{18}$$

$$\sigma_N = \gamma \sigma_{\nu^{NP}} \tag{19}$$

where α_{NP} and $\sigma_{\nu^{NP}}$ are, respectively, the mean and standard deviation of observed non-paycheck income. That the variability of stochastic non-paycheck income σ_N is only a fraction of the variability of observed non-paycheck income $\sigma_{\nu^{NP}}$ is a consequence of the presumption that a fraction of observed fluctuations in non-paycheck income are known in advance by the taxpayer.

In summary, consistent with the model, we thus treat the anticipated fraction of non-paycheck income as if it were pre-determined, paycheck income. The complement of that fraction is treated as uncertain, non-paycheck income and its standard deviation is scaled accordingly. The parameter γ is apposite for the calibration exercise: It is tightly related to the mechanism generating tax refunds, but not directly a function of the moments of the income process or tax refunds.

Table 4 gives the empirical moments used in calibrating the model. The sample mean of paycheck and non-paycheck income and the sample standard deviation of the annual innovation in non-paycheck income—all from Table 1—fully characterize the income processes necessary for calibrating the model since the only source of uncertainty at annual frequency is stochastic non-paycheck income. The mean of tax refunds conditional on receiving a tax refund from the top row of Table 2 is the targeted moment in the calibration. (Note that the model variable $\overline{T}$ is final settlements; refunds are negative final settlements.)

6.2 Functional Forms and Tax Parameters

Table 5 summarizes the functional forms and tax parameters that enter the calibrated model. We adopt a standard constant relative risk aversion form for utility $u(C) = \frac{C^{1-\theta}}{1-\theta}$. The tax liability function $\tau()$ is a numeric approximation to the tax schedule as a function of income with a fifth degree polynomial. See Appendix A.6 for details. The rate of interest $R - 1$ is 0.2 percent, the sample average of the one-year Treasury note yield over the sample period. The penalty rate ϕ is the interest rate plus the three percentage point statutory premium for underpayments of estimated tax.

Table 4: Model Calibration: Moments

Moment	Value	Notes
$\overline{\alpha}_P$	\$68,226	mean of paycheck income
$\overline{\alpha}_{NP}$	\$38,764	mean of non-paycheck income
$\overline{\sigma}_{\nu^{NP}}$	\$19,879	standard deviation of non-paycheck income
$-\overline{T}$	\$3,184	mean of tax refunds [targeted]

Notes: Refunds are negative final settlements.

Table 5: Model Calibration: Functions and Tax Parameters

Parameter/function	Value	Notes
$u(C)$	$\frac{C^{1-\theta}}{1-\theta}$	utility function
$\tau(Y+N)$	see Appendix A.6	approximate tax liability schedule
$R-1$	0.002	1-year Treasury yield
ϕ	0.032	penalty rate = (R-1) + 3%

Table 6: Model Calibration: Baseline Parameters

Parameter	Value	Fixed or Calibrated	Notes
θ	1	Fixed	risk aversion
β	0.985	Fixed	discount factor
γ	0.5674	Calibrated	fraction non-paycheck income unanticipated
α_Y	\$84,995	Implied	$\overline{\alpha}_P + (1-\gamma)\overline{\alpha}_{NP}$
α_N	\$21,995	Implied	$\gamma\overline{\alpha}_{NP}$
σ_N	\$11,279	Implied	$\gamma\overline{\sigma}_{\nu^{NP}}$

6.3 Baseline Solution and Calibration

Table 6 presents the levels of the parameters for the baseline calibration and simulation of the model. For this baseline, we assume a coefficient of relative risk aversion equal to one (log utility) to ensure the results do not depend on unusually high degrees of risk aversion. We set the time discount factor to 0.985.

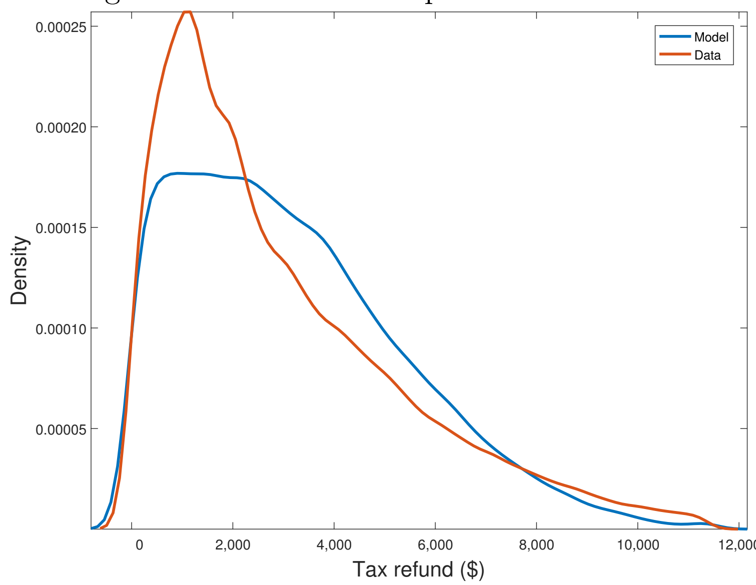
To simulate the model and estimate the distribution of tax refunds received by the agent solving the optimization problem in Section 5.1, the estimated income process, and the tax parameters, we draw 100,000 realizations of the innovation of non-paycheck income assuming the normal distribution and the moments in Table 4 given the functional form and paramater values. We then solve the model in each period for the fixed parameter values in Table 6 and for a grid of values for the calibrated parameter γ. We throw out the first 1,000 simulated observations so that simulated moments are based on the ergodic distribution. The calibrated parameter $\gamma = 0.5674$ yields a simulated average refund that exactly matches the targeted average refund equal to $\$3,184$ in the data. Note that in the data and the calibrated model, this average is conditional on refunds being positive.

