Risks, Uncertainty, and Flexibility in Gig Work: Evidence via Conditional Income Guarantee Interventions in Uganda[⋆,⋆⋆]

Mitali Mathur[a], Aleksandr Michuda[b,*], Shotaro Nakamura[c]

[a] *University of California, Davis, CA, 95618, USA*

[b] *Swarthmore College, 500 College Ave., Swarthmore, PA, 19081, USA*

[c] *Pennsylvania State University, 201 Old Main, University Park, PA, 16802, USA*

⋆Study preregistration submitted to: AEA RCT Registry, RCT ID AEARCTR-0017386

⋆⋆Acknowledgments: We thank Leo Nitczynski and the Swarthmore College Summer Scholars Program.

Funding: This work is supported by the Digital Economy Research & Impact Initiative (DERII) by IDInsight, supported by the Gates Foundation. Support is also provided by Swarthmore College.

*Corresponding author, amichud1@swarthmore.edu, 1-718-564-9741

Email addresses: mrmathur@ucdavis.edu (Mitali Mathur), amichud1@swarthmore.edu (Aleksandr Michuda), snn5225@psu.edu (Shotaro Nakamura)

Abstract

Informal workers, many of whom now work on gig-economy platforms in developing economies, lack formal contracts, social protection, and income stability. Can incentive design reduce risks and uncertainty in their earnings, while maintaining their preference for flexibility? Do such interventions improve productivity and workers' welfare? We will conduct a randomized controlled trial with drivers on a major ride-hailing platform in Kampala, Uganda, and randomly offer three bonus schemes: a fixed bonus, a conditional income guarantee, and additional flexibility to make-up bonuses. Using trip-level administrative data and survey measures, we will estimate the effects of exogenous bonus assignment on labor supply, earnings, consumption, and welfare. We will also conduct incentivized elicitation on bonus types to identify workers' preferences for risk and flexibility, and identify the welfare effects of matching contracts to workers' preferences. These findings will provide novel insights into incentive design to improve worker welfare in developing economies.

Keywords: Gig work, income variation, field experiments

JEL: D81, J33, O17

Risks, Uncertainty, and Flexibility in Gig Work: Evidence via Conditional Income Guarantee Interventions in Uganda

Timeline

Years	2025																																				2026																							
Months	Mar				Apr				May				Jun				Jul				Aug				Sept-Oct				Nov				Dec				Jan				Feb				Mar				Apr				May				Jun			
Weeks	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Project Preparation																																																												
Finalize IRB protocols (UNCST & partner institutions)																																																												
Establish partnership with SafeBoda																																																												
Pilot Study Preparation																																																												
Acquire access to SafeBoda's administrative data																																																												
Random sampling of study participants																																																												
Code questionnaire in SurveyCTO																																																												
Hiring - Hire enumerators and field assistants																																																												
Training - Train team on survey tool and procedures																																																												
Pilot Study																																																												
Baseline - collect primary data via surveys																																																												
Randomize participants - assign drivers to each arm																																																												
Mobilize - prepare the intervention with SafeBoda																																																												
Intervention - Implement the treatments for each arm																																																												
Endline - retrain enumerators on updated tool																																																												
Endline - collect primary data via surveys																																																												
Analysis - clean data from primary surveys & merge with SafeBoda administrative data, then analyze																																																												
Present - share pilot results with SafeBoda																																																												
Full-Scale Study Preparation																																																												
Brainstorm full-scale intervention tweaks																																																												
Conduct focus groups with SafeBoda drivers to understand preferences and behaviors more deeply																																																												
Finalize protocols and write a Pre-Analysis-Plan																																																												
Amend IRB																																																												
Random sampling of study participants																																																												
Code questionnaire in SurveyCTO																																																												
Hiring - Hire enumerators and field assistants																																																												
Training - Train team on survey tool and procedures																																																												
Full-Scale Study																																																												
Baseline - collect primary data via surveys																																																												
Randomize participants - assign drivers to each arm																																																												
Mobilize - prepare the intervention with SafeBoda																																																												
Intervention - Implement the treatments for each arm																																																												
Endline - retrain enumerators on updated tool																																																												
Endline - collect primary data via surveys																																																												
Intervention - continue the intervention based on preferences from endline																																																												
Analysis - clean data from primary surveys & merge with SafeBoda administrative data, then analyze																																																												
Present - share final results with SafeBoda																																																												

1. Introduction

Informal workers make up a large share of the labor market in developing countries. This sector is characterized by precarious labor opportunities that result in low wages relative to those in the formal sector, coupled with increased uncertainty around take-home pay (World Bank, 2018). On the demand side of the labor market, variations and uncertainty in earning opportunities can limit worker productivity and the ability to smooth consumption (e.g. Morduch 1995; Murthy and Deshpande 2022). On the supply side, supplied work hours may be driven by shocks and uncertainty in cash needs (Mobarak and Rosenzweig, 2014; Anonymous, 2024a). Additionally, informal workers often face personal constraints and commitments that make it difficult to maintain a steady income. These three challenges may cause welfare losses to workers, especially if they cannot fully insure against shocks.

The challenges of informal workers meeting their income needs are likely to persist with the rise of workplaces enabled by the digital gig economy. Gig workers are a substantial portion of the labor force in developing countries; a recent study of 17 developing economies estimates that the share of online gig workers in the global labor force ranges from 4.4 to 12.5 percent (Datta et al., 2023). Anecdotal and descriptive evidence shows that gig workers in developing countries face challenges similar to those of other informal workers, such as lower wages, limited protection under labor regulations, and earnings uncertainty (Anwar and Graham, 2021; Ayentimi et al., 2023; Berg et al., 2018; Heeks et al., 2020).

Regulations and contracts to increase income, reduce earnings uncertainties, and accommodate workers' commitments may affect worker labor supply and welfare. However, empirical evidence on the effects of earnings interventions on labor supply and welfare in developing gig economies is limited, with the exception of transaction-level price controls in Indonesia (Anonymous, 2024b). We will study the efficacy of a variety of conditional earnings contracts in addressing the challenges faced by informal workers in developing economies. Our research design will allow us to examine whether gig economy workers prefer working on platforms because of an inherent preference for flexible earnings at the cost of uncertainty, or whether they work under flexible contracts because of limited labor-market opportunities or limited access to insurance contracts. Understanding this difference is at the heart of several academic and policy debates over preferences for flexibility in the digital economy and for the employment of last resort (Chen et al., 2019; Glasner, 2023; Hacamo and Kleiner, 2022; Harris and Todaro, 1970).

Our research addresses the following questions. First, how do conditional earnings guarantees in the gig economy affect labor supply, earnings, and welfare? Second, how does the reduced earnings risk from such contracts affect driver outcomes and preferences for flexibility?[1]. Third, how do guaranteed income contracts affect labor supply, consumption, and financial well-being outside of the gig economy? Fourth, do earnings guarantees and additional flexibility improve worker welfare, beyond the income effects that these contracts generate?

To answer these research questions, we will conduct a randomized controlled trial in which we will randomly offer various forms of equal-value driver incentives between treatment groups to a subset of drivers in a major rideshare platform in Kampala, Uganda. We will conduct two phases of treatment.

In the first phase, we randomly assign active drivers on the platform to one of four treatment groups: T0, T1, T2, or T3. T0 is a pure control group with no financial incentives.[2] In T1, drivers are offered a set amount of extra income (bonus) when they supply enough labor to meet a threshold. In T2, drivers are offered a conditional earnings guarantee in which drivers are guaranteed at least a certain level of earnings (i.e., income floor) each day they meet the same work requirement. In T3, drivers are offered a more flexible version of the fixed bonus (T1), allowing them to "make up" missed bonuses from the previous day.

In the second phase, which will occur after the first intervention phase and the endline survey, drivers will participate in an incentivized game based on the Becker-DeGroot

[1]We define earnings risk as the ex-ante unknown variations in drivers' daily earnings that they cannot mitigate or insure

[2]The platform defines an active driver as one who has completed at least one trip in the period in question.

Marschak (BDM) method (Becker et al., 1964). In the game, drivers are assigned the fixed bonus as in T1 by default, but can bid for T2 and T3. They will receive two weeks of bonus incentives based on the outcome of the BDM game.

Our experimental design allows us to disentangle key mechanisms that may drive the effect of financial contract types in this labor market. Comparing T1 against T2 in the first phase will enable us to disentangle the income effect from the benefit of the income guarantee, which reduces downside earnings risk. Comparing T1 to T3 will allow us to identify the impacts of additional flexibility while holding the bonus amount constant. In the second phase, the combination of a random choice between two BDM games we play and the randomly drawn price creates exogenous variation in whether drivers are matched with their preferred bonus type, allowing us to identify whether the match in preferences to contract type is a key mechanism.

We prespecify primary outcomes on labor supply, earnings, expenditures, and preferences over contract types, based on a combination of the platform's administrative database, incentive-compatible preference measures from lab-in-the-field games, and stated measures from in-person surveys. With access to the ridesharing company's database, we can precisely measure labor supply and driver earnings on the platform. Through primary surveys, we collect measures of types and quantities of labor supply outside of the collaborating platform and the gig economy, credit use, consumption, and investments in productive assets. These comprehensive measures of labor supply, earnings, and expenditures allow us to identify the effects of our intervention on consumption. Furthermore, we will use the bids from the BDM games as outcome measures to reveal drivers' preferences for bonus-incentive attributes (e.g., reduced risk, flexibility).

Our research contributes to four strands of literature. First, we extend our understanding of labor supply and its mechanisms in the casual and informal economies of developing countries, drawing on insights from modern gig work. The literature has posited three models behind labor supply decisions in these contexts: neoclassical models of intertemporal substitution, behavioral models of reference dependence, and models emphasizing liquidity constraints. In Neoclassical models, individuals smooth consumption by adjusting consumption expenditures based on earnings and smooth income by diversifying employment opportunities in response to uncertainty (Farber, 2015; Stafford, 2015; Banerjee and Duflo, 2007; Morduch, 1995). Liquidity constraint models posit that individuals supply labor as a means to access short-term credit and smooth consumption (Rossi and Trucchi, 2016). Reference-dependence models suggest individuals evaluate income based on set income targets (Kőszegi and Rabin, 2006; Dupas et al., 2020; Duong et al., 2023; Buchholz et al., 2023). These labor supply models, however, do not consider the main characteristic of gig work: the salience to drivers and value of flexibility in work (Bodie et al., 1992; Chen et al., 2019). Even traditional taxi services are characterized by shift work, not a fully flexible labor supply in which a driver can turn on the application and choose passengers at will (Duong et al., 2023), with reference points shifting throughout the day Thakral and Tô (2021). Our study bridges the gap between these models by leveraging experimental variation in contract design, coupled with rich administrative data to disentangle the relative importance of income smoothing, liquidity constraints, and flexibility in shaping labor supply (Hu et al., 2025; Chen et al., 2025).

