Digitization and Demonetization in a Shadow Economy Model*

Ayushi Bajaj
Monash University

Nikhil Damodaran
Louisiana State University

February 15, 2018

Abstract

Payments made in fiat money, as opposed to other digital means are anonymous allowing the emergence of a shadow economy to avoid tax payment. In this paper, we build a model where size of the shadow sector is an equilibrium outcome and depends on the trade-offs between two means of payment. We first determine the revenue maximizing size of the shadow economy, and find that a partial shadow equilibrium may maximize tax revenue, but not output. We then analyze policies to reduce its size by comparing demonetizing legal tender, facilitating digitization of the means of payment or simply reducing tax rates. We calibrate the model to US and India, and find that subsidizing the transition to digital currency in India can reduce the size of the shadow economy and improve private welfare. Demonetizing legal tender comes at a short run cost and can potentially improve welfare but only in the presence of multiple equilibria.

*We thank Guillaume Rocheteau, Pedro Gomis-Porqueraz, Pushkar Maitra, Dilip Mookherjee, Gautam Bose and Raghbendra Jha for helpful discussions. We also thank participants at Australasian Economic Theory Workshop, ANU; Ashoka University and Delhi School of Economics for useful comments and suggestions. Ayushi Bajaj: ayushi.bajaj@monash.edu, Nikhil Damodaran: nnikhi1@lsu.edu.

1 Introduction

An important feature of cash, as opposed to other digital means of payments such as debit (or credit) cards and bank transfers is anonymity.[1] This in turn facilitates the emergence of a shadow economy, a sector that is deliberately concealed from public authorities for tax evasion but comprises legal production of goods and services. Since tax enforcement is costly or difficult, policies promoting digital payments or penalizing the use of cash can be used to boost tax revenues. Indeed, advocates of digitization and cashless payments including Rogoff (2014) emphasize the associated gains in revenue. In this paper, we build a model where size of the shadow sector is an equilibrium outcome and depends on the trade-offs between two means of payment: cash and digital. The model helps determine the size of shadow that maximizes tax revenue, aiding a comparison of policies that achieve this objective.

We adapt a class of monetary models which are explicit about the role of money as a medium of exchange viz. Lagos & Wright (2005) and Rocheteau & Wright (2005). We introduce two means of payment: cash and a digital means such as debit cards. The government is able to enforce tax payments on recorded transactions, hence only on ones using digital means. This links the size of shadow economy with the extent of digitization. Immordino & Russo (2017) use European data to find empirical evidence supporting this observation, as credit and debit card payments are found to be negatively related to VAT evasion. Along with paying taxes on digital transactions, we assume that using them incurs an additional fixed cost[2] - this captures the monetary as well as non-monetary costs associated with opening bank accounts, issuing debit cards, and also the cost incurred by sellers to set up payment machines. However, carrying cash is subject to theft, which amounts to a marginal cost on using cash. This cost also captures general portability costs associated with cash such as counting and waiting for change at check out lanes, cash bills bringing diseases and so on. We also introduce two types of buyers with a high or low preference for consumption. This helps us capture some heterogeneity in the

[1]Crypto-currencies are anonymous, but they're not yet in widespread use as media of exchange.

[2]Wang & Wolman (2016) using three years of transactions data from a discount retailer with thousands of stores in US find the evidence consistent with this theoretical framework in which consumers choose between cash and non-cash payments based on a threshold transaction size.

market for different means of payment, allowing existence of a partial shadow economy.

The model gives rise to three types of equilibria depending on the size of the shadow economy, which follows from the extent of digitization. That in turn depends on the tax policy and relative costs of the two means of payments. A full shadow equilibrium exists if the tax rate or cost of digital means of payment is too high and the cost of holding fiat money is relatively low. A partial one arises when buyers with a higher demand for consumption goods prefer to carry digital means of payment and thereby pay taxes than ones with a lower demand. This equilibrium shows that cash is more prevalent for small scale transactions, as evidenced by Wang & Wolman (2016). Finally, if taxes and cost to move to digital are low enough, then an all-digital or no shadow equilibrium exists.

We find that while the no shadow (or all-digital) equilibrium maximizes private welfare, tax revenue may be maximized in a partial shadow equilibrium. If the tax rate to incentivize everyone to pay taxes and the subsequent revenue generated is very small, having a partial shadow equilibrium is optimal for revenue maximization. In case the size of the shadow economy is larger than optimal for revenue maximization, we compare three policies to maximize revenue: penalizing cash holdings by suddenly demonetizing legal tender, directly facilitating digitized means of payment or simply reducing the tax rate.

If the equilibrium is unique and is at full shadow, a one-time demonetization while profitable for the government and imposing a one-time cost for non-tax payers has no long-run positive impact on households, as by itself the policy does not alter any fundamentals in the economy. But, if it does change the behavior of agents, for instance if coupled with changes in tax enforcements or if agents are made to believe that such an action can be undertaken again, then the effective cost of using cash increases, and if big enough, there might be a move to digital currency, and a lower shadow economy.[3] Alternatively, if the government sufficiently subsidizes use of digital means of payment (similar to India's '*Jan Dhan Yojna*'), then there would be a clear shift to lower shadow economy, and an increase in welfare. Finally, a simple

[3] A key criticism of the policy in India was also to do with the slow and haphazard remonetization, which added further to its costs with no clear benefits in sight. In the model we assume that remonetization is immediate, but since money is neutral, a fall in the money supply will have no real effects in the model.

reduction in the tax rate can also lead to a reduction in the shadow economy, by making digital means less costly on the margin for everyone, but the total revenue might be lower than simply subsidizing digitization, which depends on the cost of using the digital means.

In some regions of the parameter space two or more equilibria may co-exist, in which case the policy implications can be different. The presence of multiple equilibria implies that the economy's outcome, not just depends on fundamentals but also on beliefs. In particular, if partial and full shadow equilibrium co-exist, agents can coordinate on the full shadow economy equilibrium which in fact happens to be detrimental for both output and tax revenue. Using cash implies that the output obtained is low as at the margin cash is costlier than digital. The low output in turn means that it is not worth the cost of moving to digital means of payment and the economy remains at the full shadow equilibrium, despite the partial being Pareto improving.

If the economy's fundamentals implies the existence of multiple equilibria, and the economy is in a full/partial shadow equilibrium, then to move to a no shadow (or digital) equilibrium, a demonetization shock might work if it helps people coordinate on the "better" no shadow equilibrium. But, the model cannot shed further light on such a switch. And, in fact in this case, a subsidization policy towards digital means, may not work if agents cannot coordinate to move to the all-digital equilibrium. If the subsidy provided by the government is not large enough to effect a switch in equilibrium regimes, i.e. a move away from a region with multiple equilibria to that of unique all-digital, then policies such as the '*Jan Dhan Yojna*' would prove to be ineffective with perhaps many unused bank accounts.

Given the rich set of analytical cases we calibrate the model to Indian and US datasets to compare the two economies and derive consequent policy implications. A calibration of the model shows that a unique partial shadow equilibrium exists, leading to the conclusion that demonetization as a policy tool may not have any desired long run effect. Subsidized digitization and reducing tax rates have a similar effect on revenue generated. It is worth pointing out that the model only captures the effects of policies on the transactions side of shadow economy, and not its stock in which can take the form of real estate investments. We also do not capture income tax evasion, but only sales tax. Furthermore, this class of models assumes a quasi-linear utility which facilitates analytical characterization of results giving clean predictions. We however, are not able to capture wealth effects or distributional implications - though we include types of agents based on preferences which captures some heterogeneity and the differential effects of policy. We are also not able to study persistent or long-run effects of policy, as choice of future variables is history-independent in the model. In reality, sellers may be investment constrained, and demonetization might not only negatively affect one-period consumption but also future investment, in that sense we underestimate the costs of demonetization.