The finding that only about half of non-paycheck income is anticipated sufficiently for taxpayers to make exact tax pre-payments strikes us as reasonable given its significant variability documented in these data. We do not have data that speak directly to the fraction of non-paycheck income that is unanticipated. The calibrated value of γ provides some independent evidence on the unanticipated fraction where direct evidence is difficult to obtain.

Despite the simple structure of the model and just one degree of freedom, the model is successful at fitting the average tax refund, $3,184, with a coefficient of relative risk aversion of 1, and discount factor of 0.985. As discussed in Section 5.2.1, we expect workers to overwithhold on average in order to avoid the underwithholding penalty. These results show that, with a conventional calibration of other model parameters, the estimated level and variation in unanticipated non-paycheck income is sufficient to justify the average refund in the data. In this way, and with a sensible value of the one free parameter, the model is able to explain the average level of tax refunds using very standard values of utility function parameters. That is, the friction posited in the model that withholding and estimated tax decisions are made in advance of the realization of income for the year, together with the asymmetry in being over- versus underwithheld arising from the penalties built into the tax code, can explain the high average level of refunds.

The simulation yields an estimate of the distribution of tax refunds, not just its mean. To get a sense of how the variability of income driving the model explains the distribution of refunds, Figure 3 shows the simulated and empirical distribution of tax refunds from the model and data. Table 7 shows the predicted and empirical average and quartiles of the final tax settlement distribution. The empirical distribution is somewhat more concentrated than the model distribution. Nonetheless, the shape and the range of the simulated distribution is quite close to the empirical distribution. Note that the calibration relies just on the mean of tax refunds; that the model captures much of the variability of tax refunds derives from the modeled implication of the observed variability of income. Hence, while the targeting of the average tax refund is exact given the degree of freedom in the calibration, matching the variability of refunds is not at all baked into the simulations.

Figure 3: Distribution of positive tax refunds



Notes: This figure plots the estimated density of tax refunds from the model and data respectively. While the underlying data only include positive tax refunds, the estimated density will exhibit some mass at points below 0 due to the smoothing process.

Table 7: Distribution of tax refunds ($)

	mean	p25	p50	p75
Model	3,184	1,363	2,712	4,564
Data	3,184	1,148	2,353	4,506

Notes: This table shows statistics of the tax refund distributions plotted in Figure 3.

6.4 Alternative Parameterizations

The model's qualitative prediction of a large average tax refund is not a knife-edge case that relies on a special combination of parameters. Table 8 shows how the average size of tax refunds predicted by the model, with log utility, depends on the time discount factor, (β), and on the fraction of income volatility that is unanticipated (γ). The average tax refund is increasing in the discount factor as more patient workers place more weight on the consequences of income uncertainty for future utility and thus "save" more to avoid the costs of being underwithheld. Similarly, increases in γ raise the effective risk that workers face, and thus their incentives to guard against the tax penalties for being underwithheld due to a large, positive income realization.[22] There is a narrow range of combinations of γ and β that yield a simulated tax refund close to that found in the data.

Table 8: Average tax refund under different parameter values ($\theta = 1$)

		β				
		0.975	0.980	0.985	0.990	0.995
	0.30	1,133	1,191	1,240	1,291	1,342
	0.40	1,818	1,895	1,986	2,076	2,169
	0.50	2,489	2,590	2,692	2,801	2,906
γ	0.57	2,967	3,071	3,184	3,299	3,416
	0.60	3,226	3,342	3,464	3,593	3,723
	0.70	3,932	4,083	4,222	4,355	4,502
	0.80	4,715	4,906	5,094	5,254	5,434

Notes: This table calculates the average tax refund for 100,000 simulated observations under different parameter values.

6.5 Explaining the Level and Slope of the MPC

The previous section showed how the model explains large average tax refunds. The monetary costs of underwithholding combined with uncertainty in non-paycheck income makes it optimal to overwithhold. This section starts by estimating the MPC out of refunds using the data. This section uses the model simulations to understand why workers spend

[22] Average tax refunds are also modestly increasing in risk aversion (θ) as the usual precautionary motives also increase incentives to "save" in the form of overwithholding. See Appendix Table I.1 for the analogue of 8, but with risk aversion, θ, set to 4 instead of 1.

a large fraction of their refunds when they arrive, and why that fraction tends to rise with the size of the refund.

6.5.1 The Average MPC Out of Refunds: Estimates from the Account Data

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. 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. (2021) 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. (2021) are 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.

We estimate the MPC using the data with the following specification:

$$\tilde{C}_{it} = \alpha + MPC_{average} \times \widetilde{Refund}_{it} + \sum_{j=2}^{12} month_j + \varepsilon_{it} \tag{20}$$

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, and $month_j$ are month fixed effects.

Table 9: MPC estimates: Data

	(1)	(2)
$MPC_{average}$	0.206	
	(0.002)	
$MPC_{quantile1}$		0.187
		(0.006)
$MPC_{quantile2}$		0.192
		(0.005)
$MPC_{quantile3}$		0.206
		(0.004)
$MPC_{quantile4}$		0.210
		(0.004)
$MPC_{quantile5}$		0.232
		(0.005)
Constant	0.264	0.287
	(0.000)	(0.001)
NxT	2,248,386	2,248,386
N	49,519	49,519
R^2	0.035	0.045

Notes: Robust standard errors in parenthesis. Estimates include monthly fixed effects for (1) and (2) and quantile fixed effects for (2) (not reported). NxT represents the number of individual-month observations. N represents the number of individual observations.

