The Economics of Digital Transactions

Shilpa Aggarwal* Syed Anas Ahmed† Shashwat Alok‡

Bhagwan Chowdhry§

April 2026

Preliminary draft. Comments welcome; please do not cite without permission.

Abstract

Are digital transactions cheaper than cash? Using India's Unified Payments Interface as our laboratory, we show that digital payments dominate cash on cost grounds at virtually all transaction sizes once fixed infrastructure is amortized. A modified cash-in-advance model formalizes how rising digital adoption increases velocity—requiring the central bank to tighten money supply beyond its inflation target by a computable correction factor. Digitization further expands credit access through verifiable digital income records and generates large time savings from eliminating physical cash management. Policy implications for fee design, cross-subsidization, and currency issuance follow directly.

JEL Classification: E41, E42, E52, G28, O33

Keywords: digital payments, cash-in-advance, velocity of money, financial inclusion, UPI, merchant discount rate

*Indian School of Business. Email: Shilpa_Aggarwal@isb.edu.

†Indian School of Business. Email: Syed_Ahmed@isb.edu.

‡Indian School of Business. Email: Shashwat_Alok@isb.edu.

§Indian School of Business. Email: bhagwan@isb.edu. Corresponding author.

1 Introduction

In April 2016, a street-side *pan* vendor in Varanasi was most likely paid in cash—as were the great majority of small vendors at the time. Digital options existed: cards, bank transfers, and mobile wallets like Paytm were already in use, and some street vendors had begun to accept them. But for most low-value sellers, acceptance remained the exception rather than the norm, because the available instruments were rarely cheap or frictionless enough for a five-rupee transaction. What the cash that still dominated his trade did not produce was a record: nothing that built a credit history or made him legible to a lender. By 2023, that same vendor—and tens of millions of merchants like him across India, with roughly 6.5 crore (65 million) merchants served by some 56.86 crore deployed QR codes (approximately 8.7 per merchant) by FY2024–25 [Ministry of Finance, Government of India, 2025]—had a QR code taped to the corner of his stall. The Unified Payments Interface, launched by the National Payments Corporation of India in April 2016, has grown into India's principal retail payment channel, used by hundreds of millions of people. By the time of writing, India processes over 640 million UPI transactions per day and accounts for close to half of the world's real-time payment transactions by volume [International Monetary Fund, 2025, ACI Worldwide, 2024]—a scale that reflects both rapid adoption and the size of India's population and its high frequency of low-value transfers. Over the same period, the volume and reach of digital payments in India grew at an extraordinary pace. What did it cost? And what did it save?

These are not merely accounting questions. Cash is expensive—its costs are simply hidden, distributed so widely across the sovereign, the banking system, and the household that no single actor ever sees the full bill. Estimates of the social cost of cash for advanced economies range between 0.2 and 0.7 percent of GDP [Bech et al., 2018]; for India, the cost of printing and distributing currency through the ATM network alone reaches roughly 0.2 percent of GDP [Digital Planet, Tufts University, 2014], and that figure excludes storage, security, counterfeit detection, and the hours millions of Indians spend each year standing

in lines and counting change. Digital transactions are not free either. They require fixed infrastructure, cybersecurity systems, and regulatory compliance—costs that are heavily scale-dependent and whose distribution across transaction sizes has direct implications for fee design and financial inclusion.

This paper provides a unified framework for comparing the full costs of cash and digital payments, and derives the policy implications that follow. We ask three questions. First, what are the direct unit costs of a digital transaction relative to a cash transaction, decomposed by transaction size and counterparty type? Second, how does rising digital adoption alter the demand for real money balances, the velocity of circulation, and therefore the optimal conduct of monetary policy? Third, what are the indirect benefits of digitization—through credit access, monetary policy transmission, and time savings—that never appear in direct cost comparisons but may dwarf the direct effects in welfare terms?

On the first question, we show that the cost structure of digital payments is sharply two-part: a small fixed infrastructure cost of approximately Rs. 0.01 per transaction, plus a proportional compliance and cybersecurity load of roughly 0.1 percent of transaction value. Cash, by contrast, imposes costs that are predominantly proportional—handling, storage, and security costs scale with value—with a significant fixed component in the form of ATM infrastructure. The crossover point at which digital payments become cheaper than cash occurs at surprisingly modest transaction values, but does not reach zero-cost micro-transactions without cross-subsidization. The current Indian policy of zero merchant discount rate (MDR) on BHIM-UPI and RuPay debit card transactions—mandated by the Finance Act 2019 (Section 10A, Payment and Settlement Systems Act, 2007) effective January 2020—is not a market outcome but a deliberate welfare intervention. To partially compensate the ecosystem for operating under this mandate, the government runs a series of incentive schemes (Rs. 1,300 crore for FY2021–22, Rs. 2,600 crore for FY2022–23, and Rs. 1,500 crore for FY2024–25; Government of India, 2021, 2023, 2024) that reimburse acquiring banks for low-value person-to-merchant transactions up to Rs. 2,000 at small merchants—a targeted subsidy, not full

cost recovery. The logic of this intervention is precisely what our framework makes explicit.

On the second question, we develop a modified cash-in-advance model [Lucas and Stokey, 1987] in which a parameter λ_t captures the share of the consumption basket purchasable through digital means without holding cash in advance. In equilibrium, velocity equals $1/(1-\lambda_t)$—rising and convex in digital adoption—and the optimal money growth rate must be adjusted downward relative to the inflation target by the factor $(1-\lambda_{t+1})/(1-\lambda_t)$. At current Indian adoption rates of roughly $\lambda \approx 0.35$, this correction is modest. But if UPI adoption reaches $\lambda = 0.60$ over the next five years—a plausible trajectory given historical growth—velocity would rise 62 percent, and the optimal money-growth path along this transition lies roughly 8–12 percentage points per year below the naïve inflation-target prescription (see Section 2). This is a monetary policy puzzle hiding in plain sight inside India's digital revolution, and it has received almost no attention.

On the third question, the indirect benefits of digitization are large and, we argue, ultimately more consequential than the direct transaction cost savings. As digital payment volumes grow, they generate transaction histories that lenders can use to assess creditworthiness—particularly for households and small businesses that lack conventional credit records. A *chaiwala* who transacts entirely in cash has little verifiable income to show a lender; the same *chaiwala* accepting payments through a Paytm QR code accumulates a transaction record on which fintech lenders and NBFCs are increasingly building underwriting models (see Crouzet et al., 2023 for India-specific evidence on digital payment adoption). The foundational evidence for this mechanism comes from Kenya's mobile-money experience: Jack and Suri [2014] document risk-sharing gains from cheaper transfers, Suri and Jack [2016] estimate long-run poverty and gender effects, and Bharadwaj et al. [2021] show that digital payment histories enable consumption-smoothing credit (M-Shwari) for households previously excluded from formal lending. India's digital payment infrastructure operates at considerably larger scale, and our framework provides the theoretical foundations for analysing how these mechanisms interact with fee design and monetary policy in such a setting.

Our paper connects to several literatures. On monetary theory, we build on Lucas and Stokey [1987] and Kocherlakota [1998], the latter of whom reframes money itself as a memory system—a record of past obligations—and thereby anticipates the digital payments revolution more presciently than almost any subsequent work. Search-theoretic models of money [Lagos and Wright, 2005] provide microfoundations for money demand that are complementary to our CIA approach: in the Lagos-Wright framework, decentralized bilateral trade generates a role for money precisely because imperfect commitment and limited record-keeping make credit infeasible, and the payment technology λ would enter as an improvement in the record-keeping technology that loosens that friction in a measurable share of meetings. We view our CIA framework as a tractable reduced form of this deeper mechanism, better suited to calibration at national scale. On money demand under payment choice, Alvarez and Lippi [2009] analyze inventory models in which households choose optimally between cash and alternative payment instruments—our λ parameter is a reduced-form summary of the equilibrium adoption outcome from such a model.

On payment platform design, Rochet and Tirole [2003] establish the two-sided market foundations for interchange and MDR pricing: platforms face competing incentives from merchants and consumers, and the welfare-optimal fee structure depends on relative elasticities and network externalities. Our two-part tariff prescription in Section 5 is consistent with the Rochet-Tirole framework applied to a setting where NPCI acts as a regulated, not-for-profit platform; we note that the normative case for cross-subsidization is stronger when the platform operator is not a private profit-maximizer, since a regulated not-for-profit platform can set the fee structure to maximize total welfare rather than platform revenue, correcting the distortion that profit-maximizing interchange introduces on the less elastic side of the market.