The model can be extended to include a richer sectoral level analysis as different sectors have differential tax rates and subsequently various levels of shadow. This might also help analyze the interaction of demonetization as a policy with the new goods and services tax in India. Moreover, in the current paper the fixed cost of using digital does not depend on the digital population. We can include positive network externalities and negative congestion externalities to these costs. We can further extend the model to include a banking sector, to talk about monetary policy issues surrounding interest on reserves and negative nominal rates amidst a zero lower bound environment.

This paper builds mainly on the theoretical literature on shadow economies, while drawing important insights from its empirical counterpart. A paper working on a similar class of models where a shadow economy arises endogenously is Gomis-Porqueras et al. (2014). They however compare cash and trade credit as the two means of payment with cash being subject to a inflation tax and credit transactions subject to tax (as is the case with digital currency in the current paper). In the present paper, since both means of payment are issued as money, both are subject to inflation.

Other work includes Koreshkova (2006) which again focuses on inflation as tax on underground economy. And, Loayza (1996) focuses on Latin American economies. The recent book by Rogoff titled "Curse of the Cash" develops insights from Rogoff (2014), which has interesting examples and cases of tax evasion and makes a case for phasing out large denomination bills.

Other papers that model the two means of payment: cash and digital without discussing the shadow economy include Li (2011), David et al. (2016) and Kim & Lee (2010).

The paper is organized as follows. Section 2 lays down the model environment and Section 3 the equilibrium. Section 4 includes the key analytical results of the model in terms of the different equilibrium regimes. Section 5 draws policy conclusions from the model, Section 6 does a quantitative analysis and Section 7 concludes.

2 Model

Time is discrete and continues forever. Household's discount rate after every period is $\rho > 0$ and $\beta \equiv 1/(1+\rho)$. In each period, two markets meet sequentially: a frictionless centralized goods market denoted CM, and a decentralized goods market with frictions called DM[4]. The alternating goods market structure follows from Lagos and Wright (2005) and Rocheteau and Wright (2005). There is a unit measure of households divided between buyers indexed by b and sellers indexed by s depending on their role in DM.

There are two types of perishable and non-storable consumption goods. First, a good that is traded and consumed in CM by households (quantity denoted by x, is also the numeraire), second, a good that is produced by sellers in DM and valued by buyers in DM (quantity denoted by q). Buyers are of two types depending on the utility they get from consumption in DM. The high type buyers, indexed by h get utility $u(q)$ (and low indexed by l get utility $\epsilon u(q)$, $\epsilon \in (0,1)$) from consuming q units of the consumption good in DM which is produced at cost $c(q) = q$ by sellers. The CM good gives utility $U(x)$, and l denotes labor supply in the CM.

The period payoff for high-type buyers (pre-multiply $u(q)$ by ϵ for low types) and sellers are as follows:

$$\mathcal{U}^b(q,x,l) = u(q) + U(x) - l, \text{ and } \mathcal{U}^s(q,x,l) = -q + U(x) - l. \tag{1}$$

As usual, U and u are twice continuously differentiable with $U^{'} > 0$, $U^{''} < 0$, $u^{'} > 0$ and $u^{''} < 0$. Also, $u(0) = 0$. Define the efficient $q = q^*$ such that $u^{'}(q^*) = 1$. The quasi-linear structure

[4]We can think of the DM as the informal sector of the economy, and shadow will be part of the informal sector.

in (1) follows from Lagos and Wright (2005) which simplifies the analysis because it leads to a degenerate distribution of assets across agents of a given type at the start of each DM market, and makes CM payoffs linear in wealth.

Buyers and sellers meet bilaterally in the DM with probability α, where terms of trade are determined by a take-it-or-leave-it offer by the buyer for simplicity. Since the consumption goods are perishable and non-storable, direct barter in DM is ruled out and agents are to some extent anonymous in DM to also rule out unsecured credit. This generates a role for assets in facilitation of intertemporal exchange. The Central Bank/government supplies fiat money; supply every period is denoted by M_t which acts as a medium of exchange.

Transactions in DM are also subject to per unit taxation by the government at the rate of τ. But, since the meetings are somewhat anonymous, the government is no more powerful than private agents in tracking transactions, so only transactions that can be tracked/recorded (quantity denoted by q^t) can be taxed. The government's budget constraint is:

$$G + \tau q^t - \gamma \phi M = 0, \tag{2}$$

where the first term is government consumption of the CM good, second is the sales tax on the tracked DM good which collected in CM. Money growth rate is γ implying $M_{+1}/M = 1 + \gamma$ and ϕ is the price of money in terms of CM goods. When we focus on stationary equilibria, $z \equiv \phi M$ is constant over time which implies that money growth rate equals the inflation rate: $\phi/\phi_{+1} = M_{+1}/M = 1 + \gamma$, where subscript +1 indicates next period. Given other variables, G adjusts such that (2) holds.

The form in which the money issued by the Central Bank is carried to DM determines whether the transaction can be tracked or not. There are two such modes of payment: cash or digital (like a debit card or e-wallet). We model the choice of payment instruments as letting buyers decide to carry her fiat money holdings as cash or digital before entering the DM. Conversion to digital is assumed to incur a fixed cost κ[5] - it captures the monetary as well as non-monetary costs associated with opening bank account, issuing such prepaid cards,

[5]Wang and Wolman (2016) using three years of transactions data from a discount retailer with thousands of stores find the evidence as consistent with such a theoretical framework in which consumers choose between

and also the cost incurred by the seller to set up the payment (we assume that it falls fully on the buyer, due to our assumption of take-it-or-leave-it offer by the buyer which we make for simplicity). Moreover, while cash exchange is anonymous implying that tax payments on such transactions can be avoided, digital ones are recorded, so the government collects tax on them further imposing a marginal cost on using digital payment. But, carrying cash from DM to CM is also subject to a marginal cost, as sellers lose their money holdings with probability, η before entering the CM (this also captures general portability costs associated with cash).

3 Equilibrium

We now discuss the equilibrium of the economy. Let the CM and DM value functions be denoted by $W(\cdot)$ and $V(\cdot)$.

A buyer type $i \in \{h, l\}$ in CM solves:

$$W_i(m_i) = \max_{x,l,m_i^{'} \geqslant 0} \{U(x) - l + \beta V_i(m_i^{'})\},$$

s.t.

$$x = \phi m_i + l - \phi m_i^{'},$$

where next period values, i.e. the money balances, m_i carried to the next sub-period are denoted by $'$. Note that we have assumed that 1 unit of l produces 1 unit of x. Now, substitute for l to get,

$$W_i(m_i) = \max_{x,m_i^{'} \geqslant 0} \{U(x) - x + \phi m_i - \phi m_i^{'} + \beta V_i(m_i^{'})\}. \tag{3}$$

There is a similar problem for sellers in CM but we can assume that they carry no money to DM, as money is costly to hold. Given that the labor supply is not binding, the quasi-linear structure implies the following: $U'(x) = 1$, m_i' is independent of m_i and that $W(\cdot)$ is linear in m. The linearity simplifies the bargaining problem in DM, which we solve next.

cash and non-cash payments based on a threshold transaction size.

Recall that we assume buyer take-it-or-leave-it offers for simplicity, and before entering the DM, the buyer has to decide whether to carry her money in the form of cash or digital. Thus, there can be four types of matches, two for each type-l holding either cash, c or digital, d. Denote $g_i^c \leq \phi m_i$ and $g_i^d \leq \phi m_i - \kappa$ as the buyer's liquidity holding for each type of buyer match, $i = \{l, h\}$.