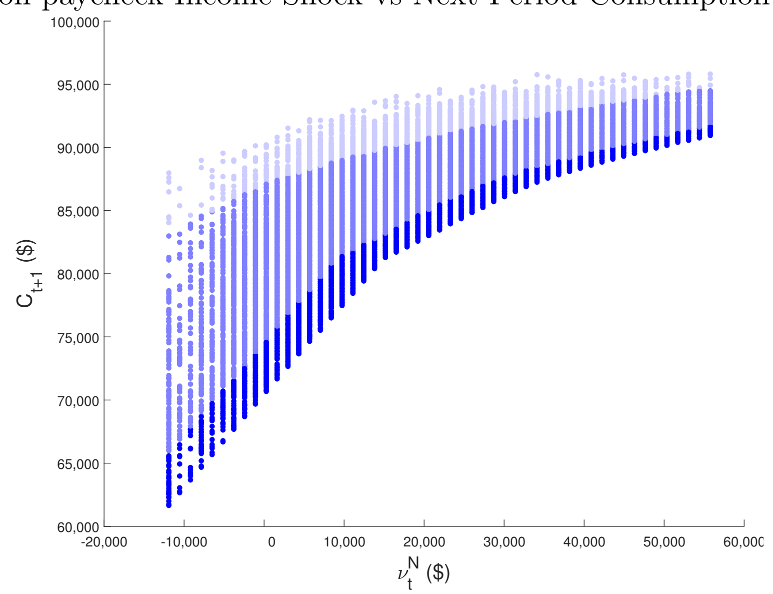
Column (1) of Table 9 shows the estimates of the MPC. The average MPC in our data of 0.21 is within range of the empirical estimates in the literature. Previous consumption models have explicitly targeted the MPC. We calibrate our model to match the average level of refunds and then deliver the MPC as a non-targeted finding.

6.5.2 The Slope of the MPC With Respect to Refunds: Model Simulations

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

To understand better the mechanism linking shocks, refunds, and expenditure, Figure 4 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 shades represent different levels of cash on hand (X_t). Darker shades 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.

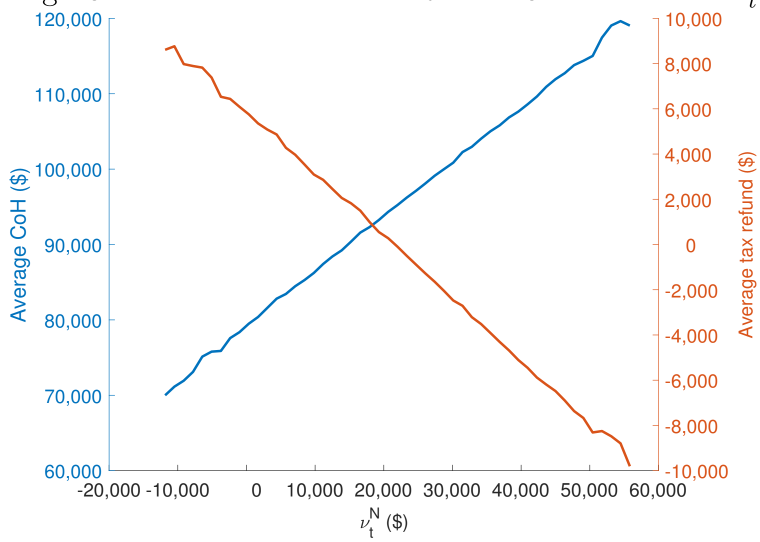
Figure 4: Non-paycheck Income Shock vs Next Period Consumption: Simulation



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 shown directly in Figure 5. When a worker experiences a negative non-paycheck income shock (ν_t^N), this directly reduces cash on hand. 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 5: Average Cash on Hand and Tax Refund Conditional on ν_t^N: Simulation

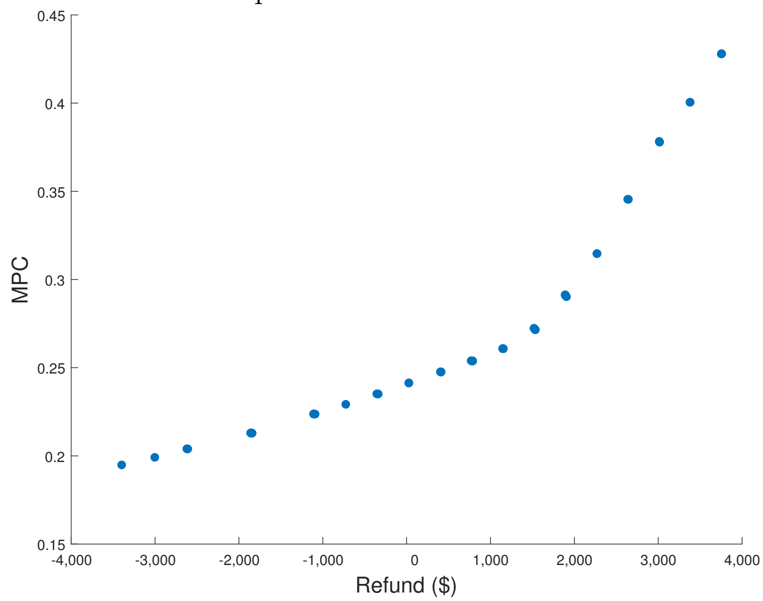


Notes: This figure plots the relationship between the non-paycheck income shock (ν_t^N) and average cash on hand on the left Y-axis and average tax refund on the right Y-axis. 100,000 simulated observations.

