Asymmetric Information and Digital Technology Adoption: Evidence from Senegal*

Deivy Houeix†

[Click here for the most recent version]

February 17, 2026

Abstract

Digital technologies promise large productivity gains, but generate data that can be made observable to others at low cost. I show that this embedded observability can be a double-edged sword: while it reduces information frictions and raises efficiency, it also threatens agents' informational rents and deters adoption. Two field experiments in Senegal's taxi industry illustrate this trade-off. When transactions are observable, owners monitor effectively and driver effort rises, but adoption falls—especially among poorer, low-performing drivers. Removing observability nearly doubles adoption and, in a relational contract framework, increases total welfare. The findings highlight how the very feature that enhances efficiency of digital tools can also hinder their diffusion.

*I am deeply grateful to David Atkin, Esther Duflo, Ben Olken, and Tavneet Suri, for their unwavering guidance and support throughout this project. For helpful discussions and suggestions, I would also like to thank Charles Angelucci, Abhijit Banerjee, Bob Gibbons, Frank Schilbach, Robert Townsend, and participants of various internal seminars at MIT, and external presentations including WEFIDEV 2023 and NBER SI 2025 Development, for valuable suggestions and comments. I am also indebted to my amazing research team, composed of over 130 people throughout the two years of this project. To mention a few: Koffi Adjassou, William Amedanou, Elisée Amewouame, Justin Chery, Martin Duflo, Cheikh Gaye, Ismael Keita, Loic Laurent, Sara Niang, Yacine Ouahioune, Francois Soleiman for their outstanding field assistance, as well as Gabriel Davis-Hollander, Alexis Yi, Lawrence Yong at MIT for their excellent research assistance. I am especially thankful to Drew Durbin, Joyce Keeley, Lincoln Quirk, and the numerous people involved at Wave Mobile Money, the implementing partner company, for trusting me to carry out this study independently. I gratefully acknowledge financial support from JPAL-DigiFI Africa, PEDL, The Weiss Fund, and Shultz and data collection support from the SDI. I received IRB approval from MIT, protocol #2012000286. This RCT was registered as AEARCTR-0009155.

†Harvard University. deivyhoueix@fas.harvard.edu

Digital technologies are often promoted as tools to lower transaction costs and improve decision-making inside firms. A core reason is that digitization converts day-to-day activity into structured records—time-stamped, searchable, and easy to summarize. Because those records are cheap to store and transmit, digitization comes with an often-implicit design choice: who gets to see the data, at what level of detail, and when. Prior work has confirmed canonical contract-theoretic predictions in settings where technologies are introduced solely to monitor workers (e.g., Hubbard, 2000; Baker and Hubbard, 2003, 2004; Kelley et al., 2024). But many of the digital tools now spreading in lower-income countries—payments, platforms, inventory and supply-chain apps—are adopted for operational reasons and only *incidentally* generate the kind of transaction-level data that can be repurposed for oversight.

This paper studies a central tension created by this embedded observability. On the one hand, making activity easier to observe can reduce information frictions and improve efficiency. On the other hand, when adoption requires workers' cooperation, greater observability can discourage adoption by threatening informational rents and shifting bargaining power within ongoing relationships. Whether embedded observability is a second-order feature or a first-order determinant of the returns to digitalization is ultimately an empirical and quantitative question—one that is hard to answer with standard evaluations that measure productivity conditional on adoption while treating diffusion as exogenous.

I provide evidence on both sides of this trade-off—productivity gains and adoption costs—within the same setting and the same technology. I run two field experiments in Senegal's taxi industry in partnership with the country's largest mobile money company. Together we developed a novel digital payment solution for the taxi industry that allows drivers to receive digital payments securely into a business wallet. Prior to this study, cash was ubiquitous, so digital payments promised immediate reductions in cash-related transaction costs. Like most digital technologies, payment systems are not designed as monitoring tools. Yet, they automatically generate a transaction trail whose visibility can be engineered. By increasing observability and reducing moral hazard, these gains will be magnified among firms that use the technology, but at the cost of potentially discouraging initial adoption.

The taxi industry offers an ideal setting to study these dynamics, as it illustrates common principal-agent problems faced by small firms in lower-income contexts. In the Senegalese taxi industry, the typical arrangement links a car owner (employer) and a single driver (employee) through a relational contract. The driver keeps the revenue exceeding a fixed weekly rental fee and may also receive a basic upfront payment. Because owners cannot observe driver effort or revenue, limited liability allows drivers to default by claiming low earnings, whether due to luck, shirking, or misreporting. This creates scope for moral hazard in both effort and output reporting, enabling drivers to capture informational rents and reducing efficiency. Default may lead to (costly) termination of the relationship as owners seek to mitigate moral hazard.

The distinctive feature of my experimental design is the ability to randomly vary and separate the "observability" of drivers' digital transactions to taxi owners, allowing me to quantify its effects on both contractual relationships *and* workers' adoption decisions. Guided by contract theory, I implemented two complementary experiments: the "impact experiment" to estimate the effect of digital payments on firm performance and contracts, and the "adoption experiment" to estimate the role of observability on technology adoption.

In the "impact experiment," I identified drivers willing to adopt and varied two dimensions: (i) access to digital payments among 1,821 drivers—including owner-operators to increase precision—and (ii) transaction observability for taxi owners (employers) among 608 owner-driver pairs. Owners were randomly assigned to one of three observability regimes: *Granular*, the default for most businesses, which revealed all digital transactions to owners; *Coarse*, which displayed only aggregate collections up to a cutoff, allowing drivers to possibly signal low-output periods, but not the full transaction history and effort; and *No Observability*, which isolated the benefits of digital payments apart from monitoring. These treatments test whether observability—arising incidentally through payments—can mitigate moral hazard, increase agent's effort, and improve contract efficiency.

In the "adoption experiment," I test whether anticipated employer observability deters adoption when worker cooperation is required. Adoption initially required drivers to share their owner's contact information so the company could register the business account, and many drivers refused—owners were difficult to reach in the absence of a formal registry. I followed up with 433 drivers who had withheld owner contacts and therefore could not adopt. I then re-offered the same technology, while randomly varying the data-visibility setting—whether owners would be able to observe the driver's digital transactions. Unlike in the impact experiment, drivers learned this information *before* making their decision, allowing me to identify the causal effect of anticipated observability on willingness to share owner contacts and adoption. More broadly, the design captures a common implementation constraint in informal firms: when employees act as gatekeepers to onboarding, visibility settings can determine diffusion.

I use three data sources. First, I conducted five survey rounds with owners and drivers over nearly two years (95% pair-level follow-up). The key innovation of these surveys is to track employer-employee informal contract data over time, enabling an analysis of how technology influences contract dynamics. This survey panel dataset offers a rare opportunity to study contracts in developing economies, where the relational, verbal, and informal nature of contracts has historically hindered progress in the literature. Second, I measured drivers' effort by conducting mystery passenger audits across Dakar. In this exercise, about twenty surveyors hailed 7,896 taxis and discreetly recorded license plates. This provides an objective measure of driver effort on the road. Third, administrative data from the payment company, covering nearly all adults in Senegal and all study participants, capture daily payment and transfer transactions at the driver level.

I have four main findings: (i) digital payments cut drivers' cash-related costs by about half; (ii) by embedding observability, they reshape contracts and raise effort and firm performance; (iii) this same observability deters adoption, especially among lower-ability workers; and (iv) while digital payments improves total welfare, they can also widen inequality across and within firms.

First, digital payments significantly reduce cash-related costs such as time spent searching for small change or losing customers who prefer to pay digitally (even absent any monitoring role). For taxis, these costs are large—about 10% of baseline profits, as self-reported by drivers.[1] Mystery passenger audits cross-validate these reports, showing a 47% drop in price distortions from small-change shortages, as drivers no longer round fares down when they use digital payments. These reductions are striking given that digital payments cover only a fraction of drivers' revenue (about two customers a day, 13%). After 7–9 months, drivers report a willingness to pay for the technology equal to roughly one week's profits, highlighting substantial benefits of the technology, not tied to its observability feature.

Second, transaction observability reduces information frictions and reshapes contracts by mitigating moral hazard. Under *Granular Observability*, drivers are observed on the road 38% more often in mystery passenger audits (p=0.001) and process 35% more digital transactions (p=0.012). Owners with observability experience 34% fewer rent defaults (p=0.068). To compensate for higher effort, they adjust contracts: owners are 16% more likely to provide an upfront monthly payment ("salary") to their drivers (p=0.006) in addition to the rent they collect. Overall, the observability effect on taxi owners' profits is imprecisely estimated but positive, with point estimates of 6%.

Furthermore, observability improves worker retention. Across the sample, 33% of pairs separated within nine months and 61% within two years—a high turnover that is a critical challenge for many industries in lower-income countries (McKenzie and Paffhausen, 2019). Observability reduces turnover by 34% after nine months (p=0.030), especially among non-family pairs. The effect is concentrated under *Granular Observability*. These results indicate that moral hazard in effort, particularly mitigated under *Granular Observability*, is a significant constraint for owners. This effect shows how digital technologies can reduce information frictions and support business growth, something particularly relevant for industries dominated by family businesses precisely due to trust challenges between employers and employees (Bertrand and Schoar, 2006).

Third, the same observability that improves performance acts as a barrier to technology adoption for the lowest-performing and poorest workers. At baseline, 50% of drivers refused to adopt the technology, by withholding owner contacts, a prerequisite for adoption. To characterize selection, I follow Karlan and Zinman (2009) and compare drivers in the "impact experiment"—who opt into potential observability but whose owners are randomly assigned not to observe—to reluctant drivers in the "adoption experiment." Reluctant drivers report 83% more stress at work, perform significantly worse, e.g., have fewer passengers (-11%) and work fewer hours (-4%), and are also significantly poorer and 28% less likely to have completed primary school. Thus, increased monitoring deters adoption by a policy-relevant group, the poorest and least educated. In the "adoption experiment", I find that randomly assuring drivers that transactions would not be observable by their employer nearly doubles adoption rates, with especially large effects among

[1] The costs involved in cash payments do not entirely disappear because customer adoption of mobile money wallets remains incomplete—see Higgins (2024) on demand-side adoption and Alvarez and Argente (2022); Crouzet et al. (2023) for the dynamics between cash and digital payments. I examine the broader diffusion of digital payments in Senegal in a companion paper (Houeix, 2025).

these disadvantaged workers.

I develop a framework that (1) generalizes the findings beyond this setting by pinning down the underlying mechanisms, and (2) quantifies the net and distributional effects of alternative data-visibility designs. I model the contract between a taxi owner (principal) and a driver (agent) in a framework with limited liability and both unobservable effort and output, which together generate informational rents for the agent. I analyze both the impact and adoption of digital payments in a relational contract where the principal cannot commit to contract terms once the agent adopts the technology.

The framework yields three predictions. First, digital payments increase observability of effort through transaction histories and timestamps, reducing moral hazard. This raises driver effort, lowers rent defaults, and induces contract changes—notably the introduction of an upfront payment to compensate drivers. The upfront nature of this payment also helps limit the risk of owner reneging under a relational contract. Second, observability of digital revenues allows the owner to distinguish bad luck from misreporting or low effort, relaxing incentive and truth-telling constraints and reducing separations. Third, when deciding whether to adopt, drivers anticipate these effects: they weigh technology's benefits (reduced cash-related costs) against contract adjustments that increase effort and reduce informational rents, knowing that the principal cannot commit to initial terms in a relational contract setting. Observability is particularly deterring for low-ability drivers, for whom high effort is costly and unprofitable for the principal to compensate.

The framework highlights the key ingredients that generate these predictions: *limited liability* and *information asymmetries*, which generate informational rents and moral hazard, and *weak contract enforcement*, which explains both the change in contractual form—upfront rather than ex-post payments—and why the principal's inability to commit to contract terms discourages adoption. These three frictions are pervasive features of lower-income economies, making the trade-off especially severe.

The results show that digital payments have nuanced welfare effects: adopters remain with employers longer but also exert more effort. To quantify drivers' effort disutility—needed for welfare analysis—I estimate the model by the generalized method of moments (GMM), leveraging the randomized observability treatments and the characteristics of non-adopters.

The quantitative results show the importance of embedded observability. In the model, the baseline introduction of the technology raises aggregate welfare modestly (0.5%) but concentrates gains among higher-ability drivers who are willing to adopt under observability.[2] Policies that mandate adoption while preserving observability can backfire: mandating adoption reduces low-ability drivers' welfare (by 8%) and yields smaller aggregate gains (0.4%) than the status quo. By contrast, a privacy-by-default redesign that removes employer observability—holding fixed the payment functionality—induces broad adoption and increases aggregate welfare by 0.7%, while sharply reducing welfare inequality between high- and low-type businesses. Consistent with

[2] Welfare calculations consider only the employer and employee, with equal weights, excluding the consumer benefits from paying digitally. I also verify a "no-deviation" condition: for the relevant range of δ in my estimation, the owner does not profitably terminate a low-type agent to search for a high-type adopter willing to adopt the technology.

these implications, the partner company later made non-observability the default in this market.

By identifying and quantifying a design trade-off in digitalization, I contribute to the literature on technology adoption and organizations in development economics. In partnership with a payment company, I randomize not only whether firms receive a digital payment technology, but also whether—and how—the transaction-level records it generates are shared with employers. This design allows me, within the same industry and technology, to estimate (i) the impacts of digitizing transactions, (ii) the incremental effects of making the resulting records visible to principals, and (iii) the extent to which visibility changes take-up when adoption requires worker cooperation. Beyond this specific setting, the paper offers a general lesson for evaluating digitalization: when technologies simultaneously change operational costs and the information environment, design choices shape both performance conditional on adoption and diffusion itself. Measuring only one side can misstate the total gains from digital tools and how those gains are distributed.

First, I contribute to the literature on digital payments and firm performance by showing that, in cash-dominant service markets, a central impact channel operates through *cash frictions*. While prior work emphasizes household transaction costs (Aker et al., 2016; Jack and Suri, 2014, 2016) and other firm channels such as customer acquisition and credit access (Agarwal et al., 2019; Dalton et al., 2023; Higgins, 2024; Riley, 2024), I provide experimental evidence that digital payments also mitigate a broad set of operational cash costs in informal services—reducing small-change issues and the resulting stepwise price rounding, as well as electronic theft and theft-related anxiety. Although small-change shortages have been documented in retail settings (Beaman et al., 2014), I show they remain quantitatively important in urban services today, and that even partial digital-payment usage can relax these constraints. The results highlight how the spread of business-facing digital payment tools can generate first-order gains by easing everyday cash-handling frictions.

Second, I bridge the literatures on payments technologies and monitoring inside firms. A separate body of work studies *purpose-built* monitoring technologies, from seminal evidence in U.S. trucking (Baker and Hubbard, 2003, 2004; Hubbard, 2000) to recent experimental evidence in developing contexts (de Rochambeau, 2021; Kelley et al., 2024; Bossuroy et al., 2025; Kala and Lyons, 2025). My distinct point is that observability is not confined to specialized monitoring tools—nor is it the lens through which payment systems are typically viewed. Yet, observability can emerge as a byproduct of many forms of digitization: transactions automatically generate time-stamped records that can be shared at low cost. I show that this embedded observability reshapes relational contracting and behavior, including effort, default, compensation, and separation.

Third, I contribute to the literature on technology adoption within firms by identifying *observability and employee gatekeeping* as mechanisms behind uneven uptake of digital technologies. Prior work on transparency and adoption is largely descriptive, focusing on health data sharing (Goldfarb and Tucker, 2012; Derksen et al., 2024) or firms' reluctance to digitize payments for tax reasons (Brockmeyer and Somarriba, 2024). Outside digital technologies, Atkin et al. (2017) shows how misaligned incentives can generate worker resistance even when owners are the adopting party. In contrast, I show that in many informal service firms the bottleneck can arise before owners can act, through implementation gatekeeping. Onboarding and daily use run through employees who are easier to reach, already operate the business, and often control access to activation inputs (e.g., an owner phone number or required KYC documents). As a result, employees can delay or block adoption. In my setting, drivers revise their adoption decisions when the product is redesigned to protect transaction privacy. This underscores that design choices—e.g., how much transaction information is visible—jointly shape the performance gains from digitization and the pattern of adoption, with direct distributional implications.

Finally, I contribute to the literature on relational contracting and organizational design by developing and estimating a framework—building on models of informal contracting and monitoring in lower-income-country labor markets (Macchiavello and Morjaria, 2015, 2021; Kelley et al., 2024)—that emphasizes a commitment problem over newly generated information. When a technology creates verifiable records, principals may be unable to credibly commit *ex ante* not to use that information to tighten control or renegotiate terms *ex post*; anticipating this, agents may resist adoption even when the technology could benefit them. I match the model to rich data I collected on informal contracting to quantify welfare and evaluate counterfactual information designs, such as restricting employers' access to transaction records. The framework clarifies a trade-off between efficiency gains from improved observability and diffusion when adoption requires agent cooperation. Departing from U.S.-based studies of firms (Lazear, 2000; Blader et al., 2019), the analysis highlights how weak enforcement makes information design central for adoption, organizational responses, and the division of surplus.

The rest of this paper proceeds as follows. Section 1 reviews the setting and key stylized facts. Section 2 outlines the experimental design. Section 3 discusses the data collection process. Section 4 details the experimental results and mechanisms. Section 5 presents the theoretical framework. Section 6 explores counterfactuals based on structural estimations. Section 7 concludes.

1 Study Background

1.1 The Digital Payment Technology

This study leverages a new digital payment technology that I developed in partnership with Wave Mobile Money, the largest payment company in Senegal and Francophone Africa's first unicorn. At the time of the study (2022), Wave operated in six countries, serving about 80% of Senegal's adult population (7.2 million active users) primarily for peer-to-peer (P2P) transfers.

Working with Wave, I helped design a peer-to-business (P2B) technology tailored to the taxi industry.[3] I contributed to market research, product adaptation, and testing. Together with engineers, we refined the business app and piloted a QR-code system in taxis—a printed code hung from the rear-view mirror—that enables passengers to securely pay via mobile money. Drivers pay a 1% fee (see Figure 1), waived for the first 50,000 CFA (USD 85). This design introduces two important features compared to P2P transfers: (i) *irrevocability*, transactions could not be reversed,

[3]Electronic merchant payments exist elsewhere, such as Safaricom's *Lipa na M-PESA* in Kenya, launched in 2013.

unlike P2P transfers; and (ii) *convenience*—passengers scan a QR code rather than enter the driver's phone number.

Partnering with a market leader was crucial, as customer familiarity with Wave created strong incentives for business adoption, and therefore an opportunity to focus on the business-side frictions. At the start of the study, Wave had just begun digitizing payments for about 10,000 non-taxi businesses and expanded to nearly 200,000 within a year, reaching over two million unique users. Despite this growth, digitalization is only partial, as cash still remains the predominant mode of transaction. Owners and drivers offered the technology received training and the app on their phones; non-driving owners were briefed at the end of a baseline survey.

Employers observe transactions by default. The payment product uses the company's standard *merchant* (P2B) account architecture, designed for small businesses in which a manager/owner may delegate payment collection to employees (e.g., cashiers). In this default configuration, the manager account can view the transaction ledger and receive automated summaries for reconciliation and fraud prevention, while employees use a linked interface to accept customer payments. In the taxi setting—where an owner delegates daily operations to a driver—the same architecture implies that owners can view the driver's business-payment history by default (Figure 1(c)). The company typically verifies taxi registration documents available through the owner to determine the best type of Wave account related to maximum daily and monthly transaction limits. More generally, such observability is a feature of many digital payment technologies, which automatically generate and retain transaction data. I therefore randomized both *access* and *observability* to isolate their effects on contracting and adoption (Section 2.2). Post-experiment, the company shifted to a no-observability configuration for this industry, as a design choice, as discussed in Section 6.5.

1.2 Stylized Facts

1.2.1 Taxi Industry in Senegal

This study focuses on Dakar's private taxi sector, which employs 4–6% of adult urban men, with about 21,000 active taxis in 2019.[4] The industry is informal and includes two main stakeholders: the taxi business owner (principal) and the driver (agent). Owners and drivers are typically men aged 30–50.

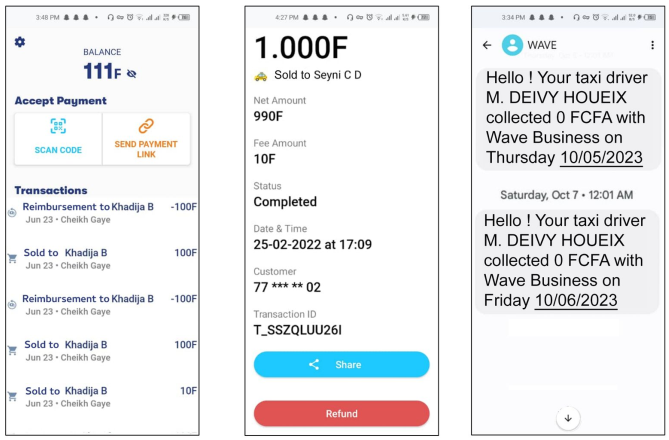
The sector is well-suited to study digital payments: (i) cash use imposes substantial costs, and (ii) owners cannot observe driver effort or revenue, creating information asymmetries. Digital payments, if adopted, promise to alleviate both of these challenges.

Stakeholders. Baseline surveys (described in Section 3.1) identify three main types of owners: sole proprietors driving their own taxis; non-driving owners employing drivers; and hybrid own-

[4]Data from CETUD, covering only licensed taxis. Drivers are virtually all men in this setting. Total employment is likely higher given informality. For comparison, New York City has 13,000 yellow taxis for 8 million residents.

(a) Taxi Driver with Technology Access

(b) Digital Payment Interface for Drivers



(c) Owner's Observability of Driver Transactions and Daily SMS Updates

Figure 1: Digital Payments and Transaction Observability for the Taxi Industry in Sénégal

Notes: Figures 1(a) and 1(b) show the peer-to-business (P2B) payment technology developed with Wave Mobile Money for Senegal's taxi industry. The QR code allows customers to identify participating taxis, and a windshield sticker enhances visibility. Drivers received training and the business app, and owners in the observability treatment were given a parallel app showing driver transactions. Passengers could pay with smartphones or via mobile money cards. Figure 1(c) shows screenshots from an owner's test account: (i) the driver's balance and transaction history, (ii) a detailed transaction view with timestamp, and (iii) the automated midnight SMS update. My name appears as I coordinated the pilot and received daily updates. Screenshots were translated from French.

ers who alternate. Most owners hold one taxi (92%) and out of the ones with an employed driver, employ only one (88%); about 54% of owner–driver pairs belong to the same family. Both owners and drivers have typically not completed primary education (67%). Half of the drivers report no savings in the past three months, most are the main household earners, and can be classified as urban poor (see Tables B2 and B3).[5]

[5] I use the Poverty Probability Index (PPI) specific to Senegal to measure wealth. The average score of 63 implies a 50% chance of living below 200% of the 2011 National Poverty Line.

Cash Payments at Baseline. Fares are negotiated upfront and usually paid in cash. While P2P mobile money transfers are possible, they are rarely used for taxi fares due to revocability concerns. Passengers can easily reverse payments after rides, which is an issue the technology solves. Consequently, drivers averaged only six P2P transactions resembling taxi payments, based on their value, in the three months prior, a negligible share of revenue. More broadly, digital merchant payments were still uncommon in Senegal pre-intervention: only about 12% of adults (15+) report having made a digital merchant payment in 2021 (Global Findex 2021).

Costs of Cash. There are substantial costs of using cash across four categories identified through interviews with taxi drivers: (i) *Any Time Lost*—86% of drivers report losing time finding small change, spending at least 10 minutes about 1.52 times weekly, with 7% of drivers experiencing this daily. (ii) *Refused Customers*—60% report turning down passengers without change, which is 3.77% of their passengers.[6] (iii) *Reduced Price*—92% report cutting fares to the nearest bill at least once weekly due to small change shortages. (iv) *Mistake Change*—41% report weekly miscalculations. In addition, drivers report anxiety about theft related to carrying cash, and electronic theft—when customers pay digitally through personal transfers and then leave the taxi—is a major concern.

Overall, weekly monetary loss averages USD 8.00, about 10% of effective profit.[7]

Absence of Monitoring. *GPS tracking* is virtually nonexistent in the taxi sector, and owner awareness of such tools is minimal. Important barriers to adoption are the high acquisition and maintenance costs—GPS trackers cost owners about a month of profit plus recurring fees—as well as strong resistance from drivers, who find them too disruptive. Moreover, most taxis are old (20-30 years) with broken *odometers*, further limiting monitoring of miles driven. At baseline, *ride-hailing* apps were rare ($< 4\%$ of drivers). Digital payments were thus often the first digital technology adopted, serving as a gateway to further digitalization (World Bank, 2024).

1.2.2 Taxi Owner-Driver Contractual Relationships

In this section, I present four facts about the taxi industry that motivate both the experimental design and the theoretical framework that follow.

Fact 1 – Owner-Driver Contract Structure. Taxi owners use informal rental contracts and in some cases provide upfront payments to drivers. The rental contract requires a weekly transfer of about CFA 60,000. Partial rent defaults are possible but may trigger termination. This structure resembles contracts in other lower-income settings, such as Kenyan minibuses (Kelley et al.,

[6] Estimating lost customers requires knowing how long it takes drivers to find another fare; estimates show drivers spend about half their working time idle, indicating that refusing a customer is a substantial cost.

[7] Losses are imputed by valuing each reported cost based on fieldwork with a subset of drivers prior to the experiment: CFA 500 (USD 0.8) per time lost or change mistake, CFA 1,500 (USD 2.5) per refused customer, and CFA 800 (USD 1.3) per reduced fare. These figures exclude harder-to-quantify costs such as theft, anxiety, and record-keeping.

2024), but differs from U.S. taxi contracts (Angrist et al., 2021), with each of the four stylized facts underscoring key differences that distinguish the two contexts.

In addition to the fares kept by drivers, about 53% of owners pay their drivers an upfront monthly payment—that they refer to as a "salary"—ranging from CFA 40,000 to 50,000 (USD 65-85). This payment is best understood as a minimum-income guarantee within the fixed-fee contract. What differs between the two payments is timing: in 90% of cases, this is paid upfront regardless of rent payments and represents about 18% of driver's total compensation (revenue collected including salary minus the rent and the costs, e.g., fuel, food consumption, police bribes, minimal maintenance costs). Most owners (84%) cite the main purpose as committing to a minimum payment even when fares are low; other reasons include discouraging poaching and reducing risk-taking. Owners also cover major maintenance costs.

Fact 2 – Limited Liability and Defaults. Drivers remit full rent only if revenue suffices, otherwise defaulting partially. Reported reasons for defaulting include low demand, accidents, or traffic. Limited liability, a typical constraint in lower-income contexts, prevents drivers from paying the rent in advance and reflects limited credit access: only 8% of drivers had any loan in the three months prior, despite a high demand for credit, and less than half were able to save money. Defaults are widespread: 69% of drivers experienced at least one in the past three months, and 47% at least monthly. However, owners cannot easily verify drivers' effort nor reports of low-output. For these two reasons—moral hazard in effort and in output reporting—defaults have been identified as a frequent source of disputes with drivers for 65% of owners and an important source of stress for 49% of drivers.

Fact 3 – Large Information Asymmetries. Owners often have incomplete and inaccurate information about their drivers' effort and output. I analyze matched owner–driver dyads and document large *pairwise* misperceptions. Owners tend to *underestimate* hours, days, revenue, and passenger counts: only 39% know their drivers' working hours (± 2 hours), 26% know earnings, and 46% know days worked, with 33% underestimating. Further, 68% of drivers park taxis away from owners' homes, preventing owners from monitoring work days.

Output is also highly variable. Predictive models using calendar and effort measures explain little of daily revenue (R^2=22%; adding driver FEs raises it only to 45%), indicating the importance of demand shocks (Table B4). These large asymmetries suggest digital observability could substantially improve owners' knowledge of drivers' work.

Fact 4 – High Turnover. The taxi industry is characterized by high turnover rates: median duration is 1.5 years, and drivers had worked for three owners on average at baseline.[8] Limited liability

[8]Comparable turnover rates are documented in other low-income industries. Fajnzylber et al. (2006); Nagler and Naudé (2017) analyze small firm exit, while McKenzie and Paffhausen (2019) reviews 16 panel surveys from 12 countries, estimating an average annual death rate of 8.2%. Evidence on employee turnover in small and medium firms remains limited though, largely due to scarce within-firm data in informal contexts.

and private information create scope for conflict with owners, sometimes resulting in the termination of the relationship. In the data, 65% of owners cite issues like drivers defaulting on rental payments as among the primary causes for separations and baseline driver default is positively correlated with subsequent turnover. However, as discussed below, separations often follow the accumulation of multiple issues rather than a single incident.

Over nine months, 33% of pairs separated, rising to 61% after two years. I manually coded owners' and drivers' open-ended responses on separation: the most common reasons were drivers quitting (29%), owners firing drivers (21%), and owners selling the taxi (21%). Separations are costly: both sides report losing 34-38 days finding replacements. In most cases (92%), separated drivers remain in the taxi industry. However, the market provides limited discipline through reputational mechanisms—qualitative responses emphasize that owners have few ways to credibly share information about driver quality because there is no formal reference system and relationships are often negotiated bilaterally. Separation rates are substantially higher in non-family businesses, which themselves are associated with lower trust, as discussed in Section 4.2.2.

2 Experimental Design

2.1 Experiments Guided by Contract Theory

I designed two experiments in Dakar's taxi industry to test the core idea that digital technologies can raise productivity but also inherently bundle observability, which may deter adoption. Within a principal–agent framework, I examine two mechanisms: (1) Observability embedded in digital payments reduces *moral hazard in effort* and *in output reporting*, raising effort and contract efficiency, in line with the canonical prediction in contract theory, and (2) because agents anticipate these changes and the principal cannot credibly commit to initial contract terms in *relational contracts*, this may discourage initial adoption.

The two experiments are: (i) the "impact experiment" that tests how digital payments affect firm performance, focusing on drivers' cash-related costs and the observability effect on employer-employee relationships (contracts, trust, and retention), and (ii) the "adoption experiment" that tests how observability influences workers' willingness to adopt. The first experiment targets drivers willing to adopt the technology by sharing owner contacts, while the second targets drivers who refused, preventing adoption. Figure 2 summarizes the design. Section 5 formalizes the theoretical mechanisms behind the treatments and guides the structural estimation.

2.2 Impact Experiment: Digital Technology Access and Observability

Drivers were listed and invited to participate in the study to adopt a newly developed digital payment technology. After drivers provided their owner's contact information, I randomized access to the technology at the *taxi business owner* level, with the sample primarily consisting of owners with only one taxi. At the time of the experiment, owners' approval is required for adoption, and

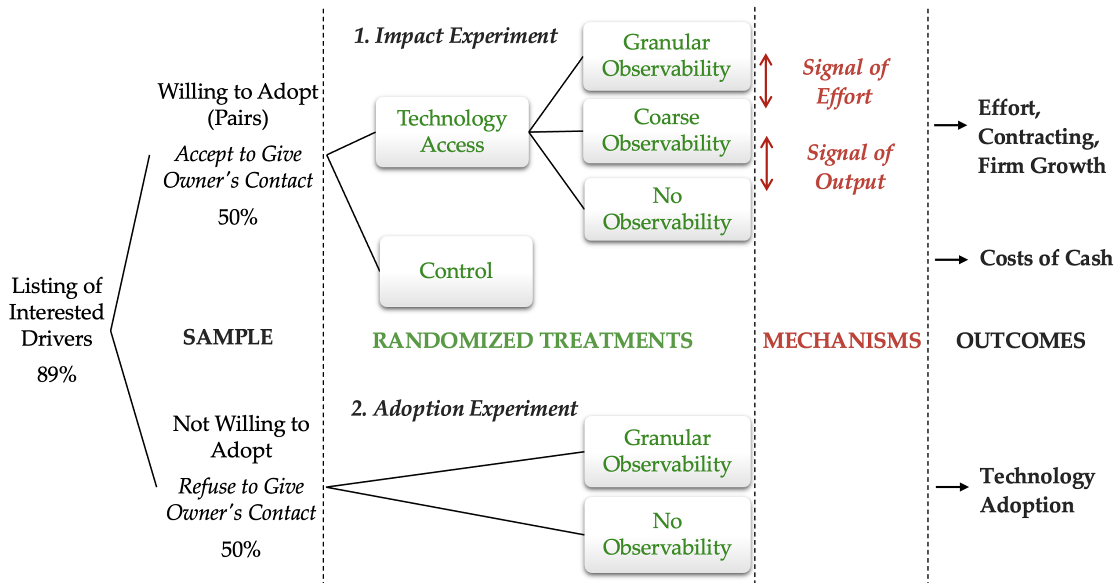


Figure 2: Experimental Design: Impact and Adoption Experiments

Notes: From the listing activity, where most drivers expressed interest in adoption, two groups emerged. The design includes: (i) the **impact experiment** (top), randomizing access to digital payment among 1,821 drivers—including owner-operators to increase precision—and specifically transaction observability for taxi owners (employers) among 608 owner-driver pairs; and (ii) the **adoption experiment** (bottom), targeting 433 drivers who refused adoption due to reluctance to share owner contact information—a prerequisite for access. Both experiments were followed by mid- and long-term surveys to track driver performance, contracts, and owner–driver relationships. This two-part design highlights the trade-off between the benefits of observability and the adoption barriers it creates.

customers ultimately choose cash or digital payment.

Technology Access. Drivers and owners were invited to three different locations in Dakar to be surveyed separately (see Section 3.1). At the end of the baseline surveys, drivers were randomized into treatment and received training on the app and QR card from field agents working at the payment company. Owners who drive their own taxis without employees were also randomized access to the technology primarily to improve the precision of cash-related cost outcomes.[9]

The Observability Treatments. The base system used by non-taxi businesses embeds observability by default, allowing employers to track employees' transactions. Partnering with Wave, I altered this feature and randomized three observability options within owner–driver pairs conditional on access. Each targets a different information asymmetry—moral hazard in effort or output reporting—modeled in Section 5. Participants were told these options were part of a pilot digital payment system. Because personal mobile money use was already widespread, all participants understood what "observability" of transaction records entailed. Appendix C provides the full

[9]As discussed later in Section 4, these owner-operators were also randomized into one of the observability treatments, which would only apply if they hired employees in the future. They are excluded from the analysis of observability effects, except when investigating how the technology's observability affects their hiring decisions.

scripts. The three randomized levels are:

1. *No Observability (N-O)*—20% of taxi businesses. Owners cannot view drivers' transactions; drivers still receive the technology.

2. *Granular Observability (G-O)*—20%. Owners access an app with the driver's full transaction history, including timestamps (Figure 1(c)), the default option typically offered to businesses. Metrics include customer count, transaction values and times, and digital revenue. Figure A2 illustrates the data available to owners. Owners also receive daily midnight SMS updates on total digital collections. This treatment provides owners with a far more comprehensive view of the driver's work, in contrast to the previous complete lack of observability. This data may allow monitoring by providing signals of driver's *effort* (hours, days worked) and *output* (digital revenue). Because digital payments cover only a fraction of total revenue,[10] the timestamps are especially valuable when drivers exert high effort, allowing owners to observe sustained activity throughout the day and enabling drivers to credibly signal effort.