The buyer's bargaining problem in a match in which she carries cash becomes (recall that cash is stolen with probability η), $i = \{l, h\}$:

$$\max_{q_i^c, g_i^c \in [0, \phi m_i]} [u_i(q_i^c) + W_i(\phi^{'} m_i - g_i^c)] \text{ s.t. } - q_i^c + W((1-\eta) g_i^c) \geq 0.$$

The buyer offers $g_i^c < \phi m_i$ units of real balances in exchange for DM goods, q_i^c subject to the seller's participation constraint. Seller produces the good at linear cost, and carries the real balances to CM with probability, $(1-\eta)$. Using linearity of the value function, $W_i(\cdot)$ we can simplify the above problem as follows:

$$\max_{q_i^c, g_i^c \in [0, \phi m_i]} [u_i(q_i^c) + g_i^c] \text{ s.t. } - q_i^c + (1-\eta) g_i^c \geq 0.$$

The solution to the above problem is given by:

$$q_i^c = \min\{\phi m_i, q_i^\eta\}. \tag{4}$$

If the buyer holds enough liquidity, the first-best level of output for the low type subject to theft, $q_i^\eta < q^*$ is obtained (which solves $\epsilon u^{'}(q_l^\eta) = 1/(1-\eta)$ or $u^{'}(q_l^\eta) = 1/(1-\eta)$ depending on the type), else the buyer gets as much as she can given her money holdings, ϕm_i.

The buyer's problem in a match in which she carries digital means of payment, d becomes (recall that since digital transactions are recorded, they are subject to tax at the rate, τ and buyers incur a cost, κ to use the digital means of payment), $i = \{l, h\}$:

$$\max_{q_i^d, g_i^d \in [0, \phi m_i - \kappa]} \{u_i(q_i^d) + W_i(\phi^{'} m_i - \kappa - g_i^d)\} \text{ s.t. } - q_i^d + W_i[(1-\tau) g_i^d] \geq 0.$$

The buyer offers $g_i^d < \phi m_i - \kappa$ units of real balances in exchange for DM goods, q_i^c subject to the seller's participation constraint. Seller produces the good at linear cost, and carries the real balances to CM where her after-tax balance is, $(1-\tau)g_i^d$. Using linearity of the value function, $W_i(\cdot)$ we can simplify the above problem as:

$$\max_{q_i^d, g_i^d \in [0, \phi m_i - \kappa]} \{u_i(q_i^d) - g_i^d - \kappa\} \text{ s.t. } -q_i^d + (1-\tau)g_i^d \geq 0.$$

Thus, output will be:

$$q_i^d = \min\{\phi m_i, q_i^\tau\}. \tag{5}$$

If the buyer holds enough liquidity the efficient level of output for the low type subject to tax, $q_i^\tau < q^*$ (which solves $\epsilon u^{'}(q) = 1/(1-\tau)$ or $u^{'}(q) = 1/(1-\tau)$ depending on the type) is obtained, else the buyer gets as much as she can given her money holdings, $\phi m_i - \kappa$.

Having solved the bargaining problem, we can now write the DM value functions for the seller and the buyer. Seller's DM value function is as follows:

$$V^S(m) = \phi m + W_d^B(0). \tag{6}$$

The value to a seller in the DM is her share of the surplus (=0) when matched with the buyer to whom she sells the DM good. But, she has her money holdings, ϕm which she carries forward to the CM. The buyer's DM value function depends on her type $i = \{l, h\}$, as follows:

$$V_i(m_i) = \alpha \max_{c,d} \left\{ \max_{q_i^c, g_i^c \in (0, \phi m_i)} [u_i(q_i^c) + W(\phi m_i - g_i^c)], \max_{q_i^d, g_i^d \in (0, \phi m_i - \kappa)} [u_i(q_i^d) + W(\phi m_i - g_i^d - \kappa)] \right\} + (1-\alpha) W_i(\phi m_i).$$

The value to a buyer in the DM is her share of the surplus when matched with the seller, which happens with probability α. Depending on the bargaining outcomes (4) or (5), the choice between cash and digital is based on which gives the higher bargaining surplus which in turn determines the quantity of DM good, q_i^j produced in return for g_i^j units of real balances, where $j \in \{c, d\}$. There is also a possibility that the buyer does not match with a seller with complementary probability, $(1-\alpha)$ in which case she carries forward her money holdings, ϕm_i to the CM. Using linearity of the value function, we can simplify the above as follows:

$$V_i(m_i) = \alpha \max_{c,d} \left\{ \max_{q_i^c, g_i^c \in (0, \phi m_i)} [u_i(q_i^c) - g_i^c], \max_{q_i^d, g_i^d \in (0, \phi m_i - \kappa)} [u_i(q_i^d) - g_i^d - \kappa] \right\} + \phi m_i + W_i(0). \tag{7}$$

We now take the buyer's DM value function in (7) one period forward and plug in to the CM value function in (3). Ignore constants to get a maximization problem that determines the money holdings of the buyer type-$i \in \{l, h\}$ given her portfolio choice $j \in \{c, d\}$.

$$\max_{m_i} \left\{ -\phi m_i + \beta \left\{ \phi' m_i + \alpha \max_{c,d} \left\{ \max_{q_i^c, g_i^c \in (0, \phi m_i)} [u_i(q_i^c) - g_i^c], \max_{q_i^d, g_i^d \in (0, \phi m_i)} [u_i(q_i^d) - g_i^d - \kappa] \right\} \right\} \right\}. \quad (8)$$

If buyers choose positive holdings of real balances, we get the following FOCs, where q_i depends on whether the i-type buyer carries cash or digital money. Differentiate (8) with respect to m_i to get:

$$\frac{\phi_m}{\phi_m'} = \beta \left\{ 1 + \alpha \max_{c,d} \left\{ u_i'(q_i^c) - \frac{1}{1-\eta}, u_i'(q_i^d) - \frac{1}{1-\tau} \right\} \right\}, \quad (9)$$

where q_i^c is given by (4) and q_i^d is given by (5). It turns out that since money is costly to hold, households do not carry anything in excess over what they intend to use in DM, so $q_i^j = \phi m_i$, for $j \in \{c, d\}$ and $i \in \{l, h\}$.

Since markets have to clear, we have that total demand for money is equal to its supply:

$$\phi(\pi_l m_l + \pi_h m_h) = \phi M. \quad (10)$$

We can now define the steady state monetary equilibrium of the economy as follows.

Definition 1. *A steady-state monetary equilibrium is a list* $(q_i^c, q_i^d, m_i, \phi)$ *for* $i \in \{l, h\}$ *that solves* (4)*,* (5)*,* (9)*, and* (10)*.*

Before we further characterize the equilibrium we can do some comparative statics on output for the different cases. Output from matches with cash and digital are decreasing in their respective marginal costs, η and τ respectively, and both are unaffected by the fixed cost, κ. The preference parameter, ϵ affects the output consumed by the l-type agents in both types of matches, and not the h-type's output. Finally, other parameters of the model, including money growth rate, γ, discount factor, β and matching probability, α affect the outputs as in a standard monetary model. Higher inflation implies lower output, as money is costlier to hold.

Higher weight on DM output and higher matching probabilities imply higher quantities. This discussion is formalized in Lemma 1 below.

Lemma 1 (Comparative Statics)**.** *Denote $q_i^j(\cdot)$ as functions of $\eta, \tau, \kappa, \epsilon, \gamma, \beta, \alpha$ where $i \in \{l, h\}$ and $j \in \{c, d\}$: (i) $\frac{q_i^c}{\partial \eta} < 0$, $\frac{q_i^d}{\partial \eta} = 0$, (ii) $\frac{q_i^c}{\partial \tau} = 0$, $\frac{q_i^d}{\partial \tau} < 0$, (iii) $\frac{q_i^j}{\partial \kappa} = 0$, (iv) $\frac{q_l^j}{\partial \epsilon} > 0$, $\frac{q_h^j}{\partial \epsilon} = 0$, (vi) $\frac{q_i^j}{\partial \gamma} < 0$, (vii) $\frac{q_i^j}{\partial \beta} > 0$, (viii) $\frac{q_i^j}{\partial \alpha} > 0$.*

4 Equilibrium Regimes

We now solve for the steady state equilibrium as defined in Definition 1. To solve (9) to get ϕ and $m_i, \in \{l, h\}$ we need to know the choice of c, d. And, to solve for the choice we need q_i^c, q_i^d from (4) and (5) which in turn depend on ϕ and m_i. So, the way we proceed is by assuming the portfolio choice of either cash or digital and then solve (9) for ϕ and m_i along with the market clearing in (10). We then plug back those to the problem that determines portfolio choice (between c, d that maximizes the surplus) to check if indeed the assumption we made about the portfolio choice holds.