On mobile money and financial inclusion, the foundational evidence is Jack and Suri [2014] on risk-sharing, Suri and Jack [2016] on long-run welfare, and Bharadwaj et al. [2021] on digital-payment-history-based credit; we extend these complementary strands to the much

larger UPI context and add the monetary policy channel they do not consider. On the cost of cash and the transition to digital, Bech et al. [2018] provide the cross-country benchmarks and Chodorow-Reich et al. [2020] provide the cleanest causal evidence of cash's real costs from India's demonetization experiment. Our contribution is to integrate these strands into a unified cost-benefit framework with explicit policy prescriptions for fee design, currency issuance, and the management of monetary policy during a digital transition.

Section 2 models the velocity of money and derives the monetary policy implications of rising digital adoption. Section 3 estimates direct costs at each node of the payment flow for B2C and B2B transactions, and constructs the crossover analysis. Section 4 quantifies indirect benefits. Section 5 derives normative implications for fee design, cross-subsidization, and central bank currency issuance. Section 6 concludes.

2 Velocity of Money and Monetary Policy

2.1 Why the Medium of Transaction Matters for Macroeconomics

Money is memory [Kocherlakota, 1998]. In a world of perfect record-keeping, where every agent's transaction history is costlessly observable and enforceable, money would be unnecessary—credit and deferred settlement would suffice. Cash exists precisely because memory is imperfect: physical currency is a self-enforcing record of purchasing power that requires no counterparty trust, no credit history, and no digital infrastructure. Its cost is the friction it introduces into the transactions technology. Digital payments reduce that friction—but do not eliminate it, and their effect on monetary equilibrium is anything but neutral.

The classic formalization of this friction is the cash-in-advance constraint of Lucas and Stokey [1987]: you must hold money before you can spend it. This constraint determines how tightly money supply governs consumption and therefore how potent monetary policy is. When the constraint is tight—when all transactions require cash in advance—every unit

of cash that households must hold to satisfy the constraint earns no nominal interest, so the inflation tax falls on the entire required money balance. When the constraint is relaxed—when some transactions can settle digitally without cash held in advance—the welfare cost of inflation falls, the demand for real money balances declines, and velocity rises. India in 2016 was a near-perfect cash-in-advance economy. India today is not. The model below makes this transition precise.

2.2 A Modified Cash-in-Advance Model

The Environment. Consider an infinite-horizon economy with a representative household that maximizes:

$$\sum_{t=0}^{\infty} \beta^t \, u(c_t), \quad \beta \in (0,1), \tag{1}$$

where c_t is consumption and $u(\cdot)$ is strictly concave and twice differentiable. Output is endowed: $y_t = 1$ in every period. The model strips away capital and labor—not because these margins are unimportant, but because we want to isolate the transactions technology mechanism without confounding it with production-side dynamics. The single critical generalization relative to Lucas and Stokey [1987] is the parameter:

$$\lambda_t \in [0,1],$$

the *digital adoption rate*—the share of the consumption basket that households can purchase using digital payments that settle at end-of-period, requiring no cash held at the start of the period. One can think of λ_t as the fraction of transactions flowing through UPI. The remaining $(1-\lambda_t)$ fraction requires physical currency available *before* the period begins.

The Modified Cash-in-Advance Constraint.

$$M_t \geq (1-\lambda_t) \cdot P_t \cdot c_t, \tag{2}$$

where M_t is nominal money balances and P_t is the price level. When $\lambda_t = 0$, this collapses to the standard Lucas-Stokey constraint: all consumption requires cash. When $\lambda_t = 1$, the constraint vanishes entirely. India in 2016 was close to $\lambda = 0$. India today sits at roughly $\lambda \approx 0.35$ for the aggregate economy—higher in urban areas, lower in rural ones—and rising.

The household also faces a standard flow budget constraint:

$$P_t c_t + M_{t+1} \leq P_t y_t + M_t + T_t, \tag{3}$$

where T_t are lump-sum transfers from the government. The government controls money supply through:

$$M_{t+1} = (1 + \mu_t) \cdot M_t, \tag{4}$$

where μ_t is the money growth rate—the central bank's instrument.

Equilibrium. In a competitive equilibrium, goods market clearing requires $c_t = y_t = 1$ in every period. The CIA constraint (2) then pins down real money demand:

$$\frac{M_t}{P_t} = (1 - \lambda_t). \tag{5}$$

Real money balances demanded are proportional to the cash-intensive fraction of the economy. Every percentage point increase in digital adoption directly and proportionally reduces the demand for real money balances.

The quantity equation $M_t V_t = P_t Y_t$ gives equilibrium velocity:

$$V_t = \frac{1}{1 - \lambda_t}. \tag{6}$$

Velocity is strictly increasing and *convex* in digital adoption. At $\lambda = 0$, $V = 1$. At $\lambda = 0.5$, $V = 2$. At $\lambda = 0.9$, $V = 10$. The convexity is not a technical artifact—it is a warning. As an economy approaches full digitization, the velocity effects of incremental adoption become

explosive, and a central bank that is slow to respond faces rapidly accelerating inflation with a fixed money supply.

2.3 Propositions

Proposition 1 (Optimal Money Supply Rule)**.** *Suppose the central bank targets a constant inflation rate π^* and the nominal interest rate satisfies $i > 0$, so the CIA constraint* (2) *binds. The optimal money growth rate is:*

$$1 + \mu_t^* = (1 + \pi^*) \cdot \frac{1 - \lambda_{t+1}}{1 - \lambda_t}. \tag{7}$$

Proof. See Appendix A. □

Intuition. The optimal money growth rate is not simply the inflation target. It must be adjusted downward by the factor $(1 - \lambda_{t+1})/(1 - \lambda_t) < 1$ whenever digital adoption is rising. Whenever UPI adoption increases—as it has, every year since 2016—the RBI must grow the money supply more slowly than the inflation target alone would prescribe. The faster the pace of adoption, the larger the required adjustment. This is the monetary policy analog of the velocity puzzle: not a puzzle at all, but a predictable consequence of a changing transactions technology.

Remark 1 (Relation to Interest-Rate Rules)*.* Proposition 1 is stated in terms of a money-growth rule, consistent with the quantity-theory/CIA skeleton. In practice, the RBI and most central banks operate interest-rate rules (Taylor-type or inflation-forecast targeting). In a New Keynesian extension with sticky prices, the same λ-shift enters as a downward shift in money demand that, for a given real interest rate and output gap, creates excess liquidity unless the policy rate is raised. The implied path is: rising $\lambda \Rightarrow$ lower money demand $\Rightarrow$ upward pressure on prices for fixed $i \Rightarrow$ central bank must raise i (or allow money supply to contract) to keep inflation at π^*. The money-growth correction in equation (7) is thus the quantity-theory analog of a rate increase in the NK environment. A full NK treatment with

a Calvo-pricing block and an explicit IS/Phillips curve is beyond the scope of this paper but would be a natural extension.

Proposition 2 (Welfare Cost of Inflation)**.** *The welfare cost of positive nominal interest rates is proportional to* $(1 - \lambda_t)$*. As* $\lambda_t \to 1$*, the welfare cost of moderate inflation vanishes: digital adoption and the Friedman rule become substitutes.*

Proof. See Appendix A. □

Intuition. The Friedman rule—setting the nominal interest rate to zero by deflating at the real rate—is the classical remedy for the welfare cost of inflation [Friedman, 1969]. Proposition 2 shows that digital adoption offers an alternative path to the same destination: a fully digitized economy is one in which the cash-in-advance constraint no longer binds at any positive interest rate, and the welfare cost of inflation disappears without requiring the central bank to commit to deflation. For a developing economy like India, where credible deflation commitments are institutionally difficult, this is a significant practical implication.

2.4 Fraud as a Wedge on Effective Digital Adoption

The CIA model as developed in Section 2.2 assumes that digital transactions are costless once the infrastructure exists. They are not. Fraud—unauthorized transactions, phishing, social engineering, and identity theft—imposes a real cost on digital payers that has no direct analog in cash transactions. We now extend the model to incorporate fraud risk and derive its implications for optimal adoption, velocity, and the crossover analysis.

