On the Welfare Effects of Phasing Out Paper Currency*

Julio Garín[†] — Claremont McKenna College

William D. Lastrapes[‡] — University of Georgia

Robert Lester[§] — Colby College

This Version: April 3, 2021

Abstract

We quantify the welfare effects of cash suppression policies within a general equilibrium model where cash reduces transactions costs and aids tax evasion in underground markets. In the model, currency suppression increases transactions costs and raises effective tax rates, but shifts resources out of costly underground markets and relaxes the government budget. When coupled with a reduction in distortionary taxes on consumption or factor inputs to ensure budget neutrality, full or partial cash suppression increases welfare in our baseline representative agent model. In a model with individual heterogeneity in cash use, suppression increases welfare, though by less for cash-intensive users.

JEL Classification: E42; E51; E52; E26.

Keywords: Currency; Cash; Currency Suppression; Monetary Policy; Demonetization.

*We are grateful to Eric Sims and Julie Smith for helpful comments and seminar participants at the Banco Central del Uruguay, Bates College, Bowdoin College, Colgate University, dECON-Facultad de Ciencias Sociales, UDELAR, University of Kentucky, Xavier University and participants at the 2018 Jornadas Anuales de Economía del Banco Central del Uruguay and the 2019 Midwest Macroeconomics Conference. The usual disclaimer applies.

[†]E-mail address: jgarin@gmail.com.

[‡]E-mail address: last@uga.edu.

[§]E-mail address: rblester@colby.edu.

1 Introduction

This paper quantifies the welfare effects of eliminating or curtailing the use of paper currency – cash – in the US. Cash transactions are anonymous and difficult to trace, giving cash a comparative advantage over other means of payment for illicit or 'underground' economic activity. Currency suppression proponents like Rogoff (1998, 2016) and Sands (2016) argue that 'demonetizing' currency or removing large denomination banknotes from circulation can effectively prohibit behavior that is deemed undesirable by policymakers. However, even if such proposals make life more difficult for criminals, tax evaders, and terrorists, cash should be restricted only if these benefits exceed the costs – sometimes unintended – of eliminating a manifestly valuable means of payments to ordinary households and businesses. Properly accounting for net effects on welfare requires a general equilibrium framework to capture the potential channels through which cash influences behavior, economic activity and well-being; yet, quantitative work in this area is scarce. Our aim in this paper is to begin to fill this gap in the literature on the overall welfare costs and benefits of cash.

The general equilibrium model we construct in this paper focuses on the tradeoff between cash as a cost-effective means of payment for 'above-ground' or legitimate transactions, and as a medium for facilitating one particular dimension of the underground economy, tax evasion. In the model, cash economizes on the fixed costs of making payments with credit-money alternatives like bank deposits, which generates a well-defined demand for cash by households to be used in small-value trades (as observed in the data). At the same time, income from anonymous cash transactions need not be reported to the tax authorities, so holding cash promotes consumption and income tax evasion in underground markets, thereby lowering effective taxes rates on labor and capital below statutory rates and increasing overall production as resources are shifted underground. Thus, using cash generates fundamental economic benefits to agents while reducing the inefficiencies of distortionary taxes, but can have adverse effects on the government's budget and fiscal policy. Section 2 of the paper takes a quick look at the data on cash usage and tax evasion to motivate the model and to aid in calibrating parameters for quantitative experiments.

Section 3 develops the representative agent version of the model, in which we ignore individual heterogeneity in cash usage. For the steady-state equilibrium of this representative agent model we conduct quantitative policy experiments in which cash is no longer accessible to households, and compute the effects on the economy and household welfare as measured by compensating variation in consumption. The results of these simulations are included in section 4. We consider the extreme case of complete cash suppression, as well as partial suppression that can be interpreted as demonetizing large bills while leaving smaller bills to circulate. In our model, the gains from currency suppression come mainly from expanding the government's budget constraint and shifting resources out of the underground economy, the latter of which we assume entails social costs. Because quantifying the benefits of government spending through the relaxed budget constraint is difficult, we consider the welfare implications of cash demonetization in combination with reductions in statutory tax rates on labor, capital and consumption. Our counterfactual simulations show that eliminating currency can enhance household welfare by non-trivial amounts, but only when such compensatory tax rate reductions accompany currency bans to ensure government budget neutrality.

The data and experience suggest that some agents will be harmed more than others if cash is restricted; however, our baseline representative agent model cannot account for such distributional effects of currency suppression. We provide evidence in section 2 that use of cash indeed varies across households. Section 4.3 of the paper extends the baseline model to one in which households vary in the costs they face to make non-cash transactions. This assumption implies that households in the model have varying degrees of cash usage for transactions as in the data. We conduct further cash suppression experiments in the context of this heterogenous agent model to determine the distribution of welfare effects. We again find that cash suppression can have positive effects on welfare across the distribution of cash users, but the welfare gains are significantly lower for high-intensity cash users.

We focus here on tax evasion as the primary source for the social costs of cash. Others, such as Rogoff (2016) and Sands (2016), have pointed to additional illicit uses of cash that we ignore – drug trafficking, prostitution, money laundering, corruption, terrorism, etc. Yet, as argued by Rogoff (2016, p. 59), tax evasion in the US is "truly massive" and is probably the most important underground use of cash in that country. And modeling the demand for cash for illegal markets is difficult because the tradeoffs are less obvious than they are for tax evasion and the government budget.[1] However, in our model we do account for the social costs of cash beyond the fiscal implications of tax evasion by assuming that labor supplied in underground markets generates higher disutility to households than above-ground labor. While this approach lacks the richness of a more detailed model, it allows cash suppression to create welfare gains by reducing the share of output produced and sold underground.[2]

Most of the recent scholarly work on currency focuses on the resilience of the demand for cash in the face of technological advances that reduce the cost of cash alternatives, and only

[1] Camera (2001) builds a model with both legal and illegal activities and shows that the expansion of currency may expand the scale of illegal activities.

[2] On the other hand, we ignore the second general rationale for eliminating currency emphasized by Rogoff (2016) – improving monetary policy in the face of effective lower bounds on policy rates due to zero nominal interest paid on cash.

a very few studies examine the welfare effects of proposals to eliminate cash. As discussed below, our model relies on the preference structure of Freeman and Kydland (2000) and transaction size to motivate a demand for cash relative to bank money, but other studies consider different mechanisms. Alvarez and Lippi (2017) model the demand for credit money conditional on cash holdings in a dynamic cash-management model. Jiang and Shao (2019) allow for substitutability between cash and credit to vary across sectors and households, while Chen et al. (2019) consider optimal cash holdings given different note denominations and the costs of receiving (heavy) coins as change. Kalckreuth et al. (2014) claim that many households use cash as a record-keeping, budgetary device. Each of these theories has some empirical support.

Gordon (1990) provides an initial effort to model the role of currency for tax evasion, but only in a partial equilibrium model of the demand for and supply of cash sales. Rogoff (1998, Appendix A) also develops a theory of currency demand and tax evasion in which effective tax rates depend on real currency balances, which is similar to our approach; however, his model too does not consider general equilibrium linkages, and neither Rogoff nor Gordon report quantitative experiments.[3] Gomis-Porqueras et al. (2014) model tax evasion and cash demand endogenously in a search-theoretic framework, ignoring the use of cash for illegal or illicit goods. Their focus is on cash and the shadow economy, and so do not consider currency suppression or welfare effects.[4]

We know of only two studies that quantify currency suppression policies on welfare. Alvarez and Lippi (2017, p. 109), noted above, force households to move from their optimal cash-holding policy to no cash – which is identical to our full-suppression policy experiment – to predict the effect on household utility. However, their simulation is partial equilibrium, only the costs to households are considered, and tax evasion plays no role. For this reason, their finding of small welfare effects of currency suppression for the US is interesting but not likely to serve as a solid basis for actual policy. Hendrickson and Park (2018) construct a dual-currency model of coins and paper money in which agents engaging in legal activity are indifferent to the medium of exchange, but illegal-goods traders prefer paper currency because it reduces transactions costs (paper is 'quieter' in exchange compared to coins). They show that eliminating paper currency can be welfare improving if the negative externality from crime is sufficiently large. Our paper is the only extant work that examines the general

[3] Rogoff (1998, Appendix B) extends his model to include multiple currency denominations which is appropriate for considering current proposals to eliminate only large bills, like \$100's and \$50's. We do not specifically model multiple denominations in this paper, but approximate large-bill elimination proposals by capping the extent of currency suppression.

[4] The theoretical literature on tax evasion without a focus on cash is extensive. The classic study is Allingham and Sandmo (1972); see Orsi et al. (2014) and Lopez (2017) for more recent efforts.

equilibrium welfare effects of currency suppression proposals in terms of their implications for fiscal policy.

Our motive for studying this issue does not derive solely from proposed policies in the US. Demonetization and currency suppression as a means for combatting the ills of society have been around for a long time, and there are many recent examples. The 'Great Indian Demonetization' of 2016 was motivated primarily to fight crime, corruption and tax evasion, and met with limited success (Lahiri, 2020; Chodorow-Reich et al., 2020). Most of the central banks in the euro area no longer issue 500 euro notes to prevent malevolent use of large bills.[5] Many European countries, such as Denmark, France, Italy, Belgium and Greece, have imposed upper limits on cash transactions (Slemrod, 2019), while Uruguay does not allow cash payments for taxes or salaries – they must be paid electronically. There have even been proposals to impose a Pigovian tax on cash holdings or withdrawals for the same ends (Benshalom, 2012). Yet, even countries like Sweden that have quickly moved away from cash and into electronic payments systems, and cities like New York, realize that eliminating currency may have significant costs and inequitable distributional effects, so are taking a cautious approach.[6] Clearly, more quantitative research is needed.

2 Motivating facts

This section documents some facts about cash usage in the US, as well as about the role of cash for tax evasion. We use these facts to motivate the theoretical model of Section 3, and to calibrate and quantify our policy experiments.