Second, we contribute to the literature on the gig economy in a developing-country context, where binding liquidity constraints and limited insurance access affect welfare and optimal platform policy. Much of the evidence on the mechanisms behind labor supply decisions has come from developed countries, where drivers often use platform work to supplement income during unemployment shocks, retirement, or in addition to primary employment Jackson (2022); Chen et al. (2019); Cook et al. (2019). In our context, we study gig workers whose outside options are other informal labor or gig work opportunities (Li et al., 2021). Unlike full-time or shift-based employees (e.g., taxi drivers), gig workers experience algorithmically determined pay, often leading to greater wage uncertainty. Previous studies in Mozambique suggest that the entry of the Uber platform increased demand for services but did not lead to an associated increase in supply, suggesting frictions to formal employment (Jones and Manhique, 2025). Our study addresses this gap by providing evidence on the constraints that shape driver decision-making, including consumption, durable goods expenditures, and loans. Not only will we gain a deeper understanding of how workers allocate their time between income-generating activities, but we will also be able to determine the extent to which their household welfare is affected by platform policy.

Third, we contribute to the literature on insurance, contract design, and algorithmic management (Lee et al., 2015). Previous work has highlighted how households in developing countries use informal mechanisms to insure against shocks (Townsend, 1994; Dercon, 2002). Past interventions have sought to aid households in smoothing income and buffer against risks. Our study evaluates one of these potential on-platform interventions: formalized contracts designed to reduce risk while plausibly improving productivity. By evaluating the impact of contracts, we contribute to debates on the flexibility-insurance trade-off, shedding light on the welfare implications of contract design in gig labor markets and on the role that algorithmic management can play in achieving development goals.

Fourth, we contribute to the literature on the connection between worker preferences and contract design by examining the mechanisms through which contracts affect worker behavior (Herbst, 2025). We build on studies that elicit worker preferences (Dizon-Ross and Zucker, 2025; Mas and Pallais, 2017) and compare outcomes under randomized contract assignment to those under workers' preferences, allowing us to isolate the role of preference alignment as a mechanism behind contract effectiveness. Our research on contract design is especially relevant in developing-country settings, where large-scale interventions are costly and logistically challenging to implement.

Overall, we provide causal evidence on how gig workers in developing countries respond to randomized financial incentives that trade off flexibility and income security. Our design allows us to test core labor supply theories while also informing policy-relevant questions on gig economy regulation and welfare.

2. Research Design

2.1. Context

The research team will collaborate with a large ridesharing platform that offers motorcycle taxi (i.e., boda) services in Kampala, Uganda. The platform matches drivers with trips and charges commission fees. The platform also addresses safety and credit issues by mandating helmets and issuing loans for drivers' motorcycles and telephones.

Rideshare drivers in Kampala face several risks and uncertainties in their earnings. First, as is the case in many other ridesharing contexts, driver earnings vary and depend on the demand conditions in the locations where they work. Second, the uncertainty around demand conditions is exacerbated by traffic and accident risk (Hamza et al., 2023). Third, they face liquidity shocks and cash needs to meet their household needs; in many cases, boda drivers are connected to rural relatives who rely on them for financial assistance (Anonymous, 2024a). A previous focus group conducted with a microcredit firm in 2019 also found that boda drivers still face liquidity constraints that can only be partially offset by adjusting supply hours on ridesharing apps.

2.2. Basic Methodological Framework

We will conduct a randomized controlled trial (RCT) to understand how drivers respond to financial incentives in the form of conditional bonuses, conditional earnings guarantees, and flexible bonuses. Doing so enables us to isolate how drivers make decisions and are impacted by the income effect, changes in uncertainty about earnings, and preference for flexibility. We will have two phases to the experiment; in the first phase, we will randomly assign drivers to one of three bonus incentive types or a control group. Drivers will receive daily bonus incentives for six weeks following the baseline survey and random assignment. In the second phase, which occurs after the endline survey, drivers bid for contract types and are assigned one of the contracts determined by the BDM method.

2.3. Intervention

2.3.1. Study Population, Inclusion Criteria, and Randomization

We will collect data from a representative sample from a list of "active drivers" on the platform, defined as drivers who have completed at least one successful trip in a given time period.[3] These drivers could be part-time or full-time workers and do not need to meet a minimum number of trips to qualify for our study.[4] We first randomly select our sample of drivers from the pool of active drivers. Next, we complete a baseline survey to understand their current driving habits and basic demographics. As part of the consent process, drivers can decline to participate in the intervention.

After the baseline survey, we randomly assign drivers who agreed to be part of the intervention, to one of the four treatment groups for the first phase, stratifying on five dimensions: whether or not a driver worked a job in addition to being on this specific platform, whether they worked above or below the median number of hours of drivers in the sample, variation in hours worked across days (defined as within-driver standard deviation measure, and turned into above- or below-median binary indicator), their stated general preference for flexibility in when and how long to work, and their stated general risk preferences (high-risk or low-risk based on a binary indicator). The stratified random assignment ensures that drivers of different characteristics are well represented across different treatment arms.

2.3.2. Phase-1 Treatment Arms

For the first phase, we will randomly select 1,000 drivers from the platform database and randomly assign them to four arms. We will evenly split the sample, with 250 drivers per treatment arm. One challenge is that our partner already has a bonus structure in place that awards drivers daily bonuses based on a points system (consisting of the total number of trips and the proportion of cashless trips). We will refer to this existing system as the "partner

[3] This is typically 1 week or 1 month

[4] We follow the platform's definition of part-time and full-time workers. In a week, drivers who work more than 3 days a week are considered full-time, while those who don't are considered part-time.

bonus" and any additional financial incentives from our study as the "study bonus." We have designed the study arms to still tease out the effects of financial incentives by selecting thresholds *below* the existing "partner bonus" thresholds. In this way, the first opportunity drivers have to receive a bonus is by reaching our "study bonus" threshold.

1. **T0—Pure Control:** No "study bonus."

2. **T1—Fixed Bonus:** Drivers who meet a particular condition are given a fixed "study bonus" V amount in addition to their daily post-commission and "partner bonus" earnings.

   - **Condition type:** labor supply
   - **Condition description:** drivers who earn X points (based on completing a certain number of trips, with preference to cashless trips) qualify.
   - **Total driver earnings:** Let V be the fixed "study bonus".
     F = post-commission earnings + partner bonus + V.

3. **T2—Earnings Guarantee:** Drivers who meet the same condition as in T1 are guaranteed a minimum earnings for the day G that is equivalent to the expected value of total earnings under T1, conditional on qualifying for the bonus, i.e., $\mathbb{E}[F|points \geq X]$.

   - **Condition type:** labor supply
   - **Condition description:** drivers who earn X points (based on completing a certain number of trips, with preference to cashless trips) qualify
   - **Total driver earnings:**
     - If a driver is set to earn more than G one day and meets the condition, they do not receive any additional bonus and earn beyond G
     - If a driver is set to earn less than G one day and meets the condition, they will earn G guaranteed. The "study bonus" will pick up the shortfall and be equal to G - "partner bonus" - post-commission earnings

4. **T3—Flexible bonus:** Drivers are guaranteed the same, fixed "study bonus" (V) for the day as in T1, but the condition for qualifying for that control is more flexible. We provide additional flexibility by allowing drivers to "recover" that target.

    - **Condition type:** labor supply
    - **Condition description:** drivers who earn X points (based on completing a certain number of trips, with preference to cashless trips) qualify for the bonus that day (t). However, if they do not meet the bonus the previous day ($t-1$), they can "recover" the bonus (V) based on the points they earn on day t.
    - **Total driver earnings:**
        - Let V be the study bonus. If a driver earns X points on day t, they earn F = post-commission earnings + "partner bonus" + V
        - If a driver earns less than X points on day t, they do not receive a bonus and earn the post-commission earnings
        - Suppose a driver earns $x_{t-1} < X$ points on day $t-1$. On day t, they can recover these points if the driver earns $2X - x_{t-1} + P$ points, where P is a (small) penalty point. They will earn $2V$ (effectively "recovering" the bonus from $t-1$) for day t
        - Points do not "double count" towards bonuses

To calculate the fixed bonus (V), guaranteed earnings (G), and labor supply condition (X), the research team will leverage data from the platform and a pilot study conducted in May 2025. We first combine data on the number of trips completed per driver, per day, with daily earnings and "partner bonus" amounts. Next, we loop through different combinations of the missing parameters and calculate both power and the proportion of drivers who would qualify under each arm. We then select thresholds that ensure we are sufficiently powered to detect effects and that have financial incentives that would increase take-up.

2.3.3. Phase-2 Intervention Based on BDM

In the second phase, we will offer all drivers opportunities to bid for different bonus types. The exact number of drivers depends on attrition at the endline survey.

At the endline survey, drivers are given the opportunity to participate in the following two BDM games. They are given an endowment of UGX 6,000. They are told that the outcome of one of the games will be selected at random to determine the bonus incentives for the two weeks after the endline survey. After the drivers make their bids, we will flip a coin to choose one of the two games. Then, we will draw a single random price from a uniform distribution between 0 and UGX 6,000. If the randomly drawn number is less than or equal to the driver's bid for the selected game, they will get the bonus type they bid for at the randomly drawn price, and be paid the difference between UGX 6,000 and the random price. If the randomly drawn price is greater than the bid, the driver will get T1 and be paid UGX 6,000. The two games are as follows:

- **Willingness to pay for T2 over T1**: Drivers are told to bid any value between 0 and UGX 6,000 to receive the T2 bonus type (conditional income guarantee). If they win the bid, they will receive the T2 bonus incentive for two weeks. If they lose, they will receive T1.

- **Willingness to pay for T3 over T1**: Drivers are told to bid any value between 0 and UGX 6,000 to receive the T3 bonus type (fixed bonus with additional flexibility). If they win the bid, they will receive the T3 bonus incentive for two weeks. If they lose, they will receive T1.

2.3.4. Timeline: Treatment duration and data collection

The timeline and durations of treatments and data collection are as follows:

<table>
<tr><td colspan="3">March</td><td colspan="4">April</td><td colspan="3">May</td></tr>
<tr><td>Week 2</td><td>Week 3</td><td>Week 4</td><td>Week 1</td><td>Week 2</td><td>Week 3</td><td>Week 4</td><td>Week 1</td><td>Week 2</td><td>Week 3</td></tr>
<tr><td>Baseline Survey</td><td colspan="6">Phase I –
Initial Intervention
(random assignment)</td><td>Endline Survey</td><td colspan="2">Phase II –
Intervention
(preferences)</td></tr>
</table>

The Phase I intervention will last 6 weeks, and the Phase II intervention will last 2 weeks. During the week in which we conduct the endline surveys (Week 1, May 2026), we will not provide bonus incentives, as the in-person surveys take drivers off the road for a significant amount of time. We will resume bonus incentives with new assignments in Phase 2 in the following week.

We will be able to access and download data from the platform's database throughout the research period, starting before our main intervention and continuing past the end of the Phase-2 intervention.

2.4. Primary Outcomes

Conditional earnings guarantees and other bonuses can affect workers' labor supply, earnings, consumption, expenditure, and contract preferences. We define primary outcomes into four categories to estimate the causal effects of the treatment on them. We also define the "first-stage" take-up variables, i.e., whether a driver achieves the labor supply condition to qualify for a bonus. The unit of observation is the driver-day for outcomes using the platform database, and the driver at baseline and endline for those based on the surveys.