Let $\phi_t \in [0, 1]$ denote the *fraud rate*—the expected loss from fraud per unit of transaction value for digital payments.[1] In India, Finance Ministry data places UPI fraud losses at

[1]The formulation $\tilde{\lambda}_t = \lambda_t(1 - \phi_t)$ is a reduced-form device that captures the first-order effect of fraud on effective digital adoption without fully micro-founding the payment choice. A more rigorous treatment would derive λ endogenously from expected-utility maximization under fraud risk, risk aversion, and a salience parameter for perceived versus true ϕ—yielding $\lambda(\phi)$ as a downward-sloping function. We leave this extension, and the implied optimal fraud-deterrence policy, for future work. The current formulation correctly signs and orders all comparative statics; the convexity result in Proposition 3 is robust to the specific functional form.

approximately Rs. 573 crore on total UPI transaction value of roughly Rs. 139 lakh crore in FY2022–23, and Rs. 1,087 crore on a larger base in FY2023–24—implying $\phi_t \approx 4\text{–}5 \times 10^{-5}$, or roughly half a basis point of transaction value. This is the *realized* fraud rate. The *perceived* fraud rate facing households may be higher—a survey by LocalCircles finds that one in five UPI user families has experienced fraud, and 51 percent of victims do not report [LocalCircles, 2025], suggesting the official figures substantially understate the true incidence.

The augmented digital transaction cost becomes:

$$C^D(v) = \underbrace{f}_{\text{fixed infra}} + \underbrace{\alpha \cdot v}_{\text{compliance}} + \underbrace{\phi_t \cdot v}_{\text{fraud loss}}, \tag{8}$$

so that the effective proportional cost of a digital transaction is $\alpha + \phi_t$ rather than α alone. The fraud-adjusted crossover condition becomes:

$$v > v^{**} \equiv \frac{f - \kappa}{\delta - \alpha - \phi_t}. \tag{9}$$

Comparing v^{**} with the fraud-free crossover v^*: since $\phi_t > 0$ reduces the denominator, $v^{**} > v^*$. Fraud shifts the crossover to a higher transaction value—digital remains cheaper than cash at that threshold, but the margin is narrower than the infrastructure cost comparison alone would suggest.

Proposition 3 (Fraud-Adjusted Adoption)**.** *Let* $\tilde{\lambda}_t \equiv \lambda_t(1 - \phi_t)$ *denote the* fraud-adjusted *effective digital adoption rate—the share of transactions that are both digitally settled and free of fraud loss. Then:*

(i) Equilibrium velocity is $V_t = 1/(1 - \tilde{\lambda}_t) < 1/(1 - \lambda_t)$*. Fraud suppresses velocity below the frictionless digital benchmark.*

(ii) The optimal money growth rate becomes $1 + \mu_t^* = (1 + \pi^*) \cdot (1 - \tilde{\lambda}_{t+1})/(1 - \tilde{\lambda}_t)$*. A central bank that uses the raw adoption rate* λ_t *overstates velocity and will undershoot the inflation target during periods of rising digital adoption whenever* $\phi_t < b\phi_{t+1}$ *where*

b refers to the following ratio - $\frac{\lambda_{t+1}(1-\lambda_t)}{\lambda_t(1-\lambda_{t+1})}$.

(iii) Welfare losses from fraud are convex *in ϕ_t, provided that the coefficient on the precautionary reversion term, γ, dominates the offsetting effect arising from velocity distortions. Specifically, convexity holds whenever $\gamma > \frac{3\lambda_t^3}{(1-\lambda_t+\lambda_t\phi_t)^4}$.*

Proof. Part (i): substitute $\tilde{\lambda}_t$ for λ_t in equation (6). Part (ii): substitute into equation (7). Part (iii) follows from the interaction of three effects. The direct fraud loss is linear in ϕ_t, while the velocity channel, operating through the convexity of $1/(1-\tilde{\lambda}_t)$ in effective digital adoption, is concave in ϕ_t and therefore exerts an offsetting effect. By contrast, the precautionary cash-reversion mechanism is convex in ϕ_t and generates second-order amplification. Welfare losses are therefore convex in ϕ_t whenever the precautionary reversion effect dominates the opposing velocity effect (see Appendix A). □

Intuition. Fraud does not merely impose a direct cost on victims—it corrupts the transactions technology itself. A household that has been defrauded once, or that *fears* being defrauded, partially reverts to cash even when digital is cheaper on expected-cost grounds. This precautionary reversion is self-defeating in aggregate: it raises the cash-intensive fraction $(1-\tilde{\lambda}_t)$, lowers velocity, and requires the RBI to expand money supply relative to the fraud-free benchmark. Fraud is, in this sense, a form of monetary policy loosening that the central bank must offset.

2.5 Calibrating the India Moment

India's nominal GDP in FY2024 was approximately Rs. 293 lakh crore.

Operationalizing λ. The parameter λ_t in the model is the share of the *consumption basket* that households can purchase digitally without holding cash in advance. It is not identical to UPI's share of total payment value, which includes P2P transfers, government disbursements, and business-to-business flows—none of which correspond to the c_t in the household's

CIA constraint. To construct λ_t properly, we take NPCI's reported share of UPI in retail payment transaction *value* and remove: (i) pure P2P transfers (approximately 37 percent of UPI volume by count and roughly 71 percent by value, reflecting higher average ticket sizes in peer-to-peer flows; National Payments Corporation of India, 2024); these shares shift year to year as P2M grows faster than P2P, so the figures should be read as representative of the recent (FY2023–24) period rather than as fixed parameters; (ii) government and salary credits; and (iii) card-to-UPI substitution that does not reduce ex-ante cash holdings since the underlying funding account is a debit card. After these exclusions, our baseline estimate for the consumption-basket share is $\lambda \approx 0.35$, with an urban household estimate of $\lambda_H \approx 0.55$–0.60 and a rural estimate of $\lambda_L \approx 0.15$–0.20. The model's velocity and policy-correction results are reported under this baseline. The corresponding monetary aggregate M in equation (5) is M1—currency in circulation plus demand deposits—which is the aggregate most tightly linked to transactions balances in the CIA framework. Results using M2 or reserve money are broadly similar but attenuate the velocity effect because broader aggregates include savings balances not primarily held for transactions purposes.

Under our approximate baseline of $\lambda \approx 0.35$ for the aggregate economy (see Appendix B), equilibrium velocity from equation (6) is $V \approx 1.54$. India's published broad-money (M3) income velocity runs in a comparable 1.5–1.8 range. We flag, however, that the aggregate specified for M above is M1, and India's M1 income velocity is materially higher (of order 4); the comparison should therefore be read as an order-of-magnitude consistency check against broad money, not as a direct match to published M1 velocity.

Now consider a plausible five-year projection. If UPI adoption rises to $\lambda = 0.60$—a trajectory consistent with historical growth rates and NPCI's stated targets for QR code merchant deployment, which reached 6.5 crore merchants by FY2024–25—cumulative velocity rises from $V \approx 1.54$ to $V = 2.5$, a 62 percent increase.

To translate this into an annual policy rule, suppose the RBI operates a money-growth rule targeting $\mu^* = \pi^* = 4$ percent (consistent with the midpoint of its 2–6 percent mandate).

In a stationary economy with no change in λ, this rule is exactly correct. Apply equation (7) year by year along a linear adoption path $\lambda_t \in \{0.35, 0.40, 0.45, 0.50, 0.55, 0.60\}$: the optimal one-period money growth in year t is $1 + \mu_t^* = (1 + \pi^*)(1 - \lambda_{t+1})/(1 - \lambda_t)$, which gives:

Year (t to t+1)	0.35→0.40	0.40→0.45	0.45→0.50	0.50→0.55	0.55→0.60
μ_t^* (%)	−4.0	−4.7	−5.5	−6.4	−7.6

Each year's optimal money growth lies roughly 8–12 percentage points below the naïve $\mu^* = \pi^* = 4$ percent rule, and the gap widens with λ because the velocity function $1/(1 - \lambda)$ is convex [Singh and Gupta, 2025]. This figure is specific to the quantity-theory/CIA framework; in a standard Taylor-rule environment, the same velocity shift would manifest as downward pressure on the neutral rate of interest—a related but distinct adjustment. Section 5 discusses the interest-rate analog. Ignoring this correction risks an unintended inflationary surge as velocity rises faster than anticipated.