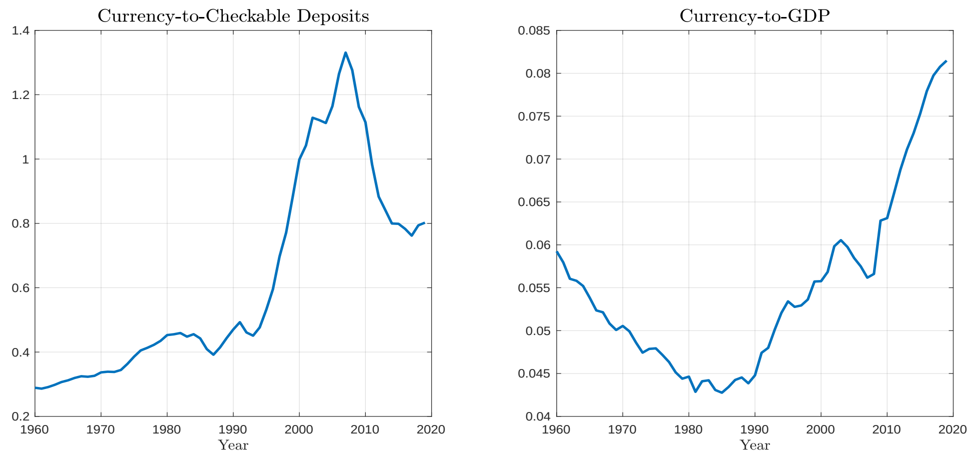
It has been well-documented that, despite technological advances for cash alternatives and a general trend toward electronic payments, paper currency remains a popular means of payment in the US. Figure 1 shows that from 1960 to the mid-1980's, currency-in-circulation relative to GDP (the inverse of the income velocity of cash) fell from around 6% to 4%, then reversed its trend, rising to 8.2% by the end of 2019, a higher percentage than in most other developed countries; see for example Jiang and Shao (2019, Figure 1) and Bech et al. (2018, Annex, p. 80). The figure also shows, up until the Covid-19 pandemic crisis, that the currency-to-checkable deposit ratio at 80%, although this is a decline from a peak of 135% in November 2007. From 2000 to 2019, this ratio averaged about one-to-one. Much of these cash dollars are held and circulate abroad – according to Judson (2017, Figure 6A), upwards of 60% of the total value of US cash in circulation.[7] Yet, even if only 40% of the value of

[5]See ECB Banknotes. Although no new notes of this denomination are to be issued, previously printed notes still circulate and remain legal tender.

[6]Bremner (2018), Shanahan and Mays (2020)

[7]Measuring currency held abroad is fraught with difficulties which can lead to wide ranges of estimates;

dollar bills is held domestically, the *per capita* stock of cash held in the US is around $2,000, a sizeable sum. Of this amount, $1,150 (57%) is estimated to be held in $100 bills. This latter fact is consistent with hoarding and our presumption that cash plays an important role in the underground economy.



Notes: Aggregate data from the St. Louis Federal Reserve Data Base.

Figure 1: Relative Importance of Currency

Our theoretical model relies on the assumption that the legitimate use of currency is primarily for small transactions. This assumption is borne out in the data. We use the 2017 Diary of Consumer Payment Choice (DCPC), an annual survey undertaken by the Federal Reserve Banks of Atlanta, Boston, Richmond, and San Francisco, to analyze the propensity to use cash conditional on transaction size. The respondents report basic demographic information as well as detailed transaction data over a three day period including the size and mode of payment for each transaction. The 2017 DCPC surveyed 2,793 individuals who made 11,380 purchases over a three day period.[8] In this sample, the proportion of the total number of transactions using cash is 32%, the average transaction size for a cash payment is $25, and the median transaction size for cash is $9.75. Figure 2 shows that people use cash on more than half of all transactions under $10. The probability of cash use declines with the size of the transaction size.

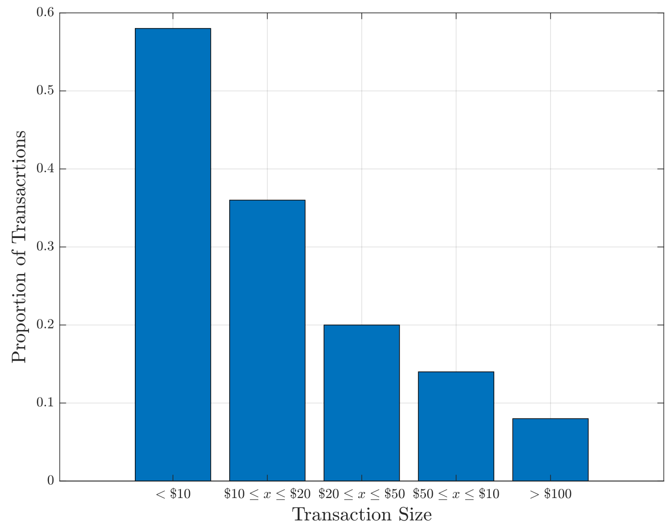
Results from the 2018 survey corroborate this finding. Greene and Stavins (2019) and Kumar and O'Brien (2019) show that the average US household made cash payments in 26% of its transactions, but only in 6% of payments by dollar value.[9] The average transaction size

for example, whereas Judson (2017) puts the number at around 50% in 2010, Feige (2012, Figure 4) provides an estimate of 30%.

[8]The DCPC also asks about transactions in which the respondent earned income or transferred income between financial accounts or into a different form, e.g. from a deposit to cash.

[9]On average, households made 43.3 payments during the month, 11.2 of which were in cash. The dollar

of a payment made in cash was $21.16 for the household sample, compared to $92.40 for all other payment methods (checks, debit cards, credit cards, electronic, etc.) 25% of the total number of payments were for transactions valued at less than $10; of these small transactions, 49% were for cash. Another 33% of payments between $10 and $25 were made in cash.



Notes: The graph shows the proportion of transactions of a given size made in cash. Data is from the 2017 DCPC. $N = 11,380$.

Figure 2: Cash Use by Amount

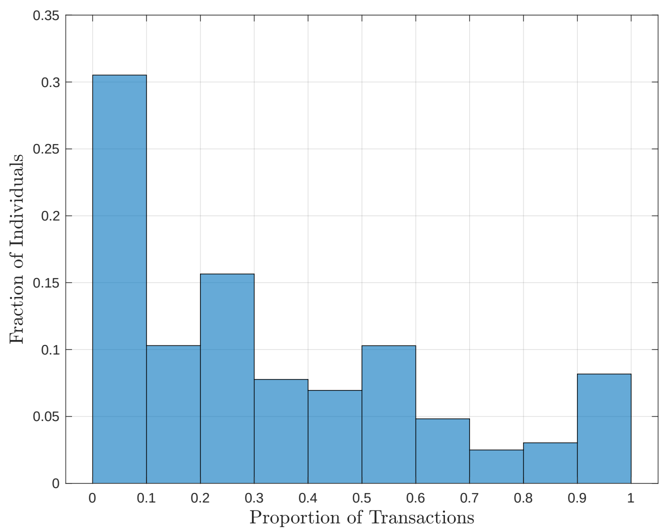
These findings are consistent with other research in this area. Using all two billion transactions for a national discount retailer over a three-year period (2010-2012), Wang and Wolman (2016) find that the proportion of total payments made by cash trends downward during this period (mirroring a rise in debit card payments), but ranges from 75% to 80% of transactions. In March 2013, 80% of transactions smaller than $10 were in cash, which exceeds the DCPC data (albeit for a different time period), and cash usage declined monotonically with transaction size. Aggregating by zip code, they also find significant variation in how cash is used across physical locations, and that the variance of cross-location payment mix increases with transaction size. The latter they argue is consistent with geographical variation in access to banking and cash alternatives. Wang and Wolman (2016, p. 98) claim that their findings on the importance of transaction size are consistent with threshold models of cash demand, as in Freeman and Kydland (2000).[10]

These results motivate us to examine what the data have to say about the *distribution* of cash usage across individuals. Figure 3 shows the frequency distribution across individual

value of all payments was $4,000, of cash payments $237. 2018 was the first year cash was not the most-used payment method: debit cards accounted for 28% of transactions in that year.

[10]See Briglevics and Schuh (2014) for estimates of currency demand using Survey of Consumer Payments choice, which show that cash withdrawal type (ATM, bank teller, etc.) affects transactions costs and cash demand elasticities.

households with respect to the proportion of the number of total transactions made with cash. To construct this distribution, we calculated the fraction of each individual's total transactions that use cash, then weight this fraction by total transactions. For instance, a person reporting ten transactions over the three day period gets twice as much weight as a person reporting five transactions. As shown in the figure, 10% of the sample do not use cash while another 20% use cash in less than 10% of their transactions. On the other end of the distribution, 8% of households use cash in more than 90% of their transactions. The median percentage of transactions using cash is around 25%.



Notes: The horizontal axis is the fraction of transactions made in cash. The vertical axis is the density of people in our sample. Each individual is weighted by number of transactions. Data is from the 2017 DCPC and $N = 2,793$.

Figure 3: Cash Use by Amount

We also looked at the role demographics play in explaining cash variation in the 2017 sample. Table 1 shows that cash users tend to have lower income, are in the extremes of the age distribution, and have less than a college education.[11] However, using each transaction in the sample as the observation unit, we regressed a binary variable equal to one for whether the transaction was for cash and zero otherwise on these demographic variables, and found essentially no explanatory power for these variables. These results are interesting because they inform us how (or how not) to model heterogeneity. In light of our results and Wang and Wolman (2016), we model heterogeneity as variation in the fixed costs of using non-cash transaction media.

The empirical literature on tax evasion is vast; see Slemrod (2019, 2007) and Alm (2012) for informative overviews. For our purposes, knowing the extent to which taxes are evaded in

[11]The results for age are consistent with Kumar and O'Brien (2019, Figure 7). They also show that older and high-income households tend to hold more cash.

the US helps us calibrate our model. The Internal Revenue Service periodically uses internal data to estimate the federal tax gap, the difference between what the IRS expects to receive in annual tax revenue and what it actually receives. According to the agency's most recent effort, which provides estimates for the years 2011-2013 (Johnson and Rose, 2019, p. 11, Table 2), the average annual gross tax gap over the period 2011 to 2013 was $441 billion, or 2.7% of US GDP. The shares of this gap due to nonfiling and under-reporting of income – types of evasion most likely related to cash transactions – were 9% and 80% respectively.[12] Thus, IRS estimates imply that unpaid annual taxes likely facilitated by using cash amount to roughly 2.4% of GDP.

Table 1: Cash Use by Demographic Characteristics

		Proportion of Transactions
Income		
	$< \$25,000$	44.1 %
	\$25,000-\$49,999	38.1 %
	\$50,000-\$99,000	29.1 %
	$> \$100,000$	26.6 %
Age		
	18-30	31.4 %
	31-40	23.9 %
	41-50	30.9 %
	51-60	35.7 %
	61-70	34.6 %
	> 70	34.1 %
Education		
	High school grad	40.0 %
	Bachelor's degree	26.3 %
	post grad degree	25.4 %

Notes: Data is from the 2017 Diary of Consumer Payment Choice. The means are unweighted. However, weighting by individual level weights, "ind_weight", does not qualitatively change the results.