0. Take-up

To identify the treatment-on-the-treated (ToT) effects of the financial incentives we provide, we define the "treatment" variables separately from the assignment variable. The treatment variables are used for our empirical specifications in Equations 3 and 4 of section 2.6.1.

The treatment variable $Q^a_{i,s,t}$ takes the value 1 if driver i assigned to Treatment $a \in$

$\{1, 2, 3\}$ qualifies for the bonus condition on day t. For T1 and T2, this means earning X points or more on that day. For T3, we also backfill the days for which drivers make up the bonus condition. In other words, if a driver does not meet the bonus threshold on day t alone (e.g., only collects $X - 10$ points) but makes up the gap on day $t+1$ (e.g., collects $X + 10$ points), then we set $Q^3_{i,s,t} = 1$, as well as $Q^3_{i,s,t+1} = 1$.[5] We set $Q^a_{i,s,t} = 0$ for all others, i.e., those in T_a that do not meet the bonus, and those in every other treatment arm, including $T0$.

1. Labor Supply

Conditional earnings guarantee and other incentives can affect workers' participation and effort. Specifically, our intervention may not only affect how much to work, but also on which income-generating activities, including the partnering ridesharing platform. To measure labor supply, we specify the two following primary outcomes as the best measures capturing the amount of work on and off the platform:

- **Outcome 1A: Daily Active Hours on the Platform** - The total number of hours a driver spends with the platform app open per day, measured via the platform database
  - We will use the platform's database that includes a measure of a driver's active hours via high-frequency pings to identify whether a driver has the app open
  - This measure captures both a driver's actual work hours *and* the time the driver spent waiting for a rider to demand a ride (i.e., idle time)
  - This outcome is available as part of the driver-day level panel data
- **Outcome 1B: Weekly Hours Spent Working outside the Platform** - The total weekly hours spent on all income-generating activities except for the ridesharing platform on which we offer incentives, measured via survey data

[5] We will also conduct robustness checks using a version of $Q^3_{i,s,t}$ that does not backfill, i.e., $Q^3_{i,s,t} = 0$ even when the driver makes up on day $t+1$.

- We will use survey data to estimate the total number of self-reported hours a driver spends in a given week on each income-generating activity, excluding the platform.[6]
- This measure captures the types of labor supply we cannot observe in the platform data. We will also ask how many hours a driver reports having worked on the platform in the survey. We will compare this measure against the platform's administrative data (Outcome 1A) to gauge the magnitude of recall bias in the survey-based labor supply measures.
- This outcome will be part of the driver-level survey dataset and captured at both baseline and endline.

Outcome measures 1A and 1B are the best available measures that capture labor supply on and off the platform, respectively. We will use these measures to determine whether the conditional earnings guarantee and other incentives affect labor supply and whether drivers substitute away from other income-generating activities outside the platform. Although we prioritize outcomes 1A and 1B in our analysis, we note additional alternative definitions of labor supply outcomes:

- **Outcome 1C: Daily Transaction Hours on the Platform** - The total number of hours a driver spends *actively* making trips on the platform, per day, measured via platform app data
  - We will use the platform data on total hours spent on trips (as opposed to Outcome 1A, which includes idle time on the app) to identify treatment effects on the quantity of paid work conducted on the platform

[6] We ask drivers about their sources of income in the past week and include the following categories as income generating activities: platform job, transportation digital platform work on competing platforms, non-transportation digital platform jobs [ex: service jobs], offline transportation work, full-time formal employment, full-time informal employment, part-time formal employment, part-time informal employment, own business, agriculture, rental/investment income, retirement, other)

- This outcome will be part of the driver-day level panel data set

- **Outcome 1D: Weekly Hours Spent Working on and off the Platform** - The total weekly hours spent on all income-generating activities, *including* for the ridesharing platform on which we offer incentives, measured via survey data
  - This measure is identical to Outcome 1B, except that we also include reported hours worked on the platform.
  - This measure captures the total hours worked, both on and off the platform, through a consistent source.

2. Earnings

Conditional earning incentives directly impact a driver's earnings by changing the total take-home pay from platforms. To measure earnings, we specify one primary outcome:

- **Outcome 2A: Daily Platform Earnings** - The total post-commission income earned via platform trips, exclusive of bonuses, measured via platform app data
  - We will use backend platform app data to aggregate post-commission earnings across each trip for each day
  - We do not incorporate any bonuses or "top-up" bonuses as our intervention is designed such that the expected value of take-home pay is consistent across arms
  - This measure is useful for capturing the income effect of each treatment arm (irrespective of any mechanical increases from bonuses/"top-ups")
  - This outcome will be part of the driver-day level panel data set

We will prioritize outcome 2A for our primary analysis. However, we also define an alternative definition from survey data to capture total take-home pay:

- **Outcome 2B: Total Weekly Earnings** - The total weekly earnings earned from all income-generating activities, measured via survey data

  - We use survey data on all of the self-reported income-generating activities that different drivers have
  - This measure captures total earnings across activities to reflect changes in earnings when the relative share of time across activities might have changed
  - This outcome will be part of the driver-level survey data set and captured during both baseline and endline

3. Consumption and Expenditures

Conditional income guarantees and other incentives may affect workers' expenditure patterns in response to changes in the income streams they expect from gig work. Changes in expenditures may come not only from consumption but also from other sources, such as paying down debt, remittances to family members, and investments in other income-generating activities. We define one primary outcome to measure expenditures:

- **Outcome 3A: Total Overall Monthly Expenditures** - The total amount of money a household spent on "overall expenditures" last month, measured via survey data
  - We use survey data to ask drivers the amount of money they spent on their household overall in the last month
  - While this measure is slightly noisy, it is nonetheless useful in noting how the intervention translates to consumption
  - This outcome will be part of the driver-level survey data set and captured during both baseline and endline

We will use total expenditures to remain agnostic about which types of spending are affected by our interventions. We plan to identify the impact on the aggregate measure of expenditures, then identify the types of expenditures most affected as secondary analysis to understand the mechanisms. We also prefer to use a continuous measure of expenditures rather than binary outcomes for meeting income needs, as it is less prone to respondents' subjective biases and more likely to capture expenditure changes with a flexible measure. Nonetheless, we define the following as alternative measures of expenditures:

- **Outcome 3B: Total Monthly Expenditures for Own Household Consumption** - The total amount of money a household spent on their own consumption last month, measured via survey data. This is a subset of total expenditures focused on immediate household needs, which may be of primary concern to drivers. This measure is also more focused on their own consumption, rather than the total expenditure measures that include debt, investment, and remittances.
  - We use survey data to ask drivers the amount of money they spent on their household on their own consumption (food, clothes, household items, etc.) in the last month
  - This measure is useful for specifically tracking changes in their own consumption
  - This outcome will be part of the driver-level survey data set and captured during both baseline and endline
- **Outcome 3C: Income Needs** - A binary indicator for whether or not a driver can meet their monthly income needs, measured via survey data
  - We use survey data to collect information on whether or not a driver met their income needs in the past month
  - This measure captures whether drivers are earning enough to meet their needs
  - This outcome will be part of the driver-level survey data set and captured during both baseline and endline

4. Preferences over Incentive Types

Eliciting driver preferences over incentive types is critical for evaluating the overall welfare effects of incentive programs, as well as the roles of attributes, such as lower earnings risk and flexibility. To properly estimate the welfare effect and its attributes, we will rely on elicited preferences between the incentive types. We will use the following measures:

- **Outcome 4A: Preference for Reduced Earnings Risk** - The bid (in UGX) driver makes to receive T2 (earnings guarantee) over T1 (fixed bonus) for two weeks. We provide details of the BDM games in Section 2.3.3, as well as in the survey instrument in the supplement.
  - We will elicit drivers' preference for T2 over T1. The drivers bid a value between UGX 0 and 6,000.
  - This measure allows us to capture drivers' relative preferences for the features of T2 (lower earnings risk via the earnings floor) relative to those of T1 (a fixed bonus).
- **Outcome 4B: Preference for Flexibility** – The bid (in UGX) driver makes to receive T3 (fixed bonus with additional flexibility) over T1 (fixed bonus) for a week
  - We will elicit drivers' preference between for T3 over T1. The drivers bid a value between UGX 0 and 6,000.
  - This measure allows us to capture drivers' preferences for the additional feature of T3 (the ability to make up the bonus the next day, i.e., flexibility) in addition to other features common across T1 and T3.

2.5. Secondary Outcomes

In addition to the primary outcomes defined earlier and their alternative definitions, we explore other outcomes that may further our understanding of the intervention's effects.

Our first set of secondary outcomes includes additional measures for how each intervention influences driver trips. Our primary outcomes focus on time spent on the platform, which more closely proxies effort. While time and trips might be correlated, trips are a more direct measure of labor supply relevant to the platform. In particular, we explore outcomes related to the number of driver trips made daily and weekly, and how time spent on platform trips relates to other income-generating activities.

Our second set of secondary outcomes characterizes the types of trips made. We investigate whether drivers adjust the features of their trips (e.g., length, distance, location, timing) to meet the conditions for receiving the incentives. These outcomes are useful to explore, as they can provide evidence of drivers' knowledge of the platform.

Our final set of secondary outcomes includes stated measures of welfare from survey data. These are useful for exploring other proxies of how drivers' finances have changed following the intervention. We specifically investigate driver savings, remittances, and loans. We also measure driver job satisfaction as another subjective measure of welfare.

A more detailed description of each of our secondary outcomes is in Appendix Appendix A.2.

2.6. Hypotheses and Inference

The randomized interventions are designed to disentangle three factors that determine the effects of conditional income guarantees and other financial incentives on driver outcomes: the income effect, changes in uncertainty about earnings, and the preference for flexibility. The primary, pre-specified hypotheses are based on the following specifications.

2.6.1. Specifications for the Phase-1 intervention

The data structure for our primary outcomes takes two forms: a) driver-day level panel data from the administrative database and b) driver-level survey data at baseline and endline. For both data types, we will estimate both the intent-to-treat (ITT) and treatment-on-the-treated (TOT) effects, where treatment "takeup" is defined as meeting the bonus threshold. However, we prioritize estimating the ITT effect. The ITT is the more policy-relevant estimate, as one of our primary objectives is to assess how workers adjust their labor supply in response to learning about the conditional earnings policy. Estimating a ToT would be less informative, as platforms cannot compel all workers to alter their labor supply decisions to meet the condition. Therefore, although we estimate both effects, we prioritize the ITT.

The statistical approach for outcomes in the administrative data is the difference-in-differences model, as follows:

$$Y_{i,s,t} = \beta_1 \times T_i^1 \times \mathbb{1}_{t>0} + \beta_2 \times T_i^2 \times \mathbb{1}_{t>0} + \beta_3 \times T_i^3 \times \mathbb{1}_{t>0} + \alpha_i + \omega_t + \varepsilon_{i,t} \tag{1}$$

The model is estimated on administrative panel data for individual i of stratum s on day t. The time period will run from at least 6 weeks before the beginning of the intervention (or as long as we can, to ensure a balanced panel) through the last day of the Phase-1 treatment period. The start of the experimental period is standardized at $t = 1$, with pre-treatment periods parameterized as $t \leq 0$. T_i^1, T_i^2, and T_i^3 are treatment assignment indicators for Treatment Groups 1, 2, and 3, respectively. We include individual (α_i) and time (ω_t) fixed effects.