We can combine the mechanisms described in Figures 4 and 5 to characterize the relationship between the MPC and tax refunds in the calibrated model. Figure 6 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 average cash on hand, a negative non-paycheck income shock leads to a large refund because tax liability is lower than expected. At the same time, the negative shock results in lower cash on hand levels and, because the consumption function is concave, lower cash on hand leads to a higher MPC. In this way, the model predicts both that workers will spend substantial fractions of tax refunds, on average, and that the larger the refund the higher the MPC.[23]

[23]This prediction of a positive correlation between refunds and the MPC is distinct from the finding in Kueng (2018) of a positive correlation between income and MPCs from Alaska Permanent Fund, where all fund payments are equal in size and higher income households have a higher MPC from them. In the next subsection, we evaluate the mechanism in our model that the larger refunds associated with smaller paycheck

Figure 6: Predicted Relationship Between the MPC and Tax Refunds: Simulation



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.

6.5.3 Testing the Link Between the MPC and Refunds

To evaluate the model's distinctive prediction of a positive association between tax refunds and the MPC, we turn again to the individual-level data.[24] Specifically, we estimate the MPC as a function of tax refund size using the following specification

$$\tilde{C}_{it} = \alpha + \sum_{j=1}^{5} MPC_j \times \widetilde{Refund}_{it} \times Q_i^j + \sum_{j=2}^{5} Q_i^j + \sum_{j=2}^{12} month_j + \varepsilon_{it} \tag{21}$$

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 $month_j$ are month fixed effects. MPC_j is the average MPC out of refunds for each quintile of estimated tax refunds.

To isolate changes in refunds attributable to the non-paycheck income mechanism in the model, we adopt a two-stage estimation strategy that, in the second stage, replaces the

shares are associated with larger MPCs.

[24]The upward sloping relationship between the level of tax refunds and the MPC is also qualitatively consistent with the results in Baugh et al. (2021) who find that the MPC out of tax refunds is much higher than the MPC out of tax payments.

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 9 and summary statistics of the predicted quintiles are provided in Appendix Table D.1.

The results of column (2) in Table 9 are consistent with the link between the MPC and tax refunds derived from the model simulations in the previous sub-section. 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.[25] The simulated MPCs rise similarly with the level of refunds. The correlation between the MPCs estimated across quintiles in Table 9 and the analogous simulated values is 0.91.

The levels of the MPC are difficult to compare across simulation and estimates because the model is stylized. In particular, the model presumes all spending is on a single non-durable good. The levels of the simulated MPCs and estimated MPCs are nevertheless quite close. 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.

Hence, in the important dimension of how the MPC varies with the level of the refund, the model simulations closely accord with the estimates from the data. This result provides strong support for the features of the model since the models calibration does not rely on the estimated MPCs. That said, we are cautious in interpreting the results since, as discussed above, the magnitudes of the levels of the MPCs from the simulation and from the data are not directly comparable.

[25] An F-test for the equality of coefficients results in a p- value < 0.0001.

7 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. 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 when 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.

Account data on income, spending, and refunds show that the average level of and variation in non-paycheck income is 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 MPCs out of 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 et al. (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 can arise 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, "Debt and the response to household income shocks: Validation and application of linked financial account data," *Journal of Political Economy*, 2017.

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

Baugh, Brian, Itzhak Ben-David, Hoonsuk Park, and Jonathan A. Parker, "Asymmetric Consumption Smoothing," *American Economic Review*, January 2021, *111* (1), 192–230.

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

Braxton, J. Carter, Gordon Phillips, and Kyle Herkenhoff, "Can the Unemployed Borrow? Implications for Public Insurance," in "2018 Meeting Papers" number 564 Society for Economic Dynamics 2018.

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

Carroll, Christopher D., "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis," *The Quarterly Journal of Economics*, 1997, *112* (1), 1–55.

_ **and Miles S. Kimball**, "On the Concavity of the Consumption Function," *Econometrica*, 1996, *64* (4), 981–992.

Carroll, Christopher, Jiri Slacalek, Kiichi Tokuoka, and Matthew N. White, "The distribution of wealth and the marginal propensity to consume," *Quantitative Economics*, 2017, *8* (3), 977–1020.

Christelis, Dimitris, Dimitris Georgarakos, Tullio Jappelli, Luigi Pistaferri, and Maarten van Rooij, "Asymmetric Consumption Effects of Transitory Income Shocks*," *The Economic Journal*, August 2019, *129* (622), 2322–2341.

Deaton, Angus, "Saving and Liquidity Constraints," *Econometrica*, 1991, *59* (5), 1221–1248.

Debacker, Jason, Bradley Heim, Vasia Panousi, Shanthi Ramnath, and Ivan Vidangos, "Rising Inequality: Transitory or Persistent? New Evidence from a Panel of U.S. Tax Returns," *Brookings Papers on Economic Activity*, 2013, *2013* (1), 67–142.

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

Fuster, Andreas, Greg Kaplan, and Basit Zafar, "What Would You Do with $500? Spending Responses to Gains, Losses, News, and Loans," *The Review of Economic Studies*, November 2020, (rdaa076).

Ganong, Peter and Pascal Noel, "Consumer Spending during Unemployment: Positive and Normative Implications," *American Economic Review*, July 2019, *109* (7), 2383–2424.

Gelman, Michael, "What drives heterogeneity in the marginal propensity to consume? Temporary shocks vs persistent characteristics," *Journal of Monetary Economics*, 2021, *117*, 521–542.

_, Shachar Kariv, Matthew D. Shapiro, Dan Silverman, and Steven Tadelis, "Harnessing naturally occurring data to measure the response of spending to income," *Science*, July 2014, *345* (6193), 212–215.