These figures square with independent research by Feige (2012) and Cebula and Feige (2012), who use currency stock ratios in the US to estimate tax gaps and underreported income. Their model allows for currency held abroad and financial innovations like sweep accounts that affect the currency-to-deposit ratio. Their best estimates put the 2010 federal income tax gap at between $400 and $550 billion, and unreported income in the range of $1.9 to $2.4 trillion – on the order of 13% to 16% of GDP (Feige, 2012, p. 259).

[12] The remaining 11% comes from underpayment of taxes from reported income, which is probably not closely related to cash transactions. The IRS reports a 'net' tax gap that accounts for late and enforced payments, but the 'gross' concept is relevant for our purposes.

In sum, these facts provide guidance for our structural model and quantitative simulations below. In particular, they suggest that above-ground uses of cash in the US depend on transaction size – the smaller the amount of the payment, the more likely household value the use of cash relative to other payment methods. We rely on this assumption in the theoretical model. We know that the use of cash varies across individuals, but focusing on variation in access to non-cash payment alternatives, rather than demographics, is likely to be most fruitful. We also know that federal income tax evasion in the US is non-trivial, with half a trillion dollars unpaid annually, and that cash is sure to play an important role in determining that magnitude.[13]

3 Representative Agent Economy

Our baseline model is of a representative agent economy in which circulating fiat currency (cash) and bank deposits are imperfect substitutes as media of exchange. Unlike deposits, cash pays no nominal interest, but allows households to avoid the fixed cost (per transaction) of deposits; thus, households have a well-defined demand for each of these payment methods based on transaction size. Households also must pay a consumption tax to the government, a proportion of which they choose to evade. We assume that cash, because of its anonymity, facilitates consumption-tax evasion through a tax-evasion 'productivity' function, which generates an additional source of household demand for cash relative to deposits.

Profit-maximizing firms produce goods sold in 'above-ground' markets, for which they use use and pay proportional taxes on labor and capital, and goods sold in the 'underground' economy, for which they can evade these taxes. Participation in the underground economy requires firms to use cash to pay for its factors of production. With some exogenous probability firms are audited by the tax authorities and pay a penalty for underground tax avoidance. These conditions imply that firms also have a well-defined demand for cash that balances the marginal benefits of avoiding taxes with the marginal costs of the tax penalty and inflation. From the household perspective, goods sold above- and below-ground are perfect substitutes (and thus have identical prices in equilibrium), so the potential relative price effects across the two markets of policies toward tax evasion and the use of cash are reflected in input prices rather than produced-goods prices.

The banking sector issues deposits as its sole source of funds, makes loans to firms for investment in capital, and holds reserves issued by the central bank. The government sector comprises a fiscal authority (the treasury) and a central bank. The fiscal authority spends,

[13]IRS figures likely underestimate aggregate tax evasion since state income and local sales taxes are ignored.

collects the taxes noted above in addition to a lump sum tax on households, and borrows from the central bank to finance budget deficits. The central bank issues base money in the form of currency (central bank notes) and reserves, which finances the bank's purchase of treasury debt. Within this representative agent framework we examine the steady-state welfare effects of the central bank's decision to require households to redeem currency notes for private bank deposits.

3.1 Households

In our model, household demand for cash and deposits is based on the framework of Freeman and Kydland (2000), as generalized by Henriksen and Kydland (2010), which we have shown above to be consistent with US data. Household utility at time t depends additively on leisure and a continuum of consumer goods, c_{jt}, $0 \leq j \leq 1$, where within-period utility from consumption is given by $u\left[\min\left(\frac{c_{jt}}{(1-\omega)j^{-\omega}}\right)\right]$, which collapses to the original formulation in Freeman and Kydland (2000) for $\omega = -1$. The Leontief form of utility implies that consumption must be allocated such that $c_{jt} = (1-\omega)j^{-\omega}c_t$, where $c_t = \int_0^1 c_{jt}dj$ is total consumption (Henriksen and Kydland, 2010, p. 472). Goods indexed by larger values of j reflect larger transactions (like large-item durable goods) than those indexed by smaller j (like a cup of coffee); thus, j is an indicator of transaction size.

Households can purchase consumer goods using either cash, which incurs no transactions cost, or bank deposits, which incur a fixed cost per transaction, γ, that is independent of the size of the transaction; i.e., independent of j. We can think of γ as accounting for the fixed resource costs of payments services by banks (check-clearing costs, security costs, and so on) that households end up paying. Banks pay competitive returns on deposits, but cash pays no nominal interest. As shown by Freeman and Kydland (2000, p. 1127), these assumptions imply that the real deposit return to households, net of transactions costs, is increasing in transactions size, which in turn implies that households optimize by paying cash for transactions below a uniquely determined size threshold, j^*, and using checks or debit cards on their deposit accounts for transactions above j^*. They buy coffee with cash and cars with deposits.

Given these conditions on the allocation of consumption and the method of payment, the household's optimization problem is to maximize

$$U = \sum_{t=0}^{\infty} \beta^t \left[\ln(c_t) - \left(\frac{\xi_0}{1+\chi}\right)(\hat{l}_t + \check{l}_t)^{1+\chi} - \left(\frac{\xi_1}{1+\chi}\right)\check{l}_t^{1+\chi}\right] \tag{1}$$

subject to

$$\begin{aligned}
&(1+x_t\tau_c)c_t+\gamma(1-j_t^*)+\hat{a}_t+\check{a}_t+d_t+m_{ht}\leq \\
&\hat{w}_t\hat{l}_t+\check{w}_t\check{l}_t+\hat{R}_t\hat{a}_{t-1}+\check{R}_t\check{a}_{t-1}+\tilde{R}_t d_{t-1}+\frac{m_{h,t-1}}{1+\pi_t}+\Pi_t^F+\Pi_t^B-T_t
\end{aligned} \tag{2}$$

$$\int_0^{j_t^*}(1-\omega)j^{-\omega}c_t = {j_t^*}^{1-\omega}c_t \leq m_{ht} \tag{3}$$

$$\int_{j_t^*}^1(1-\omega)j^{-\omega}c_t = (1-{j_t^*}^{1-\omega})c_t \leq d_t, \tag{4}$$

where all variables are in real terms and each constraint holds for $t = 0, 1, \cdots, \infty$.[14]

Households get utility from total consumption, which aggregates not only across size but across above- and underground expenditures, and disutility from supplying hours of labor to the above-ground sector ($\hat{l}_t$) and underground sector ($\check{l}_t$). Disutility in general differs across labor sectors (if $\xi_1 = 0$ the two types of labor are perfect substitutes), but the Frisch labor supply elasticity χ is the same. When $\xi_1 > 0$, underground labor imposes greater disutility on households than formal labor markets, which we think of as a proxy for real resource costs of underground activity.

The left-hand-side of the household's lifetime budget constraint in (2) shows the uses of funds. The first term is total consumption, gross of the consumption tax, where τ_c is the statutory tax rate and x_t is the proportion of consumption expenditures reported to the tax authority. The household's effective consumption tax rate thus equals $x_t\tau_c$. We imagine that the decision to avoid consumption taxes takes place at the retail transaction level by the buyer/household, and that on the margin cash transactions make it easier for households to under-report sales *via* an exponential tax evasion function:

$$x_t = e^{-\phi m_{ht}}, \tag{5}$$

where $\phi > 0$ (which ensures that x_t lies between 0 and 1) and m_{ht} are real cash holdings of households. As real cash balances rise, households are able to reduce reported sales and thus their effective consumption tax rate. For a given level of cash holdings, an increase in ϕ increases the marginal 'productivity' of cash in reducing the effective tax rate. The form of the tax evasion function imposes diminishing returns to households on the use of cash for tax evasion purposes. In addition to consumption and consumption taxes as uses of funds, households accumulate capital used in the above-ground and underground markets, $\hat{a}_t + \check{a}_t$, real bank deposits d_t and real cash balances. Households also incur a fixed cost, $\gamma(1 - j_t^*)$, for all deposit transactions.

[14]Because we focus on the steady-state, there is no loss in generality of assuming perfect foresight.

Households obtain resources by supplying labor to both sectors of the economy, earning different real wages, $\hat{w}_t$ and $\breve{w}_t$, and earn different real gross returns on capital used in both sectors, $\hat{R}_t$ and $\breve{R}_t$. They earn real returns on deposits of $\tilde{R}_t d_{t-1}$ and suffer an inflation tax on cash holdings of $(1+\pi_t)^{-1}$, where $\pi_t = \frac{P_t - P_{t-1}}{P_{t-1}}$ and P_t is the price level used to deflate nominal magnitudes. Households earn profits from their ownership of firms and banks, and must pay lump-sum taxes, T_t to the government. Equation (3) is a cash-in-advance constraint while (4) is deposit-in-advance constraint, both of which are implied by the household's optimal payment decisions; note their dependence on the threshold variable, j_t^*.

Households choose total consumption, labor allocation across sectors, money holdings in the form of deposits and cash, capital across sectors, the transaction size threshold, j_t^*, and the extent to which consumption taxes are evaded, to maximize lifetime utility subject to the budget constraint, the money-in-advance constraints, and the tax evasion technology function (5). After substituting x_t out using (5), the first order conditions are:

$$c_t: \quad \frac{1}{c_t} - \mu_{1t} - \mu_{2t} {j_t^*}^{1-\omega} - (1 - {j_t^*}^{1-\omega})\mu_{3t} = 0 \tag{6}$$

$$\hat{l}_t: \quad -\xi_0(\hat{l}_t + \breve{l}_t)^\chi + \mu_{1t}\hat{w}_t = 0 \tag{7}$$

$$\breve{l}_t: \quad -\xi_0(\hat{l}_t + \breve{l}_t)^\chi - \xi_1 \breve{l}_t^\chi + \mu_{1t}\breve{w}_t = 0 \tag{8}$$

$$\hat{a}_t: \quad -\mu_{1t} + \beta\mu_{1,t+1}\hat{R}_{t+1} = 0 \tag{9}$$

$$\breve{a}_t: \quad -\mu_{1t} + \beta\mu_{1,t+1}\breve{R}_{t+1} = 0 \tag{10}$$

$$m_{ht}: \quad -\mu_{1t} + \mu_{2t} + \mu_{1t}\phi\tau_c c_t e^{-\phi m_{ht}} + \beta\frac{\mu_{1,t+1}}{1+\pi_{t+1}} = 0 \tag{11}$$

$$d_t: \quad -\mu_{1t} + \mu_{3t} + \beta\mu_{1,t+1}\tilde{R}_{t+1} = 0 \tag{12}$$

$$j_t^*: \quad \mu_{1t}\gamma - \mu_{2t}(1-\omega){j_t^*}^{-\omega} c_t + \mu_{3t}(1-\omega){j_t^*}^{-\omega} c_t = 0 \tag{13}$$

where μ_{1t}, μ_{2t}, and μ_{3t} are the multipliers on the respective constraints.