2.6 The Distributional Extension

The representative-agent model obscures the question that matters most for welfare: not the average effect of digitization, but its distributional incidence. Poor and rural households—precisely those India most wants to reach—have lower λ than urban households. They bear a disproportionately large share of the inflation tax (roughly 80–85 percent of the per-rupee burden at the paper's rural calibration $\lambda_L = 0.15$–0.20). They stand in the longer ATM queues. They are the ones for whom cash is not a choice but a constraint.

In a two-agent extension with urban households facing λ_H and rural households facing $\lambda_L < \lambda_H$, the welfare cost of positive nominal interest rates falls disproportionately on rural households. Digital adoption, in this model, is not merely a monetary policy variable—it is a distributive justice variable. Every rural household brought onto UPI is a household whose inflation tax burden falls and whose credit access expands. This insight motivates the fee design prescriptions in Section 5.

3 Direct Costs: Cash versus Digital

3.1 The Anatomy of a Digital Transaction

Every UPI transaction passes through ten discrete steps across five entities: the Mobile App, the Payer's PSP (Payment Service Provider), UPI/NPCI, the Remitter Bank, and the Beneficiary Bank. Understanding the cost at each node is essential to any serious analysis of payment fee design—and to answering the question our red-ink collaborators keep asking: does a roadside *pan wala* pay the same transaction cost as a washing machine retailer?

The cost structure is two-part. Fixed infrastructure costs at the NPCI switch layer run approximately Rs. 0.01 per transaction once the central switch investment is amortized over the network's volume—a figure that falls with scale and is already very close to zero at current UPI volumes of 20 billion transactions per month. Indirect costs—cybersecurity monitoring, fraud detection, regulatory compliance, and settlement risk management—scale with transaction value at approximately 0.1 percent. This two-part structure has a direct implication for optimal fee design that we develop in Section 5.

3.2 B2C Transactions: Chips versus Washing Machines

Consider two stylized transactions. Rajan buys a Rs. 20 packet of chips from a roadside vendor. Priya buys a Rs. 25,000 washing machine from an appliance retailer. Both pay via UPI. Table 1 compares the direct merchant cost under alternative payment technologies.

Several observations follow immediately. First, UPI's zero-MDR regime is not a market equilibrium—it is a policy choice, sustained by taxpayer subsidy. The NPCI switch-layer cost of approximately Rs. 0.01 per transaction is only a small fraction of the fully-loaded per-transaction cost once bank-side processing, PSP operations, and acquirer compensation are included. The government incentive schemes (Rs. 1,300 crore for FY2021–22; Rs. 2,600 crore for FY2022–23; Rs. 1,500 crore for FY2024–25; Government of India, 2021, 2023, 2024) defray part of this stack, but only for low-value person-to-merchant transactions up to Rs. 2,000

Table 1: Direct Transaction Cost Comparison: B2C Transactions. *Notes:* MDR figures for debit cards based on RBI circulars and AMLEGALS (2024). UPI MDR was made zero for BHIM-UPI (and RuPay debit) P2M transactions effective January 1, 2020 via Section 10A of the Payment and Settlement Systems Act, 2007 (inserted by the Finance Act 2019); the government supports the ecosystem through low-value P2M incentive schemes (Rs. 1,300 crore for FY2021–22; Rs. 2,600 crore for FY2022–23; Rs. 1,500 crore for FY2024–25), which reimburse acquiring banks for small-merchant P2M transactions up to Rs. 2,000 [Government of India, 2021, 2023, 2024]. Cash costs are qualitative estimates; quantitative estimates require primary data collection (see Section 4).

Payment Method	**Rs. 20 (chips)**	**Rs. 25,000 (washing machine)**
UPI (bank account)	Rs. 0 (zero MDR)	Rs. 0 (zero MDR)
RuPay Debit Card	Rs. 0 (zero MDR)	Rs. 0 (zero MDR)
Visa/Mastercard Debit (small merchant, ≤Rs. 20 lakh turnover)	Rs. 0.08 (0.4%)	—
Visa/Mastercard Debit (large merchant)	—	Rs. 225 (0.9%)
Cash	Float management, small denomination handling cost	Theft risk, cash-in-transit, bank deposit cost, insurance

at small merchants; they cover neither P2P transfers nor high-value P2M, and are paid to acquiring banks for sharing with issuer banks, PSP banks, and app providers rather than as blanket cost recovery. The subsidy is therefore a partial, targeted intervention rather than full per-transaction cost recovery across the network. Second, the *pan wala* and the appliance retailer do not face the same costs under debit card systems: RBI rules mandate that small businesses (turnover up to Rs. 20 lakhs) pay a maximum MDR of 0.4 percent on debit card transactions, capped at Rs. 200 per bill, while medium and large businesses pay 0.9 percent (0.8 percent for QR-based acceptance). Under UPI, both pay zero—but this conflation of micro-merchant and large retailer in a single zero-MDR regime is a design choice with fiscal consequences. Third, for the large washing machine transaction, the debit card MDR would be up to Rs. 225 (0.9 percent for large merchants at physical PoS; 0.8 percent for QR-based acceptance)—a cost that either reduces merchant margins or is passed to consumers. UPI eliminates this cost entirely under the zero-MDR mandate, which helps explain the rapid displacement of debit cards at the point of sale since 2020. Mastercard's CFO Sachin

Mehra characterised India's UPI ecosystem as "an incredibly painful experience for ecosystem participants who all end up losing money" (October 2023). In the language of our framework, zero MDR transfers the surplus that card networks would otherwise capture to merchants and consumers, with the residual ecosystem cost partly defrayed by the government incentive schemes described above and the remainder absorbed by banks, PSPs, and app providers.

3.3 The Cost of Cash: Sovereign and Private

Cash transactions are not free. Their costs are distributed so widely that they appear free to any single actor. We decompose them into three layers. The quantitative figures in this subsection draw on Digital Planet, Tufts University [2014] and are therefore early-2010s baseline estimates; currency-in-circulation and ATM magnitudes have shifted materially since (notably after the 2016 demonetisation and the COVID-19 period), so they are retained for the cost-decomposition logic rather than as current levels (cf. Reserve Bank of India, Annual Report, various years).

The Sovereign's Costs. The RBI incurred an annual cost of approximately Rs. 2,800 crore simply to print currency notes as of FY2009–10 [Digital Planet, Tufts University, 2014]. Adding ATM deployment—each machine costing approximately Rs. 7 lakh to install and an equivalent amount to maintain annually, applied to the roughly 60,000–70,000 ATMs operating in FY2009–10 [Digital Planet, Tufts University, 2014] and amortized over a five-year horizon—brings the total cost of printing and distributing cash through the ATM network alone to roughly 0.2 percent of India's GDP. For perspective, India's currency-to-GDP ratio of 12.2 percent is higher than Russia (11.9 percent), Brazil (4.1 percent), and Mexico (5.7 percent) [Digital Planet, Tufts University, 2014], consistent with—though not by itself sufficient to establish—higher sovereign cash management costs than most peer economies.

The Banking System's Costs. Commercial banks bear the cost of cash vault management, cash-in-transit logistics, counterfeit detection, and ATM replenishment. India's

cash-in-transit industry, despite serving a population of 1.3 billion, employed only 40,000 individuals and 6,000 vehicles as of the mid-2010s—a severe under-provision of physical currency logistics that itself generated queues, shortages, and rationing [Digital Planet, Tufts University, 2014].

The Household's and Merchant's Costs. These are the hardest to measure and the most consequential for welfare. They include: time spent travelling to ATMs and bank branches, time spent counting and reconciling physical currency, theft risk borne by merchants holding end-of-day cash, the cost of safe storage for cash-intensive businesses, and—for rural households—the significant opportunity cost of distance from the nearest ATM. We estimate these time costs in Section 4.2.

3.4 The Crossover: When Does Digital Become Cheaper?

Define the total cost of a digital transaction of value v as:

$$C^D(v) = \underbrace{f}_{\text{fixed infra}} + \underbrace{\alpha \cdot v}_{\text{compliance}},$$

with $f \approx \text{Rs. } 0.01$ and $\alpha \approx 0.001$. Define the total cost of a cash transaction as:

$$C^C(v) = \underbrace{\kappa}_{\text{ATM/vault fixed}} + \underbrace{\delta \cdot v}_{\text{handling/risk}},$$

where κ includes the amortized sovereign cost of ATM infrastructure and δ reflects proportional handling and security costs. Assuming $\delta > \alpha$—cash imposes higher proportional costs than digital, which is empirically uncontroversial in the Indian setting—digital payments are cheaper when $C^D(v) < C^C(v)$, i.e., when:

$$v > v^* \equiv \frac{f - \kappa}{\delta - \alpha}.$$

When $\kappa > f$—when the fixed sovereign cost of cash infrastructure exceeds the fixed digital infrastructure cost—the crossover is immediate: digital is cheaper at all transaction values. Given that India's ATM infrastructure carries an amortized fixed cost per transaction substantially above Rs. 0.01 at current usage levels, this condition is plausibly satisfied. The crossover exists at near-zero transaction values—but only once fixed digital infrastructure is fully deployed and amortized.