_ , _ , _ , _ , and _ , "How individuals respond to a liquidity shock: Evidence from the 2013 government shutdown," *Journal of Public Economics*, 2020, *189*, 103917.

Guvenen, Fatih, Serdar Ozkan, and Jae Song, "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, June 2014, *122* (3), 621–660.

Herkenhoff, Kyle F, "The Impact of Consumer Credit Access on Unemployment," *The Review of Economic Studies*, November 2019, *86* (6), 2605–2642.

Highfill, Jannett, Douglas Thorson, and William V. Weber, "Tax Overwithholding as a Response To Uncertainty," *Public Finance Review*, 1998, *26* (4), 376–391.

Hsieh, Chang-Tai, "Do Consumers React to Anticipated Income Changes? Evidence from the Alaska Permanent Fund," *American Economic Review*, March 2003, *93* (1), 397–405.

Jones, Damon, "Inertia and Overwithholding: Explaining the Prevalence of Income Tax Refunds," *American Economic Journal: Economic Policy*, February 2012, *4* (1), 158–85.

Kaplan, Greg and Giovanni L Violante, "A model of the consumption response to fiscal stimulus payments," *Econometrica*, 2014, *82* (4), 1199–1239.

_, Giovanni L. Violante, and Justin Weidner, "The Wealthy Hand-to-Mouth," *Brookings Papers on Economic Activity*, 2014, *45* (1 (Spring)), 77–153.

Koustas, Dmitri, "Consumption insurance and multiple jobs: Evidence from rideshare drivers," *Unpublished working paper*, 2018.

Kuchler, Theresa and Michaela Pagel, "Sticking to your plan: The role of present bias for credit card paydown," *Journal of Financial Economics*, February 2021, *139* (2), 359–388.

Kueng, Lorenz, "Excess Sensitivity of High-Income Consumers," *The Quarterly Journal of Economics*, November 2018, *133* (4), 1693–1751.

Neumark, David, "Are Rising Earnings Profiles a Forced-Saving Mechanism?," *The Economic Journal*, 1995, *105* (428), 95–106.

Parker, Jonathan A., "The Reaction of Household Consumption to Predictable Changes in Social Security Taxes," *American Economic Review*, September 1999, *89* (4), 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, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, 2011, *12*, 2825–2830.

Rees-Jones, Alex, "Quantifying Loss-Averse Tax Manipulation," *The Review of Economic Studies*, April 2018, *85* (2), 1251–1278.

Shapiro, Matthew D. and Joel Slemrod, "Consumer Response to Tax Rebates," *The American Economic Review*, 2003, *93* (1), 381–396.

Souleles, Nicholas S., "The Response of Household Consumption to Income Tax Refunds," *The American Economic Review*, 1999, *89* (4), 947–958.

Stephens, Melvin, ""3rd of tha Month": Do Social Security Recipients Smooth Consumption between Checks?," *The American Economic Review*, 2003, *93* (1), 406–422.

Thaler, Richard H., "Psychology and Savings Policies," *The American Economic Review*, 1994, *84* (2), 186–192.

Wilcox, David W., "Social Security Benefits, Consumption Expenditure, and the Life Cycle Hypothesis," *Journal of Political Economy*, 1989, *97* (2), 288–304.

Zeldes, Stephen P, "Consumption and liquidity constraints: An empirical investigation," *The Journal of Political Economy*, 1989, *97* (2), 305–346.

Zeldes, Stephen P., "Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence," *The Quarterly Journal of Economics*, May 1989, *104* (2), 275–298.

A Appendix – Data Filters, Definitions

The main analysis sample is filtered appropriately 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 consequences for sample size are presented in Table A.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.[26] We condition on four statistics to ensure that these transactions represent regular paychecks.

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

A.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 2 shows

[26] 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."

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

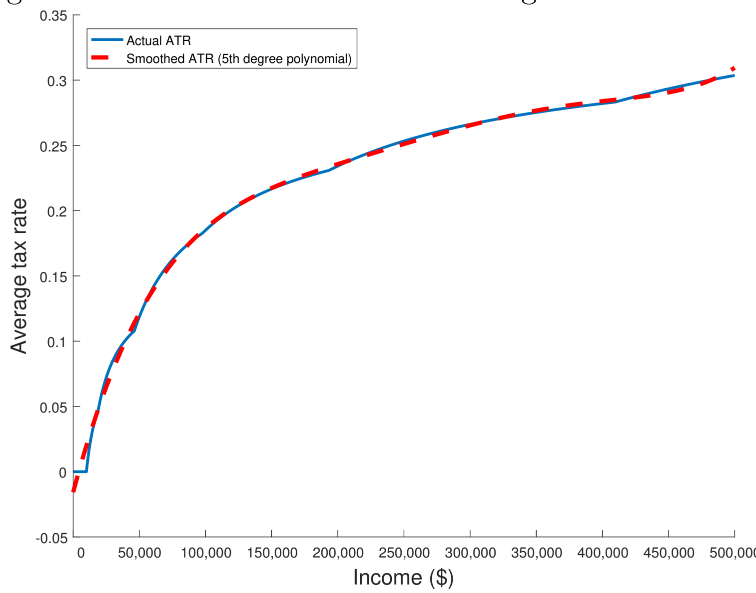
	Individuals	%
Full sample as of December 2012	883,529	100
Long time series (N $\geq$ 40)	341,841	39
Linked ratio $\geq$ 0.8	264,043	30
Linked accounts $\in$ [3,15]	197,530	22
Has regular bi-weekly paycheck	92,883	11
Has stable paycheck	62,946	7

A.6 2013 Tax Schedule

The tax function is based on 2013 average tax rates (ATR). It is calculated from the Stata package *taxliab* for income values over the range $0 to $500,000 in $100 intervals. The package calculates the ATR from the marginal tax rate schedule. We assume that individuals are single filers who claim two personal exemptions ($3,900 each) and the standard deduction ($6,100).[27] 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.