The first order conditions reveal the tradeoffs households face in choosing the extent of tax evasion and the allocation of their money holdings between cash and deposits. Assuming that the return on above-ground and underground assets are identical for the sake of simplicity, eliminating the multipliers from the last four of these conditions implies:

$$\gamma = \left[\frac{\tilde{R}_{t+1} - \frac{1}{1+\pi_{t+1}}}{R_{t+1}} - \tau_c c_t(\phi e^{-\phi m_{ht}})\right] c_{j^*t}. \tag{14}$$

An increase in the cash-deposit threshold (j^*) reduces the fixed transactions cost of deposits and allows the household to reduce consumption taxes according to its cash holdings and the tax evasion technology, but comes at the opportunity cost of the difference in real interest on deposits and cash relative to the return on real assets. The optimality conditions ensure that j^* is chosen so that the marginal benefits equal the marginal cost. Note that this condition implies that j^* depends not only on the size of transactions, as it does in Henriksen and Kydland (2010), but also on the incentives for reducing the tax burden through holding cash. Thus, in our model households determine the allocation of money between cash and deposits jointly factoring in the benefits of cash for small transactions and reduced sales tax rates.

3.2 Firms

Firms maximize profits by producing output for sale in the formal economy ($\hat{y}_t$) and in underground markets ($\check{y}_t$). Production technology is Cobb-Douglas and identical for each good: $\hat{y}_t = \hat{k}_t^{\alpha}\hat{l}_t^{1-\alpha}$ and $\check{y}_t = \check{k}_t^{\alpha}\check{l}_t^{1-\alpha}$, where k is physical capital. Because these goods are perfect substitutes in consumption, their relative price is one and total output is $y_t = \hat{y}_t + \check{y}_t$. Firms are responsible for paying taxes on labor and capital income but can choose to evade those taxes by participating in underground input markets; to do so they must pay in cash. However, with probability p firms are audited by the taxing authority and required to pay a proportional penalty, $s > 1$, on taxes not paid.

The objective function of firms is to maximize the present value of profits from above-ground and underground production less an inflation tax from holding cash (m_{ft}):

$$V = \sum_{t=0}^{\infty} \beta^t u'(c_t)\Pi_t^F \tag{15}$$

$$\Pi_t^F = \hat{\Pi}_t + \check{\Pi}_t - m_{ft} + \frac{m_{f,t-1}}{1+\pi_t} \tag{16}$$

$$\hat{\Pi}_t = [\hat{y}_t - (1+\tau_l)\hat{w}_t\hat{l}_t - (1+\tau_k)(\hat{R}_t - 1 + \delta)\hat{k}_t] \tag{17}$$

$$\check{\Pi}_t = [\check{y}_t - (1+ps\tau_l)\check{w}_t\check{l}_t - (1+ps\tau_k)(\check{R}_t - 1 + \delta)\check{k}_t] \tag{18}$$

subject to a cash-in-advance constraint motivated by tax evasion:

$$\check{w}_t\check{l}_t + (\check{R}_t - 1 + \delta)\check{k}_t \leq m_{ft}. \tag{19}$$

τ_k and τ_l are statutory flat-rate taxes on capital and labor. The first-order conditions from this problem are

$$\hat{k}_t: \quad \alpha\frac{\hat{y}_t}{\hat{k}_t} = (1+\tau_k)(\hat{R}_t - 1 + \delta) \tag{20}$$

$$\check{k}_t: \quad \alpha\frac{\check{y}_t}{\check{k}_t} = (1+ps\tau_k + \mu_{mt})(\check{R}_t - 1 + \delta) \tag{21}$$

$$\hat{l}_t: \quad (1-\alpha)\frac{\hat{y}_t}{\hat{l}_t} = (1+\tau_l)\hat{w}_t \tag{22}$$

$$\check{l}_{u,t}: \quad (1-\alpha)\frac{\check{y}_t}{\check{l}_t} = (1+ps\tau_l + \mu_{mt})\check{w}_t \tag{23}$$

$$m_{ft}: \quad 1 = \mu_{mt} + \beta\frac{u'(c_{t+1})}{u'(c_t)}\frac{1}{1+\pi_{t+1}} \tag{24}$$

where μ_{mt} is the multiplier on the firm's cash-in-advance constraint.

We model firms' tax evasion differently from households because the former use cash only for tax evasion purposes while households also care about the size of transactions at the retail level, and to specifically account for the underground economy. The effective tax rates on labor and capital not only depend on audits and penalties, but also on the share of input costs in the above-ground and underground markets. For example, the effective labor tax rate is

$$\tau_l^{eff} = \left(\frac{\hat{w}\hat{l}}{wl} + \frac{\check{w}\check{l}}{wl}ps\right)\tau_l \tag{25}$$

where wl is the firm's total wage bill. When $p \cdot s < 1$, firms have an incentive on the margin to hold cash and participate in underground markets to reduce effective input tax rates below statutory rates. In this case, firms will have a well-defined demand not only for cash but also for underground labor and capital; relative wages across above-ground and underground sectors will reflect tax advantages and household preferences for supplying labor. In the extreme case that $ps = 0$, firms pay no labor or capital taxes. If $ps = 1$, firms have no incentive for underground market participation, will not be willing to pay higher wages in the underground market to attract workers (assuming $\xi_1 > 0$), production will take place exclusively in aboveground markets, and effective tax rates will equal statutory rates.[15]

3.3 Banks, Government, and Equilibrium

The competitive banking sector's sole source of funds is checking deposits issued to households. The central bank imposes a reserve requirement on bank deposits, θ, that we assume binds. Banks also serve as intermediaries by using their deposits to finance new

[15]The extra disutility from household participation in underground labor markets ensures the model does not have a corner solution.

capital in the above-ground economy, but do not participate in the underground economy. Banks profits are then

$$\Pi_t^B = \hat{R}_t(1-\theta)d_{t-1} + \frac{\theta d_{t-1}}{1+\pi_t} - \tilde{R}_t d_{t-1}. \tag{26}$$

Banks choose deposit supply to maximize these returns, taking interest rates, prices, and θ as given.[16] Free entry implies that maximum profits are zero, which further implies that

$$\tilde{R}_t = \hat{R}_t(1-\theta) + \frac{\theta}{1+\pi_t}. \tag{27}$$

The rate of interest on bank deposits is a weighted average of the return on loans and interest earned on reserves (the latter of which we assume to be zero).

The government consists of a fiscal authority and a central bank. The central bank issues bank reserves and prints, without cost, currency notes and exchanges these notes for reserves perfectly elastically on demand and at par in terms of the economy's unit of account. Thus, the size of the monetary base – the sum of currency and reserves – is independent of household and firm demand for cash, which characterizes the actual economy. The central bank uses base money as a source of funds to purchase debt issued by the fiscal authority, and sets a constant growth rate, ψ, for the monetary base. These assumptions imply that total cash in nominal terms is currency held by households and firms, the nominal monetary base is the sum of total cash and reserves and the nominal money supply is the sum of cash and bank deposits:

$$M_t = M_{ht} + M_{ft} \tag{28}$$

$$MB_t = M_t + \theta D_t \tag{29}$$

$$M1_t = M_t + D_t. \tag{30}$$

Furthermore, the flow of (nominal) seignorage to the fiscal authority is $MB_t - MB_{t-1} = (1+\psi)MB_{t-1}$. The fiscal authority does not borrow directly from the public, so its budget constraint in real terms is

$$g_t = x_t\tau_c c_t + \tau_l(\hat{w}_t\hat{l}_t + ps\check{w}_t\check{l}_t) + \tau_k(R_t - 1 + \delta)(\hat{k}_t + ps\check{k}_t) + \psi\left(\frac{mb_{t-1}}{1+\pi_t}\right) + T_t, \tag{31}$$

where mb is the real monetary base and we have imposed the equilibrium condition that above-ground and underground capital are perfect substitutes to ease notation.[17]

[16] We ignore interest payments on bank reserves. In general, the central bank could pay gross nominal interest on reserves, Q^n, in which case bank profits would be $\Pi_t^B = R_t(1-\theta)d_{t-1} + \frac{Q^n}{1+\pi_t}\theta d_{t-1} - \tilde{R}_t d_{t-1}$.

[17] Since taxes are distortionary in the model, debt finance could matter for welfare effects. However, in our simulations real interest rate effects are small; therefore, there is no loss in generality in assuming the

Noting that the stock of capital evolves according to bank loans and direct investment by households, we have

$$k_t = \hat{k}_t + \check{k}_t \tag{32}$$

$$a_t = \hat{a}_t + \check{a}_t \tag{33}$$

$$k_{t+1} = a_t + (1-\theta)d_t \tag{34}$$

and that the aggregate resource constraint is

$$y_t = c_t + g_t + k_{t+1} - (1-\delta)k_t + \gamma(1 - j_t^*). \tag{35}$$

The model is in equilibrium when households maximize utility, firms and banks maximize profits, all resource constraints bind, and all markets for output, labor, capital, bank loans and deposits and cash clear.

4 Quantitative Analysis

4.1 Parameterization and the Role of Cash

We calibrate the model parameters to achieve a steady-state baseline that matches key features of the US economy. We then measure how much welfare changes from this baseline in the face of currency suppression experiments. Table 2 summarizes the calibrated parameter values. We assume the model period is one year and use standard parameterizations for the household discount factor ($\beta = 0.99$), Frisch labor elasticity ($\chi = 1.0$), and the production function parameters measuring the output elasticity of capital ($\alpha = 0.33$) and the rate of depreciation ($\delta = 0.08$). We set the base money growth rate ψ to be 2% which is consistent with steady-state inflation of the same rate, and the reserve ratio θ to be 10%. While the US reserve-to-deposit ratio at the time of writing is around 60%, we select a lower value to account for bank lending financed by non-deposit sources and check for sensitivity later. Average tax rates on consumption, capital and labor are from McDaniel (2007) after updating to 2015; we set the model tax rate parameters equal to their average values over the period 2000-2015: $\tau_c = 0.10$, $\tau_l = 0.203$, $\tau_k = 0.255$. We set $\omega = -1$ so that, given the other model parameters, the share of the dollar value of transactions is around 0.09, which is in line with the specification in Freeman and Kydland (2000). Lump sum taxes are exogenously set to zero, so government spending adjusts endogenously to ensure that the government's budget

government does not issue bonds directly to the public.