3. *Coarse Observability (C-O)*—20%. Owners receive only daily SMS updates up to a CFA 5,000 (USD 8) threshold, the average daily digital collection in pilot data—about 17% of total daily revenue. Surplus amounts above CFA 5,000 are not disclosed. This aims to signal low-output days without revealing customer count, hours worked, or detailed revenue. It also tests for ratchet effects if drivers strategically minimize digital payments under *Granular Observability*.

4. *Pure Control*—40%. Drivers did not receive the technology in the first nine months and were placed on a waitlist.[11]

The contrasts across treatments capture distinct effects: *G-O* vs. *C-O* aims to isolate the effect of increased information on effort (reducing moral hazard in effort), while *G-O* vs. *N-O* aims to capture the combined effect of the effort and output signals, including low-output days.

2.3 Adoption Experiment: Digital Technology Adoption

Most drivers expressed strong interest in adopting the technology during the listing (89%), but roughly half (not by design) refused to provide their owner's contact information, a necessary step for adoption. I followed up with drivers up to three times and included them in the impact experiment if they eventually agreed to provide owner contact.

For those who continued to refuse, I ran the "adoption experiment" to disentangle the reasons and isolate the role of observability. Specifically, (i) I re-offered the technology to a subsample

[10] Owners were explicitly reminded that digital transactions would represent a limited share of their driver's total revenue, given the use of cash.

[11] At midline, control drivers received the technology with randomized *Granular* or *No Observability*. Some were further assigned nudges about contractual changes for a separate study, so the control group is excluded from the long-term analysis.

of reluctant drivers and (ii) I randomized owner observability *before* drivers decided whether to adopt (Figure 2). This design tests whether adoption rises once drivers no longer fear owners seeing their digital transactions. These drivers were then tracked in mid- and long-term surveys, with eventual access to the technology after the experiment.

2.4 Randomization Procedure

Randomization for the impact experiment was conducted at the owner level in twelve batches. Batches were designed to minimize the time between listing and baseline surveys, thus reducing attrition. Stratification used three dimensions: (i) *baseline digital usage*—whether drivers made more than the median six personal transactions in the taxi price range over the past three months; (ii) *baseline relationship*—relationship length (above/below two years, the median), business type (owner-driver vs. owner only), and fleet size (one vs. multiple taxis); and (iii) *baseline risk aversion*—whether drivers operated citywide or waited at fixed locations. This ensured balance and enables heterogeneity analysis (e.g., recent relationships). Separately, 60 association presidents were randomized to either *N-O* or *G-O* to prevent implementation issues that could arise if they were assigned to the control group. They are excluded from the treatment–control comparison and constitute a separate stratum.[12] Randomization balance is reported in Tables B2 and B3.

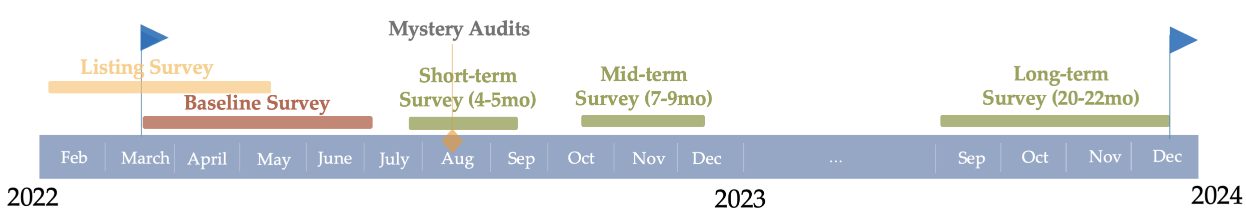
In the adoption experiment, observability was randomized among interested drivers at follow-up without stratification; robustness checks include alternative specifications and controls.

3 Data: Measuring Informal Contracts and Worker Behavior

This section details the three primary sources of data used in this study: survey responses, mystery audits, and payment data.

3.1 Survey Data

Five rounds of survey data were conducted: listing, baseline, short-term (after 4–5 months), mid-term (7–9 months), and long-term (20–22 months)—see the timeline below.



[12]Randomization was implemented in Stata using `randtreat` within each batch×stratum cell. Target assignment probabilities were 0.20 to each observability arm (N-O, C-O, G-O) and 0.40 to pure control. When a cell size made exact target shares infeasible due to integer constraints (about 40% of observations fell in such cells), `randtreat` assigns treatments as evenly as possible within the cell, and any remaining "leftover" observations (misfits) implied by integer indivisibility were independently randomized across the arms. Note that all regressions include stratum and batch FEs.

Listing Survey. Between March and May 2022, drivers and owners were listed at garages, car washes, meeting points, and taxi stands across Dakar. Non-driving owners were contacted through their drivers, as owner contact details were required for adoption. Employed drivers who refused to share owner contacts were also listed.[13]

Baseline Survey (Willing to Adopt). Starting from the listing survey, I contacted 2,072 taxi owners and asked each owner to update their current roster of drivers. Baseline data collection was conducted in twelve rolling batches across three Dakar sites. In total, 3,011 individuals (owners + drivers) were part of the baseline, with owners and drivers interviewed separately in 45-minute questionnaires. The baseline dataset can be viewed in two complementary (overlapping) units. First, it includes 2,196 unique taxi stakeholders interviewed (Column (2) in Table B1): 1,821 are drivers (including owner-drivers who operate their own taxi) and the remainder are non-driving owners who employ at least one driver. Second, it includes 608 linked owner–driver pairs—cases where both an owner and one of their drivers were successfully surveyed and can be matched—forming the main dyadic analysis sample. Overall attrition from invitation to participation was 26% in the stakeholder sample and 33% in the linked pair sample, with no significant differences across treatment arms (treatment assignments were disclosed only after baseline surveys to prevent differential attrition). Roughly half of attrition reflects baseline ineligibility (e.g., the driver lacked an Android phone, a license, or an ID), rather than refusal or no-shows.[14]

Adoption Survey (Reluctant Drivers). Drivers who refused to provide owner contacts could not be onboarded at baseline and are therefore classified as initial non-adopters. In June–July 2022, I re-contacted 433 of these drivers by phone to re-offer the technology and experimentally vary whether adoption would be bundled with owner observability, allowing me to test how observability affects adoption decisions on this important population. Because initial non-adopters withheld owner contact information and many had declined the extended baseline survey, this follow-up necessarily targets the subset we could reach and who remained engaged: the 433 drivers in the adoption experiment are those for whom we had a working phone number and successfully made contact, who were still active in the taxi market at re-contact, and who agreed to complete the short survey and still expressed willingness to consider adopting the technology.

Follow-up Surveys. Short-, mid-, and long-term surveys were conducted approximately 5, 9, and 20 months post-intervention. Attrition was low: 82.2% of drivers, 95.1% of pairs, and 84.8% of non-adopters were reinterviewed at two years (Table B1), with no differential survey attrition across treatments (see Table B5). Because many outcomes focus on owner-driver contracts, survey responses from either party still provide key information on the business. These survey rounds yield a rich panel on firm activity, worker behavior, retention, and contracts over time.

[13] Taxis were deemed ineligible if drivers lacked a smartphone (15%), were unreachable after repeated contact attempts (10%), or if owners operated more than four taxis (1%).

[14] We encountered 130 owners—mostly non-driving owners—who refused an in-person survey and were instead surveyed by phone. Five owners became drivers between listing and baseline; they are excluded and counted as attrition. Tables B2 and B3 confirm baseline balance across groups.

Survey Quality. All interviews were conducted by trained enumerators and reviewed by back-checkers who reinterviewed about half the sample on key questions. Quality checks included survey duration monitoring and targeted reviews on SurveyCTO. Follow-up phone surveys used strict callback protocols, including night/weekend shifts, WhatsApp reminders, and contact tracing through friends to recover updated phone numbers. While discrepancies between owner and driver reports were limited, any inconsistencies prompted a third review by a senior field coordinator to determine its cause.

3.2 Mystery Passenger Audits

In August 2022, I conducted mystery passenger audits to measure (i) drivers' behavior related to digital payments and pricing, and (ii) drivers' effort, proxied by road presence. Twenty trained surveyors hail taxis throughout Dakar, following a strict procedure to mimic typical price bargaining. Over two weeks, they systematically rotated across seven high-traffic locations, capturing a broad sample of drivers over a meaningful timeframe. Surveyors asked questions and discreetly recorded the license plate, later matched to the experimental sample (Figure A1). The activity was repeated a sufficient number of times to match taxi drivers with their license numbers in the experimental sample. Specifically, mystery passengers adhered to the following steps: (1) Memorize the randomized destination and pre-specified price on their data collection application, (2) Stop a taxi, (3) Ask the driver's initial price, (4) Suggest the pre-specified low price, (5) Listen to the driver's counteroffer and ask their last price, (6) Suggest a non-rounded price, (7) Ask to use digital payments. Once the taxi left, they recorded data about each step of the process. The bargaining process averaged just 1-2 minutes to minimize any negative impact on the driver's activities. In total, 7,896 taxis were audited, recovering 500 study taxis (including 41% of the taxis in pairs).[15]

3.3 Mobile Money and Payment Data

I obtained administrative mobile money data from the partner company, covering the universe of transactions for most adults in Senegal, including all study participants. These data track digital payment usage at the driver level across treatment arms and span pre-, during-, and post-study periods, with pre-treatment consisting only of personal transactions and later periods including both personal and business payments.

4 Experimental Results

4.1 Impact Experiment: Digital Payments Reduce Costs of Using Cash

This section examines the impact of digital payments on cash-related costs, including all drivers and owner-operators. To frame the observability–adoption trade-off, I first show that the technol-

[15] Roughly 30% of taxis were audited multiple times. All encounters are retained, with standard errors clustered at the business level; frequency of observation is used as a measure of effort.

ogy delivers clear benefits: it substantially reduces cash costs and, by the end of the study, drivers report a high willingness to pay for it.

4.1.1 Estimation Strategy

The estimation strategy follows the main specifications outlined in the pre-analysis plan (PAP) AEA registry ID #0009155. I employ the following driver-level regressions to assess the impact of digital payments on drivers' cash-related costs.

$$y_{ij} = \beta_0 + \beta T_{ij}^{Access} + \alpha_s + \epsilon_{ij} \quad (1)$$

where i indexes drivers, and j indexes businesses (taxi owners), T_{ij}^{Access} indicates access to digital payments, α_s are strata and batch fixed effects. No further covariates are included unless specified.

Treatment compliance was near-complete: only 9 control drivers (fewer than 1%) obtained the technology by changing phone numbers and all eligible treated firms received it. Firms that did not participate or were ineligible are simply excluded, as they were not made aware of their treatment status in advance (ensuring no differential attrition), hence ITT $\approx$ TOT. Equation (1) estimates impacts on cash-related costs, pooling employed and self-employed drivers to increase power. Standard errors are clustered at the business level, following Abadie et al. (2022).

4.1.2 Reduced-Form Results

Cash-Handling Costs Fall. In Table 1, I pool across observability treatments and show that digital payments reduce cash-related costs by roughly half. For example, 47% of drivers in the control group lost more than 10 minutes seeking small change in the past week; treated drivers were 23 percentage points (pp) less likely to report this. Similar reductions appear in refusing customers, reducing prices, and giving incorrect change. As a result, the imputed monetary loss from these frictions falls from USD 5 (about 6.4% of profit) by nearly 50%. Digital payments also reduce electronic theft by about 68%. I cross-validate some of these results by conducting mystery audits on taxi drivers to check for social desirability bias, as described below.

In magnitude, these cash-friction savings are economically meaningful compared to other profitable opportunities for firms: e.g., they are comparable to the median profit gains from a lower merchant-fee contract in Mexico (about 3% of profits; Gertler et al., 2025) and smaller than the profit losses implied by suboptimal pricing policies in large U.S. retail chains (on the order of 7% of profits; DellaVigna and Gentzkow, 2019).

Table 1: Impact of Digital Payments on Costs Associated with Cash Payments

	Any Time Lost	Refused Customers	Reduced Price	Mistakes Change	Imputed Loss	Electronic Theft
Panel A. Short-Term 5-Month Survey						
Technology Access	-0.233***	-0.195***	-0.130***	-0.048***	-2.090***	
	(0.026)	(0.023)	(0.027)	(0.016)	(0.267)	
Observations	1502	1500	1500	1497	1501	
Control Mean at Short-Term	0.47	0.33	0.51	0.12	4.91	
% Change T at Short-Term	-49.97	-58.83	-25.67	-40.76	-42.53	
Panel B. Mid-Term 9-Month Survey						
Technology Access	-0.127***	-0.171***	-0.039	-0.033**	-1.355***	-0.045***
	(0.023)	(0.025)	(0.026)	(0.014)	(0.231)	(0.011)
Observations	1352	1351	1351	1352	1352	1564
Control Mean at Mid-Term	0.27	0.34	0.32	0.09	3.28	0.07
% Change T at Mid-Term	-47.59	-50.91	-12.04	-38.84	-41.26	-67.63

Notes: Baseline data were collected March–June 2022, short-term July–September 2022, and mid-term October–November 2022 (9 months). Outcomes are regressed on treatment (technology access), with pure control as the omitted group: $y_{ij} = \beta_0 + \beta T_{ij}^{Access} + \alpha_s + \epsilon_{ij}$, where i indexes drivers and j businesses and α_s strata and batch fixed effects. Standard errors are cluster-robust at the business level; significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Percent changes are coefficients divided by control means. The sample includes all drivers surveyed at least once; values are missing when drivers did not work in the past 7 days or could not recall/refused to respond. Regressions include baseline controls when available. Outcomes are 0–1 dummies referring to the past 7 days: *Any Time Lost* (time wasted seeking small change, >10 min), *Refused Customers* (customers turned down for wanting to pay electronically), *Reduced Price* (fare cut due to small change), *Mistakes Change* (losses from miscalculation), *Imputed Loss* (cash-related losses, imputed from driver interviews), and *Electronic Theft* (electronic money stolen in past 3 months). All monetary values converted to USD (USD 1 = CFA 600).

These reductions are striking, given that the digitalization covered a limited share of revenue: drivers process about 2.2 customers per day on average (up to 6 at the 95^{th} pct.). Based on self-reports over the prior three days, digital payments account for about 13% of the total revenue among users on average (up to 40% at the 95^{th} pct.). This is consistent with cash frictions being *lumpy*: even if most rides remain cash, the option to accept a digital payment relaxes small-change constraints on the margin (e.g., avoiding refusals or stepwise rounding when change is scarce).[16]

Effects are stronger after five months than after nine, partly because the control group improved over time. Administrative data show rising acceptance of P2P transfers from passengers, despite revocability concerns, suggesting broader diffusion of digital payments among businesses in Dakar indirectly benefited control drivers.

[16] Back-of-the-envelope: using the mid-term control mean daily revenue of about USD 51 and a 6-day workweek implies weekly sales of roughly USD 306. With the current average digital revenue share of about 13% among users, a 1% transaction fee corresponds to 0.13% of sales (USD 0.4/week). The estimated profit gain from reduced cash frictions is about USD 2.09/week, or roughly 0.7% of sales—already larger than the implied fee burden at current penetration. Under the counterfactual of full digitalization (100% of fares paid digitally), the fee would be 1% of sales (about USD 3/week); so full adoption is privately worthwhile whenever cash-friction savings exceed this USD 3/week. For reference, drivers' self-reported cash losses due to cash-handling issues are on the order of USD 8.00 per week at baseline, and USD 5 at short-term.

Pricing Flexibility. Mystery audits validate survey reports. In Panel A of Table 2, treated taxis are 30 pp more likely to accept digital payments (relative to the 44% control mean). Compliance is not perfect: some drivers may still prefer cash-on-hand and the license number matching may not perfectly identify the experimental driver if the car was shared with an occasional driver on the audited day (measurement error).

Table 2: Impact of Digital Payments on Non-Rounded Prices

	Accept Digital Payments (1)	Non-Rounded Prices (2)	Non-Rounded Prices (3)
Panel A. First-Stage			
Treatment (Technology Access)	0.302***		
	(0.043)		
Panel B. OLS			
Treatment (Technology Access)		0.097**	0.081**
		(0.041)	(0.037)
Panel C. IV			
Accept Digital Payments		0.322**	0.275**
		(0.126)	(0.114)
Observations	707	707	707
# Businesses	484	484	484
Control Mean	0.442	0.325	0.325
Surveyor & OD FE	NO	NO	YES

Notes: Data were collected in August 2022 through mystery passenger audits at taxi stands across Dakar. Trained surveyors bargained with drivers and discretely recorded license plates, later matched to driver data and treatment status. Regressions are at the passenger–driver interaction level, with heteroskedasticity-robust SEs clustered at the business level. Column (1) reports the treatment effect on accepting digital payments; Columns (2)–(3) show OLS estimates for accepting non-rounded prices. All specifications include strata and batch fixed effects, and when indicated, surveyor ($Surveyor_i$) and origin–destination (OD) FEs to account for systematic price differences (e.g., higher fares for male passengers). Panel B estimates $NonRounded_i = \beta_0 + \beta T_i^{Access} + Surveyor_i + \alpha_s + \epsilon_i$. Panel C reports IV results, with the first stage in Panel A. At baseline, 18 drivers refused and 3 gave duplicate plates; these were dropped. Outcomes are: *Non-Rounded Price* =1 if the driver accepted a final offer CFA 200 (≈ USD 0.33) below their last price (not rounded to CFA 500), and *Accept Digital Payments* =1 if the driver suggested or accepted payment via QR code. About 30% of taxis were observed multiple times; all encounters are retained and later used in Section 4.2 to measure driver effort.

I examine pricing in two ways. First, OLS estimates (Panel B) show digital payments increase acceptance of non-rounded prices by 10 pp (31% relative to control). This matters because fares are negotiated verbally: if non-rounded outcomes were merely an "experiment artifact" that occurs only when surveyors or passengers explicitly propose them, we would expect little movement in acceptance. Instead, the increase indicates that non-rounded terms are feasible in ordinary negotiation but are often ruled out by the cash constraint. Second, given imperfect compliance, I estimate the local average treatment effect (LATE) on drivers accepting digital payments. The exclusion restriction states that treatment status affects the outcome "accepting a non-rounded price" only through accepting digital payments. Panel C shows substantially larger effects: drivers who accept digital payments are nearly twice as likely to accept non-rounded prices (above the 32% control mean), implying a 47% reduction in distortions from small change shortages, even with surveyor and origin–destination (OD) fixed effects. Third, and importantly for external validity, administrative data from the treated group show that non-rounded prices are not rare: about 30% of observed fares are non-rounded. This suggests that when the payment technology removes the small-change constraint, drivers and passengers frequently settle on flexible, non-stepwise prices as part of routine bargaining. Taken together, these results provide causal support that cash denomination constraints distort negotiated prices, forcing stepwise rounding, while digital payments enable flexible pricing and reduce deadweight loss from small-change shortages.

Driver Daily Profit. I next examine daily profits. Measuring profit in informal settings is notoriously difficult and subject to noise. I construct two complementary measures: (i) detailed 3-day revenue and cost accounting (revenue minus fuel, rent transfers, repairs, bribes, food, etc.), based on fieldwork, and (ii) self-reported average daily profit over the past 30 days (De Mel et al., 2009). While noisy, the two measures are strongly positively correlated.

Table B6, Columns (1)–(2), shows no detectable impact of digital payments on daily profit. Treated drivers may have briefly raised productivity (e.g., revenue or passengers per hour), but these effects are small and fade over time. The absence of detectable profit effects does not imply the technology is ineffective. Rather, the cost reductions documented above—roughly 3% of profits—are meaningful but modest relative to total earnings. Given the inherent volatility of profits in this sector, detecting such modest changes requires much larger samples. Power calculations confirm this: the minimum detectable effect (MDE) with 80% power is 6–12% (Table B7), well above the expected effect size.

Consistent with prior work on informal firms (e.g., De Mel et al., 2009), small productivity gains are often swamped by volatile profit measures. While digital payments significantly reduce cash-related frictions, their profit impact is too modest relative to sample size and noise to be detected. As discussed below, I find some increases in monthly driver profits from observability and the resulting salary effect (see Table B21).

Additional Outcomes: Theft Anxiety, Record-Keeping, and Savings. Table B8 explores other outcomes, outlined in the PAP, including theft-related anxiety, record-keeping, and savings. Digital payments significantly improve drivers' record-keeping by 4-5pp (over a control mean of 2%). Record-keeping nevertheless remains low because most fares are still paid in cash for most drivers, limiting the scope for the technology to generate comprehensive transaction records for accounting yet. Digital payments also reduce theft anxiety by 5-8%. They lower "impulsive" purchases, often made from street vendors while on the road, by 15% in the short term, although this is not enough to increase savings. Anecdotally, digital payments motivated some drivers to formalize their status by acquiring the necessary paperwork to access the technology, highlighting a potential formalization effect.

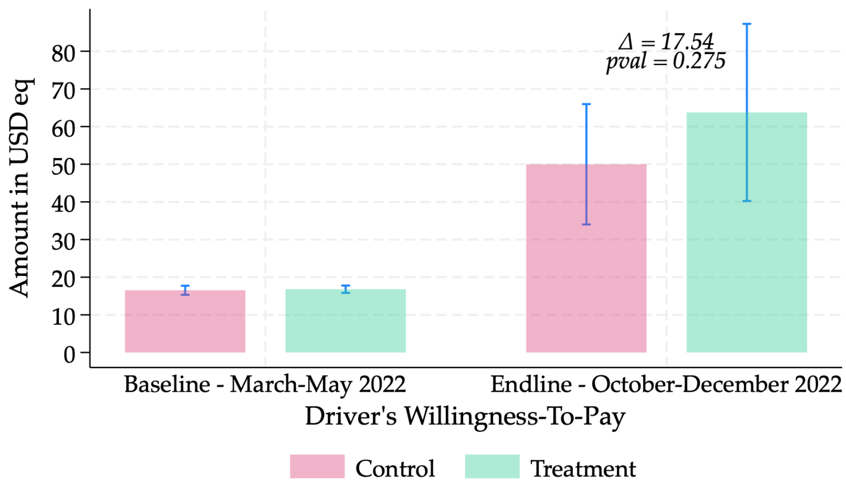


Figure 3: Driver's Willingness-To-Pay (WTP) for Digital Payments

Notes: This figure reports drivers' willingness-to-pay (WTP) for the digital payments bundle. At baseline, WTP was elicited using the Becker–DeGroot–Marschak (BDM) mechanism (Becker et al., 1964). To preserve experimental assignment, the BDM draw was implemented for a 5% lottery of treated drivers only. Following Dizon-Ross and Jayachandran (2022), WTP for a benchmark good (a bottle of water) was also elicited, which I use as a proxy for general price-stating tendencies and response noise. At mid-term, the elicitation was not incentivized because the technology could not be withdrawn from treated drivers and was diffusing outside the experimental sample; treated drivers were asked their WTP to *keep* the bundle and control drivers their WTP to *obtain* it. Values are reported in USD using an exchange rate of CFA 600 per USD.

Willingness to Pay (WTP). I elicited drivers' willingness to pay (WTP) for the digital payment bundle (QR card, business app, and onboarding/training) at baseline and again at the mid-term follow-up. At baseline, WTP was elicited using the Becker–DeGroot–Marschak (BDM) mechanism: each driver stated the maximum amount he would be willing to pay for access to the bundle; under BDM, a price is then drawn at random and the offer is implemented only if the stated WTP weakly exceeds the draw. To keep the treatment assignment essentially intact, I implemented this mechanism only for a 5% lottery of treated drivers, so that a low valuation could (rarely) translate into being placed on the waitlist rather than receiving the technology.[17]

At mid-term, full incentivization was infeasible: withdrawing the technology from treated drivers was not operationally acceptable to the payment company, and by that stage the product was diffusing beyond the experimental sample. I therefore elicited *stated* WTP: treated drivers reported the maximum one-time amount they would pay to keep the technology, and control drivers the maximum they would pay to obtain access. These mid-term valuations are informative about perceived value but are not incentive compatible, so I interpret them cautiously.[18]

Figure 3 shows three patterns. First, mid-term stated WTP is economically large—about one

[17] The draw was a predetermined random number between 0 and CFA 3,000, typically below stated baseline WTP; only one treated driver was therefore moved to the waitlist.

[18] Two opposing forces are plausible: (i) respondents may shade WTP downward if they fear that reported values could influence future pricing by the provider—from anecdotes, the dominant mechanism; (ii) respondents may inflate WTP due to experimenter-demand effects or goodwill toward a project that provided a free service.

week of profits—for a product that was provided free of charge. Second, WTP rises sharply from about USD 17 at baseline to USD 58 at mid-term (a 241% increase). Third, treated and control drivers report similar mid-term WTP ($p = 0.275$). This convergence does not imply that treated drivers failed to learn from use. Rather, it is consistent with rapid informational diffusion: by mid-term, many control drivers had substantial exposure to the technology through peers and passengers (and the broader market rollout), so valuation may reflect shared beliefs about benefits and network effects rather than purely "learning by using."

Finally, I assess whether stated valuations map to economically meaningful frictions by relating WTP to cash-related costs. Table B9 shows that cash frictions predict WTP in both waves: drivers who report larger small-change shortages and related delays also report higher WTP for digital payments (e.g., delays from small-change shortages are associated with a 20% higher baseline WTP). Controlling for benchmark-good WTP reduces noise but leaves the substantive patterns unchanged, supporting the interpretation that stated valuations track heterogeneity in perceived benefits from reducing cash frictions rather than generic overstatement.

4.2 Impact Experiment: Digital Observability Improves Efficiency

This section tests whether making digital transactions observable reduces information asymmetries, improves contract efficiency, and reshapes owner–driver relationships. The underlying mechanisms are formalized in Section 5.

4.2.1 Estimation Strategy

The estimation strategy follows the main specifications outlined in the pre-analysis plan (PAP). Here, I employ business-level regressions to measure the impact of digital observability on owner-driver relationships. For owners with multiple drivers (12%), only the longest relationship is retained; results are robust to alternative specifications, including those based on other drivers.

$$y_j = \beta_0 + \beta_1 T_j^{GranularObs} + \beta_2 T_j^{CoarseObs} + \beta_3 T_j^{NoObs} + \alpha_s + \epsilon_j \tag{2}$$

with $T^{GranularObs}$, $T^{CoarseObs}$, and T^{NoObs} denote the observability arms. No additional covariates are included unless specified.

Equation (2) captures the causal effects of observability on contracts compared to control; I also report F-tests of $\beta_1 = \beta_3$ to isolate observability effects relative to access alone and discuss both.

Survey Waves and Differential Separation Rates. Outcomes from short-, mid-, and long-term surveys are examined. A key empirical challenge is the differential pair separation rates across treatment arms, with separation itself an outcome. To address this, because owners with observability accessed transaction histories for both current and newly hired drivers, I analyze data at the business level, including both existing and newly formed pairs where drivers were replaced. The focus is on mid-term results, allowing enough time for contracts to evolve while ensuring that a significant number of owner-driver pairs remain together. In the analysis below, I show that results remain robust for surviving pairs only, and after two years—when 35% of owners exited the industry or became sole proprietors—focusing on separation as the main outcome.

4.2.2 Reduced-Form Results

This sub-section traces a simple mechanism: digital payments do not only lower cash frictions—they also act as a monitoring technology. I first show that observability increases the information owners have about drivers' activity, both in owners' self-reports and in administrative correlations between digital transactions and work. I then examine whether this additional information changes behavior and contracts. Under *Granular Observability*, drivers exert more effort, as indicated by mystery audits, lower rent default, and higher platform usage. Owners respond by adjusting compensation—most notably by increasing upfront payments—while the weekly rent target remains rigid. These changes coincide with fewer owner–driver separations and, for some owners, more hiring over time, consistent with observability alleviating trust and enforcement frictions in employment relationships. Taken together, the results suggest that even imperfect observability—making effort partially contractible—can generate meaningful efficiency gains.

Addressing Information Asymmetries. Digital payments give owners new information on drivers' effort and digital revenue. Figure A2 illustrates the data available under full observability (e.g., transaction timestamps, days worked). To test this "first-stage" more formally, I use two strategies.

First, I measure owners' reported knowledge. Table B10 shows that *Granular Observability* increases the likelihood that owners claim to know their driver's digital revenue by 14 pp (a 67% increase, as shown in Column 3). It also raises the probability that owners correctly guess days worked by 11 pp, but this estimate is imprecise and not statistically distinguishable from zero (37%). Results on *Coarse Observability* are statistically inconclusive. These results suggest digital transactions provide only *imperfect* signals of drivers' work.[19] Overall, 36% of owners with *Granular Observability* report using the technology to monitor drivers' effort, and 43% check their driver's transactions daily or weekly.

Second, I compare administrative digital payment data with drivers' self-reported effort and output. Table B11 shows strong positive correlations: digital transactions and revenue are highly predictive of days worked among treated drivers. For example, one additional digital transaction predicts 0.9 more passengers that day.

Taken together, these results show that observability provides valuable, though imperfect, signals of drivers' work, with granular observability yielding the largest information gains.

[19] A cultural and religious norm in Senegal—particularly the Islamic principle that one should not speak without certainty—led many owners to respond "don't know" when asked about their perceptions of the driver's work.

Increase in Worker Effort. Measuring effort in informal sectors is challenging due to the inherent difficulty for the principal to observe it. Moreover, self-reports are often subject to social desirability bias and recall bias: 41% of drivers found it hard to recall performance over the past three days, and hours worked are noisy.[20]

I therefore rely on three complementary measures: (a) mystery audits, (b) rent default, and (c) digital usage. Together, they consistently indicate higher effort under *Granular Observability*.

Table 3: Impact of Observability on Effort (Mystery Passengers Audit Survey)

	Count	Unique Days	Count Per Day
	(1)	(2)	(3)
Granular Observability	0.324***	0.250***	0.089***
	(0.098)	(0.090)	(0.032)
Coarse Observability	-0.067	-0.130*	0.054**
	(0.083)	(0.072)	(0.026)
No Observability	0.070	0.026	0.052*
	(0.081)	(0.073)	(0.028)
Observations	588	588	588
Control Mean	0.370	1.324	1.010
Enumerator FEs	YES	YES	YES
Chi-squared test Granular O = No O (p-value)	0.01	0.01	0.29

Notes: Business-level Poisson regressions of y_j on treatment dummies for *Granular*, *Coarse*, and *No Observability*, with strata (α_s) and enumerator fixed effects. Mystery passenger data were collected over two weeks in August 2022 by trained surveyors who bargained with drivers and discretely recorded license plates, later matched to driver data and treatment group (see Section 3.2). In Column (1), the sample is limited to businesses that provided a license plate, 18 refused and 3 gave duplicates, so these were excluded. Business-level Poisson regressions are also reported for taxis driven by employees to separate owner and driver effort, with strata, batch, and enumerator fixed effects included. Heteroskedasticity-robust standard errors are reported, with significance denoted as * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. A Chi-squared test for equality of *No Observability* and *Granular Observability* coefficients is reported at the bottom. Outcomes: *Count* = number of times a taxi (by license plate) was audited, imputed as 0 if never observed, *Unique Days* = number of distinct days a taxi was audited, *Count Per Day* = average audits per unique day.

First, mystery passenger audits provide a direct measure of effort: the more often a driver is spotted on the road, the more likely they exert higher effort. Table 3 uses Poisson regressions to show that taxis under *Granular Observability* are seen on the road 38% ($\beta = 0.32$) more often than controls (p=0.001), and significantly more than under *No Observability*. Effects are concentrated on the extensive margin (days observed) and absent under *Coarse Observability*. This suggests *Granular Observability* increases effort during the two-week audit.[21] Importantly, audits also confirm the absence of manipulation in digital payment usage: drivers consistently accepted digital payments when requested, with no evidence of gaming across treatments (Figure A3).

[20]Nonetheless, I collected self-reported data at short- and mid-term. FE models show a small increase in hours worked under *Granular Observability*, though not statistically different from *No Observability* (see Table B18). Other performance measures (customers, total revenue) show no significant differences.

[21]When disaggregated, the largest effect (+369%) appears in Dakar Plateau—an affluent district with government offices and businesses—albeit from a smaller sample of observed taxis. This suggests drivers may not only work more but also shift toward areas with higher smartphone penetration.

Second, rent defaults decline. To mitigate drivers' self-report biases, I asked both owners and drivers about the frequency of default on the rent over the past three months, and created a dummy variable for whether drivers defaulted at least once a month (31% of drivers in the mid-term). Table 4 shows that partial default decreases by 10 pp, a 34% reduction under *Granular Observability* (p=0.068). This reduction, compared to *No Observability*, highlights the direct gain for owners in monitoring their employees: they can encourage increased effort, thereby raising the frequency of high-revenue weeks and reducing default. Owner's profit under *Granular Observability* increases by 6% (or by 11% when maintenance costs are included; this alternative measure is particularly noisy because many owners were unable to report these costs, see Table B13), but the estimate is imprecise and therefore statistically insignificant.

Table 4: Impact of Observability on Default and Owner's Profit

	Monthly Default Rate	Total Transfers to Owner (USD)	Owner's Monthly Profit	Owner's Monthly Profit (Wins.)
	(1)	(2)	(3)	(4)
Granular Observability	-0.104*	28.595**	17.571	17.374
	(0.057)	(13.288)	(13.555)	(13.458)
Coarse Observability	0.038	-9.750	-13.112	-13.245
	(0.065)	(15.169)	(15.228)	(15.154)
No Observability	0.017	-4.214	-8.486	-7.393
	(0.061)	(14.353)	(14.620)	(14.320)
Observations	474	474	470	470
Control Mean	0.31	336.19	277.62	278.21
% Change Gran. Obs.	-33.80	8.51	6.33	6.25
F-t Gran. O = No O (p-val)	0.08	0.04	0.11	0.12

Notes: Each column reports coefficients from OLS regressions of the outcome on indicators for the three treatment arms, with the control group omitted. All regressions include strata and batch fixed effects and heteroskedasticity-robust standard errors. Significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Outcomes are measured at midline (7–9 months after baseline). Do not know and refusal are coded as missing. Baseline profit values are included.
All monetary values are in USD using 600 CFA/USD. The *Monthly Default Rate* equals 1 if the driver failed to remit at least one expected rent payment in a given month, based on owner or driver reports (past three months). *Total Transfers to Owner* is imputed based on information from a non-random subset of owners who reported losing, on average, 2.2 of four weekly transfers during months with a default (mid-term surveys did not collect transfer amounts).
Owner's Profit is defined as monthly transfers received (net of imputed defaults) minus the upfront payment (salary) and the lost profit when the pairs separated. Column (3) uses the baseline definition and excludes maintenance, a noisy measure that many owners could not estimate. Column (4) winsorizes this measure at the 2nd and 98th percentiles.