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

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

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

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)[28]

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

$$\tilde{p}_{i,b,t} = f(p_{i,b,t}; s, e, t) \tag{22}$$

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

Pre-tax benefits such as health insurance premiums and 401(k) contributions also lead to differences in gross and net paycheck income. We do not adjust for these benefits as we do the types of withholding listed above. We don't see this income, but equally we dont see its consumption. Moreover, these benefits are generally not subject to income taxation that we are modeling. Hence, that they are excluded from both income and spending in the data is fortuitously correct. The same argument holds for pension benefits, but with a more complicated intertemporal accounting.

C Appendix – The Withholding Function

C.1 Measuring Excess Withholding Due to High Frequency Paycheck Volatility

As noted in section 3, the rules governing paycheck tax withholding schedules, and the convexity of the tax schedule, may induce a "mechanical" relationship between within-year paycheck volatility and refunds. To quantify the magnitude of the mechanical effect we define excess withholding from high frequency paycheck volatility as:

[28] 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.

$$ExcessW_{i,y} = \sum_{b=1}^{26}(w(p_{i,b,t}; s, e, t) - w(\bar{p}_{i,t}; s, e, t)) \quad (23)$$

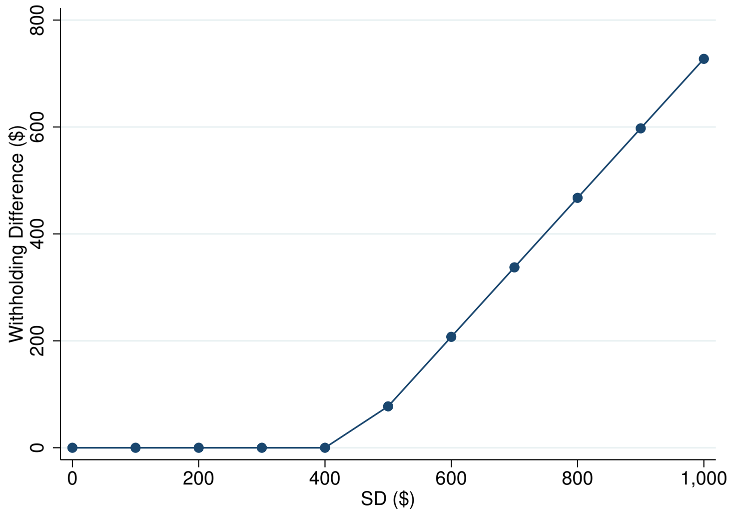
where $w(\cdot; s, e, t)$ is a periodic withholding function that takes paycheck income as its argument and is influenced by filing status s, number of exemptions e, and year t.[29] $p_{i,b,t}$ is the bi-weekly pre-withholding paycheck for individual i in bi-week b of year t, and $\bar{p}_{i,t}$ is the average bi-weekly pre-withholding paycheck for individual i in year t.[30] We assume single filing status and two exemptions in our calculations of excess withholding.

Figure C.1 illustrates the relationship between this measure of potential excess withholding and within-year paycheck volatility. The example in the figure assumes paychecks are one standard deviation above average half the time and one standard deviation below average the other half of the time. As expected, the measure of potential excess withholding, $ExcessW_{i,t}$, 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.

[29]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.

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

Figure C.1: $ExcessW_y$ as a function of within-year paycheck variation



Notes: $ExcessW_{i,y}$ 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.

C.2 Calibrating the Withholding Function

For purposes of calculating excess withholding, the withholding function is calibrated using IRS publication 15 (aka circular E).[31] Figure C.2 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.2: 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.

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

D Appendix – Predicted Refunds Statistics

This appendix shows summary staitstics for the regression reported in Table 9, column (2).

Table D.1: Summary statistics for each predicted refund quintile ($)

Q_i^j	Mean	p25	p50	p75
1	2,611	2,546	2,623	2,686
2	2,859	2,803	2,859	2,913
3	3,095	3,031	3,092	3,157
4	3,395	3,305	3,389	3,482
5	3,986	3,732	3,908	4,174
Total	3,189	2,803	3,092	3,482