The statistical model for the survey-based outcomes is ANCOVA, as follows:

$$Y_{i,s,t=endline} = \beta_1 \times T_i^1 + \beta_2 \times T_i^2 + \beta_3 \times T_i^3 + \delta_s + Y_{i,j,t=baseline} + \boldsymbol{X_i'}\boldsymbol{\psi} + \varepsilon_i \tag{2}$$

The model is estimated using survey data, in which individual i is surveyed at baseline ($t = baseline$) and endline ($t = endline$). Aside from the primary treatment indicators, we will include strata-level fixed effects (δ_s), baseline outcome variable ($Y_{i,j,t=0}$), and other covariates if we find imbalance ($\boldsymbol{X_i'}\boldsymbol{\psi}$).

To estimate the ToT effect, we use an instrumental variables approach, in which qualifying for a treatment condition is a function of the random assignment variable to the treatment arm. In the first stage, we estimate the following equation for each assignment arm a if the outcome is from the administrative data in a panel format.[7]

$$Q^a_{i,s,t} = \gamma_a \times T^a_i \times \mathbb{1}_{t>0} + \alpha_i + \omega_t + \varepsilon_{i,t} \tag{3}$$

$Q^a_{i,s,t}$ is a binary variable for whether the driver qualified for the bonus in the treatment arm T_a on day t. We, then, estimate the following equation in the second stage.

$$Y_{i,s,t} = \beta_1 \times Q^1_{i,s,t} \times \mathbb{1}_{t>0} + \beta_2 \times Q^2_{i,s,t} \times \mathbb{1}_{t>0} + \beta_3 \times Q^3_{i,s,t} \times \mathbb{1}_{t>0} + \alpha_i + \omega_t + \varepsilon_{i,t} \tag{4}$$

2.6.2. Specifications for the Phase-2 intervention

In the second phase, we are interested in identifying the causal effect of receiving the bonus type a driver prefers. We identify drivers' relative preferences among the three contract types (T1, T2, and T3) from bids in the BDM game. We denote driver i's bid amount for T2 and T3 over T1 as B^2_i and B^3_i, respectively. Based on bids and assumptions about transitivity, we determine which contract type a driver prefers most by identifying the contract for which they bid the highest amount. We denote this preference measure as $L_i = 1$ if T1 is the most preferred contract (i.e., $B^2_i < 0$ and $B^3_i < 0$), and $L_i = 2$ if T2 is the most preferred contract (i.e., $B^2_i > 0$ and $B^2_i > B^3_i$), and $L_i = 3$ if T3 is the most preferred contract (i.e., $B^3_i > 0$ and $B^3_i > B^2_i$).

We will identify the causal effect via the 2SLS method. The variables to be instrumented are whether the bonus assigned to the driver in the second phase, denoted $A_i \in \{1, 2, 3\}$, is the contract they prefer. We define these binary variables as $M^c_i = \mathbb{1}(A_i = L_i = c)$ for $c \in \{1, 2, 3\}$. There are two exogenous sources of variation on whether a driver receives the bonus type they prefer: the randomly drawn price P_i, and which of the BDM games (T1 v.s. T2, or T1 vs. T3) will be picked for consideration, denoted as $M_i = 1$ if the T1 vs. T2

[7]The specification for the survey-based outcomes is analogous.

game is picked, and zero if the T1 vs. T3 game is selected. The following is the first-stage equation, in which we predict M_i^c using the two instruments.

$$M_i^c = \gamma_{c,1} \times P_i + \gamma_{c,2} \times M_i + \varepsilon_i \tag{5}$$

The second-stage equation is estimated on the panel data from $t = 43$ (the day after the end of Phase 1) to $t = 63$ (the last day of Phase 2) as follows:

$$Y_{i,s,t} = \beta_1 \times M_i^1 \times \mathbb{1}_{t>49} + \beta_2 \times M_i^2 \times \mathbb{1}_{t>49} + \beta_3 \times M_i^3 \times \mathbb{1}_{t>49} + \alpha_i + \omega_t + \varepsilon_{i,t} \tag{6}$$

The Phase-2 treatment starts at $t = 50$. We will also estimate results based on alternative specifications, such as a) interacting M_i^c with the first-phase treatment assignment T_c and b) expanding the time period to include the Phase-1 data and identifying the effect of switching into/out of bonus types drivers prefer.

2.6.3. Hypotheses

Based on the specifications for both Phase 1 and Phase 2, we will test the following hypotheses for all primary outcomes: labor supply, earnings, expenditures, and preferences for incentive types.

- Hypothesis 1: $\beta_2 \neq 0$ would indicate that the conditional guaranteed income treatment affects workers' choices as defined through the pre-specified set of primary outcomes. The test compares the effect of T2 against the pure control: T0.
  - Similarly, we will test whether $\beta_1 \neq 0$, which would indicate the effect of the fixed-income bonus on workers' choices. $\beta_3 \neq 0$ would indicate the effect of a more flexible fixed-income bonus treatment on workers' choices.

- Hypothesis 2: $(\beta_2 - \beta_1) \neq 0$ would indicate that the effect of the conditional income guarantee treatment, holding constant the income effect, affects workers' choices as measured through the pre-specified primary outcomes. In T1, the total earnings a driver receives each day they meet the condition may differ, whereas in T2, the total earnings a driver receives each day they meet the condition are certain. The expected value of total earnings will remain the same between the two arms, but the level of earnings uncertainty differs between arms. As such, comparing $(\beta_2 - \beta_1)$ against 0 would allow us to disentangle the income effect from that of reduced risk via the conditional income guarantee.

- Hypothesis 3: $(\beta_3 - \beta_1) \neq 0$ would indicate the effect of additional flexibility in T3 on drivers' choices, as measured through the pre-specified primary outcomes. The total amount drivers could earn remains the same between T1 and T3, but the conditions differ: T1 requires meeting the labor supply target every day with relatively little flexibility, whereas T3 allows workers to distribute work hours over a longer period to accommodate their flexibility needs.

For analysis using Phase-2 variation, the hypotheses we test are about the effects of receiving bonus incentives that match the driver's preferences, and about the particular attributes of the bonus incentives they prefer.

2.6.4. Welfare Effects

Beyond the reduced-form analysis based on the hypotheses above, we plan to conduct additional analyses of welfare effects and their mechanisms by leveraging preference measures for contract types and the experimental variation in actual contracts offered in the intervention. We use BDM-based measures of preferences over contract types. We take these measures as the overall welfare impact that includes the following, each of which we can estimate:

- Income effect from bonuses and transfers from the study/platform, holding constant labor supply and productivity

- Additional income effect from changes in labor supply and productivity, on and off the platform

- Changes in the marginal cost of labor supply, identified through survey measures

- The expected benefit of T2's attributes (i.e., lower earnings risk), identified through the BDM game from those assigned to T0.

- The expected benefit of T3's attributes (i.e., more flexibility), identified through the BDM game from those assigned to T0.

- The updated expected benefit of T2's attributes (i.e., lower earnings risk), identified through the difference in BDM bids for T2 between those assigned to T2 and T0.

- The updated expected benefit of T3's attributes (i.e., more flexibility), identified through the difference in BDM bids for T3 between those assigned to T3 and T0.

2.6.5. Interpretation of Results

The results of our hypotheses, when coupled with our primary outcomes, will allow us to better understand driver behavior and disentangle competing labor supply models of behavior.

Labor Supply

Labor supply outcomes are the most essential for differentiating among labor supply models. Table 1 gives a summary of our expected predictions of how drivers would react to each treatment arm, conditional on them behaving in line with each model.

Our design also allows us to identify the sources of changes in labor supply on the platform and the effects of those adjustments on overall labor supply. We address potential measurement concerns about substitution between platform work and other employment opportunities by including an extensive labor supply module in the surveys. The combination

Theory	T1: Fixed Bonus	T2: Guarantee	T3: Flexible Bonus
Neoclassical	Increase L at the margin.	Increase L more than T1 if risk averse	Highest variance in L.
Reference Dependence	May reduce L if bonus helps hit target early.	Reduce L if income-target lower than guarantee; No effect if income-target higher than guarantee	T3 $\approx$ T1 or T3 $>$ T1 depending on time horizon of income target.
Liquidity-Constrained	No effect on L, but changes to household expenditure.	Reduces unstable labor supply when income needs high; acts as credit.	Low value if cash needs are daily; T3 $\approx$ T1.

Table 1: Summary of Theoretical Predictions Across Bonus Schemes

of survey measures and platform data will allow us to determine whether drivers adjust their overall labor supply hours and, if so, how much slack in labor supply exists, if any.

Earnings and Household Expenditures

It is reasonable to assume that the interventions will increase driver take-home pay. But how this transmits to household expenditures should give us a signal as to how drivers make decisions (whether they are neoclassical, reference-dependent or liquidity constrained). In a standard neoclassical model with perfect credit, our six-week intervention should not substantially change monthly expenditures (assuming smoothing over a long horizon). However, if there is a significant change in expenditures due to the treatments, this would support the presence of liquidity constraints among the drivers.

Revealed Preferences and BDM Games

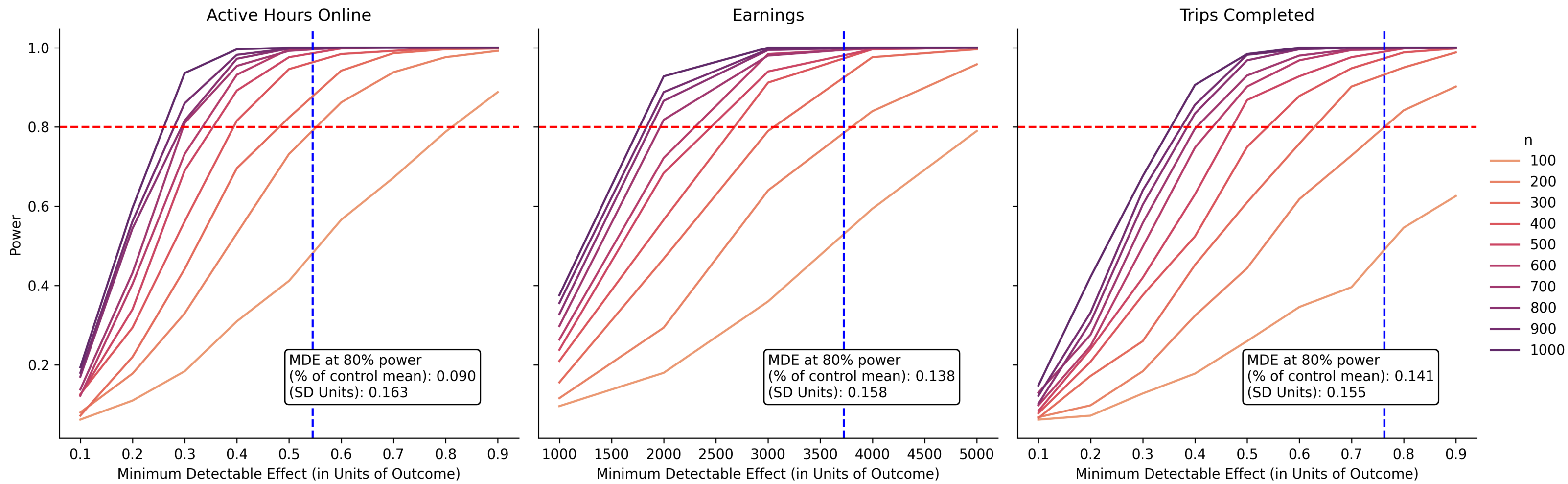
Drivers' bids provide information about the importance of lower earnings risk and flexibility. The mean bid for T2 relative to T1 represents the monetary value of the income floor relative to the fixed bonus, i.e., the value of lower earnings risk. The mean bid for T3 against

T1 would be identify the value of additional intertemporal flexibility to meet labor supply requirements.