There can potentially be four types of equilibrium: both carry cash, both carry digital, l carries cash, h carries digital and h carries cash, l caries digital. We start with the equilibrium when both types hold cash (ϕ and m_i are indexed by c). The following FOCs along with the market clearing condition given in (10) solve for three unknowns, ϕ^c, m_l^c and m_h^c:

$$\iota = \alpha\left[\epsilon u^{'}(\phi^c m_l^c) - \frac{1}{1-\eta}\right], \tag{11}$$

$$\iota = \alpha\left[u^{'}(\phi^c m_h^c) - \frac{1}{1-\eta}\right]. \tag{12}$$

The above is true if given ϕ^c, m_l^c and m_h^c above, both types $i \in \{l, h\}$ prefer to carry cash i.e. the following conditions are satisfied:

$$\epsilon u(\phi^c m_l^c) - \frac{\phi^c m_l^c}{1-\eta} > \epsilon u(\phi^c m_l^c) - \frac{\phi^c m_l^c}{1-\tau} - \kappa, \tag{13}$$

and,

$$u(\phi^c m_h^c) - \frac{\phi^c m_h^c}{1-\eta} > u(\phi^c m_h^c) - \frac{\phi^c m_h^c}{1-\tau} - \kappa. \tag{14}$$

Since, $\epsilon < 1$, we have $m_l^c < m_h^c$ from (11) and (12), so if (14) is satisfied, then (13) is also satisfied. So, both carrying cash is an equilibrium if and only if $\tau \geq \bar{\tau}^c$ where $\bar{\tau}^c$ is defined as:

$$\frac{\phi^c m_h^c}{1-\eta} - \kappa = \frac{\phi^c m_h^c}{1-\bar{\tau}^c}. \tag{15}$$

Second, consider the equilibrium when both types hold digital (we label the value of money as ϕ^d to denote the case we're considering), the following FOCs and the market clearing condition given in (10) solve for three unknowns, ϕ^d, m_l^d and m_h^d:

$$\iota = \alpha\left[\epsilon u^{'}(\phi^d m_l^d) - \frac{1}{1-\tau}\right], \tag{16}$$

$$\iota = \alpha\left[u^{'}(\phi^d m_h^d) - \frac{1}{1-\tau}\right]. \tag{17}$$

The above is true if given ϕ^d, m_l^d and m_h^d above, both types $i \in \{l, h\}$ prefer to carry digital, i.e. the following conditions are satisfied:

$$\epsilon u(\phi^d m_l^d) - \frac{\phi^d m_l^d}{1-\eta} < \epsilon u(\phi^d m_l^d) - \frac{\phi^d m_l^d}{1-\tau} - \kappa, \tag{18}$$

and,

$$u(\phi^d m_h^d) - \frac{\phi^d m_h^d}{1-\eta} < u(\phi^d m_h^d) - \frac{\phi^d m_h^d}{1-\tau} - \kappa. \tag{19}$$

Since, $\epsilon < 1$, we have $m_l^d < m_h^d$ from (16) and (17), so if (18) is satisfied then (19) is also satisfied. So, both carrying digital if and only if $\tau \leq \underline{\tau}^d$ where $\underline{\tau}^d$ is defined as:

$$\frac{\phi^d m_l^d}{1-\eta} - \kappa = \frac{\phi^d m_l^d}{1-\underline{\tau}^d}. \tag{20}$$

Finally, consider the equilibrium when l-types carry cash and h carries digital (the other case with l digital and h cash will never be an equilibrium due to the presence of a fixed cost for digital because $\epsilon < 1$). We remove the indexation from ϕ and m_i for this case. The following FOCs and the market clearing in (10) solve for three unknowns, ϕ, m_l and m_h:

$$\iota = \alpha\left[\epsilon u^{'}(\phi m_l) - \frac{1}{1-\eta}\right], \tag{21}$$

$$\iota = \alpha\left[u^{'}(\phi m_h) - \frac{1}{1-\tau}\right]. \tag{22}$$

The above is true if given ϕ, m_l and m_h above, type l prefers to carry cash and h digital, i.e. the following conditions are satisfied:

$$\epsilon u(\phi m_l) - \frac{\phi m_l}{1-\eta} > \epsilon u(\phi m_l) - \frac{\phi m_l}{1-\tau} - \kappa, \tag{23}$$

and,

$$u(\phi m_h) - \frac{\phi m_h}{1-\eta} < u(\phi m_h) - \frac{\phi m_h}{1-\tau} - \kappa. \tag{24}$$

Thus, l-types carry cash and h carries digital is an equilibrium if and only if $\underline{\tau} \leq \tau \leq \bar{\tau}$ where $\underline{\tau}$ and $\bar{\tau}$ are defined as:

$$\frac{\phi m_h}{1-\eta} - \kappa = \frac{\phi m_h}{1-\bar{\tau}}, \tag{25}$$

$$\frac{\phi m_l}{1-\eta} - \kappa = \frac{\phi m_l}{1-\underline{\tau}}. \tag{26}$$

Before we characterize conditions for the existence of each equilibrium type, we can do some comparative statics for the value of money, ϕ^i and money demand, m_i^j for the different equilibrium types, $j \in \{, c, d\}$ and buyer types, $i \in \{l, h\}$. The following lemma summarizes the key results followed by the proof.

Lemma 2 (Comparative Statics)**.** *Denote $m_i^j(\cdot)$ and ϕ^j as functions of η, ϵ where $i \in \{l, h\}$ and $j \in \{, c, d\}$: (i)* $\frac{m_i^c}{\partial\eta} < 0$, $\frac{m_i^d}{\partial\eta} = 0$, $\frac{m_l}{\partial\eta} < 0$, $\frac{m_h}{\partial\eta} > 0$, $\frac{\phi^d}{\partial\eta} = 0$, $\frac{\phi^c}{\partial\eta} > 0$, $\frac{\phi}{\partial\eta} < 0$, *(ii)* $\frac{m_l^j}{\partial\epsilon} > 0$, $\frac{m_h^j}{\partial\epsilon} < 0$, $\frac{\phi^d}{\partial\epsilon} > 0$, $\frac{\phi^c}{\partial\epsilon} > 0$, $\frac{\phi}{\partial\epsilon} > 0$.

Proof: From Lemma 1, we know that $\frac{\partial q_l^j}{\partial\epsilon} > 0$, so $\frac{\partial m_l^j}{\partial\epsilon} > 0$ and $\frac{\partial\phi^j}{\partial\epsilon} > 0$. From market clearing, $\frac{\partial m_h^j}{\partial\epsilon} < 0$, which does not contradict $\frac{\partial q_h^j}{\partial\epsilon} = 0$. Other possibilities will contradict the last equality. ■

To characterize conditions for the existence of each equilibrium type, first note that since $\kappa > 0$, the marginal cost of using digital currency, τ has to be lower than that of cash, η for a digital equilibrium to exist,[6]as characterized in the following lemma.

Lemma 3. *If a full or partial digital equilibrium exists then, the marginal cost of using digital currency must be lower than that of cash, i.e.* $\tau < \eta$.

Second, characterize the relation among the threshold tax rates which govern the equilibrium type as defined in (15), (20), (25) and (26) which helps pins down the conditions for existence of different equilibrium types.