This is precisely the policy significance of the PIDF (Payments Infrastructure Development Fund): by subsidizing QR code and POS deployment in Tier-3 to Tier-6 centers—5.45 crore PIDF-funded touch points deployed by October 2025, out of a total UPI acceptance footprint of roughly 6.5 crore merchants—the government is driving down the effective f for digital transactions in underserved areas, pushing the crossover toward zero. The economic logic is sound.

3.5 Fraud: The Hidden Cost of Digital Transactions

The two-part cost function of Section 3 accounts for infrastructure and compliance but omits the cost that is rising fastest and worrying policymakers most: fraud. India's Finance Ministry data reveal that UPI fraud cases jumped 85 percent in FY2023–24 relative to the prior year, reaching 13.42 lakh incidents totaling Rs. 1,087 crore [Ministry of Finance, Government of India, 2024]. The RBI's Annual Report for FY2023–24 records digital payment fraud surging more than fivefold to Rs. 1,457 crore ($175 million) [Reserve Bank of India, 2024a]—a figure that, while still small relative to the Rs. 200 lakh crore in UPI transaction value, is growing far faster than the underlying transaction base.

The fraud rate $\phi_t \approx 4\text{–}5 \times 10^{-5}$ looks small in basis points. But it obscures two features that matter enormously for welfare. First, fraud is not distributed uniformly: social engineering—refund scams, phishing, fake customer support, and QR code manipulation—is the dominant fraud vector [PricewaterhouseCoopers India, 2025], and these tactics disproportionately exploit users who are less digitally literate, including first-time digital payers

in rural and semi-urban areas where, as PwC notes, adoption has outpaced education. The distributional skew of fraud toward newer, less experienced users is precisely what makes it a threat to the inclusion agenda. Second, case-count fraud incidence is severely under-reported: 51 percent of UPI fraud victims file no complaint, according to a survey of over 32,000 UPI users across 365 districts [LocalCircles, 2025]; the case-count adjustment is roughly $2\times$, while the value-weighted adjustment is plausibly smaller if unreported frauds are skewed toward lower per-incident amounts.

How does digital fraud compare with the fraud and theft costs of cash? Cash is not fraud-free either. Physical theft, counterfeit notes, and robbery of cash-in-transit impose real costs—but these costs are predominantly proportional to the value of cash held, and they fall on the physical holder, not on a remote victim who can be defrauded without presence. Digital fraud introduces a qualitatively new risk: the attacker need not be physically proximate to the victim, enabling large-scale, geographically dispersed fraud operations of a kind that cash theft cannot support. This "non-proximity" property of digital fraud is what drives the extraordinary 85 percent year-on-year growth in UPI fraud cases even as the underlying payment infrastructure improves.

Table 2 augments Table 1 with the expected fraud cost per transaction, using $\phi_t = 5\times10^{-5}$ (the upper bound of the FY2023–24 range) and noting that cash fraud costs are primarily a fixed risk per cash-holding episode rather than a per-transaction charge.

Three observations stand out. First, at $\phi_t = 5 \times 10^{-5}$, UPI's fraud cost is negligible relative to infrastructure and compliance: the fraud-adjusted digital cost of roughly Rs. 26 on a Rs. 25,000 transaction remains substantially below the debit card MDR of Rs. 225. Digital's cost advantage is robust to the realized fraud rate. Second, for the Rs. 20 micro-transaction the expected fraud loss (Rs. 0.001) is one order of magnitude smaller than the fixed infrastructure cost (Rs. 0.01), so on *realized* fraud grounds infrastructure investment dominates fraud prevention investment for micro-payments. Third, the welfare-relevant fraud cost is not the realized rate but the *perceived* rate that drives precautionary cash-reversion;

Table 2: Transaction Costs Including Expected Fraud, B2C Transactions. *Notes:* Fraud cost for UPI computed as $\phi_t \cdot v$ with $\phi_t = 5 \times 10^{-5}$ (upper bound from Finance Ministry FY2023–24 data). True case-count incidence is likely higher due to under-reporting (estimated at 51% of cases); the value-weighted under-reporting correction is plausibly smaller because unreported frauds are skewed toward lower per-incident amounts. Cash fraud/theft cost is qualitative; primarily affects merchants holding end-of-day float rather than per-transaction basis.

Payment Method	**Rs. 20 transaction**	**Rs. 25,000 transaction**
UPI: infrastructure + compliance	Rs. 0.01 + Rs. 0.02	Rs. 0.01 + Rs. 25
UPI: expected fraud cost ($\phi_t \cdot v$)	Rs. 0.001	Rs. 1.25
UPI: total expected cost	**Rs. 0.03**	**Rs. 26**
Visa/Mastercard Debit (0.9% MDR, large merchant)	Rs. 0.18	Rs. 225
Cash: handling + sovereign costs	Rs. 0.05–0.15 (low per-transaction theft/counterfeit risk)	Rs. 50–200 (moderate per-transaction theft/counterfeit risk)

under-reporting and survey evidence (51 percent non-reporting; one in five families reporting a fraud experience) suggest the perceived rate is materially higher than the realized one. The case for fraud-prevention investment rests on the perception channel and on the distributional skew of victims toward first-time users (Section 5), not on the realized-rate arithmetic.

4 Indirect Benefits of Digitization

4.1 Aggregate Savings

The two-part cost function of Section 3 implies that the aggregate direct savings from UPI—the difference between what the cumulative volume transacted through UPI would have cost under cash and card alternatives and what it cost through UPI—are positive and potentially large, since UPI's proportional cost ($\alpha \approx 0.1$ percent) is an order of magnitude below the merchant discount rates on debit cards (0.4–0.9 percent at physical PoS, as regulated by RBI; QR-based acceptance carries rates 10 basis points lower, at 0.3 percent and 0.8 percent respectively) and below plausible estimates of the all-in social cost of cash. Produc-

ing a *credible point estimate*, however, requires transaction-level cost data and a defensible counterfactual payment mix that we do not attempt to pin down here: it depends on the size distribution of UPI transactions, the share of UPI volume that is genuinely incremental rather than substituting for pre-existing digital payments, and the social cost of cash, none of which we can resolve without primary data. We therefore treat the magnitude of aggregate direct savings as an open empirical question and leave its quantification—ideally applying the cost functions of Section 3 to NPCI transaction-level data under explicit counterfactual scenarios with uncertainty bounds—to future work. The indirect benefits analyzed below—credit access, time savings, and formalization—are, in our reading, likely to dwarf the direct transaction cost savings in welfare terms regardless of where the direct figure ultimately settles.

4.2 The Value of Time

The most underappreciated cost of cash is time. Every bank branch visit, every ATM queue, every minute spent counting change and reconciling the float at end-of-day is real labor diverted from productive activity. In India—where the informal economy employs roughly 90 percent of the labor market and accounts for approximately 40 percent of GDP [Digital Planet, Tufts University, 2014]—these time costs fall disproportionately on the poor and the self-employed, who have the fewest opportunities to delegate cash management.

A conservative estimate proceeds as follows. As of the mid-2010s, the value of ATM transactions in India had grown to approximately Rs. 18 lakh crore annually [Digital Planet, Tufts University, 2014]; at typical withdrawal sizes of Rs. 3,000–5,000, this implies several billion ATM visits per year. If each visit requires an average of 30 minutes of travel and transaction time, and if we value that time at the median informal sector daily wage of roughly Rs. 200 (NSSO 68th round, 2011–12), or about Rs. 25–30 per hour assuming an eight-hour day, then the aggregate time cost of ATM usage alone still runs into several thousand crore rupees annually (of order Rs. 4,500–9,000 crore at these parameters)—before

accounting for bank branch visits, in-store cash handling, and end-of-day reconciliation. A full bottom-up estimate, using CMIE Consumer Pyramids data on ATM visit frequency and wages by income quintile, remains a priority for future empirical work.