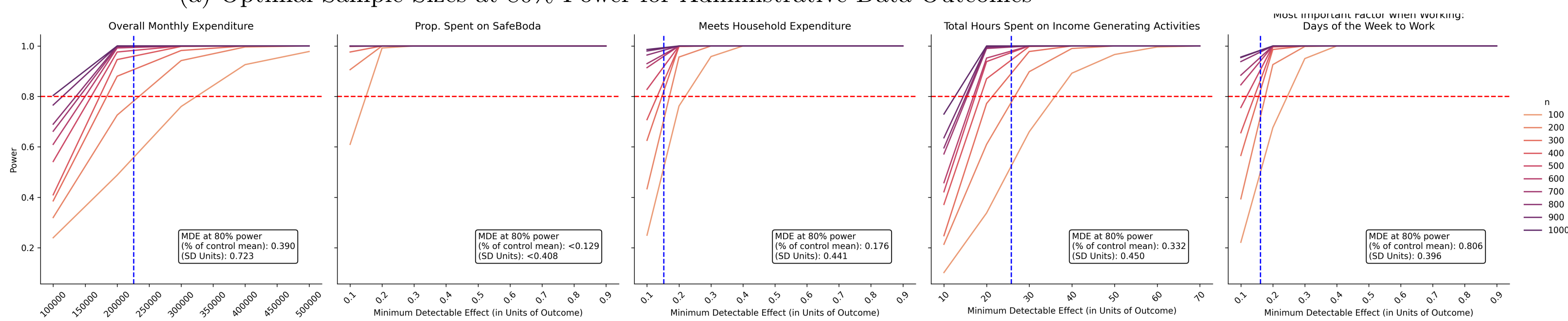
2.7. Power Analysis

We conduct power analysis using survey and administrative data provided by the platform. We follow Burlig et al. (2020). The advantage of using this method is that we can fully leverage the administrative data provided during our pilot study to incorporate the serial correlation structure of our pilot data, rather than assuming a particular standard deviation.

For administrative outcomes, we use the maximum amount of pre-intervention data available to us to ensure that our power analysis is not contaminated by the pilot study's treatment effects. This amounts to 33 days of pre-intervention data.



(a) Optimal Sample Sizes at 80% Power for Administrative Data Outcomes



(b) Optimal Sample Sizes at 80% Power for Survey Data Outcomes

Figure 1: Power Curves for Administrative and Survey Data Outcomes across Effect Sizes

Notes: Figure shows the results of power analysis based on the method in Burlig et al. (2020). Power curves are shown for administrative and survey outcomes across different assumed sample sizes. The X-axis shows assumed effect sizes in units of the outcome. For binary outcomes, changes in probability were used. The intersection of the hatched red and hatched blue lines denotes the minimum detectable effect assuming 80% power and a sample size of 200 drivers per treatment arm.

Figure 1 illustrates power curves across different sample sizes for each treatment arm, ranging from 100 to 1000. The x-axis denotes the effect size assumed in units of the outcome. Each plot indicates the minimum detectable effect at 80% power as a proportion of the control mean and in units of standard deviation. For binary outcomes, effects are expressed as increases in probability. The intersection of the red and blue hatched lines denotes the minimum detectable effect at our chosen sample size of 200 drivers per treatment arm. For administrative data outcomes, we assumed a multi-period panel, as in equation 1, using all pre-pilot time periods available. For survey data outcomes, we used the ANCOVA specification in equation 2.

For administrative data outcomes, we find that, on average, we detect a 9-14% change in the outcome relative to the control mean. This is equivalent to approximately 0.16 standard deviations. For the survey outcomes, the minimum detectable effect varies across outcomes. For the driver's overall monthly expenditure (including household expenditures, outstanding loans, school fees, and remittances), we detect an effect of approximately 0.72 standard deviations. For the proportion of time spent on the platform, we find a minimum detectable of less than 13% of the control mean or less than 0.408 units of standard deviation (meaning that we can detect less than a 10% increase in the proportion of time spent on the platform). For the binary outcome of whether the driver self-reports that they can meet their household expenditure, 80% power is achieved at a control mean of 18% (approximately 0.44 standard deviations). For the total hours spent on income-generating activities, we detect a 33% change relative to the control mean, or 0.45 standard deviations. Finally, for the outcome that the days of the week to work are most important when working on the platform, we find that 80% power is achieved with 61% control of the mean. As expected, the survey outcomes have a higher minimum detectable effect than the administrative outcomes. We lose power when collapsing the data into an ANCOVA specification, thereby reducing variation over time. This is compounded by the fact that survey outcomes tend to exhibit greater noise.

Additionally, the small effect sizes observed in our pilot data may be driven by the platform experiment design, specifically by the choice of trip conditions and earnings guarantees.

Higher trip conditions imply lower eligibility for bonuses and thus lower power. Similarly, higher earnings guarantees may lead to lower time spent on the platform and thus lower power. This is compounded by the fact that trip threshold choices and earning guarantees affect the research budget. The optimal choice of design parameters should be made with budget and power in mind. To incorporate these considerations, we use a modified version of Black et al. (2022) to conduct Monte Carlo simulations at different trip thresholds and earnings guarantee levels to explore the tradeoff between treatment conditions, power, and their impact on the budget. The results of the Monte Carlo simulations are presented in Appendix Figure A.5. We find that the range of trip condition should be between 5-7 trips and the earnings guarantee should be between 40,000-60,000 UGX. Based on the simulations, the implied bonus for T1 should be between 2,000-7,600 UGX.

3. Data

3.1. Data Collection and Processing

We will collect administrative data on drivers and trips from the platform, as well as in-person surveys at baseline and endline. We will access the administrative data via a database management tool used by the platform. The database is updated at least daily, allowing us to track both compliance with the bonus conditions and outcomes.

The in-person surveys will be conducted at the platform's campus by a respected local survey firm. The survey location was chosen based on logistical constraints; drivers regularly stop by the campus to receive updates and supplies, making it a convenient site to gather a large group of drivers. We will design the surveying procedure to be as independent of the platform as possible to minimize concerns that the platform's involvement influences drivers' responses. The surveys will be conducted in a large conference room, with no platform employees present. The survey firm also has no direct financial relationship with the platform.

3.2. *Variations from the Intended Sample Size*

We expect attrition in the experimental sample due to a) drivers' exit from the platform and b) nonresponse at the endline survey. We mitigate these concerns as follows. First, we target drivers that the platform considers "active," rather than those who have an account but do not make a meaningful number of trips. Second, we will incentivize survey participation through financial compensation, a light lunch, and repeated follow-up by the survey firm to ensure a low attrition rate at the endline.

3.3. *Pilot Data*

A pilot study on a stratified random sample of 79 drivers was conducted on May 5-12, 2025, with associated baseline and endline surveys. This pilot study provided rich signals on how drivers reacted to our intervention design, and it also streamlined data collection and offered a valuable source of data for improving our design for the scale-up. The pilot treatment designs differed from the ones presented in this report. The intervention arms for the pilot were as follows:

T0 Pure Control: No incentives or payouts are given.

T1 Low Condition, Fixed Bonus: Drivers who meet a low condition (5 trips completed) are given 16,500 UGX in addition to their daily earnings at the end of the day.

T2 High Condition, Fixed Bonus: Drivers who meet a high condition (10 trips completed) are given 16,500 UGX in addition to their daily earnings at the end of the day.

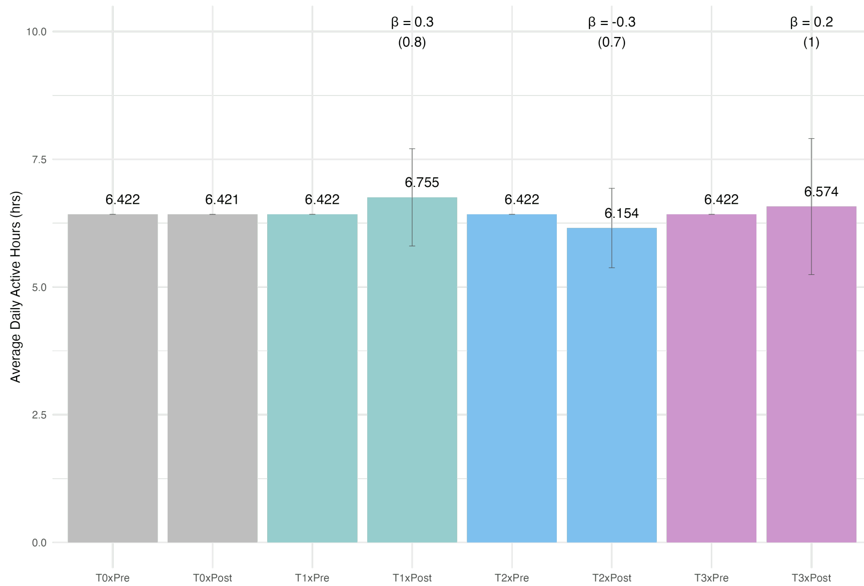
T3 High Condition, Earnings Guarantee: Drivers who meet a high condition (10 trips completed) are guaranteed 55,000 UGX for their daily earnings. They are paid the difference between 55,000 UGX and their trip earnings at the end of the day.

Compared with the current design, we have removed A1 and added our current T3. The results of the pilot study showed that drivers responded to the treatment, but the results were limited due to trip conditions being too high in A2 and A3, and the conditional guarantee being too low, given the trip conditions. Conditional upon completing the 10-trip minimum required to qualify for the earning guarantee, many drivers have already made at least UGX 55,000. The trip minimums and earnings/bonus levels were set in collaboration with the platform, and at the time, we did not have access to the administrative data.