Table 2: Baseline parameter values

Description	Parameter	Value	Calibration/target value
Discount factor	β	0.99	standard
Frisch elasticity	χ	1.00	standard
Elasticity of capital	α	0.33	standard
Depreciation rate	δ	0.08	standard
Base growth rate	ψ	0.02	2% inflation
Reserve ratio	θ	0.10	standard
Capital tax rate	τ_k	0.255	McDaniel (2007)
Labor tax rate	τ_l	0.201	McDaniel (2007)
Consumption tax rate	τ_c	0.10	McDaniel (2007)
Consumption utility	ω	-1.00	cash value share=0.09 (j^{*2})
Consumption tax evasion	ϕ	0.20	sales tax gap=0.5%
Disutility of labor	ξ_0	8.55	labor hours = 0.33
Disutility of cash labor	ξ_1	11.68	cash transactions share=0.31 (j^*)
Fixed cost of deposits	γ	0.0039	cash-deposit ratio = 0.40
Effective input tax rate	$p \cdot s$	0.2371	Input tax gap=0.10

Values below single line are jointly determined to achieve equality between target and model values in last column.

constraint holds.

Calibrating ϕ, the parameter measuring the productivity of cash for sales tax evasion, is difficult because there is no reliable evidence on sales or consumption tax gaps. In our baseline model we set $\phi = 0.2$ which yields a relatively small consumption tax gap of 0.5% in the model's baseline steady-state. However, to examine sensitivity to this assumption we also consider larger measures of ϕ that increase the consumption tax gap.

The remaining parameters – ξ_0, ξ_1, $p{\cdot}s$ and γ – are determined jointly to match steady-state labor hours, the income 'tax gap', the currency-deposit ratio and the share of transactions made in cash. We set steady-state hours to be one-third of a year or 0.33. The income tax gap is the shortfall of labor and capital tax revenues through evasion relative to potential tax revenues; our value of 0.10 lies in the range of the estimates we report in section 2. The currency-deposit ratio is 0.40 based on a total currency-deposit ratio of 0.80 and assuming that 50% of currency is held domestically. The product $p \cdot s$ determines the effective tax rates faced by firms in the underground economy, which we set to 0.24.[18] These parameter values imply an equilibrium, steady-value of j^* to be 0.31, which we interpret as the share of *transactions* using cash; the cash share of transaction *value* (j^{*2}, see Equation (3)) in the

[18]Evidence from the IRS suggest that the probability of an audit is in the range of 1%. If we think of our parameter p to be 1%, the implicit value of the penalty s is in the range of 24, which can account for large lifetime costs of getting convicted of tax evasion.

model steady state is 9%. Relative to GDP, the fixed of costs of deposits is 0.6%.[19]

In the model, cash has value to households because it allows them to reduce transactions costs and to evade consumption taxes. Cash suppression will increase these transactions costs and effective tax rates. To firms, cash has value because it can shield payments from the government in the underground economy, thereby reducing effective tax rates. Cash suppression will raise distortionary income tax rates for firms, leading to decreased output and consumption, all else the same. If the underground economy entails no additional social costs, the sole benefit of cash suppression is a potential increase in tax revenues to the government and whatever welfare that generates. However, in the calibrated model the underground economy entails additional social costs, which implies that cash suppression can also increase welfare by shifting resources out of the informal sector into the formal, and can also increase overall output.

4.2 Welfare Effects of Currency Suppression

We now examine the implications of the model of a government ban on cash. In practice, the most straightforward way to implement such a policy is for the central bank to stop issuing new banknotes and to redeem outstanding cash for bank deposits at par, perhaps over a fixed time interval. This action directly substitutes currency for reserves one-for-one and, before any general equilibrium effects kick in, alters the composition, but not the level, of the nominal monetary base. In our most extreme policy experiment, we impose a complete ban on currency by restricting j^* to be zero and $ps = 1$; the presumption is that in the steady-state all currency held by households and firms is redeemed and disappears from circulation.[20] Under this hypothetical policy, households solely use deposits for transactions, while firms cannot satisfy their cash-in-advance constraint and thus do not participate in the underground economy.[21]

Thinking of a total ban on currency assumes away real-world frictions in implementing such a ban and alternative means for suppressing the use of cash. Given the benefits of cash, agents would have the incentive to only partially redeem cash for deposits, in which case unredeemed banknotes could circulate at a premium relative to deposits in an underground market for currency.[22] Most actual proposals in the US call for eliminating only large

[19] Freeman and Kydland (2000) set $\gamma = 0.006$ to obtain a currency-deposit ratio of 11% in their model. Their value of γ is 0.41% of GDP.

[20] For any given penalty rate s we imagine that eliminating cash raises the probability of audit to such a point that firms have no incentive based on effective tax rate differentials to evade taxes; i.e. $ps = 1$.

[21] The no-cash experiment in Alvarez and Lippi (2017, p.109) is essentially the same as this extreme case here.

[22] They could also circulate at a discount, as noted by Rogoff (2016, p.95), depending on government enforcement of the demonetization. We also note that while our full suppression is extreme, according to the

bills while allowing small bills to be issued and to circulate. And governments could apply less extreme policies to promote deposits and a 'banked' public by taxing cash holdings or withdrawals, or subsidizing deposits. Instead of modeling each of these alternatives in detail, which would significantly complicate and obfuscate the results, in addition to a complete ban we examine a few partial suppression experiments in which the household's cash threshold is constrained to be below its optimal level but not zero, and ps is increased to some value less than one.

We measure the effect of cash suppression on welfare as the compensating variation to consumption in the steady state – the percentage change in household consumption required to equate utility in the (constrained) cash-suppressed economy with the (unconstrained) cash economy. Formally, compensating variation is the value of λ such that

$$\sum_{t=0}^{\infty} \beta^t \left\{ \ln \left[c_t' \left(1 + \frac{\lambda}{100} \right) \right] - LS_t' \right\} = \sum_{t=0}^{\infty} \beta^t \left(\ln c_t - LS_t \right) \tag{36}$$

$$LS_t = \left(\frac{\xi_0}{1+\chi} \right) (\hat{l}_t + \check{l}_t)^{1+\chi} + \left(\frac{\xi_1}{1+\chi} \right) (\check{l}_t)^{1+\chi} \tag{37}$$

$$LS_t' = \left(\frac{\xi_0}{1+\chi} \right) (\hat{l}_t' + \check{l}_t')^{1+\chi} + \left(\frac{\xi_1}{1+\chi} \right) (\check{l}_t')^{1+\chi} \tag{38}$$

where c_t' and l_t' denote optimal steady-state values under cash suppression policies. A positive value for λ implies that cash suppression reduces household welfare, and vice versa.

Table 3: Tax rate distortions

	λ	$\frac{g'}{g}$
$\psi' = 0$	-2.12	0.95
$\tau_c' = 0$	-8.99	0.75
$\tau_l' = 0$	-14.98	0.61
$\tau_k' = 0$	-10.81	0.83

Notes: λ is compensating variation for eliminating each tax; g'/g is the ratio of government spending with elimination relative to spending without.

Before simulating the model without cash, we use this measure of welfare to quantify the tax distortions in the model. In Table 3 λ is computed for setting each tax rate (including inflation) to zero, holding the other tax rates fixed at their calibrated levels; the table also reports the ratio of government spending when each tax rate is eliminated. Households are willing to pay 15% of their steady-state consumption to have the labor income tax eliminated, compared to over 10% for capital taxes and 9% for consumption taxes. The smallest distortion

Reserve Bank of India 99% of the demonetized cash was ultimately redeemed in the 2016 demonetization.

is for the inflation tax. The large labor tax distortion is primarily due to a wealth effect: completely eliminating either labor or capital taxes increases output in equilibrium, but household wealth rises more for the labor tax since government spending falls by 39% relative to 17% under a capital tax reduction.

Table 4 reports compensating variation for four cash suppression policies: full suppression, partial suppression where j^* is set to zero but ps is unconstrained, ps is set to 1 but j^* is unconstrained, and partial suppression that involves restrictions to both j^* and ps, which we consider to be the most relevant policy scenario. For each suppression experiment we first consider the case in which statutory tax rates remain fixed so that government spending adjusts endogenously to the relaxed budget constraint as tax evasion declines. However, since we rule out utility gains from government spending, we also adjust tax rates endogenously, one-by-one, to maintain budget neutrality, which generates potential welfare gains by reducing distortions.[23] The top row of the table shows baseline values for output, the underground share of output, government spending relative to output, the money-deposit ratio and the tax gap. In the initial steady-state the underground economy as a share of overall production is under 15%, which is a bit above common estimates for the US (around 9%) but very close to averages for advanced economies (Medina and Schneider, 2018, p. 54), (Schneider, 2007, p. 28).

Consider first the top panel on the effects of a total cash ban. The cash-to-deposit ratio goes to zero as does the tax gap – there is no tax evasion by either households or firms – and the underground economy disappears. When statutory tax rates remain constant, government spending rises both absolutely and relative to GDP, from 21.8% to 23.4%. GDP ultimately rises because the decline due to the increase in effective tax rates is offset by a shift to production in the above-ground markets; however, the increase is a small 0.7%. This shift above-ground reduces labor costs because wages are lower than underground. Overall, welfare declines – households must receive nearly a 2% increase in consumption to compensate for the losses under suppression. The main source of lost welfare is declining consumption as increased government spending, which has no value by assumption, crowds out private expenditures. Consumption also falls because setting j^* to zero has a negative wealth effect on spending as the cost of transactions rises.

[23] In principle, seigniorage could also adjust to maintain budget neutrality, but it is unlikely in the US that the growth rate of money would be set for fiscal policy purposes.