Third, digital usage rises. Drivers under *Granular Observability* process 35% more transactions (p=0.012), on 18% more weeks and 26% more days (Table B14). Controlling for pre-trends using pre-experiment peer-to-peer transactions does not substantially alter these results. Usage under *Coarse Observability* initially increases but fades after six months, suggesting drivers might have initially attempted to signal effort, but stopped when it did not lead to contract changes. I find no evidence of a ratchet effect or manipulation: drivers do not use digital transactions more under

Coarse compared to *Granular Observability*.[22]

Taken together, these findings show that drivers exert more effort under *Granular Observability*, consistent with the fact that this treatment reveals a signal of worker effort to owners, contrary to *Coarse Observability*, where I do not find such an effect.

Contract Changes. I next examine the effect of observability on contract form, focusing on two parameters: owners' upfront payments and drivers' rental transfers. Table 5 shows that owners with *Granular Observability* are significantly more likely to offer an upfront payment to their drivers, referred to as a "salary" in the taxi industry. In the mid-term, 76% of owners provide such payments, a share that rises by 12 pp (16%) under observability (p=0.006).[23] The F-test comparing *Granular* and *No Observability* confirms the difference (F-stat = 6.5). This change occurs primarily at the extensive margin, with values increasing by 19% over the control group, rather than through adjustments to existing payments. These results remain robust to contamination bias, corrections for multi-arm trials, addressing non-convex averages of other treatment effects as recommended in Goldsmith-Pinkham et al. (2024) (Table B15). Observability impacts contracts for both existing and newly hired drivers, as treated owners can now monitor transactions, see Table B16; estimates for newly formed pairs are imprecise but typically larger, though I interpret them cautiously since remaining together is itself an outcome.

The upfront payment compensates drivers for the increased effort required when adopting the technology, as formalized in the theoretical framework (Section 5). It may also serve a working-capital function—helping drivers finance day-to-day operating expenses (e.g., fuel) that rise mechanically with longer hours—so that the contract adjustment both covers higher operating needs and sustains higher effort. In a weak enforcement environment with relational contracts, this adjustment is more credible than alternatives such as an ex-post bonus or a reduction in the weekly rent target. An ex-post bonus is harder to enforce once effort has been exerted. Likewise, lowering the rent target is rarely feasible given the rigidity of industry norms—over 72% of contracts fix rent at CFA60,000 weekly. By contrast, an upfront payment prevents owners from reneging and directly addresses the commitment problem.

Consistent with this, I find no change in the agreed rental fee (Table 5, Col. 3), while rent default decreases, as previously shown. Overall, the upfront payment means drivers' monthly profits increase on average, particularly under Granular Observability (Table B21)—though this measure excludes non-monetary costs of effort. Section 6 estimates the welfare effects accounting for effort costs.

[22] As a robustness check, I examine whether drivers strategically misreport by disguising personal P2P transfers as taxi payments ("taxi-like" transactions), see Table B20. If this were the case, P2P taxi-like transfers should fall when business transactions rise. I find no such evidence: taxi-like P2P transactions, though about three times fewer than business ones on average (22.42 vs. 59.31), actually rise by 11%. While not significant, the effect is positive across outcomes and smaller than for business transactions. All other P2P transfers also rise, though this is harder to interpret since some upfront payments are sent via mobile money.

[23] The higher control mean at 9 months may reflect an increase in drivers' outside options, possibly due to the citywide bus rapid transit construction underway during the experiment.

Table 5: Impact of Observability on Contracts and Relationships

	Upfront Payment 'Salary' Dummy	Upfront Payment 'Salary' Value (USD)	Weekly Rent Target Value (USD)	Separation p
	(1)	(2)	(3)	(4)
Granular Observability	0.124***	10.464***	3.082	-0.109**
	(0.045)	(3.515)	(1.927)	(0.050)
Coarse Observability	-0.013	0.442	0.562	-0.041
	(0.052)	(4.151)	(1.552)	(0.054)
No Observability	-0.009	0.468	0.764	-0.028
	(0.050)	(4.000)	(1.570)	(0.053)
Observations	476	468	470	572
Control Mean	0.76	55.62	99.96	0.32
% Change Granular Observability	16.38	18.81	3.08	-33.98
F-test Granular O = No O (p-value)	0.01	0.02	0.29	0.18

Notes: Business-level OLS regressions of contract outcomes on indicators for the three treatment arms, with the pure control group omitted. The model is $y_j = \beta_0 + \beta_1 T_j^{GranularObs} + \beta_2 T_j^{CoarseObs} + \beta_3 T_j^{NoObs} + \alpha_s + \epsilon_j$, where y_j is the outcome and α_s the strata and batch fixed effects. Estimates use mid-term data (about 9 months post-baseline). Contract outcomes are recorded only for businesses that still employ a driver at mid-term, whereas Separation is coded for every baseline owner–driver pair, explaining the sample size difference. Standard errors are heteroskedasticity-robust. Significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Data is missing if the respondent refused to answer.
Outcomes: *Upfront Payment Dummy* = 1 if the owner provides a monthly fixed payment (salary), *Upfront Payment Value* = amount of that payment in USD (0 if none, USD 1 = CFA 600), *Weekly Rent Target* = agreed weekly rental payment (USD), *Separation* = 1 if owner and driver no longer work together at survey time.

As a robustness check, I interact observability with drivers' digital usage. Table B19 shows that the effect on upfront payments is stronger among the top 50% of users, by about 9 pp, though the interaction is not significant. While endogenous, this pattern supports the interpretation that contract changes are concentrated where digital usage is most intensive.

Reduction in Owner-Driver Separations. Separation rates in the taxi industry are high: over the first nine months, 33% of pairs split, for reasons ranging from defaults and disputes to owners selling taxis or long repairs (see Fact 4).[24]

Table 5 shows lower separation under *Granular Observability* after nine months and two years. Pairs with a driver using the technology are less likely to split, primarily driven by the *Granular Observability* arm. Pairs in this arm are 11 pp (p=0.030) less likely to break than control, a reduction of 34%. Although large, this difference is not statistically significant compared to *No Observability* (F-test p=0.18).[25] The effect of observability on separation, concentrated under *Granular Observabil-*

[24]Since separations occur for varied reasons, I focus on overall separation; analyzing specific reasons yields no significant differences across treatment arms.

[25]To shed light on mechanisms—and because I do not observe what occurs at each driver default—I examine whether effects are stronger when drivers use the digital platform intensively (Table B19). Indeed, high-usage drivers stay longer, by 3pp, though the estimate is imprecise. This regression is subject to the same endogeneity concerns as above, since the observability treatment also affects drivers' usage.

ity, can be explained by the fact that the technology enabled better monitoring of effort, reduced moral hazard, leading to fewer separations. I also find suggestive evidence that *Coarse Observability* reduces separations by 5 pp after two years (Table B17). Although highly suggestive, this may indicate that low output observability may become beneficial as the app gets used more.

Digital Observability Increases Hiring by Owners. I investigate whether observability affects hiring, to causally test the hypothesis that effort contractibility influences firm size (Baker and Hubbard, 2004). To do so, I combine the sample of owner-driver pairs with owners driving their taxis alone, who were also randomized into observability arms with treatments applying to any future hires.[26] By doing so, I increase the sample size and test how observability might influence hiring decisions, particularly for owners typically reluctant to hire drivers.

Table B22 shows that owners under *Granular Observability* are 6 pp and 2 pp more likely to hire a driver after nine months and nearly two years (40% and 10% increase), respectively, compared to those under *No Observability* (the pure control group is excluded since they received treatment after nine months). The effect is strongest at mid-term and concentrated among owners with a driver at baseline, consistent with observability alleviating mid-run hiring frictions by enabling owners to establish trust with new drivers more rapidly.[27]

Increase in Trust and Value of Relationships. I show that observability increases trust in owner-driver relationships. Given the multifaceted nature of trust, I employ two strategies: survey-based measures (World Values Survey, 2017) and heterogeneity by baseline business characteristics, pre-specified in the PAP.

Under *Granular Observability*, owners are 28.1% more likely to allow the driver to park the taxi at the driver's home (Table B12, Col. 1), a high-stakes revealed-preference measure of trust in this setting: owners who distrust drivers typically require parking at a specific monitored location as an existing monitoring strategy, to verify daily attendance and maintain oversight. Moreover, owners are 49.2% less likely to attribute low earnings to low driver effort rather than bad luck (Col. 2), consistent with a decline in perceived moral hazard as effort becomes more verifiable. Finally, self-reported owner's trust moves in the same direction but is statistically indistinguishable from zero (Col. 3).[28] Overall, the pattern across outcomes—greater delegation and lower perceived moral hazard—points to economically meaningful trust-related behavior when owners can better observe transactions.

[26]For instance, taxi owners with no employees in *Granular Observability* were told: "As taxi owner, you will have access to the digital transaction history of any driver you hire in the future and receive an SMS indicating their total daily transactions." Conversely, owners with *No Observability* were explicitly told that they would not have access to any new driver's transaction history in the future.

[27]The effect is much smaller and statistically insignificant by two years, so I interpret the longer-run hiring results cautiously. A possible explanation is that many pairs split and rematch over time, which adds noise to "hiring" measured at two-year and may attenuate detectable differences across arms in the longer run.

[28]I interpret stated trust cautiously: it is very high on average (9/10), consistent with ceiling effects and social-desirability bias that compress variation and reduce power. The same applies to drivers' stated trust in owners (Col. 4). Also, while a dictator game between owner–driver pairs was implemented at baseline, it was not replicated at endline due to practical difficulty eliciting reliable stated trust from drivers within the survey setting and time constraints.

Second, I examine heterogeneity in the effect of observability on separation rates by baseline characteristics related to trust. Table B23 focuses on three dimensions: relationship length, family ties, and risk aversion.[29] Two findings directly relate to trust. First, separation rates are generally higher in non-family businesses and in recent owner–driver relationships (top row of Table B23), both negatively correlated with trust. Second, the effect of observability is stronger in these pairs, with interaction terms of -3 pp for recent relationships, and -5 pp and -18 pp for non-family pairs at mid- and long-term, respectively.[30] These results, while suggestive, indicate that observability could help retain non-family and recent employees relative to *No Observability*.[31]

Taken together, observability strengthens trust, especially in low-trust pairs, underscoring the role of monitoring technologies in lower-income countries. Information frictions and limited access to these technologies may help explain the persistence of family businesses, which are often linked to efficiency losses (Bertrand and Schoar, 2006; Chandrasekhar et al., 2020).

Taking Stock: Digital Payments as Effective Monitoring Technologies. Digital payments provide owners with information on drivers, leading to higher effort, fewer defaults, contract adjustments, and greater retention. These effects are concentrated under *Granular Observability*, where owners observe a time-stamped transaction trail (frequency and timing), increasing the informativeness of digital activity as a proxy for work intensity. Under *Coarse Observability*, owners observe only a capped daily aggregate; at current penetration rates this summary is a low-precision, demand-driven signal that is often insufficient to distinguish low effort from low digital usage. Consistent with this weaker information content, estimated effects under coarse observability are attenuated and imprecise. With only about 13% of revenue processed digitally on average, the record is therefore most informative for high-effort drivers, for whom frequent time-stamped transactions provide a credible signal of sustained activity even when total revenue remains partially unobserved. The results point to moral hazard in effort—rather than in output reporting—as the binding constraint. Overall, digital payments expand the production frontier by reducing cash costs and moral hazard, thereby improving efficiency for adopting businesses.

4.3 Adoption Experiment: Digital Observability is a Barrier to Technology Adoption

This section tests whether the observability feature reduces the net benefits of the technology by discouraging initial adoption. The adoption experiment estimates the effect of observability on

[29] These dimensions were pre-specified in the PAP (AEA registry ID #0009155); two were also used for stratification. Other heterogeneity analyses discussed in the PAP did not yield significant differences and thus are omitted.

[30] The control group is excluded to allow clean comparisons between *Granular Observability* and *No Observability*, since they received treatment after nine months. Long-term effects are particularly useful here, as heterogeneity tends to widen over time, though limited sample size and high p-values mean results should be interpreted cautiously.

[31] Risk aversion is not directly related to trust but is reported for completeness. I elicited the risk aversion coefficient for all owners and drivers at baseline using an incentivized game with simple choice tasks à la Holt and Laury (2002), tailored to the taxi industry. Drivers were then assigned a relative risk-aversion score, with those above 1 (CRRA utility function) defined as risk-averse agents. Observability effects are larger for non-risk-averse agents (-11 pp and -12 pp at mid- and long-term). This is consistent with a similar logic: pairs with risk-averse agents are more likely to be in established low-risk contracts with low separation rates already.

the likelihood that drivers provide owner contact information. I also compare adopters and non-adopters along baseline characteristics.

Barriers to Adoption of Digital Payments. During the listing survey, most drivers expressed general interest in the technology, but 50.2% refused to provide their owner's contact information even after three follow-ups, preventing adoption. Drivers often mentioned that they needed to talk to the taxi owners before sharing contact information. In particular, 48% of drivers cited privacy concerns or the need to consult owners before sharing details, 15% said owners were unavailable or uninterested, and 20% explicitly mentioned fears that owners would gain access to their digital transaction history (indicating that many reluctant drivers anticipated owner observability as a possible implication of adoption when asked to share owner contact information). Only 5% later changed their mind and entered the impact experiment.

Table 6: Impact of Observability On Technology Adoption

	Technology Adoption (Share Owner's Information)					
	(1)	(2)	(3)	(4)	(5)	(6)
Removing Observability	0.114***	0.111***	0.102***	0.100***	0.196***	0.158***
	(0.038)	(0.038)	(0.037)	(0.037)	(0.049)	(0.060)
Observations	433	433	433	433	204	159
Mean Under Observability	0.143	0.143	0.143	0.143	0.069	0.095
Enumerator FE	NO	YES	NO	YES	NO	NO
Privacy Concern Controls	NO	NO	YES	YES	NO	NO
Relationship Length Control	NO	NO	NO	YES	NO	NO
Sample	All	All	All	All	Poorest	Worst-Performing
% Change Removing Observability	80	77	71	70	284	166

Notes: Survey data were collected June 15–July 7, 2022, from drivers who refused to provide their owner's contact during listing. Driver-level regressions are estimated with heteroskedasticity-robust standard errors, significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The outcome *Adoption* equals one if the driver provided the owner's contact to the surveyor, thereby enabling adoption of digital payments. Controls include enumerator fixed effects, length of the driver–owner relationship (years), and privacy concern dummies (listed at the bottom of the table).
Privacy concern controls are: (i) acceptance of a friend's phone number for follow-up, (ii) acceptance of a garage/taxi association name (when relevant), and (iii) acceptance of a license number. These capture general information-sharing concerns beyond observability of contracts. "Poorest" drivers are those below the median wealth PPI score, "worst-performing" drivers are those with an average productivity index z-score below zero.

Adoption Experiment: The Role of Observability in Technology Adoption. To isolate the role of digital observability, I re-offered the technology to reluctant drivers about a month after their initial refusal (see Section 2.3, Figure 2).

Table 6 shows that removing observability has a large positive effect on adoption. In the control group, where owners could see transactions, only 14% of drivers changed their minds and provided owner contacts. Removing observability almost doubles this share (+80%, p=0.003). The result is robust to surveyor fixed effects, privacy controls (e.g., willingness to share alternative contacts like that of their association president or closest friend), and relationship length.

These findings highlight a trade-off: while observability improves profits and contracts, it also deters adoption. This adoption response is central for interpretation: it shows that observability is not merely correlated with take-up, but causally shifts the adoption margin in this setting. The next subsection characterizes the drivers for whom this observability-sensitive margin is most salient. This is particularly relevant in sectors where employees have some degree of autonomy in deciding whether to adopt the technology. For instance, employees of small businesses may play a crucial role in informing their employers about new technologies available in the market. Assurances that adoption does not entail data sharing substantially increase take-up.

The effect may be underestimated, as some drivers may not have fully trusted the research team that observability would be removed. Indeed, regressions by reason for refusal (Table B25) show significant effects even among drivers who had not cited observability, while adoption remained incomplete for those most concerned. Consistent with this, administrative data show that 88% of these reluctant drivers eventually adopted the technology by 2024, once the payment company made non-observability the default (as discussed in Section 6.4.3).

Profile of Reluctant Drivers: High-Disutility of Work, Low-Performing, and Poorest Drivers. To characterize selection, I follow the spirit of Karlan and Zinman (2009) and compare drivers in the "impact experiment"—who accepted potential observability but whose owners were randomly assigned not to observe (*No Observability* or *Control*; hence "willing to adopt") to "reluctant drivers," who initially refused to share owner contacts. I further exclude the 14% of "reluctant drivers" who later provided owner contacts under *Granular Observability*. This approach is used because many reluctant drivers refused the extended baseline survey, so I rely primarily on mid-term data for this comparison (I also estimate specifications that control for whether the reluctant driver has the technology at mid-term, and the patterns are unchanged). The goal is to characterize selection into adoption.[32]

Their wealth index (IPA, tailored to Senegal) is significantly lower (-8%), and they are less likely to have completed primary school (28%) or be literate (10%) (Panel D of Table B24). They are also more stressed about meeting rental payments (+83%) and more risk-averse. In the three days before the survey, they report fewer passengers (-11%), lower revenue (-4%), and fewer hours worked (-4%), resulting in a lower z-score performance index (Panel A). They are more likely to already receive an upfront payment—consistent with tighter liquidity constraints—while their mid-term separation rate is similar to adopters (Panel B).

[32] Accordingly, the estimates describe selection among *initial refusers* of the bundled offer, not selection in the full driver population. This is a relevant margin for adoption in this setting: drivers in the impact experiment accepted the technology when it was offered with potential owner observability (and were then randomized into whether observability was actually provided). By doing so, these drivers reveal that adoption is not a binding constraint for them: removing observability cannot raise their take-up because they already adopted. The key selection is therefore concentrated among those who initially refuse when observability is bundled. Re-offering the same technology without observability to these refusers isolates the observability-sensitive adoption margin.

These differences admit at least two interpretations. One is that improved observability would compress informational rents by making effort and/or revenue more verifiable, lowering the attractiveness of adopting a technology that bundles monitoring with payments. A second interpretation is that more risk-averse or less educated drivers could be more uncertain about how a new technology—and especially owner observability—will affect them, and therefore adopt more cautiously in a "wait-and-see" manner. The adoption experiment helps discipline this alternative: holding fixed the payment technology and varying only the observability regime at onboarding, removing owner access increases adoption sharply, with effects more than twice as large among the poorest and worst-performing drivers (Col 5 and 6 of Table 6). This pattern is hard to explain by generic technology reluctance alone, and instead suggests that anticipated data sharing with owners is a first-order determinant of non-adoption among disadvantaged drivers.

These patterns are particularly relevant for welfare: those with possibly the highest marginal utility from reducing the hassle costs of cash are precisely the least willing to adopt when observability is bundled with the technology.

High-Performing Drivers Prefer Observability. I complement the selection patterns above with stated-preference evidence among adopters, which helps assess whether attitudes toward observability covary with objective performance and with owner misperceptions—addressing the concern that reactions to observability may reflect mistaken beliefs. Among adopters (preferences were not elicited from reluctant drivers to limit survey length), both owners and drivers ranked Granular, Coarse, and No Observability without knowing whether rankings would be used, reducing bias; these rankings did not affect random assignment. Most drivers preferred *No Observability* (59%), but 23% favored *Granular Observability*. Owners were evenly split, often citing concern about drivers' reactions to monitoring.

I find that high-performing characteristics significantly predict drivers' preferences for observability, consistent with the idea that such drivers expect observability to benefit them. Table B26, Panel A, shows that drivers with higher daily revenue, more days worked per week, and fewer defaults significantly more likely to prefer observability. Drivers with more high-performance days are 10 pp more likely to prefer observability, while low performers are 8 pp less likely. Consistent with this, drivers preferring observability have longer relationships with their employers. Panel B shows that drivers are more likely to prefer observability when their employers underestimate their work. This pattern suggests that drivers anticipate observability will help correct owners' biased beliefs, thereby building trust and improving retention.[33]

Long-Term Worker Retention Across Groups. Figure A4 shows that reluctant drivers—those explicitly refusing adoption due to observability concerns but later adopted when the technology was re-offered without owner observability—exhibit the highest separation rate after nearly two years (68%), above the overall mean of 60%. Pairs under *Granular Observability* have the low-

[33]In addition, drivers favoring observability at baseline exhibit larger treatment effects of observability on retention.

est turnover (55%), followed by *Coarse Observability*.[34] These retention gaps across groups have important welfare implications, explored in Section 6.

5 Theoretical Framework: Impact and Adoption of Digital Payments

The experiments establish two facts about the same digital tool. Among adopters, making digital transactions observable changes incentives and contracts: effort rises, defaults fall, and compensation adjusts. Meanwhile, anticipated observability discourages adoption when worker cooperation is required, especially among disadvantaged drivers. These findings pose a central evaluation problem: how large are the efficiency gains from embedded observability relative to the welfare losses from reduced diffusion—and how are these effects distributed across workers and firms?

This section develops a contracting model that (i) formalizes the mechanisms behind the main empirical facts and (ii) provides a basis for the quantitative welfare and distributional accounting in the structural estimation. Drawing on insights from relational contracting (Baker et al., 2002; Levin, 2003) and the sharecropping literature (Banerjee et al., 2002), I model the owner–driver relationship as a simple principal–agent problem in which effort and output are imperfectly observed, the agent faces limited liability, and contracts are enforced relationally.

Three mechanisms emerge. First, *effort becomes more contractible*: when digital usage is sufficiently frequent, the transaction trail provides a credible signal of high work intensity, reducing moral hazard in effort. Second, *reported revenue becomes more verifiable*: the digital component limits the scope for under-reporting. Third, *adoption becomes selective*: because improved observability makes it privately optimal for owners to demand higher effort (and because owners cannot credibly commit not to "ratchet" future effort requirements), some drivers anticipate a shift toward a more demanding relationship and therefore prefer not to adopt.

The model's value lies in the interaction of these mechanisms in an environment with weak enforcement, limited liability, and asymmetric information. *Weak enforcement* limits the use of ex post bonuses and prevents credible commitment to future terms. *Limited liability* and *information asymmetries* about effort and revenue create informational rents that owners compress once monitoring improves. Together, these features rationalize why observability induces upfront compensation and changes in retention among adopters, rather than simply lower ex post rents—and why the same information that strengthens incentives can deter adoption for a subset of drivers.

5.1 Setup

Consider an environment with an infinite discrete time horizon, with periods indexed by $0, 1, \ldots, \infty$. Both principal and agent share a common discount factor, $\delta < 1$. The principal aims to incentivize the agent to exert effort. Effort takes discrete values, $e \in \{0, 1, 2\}$, and is unobservable to the principal at baseline. Output $y(e)$ is assumed binary as follows:

[34] Reluctance is here defined using drivers' stated reasons for refusing to share owner contact. Consistent with an observability mechanism, drivers who refused for reasons unrelated to observability have lower separation rates.

$$y(e) = \begin{cases} Y & \text{with probability } q_e \\ X & \text{with probability } 1 - q_e \end{cases} \tag{3}$$

with $X < Y$, and $q_2 > q_1 > q_0$. This production function captures the high output uncertainty in the taxi industry, where even high effort does not guarantee high revenue (see Section 1.2.2).

Agents differ by type $\theta \in l, h$, reflecting the disutility of effort $\phi^{\theta}(e)$, with $\phi^{l}(e) > \phi^{h}(e)$ and $\phi^{\theta}(0) = 0$. Types are assumed public because adoption decisions are made largely within existing owner–driver relationships; this shows that adverse selection is not required for the main results, though the framework could be extended to incorporate private types and screening. Agents are referred to as low- and high-ability types.[35] The agent's utility is denoted by U^{θ}, with per-period component agent's revenue collected y, minus transfer t and disutility cost of effort $\phi^{\theta}(e)$.

The baseline contract revolves around two endogenous variables: the transfer $t(\tilde{y})$ the agent remits at the end of the period and the continuation probability $p(\tilde{y})$. Both are functions of the agent's reported output, $\tilde{y}$. If the relationship ends, both the principal and the agent incur one-time replacement costs, K_p and K_a, respectively, and are rematched from an infinite pool of players. The pool of unmatched agents consists of a fraction μ of high types, $1 - \mu$ of low types, with μ known to the principal. To keep the exposition simple, I assume μ is constant over time, reflecting a setting where the stock of new agents is large. The agent can always exit the taxi industry and take an outside option $\bar{u} > 0$.[36] The principal always has the outside option to sell the car and stop working in the taxi industry.

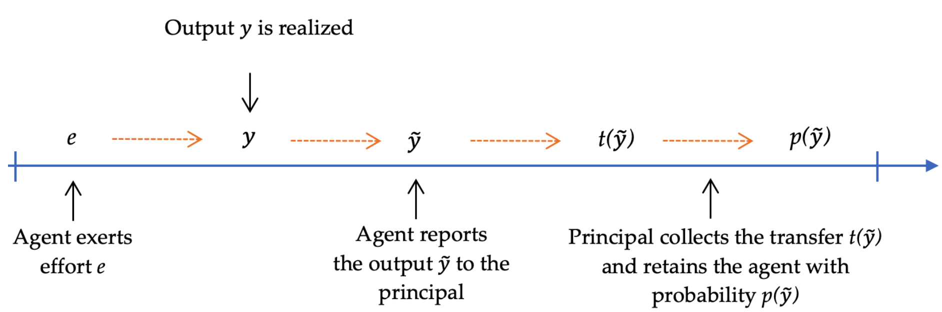


Figure 4: Timing of a Principal-Agent Relational Contract Period

Notes: Timeline of events within one period of the dynamic game.

5.2 Assumptions

Drawing on survey evidence and empirical facts documented in Section 1.2.2, I make the following assumptions:

[35] For simplicity, this section does not delve into what determines these types. However, it is empirically observed that low-ability drivers are typically poorer, a population of high policy relevance given the welfare implications discussed in the results.

[36] For simplicity, I assume $\bar{u}$ is identical across types. In the structural estimation section, I relax this assumption and allow for $\bar{u}^{h} > \bar{u}^{l}$.

Assumption 1. ***Unobservability.*** *Agent's effort e and output y are unobservable to the principal.*

While unobservable effort is standard in contract theory, I also assume output is unobservable—a prevalent feature in many informal settings. As shown in Fact 3, the principal lacks direct information about the agent's actions, relying on reported output $\tilde{y}$. This is discussed in static contract models such as Townsend (1979); Lacker and Weinberg (1989); de Janvry and Sadoulet (2007).

Assumption 2. ***Limited Liability.*** *The agent faces a constraint $t(\tilde{y}) \leq \tilde{y} \leq y(e)$, ensuring that the transfer t does not exceed reported output and reported output does not exceed actual output.*

This reflects Fact 2: many drivers default, lack savings, and cannot access credit. By assumption, they cannot report more than what they collected (see Innes (1990)).

Assumption 3. ***Stationary Equilibria.*** *When deciding the contract for the next period, the principal relies exclusively on current period reported output.*

The contract retains the same contingent compensation and termination scheme each period, abstracting from history dependence. Given the complexities of optimal relational contracting under assumptions suited to lower-income settings—such as limited liability and unobservable output—I restrict attention to stationary equilibria.[37] I thus omit the time subscripts for simplicity. This matches practice: rental target payments are typically fixed over time.

Assumption 4. ***Risk-neutrality.*** *Both principal and agent are risk-neutral.*

This assumption is made for tractability, as the theory literature has not extensively explored risk aversion within relational contracts.

The framework is a two-stage game solved by backward induction. In Stage 1 ("adoption"), *the agent* decides whether to adopt the technology. In Stage 2 ("impact"), *the principal* offers a contract $t(\tilde{y}), p(\tilde{y})$, subject to the agent's constraints. In line with practices in this industry, I assume that contracts are "take-it-or-leave-it" offers, and both the owner's and driver's participation constraints must be satisfied for the relationship to form.

5.3 Baseline Contract Without Digital Payments

Figure 4 shows the timing of events in one period of this dynamic game.[38] The principal maximizes expected transfers and the discounted value of the relationship. The objective functions of the principal V^θ when matched with agent of type θ can thus be written:

[37] This can be interpreted as a reduced form representation of richer history-dependent strategies that condition on past performance yet generate similar *average* separation hazards following low realizations. Fully characterizing optimal non-stationary contracts would require (i) taking a stand on additional institutional features governing how owners use retrospective information in practice, and (ii) enlarging the state to include history-dependent objects—e.g., promised continuation utilities, cumulative performance histories, and (under incomplete information) posterior beliefs—which would substantially expand both the theoretical solution and the empirical objects required for estimation. See, e.g., Andrews and Barron (2016); Fong and Li (2017) for non-stationary relational contracting environments.

[38] This sub-section draws partly from Kelley et al. (2024), but departs in two ways: (1)I model a two-stage game with heterogeneous agent types θ, to study selection and (2) I relax the role of risk-taking to focus on dynamics around effort.

$$V^{\theta} = \max_{t,p,e} \mathbb{E}[t(\tilde{y}) + \delta[p(\tilde{y})V^{\theta} + (1 - p(\tilde{y}))(-K_p + \mu V^h + (1 - \mu)V^l)]|e] \tag{4}$$

This optimization is subject to the following constraints:

$$\begin{cases} \mathbb{E}[(y(e) - t(\tilde{y})) - \phi^{\theta}(e) + \delta U^{\theta} - \delta K_a(1 - p(\tilde{y}))|e] \geq \max\{-\delta K_a + \delta U^{\theta}; \bar{u}\} & \text{Participation Constraint } \textit{(IR)} \\ e \in \arg\max_{\tilde{e}\in\{0,1,2\}} \mathbb{E}[y(e) - t(\tilde{y}) + \delta U^{\theta} - \delta K_a(1 - p(\tilde{y}))|\tilde{e}] - \phi^{\theta}(\tilde{e}) & \text{Incentive Compatibility } (IC) \\ t(\tilde{y}) \leq \tilde{y} \leq y(e) & \text{Limited Liability } \textit{(LL)} \\ Y - t(Y) + \delta(U^{\theta} - K_a(1 - p(Y))) \geq Y - t(X) + \delta(U^{\theta} - K_a(1 - p(X))) & \text{Truth-Telling } \textit{(TT)} \\ y(e) - t(\tilde{y}) + \delta(U^{\theta} - (1 - p(\tilde{y}))K_a) \geq y(e) + \delta(U^{\theta} - K_a) & \text{Dynamic Enforceability } \textit{(DE)} \end{cases}$$

Here (IR) ensures participation, (IC) incentive compatibility, (LL) limited liability, (TT) truthful reporting of output, and (DE) dynamic enforceability. The LL constraint implies transfers occur ex-post (end of the period) rather than upfront. (DE) ensures the agent prefers to continue the relationship rather than walk away with the output; a symmetric constraint applies to the principal. Unlike standard models, (TT) ensures truthful reporting of collected output because output is unobservable.[39]

Lemma 1 (in Appendix) shows that under full observability the owner would induce optimal effort, pay an upfront fixed compensation that makes the agent indifferent between working and his outside option, with no termination occurring in equilibrium. With both effort and output unobservable, the best stationary contract is:

Result 1. (Baseline Contract Without Digital Payments) *Under Assumptions 1–4, $\exists\, \underline{K}_p < \bar{K}_p$, $\underline{\delta} < \bar{\delta}$, s.t. when $K_p \in (\underline{K}_p, \bar{K}_p)$ and $\delta \in (\underline{\delta}, \bar{\delta})$, the principal's best type-dependent stationary contract is:*

$$\bar{t}^{\theta} = \begin{pmatrix} t(Y) = R^{\theta} \\ t(X) = X \end{pmatrix} \quad \text{and} \quad \bar{p}^{\theta} = \begin{pmatrix} p(Y) = 1 \\ p(X) = \bar{p}^{\theta} \end{pmatrix}$$

where $\bar{p}^{\theta}$ is the continuation probability for a low-output outcome and R^{θ} is the rental transfer for a high-output outcome for agent θ, with $\bar{p}^h > \bar{p}^l$, $R^h > R^l$. The agent induced effort is $e^l = e^h = 1 < 2$.

The contract parameters $(R^{\theta}, \bar{p}^{\theta})$ depend on whether the incentive compatibility (IC) or truth-telling (TT) constraint binds for type θ (see derivations and proof in Appendix D.1). Intuitively, the owner always retains the driver when high output is reported ($p(Y) = 1$), since firing after good performance is costly. By contrast, when low output is reported, the limited liability constraint prevents extracting further transfers, so the only way to discipline incentives is through inefficient termination with probability $\bar{p}^{\theta} < 1$ (Fuchs, 2007).[40] Limited liability also renders the participation constraint slack, giving the agent informational rents at baseline (see Appendix D.1).

[39] LL implies the agent cannot report more than realized output, so truth-telling for low-output X holds on path.

[40] The principal's optimization implies that the agent's payoff should be minimized during low-output periods. To prevent reneging, I assume—following Mailath and Samuelson (2006), Chapter 7—that both players observe a public randomization device for p at the end of each period. The deviation, in which the principal does not follow through with the randomization device, would unravel the equilibrium by inducing misreporting and zero effort.

The derived transfer schedule is in line with Fact 1– Fact 4, rationalizing the transfer from driver to owner, the possibility to default, and the high turnover in the taxi industry. In addition, heterogeneity in liquidity constraints under limited liability can explain why some drivers receive upfront salaries at baseline, as these payments are best interpreted as working-capital advances.[41]

Overall, the baseline contract is inefficient for two reasons: (a) moral hazard in effort e and (b) in output reporting $\tilde{y}$. Digital payments, by making transactions partially observable, offer signals on both effort and output to the principal, reducing these information frictions and raising total surplus. I now examine this possibility.

5.4 Stage 2: Impact of Digital Observability

This section derives comparative statics on the impact of digital payments for adopters by considering various information benchmarks, relaxing Assumption 1. For simplicity, I assume that digital payments provide no direct benefits to drivers in this section (e.g., lower cash-handling costs); these are additive and simply incorporated in the structural estimation (Section 6). This framework is in partial equilibrium: the share of high-type agents, μ, is held constant before and after the introduction of the technology.[42]

In practice, digital payments provide only *imperfect* information to principals because (i) cash remains widely used, and (ii) timestamps and values of digital transactions only partially reflect effort and output. The signal becomes informative only when agents exert sufficiently high effort and process a large share of transactions digitally. I show that such partial information can alter contracts. Let s denote a high-effort signal observed with probability $\kappa = P(s \mid e = 2)$, and assume $P(s \mid e = 1) = P(s \mid e = 0) = 0$. The technology thus introduces partial observability by revealing a discrete signal of high effort—which, as I will show, can be worse than no observability for low-type agents.

Result 2. (Imperfect Information on Effort) *Under Assumptions 1–4, when (IC) binds, for $K_p \in (\underline{K}_p, \bar{K}_p)$ and $\delta \in (\underline{\delta}, \bar{\delta})$, $\kappa > \bar{\kappa}$ for $\bar{\kappa} < 1$, and $\phi^{\theta}(2) < \tilde{\phi}\ \forall\ \theta$, the principal's best type-dependent stationary contract is:*

[41] Baseline upfront payments ("salaries") need not be monotone in worker type. In this setting, they primarily operate as working-capital advances (e.g., to finance fuel and daily operating expenses needed to start working; Fact 1). When limited liability binds and the participation constraint is slack in the baseline relational contract, outside options do not pin down equilibrium transfers; liquidity needs can therefore make baseline advances more prevalent among poorer drivers (often lower types). Because such state-independent transfers enter continuation values as level shifters, the incentive analysis abstracts from them, and the empirical analysis focuses on the *incremental* upfront compensation induced by observability.