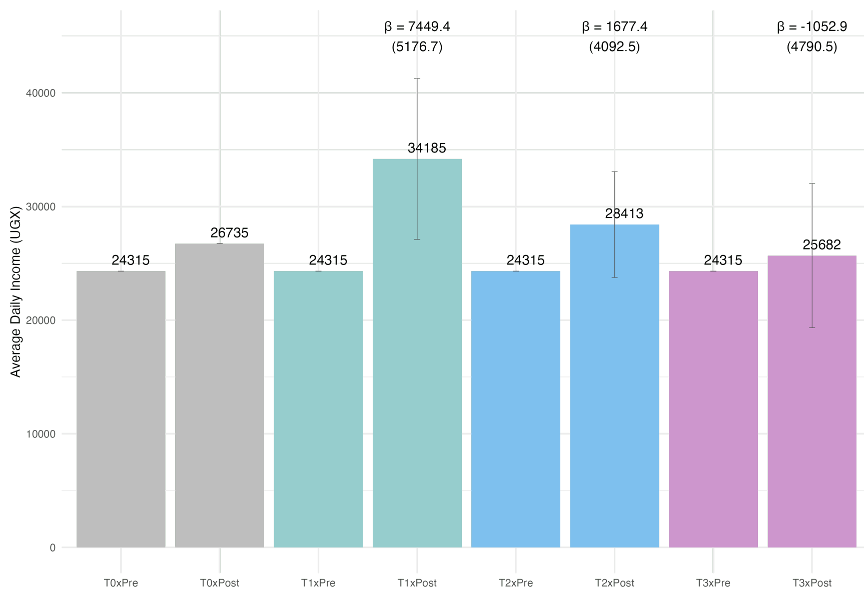
Figure 2 shows the results of a 2-period difference-in-difference estimation on active hours and earnings on the platform for the drivers in each treatment group. T1 had a positive effect on both hours and earnings, and the impact on earnings was significant at the 5% level. For T2, there was a decrease in hours and an increase in earnings, while for T3, there was an increase in both earnings and active hours. Effects are not significant for T2 and T3 and are likely partially due to our design being underpowered and to suboptimal choices of the earnings guarantee and trip minimums. Figure 3 shows a scatter plot of trips vs. earnings and highlights which trips qualified for their respective treatment arms. Overall, 85% of T1 drivers qualified for their bonus, and 40% of T2 drivers qualified. The high take-up for T1 drivers was by design, as T1 was intended to be an as-if unconditional bonus, which, compared to T2, was meant to test the effect of the condition on driver behavior. However, only 4.8% of drivers qualified for the earnings guarantee in T3. As shown in Figure 2, the reason was that most drivers who completed their trip condition (10 trips) had already earned at least their earnings guarantee, making them ineligible for any additional funds.

Additionally, the pilot study exhibited low driver recall and low treatment salience. Despite being told they were part of a treatment group and receiving notifications afterwards that they were in the intervention, drivers found it difficult to determine which treatment arm they were in and what their payout would be. Figure 4 shows the proportion of drivers who answered that they had no recall, partial recall, or full recall of their treatment arm. Although 62.5% of T1 drivers reported receiving a bonus, this decreased to 38.9% among T2 drivers. Only 25% of drivers in T3 reported a full recall of their treatment.

We have addressed power issues by increasing the sample size for the full-scale intervention and using administrative data directly to estimate suitable trip conditions and earnings guarantees (See Appendix Section Appendix A.1). To address the recall issue, we are leveraging a new user-experience system implemented by the platform that reformats trips into



(a) Optimal Trip Conditions



(b) Optimal Earnings Guarantees

Figure 2: Pilot Results: Effect of Treatment on Active Hours and Earnings

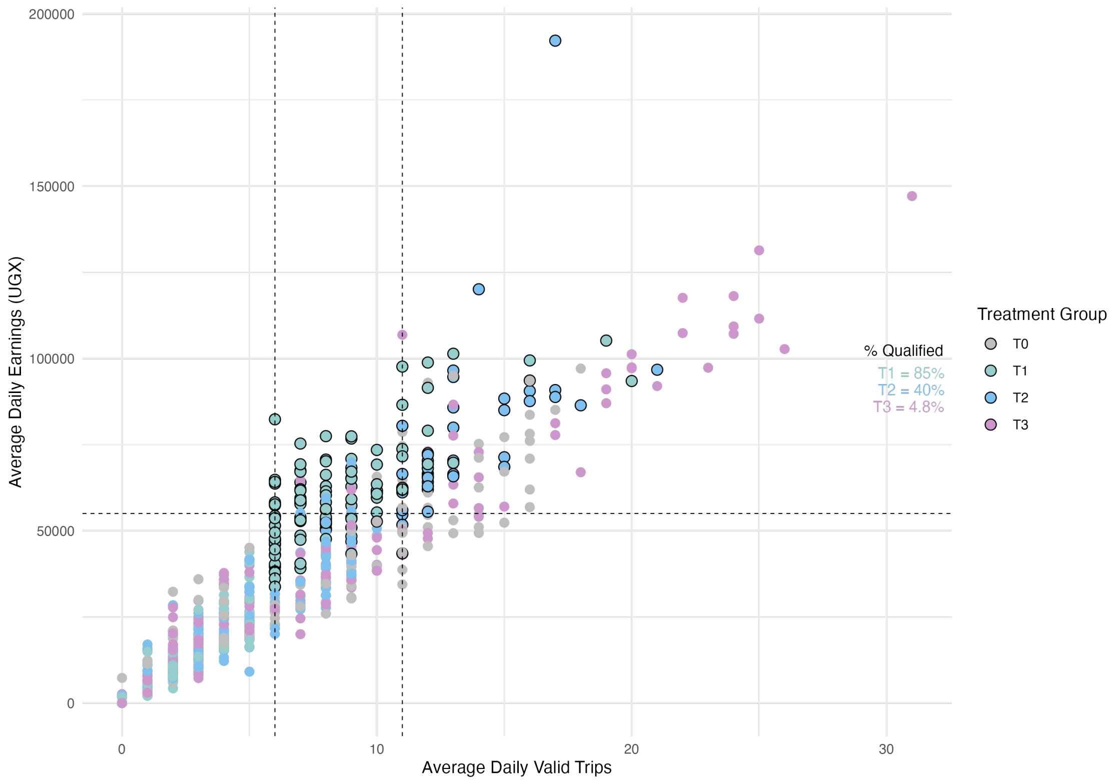


Figure 3: Pilot Results: Scatterplot of Takeup by Treatment Arm

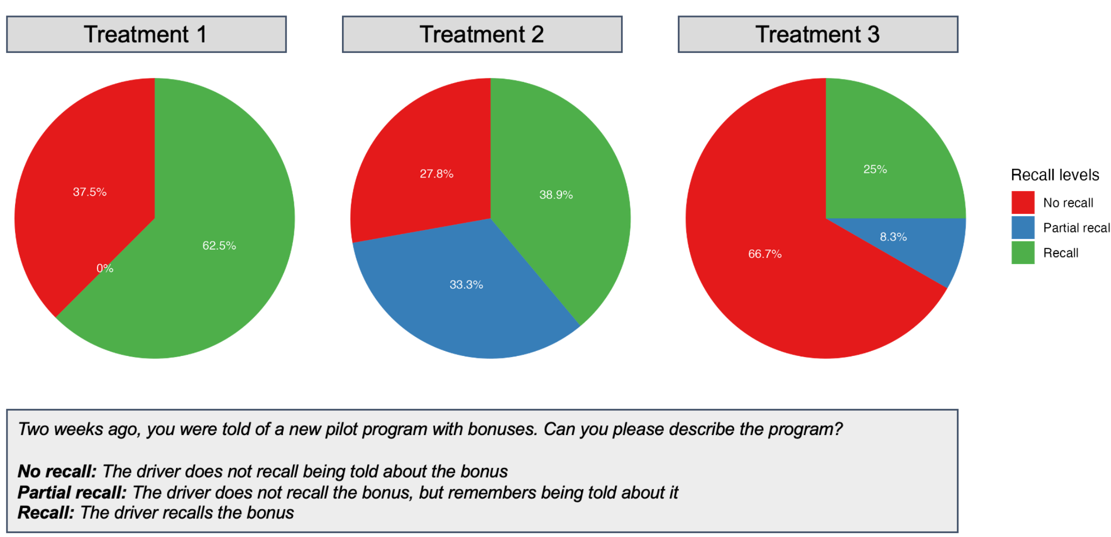


Figure 4: Pilot Results: Recall by Treatment Arm

a point system and shows a growing bar plot to drivers as they get closer to completing a bonus goal. To use this system, we are redesigning our treatment in terms of points rather than trips. This will increase saliency for drivers and improve recall of our study.

4. Analysis

4.1. Statistical Methods and Models

Our primary goal is to establish causal claims about how bonus contracts and variations in their attributes affect workers' labor supply, earnings, productivity, and welfare, using the gig economy as the context for a randomized controlled trial. The underlying assumption is that random assignment to treatment arms enables a causal interpretation of the changes in outcomes we estimate within treatment groups. We also leverage randomness in identifying drivers' willingness to pay for the bonus incentive types in the BDM games. The key statistical methods we plan to employ are:

- Ordinary Least Squares (OLS) regressions to identify the intent-to-treat (ITT) effects of the bonus-type assignment in the first phase (i.e., T1 through T3) against the control (T0) on all pre-specified primary outcomes.

- Two-stage Least Squares (2SLS) regressions to identify the treatment-on-the-treated (ToT) effects of the bonus-type assignment in the first phase (i.e., T1 through T3) against the control (T0) on all pre-specified primary outcomes.

- Two-stage Least Squares (2SLS) regressions to identify the treatment-on-the-treated (ToT) effects of being assigned to the bonus type of the drivers' preference in the second phase (i.e., T1 through T3) on all pre-specified primary outcomes based on the administrative database.

We provide further details on the statistical approach, including specifications and the 2SLS approach, in Sections 2.6.1 and 2.6.2.

4.2. Multiple Outcomes and Multiple Hypothesis Testing

This study involves multiple hypotheses tested on several primary outcomes. As such, we will take the false discovery rate (FDR) approach and report the Benjamini et al. (2006) sharpened q-values for the tests across all primary outcomes for each type of hypotheses specified in Section 2.6.3.

4.3. Heterogeneous Effects

We plan on investigating heterogeneous treatment effects based on key driver characteristics that we hypothesize our treatments interact with: a) risk preference, b) preference for flexibility, and c) other behavioral characteristics and income needs. We solicit these characteristics during the baseline survey and using pre-intervention administrative data. We plan to explore heterogeneous effects using our ITT specification.

First, the intervention, and T2 in particular, provides financial incentives to reduce risk and uncertainty in earnings schedules. As such, we are interested in heterogeneous effects based on the survey measure as follows:

- **Stated risk preference:** In the baseline, we elicit a subjective measure of how willing a worker is to take risks. The various earnings contracts tested in the intervention may have differing impacts depending on how risk-averse a driver is (i.e., are they more likely to value earnings guarantees to mitigate risk). We use our baseline measures of risk tolerance to construct an index and divide workers into those that fall above and below its median.

Second, the intervention, and T3 in particular, is designed to address preferences for flexibility by conditioning driver earnings on the amount of work required and exogenously varying the flexibility in meeting that requirement. As such, we may find heterogeneous effects based on workers' baseline preference for flexibility and labor supply schedule.

- **Stated preference for flexibility:** In the baseline survey, we ask drivers how much they prefer the feature of gig work that allows them to have a flexible schedule. We will collect a series of Likert-scale measures of flexibility, construct an index, and divide the drivers into those above and below the median.

- **Part-time vs. Full-time:** Part-time drivers likely have alternative employment, so their labor supply elasticity would be higher than for full-time workers. This dimension of heterogeneity, coupled with our survey results on off-platform work sources, would allow us to observe the extensive margin effects of our interventions in the presence of heterogeneous outside options.