Table 4: Compensating variation of cash suppression

	τ_i'	λ	y	$\frac{\check{y}}{y}$	$\frac{g}{y}$	$\frac{M}{D}$	Tax Gap
Baseline: $j^* = 31\%$, $ps = 24.71\%$							
	base	0.00	1.00	14.80	21.79	39.99	10.00
Complete Suppression: $j^* = 0$, $ps = 1$							
τ_i	base	1.97	1.007	0.00	23.36	0.00	0.00
τ_c'	6.5	−1.28	1.007	0.00	21.64	0.00	0.00
τ_l'	16.3	−1.31	1.008	0.00	21.63	0.00	0.00
τ_k'	13.9	−3.44	1.054	0.00	20.67	0.00	0.00
Partial Suppression: $\bar{j}^* = 0$							
τ_i	base	0.26	1.002	14.80	21.72	25.73	9.91
τ_c'	10.1	0.33	1.002	14.80	21.75	25.75	9.89
τ_l'	20.4	0.34	1.002	14.86	21.75	25.86	9.96
τ_k'	26.0	0.44	1.000	14.94	21.79	25.96	10.02
Partial Suppression: $ps = 1$							
τ_i	base	1.69	1.005	0.00	23.44	11.68	0.14
τ_c'	6.4	−1.55	1.005	0.00	21.68	8.85	0.08
τ_l'	16.2	−1.62	1.006	0.00	21.67	11.08	0.16
τ_k'	13.6	−3.77	1.053	0.00	20.69	10.77	0.16
Partial Suppression: $ps = 0.48$, $\bar{j}^* = 0.16$							
τ_i	base	1.05	1.004	9.32	22.69	18.73	4.47
τ_c'	8.0	−0.79	1.004	9.32	21.71	18.44	4.66
τ_l'	17.7	−0.95	1.005	8.37	21.68	16.91	3.92
τ_k'	17.3	−2.50	1.036	7.74	21.03	16.12	3.59

Notes: All values in percent. Baseline tax rates are $\tau_c = 10\%$, $\tau_l = 20.3\%$, and $\tau_k = 25.5\%$. τ_i' is the adjusted tax rate of type i, holding constant the other tax rates, that keeps government spending the same across baseline and currency suppression experiments.

However, when tax rates adjust to ensure budget neutrality, welfare unambiguously *increases* when currency is eliminated. Allowing tax rates to decline as an offset to cash suppression eliminates the crowding out effect and pushes up consumption, leading to utility gains. The biggest welfare gain comes from reducing the capital tax rate from 25.5% to 13.9%; households must give up 3.44% of their consumption in the new steady-state to remain indifferent, and output increases by 5.4%. The gain in welfare of 3.44% is more than a third of the potential gains from completely eliminating the capital tax without budget neutrality (Table 3). Government spending as a share of output falls in this case to 20.67%.

The second panel indicates that in our model restricting households from using cash has only small negative welfare effects (regardless of tax rate adjustment), since there is essentially no change in GDP, the underground share of GDP, government spending or the tax gap. The welfare loss comes from the increase in resource costs incurred by being forced to use deposits rather than cash for small transactions. According to the third panel, on the other hand, cash restrictions on firms essentially drive the results of the complete suppression. As in that case, welfare rises with tax rate adjustments to eliminate the crowding out effect.

The most relevant simulation is in the bottom panel of the table, where we cap cash usage for both households and firms, setting $ps = 0.48$ and $j^* = 0.16$. We interpret this policy change as tantamount to eliminating large-denomination notes (fifties and hundreds) while allowing small bills to circulate, as proposed by Rogoff (2016, p. 95) and others. In our model's original steady-state the currency-deposit ratio of 40% assumes a total currency-deposit ratio equal to 0.8 but with 50% of currency in circulation held in the US. Judson (2012, pp. 25-26) estimates that \$340 billion large-denomination notes were held domestically in the US in 2011 out of a total value of currency held by the public of \$965. If we interpret our partial policy experiment as eliminating only these large bills, then currency falls by \$340 while bank reserves rise by the same amount. If the reserve ratio ranges between 10% and 15%, we obtain currency-deposit ratios in the range of 14% to 19%, which is consistent with the steady-state M/D ratio in the penultimate column of the last panel.

As with complete suppression and suppression of firms' cash, without tax rate adjustments the policy harms welfare even though output rises (slightly) and the share of output in underground economy falls from 14.8% to 9.32%. The main culprit again is crowding out by government spending and the resource costs of not using cash. As before, eliminating large bills is welfare improving when distortionary tax rates are lowered. The gains are small when consumption and labor taxes are adjusted, but there is a more noticeable advantage to altering labor taxes than consumption taxes compared to full suppression. Gains are significant for capital taxes, however; households are willing to give up 2.5% of consumption to eliminate one-hundred and fifty dollar bills. These gains come from a 3.6% increase in output and a decrease by half in the underground economy's share. To maintain budget neutrality, capital tax rates decline from 25.5% to 17.3%. These gains are one-quarter of those from totally eliminating capital taxes.

As the table indicates, the biggest welfare and output gains occur when capital taxes are reduced to ensure budget neutrality in the face of cash suppression. Given the smaller capital share in production and the relative elasticities of capital demand and supply, the decline in capital taxes must be greater than that of labor taxes, which leads to a large increase in equilibrium capital input and large boost of GDP. It is therefore worth considering whether it is welfare-improving in our model to reduce capital taxes while raising labor taxes to maintain budget neutrality, but leaving cash usage unconstrained. To examine this question,

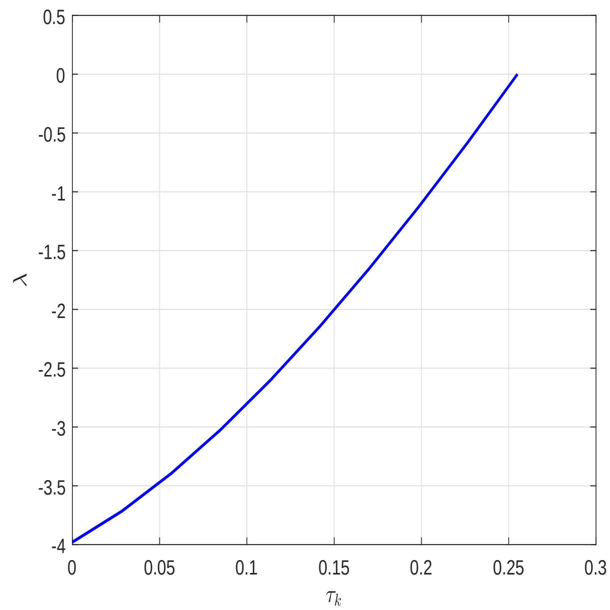


Figure 4: Compensating variation for reducing capital tax rates.

we lowered capital taxes from 25.5% (the calibrated value) down to zero at 5 percentage point intervals, in each case raising labor tax rates to maintain budget neutrality in the baseline case with no cash suppression. The result is shown in Figure 4. As the capital tax rate is lowered and labor taxes raised, compensating variation in consumption monotonically declines from 0 to −3.9, indicating that the capital taxes are more distortionary than labor income taxes. This welfare improvement as capital taxes are lowered is about a half percentage point greater than from the full cash suppression policy, and one and half percentage points greater than eliminating large bills.[24]

Finally, Figure 5 provides a check for the robustness of our welfare effects in the large-bill suppression case to variations in key model parameters. Each panel corresponds to a different parameter; within each panel we plot λ for each adjusted statutory tax rate individually. In all cases, the dominance of the revenue-neutral capital taxes is obvious regardless of parameter values. Doubling ϕ, which determines the productivity of cash for household tax evasion reduces welfare gains slightly, while increasing the reserve ratio θ increases those gains. Welfare gains tend to decline with increases in the cost of deposits (γ), the steady-state inflation rate (ψ) and the costs of the underground economy (ξ_1).[25]

The utility parameter ω affects the skewness of the transaction-size distribution: as ω

[24]These results are not inconsistent with Table 3 because here we are considering compensating tax rate changes to maintain budget neutrality.

[25]The robustness checks for the complete suppression case, not reported here, show similar sensitivity to changing parameter values.

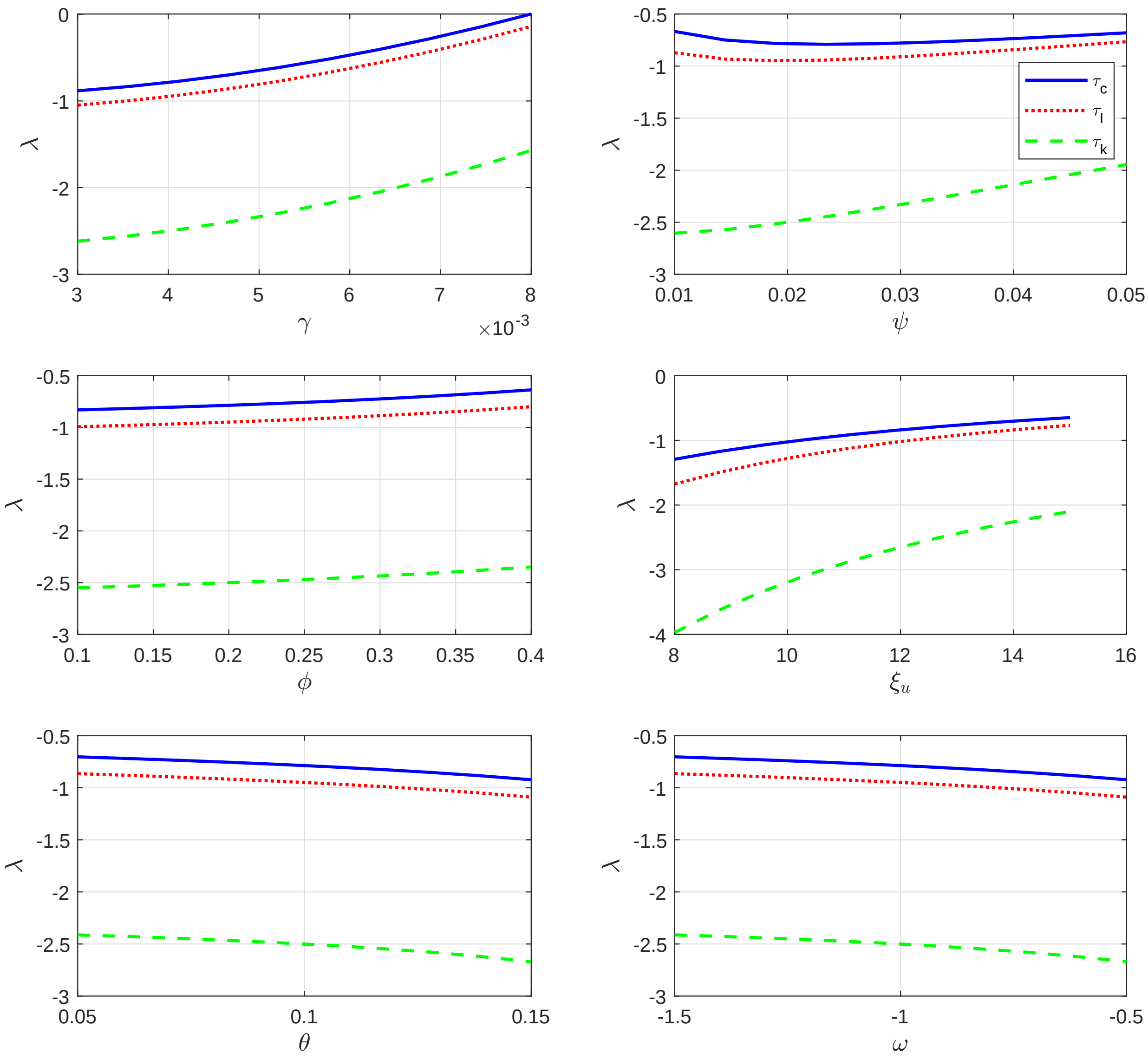


Figure 5: Robustness to parameter values in under partial (large bills) suppression

falls from the baseline value of -1 to -1.5 j^* will rise for any given cash expenditure share. In our unconstrained baseline where the cash's share of total spending is around 10%, the share of the number of transactions in cash rises from 30% to 40% as ω falls from -1 to -1.50. The last panel in Figure 5 shows that welfare gains are smaller for the skewed distribution – because of a larger loss of payment efficiency – but only slightly so.