Lemma 4. *The thresholds on tax rates,* $\bar{\tau}^c$, $\underline{\tau}^d$, $\bar{\tau}$ *and* $\underline{\tau}$ *given in* (15), (20), (25) *and* (26) *have the following relations: (i) for all parameter values,* $\bar{\tau} > \bar{\tau}^c > \underline{\tau}$, $\bar{\tau} > \underline{\tau}^d > \underline{\tau}$ *and, (ii) for sufficiently high* η *and high* ϵ, $\bar{\tau}^c < \underline{\tau}^d$.

Proof: (i) The thresholds for τ are decreasing in $q = \phi m$. Since $\phi^d m_l^d > \phi m_l$ (as τ is lower than η for the inequality to hold), we get $\underline{\tau}^d > \underline{\tau}$ and since $\phi m_h > \phi^c m_h^c$ we get $\bar{\tau} > \bar{\tau}^c$.

(ii) The relation between $\phi^d m_l^d$ and $\phi^c m_h^c$ is less straightforward, so we can get $\bar{\tau}^c < \underline{\tau}^d$ (if $\phi^d m_l^d > \phi^c m_h^c$) or $\bar{\tau}^c > \underline{\tau}^d$ (if $\phi^d m_l^d < \phi^c m_h^c$).

We know that $\phi_d m_h^d > \phi^d m_l^d > \phi^c m_l^c$ and $\phi_d m_h^d > \phi^c m_h^c > \phi^c m_l^c$. If ϵ is close to 1 such that $\phi_d m_h^d$ and $\phi_d m_l^d$ are close enough and η (or τ given η) is large such that $\phi_d m_h^d$ and $\phi^c m_h^c$ are far enough, we get that $\phi^d m_l^d > \phi^c m_h^c$ and $\bar{\tau}^c < \underline{\tau}^d$ (or the η effect dominates the ϵ effect).

■

Since, we are interested in the size of shadow economy that emerges, we characterize the different equilibrium regimes in terms of its size, as given in Proposition 1 below. The model is constructed such that the size of the shadow economy follows from the extent of digitization which is based on the observation that there is a negative relation between the shadow economy and digitization. Immordino and Russo (2017) use European data to find empirical evidence supporting this observation. They find that credit and debit card payments are negatively

[6]While this may seem like a strong assumption in reality one can think of different sectors with different tax rates or enforcement probabilities where for some it is optimal to operate in the shadow economy.

related to VAT evasion. They also find that using these electronic cards to withdraw cash at ATMs - which makes cash more abundant - fosters VAT evasion.[7]

Proposition 1 (Shadow Economy)**.** *From Lemma 4, three broad equilibrium regimes emerge that characterize existence and size of the shadow economy:*

1: $0 \leqslant \tau \leqslant \underline{\tau}$*: No shadow economy*

2: $\underline{\tau} < \tau < \bar{\tau}$*: Part shadow, all or no*

3: $\bar{\tau} \leqslant \tau$*: All shadow*

Moreover, under regime 2, we have two possible cases:

2(i): if $\bar{\tau}^c < \underline{\tau}^d$ *(high* η*, or low* τ *given* η *and high* ϵ*) then three multiple equilibria regimes,*

2(ii): if $\bar{\tau}^c > \underline{\tau}^d$ *(low* η *and low* ϵ*) then one unique partial and two multiple regimes.*

It is straightforward to note that for $\tau > \eta$, a full shadow equilibrium will always exist. But, it also exists for $\bar{\tau}^c < \tau < \eta$, since for any tax rate above $\bar{\tau}^c < \eta$ all agents are better off using cash, as using digital also involves a fixed cost. Similarly, for all $\tau < \underline{\tau}^d$, no one uses cash or a no shadow equilibrium exists. There are two other thresholds such that for $\underline{\tau} < \tau < \bar{\tau}$, a partial shadow equilibrium exists. Now putting it all together, from Lemma 4 (i), there will always be a region where partial shadow co-exists with no shadow or full shadow equilibria. But, from Lemma 2, under certain conditions, when $\bar{\tau}^d > \underline{\tau}^c$, all three equilibria also co-exist else there is a region with a unique partial shadow equilibrium. Figure 1 illustrates the different cases. In the above panel, $\underline{\tau}^d < \bar{\tau}^c$, so a unique partial shadow equilibrium exists for $\underline{\tau}^d < \tau < \bar{\tau}^c$, and in the bottom panel $\underline{\tau}^d > \bar{\tau}^c$, so all three equilibria co-exist for $\bar{\tau}^c < \tau < \underline{\tau}^d$.

The thresholds on tax rates which determine the size of different regimes themselves depend on other parameters as characterized in the following proposition:

Proposition 2 (Size of Equilibrium Regimes)**.** *Denote* $\bar{\tau}(\cdot)$ *and* $\underline{\tau}(\cdot)$ *as functions of* $\eta, \epsilon, \kappa, \gamma, \beta, \alpha$ *to get: (i)* $\frac{\partial \underline{\tau}^d}{\partial \eta} > 0, \frac{\partial \bar{\tau}}{\partial \eta} > 0$*, (ii)* $\frac{\partial \underline{\tau}}{\partial \epsilon} > 0$, $\frac{\partial \bar{\tau}}{\partial \epsilon} = 0$, $\frac{\partial \underline{\tau}^d}{\partial \epsilon} > 0$, $\frac{\partial \bar{\tau}^c}{\partial \epsilon} = 0$*, (iii)* $\frac{\partial \underline{\tau}}{\partial \kappa} < 0$, $\frac{\partial \bar{\tau}}{\partial \kappa} < 0$*, (iv)*

[7] But, there are examples where this need not be the case in the real world (counter examples are Japan with low shadow and digitization and Kenya with high digitization and high shadow). We can also introduce another parameter on the degree of enforcement by the government, or the efficiency of the tax authority in enforcing tax payments to capture these cases.

$\frac{\partial \underline{\tau}}{\partial \alpha} > 0$, $\frac{\partial \bar{\tau}}{\partial \alpha} > 0$, *(v)* $\frac{\partial \underline{\tau}}{\partial \beta} < 0$, $\frac{\partial \bar{\tau}}{\partial \beta} < 0$, *(vi)* $\frac{\partial \underline{\tau}}{\partial \gamma} < 0$, $\frac{\partial \bar{\tau}}{\partial \gamma} < 0$.

Sketch of Proof: (i) It's hard to show other cases, but since $\phi^d m_l^d$ and ϕm_h are independent of η these two cases are easy to see.

(ii) The thresholds are increasing in ϕm, so the sign of these derivatives depends on ϕm responds to ϵ - $\phi^j m_l^j$ is increasing in ϵ and $\phi^j m_h^j$ is independent.

(iii) straightforward

(iv) - (vi) via ϕm

■

Figure 2 shows the different regimes based on two parameters: the tax rate, τ and fixed cost of using digital, κ. From Lemma 3, we assume that η is the maximum tax rate, as for any τ above η, it is always better to hold cash at the margin, and κ becomes irrelevant.[8] As discussed, we can see for low τ and κ, there is an equilibrium with only digital, for high we have only cash and the intermediate cases include regimes with multiple equilibria or one with a unique partial shadow equilibrium.

Compare Output and Surplus in Equilibrium Regimes

We can now find DM output and aggregate DM surplus (defined as the weighted sum of surpluses for the high and low type buyer, as seller gets zero surplus) for the different equilibrium regimes. Given that $\tau < \eta$ (for existence of digital equilibrium from Lemma 3), output is always higher under a digital equilibrium. So at the margin it is always better to hold digital and q is higher under digital. But since there is a fixed cost of holding digital currency, achieving the higher output necessarily involves some waste of resources, so the question is whether digital still gives a higher surplus, and the answer is yes.