E Appendix – Correlates of Tax Refunds by Terciles of Income

Table E.1: Tax refunds and income volatility by income tercile: $Log(Refund)_{it}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Top							
$payshare_i$	-1.182				-0.903	-1.170	
	(0.0396)				(0.0588)	(0.0456)	
$Log(\sigma^2_{\nu_i^{NP}})$		0.0856	0.0902		0.0424		0.0928
		(0.00462)	(0.00449)		(0.00665)		(0.00550)
$Log(\sigma^2_{\nu_i^{P}})$			-0.0220				
			(0.00428)				
$Log(ExcessW_{it-1})$				0.0143	0.0123		
				(0.00259)	(0.00256)		
NxT	46,380	46,380	46,380	29,405	29,405	29,405	29,405
N	26,418	26,418	26,418	21,391	21,391	21,391	21,391
R^2	0.051	0.038	0.039	0.015	0.051	0.047	0.038
Panel B: Middle							
$payshare_i$	-0.829				-0.945	-0.881	
	(0.0350)				(0.0506)	(0.0417)	
$Log(\sigma^2_{\nu_i^{NP}})$		0.0419	0.0433		-0.00708		0.0476
		(0.00353)	(0.00355)		(0.00503)		(0.00418)
$Log(\sigma^2_{\nu_i^{P}})$			-0.0165				
			(0.00394)				
$Log(ExcessW_{it-1})$				0.00997	0.0173		
				(0.00286)	(0.00287)		
NxT	48,566	48,566	48,566	31,061	31,061	31,061	31,061
N	29,335	29,335	29,335	23,250	23,250	23,250	23,250
R^2	0.029	0.016	0.017	0.015	0.035	0.034	0.021
Panel C: Bottom							
$payshare_i$	-0.295				-0.132	-0.341	
	(0.0374)				(0.0542)	(0.0423)	
$Log(\sigma^2_{\nu_i^{NP}})$		0.0469	0.0422		0.0423		0.0524
		(0.00375)	(0.00375)		(0.00541)		(0.00426)
$Log(\sigma^2_{\nu_i^{P}})$			0.0419				
			(0.00432)				
$Log(ExcessW_{it-1})$				0.0377	0.0374		
				(0.00388)	(0.00392)		
NxT	39,806	39,806	39,806	27,246	27,246	27,246	27,246
N	21,142	21,142	21,142	18,025	18,025	18,025	18,025
R^2	0.006	0.010	0.014	0.008	0.016	0.007	0.012

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 (4) and (5) are based on one fewer year's observations to allow for the lagged variable. Columns (6) and (7) repeat the estimates of columns (1) and (2) with this sample.

F 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 $\widehat{W_t}$. $u'(C_t(\widehat{W_t})) = \beta \int_\nu u'(C_{t+1}(\widehat{W_t}))\tilde{\Phi} d\nu$

   (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 back out current period $C_t = u'^{-1}(\beta R \int_\nu u'(C_{t+1}) d\nu)$

   (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

G Appendix – Estimating the Parameters of the Income Process

The following equations derive expressions for each of our income parameters as functions of the theoretical moments. Upper case variables represent annual variables and lower case variables represent bi-weekly variables. The theoretical moments are then estimated using sample moments calculated from within-individual variation across time. Lastly, the model parameters are calculated from the individual-level parameters by averaging across individuals.

$\alpha_{i,P}$

$$\mathbb{E}[p_{i,t,b}] = \mathbb{E}\left[\frac{P_{i,t}}{26}\right] + \mathbb{E}[\epsilon^P_{i,t,b}] \tag{24}$$

$$\alpha_{i,P} = \mathbb{E}[p_{i,t,b}]26 \tag{25}$$

$$\bar{\alpha}_{i,P} = \frac{\sum_t p_{i,t,b}}{T}26 \tag{26}$$

$$\bar{\alpha}_P = \frac{\sum_i \bar{\alpha}_{i,P}}{N} \tag{27}$$

σ^2_{i,ϵ^P}

$$p_{i,t,b} - \bar{p}_{i,t} = \epsilon^P_{i,t,b} \tag{28}$$

$$\sigma^2_{i,\epsilon^P} = \mathbb{V}[\epsilon^P_{i,t,b}] \tag{29}$$

$$\overline{\sigma^2}_{i,\epsilon^P} = \frac{\sum_t (\epsilon^P_{i,t,b})^2}{T} \tag{30}$$

$$\overline{\sigma^2}_{\epsilon^P} = \frac{\sum_i \overline{\sigma^2}_{i,\epsilon^P}}{N} \tag{31}$$

σ^2_{i,ν^P}

$$\mathbb{V}[p_{i,t,b}] = \mathbb{V}\left[\frac{P_{i,t}}{26}\right] + \mathbb{V}[\epsilon^Y_{i,t,b}] \tag{32}$$

$$\mathbb{V}[p_{i,t,b}] = \frac{\sigma^2_{i,\nu^P}}{26^2} + \sigma^2_{i,\epsilon^P} \tag{33}$$

$$\overline{\sigma^2}_{i,\nu^P} = \left(\frac{\sum_t (p_{i,t,b})^2}{T} - \overline{\sigma^2}_{i,\epsilon^P}\right)26^2 \tag{34}$$

$$\overline{\sigma^2}_{\nu^P} = \frac{\sum \sigma^2_{i,\nu^P}}{N} \tag{35}$$

$\alpha_{i,NP}$

$$\mathbb{E}[np_{i,t,b}] = \mathbb{E}\left[\frac{NP_{i,t}}{26}\right] \tag{36}$$

$$\alpha_{i,NP} = \mathbb{E}[np_{i,t,b}]26 \tag{37}$$

$$\bar{\alpha}_{i,NP} = \frac{\sum_t np_{i,t,b}}{T}26 \tag{38}$$

$$\bar{\alpha}_{NP} = \frac{\sum_i \bar{\alpha}_{i,NP}}{NP} \tag{39}$$

$\sigma^2_{i,\epsilon^{NP}}$

$$np_{i,t,b} - \overline{np}_{i,t} = \epsilon^{NP}_{i,t,b} \tag{40}$$

$$\sigma^2_{i,\epsilon^{NP}} = \mathbb{V}[\epsilon^{NP}_{i,t,b}] \tag{41}$$

$$\overline{\sigma^2}_{i,\epsilon^{NP}} = \frac{\sum_t (\epsilon^{NP}_{i,t,b})^2}{T} \tag{42}$$

$$\overline{\sigma^2}_{\epsilon^{NP}} = \frac{\sum_i \overline{\sigma^2}_{i,\epsilon^{NP}}}{N} \tag{43}$$