Third, drivers may respond differentially based on factors other than their preferences for risk and flexibility, such as a) the extent to which they target their daily earning levels and b) their income needs. We investigate heterogeneous effects based on the following dimensions.

- **Targeting:** We directly elicit income and hours targeting from drivers in the baseline survey. Drivers who target may have a different sign of the effect of the treatments, since increased income would allow them to hit their target more quickly.

- **Household Expenditures:** Our survey will elicit drivers' household expenditures such as rent, agricultural activities, school fees, active loans/borrowing, and food costs. Drivers with higher expenditures may work more hours immediately before a loan payment is due or before school fees are due. The randomized interventions may give more income stability to drivers, allowing them to smooth hours across time.

- **Tenure on the Platform:** Newer drivers often need more time to optimize over a given platform policy. But drivers who have been on the platform longer would be able to understand the different contract structure more quickly. We would expect that earnings would increase more quickly for drivers with longer tenure.

5. Limitations and Challenges

We anticipate a few challenges in the implementation of our study:

- **Attrition:** Not all drivers who participate in the baseline survey will join in the endline survey. To account for this, we will (1) oversample drivers based on the attrition rate of the pilot, (2) offer small tokens/meals and cover the costs to drivers who visit the platform's campus for surveys, (3) test if attrition is balanced.

- **Recall:** We may not be able to detect effects of drivers who do not remember and understand the arm they are assigned to. Based on our experience during the pilot, in which driver recall was low, the platform implemented a new *notification feature* that enables us to schedule and send push notifications to drivers more directly. The update also includes a new UI that shows drivers their point total and how far they are from each bonus threshold. The combination of notifications and visual salience can improve driver recall.

- **BDM-comprehension:** The BDM game during the endline might be complicated for drivers. Drivers may bid overwhelmingly at the extremes if they do not understand it well. We plan on piloting the BDM to refine the explanations. We will also include other preference-related questions in the survey as an additional check.

- **Experiment Demand Effects:** Experimenter demand effects may be present in our surveys and BDM games.

  - *Surveys:* Our questions on off-platform work and platform sentiment may be skewed if drivers believe that we are affiliated with the platform. Our survey consent forms will make it clear that we are an independent set of researchers unaffiliated with the platform and that we collaborate only with it.

  - *BDM Games:* When playing the BDM games, drivers may think that we "want" them to have a high willingness to pay for switching to another treatment. This may create bias in the bids made by drivers. Our explanation of the BDM games will make it clear that drivers should bid based on how much they value switching to another treatment.

6. Policy Relevance

The growth and expansion of digital platforms in developing countries raise questions about the optimal way to regulate gig work, given the unique constraints drivers face. Informal gig workers often face binding household liquidity constraints and limited access to formal insurance products, assets, or alternative employment opportunities. As such, regulating gig work may require an approach different from that attempted in developed economies. Our research is relevant for two actors: drivers and platforms.

On the driver front, in developing countries, it is unclear whether platform workers select into gig work because of a preference for flexibility or because of a lack of stable income opportunities. We provide additional contextual evidence on the characteristics of gig workers, the challenges they face, and how they allocate labor across activities. Through our intervention, we evaluate the impact of different earnings contracts on driver welfare. Such findings are relevant to policymakers seeking to understand how platform regulations can address the challenges faced by workers.

On the platform front, platforms in developing countries are constrained by a less stable labor force, making it difficult to streamline their workforces. We test whether the approach to algorithmic management undertaken by platforms can be modified to increase driver welfare, productivity, and performance. By altering contract structure, we posit that we can reduce earnings risk while increasing productivity. Additionally, we evaluate the merits of different contract structures independently, accounting for heterogeneous driver preferences. Given that drivers may prefer additional flexibility or reduced earnings uncertainty, incentives may be effective only when contracts align with those preferences. Our research directly tests this mechanism, underscoring the importance of designing incentive-compatible policies between workers and platforms.

Disentangling the effects of the treatment would yield different policy recommendations for the platform and for development policy more broadly. If, for instance, liquidity constraints dominate driver behavior, policy should focus on alleviating these constraints by expanding credit access. However, if reference dependence is the dominant factor in driver behavior, behavioral interventions would be recommended to increase driver productivity through alternative algorithmic management. Moreover, if the results of the BDM shows that the mean bid for T2 is high, and labor supply productivity increases as a result of T2, this would be evidence for a demand for insurance and a need for tools for smoothing income.

Overall, our research informs the design of policies that can be implemented more locally by specific platforms and broadly by governments aiming to regulate gig work. Our findings are particularly relevant in developing countries. Given that market-level policy changes can be challenging and costly, platform-level policy changes offer a valuable tool to circumvent existing market failures and improve worker welfare.

7. Administrative Information

7.1. Ethics Approval

Ethics approval was granted by the Uganda National Council on Science and Technology (UNCST) (MUREC-2025-796). We are currently seeking an amendment to this protocol. Ethics approval is also in process at the Institutional Review Board at Swarthmore College.

7.2. Declaration of Interest

We do not have any conflict of interest to declare.

References

Anonymous, 2024a. Details omitted for double-anonymized reviewing.

Anonymous, 2024b. Details omitted for double-anonymized reviewing.

Anwar, M.A., Graham, M., 2021. Between a rock and a hard place: Freedom, flexibility, precarity and vulnerability in the gig economy in africa. Competition Change 25, 237–258.

Ayentimi, D.T., Abadi, H.A., Burgess, J., 2023. Decent gig work in sub sahara africa? Journal of Industrial Relations 65, 112–125.

Banerjee, A.V., Duflo, E., 2007. The economic lives of the poor. Journal of Economic Perspectives 21, 141–167.

Becker, G.M., DeGroot, M.H., Marschak, J., 1964. Measuring utility by a single-response sequential method. Behavioral science 9, 226–232.

Benjamini, Y., Krieger, A.M., Yekutieli, D., 2006. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93, 491–507.

Berg, J., Furrer, M., Harmon, E., Rani, U., Silberman, M.S., 2018. Digital Labour Platforms and the Future of Work: Towards Decent Work in the Online World. International Labour Organization.

Black, B., Hollingsworth, A., Nunes, L., Simon, K., 2022. Simulated power analyses for observational studies: An application to the affordable care act medicaid expansion. Journal of Public Economics 213, 104713.

Bodie, Z., Merton, R.C., Samuelson, W.F., 1992. Labor supply flexibility and portfolio choice in a life cycle model. Journal of Economic Dynamics and Control 16, 427–449.

Buchholz, N., Shum, M., Xu, H., 2023. Rethinking reference dependence: Wage dynamics and optimal taxi labor supply. Technical Report. working paper, Princeton University Economics Dept.

Burlig, F., Preonas, L., Woerman, M., 2020. Panel data and experimental design. Journal of Development Economics 144, 102458.

Chen, K.M., Ding, N., List, J.A., Mogstad, M., 2025. Reservation wages and workers' valuation of job flexibility: Evidence from a natural field experiment. Journal of the European Economic Association , jvaf022.

Chen, M.K., Rossi, P.E., Chevalier, J.A., Oehlsen, E., 2019. The value of flexible work: Evidence from uber drivers. Journal of political economy 127, 2735–2794.

Cook, C., Diamond, R., Oyer, P., 2019. Older workers and the gig economy, in: AEA Papers and Proceedings, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203. pp. 372–376.

Datta, N., Rong, C., Singh, S., Stinshoff, C., Iacob, N., Nigatu, N.S., Nxumalo, M., Klimaviciute, L., 2023. Working Without Borders: The Promise and Peril of Online Gig Work. Technical Report 32573393. The World Bank.

Dercon, S., 2002. Income risk, coping strategies, and safety nets. The World Bank Research Observer 17, 141–166.

Dizon-Ross, R., Zucker, A.D., 2025. Mechanism design for personalized policy: A field experiment incentivizing exercise. NBER Working Paper Series 33624.

Duong, H.L., Chu, J., Yao, D., 2023. Taxi drivers' response to cancellations and no-shows: New evidence for reference-dependent preferences. Management Science 69, 179–199.

Dupas, P., Robinson, J., Saavedra, S., 2020. The daily grind: Cash needs and labor supply. Journal of Economic Behavior & Organization 177, 399–414.

Farber, H.S., 2015. Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. The Quarterly Journal of Economics 130, 1975–2026.

Glasner, B., 2023. The minimum wage, self-employment, and the online gig economy. Journal of Labor Economics 41, 103–127.

Hacamo, I., Kleiner, K., 2022. Forced entrepreneurs. The Journal of Finance 77, 49–83.

Hamza, S., Yahya, U., Kasule, A., Usama, K., Fahad, P., Ibrahim, K., 2023. Characterizing peak-time traffic jam incidents in kampala using exploratory data analysis. Journal of Transportation Technologies 13, 61–86.

Harris, J.R., Todaro, M.P., 1970. Migration, unemployment and development: a two-sector analysis. The American economic review 60, 126–142.

Heeks, R., Graham, M., Mungai, P., Van Belle, J.P., Woodcock, J., 2020. Systematic Evaluation of Platform Work Against Decent Work Standards: Development of a New Framework and Application in the Global South. Technical Report.

Herbst, D., 2025. Asymmetric Information in Wage Contracts: Evidence from an Online Experiment. Technical Report. working paper, University of Arizona.

Hu, S., Feng, Y., Keppo, J., 2025. Understanding gig economy labor supply: Evidence from a logistics platform. Available at SSRN 5168285.

Jackson, E., 2022. Availability of the gig economy and long run labor supply effects for the unemployed, in: 2021 APPAM Fall Research Conference, APPAM.

Jones, S., Manhique, I., 2025. Digital labour platforms as shock absorbers: Evidence from the covid-19 pandemic in mozambique. Journal of African Economies 34, 116–141.

Kőszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. The Quarterly Journal of Economics 121, 1133–1165.

Lee, M.K., Kusbit, D., Metsky, E., Dabbish, L., 2015. Working with machines: The impact of algorithmic and data-driven management on human workers, in: Proceedings of the 33rd annual ACM conference on human factors in computing systems, pp. 1603–1612.

Li, Z., Hong, Y., Zhang, Z., 2021. The empowering and competition effects of the platform-based sharing economy on the supply and demand sides of the labor market. Journal of Management Information Systems 38, 140–165.

Mas, A., Pallais, A., 2017. Valuing alternative work arrangements. American Economic Review 107, 3722–3759.

Mobarak, A.M., Rosenzweig, M., 2014. Risk, insurance and wages in general equilibrium. Technical Report. National Bureau of Economic Research.

Morduch, J., 1995. Income smoothing and consumption smoothing. Journal of Economic Perspectives 9, 103–114.

Murthy, G., Deshpande, R., 2022. How Can Financial Services Support Platform Work? Insights from Five Emerging Markets. Focus Note. CGAP. Washington, D.C.

Rossi, M., Trucchi, S., 2016. Liquidity constraints and labor supply. European Economic Review 87, 176–193.

Stafford, T.M., 2015. What do fishermen tell us that taxi drivers do not? an empirical investigation of labor supply. Journal of Labor Economics 33, 683–710.