First, consider the case with multiple equilibria where we can find the 'best' equilibrium. Compare the surplus from all digital and all cash for the low type first. For both equilibria to co-exist we need that the low type prefers cash if everyone uses cash and digital if everyone

[8]It is assumed that $u(q) = q^{(1-\sigma)}/(1-\sigma)$, and $\sigma = 0.5$. Other parameters are set at: $\eta = 0.5$, $\alpha = 0.5$, $\iota = 0.02$, $\epsilon = 0.6$, $\pi_l = 0.5$, $M = 0.26$.

goes digital, i.e. both (13) and (18) have to be satisfied. Since, we know that $\phi^c m_l^c < \phi^d m_l^d$ as $\tau < \eta$, so it follows that:

$$\epsilon u(\phi^d m_l^d) - \frac{\phi^d m_l^d}{1-\tau} - \kappa > \epsilon u(\phi^d m_l^d) - \frac{\phi^d m_l^d}{1-\eta} > \epsilon u(\phi^c m_l^c) - \frac{\phi^c m_l^c}{1-\eta}.$$

The derivation follows exactly for the high type, just replace l with h in all equations. In words, for the two equilibria to co-exist the output under digital for both types has to be much higher than under cash which implies that the net surplus under digital turns out to be higher. The low output in cash equilibrium means that going digital is not worth the fixed cost, κ, i.e. if everyone is pessimistic and thinks that others will hold cash, then ϕ is low and they don't expect to get high output in the matches and hence they continue to hold cash. Whereas κ is worth it if everyone else goes digital, as ϕ is high enough.

Now compare the surplus from all digital and partial shadow equilibrium when they co-exist. In case of partial shadow equilibrium, we have that the low type prefers cash and high digital if everyone else does so, i.e. (23) and (24) are satisfied. For the partial to co-exist with all digital we also need (18) and (19) to be satisfied simultaneously. Since, we know that $\phi m_i < \phi^d m_i^d$ for $i \in \{l, h\}$ as $\tau < \eta$, so it follows that:

$$u(\phi^d m_h^d) - \frac{\phi^d m_h^d}{1-\tau} - \kappa > u(\phi m_h) - \frac{\phi m_h}{1-\tau} - \kappa,$$

and

$$\epsilon u(\phi^d m_l^d) - \frac{\phi^d m_l^d}{1-\tau} - \kappa > \epsilon u(\phi^d m_l^d) - \frac{\phi^d m_l^d}{1-\eta} > \epsilon u(\phi m_l) - \frac{\phi m_l}{1-\eta}.$$

Thus, both the high and low types get a higher surplus in the no shadow than in the partial shadow equilibrium.

Second, consider the cases with unique pure equilibrium. Compare the surplus for low type between full and no shadow first. Since, we know that $\phi^c m_l^c < \phi^d m_l^d$ as $\tau < \eta$, and (18) holds so it follows that:

$$\epsilon u(\phi^d m_l^d) - \frac{\phi^d m_l^d}{1-\tau} - \kappa > \epsilon u(\phi^d m_l^d) - \frac{\phi^d m_l^d}{1-\eta} > \epsilon u(\phi^c m_l^c) - \frac{\phi^c m_l^c}{1-\eta}.$$

The derivation follows exactly for the high type, just replace l with h in all equations. And, similarly for comparison between the partial shadow and no shadow equilibrium. Finally, since taxes are distortionary, $\tau = 0$ is the optimal tax rate for surplus maximization.

Tax Revenue

We now consider the size of tax revenue, τq depending on the equilibrium regimes. We assume that $\bar{\tau} < \tau^*$, where τ^* is the optimal tax rate that maximizes tax revenue given that taxes can be enforced perfectly. It solves the following maximization problem:

$$\tau^* = \arg\max_{\tau}[\pi_h q_h^d(\tau) + \pi_l q_l^d(\tau)] \tag{27}$$

Given that $\bar{\tau} < \tau^*$, revenue is increasing in τ in the no shadow regime, falls when the switch to partial shadow takes place but is rising hence till the regime switches to partial shadow during which it is of course flat.

If the objective of the government is to set τ such that the tax revenue is maximized then it may be optimal to have a small shadow economy, as the partial shadow economy regime has a higher tax rate, even though lower enforcement. Clearly, $\tau \geqslant \bar{\tau}$ yields zero revenue, so that is not optimal. If the following condition holds, then it may be optimal to have some shadow economy from a revenue maximization perspective:

$$\bar{\tau}\pi_h \phi(\bar{\tau}) m_h(\bar{\tau}) > \underline{\tau}\phi^d(\underline{\tau})M,$$

$$\left(1 - \frac{1}{\frac{1}{1-\eta} - \frac{\kappa}{\phi m_h}}\right)\pi_h \phi m_h > \left(1 - \frac{1}{\frac{1}{1-\eta} - \frac{\kappa}{\phi m_l}}\right)\phi^d M \tag{28}$$

We can derive a condition on ϵ such that the above holds, as formalized in the proposition below. It turns out that it is true for low ϵ, because if ϵ is low then m_l is much lower compared to m_h and $\bar{\tau}$ is much higher than $\underline{\tau}$. Also, if m_l^d is lower in absolute terms means that we don't lose much revenue. We also need to check for ϕm^h and $\phi^d m_h^d$, if ϵ is low then these two values will be closer than if ϵ were high.

Proposition 3 (Optimal Shadow)**.** *(i) There exists a threshold on l-type's utility weight,* $\tilde{\epsilon}$ *such that for all* $\epsilon < \tilde{\epsilon}$*, (28) holds and it is optimal to have some shadow economy to maximize revenue.*

Proof: At $\epsilon = 0$, LHS > RHS, and at $\epsilon = 1$, RHS > LHS. Furthermore, RHS is increasing in ϵ and LHS is independent, so there is a unique $\tilde{\epsilon}$ where LHS = RHS and below which LHS > RHS. ■

Intuitively, if the low value transactions are so small[9] that they will be part of the shadow economy (as κ is relatively larger for them). And, it is also not worth it for the government to incentivize them to pay taxes for which the rates have to be lowered by a lot, i.e. $\underline{\tau}$ is very small which means the revenue obtained form the h-type also falls. Instead it is better to just collect from the high value transactions who would pay taxes even at a higher rate.

Assuming $\bar{\tau} < \tau^*$ and that when $\tau \in \{\underline{\tau}, \bar{\tau}\}$, the asymmetric equilibrium exists, the optimal tax rate for maximizing revenue, τ^o is given by[10]:

$$\tau^o = \begin{cases} \underline{\tau} & \text{if } \epsilon \geqslant \tilde{\epsilon}, \\ \bar{\tau} & \text{if } \epsilon < \tilde{\epsilon}. \end{cases}$$

5 Policy

We can now do some policy experiments, given that the objective of the policy authority is to maximize revenue (aggregate welfare analysis is less straightforward). Under perfect enforcement - a first best case the government should set $\tau = \tau^*$, but we are in the second best world where cash transactions cannot be enforced so the policy maker has to take other policy measures.

We characterized the equilibrium regimes based on two parameters: τ and κ as we will treat them as policy variables. Tax rate is a policy variable for obvious reasons, but so is κ in so far

[9]This is also true for low π_l for a similar reasoning. For $\pi_l \approx 0$, LHS > RHS, and for $\pi_l \approx 1$, LHS < RHS. And, RHS is independent of π_l, while LHS is increasing in π_l.

[10]Note that the optimal tax rate under perfect enforcement as defined earlier is τ^*

as the government can subsidize cost of digitization by facilitating opening of bank accounts, issuing debit cards and such. Since we do not model the digital platform, taking the cost of going digital as given, the subsidy works as the government pays part of the fixed cost in effect lowering κ for buyers. So, like we defined thresholds on τ, we can do so on κ following from our discussion on equilibrium regimes.