$\sigma^2_{i,\nu^{NP}}$

$$\mathbb{V}[np_{i,t,b}] = \mathbb{V}\left[\frac{NP_{i,t}}{26}\right] + \mathbb{V}[\epsilon^{NP}_{i,t,b}] \tag{44}$$

$$\mathbb{V}[np_{i,t,b}] = \frac{\sigma^2_{i,\nu^{NP}}}{26^2} + \sigma^2_{i,\epsilon^{NP}} \tag{45}$$

$$\overline{\sigma^2}_{i,\nu^{NP}} = \left(\frac{\sum_t (np_{i,t,b})^2}{T} - \overline{\sigma^2}_{i,\epsilon^{NP}}\right) 26^2 \tag{46}$$

$$\overline{\sigma^2}_{\nu^{NP}} = \frac{\sum \sigma^2_{i,\nu^{NP}}}{N} \tag{47}$$

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.[32] 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 $\hat{Y}_0 = \hat{f}(X_0)$.

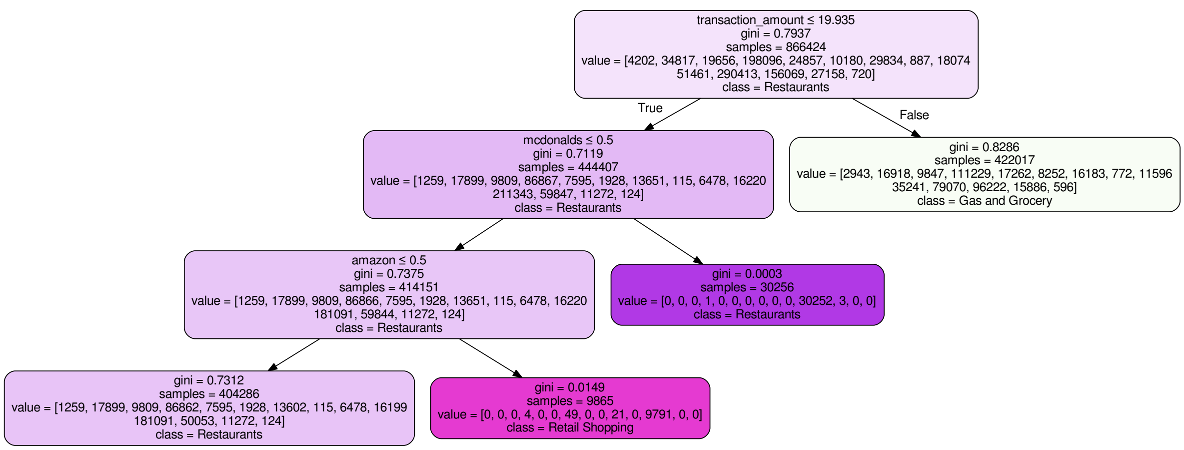
One option for the mapping is to use the multinomial logit model since the dependent variable is a categorical variable with no cardinal meaning. However, this approach is not well suited to textual data because each word would need its own dummy variable. Furthermore, interactions may be important for 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 we use a machine learning algorithm called random forest. A random forest model is composed of many decision trees that map transaction data to MCCs. This mapping is created by splitting the sample up into nodes depending on the features of the data. For example, for transactions that have the keyword "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.1 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 shows 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

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

$\leq$ 19.935 are more likely to be "Restaurant" 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.1: Decision tree example



We 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

$$Q_{left}(\theta) = (X, y)|x_j \leq t_m \tag{48}$$

$$Q_{right}(\theta) = (X, y)|x_j > t_m \tag{49}$$

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

$$H(X_m) = \sum_k p_{mk}(1 - p_{mk}) \tag{50}$$

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[33] and maximized when the transactions are evenly split between the two categories.[34]

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

$$\theta^* = argmin_\theta \frac{n_{left}}{N_m} H(Q_{left}(\theta)) + \frac{n_{right}}{N_m} H(Q_{right}(\theta)) \tag{51}$$

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.

Table H.1 shows our goodness of fit measures when we train the model on 70% of the

[33]because 0*1 + 1*0 = 0.
[34]because 0.5*0.5 + 0.5*0.5 = 0.5.

data and use the remaining 30% as the testing data set. We calculate the measures at the category level as well as our aggregated measure. The aggregate measure has higher precision and recall because a transaction is still coded as correct as long as it is identified as one of the four non-durable consumption categories. Accuracy can only be calculated for the aggregate measure.

Table H.1: Goodness of fit measures

	Precision	Recall	Accuracy	Share
Restaurants	0.92	0.94	-	0.51
Gas and Grocery	0.92	0.94	-	0.38
Entertainment	0.90	0.78	-	0.06
Misc. Services	0.92	0.77	-	0.05
Aggregate	0.96	0.96	0.95	1.00

Notes: Precision measures the fraction of predicted consumption transactions that are correctly predicted. Recall measures the fraction of actual consumption transactions that are correctly predicted. Accuracy calculates the fraction of total observations that are correctly predicted. The last column shows the share of transactions in each category.

I Appendix – Alternate parameter values

Table I.1: Average tax refund under different parameter values ($\theta = 4$)

		β				
		0.975	0.980	0.985	0.990	0.995
	0.30	1,267	1,292	1,318	1,343	1,366
	0.40	1,935	1,973	2,043	2,128	2,203
	0.50	2,722	2,761	2,811	2,884	2,947
γ	0.57	3,196	3,233	3,293	3,373	3,451
	0.60	3,490	3,531	3,591	3,676	3,764
	0.70	4,223	4,268	4,330	4,420	4,530
	0.80	5,118	5,154	5,194	5,312	5,437

Notes: This table calculates the average tax refund for 100,000 simulated observations under different parameter values.