[42] General-equilibrium effects—entry/exit, and market-wide contract redesign—are beyond the scope of the paper given the study horizon and the institutional setting. In the model, equilibrium responses are possible under observability: low-type drivers may be screened out as outside options shift, and different workers could select into driving if digital payments change job amenities. Pinning down these changes would require an equilibrium model with entry and matching and data on contract offers and workforce composition over a longer post-adoption horizon. The experiment is not designed for this: the payment company did not retain observability as the default after the study and owner demand for it at that time was limited, making short-run market-wide transformation unlikely in this setting.

$$\bar{t}^{\theta} = \begin{pmatrix} t(Y) = R^{\theta} - W^{\theta}_{\tilde{e}=2} \\ t(X) = X - W^{\theta}_{\tilde{e}=2} \end{pmatrix} \quad \textit{and}$$

$$p(\tilde{y}, s) = \begin{cases} 1 & \textit{if } \tilde{y} = Y, \\ \bar{p}_{TT} & \textit{if } \tilde{y} = X \textit{ and } s \textit{ is observed} \\ \bar{p}^{\theta\prime} < \bar{p}^{\theta} & \textit{if } \tilde{y} = X \textit{ and } s \textit{ is not observed} \end{cases}$$

The agent θ induced effort is $e^{\theta} = 2$.

In equilibrium, the principal induces high effort by offering an upfront payment $W^{\theta}_{\tilde{e}=2}$ each period when the agent adopts the technology. To do so, continuation probabilities rise in low-output states when the high-effort signal s is observed, but fall otherwise. Because reneging is a concern in relational contracts, the principal compensates effort *ex-ante* with $W^{\theta}_{\tilde{e}=2}$ rather than adjusting transfers *ex-post*. See Appendix D.4 for proof.

Appendix Sections D.3, D.4, and D.5 extend the framework to alternative information benchmarks. Lemmas 2 and 3 show how (imperfect) information on output relaxes the truth-telling constraint, benefiting both principal and agent without directly revealing effort. This captures the logic of the *Coarse Observability* treatment, where low-output signals matter most. Lemma 4 shows that agents have little incentive to manipulate either output or effort signals, since doing so yields no contract changes or increases the risk of termination.

5.5 Stage 1: Differential Technology Adoption

Stage 1 examines the adoption decision by different agent types. In line with the experiment, I assume the agent (driver) makes the one-time decision whether to adopt the technology and retains it upon termination. The framework could, however, be extended to the alternative case where the principal adopts first and screens drivers by willingness to use it, achieving the same separation.

Result 3. (Differential Adoption) *Under Assumptions 1–4, $\exists\ \underline{K}_p < \bar{K}_p, \underline{\delta} < \bar{\delta}^{tech}$ and $\bar{\phi}^h < \bar{\phi}^l$ s.t. if $K_p \in (\underline{K}_p, \bar{K}_p)$ and $\delta \in (\underline{\delta}, \bar{\delta}^{tech})$, $\phi^l(2) > \bar{\phi}^l$, and $\phi^h(2) < \bar{\phi}^h$, then only high-ability agents adopt the technology, while low-ability agents opt not to adopt it.*

Intuitively: (i) high types adopt because the new contract compensates higher effort, lowers termination risk, and leaves them indifferent when $\phi^h(2) < \bar{\phi}^h$ (Result 2); (ii) low types refuse because the disutility of high effort, $\phi^l(2)$, exceeds the surplus it generates, so no transfer can make adoption profitable for such pairs (only high types experience a surplus gain from adoption). If principals could credibly commit not to demand high effort, low types could adopt at an intermediate effort level; absent commitment, however, observability ratchets effort up and deters adoption (proof in Appendix D.6). This result relies on the technology revealing only high-effort signals—if all effort levels were observable, the outcome would differ—that is exactly what the technology is about, as discussed in Section 4.2. With only partial digitization, low or intermediate effort generates too few time-stamped transactions to be informative, while high effort produces frequent, dispersed transactions that are hard to fake and easy to interpret.

Both types co-exist in equilibrium when the share of low-types is large or the discount factor is low. The "no-deviation" condition D19 on $\delta < \bar{\delta}^{tech}$ ensures owners have no incentive to deviate by terminating the low-type agents, incur the replacement cost K_p, and recruit a new agent, with probability μ of being matched with a high-type accepting the technology.

Parameter Restrictions and Generality. Results 1-3 are existence results that characterize an open region of primitives under which digitized records improve incentives among adopters yet can deter adoption for a subset of workers. Selective adoption arises when increased data visibility is sufficiently informative to induce principals to ratchet contracts toward higher effort, while (for some workers) the resulting increase in effort disutility outweighs the additional surplus generated, so no transfer can make adoption privately optimal. The relevant thresholds depend on primitives—including signal precision, direct operational gains from digitization, outside options, replacement costs, and type composition—so alternative environments could generate different equilibria. In Section 6 and Appendix E, I verify that the estimated parameters satisfy the required inequalities (e.g., Table B28) and show—via bootstrap and moment-sensitivity analyses—that the qualitative welfare conclusions are robust in a neighborhood of the estimated environment.

5.6 Comparative Statics: Impact and Adoption of Digital Payments

The framework's predictions align with the experimental outcomes, supporting the modeling assumptions and showing its suitability for the structural estimation that follows.

1. Observability Effects on Contract for Adopters

1a *Effort Effect* $e \uparrow$: Digital payments generate effort signals, enabling the principal to incentivize higher effort, reduce default, and increase average transfers from the agent.

1b *Contract Effect* $W_{\tilde{e}=2}$: The principal compensates the agent for higher effort using an upfront payment $W_{\tilde{e}=2}$.

1c *Retention Effect* $\bar{p}^{\theta} \uparrow$: Imperfect but informative signals on effort/output reduce moral hazard, raising continuation probabilities in low-output periods.

2. Observability Effects on Adoption

2a *Characterization of Low-Types* θ^{l}: Non-adopters can be identified by characteristics linked to high disutility of effort.

2b *Technology Adoption*: Observability and subsequent contractual changes requiring higher effort create a barrier to technology adoption for low-type agents.

Beyond the taxi owner-driver relationship, this framework highlights how information asymmetries shape employment relationships in lower-income contexts, where limited liability and weak contract enforcement, as formalized here, play a central role. Combined with the reduced-form results, it provides the foundation for quantifying the distributional consequences of digital payments and running policy counterfactuals in Section 6.

6 Welfare Impacts of Digital Payments

I combine the theoretical framework with structural estimates to evaluate the welfare impacts of digital payments, focusing on two objectives.

First, I quantify welfare gains from (i) cost savings through reduced cash use and (ii) the "observability effect" from digital transactions. This requires estimating drivers' disutility of effort and computing the relationship values for owners (V) and drivers (U^{θ}).

Second, I simulate counterfactuals to assess alternative policy and design choices: (1) mandating driver adoption, (2) redesigning the technology without observability (now implemented by the payment company), and (3) a full-information benchmark. These exercises quantify the role of information frictions in shaping owner–driver contracts and outcomes in an informal labor market. The analysis aims to provide a framework for managers, policymakers, and innovators to weigh trade-offs in information disclosure and guide technology design.

6.1 Inputs Calibration

I calibrate the model using survey data, reduced-form estimates, and parameters from the literature. Some parameters are taken directly from the data, while others are estimated via GMM using moments from the experiment and surveys. Details are in Appendix E.1 (Table B27).

From the Survey Data. The survey was designed to provide key inputs for estimation and welfare analysis. I use baseline data on hours and earnings to estimate the binary production function. To follow the simple two-output, three-effort framework, I define high and low output, Y and X, as the average earnings above or below the median, calibrated from the groups where the information frictions remain unchanged—*Control* and *No-Observability*. The share of high-type drivers μ and the owner's replacement cost K_p come from owner surveys. I allow the outside options to differ by type, $\bar{u}^h$ and $\bar{u}^l$, calibrated using a representative survey of small vendors I conducted in September 2022 (a common outside option for drivers) and differences in the wealth index. I vary model inputs, compute bootstrapped standard errors, and assess the sensitivity of parameter estimates to each moment, following Andrews et al. (2017) (Appendix E.4).

From the Reduced-Form Estimates. I use reduced-form estimates to calibrate the following parameters: the driver's gains from reducing cash-related costs, G, for drivers with technology access (pooling across observability treatments)[43]; the production function for drivers under *Granular Observability*; and the contract characteristics under *Granular Observability*, notably the observed upfront payment $W_{\tilde{e}=2}$ and retention probability $p_{\tilde{e}=2}$.

From the Literature. I calibrate the discount rate δ using the closest relevant estimate, Yesuf and Bluffstone (2019) in Ethiopia, which reports a weekly discount rate of 0.99.

6.2 Estimation Procedure

In Section 5, I derive the best stationary contracts under different information benchmarks, from no observability to full information. I now use these derivations to estimate key parameters and quantify contract valuations for owners and drivers, thereby assessing the welfare impact of digital payments and running counterfactuals.

Identification. The structural estimation exploits the experimental variation to recover three parameters: the disutility of effort for low- and high-type drivers at baseline, $\phi^l(1)$ and $\phi^h(1)$, and for high-types exerting high effort under *Granular Observability*, $\phi^h(2)$. Identification rests on eight empirical moments: (i) rehiring rates in the *Control*, *No Observability*, and *Granular* arms, together with the adoption experiment that separates high- from low-ability drivers (reluctant drivers); (ii) perceived replacement cost; (iii) drivers' reported contract valuation; (iv) target transfers for both types; and (v) upfront payments under *Granular*. Intuitively, when work becomes more demanding, continuation probabilities must fall to sustain incentives, while upfront payments under observability reveal the compensation required for high effort. The model is over-identified, and Appendix E.2 details the mapping from each empirical moment to its theoretical counterpart.

Parameter Estimation. Parameters are estimated by GMM, minimizing the distance between empirical and structural moments. My data $\mathbf{X}i$ includes the eight empirical targets. The weighting matrix $\mathbf{W}$ is the inverse variance of the moments. Appendix E provides detailed derivations. Standard errors are obtained resampling with 1,000 bootstrap replications of the survey data.

[43] An alternative to calibrate G would be to use drivers' stated WTP (Section 4.1.2). I do not pursue this approach: because the elicitation was unincentivized and the technology was provided for free, stated WTP likely understates true value. Respondents may struggle to quantify lifetime WTP for a good they are not purchasing. With $\delta = 0.99$, the average stated WTP of USD 58 implies weekly utility gains below the treatment effect from the reduced form.

Table 7: Structural Estimation: Matched Moments and Parameter Estimates

Panel A: Reduced Form, Structural, and Matched Moments

Control group outcome	Reduced form	Structural	Difference
Targeted moments:			
Probability $\bar{p}^l$ (Reluctant Drivers)	0.965	0.987	-0.022
	(0.003)	(0.010)	(0.010)
Probability $\bar{p}^h$ (Control and No Observability)	0.968	0.987	-0.019
	(0.005)	(0.009)	(0.010)
Probability $\tilde{p}_{\tilde{e}=2}$ (Granular Observability)	0.966	1.000	-0.034
	(0.011)	(0.000)	(0.011)
Driver's replacement cost K_a	438.78	438.78	0.000
	(59.29)	(0.00)	(59.286)
High-θ Driver's contract value U^h	4111.82	4264.48	-152.651
	(130.12)	(33.43)	(131.023)
Transfer R_l	100.027	100.040	-0.013
	(0.345)	(0.343)	(-0.013)
Transfer R_h	100.027	100.035	-0.009
	(0.345)	(0.343)	(-0.009)
Salary of Adopters $W_{\tilde{e}=2}$	11.03	11.03	0.000
	(0.50)	(0.50)	(0.000)
Untargeted moment:			
Low-θ Driver's contract value U^l	4002.03	3664.59	337.443
	(116.54)	(33.43)	(118.735)
Baseline welfare estimates:			
Owner's contract value V_h	—	6777.24	—
		(297.92)	

Panel B: GMM Parameter Estimates

Input	Value	Interpretation
Low-θ Baseline driver disutility $\hat{\phi}^l(1)$	18.18	Driver disutility in USD
	(5.32)	
High-θ Baseline driver disutility $\hat{\phi}^h(1)$	12.19	Driver disutility in USD
	(5.41)	
High-θ Endline driver disutility $\hat{\phi}^h(2)$	30.74	Driver disutility in USD
	(6.31)	

Panel C: Computed Parameter Estimate

Input	Value	Interpretation
Low-θ Baseline driver disutility $\hat{\phi}^l(2)$	39.91	Driver disutility in USD 39.91
	(8.74)	

Notes: In Panel A, I use GMM with the eight targeted moments. The reduced form consists of observed empirical data, while structural represents the corresponding model predictions. Reduced-form weekly continuation probabilities are computed by deriving observed mid-term probabilities (28 weeks after baseline) while accounting for effort differences q_1 and q_2. It is considered that the agent has a probability $(1-q_1)(1-\bar{p})$ of leaving each period when exerting effort $e=1$. The un-targeted moment uses empirical contract valuation for adopting drivers who initially would have preferred not to have *Granular Observability* at baseline. This approach is used because baseline valuations for low-type drivers were not collected.
In Panel B, I use GMM to estimate driver disutilities ϕ for each type of driver with $e=1$, and $e=2$ for a high-type driver (upon adoption of the technology).
In Panel C, I estimate the counterfactual lower bound for the driver's disutility of effort for $e=2$, $\phi^l(2)$. Since this parameter is empirically unobserved, as low-type drivers did not adopt the technology, I obtain a lower bound using the following model intuition: a low-type θ driver would need a high enough salary $W^l_{\tilde{e}=2} > W$, (for a given $\tilde{p}$) to exert high effort $e=2$. I compute the minimum value for $\phi^l(2)$ such that at $W^l_{\tilde{e}=2}$, the owner would be better off in the status quo.
Standard errors for each parameter are shown in brackets, estimated using a bootstrap procedure with 1000 replications based on the empirical distributions of the framework inputs.

6.3 Parameter Estimates and Owner-Driver Contract Valuations

Table 7 summarizes the estimation results and matched moments. Panel A shows that the model fits the data well: baseline continuation probabilities for low- and high-type drivers are matched within 1–2 pp, while the driver's replacement cost K_a, the target transfers R^h and R^l, and the upfront payment under observability $W_{\tilde{e}=2}$ match almost exactly. The estimated driver's replacement cost is $439, which corresponds to about 33 days of lost profit. Although I did not collect contract valuations for low-type drivers, I compare the structural moment with valuations from drivers willing to adopt (in the impact experiment), but who would have preferred not to have *Granular Observability*, and this comparison yields a close match.

Panel B reports GMM estimates of the disutility of work: $18 for low-type drivers and a lower $12 for high-types. For high-types at $e = 2$, the disutility rises to $\phi^h(2) =$ $31. Panel C computes a counterfactual lower bound for $\phi^l(2)$—the disutility level that would make upfront compensation unprofitable for owners—estimated at $40, above $\phi^h(2)$, as expected.

I use these estimates to calculate contract valuations and total welfare, assuming that the social planner maximizes a social welfare function simply equals the sum of the owner's and the driver's welfare (equal weight). At baseline, without the technology, the owner's present-discounted contract value is about $6,777,[44] and the driver's is $3,915, or about 37% of total welfare. Specifically, the high-type driver's contract value is $4,264, while the low-type's is lower ($3,665) reflecting higher disutility of effort. These values broadly align with the emerging literature estimating the value of the contractual relationship in lower-income contexts: Kelley et al. (2024) estimates the value of the relationship to be between $1,794 and $2,753 on average for a minibus owner in Kenya, and $507 for drivers. In the rose market in Kenya, Macchiavello and Morjaria (2015) finds higher valuations ($13,872 and $22,127 for sellers and buyers).[45]

6.4 Welfare and Distributional Consequences Under Counterfactuals

6.4.1 Without Policy Intervention

I first analyze the status quo without policy intervention, where only high-type drivers adopt the technology. Figure A5(a) plots contract valuations for owners matched with high- versus low-type drivers, with and without the technology. Low-types do not adopt, so their welfare remains unchanged. This analysis restricts attention to production-side welfare, weighting owners and drivers equally, and omits consumer surplus (as a result, the estimated welfare gains may be understated, though this remains speculative).

[44] Owner's valuations with high- and low-type drivers are nearly identical here, though this need not be the case. This similarity arises because the outside option for low-type drivers is lower such that the owner optimally offers a comparable baseline contract to both types.

[45] Kelley et al. (2024) assume a lower daily discount factor of 0.99 (implying a dollar today is worth only 2 cents in a year), whereas I use a weekly rate of 0.99, relying on Yesuf and Bluffstone (2019) in Ethiopia. Adjusting for this difference, the relationship values are comparable across studies. In Macchiavello and Morjaria (2015), the relationship value is given by the maximum temptation to deviate between the Kenyan rose seller and the Dutch buyers, which I computed from its Table 1 (using replication data).

In this structural estimation, I assume that the principal does not capture the direct reduced cash-related costs from accessing digital payments. This aligns with empirical evidence: as shown in Section 4.2, while the technology without observability benefited drivers, these gains did not alter the contract between owners and drivers (comparing *No Observability* to control). Thus, the technology increases high-type drivers' welfare through reduced cash costs, as quantified in Section 4.1. In addition, it generates efficiency gains via reduced moral hazard captured by owners, given the structure of the model. Overall, 73% of the technological gains flow to high-type drivers.

Without social planner intervention, the introduction of the technology exacerbates welfare inequality between high- and low-type businesses, as only high-type pairs benefit from its adoption. Figure 5, Col (2), shows aggregate welfare for owners matched with high- and low-type drivers. The top of each bar represents the overall welfare increase, accounting for the distribution of high- and low-type drivers as reported by the owners. In the model, total welfare rises by 0.5%, but the gains are concentrated among high-type pairs.

Figure 5: Total Welfare $U + V$ in the Economy Under Different Counterfactuals

Notes: This figure plots total contract value (welfare) for owners and drivers under different counterfactuals, combining the owner with the high-type θ driver (blue) and the low-type θ driver (red). Bars represent: (i) information frictions without technology; (ii) no policy intervention, where only high-type drivers adopt; (iii) a mandate requiring all drivers to adopt, giving owners access to the high-effort signal; (iv) a redesigned technology without observability, allowing universal adoption; and (v) a full-information benchmark (wage employment), where effort and output are observable. The top of each bar shows aggregate welfare gains, weighted by the share of high-type μ and low-type pairs $1 - \mu$.

6.4.2 Counterfactual 1: Mandating Digital Payment Adoption

Consider a mandate requiring all drivers to adopt digital payments. With some governments contemplating the enforcement of the adoption of digital technologies (e.g., Ghana, Nigeria), this counterfactual examines the potential impact of such policies. Another way to view this counterfactual is as an "owner-mandate" environment: the platform retains observability and the owner has the ability to make adoption a condition of employment, and—given the calibrated outside options—both driver types remain in the market.

I find that this mandate has significant welfare and distributional consequences (Figure A5(b)). For high-type pairs, the mandate redistributes gains from drivers to owners. By requiring drivers to adopt, the mandate eliminates the need for owners to offer compensation for adoption, leaving the high-types only with the lower cash-related costs.[46]

The mandate substantially reduces the welfare of low-type drivers by 8%. Adoption forces higher effort through the observability signal, while the cash savings are too small to offset the extra disutility. In contrast, the owner's welfare in low-type matches increases by 4%, benefiting from reduced moral hazard in effort.

At the aggregate level, the mandate further exacerbates welfare inequality between and within high- and low-type businesses (see Figure 5, Col (3) and Figure A6 for a further decomposition of owner/driver welfare). The overall welfare gain is lower than under the no-intervention scenario (0.4% compared to 0.5%). This suggests that, although the mandate increases adoption, it may have adverse welfare and distributional consequences, by pushing low-ability drivers to exert inefficiently high effort, benefiting owners but not drivers.

6.4.3 Counterfactual 2: Redesigning the Technology to Remove Observability

This counterfactual explores the impact of redesigning the technology to remove its observability feature. This scenario is relevant for at least two reasons. First, this is what the company later did by default in the taxi industry, based on the study's findings.[47] Second, many technologies, beyond payment systems, default to observability but can often be redesigned to exclude it.

I find that removing observability fundamentally changes the welfare effects of the technology (Figure A5(c)). Both low- and high-type drivers adopt the technology and benefit from reduced cash-related costs, thereby shifting the production possibility frontier upward. All welfare gains now accrue to drivers, who retain the informational rent, while owners see no direct benefit. For example, the share of total welfare within pairs captured by low-type drivers rises from 35% to 36%, thereby slightly reducing within-firm inequality. However, this design choice introduces an efficiency trade-off for high-type businesses: welfare gains for high-type pairs are smaller, with a lower possibility frontier compared to the status quo with observability (Section 6.4.1).[48]

At the aggregate level, this policy sharply reduces welfare inequality between high- and low-type businesses, as all drivers now adopt the technology (Figure 5, Col (4)). The company chose to implement a version of this counterfactual to increase driver access, prioritizing adoption over

[46]If I relax the assumption that high-type drivers retain the technological gains, the result—redistribution of gains from drivers to owners—is even stronger.

[47]I assume that removing observability leads all drivers to adopt the technology. Empirically, 88% of previously reluctant drivers adopted the technology after the company made non-observability the default post-experiment.

[48]I focus on the non-observability-by-default redesign because it is the design the company actually implemented in this setting. An alternative is one where drivers decide whether to adopt and, conditional on adopting, can directly opt in or out of observability. In the model, this option may increase aggregate welfare: high-type drivers would adopt with observability (unlocking efficiency gains), while low-type drivers would adopt without observability (preserving informational rents while still benefiting from reduced cash frictions).

efficiency coming from reduced moral hazard. The aggregate welfare gain is higher than under the no-intervention scenario (0.7% compared to 0.5%), providing a clear rationale for implementation.

6.4.4 Counterfactual 3: Full-Information Benchmark

I examine the welfare implications of the full-information benchmark. This scenario sheds light on the extent to which moral hazard impacts welfare in this economy and the distributional effects of fully removing it. Specifically, I consider a hypothetical scenario where the technology provides the same cost-saving benefits (i.e., reduced cash-related costs) and is universally adopted, but now fully reveals the agent's effort at each level, not just high effort as in the previous cases. This counterfactual can also be interpreted as a mandate requiring digital technology that provides employers full information on effort. Under full information, the best stationary equilibrium is wage employment, as outlined in Lemma 1, where owners compensate drivers just enough to cover their disutility of effort and outside option.

Moving to full information raises owner welfare by 20%, and yields the highest efficiency gains of any counterfactual. But drivers experience a significant welfare reduction, as they lose their informational rents. Figure A5(d) illustrates this trade-off: welfare decreases by 17% for high-type drivers and 20% for low-type drivers, while overall welfare rises by 2.3% (Figure 5, Col (5)). Thus, while the production frontier expands, the outcome is not Pareto-improving. This raises a policy concern: as technologies approach full observability, employers may be unable to raise wages enough to offset drivers' lost rents (e.g., due to factors outside the model, such as credit constraints), dampening adoption of otherwise welfare-enhancing technologies.[49]

6.5 Discussion: Trade-off Between Observability and Adoption

The structural estimation, together with the experimental results, has important policy implications. While both principals and agents can benefit from the technology on some dimensions—through reduced cash-related costs and lower information frictions—low-type agents are not adopting it. As a result, observability increases inequality between high- and low-type workers and lowers aggregate welfare relative to a design without observability. To reduce this gap and increase both technology access and overall welfare, a social planner or technology designer with reasonable welfare weights on low-type businesses may thus prefer to limit the observability embedded in digital technologies. This would prevent information from being used against low-type agents, encouraging adoption and increasing overall welfare. This could be achieved either through a total removal of observability (as in Counterfactual 2) or possibly by giving drivers the option to select transaction observability.[50]

[49]This analysis assumes risk neutrality, as the theory literature has not extensively explored risk aversion within relational contracts. Conclusions might differ quantitatively since risk-averse agents may particularly value salaried employment due to reduced income volatility. I leave this theoretical and empirical consideration for future research.

[50]The latter might work if types were perfectly known at baseline, but it could also introduce adverse selection. Because this channel is not directly modeled and lies beyond the scope of this paper, I do not explore it further.

Reflecting these insights, the partner payment company shifted to *No Observability* as the default for taxi drivers (akin to Counterfactual 2), enabling widespread adoption in the taxi industry, precisely due to the considerations highlighted in this structural estimation. By early 2024, over 16,000 drivers—roughly 75% of Dakar's taxi industry—were using the technology, and the company has since expanded it to Côte d'Ivoire. At the same time, observability was retained in sectors with formal contracts (e.g., supermarkets), where adoption frictions are minimal and managers now use digital transaction observability to monitor cashiers.

7 Conclusion

This study investigates the relationship between technology adoption and within-firm contracts in a lower-income setting. The global proliferation of digital technologies has drawn considerable interest from policymakers and the private sector due to their potential to enhance firm productivity and firm growth. Academic research examining their influence on private sector development and intra-firm organization is key to informing this discussion.

I conducted two randomized experiments in Senegal's taxi industry, over nearly two years in partnership with the country's largest payment company, to measure the effects of digital technologies like payments on businesses. Relying on contract theory, the experiments isolate how unintended observability, embedded in digital payments, reduces information frictions and affects worker behavior, within-firm contracts, and adoption.

The study has four key findings. First, digital payment technologies benefit businesses by significantly reducing the costs associated with using cash, enhancing security, and improving earnings tracking. Second, business owners leverage digital payments as a monitoring tool, enabling contract changes and increased employee effort. As a result, these owner-driver relationships last significantly longer, and owners' trust in their drivers increases. Third, this same observability acts as a barrier to adoption for low-ability drivers, with differential adoption among workers. Fourth, the technology increases overall welfare by providing employers with better information on employee actions, although it exacerbates welfare inequality between adopters and non-adopters.

Taken together, these findings show that while digital technologies can expand the production possibility frontier and increase total surplus, they may not be adopted in the first place due to the observability that can be generated at low cost, suggesting the need for policy interventions. These insights may extend to informal sectors often characterized by weak contract enforcement and limited liability, where digital tools are often not adopted, or to contexts where they raise data privacy concerns.

Three features of this paper—the dual randomization of technology access and observability features, the comprehensive two-year panel data on employers and employees in informal firms, and the analysis of the interplay between technology adoption and within-firm interactions grounded in economic theory—aim to contribute to the literature that studies organizations in developing contexts. This study suggests the importance of further investigating how technology design shapes organizational structures within and across firms at various stages of economic development. Understanding these dynamics is an exciting avenue for future research.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge, "When Should You Adjust Standard Errors for Clustering?*," *The Quarterly Journal of Economics*, 10 2022, *138* (1), 1–35.

Agarwal, Sumit, Wenlan Qian, Bernard Y. Yeung, and Xin Zou, "Mobile Wallet and Entrepreneurial Growth," *AEA Papers and Proceedings*, May 2019, *109*, 48–53.

Aker, Jenny, McClelland Amanda Boumnijel Rachid, and Niall Tierney, "Payment Mechanisms and Anti-Poverty Programs: Evidence from a Mobile Money Cash Transfer Experiment in Niger," *Economic Development and Cultural Change*, 2016, *65* (1), 1–37.

Alvarez, Fernando and David Argente, "On the Effects of the Availability of Means of Payments: The Case of Uber.," *The Quarterly Journal of Economics*, 2022, *137* (3), 1737–1789.

Andrews, Isaiah and Daniel Barron, "The Allocation of Future Business: Dynamic Relational Contracts with Multiple Agents," *American Economic Review*, September 2016, *106* (9), 2742–59.

__, Matthew Gentzkow, and Jesse M. Shapiro, "Measuring the Sensitivity of Parameter Estimates to Estimation Moments*," *The Quarterly Journal of Economics*, 06 2017, *132* (4), 1553–1592.

Angrist, Joshua D., Sydnee Caldwell, and Jonathan V. Hall, "Uber versus Taxi: A Driver's Eye View," *American Economic Journal: Applied Economics*, July 2021, *13* (3), 272–308.

Atkin, David, Azam Chaudhry, Shamyla Chaudry, Amit Khandelwal, and Eric Verhoogen, "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan," *The Quarterly Journal of Economics*, 2017, *132* (3).

Baker, George and Thomas Hubbard, "Make Versus Buy in Trucking: Asset Ownership, Job Design, and Information," *American Economic Review*, June 2003, *93* (3), 551–572.

__ and __ , "Contractibility and Asset Ownership: On-Board Computers and Governance in U.S. Trucking," *The Quarterly Journal of Economics.*, 2004, *119* (4).

__, Robert Gibbons, and Kevin J. Murphy, "Relational Contracts and the Theory of the Firm," *The Quarterly Journal of Economics*, 2002, *117* (1), 39–84.

Banerjee, Abhijit V., Paul J. Gertler, and Maitreesh Ghatak, "Empowerment and Efficiency: Tenancy Reform in West Bengal," *Journal of Political Economy*, 2002, *110* (2), 239–280.

Beaman, Lori, Jeremy Magruder, and Jonathan Robinson, "Minding Small Change among Small Firms in Kenya.," *Journal of Development Economics*, 2014, *108*, 69–86.

Becker, GM, MH DeGroot, and J Marschak, "Measuring utility by a single-response sequential method," *Behav Sci*, 1964, *9* (3), 226–32.

Bertrand, Marianne and Antoinette Schoar, "The Role of Family in Family Firms.," *Journal of Economic Perspectives*, 2006, *20* (2), 73–96.

Blader, Steven, Claudine Gartenberg, and Andrea Prat, "The Contingent Effect of Management Practices," *The Review of Economic Studies*, 07 2019, *87* (2), 721–749.

Bossuroy, Thomas, Clara Delavallade, and Vincent Pons, "Biometric Monitoring, Service Delivery and Misreporting: Evidence from Healthcare in India," *Review of Economics and Statistics*, 2025.

Brockmeyer, Anne and Magaly Sáenz Somarriba, "Electronic Payment Technology and Tax Compliance: Evidence from Uruguay's Financial Inclusion Reform," *American Economic Journal: Economic Policy*, 2024, (9947).

Chandrasekhar, Arun G, Melanie Morten, and Alessandra Peter, "Network-Based Hiring: Local Benefits; Global Costs," February 2020.

Crouzet, Nicolas, Apoorv Gupta, and Filippo Mezzanotti, "Shocks and Technology Adoption: Evidence from Electronic Payment Systems," *Journal of Political Economy*, 2023.

Dalton, Patricio, Haki Pamuk, Ravindra Ramrattan, Burak Uras, and Daan van Soest, "Electronic Payment Technology and Business Finance: A Randomized, Controlled Trial with Mobile Money," *Management Science*, 2023.

de Janvry, Alain and Elisabeth Sadoulet, "Optimal share contracts with moral hazard on effort and in output reporting: managing the double Laffer curve effect," *Oxford Economic Papers*, 04 2007, *59* (2), 253–274.

de Rochambeau, "Monitoring and Intrinsic Motivation: Evidence from Liberia's Trucking Firms," *Working Paper*, 2021.

DellaVigna, Stefano and Matthew Gentzkow, "Uniform Pricing in U.S. Retail Chains," *The Quarterly Journal of Economics*, 2019, *134* (4), 2011–2084.

Derksen, Laura, Anita McGahan, and Leandro Pongeluppe, "Privacy at What Cost? Saving the Lives of HIV Patients With Electronic Medical Records," *Working Paper*, 2024.

Dizon-Ross, Rebecca and Seema Jayachandran, "Improving Willingness-to-Pay Elicitation by Including a Benchmark Good.," *AEA Papers and Proceedings*, 2022, *112* (551-55).

Fajnzylber, Pablo, William Maloney, and Gabriel Montes Rojas, "Microenterprise Dynamics in Developing Countries: How Similar are They to Those in the Industrialized World? Evidence from Mexico," *World Bank Economic Review*, 2006, *20* (3), 389–419.

Fong, Yuk-Fai and Jin Li, "Relational Contracts, Limited Liability, and Employment Dynamics," *Journal of Economic Theory*, 2017, *169*, 270–293.

Fuchs, William, "Contracting with Repeated Moral Hazard and Private Evaluations," *American Economic Review*, September 2007, *97* (4), 1432–1448.

Gertler, Paul, Sean Higgins, Ulrike Malmendier, and Waldo Ojeda, "Do Behavioral Frictions Prevent Firms from Adopting Profitable Opportunities?," January 2025, (33387).

Ghatak, Maitreesh and Priyanka Pandey, "Contract choice in agriculture with joint moral hazard in effort and risk.," *Journal of Development Economics*, 2000, *63* (2), 303–326.

Goldfarb, Avi and Catherine Tucker, "Privacy and Innovation," *Innovation Policy and the Economy*, 2012, *12* (12).

Goldsmith-Pinkham, Paul, Peter Hull, and Michal Kolesár, "Contamination Bias in Linear Regressions," *NBER Working Paper*, 2024, (30108).

Higgins, Sean, "Financial Technology Adoption: Network Externalities of Cashless Payments in Mexico," *American Economic Review*, November 2024, *114* (11), 3469–3512.

Holmström, Bengt, "Moral Hazard and Observability," *The Bell Journal of Economics*, 1979, *10* (1), 74–91.

Holt, Charles A and Susan K Laury, "Risk aversion and incentive effects," *American Economic Review*, 2002, *92* (5), 1644–1655.

Houeix, Deivy, "Nationwide Diffusion of Technology Within Firms' Social Networks.," *Working Paper*, 2025.

Hubbard, Thomas, "The Demand for Monitoring Technologies: The Case of Trucking*," *The Quarterly Journal of Economics*, 05 2000, *115* (2), 533–560.

Innes, Robert, "Limited liability and incentive contracting with ex-ante action choices," *Journal of Economic Theory*, 1990.

Jack, William and Tavneet Suri, "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution," *American Economic Review*, January 2014, *104* (1), 183–223.

__ **and** __ , "The Long-run Poverty and Gender Impacts of Mobile Money," *Science*, 2016, *354* (6317), 1288–1292.

Kala, Namrata and Elizabeth Lyons, "The Effects of Digital Surveillance and Managerial Clarity on Performance," *Working Paper*, 2025.

Karlan, Dean and Jonathan Zinman, "Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment," *Econometrica*, 2009, *77* (6), 1993–2008.

Kelley, Erin, Gregory Lane, and David Schönholzer, "Monitoring in Small Firms: Experimental Evidence from Kenyan Public Transit," *American Economic Review*, 2024.

Lacker, Jeffrey M. and John A. Weinberg, "Optimal Contracts under Costly State Falsification," *Journal of Political Economy*, 1989, *97* (6), 1345–1363.

Lazear, Edward P., "Performance Pay and Productivity," *American Economic Review*, December 2000, *90* (5), 1346–1361.

Levin, Jonathan., "Relational Incentive Contracts," *American Economic Review*, 2003, *93* (3), 835–857.

Macchiavello, Rocco and Ameet Morjaria, "The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports.," *American Economic Review*, 2015, *105* (9).