The key threshold we will be interested in for policy analysis is the cost of digitization which is low enough for high types to go digital, but not for low types - we call that $\bar{\kappa}$. It is defined from (25) as:

$$\bar{\kappa} = \frac{\phi m_h}{1-\eta} - \frac{\phi m_h}{1-\tau} = \phi m_h \frac{\eta - \tau}{(1-\eta)(1-\tau)}. \tag{29}$$

Thus, for $\kappa > \bar{\kappa}$, the economy is in a full shadow regime, and below it there can be multiple cases. We skip the discussion of those, as they're similar to that of τ.

Let the economy be in a full shadow regime, i.e. $\tau > \bar{\tau}$ for $\kappa > \bar{\kappa}$. The objective is to maximize tax revenue and reduce the size of the shadow economy, let (28) hold, so the optimal tax rate is $\bar{\tau}$ given κ or $\bar{\kappa}$ given τ. There are two options for a policy-maker:

(i) reduce tax rate to $\bar{\tau}$, revenue $= \bar{\tau}\pi_h\phi(\bar{\tau})m_h(\bar{\tau})$,

(ii) subsidize κ/facilitate digitization: lower κ to $\bar{\kappa}$, revenue $= \tau\pi_h\phi(\tau)m_h(\tau) - \pi_h(\kappa - \bar{\kappa})$.

Both policy tools bring the economy from full shadow to partial shadow equilibrium. The one which is least costly depends on the size of revenue generated. (ii) is better than (i) if $\kappa - \bar{\kappa} < \tau\phi(\tau)m_h(\tau) - \bar{\tau}\phi(\bar{\tau})m_h(\bar{\tau})$.

Lemma 5 (Policy)**.** *For $\kappa < \bar{\kappa} + \tau\phi(\tau)m_h(\tau) - \bar{\tau}\phi(\bar{\tau})m_h(\bar{\tau})$, subsidizing cost of digital currency generates larger revenue than reducing tax rate to, $\bar{\tau}$.*

Proof: Straightforward, and since we assume that $\bar{\tau} < \tau^*$, RHS is always positive.

■

So, if the cost is low enough or current tax is much higher than the $\bar{\tau}$ then subsidizing cost of digital currency generates larger revenue than reducing tax rate. However, if the economy is in the multiple equilibria regime - but is coordinating on the cash/full shadow equilibrium, then, these policies will not work if the economy stays in the multiple equilibrium regime. In particular, even if tax rate is lowered or κ subsidized, the coordination failure cannot be overcome unless the regime switches.

5.1 Demonetization

The model is well-suited to shed some light on the recent policy intervention in India where high denomination currency bills were demonetized (and subsequently remonetized).[11] The way we model such an intervention is by deeming useless any currency bills brought to CM on which taxes were not paid (which happens to be all the cash carried by sellers). So, they cannot be converted to consumption goods in CM if taxes were not paid on the corresponding transaction in DM. While profitable for the government (as $M = 0$ in (2)) it incurs a one-time loss for the sellers. Since, other parameters of the model remain unchanged, and in particular since remonetization is immediate, i.e. buyers who demand money to carry forward to the next can get it immediately, there is no other effect on prices or quantities. So, everything returns to business as usual.

One potential way such a policy can have a positive effect is if the economy starts in a multiple equilibrium regime. Such a policy shock, can change the beliefs of agents and push them to coordinate on the"better" equilibrium even though the policy itself does not change the fundamentals/parameters of the economy. It in fact comes at a one-time loss for sellers (redistribution from sellers to government/Central Bank). Since, our model does not allow for history dependence or persistence for tractability reasons, another potential perverse effect of such an intervention can be on the level of investment by sellers, which is missed by the model. So, the model gives a lower bound on the cost of this intervention, which can be larger than just the redistribution from sellers to government.

Another way this policy might work is by having a behavioral change in agents which actually

[11] The policy was criticized primarily because the remonetization process was very slow, which led to a drying up of liquidity in the economy. In the model we however assume that remonetization is immediate, to analyze solely the effects of demonetization; also if currency shrinks prices adjust immediately leading to no real changes in the model as money is neutral.

alters some parameters. For instance, if agents believe that the government can undertake such an intervention again - the threat of future demonetizations - the marginal cost of holding cash increases. Either the probability of losing the cash balances, η increases or they would either start paying on cash transactions too, so cost of using cash gets augmented by τ. For both these cases, there would be an increase in the region with digital currency/no-shadow, and depending on the size of the change, there can be a change in equilibrium.

It is worth pointing out that the model only captures the effects of demonetization on the transactions side of shadow economy, and not on its stock, which can take the form of real estate investments, and a consequent drop in real estate activity. We also do not capture income tax evasion, but only sales tax.

6 Quantitative

We now calibrate the model to a quarter using Indian and US data - the former helps shed light on demonetization and the latter as a check as similar models have been calibrated to US data.

We assume some functional forms, as is standard in the literature. For the DM goods market, assume a generalized version of the standard constant relative risk aversion preferences, as $u(q) = [(q+b)^{(1-\sigma)} - b^{(1-\sigma)}]/(1-\sigma)$ as in Lagos and Wrigt (2005). This forces $u(0) = 0$; which does not matter much as we set $b \approx 0$ and $u'(q^*) = 1$. Next, assume that the CM goods market utility, $U(x) = Alog(x)$, which implies that $x^* = A$, so CM output is captured by A.

The parameters (σ, α, B) typically are calibrated to match the relationship between money demand (M/pY) and ι in data. We set $\sigma = 2$ following macro literature and pick α, B to match the average M/pY and its empirical elasticity with ι. In the model this is:

$$\frac{M/p}{Y} = \frac{\pi_l q_l + \pi_h q_h}{\alpha(\pi_l q_l + \pi_h \alpha q_h) + A},$$

where q_l, q_h are functions of ι from (9). To proceed we also need ϵ and π_l. We use the consumption distribution data to back out ϵ and π_l. We set $\pi_l = 0.5$, and so divide the distributions into equal halves and get ratio between q_l and q_h, which is then used to obtain

$\epsilon = 0.18$ for US and $\epsilon = 0.13$ for India. For US, the first to fifth decile's annual consumption was $33,553 and the sixth to tenth was $81,092. For India the corresponding numbers are Rs. 5,036 and Rs. 14,808 respectively. So, $q_l/q_h = 0.41$ for US and 0.34 for India.

The remaining parameters are: (τ, η, κ). The average sales tax for the US is 6% and for India 18%. Given Lemma 3, we set $\eta = 0.3$. We do not need the exact value of κ for the obtaining α, B but we assume that the small transactions take place in cash and the high in digital, as is evidenced by Wang and Wolman (2016) where they find that the fraction of cash payments declines with transaction size. Given this assumption, we can later back out κ/pY which is between $[3.4\%, 8.5\%]$ for US and $[8.4\%, 21.4\%]$ for India, which seems somewhat reasonable.

Results

The objective of the quantitative exercise is to get some quantifiable predictions for the shadow economies in India and US and the consequent policy predictions. The first check of the model is by the size of shadow economy it generates. Given that both the economies are in partial shadow, India has 2.8% of GDP in the shadow while US has 1.2%. The numbers though small in magnitudes, indicates the right relative shares.[12] Schneider et al. (2010) estimate India's shadow economy at 22.2% and US at 8.6% for 1997-2007.

The equilibrium regions for India and US are given in Figures 3 and 4. There is a region where the partial equilibrium exists uniquely. This leads to the conclusion that demonetisation would be ineffective as a policy to switch equilibria unless the economy is in the region where partial and full shadow co-exists and the behavioral implications are substantial such as the effective η reduces significantly.

The other implication from the quantitative exercise is that, having a partial shadow is better than a no shadow equilibrium for tax revenue assuming that κ is at its lower range for both the economies as shown in Figure **??**. This is primarily because ϵ is small enough as characterized in Proposition 3 and the size of the shadow is so small that it's better to charge a higher tax rate than to lower it.