4.3 Access to Credit: Two Channels

Digitization expands credit access through two mechanisms that are analytically distinct and empirically separable.

The Deposit Channel. When small savers shift from cash under the mattress to UPI-linked bank accounts, loanable funds in the formal banking system increase. The magnitude is primarily driven by *small* savers—large savers already hold their wealth in deposits. India's M0-to-M2 ratio of 51 percent is dramatically higher than Egypt (29.3 percent), South Africa (8.9 percent), and Mexico (8.7 percent) [Digital Planet, Tufts University, 2014], indicating a large pool of cash that could potentially be intermediated into credit if digitization draws it into the banking system. Demonetization's legacy supports this channel: 99.3 percent of demonetized notes returned to the banking system, temporarily channeling previously hoarded cash into deposits—providing a quasi-experimental window into the deposit channel's magnitude.

The Creditworthiness Channel. Perhaps more consequential than the deposit channel is the information channel. A household that transacts digitally generates a verifiable income record. A street vendor in Surat who conducted most of her business in cash before 2019 had little formal documentation to show a lender—limited credit history, no systematic income record, and few assets to pledge. After two years of regular digital transactions, including through UPI, her daily revenue pattern, her seasonal cash flows, and her payment regularity become part of a digital trail that fintech companies and NBFCs are exploring as an input for underwriting borrowers who lack conventional credit records. India-specific evidence on related mechanisms is emerging: Crouzet et al. [2023], using data from a large Indian

electronic-wallet provider around the 2016 demonetization, document that the cash shock caused a persistent shift to digital payments—precisely the kind of adoption that generates the transaction records lenders can use. Chodorow-Reich et al. [2020] complement this by documenting the large real-output costs of the cash contraction itself, reinforcing that cash dependence carries measurable economic costs. The digital footprint is the collateral—a record that a purely cash-based livelihood does not generate. The underwriting mechanism here is a property of digital payment histories in general rather than of UPI specifically: mobile-wallet and card histories generated comparable footprints before UPI, and the Kenyan parallel discussed below operated on a mobile-money wallet. What the government's zero-MDR mandate and sustained push for UPI adoption have done is dramatically expand the volume of digital transactions, creating the conditions for this footprint to emerge at mass scale.

The foundational evidence for this mechanism comes from Kenya's mobile-money experience. Jack and Suri [2014] document that M-PESA reduces transfer frictions enough to enable geographically dispersed kin to smooth consumption against idiosyncratic shocks; Bharadwaj et al. [2021] show that digital-payment histories generated by M-PESA were subsequently used by lenders (M-Shwari) to extend short-term credit at scale to households with no formal banking record, and that take-up materially improved household resilience. India's growing digital payment infrastructure creates the conditions for the same mechanism to operate at considerably larger scale, though rigorous evidence on its realised magnitude remains limited. The institutional infrastructure is evolving to support this channel: as of September 2023, the RBI expanded UPI to include pre-sanctioned credit lines from scheduled commercial banks as a funding account, a step toward integrating payments and credit access on a single rail.

4.4 Enabling Micro-Transactions and Business Formalization

The zero-MDR policy on UPI and RuPay transactions is what makes micro-transactions economically viable at scale. Before this policy, a Rs. 5 digital merchant payment was uneconomical on card rails—the MDR alone would have consumed the transaction value, and POS terminal costs rendered most micro-merchants ineligible for acceptance. It was not, however, impossible to transact digitally at small values: NPCI's Immediate Payment Service (IMPS), launched in November 2010, already provided instant, 24×7 interbank transfers including person-to-merchant remittances [National Payments Corporation of India, 2010, World Bank, 2021], and closed-loop mobile wallets (Paytm, Mobikwik, Oxigen) were enabling small-value digital payments from around 2014–15. What these earlier options lacked was interoperability across providers. UPI addressed this, but it is the government's zero-MDR mandate—eliminating merchant costs entirely—that changed the economics of micro-payments and drove mass merchant adoption.

The macro-economic implication extends beyond payment convenience. Micro-transactions support market structures—pay-per-use content, micro-insurance premia collected daily, small-lot agricultural commodity trades, toll collection at rural road crossings—that existed previously in cash or via closed-loop wallets but at limited scale, and that interoperable, zero-merchant-cost digital rails have the potential to scale significantly. Low-value digital payments were already possible through prepaid wallets, which stored balances locally and required minimal connectivity; UPI Lite—on-device, additional-factor-authentication-free payments up to Rs. 1,000 per transaction (raised from Rs. 500 in December 2024)—extends similar functionality within the UPI ecosystem, while fully connectivity-free payments via UPI Lite X / NFC remain at an early stage of deployment [Reserve Bank of India, 2024b].

Formalization is a potential second-order effect. A merchant who previously conducted most business in cash—whether below the tax threshold or simply outside the formal system—generates, when transacting digitally, a trail in the banking system that is in principle visible to tax authorities. The growth of digital transactions across all channels—UPI, cards, wal-

lets, and bank transfers—may over time contribute to expanding the formal tax base, though the magnitude of this effect remains an active empirical question.

5 Policy Implications

5.1 Fee Design: The Case for a Two-Part Tariff

Should payment platforms charge a fixed fee per transaction, a proportional fee based on transaction value, or some combination? The answer follows directly from the cost structure established in Section 3. Digital payment costs are genuinely two-part—fixed infrastructure plus proportional compliance—and therefore the fee structure should be two-part as well:

$$\text{Merchant fee}(v) = f + \alpha \cdot v,$$

where f covers fixed NPCI infrastructure costs and $\alpha \cdot v$ covers proportional compliance and cybersecurity costs. This design is not merely analytically appealing—it is welfare-superior to either a pure fixed fee (which penalizes small merchants for micro-transactions) or a pure proportional fee (which under-recovers infrastructure costs on high-value transactions).

The critical design question is cross-subsidization. Given that micro-transactions—the Rs. 20 chips purchase—have a ratio of fixed infrastructure cost to transaction value that is extremely high, a break-even fee would be prohibitive and would price the *pan wala* out of the digital economy entirely. The two-part tariff should therefore incorporate an explicit cross-subsidy: large B2B and high-value B2C transactions pay above their marginal compliance cost, generating surplus that finances near-zero pricing for micro-transactions. This is not a subsidy in any pejorative sense—it is the standard Ramsey pricing solution to a natural monopoly with heterogeneous transaction sizes.

5.2 How Much Currency Should the RBI Print?

From Proposition 1, the answer is: less, and decreasingly so as UPI adoption rises. The optimal currency in circulation satisfies equation (5): real money balances should equal $(1 - \lambda_t)$ times nominal output. As λ_t rises from 0.35 today toward 0.60 over five years, the RBI's target for currency in circulation should fall by approximately $(0.60 - 0.35)/(1 - 0.35) \approx 38$ percent in real terms—a large reduction that must be managed gradually to avoid disruptive cash shortages in rural areas where λ remains low.

This creates an important distributional constraint on currency issuance policy. The optimal aggregate target may call for monetary contraction that outpaces rural digitization, temporarily leaving rural households cash-short. The two-agent extension of Section 2 formalizes this tension: the RBI should target a distribution-adjusted digital adoption rate $\bar{\lambda}_t$ that weights rural and urban adoption by their respective consumption shares, and should phase currency reduction to track $\bar{\lambda}_t$ rather than the urban-weighted aggregate.

5.3 Fraud Liability, Detection Investment, and the Case for Zero-MDR

Fraud creates a policy design problem that the MDR debate has not adequately addressed. Under the current zero-MDR regime, banks and NPCI bear infrastructure costs but recover nothing from merchants on a per-transaction basis—making it financially difficult to invest in the fraud monitoring systems that would reduce ϕ_t. This creates a perverse dynamic: the policy that maximizes adoption (zero MDR) also minimizes the revenue available for fraud prevention, which in turn raises ϕ_t and depresses adoption among risk-averse users. The optimal policy must internalize this feedback.

Three principles follow from Proposition 3. First, *liability should sit with the party best positioned to reduce fraud risk*—typically the bank or PSP, not the end user. The RBI's current zero-liability policy (full refund if fraud is reported within three days) is welfare-improving

in principle, but its impact is blunted by the 51 percent non-reporting rate. Streamlining complaint mechanisms—the RBI's proposed Digital Payments Intelligence Platform and MuleHunter.AI initiative are steps in this direction—raises the effective coverage of the zero-liability guarantee without raising its fiscal cost.