__ **and** __ , "Competition and Relational Contracts in the Rwanda Coffee Chain.," *The Quarterly Journal of Economics.*, 2021, *136* (2).

Mailath, George J. and Larry Samuelson, *Repeated Games and Reputations: Long-Run Relationships*, Oxford University Press, 09 2006.

McKenzie, David and Anna Luisa Paffhausen, "Small Firm Death in Developing Countries," *Review of Economics and Statistics*, 2019, *101* (4), 645–57.

Mel, Suresh De, David McKenzie, and Christopher Woodruff, "Measuring Microenterprise Profits: Must We Ask How the Sausage Is Made?," *Journal of Development Economics*, 2009, *88:19-31.*

Nagler, Paula and Wim Naudé, "Non-farm entrepreneurship in rural sub-Saharan Africa: New empirical evidences," *Food Policy*, 2017, *67*, 175–191.

Riley, Emma, "Resisting Social Pressure in the Household Using Mobile Money: Experimental Evidence on Microenterprise Investment in Uganda," *American Economic Review*, May 2024, *114* (5), 1415–47.

Shavell, Steven, "Risk Sharing and Incentives in the Principal and Agent Relationship.," *The Bell Journal of Economics*, 1979, *10* (1), 55–73.

Shetty, Sudhir, "Limited Liability, wealth differences and tenancy contracts in agragian economies," *Journal of Development Economics*, 1988, *29*, 1–22.

Townsend, Robert M, "Optimal contracts and competitive markets with costly state verification," *Journal of Economic Theory*, 1979, *21*, 265 – 293.

World Bank, *Digital Opportunities in African Businesses*, Washington, DC: World Bank, IFC, 2024.

Yesuf, Mahmud and Randall Bluffstone, "Consumption Discount Rates, Risk Aversion and Wealth in Low-Income Countries: Evidence from a Field Experiment in Rural Ethiopia," *Journal of African Economies*, 06 2019, *28* (1), 18–38.

Supplemental Appendix

Table of Contents

A Additional Figures

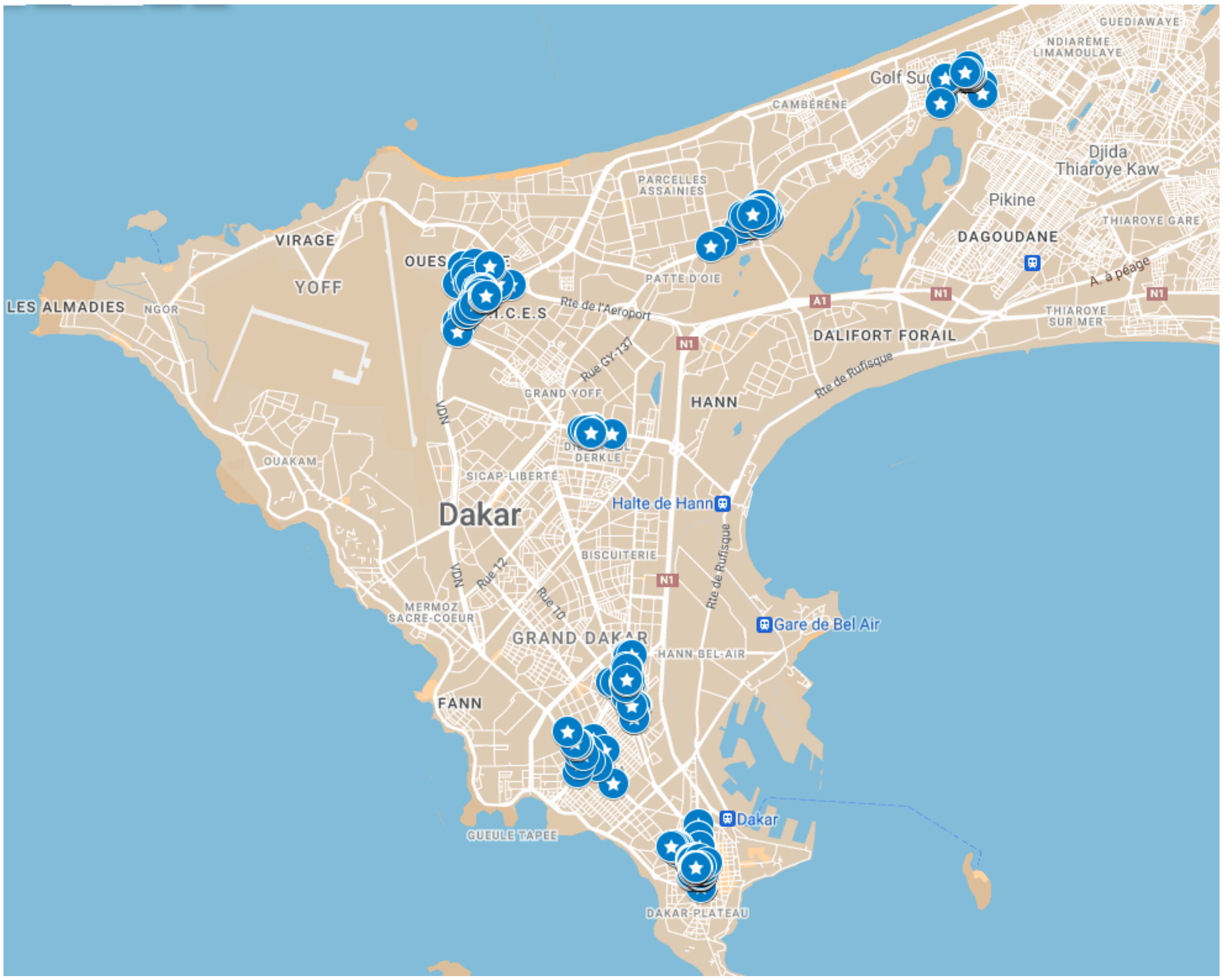


Figure A1: Locations of the Mystery Passenger Audits - August 2022

Notes: The figure displays a map of Dakar, the capital city of Senegal. Each blue dot is the GPS location of the mystery passenger audit activity for a random subset of data points. The goal was to measure (i) drivers' behavior related to digital payments and pricing, and (ii) drivers' effort based on their presence on the road. In particular, in August 2022, I trained twenty mystery passengers to hail taxis throughout Dakar, following a strict procedure to mimic typical price bargaining. Over two weeks, they systematically rotated across seven high-traffic locations each day, capturing a broad sample of taxis and driver behavior over a meaningful timeframe. Surveyors asked questions and secretly recorded taxi license numbers. They pretended they had to leave after a pre-set bargaining process—primarily to increase the sample size and reduce field costs. The activity was repeated a sufficient number of times to match taxi drivers with their license numbers in the experimental sample. Specifically, mystery passengers adhered to the following steps: (1) Memorize the randomized destination and pre-specified price on their data collection application, (2) Stop a taxi, (3) Ask the driver's initial price, (4) Suggest the pre-specified low price, (5) Listen to the driver's counteroffer and ask their last price, (6) Suggest a non-rounded price, (7) Ask to use digital payments. Detailed data were recorded on a tablet once the taxi left about each step of the process.

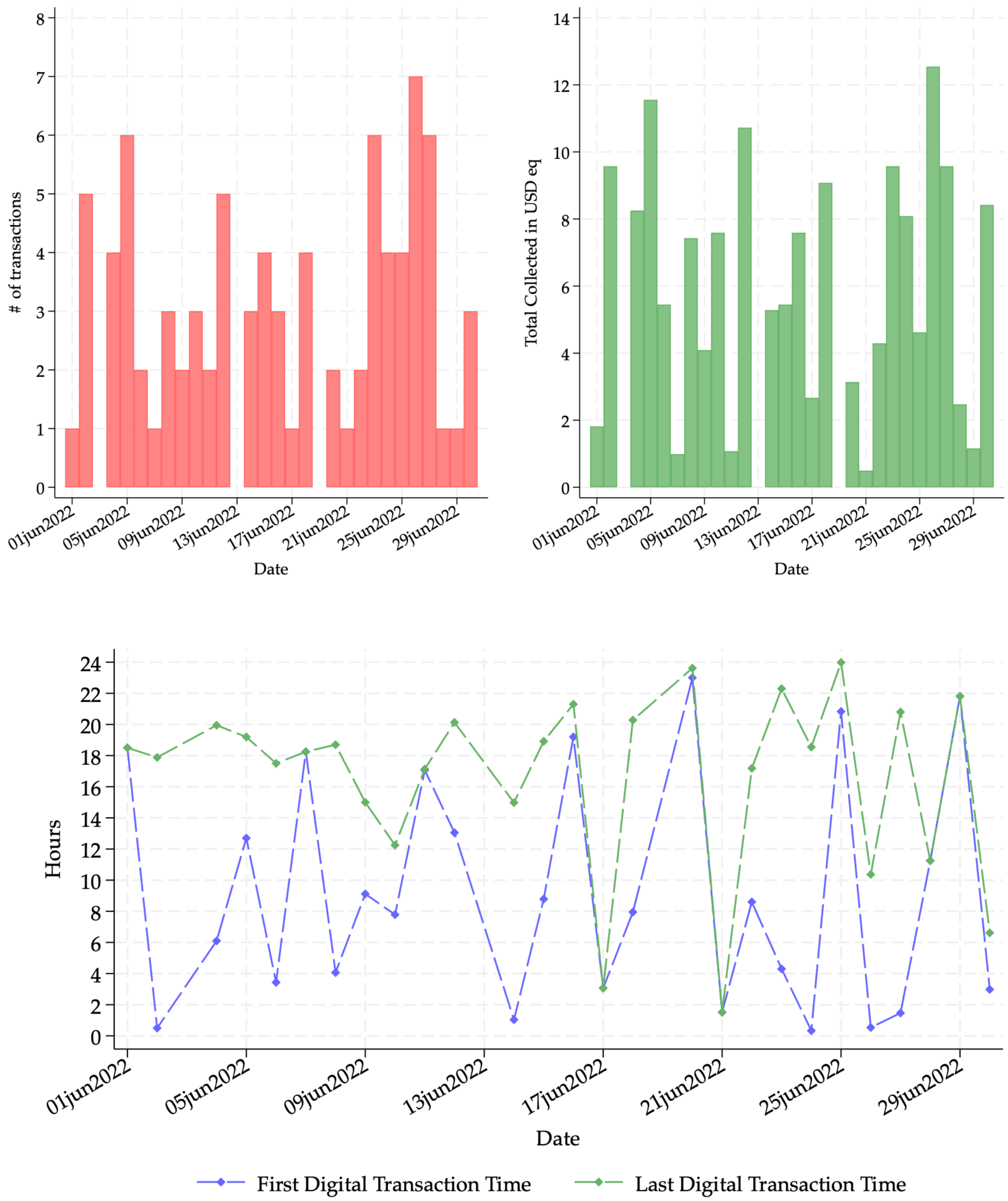


Figure A2: Illustrations of Digital Observability under *Granular Observability*

Notes: These panels show the information available to taxi owners under *Granular Observability*. Panel (a) illustrates daily transaction counts and total amounts collected. Panel (b) illustrates start and end working hours inferred from transaction timestamps, providing a signal of driver effort.

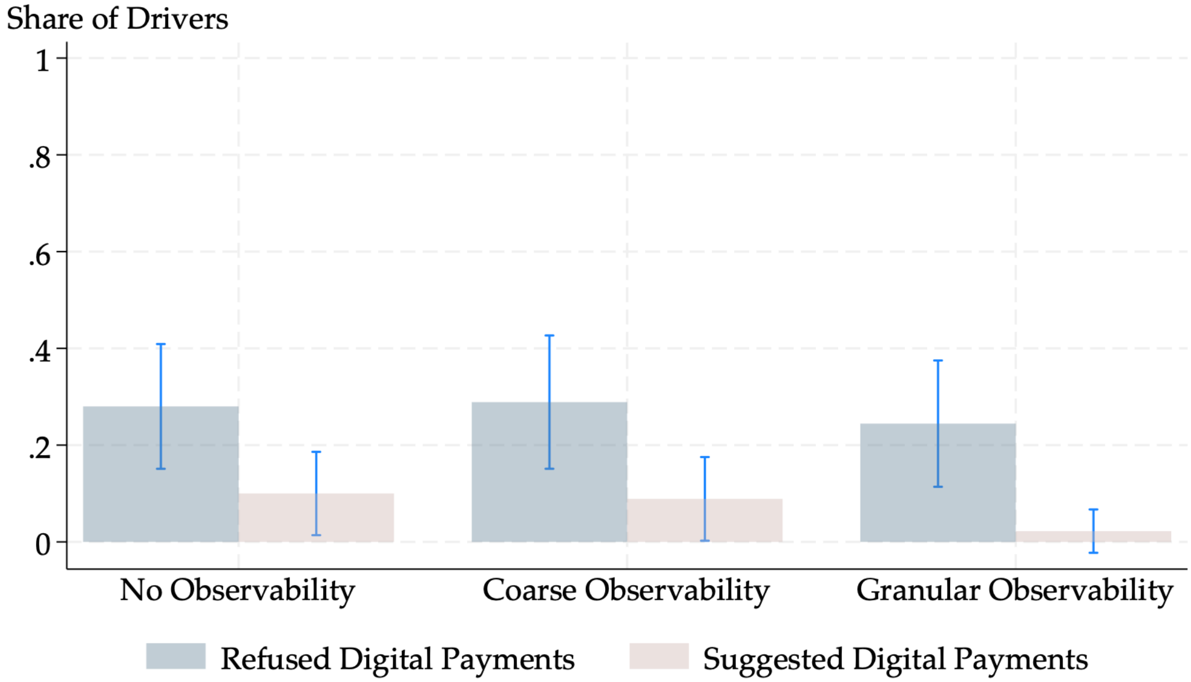


Figure A3: No Manipulation of the Effort and Output Signals - Mystery Passenger Audits

Notes: This figure shows no evidence of manipulation of the output of effort signals by the taxi drivers during the mystery passenger audits. It shows the share of drivers who refused digital payments or suggested paying digitally to the mystery passengers during the audit activity. In general, the agent can mostly manipulate the share of digital transactions downward (i.e., processing more cash than digital payments) and not the reverse in this setting. In most cases, passengers have the choice of whether to pay digitally or in cash. There were no cases in which drivers *demanded* digital payments from customers: this is understandable in an economy where cash remains the dominant form of transaction and it is practically difficult to compel customers to pay digitally. This figure shows no differential propensity to engage in manipulating the effort or output signal in the data from mystery passenger audits, neither upward nor downward. The absence of manipulation may be explained by the competitive pressure drivers face to secure passengers, which discourages them from manipulating digital payment usage. Drivers may perceive the short-term loss of passengers—resulting from pressuring or discouraging them to pay digitally to signal something to their owners—as too costly to justify such actions.

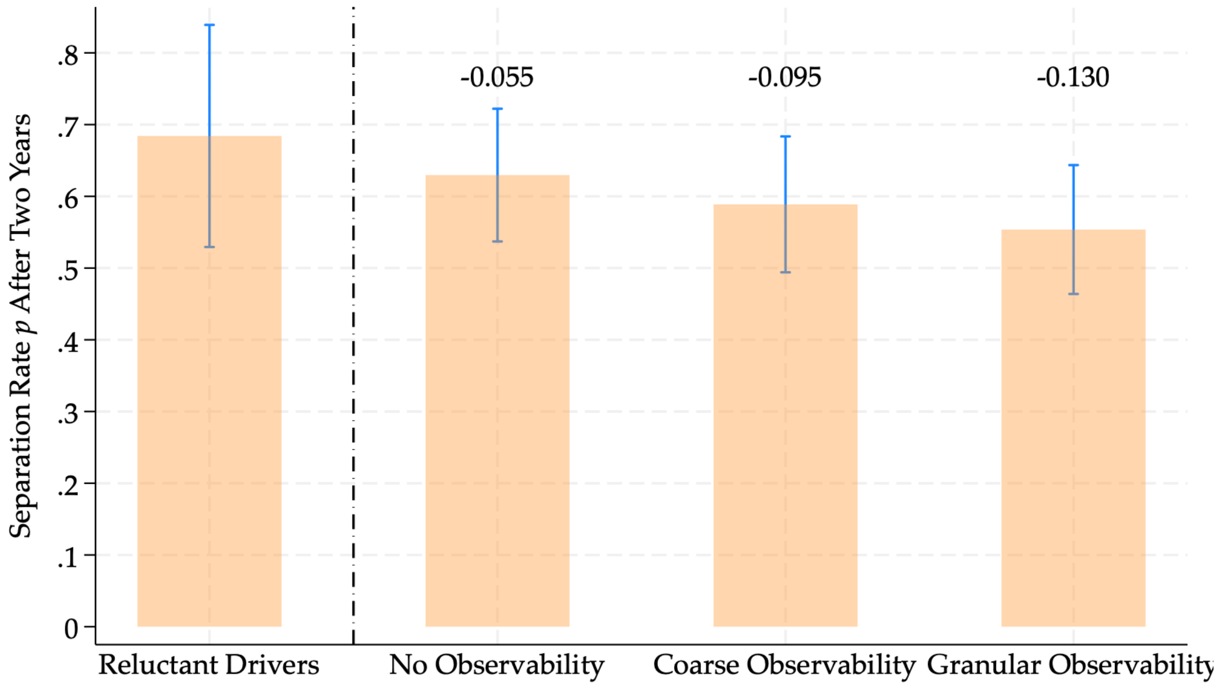


Figure A4: Separation Rates After Nearly Two Years Across Experimental Groups

Notes: This figure shows the separation rates or shares of employment turnover across the different experimental groups. 95% confidence intervals are displayed. From the left bar in order, the first group defines the "low-types" drivers who refused to give their employer's contact in the first place due to concerns regarding transaction observability but that later adopted when the technology was re-offered without owner observability (removing the possible treatment effect of getting the technology). The intuition is as follows: the low-type drivers are the ones that refused to give the owner's contact due to fears of observability (self-reported by drivers), while the ones who refused for other reasons (self-reported by drivers) are less likely to be low-types. Drivers who refused due to observability concerns tend to have shorter relationships. In addition, owner-driver pairs randomized into *Granular Observability* tend to stay together significantly longer. The model is defined as $y_j = \beta_0 + \beta_1 T_j^{GranularObs} + \beta_2 T_j^{CoarseObs} + \beta_3 T_j^{NoObs} + \alpha_s + \epsilon_j$, where y_j indicates the separation rate after almost two years, and the omitted category being the reluctant drivers ("low-types" drivers described above). The bars display raw separation rates, while the numbers above the treatment bars report the estimated coefficients $\beta_1, \beta_2, \beta_3$, i.e., differences relative to the reluctant-driver baseline, as described.

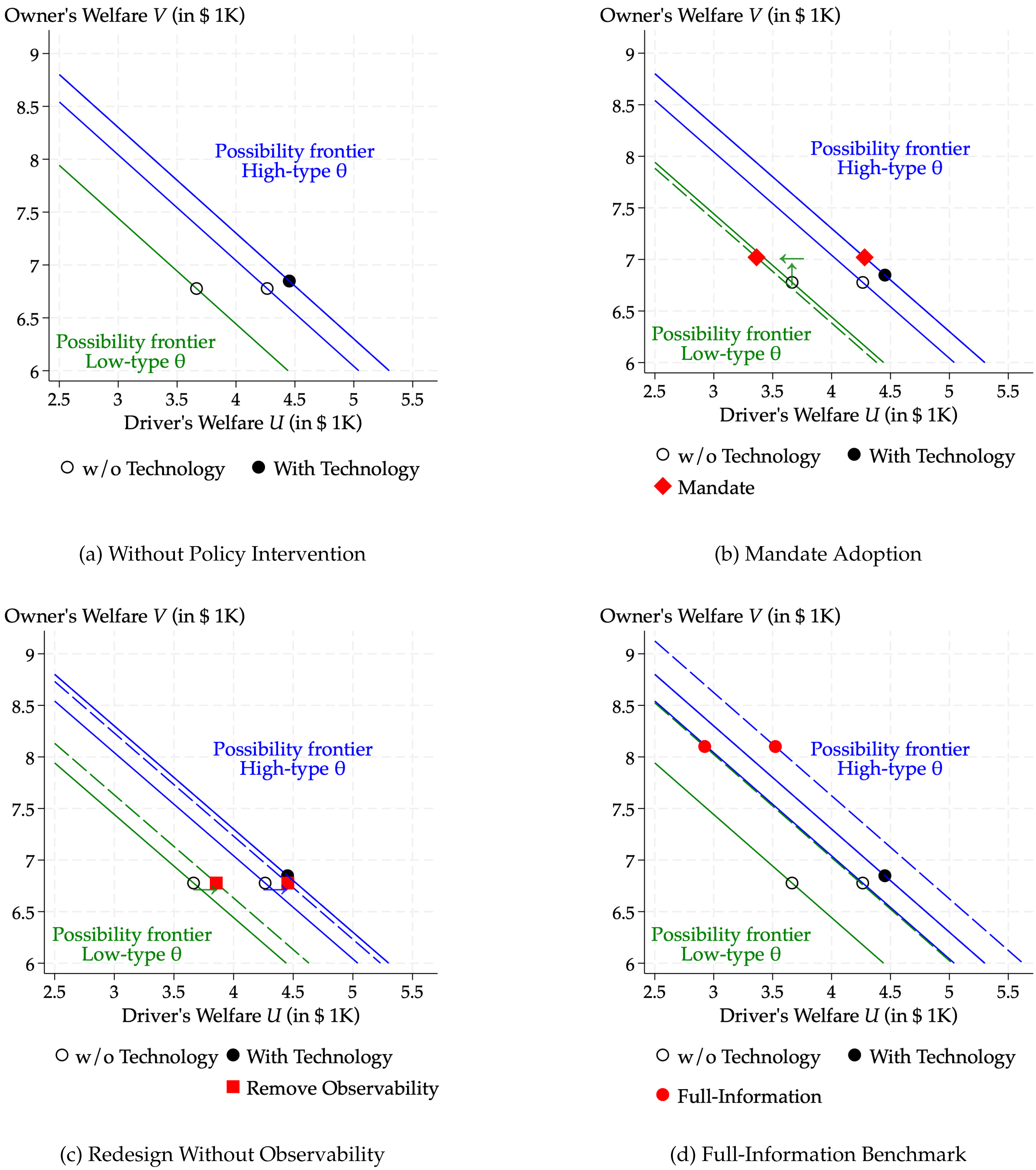


Figure A5: Owner's and Driver's Welfare Under Different Counterfactuals

Notes: These figures present counterfactual analyses by plotting the contract valuations (welfare) for both the owner and the driver on the same graph. The solid lines depict the utility possibility frontiers for low- and high-type drivers, before and after the introduction of the technology. Dashed lines represent the utility frontiers under various counterfactual scenarios. The graph is re-scaled to zoom into the area of interest. In constructing these graphs, I assume a social planner who maximizes total welfare, defined as the sum of the owner's and driver's welfare (with equal weight).
Panel (a) contrasts the baseline contract (without digital payments) with the *Granular Observability* group (with digital payments). In Panel (b), I analyze the effect of mandating digital payment adoption, which requires both low- and high-type drivers to adopt the technology and exert high-effort. Panel (c) examines a counterfactual where the technology is redesigned to remove the observability feature, allowing both types of drivers to adopt. Panel (d) explores a full-information benchmark, where the technology remains unchanged, is universally adopted, but now fully reveals the driver's effort level (and not only a high-effort signal) and output.

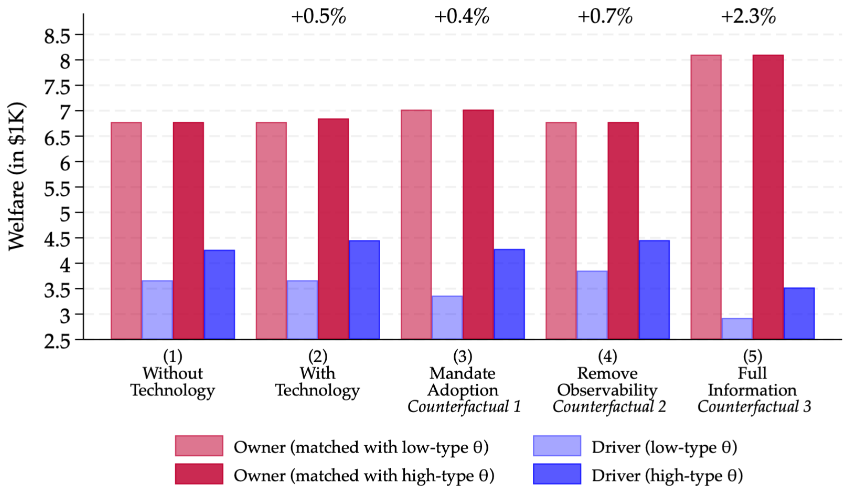


Figure A6: Welfare Decomposition by Owner/Driver and Driver Type Across Counterfactuals

Notes: This figure decomposes welfare (contract value) into four components for each counterfactual: owner welfare and driver welfare, separately for low-type ($\theta = L$) and high-type ($\theta = H$) matches. In each group of bars, the first two bars correspond to the low-type match (owner, then driver), and the next two bars correspond to the high-type match (owner, then driver). Counterfactuals are: (i) information frictions without technology; (ii) baseline with technology, where only high-type drivers adopt; (iii) a mandate requiring all drivers to adopt; (iv) a redesigned technology without observability allowing universal adoption; and (v) a full-information benchmark (wage employment) in which effort and output are observable. The annotations above each group report the change in aggregate welfare relative to (i), weighted by the share of high-type matches μ and low-type matches $1 - \mu$.

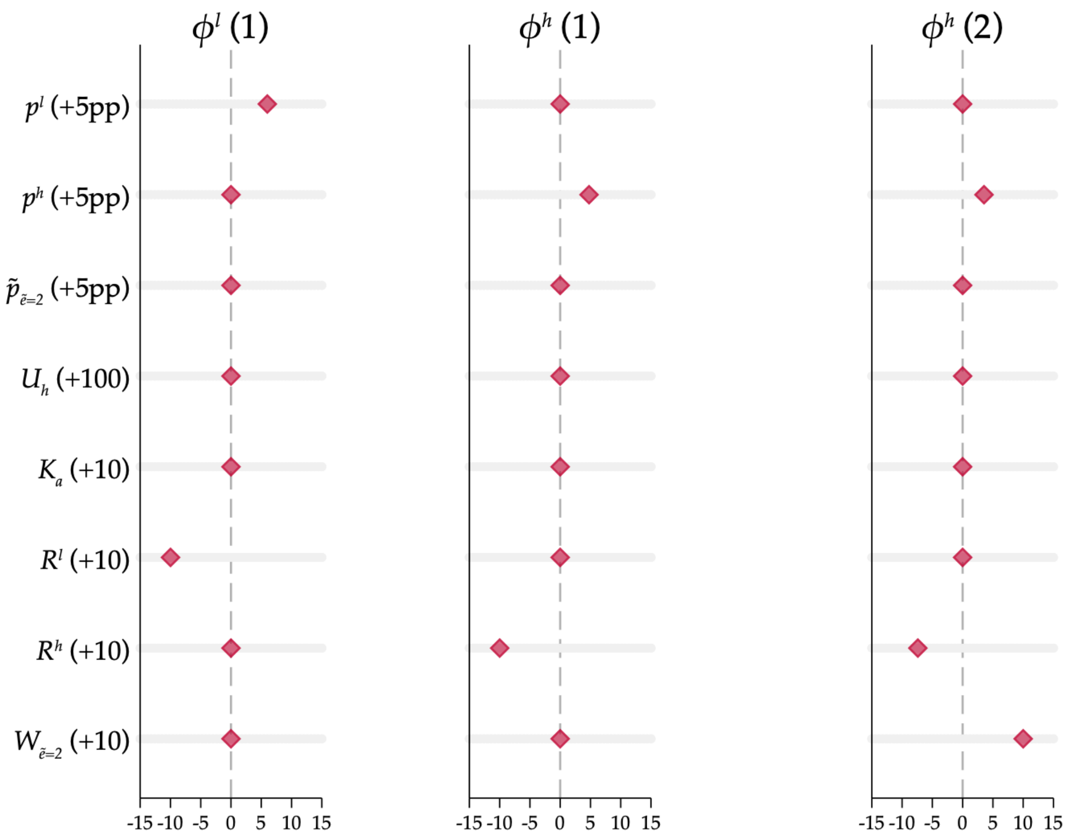


Figure A7: Sensitivity Λ in USD of Parameter Estimates to Estimation Moments

Notes: This figure plots estimated values of $\Lambda = (\mathbf{J}'\mathbf{W}\mathbf{J})^{-1}\mathbf{J}'\mathbf{W}$, where $\mathbf{J}$ is the Jacobian matrix of derivatives of the 8 moments with respect to each of the 3 theoretical parameters ϕ_1^l, ϕ_1^h, and ϕ_2^h, each represented in a different panel, and $\mathbf{W}$ is the weighting matrix. It follows the methodology proposed by Andrews et al. (2017) to measure the sensitivity of parameter estimates to moments. It shows the sensitivity, in dollars, of each parameter estimate to a unit change in each moment (the rows of Λ), displayed on the left y-axis. The values are transformed to facilitate interpretation of changes in moment values, as indicated in parentheses (e.g., +5pp).

B Additional Tables

Table B1: Survey Rounds and Follow-up Rates

	Baseline	Short-term (5mo)		Mid-term (9mo)		Long-term (20mo)	
	N	#	Rate (%)	#	Rate (%)	#	Rate (%)
Panel A. Taxi Businesses (business-level follow-up)							
Businesses surveyed (any owner or driver surveyed)	2196	1960	89.3	1879	85.6	1758	80.1
Drivers surveyed (among baseline drivers)	1821	1652	90.7	1611	88.5	1496	82.2
Panel B. Owner–Driver Pairs (pair-level follow-up)							
Pairs observed (owner or driver surveyed)	608	607	99.8	572	94.1	578	95.1
Owner surveyed	608	535	88.0	493	81.1	453	74.5
Driver surveyed	608	577	94.9	546	89.8	496	81.6
Both owner and driver surveyed	608	505	83.1	467	76.8	371	61.0
Relationship outcomes available	608	548	90.1	476	78.3	363	59.7
Panel C. Initial Non-Adopters (driver-level follow-up)							
Drivers observed	433			366	84.5	367	84.8

Notes: The survey data collection process is detailed in Section 3.1. Rates are computed relative to the baseline N shown in column (2). A taxi business is counted as observed in a wave if the owner or any driver from that business completed the survey. Pair-level rows report whether the baseline owner, the baseline driver, or both responded; relationship outcomes can be recovered when the pair is still together and at least one responds, or when the owner responds about the current/new driver(s). Short-term surveys were collected July–September 2022, mid-term October–December 2022, and long-term September–December 2023. Respondents consent independently in each wave. Baseline non-adopters were administered an adoption survey and followed up in subsequent rounds.

Table B2: Balance Table - Experimental Sample of Taxi Businesses

	Control (1)	Treatment (2)	t-stat (3)	N (4)
Panel A. Taxi Businesses				
Owners Not Driving Their Taxi	0.17	0.17	(-0.03)	2196
	(0.38)	(0.38)		
Owners Driving Their Taxi	0.49	0.51	(0.77)	2196
	(0.50)	(0.50)		
Taxi Drivers (Non-Owners)	0.28	0.28	(-0.42)	2196
	(0.45)	(0.45)		
Part of a Taxi Association	0.38	0.38	(-0.12)	2196
	(0.49)	(0.49)		
Daily Hours Worked	10.62	10.62	(-0.00)	2196
	(2.55)	(2.69)		
Panel B. Individual Characteristics				
Male Respondent	1.00	1.00	(0.78)	2196
	(0.07)	(0.05)		
Education Level: Less Than Primary	0.67	0.68	(0.43)	2196
	(0.47)	(0.47)		
Literacy (Read And Write)	0.70	0.70	(-0.16)	2196
	(0.46)	(0.46)		
Senegalese National	0.93	0.95	(1.56)	2196
	(0.26)	(0.23)		
Wealth Index - PPI Poverty Line 2011	62.83	63.96	(1.38)	2196
	(18.78)	(16.93)		
Saved Money In The Past 3 Months	0.57	0.56	(-0.44)	2196
	(0.50)	(0.50)		
Household Head	0.88	0.87	(-0.84)	2196
	(0.33)	(0.34)		
Omnibus: covariates jointly predict Treatment (p)			(0.23)	
Number of Obs	918	1278		2196

Notes: The following regression is run: $Y_i = \alpha + \beta T_i^{Access} + \epsilon_i$. Heteroskedasticity-robust standard errors are clustered at the business level. The t-Test is reported with the following significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Omnibus p-values report an F-test from regressing the treatment indicator on all baseline covariates (including strata and batch fixed effects), testing that covariates jointly do not predict assignment. All variables are collected during the baseline survey. The PPI Index is an aggregated wealth index specific to Senegal, as described in Poverty Probability Index (PPI).

Table B3: Balance Table - Experimental Sample of Owner-Driver Pairs

	G-O (1)	C-O (2)	N-O (3)	Control (4)	F-test p-value (5)	N (6)
Panel A. Business Setup and Contract						
Owners Not Driving *	0.67	0.58	0.64	0.66	[0.40]	608
	(0.47)	(0.50)	(0.48)	(0.47)		
Owns Only One Taxi *	0.94	0.88	0.93	0.93	[0.27]	608
	(0.23)	(0.33)	(0.26)	(0.26)		
Long Relationship (> 2 Years) *	0.44	0.47	0.39	0.46	[0.52]	608
	(0.50)	(0.50)	(0.49)	(0.50)		
Age of the Relationship	3.52	3.19	2.78	3.63	[0.28]	608
	(4.30)	(3.48)	(3.35)	(4.26)		
Proxy for Risk Aversion (Lining in Garages) *	-6.20	-13.31	-14.15	-15.20	[0.87]	608
	(70.07)	(101.59)	(104.78)	(108.16)		
Upfront Payment / Salary W	0.55	0.50	0.49	0.55	[0.68]	608
	(0.50)	(0.50)	(0.50)	(0.50)		
Upfront Payment / Salary Value W - Unconditional	41.13	41.01	36.70	43.35	[0.62]	608
	(38.53)	(47.87)	(38.54)	(45.30)		
Weekly Rent Target Value R	102.27	104.72	102.97	103.52	[0.54]	608
	(14.26)	(12.03)	(13.79)	(13.15)		
Family Business	0.56	0.56	0.59	0.51	[0.52]	608
	(0.50)	(0.50)	(0.49)	(0.50)		
Owner also President of Taxi Association	0.03	0.00	0.02	0.00	[0.01]	608
	(0.18)	(0.00)	(0.13)	(0.00)		
Panel B. Driver's Effort						
End-Start Work Time	12.05	12.65	12.58	12.65	[0.29]	608
	(2.87)	(2.77)	(3.05)	(2.93)		
Daily Hours Worked	10.38	10.84	10.91	10.95	[0.26]	608
	(2.64)	(2.56)	(2.76)	(2.58)		
Driver Defaulted at Least Once in the Past Month	0.40	0.45	0.50	0.51	[0.16]	608
	(0.49)	(0.50)	(0.50)	(0.50)		
Avg Daily Revenue Collected	48.05	48.02	48.21	48.16	[1.00]	608
	(9.51)	(9.78)	(9.91)	(9.62)		
Avg # of Daily Customers	15.77	15.84	16.88	16.55	[0.41]	608
	(5.56)	(5.02)	(6.56)	(6.11)		
Panel C. Driver's Characteristics						
Education level: less than primary	0.69	0.70	0.74	0.68	[0.70]	608
	(0.46)	(0.46)	(0.44)	(0.47)		
Wealth Index - PPI Poverty Line 2011	64.78	65.11	63.60	63.45	[0.78]	608
	(16.82)	(15.88)	(16.14)	(17.26)		
High Digital Users (> 6 Taxi-like Transactions) *	0.52	0.52	0.43	0.47	[0.39]	608
	(0.50)	(0.50)	(0.50)	(0.50)		
Omnibus: covariates predict G-O vs Control (p)					(0.42)	
Omnibus: covariates predict C-O vs Control (p)					(0.44)	
Omnibus: covariates predict N-O vs Control (p)					(0.47)	
Omnibus: covariates predict G-O vs N-O (p)					(0.66)	
Number of Obs	127	116	111	254		608

Notes: This table compares owners' and drivers' baseline characteristics across observability treatment groups. The treatment provided a digital payment technology to the taxi owner and their driver(s). Among treated pairs, observability was randomized as follows: G-O (Granular Observability), C-O (Coarse Observability), N-O (No Observability); Control received no technology. Column (5) reports p-values from one-way ANOVA tests of equality of means across the four arms for each covariate (with significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$). The omnibus rows at the bottom report p-values from F-tests of whether baseline covariates jointly predict assignment to each treatment arm versus Control or N-O, from linear probability models that include the randomization strata and batch fixed effects (robust standard errors).
Variables used to stratify are described in Section 2.4, indicated with a star (*) in the Table. Digital personal transactions are observed in the administrative data. All other variables are collected during the baseline survey and averaged, and missing responses (refused to answer or don't know) are dummied out from the joint F-test. The PPI Index is an aggregated wealth index, specific to Senegal, to measure poverty likelihood, as described in Poverty Probability Index (PPI). All values are converted to USD (USD 1 = CFA 600).