Thakral, N., Tô, L.T., 2021. Daily labor supply and adaptive reference points. American Economic Review 111, 2417–2443.

Townsend, R.M., 1994. Risk and insurance in village india. Econometrica , 539–591.

World Bank, 2018. World Development Report 2019: The Changing Nature of Work. The World Bank.

Appendix A. Appendix

Appendix A.1. Optimal Design Parameters and Power

We use block random sampling with replacement to choose a sample of 800 drivers. They are randomly assigned to treatment status, and outcomes are modified according to the treatment design to provide them with either an earnings guarantee or a bonus (T1 or T2). Only administrative data outcomes were considered. Hours were shocked in the post-period for those in the treated groups, and effects on trips and earnings were estimated using the median trip and median earnings per hour. An implied assumption is that the only causal channel through which drivers are affected by the treatment is an increase in their active hours. This assumption is defensible because drivers choose their online active hours. In contrast, the number of trips a driver completes or the driver's earnings may be confounded by demand effects.

Subsequently, the Monte Carlo simulations are used to calculate optimal trip conditions and earnings guarantees as those that solve the following optimization problem:

$$\min_{q_H, F} C(q_H, F, t_2, t_3) \tag{A.1}$$

$$s.t.$$

$$power \geq 0.8$$

where $C(.)$ is the monetary cost of the project, dependent on the trip minimum (q_H), earnings guarantee (F), and effect sizes of each treatment (t_1, t_2). We omit T3, as it was not longer used in our pilot intervention. Since raising the trip minimum reduces the number of drivers who qualify, costs are decreasing in q_H, and since the earnings guarantee increases the driver's payout, costs are rising in F. *power* is defined by several conditions, depending on the test considered. This can depend on the ability to detect each treatment independently, or detecting differences across treatment arms. For the following simulations, we set the power to at least 0.8 to detect each treatment arm independently.

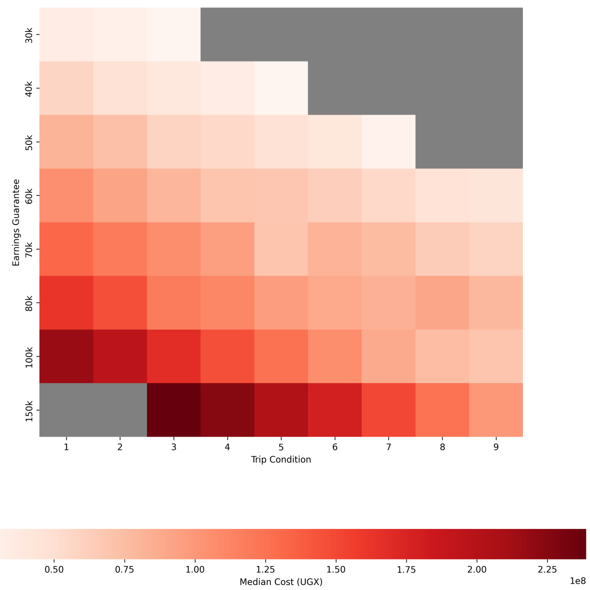
For the Monte Carlo simulations, we assume 200 drivers per treatment arm. Figure A.5 presents the results of the Monte Carlo simulations in two dimensions. In Figure A.5a, we show the cost of combinations of trip conditions and earnings guarantees. In Figure A.5b, we show the minimum detectable effects, in standard deviation units, for detecting a treatment effect for T1 and T2. The grey areas in the figure denote combinations which are infeasible. The top-right of the figures are infeasible due to insufficient power. This area represents combinations that are both high condition and low earnings guarantee areas, leading to low take-up. Additionally, high trip conditions generate higher income making it less likely that a driver will be below the earnings guarantee once they achieve the condition. In the bottom-left region, the grey cells represent areas that are infeasible due to budget. Low trip conditions and high earnings guarantees imply high takeup and high payout to drivers.

As expected, the intervention costs decrease in the north-east direction. Minimum detectable effect sizes increase with the trip condition, with less sensitivity to the earnings guarantee. This holds for both T1 and T2, despite T1 not being subject to the earnings guarantee. Earnings, however, are more sensitive to changes in the guarantee amount and the trip condition. The minimum detectable effect sizes are highest for active hours worked, compared to earnings and trips completed.

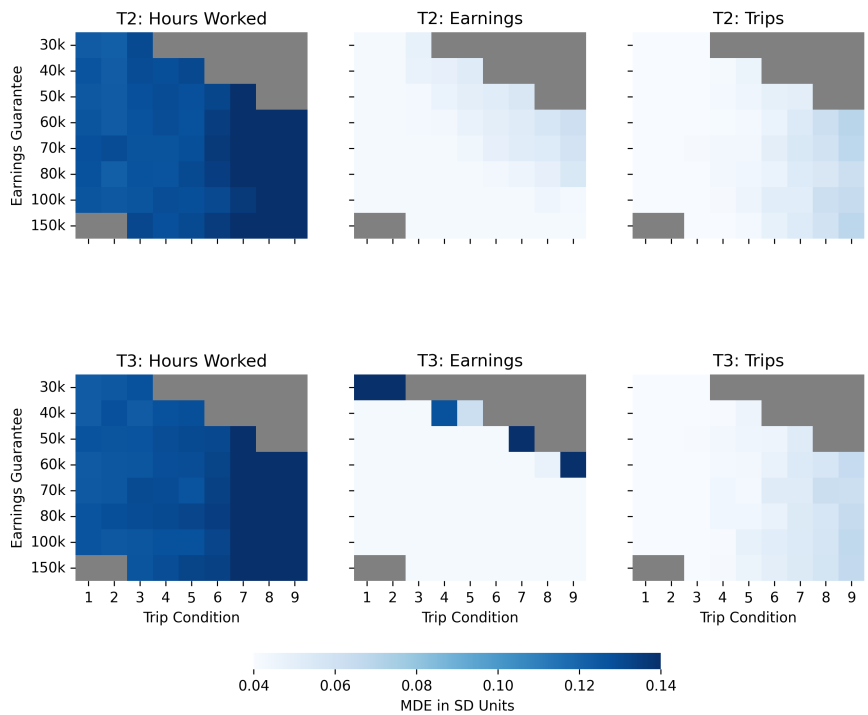
In summary, when considering the set of feasible combinations and the minimum detectable effect sizes, the cost-minimizing combination is somewhere in the range of a trip condition of 5-7 trips and an earnings guarantee of 40,000-60,000 UGX. Based on the simulations, this implies that, on average, 81% of drivers will be eligible for the bonus, and the implied bonus for T1 should be between 2,000-7,600 UGX.

Appendix A.2. Secondary Outcomes

Below are more details of the secondary outcomes we will explore:



(a) Budget Feasible Combinations of Trip Condition and Earnings Guarantee



(b) Minimum Detectable Effects by Combination of Trip Condition and Earnings Guarantee

Figure A.5: Results of Monte Carlo Simulations: Budget Feasibility and Minimum Detectable Effects

Notes: Heatmaps shows the results of Monte Carlo simulations inspired by Black et al. (2022). Assumed trip conditions and earnings guarantees on the x- and y-axes, respectively. Colors of the heatmap correspond to budget cost in Ugandan Shilling (UGX) in Figure (a) and minimum detectable effects in standard deviation units in Figure (b). Cells in grey represent trip condition-earnings guarantee combinations that are infeasible due to either insufficient power (top-right) or budget (bottom-left).

Appendix A.2.1. 5. Labor Supply

We define additional measures of labor supply:

- **Outcome 5A: Proportion of Time on Platform** - The proportion of total weekly hours spent on the platform relative to weekly hours spent on all income-generating activities, measured via survey data.
  - We will use survey data to collect the total number of self-reported hours a driver spends in a given week on each income-generating activity they do. We will then divide the time they spend on the platform by total hours spent working (sum of hours across all activities) to obtain the proportion of time spent on the platform.
  - This measure is useful as it can highlight switching behavior across different income-generating activities, highlighting how drivers respond to different types of financial incentives in relative terms
  - This outcome will be part of the driver-level survey data set and captured during both baseline and endline
- **Outcome 5B: Active Days per Week** - The number of days each week a driver makes at least one trip
  - This outcome will be part of the driver-week level panel data set
  - Measured via app data
- **Outcome 5C: Trips per Day** - The number of trips completed each day
  - This outcome will be part of the driver-day level panel data set
  - Measured via app data
- **Outcome 5C: Income Targeting Behavior** - An indicator for whether or not a driver chooses how much to work by targeting a particular income
  - This outcome will be part of the driver-level data set

- Measured via survey data

- **Outcome 5D: Income Targets** - The level of income a driver reports they target (conditional on income-targeting behavior)
  - This outcome will be part of the driver-level data set
  - Measured via survey data

Appendix A.2.2. 6. Trip Changes

Drivers might respond to the intervention by selecting the trips they are willing to accept. We define a few measures to capture on what margins drivers might change the types of trips or the variety of trips they take:

- **Outcome 6A: Earnings Variance** - The variance of post-commission, pre-bonus earnings across all days
  - This outcome will be part of the driver-level data set
  - Measured via app data
- **Outcome 6B: Trip Duration** - The number of minutes of each trip
  - This outcome will be part of the driver-trip level panel data set
  - Measured via app data
- **Outcome 6C: Trip Distance** - The number of Kilometers of each trip
  - This outcome will be part of the driver-trip level panel data set
  - Measured via app data
- **Outcome 6D: Trips in the Morning** - The proportion of trips per day between 7 am and 11 am
  - This outcome will be part of the driver-day level panel data set

  - Measured via app data

- **Outcome 6E: Trips in the Afternoon** - The proportion of trips per day between 11 am-4 pm
  - This outcome will be part of the driver-day level panel data set
  - Measured via app data
- **Outcome 6F: Trips in the Evening** - The proportion of trips per day between 4 pm and 8 pm
  - This outcome will be part of the driver-day level panel data set
  - Measured via app data
- **Outcome 6G: Trips in the Night** - The proportion of trips per day between 8 pm and 12 am
  - This outcome will be part of the driver-day level panel data set
  - Measured via app data
- **Outcome 6H: Trips in the Early Morning** - The proportion of trips per day between 12 am and 7 am
  - This outcome will be part of the driver-day level panel data set
  - Measured via app data
- **Outcome 6I: Trips in Kampala City Center** - The proportion of trips per day in the business district of Kampala
  - This outcome will be part of the driver-day level panel data set
  - Measured via app data

Appendix A.2.3. 7. Welfare

We define further measures captured in survey data that proxy for several other components of a driver's welfare:

- **Outcome 7A: Savings** - Total savings a driver has this month
  - This outcome will be part of the driver-level data set
  - Measured via survey data
- **Outcome 7B: Remittances** - Total amount a driver has remitted in the past month
  - This outcome will be part of the driver-level data set
  - Measured via survey data
- **Outcome 7C: Job Satisfaction** - Binary indicator for whether or not the driver ranked their job satisfaction as satisfactory or higher
  - This outcome will be part of the driver-level data set
  - Measured via survey data
- **Outcome 7D: Active Loans** - Total amount a driver owes in loans this month
  - This outcome will be part of the driver-level data set
  - Measured via survey data