Second, *fraud prevention investment has public good properties* that justify public subsidy. NPCI's AI/ML-based real-time fraud monitoring benefits all participants on the network simultaneously—a classic positive externality. Private banks will under-invest in network-level fraud monitoring relative to the social optimum. The case for NPCI's centralized fraud registry (CPFIR) and its mandatory reporting requirements rests precisely on this externality.

Third, *the distributional incidence of fraud reverses the distributional promise of digitization* unless actively corrected. If fraud disproportionately affects first-time digital payers—often rural, older, or financially inexperienced users, who are most susceptible to the social engineering tactics that dominate UPI fraud—then the expansion of UPI to underserved populations is simultaneously expanding the pool of the most vulnerable fraud targets. A financial inclusion agenda that brings rural households onto digital payments without investing in commensurate fraud protection is not merely incomplete: it is potentially harmful. The PIDF subsidy framework should be extended beyond QR code and POS deployment to fund fraud literacy campaigns and localized customer support in Tier-3 to Tier-6 centers.

5.4 The CBDC Question

India's CBDC experiment—the e-Rupee, launched by the RBI in 2022—raises the question of whether a central bank digital currency is superior to, complementary to, or in tension with UPI as a payments architecture. Our framework suggests the following. On cost grounds, Khiaonarong and Humphrey [2019] note that digital cash would markedly reduce the expense of printing, storage, and distribution of physical currency; the BIS Finternet vision [Carstens and Nilekani, 2024] sketches a longer-run architecture in which tokenized

money and assets settle on shared programmable ledgers. On monetary policy grounds, a CBDC that earns interest introduces a new instrument—the rate of return on digital money—that can substitute for the inflation tax as a tool for managing money demand. On financial stability grounds, a CBDC that competes with bank deposits for household savings creates disintermediation risk that must be weighed against the credit access benefits documented in Section 4.

On balance, the comparison between UPI and a CBDC depends on considerations that our framework—which does not formally model central bank digital currency—can illuminate but not resolve. UPI's bank-account-linked design, operating under the government's zero-MDR mandate, already delivers low-cost digital payments at scale; the incremental cost benefit of a CBDC in this environment may be modest. However, a CBDC offers properties that UPI does not: it is a direct central bank liability rather than a claim intermediated through commercial banks, it could in principle reach users without bank accounts, and programmable money—conditional government transfers, auto-executing subsidy rules [Uña et al., 2023]—represents functionality beyond what the current UPI architecture supports. Whether these additional properties justify the cost and complexity of a parallel system, and how a CBDC would interact with the existing banking and payments infrastructure, remain open questions that warrant careful empirical and institutional analysis rather than a prior judgement based on UPI's current reach.

5.5 Private Currencies: Physical and Digital

Should private currencies—physical scrip or cryptocurrency—be permitted and regulated? Our model provides a clean answer in the two-part framework. Private currencies that compete with the RBI rupee for transactions balances erode the government's seigniorage revenue while potentially reducing transaction costs for users. The net welfare effect depends on the competition between private currency cost efficiency and the loss of monetary policy traction. In a cash-in-advance economy with $\lambda \approx 0.35$, the government still retains significant

seigniorage from the cash-intensive fraction of transactions. Permitting private currencies to displace this fraction imposes fiscal costs that must be weighed against efficiency gains. Our framework does not resolve this tradeoff definitively—but it establishes that the relevant metric is the fraction of total transactions that private currencies would capture and the cost differential relative to UPI, which is already near-zero. The case for permitting private currencies to compete with a zero-cost public payments rail is correspondingly weak.

6 Conclusion

The *pan* vendor with the QR code on his stall is not merely a symbol of India's digital ambition. He is a data point in one of the most significant transitions in payment systems that any large economy has undertaken. Over the past decade, India's daily digital payment volume has grown to over 640 million UPI transactions—and in doing so has generated evidence on questions that monetary economists have theorized about for decades without the data to answer them.

This paper provides a framework for reading that evidence. The cost structure of digital payments is two-part: small fixed infrastructure costs that fall rapidly with scale, plus proportional compliance costs that rise with transaction value. Cash imposes costs that are both fixed—in ATM infrastructure, printing, and distribution—and proportional—in handling, security, and the time of millions of households who must physically manage money. The crossover point at which digital becomes cheaper than cash occurs at surprisingly modest transaction sizes, with the exception of true micro-transactions where cross-subsidization is needed.

Rising digital adoption is not, however, a free monetary policy lunch. Velocity rises—convexly—as the cash-in-advance constraint is relaxed. A central bank that ignores this dynamic and targets inflation using a rule calibrated for a cash economy will find itself repeatedly surprised by inflationary pressure that has nothing to do with the business cycle

and everything to do with the changing transactions technology. Our model provides the correction: the optimal money growth rate must be adjusted downward by the digital adoption factor $(1 - \lambda_{t+1})/(1 - \lambda_t)$ in every period. At current Indian adoption rates, this correction is moderate; five years from now, it may not be.

The indirect benefits of digitization are potentially large. The direct transaction cost savings from digital payments are positive and plausibly substantial, but—as we argue in Section 4.1—they may prove to be the smaller part of the welfare gain. The credit access channel, the formalization channel, and the time savings channel could together prove significant, and would be concentrated among the households that India's development agenda most seeks to reach. Digital adoption, in our model, is a distributive justice variable: every rural household brought onto digital payments is a household whose inflation tax burden falls, whose potential access to credit improves, and whose economic activities become more visible to the formal economy.

Four policy prescriptions follow with particular clarity. First, the optimal fee structure for digital payment platforms is a two-part tariff with explicit cross-subsidization from high-value to micro-transactions. The current zero-MDR regime achieves this through government subsidy; a more sustainable design would internalize the cross-subsidy within the platform itself. Second, the RBI should calibrate currency issuance targets to a fraud-adjusted, distribution-weighted digital adoption rate $\tilde{\bar{\lambda}}_t$—not the raw urban adoption figure—phasing currency reduction to track realized effective digitization. Third, fraud liability should be assigned to banks and PSPs rather than end users, with mandatory investment in centralized fraud detection infrastructure as a condition of operating on the UPI network; the current zero-liability policy is the right principle, but its 49 percent effective coverage rate must improve dramatically. Fourth, the CBDC experiment raises questions that our framework can illuminate but not resolve. A CBDC offers properties that the current bank-intermediated UPI architecture does not—direct central bank liability, potential access without a bank account, and programmable money for conditional transfers—and these should be weighed

against the costs and complexity of a parallel system, rather than dismissed on the grounds that UPI's current reach is already extensive.

The deeper lesson is this: money has always been memory—a record of who owes what to whom [Kocherlakota, 1998]. But memory can be corrupted. The question is not only whether the ledger is cheaper than paper—it demonstrably is—but whether the ledger is trustworthy enough that the households it is supposed to serve will choose to use it. India has spent seven years demonstrating that the ledger can be cheap. The next seven years must demonstrate that it can also be safe. Those two properties together—not either one alone—are what will determine whether UPI's revolution endures.

References

ACI Worldwide. Prime time for real-time 2024. Technical report, ACI Worldwide / GlobalData, 2024.

Fernando Alvarez and Francesco Lippi. Financial innovation and the transactions demand for cash. *Econometrica*, 77(2):363–402, 2009.

Morten Bech, Umar Faruqui, Frederik Ougaard, and Cristina Picillo. Payments are a-changin' but cash still rules. *BIS Quarterly Review*, pages 67–80, March 2018.

Prashant Bharadwaj, William Jack, and Tavneet Suri. Fintech and household resilience to shocks: Evidence from digital loans in Kenya. *Journal of Development Economics*, 153: 102697, 2021.

Agustín Carstens and Nandan Nilekani. Finternet: The financial system for the future. Working Paper 1178, Bank for International Settlements, 2024.

Gabriel Chodorow-Reich, Gita Gopinath, Prachi Mishra, and Abhinash Narayanan. Cash and the economy: Evidence from India's demonetisation. *Quarterly Journal of Economics*, 135(1):57–103, 2020.

Nicolas Crouzet, Apoorv Gupta, and Filippo Mezzanotti. Shocks and technology adoption: Evidence from electronic payment systems. *Journal of Political Economy*, 131(11):3003–3065, 2023.

Digital Planet, Tufts University. Cost of cash: India. Technical report, Fletcher School, Tufts University, 2014.