4.3 Suppression with Heterogenous Agents

The data we presented in Section 2 document significant heterogeneity in cash use among individuals in the US economy. In this section, we generalize the representative agent model above to allow for heterogeneity in cash usage across households by assuming that they differ regarding access to payments technologies. We impose this variation by assuming exogenous differences across households in γ, the model parameter measuring the transactions costs of deposits.

In our heterogeneous agent model the household sector is still described by equations (1) through (5); however, we now allow γ_i to be different for each individual household $i = 1, \cdots, n$. The structure of the model for firms, banks and the government sector is identical to the representative agent model. For production, firms treat these different groups as perfect substitutes in terms of labor and capital, and consumption, labor and capital taxes are also the same. The steady-state outcome of the model collapses to that of the restricted representative agent case for $\gamma_i = \gamma$.[26]

To simulate the model we set $n = 5$ to consider five household groups. We calibrate γ using the DCPC survey data reported in Figure 3, which shows the distribution across individuals of j^*, the cash share of the number of transactions. We then choose γ_i so that j_i^* equals the i^{th} percentile of this distribution, where $i = 10, 30, 50, 70$ and 90, assuming all other parameter values are at the baseline values in the representative agent model. The top panel of Table 5 reports the calibrated values of γ_i. These values closely match our target for labor hours and input tax gap in Table 2; the cash-deposit ratio is somewhat higher than the target value of 40%, but remains in a reasonable range.

[26]For technical reasons, we add a term to households' budget constraints to impose an upward-sloping supply curve of capital. Given the structure of the model, this term is needed to pin down the distribution of capital in the steady-state. We assume this term is very small and equal across households, so there is no heterogeneity in wealth across individuals. When we simulate the heterogeneous agent model in the baseline case assuming equal γ's, the outcome is the same as in the representative agent case.

Table 5: Compensating variation of cash suppression for the heterogeneous agent model

%-tile		10^{th}	30^{th}	50^{th}	70^{th}	90^{th}
j_i^*		0.00	0.09	0.25	0.44	0.80
$\gamma_i(\times 10^{-2})$		0.00	0.12	0.32	0.56	1.01
	Tax rate	λ_{10}	λ_{30}	λ_{50}	λ_{70}	λ_{90}
τ_i	base	1.81	1.84	1.96	2.20	3.08
$\tau_c^{'}$	0.066	-1.38	-1.36	-1.24	-1.01	-0.15
$\tau_l^{'}$	0.164	-1.37	-1.36	-1.26	-1.05	-0.24
$\tau_k^{'}$	0.140	-3.45	-3.44	-3.36	-3.16	-2.39

Notes: Top panel – Targeted values for j_i^* based on cash-share-of-transactions percentiles from the DCPC survey data reported in Figure 3, and corresponding calibrated values for γ_i. Bottom panel – compensating variation (in percent) for each percentile for full cash suppression experiment. $\tau_k^{'}$ is the adjusted tax rate of type k, holding constant the other tax rates, that keeps government spending the same across baseline and currency suppression experiments.

The bottom panel of Table 5 reports the compensating variation in consumption for the full cash-suppression simulation for the heterogeneous agent model. As in the representative agent case, all households lose welfare when cash is eliminated without compensating adjustments to ensure government budget neutrality. These losses monotonically increase as cash usage rises across individuals. However, we find that even high-intensity cash users benefit from cash suppression when tax rates adjust. Compensating variation for low-cash users is the same as in the representative agent model, but high-cash households are willing to give up more than a full percentage point less in consumption for cash suppression than low-cash users. Because the welfare gains for low-cash users differs little from the representative agent model, average welfare gains in the heterogeneous agent model are lower than those for the former model.

5 Conclusion

In this paper we have attempted to obtain plausible quantitative projections of the effects of currency suppression policies on welfare. Our focus has been on the fiscal implications of such policies through our emphasis on the role of currency as a means for tax evasion. In both the baseline model of a representative agent and a more general model that allows for heterogeneity across households in cash usage, we find that currency suppression proposals – even those that do not propose complete elimination – can have positive welfare effects as long as distortionary tax rates are lowered to compensate for the increased effective tax rates from demonetization. For example, in the representative agent model households would be willing to give up almost 3.5% of consumption in return for complete elimination of cash. Part of this gain derives from shifting resources out of the costly underground economy. The welfare gains are smaller, however, for high cash-use households. Our findings provide stronger support for the welfare benefits of cash suppression than other studies, such as Alvarez and Lippi (2017).

There are at least three fruitful directions for future research on the welfare implications of eliminating currency. First, we have made comparisons across steady-states and have ignored welfare along the transition to a new steady-state. The framework we have developed here can be used to examine such transitory dynamics for specific proposals. Second, our model accounts for the costs of the underground economy, but only in a preliminary way. Since, for example cash is instrumental in criminal economic activity because of the relatively high costs of using bank money there (due to its lack of anonymity), specifically accounting for the economic impacts of crime, resources used to prevent crime, and the interaction with tax evasion in our model is likely to lead to better and more precise estimates of welfare effects. And third, our model assumes flat rate taxes that are imposed uniformly across individuals. Given the important role of statutory tax rates in the quantitative exercises, generalizing the model's tax structure can also help refine welfare estimates, especially regarding the distributional effects. For example, to the extent that marginal tax rates are correlated with the intensity of cash use across individuals (e.g., low income cash users paying low marginal tax rates), the difference in welfare effects across individuals could be altered. We leave these extensions for future work.

References

Allingham, M. G. and A. Sandmo (1972): "Income tax evasion: a theoretical analysis," *Journal of Public Economics*, 1, 323 – 338.

Alm, J. (2012): "Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies," *International Tax and Public Finance*, 19, 54–77.

Alvarez, F. and F. Lippi (2017): "Cash burns: An inventory model with a cash-credit choice," *Journal of Monetary Economics*, 90, 99 – 112.

Bech, M., U. Faruqui, F. Ougaard, and C. Picillo (2018): "Payments are a-changin but cash still rules," *BIS Quarterly Review*, 67–80.

Benshalom, I. (2012): "Taxing Cash," *Columbia Journal of Tax Law*, 4, 65–93.

Bremner, C. (2018): "Cashless society would hurt Sweden's elderly," *The Times*.

Briglevics, T. and S. Schuh (2014): "U.S. Consumer Demand for Cash in the Era of Low Interest Rates and Electronic Payments," Tech. Rep. 1660, ECB Working Paper Series.

Camera, G. (2001): "Dirty money," *Journal of Monetary Economics*, 47, 377–415.

Cebula, R. J. and E. L. Feige (2012): "Americas unreported economy: measuring the size, growth and determinants of income tax evasion in the US," *Crime, Law and Social Change*, 57, 265–285.

Chen, H., K. P. Huynh, and O. Shy (2019): "Cash versus card: Payment discontinuities and the burden of holding coins," *Journal of Banking and Finance*, 99, 192 – 201.

Chodorow-Reich, G., G. Gopinath, P. Mishra, and A. Narayanan (2020): "Cash and the Economy: Evidence from India's Demonetization," *The Quarterly Journal of Economics*.

Feige, E. L. (2012): "New Estimates of U.S. Currency Abroad, the Domestic Money Supply and the Unreported Economy," *Crime, Law and Social Change*, 57, 239–63.

Freeman, S. and F. E. Kydland (2000): "Monetary Aggregates and Output," *American Economic Review*, 90, 1125–1135.

Gomis-Porqueras, P., A. Peralta-Alva, and C. Waller (2014): "The shadow economy as an equilibrium outcome," *Journal of Economic Dynamics and Control*, 41, 1 – 19.

GORDON, J. P. F. (1990): "Evading Taxes by Selling for Cash," *Oxford Economic Papers*, 42, 244–255.

GREENE, C. AND J. STAVINS (2019): "The 2018 Diary of Consumer Payment Choice," Tech. rep., Federal Reserve Bank of Atlanta Research Data Reports.

HENDRICKSON, J. R. AND J. PARK (2018): "Breaking the Curse of Cash," .

HENRIKSEN, E. AND F. E. KYDLAND (2010): "Endogenous money, inflation, and welfare," *Review of Economic Dynamics*, 13, 470–486.

JIANG, J. H. AND E. SHAO (2019): "The cash paradox," *Review of Economic Dynamics*.

JOHNSON, B. W. AND P. J. ROSE (2019): Tech. Rep. Publication 1415, IRS Research, Applied Analytics & Statistics.

JUDSON, R. (2012): "Crisis and Calm: Demand for U.S. Currency at Home and Abroad from the Fall of the Berlin Wall to 2011," *International Finance Discussion Papers*.

——— (2017): "The Death of Cash? Not So Fast: Demand for U.S. Currency at Home and Abroad, 1990-2016," Tech. rep., International Cash Conference 2017 – War on Cash: Is there a Future for Cash?

KALCKREUTH, U. V., T. SCHMIDT, AND H. STIX (2014): "Using Cash to Monitor Liquidity: Implications for Payments, Currency Demand, and Withdrawal Behavior," *Journal of Money, Credit and Banking*, 46, 1753–1786.