Table B4: Mapping From Effort to Output

	Daily Output			
	(1)	(2)	(3)	(4)
Driver's Hours			2.476***	2.396***
Worked That Day			(0.063)	(0.076)
# of Taxis	1662	1662	1662	1662
Observations - Days	9177	9177	9154	9154
Adjusted R2	0.008	0.015	0.222	0.455
Days of the Week / Days of the Month FEs	Yes	Yes	Yes	Yes
Calendar Date FEs	No	Yes	Yes	Yes
Hours Worked	No	No	Yes	Yes
Driver FEs	No	No	No	Yes

Notes: This table reports regressions that predict driver's daily output using calendar controls and measures of effort. Values are converted into USD (USD 1 = CFA 600). 'Days of Week/Month FEs' specifications control for the day of the week (e.g., Monday) and the day of the month (e.g., the 10th), as these are considered to influence part of the demand variation in this setting. The location is the same for all drivers as they all operate in Dakar, Senegal. The sample includes taxi drivers that have been surveyed either at short-term *or* at mid-term and regressions are estimated at the driver-day level using all non-missing observations. Heteroskedasticity-robust standard errors are provided in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B5: Non-Differential Attrition Rates Across Observability Treatments

		Treated		Control	p-value	N
		(1)		(2)	(3)	(4)
Panel A. Taxi Businesses Overall						
Short-Term Survey		0.12		0.10	(0.31)	2196
		(0.32)		(0.30)		
Mid-Term Survey		0.15		0.14	(0.67)	2196
		(0.36)		(0.35)		
Long-Term Survey		0.19		0.21	(0.44)	2196
		(0.39)		(0.40)		
Number of Obs		918		1278		2196
	G-O	C-O	N-O	Control	p-value	N
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B. Taxi Businesses with an Employee						
Short-Term Survey	0.01	0.00	0.00	0.00	(0.29)	608
	(0.09)	(0.00)	(0.00)	(0.00)		
Mid-Term Survey	0.07	0.07	0.05	0.05	(0.84)	608
	(0.26)	(0.25)	(0.23)	(0.22)		
Long-Term Survey	0.05	0.09	0.03	0.04	(0.19)	608
	(0.21)	(0.28)	(0.16)	(0.20)		
Number of Obs	127	116	111	254		608
Panel C. Taxi Drivers Non-Adopters						
Mid-Term Survey	0.13		0.17		(0.24)	433
	(0.34)		(0.38)			
Long-Term Survey	0.15		0.15		(0.99)	433
	(0.36)		(0.36)			
Number of Obs	203		230			433

Notes: This table compares the attrition rate within each survey round across observability treatment groups. For owner–driver pairs, attrition is defined as cases where both the owner and the driver refused to respond (no information on the contract and relationship). The analysis shows no differential attrition across groups for all survey rounds. The treatment involved providing the digital payment technology to the taxi owner and their driver(s). Among the treated pairs, the following observability treatments were randomized: G-O (Granular Observability), C-O (Coarse Observability), and N-O (No Observability) of the owner on their driver's transactions. For the reluctant drivers (non-adopters), only G-O and N-O were randomized.
p-values of the differences are displayed. In panel B, p-values from the joint F-test (ANOVA) are reported. All with the following significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B6: Impact of Digital Payments on Driver's Own Revenue and Profit

	Avg Profit	Daily Profit	Revenue	Number of Customers	Avg Price	Customer per Hour	Revenue per Hour
Panel A. Short-Term 5-Month Survey							
Technology Access	0.047	0.494	-0.166	-0.475*	0.026	0.045	0.035
	(0.476)	(0.734)	(0.627)	(0.279)	(0.221)	(0.069)	(0.085)
Observations	1298	1596	1582	1489	1472	1489	1593
Control Mean at Short-Term	14.36	17.73	51.42	16.00	3.91	1.65	5.37
% Change T at Short-Term	0.32	2.79	-0.32	-2.97	0.66	2.72	0.65
Panel B. Mid-Term 9-Month Survey							
Technology Access	-0.313	-0.487	-0.287	-0.067	-0.290	-0.044	-0.140*
	(0.427)	(0.769)	(0.617)	(0.232)	(0.230)	(0.028)	(0.082)
Observations	1475	1478	1459	1391	1375	1390	1477
Control Mean at Mid-Term	19.35	22.56	50.82	14.53	4.29	1.52	5.31
% Change T at Mid-Term	-1.62	-2.16	-0.56	-0.46	-6.77	-2.91	-2.64

Notes: Baseline data were collected March–June 2022, short-term July–September 2022, and mid-term October–November 2022 (9 months). Outcomes are regressed on treatment (technology access), with pure control as the omitted group: $y_{ij} = \beta_0 + \beta T_{ij}^{Access} + \alpha_s + \epsilon_{ij}$, where i indexes drivers and j businesses and α_s strata and batch fixed effects. Standard errors are cluster-robust at the business level; significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Percent changes are coefficients divided by control means. The sample includes all drivers surveyed at least once (about 90%); values are missing when drivers could not recall/refused to respond. Regressions include baseline controls when available. The outcomes, except in Column 1, are averaged from the last 3 days prior to the survey date. They are constructed from the following questions:
- *Avg Profit*: Over the past 30 days, what is the average amount of money you earn per working day, after paying all your expenses (fuel, payments, repairs, police, contributions, food, etc.)? — à la De Mel et al. (2009).
-*Daily Profit*: For each of the 3 days prior to the survey date, profit is computed from the taxi drivers' production function is composed of the revenue collected, coming from a unique source (passengers), minus each cost category can be listed as follows: (i) fuel, (ii), transfer to the owner, (iii) side expenses (small repairs, police, washing, toll, association contribution excluding food), and (vi) food/drinks/stimulants throughout the working days.
- *Revenue*: For each day: On that day, over working hours, how much money did you collect? (from passengers' payment)?
-*Customers*: For each day: On that day, over working hours, how many customers did you have?
- *Average Price*: Daily collection divided by the number of customers.
- *Customer/Hour*: Drivers' productivity in terms of number of customers: number of customers divided by the working hours (start-end).
- *Revenue/Hour*: Drivers' productivity in terms of revenue: total revenue collected divided by the working hours (start-end).
All values are converted in USD when relevant (USD 1 = CFA 600).

Table B7: Impact of Digital Payments on Revenue and Profit - Power Calculations and Minimum Detectable Effects (MDE)

	Avg Profit (1)	Daily Profit (2)	Revenue (3)	# Customers (4)	Avg Price (5)	Customer/Hour (6)	Revenue/Hour (7)
Panel A. Short-Term 5-Month Survey							
MDE (in %)	9.57	12.20	3.51	4.82	16.02	12.04	4.33
Observations	1652	1652	1652	1652	1652	1652	1652
Control Mean$_{Avg}$	14.36	17.73	30853.03	16.00	3.91	1.65	5.37
Panel B. Mid-Term 9-Month Survey							
MDE (in %)	6.49	9.63	3.51	4.44	16.29	5.11	4.31
Observations	1611	1611	1611	1611	1611	1611	1611
Control Mean$_{Avg}$	19.35	22.56	30489.74	14.53	4.29	1.52	5.31

Notes: This table computes the minimal detectable effects (MDEs) for the profit outcomes of interest for both survey rounds. It performs a simple two-sample means test between the control and treatment group to quantify the power of the experiment on the profit dimensions, using the mean and standard deviation of the control and treatment groups. All variables, except in column 1, are averaged at the daily level from the last 3 days prior to the survey, as in the main results. MDEs were computed using analytical methods for a power of 0.8 and a significance level of 0.05.
The outcomes are constructed from the following questions:
- *Avg Profit*: Over the past 30 days, what is the average amount of money you earn per working day, after paying all your expenses (fuel, payments, repairs, police, contributions, food, etc.)? — a la De Mel et al. (2009).
-*Daily Profit*: For each of the 3 days prior to the survey date, profit is computed from the taxi drivers' production function is composed of the revenue collected, coming from a unique source (passengers), minus each cost category can be listed as follows: (i) fuel, (ii), transfer to the owner, (iii) side expenses (small repairs, police, washing, toll, association contribution excluding food), and (vi) food/drinks/stimulants throughout the working days.
- *Revenue*: For each day: On that day, over working hours, how much money did you collect (from passengers' payment)?
-*Customers*: For each day: On that day, over working hours, how many customers did you have?
- *Average Price*: Daily collection divided by the number of customers.
- *Customer/Hour*: Drivers' productivity in terms of number of customers: number of customers divided by the working hours (start-end).
- *Revenue/Hour*: Drivers' productivity in terms of revenue: total revenue collected divided by the working hours (start-end).
All values are converted to USD when relevant (USD 1 = CFA 600).

Table B8: Impact of Digital Payments on Additional Outcomes

	(1) Theft Anxiety	(2) Keep Records	(3) Impulsive Purchases	(4) Able to Save
Panel A. Short-Term 5-Month Survey				
Technology Access	-0.038*	0.054***	-0.071***	-0.016
	(0.023)	(0.010)	(0.026)	(0.028)
Observations	1629	1571	1502	1349
Control Mean at Short-Term	0.76	0.02	0.46	0.61
% Change T at Short-Term	-5.00	287.30	-15.39	-2.62
Panel B. Mid-Term 9-Month Survey				
Technology Access	-0.048*	0.037***		-0.022
	(0.027)	(0.010)		(0.028)
Observations	1570	1514		1268
Control Mean at Mid-Term	0.57	0.02		0.65
% Change T at Mid-Term	-8.37	177.40		-3.32

Notes: Baseline data were collected March–June 2022, short-term July–September 2022, and mid-term October–November 2022 (9 months). Outcomes are regressed on treatment (technology access), with pure control as the omitted group: $y_{ij} = \beta_0 + \beta T_{ij}^{Access} + \alpha_s + \epsilon_{ij}$, where i indexes drivers and j businesses and α_s strata and batch fixed effects. Standard errors are cluster-robust at the business level; significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Percent changes are coefficients divided by control means. The sample includes all drivers surveyed at least once; values are missing when drivers did not work in the considered time period or could not recall/refused to respond. Regressions include baseline controls when available. All outcomes are dummies (0-1) constructed from the following questions:
- *Theft Anxiety*: In the last 3 months, how often are you worried that part of your collection is robbed? — Dummy equal to 1 from Always (every day) to sometimes. 0 if almost never and never.
- *Keep Records*: During the last 3 months, have you kept a written or digital history of your transactions to do the accounting?
- *Impulsive Purchases*: Consider the past 7 days: did you buy perfume, deodorant, clothing, or a pillow while driving or working? *Note:* This list was constructed from qualitative fieldwork with drivers and reflects common impulsive purchases often made from street vendors while on the road.
- *Able to Save*: During the last 3 months, how much have you been able to save? — A dummy is set to 1 if more than CFA 100,000 (USD 170) was saved, which is considered a significant amount. Robustness checks also performed with other values show no effect.

Table B9: Transaction Costs Predict Willingness-To-Pay

	log(Willingness-To-Pay + 1)				
	(1)	(2)	(3)	(4)	(5)
Panel A: Baseline					
TC	0.197***	0.148***	0.272***	0.156***	0.096***
	(0.043)	(0.046)	(0.043)	(0.044)	(0.027)
Obs	1654	1630	1654	1654	1604
Benchmark Good Control	Yes	Yes	Yes	Yes	Yes
TC Mean	0.482	0.599	0.533	0.414	1.916
Panel B: Endline					
TC	0.301***	0.228**	0.492***	0.482***	0.051***
	(0.091)	(0.091)	(0.077)	(0.150)	(0.010)
Obs	1311	1310	1310	1311	1309
Treatment Status Control	Yes	Yes	Yes	Yes	Yes
TC Mean	0.193	0.236	0.299	0.066	3.639
TC =	*Any Time Lost*	*Refused Customers*	*Reduced Price*	*Mistakes Giving Change*	*log(Imputed Loss+1)*

Notes: Baseline specifications control for willingness-to-pay (WTP) for a benchmark good (a bottle of water), following Dizon-Ross and Jayachandran (2022). Endline specifications additionally control for treatment assignment (removing these controls doesn't affect any of the result). Baseline survey data were collected from March to May 2022 and refer to the 7 days prior to the interview. Baseline WTP for the technology was elicited using an incentivized Becker–DeGroot–Marschak (BDM) mechanism (Becker et al., 1964). To preserve the original randomization, the purchase lottery was implemented for a randomly selected 5% of treated drivers: a price was drawn from a uniform distribution and the product was provided when the draw was below the driver's stated WTP, and the price distribution was such that most of the 5% of treated drivers actually were given the product below their WTP.

Endline survey data refer to the 7 days prior to the interview. Endline WTP was elicited without financial incentives because withdrawing the technology from drivers was not feasible.

Standard errors are clustered at the business level. The dependent variable is in logs, so coefficients can be interpreted approximately as percent changes in WTP. Column (5) is a log–log specification, so the coefficient is an elasticity. Refused or don't know are coded as missing.

The regressors TC (Transaction Costs), displayed at the bottom of the table, are dummies (0-1) coming from survey data about the past 7 days prior to the survey date. They are constructed from the following questions:

- *Any Time Lost*: How many times have you wasted time (more than 10 minutes) looking for small-change during your work?
- *Refused Customers*: How many customers have you turned down because they only wanted to pay with electronic money, and not cash?
- *Reduced Price*: How many times have you reduced the price of the ride because of the small change problem?
- *Mistakes Giving Change*: How many times have you lost part of your earnings with customers due to giving change?
- *Imputed Loss*: how much money lost in total, imputed from driver's interviews. Converted into a log scale for interpretability.

All values are converted in USD when relevant (USD 1 = CFA 600).

Table B10: First-Stage: Observability Impact on Information Frictions

	Owner's Knowledge				
	Work Days (Cross-Checked)	Hours Worked	Digital Revenue	Use SMS To Observe Effort	Technology Monitoring Daily/Weekly
	(1)	(2)	(3)	(4)	(5)
Granular	0.106	0.105	0.139*	0.358***	0.435***
Observability	(0.087)	(0.086)	(0.072)	(0.056)	(0.074)
Coarse Observability	0.051	-0.042	0.062	0.174***	
	(0.095)	(0.090)	(0.070)	(0.051)	
Observations	229	215	289	271	46
Baseline Control	YES	YES	YES		
No Observability Mean	0.29	0.59	0.21	0.00	0.00
% Change Granular Observability	36.8	17.9	67.2		

Notes: Sample of owners surveyed at mid-term (around 9 months), all in the groups of drivers with access to the technology. The group of control drivers is not included since their owners were not asked about their (nonexistent) digital revenue. Robust standard errors are shown in parentheses with the following significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Owners' knowledge in the 7 days prior to the interview was elicited for each of the outcomes in Columns 1 to 3. Respectively, these columns represent whether the owner claimed to know the driver's effort (work days and hours) and the driver's total revenue collected digitally. Since 92% of owners in the 'no observability' group claimed to know their drivers' days worked, all reports were cross-checked by comparing the days reported by the owners to the actual work days of the drivers (as reported by the drivers in a separate survey). A substantial share of owners were incorrect. I report a dummy for whether the two reports match (i.e., both owners and drivers report the same number of days worked in the past 7 days). Refusals to answer are coded as missing.

Column 4 displays the share of owners reporting the use of the SMS specifically (received in both G-O and C-O) to observe their driver's working hours and days, with a dummy equal to 0 for the 'no observability' group due to treatment compliance (no SMS provided). The last column shows the share of owners reporting that they check the driver's transactions daily or weekly (and 0 for the 'no observability' and 'coarse observability' groups since owners cannot observe transactions under these groups). It is important to note that the sample size for this particular question is smaller because it was only added at the end of the mid-term data collection survey.

Table B11: Positive Correlation Between Self-Reported Effort and Output And Digital Payments Usage

	Day Worked		# Hours Worked		Total Revenue Collected		# Of Passengers	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
# of Digital Transactions	0.105*** (0.004)		0.213*** (0.082)		2.731*** (0.513)		0.905*** (0.187)	
Revenue Collected Digitally		0.028*** (0.001)		0.048** (0.022)		0.789*** (0.128)		0.143*** (0.049)
# of Taxis	979	979	979	979	979	979	979	979
Mean No Digital Activity	0.75	0.75	9.90	9.90	50.08	50.08	14.94	14.94
Sample of Days Considered	All		Used Digital		Used Digital		Used Digital	

Notes: Regressions are performed at the day level, controlling for day fixed effects and survey timing (mid-term or long-term) for treated taxis. The outcome variables are self-reported by the drivers for the past 3 days prior to the survey, at both mid-term and long-term. Drivers are asked about specific dates, and those self-reports are compared to the actual digital transactions recorded by the mobile money partner company. The two dependent variables are the number of digital transactions on a given day and the total money collected. The first two columns leverage the fact that, by asking drivers to report on the past 3 days worked, the responses also reveal information about days not worked. In Columns 3 to 8, the sample size is restricted to days where at least one digital transaction was made. Robust standard errors are provided in parentheses, with the following significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The 'mean no digital activity' represents the mean outcome for the full sample of excluded days with no digital activity (i.e., the number of transactions was zero on those days). Taxi drivers who refused to respond to these questions are excluded. Values are converted from CFA to USD, with CFA 600 = USD 1.

Table B12: Impact of Observability on Taxi Owner's Trust

	Revealed-preference trust	Stated beliefs and trust		
	Taxi Parked at Driver's House	Perceived Moral Hazard	Owner's Trust	Driver's Trust
	(1)	(2)	(3)	(4)
Granular Observability	0.143**	-0.142***	-0.034	-0.086
	(0.059)	(0.053)	(0.206)	(0.154)
Coarse Observability	-0.064	-0.077	-0.316	0.016
	(0.065)	(0.062)	(0.234)	(0.167)
No Observability	0.042	0.079	-0.582**	-0.176
	(0.066)	(0.069)	(0.239)	(0.180)
Observations	410	418	371	423
Control Mean	0.509	0.289	8.901	9.147
Relative Scale Control			YES	YES
% Change Granular Observability	28.1	-49.2	-0.4	-0.9
F-test Granular O = No O (p-value)	0.18	0.00	0.05	0.65

Notes: Mid-term survey data collected from October-December 2022, approximately 9 months later. Questions were asked to the owner (columns (1)–(3)) or the driver (column (4)), hence the smaller sample size. The model is defined as $y_j = \beta_0 + \beta_1 T_j^{GranularObs} + \beta_2 T_j^{CoarseObs} + \beta_3 T_j^{NoObs} + \alpha_s + \epsilon_j$, where y_j is the outcome variable displayed at the top of the column and α_s are the strata fixed effects. Heteroskedasticity-robust standard errors are used. The F-test comparing *Granular Observability* to *No Observability* is shown at the bottom of the table. Respondents who refused to respond or said they did not know are coded as missing. Relative Scale Control measures general trust in the opposite profession overall (owners' trust in drivers, or drivers' trust in owners), rather than trust in their own employer/employee.
The following questions were asked for each of the outcomes displayed:
- *Park at Driver*: Where is the taxi generally parked outside working hours? Whether the driver is allowed to park the taxi at their own house. In this context, owners who lack trust in their drivers often require them to park the car at a specific location they can monitor each day.
- *Perceived Moral Hazard*: If your driver collects less than CFA 25,000 (USD 40) during a full day of work, do you believe this is due to bad luck, God's will, unpredictable aspects of the job, or the driver's low effort?
- *Owner's Trust*: Generally, if you had to rate your trust in *taxi drivers* (not your own, but *taxi drivers* more broadly), from 0 to 10, what score would you give them?
- *Driver's Trust*: How would you rate your trust in your current owner from 0 to 10?

Table B13: Impact of Observability on Owner's Profit (Including Maintenance)

	Owner's Monthly	
	Profit (w Maint.)	Profit (w Maint., Wins.)
	(1)	(2)
Granular Observability	16.960	14.583
	(24.911)	(23.958)
Coarse Observability	-16.962	-16.393
	(26.124)	(25.873)
No Observability	-8.392	-0.537
	(31.802)	(27.688)
Observations	395	395
Control Mean	151.07	152.42
% Change Gran. Obs.	11.23	9.57
F-t Gran. O = No O (p-val)	0.47	0.62

Notes: Same specification as Table 4, but profit subtracts maintenance costs. Column (2) winsorizes at the 2nd and 98th percentiles.

Table B14: Drivers Under Observability Have Higher Digital Usage

	# of transactions	# of transactions	Amount (USD)	# of active weeks	# of active days
	(1)	(2)	(3)	(4)	(5)
Granular Observability	0.401***	0.351**	83.250**	0.177**	0.256**
	(0.146)	(0.140)	(40.574)	(0.086)	(0.110)
Coarse Observability	0.124	0.039	2.242	0.088	0.042
	(0.139)	(0.136)	(32.388)	(0.086)	(0.109)
Obs	354	354	354	354	354
Mean No Observability	59.31	59.31	220.81	15.24	38.59
Baseline P2P Control	NO	YES	YES	YES	YES

Notes: Administrative data provided by the mobile money partner company, collected from April to December 2022, the end of the mid-term survey. Zeros (drivers not using the technology at all) are included and account for approximately 15% of drivers. Robust standard errors. The significance of each coefficient is denoted as follows: $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***. Columns 1, 2, 4, and 5 display coefficients from Poisson regressions (their outcomes are counts), while Column 3 represents an OLS regression. The pure control group sample is excluded because these drivers did not have access to the technology.

The variable # of transactions represents the total number of transactions, while the amount corresponds to the total value of those transactions (in USD). The number of active weeks/days refers to the count of time periods with at least one transaction received.

Baseline P2P Control denotes the number of transactions received before the experiment launch that resemble taxi P2P transactions. These are defined as P2P transactions with values between 1,000 and 3,500 FCFA (the 5th and 95th percentiles of business transactions received). This control is included to mitigate variations caused by potential pre-trends in digital usage.

B.1 Observability Impacts Robustness: No Contamination Bias With Multiple Treatments

Table B15: Impact of Observability on Contracts — Contamination Bias With Multiple Treatments (Robustness)

	Upfront Payment 'Salary' Dummy	Upfront Payment 'Salary' Value (USD)	Weekly Rent Target Value (USD)	Separation p
	(1)	(2)	(3)	(4)
Granular Observability	0.124***	10.496***	3.129*	-0.109**
	(0.043)	(3.307)	(1.849)	(0.048)
Coarse Observability	-0.013	0.393	0.612	-0.041
	(0.050)	(3.910)	(1.454)	(0.052)
No Observability	-0.009	0.272	0.763	-0.028
	(0.047)	(3.728)	(1.493)	(0.051)
Observations	476	476	476	572
Control Mean	0.76	55.62	99.96	0.32
% Change Granular Observability	16.38	18.87	3.13	-33.98
Chi-squared test Granular O = No O (p-value)	0.01	0.01	0.26	0.16

Notes: This table estimates the effect of a single as-good-as-randomly assigned treatment using a partially linear model, following Goldsmith-Pinkham et al. (2024), so that β identifies a convex average of heterogeneous treatment effects. This analysis accounts for potential contamination bias in randomized experiments with multiple treatments and strata and batch fixed effects. Results remain unchanged. The table shows business-level regressions of the contract outcomes on the three treatment arms, with the omitted category the pure control group. The outcome variable is displayed at the top of the column. These regressions are conducted for mid-term period (approximately 9 months after the baseline survey). Heteroskedasticity-robust standard errors and the significance of each coefficient is denoted by asterisks: $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***. Each regression considers strata fixed effects. Outcomes: *Upfront Payment Dummy* = 1 if the owner provides a monthly fixed payment (salary), *Upfront Payment Value* = amount of that payment in USD (0 if none, USD 1 = CFA 600), *Weekly Rent Target* = agreed weekly rental payment (USD), *Separation* = 1 if owner and driver no longer work together at survey time.

Table B16: Impact of Observability on Contracts — Pairs Remained Together at Mid-Term

	Upfront Payment 'Salary' Dummy	Upfront Payment 'Salary' Value (USD)	Weekly Rent Target Value (USD)	Default Rate
	(1)	(2)	(3)	(4)
Granular Observability	0.104**	9.740***	2.745	-0.099
	(0.046)	(3.599)	(2.109)	(0.061)
Coarse Observability	-0.035	-1.277	0.033	0.055
	(0.056)	(4.444)	(1.691)	(0.069)
No Observability	-0.018	-0.659	0.237	0.021
	(0.052)	(4.192)	(1.859)	(0.067)
Observations	406	400	401	406
Control Mean	0.77	57.11	100.28	0.30
% Change Granular Observability	13.41	17.05	2.74	-32.48
F-test Granular O = No O (p-value)	0.03	0.02	0.34	0.11

Notes: Tables restricted to pairs that remained together at mid-term (endogenous restriction). Business-level regressions of the contract outcomes on the three treatment arms, with the omitted category the pure control group. The model is defined as $y_j = \alpha + \beta_1 T_j^1 + \beta_2 T_j^2 + \beta_3 T_j^3 + \epsilon_j$, where y_j is the outcome variable displayed at the top of the column. These regressions are conducted for mid-term period (approximately 9 months after the baseline survey). Heteroskedasticity-robust standard errors are used and the significance of each coefficient is denoted by asterisks: p $<$ 0.1 *, p$<$ 0.05 **, p$<$ 0.01 ***. Each regression includes strata and batch fixed effects.

- *Upfront Payment 'Salary' Dummy*: Whether the owner provides a monthly upfront fixed payment to their driver (also referred to as a 'salary' in the industry).
- *Upfront Payment 'Salary' Value (USD)*: Value of the monthly upfront fixed payment to the driver, including 0 if no payment is provided. Values converted in USD (USD 1 = CFA 600).
- *Weekly Rent Target Value*: Weekly rental target amount from driver to owner, as agreed between owner and driver. Values converted in USD (USD 1 = CFA 600).
- *Default Rate*: The proportion of drivers who default at least once a month (in the past three months), according to either the owner or their driver's reports.

Table B17: Impact of Observability on Contracts and Relationships - Long-Term (Nearly 2 Years)

	Upfront Payment 'Salary' Dummy	Upfront Payment 'Salary' Value (USD)	Weekly Rent Target Value (USD)	Separation
	(1)	(2)	(3)	(4)
Granular Observability	0.102	7.830	-0.578	-0.057
	(0.077)	(6.297)	(4.113)	(0.069)
Coarse Observability	-0.051	-4.543	4.079	-0.053
	(0.078)	(6.516)	(6.855)	(0.069)
Observations	209	203	162	334
Control Mean	0.74	56.77	99.13	0.63
% Change Granular Observability	13.84	13.79	-0.58	-9.06

Notes: Business-level OLS regressions of the contract outcomes on the three treatment arms, with the omitted category the No Observability group (the control group received the technology after nine months). The model is defined as $y_j = \beta_0 + \beta_1 T_j^{GranularObs} + \beta_2 T_j^{CoarseObs} + \alpha_s + \epsilon_j$, where y_j is the outcome variable displayed at the top of the column and $alpha_s$ are the strata fixed effects. These regressions are conducted at long-term (nearly 2 years after the baseline survey). Contract outcomes are recorded only for businesses that still employ a driver at the two-year survey, whereas "Separation" is coded for every baseline owner–driver pair, explaining the sample size difference. Heteroskedasticity-robust standard errors are used and the significance of each coefficient is denoted by asterisks: $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***.

- *Upfront Payment 'Salary' Dummy*: Whether the owner provides a monthly upfront fixed payment to their driver (also referred to as a 'salary' in the industry).
- *Upfront Payment 'Salary' Value (USD)*: Value of the monthly upfront fixed payment to the driver, including 0 if no payment is provided. Values converted in USD (USD 1 = CFA 600).
- *Weekly Rent Target Value*: Weekly rental target amount from driver to owner, as agreed between owner and driver. Values converted in USD (USD 1 = CFA 600).
- *Separation*: Owner and driver are not working together at the time of the survey.

Table B18: Impact of Observability on Self-Reported Effort

	Hours Worked	End-Start Time	Revenue	# Customers
	(1)	(2)	(3)	(4)
Granular Observability	0.696**	0.675**	0.051	0.351
	(0.291)	(0.314)	(1.452)	(0.689)
Coarse Observability	0.126	0.021	-2.145	0.142
	(0.312)	(0.339)	(1.456)	(0.653)
No Observability	0.300	0.397	-0.169	-0.516
	(0.343)	(0.395)	(1.327)	(0.804)
Observations	593	593	593	593
Control Mean	10.21	11.77	50.70	14.88
% Change Granular Observability	6.8	5.7	0.1	2.4
F-test Granular O = No O (p-value)	0.30	0.52	0.89	0.32

Notes: Short-term survey data were collected from July to September 2022, and mid-term survey data from October to November 2022 (after 9 months). The outcome is the self-reported effort by the driver in the past 3 days, averaged before the survey. Difference-in-difference regressions were conducted, controlling for individual fixed effects. Baseline controls include the same variable, but instead of using the past 3 days, they average over the last 3 days worked to avoid specific variations during Ramadan (which occurred during part of the baseline phase). Robust standard errors are used. The sample includes all drivers surveyed at least once at short and/or mid-term, with missing values otherwise. The F-stat testing for the difference between the estimates of No Observability and Granular Observability is shown at the bottom of the table.

- *Hours Worked*: On day X, please indicate what time you started driving, what time you finished driving, and how many hours of break you took in between.
- *End-Start Time*: On day X, please indicate what time you started driving and what time you finished driving.
- *Customers*: On day X, between Xh and Yh, how many customers did you have?
- *Revenue*: On day X, between Xh and Yh, how much money did you collect in total? — converted to USD, using USD 1 = CFA 600.

Table B19: Heterogeneous Impact of Observability on Contracts Based on Driver's Digital Usage

Top 50% Number of Daily Digital Transactions

	Upfront Payment 'Salary' Dummy	Upfront Payment 'Salary' Value (USD)	Weekly Rent Target Value	Separation p
	(1)	(2)	(3)	(4)
Granular Observability	0.079	6.676	0.775	-0.047
	(0.072)	(5.887)	(4.276)	(0.088)
Granular Observability × X	0.086	4.557	2.711	-0.035
	(0.119)	(9.666)	(4.837)	(0.120)
Coarse Observability	-0.016	-0.000	1.738	-0.017
	(0.091)	(7.370)	(3.422)	(0.087)
Coarse Observability × X	0.034	-0.752	-3.276	0.092
	(0.144)	(11.668)	(4.799)	(0.125)
X = Proxy for Digital Intensity: 50p	-0.045	0.347	0.359	-0.151*
	(0.099)	(8.067)	(3.061)	(0.086)
Observations	276	270	275	331
Mean No Observability	0.75	56.56	100.82	0.34

Notes: Business-level regressions of the contract outcomes on the treatment arms, with no observability as the omitted category. The outcome variable is displayed at the top of each column. These regressions are conducted for the mid-term period until December 2022 (approximately 9 months after the baseline survey). They include interaction terms to study heterogeneity based on digital usage. The proxy for Digital Intensity is defined as the 50th percentile across drivers based on the average number of transactions received daily throughout the experiment, for days with at least one transaction. Although digital usage is also an endogenous variable, this table aims to provide suggestive evidence linking digital usage to contract changes.

Robust heteroskedasticity-consistent standard errors are reported, and the significance of each coefficient is denoted by asterisks: $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***. The pure control group is not included because their digital usage is zero. Each regression includes controls for strata fixed effects.

- *Upfront Payment 'Salary' Dummy*: Whether the owner provides a monthly upfront fixed payment to their driver (also referred to as a 'salary' in the industry).
- *Upfront Payment 'Salary' Value (USD))*: Value of the monthly upfront fixed payment to the driver, including 0 if no payment is provided. Values converted to USD (USD 1 = CFA 600).
- *Weekly Rent Target Value*: Weekly rental target amount from driver to owner, as agreed between owner and driver. Values converted to USD (USD 1 = CFA 600).
- *Separation*: Owner and driver are not working together at the time of the survey.

Table B20: Similar Observability Impact on Peer-to-peer Transactions Received by Drivers

	# of transactions	# of transactions	Amount (USD)	# of active weeks	# of active days
	(1)	(2)	(3)	(4)	(5)
Panel A. P2P transactions Within Taxi Value Ranges					
Granular Observability	0.194*	0.108	7.982	0.023	0.046
	(0.100)	(0.089)	(7.121)	(0.060)	(0.076)
Coarse Observability	0.282***	0.154*	14.673**	0.026	0.090
	(0.101)	(0.090)	(7.454)	(0.064)	(0.078)
Obs	354	354	354	354	354
Mean No Observability	22.42	22.42	70.15	10.33	18.28
Baseline P2P Taxi Control	NO	YES	YES	YES	YES
Panel B. P2P transactions Outside Taxi Value Ranges					
Granular Observability	0.194	0.095	-27.772	-0.060	0.006
	(0.167)	(0.140)	(127.027)	(0.065)	(0.114)
Coarse Observability	0.226	0.063	105.249	0.090	0.110
	(0.152)	(0.141)	(142.785)	(0.070)	(0.116)
Obs	354	354	354	354	354
Mean No Observability	28.62	28.62	784.81	11.91	22.44
Baseline P2P Non-Taxi Control	NO	YES	YES	YES	YES

Notes: Administrative data provided by the mobile money partner company, collected from April to December 2022, the end of the mid-term survey. Zeros (drivers not using the technology at all) are included and account for approximately 15% of drivers. Robust standard errors. The significance of each coefficient is denoted as follows: $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***. Columns 1, 2, 4, and 5 display coefficients from Poisson regressions (their outcomes are counts), while Column 3 represents an OLS regression. The pure control group sample is excluded because these drivers did not have access to the technology.