[12]It relies a lot on the consumption distribution, so a richer distribution structure might give a better approximation. Moreover, the estimated size of shadow is the part owing to digital, other factors can explain the rest.

Finally, if we conduct the counterfactual poiicy exercise from Section 5 for the calibrated version of the Indian economy as characterized in Lemma 5 we get that the least costly digitization gives a slightly higher revenue boost (by 0.121) than the minimum reduction in taxes (by 0.117) to move from no-shadow to partial shadow.

Alternative Specifications

We can estimate the model with different costs of cash for India and US: India higher as more theft, foregone interest on savings account but in US tax enforcement is better, so cash is costlier. Use a richer consumption distribution, as the current results (of uniqueness, partial better for revenue) relies on low ϵ.

7 Conclusion

In this paper, we built a model where a shadow economy could arise endogenously depending on the trade-off between two means of payment: cash and digital. We then compare three policy alternatives with the objective of revenue maximization. This helped us shed light on the recent demonetization episode in India when 86% of currency in circulation was demonetized to be replaced with new notes. The stated objective of this policy at least initially was to penalize the shadow economy which does not pay its fair share in taxes and presumably hoards cash. Since people who are part of the shadow economy will be unable to change their currency for the new notes as it is black money they will be penalized and this would reduce the size of the shadow economy. Alternatively, (or in conjunction) a move to digitization of currency i.e. a move to electronic payment instruments such as debit card and credit card also leads to a lower shadow economy as transactions are recorded and taxes can be collected.

We can also extend the model by including a banking sector to talk about monetary policy issues surrounding interest on reserves and negative nominal rates amidst a zero lower bound environment. We can also include a richer denomination structure and income taxes to enrich the analysis. Also, the seller's side can be made more active: but will not change the analytical results but can help in targeting an additional variable i.e. mark-up by producers to get seller's bargaining power - there might be reason to believe that surplus to sellers is different in low value vs. high value transactions - need micro studies. Also, we can include the probability of collecting taxes, and make the relationship between digitization and shadow economy less isomorphic. And, we can include a richer distribution of types to match the data better. The model can be extended to do a richer sectoral analysis as different sectors have differential tax rates and subsequently the size of the shadow will vary. This might also help analyze the interaction of demonetization as a policy with the new Goods and Services tax in India.

It is also worthwhile mentioning that the model abstracts away from political economy aspect of the problems and its distributional implications. We can also extend the model to include negative congestion versus positive network externalities on cost of going digital, i.e. have convex cost function associated with digital currency. We leave this and other extensions for future work.

References

David, B., Abel, F., & Patrick, W. (2016). Debit card and demand for cash. *Journal of Banking and Finance*, *73*, 55–66. Retrieved from http://dx.doi.org/10.1016/j.jbankfin.2016.08.009 doi: 10.1016/j.jbankfin.2016.08.009

Gomis-Porqueras, P., Peralta-Alva, A., & Waller, C. J. (2014). The shadow economy as an equilibrium outcome. *Journal of Economic Dynamics and Control*, *41*, 1–19. Retrieved from http://dx.doi.org/10.1016/j.jedc.2014.02.006 doi: 10.1016/j.jedc.2014.02.006

Immordino, G., & Russo, F. F. (2017). Cashless payments and tax evasion. *European Journal of Political Economy*(June), 1–8. Retrieved from https://doi.org/10.1016/j.ejpoleco.2017.11.001 doi: 10.1016/j.ejpoleco.2017.11.001

Kim, Y. S., & Lee, M. (2010). A model of debit card as a means of payment. *Journal of Economic Dynamics and Control*, *34*(8), 1359–1368. Retrieved from http://dx.doi.org/10.1016/j.jedc.2010.03.003 doi: 10.1016/j.jedc.2010.03.003

Koreshkova, T. A. (2006). A quantitative analysis of inflation as a tax on the underground economy. *Journal of Monetary Economics*, *53*(4), 773–796. doi: 10.1016/j.jmoneco.2005.02.009

Lagos, R. (2010). Asset prices and liquidity in an exchange economy. *Journal of Monetary Economics*, *57*(8), 913–930. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0304393210001285 doi: 10.1016/j.jmoneco.2010.10.006

Lagos, R., & Wright, R. (2005). A Unified Framework for Monetary Theory and Policy Analysis. *Journal of Political Economy*, *113*(3), 463–484.

Li, Y. (2011). Currency and checking deposits as means of payment. *Review of Economic Dynamics*, *14*(2), 403–417. Retrieved from http://dx.doi.org/10.1016/j.red.2009.11.002 doi: 10.1016/j.red.2009.11.002

Loayza, N. V. (1996). The economics of the informal sector: a simple model and some empirical evidence from Latin America. *Carnegie-Rochester Conference Series on Public Policy*, *45*, 129–162. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167223196000218 doi: 10.1016/S0167-2231(96)00021-8

Nosal, E., & Rocheteau, G. (2011). *Money, Payments and Liquidity.* MIT Press.

Rocheteau, G., & Wright, R. (2005). Money in search equilibrium, in competitive equilibrium, and in competitive search equilibrium b. *Econometrica*, *73*(1), 175–202.

Rogoff, K. S. (2014). Costs and Benefits to Phasing out Paper Currency. *NBER Working Paper*(1), 1–5. doi: 10.1007/s13398-014-0173-7.2

Wang, Z., & Wolman, A. L. (2016). Payment choice and currency use: Insights from two billion retail transactions. *Journal of Monetary Economics*, *84*, 94–115. Retrieved from http://dx.doi.org/10.1016/j.jmoneco.2016.10.005 doi: 10.1016/j.jmoneco.2016.10.005

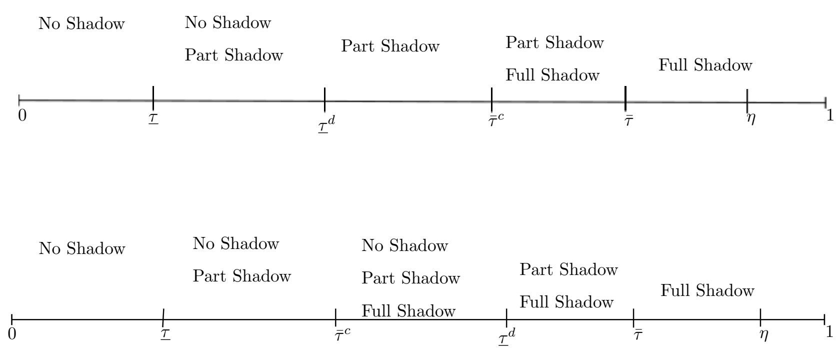

Figure 1: Equilibrium regions for different τ

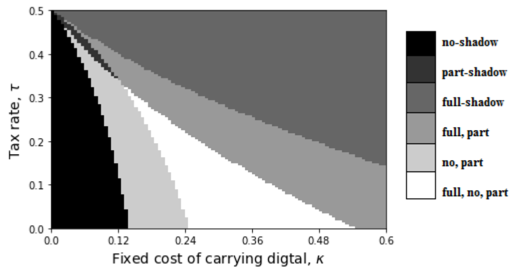

Figure 2: Equilibrium regions: Example

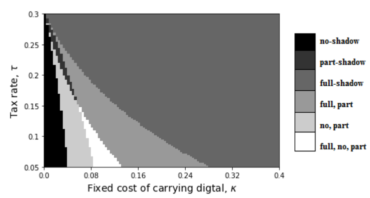

Figure 3: Equilibrium regions: India

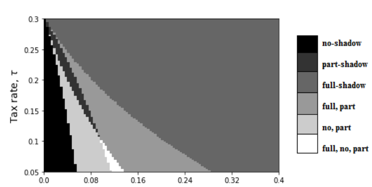

Figure 4: Equilibrium regions: US