KUMAR, R. AND S. O'BRIEN (2019): "2019 Findings from the Diary of Consumer Payment Choice," Tech. rep., Cash Product Office Federal Reserve System.

LAHIRI, A. (2020): "The Great Indian Demonetization," *Journal of Economic Perspectives*, 34, 55–74.

LOPEZ, J. J. (2017): "A quantitative theory of tax evasion," *Journal of Macroeconomics*, 53, 107 – 126.

MCDANIEL, C. (2007): "Average tax rates on consumption, investment, labor and capital," in *Arizona State University*.

MEDINA, L. AND F. SCHNEIDER (2018): "Shadow Economies Around the World: What Did We Learn Over the Last 20 Years?" .

Orsi, R., D. Raggi, and F. Turino (2014): "Size, trend, and policy implications of the underground economy," *Review of Economic Dynamics*, 17, 417–436.

Rogoff, K. (1998): "Blessing or Curse? Foreign and underground demand for euro notes," *Economic Policy*, 262 – 303.

Rogoff, K. S. (2016): *The Curse of Cash*, Princeton University Press.

Sands, P. (2016): "Making it Harder for the Bad Guys: The Case for Eliminating High Denomination Notes," *M-RCBG Associate Working Paper Series No. 52*.

Schneider, F. G. (2007): "Shadow Economies and Corruption All Over the World: New Estimates for 145 Countries," *EconStor*, 1–66.

Shanahan, E. and J. C. Mays (2020): "New York City Stores Must Accept Cash, Council Says," *The New York Times*.

Slemrod, J. (2007): "Cheating Ourselves: The Economics of Tax Evasion," *Journal of Economic Perspectives*, 21, 25–48.

——— (2019): "Tax Compliance and Enforcement," *Journal of Economic Literature*, 57, 904–54.

Wang, Z. and A. L. Wolman (2016): "Payment choice and currency use: Insights from two billion retail transactions," *Journal of Monetary Economics*, 84, 94 – 115.

6 Model With Heterogeneity

In this section we go over the heterogeneous agent model in detail and show how the model aggregates. The optimization problems facing banks and firms are exactly the same as in the representative agent model, so we do not go over them again.

The household optimization problem is to maximize

$$U = \sum_{t=0}^{\infty} \beta^t \left[\ln c_{i,t} - \xi_0 \frac{(l_{m,i,t} + l_{u,i,t})^{1+\chi}}{1+\chi} - \xi_1 \frac{l_{u,i,t}^{1+\chi}}{1+\chi} \right]$$

subject to the constraints

$$\begin{aligned} & c_{i,t}(1 + x_{i,t}\tau_c) + m_{h,i,t} + a_{i,t} + d_{i,t} + \gamma_i(1 - j^*_{i,t}) = \\ & w_{m,t} l_{i,m,t} + w_{u,t} l_{i,u,t} + R_t a_{i,t-1} - \frac{\epsilon}{2} a^2_{i,t-1} + \tilde{R}_t d_{i,t-1} + \frac{m_{h,i,t-1}}{1+\pi_t} + \frac{\Pi^f_t}{N} + \frac{\Pi^b_t}{N} - \frac{T_t}{N} \end{aligned}$$

$$\begin{aligned} x_{i,t} &= e^{-\phi m_{h,i,t}} \\ {j^*_{i,t}}^{1-\omega} c_{i,t} &\leq m_{h,i,t} \\ (1 - {j^*_{i,t}}^{1-\omega}) c_{i,t} &\leq d_{i,t}. \end{aligned}$$

A few differences relative to the representative agent model are worth noting. The most substantive is that we assume a quadratic portfolio management cost, $\frac{\epsilon}{2} a^2_{i,t-1}$, associated with holding a capital stock of $a_{i,t-1}$. Having this cost allows us to pin down the distribution of capital holdings. In equilibrium, the relative capital holdings between two agents i and j is equal to $\frac{\epsilon_j}{\epsilon_i}$. Since we are not concerned about the distribution of wealth, we give each agent the same ϵ of 0.0001. This value of ϵ implies that aggregate portfolio costs are slightly less than 0.2 percent of GDP. Second, we assume that lump sum taxes and profits are distributed equally across households. Since we set lump sum taxes equal to 0 and bank profits are 0 in equilibrium, the distribution of Π^f_t is the only variable where the distribution would matter. Third, we give each agent a different γ which leads households to choose different levels of cash holdings. This is the only way households are different. Finally, since interest rates on market capital and above ground capital are the same in equilibrium, we have just gone ahead and imposed that from the beginning.

The first-order conditions are

$$
\begin{aligned}
c_{i,t}:&\quad \frac{1}{c_{i,t}} - \lambda_{1,i,t}(1 + x_t\tau_c) - \lambda_{2,i,t} {j^*_{i,t}}^{1-\omega} - (1 - {j^*_{i,t}}^{1-\omega})\lambda_{3,i,t} = 0 \\
l_{m,i,t}:&\quad -\xi_0(l_{m,i,t} + l_{u,i,t})^\chi + \lambda_{1,t} w_{m,t} = 0 \\
l_{u,i,t}:&\quad -\xi_0(l_{m,t} + l_{u,t})^\chi - \xi_1 l^\chi_{u,i,t} + \lambda_{1,t} w_{u,t} = 0 \\
a_{i,t}:&\quad -\lambda_{1,i,t} + \beta\lambda_{1,i,t+1} R_{t+1} = 0 \\
m_{h,i,t}:&\quad -\lambda_{1,i,t} + \lambda_{2,i,t} + \lambda_{1,i,t}\phi\tau_c c_{i,t} e^{-\phi m_{h,i,t}} + \beta\frac{\lambda_{1,i,t+1}}{1+\pi_{t+1}} = 0 \\
d_{i,t}:&\quad -\lambda_{1,i,t} + \lambda_{3,i,t} + \beta\lambda_{1,i,t+1}\tilde{R}_{t+1} = 0 \\
j^*_{i,t}:&\quad \lambda_{1,i,t}\gamma - \lambda_{2,i,t}(1-\omega){j^*_{i,t}}^{-\omega} c_{i,t} + \lambda_{3,i,t}(1-\omega){j^*_{i,t}}^{-\omega} c_{i,t} = 0.
\end{aligned}
$$

Summing the budget constraints across i gives

$$
\begin{aligned}
&\sum_i c_{i,t}(1 + x_{i,t}\tau_c) + m_{h,t} + a_t + d_t + \sum_i \gamma_i(1 - j^*_{i,t}) = \\
&w_{m,t} l_{m,t} + w_{u,t} l_{u,t} + R_t a_{t-1} - \frac{\epsilon}{2}\sum_i a^2_{i,t-1} + \tilde{R}_t d_{t-1} + \frac{m_{h,t-1}}{1+\pi_t} + \Pi^f_t + \Pi^b_t - T_t.
\end{aligned}
$$

Bank profits are $\Pi^b_t = R_t(1-\theta)d_{t-1} + \theta\frac{d_{t-1}}{1+\pi_t} - \tilde{R}_t d_{t-1}$. Substituting this into the above equation gives

$$
\begin{aligned}
&\sum_i c_{i,t}(1 + x_{i,t}\tau_c) + m_{h,t} + a_t + d_t + \sum_i \gamma_i(1 - j^*_{i,t}) = \\
&w_{m,t} l_{m,t} + w_{u,t} l_{u,t} + R_t k_{t-1} - \frac{\epsilon}{2}\sum_i a^2_{i,t-1} + \theta\frac{d_{t-1}}{1+\pi_t} + \frac{m_{h,t-1}}{1+\pi_t} + \Pi^f_t - T_t
\end{aligned}
$$

where we have used the fact that $k_t = a_t + (1-\theta)d_t$. Firm profits are $\Pi^f_t = y_{m,t} + y_{u,t} - w_{m,t}(1+\tau_l)l_{m,t} - w_{u,t}(1 + ps * \tau_l)l_{u,t} - (1+\tau_k)k_{m,t}(R_t - 1 + \delta) - (1 + ps\tau_k)k_{u,t}(R_t - 1 + \delta) - m_{f,t} + \frac{m_{f,t-1}}{1+\pi_t}$. Substituting this into the last equation gives

$$
\begin{aligned}
&\sum_i c_{i,t}(1 + x_{i,t}\tau_c) + m_{h,t} + a_t + d_t + \sum_i \gamma_i(1 - j^*_{i,t}) = \\
&y_{m,t} + y_{u,t} - \tau_l w_{m,t} l_{m,t} - \tau_l ps w_{u,t} l_{u,t} \\
&- (R_t - 1 + \delta)(\tau_k k_{m,t} + ps\tau_k k_{u,t}) - m_{f,t} + \frac{m_{f,t-1}}{1+\pi_t} - \frac{\epsilon}{2}\sum_i a^2_{i,t-1} + \theta\frac{d_{t-1}}{1+\pi_t} + \frac{m_{h,t-1}}{1+\pi_t} - T_t.
\end{aligned}
$$

Using the fact that $mb_t = m_{h,t} + m_{f,t} + \theta d_t$, we can write the above equation as

$$\begin{aligned}
&\sum_i c_{i,t}(1 + x_{i,t}\tau_c) + mb_t + k_{t+1} + \sum_i \gamma_i(1 - j^*_{i,t}) = \\
&y_{m,t} + y_{u,t} - \tau_l w_{m,t} l_{m,t} - \tau_l psw_{u,t} l_{u,t} \\
&- (R_t - 1 + \delta)(\tau_k k_{m,t} + ps\tau_k k_{u,t}) + (1 - \delta)k_t + \frac{mb_{t-1}}{1 + \pi_t} - \frac{\epsilon}{2}\sum_i a^2_{i,t-1} - T_t.
\end{aligned}$$

Finally, using the fact that $T_t = g_t - \sum_i \tau_c x_{i,t} c_{i,t} - (R_t - 1 + \delta)(\tau_k k_{m,t} - ps\tau_k k_{u,t}) - \tau_l w_{m,t} l_{m,t} - \tau_l psw_{u,t} l_{u,t} - mb_t + \frac{mb_{t-1}}{1+\pi_t}$, we end up with the resource constraint

$$\sum_i c_{i,t}(1 + x_{i,t}\tau_c) + k_{t+1} + \sum_i \gamma_i(1 - j^*_{i,t}) + \frac{\epsilon}{2}\sum_i a^2_{i,t-1} + g_t = y_{m,t} + y_{u,t} + (1 - \delta)k_t.$$

The representative agent model is just a special case of this.