The variable # of transactions represents the total number of transactions, while the amount corresponds to the total value of those transactions (in USD). The number of active weeks/days refers to the count of time periods with at least one transaction received.

Baseline P2P Control denotes the number of p2p transactions, before the experiment launch, received that are within the taxi value ranges and outside the taxi values ranges, respectively. This control aims to attenuate variations coming from potential pre-trends in digital usage.

The taxi value ranges are within the CFA 1000-3500 range, USD 1.5-6. These bounds are the 5pp and 95pp of taxi p2b transactions received.

Table B21: Impact of Observability On Driver's Profit

	Driver's Profit	Driver's Profit	Driver's Profit
	(1)	(2)	(3)
Granular Observability	9.865***	10.748***	9.977***
	(3.471)	(3.302)	(3.309)
Coarse Observability	3.789	4.710	4.796
	(4.434)	(4.211)	(4.210)
No Observability	3.074	4.646	4.243
	(5.031)	(4.857)	(4.780)
Observations	519	515	515
Control Mean	440.82	440.82	440.82
% Change Granular Observability	2.24	2.44	2.26
F-test Granular O = No O (p-value)	0.19	0.23	0.26
Baseline Control	NO	YES	YES +

Notes: Outcomes are regressed on the three treatment arms, with the control group as the omitted category. Robust heteroskedasticity-consistent standard errors (SE) are used, and the significance of each coefficient is denoted by asterisks: $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***. The regressions control for strata and batch fixed effects, and the outcome is measured at the mid-term point, 7 to 9 months after the baseline survey. For respondents who refused to answer or indicated uncertainty, their responses are missing. All monetary values are converted to USD (USD 1 = CFA 600). Driver profits are self-reported average earnings over the past 30 days, following De Mel et al. (2009). Specifically, drivers were asked: 'Over the last 30 days, what is the average income that you managed to keep per worked day, after paying all your work-related expenses, including fuel, rental payment, repair, police, contributions, and food?' This figure was then multiplied by the number of days worked in a typical month, with the addition of the monthly salary value paid by the taxi owner, if applicable. Baseline controls in Column (3) include not only average profit at baseline but also a dummy variable for whether the driver had a salary at baseline and the number of days worked per week at baseline.

Table B22: Impact of Observability on Owner's Hiring Decision

	Mid-term (9mo)	Long-term (2y)	Mid-term (9mo)	Long-term (2y)
	(1)	(2)	(3)	(4)
Granular Observability	0.061*	0.021	0.007	0.015
	(0.033)	(0.038)	(0.030)	(0.039)
Granular Observability X Baseline Pairs			0.140*	0.008
			(0.076)	(0.086)
Coarse Observability	0.019	0.007	0.016	-0.005
	(0.032)	(0.037)	(0.030)	(0.039)
Coarse Observability X Baseline Pairs			0.007	0.015
			(0.076)	(0.085)
Observations	817	743	817	743
No Observability Mean	0.154	0.214	0.154	0.214
% Change Granular Observability	40	10	91	4

Notes: Survey data includes all taxi owners, including those driving their taxi alone (i.e., with no employed driver) at baseline. Includes strata fixed effects. Robust standard errors are used. p-values are displayed in brackets, with the significance of each coefficient denoted as follows: $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***. For comparability, the pure control group is excluded from the analysis as they received treatment after 9 months. The mean reported in the bottom row is the mean in the 'No Observability' treatment arm.

Hiring Decision: Whether the owner had effectively hired a driver at the time of the follow-up survey (after 9 months and about 2 years, i.e., mid- and long-term).

The % Change Granular Observability in the last two columns shows the change in the baseline pairs relative to the mean in No Observability.

Single firm owners under the Granular Observability treatment were told the following: 'Your option with this technology is 'Granular Observability.' As a taxi owner, you will have access to the transactions of your drivers, be able to observe the driver's transaction history, and receive an SMS indicating the total daily transactions. If you later hire a driver, this same visibility option will apply.' Owners in the 'No Observability' treatment were explicitly told they would not have access to the driver's transaction data. Appendix C provides the full scripts.

Table B23: Heterogeneous Treatment Effect of Observability on Separation Rate

	Separation Rate							
	Mid-term	Long-term	Mid-term	Long-term	Mid-term	Long-term	Mid-term	Long-term
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$X =$			*Recent Relationship*		*Non-Family Business*		*Non-Risk-Averse Agent*	
X			0.101	0.009	0.030	0.247**	0.126	0.062
			(0.090)	(0.102)	(0.087)	(0.096)	(0.095)	(0.107)
Granular Observability	-0.089	-0.057	-0.068	-0.047	-0.060	0.018	-0.033	-0.081
	(0.061)	(0.069)	(0.093)	(0.107)	(0.081)	(0.093)	(0.099)	(0.111)
Granular Observability $\times X$			-0.031	-0.016	-0.053	-0.185	-0.114	-0.118
			(0.125)	(0.144)	(0.120)	(0.141)	(0.135)	(0.156)
Coarse Observability	0.010	-0.053	0.037	-0.088	-0.018	-0.031	0.160	0.031
	(0.063)	(0.069)	(0.094)	(0.105)	(0.082)	(0.091)	(0.101)	(0.107)
Coarse Observability $\times X$			-0.035	0.071	0.061	-0.078	-0.221	-0.212
			(0.126)	(0.146)	(0.125)	(0.138)	(0.137)	(0.151)
Observations	331	334	331	334	331	334	331	334
No Observability Mean	.34	.63	.34	.63	.34	.63	.34	.63
Mean X			.56	.56	.43	.43	.5	.5

Notes: Business-level regressions estimate the effect of observability on separation rates at mid-term (9 months) and long-term (about two years). The omitted category is the 'No Observability' group, and the analysis includes interaction terms to study heterogeneity. The pure control group is excluded to ensure comparability across columns, as all pairs were treated after the mid-term survey. Robust heteroskedasticity-consistent standard errors (SE) are used, and regressions control for strata fixed effects. The heterogeneity analysis includes the following variables:

- *Recent Relationship:* Relationship lasted less than or equal to 2 years at baseline, before the start of the experiment.
- *Non-Family Business:* Relationship identified as non-family (e.g., close friends, friends, or neutral relations), as opposed to family members.
- *Non-Risk-Averse Agent:* Driver with a coefficient of relative risk aversion below or equal to 1 (CRRA utility function), as elicited in the field using an incentivized game.

Table B24: Reluctant Drivers Tend to Be Lower-Performing and Poorer

	Willing To Adopt (1)	Reluctant Drivers (2)	Difference (3)
Panel A. Driver's Performance			
Performance Index (Z-Score)	0.090	-0.126	-0.216***
	(0.703)	(0.858)	(0.063)
Number of Passengers (3 Days)	43.931	39.128	-4.803***
	(12.951)	(13.587)	(1.102)
Total Collection (3 Days, USD)	153.419	147.571	-5.848**
	(30.001)	(38.349)	(2.785)
Effective Hours Worked (Avg 3 Days)	10.139	9.764	-0.374**
	(2.104)	(2.325)	(0.177)
Total Work Time (End to Start - Avg 3 Days)	11.713	11.182	-0.530**
	(2.461)	(2.749)	(0.208)
Panel B. Relationship and Contracts			
Monthly Upfront Payment 'Salary' (Dummy) W	0.747	0.883	0.136***
	(0.435)	(0.322)	(0.032)
Weekly Rent Value R (USD)	101.375	101.806	0.430
	(16.491)	(11.280)	(1.211)
Separation Rate p	0.351	0.356	0.005
	(0.478)	(0.479)	(0.036)
Owner-Driver Relationship $>$ Two Years	0.425	0.415	-0.010
	(0.495)	(0.493)	(0.038)
Panel C. Driver's Characteristics			
Risk-Averse Agents (CRRA$>$ 1)	0.457	0.590	0.133***
	(0.499)	(0.493)	(0.039)
Often Stressed About Rent Transfers	0.081	0.148	0.067***
	(0.273)	(0.356)	(0.026)
Panel D. Demographics and Poverty			
Education (At Least Primary)	0.301	0.216	-0.085**
	(0.459)	(0.412)	(0.034)
Literacy (Reading and Writing)	0.649	0.587	-0.062*
	(0.478)	(0.493)	(0.038)
Wealth Index (PPI-IPA)	63.494	58.141	-5.353***
	(16.902)	(16.484)	(1.262)
Additional Revenue Source	0.160	0.220	0.059
	(0.368)	(0.415)	(0.038)
Observations	365	340	

Notes: The table summarizes the characteristics of drivers who were willing to adopt the digital payment technology compared to those who did not adopt because they refused to share their owner's contact information. Data, except for demographics and risk-aversion coefficients, were collected during the mid-term survey with drivers from October to December 2022. The first two columns present the mean values of each variable for drivers willing to adopt and non-adopters, with standard deviations in brackets below. To characterize selection, I compare drivers in the impact experiment—who accepted potential observability but whose owners were randomly assigned not to receive it (*No Observability* or *Control*, hence 'willing to adopt') to 'reluctant drivers' (results are robust comparing only drivers randomized to *No Observability*). For reluctant drivers, I exclude the 14% of drivers who later provided owner contacts under *Granular Observability*. The third column shows the estimate from the regression of the variable on being a non-adopter, that is $Y_i = \beta ReluctantDrivers + \epsilon_i$. Heteroskedasticity-robust standard errors and the significance of each coefficient is denoted by asterisks: p $<$ 0.1 *, p $<$ 0.05 **, p$<$ 0.01 ***.
Values are converted to USD (USD 1 = CFA 600). In particular: The Z-Score Productivity Index combines the z-scores of the number of passengers, total collected, and hours worked. Risk-averse agents are drivers with a coefficient of relative risk aversion above 1 (CRRA utility), as computed in an incentivized game. The Wealth Index is defined using the methodology developed by IPA in Senegal to measure household wealth, referred to as the PPI based on the poverty survey (ESPS-II) developed in 2011 in Senegal.

Table B25: Impact of Transaction Observability on Technology Adoption Amid Privacy Concerns

	Technology Adoption (Willing to Share Owner's Information)	
	(1)	(2)
Removing Observability	0.281*** (0.090)	0.073* (0.042)
Observations	87	346
Control Mean	0.119	0.149
Observability Concerns Cited	YES	NO
% Change Removing Observability	236	49

Notes: Survey data were collected from June 15 to July 7, 2022, on drivers who refused to provide their owner's contact information during the listing. Driver-level regressions are performed. The outcome *Adoption* is whether the driver provided the owner's contact information to the surveyor, thus enabling them to adopt the digital payment technology. Heteroskedasticity-robust standard errors are reported, and the significance of each coefficient is denoted by asterisks: $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***. The random assignment of removing the owner's observability of drivers' digital transactions was not stratified.

I run two separate regressions for two distinct groups of individuals: (1) those who cited transaction observability as the reason for not providing their owner's contact information, and (2) those who cited other reasons during the listing process. These regressions aim to demonstrate two key points: (1) the treatment effect is significant for both groups, including those who did not raise observability as an issue, and (2) even for the group where observability was a concern, adoption remains incomplete when observability is removed. This suggests that the treatment effect may be underestimated, as some drivers might be uncertain whether observability will actually be removed.

Table B26: Predicting Driver's Preferences for Observability At Baseline

	Preference For Observability (1)
Panel A. Driver's Characteristics	
Avg Daily Revenue (Past 3 Days, Tens of USD)	0.097***
	(0.020)
Avg Number of Days Worked in a Week	0.044***
	(0.017)
Default at Least Once a Month	-0.054
	(0.034)
Productivity - Value/Hour	0.018
	(0.017)
Hours Worked in a Day	-0.004
	(0.006)
High-Types: Above Median Top Performers Among Drivers	0.099***
	(0.035)
Low-Types: Above Median Low Performers Among Drivers	-0.081**
	(0.035)
Owner-Driver Relationship > Two Years	0.086**
	(0.036)
Panel B. Owner's Beliefs	
Underestimates Number of Days Worked in a Week	0.030
	(0.047)
Underestimates Avg Hours Worked in a Day	0.044
	(0.046)
Underestimates Avg Daily Revenue	0.190***
	(0.051)
Observations	594
Mean of Drivers Preferring Observability	0.231

Notes: Baseline survey data from March to June 2022. Drivers were asked what level of transaction observability they would prefer, if they had to choose (before the treatment was randomized). The outcome variable is defined as 'Preference for Observability' if 'Granular Observability' was the driver's top choice. This variable is regressed on different X baseline variables, each in separate regressions. The regression model is specified as $PreferenceForObservability_i = \alpha + \beta X_i + \epsilon_i$, with significance levels denoted as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Heteroskedasticity-robust standard errors are used. Drivers who refused to respond to this question are excluded. Note that belief questions were only answered by a subset of owners to limit survey length and increase response rates among owners. Questions regarding driver types were framed as follows:
High-Types: Considering a typical 30-day period in the last 3 months, how many days did you collect more than 40,000 FCFA (USD 66)? The median across drivers was *2.0*, and drivers above this median were classified as high types, or top performers.
Low-Types: Considering a typical 30-day period in the last 3 months, how many days did you collect less than 25,000 FCFA (USD 41)? The median across drivers was *8.0*, and drivers above this median were classified as low types, or below-median performers.
An owner is defined as underestimating their driver's work if the value they provide is below the actual value reported by the driver.

Table B27: Framework Inputs

Input	Value (USD)	Source
From the literature:		
Discount factor δ	0.99	Calibrated from relevant setting, Ethiopia, Yesuf and Bluffstone (2019).
Survey Calibration:		
Owner's replacement cost K_p	256.92	Survey question - about 38 days of profit lost.
Share of high-θ μ	0.42	Survey question to taxi owners.
Baseline upfront payment W	9.31	Average salary in control and N-O groups.
Target transfer R^l and R^h	100.03	Median transfer - 91% have this exact target.
High, low output: (Y, X)	(184.18, 94.27)	Avg output above/below median output with high effort in control and N-O groups.
Production function q_0 at $e = 0$	0.28	Likelihood of above median output with low effort.
Production function q_1 at $e = 1$	0.57	Likelihood of above median output with high effort in control and N-O groups.
Owner's Maintenance Cost MC	19.03	Average maintenance cost self-reported by owners at baseline.
Agent's Outside Options $(\bar{u}^l, \bar{u}^h)$	(27.33, 33.33)	Average profit of a small merchant in Dakar from a representative survey I conducted.
Reduced-Form Estimates:		
Technological Gains for Drivers G	1.90	Treatment effect on imputed loss associated with costs of using cash at short-term.
Production function q_2 at $e = 2$	0.74	Likelihood of above median output with high effort in Granular Observability group.
Upfront Payment with Observability $W_{\tilde{e}=2}$	11.03	Unconditional average in the Granular Observability group.
Probability to keep with Observability $p_{\tilde{e}=2}$	0.97	Share remaining together under Granular Observability group.

Notes: All parameters, except the discount rate, are calibrated from the survey data. Specifically, to calibrate q_1, I use the proportion of times drivers achieved upper median output when working more than the median hours in the 'No Observability' (N-O) and control groups at mid-term. Similarly, to calibrate q_0, I use the proportion of times drivers in the entire sample achieved upper median output when working less than the 25^{th} percentile of hours worked. High and low outputs (Y and X) were determined based on drivers' revenue minus expenses, which include weekly food and beverages, all measured in the survey. I define high and low outputs, Y and X, as the average output above or below the median output when working more than the median hours, respectively.
I infer the proportion of high-type drivers (μ) by directly asking owners for their estimates and averaging their responses. The exact survey question was: 'We are trying to understand your perception of drivers. There are good and bad taxi drivers in terms of work. Imagine that 10 drivers present themselves to you at random. Out of 10 drivers, how many do you consider to be 'good' drivers?' The owner's replacement cost K_p was calibrated using the response to the question: 'If you were to lose your current taxi driver, how long do you estimate it would take for you to find another similar taxi driver?'
To estimate outside options, I conducted a representative survey of merchants—common outside options for drivers—in September 2022. From this survey, I calculated the median profit for merchants with 0 or 1 employee. To distinguish between high-type and low-type outside options, I used the difference in poverty likelihood, based on the 200% National Poverty Line as computed by IPA. This measure allows me to estimate variation in outside options: I assume high-type drivers can earn the median profit level observed among merchants, while low-type drivers can earn a proportion of this median, adjusted to reflect the wealth index gap between the two empirical samples.
Then, I use reduced-form estimates to obtain the following parameters: the private technology gains for drivers (G), which are assumed to equal the total transaction cost of using cash in the control group at mid-term, reflecting anticipated technological gains as described in Section 4.1; the production function for drivers under 'Granular Observability'. Specifically, to calibrate q_2, I use the proportion of times drivers achieved upper median output when working more than the median hours in the 'Granular Observability' treatment group at mid-term; and the contract characteristics under 'Granular Observability', particularly the observed salary $W_{\tilde{e}=2}$ and the probability of retaining the driver $p_{\tilde{e}=2}$.

Table B28: Threshold Values of the Discount Rate in the Structural Estimation

Discount Rate and Thresholds	Value (1)
$\underline{\delta}$	0.23
δ	0.99
$\bar{\delta}$	1.04
$\bar{\delta}^{tech}$	1.03
$\bar{\delta}^{FI}$	1.03

Notes: This table presents the structural threshold values for the discount rate from the no-deviation conditions, under which the various results hold in the structural estimation.

C Treatment Descriptions: Owner Access to Driver Transactions

This appendix reproduces the scripts (translated from French) that were read to owners and drivers. The first script was used during the listing survey, and the second at the end of the baseline survey to explain treatment assignment.

Listing Survey Script

"We are studying the technology *"Pay with Wave"* for taxis, which enables drivers to accept secure digital payments using a QR code displayed inside the vehicle. Passengers can pay via Wave instead of using cash. The driver receives the money in a digital account, paying a 1% fee only once total collections exceed 50,000 CFA. Drivers can withdraw funds at any Wave mobile money agent. In addition, we will enable free money transfers between drivers and vehicle owners.

To proceed with enrollment during the pilot phase, we require the contact information of your vehicle's owner. This is needed to inform both drivers and owners about the pilot phase; to set up free Wave transfers between owners and drivers; to assess which documents are available through the owner to determine the best type of Wave account (e.g., transaction limits).

If you choose not to provide your owner's contact information, you will not have access to the product at this time."

Do you agree to provide the name and contact of your vehicle's owner? If not, why not? Possible reasons included: not wanting the owner to have visibility into transactions; the owner being unavailable or uninterested; not wanting the owner to know about digital payment use; needing to speak with the owner first; lack of trust in the research team; or other reasons (including refusal to answer).

Baseline Survey and Treatment Assignment Script

At the end of the baseline survey, drivers were reminded of the following:

We are testing the *Pay with Wave* technology for taxis, which allows drivers to securely accept digital payments through a QR code. Passengers can pay with Wave instead of cash. The driver receives funds in a digital account, paying a 1% fee only once collections exceed 50,000 CFA, and can withdraw money at any Wave agent.

Benefits include secure payments that cannot be canceled, free transfers between drivers and owners, no need for small change, improved passenger satisfaction, easier savings, digital transaction records for bookkeeping, and reduced risk of theft.

To better understand what features work best for both owners and drivers, you were randomly assigned to one of the following options for this pilot phase:

Treatment: No Observability Your assigned option is the **No Observability** condition. Under this setup, the vehicle owner does not have access to any of the driver's transaction information.

Treatment: Coarse Observability Your assigned option is the **Coarse Observability** condition. Each evening at midnight, the vehicle owner receives an SMS indicating the amount collected that day—up to a maximum of 5,000 CFA. For example: If the driver collects 12,000 CFA, the SMS will report "at least 5,000 CFA." If the driver collects only 3,000 CFA, the SMS will report "3,000 CFA."

Example SMS: Hello! Your taxi driver Mr. NAME collected at least 5,000 CFA using Wave Business on Sunday, 12/03/2022.

This option allows the owner to know when a driver has had a low-income day, but does not reveal the full amount collected or individual transaction details.

Note for owners: If you later hire a driver, this same visibility option will apply.

Treatment: Granular Observability Your assigned option is the **Granular Observability** condition. Under this setup, the vehicle owner has full visibility into the driver's transactions and receives a daily SMS with the total amount collected each day.

Note for owners: If you later hire a driver, this same level of visibility will also apply.

D Theoretical Framework: Proofs and Derivations

D.1 Baseline Contract and Agents' Informational Rents

The principal maximizes expected transfers and the discounted value of the relationship:

$$V^{\theta} = \max_{t,p,e} \mathbb{E}[t(\tilde{y}) + \delta[p(\tilde{y})V^{\theta} + (1 - p(\tilde{y}))(-K_p + \mu V^h + (1-\mu)V^l)]|e] \tag{D1}$$

This optimization is subject to the following constraints:

$$\begin{cases} \mathbb{E}[(y(e) - t(\tilde{y})) - \phi^{\theta}(e) + \delta U^{\theta} - \delta K_a(1 - p(\tilde{y}))|e] \geq \max\{-\delta K_a + \delta U^{\theta}; \bar{u}\} & \text{Participation Constraint } (IR) \\ e \in \arg\max_{\tilde{e} \in \{0,1,2\}} \mathbb{E}[y(e) - t(\tilde{y}) + \delta U^{\theta} - \delta K_a(1 - p(\tilde{y}))|\tilde{e}] - \phi^{\theta}(\tilde{e}) & \text{Incentive Compatibility } (IC) \\ t(\tilde{y}) \leq \tilde{y} \leq y(e) & \text{Limited Liability } (LL) \\ Y - t(Y) + \delta(U^{\theta} - K_a(1 - p(Y))) \geq Y - t(X) + \delta(U^{\theta} - K_a(1 - p(X))) & \text{Truth-Telling } (TT) \\ y(e) - t(\tilde{y}) + \delta(U^{\theta} - (1 - p(\tilde{y}))K_a) \geq y(e) + \delta(U^{\theta} - K_a) & \text{Dynamic Enforceability } (DE) \end{cases}$$

Result 1. (Baseline Contract Without Digital Payments) *Under Assumptions 1–4, $\exists\ \underline{K}_p < \bar{K}_p$, $\underline{\delta} < \bar{\delta}$, s.t. when $K_p \in (\underline{K}_p, \bar{K}_p)$ and $\delta \in (\underline{\delta}, \bar{\delta})$, the principal's best type-dependent stationary contract is:*

$$\bar{t}^{\theta} = \begin{pmatrix} t(Y) = R^{\theta} \\ t(X) = X \end{pmatrix} \quad \text{and} \quad \bar{p}^{\theta} = \begin{pmatrix} p(Y) = 1 \\ p(X) = \bar{p}^{\theta} \end{pmatrix}$$

where the continuation probability for a low-output outcome, $\bar{p}^{\theta}$, is given by:

$$\bar{p}^{\theta} = \min \begin{Bmatrix} \bar{p}^{\theta,IC} = 1 + \dfrac{\bar{u}}{\delta K_a} - \dfrac{q_0 \phi^{\theta}(1)}{\delta K_a(q_1 - q_0)}, \\ \bar{p}^{\theta,TT} = 1 - \dfrac{1}{\delta K_a}\left[q_1(Y - X) - \phi^{\theta}(1) - \bar{u}\right] \end{Bmatrix}$$

and the rental transfer for a high-output outcome, R^{θ}, is given by:

$$R^{\theta} = \min \begin{Bmatrix} R^{\theta,IC} = Y - \bar{u} - \dfrac{(1 - q_0)\phi^{\theta}(1)}{q_1 - q_0}, \\ R^{\theta,TT} = q_1(Y - X) + X - \phi^{\theta}(1) - \bar{u} \end{Bmatrix}$$

with $\bar{p}^h > \bar{p}^l$, $R^h > R^l$. The agent induced effort is $e^l = e^h = 1 < 2$.

The contract values $(R^{\theta}, \bar{p}^{\theta})$ depend on which constraint binds, (IC) or (TT).

Proof. The proof has two parts: (i) derive optimal terms of the best stationary contract depending on which constraint binds; (ii) derive conditions under which such a contract is principal optimal.

Best stationary contract terms. First, I demonstrate that $p(Y) = 1$, by contradiction. Suppose $p(Y) < 1$. The principal's continuation value increases in $p(Y)$; thus raising it to 1 strictly improves the objective. Moreover, higher $p(Y)$ strengthens the agent's incentives for effort and truth-telling. Since firing costs $K_p > 0$, it is never optimal for the principal to choose $p(Y) < 1$. Contradiction. Hence $p(Y) = 1$.

Second, I show that $t(X) = X$. Suppose, for contradiction, that $t(X) < X$. Then, (TT) implies $p(X) \leq 1 - \frac{t(Y)-t(X)}{\delta K_a}$. Because punishment $p(X)$ is costly ($K_p > 0$) for the principal, lowering $t(X)$ unnecessarily imposes additional constraints on $p(X)$ and reduces transfers without improving incentives. Hence, the principal is strictly better off with $t(X) = X$.[51] Under (LL) we must have $t(X) \leq X$.

Define $t(Y) = R^\theta$ and $p(X) = \bar{p}^\theta$, with $R^\theta \in [0, Y]$, $\bar{p}^\theta \in [0, 1]$. Both parties observe a public randomization device at the end of the stage game (Mailath and Samuelson, 2006), which the principal can use to implement $\bar{p}^\theta$ when $y = X$. Deviating from this device would unravel incentives, so the principal has no incentive to renege.

Upon low output, the principal must terminate on path, with $\bar{p}^\theta < 1$. This inefficient punishment arises from (LL): following X, the principal cannot extract additional transfers, so incentives must be provided via termination (Fuchs, 2007). Thus (LL) implies $t(X) = X$, as the principal cannot demand more than what the agent collects.

To incentivize $e = 1$, the principal chooses $\bar{p}$ to satisfy (IC):

$$\frac{q_1(Y - t(Y)) + \delta(-K_a(1-\bar{p})(1-q_1)) - \phi^\theta(1)}{1-\delta} \geq \frac{q_0(Y - t(Y)) + \delta(-K_a(1-\bar{p})(1-q_0))}{1-\delta} \iff$$
$$\bar{p} \leq 1 + \frac{Y - t(Y)}{\delta K_a} - \frac{\phi^\theta(1)}{\delta K_a(q_1 - q_0)} \equiv \bar{p}^\theta \tag{D2}$$

From (TT), we get: $\bar{p}^\theta \leq 1 - \dfrac{R^\theta - X}{\delta K_a} \;\; \forall \;\; \theta$

The driver's dynamic constraint does not bind in a low-output period if $\bar{p}^\theta \geq \dfrac{X}{\delta K_a} \;\; \forall \;\; \theta$.
To summarize:

$$\begin{cases} \bar{p}^\theta & \leq 1 + \dfrac{Y - R^\theta}{\delta K_a} - \dfrac{\phi^l(1)}{\delta K_a(q_1 - q_0)} \quad \text{implied from } (IC^\theta) \\ \bar{p}^\theta & \leq 1 - \dfrac{R^\theta - X}{\delta K_a} \quad \text{implied from } (TT) \\ \bar{p}^\theta & \geq \dfrac{X}{\delta K_a} \quad \text{implied from } (DE) \end{cases} \tag{D3}$$

[51] One may allow a minimum payment in low-output periods, so that the agent receives more than 0.

Either (IC) or (TT) binds, since termination is costly but required to discipline the agent.

Finally, consider the principal's choice of $t(Y)$. The derivative of the value function with respect to $t(Y)$ is

$$\frac{\partial V^\theta}{\partial t(Y)} > 0 \iff K_p < \frac{q_1 K_a}{(1-q_1)} \equiv \bar{K_p} \tag{D4}$$

$$\text{with } V^\theta = \frac{q_1 t(Y) + (1-q_1)X - \delta[K_p(1-\bar{p}^\theta)(1-q_1)]}{1-\delta}$$

Intuitively, this suggests that the replacement cost for the principal should be sufficiently low, ensuring that termination is an effective tool for the principal. Otherwise, the principal never fires the agent ($\bar{p} = 1$), and the agent always reports low-output—an equilibrium I rule out.

Case 1. Incentive compatibility constraints binds. When (IC) binds, then:

$$\bar{p}^{\theta,IC} = 1 + \frac{Y - R^{\theta,IC}}{\delta K_a} - \frac{\phi^\theta(1)}{\delta K_a(q_1 - q_0)}$$

(LL) implies that the participation constraint is slack and agents have an informational rent at baseline. In particular, the principal sets R^θ such that the agent's present discounted utility binds:

$$\frac{q_1(Y - t(Y)) - \phi^\theta(1) - \delta(K_a(1-\bar{p}^\theta)(1-q_1))}{1-\delta} = \frac{\bar{u}}{1-\delta}$$

From (IC), I replace $\bar{p}^{\theta,IC} = 1 + \dfrac{Y - t(Y)}{\delta K_a} - \dfrac{\phi^\theta(1)}{\delta K_a(q_1 - q_0)}$ to obtain:

$$\begin{cases} t(Y) = R^{\theta,IC} = Y - \bar{u} - \dfrac{(1-q_0)\phi^\theta(1)}{q_1 - q_0} \\ \bar{p}^{\theta,IC} = 1 + \dfrac{\bar{u}}{\delta K_a} - \dfrac{q_0 \phi^\theta(1)}{\delta K_a(q_1 - q_0)} \end{cases}$$

Informational Rent. Hence, in each period, the agent captures some surplus above their outside option $\bar{u}$ of walking away and thus receive the following *informational rent*:

$$q_1(Y - R^{\theta,IC}) - \phi^\theta(1) - \bar{u} = \frac{\phi^\theta(1) q_0 (1-q_1)}{q_1 - q_0} - \bar{u}(1-q_1)$$

In addition, the following condition needs to hold for the principal to be better off incentivizing $e = 1$ instead of $e = 2$ for the two agents. Specifically, to incentivize $e = 2$, the principal would require $\bar{p}_2^{\theta,IC} = 1 + \dfrac{Y - R^{\theta,IC}}{\delta K_a} - \dfrac{\phi^\theta(2) - \phi^\theta(1)}{\delta K_a(q_2 - q_1)}$, with $\bar{p}_2^{\theta,IC} < \bar{p}^{\theta,IC} \quad \forall \quad \theta$.

And using the same reasoning as before, we would get:

$$\begin{cases} R^{\theta,IC}_{e=2} = Y - \bar{u} - \dfrac{\phi^\theta(2)(1-q_1) - \phi^\theta(1)(1-q_2)}{q_2 - q_1} \\ \bar{p}^{\theta,IC}_{e=2} = 1 + \dfrac{\bar{u}}{\delta K_a} - \dfrac{q_1\phi^\theta(2) - q_2\phi^\theta(1)}{\delta K_a(q_2 - q_1)} \end{cases}$$

The principal is better off incentivizing $e = 1$ if:

$$\begin{aligned} V^\theta_{e=1}(R^{\theta,IC}_{e=1}, \bar{p}^{\theta,IC}_{e=1}) > V^\theta_{e=2}(R^{\theta,IC}_{e=2}, \bar{p}^{\theta,IC}_{e=2}) \iff \\ K_p > K_p^{\theta,IC} \end{aligned} \tag{D5}$$

Intuitively, if replacement costs are high, disciplining $e = 2$ is too costly, so the principal optimally induces $e = 1$.

Case 2. Truth-telling constraints binds. When (TT) binds, then:

$$\bar{p}^{\theta,TT} = 1 - \frac{R^{\theta,TT} - X}{\delta K_a} \tag{D6}$$

With the same reasoning as before, we can recover R^{TT} such that:

$$\begin{cases} R^{\theta,TT} = q_1(Y - X) + X - \phi^\theta(1) - \bar{u} \\ \bar{p}^{\theta,TT} = 1 - \dfrac{1}{\delta K_a}[q_1(Y-X) - \phi^\theta(1) - \bar{u}] \end{cases} \tag{D7}$$

The agent gets the following *informational rent*:

$$q_1(Y - R^{\theta,TT}) - \phi^\theta(1) - \bar{u} = (1 - q_1)\big[q_1(Y - X) - \phi^\theta(1) - \bar{u}\big].$$

As before, $e = 1$ is optimal if $V^\theta_{e=1}(R^{\theta,TT}, \bar{p}^{\theta,TT}) > V^\theta_{e=2}(R^{\theta,IC}_{e=2}, \bar{p}^{\theta,IC}_{e=2}) \iff K_p > K_p^{\theta,TT}$

Thus $\underline{K}_p = \max(K_p^{h,IC}, K_p^{h,TT})$, ensuring that the principal is better off inducing $e = 1$ for both types. Intuitively, when K_p is high, the cost of disciplining $e = 2$ dominates.
Finally, (TT) binds whenever

$$\begin{aligned} \bar{p}^{\theta,TT} < \bar{p}^{\theta,IC} \iff \\ (Y - X) - \frac{\phi^\theta(1)}{q_1 - q_0} > 0 \end{aligned} \tag{D8}$$

This condition is type-dependent: high-type agents are easier to incentivize (lower punishment needed), such that (TT) may bind, whereas low-type agents are more likely constrained by (IC).

Conditions for the discount factor δ. I check whether (DE) holds, such that both agents value the future enough to come back to the owner at the end of a low-output period:

$$\delta > \max\Big(\frac{(q_1 - q_0)(X - \bar{u}) + q_0\phi^l(1)}{(q_1 - q_0)K_a}, \frac{X + q_1(Y - X) - \phi^h(1) - \bar{u}}{K_a}\Big) \equiv \underline{\delta} \tag{D9}$$

I also derive an upper bound $\bar{\delta}$ so Result 1 holds only if the principal is not too patient, i.e., $\underline{\delta} < \delta < \bar{\delta}$, or the share of low-types is sufficiently high; this follows from a "no-deviation" condition:

- At equilibrium: $\mu < \dfrac{V_l(\frac{1}{\delta}-1)+K_p}{V_h - V_l}$ or $\delta < \dfrac{V_l}{-K_p + \mu V_h + (1-\mu)V_l} \equiv \bar{\delta}$ with V_h and V_l equilibrium values. That is, it must be that the share of high-type agents is sufficiently low, such that it is not profitable for the principal to fire low-type in the hope of being rematched with a high-type agent. This condition is then verified in the structural estimation.