Milton Friedman. The optimum quantity of money. *The Optimum Quantity of Money and Other Essays*, 1969.

Government of India. Cabinet approves incentive scheme for promotion of RuPay debit cards and low-value BHIM-UPI transactions (p2m), FY2021–22 (rs. 1,300 crore). Press Information Bureau, December 2021.

Government of India. Cabinet approves incentive scheme for promotion of RuPay debit cards and low-value BHIM-UPI transactions (p2m), FY2022–23 (rs. 2,600 crore). Press Information Bureau, PRID 1890314, January 2023.

Government of India. Cabinet approves incentive scheme for promotion of low-value BHIM-UPI transactions (p2m), FY2024–25 (rs. 1,500 crore). Press Information Bureau, PRID 2112771, 2024.

International Monetary Fund. Growing retail digital payments: The value of interoperability. Technical report, International Monetary Fund, 2025.

William Jack and Tavneet Suri. Risk sharing and transactions costs: Evidence from Kenya's mobile money revolution. *American Economic Review*, 104(1):183–223, 2014.

Tanai Khiaonarong and David Humphrey. Cash use across countries and the demand for central bank digital currency. Working Paper WP/19/46, International Monetary Fund, 2019.

Narayana Kocherlakota. Money is memory. *Journal of Economic Theory*, 81(2):232–251, 1998.

Ricardo Lagos and Randall Wright. A unified framework for monetary theory and policy analysis. *Journal of Political Economy*, 113(3):463–484, 2005.

LocalCircles. One in five UPI user families has experienced fraud. Survey of 32,000 respondents across 365 districts, June 2025.

Robert E. Lucas and Nancy L. Stokey. Money and interest in a cash-in-advance economy. *Econometrica*, 55(3):491–513, 1987.

Ministry of Finance, Government of India. Data on UPI fraud presented in Lok Sabha. Parliamentary response, winter session, November 2024.

Ministry of Finance, Government of India. UPI recognised as world's largest real-time payment system; PIDF and merchant acceptance statistics. Government update, December 2025.

National Payments Corporation of India. Immediate payment service (IMPS): Product overview. NPCI, November 2010.

National Payments Corporation of India. UPI ecosystem statistics: P2P and P2M volume and value split. NPCI, UPI Ecosystem Statistics (FY2023–24), 2024.

PricewaterhouseCoopers India. Combating payments fraud in India's digital payments landscape. Technical report, PwC India, April 2025.

Reserve Bank of India. Annual report FY2023–24. *Reserve Bank of India Publications*, 2024a.

Reserve Bank of India. Enhancing limits for UPI lite. RBI circular, December 2024b.

Jean-Charles Rochet and Jean Tirole. Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4):990–1029, 2003.

Priya Singh and Rahul Gupta. Impact of digital payments on monetary policy transmission in India: Evidence from the post-demonetization era. *Frontiers in Human Dynamics*, 2025.

Tavneet Suri and William Jack. The long-run poverty and gender impacts of mobile money. *Science*, 354(6317):1288–1292, 2016.

Gerardo Uña, Neil Griffin, Anish Verma, and Majid Bazarbash. FinTech payments in public financial management: Benefits and risks. Working Paper WP/23/20, International Monetary Fund, 2023.

World Bank. Fast payments—India: IMPS and UPI case study. Technical report, World Bank Fast Payments Toolkit, 2021.

A Proofs

Proof of Proposition 1. Step 1 (CIA binding condition). When the nominal interest rate $i_t > 0$, holding idle money is costly. Households minimize cash holdings subject to the CIA constraint, so the constraint binds with equality. The Lagrangian on the CIA constraint has multiplier $\mu_t^{\text{CIA}} > 0$ whenever $i_t > 0$; at $i_t = 0$ (the Friedman rule), money is satiated and the constraint may not bind.

Step 2 (Equilibrium money demand). In equilibrium, goods market clearing implies $c_t = y_t = 1$. Substituting into the binding CIA constraint (2):

$$M_t = (1 - \lambda_t)P_t. \tag{A.1}$$

This is the money demand function: real balances $(1 - \lambda_t)$ are the cash-intensive share of the consumption basket.

Step 3 (Optimal money growth). Advancing equation (A.1) one period: $M_{t+1} = (1 - \lambda_{t+1})P_{t+1}$. Taking the ratio and substituting the inflation target $P_{t+1}/P_t = 1 + \pi^*$:

$$\frac{M_{t+1}}{M_t} = \frac{(1 - \lambda_{t+1})}{(1 - \lambda_t)} \cdot (1 + \pi^*).$$

Since $M_{t+1}/M_t \equiv 1 + \mu_t^*$ by the money growth rule (4), this yields equation (7).

Step 4 (Uniqueness). The solution is unique conditional on the inflation target π^* and the adoption path $\{\lambda_t\}$. There is no indeterminacy because output $y_t = 1$ is exogenous and the price level is pinned by the money supply rule together with (A.1). □

Proof of Proposition 2. The welfare cost of positive nominal interest rates in a standard CIA model is the area of the triangle under the money demand curve above the Friedman-rule benchmark [Friedman, 1969]. In the modified model, money demand at interest rate i is:

$$L(i, \lambda_t) = (1 - \lambda_t) \cdot \ell(i),$$

where $\ell(i)$ is the standard money demand function satisfying $\ell'(i) < 0$ and $\ell(0) = 1$. The welfare triangle scales linearly with $(1 - \lambda_t)$, since the demand curve shifts inward by that factor. In the limit $\lambda_t \to 1$, money demand goes to zero at all positive interest rates and the welfare cost of inflation vanishes. □

Proof of Proposition 3, Part (iii) – Convexity. Parts (i) and (ii) follow by direct substitution as noted in the main text. For part (iii), define total welfare loss from fraud as:

$$W^F(\phi_t) = \underbrace{\phi_t \cdot \lambda_t \cdot P_t y_t}_{\text{direct loss}} + \underbrace{\frac{dV}{d\tilde{\lambda}} \cdot \frac{d\tilde{\lambda}}{d\phi_t}}_{\text{velocity distortion}} + \underbrace{\gamma \cdot \phi_t^2}_{\text{precautionary reversion}},$$

where $\gamma > 0$ parameterizes the feedback from perceived fraud risk to precautionary cash-

holding. The direct loss is linear in ϕ_t; the velocity distortion term $dV/d\tilde{\lambda} = 1/(1-\tilde{\lambda})^2 > 0$ is increasing in $\tilde{\lambda}$ and hence in ϕ_t (since $\partial\tilde{\lambda}/\partial\phi_t = -\lambda_t < 0$ implies the distortion grows as $\tilde{\lambda}$ falls toward the cash-intensive boundary); the precautionary term is strictly convex. The sum is therefore convex in ϕ_t. □

B Data Sources and Estimation Details

MDR Figures (Table 1). Merchant discount rates for UPI are per the Payment and Settlement Systems Act amendment effective January 1, 2020. RuPay debit card MDR is per the NPCI/RBI zero-MDR circular. Visa/Mastercard debit card MDR range (0.25–0.4 percent) are per RBI's MDR rationalization circular effective January 2018. Credit card MDR, which is not regulated by RBI and is not directly comparable to an instant account-to-account payment system, is excluded from the comparison. The small-business cap of 0.4 percent at physical PoS (0.3 percent for QR-based acceptance) and large-business rate of 0.9 percent at physical PoS (0.8 percent for QR-based acceptance) for debit transactions are per RBI guidelines on differential MDR by merchant turnover category and acceptance mode. Tables 1 and 2 report the physical PoS rates (0.4 percent and 0.9 percent); QR-based rates are 10 basis points lower (0.3 percent and 0.8 percent respectively).

Cash Cost Figures. The Rs. 2,800 crore annual printing cost and 0.2 percent of GDP estimate for cash distribution are from Digital Planet, Tufts University [2014], based on RBI FY2009–10 data. These are likely lower bounds on current costs given nominal GDP growth since then.

Digital Adoption Rate Calibration. The estimate $\lambda \approx 0.35$ is derived as follows. UPI's share of total retail payment value in India was approximately 35–40 percent as of FY2023–24 (NPCI data). We use the midpoint of this range as our baseline. Urban households likely face $\lambda_H \approx 0.55$–0.65; rural households face $\lambda_L \approx 0.15$–0.25. The aggregate $\lambda \approx 0.35$ reflects

the population-weighted average.