This completes the proof of Result 1. □

D.2 Full Information Benchmark and Fixed Wage Contract

Lemma 1. (Full Information Benchmark) *Under Assumptions 2-4 and $\delta < \bar{\delta}^{FI}$, the principal's best stationary equilibrium is a wage contract that guarantees re-hiring when the agent exerts the optimal effort level, with no termination occurring on the equilibrium path.*

Wages are set such that both high- and low-type agents' participation constraints bind.[52] In this benchmark, the owner must sufficiently compensate the agent for his optimal level of effort so that the agent is indifferent between working and his outside option. The first-best level of effort will be $e = 1$ for agent θ if:

$$
\begin{aligned}
V^{FI}_{e=1} > V^{FI}_{e=2} &\iff q_1 Y + (1-q_1)X - \Big(\phi^{\theta}(1) + \bar{u}\Big) > q_2 Y + (1-q_2)X - \Big(\phi^{\theta}(2) + \bar{u}\Big) \\
&\iff \phi^{\theta}(2) > (q_2 - q_1)(Y - X) + \phi^{\theta}(1)
\end{aligned}
$$

This implies re-hiring with $\bar{p}^{\theta} = 1$, since termination is costly. Various compensation schemes could implement this equilibrium. Here, I provide the formal argument showing why a fixed wage is weakly preferable from the principal's perspective compared to a bonus payment.

Proof. Suppose instead the principal uses a bonus scheme: $\bar{W}$ in high-output periods and $\underline{W}$ in low-output periods, ensuring the agent exerts $e = 2$ and is always re-hired. The principal's value is:

$$V^{\theta} = \max_{W}(qY + (1-q)X) - W + \delta V^{\theta} \tag{D10}$$

Given the relational nature of the contract, the principal must not renege on the promised bonus in high-output periods. The non-reneging constraint can be expressed as:

[52] Similar full-information benchmarks are discussed in a range of papers with different environments, such as Shetty (1988) with limited liability; in *Proposition 5* of Shavell (1979) with risk averse agents; and in Ghatak and Pandey (2000) with joint moral hazard in effort and risk.

$$Y - \bar{W} + \delta V > Y - \underline{W} + \delta(V - h),$$
$$\delta > \frac{\bar{W} - \underline{W}}{K_p}.$$

This condition indicates that the discount factor δ must be sufficiently large to deter the principal from paying a lower bonus $\underline{W}$ in high-output periods (instead of the promised large bonus). By contrast, this constraint does not apply if a fixed wage is paid directly to the agent before knowing the output, effectively eliminating the incentive for the principal to renege. Hence a fixed wage is feasible over a wider range of δ and dominates a bonus scheme.

Second, because the principal can be re-matched to agents of either type, I verify the condition under which the principal is better off keeping the low-type drivers in equilibrium, rather than firing them in the hope of being re-match with a high-type (probability μ). Let W^l and W^h denote the fixed wage of the low- and high-type agents. In the full-information benchmark, when $e = 1$ is the optimal level of effort, $W_{e=1} = \phi^\theta(1) + \bar{u}$, since the agent are paid at their outside option when leaving the taxi industry. For both types to remain in the market, the following "no-deviation" condition must hold, meaning the share of low-types $(1 - \mu)$ should be high enough.

$$\delta < \frac{V_l^{FI}}{-K_p + \mu V_h^{FI} + (1 - \mu) V_l^{FI}} \equiv \bar{\delta}^{FI}$$

□

D.3 Observability of Agent's Output Only

Consider a scenario where only the employee's output y is verifiable, not his effort e.

Lemma 2. (Information on Output) *Under Assumptions 1–4, complete output information for the principal relaxes the truth-telling constraint* (TT)*. When* (TT) *binds, this information leads to an increase in the continuation probability* $\bar{p}$ *and* R *to satisfy the incentive compatibility constraint* (IC)*.*

This lemma shows how output observability affects the principal–agent relationship. It allows the principal to raise p and R without learning effort, since verifying output removes the need to penalize agents for truthful reporting. In other words, once output is observable, the principal no longer relies on punishment to enforce truth-telling.

D.4 Stage 2: Imperfect Information on Agent's Effort and Output

Digital payments provide the principal with *imperfect* information on their agent for at least two reasons empirically. First, cash is still being used, meaning the digitalization of payments is not complete. Second, while digital transactions include timestamps and transaction values, they only partially reflect the agent's effort and output. This section shows that such imperfect information leads to similar predictions to the benchmarks detailed above, using a similar idea as the informativeness principle (Holmström, 1979). Let's define s the high-effort signal and $\kappa = P(s|e=2)$ and assume $P(s|e=1) = P(s|e=0) = 0$.

Result 2. (Imperfect Information on Effort) *Under Assumptions 1–4, when (IC) binds, for $K_p \in (\underline{K}_p, \bar{K}_p)$ and $\delta \in (\underline{\delta}, \bar{\delta})$, $\kappa > \bar{\kappa}$ for $\bar{\kappa} < 1$, and $\phi^\theta(2) < \tilde{\phi}\ \forall\ \theta$, the principal's best type-dependent stationary contract is:*

$$\bar{t}^\theta = \begin{pmatrix} t(Y) = R^\theta - W^\theta_{\tilde{e}=2} \\ t(X) = X - W^\theta_{\tilde{e}=2} \end{pmatrix} \quad \textit{and}$$

$$p(\tilde{y}, s) = \begin{cases} 1 & \textit{if } \tilde{y} = Y, \\ \bar{p}_{TT} & \textit{if } \tilde{y} = X \textit{ and } s \textit{ is observed} \\ \bar{p}^{\theta\prime} < \bar{p}^\theta & \textit{if } \tilde{y} = X \textit{ and } s \textit{ is not observed} \end{cases}$$

The agent θ induced effort is $e^\theta = 2$.

Each period, the principal uses $W^\theta_{\tilde{e}=2}$ to provide incentives for adoption . To mitigate concerns about reneging in this relational contract framework, the principal incentivizes the agent to adopt the technology by offering an upfront payment, $W^\theta_{\tilde{e}=2}$, rather than reducing the target rental payment (ex-post). More formally, the upfront payment is preferred over a promised reduction in rent because otherwise the principal would have incentives to renege on his promise. A similar logic applies when the principal obtains *imperfect* information on output, relaxing the truth-telling constraint (TT) when the latter is binding at baseline. Consequently, imperfect information can be advantageously incorporated into the contract under minimal conditions.

The proof has two steps: first, showing why the principal prefers to pay *upfront* in a relational contract; second, rationalizing the new contract structure and induced effort.

Proof. Consider the contract structure where the transfer $t(Y)$ is defined as

$$t(Y) = R^\theta - W^\theta_{\tilde{e}=2},$$

For contradiction, assume $W^\theta_{\tilde{e}=2}$ represents the *promised* reduction in rental transfer at the end of the period, based on the high-effort signal.

At the end of the work period, if high output Y is achieved, the principal has an incentive to deviate by increasing the transfer by ϵ:

$$t(Y) = R^\theta - W^\theta_{\tilde{e}=2} + \epsilon,$$

where $\epsilon > 0$ is the marginal increase in the transfer. The agent would still be willing to accept this adjusted transfer, as it remains above their outside option. A similar reneging concern exists with low output X, allowing the principal to extract

$$t(X) = X - W^{\theta}_{\tilde{e}=2} + \epsilon$$

without breaking the agent's participation constraint.

This leads to a contradiction, as there exists a profitable one-shot deviation for the principal. To prevent such ex-post reneging concerns, the principal would prefer to provide $W^{\theta}_{\tilde{e}=2}$ upfront, eliminating the temptation to deviate in this relational contract framework. $\square$

The second part of the proof is to rationalize the new contract structure and resulting incentives for effort on the agent side.

Proof. Given that $\kappa = P(s|e = 2)$ and $P(s|e = 1) = P(s|e = 0) = 0$, with s the high-effort signal. The higher κ, the more valuable the information. Consider a scenario where the incentive compatibility constraint (IC^{θ}) binds at baseline for agent θ, making imperfect information on effort valuable to relax this constraint. Specifically, the probability of retaining the agent becomes $\bar{p}_{TT}$ in a low-output period if the owner observes high effort $e = 2$, and $\bar{p}^{\theta\prime} < \bar{p}^{\theta}$ otherwise. This can be expressed as follows:

$$p(\tilde{y}, s) = \begin{cases} 1 & \text{if } \tilde{y} = Y, \\ \bar{p}_{TT} & \text{if } \tilde{y} = X \text{ and } s \text{ is observed} \\ \bar{p}^{\theta\prime} < \bar{p}^{\theta} & \text{if } \tilde{y} = X \text{ and } s \text{ is not observed} \end{cases}$$

For the sake of the argument, let's assume $\bar{p}_{TT} = 1$ meaning the truth-telling constraint does not bind in this setting. This assumption will simplify the following derivations.

The upfront payment or 'salary' must be sufficiently high to ensure that:

$$U^{\theta}_2 + W^{\theta}_{\tilde{e}=2} \geq U^{\theta}_{notech} \tag{D11}$$

Note that $W^{\theta}_{\tilde{e}=2} > 0 \iff$
$\phi^{\theta}(2) > (q_2 - q_1)(Y - R^{\theta}) + \phi^{\theta}(1) - \delta K_a((1 - \bar{p}^{\theta})(1 - q_1) - (1 - \kappa)(1 - \bar{p}^{\theta\prime})(1 - q_2))$

The IC^{θ} constraint requires that the agent θ must then be incentivized to exert high effort $(e = 2)$. When the principal does not observe the signal, with probability $1 - \kappa$, he must terminate the relationship with some sufficiently low probability $\bar{p}^{\theta\prime}$ to incentivize effort. Intuitively, the high-effort signal enables the principal to incentivize effort at a lower cost (punishment) when the signal is valuable enough.

Mathematically, the IC^{θ} constraint can be written as:

$$
\begin{aligned}
&\frac{W^{\theta}_{\tilde{e}=2} + q_2(Y - R^{\theta}) - \phi^{\theta}(2) - \delta K_a(1-q_2)(1-\kappa)(1-\bar{p}^{\theta\prime})}{1-\delta} \quad \text{(Agent exerts } e=2\text{)}\\
&> \frac{W^{\theta}_{\tilde{e}=2} + q_1(Y - R^{\theta}) - \phi^{\theta}(1) - \delta K_a(1-q_1)(1-\bar{p}^{\theta\prime})}{1-\delta} \quad \text{(Agent exerts } e=1\text{)}
\end{aligned}
$$

$$
\iff \bar{p}^{\theta\prime} \leq 1 - \frac{\phi^{\theta}(2) - \phi^{\theta}(1) - (q_2 - q_1)(Y - R^{\theta})}{((1-q_1) - (1-q_2)(1-\kappa))\delta K_a} \tag{D12}
$$

Now I turn to the equilibrium and examine the conditions under which the owner is better off with this new contract.

$$
V^{\theta}_{e=2} = \quad q_2 R^{\theta} + (1-q_2)X - W^{\theta}_{\tilde{e}=2} + \delta(V - K_p(1-\kappa)(1-q_2)(1-\bar{p}^{\theta\prime}))
$$

The principal will be better off offering this upfront payment if $V^{\theta}_{e=2} > V^{\theta}_{baseline}$. Let's have $\bar{p}^{\theta\prime}$ such that the IC binds. One can show that

$$
\frac{\partial \bar{p}^{\theta\prime}}{\partial \kappa} > 0 \tag{D13}
$$

And since we know that:

$$
\frac{\partial V}{\partial \bar{p}^{\theta\prime}} > 0, \tag{D14}
$$

There exists a threshold $\exists\ \kappa \leq \bar{\kappa}$ such that the principal would not want to incentivize the agent and would instead retain the initial contract upon adoption. In other words, the precision of information κ must be sufficiently high to induce a profitable change in the contract.

Using the same reasoning,

$$
\frac{\partial \bar{p}^{\theta\prime}}{\partial \phi^{\theta}(2)} > 0 \tag{D15}
$$

There exists a threshold $\exists\ \phi^{\theta}(2) < \tilde{\phi}\ \forall\ \theta$ such that the principal would want to incentivize both agent's types to adopt. In other words, the disutility of high-effort $\phi^{\theta}(2)$ must be sufficiently low to allow a profitable change in the type-dependent contract.

□

Then, I turn to examine how imperfect information on output impacts the principal-agent relationship. Empirically, the digital payment technology provides only a low-output signal to business owners because only a fraction of driver's transactions were digital during the experiment. Therefore, I define s' the low-output signal and $\xi = P(s'|y = X)$ and assume that $P(s'|y = Y) = 0$.

Lemma 3. (Imperfect Information on Output) *Under Assumptions 1–4, when* (TT) *binds, for* $K_p \in (\underline{K}_p, \bar{K}_p)$ *and* $\delta \in (\underline{\delta}, \bar{\delta})$, *and* $\xi > \bar{\xi}$ *for* $\bar{\xi} < 1$, *the principal's best type-dependent stationary contract is:*

$$\bar{t}^\theta = \begin{pmatrix} t(Y) = R^\theta \\ t(X) = X \end{pmatrix} \quad \textit{and}$$

$$p(\tilde{y}, s') = \begin{cases} 1 & \textit{if } \tilde{y} = Y, \\ \bar{p}_{IC} & \textit{if } \tilde{y} = X \textit{ and } s' \textit{ is observed} \\ \bar{p}^\theta & \textit{if } \tilde{y} = X \textit{ and } s' \textit{ is not observed} \end{cases}$$

The imperfect information on output enables the principal to relax the truth-telling constraint (TT) and not punish the agent for misreporting when the signal indicates low output. As before, since $\frac{\partial V}{\partial \bar{p}^\theta} > 0$, the principal is better off using the low-output signal when the information is accurate enough—that is, when the probability of low output given the signal is sufficiently high, $\xi > \bar{\xi}$. Specifically, the principal benefits from increasing the continuation probability following a low-output signal. The target rental transfer R^θ remains unchanged, because the absence of a signal does not perfectly reveal high output: raising R^θ would encourage agents to misreport revenues.

The experiment explicitly tests this mechanism under the *Coarse Observability* treatment. In this arm, owners observe only whether daily digital collections are below a threshold, meaning the agent can signal only *low output* through the app. This is expected to reduce the firing punishment for misreporting revenue when the driver fails to pay the rent, while still preserving informational rents by not fully revealing effort or high-output realizations.

D.5 Agent's Manipulation of Effort and Output Signals

Lemma 4. (No Manipulation) *Under Assumptions 1–4, the agent has no incentive to manipulate the imperfect information on effort or output provided to owners since inaccurate reporting or distorted information would limit the overall impact of the technology.*

Proof. The proof proceeds separately for the information on effort and output. In this context, passengers typically decide between cash and digital payments, so drivers cannot force more digital use but could attempt to reduce it. I assume that drivers can only lower the share of digital transactions, consistent with the empirical setting: there is no evidence of drivers enforcing digital payments during mystery passenger audits, and passengers always expect the option to pay with cash.

1. *Manipulating the effort information.* Suppose, for contradiction, that the agent reduces digital usage to signal low effort to the principal. By Result 2, this lowers the likelihood of observing the high-effort signal s and increases the probability of termination. Hence, the agent's

continuation value falls, and downward manipulation is strictly dominated. Upward manipulation is not feasible, as the agent cannot credibly increase the probability of generating s beyond what effort already determines.

2. *Manipulate the output information.* Consider the high-output state $y = Y$ (since limited liability rules out manipulation at $y = X$). The agent could misreport by triggering the low-output signal s', thereby retaining more surplus today but facing punishment with probability $1 - \bar{p}_{IC}$ (Lemma 3). The agent will not manipulate if:

$$\underbrace{Y - X + \delta[U - K_a(1 - \bar{p}_{IC})]}_{\text{Agent's value from manipulation}} \leq \underbrace{Y - R^{\theta} + \delta U}_{\text{Agent's value without manipulation}} \tag{D16}$$
$$R^{\theta} - X \leq \delta K_a(1 - \bar{p}_{IC})$$

Condition (D16) requires that the principal set termination probabilities sufficiently high to deter manipulation. If anything, concerns about manipulation reduce the positive effect of output signals on retention rather than reversing it.

□

The experiment is designed to empirically test manipulation by comparing drivers' digital usage under *Coarse Observability* and *No Observability*. If drivers engaged in manipulation, we would expect digital usage to fall under *Coarse Observability*. Instead, administrative data show slightly higher usage in this arm. Additionally, mystery passenger audits confirm this absence of manipulation, as drivers in both treatment groups respond similarly when passengers request to pay digitally (see Figure A3). The absence of manipulation may be explained by the competitive pressure drivers face to secure passengers, which discourages them to manipulate digital payment usage. Drivers may perceive the short-term loss of passengers—resulting from pressuring or discouraging them to pay digitally—as too costly to justify such actions.

D.6 Stage 1: Differential Digital Technology Adoption

After having established that the technology, once adopted, may increase owner's utility from the (imperfect) information on either the agent's output or effort, this section shows how agents' type influences who adopts the technology in the first place, see Result 3. This demonstration provides bounds for the disutility of work of the low-type agent. Augmenting the framework with direct gains for agents to adopt the technology (e.g., cash-handling costs) is straightforward and done in the structural estimation section.

Result 3. (Differential Adoption) *Under Assumptions 1–4, $\exists\, \underline{K}_p < \bar{K_p}, \underline{\delta} < \bar{\delta}^{tech}$ and $\bar{\phi^h} < \bar{\phi^l}$ s.t. if $K_p \in (\underline{K}_p, \bar{K_p})$ and $\delta \in (\underline{\delta}, \bar{\delta}^{tech})$, $\phi^l(2) > \bar{\phi^l}$, and $\phi^h(2) < \bar{\phi^h}$, then only high-ability agents adopt the technology, while low-ability agents opt not to adopt it.*

Proof. The proof proceeds by solving the two-stage game using backward induction. In Stage 1 ("adoption"), *the agent* decides whether to adopt the new technology based on expected utility in

Stage 2 ("impact"). In Stage 2, as shown in Result 2, *the principal* would offer a new contract to the agent. The goal of this proof is to show which type of agents would adopt the technology in Stage 1 and under which conditions. For simplicity, we assume throughout this proof that $\kappa = 1$, meaning the high-effort signal perfectly reveals high effort $e = 2$. While the comparative statics still hold for $\bar{\kappa} < \kappa < 1$, the derivations become more complex.

There exists values of $\phi^h(2)$ and $\phi^l(2)$ such that the agents can be better off once they adopt the technology. To make the analysis more interesting, let's focus on the case where $\phi^h(2)$ and $\phi^l(2)$ are both too high such that no agents would adopt the technology absent changes to the contract. The principal can incentivize agents to adopt by offering a compensation upfront. As described before, this minimum upfront payment or 'salary' is denoted by w. In particular, the 'salary' should be sufficiently high such that, when IC binds:

$$\begin{gathered} U_2^\theta + w \geq U_{notech}^\theta \iff \\ w \geq \phi^\theta(2) + \bar{u} - q_2(\bar{u} + \frac{(1-q_0)\phi^\theta(1)}{q_1 - q_0}) \equiv W_{\tilde{e}=2}^\theta \end{gathered} \tag{D17}$$

Note that $W_{\tilde{e}=2}^\theta > 0 \iff \phi^\theta(2) > q_2(\bar{u} + \dfrac{(1-q_0)\phi^\theta(1)}{q_1 - q_0}) - \bar{u} \equiv \underline{\phi}^\theta$

The principal is better off offering a positive payment $W_{\tilde{e}=2}^\theta$ if the present discounted value to be matched with an agent adopting the technology $V_{e=2}^\theta$ minus this upfront payment is greater than the principal's objective function at baseline $V_{e=1,notech}^\theta$.

$$\frac{V_{e=2}^\theta - W_{\tilde{e}=2}^\theta}{1-\delta} \geq \frac{V_{e=1,notech}^\theta}{1-\delta} \tag{D18}$$

In Stage 1, there exists a range of low enough of disutility of high-effort for high-type $\phi^h(2) < \bar{\phi}^h$ and high enough disutility of high-effort for low-type $\phi^l(2) > \bar{\phi}^l$ such that the principal would prefer only the high-type to adopt and exert $e = 2$. The principal matched with low-types would be worse off offering $W_{\tilde{e}=2}^l$, so they have no incentive to offer it in the first place, given the low-type agent's cost of high effort. The reason is that the low-type agent is already exerting the "first-best" effort level, $e = 1$, from the social planner's point of view. Adopting the technology, which only reveals "high-effort", would push the agent to exert high effort $e = 2$. In other words, the principal would find it unprofitable to fully compensate the agent given his high disutility. Due to the lack of formal commitment, the principal cannot credibly commit to not demanding high effort once the agent adopts the technology, and would fire the agent if the effort signal is not provided.

As before, the following "no-deviation" condition needs to hold in equilibrium for the principal of a low-type agent to keep working with him once the technology is introduced:

$$\delta < \frac{V_{tech}^l}{-K_p + \mu V_{tech}^h + (1-\mu)V_{tech}^l} \equiv \bar{\delta}^{tech} \tag{D19}$$

with V^l_{tech} the new value function upon introduction of the technology for owners matched with a low-type not adopting the technology (where the outside option of the principal increases as they can now match with a high-type adopter). This will always be true for sufficiently high K_p. This "no-deviation" condition is ultimately verified in the structural estimation, see Table B28.

Note that I assume that the agent can keep the technology with them upon termination, meaning the principal would still need to incentivize the next matched low-type agent for adoption.

This completes the proof of Result 3. $\square$

E Structural Estimation

E.1 Framework Inputs

E.1.1 From Survey Data

The survey was carefully designed to provide all necessary inputs for structural estimation and welfare analysis. Except for the discount factor, all parameters are either calibrated from survey responses or used directly as moments.

For the production function, I calibrate q_1, the probability of high output when effort is $e = 1$, as the proportion of instances where drivers in the *No Observability* (N-O) and *Control* groups (at midline) achieved above-median output when working more than the median number of hours. Likewise, q_0, the probability of high output when effort is $e = 0$, is calibrated as the proportion of instances across the full sample where drivers achieved above-median output while working less than the 25th percentile of hours. This survey-based calibration aims to map the discrete effort-output production function, and the sensitivity of the estimation is discussed in Section E.4.

High and low outputs, Y and X, are defined as drivers' revenues net of expenses (including weekly food and beverage expenditures), measured in the survey. Consistent with the calibration of q_1, I set Y as the average revenue above the median and X as the average revenue below the median, conditional on working more than the median hours.

The share of high-type drivers μ is obtained directly from owners' beliefs. Owners were asked: "We are trying to understand your perception of drivers. There are good and bad taxi drivers in terms of work. Imagine that 10 drivers present themselves to you at random. Out of 10 drivers, how many do you consider to be 'good' drivers?" I use the average response across owners as the calibration of μ.

Drivers' outside options $\bar{u}^l, \bar{u}^h$ are allowed to vary by type. To estimate them, I conducted a representative survey of traders in September 2022, since trading is a common outside option for drivers who leave the taxi sector (as observed during the two-year experiment). Average profits of small traders (with 0–1 employees) provide the baseline outside option. To distinguish between high- and low-type drivers, I use differences in poverty likelihood (at 200% of the national poverty line) as measured by IPA, mapping these differences into heterogeneous outside options.

Finally, the owner's replacement cost K_p is calibrated from owners' answers to: "If you were to lose your current taxi driver, how long would it take you to find another similar driver?" Responses are converted into monetary terms to provide K_p.

E.1.2 Reduced-form Estimates

I use the reduced-form estimates to calibrate the following parameters:

- The private technological gains for the drivers, G, which is the treatment effect on imputed loss associated with costs of using cash at short-term, as described in Section 4.1.
- The production function for high-type drivers relies on the driver's effort and production under *Granular Observability*. Specifically, to calibrate q_2, I use the proportion of times drivers achieved an upper median output when they work more than the median hours in the *Granular Observability* group at mid-term.
- The contract characteristics rely on the owner-driver pairs under *Granular Observability*, in particular the upfront payment $W_{\tilde{e}=2}$ and the probability of retaining the driver $p_{\tilde{e}=2}$.

To test the sensitivity of the estimation to the framework inputs, I estimate the standard errors for each parameter, resampling with 1,000 bootstrap replications of the survey data.

E.2 Matched Moments: Description

Enriching the Framework. To better align the framework with the empirical setting, I introduce three main modifications. First, I incorporate the fact that the technology provides benefits to drivers through reduced cash-related costs, denoted by G, which accrue entirely to drivers. Second, I account for outside options to differ by driver type, $\theta \in l, h$. Third, I include an upfront payment, $W > 0$, given to drivers at baseline and received in both states of the world, such that this payment W does not affect the drivers' incentive compatibility or truth-telling constraints but influences their utility. These modifications are further discussed in relation to the specific moments they affect below.

The following empirical moments are then matched to their theoretical counterparts in order to recover the disutility parameters: $\phi^h(1)$, $\phi^l(1)$, and $\phi^h(2)$.

Moment 1: Baseline Continuation Probabilities $\bar{p}^l$

$$\bar{p}^l = \min \begin{Bmatrix} \bar{p}^{l,IC} = 1 + \dfrac{\bar{u}}{\delta K_a} - \dfrac{q_0 \phi^l(1)}{\delta K_a (q_1 - q_0)}, \\ \bar{p}^{l,TT} = 1 - \dfrac{1}{\delta K_a} \left[q_1 (Y - X) - \phi^l(1) - \bar{u} \right] \end{Bmatrix}$$

The principal selects the minimum between the two cutoffs to incentivize both effort and truth-telling. The low-type $\theta = l$ empirical moment is considered to be the re-hiring probabilities from the reluctant drivers (adoption experiment), as they were not influenced by the technology, and by not adopting the technology, reveal their type to the researcher.

Moment 2: Baseline Continuation Probability $\bar{p}^h$

$$\bar{p}^h = \min \begin{Bmatrix} \bar{p}^{l,IC} = 1 + \dfrac{\bar{u}}{\delta K_a} - \dfrac{q_0 \phi^h(1)}{\delta K_a(q_1 - q_0)}, \\ \bar{p}^{l,TT} = 1 - \dfrac{1}{\delta K_a}\left[q_1(Y - X) - \phi^h(1) - \bar{u}\right] \end{Bmatrix}$$

I use the re-hiring rate in the *Control* and *No Observability* groups, which consist of high-type drivers willing to adopt but unaffected by observability. This allows me to recover $\bar{p}^h$ for agents willing to adopt the technology without being influenced by monitoring.

Moment 3: Continuation Probability under Granular Observability $\tilde{p}_{\tilde{e}=2}$ The theoretical moment is derived in Appendix D.6, see Equation D12, and reported below.

$$\bar{p}^{\theta\prime} = \bar{p}^{\theta\prime} \leq 1 - \frac{\phi^\theta(2) - \phi^\theta(1) - (q_2 - q_1)(Y - R^\theta)}{((1 - q_1) - (1 - q_2)(1 - \kappa))\delta K_a} \tag{E1}$$

In the lens of the framework, it is implied that $\tilde{p}_{\tilde{e}=2} = (1-(1-\kappa)*(1-q_2)*(1-\bar{p}^{\theta\prime}))$. To simplify, I assume that $\kappa = 1$ such that the owner receives a perfect signal of high-effort, and that low-output is also perfectly observable, with no change in the rental payment R (which closely maps what is observed empirically). This makes $\tilde{p}_{\tilde{e}=2} = 1$. Empirically, this maps to the reduced-form re-hiring rate in the *Granular Observability* group, when the high-type agent is induced to exert $e = 2$. This treatment arm is expected to mitigate moral hazard in effort and in output reporting, as discussed in Section 5.4.

Moment 4: High-Type Agent's Contract Valuation U^h

$$U^h = \frac{W + q_1(Y - R^h) - \phi^h(1) - \delta K_a(1 - \bar{p}^h)(1 - q_1)}{1 - \delta} \tag{E2}$$

The empirical moment is the driver's self-reported contract valuation. Specifically, I asked drivers who adopted the technology the following question to approximate the driver's value of the relationship: "Imagine the following: how much would another taxi owner need to pay you to leave your current relationship with your owner and work for them in their taxi?" Here, I enrich the framework by considering that this contract valuation incorporates the calibrated baseline up-front payment W.

Moment 5: Agent's replacement cost K_a

This is an upfront search cost to be rematched to a principal. It is derived from the baseline survey question: "If you were to lose your current job as a taxi driver, how long would it take to find another similar job?" The reported number of days is multiplied by daily profits. The parameter K_a is matched exactly.

Moments 6 and 7: Agent's Transfers in High-Output Periods R^l and R^h

These empirical moments are the agents' median transfers in high-output periods, from the baseline survey, using the control and reluctant drivers, as with moments 1 and 2. Empirically, 91% of pairs have the same target transfer R^θ at baseline, resulting in a similar median across the two types of drivers. In the framework, this can be explained by the higher disutility of work and the lower outside option for low-types, which may offset each other and lead to a similar R^θ across types. The formula for each agent type $\theta \in l, h$ is as follow (see Section D.1):

$$R^\theta = \min \begin{Bmatrix} R^{\theta,IC} = Y - \bar{u} - \dfrac{(1-q_0)\phi^\theta(1)}{q_1 - q_0}, \\ R^{\theta,TT} = q_1(Y - X) + X - \phi^\theta(1) - \bar{u} \end{Bmatrix}$$

Moment 8: Upfront payment/Salary under Granular Observability $W_{\tilde{e}=2}$ The empirical moment is the upfront payment "salary" offered to drivers under *Granular Observability* when the high-type agent exerts $e = 2$. The rationale for this payment is to incentivize adoption and compensate the agent for exerting a higher level of effort. The estimation takes into account that some drivers were already receiving an upfront payment $W > 0$ at baseline.

Intuitions Behind Each Moment These eight empirical moments are matched to the theoretical moments to recover the parameters: $\phi^h(1)$, $\phi^l(1)$, $\phi^h(2)$. The model is over-identified.

The disutility of work of the low-types for $e = 1$, $\phi^l(1)$, is identified using **Moment 1** and **Moment 6**. Intuitively, if the job becomes tougher, then the continuation probability in a low-output period or the transfer in a high-output period would need to be lower to incentivize effort at baseline. Using the same intuition, I estimate the disutility of work for $e = 1$ for the high-types, $\phi^h(1)$, using **Moment 2**, **Moment 4**, and **Moment 7**.

The disutility of work of the high-types with $e = 2$, $\phi^h(2)$, is estimated using **Moment 3** and **Moment 8**. The intuition is that the owner must compensate the driver for increased effort under *Granular Observability* by offering both a higher upfront payment, $W_{\tilde{e}=2} > W$, and a higher continuation probability, $\tilde{p}_{\tilde{e}=2}$. Finally, **Moment 5** contributes to the overall identification of the parameter vector.

On the other hand, I estimate the low-type driver's disutility of work with $e = 2$, $\phi^l(2)$. Since the low-type drivers did not adopt the technology, I can only obtain a lower bound for this parameter using the following theoretical intuition: a low-type θ driver would require a sufficiently high upfront payment, $W^l_{\tilde{e}=2} > W$, (for a given $\tilde{p}_{\tilde{e}=2}$) to exert high effort $e = 2$. I compute the minimum value for $\phi^l(2)$ such that, at $W^l_{\tilde{e}=2}$, the owner would actually prefer to maintain the (baseline) status quo. Appendix D.6 details the derivations used to compute the lower bound for $\phi^l(2)$.[53]

[53] I also specify and check in the data the "no-deviation" condition, which states that, upon adoption of the technology by high-type agents, the owners matched with low-types should have no incentive to deviate by terminating the low-type agents, incurring the replacement cost K_p, and recruiting a new agent, with probability μ of being matched with a high-type who accepts the technology. See Table B28.

Untargeted Moment: Low-Type Driver's Contract Valuation U^l. To test the framework's fit, I consider an untargeted moment: the present-discounted contract valuation for low-type drivers at baseline. This moment was not included in the estimation procedure because I did not collect baseline contract valuation data from drivers who refused the technology, as their survey was intentionally made shorter, as discussed before. However, I use the empirical contract valuation from the group of adopting drivers who initially preferred their employer not to observe their transactions at baseline. I discuss in the main text (Section 6.3) how well the framework's predicted structural components align with both the targeted and untargeted empirical moments.

Finally, the owner's present-discounted contract valuation is defined with the following formula. The owner receives the driver's rental transfer, while the owner's costs include the upfront payment W offered at baseline to some drivers and the maintenance cost MC, along with the calibrated replacement cost K_p.

$$V^h = \frac{q_1 R + (1-q_1)X - W - MC - \delta K_p (1-\bar{p}^h)(1-q_1)}{1-\delta} \tag{E3}$$

E.3 Parameter Estimation Details

I estimate the parameters of interest using a GMM approach, which minimizes the distance between the structural and reduced-form components. My data $\mathbf{X}_i$ comprises eight empirical moments as described above. The inputs form the structural component. The GMM estimator minimizes the following objective function:

$$\hat{\beta} = \arg\min_{\beta \in \Theta} \left(\frac{1}{T} \sum_{i=1}^{T} g(\mathbf{X}_i, \beta) \right)^{\intercal} \hat{\mathbf{W}} \left(\frac{1}{T} \sum_{i=1}^{T} g(\mathbf{X}_i, \beta) \right)$$

Here, $g(\mathbf{X}_i, \beta)$ represents the difference between the vector of empirical moments ($\bar{p}$, p^h, $\tilde{p}$, U^h, K, R^l, R^h, $W_{\tilde{e}=2}$) and the vector of structural moments described above. Each empirical moment corresponds to a structural moment predicted by the model, allowing the GMM estimator to match observed and theoretical behavior. The weighting matrix $\mathbf{W}$ consists of the inverse variance of the estimation moments.

I also verify the "no-deviation" conditions specified in the theoretical framework for the results to hold (see Table B28) and that the replacement cost ensures the principal is better off requiring $e = 1$ at baseline with no information on effort and output (and not $e = 2$) and setting the transfer to the maximum in high-output periods.

E.4 Sensitivity of Parameter Estimates to Estimation Moments

I assess the sensitivity of the parameter estimates derived above to the matched moments, following Andrews et al. (2017). In particular, I derive the following sensitivity parameter:

$$\Lambda = (\mathbf{J}'\mathbf{W}\mathbf{J})^{-1}\mathbf{J}'\mathbf{W}$$

where $\mathbf{J}$ is the Jacobian matrix of derivatives of the 8 moments with respect to the 3 parameters ϕ_1^l, ϕ_1^h, and ϕ_2^h; and $\mathbf{W}$ is the weighting matrix, as described above. The sensitivity measures the asymptotic bias of the parameter estimates under local perturbations when all other parameters are held fixed. More specifically, the columns of Λ represent the sensitivity in dollars of a given parameter estimate to a one-unit change in each of the moments; the rows of Λ represent the moments.

To simplify interpretation, I convert the sensitivity values as follows: a 5-percentage-point change in the probability by the end of the experiment (after 28 weeks), the contract valuation U_h as a USD 100 change, and the other parameters—the replacement cost K_a, the transfers in high-outcome periods R^h and R^l, and the upfront payment $W_{\tilde{e}=2}$—as USD 10 changes.

Figure A7 displays the sensitivity matrix Λ in four panels, each corresponding to one parameter. I observe relative differences in parameters' sensitivity to the estimated moments. I find limited sensitivity to most moments for the disutility of work for low- and high-types. The three disutilities of work are primarily determined by the continuation probabilities p and the transfers R, consistent with the economic intuition that these primarily set the incentive compatibility constraints of the agent. The disutilities are not very sensitive to the contract valuation or replacement cost. Overall, the sensitivity is within reasonable dollar ranges, supporting the robustness of the structural results.